MASTERARBEIT | MASTER'S THESIS

Titel | Title

Synthesis of phosphorylated isoforms of the Survival Motor Neuron (SMN) protein

verfasst von | submitted by

Michael Palka BSc

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

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 862
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Chemie
Betreut von | Supervisor:	Univ.-Prof. Dr. Christian Friedrich Wilhelm Becker

Zusammenfassung

Die spinale Muskelatrophie (SMA) wird durch funktionsbeeinträchtigende Mutationen des SMN-Proteins ausgelöst. Zum besseren Verständnis der Entstehung von SMA müssen sowohl die Funktionsweise als auch die unterschiedlichen Regulierungswege des SMN-Proteins untersucht werden. Posttranslationale Modifikationen (PTMs), insbesondere Phosphorylierungen, spielen hierbei eine entscheidende Rolle. Um deren Auswirkungen zu untersuchen, werden homogene, modifizierte SMN-Proteinvarianten in ausreichenden Mengen benötigt. Diese Arbeit konzentriert sich daher auf die Entwicklung einer vollständig synthetischen Strategie zur Synthese der *N*-terminalen Domäne des SMN-Proteins (Aminosäuren 1–160) und die anschließende Herstellung von SMN(1–160)-Varianten mit unterschiedlichen Phosphorylierungsmustern.

Um einen vollsynthetischen Ansatz zu ermöglichen, wurde die Zielsequenz zunächst in vier Segmente unterteilt, um die Größenbeschränkung der Festphasenpeptidsynthese (SPPS) zu überwinden. Das Segmentdesign basierte auf der Position nativer Cysteine innerhalb der SMN-Aminosäuresequenz sowie auf der Länge und den Löslichkeitseigenschaften der resultierenden Segmente, wodurch eine effiziente Verknüpfung mittels nativer chemischer Ligation (NCL) gewährleistet wurde. Alle Peptidsegmente wurden zunächst mittels SPPS hergestellt, inklusive phosphorylierter Varianten der zwei *N*-terminalen Segmente, um die Erzeugung unterschiedlich modifizierter SMN-Varianten zu ermöglichen. Die einzelnen Peptide wurden anschließend durch konvergierende Ligationen zusammengefügt. Zunächst wurden die Peptide 1–59 und 60–97 sowie 98–122 und 123–160 ligiert. Die anschließende Ligation der beiden resultierenden Segmente 1–97 und 98–160 ergab SMN(1–160). Die phosphorylierten Varianten wurden durch unterschiedliche Kombination der modifizierten und unmodifizierten Peptide 1—59 und 60—97 synthetisiert. Insgesamt wurden eine nicht phosphorylierte, zwei monophosphorylierte sowie eine diphosphorylierte Variante von SMN(1–160) erfolgreich hergestellt. Nach der erfolgreichen Synthese wurden alle Varianten mittels HPLC und HR-MS charakterisiert. Um die Auswirkungen der Phosphorylierungen auf die sekundären Strukturelemente sowie die Faltung des Proteins zu untersuchen, wurde Zirkulardichroismus-Spektroskopie (CD) durchgeführt. Die Messungen ergaben nur geringfügige Unterschiede zwischen den Varianten und zeigten eine insgesamt geringe Ausprägung sekundärer Strukturelemente, was auf suboptimale Faltungsbedingungen hindeutet. Insgesamt stellt die hier vorgestellte Synthesestrategie einen Ausgangspunkt für die Erzeugung weiterer SMN-Proteinvarianten dar und bildet die Grundlage für zukünftige Studien zu SMA und den ihr zugrunde liegenden molekularen Mechanismen.

Abstract

Spinal muscular atrophy (SMA) is caused by function-impairing mutations in the SMN protein. To gain a better understanding of the molecular mechanisms underlying SMA, both the function and the regulatory pathways of the SMN protein require further investigation. Posttranslational modifications (PTMs), particularly phosphorylation, play a crucial role in this context. To study their effects, homogeneous, modified SMN protein variants are needed in sufficient quantities. This work therefore focuses on the development of a fully synthetic strategy for the synthesis of the *N*-terminal domain of the SMN protein (amino acids 1—160) and the subsequent production of SMN(1—160) variants with distinct phosphorylation patterns.

To enable a fully synthetic approach, the target sequence was divided into four segments to overcome the size limitations of solid phase peptide synthesis (SPPS). The segment design was based on the positions of native cysteines within the SMN sequence, as well as on the length and solubility of the resulting segments, ensuring efficient assembly via native chemical ligation (NCL). All peptide segments were first synthesized using SPPS, including two phosphorylated variants of the two *N*-terminal segments, to allow for the generation of differentially modified SMN variants. The individual peptides were assembled through convergent ligation steps. Initially, peptides 1–59 and 60–97, as well as segments 98–122 and 123–160, were ligated. Subsequent ligation of the two resulting segments 1—97 and 98–160 yielded SMN(1—160). The phosphorylated variants were obtained by assembling the modified peptides 1–59 and 60—97 in different configurations. In total, one non-phosphorylated, two monophosphorylated, and one diphosphorylated variant of SMN(1–160) were successfully produced. Following successful synthesis, all variants were characterized using HPLC and HR-MS. The influence of phosphorylation on secondary structure and protein folding was subsequently assessed by circular dichroism (CD) spectroscopy. The measurements revealed only minor differences among the variants and indicated a low overall content of secondary structural elements, suggesting suboptimal folding conditions. Overall, the herein presented synthetic strategy provides a basis for the generation of additional SMN protein variants and for future studies on SMA and its underlying molecular mechanisms.

Acknowledgments

First and foremost I would like to express my sincere gratitude to my supervisor, Univ-Prof. Dr. Christian Friedrich Wilhelm Becker, for giving me the opportunity to conduct my master's thesis in his research group. I am very grateful for his guidance throughout both the experimental work and the writing of this thesis.

I would also like to thank Dr. Claudia Cobos Caceres and Herwig Weißinger, MSc., for their guidance throughout the project. This thesis would not have been possible without their expertise and support.

Furthermore, I am thankful to all members of the Becker group. Thank you for welcoming me (and Pascal) into your office, for sharing your scientific knowledge, and for creating such an enjoyable working environment. I greatly appreciate the fun and great times we had both inside and outside the laboratory.

Finally, I would like to express my heartfelt gratitude to my family and friends for their unwavering support throughout my studies. Their encouragement and belief in me have been instrumental in helping me complete not only this thesis but also my academic journey. I will always be thankful for their support.

Contents

List of Abbreviations

Abbreviation	Definition
2-CTC	2-Chlorotrityl chloride
Acm	Acetamidomethyl
ACN	Acetonitrile
Boc	*Tert*-butyloxycarbonyl
Calcd	Calculated
CD	Circular dichroism
DBU	Diazabicyclo[5.4.0]undec-7-ene
DCM	Dichloromethane
ddH_2O	Double distilled water
DIPEA	N,N-diisopropylethylamine
DMF	Dimethylformamide
DTT	Dithiothreitol
EDT	1,2-Ethanedithiol
ESI	Electrospray ionization
FA	Formic acid
Fmoc	9-Fluorenylmethoxycarbonyl
Gems	Gemini of Cajal bodies
Gnd · HCl	Guanidinium hydrochloride
HBTU	2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HPLC	High-performance liquid chromatography
HR-MS	High resolution mass spectrometry
LC-MS	Liquid chromatography mass spectrometry
MESNa	Sodium 2-mercaptoethanesulfonate
MPAA	4-Mercaptophenylacetic acid
MRE	Mean residue ellipticity
MS	Mass spectrometry
MW	Molecular weight
NCL	Native chemical ligation
Obs	Observed
PA	Peak area
pre-mRNA	Precursor messenger ribonucleic acid
PRMT5	Protein arginine methyltransferase 5
PTM	Posttranslational modification
RNA	Ribonucleic acid
RP-HPLC	Reversed phase high-performance liquid chromatography

Abbreviation	Definition
RT	Retention time
SMA	Spinal muscular atrophy
SMN	Survival motor neuron
snRNA	Small nuclear ribonucleic acid
snRNP	Small nuclear ribonucleoprotein
SPPS	Solid phase peptide synthesis
Tbeoc	2-(tert-butyldisulfanyl)ethyloxycarbonyl
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
Thz	Thiazolidine
TIC	Total ion count
TIPS	Triisopropylsilane
UHPLC	Ultra high-performance liquid chromatography
unrip	Unr-interacting protein

1 Introduction

1.1 Spinal muscular atrophy (SMA)

First documented cases of spinal muscular atrophy (SMA) date back to descriptions by the Austrian physician Guido Werdnig in 1891. Werdnig described a child with progressive paralysis and muscle weakness, attributed to degeneration of spinal motor neurons.[1] Subsequent characterisation of the pathophysiology of SMA was performed by Johann Hoffmann at the University of Heidelberg.[2] In modern medicine, SMA comprises a group of neuromuscular diseases characterized by the degeneration of alpha motor neurons in the spinal cord, leading to progressive muscle weakness, skeletal muscle atrophy and paralysis. SMA encompasses a broad spectrum of clinical phenotypes, classified into different types based on the age of onset and the severity of motor impairment.[3] Earlier onset generally corresponds to greater disease severity. With an onset of the disease occurring within the first six months of life, type 1 represents the most severe form of SMA. Affected infants are unable to sit independently and have a life expectancy of less than two years. Type 2 surfaces between the ages of seven and 18 months with affected individuals never able to stand unaided. With an age of onset after 18 months, type 3 patients typically achieve independent walking. The least severe type 4 cases emerge in adulthood and exhibit very limited impairment of motor function.[4]

The different forms of SMA are caused by genetic mutations in the survival motor neuron (SMN) genes. Two copies of the SMN gene are located on chromosome 5q13, both consisting of nine exons and eight introns: the telomeric SMN1 gene and the centromeric SMN2 gene.[5,6] Although mostly identical, only genetic impairments of SMN1 are crucial for the development of SMA. Homozygous deletions or loss-of-function mutations of SMN1 are present in more than 98% of SMA patients.[5] Whereas SMN1 produces a fully functional full-length SMN protein, the protein products derived from SMN2 are largely impaired.[5] A single point mutation in exon 7 of SMN2 results in an alternate splicing pattern, leading to a predominant production of functionally defective SMN protein variants.[7] Although SMN2 is not directly responsible for the development of SMA, disease severity correlates with the level of functional SMN protein and consequently with the ability of SMN2 to rescue SMN1 function.[8] Despite the known genetic cause of SMA, the physiological role of the SMN protein, especially its link to the survival of alpha motor neurons, is not fully understood.

1.2 Survival motor neuron protein (SMN)

The SMN protein consists of 294 amino acids with a total mass of approximately 32 kDa.[5] SMN features some known structural domains within its sequence, including a Tudor domain,[9] a polyproline motif,[10] and a YG-box.[11] SMN is expressed in multiple tissues throughout the human body, including the brain, heart, liver, and muscles.[5] It is ubiquitous in the cytoplasm and is also found at high abundance in discrete nuclear bodies called "Gemini of Cajal bodies" (gems), which are similar in both size and number to Cajal bodies and are often associated with them. However, their specific function is unclear.[12]

SMN is prone to oligomerization and self-association, mediated by a 30-residue region between amino acids 249 and 278 in its *C*-terminal domain. SMN mutations found in patients with SMA significantly decrease self-association efficiency, which may relate to disease severity.[13] In addition to hindering oligomerization, SMA-linked mutations in SMN also disrupt the proper formation of the multiprotein SMN complex.[14] To fully understand the significance of the SMN protein, it is necessary to understand the structure and function of the SMN complex.

1.2.1 The SMN complex

The SMN complex consists of nine core elements: SMN, Gemin2 to 8 and the unr-interacting protein (unrip).[15–22] Extensive research has revealed the modular details of the complex. SMN together with Gemin8 and Gemin7 form the backbone of the complex, interacting with the remaining components.[14] **Figure 1.1** shows a simplified structural overview of the core components of the SMN complex and their major interaction partners. Gemin proteins are closely associated with SMN inside the cell and co-localize with it, both in the cytoplasm and in the nucleus.[15–21] Research also indicates that Gemin4 plays a unique role in the localization of the complex, since overexpression increases nuclear translocation.[23]

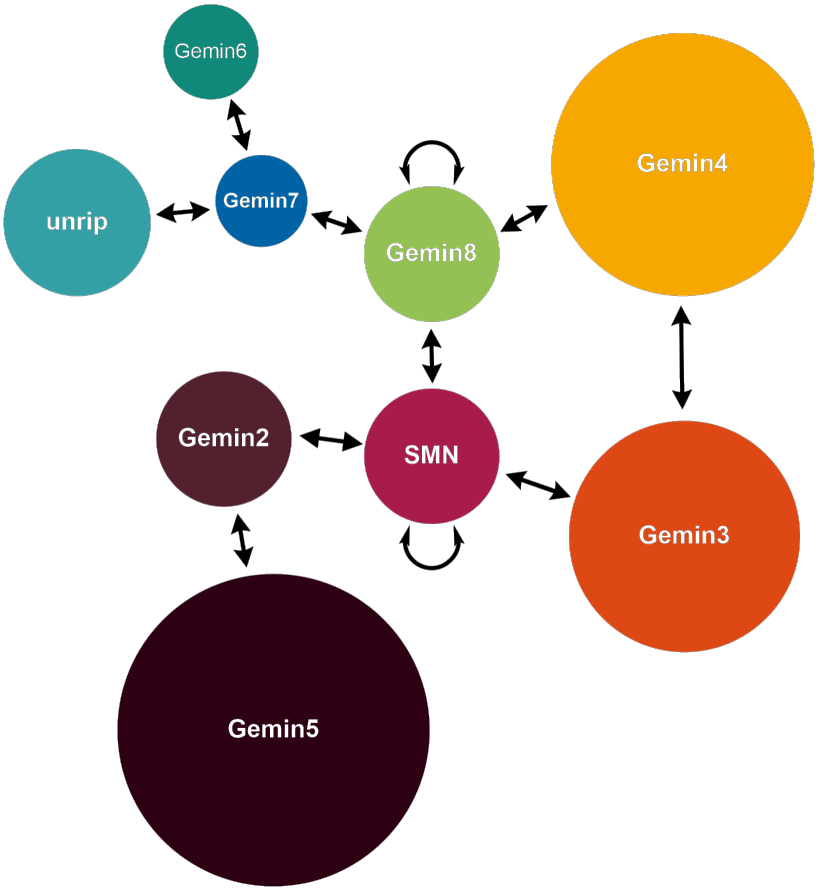

Figure 1.1. Comprehensive interaction map of the SMN complex as described by Otter *et al.*, 2007.[14]

Although the individual components of the complex and their probable positions are known, the interactions that drive its assembly are still poorly understood. The Gemin2-SMN interaction is mediated by an α-helix located between residues 37–51 of SMN and helices located in the *C*-terminal domain of Gemin2.[24] The binding of Gemin5 to Gemin2 is attributed to interactions of its *C*-terminal residues 721–1508 and Gemin2.[14] Gemin3 relies on its residues 456–547 for its interaction with SMN.[16,25] The Gemin3-SMN interaction also proved essential for a successful association of Gemin4 with the complex.[14] More research is needed to fully unravel the underlying interactions that drive the assembly of the complex. Nevertheless, the assembled complex plays a central role in cellular function, and to fully grasp its significance and therefore the significance of the SMN protein, one must take a look at the mechanism of pre-mRNA splicing.

1.2.2 Precursor messenger RNA (pre-mRNA) splicing and the SMN complex

Precursor messenger RNA (pre-mRNA) contains both coding (exon) and noncoding (intron) segments. To remove introns and generate a fully coding mRNA sequence, splicing is required. The splicing pathway was independently discovered in 1977 by Richard Roberts[26] and Phillip Sharp.[27] From a mechanistic perspective, splicing can be described as a two-step S_N2-type transesterification reaction that links two adjacent exons while eliminating introns in the process.[28] A schematic overview of the splicing reaction is shown in **Figure 1.2**. Splicing of pre-mRNA usually occurs only after its assembly into a ribonucleoprotein complex called the spliceosome.[29,30]

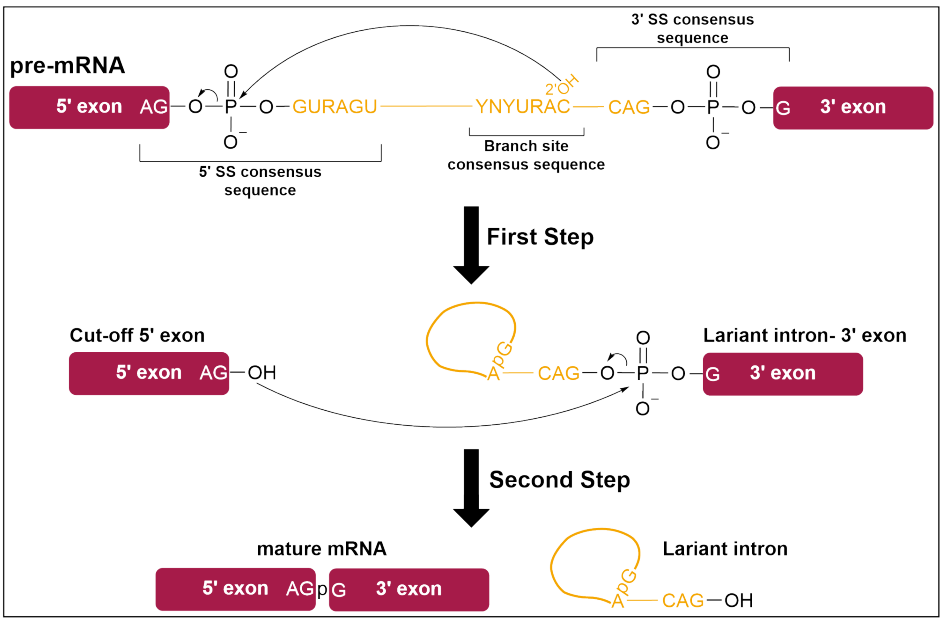

Figure 1.2. Mechanism of pre-mRNA splicing. Splicing involves two S_N2-type transesterification steps yielding the linkage of adjacent exons. This illustration was redrawn from Morais *et al.*, 2021.[31]

The spliceosome is composed of five small nuclear ribonucleoprotein particles (snRNPs). Consensus sequences within the pre-mRNA mediate interactions with snRNPs and other splicing factors, allowing transesterification reactions to occur.[31,32] Each snRNP contains one of five small nuclear ribonucleic acids (snRNAs): U1, U2, U4, U5, or U6. A set of proteins accompanies these snRNAs, with some being

specific to individual snRNAs and others common to all of them. The common proteins of snRNPs comprise eight members (B, B', D1, D2, D3, E, F, G) and are known as Sm proteins.[33] Although individual Sm proteins do not contain known RNA-binding domains and cannot interact with snRNA in a stable manner,[34] interactions among different Sm proteins can be observed.[35] These interactions lead to the stepwise formation of a heptameric Sm protein ring and the incorporation of an snRNA in its centre, producing the typical snRNP structure.[36,37] However, because Sm proteins lack specificity for unique snRNAs,[38] the stepwise snRNP assembly process requires assistance. This assistance is provided by the SMN complex.[22]

The assembly of snRNPs occurs in stages and involves several additional proteins. In the early phase, a five-membered Sm complex forms in association with the pICln protein, a component of the PRMT5 complex that acts as an assembly chaperone to prevent early binding of snRNA.[39] The SMN complex subsequently replaces pICln, with Gemin2 recognizing the Sm structure and regulating access to the RNA-binding pocket.[24,39] Once Gemin5 delivers the correct snRNA, the binding pocket becomes accessible, allowing its incorporation.[40] The subsequent addition of the remaining two Sm proteins completes the snRNP assembly process.[39] The SMN complex also accompanies snRNPs back to the nucleus, where further maturation steps occur.[41] Although interactions between the remaining components of the SMN complex and several Sm proteins have been identified,[15–22] little is known about their specific roles in the assembly process. Impaired snRNP formation has been associated with SMA-causing mutations in SMN that compromise its ability to bind Sm proteins.[42]

Given the physiological role of the SMN complex, understanding its regulatory mechanisms is equally important. Exploring posttranslational modifications of both the SMN complex and SMN protein itself is essential.

1.2.3 Posttranslational modifications (PTMs) of the SMN protein

Posttranslational modifications (PTMs) are protein modifications that occur following RNA translation. PTMs include covalent attachments to amino acid side chains or to the *C*- or *N*-termini of proteins, as well as the removal of individual amino acids or peptide segments.[43] Covalent attachments range from small moieties to large biomolecules, with phosphorylation, acetylation, N-linked glycosylation, amidation, and hydroxylation being the most frequently observed.[44] PTMs perform various tasks, such as controlling enzyme activity, organizing protein localization, and regulating protein–protein interactions.[43] Phosphorylation is the most prevalent modification in eucariotic cells and typically occurs on serine, threonine, and tyrosine residues.[44] Protein kinases mediate phosphorylation by catalyzing the transfer of a phosphate group from ATP to a target protein. More than 500 kinase genes have been identified, representing approximately 1.7% of the human genome.[45] Phosphorylations are reversible and can be removed by phosphatases.[46]

Phosphorylation appears to be important for the activity and protein-protein interactions of the SMN protein and the entire SMN complex. SMN and Gemin2–6 proteins contain multiple phosphorylation sites, with SMN and Gemin3 harbouring the highest total number.[47] SMN contains a total of

50 phosphorylation sites, the majority of which are located within its *N*-terminal domain. Twenty-six phosphorylation sites have been found by mass spectrometry, and specific functions have already been characterized for a few of them (**Figure 1.3**).[47] Phosphorylation at serine residues 28 and 31 of SMN has proven essential for the successful assembly of snRNPs. However, although both residues are within the Gemin2-binding domain of SMN, they do not hinder this interaction.[48] Phosphorylation at serine residues 49 and 63 modulate the condensation behaviour of SMN and the SMN complex inside Cajal bodies. These phosphorylation-dependent effects also influence SMN function and snRNP assembly. Given their proximity to the Gemin2 binding site of SMN, S49 and S63 might affect this interaction as well.[49] Phosphorylation of S290 within the *C*-terminal domain was found to decrease both stability of SMN as well as its ability to oligomerize.[50] Phosphorylation also influences localization of the SMN protein and the SMN complex. Highly phosphorylated SMN isoforms occur more frequently in the cytoplasm, whereas less modified versions are predominantly found in the nucleus.[48]

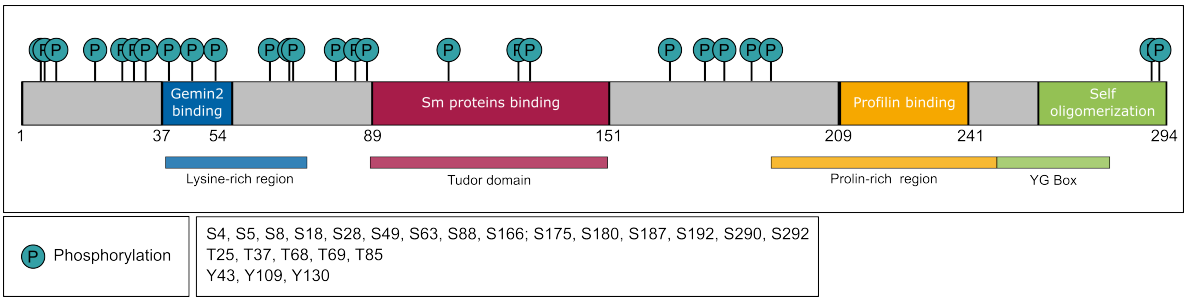

Figure 1.3. SMN domain structure with confirmed phosphorylation sites. The image was redrawn in a simplified manner from Detering *et al.*, 2022.[47]

Investigating how specific phosphorylations affect protein-protein interactions or other functional properties can be challenging. These studies rely on access to protein variants carrying a defined set of modifications, which can be obtained by various means. One strategy involves substituting phosphorylated serine or threonine residues with negatively charged amino acids such as glutamate to approximate the electrostatic effect of the phosphate group.[51,52] However, these phosphomimetic subsitutions do not fully reproduce the chemical properties of native phosphorylation.[53] Enzymatic phosphorylation using kinases offers another route to to generate phosphopeptides, but identifying a kinase with sufficient site specificity for a target site can be challenging.[52,54] Alternatively, phosphorylated amino acids can be introduced either through genetic code expansion during recombinant protein expression or via entirely chemical methods, such as solid phase peptide synthesis (SPPS).[52,55]

1.3 Solid phase peptide synthesis (SPPS)

From a chemical perspective, solid phase peptide synthesis (SPPS) is the most fundamental technique for peptide production. The peptide sequence is assembled in a stepwise manner from the *C*-terminus towards the *N*-terminus. The synthesis starts with the attachment of the first amino acid onto a solid support, followed by the sequential coupling of amino acids.[56] The use of a solid support allows excess reagents and side products to be efficiently removed through filtration and washing steps.[57] Side chain functional groups are protected with permanent protecting groups, whereas the *N*-terminus of

each amino acid building block carries a temporary protecting group.[57] Temporary protection of the *N*-terminus can be achieved using the Boc or the Fmoc strategy. The Boc strategy is based on the *tert*-butyloxycarbonyl protecting group, which is removed using trifluoroacetic acid (TFA), while global deprotection of side chains requires strong acids such as hydrofluoric acid.[57] In contrast, the Fmoc strategy relies on the base-labile 9-fluorenylmethoxycarbonyl protecting group, whereas acid-labile groups are used for side chains and are typically removed during global deprotection with TFA. Although both methods use orthogonal protection strategies, the Fmoc scheme enables deprotection under relatively milder conditions due to its distinct deprotection mechanisms.[57,58] SPPS also provides a comparatively simple method for the site-specific incorporation of PTMs into peptide sequences.[59] Single or multiple PTMs can be introduced using commercially available building blocks, allowing the synthesis of proteins with a defined and homogenous modification pattern.[60]

Appropriate coupling conditions and an effective protection strategy are essential for successful peptide synthesis.[56] Incomplete coupling or deprotection reactions can lead to the accumulation of truncated or modified byproducts, limiting the efficiency and applicability of SPPS for long peptides.[61,62] Typically, SPPS is suitable for peptides of approximately 40–50 amino acids in length, although this limit depends on the specific sequence and the synthesis strategy employed.[59] The synthesis of large proteins beyond this range requires the combination of SPPS with peptide ligation techniques, such as native chemical ligation (NCL).

1.4 Native chemical ligation (NCL)

Proteins and large peptides can be divided into smaller segments and sequentially assembled through a process known as ligation. Among the many known ligation strategies, native chemical ligation (NCL) is arguably the most successful method. NCL enables the chemoselective conjugation of an unprotected peptide bearing a *C*-terminal thioester with another unprotected peptide containing an *N*-terminal cysteine.[63] The reaction proceeds via a reversible trans-thioesterification between the *N*-terminal cysteine residue and the peptide thioester, followed by an irreversible S → N-acyl shift that results in the formation of a native peptide bond (**Figure 1.4**).[63]

NCL reactions are typically performed in aqueous solution at neutral pH in presence of a chaotropic agent such as guanidine (Gnd). These mild reaction conditions enhance the solubility of the peptide reactants, while the chaotropic additive suppresses the formation of secondary structural elements.[59,62] Ligation efficiency strongly depends on the structure of the ligation site, with reduced efficiency observed for sterically hindered residues at the *C*-terminus of peptide thioesters.[64] Changing the identity of the peptide thioester can also improve ligation efficiency.[63] In addition, the use of exogenous thiols in the reaction mixture has been shown to improve ligation rates.[65] Total chemical synthesis using SPPS and NCL has enabled the preparation of proteins up to 358 amino acids long and ubiquitinated proteins of up to 472 amino acids.[66–68] However, its application to larger proteins remains limited, as preparing peptides of this size is very labour-intensive. Only a limited number of ligation and subsequent reaction steps can be combined while maintaining sufficient yields.[59]

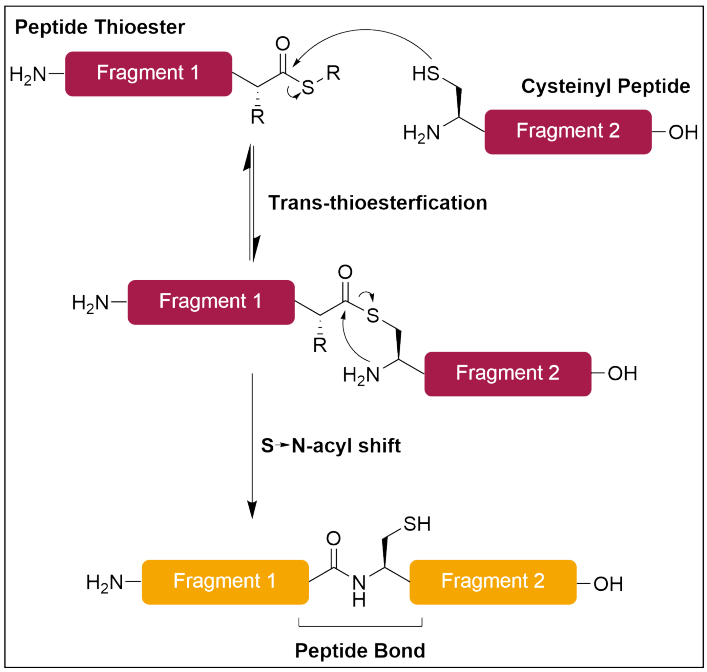

Figure 1.4. Native chemical ligation reaction mechanism. The reversible trans-thioesterfication is followed by an irreversible S → N-acyl shift leading to the formation of a peptide bond. The illustration was inspired by Conibear *et al.*, 2018.[59]

A major limitation of the NCL approach is its dependence on native cysteine residues. However, several strategies have been developed to address this issue, including desulfurization strategies to convert cysteine residues to alanine post-ligation.[69,70] Desulfurization techniques also inspired the development of β-mercapto amino acids to act as cysteine surrogates for NCL reactions.[71] Although successful in NCL applications, the limited commercial availability of β-mercapto amino acids restricts their broader use.[59] An alternative strategy to overcome the dependence on the terminal amino acid involves the use of *N*-terminal auxiliaries. Auxiliaries are equipped with a mercapto group necessary for NCL reactions and cleaved off post-ligation.[62] Although they expand the scope and flexibility of peptide synthesis, auxiliary-mediated ligations often suffer from prolonged reaction times, particularly at sterically hindered ligation sites.[72]

NCL reactions also require the generation of peptide thioesters. Over the years, several methods have been established to make them readily accessible. While Boc-SPPS enables the straightforward synthesis of peptide thioesters using mercaptoalkyl acid linkers, Fmoc-SPPS requires more elaborate strategies due to the instability of thioester bonds under basic conditions.[64,73] However, the limited compatibility of Boc-SPPS with the incorporation of PTMs necessitates methods for generating thioesters that are compatible with Fmoc-SPPS. Early Fmoc-compatible strategies involved activation of the *C*-terminus of fully protected peptides, followed by thioester formation.[73] Alternative strategies include the use of special linkers for peptide synthesis, such as the backbone amide linker (BAL) or sulfonamide-based safety-catch linkers.[74,75]. *N*-acylurea peptides derived from 3,4-diaminobenzoic acid offer an alternative approach.[76] After SPPS, acylation with p-nitrophenyl chloroformate followed by base-induced cyclization and subsequent global deprotection then affords the *N*-acylurea pep-

tide, which can be used either in transthioesterification or directly in NCL in the presence of a thiol catalyst.[76,77] Another common approach involves their generation from *C*-terminal acyl hydrazides.[78] A key advantage of *C*-terminal peptide hydrazides is their straightforward preparation via prefunctionalized solid supports during SPPS.[59] Cleavage from the resin yields a hydrazide moiety, which can subsequently be oxidized with $NaNO_2$ into the corresponding azide. Upon treatment with a thiol, the azide is further transformed into the desired peptide thioester.[78] *C*-terminal hydrazides are stable under NCL conditions, making them suitable for the assembly of small to medium sized proteins through iterative cycles of ligation and thioester conversion steps.[59]

Peptide thioesters containing an *N*-terminal cysteine residue can undergo intramolecular cyclisation under NCL conditions, necessitating the use of cysteine protecting groups, especially in sequential ligation strategies. Acetamidomethyl (Acm) was first introduced as a cysteine protecting group by Veber et al. in 1968.[79] Acm is stable both under common SPPS and NCL conditions and displays orthogonality to many other commonly used cysteine protecting groups.[80] These properties have contributed to its widespread application in peptide synthesis. A variety of Acm deprotection strategies have been developed, using various metal compounds. Following ligations, Acm removal can be achieved in a one-pot manner by treatment with $PdCl_2$, without prior purification.[81] Alternatively, treatment with AgOAc in acetic acid also provides an efficient deprotection method.[82] An additional protecting group, hydroxyquinoline-acetamidomethyl (Hqm), offers orthogonality to Acm and can be selectively cleaved by hydrazine treatment within approximately eight hours.[83,84] Simultaneous protection of both the amino and thiol functionalities can be achieved using thiozolidine (Thz).[85] Thz can be removed with oxidants such as hydrogen peroxide or by treatment with excess methoxyamine at pH 4.[84] However, Thz is unstable under $NaNO_2$ treatment at pH 3–4, which is required for the activation of peptide hydrazides.[86] This limitation can be adressed by installing a 2-(tert-butyldisulfanyl)ethyloxycarbonyl (Tbeoc) moiety to form Tbeoc-Thz, which can be converted back to Thz using tris(2-carboxyethyl)phosphine (TCEP).[86]

2 Research objectives

This thesis focuses on the development of a fully synthetic strategy to generate the *N*-terminal domain of the SMN protein. Previous work in the Becker group has already addressed the synthesis of the *C*-terminal domain of SMN (residues 201–294); therefore, the present work concentrates on the *N*-terminal region (residues 1–160). Studies by Chari *et al.* identified SMN, together with the Gemin2 protein, as the minimal functional unit for Sm core assembly *in vitro*.[39] As the essential binding regions for both Gemin2 and the Sm proteins are located within this target sequence, the SMN(1–160) segment is expected to be sufficient for further investigations into Sm core assembly.

The synthetic approach is based on a combination of solid phase peptide synthesis (SPPS) and native chemical ligation (NCL). The target region, SMN(1–160), is divided into four segments, each synthesized by SPPS and subsequently assembled through convergent NCL reactions. Following optimization of the synthetic procedures and successful preparation of the SMN(1–160) peptide, site-specific introduction of phosphorylation into the target sequence is envisioned. Three phosphorylated SMN(1–160) variants will be synthesized, each bearing a different phosphorylation pattern at serine residues S49 and S63. Schilling *et al.* identified these residues as regulators of the condensation behaviour of the SMN complex, with additional evidence suggesting a potential role in Sm core assembly and snRNP biogenesis.[49]

After successful synthesis of the four variants, further investigations will be undertaken. Circular dichroism (CD) spectroscopy will be performed to evaluate possible differences in secondary structure among the variants. In future studies, the phospho-isoforms are intended for structural analysis to investigate the role of their PTMs in Sm protein interactions and snRNP assembly, as well as to examine their binding properties with Gemin2 and Sm proteins.

3 Materials

3.1 Chemicals and solvents

All materials were purchased from commercial suppliers and used without further purification unless otherwise stated. Water used for buffer preparation and purification procedures was purified prior to use. Deionization was performed using a Professional G7895 Aqua Purificator manufactured by Miele GmbH. Further purification was carried out using a Milli-Q Reference A+ water purification system by Merck GmbH. The solvents acetonitrile (ACN) and N,N-dimethylformamide (DMF) were purchased from Sigma-Aldrich. Dichloromethane (DCM) and diethyl ether (Et_2O) were purchased from VWR Chemicals (Darmstadt, Germany).

Chemicals used for buffer preparation, including tris(2-carboxyethyl)phosphine (TCEP), sodium phosphate monobasic (NaH_2PO_4), and sodium phosphate dibasic (Na_2HPO_4), were purchased from Sigma-Aldrich. Guanidine hydrochloride (Gnd · HCl) and sodium hydroxide (NaOH) pellets were purchased from Merck. Chemicals used in solid phase peptide synthesis and peptide cleavage were purchased from Sigma-Aldrich: 1,2-ethanedithiol (EDT), N,N-diisopropylethylamine (DIPEA), hydrazine (N_2H_4), phenol, piperidine, thioanisole, and triisopropylsilane (TIPS). Piperazine was obtained from Merck, 1,8-diazabicyclo(5.4.0)undec-7-ene (DBU) from Alfa Aesar, formic acid (FA) from Fisher, and 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) from Iris Biotech (Marktredwitz, Germany). Methanol (MeOH) and trifluoroacetic acid (TFA) were supplied by VWR Chemicals (Darmstadt, Germany). Chemicals used for thioester conversion of peptide hydrazides and for ligation procedures were purchased from Sigma-Aldrich and included sodium 2-mercaptoethanesulfonate (MESNa), 4-mercaptophenylacetic acid (MPAA), sodium nitrite ($NaNO_2$), and sodium chloride (NaCl). Acetic acid (AcOH) used for deprotection procedures was purchased from Sigma-Aldrich, dithiothreitol (DTT) from Carl Roth GmbH + Co. KG., and silver acetate (AgOAc) was purchased from Alfa Aesar.

3.2 Instruments for purification procedures

Reverse-phase high-performance liquid chromatography (RP-HPLC) purifications were performed using either a Waters LC-prep 150 system or a Shimadzu system. The Waters system consisted of a Waters 2545 binary gradient module, a Waters 2489 UV/visible detector, and a Waters Fraction Collector III. The Shimadzu system comprised an LC-20AP solvent delivery unit, a FCV-200AL low-pressure gradient unit, a CBM-20ACL system controller, a FRC-10A autosampler, and a SPD-10A UV/visible detector. Separation of compounds was achieved using semi-preparative or preparative C4 and C18 columns. Semi-preparative purifications were performed using either a PerfectSil 300 C4 column (10 µm, 300 Å, 250 × 10 mm) or a PerfectSil 300 C18 column (10 µm, 300 Å, 250 × 10 mm). Preparative purifications were carried out using a PerfectSil 300 C4 column (10 µm, 300 Å, 250 × 20 mm) or a Kromasil 300-10-C18 (10 µm, 300 Å, 21.2 × 250 mm).

3.3 Instruments for analytical procedures

Mass spectrometry (MS) was performed using either a Waters analytical system or a Waters Arc system. The Waters analytical system was equipped with a 2767 Sample Manager, a 515 HPLC pump, a 2545 Binary Gradient Module, a System Fluidics Organizer, a 2489 UV/Vis detector, and a 3100 Mass Detector. The Waters Arc system comprised a SQ detector 2, a quaternary Solvent Manager-R, a Sample Manager FTN-R, a 2489 UV/Vis Detector, and a Zspray ESI source. Both systems were operated in positive ion mode using electrospray ionisation. The Waters analytical system acquired data over an m/z range of 400–2000, whereas the Waters Arc system monitored a range of 400–3000. Liquid chromatography-mass spectrometry (LC-MS) was performed on the Waters Arc system using an XBridge C4 column (3.5 µm, 400 Å, 2.1×100 mm) for analyte separation.

Analytical RP-HPLC-analyses were performed using an UltiMate 3000 Dionex UHPLC system produced by Thermo Fisher Scientific. Analyte separation was achieved using a Kromasil 300-5-C4 column (5 µm, 300 Å, 4.6× 150 mm).

CD spectroscopy was performed on a Chirascan Plus CD spectrometer manufactured by Applied Photophysics. Absorbance measurements at 301 nm for resin loading determinations were carried out with a Bormac ONDA UV-30 photometer (**Section 4.2.2**). The concentration of purified protein samples was determined using a Thermo Scientific NanoDrop^{TM} 200c (**Section 4.3**).

4 Methods

4.1 General analytical procedures

4.1.1 Analytical RP-HPLC

Samples for reaction monitoring were diluted with 6 M Gnd · HCl to a total peptide concentration of 1 mg/mL, whereas purified product samples were dissolved in either 6 M Gnd · HCl or ddH_2O depending on their solubility. TCEP was added to prevent disulfide formation, and all samples were centrifuged (5 min, 14000 rpm) prior to analysis. Elution was performed using a 5–65% gradient of buffer B (0.8% TFA in ACN) in buffer A (0.1% TFA in ddH_2O) over 30 minutes at a flow rate of 1 ml/min. The purity of all products was assessed using the aforementioned method in combination with MS analysis, and a purity exceeding 90% was regarded as high and satisfactory. Retention times (RT) and peak areas (PA) were used for compound characterisation and reaction monitoring. Due to peak broadening and limited chromatographic resolution, the calculated peak areas should be considered as rough estimates rather than absolute values.

4.1.2 Mass spectrometry

Direct injection mass spectrometry was used for purified products. Samples were dissolved in either ddH_2O or a 1:1 (v/v) mixture of ACN and ddH_2O. Samples were subsequently centrifuged (5 min, 14000 rpm) and directly injected using 50% buffer B (0.05% TFA in ACN) in buffer A (0.05% TFA in ddH_2O) over 1 minute at a flow rate of 1 ml/min. Analysis was performed in positive-ion mode using ESI. For reaction monitoring, samples were diluted with 6 M Gnd · HCl to a total peptide concentration of 1 mg/mL, centrifuged (5 min, 14000 rpm) and subsequently analysed by LC-MS. Analyte separation was performed on a C4 column using a gradient of 5–65% buffer B (0.05% TFA in ACN) in buffer A (0.05% TFA in ddH_2O) over 7 or 12 minutes at a flow rate of 0.5 mL/min.

4.2 General synthetic procedures

4.2.1 Resin preparation - Hydrazine modification of 2-CTC resin

For the SPPS of SMN(60–97) and SMN(98–122), appropriate preloaded resins were not consistently available. As an alternative, a 2-chlorotrityl chloride (2-CTC) resin with a loading of 1.60 mmol/g was chemically modified to introduce a hydrazide functionality. To reduce the effective resin loading and improve the synthetic outcome, the first amino acid of each peptide was coupled manually. About 300 mg of resin (corresponding to a 0.48 mmol synthesis scale) was transferred to a filter syringe and washed with DMF (3 × 5 mL), DCM (3 × 5 mL) and DMF (3 × 5 mL). The resin was then allowed to swell in a 1:1 (v/v) mixture of DMF and DCM for 30 minutes, after which it was treated with 10 vol% $NH_2NH_2 \cdot H_2O$ in DMF for 30 minutes. Following washing with DMF (3 × 5 mL) and DCM (3 × 5 mL), the hydrazine treatment was repeated. After an additional washing step identical to the previous ones with DMF and DCM,

a capping step was performed by treating the resin with 5 vol% MeOH in DMF for 40 minutes to block any remaining reactive sites. The resin was then washed with DMF (3 × 5 mL) and DCM (3 × 5 mL) and dried under vacuum overnight. On the following day, the first amino acid was coupled to reduce the resin loading. Four equivalents of protected amino acid (two parts Boc-protected and one part Fmoc-protected) were dissolved in DMF to a total amino acid concentration of 5 mM. The resin was suspended in this solution, followed by the addition of 3.8 equivalents of HBTU and 8.0 equivalents of DIPEA. The reaction was shaken for 30 minutes at room temperature. Excess reagents were removed by washing the resin with DMF (3 × 5 mL) and DCM (3 × 5 mL), after which the resin was dried under vacuum overnight. Before being used in SPPS, the resin loading was determined the following day as described in **Section 4.2.2**.

4.2.2 Loading determination

10 mg of resin was transferred to a 10 mL volumetric flask and suspended in 10 mL 20% piperidine solution in DMF. Aliquots of 100 µL were taken after approximately 30 minutes and diluted with 900 µL of 20% piperidine. Absorbance at 301 nm was measured using a Bormac ONDA UV-30 photometer. A 20% piperidine solution was used as a blank. Loading was calculated according to the Beer-Lambert law (**Equation 4.1**) using the mean of triplicate measurements.

$$Loading(mmol/g) = \frac{A \cdot d}{\varepsilon \cdot l \cdot m} \cdot V \cdot 10^3 \tag{4.1}$$

Equation 4.1, based on the Beer-Lambert law, relates the resin loading to the measured absorbance A, the dilution factor d (10), the molar extinction coefficient ε (7731.9 mol^{-1} L cm^{-1}), the cuvette path length l (1 cm), the mass of the resin m (10 mg) and the volume of piperidine solution V (10 mL). An adjustment factor accounts for unit conversions (10^3).

4.2.3 General solid phase peptide synthesis procedure

All peptides were synthesized via automated microwave-assisted Fmoc-based SPPS on a Liberty PRIME peptide synthesizer manufactured by CEM GmbH. Owing to differences in the properties and specific characteristics of each peptide, modifications to this general strategy were required and are elaborated in **Section 5.2**. Upon completion of the synthesis, each resin was transferred to a 10 mL syringe equipped with a polyethylene frit. The resin was subsequently washed with DMF (3 × 5 mL) and DCM (3 × 5 mL) to remove excess reagents and residual byproducts of the synthesis. The resin was left to dry under vacuum overnight.

The dried resin was then suspended in 5 mL of a cleavage solution and incubated at room temperature for up to two hours. The cleavage conditions varied among peptides and were chosen based on results from small-scale test cleavages. Peptides containing oxidation-sensitive residues like tryptophan or methionine were cleaved using cleavage cocktail A (82.5% TFA, 5% ddH_2O, 5% phenol, 5% thioanisole, 2.5% EDT), whereas less sensitive peptides were cleaved with cocktail B (95% TFA, 2.5% ddH_2O, 2.5% TIPS). The cleavage solution was then separated and diluted tenfold with cooled,

4°C Et_2O to precipitate the peptide. The precipitate was pelleted through centrifugation (5–10 min, 4500–5500 rpm) and dried with N_2 gas, before being redissolved in either ddH_2O or a 1:1 (v/v) mixture of ACN/ddH_2O and lyophilized.

The lyophilized product was dissolved in 6 M Gnd · HCl, followed by centrifugation (5–10 min, 4500 rpm) and filtration through a 0.2 µm syringe filter. Preparative RP-HPLC was performed on the Shimadzu or Waters HPLC systems using C4 or C18 columns. A solvent gradient was used, ranging from 5–65% of buffer B (ACN + 0.08% TFA) in buffer A (ddH_2O + 0.10% TFA) over 60 minutes, with a 10-minute pre-run for desalting.

4.2.4 Thioester conversion

Thioester conversions of the hydrazide moieties were carried out for peptides SMN(1–59), SMN(1–59)pS49, and SMN(98–122). Following the ligation with their respective neighbouring segments, the intermediate ligation product SMN(1–97) and its phospho variants SMN(1–97)pS49, SMN(1–97)pS63, and SMN(1–97)pS49pS63 were also converted into thioesters. All conversions were performed following a general procedure illustrated in **Figure 4.1** with minor modifications depending on the peptide variant. All peptide hydrazides were converted into MESNa thioesters, due to their increased stability in aqueous solutions compared to MPAA thioesters.[63]

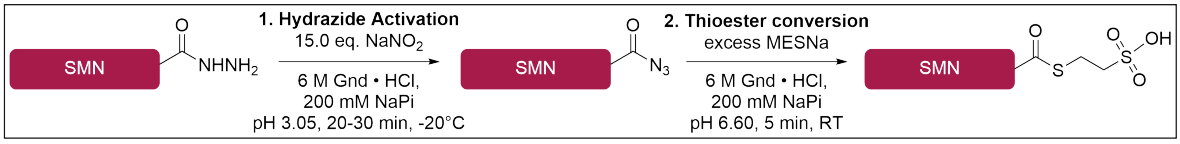

Figure 4.1. Synthetic procedure for thioester conversions. The excess of MESNa varied among the conducted conversions.

The peptide hydrazide was dissolved in a buffer containing 6 M Gnd · HCl, 0.2 M NaH_2PO_4 (pH 3.05) to a concentration of 5 mM. The solution was cooled to -20 °C using an ice-NaCl bath, after which aqueous $NaNO_2$ (0.5 M, 15.0 eq.) was added. The reaction was stirred at -20 °C for 20–30 minutes, depending on the peptide hydrazide. An excess of MESNa was then added, and the pH was adjusted to 6.60 using NaOH. The amount of MESNa used varied between conversions and is specified in **Sections 5.3** and **5.5**. The reaction mixture was subsequently stirred for an additional 10 minutes at room temperature. Products were purified by first diluting the reaction mixture with 6 M Gnd · HCl, followed by filtration through a 0.2 µm syringe filter. The compounds were then isolated by RP-HPLC using the Waters HPLC system equipped with a semi-preparative C4 column. A solvent gradient ranging from 5–65% of buffer B (ACN + 0.08% TFA) in buffer A (ddH_2O + 0.10% TFA) was applied over 60 minutes, with a 10-minute pre-run for desalting.

4.2.5 Ligation reactions

The thioesters of peptides SMN(1–59), SMN(1–59)pS49 and SMN(98–122) were used for ligation with their neighbouring peptides according to the proposed synthetic strategy (**Figure 5.2**). The thioester of the ligation product SMN(1–97), its phospho-variants, and the deprotected ligation product SMN(98–160) were treated in the same way. All ligations were performed following a general procedure. The *C*-terminal thioester peptide and the *N*-terminal cysteine peptide were dissolved in a buffer containing 6 M Gnd · HCl and 0.2 M NaH_2PO_4 (pH 3.05), with total peptide concentrations ranging from 5.0 mM to 7.0 mM, depending on the peptides used. Appropriate amounts of MPAA and TCEP were added, and the pH was adjusted to 6.80. After addition of all compounds, the reaction was flushed with N_2 and stirred at room temperature. Ligation progress was monitored by analytical RP-HPLC and LC-MS analysis. Ligations were generally allowed to proceed for about 40 hours before purification. To isolate the products, the reaction mixture was diluted with 6 M Gnd · HCl and filtered through a 0.20 µm syringe filter prior to semipreparative RP-HPLC on either the Waters or the Shimadzu system. Purification was performed on C4 columns using a gradient of 5–65% buffer B (ACN + 0.08% TFA) in buffer A (ddH_2O + 0.10% TFA) over 60 minutes with a 10-minute pre-run for desalting.

4.3 Circular dichroism spectroscopy

Samples of the SMN(1–160) variants, as well as the SMN(1–59) and SMN(1–59)pS49 peptide segments were weighed and dissolved in 10 mM sodium phosphate buffer to a final concentration of 0.25 mg/mL. The solutions were placed on a shaker overnight (1040 rpm, 25°C) and measured the following day. SMN(1–160) did not fully dissolve in the buffer. Therefore, its concentration was determined using a Thermo Scientific NanoDrop™ 2000c. The extinction coefficient of SMN(1–160) was estimated using Expasy - ProtParam.[87] The extinction coefficient at 280 nm was calculated to be 22710 M^{-1} cm^{-1}, yielding a measured concentration of 0.087 mg/mL. CD measurements were performed in a quartz microcuvette (1 mm pathlength, 400 µL) with the 10 mM sodium phosphate buffer as a blank. For each sample, five replicates were recorded between 180 nm and 260 nm at 22°C. The replicates were averaged and the resulting ellipticity values (θ, in millidegrees) were converted to mean residue ellipticity (MRE, deg·cm^2·$dmol^{-1}$). A concentration of 0.25 mg/mL was assumed for all samples, except for SMN(1–160), for which the measured concentration was used.

$$MRE = \frac{\theta \cdot 10^6}{c \cdot l \cdot n} \tag{4.2}$$

Equation 4.2 allows the conversion of the observed ellipticity θ into the corresponding MRE values, using the protein concentration c (mol/L), the optical path length l (cm), the number of amino acid residues n, and an adjustment factor (10^6).

5 Results and Discussion

5.1 Synthetic strategy

The *N*-terminal region of the SMN protein, comprising the first 160 amino acids, was synthesized using a fully synthetic approach. Due to the 40–50 amino acid size limitation of SPPS,[59] the target sequence was divided into four segments to facilitate synthesis. Each segment was designed considering the location of native cysteines, the length of each peptide and their predicted solubility. The presence of native cysteines at segment interfaces allowed the linking of neighbouring segments via NCL reactions. The target sequence contains five cysteine residues, allowing for several possible segmentation strategies.

In previous synthesis attempts of SMN(1–160), performed by Dr. Claudia Cobos Caceres, the target region was divided at cysteine residues C60, C98 and C146, resulting in peptide segments SMN(1–59), SMN(60–97), SMN(98–145) and SMN(146–160). SMN(98–145) caused significant challenges due to its high hydrophobicity, leading to low yields, poor solubility, and challenging ligation reactions. In contrast, SMN(146–160) exhibited high solubility and no solubility issues in aqueous solvents. To improve the synthesis outcome, the ligation site between segments 3 and 4 was shifted from C146 to C123, thereby transferring 24 amino acids from segment 3 to segment 4, while peptides SMN(1–59) and SMN(60–97) remained unchanged (**Figure 5.1**).

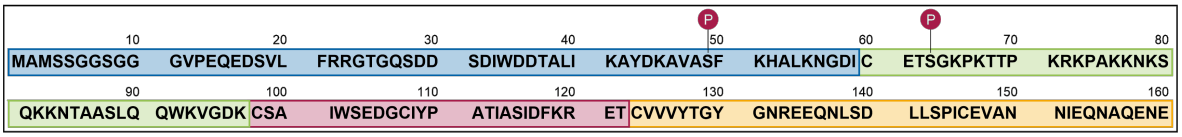

Figure 5.1. Amino acid sequence of the SMN target region 1–160. Incorporation of phosphorylations is planned at serine residues S49 and S63. Peptide segments are indicated by colour.

This new approach was explored in this thesis. The new segmentation strategy, based on cysteines at positions C60, C98 and C123, led to the peptides SMN(1–59), SMN(60–97), SMN(98–122) and SMN(123–160). The first segment, SMN(1–59), comprises 59 amino acids and is the longest peptide. Although it exceeds the commonly accepted size limit for SPPS, the absence of cysteines within the segment would otherwise have required additional synthetic steps to preserve the native sequence. The second segment, SMN(60–97), comprises 38 amino acids, while the third segment, SMN(98–122), contains 25 amino acids, making it the shortest. The fourth segment, SMN(123–160), also comprises 38 amino acids. SMN(1–59), SMN(60–97) and SMN(98–122) were synthesized as peptide hydrazides to enable their conversion into peptide thioesters required for NCL reactions.[88] The first ligations were performed simultaneously in both the $N \rightarrow C$ direction for peptides 1 and 2, and in the $C \rightarrow N$ direction for peptides 3 and 4. The *N*-terminal cysteine (C98) of SMN(98–122) was protected with an acetamidomethyl (Acm) group to prevent intramolecular cyclization during ligation with SMN(123–160) and to enable selective ligation.[80] The ligation strategy is shown in **Figure 5.2**.

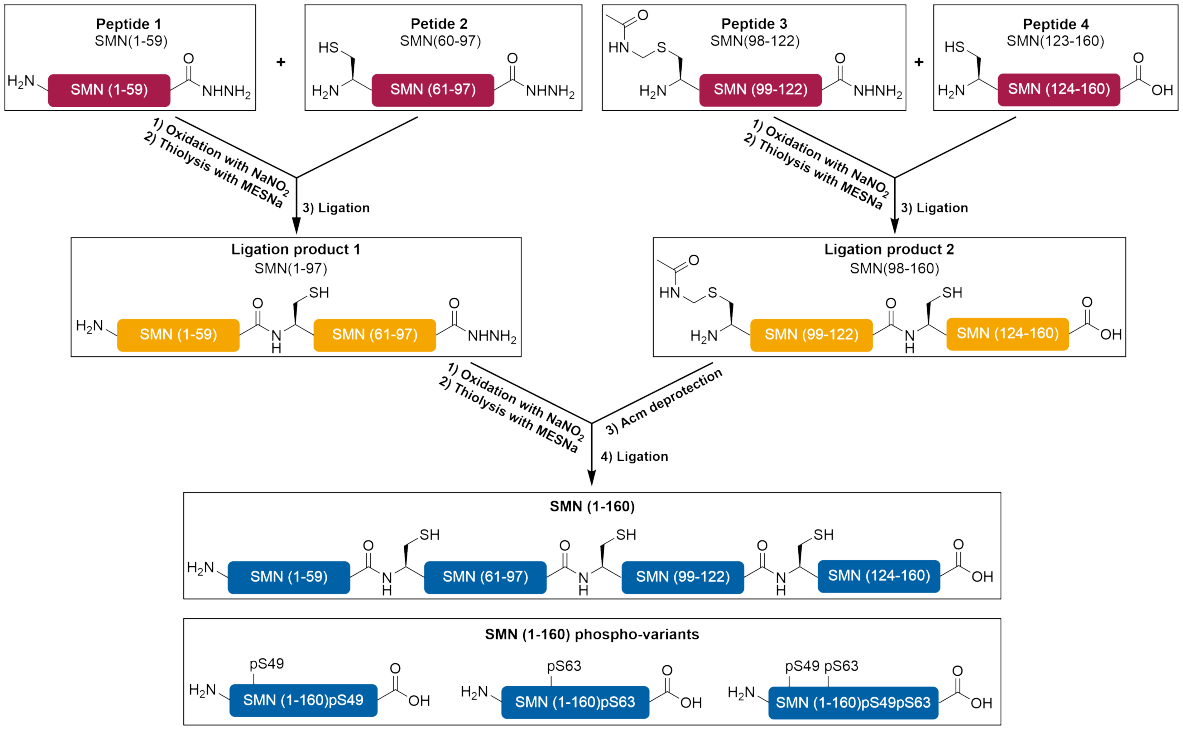

Figure 5.2. Ligation strategy for the synthesis of SMN(1–160). The sequence was divided into four peptide segments, each produced via SPPS. Three thioester conversions, three native chemical ligations, and one Acm-removal step were required to generate each SMN variant.

SMN(1–59) was first converted into a thioester and ligated with SMN(60–97). SMN(98–122) was also converted into a thioester and ligated with SMN(123–160). After the first ligations, the *C*-terminal hydrazide of the intermediate SMN(1–97) segment was converted into a thioester, while the *N*-terminal cysteine (C98) of SMN(98–160) was deprotected by removal of the Acm group. The two intermediate ligation products were subsequently coupled to afford the full-length target segment SMN(1–160). The developed strategy was then applied to synthesize phosphorylated SMN variants. The full synthetic approach enables site-selective incorporation of posttranslational modifications into individual peptide segments. Phosphoserine was incorporated into peptide 1 (pS49) and peptide 2 (pS63), and the combination of modified and unmodified versions of these peptides were assembled to generate the different SMN variants.

Various experimental procedures were explored to synthesize the intermediates of this strategy. The most reliable and successful methods leading to the synthesis of all SMN(1–160) variants are described in the following sections.

5.2 Solid phase peptide synthesis of SMN segments

The synthesis of the SMN(1–160) variants commenced with the generation of the necessary peptide segments via Fmoc-based SPPS (see **Section 4.2.3**). All six peptides were obtained in sufficient quantities and with high purity, as confirmed by analytical HPLC and MS. Detailed results for each SPPS are presented in the following sections.

5.2.1 Synthesis of SMN(1–59)

Peptide SMN(1–59) was synthesized on a 0.05 mmol scale using a Hydrazino TentaGel TRT resin with a loading of 0.22 mmol/g. The SPPS protocol included residue double couplings, the use of pseudoprolines, and capping steps (**Table 5.1**). Cleavage was performed using cleavage cocktail A due to the presence of methionine residues. The resin was treated with the cleavage cocktail for two hours. The crude peptide was dissolved in a 1:1 (v/v) mixture of ACN/ddH_2O before lyophilization, although dissolution in ddH_2O was also possible due to the high solubility of the peptide. Purification by RP-HPLC was performed on the Shimadzu system using a preparative C4 column.

Table 5.1. SPPS protocol of SMN(1–59)

SMN(1–59)	
Coupling Conditions	**Amino Acids**
Single coupling	F50, H52 to I59
Double coupling	M1 to S49, K51
Pseudoprolin double coupling	G7-S8, D17-S18, D30-S31, D36-T37
Capping step	V19, R22, W34, I40, A42, K45

SPPS of SMN(1–59) afforded a yield of 18% (57 mg) with high purity. The calculated molecular weight of 6217 Da matched its observed value (**Figure 5.3**). SMN(1–59) showed high solubility in aqueous solvents. The SPPS procedure of SMN(1–59) required extensive optimization to obtain the peptide in sufficient amounts. This investigation was conducted by Dr. Claudia Cobos Caceres in earlier phases of the project (**Section 5.1**). Most amino acids were double-coupled. In addition, capping steps were applied after residues that are prone to deletion. Pseudoprolines were used at serine and threonine residues to further improve the synthesis outcome, as their use is known to facilitate the synthesis of difficult peptide sequences.[89] The optimized protocol effectively minimized the formation of truncated byproducts and yielded a crude peptide that was easy to purify. Considering the length of the peptide, which reaches the practical limits of SPPS, the obtained yield is very satisfactory.

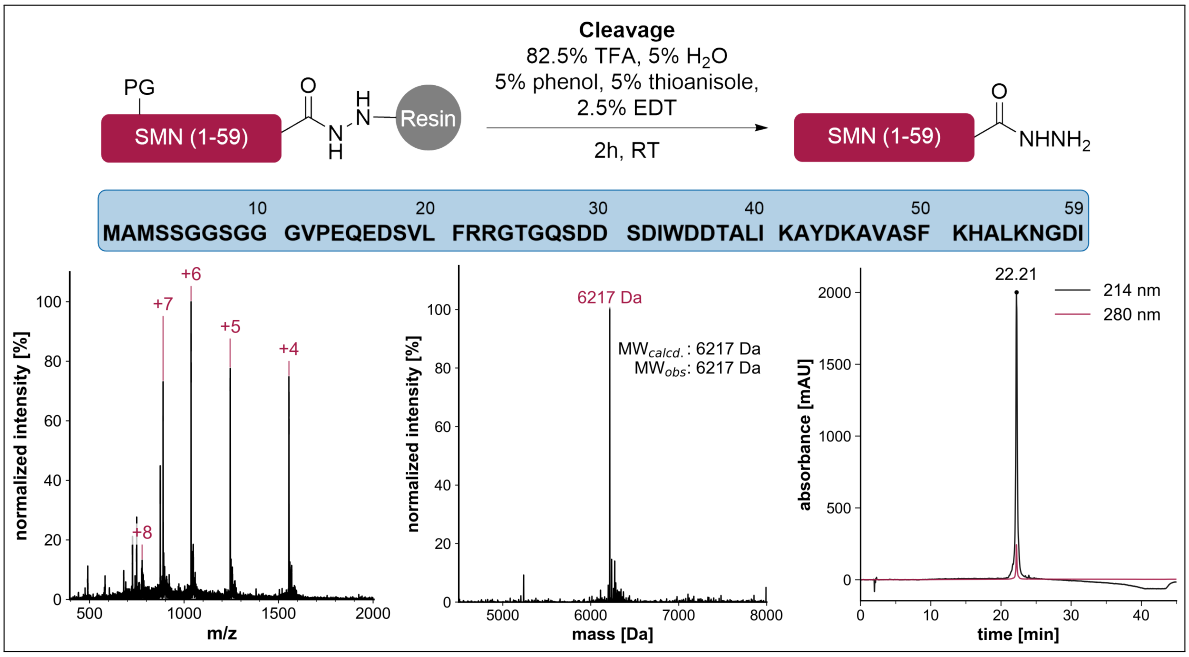

Figure 5.3. Cleavage conditions and analytical data of SMN(1–59). Top: Cleavage conditions and amino acid sequence of SMN(1–59). **Bottom:** Analytical data of purified SMN(1–59)-hydrazide. **Left:** m/z spectrum. **Middle:** deconvoluted mass spectrum. **Right:** analytical RP-HPLC UV/Vis chromatogram recorded at 214 nm and 280 nm.

5.2.2 Synthesis of SMN(1–59)pS49

The synthesis of segment SMN(1–59)pS49 was performed twice due to insufficient yields from the first synthesis. Both were performed on a scale of 0.05 mmol using a Hydrazino TentaGel TRT resin with a loading of 0.22 mmol/g. The SPPS protocol for SMN(1–59)pS49 was adapted from the SMN(1–59) method (**Table 5.1**). Cleavage was performed with cleavage cocktail A at room temperature for two hours. The crude peptide was dissolved in ddH_2O prior to lyophilization. Preparative RP-HPLC purification was performed on the Waters system using a preparative C4 column in both SPPS runs.

In the first synthesis, phosphoserine (pS49) was coupled manually to minimize amino acid usage. Residues 50—59 were assembled automatically on the synthesizer, after which the SPPS cycle was paused. The resin was washed with DMF (3 × 5 mL) and DCM (3 × 5 mL) and dried under vacuum. It was then swollen in a 1:1 (v/v) mixture of DMF and DCM for 30 minutes. Fmoc-Ser(PO(OBzl)OH)-OH (65.9 mg, 0.13 mmol, 2.6 eq.) and HBTU (45.0 mg, 0.12 mmol, 2.3 eq.) were dissolved in DMF to a concentration of 5 mM and added to the resin. After addition of DIPEA (41 µL, 0.24 mmol, 4.7 eq.), the reaction was incubated at room temperature for 40 minutes. The resin was then washed with DMF (4 × 5 mL) and DCM (4 × 5 mL) and dried under vacuum before resuming automated coupling of amino acids 1–48. Microwave-assisted deprotection was omitted for the remaining residues to avoid microwave-promoted β-elimination of the phosphate group.[90] Additionally, Fmoc deprotection was performed with piperazine (5%), DBU (1%), and FA (1%) in DMF to prevent piperidine-induced β-elimination.[91]

For the second synthesis of SMN(1–59)pS49, the protocol was modified to improve the overall outcome. SPPS was performed fully automated, as a sufficient supply of phosphoserine was available. Furthermore, microwave-assisted deprotection was omitted only for residues 46–48, for which Fmoc removal was also performed using the piperazine-based deprotection mixture. These measures were intended to suppress β-elimination of the phosphoserine residue while reducing overall synthesis time and improving deprotection efficiency, thereby minimizing byproduct formation.

The first synthesis of SMN(1–59)pS49 yielded 5% (17 mg). The second fully automated SPPS, omitting the microwave function only for residues 46–48, gave a yield of 6% (18 mg). Both syntheses produced SMN(1–59)pS49 of high purity, and the determined molecular weight of 6297 Da matched the calculated value (**Figure 5.4**). SMN(1–59)pS49 showed high solubility in aqueous solvents.

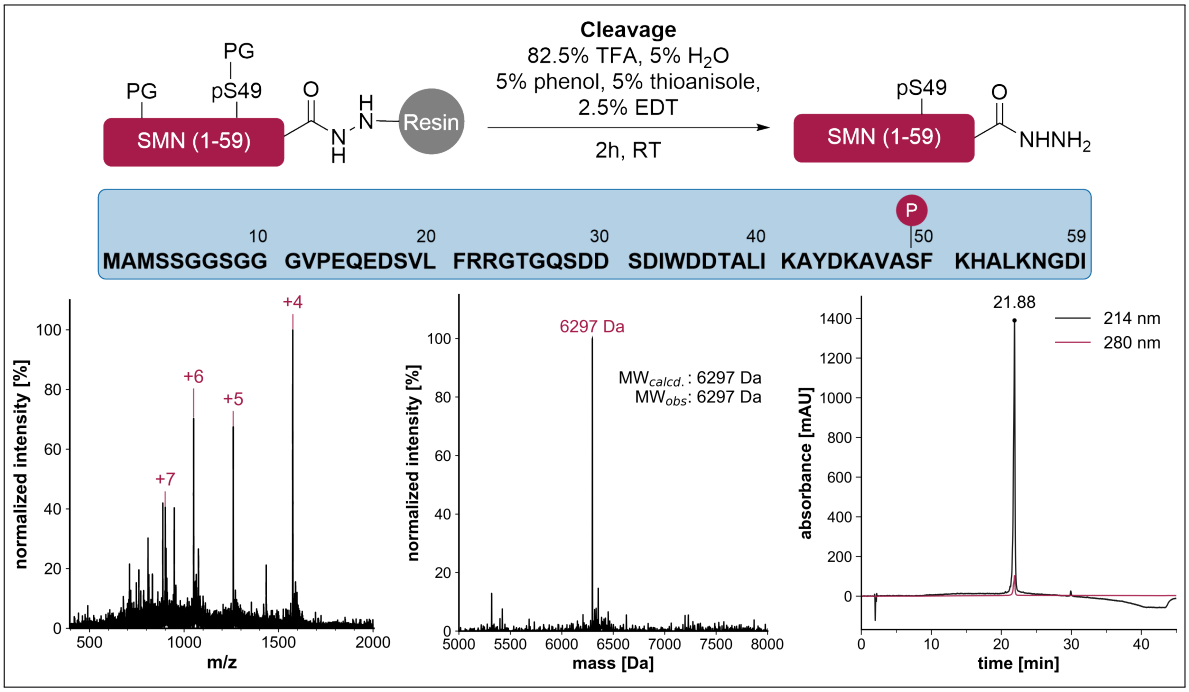

Figure 5.4. Cleavage conditions and analytical data of SMN(1–59)pS49. Top: Cleavage conditions and amino acid sequence of SMN(1–59)pS49. **Bottom:** Analytical data of purified SMN(1–59)pS49-hydrazide. **Left:** m/z spectrum. **Middle:** deconvoluted mass spectrum. **Right:** analytical RP-HPLC UV/Vis chromatogram recorded at 214 nm and 280 nm.

The synthesis of SMN(1–59)pS49 proved less efficient compared to SMN(1–59), despite similar SPPS protocols. Omitting microwave-assisted deprotection and limiting piperazine-based Fmoc removal to only the three residues following pS49, rather than all 48 residues, did not lead to a significant improvement in yield, although it reduced overall synthesis time. LC-MS data of crude products from both syntheses did not indicate β-elimination of the phosphoserine residue. Purification of both synthesis runs proved challenging due to the presence of unidentified byproducts with similar elution behaviour. To evaluate this, LC-MS analysis of the crude was performed (**Figure 5.5**).

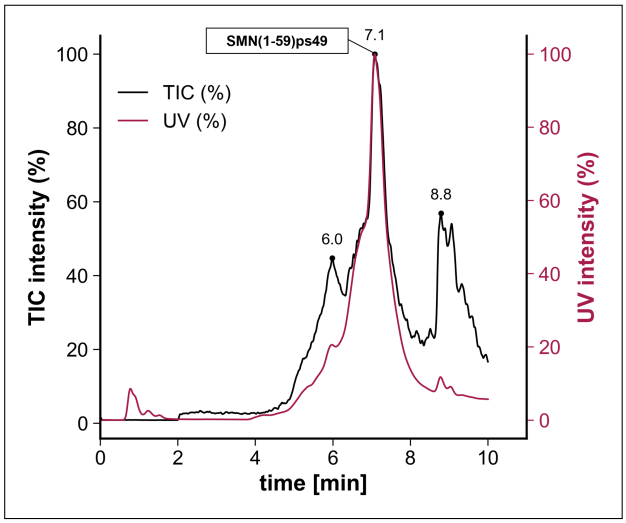

Figure 5.5. LC-MS analysis of crude SMN(1–59)pS49. Total ion count (TIC) and UV intensity of crude SMN(1–59)pS49. The crude was dissolved in a 1:1 (v/v) mixture of ACN and ddH_2O. SMN(1–59)pS49 shows a retention time of 7.1 minutes.

Species with masses of 5657 Da, 5346 Da and 2594 Da were detected during purification; however, further characterisation was unsuccessful. Their identification could provide valuable insights for further optimisation, although fine-tuning of the solvent gradient may also improve the overall yield in a less labour-intensive manner. The overall low yield compared to the unmodified SMN(1–59) peptide is likely attributable to multiple factors. In addition to purification difficulties, the increased steric hindrance of the phosphoserine building block reduces coupling efficiency, while the necessary adjustment of deprotection conditions to prevent β-elimination may also lead to less efficient coupling reactions, thereby promoting the formation of byproducts.

5.2.3 Synthesis of SMN(60–97)

The synthesis of SMN(60–97) was performed on a 2-CTC resin (299.1 mg), prepared according to the general procedure described in **Section 4.2.1**. The loading of the resin was reduced by coupling a mixture of Boc-L-Lys(2ClZ) and Fmoc-L-Lys(Boc)-OH, and the modified loading, determined as described in **Section 4.2.2**, decreased from 1.60 to 0.118 mmol/g. Based on the resin loading and the amount of material prepared, the SPPS scale was limited to 0.04 mmol. Residues C60–A86 and G90 were double-coupled, while the A87–S88 segment was incorporated as a pseudoproline. All other residues were single-coupled. After automated SPPS, cleavage was performed at room temperature for two hours using cleavage cocktail A. The crude peptide was dissolved in a 1:1 (v/v) mixture of ACN and ddH_2O and lyophilized. Purification was performed by preparative RP-HPLC on the Waters system using a preparative C18 column. Purification of collected side pools of SMN(60–97) indicated that C4 columns were equally suitable for this procedure.

SPPS of SMN(60–97) achieved a yield of 27% (40 mg) with high purity confirmed by analytical RP-HPLC and MS. The calculated mass of 4240 Da matched the observed value of 4240 Da (**Figure 5.6**). SMN(60–97) was soluble in aqueous solutions.

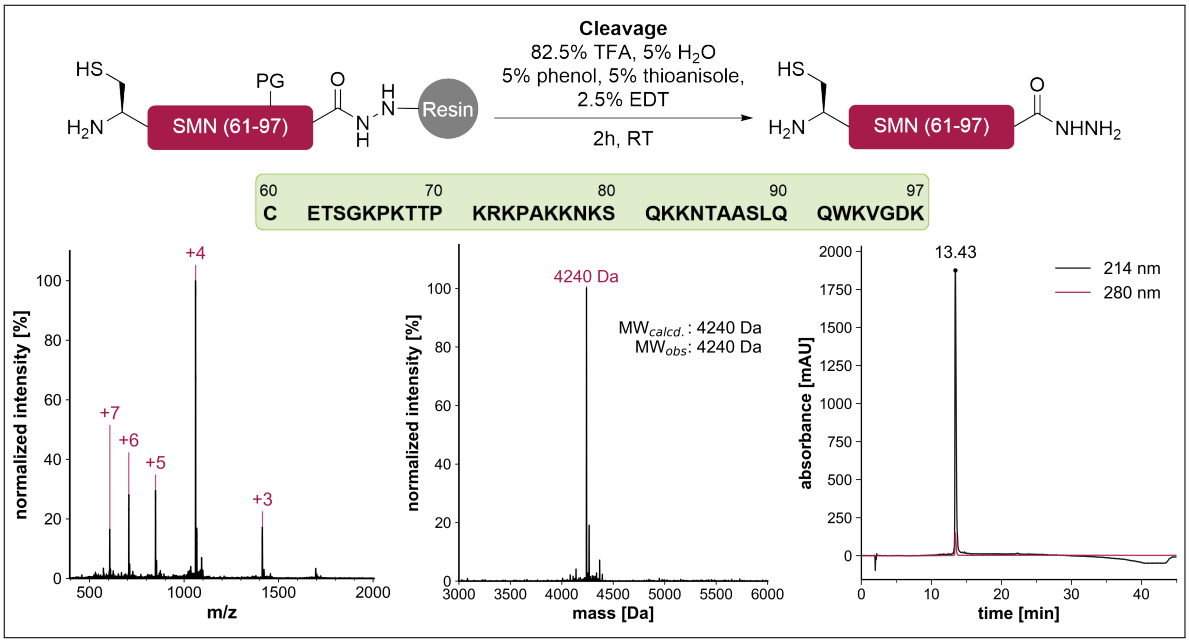

Figure 5.6. Cleavage conditions and analytical data of SMN(60–97). Top: Cleavage conditions and amino acid sequence of SMN(60–97)-hydrazide. **Bottom:** Analytical data of purified SMN(60–97). **Left:** m/z spectrum. **Middle:** deconvoluted mass spectrum. **Right:** analytical RP-HPLC UV/Vis chromatogram recorded at 214 nm and 280 nm.

Initial test cleavages of SMN(60–97) using cleavage cocktail B showed an increased mass of +16 Da, suggesting oxidation. Cleavage cocktail A largely prevented this modification, although it did not completely suppress it. This effect is likely attributable to its higher scavenger content and the broader range of scavengers employed compared with cocktail B. Previous studies have shown that the combined use of different scavengers can effectively minimize side reactions by providing concurrent protection of susceptible residues.[92] During the purification of SMN(60–97), a byproduct was isolated with an observed mass of 4266 Da. The mass increase of +26 Da is hypothesized to result from a reaction between the *C*-terminal cysteine residue of the peptide and acetaldehyde, forming a 2-methyl-1,3-thiazolidine moiety. Although acetaldehyde was not intentionally added, it may be generated through the decomposition of ACN.[93] Treatment of the suspected thiazolidine byproduct with excess methoxyamine at pH 4, followed by a 10-minute incubation at room temperature, removed the modification and confirmed our hypothesis. To further improve synthetic yield, it is therefore recommended to either reduce ACN exposure of SMN(60–97) through shorter purification times or limited ACN use, or to add methoxyamine prior to purification.

5.2.4 Synthesis of SMN(60–97)pS63

SPPS of SMN(60–97)pS63 was performed fully automated on a scale of 0.05 mmol using a Tentagel Hydrazino Protide resin with a loading of 0.22 mmol/g. The SPPS protocol was identical to that used for the unmodified SMN(60–97) peptide; however, microwave-assisted deprotection was omitted for residues 60–63 to avoid enhanced β-elimination of the phosphoserine residue induced by microwave irradiation.[90] In addition, Fmoc deprotection of residues 60–63 was carried out using piperazine, as previously described for the synthesis of SMN(1–59)pS49 (**Section 5.2.2**). Peptide cleavage was performed with cleavage cocktail A at room temperature over two hours. RP-HPLC purification was performed using the Waters system. The analyte was isolated on a preparative C4 column.

The phosphorylated variant SMN(60–97)pS63 achieved a yield of 30% (64 mg) with high purity and a molecular mass of 4321 Da consistent with calculated values (**Figure 5.7**). The peptide was soluble in aqueous solutions.

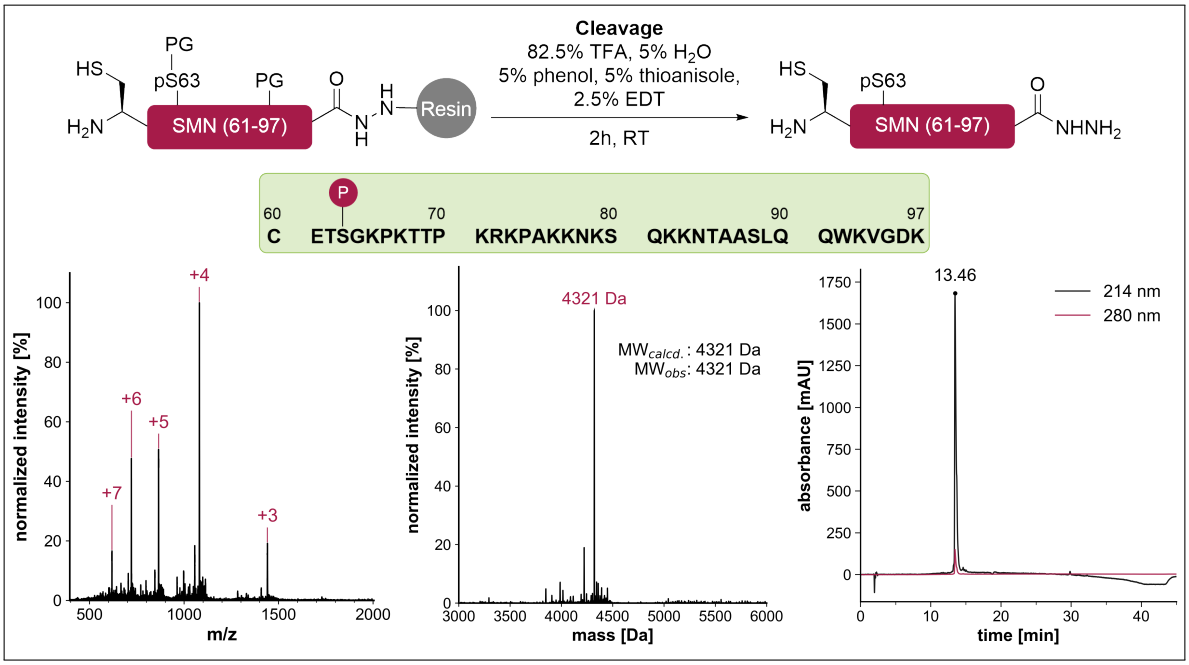

Figure 5.7. Cleavage conditions and analytical data of SMN(60–97)pS63. Top: Cleavage conditions and amino acid sequence of SMN(60–97)pS63. **Bottom:** Analytical data of purified SMN(60–97)pS63-hydrazide. **Left:** m/z spectrum. **Middle:** deconvoluted mass spectrum. **Right:** analytical RP-HPLC UV/Vis chromatogram recorded at 214 nm and 280 nm.

Neither oxidized species nor the thiazolidine byproduct observed during purification of unmodified SMN(60–97) were detected during the production of SMN(60–97)pS63. Incorporation of phosphoserine did not reduce overall peptide yield compared to the synthesis of SMN(1–59)pS49, and no β-elimination was observed. This may be attributed to the late incorporation of pS63, which required minimal changes to the SPPS protocol, affecting only four couplings versus 43 for SMN(1–59)pS49. Overall, the synthesis yielded satisfactory amounts of the peptide, sufficient for the production of all SMN(1—160) variants.

5.2.5 Synthesis of SMN(98–122)

SMN(98–122) was synthesized twice, since the first synthesis afforded satisfactory yields but did not provide sufficient material for the synthesis of all variants. The SPPS protocol was identical for both syntheses, with all residues subjected to double coupling. The D105–G106 segment was incorporated as a dipeptide, and A114–S115 was introduced as a single-coupled pseudoproline. The first synthesis was performed on a modified 2-CTC resin due to availability constraints regarding resins with a hydrazine functionality. The 2-CTC (292.3 mg, loading 1.60 mmol/g, corresponding to a 0.47 mmol scale) was prepared for automated SPPS according to the general resin preparation procedure described in **Section 4.2.1**. The loading was reduced using a mixture of Boc-Thr(Bzl) and Fmoc-Thr(tBu)-OH. After determining the resin loading (**Section 4.2.2**), SPPS was performed using a resin loading of 0.136 mmol/g, giving a synthesis scale of 0.04 mmol. Cleavage was performed at room temperature for two hours using cleavage cocktail A. The crude product was redissolved in a 1:1 (v/v) mixture of ACN and ddH_2O before lyophilization. RP-HPLC purification was performed on the Waters system using a preparative C18 column.

The second synthesis of SMN(98–122) was performed on a preloaded Tentagel Hydrazino Protide resin with a loading of 0.22 mmol/g on a synthesis scale of 0.05 mmol. Peptide cleavage was performed under the same conditions as in the first synthesis. Preparative RP-HPLC was performed on the Waters system using a C4 column. As observed for SMN(60–97), purification of side pools confirmed that C4 columns were equally effective as C18 columns for peptide isolation. Analytical data of the purified product is shown in **Figure 5.8**).

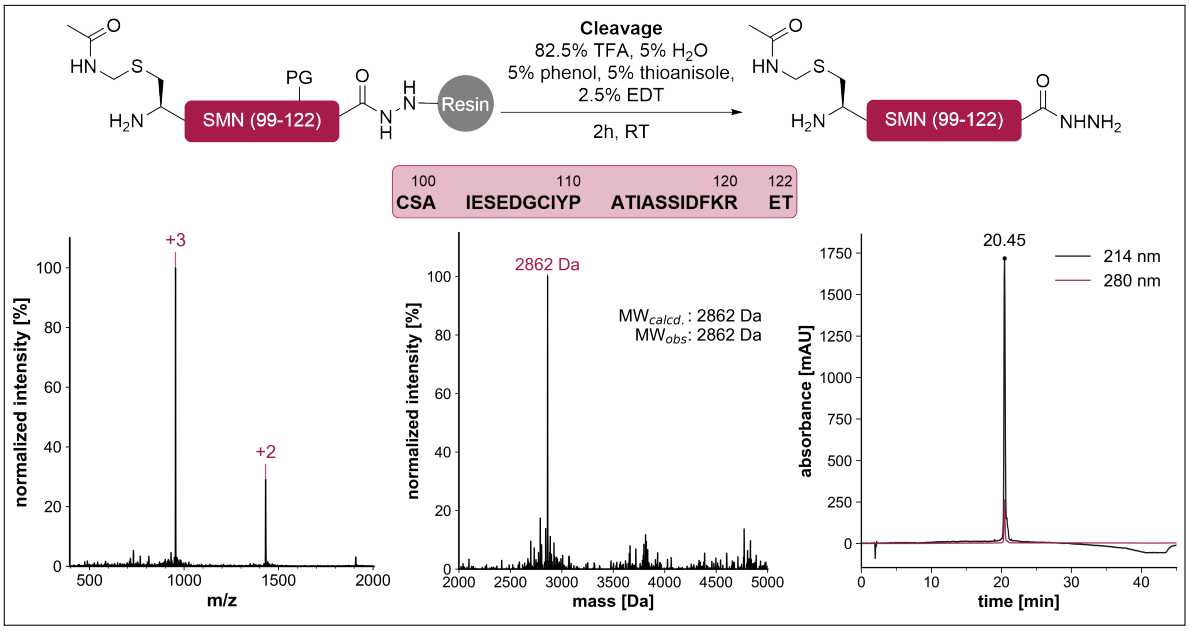

Figure 5.8. Amino acid sequence and analytical data of SMN(98–122). Top: Amino acid sequence and cleavage conditions of SMN(98–122). **Bottom:** Analytical data of peptide SMN(98–122)-hydrazide. **Left:** m/z spectrum. **Middle:** deconvoluted mass spectrum. **Right:** analytical RP-HPLC UV/Vis chromatogram recorded at 214 nm and 280 nm.

The 2-CTC resin yielded 31% (53 mg), while the Protide resin gave 22% (46 mg). Both syntheses produced peptides of high purity, as confirmed by RP-HPLC and MS. The measured mass of 2862 Da matched the calculated value. SMN(98–122) is soluble in ddH_2O and was found to be easier to purify than SMN(98–145), the peptide used in the previous segmentation strategy.

Both synthetic approaches yielded sufficient amounts of SMN(98–122). Although the 2-CTC resin afforded a slightly higher yield, the additional time and effort required for resin preparation makes the use of a commercially available, preloaded resin more attractive. The shift of ligation sites between peptides 3 and 4, which decreased the length of peptide 3 by 24 amino acids, proved beneficial for the synthesis, improving the overall synthetic outcome and simplifying purification procedures due to easier handling of the peptide.

5.2.6 Synthesis of SMN(123–160)

SMN(123–160) exhibited low solubility, which resulted in challenging purification and reduced overall yields, necessitating three separate syntheses to obtain sufficient material for all SMN(1–160) variants. All syntheses were performed using a preloaded Wang Glu-LL-resin at a synthesis scale of either 0.05 mmol or 0.10 mmol. Details of the SPPS protocol are listed in **Table 5.2**. Cleavage was performed using cleavage cocktail B at room temperature with an incubation time of two hours. The collected precipitate after cleavage was redissolved in a 1:1 (v/v) solution of ACN and ddH_2O prior to lyophilization.

Table 5.2. SPPS protocol of SMN(123–160)

SMN(123–160)	
Coupling Conditions	**Amino Acids**
Single coupling	D140, N151, I152, N155, A156, N159
Double coupling	C123 to N137, L141 to N150, E153, Q154, Q157, E158
Pseudoprolin single coupling	L138-139
Capping step	V124, V125, Y127, Y130, E147

Further purification procedures required modification compared to other peptides due to the low solubility of SMN(123–160). Crude product from a 0.10 mmol synthesis scale had to be split in half for proper purification, whereas product from a 0.05 mmol synthesis could be purified in a single run. Two approaches were found to be effective for purification. The lyophilized peptide was either dissolved in large volumes of 6 M Gnd · HCl (30–50 mL) or the solution pH was adjusted to 8 to enhance product solubility and enable the dissolution in smaller volumes (>10 mL). However, the product could not be fully dissolved with either method. The peptide solution was then centrifuged (15 min, 4000 rpm) and filtered (0.2 µm filter). The remaining precipitate was washed again with 6 M Gnd · HCl, centrifuged (5 min, 4000 rpm), filtered, and combined with the remaining peptide solution. Purification was performed on either the Shimadzu or the Waters system using a preparative C4 column. The Shimadzu system was used if the lyophilized product was dissolved in large volumes of 6 M Gnd · HCl, as no injection limit had to be considered. The Waters system allowed injection

volumes of up to 10 mL only, which would have required multiple injections for larger volumes. However, given the already difficult purification procedure, multiple injections were not considered feasible.

Two SPPS were conducted on a 0.05 mmol scale, yielding 8% (17 mg) and 6% (12 mg). The 0.10 mmol scale syntheses achieved a total of 7% (28 mg). Taking all syntheses into account, the average yield is 7% (14 mg), with respect to a synthesis scale of 0.05 mmol. All isolated products were of high purity as determined by RP-HPLC and MS. The calculated mass of 4256 Da was in agreement with the determined mass of 4256 Da (**Figure 5.9**). Purified SMN(123–160) was soluble in 6 M Gnd · HCl, but insoluble in ddH_2O.

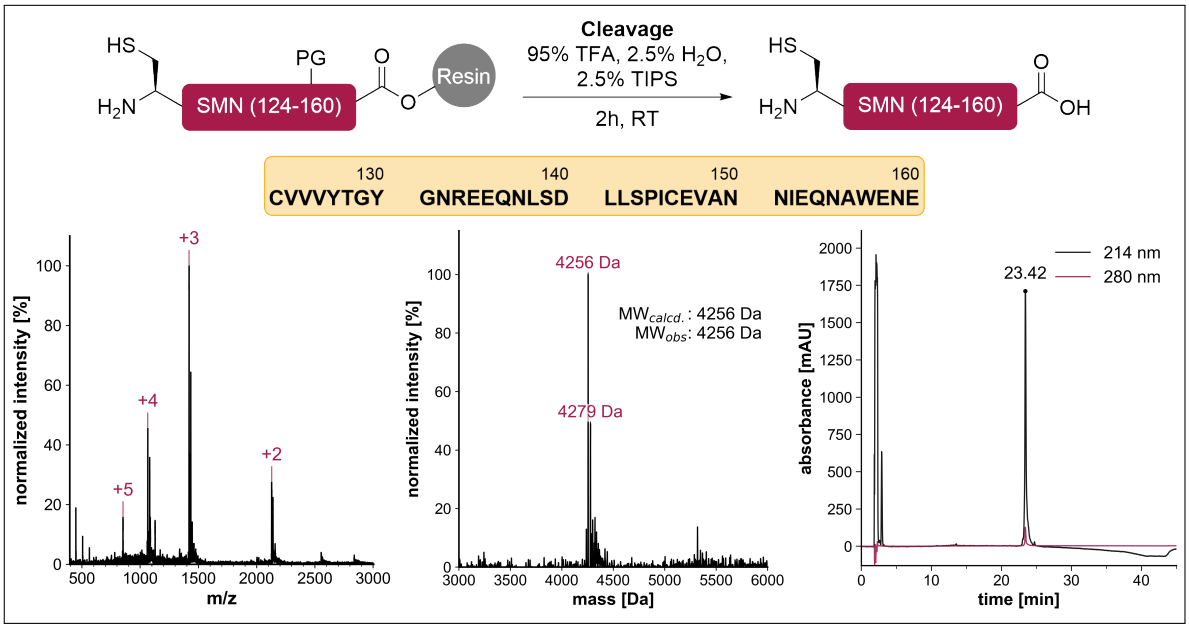

Figure 5.9. Cleavage conditions and analytical data of SMN(123–160). Top: Cleavage conditions and amino acid sequence of SMN(123–160). **Bottom:** Analytical data of purified SMN(123–160). **Left:** m/z spectrum. **Middle:** deconvoluted mass spectrum; the 4279 Da peak corresponds to a Na^+ adduct (+22 Da). **Right:** analytical RP-HPLC UV/Vis chromatogram measured at 214 nm and 280 nm.

Among all peptides, SMN(123–160) consistently showed the lowest yields, likely due to challenges associated with both the synthesis and purification process. Aggregation of hydrophobic residues during SPPS may hinder efficient coupling reactions, reducing overall yields and promoting the formation of side products.[94] These impurities further complicate purification, especially given the low solubility of all compounds. Although both described isolation strategies yielded comparable results, the overall purification efficiency remained suboptimal. While optimization of the RP-HPLC solvent system may improve performance, the high abundance of synthesis-derived byproducts appears to represent the more substantial limitation as evaluated by analytical RP-HPLC of the crude peptide dissolved in DMF (**Figure 5.10**). Purification of SMN(123–160) revealed two prominent byproducts with mass differences of approximately +55 Da and +85 Da, which were not detected during initial test cleavages leading to their late discovery. Both species exhibited very low solubility in 6 M Gnd · HCl. The +55 Da species is hypothesized to correspond to SMN(123–160) carrying an attached *tert*-butyl group, whereas the identity of the +85 Da compound remains unclear. Re-exposure to cleavage cocktail B confirmed that both com-

pounds are stable and do not contain residual protecting groups. The +55 Da byproduct likely forms during peptide cleavage due to insufficient scavengers, causing irreversible reattachment of protecting groups. Its absence during test cleavages may be explained by poor solubility and removal during centrifugation prior to MS analysis. Adjusting cleavage conditions could therefore reduce this byproduct. However, given the overall low yields, a broader review of the SPPS protocol may be necessary to achieve substantial improvements.

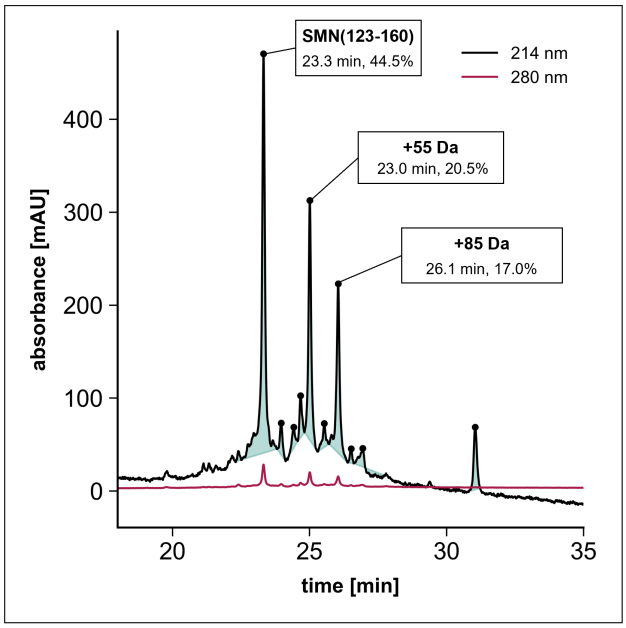

Figure 5.10. Analytical RP-HPLC UV/Vis chromatogram of crude SMN(123–160). The crude was dissolved in DMF and measured at 214 nm and 280 nm. Compound retention times (RT): SMN(123–160) 23.3 min, +55 Da species 25.0 min, +85 Da species 26.05 min. Remaining peaks were not characterized further.

SMN(123–160) gained an additional 24 residues due to the shift of the ligation site between peptides 3 and 4. While the previous peptide 4, SMN(146–160), was highly soluble and easy to purify, SMN(123–160) exhibited the expected decrease in solubility. Although not ideal, peptides SMN(98–122) and SMN(123–160) showed overall improved properties compared to their predecessors, and their properties are adequate for the subsequent generation of the SMN(1–160) variants.

5.3 Thioester conversion

Following the successful synthesis and isolation of all peptide segments in good purity, thioester conversions were performed to convert SMN(1–59), SMN(1–59)pS49 and SMN(98–122) into their respective MESNa thioesters, enabling subsequent ligation with their neighbouring peptide segments. Results for each conversion are provided in the following sections.

5.3.1 MESNa thioester conversion of SMN(1–59) and SMN(1–59)pS49

The thioester conversions of the SMN(1–59)-hydrazide were performed on a scale of 9.6 µmol (59.5 mg). The reaction was stirred at -20°C for 20 minutes after which 22.0 equivalents of MESNa were added. The conversion of SMN(1–59)pS49 was performed twice due to the low yield obtained during the initial SPPS of SMN(1–59)pS49. The first conversion was carried out using 2.8 µmol (17.7 mg) of SMN(1–59)pS49, and the second using 2.9 µmol (18.3 mg). In both cases about 30.0 equivalents of MESNa were added. The exact excess of MESNa did not appear to influence the conversion efficiency.

The MESNa thioester conversion of SMN(1–59) yielded 64%. The calculated molecular weight of the product was 6328 Da, in good agreement with the observed value of 6327 Da. SMN(1–59)pS49-MESNa was synthesized twice. The first conversion yielded 21%, whereas the second achieved a yield of 50%. The observed molecular weights of 6408 Da were in good agreement with the calculated mass of 6407 Da. All products were obtained in high purity, as assessed by RP-HPLC and MS (**Figure 5.11**).

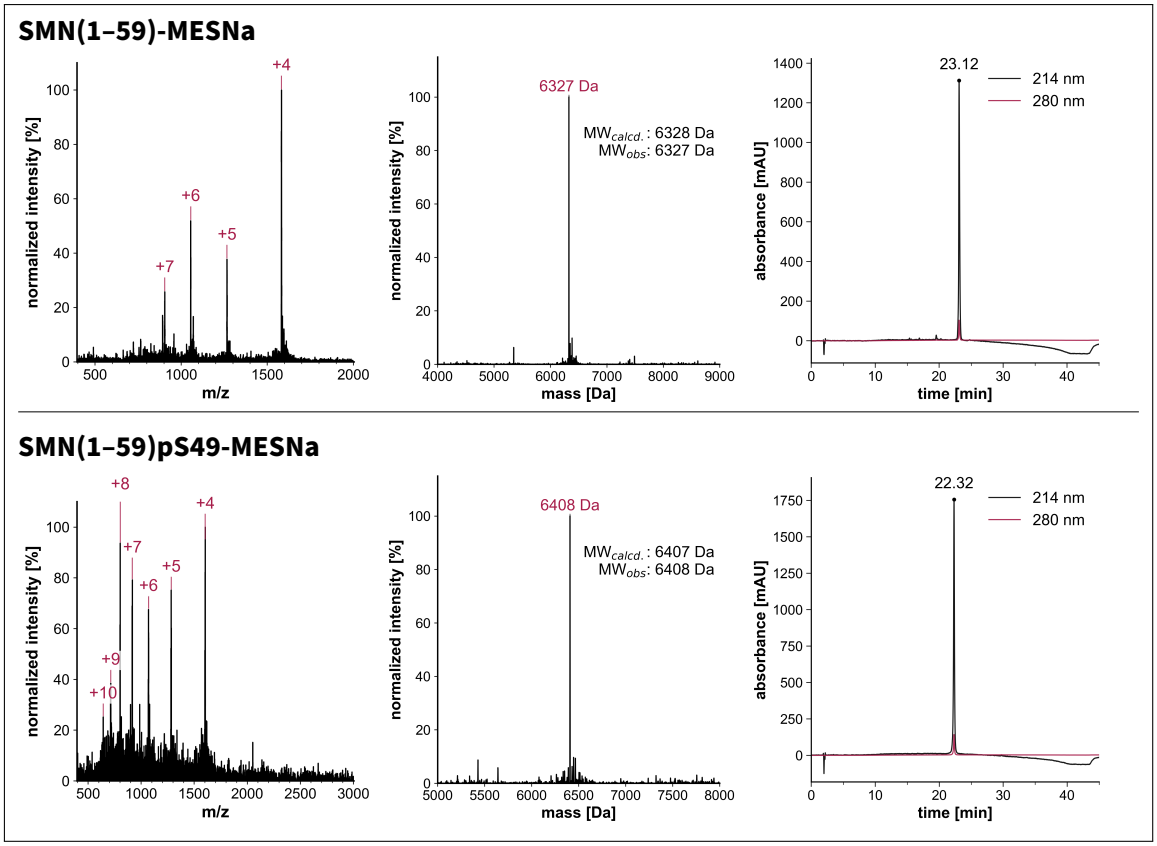

Figure 5.11. Analytical data of the SMN(1–59)- and the SMN(1–59)pS49-MESNa thioesters. Left: m/z spectra of the products. **Middle:** processed mass spectra. **Right:** analytical RP-HPLC UV/Vis chromatograms recorded at 214 nm and 280 nm.

The MESNa conversions of SMN(1–59) and SMN(1–59)pS49 proceeded to completion, with no starting material detectable by LC-MS prior to purification. The reduced yields are therefore likely due to the purification procedure itself. The particularly low yield of the first SMN(1–59)pS49-MESNa synthesis resulted from an HPLC error that caused product loss. With complete conversion of both

SMN(1–59)-MESNa and SMN(1–59)pS49-MESNa, further optimization of the purification method can be envisaged. The applied 60 min solvent gradient could be shortened to reduce solvent consumption while maintaining product purity.

SMN(1–59) thioester conversion was also attempted using MPAA, followed by ligation with SMN(60–97) in a one-pot manner. Adjusting the pH to 6.60 upon MPAA addition caused significant hydrolysis. While lowering the pH of the conversion and careful adjustment with less concentrated NaOH may reduce hydrolysis, the absence of detectable hydrolysis with MESNa made it the preferred choice.

5.3.2 MESNa thioester conversion of SMN(98–122)

Thioester conversions of the SMN(98–122)-hydrazide were conducted on scales of 1.9 µmol (5.5 mg), 5.3 µmol (15.3 mg), 8.5 µmol (24.2 mg) and 10.9 µmol (31.2 mg), depending on peptide availability and the required amount of thioester. Initial optimization of the reaction conditions revealed that an extended oxidation time was necessary following the addition of $NaNO_2$. The oxidation time was therefore increased to 30 minutes. Approximately 30.0 equivalents of MESNa were used for each conversion.

The thioester conversions of SMN(98–122) achieved an average yield of 68%. Product purity was assessed by RP-HPLC upon treatment with TCEP to reduce any potential disulfides, in combination with MS analysis. The product exhibited high purity. The calculated mass of the product was 2972 Da, matching the measured value. Analytical data is shown in **Figure 5.12**.

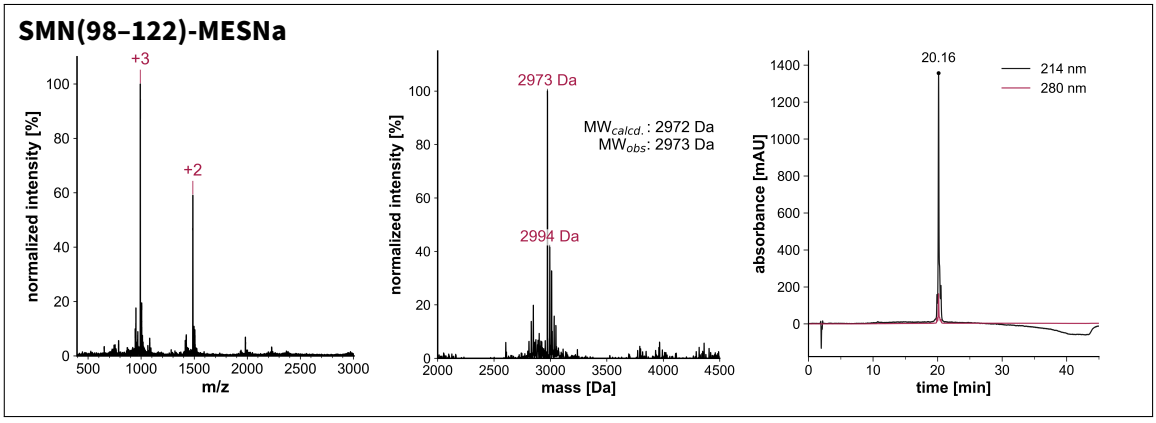

Figure 5.12. Analytical data of the SMN(98–122)-MESNa thioester. Left: m/z spectrum. **Middle:** deconvoluted mass spectrum, the 2994 Da peak corresponds to a Na^+ adduct (+22 Da). **Right:** analytical RP-HPLC UV/Vis chromatogram.

Two byproducts with masses of 3112 Da and 2830 Da were detected both prior and during purification of the reaction. The 3112 Da species corresponds to a MESNa-disulfide formed with the internal cysteine residue at position 107 and could be readily removed upon addition of TCEP. The 2830 Da compound is hypothesized to be a thiolactone between cysteine 107 and the *C*-terminal carbonyl group, consistent with its expected mass, however, isolation of this species was unsuccessful. Increasing the MESNa excess was insufficient to prevent its formation, which is unsurprising given the

intramolecular nature of the proposed reaction. The presumed thiolactone was also observed during the ligation of SMN(98–122) and SMN(123–160), discussed in **Section 5.4.2**. Incorporation of an Acm protecting group at cysteine 107 could both prevent thiolactone formation and help confirm the identity of the compound, while no additional synthetic steps would be necessary. However, if the thiolactone reacts with SMN(123–160) to form SMN(98–160) in a timely manner, or is only formed in low amounts, efforts to prevent its formation may be unnecessary.

A one-pot strategy was also investigated in which SMN(98–122) was first converted to the corresponding thioester and then directly ligated to SMN(123–160) in the same reaction vessel using MPAA, without intermediate purification. As observed previously for the MPAA-mediated thioester conversion and ligation of SMN(1–59) and SMN(60–97), a high degree of hydrolysis was detected upon adjusting the pH to 6.60. Therefore, MESNa was selected as the thioester of choice, and no further one-pot experiments were performed.

5.4 First ligation reactions - Synthesis of SMN(1–97) variants and SMN(98–160)

Thioesters SMN(1–59)-MESNa, SMN(1–59)pS49-MESNa and SMN(98–122)-MESNa were subsequently used to generate the intermediate segments SMN(1–97), the SMN(1–97) phospho-variants as well as SMN(98–160) *via* NCL reactions. Results of the ligations are provided in the following section.

5.4.1 Synthesis of SMN(1–97) variants

A small-scale ligation was performed to evaluate suitable ligation conditions and to monitor the progress of the reaction. SMN(1–59)-MESNa (1.0 eq.) and SMN(60–97) (1.0 eq.) were dissolved in ligation buffer to a total concentration of 5.0 mM. MPAA (200 mM) and TCEP (160 mM) were added, and the pH was adjusted to 6.80. The reaction was monitored over 96 hours using analytical RP-HPLC and LC-MS analysis (**Figure 5.13**). SMN(1–59)-MESNa and hydrolysed SMN(1–59) are reported jointly, as no clear separation was achieved.

The initial measurement at 0 hours showed two dominant peaks corresponding to SMN(60–97) (13.6 min, 34.1%) and SMN(1–59)-MESNa/hydrolysis (23.2 min, 61.9%). SMN(1–59)-MPAA was also detected (24.3 min, 4.0%). After 22 hours, formation of the ligation product SMN(1–97) became apparent (21.4 min, 55.7%), accompanied by a decrease in SMN(60–97) (13.7 min, 13.5%) and SMN(1–59)-MESNa/hydrolysis (23.1 min, 13.0%). In parallel, SMN(1–59)-MPAA increased (24.3, 17.8%). At 46 hours, SMN(1–97) further increased (21.5 min, 62.5%), while SMN(60–97) (13.8 min, 11.5%), SMN(1–59)-MESNa/hydrolysis (23.0 min, 10.6%), and SMN(1–59)-MPAA (24.3 min, 15.5%) decreased. Additional SMN(60–97) was added at this time point to promote ligation. The 95 hour sample showed a further increase in SMN(1–97) (21.4 min, 68.3%), alongside an increase in SMN(60–97) (13.7 min, 19.1%). No SMN(1–59)-MPAA was detected, whereas SMN(1–59)-MESNa/hydrolysis (23 min, 12.6%) showed a slight increase.

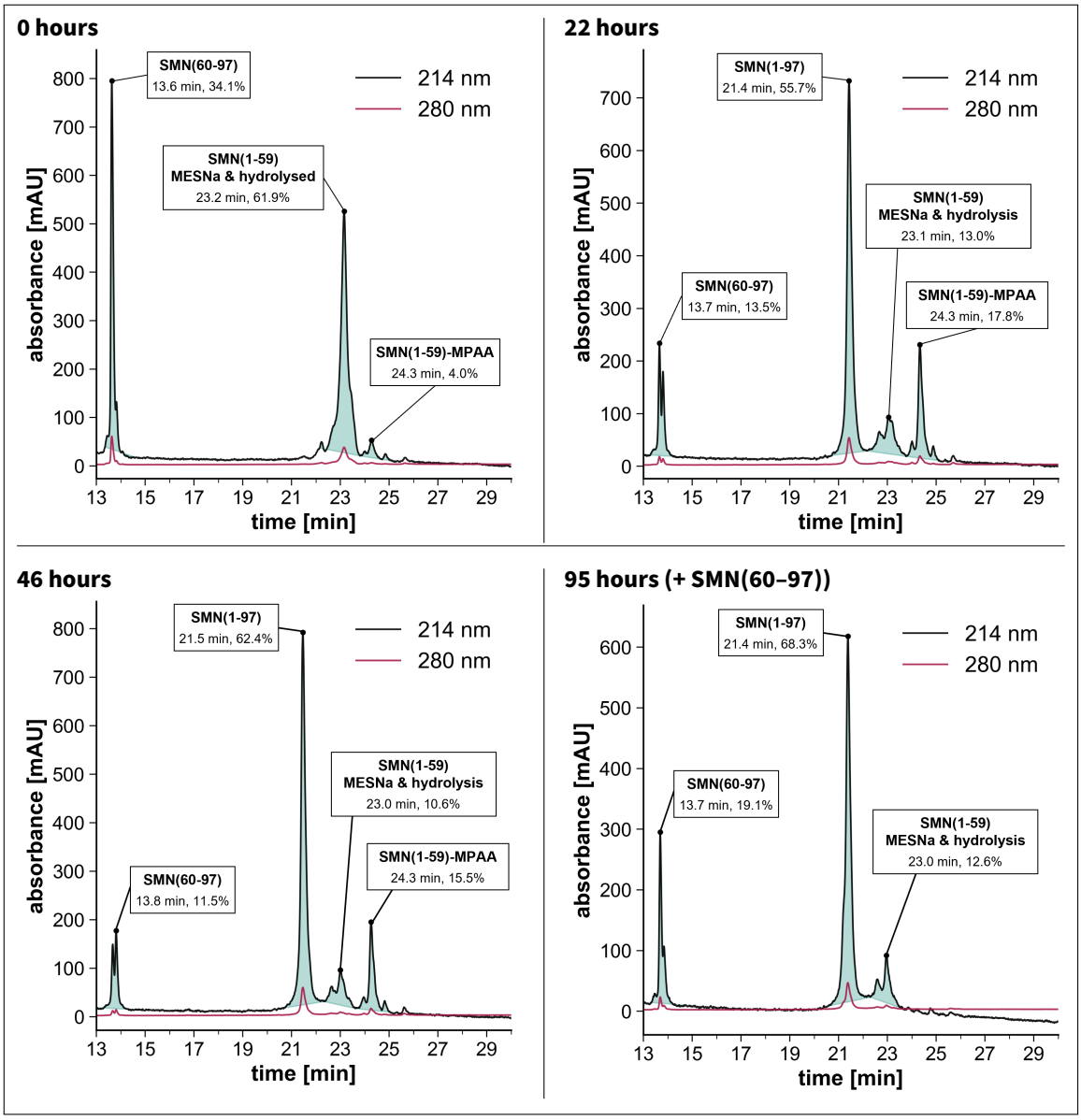

Figure 5.13. Monitoring of the NCL reaction between the SMN(1–59)-MESNa thioester and SMN(60–97). Average compound retention times (RT): SMN(60–97) 13.7 min, SMN(1–97) 21.4 min, SMN(1–59)-MESNa/hydrolysis 23.1 min, SMN(1–59)-MPAA 24.3 min. Additional SMN(60–97) was added after 46 hours. No SMN(1-95)-MPAA was detected at 95 hours.

Reaction monitoring showed that most of the ligation progress occurred within the first 24 hours, reaching a conversion of approximately 55.7%. Prolonged reaction times led to only a modest increase to around 62.4%. After addition of SMN(60–97) at 46 hours, the conversion further increased to 68.3% at 95 h. However, it remains unclear whether this increase resulted from the additional SMN(60–97) or from the additional reaction time. Additional measurements would have been required to determine whether the conversion had already reached a plateau. Nevertheless, an excess of SMN(60–97) was not considered disadvantageous. Hydrolysis was not identified as a significant side process, as no notable accumulation was observed during monitoring. Based on the reaction progress, ligation times exceeding 46 hours were considered unnecessary. Given the modest increase between 24 and

46 hours and accounting for measurement uncertainties, a ligation time of 24 hours is also expected to be sufficient without significant loss of product. The promising outcome of this initial testing led to similar conditions being adopted for the synthesis of all SMN(1–97) variants.

For the unmodified variant, SMN(1–59)-MESNa (2.4 µmol, 15.1 mg, 1.0 eq.) and SMN(60–97) (2.2 µmol, 9.4 mg, 1.0 eq.) were dissolved in ligation buffer to a total peptide concentration of 5.0 mM. MPAA (200 mM) and TCEP (160 mM) were added, and the pH was adjusted to 6.80. The reaction was stirred at room temperature for 44 hours before purification.

Due to the more challenging synthesis and lower yield of SMN(1–59)pS49 and its corresponding MESNa thioester, the ligation conditions for the phospho-variants were slightly adjusted to maximize the efficiency of peptide utilization. For the synthesis of SMN(1–97)pS49, an excess of SMN(60–97) was used to improve the conversion rate of the ligation. The total peptide concentration was also increased to further promote ligation efficiency. SMN(1–59)pS49-MESNa (0.7 µmol, 4.4 mg, 1.0 eq.) and SMN(60–97) (1.0 µmol, 4.3 mg, 1.5 eq.) were dissolved in ligation buffer to a total concentration of 7.4 mM. MPAA (210 mM) and TCEP (160 mM) were added to the reaction and the pH was adjusted to 6.80. The reaction was stirred for 44 hours before purification.

The synthesis of SMN(1–97)pS63 also used a small excess of SMN(60–97)pS63 to improve ligation efficiency and to take advantage of the high availability of this peptide. SMN(1–59)-MESNa (1.4 µmol, 8.8 mg, 1.0 eq.) and SMN(60–97)pS63 (1.6 µmol, 7.1 mg, 1.2 eq.) were dissolved in ligation buffer to a total concentration of 6.0 mM. MPAA (210 mM) and TCEP (160 mM) were added and the pH was adjusted to 6.80. The reaction was stirred at room temperature for 44 hours before purification.

The diphosphorylated variant SMN(1–97)pS49pS63 was synthesized with a small excess of SMN(60–97)pS63, as similar adjustments in previous ligations of the monophosphorylated variants did not significantly improve the ligation outcome. SMN(1–59)pS49-MESNa (1.4 µmol, 8.7 mg, 1.0 eq.) and SMN(60–97)pS63 (1.5 µmol, 6.3 mg, 1.1 eq.) were dissolved in ligation buffer to a total concentration of 6.0 mM. MPAA (210 mM) and TCEP (180 mM) were added and the pH was adjusted to 6.80. The reaction was purified after 27 hours.

SMN(1–97) yielded 47%, while the monophosphorylated variant SMN(1–97)pS49 was obtained in 52% yield using an excess of SMN(60–97) and a higher total peptide concentration. SMN(1–97)pS63 was synthesized at 34% yield using a smaller excess of SMN(60–97)pS63, at a similar peptide concentration. SMN(1–97)pS49pS63 achieved a yield of 45%, with a reduced reaction time compared to the other variants. All peptides showed high purity by RP-HPLC and MS, with observed masses in good agreement with calculated values (**Figure 5.14**). For SMN(1–97), the observed mass of 10425 Da closely matched the calculated mass of 10426 Da. SMN(1–97)pS49 displayed a deviation of +3 Da, with an observed mass of 10509 Da compared to the calculated 10506 Da. The observed mass of SMN(1–97)pS63 matched the calculated value of 10506 Da. A discrepancy of +6 Da was observed for SMN(1–97)pS49pS63, likely due to calibration issues.

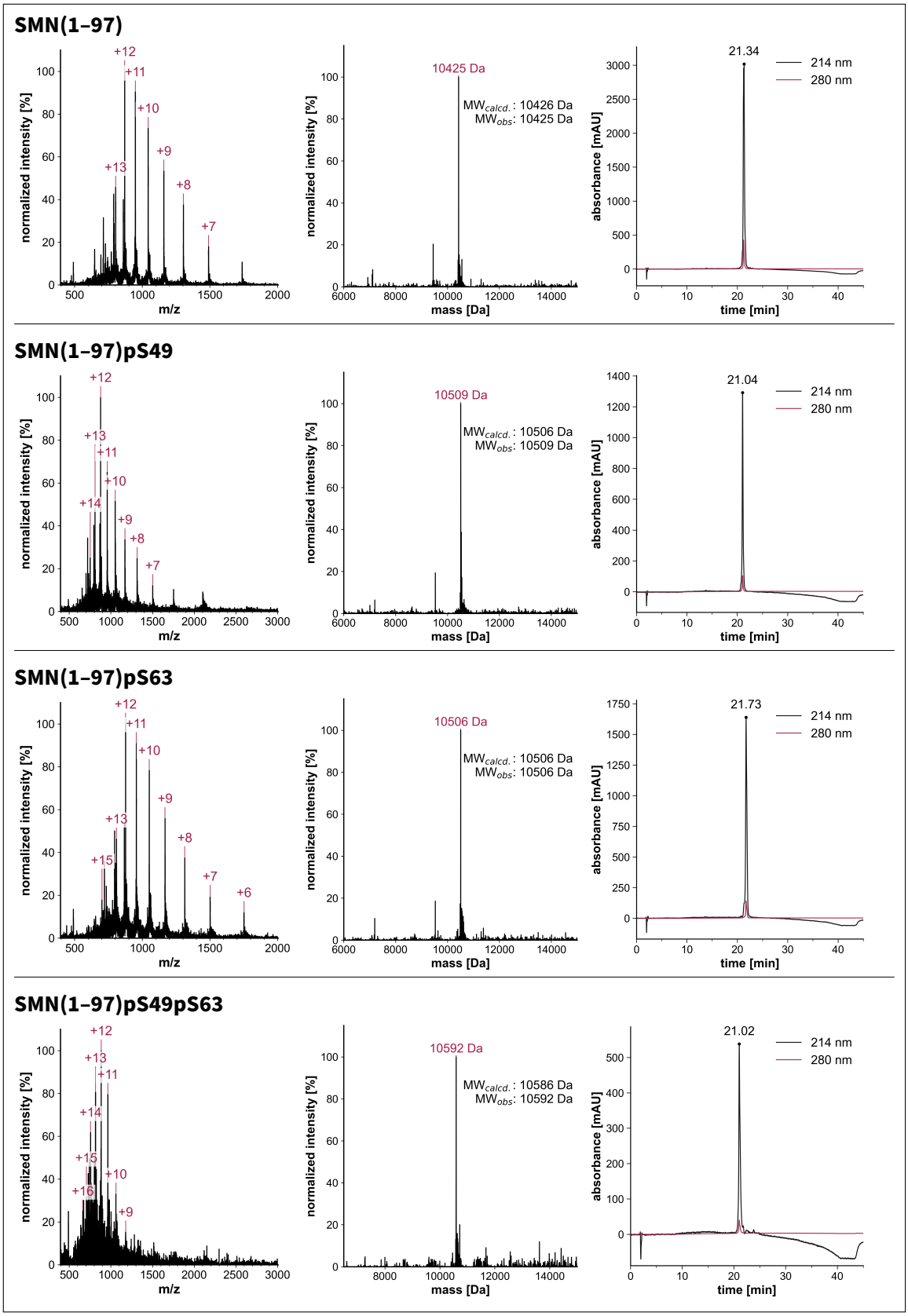

Figure 5.14. Analytical data of SMN(1–97) ligation products. Left: m/z spectra. **Middle:** deconvoluted mass spectra. **Right:** analytical RP-HPLC UV/Vis chromatograms recorded at 214 nm and 280 nm.

Except for SMN(1–97)pS63, all variants were obtained in comparable yields of about 48%. RP-HPLC peak area analysis of SMN(1–97)pS63 indicated a conversion rate of approximately 57.3%. The reduced yield is therefore likely attributable to losses during purification. However, no specific source of error could be identified. The use of an excess of SMN(60–97) or its phosphorylated analogue SMN(60–97)pS63 did not significantly affect the overall reaction outcome. Despite a shortened reaction time employed for SMN(1–97)pS49pS63, its yield was comparable to those of the other variants, supporting conclusions drawn from initial reaction monitoring data. Future syntheses are therefore expected to yield similar results with a reaction time of about 24 hours.

The ligation of SMN(1–59) with SMN(60–97), as well as with their respective phosphorylated variants, generally proceeded efficiently, exhibiting good ligation rates and minimal thioester hydrolysis. Alternative ligation strategies were also investigated. A one-pot approach was tested, combining the thioester conversion of SMN(1–59) with its subsequent ligation to SMN(60–97) without prior purification of the thioester. Both MESNa and MPAA thioesters of SMN(1–59) were tested. However, the method was unsuccessful due to substantial SMN(1–59) hydrolysis in both MESNa and MPAA trials. Combined with lower ligation efficiency compared to the method described herein, one-pot approaches were not pursued further. These findings further highlight the importance of thorough peptide purification prior to ligation.

The *C*-terminal hydrazides of the SMN(1–97) variants were subsequently converted into thioesters to enable ligation with Acm-deprotected SMN(98–160) (see **Section 5.5.1**). Prior to the second ligation step, the SMN(98–160) segment was synthesized and the Acm group removed.

5.4.2 Synthesis of SMN(98–160)

A test ligation for the synthesis of SMN(98–160) was conducted and monitored by RP-HPLC and LC-MS over 43 hours (**Figure 5.15**). SMN(98–122)-MESNa (1.2 eq.) and SMN(123–160) (1.0 eq.) were dissolved in ligation buffer to a total peptide concentration of 5.0 mM. MPAA (250 mM) and TCEP (100 mM) were added, and the pH was adjusted to 6.80. SMN(98–122)-MESNa and hydrolysed SMN(98–122) are reported jointly due to their similar elution behaviour.

At 0 h, SMN(98–122)-MESNa/hydrolysis was detected (20.7 min, 36.9%), together with SMN(98–122)-MPAA (21.9 min, 15.0%) and a putative SMN(98–122) thiolactone (21.3 min, 2.6%). SMN(123–160) was identified (23.4 min, 31.6%) together with the ligation product SMN(98–160) (24.7 min, 11.4%). A persistent impurity was detected throughout all measurements (25.7 min, 2.2%). After 2 hours, SMN(98–160) increased (24.7 min, 42.3%), while SMN(98–122)-MESNa/hydrolysis (20.8 min, 26.9%), SMN(98–122)-MPAA (21.9 min, 10.2%) and SMN(123–160) (23.4 min, 14.2%) decreased. The thiolactone remained at comparable levels (21.3 min, 3.2%). After 25 hours, SMN(98–160) further increased (24.7 min, 56.6%), accompanied by a decrease in SMN(98–122)-MESNa/hydrolysis (21.0 min, 25.8%), SMN(98–122)-MPAA (21.9 min, 4.6%), and SMN(123–160) (23.5 min, 6.9%). SMN(98-123) thiolactone remained stable (21.3 min, 3.5%). The reaction profile remained unchanged at 43 hours relative to 25 hours, with all species at comparable levels.

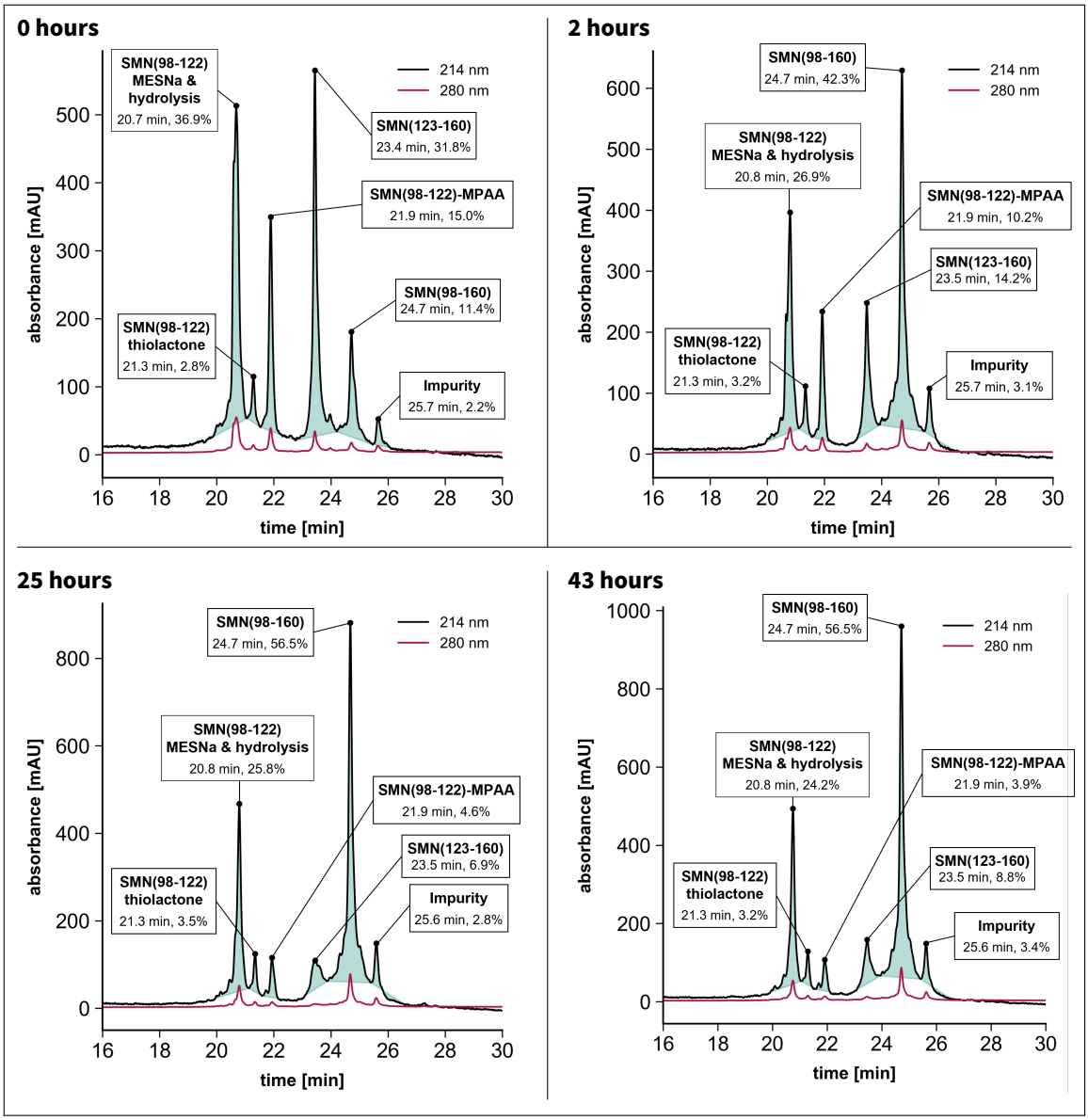

Figure 5.15. Monitoring of the NCL reaction between the SMN(98–122)-MESNa thioester and SMN(123–160). Average component retention times (RT): SMN(98–122)-MESNa and hydrolysed SMN(98–122) 20.8 min, SMN(98–122)-thiolactone 21.3 min, SMN(98–122)-MPAA 21.9 min, SMN(123–160) 23.5 min, SMN(98–160) 24.7 min. An impurity is present in all chromatogramms at 25.7 min.

A high ligation activity is observed within the first minutes. After 2 hours, the conversion reaches approximately 42.3%, with a further increase to the observed maximum of 56.6% by the 25 hours mark. This implies that a reaction time of a few hours may be sufficient to achieve significant conversion. However, additional measurements at intermediate time points (e.g., 4, 8, and 12 hours) are required to determine when the reaction reaches a plateau and to optimize the reaction time for balancing efficiency and yield. No significant increase in product formation was observed beyond 25 hours, indicating no substantial improvement from prolonged reaction time. While increasing reactant concentration may further enhance the conversion rate, the solubility of SMN(123—160) imposes a practical limitation.

Evaluating SMN(98–122) hydrolysis proved challenging due to its similar elution behaviour to SMN(98–122)-MESNa. Although hydrolysis is occurring, the observed ligation conversion of up to 56.5% was deemed sufficient to yield an adequate amount of the desired ligation product in large-scale reactions. The presumed SMN(98–122)-thiolactone, previously mentioned in the context of the thioester conversion of SMN(98–122) in **Section 5.3.2**, was also detected during the ligation. No significant changes in peak area were observed during reaction monitoring, suggesting that the thiolactone is consumed and formed at similar rates. Given its low proportion relative to the total amount of SMN(98–122) thioesters, its formation is not expected to interfere with the ligation.

Following the successful test reaction, its conditions were used for the large scale production of SMN(98–160). The reaction was performed using an excess of SMN(98–122)-MESNa thioester to compensate for thioester hydrolysis and to enhance the ligation rate. SMN(98–122)-MESNa thioester (5.6 µmol, 16.7 mg, 1.2 eq.) and SMN(123–160) (5.0 µmol, 21.1 mg, 1.0 eq.) were dissolved in ligation buffer to a total peptide concentration of 5.0 mM. MPAA (250 mM) and TCEP (100 mM) were added, and the pH was adjusted to 6.80. The reaction was left stirring for 48 hours before purification, due to time management reasons. The ligation was repeated four times at a similar scale to produce sufficient material for the synthesis of all SMN(1–160) variants.

The synthesis of SMN(98–160) achieved a yield of up to 73%, with an average yield of 65%. The observed molecular weight of 7086 Da was in good agreement with the calculated mass of 7086 Da. RP-HPLC and MS confirmed a high purity of the product. While SMN(98–160) is poorly soluble in ddH_2O, it dissolved readily in 6 M Gnd · HCl buffer. Analytical data of the purified product is provided in **Figure 5.16**.

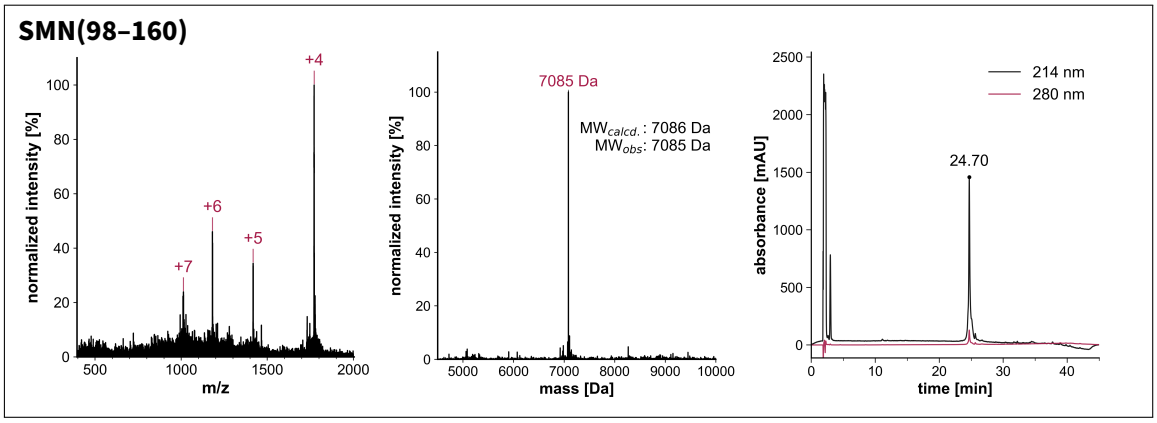

Figure 5.16. Analytical data of Acm-protected SMN(98–160). Left: m/z spectrum. **Middle:** deconvoluted mass spectrum. **Right:** analytical RP-HPLC UV/Vis chromatogram recorded at 214 nm and 280 nm.

The low solubility of SMN(123–160) did not cause problems during the reaction and the purification process. However, increasing the total peptide concentration above 5.0 mM is not recommended. Concentrations of 6.0 mM resulted in the precipitation of compounds, increasing the difficulty of product isolation. Similar to the production of SMN(1–97), one-pot ligations were also attempted for SMN(98–160). Testing was performed using MPAA and MESNa thioesters, with both showing significantly higher hydrolysis of the SMN(98–122) thioester. Similar to the synthesis of SMN(1–97), one-pot approaches could not be optimized to a satisfying degree and were therefore not pursued further. Purification of the SMN(98–122)-MESNa thioester prior to its ligation with SMN(123–160) was therefore considered essential for the success of the synthetic strategy.

5.5 Thioester conversion and Acm removal of 1st ligation products

Upon successful isolation of both SMN(1–97) and SMN(98–160), the full-length synthesis of SMN(1–160) was almost complete. SMN(1–97) and its variants were next converted into their corresponding MESNa thioesters, while the Acm protecting group on the *N*-terminal cysteine of SMN(98–160) had to be removed.

5.5.1 MESNa-thioester conversion of SMN(1–97) variants

The thioester conversion of SMN(1–97) was conducted on a scale of 0.7 µmol (6.8 mg). Following activation with $NaNO_2$ at -20 °C for 20 minutes, 35.0 equivalents of MESNa were added. The previous ligation of SMN(1–59) with SMN(60–97) resulted in the formation of an internal cysteine, which is prone to form disulfide bonds. Consequently, formation of a disulfide containing product was observed for SMN(1–97)-MESNa. A small amount of TCEP was therefor added prior to RP-HPLC purification of the remaining variants to reduce any disulfides formed during the synthesis.

The conversion of SMN(1–97)pS49-hydrazide was performed on a scale of 0.4 µmol (3.7 mg). The reaction was stirred at -20°C for 20 minutes, followed by the addition of about 50.0 equivalents of MESNa. Conversion of SMN(1–97)pS63-hydrazide was performed on a scale of 0.9 µmol (8.9 mg). After incubation at -20°C for 20 minutes, 50.0 equivalents of MESNa were added. Finally, 0.6 µmol (6.5 mg) of SMN(1–97)pS49pS63-hydrazide was used for the conversion of the last variant. After stirring at -20°C for 20 minutes, 45.0 equivalents of MESNa were added. The large excess of MESNa used for the synthesis of the SMN(1–97)-MESNa variants was likely unnecessary, as 35.0 equivalents were sufficient to achieve complete conversion of the unmodified SMN(1–97)-hydrazide.

The MESNa thioester conversions of all SMN(1–97) proceeded to completion with no starting material detectable prior to purification. MESNa thioester formation of SMN(1–97) proceeded with a yield of 93%. SMN(1–97)pS49 was obtained with a yield of 95%, whereas SMN(1–97)pS63 achieved a yield of 50%. Conversion of the diphosphorylated SMN(1–97)pS49pS63 resulted in a yield of 68%. Analytical data of all variants is depicted in **Figure 5.17**.

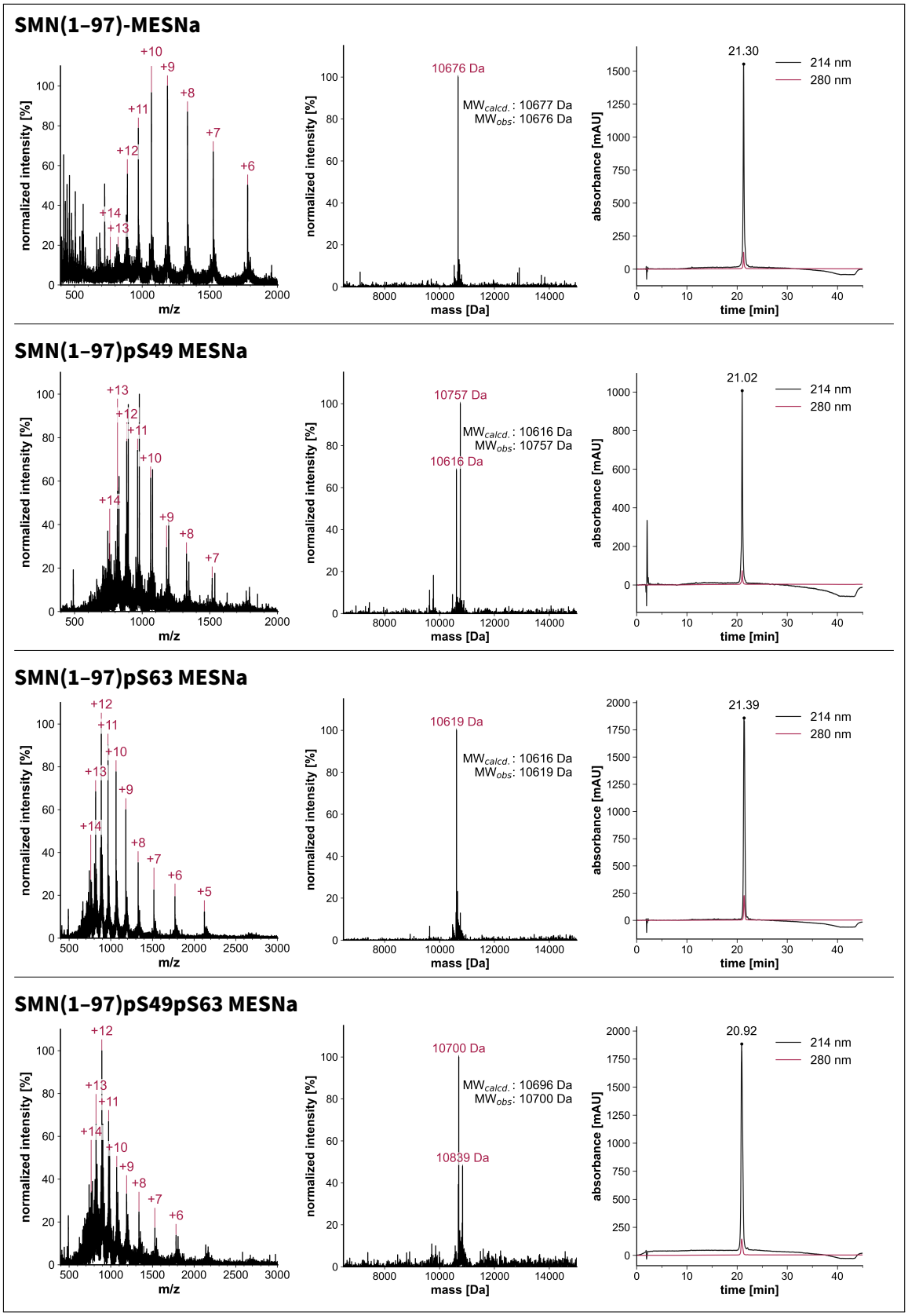

Figure 5.17. Analytical data of SMN(1–97)-MESNa thioesters. SMN(1–97), SMN(1–97)pS49 and SMN(1–97)pS49pS63 contain disulfides. **Left:** m/z spectra. **Middle:** deconvoluted mass spectra. **Right:** analytical RP-HPLC UV/Vis chromatograms recorded at 214 nm and 280 nm upon addition of TCEP.

SMN(1–97) was fully converted into its disulfide form, with a calculated mass of 10677 Da, that closely matched the observed mass of 10676 Da. The corresponding reduced species without a disulfide bond has a calculated mass of 10536 Da. Addition of TCEP prior to purification effectively prevented disulfide formation in SMN(1–97)pS63, yielding a product with an observed mass of 10619 Da, in good agreement with the calculated value of 10616 Da. However, this was only partially successful for SMN(1–97)pS49 and SMN(1–97)pS49pS63. In case of SMN(1–97)pS49-MESNa, both the disulfide (observed and calculated mass of 10757 Da) and the reduced form (observed and calculated mass of 10616 Da) were detected. Similarly, SMN(1–97)pS49pS63 exhibited both the disulfide species, with an observed mass of 10839 Da in close agreement with the calculated mass of 10837 Da, and the reduced form, which showed a small deviation of +4 Da between the observed mass of 10700 Da and the calculated mass of 10696 Da. Since disulfides can be readily reduced by TCEP, no complications were expected for subsequent ligation with SMN(98–160). To assess product purity, TCEP was added prior to RP-HPLC analysis. All products showed high purity under these conditions.

The origin of the yield discrepancies among the different variants remains unclear. Considering that full conversion was observed for all variants, the purification procedure may be responsible for the reduced yields. In contrast, the unexpectedly high yields obtained for SMN(1–97)pS49 and SMN(1–97)pS63 are most likely due to inaccuracies in the weighing of both products and starting materials, rather than improved performance of the purification method for these variants.

5.5.2 Acm removal

Following the synthesis of the SMN(1–97)-MESNa variants and the ligation of SMN(98–122) and SMN(123–160), the Acm protecting group on the *N*-terminal cysteine of SMN(98–160) had to be removed for the second ligation. SMN(98–160) was dissolved in degassed 85% AcOH to a final concentration of 10 mg/mL. Silver acetate (1 mg per mg of peptide) was added, and the reaction was stirred at room temperature. The reaction was monitored by LC-MS. Analysis prior to purification confirmed complete removal of the Acm group, with no starting material detected. The reaction mixture was then diluted 1:1 (v/v) with a degassed solution of 6 M Gnd · HCl containing 1 M dithiothreitol (DTT) and stirred for an additional 30 minutes. The mixture was then centrifuged (5 min, 14000 rpm) and the supernatant was collected. The remaining precipitate was washed with DTT buffer three times, with a centrifugation step (5 min, 14000 rpm) in between each washing. All supernatants were combined and filtered (0.20 µm). Purification was performed by semipreparative RP-HPLC on the Waters system using a C4 column. A gradient of 5–65% buffer B (ACN + 0.08% TFA) in buffer A (ddH_2O + 0.10% TFA) over 60 minutes was applied, preceded by a 10-minute pre-run for desalting.

The procedure concluded with a yield of 72%. The calculated mass of 7017 Da matched the observed value of 7016 Da. Analytical RP-HPLC and MS confirmed that the product had sufficient purity for subsequent ligation with the SMN(1–97)-MESNa thioester and its phospho-variants. Analytical data is shown in **Figure 5.18**. SMN(98–160) proved poorly soluble in ddH_2O, while its solubility in 6 M Gnd · HCl was sufficient for both analytical purposes and subsequent ligation reactions.

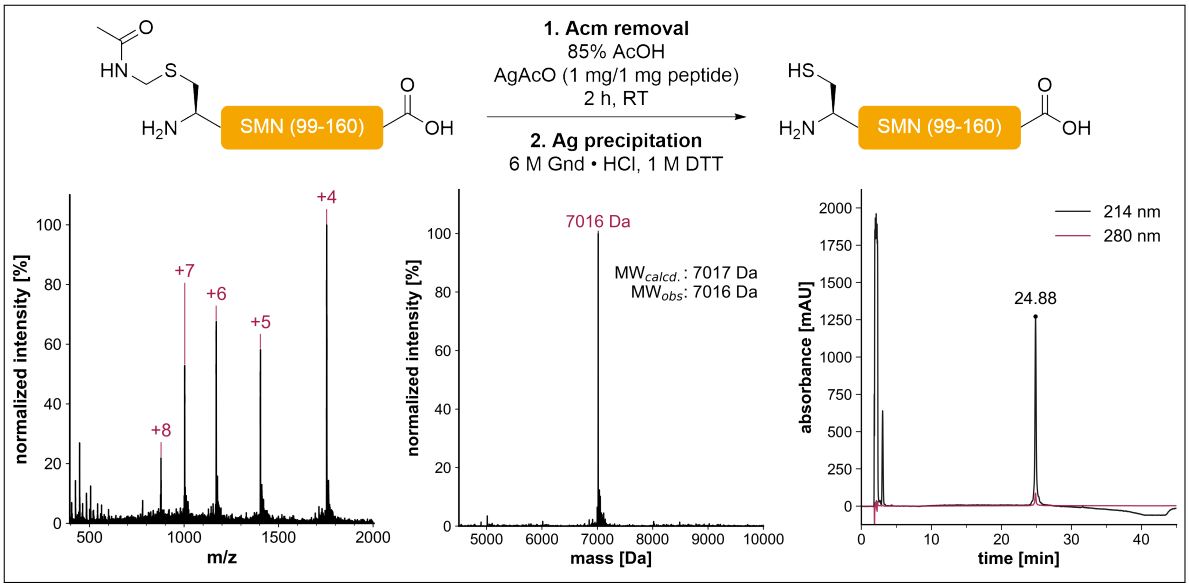

Figure 5.18. Acm removal of SMN(98–160). Top: Synthetic procedure for Acm removal of SMN(98–160). **Bottom:** Analytical data of purified SMN(98–160) after Acm removal. **Left:** m/z spectrum. **Middle:** deconvoluted mass spectrum. **Right:** analytical RP-HPLC UV/Vis chromatogram measured at 214 nm and 280 nm.

Deprotection was also successfully performed using $PdCl_2$, following a procedure reported by Maity *et al.*, 2016.[95] This protocol describes Acm removal directly after peptide ligation without prior purification of the ligation product and may offer a promising opportunity to improve the synthetic strategy presented herein. Combining the ligation and deprotection steps into a single workflow, and thereby reducing the number of purification steps, could represent a useful approach to increasing the overall yield, particularly for the difficult production of SMN(123–160).

Following the successful isolation of deprotected SMN(98–160), all compounds required for the final ligation step were assembled. This second ligation step, combining the SMN(1–97) thioester variants with SMN(98–160), was expected to yield the anticipated SMN(1–160) isoforms.

5.6 Second Ligation reactions - Synthesis of SMN(1–160) variants

The ligation of SMN(1–97)-MESNa with SMN(98–160) was first performed on a small scale and monitored by analytical RP-HPLC and LC-MS over 26 hours (**Figure 5.19**). SMN(1–97)-MESNa thioester (1.0 eq.) and SMN(98–160) (1.0 eq.) were dissolved in ligation buffer to a total concentration of 5.0 mM. MPAA (350 mM) and TCEP (100 mM) were added to the reaction and the pH was adjusted to 6.80.

At 0 hours, SMN(1–97) thioester (21.1 min, 52.3%) and SMN(98–160) (24.5 min, 30.4%) were detected. The MESNa and MPAA thioester of SMN(1–97) co-eluted and are therefore reported collectively. The ligation product SMN(1–160) was already present at this early stage of the reaction (23.4 min, 17.3%). After 2 hours, both SMN(1–97) thioesters (21.6 min, 30.6%) and SMN(98–160) (24.9 min, 18.1%) decreased significantly, while SMN(1–160) content increased (24.1 min, 51.3%). After 23 hours,

the levels of SMN(1–97) thioesters (21.5 min, 32.0%) and SMN(98–160) (24.7 min, 15.5%) remained comparable to those observed at the previous time point. The final measurement at 26 hours showed a further increase in SMN(1–160) (23.8 min, 56.9%), accompanied by a modest decrease in SMN(1–97) thioesters (21.4 min, 29.7%) and SMN(98–160) (24.7 min, 13.5%).

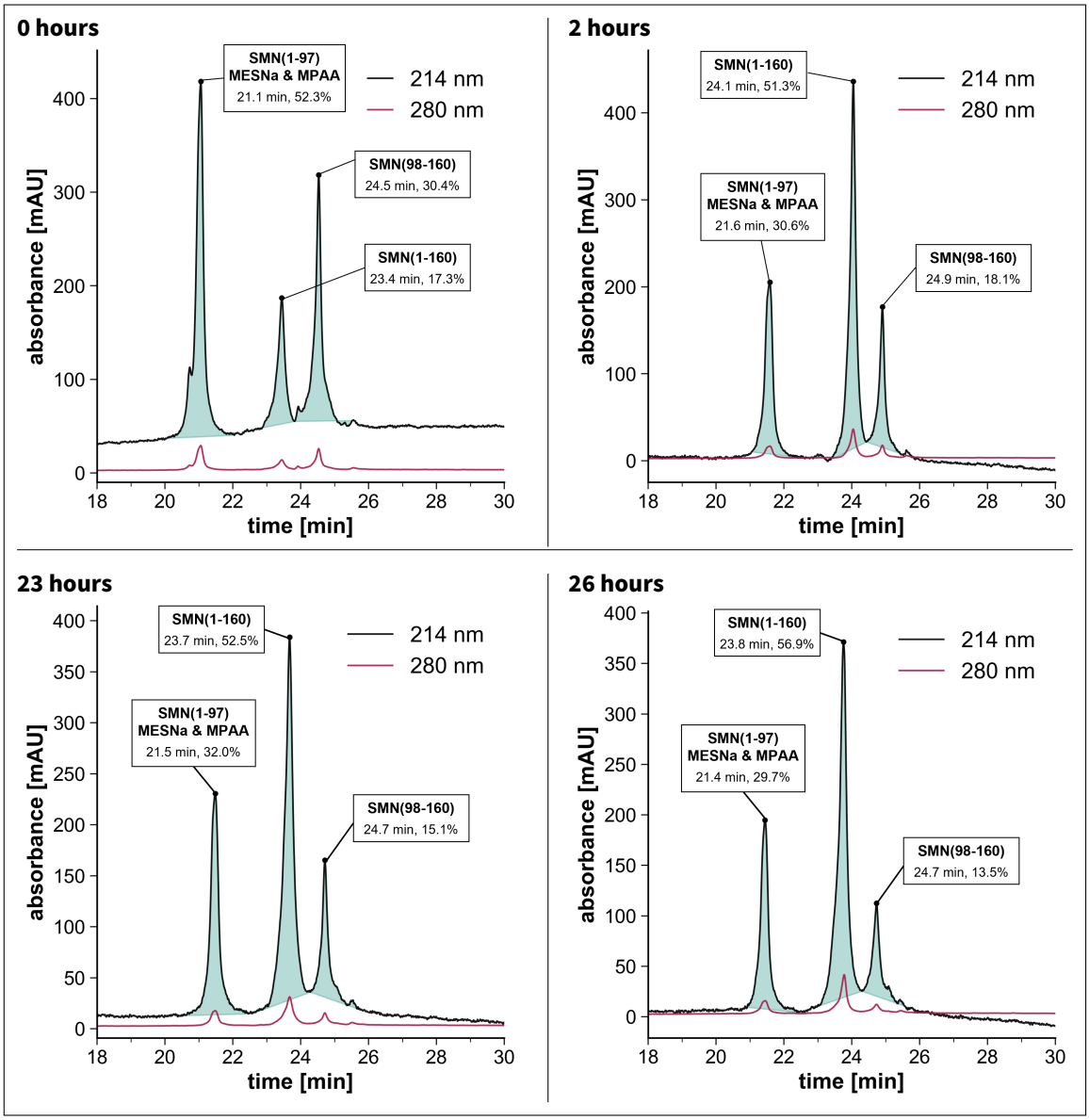

Figure 5.19. Monitoring of the NCL reaction between the SMN(1–97)-MESNa thioester and SMN(98–160). Average retention times (RT) of the compounds: SMN(1–97)-MESNa and SMN(1–97)-MPAA 21.5 min, SMN(1–160) 23.7 min, SMN(98–160) 24.7 min. Peaks are shifted to slightly lower values RT values at 0 hours, while slightly higher values are detected at 26 hours. The 0 hour measurement also shows an overall increased baseline.

The reaction proceeded rapidly, reaching a ligation rate of up to 51.3% within the first two hours. Extending the reaction time to 26 hours resulted in only minor additional conversion, indicating that a ligation period of a few hours is sufficient for the synthesis of SMN(1–160) variants. Additional measurements within the first 12 hours of the reaction would help identify the optimal reaction time,

further improving the overall protocol. However, all reactions were still carried out for up to two days due to scheduling reasons. Hydrolysis of the SMN(1–97) thioesters co-eluted with both thioester species, which complicates accurate quantification. Considering the rapid reaction kinetics and the satisfying conversion rate, hydrolysis is not expected to pose a significant issue.

Following the promising results of the test ligation, SMN(1–160) variants were synthesized based on the conditions of the initial experiment. SMN(1–97)-MESNa thioester (0.4 µmol, 4.5 mg, 1.0 eq.) and SMN(98–160) (0.4µmol, 3.1 mg, 1.0 eq.) were dissolved in ligation buffer to a total concentration of 5.3 mM. MPAA (350 mM) and TCEP (110 mM) were added to the reaction and the pH was adjusted to 6.80. The ligation was stirred at room temperature for 48 hours.

The synthesis of SMN(1–160)pS49 was performed with a small excess of SMN(98–160) and an increased total peptide concentration compared to the unmodified variant to improve ligation efficiency. SMN(1–97)pS49-MESNa thioester (0.3 µmol, 3.1 mg, 1.0 eq.) and SMN(98–160) (0.3µmol, 2.3 mg, 1.1 eq.) were dissolved in ligation buffer to a total concentration of 6.4 mM. MPAA (310 mM) and TCEP (110 mM) were added, and the pH was adjusted to 6.80. The reaction was left stirring at room temperature for 40 hours before purification.

SMN(1–160)pS63 was synthesized under similar conditions. The concentration of MPAA was reduced, as previously used amounts were considered higher than necessary. SMN(1–97)pS63-MESNa thioester (0.4 µmol, 3.8 mg, 1.0 eq.) and SMN(98–160) (0.4 µmol, 2.7 mg, 1.1 eq.) were dissolved in ligation buffer to a total concentration of 6.4 mM. MPAA (260 mM) and TCEP (120 mM) were added, and the pH was adjusted to 6.80. The reaction was left stirring at room temperature for 40 hours prior to purification.

The diphosphorylated variant SMN(1–160)pS49pS63 was synthesized using a slightly larger excess of SMN(98–160). SMN(1–97)pS49pS63-MESNa thioester (0.4 µmol, 4.5 mg, 1.0 eq.) and SMN(98–160) (0.5 µmol, 3.4 mg, 1.2 eq.) were dissolved in ligation buffer to a total concentration of 6.5 mM. MPAA (300 mM) and TCEP (120 mM) were added to the reaction and the pH was adjusted to 6.80. The reaction was left stirring for 44 hours prior to purification. The purity of each compound was confirmed by RP-HPLC, as depicted in **Figure 5.20**.

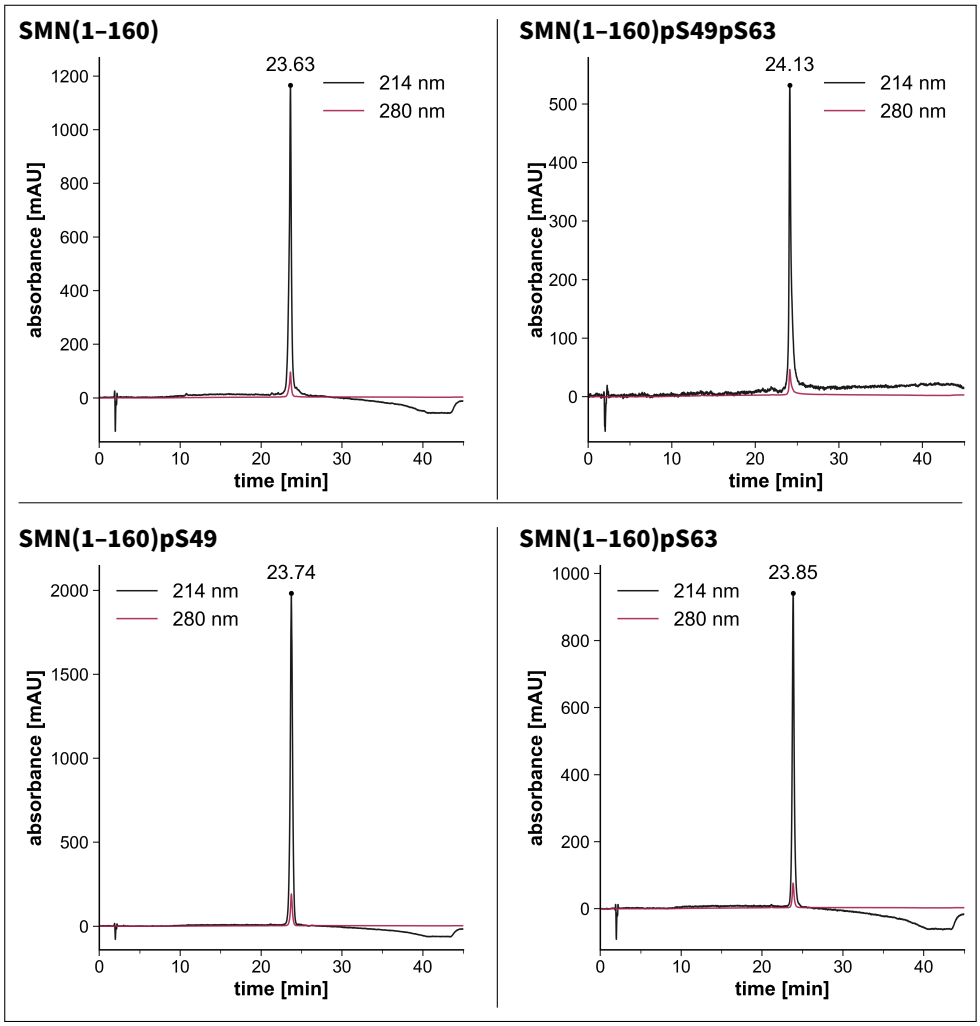

Figure 5.20. RP-HPLC analysis of purified SMN(1–160) variants. Chromatograms were measured at 214 nm and 280 nm.

RP-HPLC analysis indicated that all variants were obtained with high purity (>95%) and no detectable contaminants. Ligation of the unmodified SMN(1–160) variant achieved a yield of 36% (3 mg). A comparable yield of 39% (2 mg) was obtained for the monophosphorylated SMN(1–160)pS49. The highest yield was observed for SMN(1–160)pS63 at 53% (3 mg), whereas the diphosphorylated SMN(1–160)pS49pS63 afforded a substantially lower yield of 17% (1 mg).

Conversion rates of all variant syntheses averaged at approximately 54.2%, as determined by RP-HPLC peak area analysis. Reducing the MPAA concentration from 310 mM for SMN(1–160)pS49 to 260 mM for SMN(1–160)pS63 did not result in a decrease in ligation efficiency. Given the relatively high ligation extent of SMN(1–160)pS49pS63 (approximately 50.0%), its low yield is likely attributable to losses during purification. Further optimization of this final ligation was not pursued due to the limited availability of the intermediate ligation segments SMN(1–97) and SMN(98–160).

The masses of all variants were determined by high resolution mass spectrometry (HR-MS). SMN(1–160) showed a good agreement between the calculated mass of 17410.35 Da and the observed value of 17409.56 Da. An impurity with a mass shift of +55.14 Da was also detected ($MW_{obs.}$: 17465.49 Da) (**Figure 5.21, Top**). The calculated mass of 17568.3 Da of the diphosphorylated SMN(1–160)pS49pS63 was in good agreement with the observed value of 17569.4 Da. Three additional species, with mass differences of -16.9 Da ($MW_{obs.}$: 17552.5 Da), +23.1 Da ($MW_{obs.}$: 17592.5 Da), and +57.2 Da ($MW_{obs.}$: 175626.6 Da), were also detected (**Figure 5.21, Bottom**).

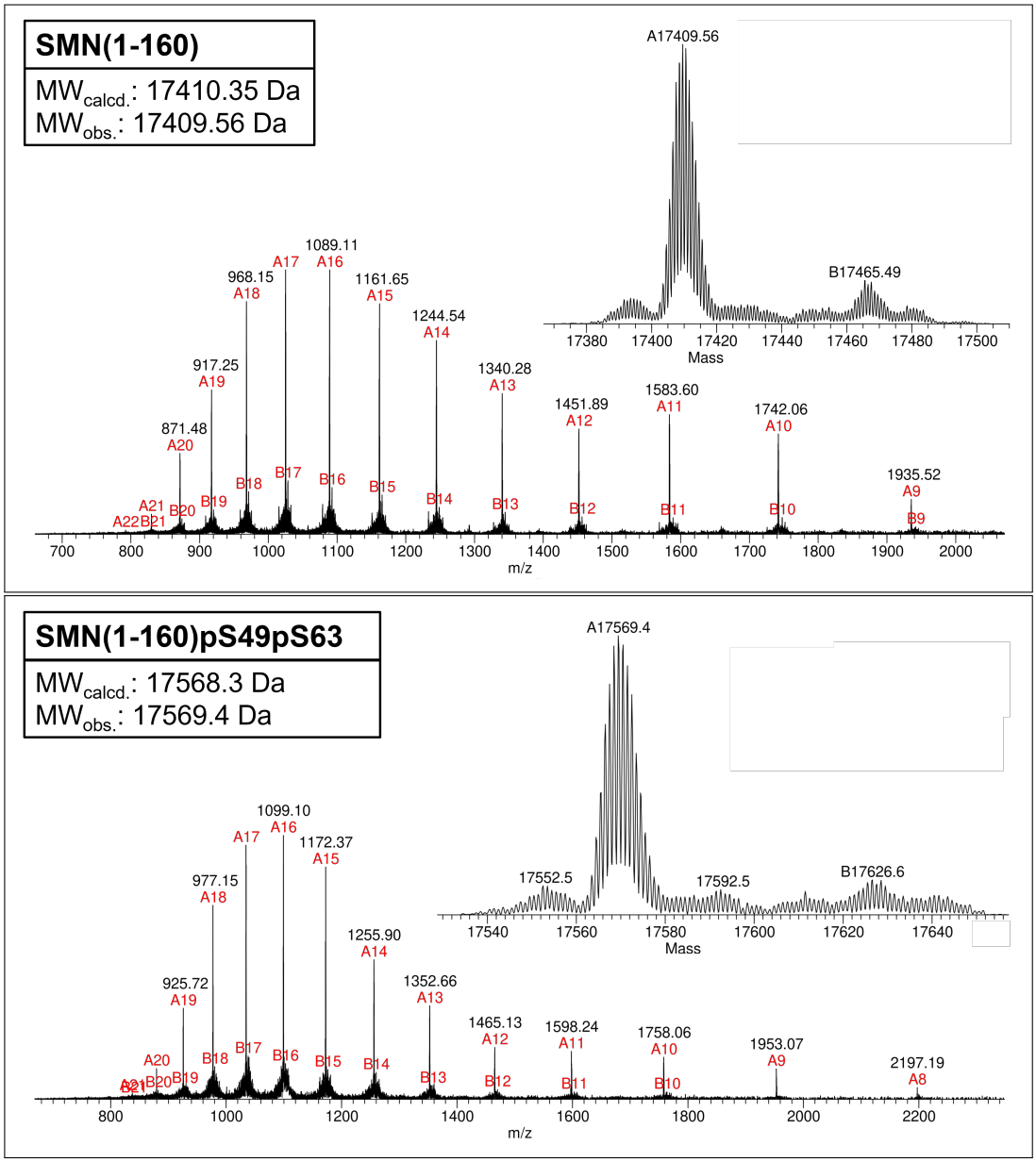

Figure 5.21. HR-MS of SMN(1–160) and SMN(1–160)pS49pS63. The deconvoluted mass spectra are shown above the corresponding m/z spectra for each variant. **Top:** SMN(1–160), **Bottom**: SMN(1–160)pS49pS63.

The observed mass of SMN(1–160)pS49 with 17489.55 Da closely matched the calculated value of 17489.26 Da (**Figure 5.22, top**). Two impurities were detected, showing mass deviations of -17.00 Da ($MW_{obs.}$: 17472.55 Da) and +56.97 Da ($MW_{obs.}$: 17546.52 Da). SMN(1–160)pS63 showed an observed mass of 17489.55 Da, in close agreement with the calculated value of 17489.36 Da (**Figure 5.22, bottom**). An impurity with a mass difference of +56.97 Da ($MW_{obs.}$: 17546.52 Da) was also detected.

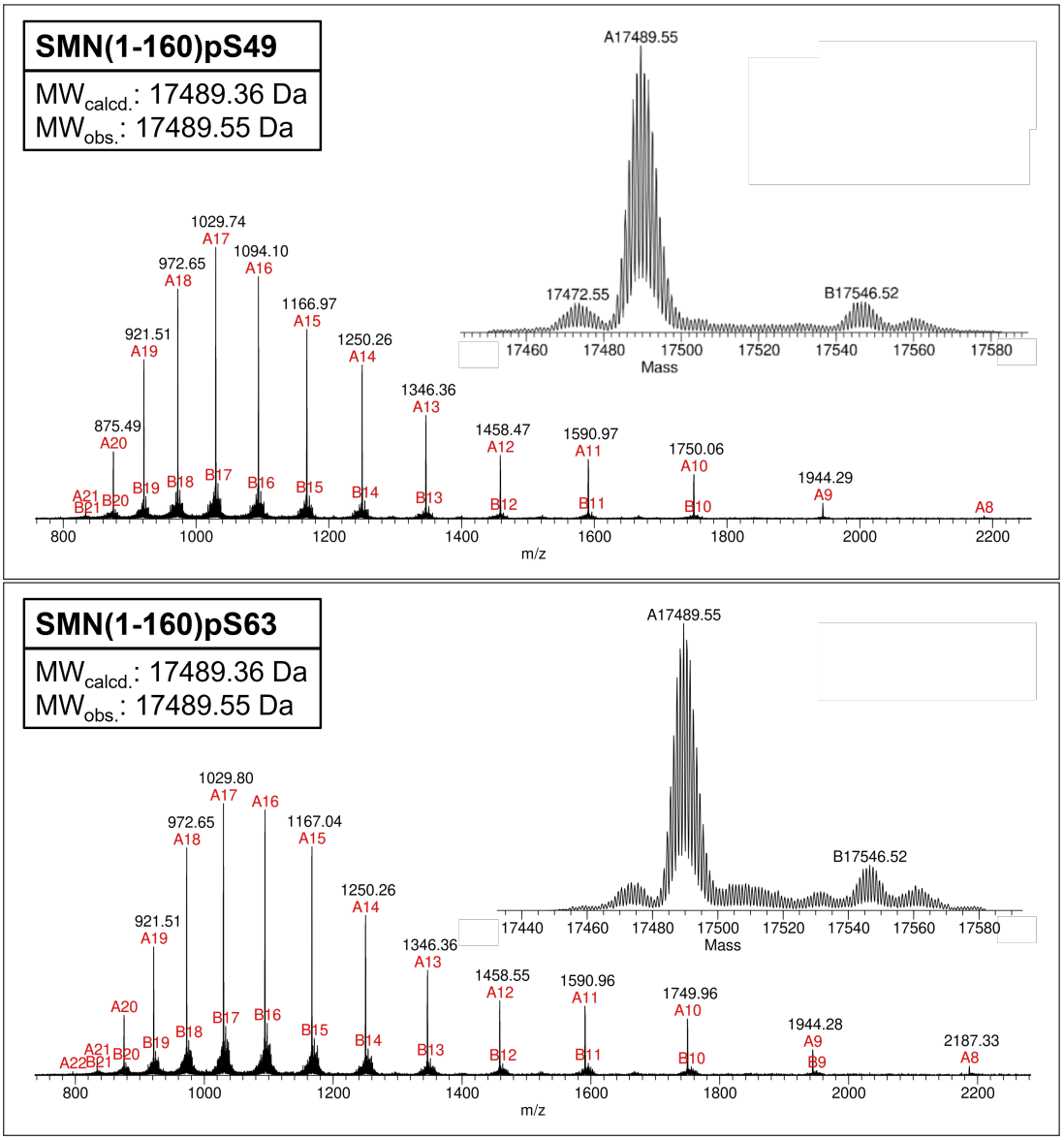

Figure 5.22. HR-MS of SMN(1–160)pS49 and SMN(1–160)pS63. The deconvoluted mass spectra are shown above the corresponding m/z spectra for each variant. **Top:** SMN(1–160)pS49, **Bottom**: SMN(1–160)pS63.

Although HR-MS revealed contaminants in all variants that were not detected previously, the overall purity was considered sufficient for CD spectroscopy and subsequent biochemical assays. Attempts to identifying these impurities were unsuccessful. The observed masses of all variants were in good agreement with their calculated values.

5.7 Discussion of the synthetic strategy

All peptides were synthesized using SPPS, employing a modified segmentation strategy of the target sequence compared to previous attempts. Although high purity was achieved for all peptides, the overall yields of SMN(1–59)pS49 and SMN(123–160) remained low, limiting efficient production of full-length SMN variants. The incorporation of phosphorylations proved particularly challenging, even under optimized SPPS conditions, indicating that further methodological optimization will be required to improve production yields. While shifting the ligation site between segments 3 and 4 from C146 to C123 improved the synthetic outcome, the pronounced hydrophobicity of residues 98–160 continued to complicate procedures. In particular, the synthesis of SMN(123—160) requires further refinement to reduce the formation of the +55 Da and +85 Da byproducts that significantly hinder purification.

In contrast to the challenges encountered during peptide synthesis, thioester conversions and subsequent ligations proceeded efficiently. MESNa was identified as the most suitable thioester to form peptide thioesters and to be added as a ligation mediator, as it minimizes hydrolysis while maintaining sufficient ligation rates. Given the total conversion of starting material, steeper solvent gradients during chromatographic purification could reduce both purification time and solvent consumption without substantially compromising product purity. Purification after each reaction step is crucial for the success of the overall strategy, as attempts at one-pot procedures failed to produce satisfactory outcomes. However, the cumulative number of purification steps represents a major bottleneck for overall yield. Substantial improvements in the synthesis of SMN(1–160) variants will therefore likely depend on reducing the number of purification steps. Although not yet successfully implemented, one-pot approaches represent a promising avenue toward achieving this goal. Additional improvements may also be achieved by shortening ligation times to 24 hours, potentially reducing the total synthesis time while maintaining comparable reaction outcomes. Given that most currently known phosphorylation sites of the SMN protein are located within the target sequence investigated in this thesis,[47] the synthetic strategy developed here provides a solid foundation for the preparation and future investigation of a broad range of phosphorylated and otherwise modified SMN variants.

5.8 Circular dichroism spectroscopy

The successful synthesis of the four planned SMN(1–160) variants in high purity now enables initial characterisation of the individual isoforms. To assess potential differences in their solution-state structures, CD spectroscopy was employed. Samples of the SMN(1–160) variants, as well as SMN(1–59) and SMN(1–59)pS49 peptide segments, were dissolved in 10 mM sodium phosphate buffer to a final concentration of 0.25 mg/mL and incubated overnight (1040 rpm, 25 °C). SMN(1–160) did not fully dissolve, so its concentration was determined using a NanoDrop 2000c, resulting in 0.087 mg/mL. CD spectra were recorded in a 1 mm quartz cuvette using buffer as blank. More details on the sample preparation are listed in **Section 4.3**.

For an all-α-helical protein, two distinct negative bands of similar magnitude are expected at 208 nm and 222 nm, together with a positive band near 190 nm. An all-β-sheet protein typically shows a broad negative band between 210 nm and 220 nm and a positive signal around 195 nm to 200 nm. Disordered proteins generally display a pronounced negative band around 200 nm.[96] The recorded spectra are shown in **Figure 5.23**

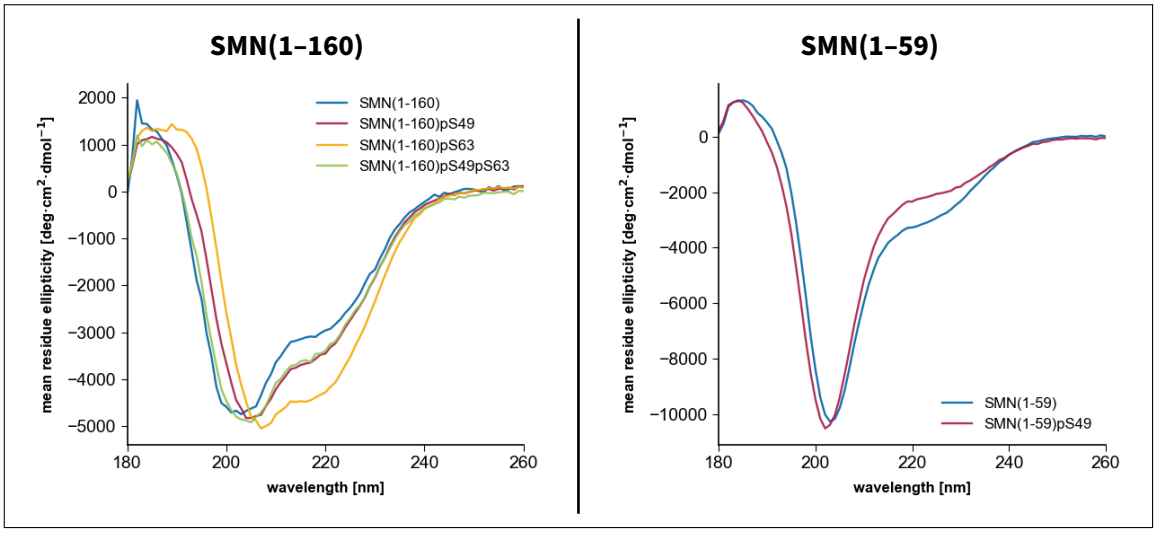

Figure 5.23. CD spectra of SMN(1–160) and SMN(1–59) variants. Spectra were recorded between 180–260 nm.

The SMN(1–160) variants show only minor differences in their CD spectra. However, limitations in sample preparation and uncertainties in concentration determination preclude a detailed comparison of the compounds. The spectra of SMN(1–160), SMN(1–160)pS49 and SMN(1–160)pS49pS63 share many similarities, with negative bands around 200–205 nm and a less pronounced band near 215 nm. In contrast, SMN(1–160)pS63 shows a slight shift of these bands toward 208 nm and 220 nm, indicating an α-helix. Both SMN(1–59) variants exhibit pronounced negative bands at 202 nm, signaling a predominantly disordered structure.

None of the SMN(1–160) or SMN(1–59) variants shows clear evidence of prominent structural elements. With the exception of SMN(1–160)pS63, neither characteristic features of α-helical nor β-sheet content are observed. Structural information on the SMN protein remains limited. Only few structural features have been characterized, including an α-helix spanning residues 37-51 and the barrel-like Tudor domain located between residues 89 and 151.[9,24] The α-helix is located within the SMN(1–59) segment and, although it comprises only a quarter of it, α-helices are generally considered to produce strong CD-signals and were therefore expected to be detectable to some extent.[97] Its apparent absence may indicate inappropriate sample preparation, highlighting the need for more refined experimental methods. Improved folding strategies may involve redissolving the samples in denaturing buffers, followed by the gradual removal of the denaturant through stepwise dilution or dialysis. Additionally, additives such as DTT or mixtures of reduced and oxidized glutathione can be employed to promote the rearrangement of disulfide bonds and thereby facilitate correct protein folding.[98]

Due to the limited structural data available for the SMN protein, structure prediction using AlphaFold may be useful. AlphaFold 3 predicts the presence of both the α-helix and the Tudor domain (**Figure 5.24**).[99] Interestingly, simulations restricted to SMN(1–59) predict an additional α-helix in close proximity to the *N*-terminus. The remaining residues within SMN(1–160) are expected to be disordered. Further studies are required to obtain a more complete understanding of the three-dimensional structure of the SMN protein and its *N*-terminal domain.

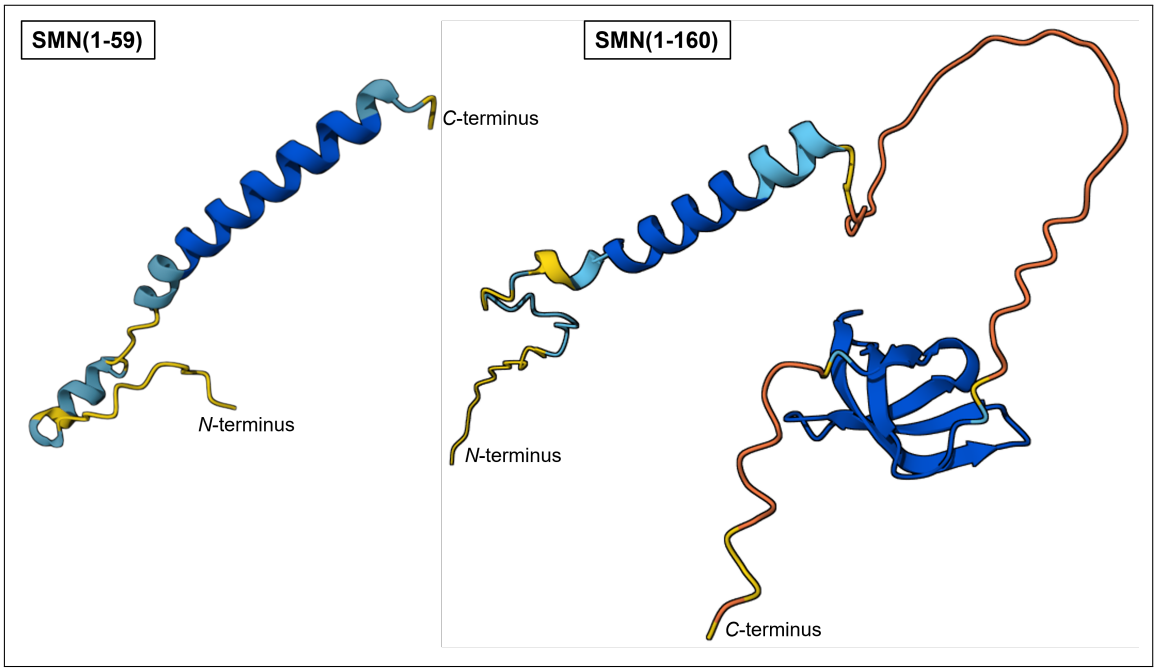

Figure 5.24. AlphaFold 3 prediction of SMN(1–59) and SMN(1–160).[99]

6 Conclusion and Outlook

In this thesis, a synthesis strategy was developed to produce site-selectively phosphorylated variants of the *N*-terminal domain (amino acids 1–160) of the SMN protein. Based on previous research of the Becker group, the target region of the SMN protein was split into four peptide segments, each generated through solid phase peptide synthesis. While this approach proved successful, challenges persist concerning the difficult synthesis of the hydrophobic fourth segment SMN(123–160) as well as the phosphorylated segments that give lower yields compared to their unmodified counterparts. Although detailed investigations of the synthetic procedures might lead to further improvements, the described process still supplies sufficient amounts of each peptide for obtaining SMN(1–160) variants.

Chemical transformations of the peptide segments throughout the strategy, including thioester conversions, peptide segment ligations, and Acm removal, achieved high conversion rates and short reaction times. Purification procedures were identified as the main source of product loss and, although essential for the success of the overall strategy, significantly limited overall yields. Consequently, developing more robust and efficient purification methods, or alternatively, reducing or eliminating purification steps through alternative protocols, will be crucial for improving overall yields and advancing future applications.

A total of four variants of the SMN protein were synthesized: one unmodified, two monophosphorylated, and one diphosphorylated variant bearing modifications on S49 and S63. All were successfully obtained on a milligram scale (1–3 mg) and subsequently characterized. The developed pathway, relying solely on solid phase peptide synthesis and native chemical ligation, forms a solid foundation for the synthesis and subsequent investigation of additional phosphorylation patterns of the SMN protein. This is especially relevant, considering the notoriously difficult nature of obtaining SMN through alternative methods such as recombinant protein expression. A method that also limits access to phosphorylated SMN-variants.[47,100]

Future studies can build on this work to investigate the physiological relevance and functional significance of not only pS49 and pS63 but also additional phosphorylation sites. Although circular dichroism spectroscopy was employed to probe structural changes induced by pS49 and pS63, the results were inconclusive. Further research is therefore needed to assess the potential effects of these phosphorylations and to obtain a more complete understanding of the structure of SMN beyond its known structural motifs and in the context of the SMN complex.[9,24] At a physiological level, subsequent experiments may also explore how these modifications influence the interaction of SMN with Gemin and Sm proteins. Since the Gemin2- and Sm protein-binding domains of SMN are contained within the synthesized target region,[39] *in vitro* investigations of Sm core assembly are expected to be feasible, providing insights into the pivotal role of phosphorylation in regulating this process.

7 References

(1) Werdnig, G. *Archives of Neurology* **1971**, *25*, 276–278.

(2) Hoffmann, J.; Hoffmann, J. *Archiv für Psychiatrie und Nervenkrankheiten* **1892**, *24*, 644–646.

(3) Mercuri, E. et al. *Neuromuscular Disorders* **2018**, *28*, 103–115.

(4) Wang, C. H.; Finkel, R. S.; Bertini, E. S.; Schroth, M.; Simonds, A.; Wong, B.; Aloysius, A.; Morrison, L.; Main, M.; Crawford, T. O.; Trela, A. *Journal of Child Neurology* **2007**, *22*, 1027–1049.

(5) Lefebvre, S.; Bürglen, L.; Reboullet, S.; Clermont, O.; Burlet, P.; Viollet, L.; Benichou, B.; Cruaud, C.; Millasseau, P.; Zeviani, M. *Cell* **1995**, *80*, 155–165.

(6) Bürglen, L.; Lefebvre, S.; Clermont, O.; Burlet, P.; Viollet, L.; Cruaud, C.; Munnich, A.; Melki, J. *Genomics* **1996**, *32*, 479–482.

(7) Lorson, C. L.; Hahnen, E.; Androphy, E. J.; Wirth, B. *Proceedings of the National Academy of Sciences* **1999**, *96*, 6307–6311.

(8) Lefebvre, S.; Burlet, P.; Liu, Q.; Bertrandy, S.; Clermont, O.; Munnich, A.; Dreyfuss, G.; Melki, J. *Nature Genetics* **1997**, *16*, 265–269.

(9) Selenko, P.; Sprangers, R.; Stier, G.; Bühler, D.; Fischer, U.; Sattler, M. *Nature Structural Biology* **2001**, *8*, 27–31.

(10) Giesemann, T.; Rathke-Hartlieb, S.; Rothkegel, M.; Bartsch, J. W.; Buchmeier, S.; Jockusch, B. M.; Jockusch, H. *Journal of Biological Chemistry* **1999**, *274*, 37908–37914.

(11) Martin, R.; Gupta, K.; Ninan, N. S.; Perry, K.; Van Duyne, G. D. *Structure* **2012**, *20*, 1929–1939.

(12) Liu, Q.; Dreyfuss, G. *The EMBO Journal* **1996**, *15*, 3555–3565.

(13) Lorson, C. L.; Strasswimmer, J.; Yao, J.-M.; Baleja, J. D.; Hahnen, E.; Wirth, B.; Le, T.; Burghes, A. H. M.; Androphy, E. J. *Nature Genetics* **1998**, *19*, 63–66.

(14) Otter, S.; Grimmler, M.; Neuenkirchen, N.; Chari, A.; Sickmann, A.; Fischer, U. *Journal of Biological Chemistry* **2007**, *282*, 5825–5833.

(15) Liu, Q.; Fischer, U.; Wang, F.; Dreyfuss, G. *Cell* **1997**, *90*, 1013–1021.

(16) Charroux, B.; Pellizzoni, L.; Perkinson, R. A.; Shevchenko, A.; Mann, M.; Dreyfuss, G. *Journal of Cell Biology* **1999**, *147*, 1181–1194.

(17) Charroux, B.; Pellizzoni, L.; Perkinson, R. A.; Yong, J.; Shevchenko, A.; Mann, M.; Dreyfuss, G. *Journal of Cell Biology* **2000**, *148*, 1177–1186.

(18) Gubitz, A. K.; Mourelatos, Z.; Abel, L.; Rappsilber, J.; Mann, M.; Dreyfuss, G. *Journal of Biological Chemistry* **2002**, *277*, 5631–5636.

(19) Pellizzoni, L.; Baccon, J.; Rappsilber, J.; Mann, M.; Dreyfuss, G. *Journal of Biological Chemistry* **2002**, *277*, 7540–7545.

(20) Baccon, J.; Pellizzoni, L.; Rappsilber, J.; Mann, M.; Dreyfuss, G. *Journal of Biological Chemistry* **2002**, *277*, 31957–31962.

(21) Carissimi, C.; Saieva, L.; Baccon, J.; Chiarella, P.; Maiolica, A.; Sawyer, A.; Rappsilber, J.; Pellizzoni, L. *Journal of Biological Chemistry* **2006**, *281*, 8126–8134.

(22) Meister, G.; Bühler, D.; Pillai, R.; Lottspeich, F.; Fischer, U. *Nature Cell Biology* **2001**, *3*, 945–949.

(23) Meier, I. D.; Walker, M. P.; Matera, A. G. *Biology Open* **2018**, *7*, bio032409.

(24) Zhang, R.; So, B. R.; Li, P.; Yong, J.; Glisovic, T.; Wan, L.; Dreyfuss, G. *Cell* **2011**, *146*, 384–395.

(25) Campbell, L.; Hunter, K. M. D.; Mohaghegh, P.; Tinsley, J. M.; Brasch, M. A.; Davies, K. E. *Human Molecular Genetics* **2000**, *9*, 1093–1100.

(26) Chow, L. T.; Gelinas, R. E.; Broker, T. R.; Roberts, R. J. *Cell* **1977**, *12*, 1–8.

(27) Berget, S. M.; Moore, C.; Sharp, P. A. *Proceedings of the National Academy of Sciences* **1977**, *74*, 3171–3175.

(28) Ruskin, B.; Krainer, A. R.; Maniatis, T.; Green, M. R. *Cell* **1984**, *38*, 317–331.

(29) Grabowski, P. J.; Seiler, S. R.; Sharp, P. A. *Cell* **1985**, *42*, 345–353.

(30) Frendewey, D.; Keller, W. *Cell* **1985**, *42*, 355–367.

(31) Morais, P.; Adachi, H.; Yu, Y.-T. *Frontiers in Genetics* **2021**, *12*, 652129.

(32) Will, C. L.; Lührmann, R. *Cold Spring Harbor Perspectives in Biology* **2011**, *3*, a003707.

(33) Lerner, M. R.; Steitz, J. A. *Proceedings of the National Academy of Sciences* **1979**, *76*, 5495–5499.

(34) Raker, V. A.; Plessel, G.; Lührmann, R. *The EMBO Journal* **1996**, *15*, 2256–2269.

(35) Hermann, H.; Fabrizio, P.; Raker, V.; Foulaki, K.; Hornig, H.; Brahms, H.; Lührmann, R. *The EMBO Journal* **1995**, *14*, 2076–2088.

(36) Pomeranz Krummel, D. A.; Oubridge, C.; Leung, A. K. W.; Li, J.; Nagai, K. *Nature* **2009**, *458*, 475–480.

(37) Raker, V. A.; Hartmuth, K.; Kastner, B.; Lührmann, R. *Molecular and Cellular Biology* **1999**, *19*, 6554–6565.

(38) Pellizzoni, L.; Yong, J.; Dreyfuss, G. *Science* **2002**, *298*, 1775–1779.

(39) Chari, A.; Golas, M. M.; Klingenhäger, M.; Neuenkirchen, N.; Sander, B.; Englbrecht, C.; Sickmann, A.; Stark, H.; Fischer, U. *Cell* **2008**, *135*, 497–509.

(40) Yong, J.; Kasim, M.; Bachorik, J. L.; Wan, L.; Dreyfuss, G. *Molecular Cell* **2010**, *38*, 551–562.

(41) Narayanan, U.; Achsel, T.; Lührmann, R.; Matera, A. G. *Molecular Cell* **2004**, *16*, 223–234.

(42) Pellizzoni, L.; Charroux, B.; Dreyfuss, G. *Proceedings of the National Academy of Sciences* **1999**, *96*, 11167–11172.

(43) Walsh, C. T.; Garneau-Tsodikova, S.; Gatto Jr., G. J. *Angewandte Chemie International Edition* **2005**, *44*, 7342–7372.

(44) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. *Scientific Reports* **2011**, *1*, 90.

(45) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. *Science* **2002**, *298*, 1912–1934.

(46) Rabinowitz, M.; Lipmann, F. *Journal of Biological Chemistry* **1960**, *235*, 1043–1050.

(47) Detering, N. T.; Schüning, T.; Hensel, N.; Claus, P. *Cellular and Molecular Life Sciences* **2022**, *79*, 497.

(48) Grimmler, M.; Bauer, L.; Nousiainen, M.; Körner, R.; Meister, G.; Fischer, U. *EMBO Reports* **2005**, *6*, 70–76.

(49) Schilling, M.; Prusty, A. B.; Boysen, B.; Oppermann, F. S.; Riedel, Y. L.; Husedzinovic, A.; Rasouli, H.; König, A.; Ramanathan, P.; Reymann, J.; Erfle, H.; Daub, H.; Fischer, U.; Gruss, O. J. *Cell Reports* **2021**, *35*, 109277.

(50) Rademacher, S.; Detering, N. T.; Schüning, T.; Lindner, R.; Santonicola, P.; Wefel, I.-M.; Dehus, J.; Walter, L. M.; Brinkmann, H.; Niewienda, A.; Janek, K.; Varela, M. A.; Bowerman, M.; Di Schiavi, E.; Claus, P. *Cells* **2020**, *9*, 2405.

(51) Maciejewski, P. M.; Peterson, F. C.; Anderson, P. J.; Brooks, C. L. *Journal of Biological Chemistry* **1995**, *270*, 27661–27665.

(52) Bressler, S. G.; Grunhaus, D.; Hurevich, M.; Friedler, A. *Biochemical Society Transactions* **2026**, *54*, 281–297.

(53) Kozeleková, A.; Náplavová, A.; Brom, T.; Gašparik, N.; Šimek, J.; Houser, J.; Hritz, J. *Frontiers in Chemistry* **2022**, *10*, DOI: 10.3389/fchem.2022.835733.

(54) Yoshida, H.; Goedert, M. *Journal of Neurochemistry* **2006**, *99*, 154–164.

(55) Park, H.-S.; Hohn, M. J.; Umehara, T.; Guo, L.-T.; Osborne, E. M.; Benner, J.; Noren, C. J.; Rinehart, J.; Söll, D. *Science* **2011**, *333*, 1151–1154.

(56) Merrifield, R. B. *Journal of the American Chemical Society* **1963**, *85*, 2149–2154.

(57) Amblard, M.; Fehrentz, J.-A.; Martinez, J.; Subra, G. *Molecular Biotechnology* **2006**, *33*, 239–254.

(58) Carpino, L. A.; Han, G. Y. *The Journal of Organic Chemistry* **1972**, *37*, 3404–3409.

(59) Conibear, A. C.; Watson, E. E.; Payne, R. J.; Becker, C. F. W. *Chemical Society Reviews* **2018**, *47*, 9046–9068.

(60) Conibear, A. C. *Nature Reviews Chemistry* **2020**, *4*, 674–695.

(61) Dawson, P. E.; Kent, S. B. H. *Annual Review of Biochemistry* **2000**, *69*, 923–960.

(62) Kulkarni, S. S.; Sayers, J.; Premdjee, B.; Payne, R. J. *Nature Reviews Chemistry* **2018**, *2*, 0122.

(63) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. H. *Science* **1994**, *266*, 776–779.

(64) Hackeng, T. M.; Griffin, J. H.; Dawson, P. E. *Proceedings of the National Academy of Sciences* **1999**, *96*, 10068–10073.

(65) Johnson, E. C. B.; Kent, S. B. H. *Journal of the American Chemical Society* **2006**, *128*, 6640–6646.

(66) Jiang, W.; Zhang, B.; Fan, C.; Wang, M.; Wang, J.; Deng, Q.; Liu, X.; Chen, J.; Zheng, J.; Liu, L.; Zhu, T. F. *Cell Discovery* **2017**, *3*, 17037.

(67) Sun, H.; Brik, A. *Accounts of Chemical Research* **2019**, *52*, 3361–3371.

(68) Sun, H.; Mali, S. M.; Singh, S. K.; Meledin, R.; Brik, A.; Kwon, Y. T.; Kravtsova-Ivantsiv, Y.; Bercovich, B.; Ciechanover, A. *Proceedings of the National Academy of Sciences* **2019**, *116*, 7805–7812.

(69) Wan, Q.; Danishefsky, S. J. *Angewandte Chemie International Edition* **2007**, *46*, 9248–9252.

(70) Yan, L. Z.; Dawson, P. E. *Journal of the American Chemical Society* **2001**, *123*, 526–533.

(71) Crich, D.; Banerjee, A. *Journal of the American Chemical Society* **2007**, *129*, 10064–10065.

(72) Loibl, S. F.; Dallmann, A.; Hennig, K.; Juds, C.; Seitz, O. *Chemistry – A European Journal* **2018**, *24*, 3623–3633.

(73) Futaki, S.; Sogawa, K.; Maruyama, J.; Asahara, T.; Niwa, M.; Hojo, H. *Tetrahedron Letters* **1997**, *38*, 6237–6240.

(74) Alsina, J.; Yokum, T. S.; Albericio, F.; Barany, G. *The Journal of Organic Chemistry* **1999**, *64*, 8761–8769.

(75) Ingenito, R.; Bianchi, E.; Fattori, D.; Pessi, A. *Journal of the American Chemical Society* **1999**, *121*, 11369–11374.

(76) Blanco-Canosa, J. B.; Dawson, P. E. *Angewandte Chemie International Edition* **2008**, *47*, 6851–6855.

(77) Blanco-Canosa, J. B.; Nardone, B.; Albericio, F.; Dawson, P. E. *Journal of the American Chemical Society* **2015**, *137*, 7197–7209.

(78) Fang, G.-M.; Li, Y.-M.; Shen, F.; Huang, Y.-C.; Li, J.-B.; Lin, Y.; Cui, H.-K.; Liu, L. *Angewandte Chemie International Edition* **2011**, *50*, 7645–7649.

(79) Veber, D. F.; Milkowski, J. D.; Denkewalter, R. G.; Hirschmann, R. *Tetrahedron Letters* **1968**, *9*, 3057–3058.

(80) Veber, D.; Milkowski, J.; Varga, S.; Denkewalter, R.; Hirschmann, R. *Journal of the American Chemical Society* **1972**, *94*, 5456–5461.

(81) Laps, S.; Sun, H.; Kamnesky, G.; Brik, A. *Angewandte Chemie International Edition* **2019**, *58*, 5729–5733.

(82) Liu, S.; Pentelute, B. L.; Kent, S. B. H. *Angewandte Chemie International Edition* **2012**, *51*, 993–999.

(83) Shen, F.; Zhang, Z.-P.; Li, J.-B.; Lin, Y.; Liu, L. *Organic Letters* **2011**, *13*, 568–571.

(84) J. Spears, R.; McMahon, C.; Chudasama, V. *Chemical Society Reviews* **2021**, *50*, 11098–11155.

(85) Ratner, S.; Clarke, H. T. *Journal of the American Chemical Society* **1937**, *59*, 200–206.

(86) Fang, G.-M.; Wang, J.-X.; Liu, L. *Angewandte Chemie International Edition* **2012**, *51*, 10347–10350.

(87) Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A. *Nucleic Acids Research* **2003**, *31*, 3784–3788.

(88) Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. *Nature Protocols* **2013**, *8*, 2483–2495.

(89) García-Martín, F.; White, P.; Steinauer, R.; Côté, S.; Tulla-Puche, J.; Albericio, F. *Peptide Science* **2006**, *84*, 566–575.

(90) Attard, T. J.; O'Brien-Simpson, N. M.; Reynolds, E. C. *International Journal of Peptide Research and Therapeutics* **2009**, *15*, 69–79.

(91) Ralhan, K.; KrishnaKumar, V. G.; Gupta, S. *RSC Advances* **2015**, *5*, 104417–104425.

(92) King, D. S.; Fields, C. G.; Fields, G. B. *International Journal of Peptide and Protein Research* **1990**, *36*, 255–266.

(93) Provatas, A. A.; Epling, G. A.; Stuart, J. D.; Yeudakimau, A. *Toxicological & Environmental Chemistry* **2014**, *96*, 353–361.

(94) Tamás, B.; Alberts, M.; Laino, T.; Hartrampf, N. *Nature Chemistry* **2026**, *18*, 677–685.

(95) Maity, S. K.; Jbara, M.; Laps, S.; Brik, A. *Angewandte Chemie International Edition* **2016**, *55*, 8108–8112.

(96) Corrêal, D. H. A.; Ramos, C. H. I. *African Journal of Biochemistry Research* **2009**, *3*, 164–173.

(97) *Circular Dichroism and the Conformational Analysis of Biomolecules*; Fasman, G. D., Ed.; Springer US: Boston, MA, 1996.

(98) Vallejo, L. F.; Rinas, U. *Microbial Cell Factories* **2004**, *3*, 11.

(99) Abramson, J. et al. *Nature* **2024**, *630*, 493–500.

(100) Anderton, R. S.; Meloni, B. P.; Mastaglia, F. L.; Boulos, S. *Translational Neuroscience* **2014**, *5*, 8–16.

List of aids

This master thesis was written using artificial intelligence (AI) based auxiliary programs. All programs used are listed in **Table 7.1**. AI was used solely to improve wording, readability, and to support data visualisation performed using Python scripts. All changes suggested by AI were reviewed before being implemented in the work.

Table 7.1. List of AI based tools during the writing process of this thesis

Program	Area of use
deepL	Translation of text passages and improvement of wording
OpenAI ChatGPT	Improvement of wording, support in Python coding for visualisation of HPLC, MS and CD data

List of Figures