universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Semi-synthesis as a tool for the exploration of protein structure and function“

verfasst von / submitted by

Oliver Gajsek BSc

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

Master of Science (MSc)

Wien, 2019 / Vienna 2019

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 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
Mitbetreut von / Co-Supervisor:	Anne Conibear, PhD

Table of contents

List of Abbreviations

ACN	Acetonitrile
Ac_2O	Acetic anhydride
ATP	Adenosine triphosphate
Boc	*tert*-Butyloxycarbonyl
Bpa	4-Benzoylphenylalanine
*t*Bu	*tert*-Butyl
CBD	Chitin binding domain
DCM	Dichloromethane
DIC	*N*,*N*′-Diisopropylcarbodiimide
DMF	Dimethylformamide
DMS	Dimethyl sulfide
EPL	Expressed protein ligation
Gdn·HCl	Guanidine hydrochloride
MS	Mass Spectrometry
MPAA	4-Mercaptophenylacetic acid
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl
RP-HPLC	Reverse phase high pressure liquid chromatography
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TIPS	Triisopropyl silane
Trt	Trityl
TCEP	Tris(2-carboxyethyl)phosphine
TFA	Trifluoroacetic acid
Tris	Tris(hydroxymethyl)-aminomethane

Abstract

It is of fundamental importance for pharmacological studies to understand the structure and the associated functions of proteins, which play a key role in eukaryotic cells. This thesis is divided in two sub-projects focusing on semi-synthetic approaches that combine recombinant DNA technology and solid-phase peptide synthesis (SPPS) for the generation of specifically modified protein domains. Semi-synthetic access to Heat shock protein 90 (Hsp90) and Immunoglobulin M (IgM) will provide new insights into the role of specific modifications and residues of the proteins and can therefore help to understand structure function relations.

Hsp90 is the subject of a variety of modifications after the translation process (post-translation modifications, PTMs) of which their specific impact on the structure and associated protein dynamics is still widely unknown. In this project a synthesis strategy was developed to obtain the specifically post-translationally modified Hsp90 N-terminal domain (NTD) based on the native chemical ligation (NCL) principle. Therefore, 19 amino acid long N-terminal peptide hydrazides, bearing PTMs at specific positions, were synthesised using SPPS and analysed via circular dichroism (CD) and nuclear magnetic resonance (NMR) spectroscopy. The synthetic peptide hydrazides were ligated to the remaining part of the domain, which was generated via recombinant expression, using *in situ* peptide hydrazide to thioester conversion. The recombinant polypeptide carries an N-terminal cysteine required to perform NCL. Once the synthesis is accomplished, including postligation-desulfurization of the cysteine to native alanine and isotope labelling of the recombinant peptide, the influences of the PTMs on the structure and function of Hsp90 NTD will be studied via NMR spectroscopy and functional assays.

The antibody IgM undergoes a complex oligomerisation process, generating its exceptional topology, which is still not fully understood but essential for the protein's activity. The IgM sub-project also included a semi-synthetic approach but focused on the generation of the C-terminal constant heavy domain 4 (CH4) with crosslinkers to define the residues critical for oligomer formation. To this end, a recombinant thioester was prepared via intein chemistry, which was ligated to a 41 amino acid long C-terminal polypeptide bearing a native N-terminal cysteine made by Fmoc-SPPS. The different variants of the synthetic peptide, contained specific hydrophobic amino acid residues, which are suggested by previous studies to significantly contribute to the complex oligomerisation process of IgM and that are exchanged with a photo

activatable crosslinker which allows for studies via crosslinking/mass spectrometry in order to gain a deeper understanding of the assembly of IgM oligomers.

Both projects illustrate the diversity and flexibility of protein semi-synthesis allowing the generation of homogeneous protein domains with specific natural or unnatural modifications in either N- or C-termini.

Zusammenfassung

Die dreidimensionale Struktur und die damit korrelierende Funktion von Proteinen ist von außerordentlichem biochemischem und medizinischem Interesse. Sie ist nicht nur für die gezielte Entwicklung von pharmazeutischen Wirkstoffen, sondern auch für das Verständnis grundlegender, oft auch pathologischer Prozesse, in eukaryotischen Zellen unentbehrlich. Das hier beschrieben Projekt, welches sich in zwei Teile gliedert, zielt auf die Semisynthese spezifisch modifizierter Domänen der Proteine Hsp90 (Hitzeschock Protein 90) und Immunoglobuline M (IgM) ab, um neue Einblicke in deren Dynamiken und Strukturen zu erhalten.

Das, zur Gruppe der Chaperone gehörige, Protein Hsp90 durchläuft eine Reihe von Modifikationen nach der Translation (post-translationale Modifikationen), deren genaue Relevanz in Bezug auf Struktur und Funktion nicht vollkommen bekannt sind. Die hier angewandte Strategie für die Semisynthese der N-terminalen Domäne (NTD) von Hsp90 verwendet die native chemische Ligation (NCL) eines spezifisch PTM modifizierten und mittels Festphasensynthese generierten 19 Aminosäure langen Peptides an ein rekombinant hergestelltes Protein. Die Auswirkungen der eingeführten post-translationalen Modifikationen sollen, nach Isotopenmarkierung des rekombinanten Segments, mittels nuklearer magnetischer Resonanzspektroskopie und anderen Verfahren zur Charakterisierung untersucht werden.

IgM Antikörper weisen in ihrer aktiven, antigenbindenden Form eine äußert komplexe Topologie auf, bestehend aus Penta- oder Hexameren, deren Oligomerisationprozess nicht komplett aufgeklärt ist. Die hier angewandte Synthesestrategie beinhaltet die Ligation eines, mittels Festphasensynthese generierten, 41 Aminosäuren langen C-terminalen Peptides der konstanten Domäne 4 (CH4), an den restlichen, mittels rekombinanter Expression erstellten, Teil der Domäne durch Expressed Protein Ligation (EPL). Spezifische hydrophobe Aminosäuren in der natürlichen Sequenz des synthetisch erstellten Peptides sollen hierbei durch lichtaktivierbare quervernetzende Aminosäuren ersetzt werden, um so neue Erkenntnisse in den Oligomerisierungsprozess des Antikörpers mittels Quervernetzung/Massenspektrometrie zu erhalten.

Die hiermit generierten Protein Domänen sind gute Beispiele für die Vielfältigkeit und Flexibilität der Protein Semi-synthese, welche die Herstellung homogener Proteine mit spezifischen natürlichen oder unnatürlichen Modifikationen an entweder C- und/oder N-Termini erlaubt.

1 Introduction

1.1 Development and applications of protein semi-synthesis

Often considered as “molecules of life”, proteins are complex high molecular weight organic compounds that are a primary constituent of living matter. The desire to understand protein function and structure is not only motivated by a fundamental interest in the way things work in living organisms but can also lead to solutions to a multitude of problems in biology and medicine and is therefore of high interest. Studying natural proteins requires the ability to access and often specifically manipulate proteins with atomic precision, which has been facilitated by advances made in chemical synthesis of proteins and recombinant DNA technology in the last century [1].

Chemical synthesis of peptides and therefore also of proteins can be traced back to the first efforts by Emil Fischer and others in the late 19th century. With the emergence of protection group strategies, particularly the introduction of the first urethane protecting group for the α-amino group by Bergmann & Zervas in 1932, small peptides became accessible by solution-phase synthesis [2,3]. The introduction of solid-phase peptide synthesis (SPPS) by Merrifield in 1963 revolutionised the field and provided the basis for automated SPPS procedures [4]. Whereas the generation of large polypeptides and proteins is usually achieved through biological expression, SPPS remains the most efficient method for the generation of polypeptides up to about 50 amino acid residues. SPPS allows site-selectively introduction of tailored modifications, providing homogenous peptides, e.g. with post-translational modifications of specific residues, which would be challenging for recombinant DNA methods [5]. The restriction of peptide length for the SPPS method originates from its linear nature and the limitations of reaction yields, which are associated with peptide chain aggregation and difficulties in separating the target peptides from accumulated by-products as the chain length increases [6]. However, this limitation led to the development of chemoselective ligations methods that convergently assemble smaller more manageable peptide fragments in aqueous media, to generate larger peptides and small proteins by chemical synthesis [7]. These methods evolved as the originally developed approaches based on the condensation of fragments with protected sidechains generated too many problems based on peptide solubility. The use of unprotected peptide fragments that can be linked in aqueous media made numerous peptide and protein targets accessible for the first time.

The introduction of native chemical ligation (NCL) by Kent and co-workers in 1994 is a key in the advances of ligation methods and represents the most widely used method today [8]. The reaction, which requires a C-terminal thioester peptide and an N- terminal cysteine peptide as coupling partners, first involves a reversible trans-thioesterfication initiated by a nucleophilic attack of the cysteine sulfhydryl moiety on the thioester, resulting in a thioester linkage, followed by a rapid and irreversible intramolecular S→N acyl shift to generate a native amide bond (Figure 1). The fact that the reaction can be implemented in aqueous media at neutral pH using completely unprotected peptides makes it a powerful tool for the total synthesis of proteins and paved the way for interfacing it with recombinantly produced proteins in a procedure termed protein semi-synthesis [9,6,10].

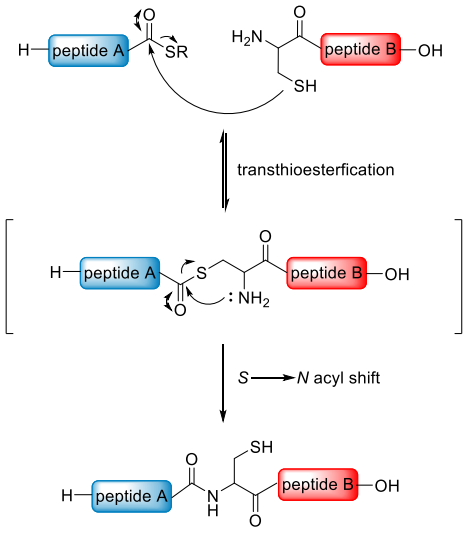


Figure 1: Mechanism of the native chemical ligation reaction, involving the nucleophilic attack on a C-terminal thioester peptide by the sulfhydryl group of an N-terminal cysteine peptide, followed by an intramolecular S→N acyl shift which results in a native amide bond at the ligation site.

In protein semi-synthesis, a synthetic peptide bearing a C-terminal thioester is ligated to a recombinantly expressed protein bearing an N-terminal cysteine (Figure 2B), which makes large proteins with site-specific N-terminal modifications accessible. This concept was further expanded by the introduction of an expression system which makes the thioester moiety on recombinant proteins accessible and was later termed Expressed Protein Ligation (EPL) (Figure 2A) [11]. The method is based on the posttranslationally autocatalytic cleavage of a protein domain named inteins in a process termed protein splicing [12]. Similar to RNA self-slicing, protein splicing is an autocatalytic process that involves an intervening protein domain (intein) which removes itself from a polypeptide, simultaneously creating a new peptide bond between its two flanking regions called exteins [13, 14]. Since an α-thioester was identified as a crucial

intermediate during this process, inteins were specifically engineered that allow the thiolysis of the intermediate and therefore generate the desired recombinant C-terminal thioester protein segment. In practical terms, this is achieved by the expression of the polypeptide of interest as a fusion protein in which its C-terminus is attached to an intein-affinity tag domain. The expressed fusion protein is purified using the used affinity tag, usually a chitin binding domain (CBD) or a hexahistidine tag(His_6), and the desired recombinant thioester is obtained by treating the fusion protein with an appropriate thiol (e.g. sodium 2-mercaptoethanesulfonate), either immobilized on the affinity beads or in solution after the elution (Figure 2A) [11].

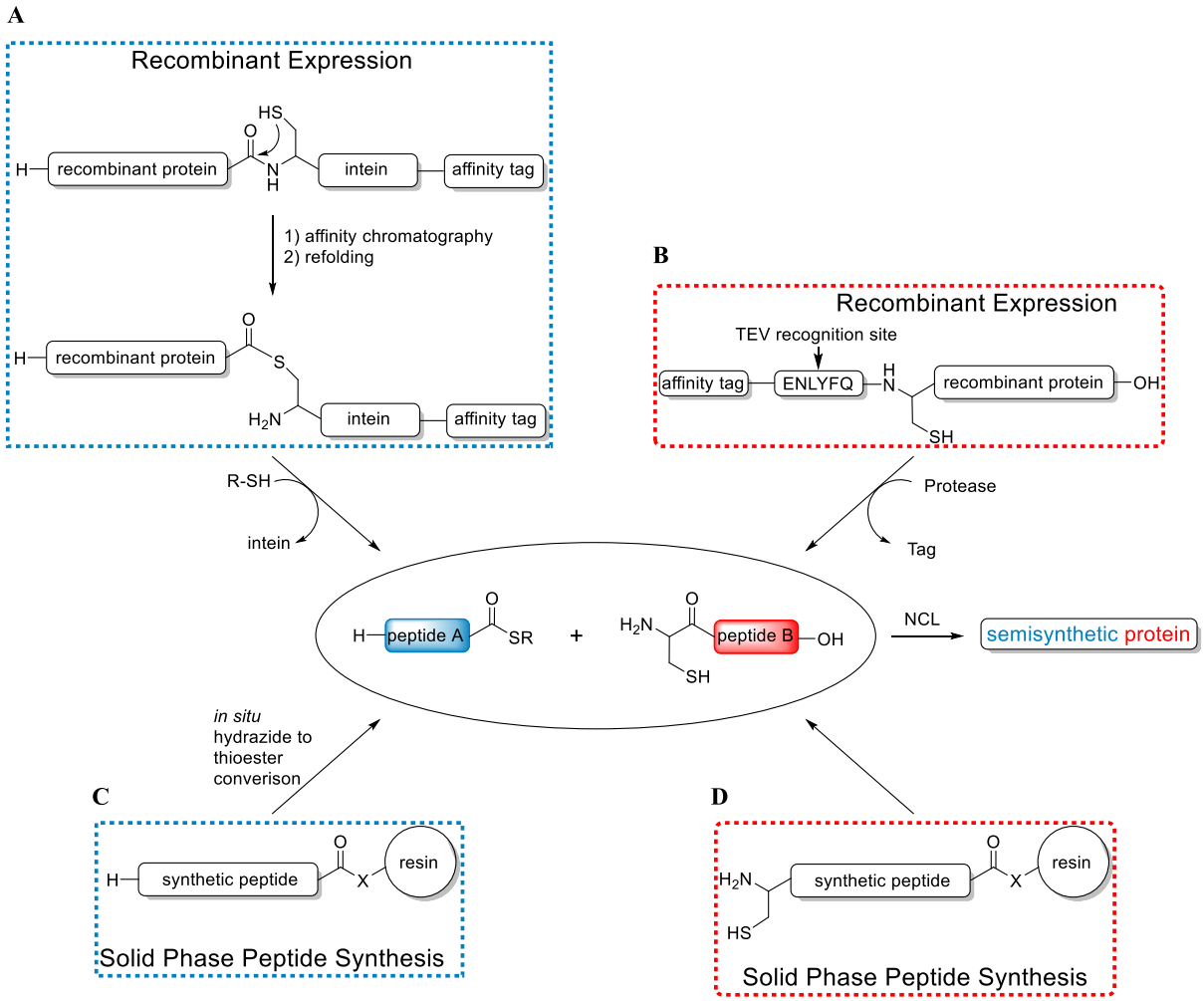


Figure 2: Pathways for the semi-synthesis of proteins by Expressed Protein Ligation (EPL). The method is based on the ligation of a recombinant protein to a synthetic peptide via native chemical ligation. The necessary α-thioester can be generated either by expression of the protein segment fused to an intein (**A**) or by synthesis (**C**) using, e.g. a hydrazide to thioester conversion protocol (X= NH−NH). The N-terminal cysteine ligation partner can be expressed (**B**) or synthesised (**D**) via standard Fmoc-based SPPS (X= appropriate linker).

A remaining challenge that needed to be overcome was the requirement of a cysteine residue at the ligation site, since this limited the possible disconnection sites severely, especially considering the relative scarcity of cysteine (1.8% abundance) in proteins. A solution to this problem was the introduction of post-ligation, desulfurization protocols initially using Raney Ni or Pd on Al_2O_3 as published by Yan and Dawson, which allowed the selective reduction of the cysteine sulfhydryl group [15]. This permitted the generation of the more abundant alanine (8.9% abundance) and thus substantially expanded the possible ligations sites by using cysteine as a surrogate for a native occurring alanine [16]. The concept was later expanded by a milder, metal free desulfurization based on using the radical initiator 2,2´-azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) in combination with tris(2-carboxyethyl)phosphine (TCEP) [17]. This concept also lead to the development of other strategies which involve thiol-derived variants of canonical amino acids as cysteine surrogates in NCL followed by post-ligation desulfurization to native residues [6]. Together with the continuous development of strategies for the generation of synthetic peptide α-thioesters based on SPPS, such as the *in situ* hydrazide to thioester conversion shown in Figure 2C, this created a plethora of possibilities for the synthesis of specifically modified (semi-synthetic) proteins [9,18].

Since the very labile thioester-bond cannot withstand the basic conditions of standard Fmoc-based SPPS, the introduction of new strategies for the generation of synthetic thioester peptides was required. The applied synthesis strategy for the generation of posttranslationally modified peptides bearing the thioester moiety is based on the published protocol by Zheng *et al.* and includes the synthesis of the peptides as peptide hydrazides via the attachment of hydrazine to the 2-Cl-Trt resin before the standard Fmoc-based SPPS (Figure 3) [18]. The cleaved peptide hydrazide can then be readily converted *in situ* by $NaNO_2$ activation and ensuing thiolysis of the azide intermediate in order to receive the thioester for native chemical ligation.

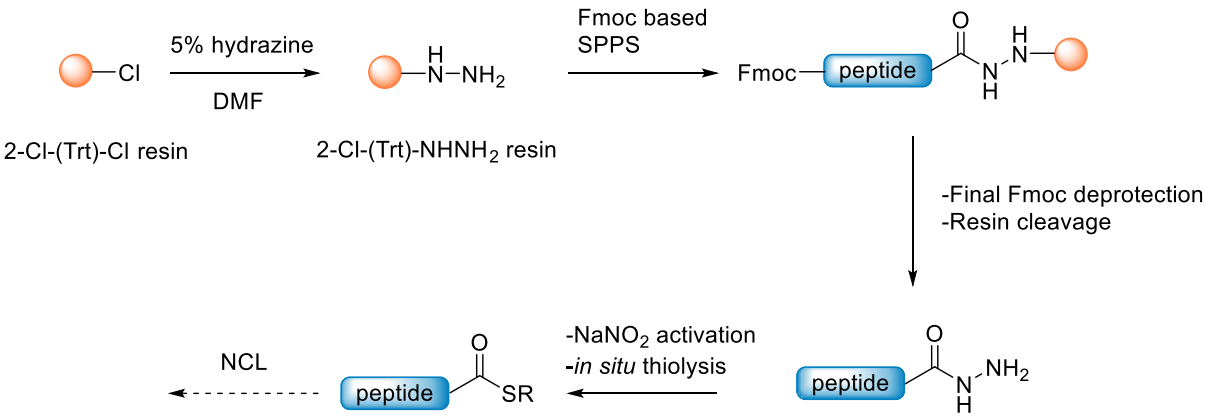


Figure 3: Synthesis strategy for preparation of modified α-thioester peptides according to the protocol of Zheng *et al.* [18]. The desired PTMs can be introduced via the use of modified amino acids during standard Fmoc-based SPPS.

In the herein described projects all of the above mentioned EPL synthesis strategies (Figure 2) were applied. For the generation of semi-synthetic Hsp90 N-terminal domain, modified peptide hydrazides were synthesised and subsequently ligated to a recombinantly expressed polypeptide bearing an N-terminal cysteine using *in situ* hydrazide to thioester conversion (Figure 2B and C). The other project involved the generation of a semi-synthetic C-terminal constant domain of IgM (CH4). Therefore, an intein-fusion protein was expressed and converted to a thioester before ligation to a synthetic peptide bearing an N-terminal cysteine (Figure 2A and D). Besides the introduction of C- or N-terminal PTMs or unnatural amino acids, the methods of Expressed Protein Ligation also allow for isotope labelling of protein domains within large proteins which can be used for structural studies using nuclear magnetic resonance (NMR) spectroscopy [19]. This is achieved by expressing the recombinant part of the semi-synthetic protein in ^{13}C and/or ^{15}N labelled media, which forces the organisms of expression to include the labelled isotopes in the recombinant protein.

The following project utilizes the above described synthesis strategies for the generation of specifically modified semi-synthetic proteins in order to get new insights into their structure and function via nuclear magnetic resonance spectroscopy and crosslinking-mass spectrometry. A fundamental step during the strategic planning of a semi-synthetic pathway is the choice of the ligation site and thereby defining the peptide fragments that need to be prepared for the ligation reaction. A detailed explanation of the synthesis strategy with respect to the chosen synthetic and recombinant peptide fragments and the selected ligation sites is provided in sections 3.1 and 3.2. The following sections (1.2 and 1.3) introduce the two proteins Hsp90 and IgM and highlight the motivations behind further studies of the proteins.

1.2 Function and structure of the molecular chaperone Hsp90

Heat shock proteins (Hsp) belong to the most highly expressed cellular proteins and play a pivotal role in retaining cellular function by providing assistance during the folding, transport, maintenance and degradation of proteins under stressed and normal conditions [20]. The 90 kDa molecular chaperone heat shock protein 90 (Hsp90) is one of the most abundant heat-shock proteins and is involved in the processing of a wide variety of proteins including transcription factors, protein kinases and E3 ubiquitin ligases [21]. Several hundred of these proteins, called "clients", are known today with most of them being involved in signal transduction pathways. However, among these the function and stability of many oncogenic mutant proteins also depend on the Hsp90 machinery [22]. This has driven the exploration of Hsp90 as an anticancer target for the last two decades and therefore asks for a deeper understanding of its complex ATPase conformation cycle, its co-chaperones and post-translation modifications [23].

The domain structure of Hsp90 consists of three major domains. The β-strand rich N-terminal domain (NTD) bears a nucleotide binding pocket and a lid region that closes upon ATP binding. The NTD is connected via a flexible linker to the to the middle domain (MD), which carries the binding sites for client proteins and co-chaperones. The MD is directly connected to the C-terminal domain (CTD), which allows for the dimerization of the protein, generating the "V-shaped" active homodimer (Figure 4A) [24]. Additionally, the CTD contains a carboxy terminal MEED motif, which is only present in cytosolic paralogs. This motif mediates binding to a specific group of so called "co-chaperones" which play an important role during the chaperone cycle. Hsp90 undergoes a complex conformational cycle that starts with the binding of ATP in the open "V-shaped" conformation followed by the closing of the nucleotide lid region and a large structural rearrangement that leads to the dimerization of the N-terminal domain (closed state 1) (Figure 4B) [24]. This is followed by a side swap of the NTD segments which leads to the compression of the MD domains and the subsequent ATP hydrolysis. The release of ADP and phosphate and the simultaneous dissociation of the NTDs mark the completion the chaperone cycle. During this cycle, more than 20 different co-chaperones have been identified for the eukaryotic system that associate during specific conformational states with Hsp90 and regulate its function in different ways, e.g. by inhibition or activation of its ATPase activity or by recruiting specific client proteins [25].

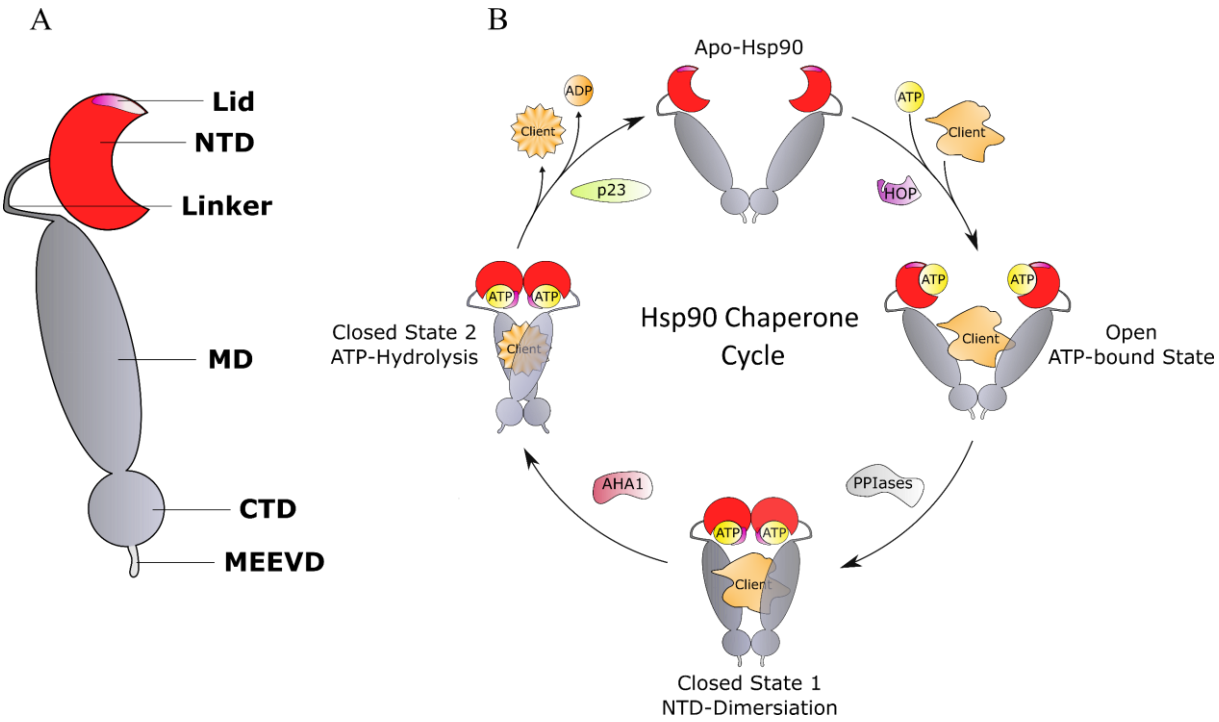


Figure 4: A) Domain structure of the monomeric Hsp90 consisting of the N-terminal domain (NTD), a flexible linker which connects the NTD to the client binding middle domain (MD), and the dimerising carboxy-terminal domain (CTD) including the MEEVD co-chaperone binding motif. **B)** Simplified model for the Hsp90 chaperone cycle adapted from Schopf *et al.* [24]. The cycle initiates with the binding of ATP and the client protein to the "V-shaped" homodimer in an open conformation followed by the dimerization of the NTD. The following large conformational changes of the MD assist the client protein during its folding process and the cycle is finished with the release of ADP, phosphate and the client as well as the dissociation of the NTD. HOP, PPIase, AHA1 and p23 represent important co-chaperones that bind during the different cycle steps.

The human genome encodes for two major isoforms of the protein, the constitutively expressed Hsp90β and the heat-induced Hsp90α, which share ~85% sequence identity [26]. Although the general mechanism of action of the Hsp90 ATPase cycle is applicable to all Hsp90 paralogs and isoforms, isoform-specific differences exist and it is assumed that some clients preferentially bind to one of the isoforms [27]. Another key factor in the regulation of Hsp90 besides co-chaperones are post-translational modifications (PTMs) including phosphorylation, acetylation, S-nitrosylation, and ubiquitination [28]. Although detailed influences of single PTMs are not understood yet, earlier work shows the dynamics of the regulation of Hsp90 by PTMs like phosphorylation: hyperphosphorylation negatively influences the activity of Hsp90, causing a weaker association with client proteins, while it was also reported that phosphorylation positively affects the folding of some clients [29,30].

This project specifically focuses on the highly conserved amino-terminal domain (NTD) of human Hsp90α. The domain forms a nucleotide-binding pocket, a so called "Bergarat fold", which is responsible for the binding of ATP. Upon binding of ATP, a lid region closes over the bound ATP resulting in an intermediate state. The "split ATPase" nature of the domain makes further rearrangements necessary which triggers the reposition of the NTD and yields a dimerized state of the domain with swapped segments in the NTDs and a compression of the MDs [21]. The 27 kDa NTD undergoes a variety of PTMs including the phosphorylation of specific serine, threonine and tyrosine residues as well as four known positions of lysin acetylation [28]. Previous work has shown that the phosphorylation of this domain by DNA-activated protein kinase (DNA-PK), casein kinase 2 (CK2) and the kinase Wee1 can provide a regulatory tool for the chaperoning of different classes of client proteins by Hsp90 [28]. For example, Mollapour *et al.* showed that CK2-dependent phosphorylation of a conserved threonine residue (T22) in yeast Hsp90 (yHsp90) NTD significantly reduced its ATPase activity and the chaperoning of several kinase clients. In another study by the same author, Swe1 (yeast Wee1) dependent phosphorylation of a conserved tyrosine residue (Y24) triggered cytosolic relocation and proteasomal degradation of phosphotyrosyl yHsp90 [31,32]. These examples show the complexity of the role of PTMs and demonstrate the importance of further studies to fully understand the significant impact of these modifications.

The semi-synthesis of a selectively posttranslationally modified and isotope labelled Hsp90 N-terminal domain should help to get new insights into the influence and significance of specific modified residues on the structure and function of Hsp90.

1.3 Function and structure of the polymeric antibody IgM

The multivalent polymeric antibody immunoglobulin M (IgM) represents the "first-responder" in the defence against microbial infections and other pathogens and is found in all vertebrates that have been studied [33]. Of the five major isoforms of immunoglobulins produced by vertebrates, named IgM, IgD, IgG, IgA and IgE, immunoglobulin M represents the largest class by molar mass and was therefore initially named γ-macroglobulin (M- for 'macro') [33]. IgM antibodies classified into two classes: The first is predominantly produced by B1 cells prior to the infection and is denoted as natural IgM; The second class, which is secreted early after antigen exposure, is denoted as immune IgM [34,35]. These antibodies play an important and complex role in the immune response, in a direct way via the activation of the compliment pathway and indirectly as regulatory tools [33].

Due to its multimeric structure, natural and immune IgM represents an extremely strong activator of the complement pathway and enables efficient binding of antigens and complement receptors. However, the complex topology of polymeric IgM and the associated assembly process of the polymeric antibody is still not fully understood and deservers further attention. In the presence of the joining (J) chain protein, the antibody forms pentameric structures composed of five monomeric subunits, where each subunit is composed of two heterodimers, sometimes also referred as 'halfmeres' connected via disulphide bridges (Figure 5A and D) [36,37]. The multimeric structure is stabilized by the availability of three different cysteine residues (C337 Cμ2, C414 Cμ3, and C575 Cμ4) in the μ-heavy chain of IgM, which allow for inter-μ chain disulphide bonds as well as the binding to the J chain protein. However in the absence of the J chain the formation of hexameric IgM is observed, which was reported to be up to 20 times more active than pentameric IgM in initiating the complement pathways (Figure 5B) [38]. The IgM heterodimers contain, similar common antibodies, a light chain (L), and a heavy chain (μ or H chain) connected via a disulphide bonds [39]. The light chain contains a variable region (VL) and a constant region (CL) whereas the heavy chain is separated into four constant regions (Cμ1, Cμ2, Cμ3 and Cμ4) and a variable region (VH) (Figure 5C).

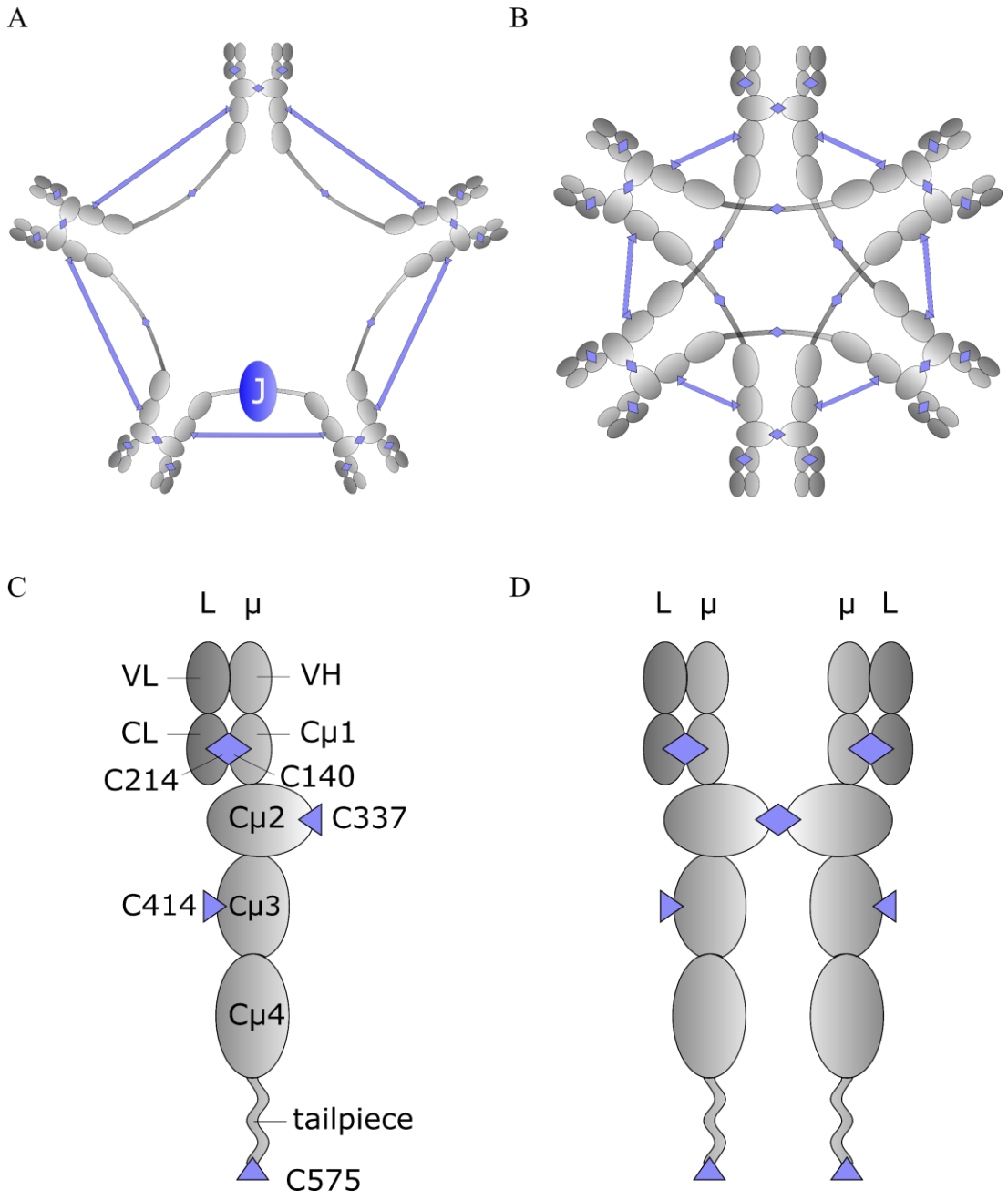


Figure 5: Schematic models of monomeric and polymeric IgM. Cysteine residues are labelled with the corresponding positions (e.g. C575) and displayed as blue triangles. Disulphide bonds in the heterodimer and monomeric models are depicted as blue rhombuses or as long blue double arrowheads between monomeric subunits. **A)** Pentameric IgM including the J chain protein. **B)** Hexameric structure of IgM. **C)** Structure of the IgM μL heterodimer (halfmer) with labelled constant (Cμ1-Cμ4, CL) and variable regions (VH, VL) as well as the tailpiece (tp). **D)** Monomeric IgM (μL)2.

Of specific interest for this project is a 18 amino acid long C-terminal peptide extension of the Cμ4 domain of the heavy chain, referred to as the tailpiece (tp), which plays a pivotal role during the assembly process of the polymeric antibody and is able to direct the formation of the

hexameric IgM complex [40]. Alanine scanning (systematic, sequential exchange of each single amino acid to alanine by site-directed mutagenesis) of the tp revealed the specific importance of four hydrophobic residues (Y562, V564, L566, and I567) for the assembly of the polymeric antibody [41]. When mutated to Ala the resulting proteins were exclusively secreted as μ_2-L_2 monomers and did not form any complexes of higher molecular weight indicating the importance of hydrophobic interactions between the monomers during the assembly process. In order to get new insights into the assembly process of the polymeric antibody, a strategy for the semi-synthesis of the Cμ4 domain where these specific hydrophobic residues are exchanged by photoactivatable crosslinkers was developed and is described in section 3.2.

1.4 Objectives

This thesis describes semi-synthetic approaches for the generation of specifically modified domains of Hsp90 and IgM in order gain new insights into the structure and function of these proteins. Both synthesis strategies developed here are based on native chemical ligation (NCL), involving the ligation of a smaller modified synthetic segment of the protein to a larger recombinant protein segment, but differ in the generation of the pivotal α-thioester and cysteine peptide. The applied synthesis strategies illustrate the flexibility and applications of NCL for the generation of homogeneous semi-synthetic proteins with natural or unnatural modifications in either N- or C-termini.

In the case of Hsp90, the N-terminal domain (NTD) and the influences of post-translational modifications (PTMs) on its function and structure are of interest. Therefore, a strategy was developed which allowed for the generation of modified synthetic α-thioester peptides and the recombinant expression of the remaining protein segment bearing an N-terminal cysteine (Figure 2B and C). The overall impact of the modifications on the semi-synthetic domain are meant to be studied via NMR spectroscopy.

In contrast, the strategy for the semi-synthesis of the heavy constant domain (Cμ4 or CH4) of IgM comprised the synthesis of modified peptides bearing an N-terminal cysteine and the recombinant expression of the remaining protein segment as an intein fusion protein that yields the C-terminal thioester (Figure 2A and D). Specific hydrophobic residues of the C-terminal synthetic peptide, that are known to play an important role during the assembly of the polymeric antibody, were replaced with a photocrosslinker which allows for covalent linkage to amino acid residues in spatial proximity. The semi-synthetic domain containing the crosslinker modification is therefore meant to be studied via mass spectrometry after the oligomerisation and photocrosslinking in order to gain new insights into the assembly of this important antibody.

2 Materials and Methods

2.1 Chemicals

All solvents and chemicals were purchased from commercial suppliers and used without further purification unless otherwise noted. Standard Fmoc- protected amino acids were from either Novabiochem (Merck KGa, Darmstadt, Germany) or Iris Biotech GmbH (Marktredwitz, Germany), Fmoc-*p*-Bz-Phe-OH was from Bachem Holding AG (4416 Bubendorf Switzerland). Resins for solid phase peptide synthesis were ordered from Novabiochem [2-Cl-(Trt)-Cl] or CEM [Cl-TCP(Cl) ProTide]. Activating agents (HATU, HBTU) and DIPEA were from Novabiochem, Oxyma and DIC from Iris Biotech GmbH. Solvents (DCM, DMF) and peptide cleavage chemicals (DMS, TIPS, TFA) as well as additives used for ligation reactions (MPAA, TCEP, MesNa, thiophenol) were from Sigma-Aldrich. Buffers were prepared with water purified from a Millipore MilliQ Reference A+ water purification system.

2.2 Peptide Synthesis

2.2.1 General procedures

2.2.1.1 Manual solid phase peptide synthesis

Manual solid-phase peptide synthesis (SPPS) was carried out in 10 or 20 mL polypropylene syringes equipped with polypropylene frits according to the Fmoc strategy. Fluorenylmethoxycarbonyl (Fmoc) protected L- amino acids and standard side-chain protecting groups [Arg(Pbf), Asn(Trt), Asp(*t*Bu), Cys(Trt), Gln(Trt), Glu(*t*Bu), His(Trt), Lys(Boc), Ser(*t*Bu), Thr(*t*Bu), Trp(Boc) and Tyr(*t*Bu)] were used. Fmoc-L-*p*-benzoylphenylalanine (Fmoc-p-Bz-Phe-OH) was used for the introduction of the benzophenone functionality and Fmoc-O-benzylphospho-L-threonine for the introduction of phosphothreonine. Amino acid residues (4 eq.) were coupled with 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU, 3.9 eq., 0.5 M in DMF) and diisopropylethylamine (DIPEA, 8 eq.) for 25 – 30 min, if not otherwise stated. Subsequent Fmoc removal was performed with 20% piperidine in DMF (v/v). After the final deprotection the resin was washed with DMF and DCM and dried under vacuum overnight. Peptide cleavage was performed using TFA/TIPS/DMS/H_2O (90% : 2.5% : 2.5% : 5%, v/v/v/v) for 2-3 h (5x resin volume) and the cleaved peptide precipitated by the addition cold diethylether (8x cleavage solution volume).

Precipitated peptides were pelleted by centrifugation (4500 rpm, 5 min, 4°C) and dried under ambient conditions before dissolving in 50% water/acetonitrile (v/v, + 0.1% TFA) and lyophilizing.

2.2.1.2 Automated peptide synthesis

Automated peptide synthesis was carried out on a CEM Liberty Blue Automated Microwave Peptide Synthesiser using Cl-TCP(Cl) ProTide resin with a loading of 1.51 mmol/g for CH4 92-131 peptides (**M1**, **M2**) and 2-Cl-(Trt)-NHNH-F resin with a determined loading of 0.54 mmol/g for the Hsp90 peptide **H2** via Fmoc strategy. Fmoc protected L- amino acids and standard side-chain protecting groups [Arg(Pbf), Asn(Trt), Asp(tBu), Cys(Trt), Gln(Trt), Glu(tBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc) and Tyr(tBu)] were used. Double couplings were performed for each residue after the coupling of proline and after ten coupled residues. Additionally, the pseudoproline dipeptide Fmoc-Lys(Boc)-Ser(Ψ(Me,Me)pro)-OH was incorporated in position 11 of CH4 peptides **M1** and **M2** to reduce aggregation and maximize synthesis quality. Amino acids (5 eq., 0.2M) were dissolved in DMF and coupling steps performed under microwave irradiation (74°C) using 0.2M DIPEA in 0.5M Ethyl (2Z)-2-cyano-2-hydroxyiminoacetate (Oxyma) and 0.25M *N,N′*-Diisopropylcarbodiimide (DIC). Fmoc removal was performed using 10% piperazine in NMP/EtOH (90:10, v/v) for peptides **M1**, **M2** and 20% piperidine in DMF (v/v) for peptide **H2.** Peptide cleavage was carried out as described in section 2.2.1.1, if not otherwise stated.

2.2.1.3 Peptide purification and analysis

Crude peptides were dissolved in either 10 mL of H_2O or 10 mL 6 M Gdn·HCl (pH 4.7) and filtered through a sterile 0.22 μm filter before purification by RP-HPLC on Waters AutoPurification HPLC/MS System using a Kromasil C18 column (300-10-C18, 250 x 21.2 mm, 10 μm particle size). Peptides were eluted using a gradient of 5-45% solvent B (CAN + 0.05% TFA) in solvent A (ddH_2O + 0.05% TFA) in 40 minutes with a flow rate of 20 mL/min. Fractions were analysed by electrospray ionisation mass spectrometry (ESI-MS), pooled according to their purity and lyophilized. Purified peptides were characterized using ESI-MS in positive ion operation mode on Waters AutoPurification HPLC/MS System and analytical RP-HPLC analysis on a Dionex Ultimate 3000 instrument using a gradient of 5-65% solvent B (CAN + 0.08% TFA) in solvent A (ddH_2O + 0.1% TFA), a Kromasil C4 column (300-5-C4, 150 x 4.6 mm, 5 μm particle size) and flow rate of 1 mL/min over 30 minutes. Chromatograms were recorded over 45 minutes measuring the absorbance at 214 and 280 nm.

2.2.2 Synthesis of Hsp90 2-20 peptide hydrazides

2.2.2.1 Preparation of 2-Cl-(Trt)-$NHNH_2$ resin and determination of loading

2-Cl-(Trt)-$NHNH_2$ resin was prepared according to the protocol of Zheng *et al.* [18]. 500 mg of 2-Cl-(Trt)-Cl resin with a loading of 1.51 mmol/g was washed with DMF, DCM and again DMF and subsequently swelled for 30 min in 4.5 mL 50% DMF/DCM (v/v). After the addition of 4.4 mL of 10% NH_2-NH_2 in 50% DMF/DCM (v/v) the resin was agitated for 30 min at room temperature, washed with DMF and the coupling step with 4.4 mL 10% NH_2-NH_2 repeated. The resin was washed with DMF and DCM and unreacted resin capped by the addition of 4.4 mL 5% MeOH/DMF (v/v). The resin was washed once more with DMF and DCM and the first amino acid residue (Fmoc-Phe-OH) coupled manually assuming a loading of 1 mmol/g using double coupling and the conditions mentioned in the general procedures section. Loading of 2-Cl-(Trt)-NHNH-F resin was determined by stirring 6 mg resin with 4 mL of 20 % piperidine in DMF (v/v) for 30 min prior to measuring the UV absorbance of the cleaved dibenzofulvene–piperidine adduct in the deprotection solution. The following formula was used to obtain the loading from the measured absorbance of the deprotection solution:

$$Loading = \frac{Abs_{301\,nm} \cdot volume\ \ 20\%\ piperidine\ in\ DMF\ (mL)}{extinction\ coefficient \cdot mass\ of\ resin\ (mg)}$$

Extinction coefficient dibenzofulvene-piperidine adduct: 7.8

In order to be in the linear absorbance range, the deprotection solution was diluted 1:10 with 20% piperidine in DMF (v/v) before measuring the absorbance and 20% piperidine in DMF (v/v) used as a blank. Fmoc- cleavage and UV absorption measurement was repeated three times and the average used to determine the loading.

2.2.2.2 Synthesis of Hsp90 2-20 peptide hydrazides H1, H2, H3 and H4

Synthesis of phosphothreonine modified Hsp90 2-20 peptides **H1**, **H2**, **H3** and unmodified peptide **H4** were performed on a 0.05 mmole scale on preloaded 2-Cl-(Trt)-NHNH-F resin (for the resin preparation see section 2.2.2.1). All peptides were synthesized manually except for **H2**, which was synthesized on CEM Liberty Blue Automated Microwave Peptide Synthesiser up to glutamic acid (Q) at position 6 and finished manually using the SPPS procedure described in section 2.2.1.1. Double couplings were performed after proline (P), glutamine (Q) and valine (V) residues.

2.2.2 Synthesis of IgM CH4 92-132 peptides

Synthesis of peptide **M1** with the benzophenone functionality in position 120 was performed on a 0.15 mmole scale and synthesis of peptide **M2,** with the benzophenone functionality in position 118 on a 0.05 mmole scale. Automated peptide synthesis as described in section 2.2.1.1 was used up to the position of the benzophenone moiety (Bpa), which itself and the following amino acid residue was introduced by manual SPPS. During and after the incorporation of the benzophenone moiety the resin was protected from light as far as possible by coverage of the reaction vessel with aluminium foil. For the introduction of the benzophenone amino acid residue and the following amino acid residue (Fmoc-N-OH) in peptide **M1**, 1-[bis(Dimethylamin)methylen]-1H-1,2,3-triazol[4,5-b]pyridinium-3-oxid-hexafluorophosphate (HATU, 2.4 eq., 0.5M in DMF), DIPEA (5 eq.), double coupling with a coupling time of 1 h and 2.5 eq. of the amino acid (Fmoc-Bpa-OH) was used. For the introduction of the benzophenone functionality and the following amino acid residue (Fmoc-L-OH) in peptide **M2**, Ethyl (2Z)-2-cyano-2-hydroxyiminoacetate (Oxyma, 4 eq., 0.5 M in DMF), DIC (4 eq.), double coupling with a coupling time of 30 min and 4 eq. amino acid were used. Subsequently, unreacted amines were capped in Ac_2O/DIPEA/DCM (10%: 5%: 85%, v,v,v) for 15 min and peptide synthesis completed using automated peptide synthesis as described in section 2.2.1.2.

2.3 Protein expression and purification

2.3.1 Transformation, expression and purification of Hsp90 constructs

2.3.1.1 General Procedures

Human Hsp90-alpha (UniProt entry: P07900; isoform 2) N-terminal domain (NTD) constructs were ordered in a pET-21a+ bacterial expression vector from BioCat GmbH (69120 Heidelberg, Germany) and transformed into *E. coli* BL21(DE3), XL1 (Blue) and Rosetta 2 (DE3) competent cells via heat shock method described in section 2.3.1.2. Codon optimisation for *E. coli* was performed using the free online tool from Integrated DNA Technologies, Inc. (Iowa 52241, USA). Small scale test expressions (200 mL LB medium) were performed to identify optimal expression conditions and strains. Solely *E. coli* BL21 strains were used for large scale expressions. Large scale expressions were performed by inoculating 2 L of LB medium containing 100 µg/mL ampicillin with 50-80 mL overnight culture (starting OD_{600} ~0.2) and expression induced with 0.5 mM isopropyl-*β*-D-1-thiogalactopyranoside (IPTG) at an OD_{600} ~0.8 for 4 h at 30°C and 180 rpm. Subsequently, cells were harvested by centrifugation (6000

rpm, 10°C, 20 min) and the cell pellets resuspended for 45 min – 1 h in either TBS (150 mM NaCl, 50 mM Tris, pH 7.5) or PBS buffer (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) on ice. The resuspended cell solutions were homogenized using an ultra turrax (IKA GmbH , 79219 Staufen, Germany) and cells lysed by disruption (Constant Systems Limited, Northants United Kingdom) at 1.93 kbar. The cell lysate was cleared by centrifugation (24000 rpm, 45 min, 10°C) and samples from the supernatant and pellet taken to identify if the protein was expressed as inclusion bodies (IB) or in the soluble fraction via SDS-PAGE. Ensuing purification was performed by Ni-NTA affinity chromatography using a 5 mL HisTrap™ High Performance column (GE Healthcare) on an ÄKTAprime plus (GE Healthcare) FPLC system at room temperature.

2.3.1.2 Transformation of Hsp90 NTD constructs

Plasmids for Hsp90 NTD constructs, His_6-TEV-21-235 ($Hsp90^{C1}$) and 1-20-TEV-21-235-His_6 ($Hsp90^{C2}$), both bearing Cysteine at position 21, as well as wild-type Hsp90 NTD 1-235-His_6 ($Hsp90^{WT}$) were dissolved in 25 µL of ddH_2O and 100 µL aliquots of competent cells incubated with 5 µL (BL21, Rosetta 2) or 3 µL (XL1) plasmid solution. Subsequently, cells were incubated on ice for 10 min before exposing to heat-shock for 90 seconds at 42°C. After an incubation on ice for 5 min, 1 mL of sterile LB media was added to the cells before incubating at 37°C for 45 min with gentle shaking. Cells were pelleted via centrifugation and the supernatant removed leaving 50 µL for resuspension. The resuspended cell solutions were spread over freshly prepared agar plates containing the corresponding antibiotics under semi-sterile conditions and the plates incubated at 37°C overnight under humid conditions. Single colonies were picked and added to 4 mL sterile LB medium containing either 100 µg/mL ampicillin (BL21, XL1) or 100 µg/mL ampicillin + 30 µg/mL chloramphenicol (Rosetta 2). Cells were grown at 37°C overnight at 200 rpm and the cultures used to prepare 20% glycerol stocks which were stored at -80°C.

2.3.1.3 Expression and purification of protein constructs $Hsp90^{C1}$ and $Hsp90^{C2}$

Large scale expression of constructs $Hsp90^{C1}$ and $Hsp90^{C2}$ followed by cell lysis was performed as described in section 2.3.1. Inclusion body pellets were solubilized in 6M Gdn·HCl, TBS pH 7.5 overnight and subsequently centrifuged at 24000 rpm for 1 h at 10 °C. The supernatant was loaded onto a Ni-NTA affinity column, washed with 25 mM imidazole, 6M Gdn·HCl, TBS pH 7.5 and the target protein eluted using a linear gradient from 25 – 500 mM imidazole, 6M

Gdn·HCl, TBS, pH 7.5 in 10 column volumes at a flow rate of 5 mL/min ($Hsp90^{C1}$) or 1 mL/min ($Hsp90^{C2}$). Peak fractions were pooled and submitted for LC-MS analysis.

2.3.1.4 Expression and Purification of wild-type Hsp90 NTD ($Hsp90^{WT}$)

Large Scale expression of wild-type Hsp90 NTD was performed as described in section 2.3.1.1. Supernatant of the cleared cell lysate was loaded onto a Ni-NTA affinity column for purification and the protein was eluted using a linear gradient from 25 - 500 mM imidazole, PBS, pH 7.4 in 10 column volumes at a flow rate of 1 mL/min. Peak fractions were pooled and desalted by dialysis against TBS pH 7.5 at 4°C. The resultant pure wild-type Hsp90 NTD was submitted for LC-MS analysis, shock frozen in liquid nitrogen and stored at -80°C.

2.3.1.5 Refolding and TEV-Cleavage of $Hsp90^{C2}$

Ni-NTA purified peak fractions were dialysed against 6M Urea, TBS pH 7 at 4°C overnight and subsequently desalted by dialysis against TBS, pH 7.5 at 4°C. Segment 1-20-TEV-tag was removed by TEV protease cleavage at 4°C overnight using a protein:TEV ratio of 3:1 (w/w) and 1 mM 1,4-Dithio-D-threitol. The cleaved protein was separated by RP-HPLC purification using a Kromasil C4 column (300-10-C4, 250 x 21.2 mm, 10 μm particle size) and a gradient of 5-55% solvent B (ACN + 0.08% TFA) in solvent A (ddH_2O + 0.1% TFA) in 30 min at a flow rate of 20 mL/min. Fractions containing the desired cleaved protein were pooled, lyophilized and characterised via ESI-MS.

2.3.2 Expression, purification and intein cleavage of IgM fusion protein

2.3.2.1 Expression and purification of IgM CH4 1-91-*Mxe*-His_7-CBD

The *E. coli* BL21(DE3) glycerol stock of the competent cells for the expression of IgM CH4 1-91 fusion protein and the established protocol used for the expression, purification and intein cleavage of the protein was kindly provided by Dr. Alexander Kravchuk. *Mus muculus* (mouse) IgM heavy constant domain CH4 1-91 (UniProt entry: P01872) was cloned into an pTXB1 expression vector including a C-terminal *Mxe* GyrA-His_7-CBD (CBD, chitin binding domain) fusion intein. Large scale expressions (2 L) were performed by inoculating 2 L of LB medium containing 100 μg/mL ampicillin with 50 mL overnight culture, followed by growth at 37°C to an OD_{600} of ~0.8. Protein expression was induced with 1 mM IPTG at 37°C overnight at 180 rpm and cells collected by centrifugation (8000 rpm, 10°C, 15 min). Cell pellets were resuspended by stirring in PBS buffer on ice for 45 min and the solution homogenized using an ultra turrax homogenizer. Cell were lysed by disruption (3 cycles, 1.93 kbar) and the lysate

cleared by centrifugation (24000 rpm, 1 h). The obtained pellet containing the protein as inclusion bodies, was washed two times by resuspension with 1% Triton in PBS (v/v) and one time with PBS only. The fusion protein was solubilized using 6 M Gdn·HCl, 100 mM Tris, 20 mM imidazole, 10 mM DTT, pH 7.5 buffer over 48 h with continuous stirring and subsequently centrifuged for 1 h at 24 000 rpm to separate undissolved debris. The supernatant was diluted 1:1 with loading buffer (6M Gdn·HCl, 100 mM Tris pH 7.5, 20 mM imidazole) and loaded onto a 5 mL HisTrap™ Ni-NTA affinity column (GE Healthcare) using a ÄKTAprime plus (GE Healthcare) FPLC system at room temperature. The fusion protein was eluted using a linear gradient of 20 – 500 mM imidazole in 10 column volumes at a flow rate of 5 mL/min. Peak fractions were pooled and dialysed against 4M Urea, 100 mM Tris pH 7.5 at 4°C overnight.

2.3.2.2 MesNa induced intein-mediated cleavage of IgM CH4 1-91-*Mxe*-His_7-CBD

Ni-NTA purified IgM CH4-91-*Mxe*-His7-CBD fusion protein was concentrated using Amicon Ultra Centrifugal filters to a volume of ~40 mL and an equal volume of 1 M sodium 2-mercaptoethanesulfonate (MesNa) in ddH_2O was dropwise added to the solution while stirring at room temperature. The thiol-induced intein cleavage of the fusion protein was monitored by SDS-PAGE and after 48 h the solutions was passed through a 0.23 µm syringe filter and purified via RP-HPLC using a C4 Kromasil column and a gradient of 5-55% solvent B (ACN + 0.08% TFA) in solvent A (ddH_2O + 0.1% TFA) at a flow rate of 20 mL/min. Fractions containing the cleaved protein were pooled, lyophilized and stored at -80°C.

2.4 Circular dichroism spectroscopy

CD spectra for Hsp90 peptide hydrazides **H1**, **H2**, **H3**, **H4** and Hsp90 wild-type N-terminal domain were recorded on a Chirascan Plus CD spectrometer (Applied Photophysics, UK) in micro-cuvettes (1 mm path length, Hellma Analytics, Germany). CD spectra of peptides **H1**-**H4** were recorded in PBS (pH 7.4) at 25 °C from 190 nm to 280 nm in 1 nm steps at a peptide concentration of 100 µM. Peptide concentrations were determined by the absorbance at 214 nm using a Nanodrop 2000 c (Thermofischer Scientific) and the extinction coefficient $\varepsilon_{214\,nm}$ calculated using the formula

$$\varepsilon_{214\,nm} = (n_{AA} - 1 + n_N + n_Q)2846 + n_F 7200 + n_H 6309 + n_W\ 22735 + n_Y\ 5755$$

n_{AA} = Total number of amino acid residues

n_X = Number of amino acid residue X

according to Gill *et al.* [42]. Five measurements were averaged and the background subtracted for each spectrum. CD spectra of $Hsp90^{WT}$ were performed in TBS (50 mM Tris, 150 mM NaCl, pH 7), PBS (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) and NaPi (40 mM Na_2HPO_4/NaH_2PO_4, 100 mM NaCl, pH 6.8) buffer at 25°C between 190 nm and 280 nm in 1nm steps and a protein concentration of ~180 µM. Protein concentration was determined using the absorption at 280 nm (Nanodrop 2000c) and the absorption coefficient $\varepsilon_{280\,nm}$ calculated using the ExPASy ProtParam online tool. Three measurements were averaged and the background subtracted for each spectrum.

2.5 NMR spectroscopy

Nuclear magnetic resonance (NMR) spectra of Hsp90 peptide hydrazides **H1**-**H4** were acquired on a Bruker Avance III HDX 700 MHz NMR spectrometer equipped with an inverse helium cooled cryoprobe and a sample concentration of 2 – 3.5 mM in 0.6 mL NMR buffer (20 mM NaPi, 10 µM DSS, 5% D_2O in ddH_2O, pH 6) at 25°C. Standard Bruker pulse sequences were used with water suppression by presaturation or 3-9-19 WATERGATE. Spectra acquired included: ^1H, ^1H-^{15}N HSQC, ^1H-^1H TOCSY (isotropic mixing time D9 = 100 ms), ^1H-^1H NOESY (noesy mixing time D8 = 400 ms), ^1H-^{13}C HSQC and ^1H-^{13}C HMBC. Spectra were processed with Bruker TopSpin and assigned with CcpNmr Analysis [43]. Side chain assignments were made using ^1H-^1H total correlation spectroscopy (TOCSY) and nuclear overhauser effect spectroscopy (NOESY) spectra. Backbone resonance assignments were made using ^1H-^{15}N, ^1H-^{13}C HSQC and HMBC spectra. Spectra were referenced to DSS in the ^1H dimension and indirectly in the ^{13}C or ^{15}N dimensions, according to IUPAC recommendations using a ratio of 0.251449530 for ^{13}C and 0.101329118 for ^{15}N [44]. Secondary Hα and Cα shifts were calculated by subtraction of random coil Hα and Cα shifts from the observed Hα and Cα chemical shifts (δ, ppm) for each peptide and each assigned residue. Values for random coil shifts of standard amino acids were taken from Wishart *et al.* [45] and for phosphorylated threonine from Conibear *et al.* [46]. Full spectra and chemical shift tables can be found in the supplementary data (S6-S13 and ST1-ST4).

2.6 Ligation Reactions

2.6.1 EPL of IgM CH4-thioester to peptide M1 and peptide M3

IgM CH4 92-132 wild-type peptide (**M3**) was kindly provided by Dr. Alexander Kravchuk. Expressed protein ligations (EPLs) of IgM CH4 1-91 thioester (0.4 mM) to wild-type and modified IgM CH4 92-132 (0.8 mM) were performed in 6 M Gdn·HCl, 150 mM NaPi (pH 7.4) an either 100 mM MPAA (wild-type ligation) or 100 mM thiophenol (**M3**) at room temperature. Reactions were monitored by LC-MS and SDS-PAGE and quenched by dilution with equal volumes of 6 M Gdn·HCl in ddH_2O (pH 4.7) and the addition of β-mercaptoethanol (20%, v/v) and a few crystals of TCEP. Preparative purification of the ligation product was performed on a C4 Kromasil column using a linear gradient of 5-45% solvent B (ACN + 0.08% TFA) in solvent A (ddH_2O + 0.1% TFA) at a flow rate of 20 mL/min at room temperature (ligation to modified peptide **M1**) or 60°C (wild-type ligation **M3**).

2.6.2 EPL of Hsp90^{C2} to peptide hydrazide H3

Small scale expressed protein ligation of Hsp90^{C1} (0.13 mM) to Hsp90 2-20 peptide hydrazide **H3** (1.34 mM) was performed in degassed ligation buffer consisting of 6 M Gdn·HCl, 150 mM NaPi (pH 7.4), 70 mM MPAA, at room temperature using in situ hydrazide to thioester conversion according to the protocol of Zheng *et al*. [18]. First, the peptide hydrazide to peptide azide activation was performed in 6M Gdn·HCl, 150 mM NaPi (pH 3) at -13°C (ice/salt bath) for 15 min by the addition of $NaNO_2$ (136 mM) and subsequently the reaction mixture was combined with the N-terminal cysteine peptide (Hsp90^{C1}) dissolved in the ligation buffer, brought to room temperature and the pH adjusted to 7.4. The ligation was stirred at room temperature, monitored via LC-MS and SDS-PAGE, quenched after 5 hours by addition of TCEP and acidification (HCl) and stored at -20°C. The reaction mixture was centrifuged after thawing and the supernatant loaded on a Kromasil C8 column (300-5-C4, 150 x 4.6 mm, 5 µm particle size) for purification by RP-HPLC (Dionex Ultimate 3000 HPLC) at 60°C using a gradient of 5-95% solvent B (ACN + 0.08% TFA) in solvent A (ddH_2O + 0.1% TFA) at a flow rate of 1 mL/min. Peak fractions were collected manually and analysed via ESI-MS.

3 Results and Discussion

3.1 Semi-synthesis of Hsp90 NTD

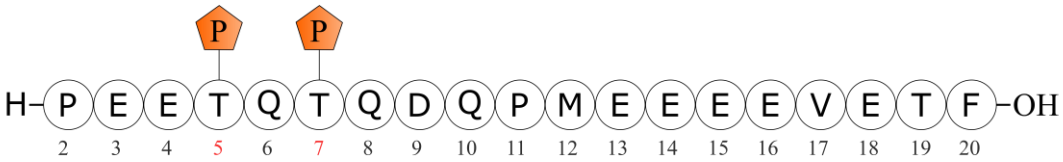
A key point in the design of a synthesis strategy for a semi-synthetic protein based on the native chemical ligation reaction is the choice of a suitable ligation site. A ligation site needs to be selected that keeps the synthetic fragment bearing the PTMs within reach of SPPS.

Furthermore, the acyl donor-containing fragment (thioester peptide) should not include residues which are prone to cyclise (Asp, Glu and Lys), electronically deactivating (Pro) or sterically hindered (Val, Ile) on the C-terminus which could lead to isopeptide bond formation or very low ligation yields [9]. Since the Hsp90 NTD does not harbour any natural cysteine residues, a native alanine was selected by mutation into a cysteine as ligation site. Post ligation desulfurization would then yield the protein with the native sequence. Therefore, alanine at position 21 was selected as ligation site in the α isoform of human Hsp90 NTD. The synthetic fragment includes Hsp90α NTD residues 2-20, excluding the starting methionine residue (position 1) in order to resemble the natural occurring protein in eukaryotic cell systems [47]. Besides the fulfilment of the mentioned requirements, this strategy would allow the examination of two specific PTMs of this domain. The phosphorylation of threonine residues at position 5 and 7, is reported to occur solely in the α-isoform of human Hsp90 NTD and therefore a deeper understanding of the impact of these modifications could lead to new insights into the different tasks and functions of the stress responding α-isoform of Hsp90 [48].

Based on this a strategy for the semi-synthesis of modified Hsp90 NTD was developed which included different variants of phosphothreonine modified Hsp90 2-20 peptide hydrazides (bearing single, double or no phosphothreonine modifications) and the generation of the remaining NTD fragment by recombinant expression. The peptide hydrazides function as precursors for the ligation to the recombinant part of the domain using *in situ* hydrazide to thioester conversion. Three different recombinant constructs were designed including the initial truncated mutant $Hsp90^{C1}$ bearing a N-terminal His_6-tag followed by a TEV recognition site and the sequence 21-235 (C21A). The second construct ($Hsp90^{C2}$) was designed with a TEV-cleavage site mutated between position 20 and 21 (C21A) and a C-terminal His_6-tag. The TEV-cleavage site allows for the release of the mutated cysteine at position 21, which is essential for the native chemical ligation. The third construct represent the full length wild-type Hsp90α NTD bearing a C-terminal His_6-tag. The sequence of the Hsp90 2-20 peptide with positions of phosphothreonine modification as well as an overview of the synthesis strategy are depicted in

Figure 6. Amino acid and nucleotide sequences (S1, S2 and S3) as well as a transformation vector map (S4) for the recombinant constructs including wild-type Hsp90α NTD ($Hsp90^{WT}$) are displayed in the supplementary data section.

A



B

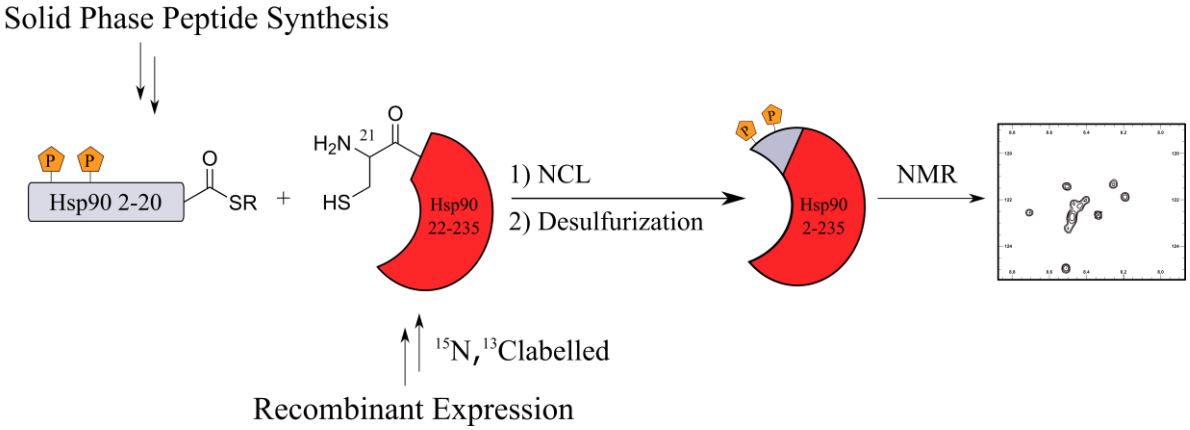


Figure 6: A) Amino acid sequence of human Hsp90α 2-20 NTD with indicated positions of phosphothreonine modifications. **B)** Schematic overview of the synthesis strategy for the preparation of specifically modified and segmentally isotope labelled Hsp90 NTD.

3.1.1 Solid phase peptide synthesis of Hsp90 2-20 peptide hydrazides

The Fmoc based solid phase peptide synthesis was used to generate Hsp90 2-20 peptide hydrazides **H1**, **H2**, **H3**, **H4** bearing either single (**H2**, **H3**), double (**H1**), or no phosphothreonine modification (**H4**) in positions 5 and 7, according to the protocol of Zheng *et al.* [18]. Purification and separation from deletion and addition products was challenging, suggesting that the synthesis could be further optimised. Peptides were obtained with a yield of 6% (**H1**, 6.8mg), 13% (**H2**, 15 mg), 15% (**H3**, 18.5 mg) and 15% (**H4**, 17 mg) based on the theoretical yield calculated from the synthesis scale (0.05 mmol). Analytical HPLC chromatograms and ESI-MS spectra of peptides **H1** – **H4** are shown in Figure 7 and Figure 8.

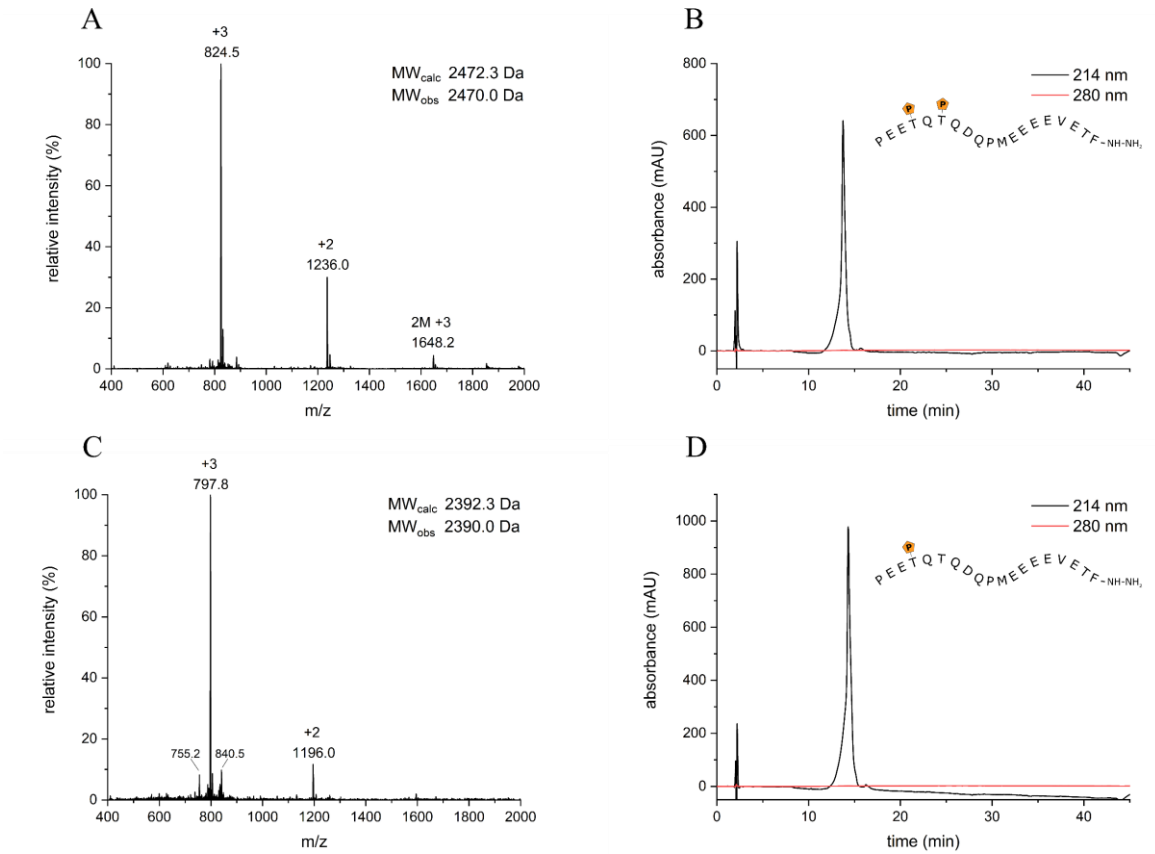


Figure 7: Characterisation of purified Hsp90 2-20 peptide hydrazides. **A)** ESI-MS of **H1** (phosphothreonine in position 5 and 7). **B)** RP-HPLC of **H1**. **C)** ESI-MS of **H2** (phosphothreonine in position 5). Observed masses 755.2 Da and 840.5 can be assigned to $[M+3H]^{3+}$ glutamic acid (E) deletion and addition products, respectively. **D)** RP-HPLC of **H2**. Dimeric masses (2M+3) can be attributed to the formation of adducts during the ESI process.

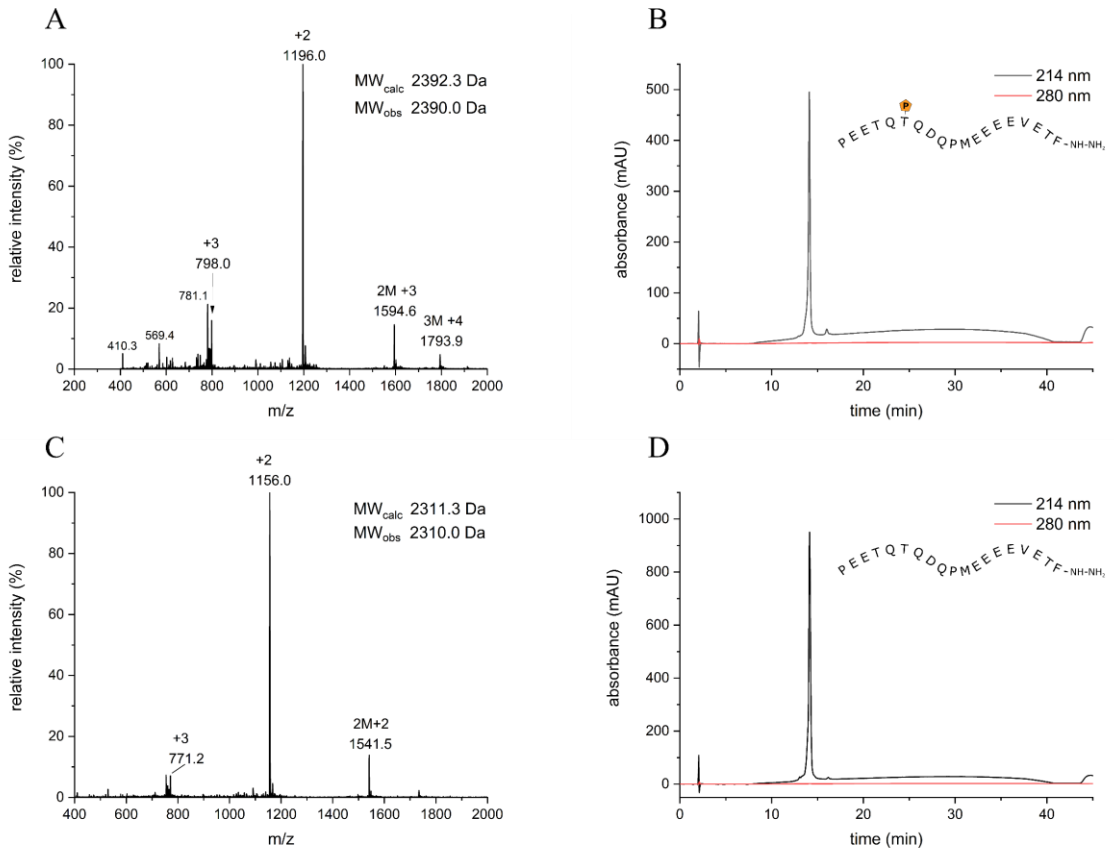


Figure 8: Characterisation of purified Hsp90 2-20 peptide hydrazides **H3** and **H4**. **A)** ESI-MS of **H3** (phosphothreonine in position 7). **B)** RP-HPLC of **H3**. **C)** ESI-MS of **H4** (unmodified). **D)** RP-HPLC of **H4**. Dimeric masses (2M+2, 2M+3) can be attributed to the formation of adducts during the ESI process.

Besides various other known SPPS based methods for the generation of C-terminal alpha thioester peptides, the synthesis of peptides hydrazides and the subsequent activation and *in situ* conversion to the alpha-thioester displays a convenient and relatively low-cost method that allowed the use of a common SPPS resin (2-Cl-Trt) and chemicals [49]. The relatively short peptide (19 amino acids) with a proline occurring approximately in the middle of the sequence allowed for a straightforward synthesis without the necessity of using pseudoprolines or other aids to prevent aggregation. Challenges included the separation of peptides containing oxidised methionine and deletion/addition side products which appeared during the synthesis (Figure 7C and Figure 8A). The oxidation of methionine and therefore the generation of methionine sulfoxide is a common side product during SPPS, which we tried to reduce by minimizing the amount and time of direct air exposure and by choosing a suitable peptide cleavage cocktail that allowed for the reduction of methionine sulfoxide during the cleavage reaction [50]. Although separation of methionine sulfoxide side product can be held to account for the modest

peptide yields between 6 and 15%, the generation of glutamic acid (E) deletion and addition products displayed a much larger obstacle to overcome during the purification process. With seven out of 19 residues being glutamic acids, four of them in a row between position 13 and 16, the accumulation of deletion products due to incomplete couplings of the amino acids displayed a severe problem during the purification process (Figure 7C, Figure 8A). Furthermore, the mass of a glutamic acid addition product (+129 Da) was observed (Figure 7C), which could originate from premature Fmoc removal between double coupling steps of glutamic acid. A possible optimisation for the synthesis protocol could include additional capping steps for free primary amine groups between the amino acid couplings that would block the chain elongation of deletion products and therefore improves separation during purification via RP-HPLC. However, since the synthesis yielded the desired modified and unmodified peptides in satisfactory yield and purity for first test ligations the next focus of the project was on the generation of the C-terminal cysteine modified recombinant part of Hsp90 NTD instead of improving these syntheses.

3.1.2 Protein expression, purification and TEV cleavage

3.1.2.1 Expression and purification of Hsp90^{C1}

Small scale (200 mL LB) test expressions verified the successful transformation of construct Hsp90^{C1} into *E. coli* BL21 (DE3) and *E. coli* Rosetta 2 (DE3). Expression screens showed that *E. coli* BL21 (DE3) and 0.5 mM IPTG at 30°C for 4 h give optimal expression. Expressions were monitored by SDS-PAGE by taking samples at OD_{600} ~0.8 before inducing with IPTG (t=0), and during expression t=1-4 h. Expression level monitoring of the strain BL21 and Rosetta 2 shows the strong expression of the construct Hsp90^{C1} (MW_{calc} 25815 Da) as a single band at an apparent molecular weight of ~29 kDa and is depicted in Figure 9A. SDS-PAGE analysis of the supernatant and pellet of the cell lysate shows that both strains express the protein exclusively in inclusion bodies as depicted in Figure 9B. Although both strains were suitable for large scale expressions, BL21 was chosen for further large-scale expressions due to the apparent higher expression level as well as overall expression purity.

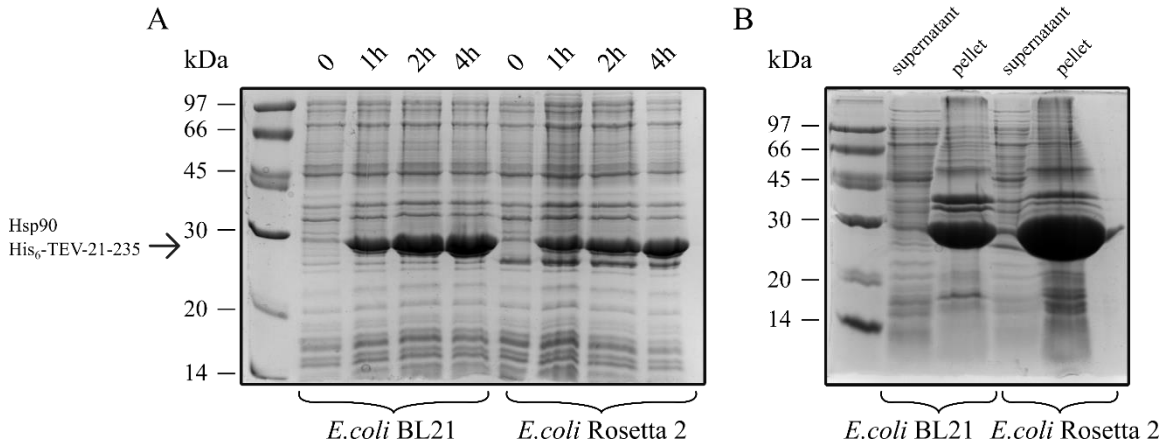


Figure 9: A) Expression of Hsp90^{C1} in *E. coli* BL21 (DE3) and *E. coli* Rosetta 2 (DE3). The gel shows the expression of the construct as a strong single band at MW_{app} of ~29 kDa. **B)** SDS-PAGE analyse of BL21 and Rosetta 2 cell lysate (supernatant and pellet) shows that the construct Hsp90^{C1} is exclusively expressed in inclusion bodies.

The protein expressed in inclusion bodies was solubilized in TBS buffer containing 6M Gnd·HCl (pH 7.5) over night and the supernatant loaded onto a Ni-NTA affinity column for purification. SDS-PAGE analysis of the collected peak fractions showed the successful binding of the protein to the Ni-affinity column and the high amount of protein that was loaded and eluted from the column (Figure 10A). The high amounts of unbound protein in the flow through can be explained by column overload which could be avoided by either increasing the column

volume and therefore the capacity or decrease the amount of protein loaded onto the column before elution to optimize the yield of the purification. The pooled peak fractions were submitted for LC-MS analysis (Figure 10B), which verified the correct molecular weight of the expressed construct. The average yield of the obtained protein was 26 mg/L of LB expression medium.

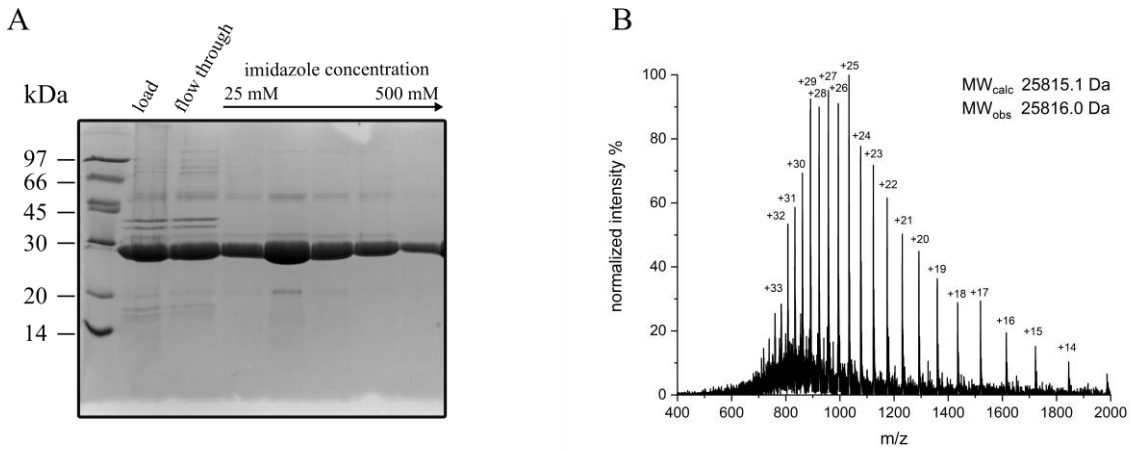


Figure 10: A) Ni-NTA affinity chromatography of Hsp90^{C1} on an ÄKTAprime Plus System. The protein was eluted at flow rate of 5 mL/min using a gradient of 25-500 mM imidazole (lanes of the samples taken during the elution are labelled with an arrow indicating the gradient) in 6M Gdn·HCl, TBS, pH 7.5 buffer. The strong elution band of the protein already at low imidazole concentrations and the high amount of unbound protein in the flow through indicate that the Ni-NTA column was overloaded and its capacity exceeded. **B)** Mass spectrum of Hsp90^{C1} after the purification via Ni-NTA affinity chromatography (MW$_{calc}$ 25815 Da, MW$_{obs}$ 25816 Da).

3.1.2.2 Refolding attempts of Hsp90^{C1}

In order to release the essential N-terminal cysteine for ligation of the recombinant protein to C-terminal thioester peptides, the denaturing reagent had to be removed so that TEV cleavage (releasing cysteine in position 21) could be performed. All attempts at finding conditions that allow for sufficient TEV cleavage while keeping both proteins in solution were unsuccessful. Methods for desalting and protein refolding included conventional (stepwise) dialysis and stepwise dilution using different temperatures, concentrations and additives as summarized in Table 1.

Table 1: Overview of the desalting attempts of Hsp90^{C1}. All attempts where performed in a buffered system (TBS) at pH 7.5.

From →	***To***	***Method***	***Temperature***	***Additive***	***Precipitation***
6M Gdn·HCl	4M Gdn·HCl	dialysis	4°C	×	×
4M Gdn·HCl	2M Gdn·HCl	dialysis	4°C	×	Partially
2M Gdn·HCl	0M Gdn·HCl	dialysis	4°C	×	entirely
4M Gdn·HCl	2M Gdn·HCl	dialysis	RT	×	entirely
6M Gdn·HCl	8M Urea	dialysis	4°C	×	×
8M Urea	4M Urea	dialysis	4°C	×	×
4M Urea	3M Urea	dialysis	4°C	×	Partially
3M Urea	2M Urea	dialysis	4°C	×	Partially
2MUrea	0M Urea	dialysis	4°C	×	entirely
8M Urea	0.8M Urea	stepwise dilution (1:10)	4°C	×	entirely
8M Urea	0.8M Urea	stepwise dilution (1:10)	RT	×	entirely
8M Urea	0.8M Urea	stepwise dilution (1:10)	4°C	10% glycerol (v/v)	entirely
8M Urea	1.6 M Urea	stepwise dilution (1:5)	4°C	×	Partially

Expression and purification conditions for the well-studied Hsp90 NTD in its α and β isoform are commonly found in the literature and usually describe the purification of the protein from the soluble fraction of the cell lysate without issues of refolding or precipitation [51,52]. The

different behaviour of this truncated construct compared to the wild-type Hsp90 NTD indicates the importance of the highly conserved N-terminal region and that modifying this segment by cutting a comparatively short sequence of 19 amino acids drastically influences the protein properties in terms of solubility. Since all efforts trying to release the N-terminal cysteine in position 21 (essential for ligation) by TEV cleavage failed due to solubility problems, we decided to design a new construct incorporating the N-terminal residues before the TEV cleavage site ($Hsp90^{C2}$) and therefore resembling the wild-type Hsp90 NTD ($Hsp90^{WT}$) as far as possible (Figure 11).

A

MHHHHHH ENLYFQ **C**FQAEIAQLM SLIINTFYSN KEIFLRELIS NSSDALDKIR YESLTDPSKL DSGKELHINL IPNKQDRTLT IVDTGIGMTK ADLINNLGTI AKSGTKAFME ALQAGADISM IGQFGVGFYS AYLVAEKVTV ITKHNDDEQY AWESSAGGSF TVRTDTGEPM GRGTKVILHL KEDQTEYLEE RRIKEIVKKH SQFIGYPITL FVEKERDKEV SDDEA

B

MPEETQTQDQ PMEEEEVETF ENLYFQ **C**FQAEIAQLM SLIINTFYSN KEIFLRELIS NSSDALDKIR YESLTDPSKL DSGKELHINL IPNKQDRTLT IVDTGIGMTK ADLINNLGTI AKSGTKAFME ALQAGADISM IGQFGVGFYS AYLVAEKVTV ITKHNDDEQY AWESSAGGSF TVRTDTGEPM GRGTKVILHL KEDQTEYLEE RRIKEIVKKH SQFIGYPITL FVEKERDKEV SDDEA HHHHHH

Figure 11: Amino acid sequence of constructs $Hsp90^{C1}$ (**A**) and $Hsp90^{C2}$ (**B**). The His_6-affinity tag is written in red, the TEV recognition site in blue and the cysteine essential for the ligation reaction in bold.

3.1.2.3 Expression and purification of $Hsp90^{C2}$

SDS-PAGE monitoring of the expression of $Hsp90^{C2}$ in *E. coli* BL21 (DE3) and *E. coli* Rosetta 2 (DE3) is depicted in Figure 12A. The gel shows the unexpected overexpression of two different constructs with apparent molecular weights of ~31 kDa (red box) and ~27 kDa (green box). The desired construct was expected to be the upper of the two bands since the construct $Hsp90^{C1}$ (MW_{calc} 25815 Da) was running at an apparent molecular weight of ~29 kDa and therefore the marginally heavier construct $Hsp90^{C2}$ (MW_{calc} 27962 Da) would likely be observable at an apparent molecular weight of ~31 kDa.

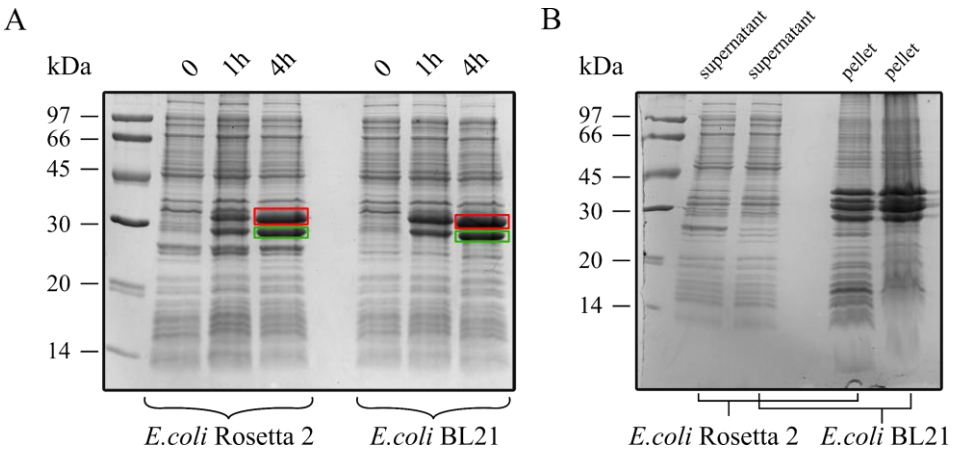


Figure 12: A) Expression monitoring of Hsp90^{C2} (MW_{calc} 27962 Da) in *E. coli* BL21 and *E. coli* Rosetta 2 via SDS-PAGE. The overexpression of at least two different proteins was observed with apparent molecular masses of ~31 kDa (red box) and ~29 kDa (green box). **B)** SDS-PAGE analysis of the supernatant and pellet of Hsp90^{C2} expression cell lysate. The gel shows that both strains (Rosetta 2 and BL21) expressed the construct Hsp90^{C2} in inclusion bodies.

The inclusion body pellet was solubilized in TBS, pH 7.5 containing 6M Gdn·HCl overnight and any undissolved cell debris cleared from the supernatant via centrifugation. Further purification via Ni-NTA affinity chromatography showed that both expressed protein constructs were likely to bear a His_6-affinity tag (Figure 13). The high amount of protein in the flow-through with a corresponding weak band resulting from the eluted proteins in the peak fractions could be explained by the usage of an old Ni-NTA column with low capacity.

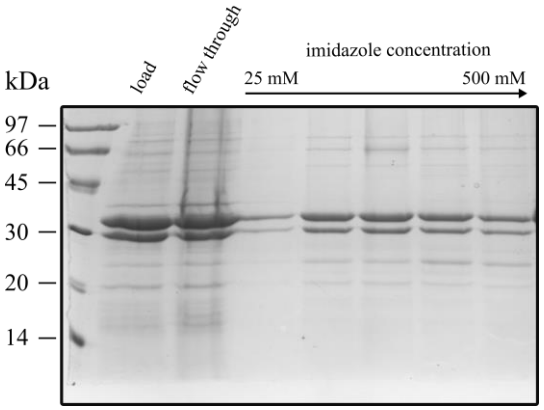


Figure 13: Ni-NTA affinity chromatography of Hsp90^{C2} on an ÄKTAprime Plus system. The protein was eluted at flow rate of 1 mL/min using a gradient of 25-500 mM imidazole in 6M Gdn·HCl, TBS, pH 7.5 buffer. The gel shows that two proteins were eluted with increasing imidazole concentration (lanes of the samples taken during the elution are labelled with an arrow indicating the gradient), indicating the presents of multiple overexpressed constructs which could bear a His_6-affinity tag.

The proteins were eluted as a single peak and the peak fractions pooled and analysed via LC-MS, as displayed in Figure 14.

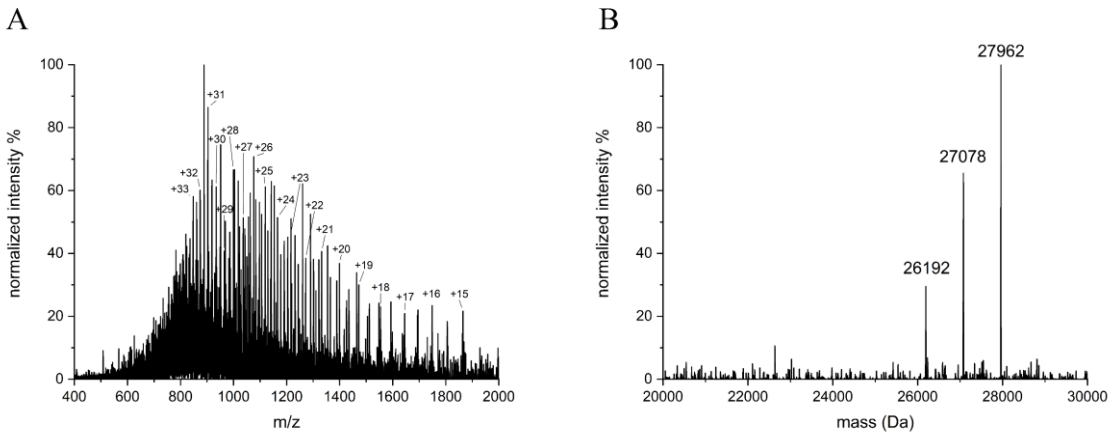


Figure 14: A) Mass spectrum of Hsp90^{C2} after Ni-NTA affinity purification. **B)** Deconvoluted mass spectrum of Hsp90^{C2} (MW_{calc} 27962 Da, MW_{obs} 27962; 27078; 26192 Da).

The observed spectra showed the expected mass of Hsp90^{C2} (MW_{calc} 27962 Da, MW_{obs} 27962 Da) and the appearance of two other overexpressed constructs with a mass differences of -884 Da (MW_{obs} 27078 Da) and -1770 Da (MW_{obs} 26192 Da). It was not possible to determine the origin of the other overexpressed constructs since no meaningful deletion products that could be responsible for the appearing masses could be found. Total protein yield was 13 mg/L LB expression medium. Since it was not possible to separate the different expressed protein constructs via Ni-NTA affinity chromatography the exact yield of Hsp90^{C2} couldn't be determined. Although the expression of the construct Hsp90^{C2} in inclusion bodies as well as the overexpression of other not identifiable constructs were not very promising first results for subsequent use of the construct, the protein was further used for refolding and TEV cleavage attempts as described in the following section.

3.1.2.4 Refolding and TEV cleavage of Hsp90^{C2}

In order to obtain the pivotal N-terminal cysteine required for the ligation reaction, a cleavage by TEV protease, which specifically recognizes the amino acid sequence ENLYFQ inserted in between position 20 and 21 of the construct Hsp90^{C2} (see section 3.1.1), was performed. The protein was purified under denaturing conditions using high concentrations of chaotropic salt (6M Gdn·HCl) had to be desalted to prevent denaturation of the TEV protease. Removal of the chaotrope was achieved in a single dialysis step by conventional dialysis to TBS, pH 7.5 at 4°C which lead to minor precipitation of the protein. The TEV cleavage was monitored via SDS-

PAGE and performed as described in section 2.3.1.5. RP-HPLC chromatogram of the purification of the protein shows the elution of uncleaved Hsp90 at 35.6 min and the cleaved protein as an emerging broad shoulder at 35.2 min (Figure 15A).

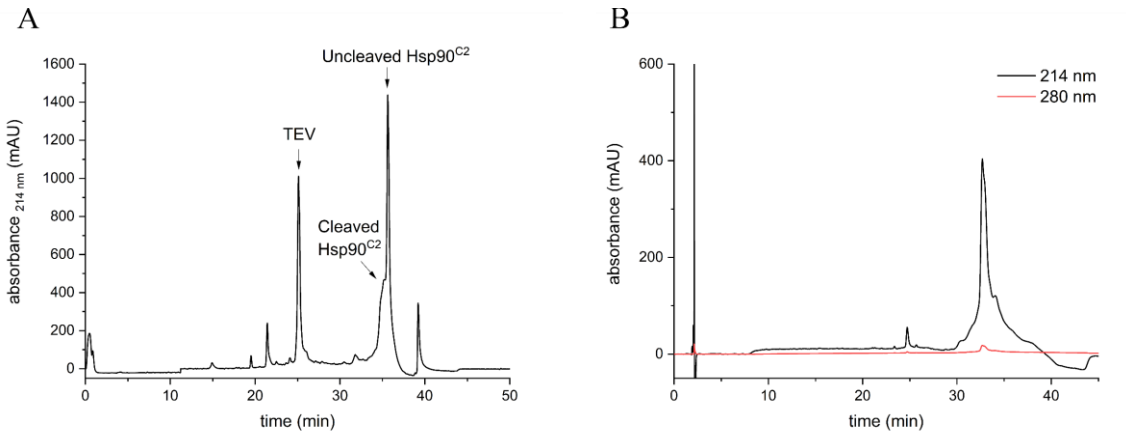


Figure 15: A) Purification of Hsp90^{C2} via RP-HPLC after TEV cleavage. The desired cleaved protein was eluted as a broad shoulder at 35.2 min. The sharp peak next to the eluted cleaved protein represents uncleaved Hsp90^{C2}. **B)** RP-HPLC chromatogram of purified Hsp90^{C2} after TEV cleavage.

Fractions were analysed by ESI-MS and collected according to their masses, combined and lyophilized. RP-HPLC chromatogram and MS spectra of purified Hsp90^{C2} after TEV cleavage are depicted in Figure 15B and Figure 16 respectively. The protein yield was 0.9 mg in an acceptable purity for test ligations.

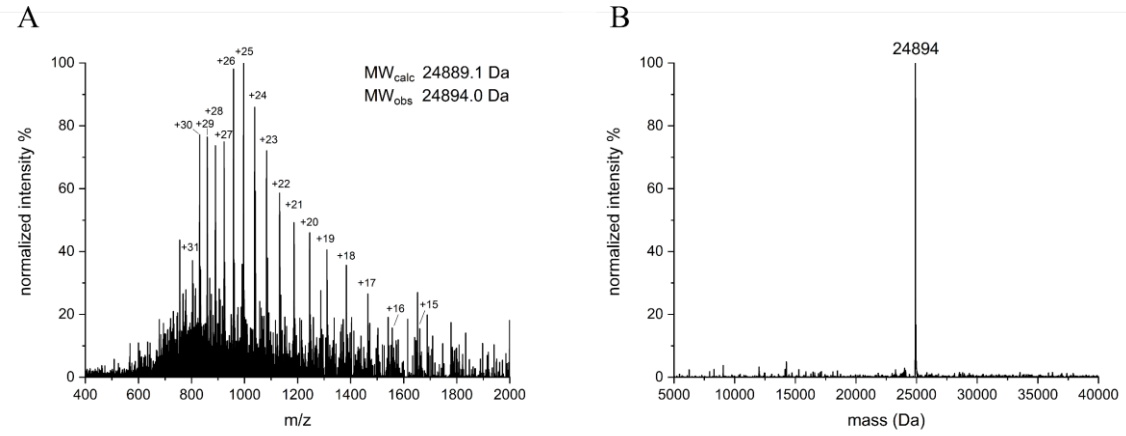


Figure 16: A) Mass spectrum of purified Hsp90^{C2} after the TEV cut. The observed mass (MW_{obs} 24894 Da) is in good agreement with the calculated mass (MW_{calc} 24889 Da). **B)** Deconvoluted mass spectrum of purified Hsp90^{C2} after the TEV cut.

Although the construct Hsp90^{C2} was also expressed in inclusion bodies and not, as might be desired, in a soluble form, the successful single step refolding of the protein from high denaturing conditions (6M Gdn·HCl) to a buffer system without any chaotropic salts or solubilization reagents can be considered a major improvement compared to the handling of construct Hsp90^{C1}. The purification chromatogram of the Hsp90^{C2} TEV cleavage solution (Figure 16A) indicates the incompleteness of the reaction after 24 h and asks for an optimisation of the reaction conditions in the future, especially considering the high ratio of TEV:protein. A low concentration of denaturing reagents during the cleavage reaction could partially unfold the protein and therefore increase the accessibility of the cleavage site for the TEV protease and improve the reaction yield. Additionally, raising the temperature during the cleavage reaction and a prolonged reaction time should be considered as well as an optimised solvent gradient during the purification process, which could lead to an overall improved yield.

3.1.2.5 Expression and purification of Hsp90^{WT}

Expression of Hsp90^{WT} in *E. coli* BL21 and *E .coli* Rosetta 2 was performed on a 2 L scale as described in section 2.3.1.3 and monitored via SDS-PAGE (Figure 17A). The protein with a calculated molecular weight of 27266 Da was observed as a single band at an apparent molecular weight of ~30 kDa. Subsequent SDS-PAGE analysis of the cell lysate revealed that significant amount of protein was expressed in a soluble form, both in *E. coli* BL21 and *E. coli* Rosetta 2 (Figure 17B) and therefore the supernatant of *E. coli* BL21 cell lysate was selected for further purification.

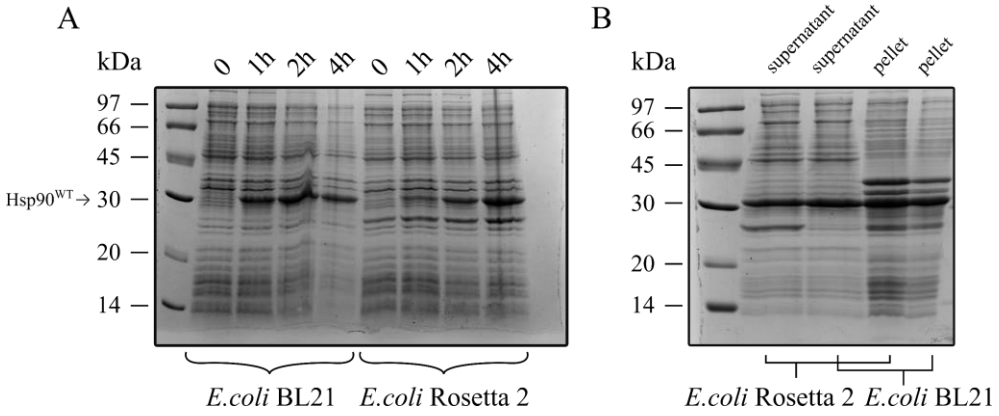


Figure 17: A) Expression monitoring of Hsp90^{WT} in *E. coli* BL21 and *E. coli* Rosetta 2 via SDS-PAGE. **B)** SDS-PAGE analysis of *E. coli* BL21 and *E. coli* Rosetta 2 cell lysate revealed that of Hsp90^{WT} was expressed in inclusion bodies as well as in a soluble form.

The successful purification of Hsp90^{WT} from the supernatant of *E. coli* BL21 cell lysate via Ni-NTA affinity chromatography was monitored via SDS-PAGE and is depicted in Figure 18. The

peak fractions were pooled and dialysed against TBS, pH 7.5 at 4°C overnight and the protein was stored at –80°C.

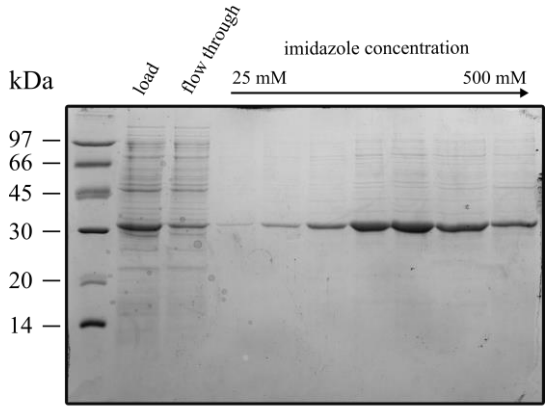


Figure 18: SDS-PAGE monitoring of the purification of Hsp90^{WT} via Ni-NTA affinity chromatography. The soluble fraction of *E. coli* BL21 cell lysate was loaded onto the Ni-NTA column and the protein eluted with a flow rate of 1 mL/min using a gradient of 25-500 mM imidazole (lanes of the samples taken during the elution are labelled with an arrow indicating the gradient) in PBS, pH 7.4.

LC-MS analysis of the purified Hsp90^{WT} confirmed the correct mass of the expressed construct and revealed that the starting methionine is cleaved after the expression (Figure 19A and B). The average yield of the obtained protein was 11 mg/L of LB expression medium.

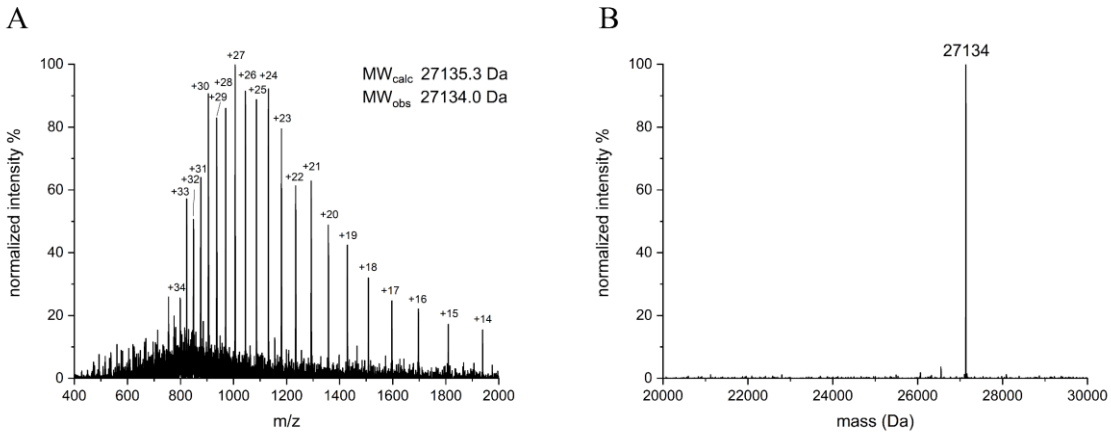


Figure 19: A) Mass spectrum of purified Hsp90^{WT}. The observed mass of 27134 Da is in good agreement with the calculated mass of 27135 Da after the subtraction of methionine (131 Da) indicating that the start methionine gets cleaved after the expression. **B)** Deconvoluted mass spectrum of Hsp90^{WT}.

The expression monitoring of wild-type Hsp90 (Figure 17A) indicated weaker expression level of the construct when compared to the mutants Hsp90^{C1} and Hsp90^{C2} but also revealed that the

wild-type domain, in contrast to the mutants, is in significant amounts expressed as a soluble protein (Figure 17B) and not solely in inclusion bodies. The comparison with $Hsp90^{C2}$ (Figure 12A and B) clearly indicates the sensitivity of the N-terminal region of the NTD since the insertion of ENLFQC between position 20 and 21 of $Hsp90^{C2}$ changes the solubility properties and the expression levels severely. On the other hand, the C-terminal His_6-affinity tag on the NTD does not influence the solubility and expression properties of the wild-type protein, which can be explained by the anyhow flexible linker region that would follow at the C-terminal end to connect the NTD with the MD domain in the full-length sequence of the protein.

3.1.3 Protein characterisation

3.1.3.1 CD spectroscopy of $Hsp90^{WT}$ and Hsp90 2-20 peptide hydrazides

3.1.3.1.1 Circular dichroism spectroscopy of $Hsp90^{WT}$

In order to determine whether refolding of the expressed construct $Hsp90^{WT}$ was successful, circular dichroism (CD) spectra were recorded in several buffer systems and compared with CD spectra in the literature. Overlaid spectra of the recorded CD spectra are depicted in Figure 20.

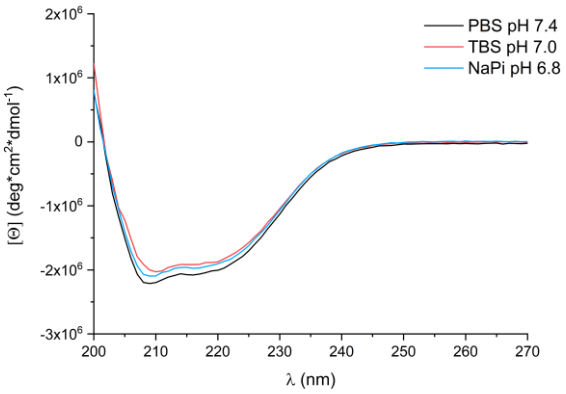


Figure 20: Overlaid Far-UV circular dichroism spectra of $Hsp90^{WT}$ in three different buffer systems. Spectra were recorded at a protein concentration of 180 µM, using a 1.0 mm path-length cell at 25°C and a scanning interval of 1 nm.

CD analysis of the recombinant wild-type Hsp90 NTD indicate the successful folding of the protein domain in several buffer systems (Figure 20). The spectra of the highly conserved domain shows the expected predominantly helical conformation indicated by the minima at 209 and 222 nm and is in good agreement with the predicted CD spectra from the PDB2CD prediction tool (PDB entry: 6GPW) as well as published far-UV CD data for recombinant chicken Hsp90 NTD which shares 91 % amino acid sequence identity [53,54].

3.1.3.1.2 Circular dichroism spectroscopy of Hsp90 2-20 peptide hydrazides

Far-UV circular dichroism spectroscopy of synthesised Hsp90 2-20 peptide hydrazides **H1**, **H2**, **H3** and **H4** was performed in order to determine if the peptides exhibit secondary structure elements. The overlaid spectra of the peptides, which were analysed in a PBS buffer system resembling physiological conditions (pH 7.4 at 25°C), are depicted in Figure 21.

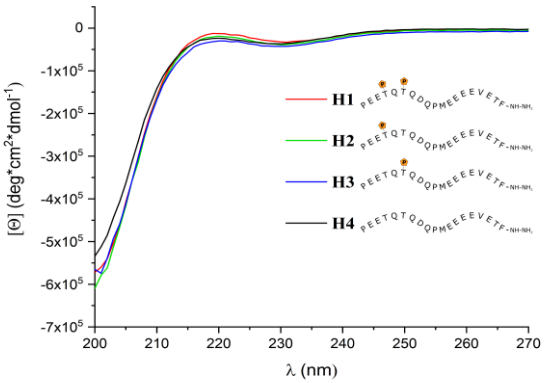


Figure 21: Overlaid far-UV circular dichroism spectra of Hsp90 2-20 peptide hydrazides **H1**, **H2**, **H3** and **H4**. Spectra were recorded at peptide concentration of 100 µM in PBS, pH 7.4 at 25°C using a 1.0 mm path-length cell and a data interval of 1 nm.

The recorded spectra revealed that the synthetic peptides exhibit no secondary structure elements and no significant influence of threonine phosphorylation was observed.

3.1.3.2 NMR spectroscopy of Hsp90 peptide hydrazides

The overlaid ^{1}H-^{1}H NOESY and ^{1}H-^{1}H TOCSY spectra of the NH/Hα region of Hsp90 peptide hydrazide **H4** are depicted in Figure 22A. Successful HN/Hα crosspeak assignment was performed for each amino acid residue except glutamic acid at position 3 and glutamic acids between positions 13 and 16 due to overlapping peaks (marked with a rectangle in Figure 22A). An overlay of the NH/Hα regions of the ^{1}H-^{1}H TOCSY spectra of peptide hydrazides **H1**-**H4** is depicted in Figure 22B. Black arrows and dotted lines indicate significant chemical shifts of the resonances marked in phosphorylated peptides (Figure 22B). Similarly, the ^{1}H-^{13}C HSQC spectrum of peptide **H4** and the overlaid ^{1}H-^{13}C HSQC spectra of peptides **H1**-**H4,** as well as the ^{1}H-^{15}N HSQC spectrum of **H4** and the overlaid ^{1}H-^{15}N HSQC spectra of peptides **H1**-**H4,** are depicted in Figure 23, Figure 24 and Figure 25A and B respectively**.**

The NMR spectra of the peptide hydrazides allow for first insights into the influences of PTMs on structural elements of the peptide and can be compared to literature chemical shifts of modified and unmodified random coil peptides. The ^{1}H-^{1}H NOESY spectrum of the unmodified peptide **H4** and the modified peptides **H1**-**H3** only show strong (*i*, *i*+1) inter-residue crosspeaks which is an indication that the peptide exhibits no secondary structure and resembles a random coil peptide [55,56]. This agrees with the results obtained from the CD measurements of the peptide hydrazides **H1**-**H4** (see section 3.1.3.1.2).

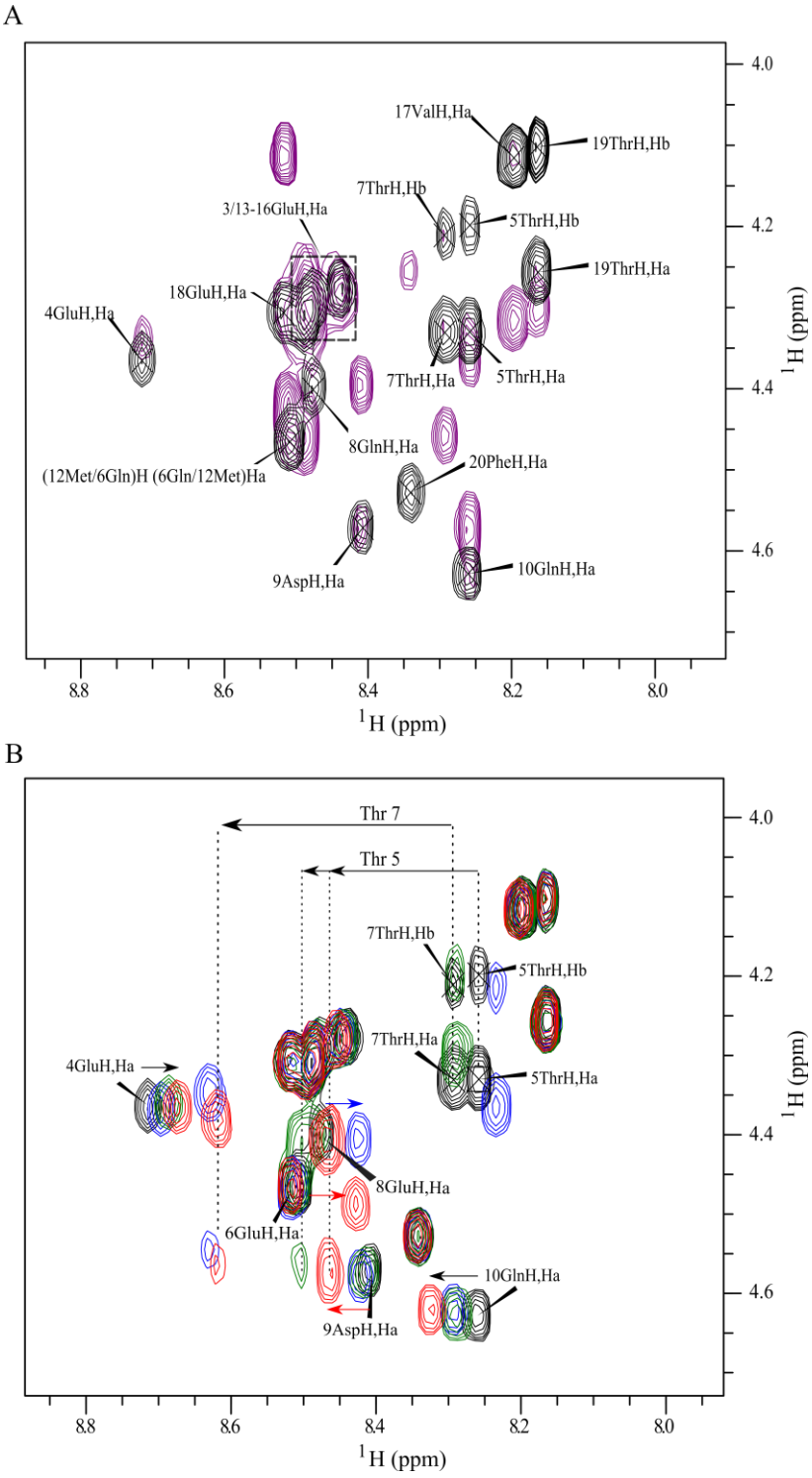


Figure 22: A) Overlaid ^{1}H-^{1}H NOESY (purple) and ^{1}H-^{1}H TOCSY (black) spectra of unmodified Hsp90 2-20 peptide hydrazide **H4**. **B)** Overlaid ^{1}H-^{1}H TOCSY spectra of modified and unmodified Hsp90 peptide hydrazides. Signals in red, green, blue and black represent the peptides **H1**, **H2**, **H3** and **H4**, respectively. Arrows and dotted lines indicated significant changes in chemical shifts due to phosphorylation on threonine residues 5 and/or 7.

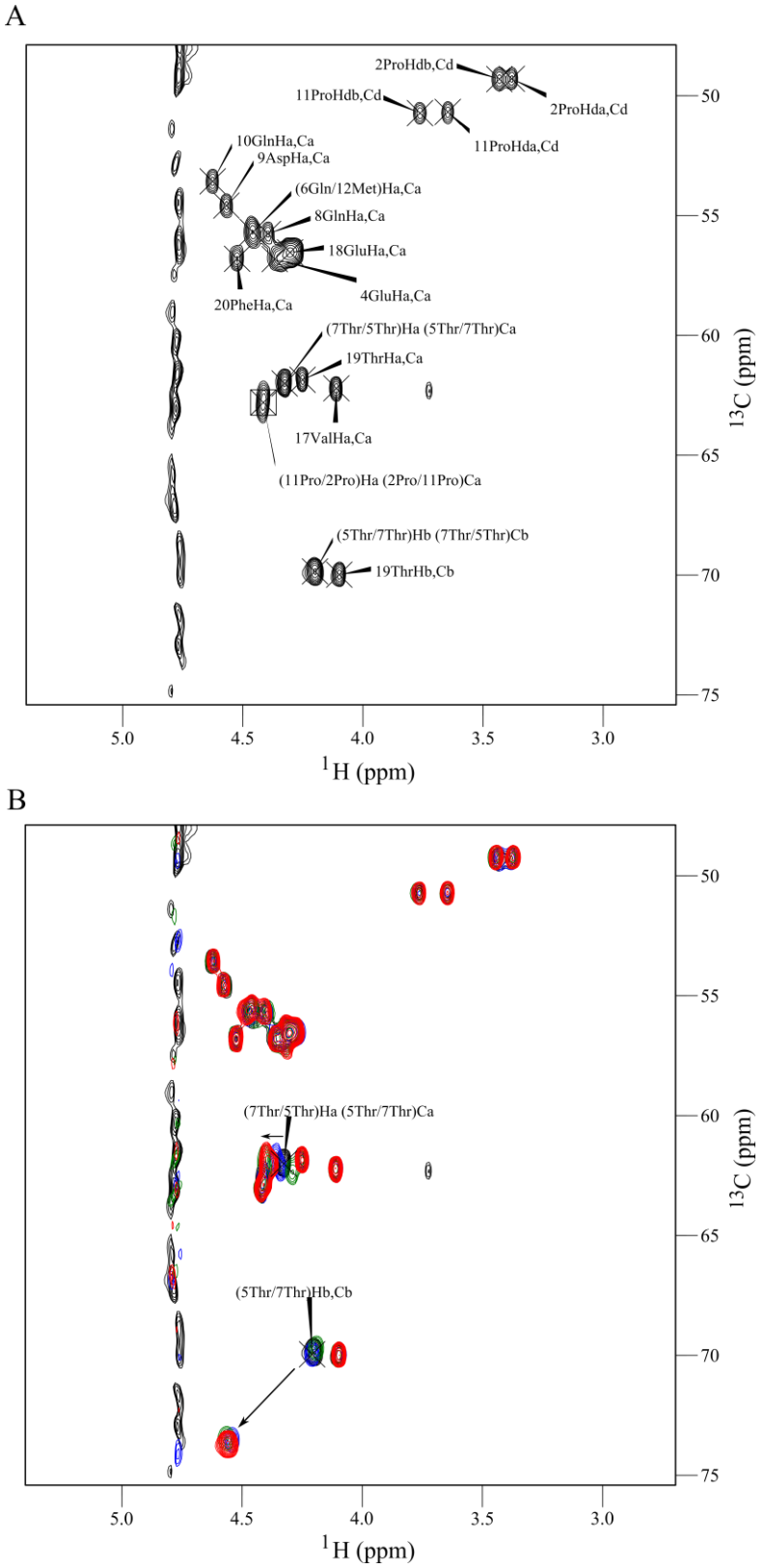


Figure 23: A) C-H region of ^{13}C-HSQC spectrum of Hsp90 peptide hydrazide **H4**. **B)** Overlaid C-H regions of 13C-HSQC spectra of Hsp90 peptide hydrazides **H1**, **H2**, **H3** and **H4**. Peptide **H1**, **H2**, **H3** and **H4** are represented in the colours red, green, blue and black, respectively. Arrows indicated significant changes in chemical shifts due to phosphorylation on threonine residues 5 and/or 7.

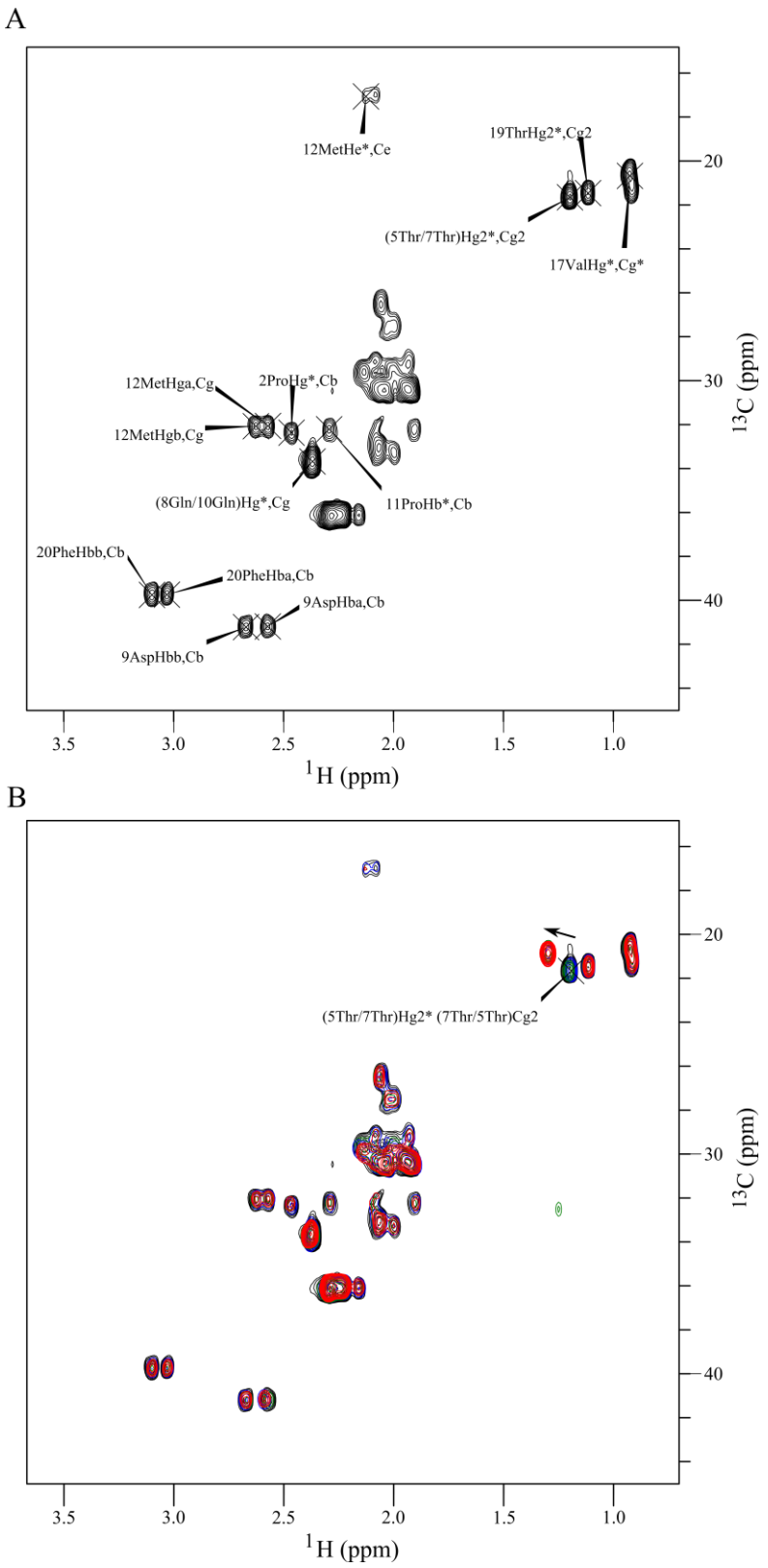


Figure 24: A) CH_2 and CH_3 region of ^{13}C-HSQC spectrum of Hsp90 peptide hydrazide **H4**. **B)** Overlaid ^{13}C-HSQC spectra of Hsp90 peptide hydrazides **H1**, **H2**, **H3** and **H4** in the CH_2 and CH_3 region. Peptides **H1**, **H2**, **H3** and **H4** are represented in the colours red, green, blue and black, respectively. Arrows indicated significant changes in chemical shifts due to phosphorylation on threonine residues 5 and/or 7.

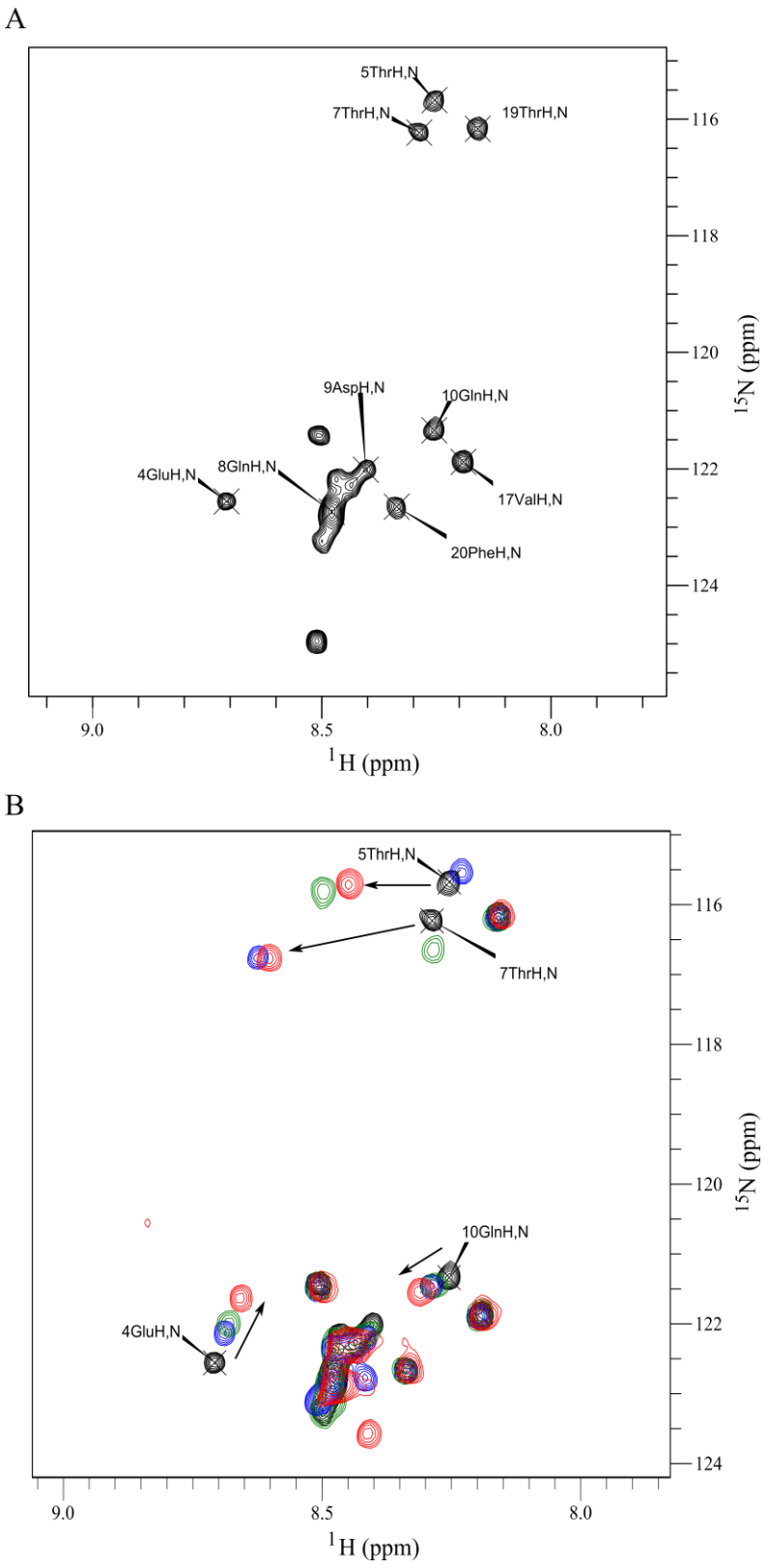


Figure 25: A) ^{15}N HSQC spectrum of Hsp90 peptide hydrazide **H4**. **B)** Overlaid ^{15}N HSQC spectra of Hsp90 hydrazide peptides **H1**, **H2**, **H3** and **H4**. Signals in red, green, blue and black represent the peptides **H1**, **H2**, **H3** and **H4**, respectively. Arrows lines indicated significant changes in chemical shifts due to phosphorylation on threonine residues 5 and/or 7.

Furthermore, differences between successfully assigned Hα, Cα and HN atom shifts and random coil shifts for the Hsp90 peptide hydrazides **H1**-**H4** are shown as secondary shift plots in Figure 26. Secondary shift plots allow for a better visualisation of the differences between the measured chemical shifts and random coil shifts which can give insights into protein secondary structural elements and the influences of PTMs on them. Since the secondary shift differences all peptide hydrazides are small (< 0.1 ppm for ^{1}H and < 1.0 ppm for ^{13}C) and there are no major structural changes on phosphorylation of Thr at positions 5 and 7 besides minor changes in shifts of neighbouring residues, this results support the lack of a secondary structure.

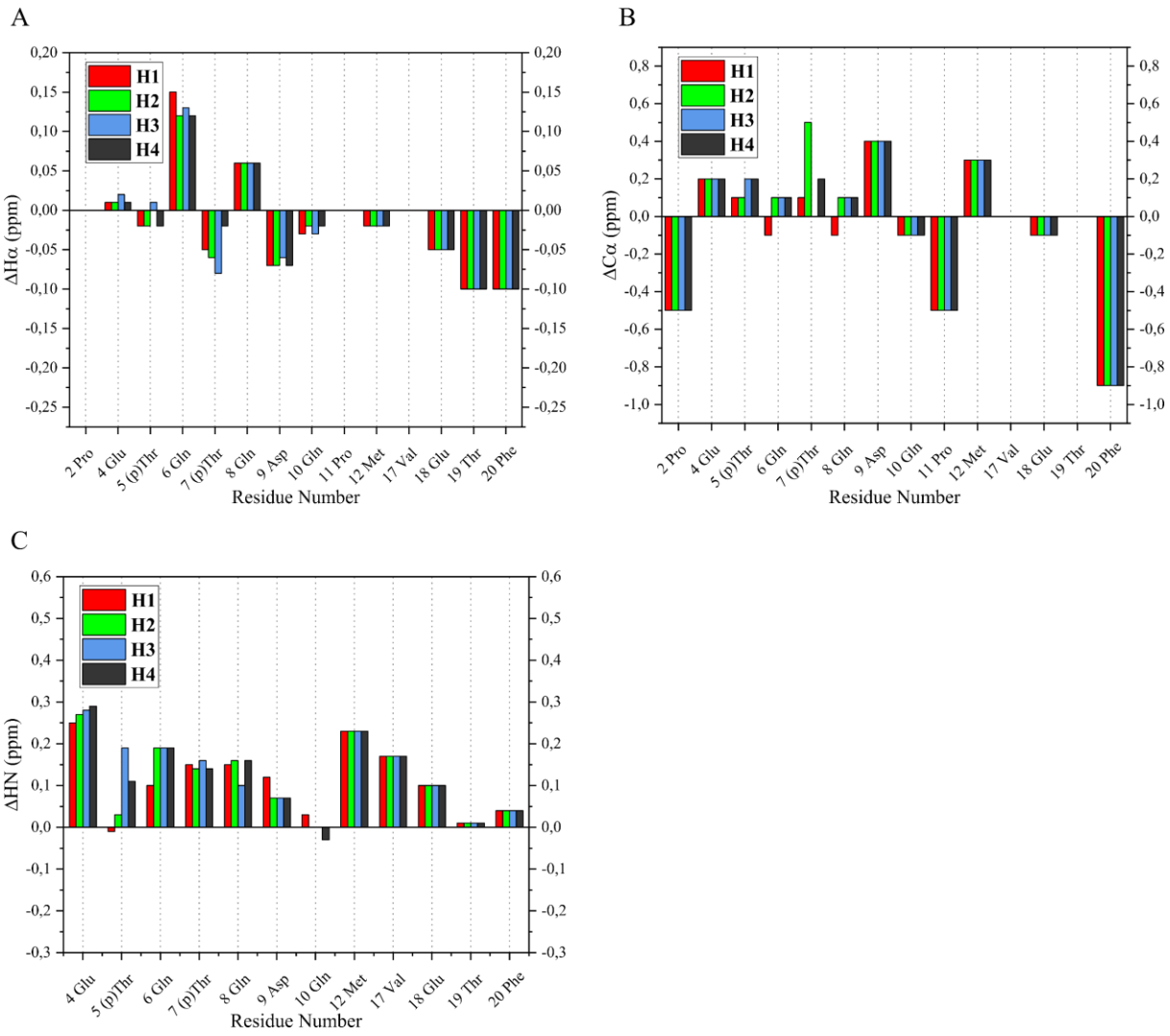


Figure 26: Hsp90 peptide hydrazides **H1**-**H4** chemical shifts (Δδ ppm) as a function of the residue number. The plots visualize the modulation of shifts caused by phosphothreonine and the overall deviation of the peptide resonance from a random coil peptide **A)** Δδ Hα backbone atom chemical shift differences. **B)** Δδ Cα atom chemical shift differences. **C)** ΔHN atom chemical shift differences.

However, according to the literature, the addition of most PTMs to a random coil peptide make little or no difference to the backbone chemical shifts of the modified residue (NH, HN, Hα, Cα

and CO) with the exception of serine and threonine phosphorylation [46]. Indeed, the observed downfield shifts (~0.3 ppm) for HN values of phosphorylated threonine in position 5 and 7 match the literature values of 0.25-0.3 ppm which are assumed to be the result of intra-residue hydrogen bonding between phosphate and backbone amide groups in unstructured regions [55, 57]. The observed downfield shifts of Hβ/Cβ resonances (0.36 / 3.8 ppm) are also in good agreement with the literature values (+~0.3 / +~3-4 ppm) (Figure 23B) and are due to the proximity of these nuclei to the phosphate group [46].

Additional shifts of residues in proximity to the modified threonines are also observed (Figure 22B and Figure 25B). These include the HN upfield shifts of glutamic acid residue 4 (0.1 ppm) and the downfield shift of and glutamine 10 (0.06 ppm). Shifts of the peptide **H1**, with pThr in positions 5 and 7 additionally included the downfield shift of aspartate 9 (0.06 ppm,) as well as the upfield shift (0.08 ppm) of glutamic acid 6 (Figure 22B, red arrows). Noticeably, the 0.05 ppm upfield HN shift of glutamic acid 8 was solely observed in the spectrum of peptide **H3** with the phosphothreonine in position 7 (Figure 22B, blue arrow). The observed shifts of non-modified residues can be seen as indications of a structured or hairpin region where the phosphate group might interact with the backbone HN of the residues which are in proximity of the phosphorylated threonine [46,57,58]. It can be concluded that the peptides seem not to exhibit an overall secondary structure, based on the similarity of the chemical shifts with those of random coil peptides and the lack of long-range NOE correlations, which is not unexpected since published structure data shows that part of the sequence contributes to the start of a beta-strand region [59]. However, the influences of phosphorylation onto the chemical shifts of neighbouring residues indicate that the PTMs might have an impact on the local conformation. Further NMR studies of the complete protein domain are needed to determine whether phosphorylations at positions 5 and 7 have long-range interactions with other structural elements in the NTD.

3.1.4 EPL of Hsp90^{C2} to peptide hydrazide H3

The small scale ligation of recombinant protein Hsp90^{C2} to synthetic hydrazide peptide **H3** (10-fold excess), using *in situ* hydrazide to thioester conversion, was monitored by LC-MS and SDS-PAGE and is depicted in Figure 27. The emergence of the expected mass of the ligation product (MW_{calc} 27247 Da; MW_{obs} 27249 Da) indicated the successful ligation, however attempts to isolate the product by RP-HPLC purification were not successful (RP-HPLC chromatogram of the attempted purification can be found in the supplementary data, S15).

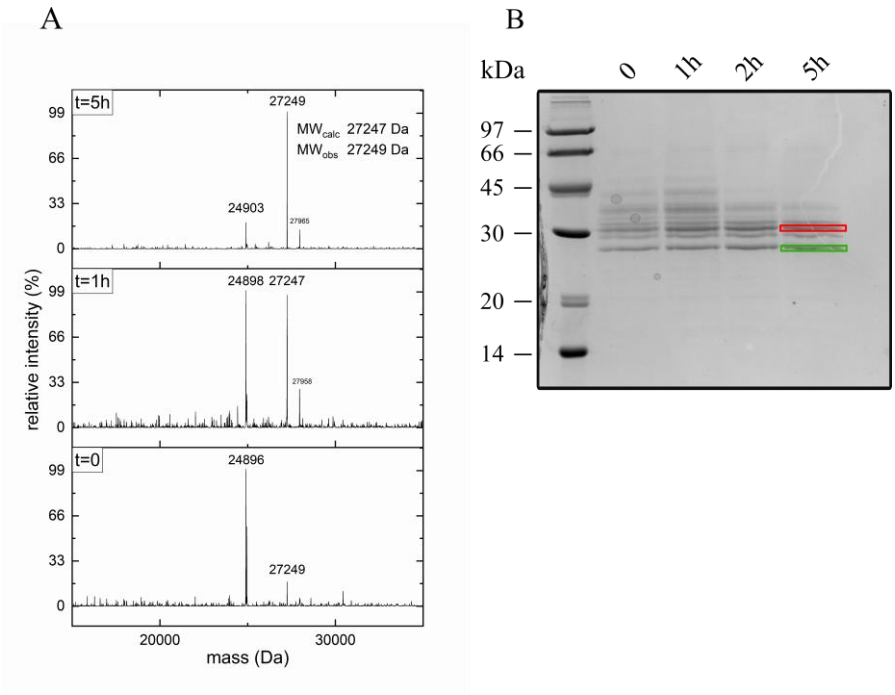


Figure 27: Monitoring of the reaction between recombinant Hsp90^{C2} and synthetic Hsp90 2-20 peptide hydrazide **H3** via LC-MS and SDS-PAGE. Samples for the time point zero (t=0), were taken after reactants were dissolved and the pH of the reaction mixture adjusted. **A)** LC-MS monitoring taken at the time points t=0 (after combination of the reactants and pH adjustment), t=1 h and t=5 h. **B)** Reaction monitoring via SDS-PAGE at the time points t=0, t=1 h, t=2 h and t=5 h. The ligation product is expected to run at an apparent molecular weight of ~31 kDa (red box) while Hsp90^{C2} is running at an apparent molecular weight of ~28 kDa (green box).

For this first test reaction, standard conditions such as a buffer system containing 6M Gdn·HCl, 150 mM NaPi pH 7.5, as well as the thiol additive MPAA (100 mM) were used. MPAA was chosen as a thiol additive since it proofed to be an efficient additive for ligation reactions, which gives rise to fast reactions rates and is easy to handle, although it doesn't allow one pot ligation-desulfurization since it quenches the radical reaction initiated by VA-044. Therefore, it will most likely be exchanged with methyl thioglycolate (MTG) in future ligations. The fast reaction rate is illustrated by the formation of product at time point 0, which was taken after the ligation

mixture was combined and the pH of the reaction mixture adjusted (5-10 min) (Figure 27A). The reaction was considered as completed after 5 h although a longer reaction time could have been beneficial for the formation of product and should therefore be extended in future ligations. The failed isolation of the product can be explained by the small scale of the reaction together with product loss during the reaction work up. Quenching of the reaction by acidification was most likely the major reason for product loss since a change in colour from clear to brown as well as the formation of a precipitant which contained MPAA and ligation product was shortly after observed. Although the product mass was still observable during LC-MS analysis of the resultant supernatant, a purification attempt via RP-HPLC only delivered the separation of the peptide hydrazide starting material without a separated peak for the ligation product or the recombinant $Hsp90^{C2}$ fragment. The commonly used reducing agent TCEP was only added during the reaction workup since the formation of MPAA adducts have been observed after the storage of the quenched reaction overnight. However, the initial ligation results show that human Hsp90α NTD with site-specific PTMs can be generated using the developed semi-synthesis protocol. Further optimisation of the ligation and purification conditions would enable larger scale production and future segmental labelling of the semi-synthetic domain

3.2 Semi-synthesis of IgM CH4

The same strategic principles as described above were applied for the selection of the ligation site within the IgM CH4 domain. A natively occurring cysteine at position 92 marked an ideal choice. Since the ligation site is close to the C-terminal end of the domain, the synthetic fragment bearing the cysteine at its N-terminus is ligated to the larger recombinant thioester segment, which is obtained via expression of an intein fusion protein using an pTXB1 expression system (for the vector map see S5). Benzophenone was chosen as the photoreactive crosslinker due to its stability which allows manipulation in ambient light and the commercial availability as amino acid (Fmoc-*p*-benzoylphenylalanine) suitable for standard Fmoc-based SPPS. The sequence of the 41-residue C-terminal peptide with positions fore photocrosslinker incorporation as well as a schematic overview of the overall project strategy is depicted in Figure 28.

A

B

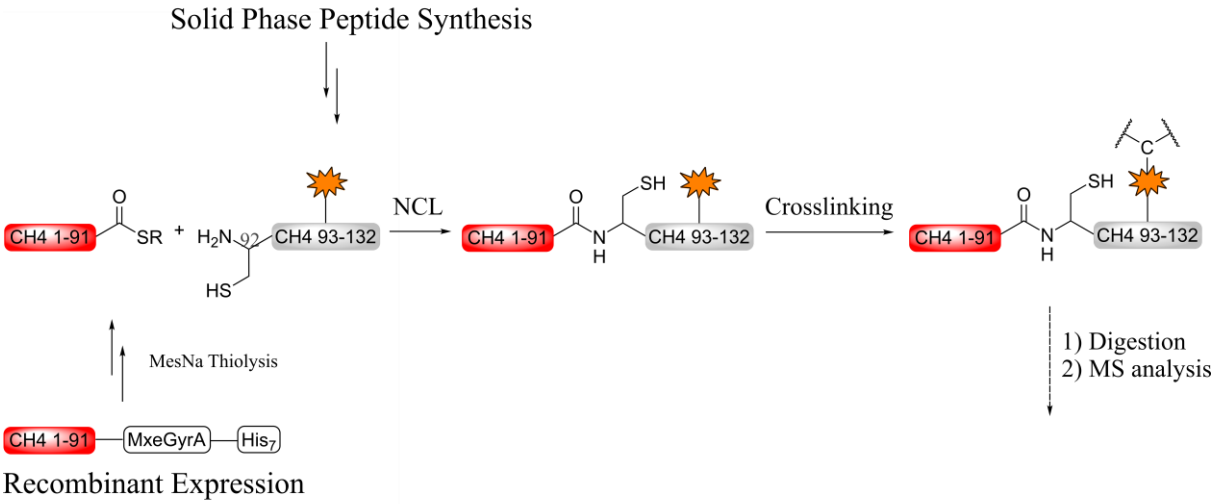


Figure 28: A) Sequence of IgM CH4 92-131. Amino acids that will be replaced by photocrosslinkers are highlighted in yellow. **B)** Schematic overview of the strategy for the semi-synthesis of IgM CH4 domain with incorporated photocrosslinker for studies via mass spectrometry.

3.2.1 Solid phase peptide synthesis of modified IgM CH4 92-132 peptides

Analytical RP-HPLC chromatograms and ESI-MS spectra of purified peptides **M1** and **M2** are shown in Figure 29. Peptide **M1** was obtained with a yield of 4% (8.3 mg) and peptide **M2** with a yield of 2% (15 mg), based on the synthesis scale.

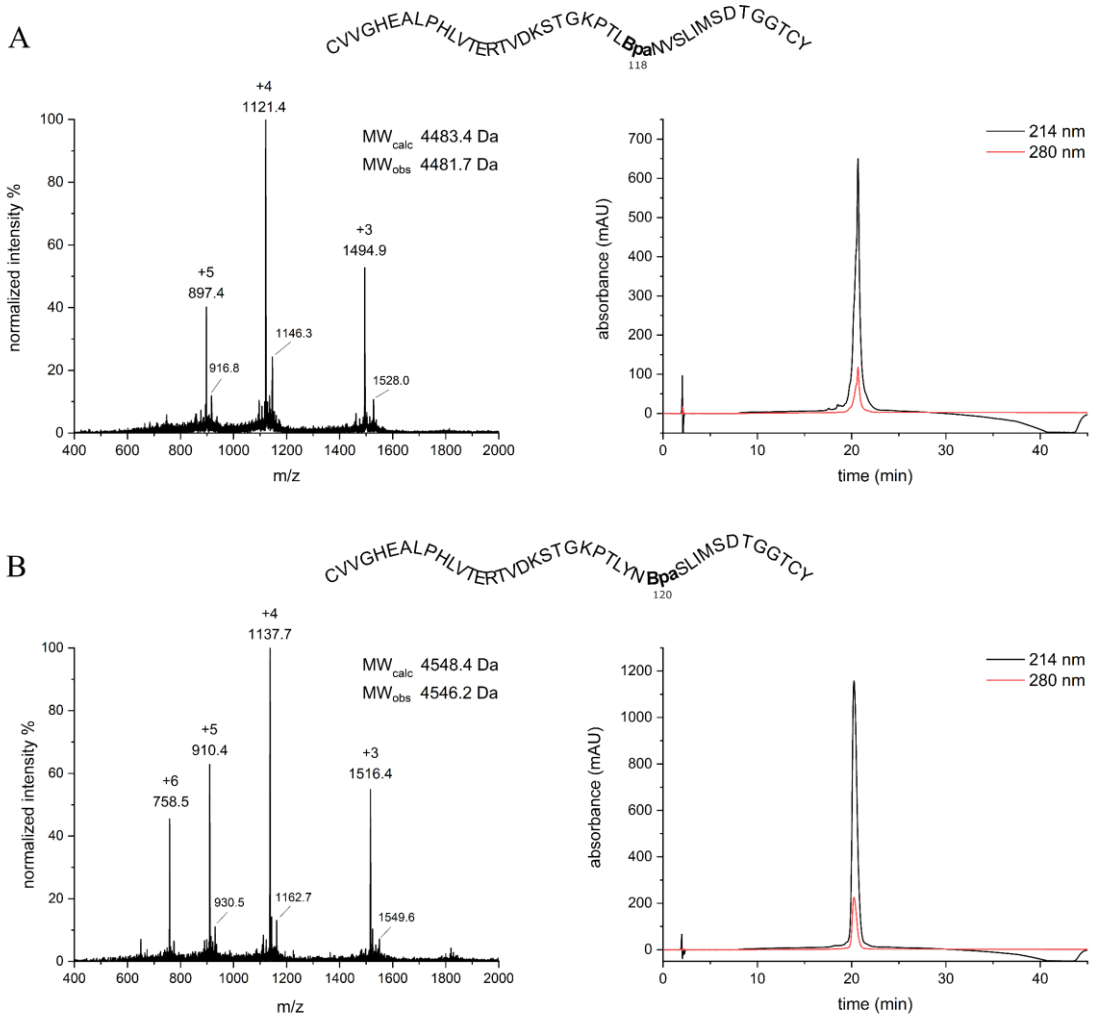


Figure 29: Characterisation of modified IgM CH4 peptides 92-132. **A)** ESI-MS spectrum and RP-HPLC chromatogram of modified peptide **M1** bearing the benzophenone moiety (Bpa) in position 118. The abundant labelled side product with the additional mass of +99 Da can be ascribed to a valine addition product. **B)** ESI-MS spectrum and RP-HPLC chromatogram of modified peptide **M2** bearing the benzophenone moiety (Bpa) in position 120.

The combination of manual and automated Fmoc-based SPPS has proven to be a suitable method for the synthesis of Bpa modified IgM CH4 peptides (**M1** and **M2**), which allowed the monitoring of the Bpa-residue incorporation and generated the peptides in a sufficient purity and yield especially when considering the length and complexity of the sequence (41 amino

acid residues). The benzophenone (Bpa) moiety was an excellent choice as a photoactivatable crosslinking reagent since its stability allows handling under ambient light as well as the use of microwave assisted SPPS. In addition, this crosslinker is readily commercially available as Fmoc-protected amino acid (Fmoc-Bpa-OH) and affordable compared to other photocrosslinking reagents [60]. Notably, the incorporation of the Bpa residue in position 118 was much more challenging than in position 120, which could be explained by steric hindrance through the relatively large trityl- protection groups of the neighbouring asparagine side chain compared to the coupling next to serine and its oxygen bound *t*-butyl protection group in position 120. All efforts trying to couple the Fmoc-Bpa-OH to asparagine using the activating agents HBTU or HATU and double couplings failed and resulted in only minor product and a majority of Fmoc deprotected starting material. Switching to Oxyma and DIC as coupling reagents as well as increasing the equivalent of amino acid (4 instead of 2.5) turned out to be a successful method for the coupling of Fmoc-Bpa-OH to asparagine. Additionally, the appearance of a rather prominent valine addition product (+99 Da), which was unfortunately not separable from the main product, was observed during the synthesis of **M1** (Figure 29A) which contributed to the difficulty of a sufficient synthesis.

3.2.2 Expression and purification of IgM CH4 1-91 fusion protein

Expression of IgM CH4 intein fusion protein was performed according to the protocol of Dr. Aleksandr Kravchuk which utilizes the commercially available IMPACT^{TM} Kit for the generation of Mxe GyrA (Mycobacterium xenopi gyrA) intein fusion protein [61,62]. The expression on a 4 L scale yielded the desired C-terminal cysteine peptide in satisfactory purity. Monitoring of the expression by SDS-PAGE is depicted in Figure 30A. The overexpression of the fusion protein with the molecular weight of ~39 kDa was observed as a single band at an apparent molecular weight of ~48 kDa. The less intense bands after 24 h of overexpression indicate that the expression time should be reduced due to partial degradation of the fusion protein. Since previous expressions showed that the construct was expressed in inclusion bodies, the obtained pellets from clearing the cell lysate via centrifugation where washed and solubilized in guanidinium hydrochloride containing buffer as described in section 2.3.2.1. The subsequent purification of the fusion protein via Ni-NTA affinity chromatography was verified via SDS-PAGE (Figure 30B) and the obtained peak fractions pooled and dialysed against 4 M Urea, TBS before the MesNa induced intein cleavage.

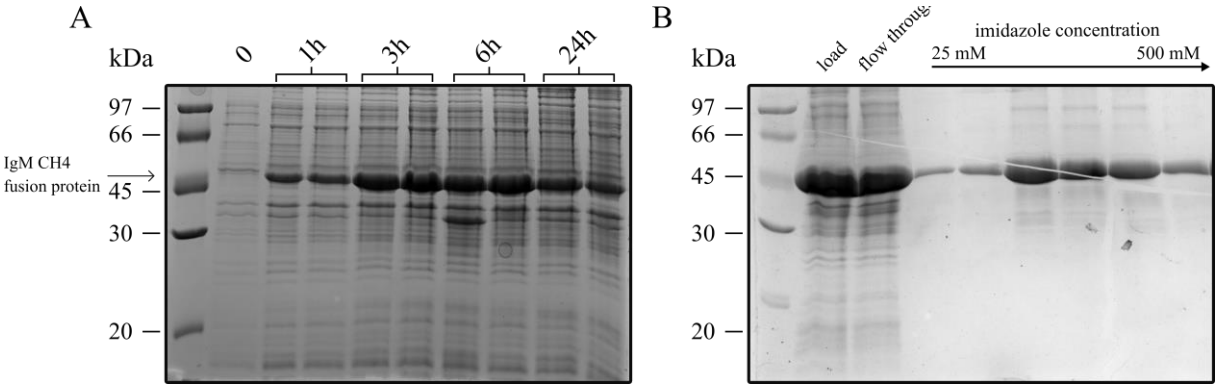


Figure 30: A) Expression monitoring of IgM CH4 1-91 fusion protein via SDS-PAGE. The two lanes for the time points t=1 h, t=3 h, t=6 h and t=24 h represent the expression in two separated 2L cultures. The strong overexpression of the fusion protein (39 kDa) is clearly visible as a band at an apparent molecular weight of ~48 kDa. **B)** SDS-PAGE analysis of the purification of IgM CH4 1-91 fusion protein by Ni-NTA affinity chromatography. The protein was eluted at a flow rate of 5 mL/min using a linear gradient from 20-500 mM imidazole (lanes of the samples taken during the elution are labelled with an arrow indicating the gradient) in 6M Gdn·HCl, 100 mM Tris, pH 7.5.

3.2.3 MesNa induced intein-mediated cleavage of IgM CH4 1-91 fusion protein

The fusion protein purified via Ni-NTA affinity chromatography was dialysed against 4 M Urea TBS buffer as described in section 2.3.2.1 without any observable precipitation of the protein. The MesNa induced intein-mediated cleavage, performed by the dropwise addition of 1 M aqueous MesNa solution to a final MesNa concentration of 0.5 M while stirring, lead to no or minor precipitation of the fusion protein. The cleavage was considered as completed after 48 hours resulting in ~70% cleaved protein based on the SDS-PAGE analysis (Figure 31).

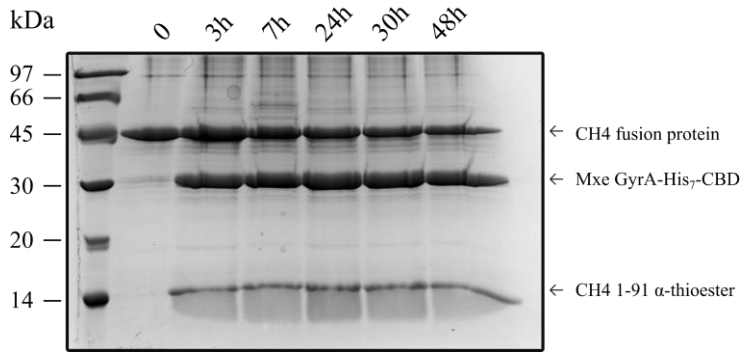


Figure 31: Monitoring of MesNa induced intein-mediated cleavage of CH4 1-91 fusion protein by SDS-PAGE. The cleaved Mxe GyrA-His_7-CBD intein (28 kDa) is running at an apparent molecular weight of ~30 kDa while the cleaved CH4 1-91 thioester (10 kDa) is visible as a thin band at an apparent molecular weight of ~15 kDa.

Purification of the CH4 1-91 α-thioester was achieved by RP-HPLC and the fractions analysed via ESI-MS. The fractions were pooled according to their masses and lyophilized. Characterisation of the final CH4 1-91 *α*-thioester by ESI-MS and RP-HPLC analysis is depicted in Figure 32**.** The protein yield was 6 mg/L expression medium. Yield loss of the highly overexpressed fusion protein can be attributed to the incomplete MesNa induced intein-mediated cleavage that only takes place up to 60 – 70% (Figure 31). This can be explained by the properties of the intein, which has to be correctly folded in order to undergo a thiol induced cleavage. Since cleavage conditions had to include a chaotropic salt (2M urea) for solubility reasons, it can be assumed that only a fraction of the fusion protein exhibits the correct folding for the thiol induced reaction. All efforts to improve the yield of the thiol induced intein cleavage where unsuccessful so far. The focus of the project was set on the generation of the synthetic peptide fragments and the ligation to the recombinant protein.

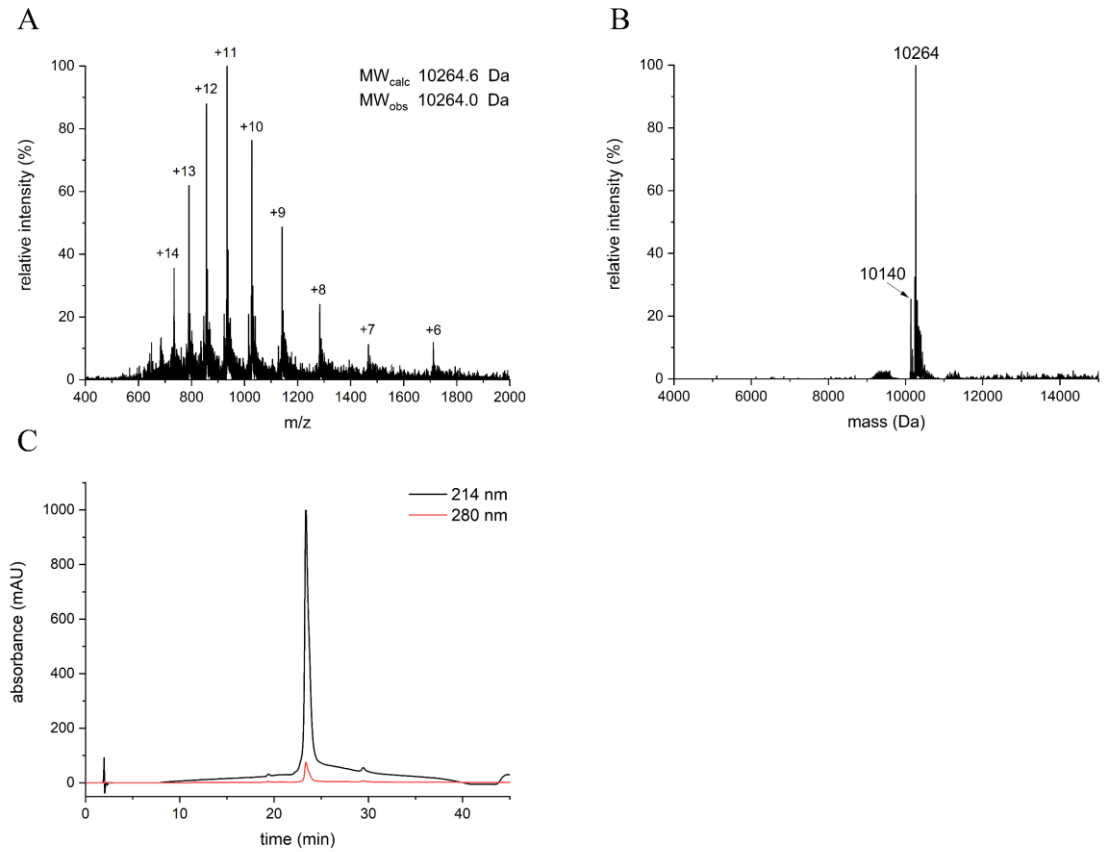


Figure 32: Characterisation of IgM CH4 1-91 α-thioester. **A)** ESI-MS spectrum of purified CH4 thioester shows that the observed mass (MW_{obs} 10264.0 Da) is in good agreement with the calculated mass (MW_{calc} 10264.6 Da). **B)** Deconvoluted mass spectrum of purified IgM CH4 1-91 α-thioester. The additional observed mass of 10140.0 Da can be assigned to the hydrolysed thioester (MW_{calc} 10139.4 Da). **C)** RP-HPLC chromatogram of purified CH4 α-thioester.

3.2.4 EPL of IgM CH4 to peptide M1 and M3

3.2.4.1 Expressed protein ligation of IgM CH4 1-91 α-thioester to synthetic peptide M1

Successful ligation of IgM CH4 1-91 α-thioester to Bpa-modified peptide **M1** was verified via LC-MS and the reaction quenched after 48 h. A detailed description of the ligation and purification conditions can be found in section 2.6.1 EPL **of IgM CH4-thioester to peptide M1 and peptide M3**During the purification the product was eluted as multiple peaks between 42.2 and 49.2 min (Figure 33A). Fractions were analysed via ESI-MS, combined according to their masses and lyophilized. Final ESI-MS and RP-HPLC characterisation of the ligation product is depicted in Figure 33. The product was obtained with a yield of 53% (7mg) based on the theoretical yield.

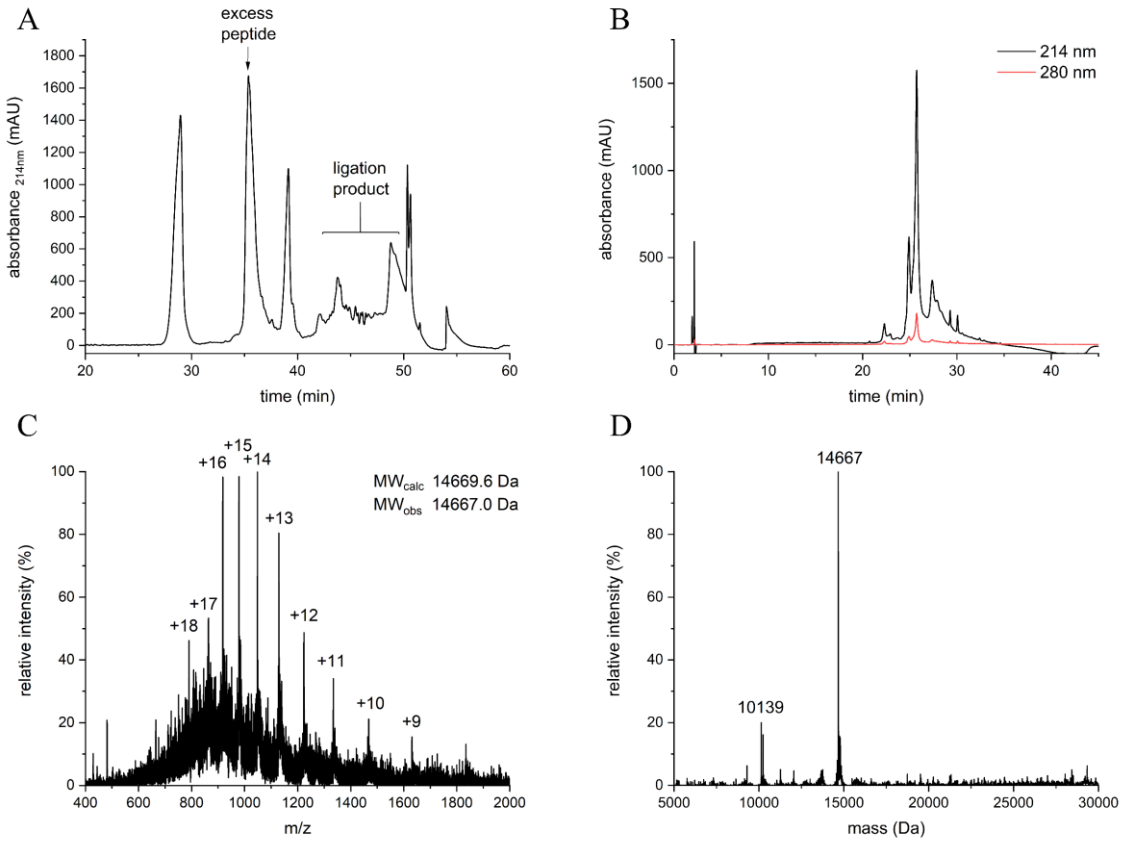


Figure 33: A) RP-HPLC purification of the ligation of peptide **M1** to 1-91 CH4 α-thioester. Purification was performed on a C4 Kromasil column at a flow rate of 20 mL/min using a linear gradient of 5-45% buffer B (ACN + 0.08% TFA) in buffer A (ddH_2O + 0.1% TFA). The product mass was observed in range of multiple peaks between 42.2 and 49.2 min. **B)** RP-HPLC chromatogram of purified ligation product IgM CH4 1-132. **C)** ESI-MS spectra. The observed mass (MW_{obs} 14667.0) is in good agreement with the calculated mass (MW_{calc} 14669.6 Da). **D)** Deconvoluted ESI-MS spectra. The observed mass of 10139.0 Da can be assigned to the hydrolysed CH4 α-thioester (MW_{calc} 10139.4 Da).

The elution of the product as a series of broad peaks made purification impossible and asked for an optimisation of the purification process. Initial thoughts that the peptide could have undergone accidental photocrosslinking due to UV-light exposure during the purification via RP-HPLC were discarded since the wild-type peptide not containing Bpa showed a similar elution profile (S14). The optimisation of the purification process was performed using ligations of the wild-type peptide, without the presence of the photocrosslinker residue, and is discussed in the following section (3.2.4.2 ***Expressed protein ligation of IgM CH4 1-91 to wild-type peptide M3***).

3.2.4.2 Expressed protein ligation of IgM CH4 1-91 to wild-type peptide M3

Ligation and purification conditions were optimised by changing the thiol additive from thiophenol to MPAA and performing the RP-HPLC purification at 60°C (see section 2.6.1 ***EPL of IgM CH4-thioester to peptide M1 and peptide M3*** for a detailed description of ligation and purification conditions). The reaction was monitored via LC-MS and SDS-PAGE (Figure 34) and quenched after 21 h. The reaction and subsequent purification yielded the product (15%, 1.5 mg) in satisfactory purity. ESI-MS and RP-HPLC characterisation of the purified ligation product are depicted in Figure 35.

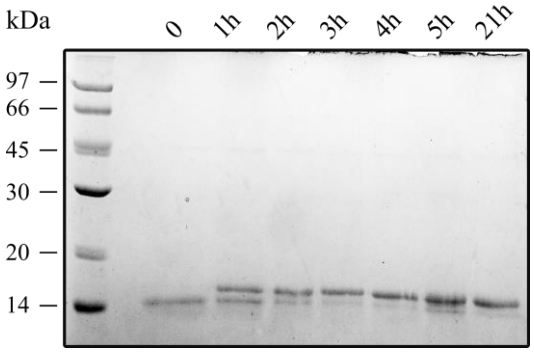


Figure 34: SDS-Page monitoring of the ligation reaction between IgM CH4 α-thioster (MW_{calc} 10139.4 Da) and wild-type peptide **M3** (MW_{calc} 4394.0 Da). Calculated molecular weight for the ligation product is 14515.5 Da. The reaction was considered as completed after 5 h but left overnight before quenching and purification.

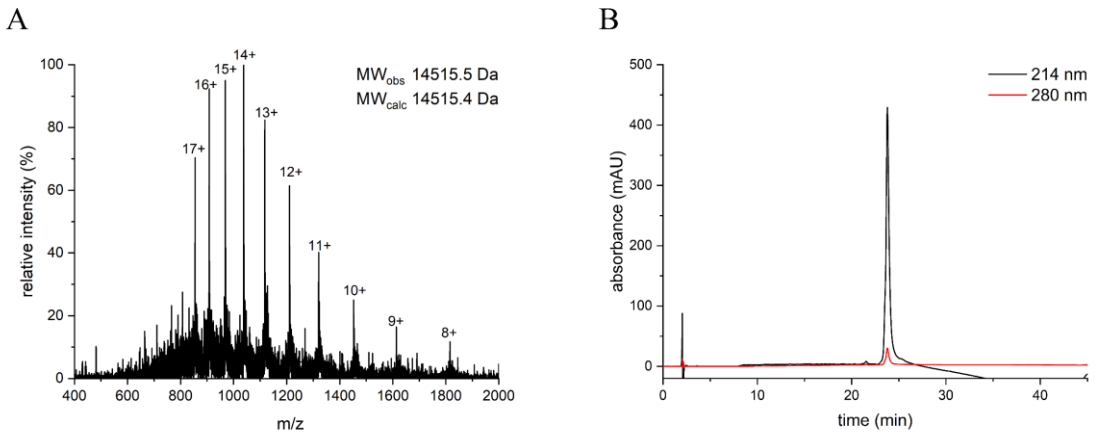


Figure 35: Characterisation of semisynthetic unmodified IgM CH4 1-132. **A)** ESI-MS spectrum. The observed molecular weight of 14515.5 Da is in agreement with the calculated value of 14515.4 Da. **B)** RP-HPLC chromatogram of the purified ligation product IgM CH4 1-132.

Successful ligations of IgM CH4 α-thioester to both the Bpa modified (**M1)** and unmodified (**M3**) peptide CH4 92-132 were accompanied by challenging purification processes since the product eluted as a range of broad peaks, which made purification impossible (Figure 33, S14). The fact that both ligations showed a similar RP-HPLC purification profile indicated that the peak smearing effects can't be attributed to the incorporated photocrosslinker and is an intrinsic behaviour of the ligation product. A reasonable explanation for this unusual behaviour during the RP-HPLC separation process would be the formation of oligomers. These higher molecular weight species of different sizes, which could dynamically form and dissociate during the elution process, would exhibit different retention times and give rise to a chromatogram that resembles the observed one. This would also be supported by the fact that the CH4 domain bearing the pivotal tail piece peptide at its C-terminus is known to be the smallest subdomain that can form higher-order oligomers and is therefore crucial for the topology of IgM [41]. The significant improvement of the elution behaviour of the wild-type ligation product at 60°C additionally supports this theory since higher temperature could prevented oligomer assembly (S14). Furthermore, care had to be taken that the modified ligation product is only minimally exposed to higher energy irradiation since excitation of the photocrosslinker by the UV detector of a RP-HPLC system could cause accidental premature crosslinking. Therefore, the purification protocol was optimised on a WatersAutopurification LC-MS system which provides a flow splitter and uses minimal amounts of sample for the LC-MS monitoring.

4 Outlook

Major progress was made in the recently started semi-synthetic approach for the generation of site-specifically phosphorylated and segmentally isotope labelled Hsp90 NTD. Next steps would be focused on the optimisation of the recombinant expression of $Hsp90^{C2}$ in order to ensure that a sufficient yield can be obtained during the expression in ^{15}N and ^{13}C containing minimal medium. Since the overexpression of various undesired proteins was observed during the expression of $Hsp90^{C2}$, which made purification very difficult, a change on the genetic level could be considered. This would include the design of a new plasmid with an altered codon sequence and the generation of a new cell line stock for the expression of $Hsp90^{C2}$ as the next step for a fundamental improvement of the generation of the recombinant part. Additionally, the optimisation of the TEV cut and the ensuing purification via RP-HPLC is desirable since it could increase the yield significantly. Once sufficient amounts of the labelled recombinant part can be generated in a reliable manner, and an optimisation of the NCL and desulfurization reaction conditions was performed, the desired modified and labelled Hsp90 NTD can be yielded and used for NMR studies.

Although the project focusing on the semi-synthesis of IgM CH4 domain already had an established protocol for the recombinant expression of the CH4 intein-fusion protein and ligation to the synthetic unmodified tailpiece peptide, crucial optimisation steps had to be performed in order to obtain sufficient amounts of the desired protein variants. The optimised ligation and purification conditions found can be applied to the ligation of the crosslinker-modified peptides to the recombinant α-thioester, in order to obtain the modified IgM CH4 domain which can be submitted for photocrosslinking-MS studies.

5 Conclusion

5.1 Semi-synthesis of Hsp90 NTD bearing PTMs

A strategy for the semi-synthesis of human α-Hsp90 NTD was devised that successfully used SPPS for the generation of specifically PTM modified peptides of the N-terminus and recombinant expression for the generation of the major part of the domain bearing the pivotal C-terminal cysteine for the NCL. The challenging generation of the recombinant part gave intriguing insights into the sensitivity of this domain since the truncation or alteration of a short N-terminal sequence of 19 amino acids, drastically changed the solubility properties of the protein and indicates an important role of this specific sequence. Successful characterisation of the synthetic peptides via CD and NMR spectroscopy displays another achievement and can be considered a building block towards understanding the complex influences of PTMs on the function and structure of this protein, into which we will be able to gain important insights once the fully segmentally labelled and modified variants of the semi-synthetic domain become available.

5.2 Semi-synthesis of modified IgM CH4 domain

C-terminal peptides, bearing the photocrosslinker in two specific positions, of the IgM CH4 domain were successfully synthesised via SPPS and the ligation to the recombinant α-thioester segment optimised using the wild-type peptide. Challenges during the purification of the ligation product, which can be attributed to the formation of higher-order oligomers or aggregates, were resolved by performing RP-HPLC at higher temperatures. When this optimized strategy is applied to a larger scale ligation using the Bpa modified peptide it should enable the generation of larger amounts of the desired modified ligation product, which will allow further insights into oligomer assembly of this important class of immunoglobins.

5.3 General words

Both projects successfully utilize the diversity of synthesis strategies for the generation of N- or C-terminal modified semi-synthetic proteins based on the native chemical ligation principle. In both cases the generation of the synthetic peptide fragment displayed a manageable task in contrast to the demanding generation of the recombinant domains, which illustrates the reliability and robustness of chemical peptide synthesis methods.

While the underlying principle for the generation of the Hsp90 NTD recombinant fragment, based on the expression of a truncated mutant with cysteine at the ligation site, may seem quite simple and straightforward, the actual implementation was due to the unexpected altered protein properties extremely challenging. The expression of the recombinant IgM CH4 domain as an intein fusion protein and the ensuing thiol mediated cleavage delivered satisfactory results, even though yield losses had to be accepted due to incomplete thiol-induced intein cleavage.

6 References

1. Nilsson, B. L., Soellner, M. B. & Raines, R. T. Chemical Synthesis of Proteins. *Annu Rev Biophys Biomol Struct* **34**, 91–118 (2005).
2. Fischer, E. & Fourneau, E. Ueber einige Derivate des Glykocolls. *Berichte der deutschen chemischen Gesellschaft* **34**, 2868–2877 (1901).
3. Bergmann, M. & Zervas, L. Über ein allgemeines Verfahren der Peptid-Synthese. *Berichte der deutschen chemischen Gesellschaft (A and B Series)* **65**, 1192–1201 (1932).
4. Merrifield, R. B. Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **85**, 2149–2154 (1963).
5. Behrendt, R., White, P. & Offer, J. Advances in Fmoc solid-phase peptide synthesis. *J. Pept. Sci.* **22**, 4–27 (2016).
6. Kulkarni, S. S., Sayers, J., Premdjee, B. & Payne, R. J. Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. *Nature Reviews Chemistry* **2**, 0122 (2018).
7. Hackenberger, C. P. R. & Schwarzer, D. Chemoselective ligation and modification strategies for peptides and proteins. *Angew. Chem. Int. Ed. Engl.* **47**, 10030–10074 (2008).
8. Dawson, P. E., Muir, T. W., Clark-Lewis, I. & Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
9. Conibear, A. C., Watson, E. E., Payne, R. J. & Becker, C. F. W. Native chemical ligation in protein synthesis and semi-synthesis. *Chem. Soc. Rev.* **47**, 9046–9068 (2018).
10. Agouridas, V. *et al.* Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **119**, 7328–7443 (2019).
11. Muir, T. W., Sondhi, D. & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *PNAS* **95**, 6705–6710 (1998).
12. Kane, P. M. *et al.* Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* **250**, 651–657 (1990).
13. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).
14. Perler, F. B. Protein Splicing Mechanisms and Applications. *IUBMB Life* **57**, 469–476 (2005).
15. Yan, L. Z. & Dawson, P. E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **123**, 526–533 (2001).
16. Rohde, H. & Seitz, O. Ligation-desulfurization: a powerful combination in the synthesis of peptides and glycopeptides. *Biopolymers* **94**, 551–559 (2010).
17. Wan, Q. & Danishefsky, S. J. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. *Angew. Chem. Int. Ed. Engl.* **46**, 9248–9252 (2007).
18. Zheng, J.-S., Tang, S., Qi, Y.-K., Wang, Z.-P. & Liu, L. Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nature Protocols* **8**, 2483–2495 (2013).
19. Liu, D. & Cowburn, D. Segmental Isotopic Labeling of Proteins for NMR Study Using Intein Technology. *Methods Mol. Biol.* **1495**, 131–145 (2017).
20. Csermely, P., Schnaider, T., So″ti, C., Prohászka, Z. & Nardai, G. The 90-kDa Molecular Chaperone Family. *Pharmacology & Therapeutics* **79**, 129–168 (1998).
21. Biebl, M. M. & Buchner, J. Structure, Function, and Regulation of the Hsp90 Machinery. *Cold Spring Harb Perspect Biol* (2019) doi:10.1101/cshperspect.a034017.

22. Jaeger, A. M. & Whitesell, L. HSP90: Enabler of Cancer Adaptation. *Annu. Rev. Cancer Biol.* (2018) doi:10.1146/annurev-cancerbio-030518-055533.
23. Kumalo, H. M., Bhakat, S. & Soliman, M. E. Heat-shock protein 90 (Hsp90) as anticancer target for drug discovery: an ample computational perspective. *Chem Biol Drug Des* **86**, 1131–1160 (2015).
24. Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nature Reviews Molecular Cell Biology* **18**, 345–360 (2017).
25. Li, J., Richter, K. & Buchner, J. Mixed Hsp90-cochaperone complexes are important for the progression of the reaction cycle. *Nat. Struct. Mol. Biol.* **18**, 61–66 (2011).
26. Chen, B., Piel, W. H., Gui, L., Bruford, E. & Monteiro, A. The HSP90 family of genes in the human genome: Insights into their divergence and evolution. *Genomics* **86**, 627–637 (2005).
27. Sreedhar, A. S., Kalmár, E., Csermely, P. & Shen, Y.-F. Hsp90 isoforms: functions, expression and clinical importance. *FEBS Lett.* **562**, 11–15 (2004).
28. Mollapour, M. & Neckers, L. Post-translational modifications of Hsp90 and their contributions to chaperone regulation. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1823**, 648–655 (2012).
29. Xu, W. *et al.* Hsp90 middle domain phosphorylation initiates a complex conformational program to recruit the ATPase-stimulating cochaperone Aha1. *Nature Communications* **10**, 2574 (2019).
30. Bachman, A. B. *et al.* Phosphorylation induced cochaperone unfolding promotes kinase recruitment and client class-specific Hsp90 phosphorylation. *Nat Commun* **9**, (2018).
31. Mollapour, M., Tsutsumi, S., Kim, Y. S., Trepel, J. & Neckers, L. Casein kinase 2 phosphorylation of Hsp90 threonine 22 modulates chaperone function and drug sensitivity. *Oncotarget* **2**, 407–417 (2011).
32. Mollapour, M. *et al.* Swe1Wee1-dependent tyrosine phosphorylation of Hsp90 regulates distinct facets of chaperone function. *Mol. Cell* **37**, 333–343 (2010).
33. Heyman, B. & Shulman, M. J. Structure, Function, and Production of Immunoglobulin M (IgM). in 1–14 (Elsevier, 2016). doi:10.1016/b978-0-12-374279-7.05001-3.
34. Ehrenstein, M. R. & Notley, C. A. The importance of natural IgM: scavenger, protector and regulator. *Nat Rev Immunol* **10**, 778–786 (2010).
35. Boes, M. Role of natural and immune IgM antibodies in immune responses. *Molecular Immunology* **37**, 1141–1149 (2000).
36. Frutiger, S., Hughes, G. J., Paquet, N., Luethy, R. & Jaton, J. C. Disulfide bond assignment in human J chain and its covalent pairing with immunoglobulin M. *Biochemistry* **31**, 12643–12647 (1992).
37. Niles, M. J., Matsuuchi, L. & Koshland, M. E. Polymer IgM assembly and secretion in lymphoid and nonlymphoid cell lines: evidence that J chain is required for pentamer IgM synthesis. *Proc Natl Acad Sci U S A* **92**, 2884–2888 (1995).
38. Randall, T. D., Brewer, J. W. & Corley, R. B. Direct evidence that J chain regulates the polymeric structure of IgM in antibody-secreting B cells. *J. Biol. Chem.* **267**, 18002–18007 (1992).
39. Wiersma, E. J. & Shulman, M. J. Assembly of IgM. Role of disulfide bonding and noncovalent interactions. *J. Immunol.* **154**, 5265–5272 (1995).
40. Müller, R. *et al.* High-resolution structures of the IgM Fc domains reveal principles of its hexamer formation. *PNAS* **110**, 10183–10188 (2013).
41. Pasalic, D. *et al.* A peptide extension dictates IgM assembly. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E8575–E8584 (2017).
42. Gill, S. C. & von Hippel, P. H. Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182**, 319–326 (1989).

43. The CCPN data model for NMR spectroscopy: Development of a software pipeline - Vranken - 2005 - Proteins: Structure, Function, and Bioinformatics - Wiley Online Library. https://onlinelibrary.wiley.com/doi/abs/10.1002/prot.20449.
44. Aeschbacher, T., Schubert, M. & Allain, F. H.-T. A procedure to validate and correct the 13C chemical shift calibration of RNA datasets. *J. Biomol. NMR* **52**, 179–190 (2012).
45. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. & Sykes, B. D. 1H, 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR* **5**, 67–81 (1995).
46. Conibear, A. C., Rosengren, K. J., Becker, C. F. W. & Kaehlig, H. Random coil shifts of posttranslationally modified amino acids. *J Biomol NMR* (2019) doi:10.1007/s10858-019-00270-4.
47. Lees-Miller, S. P. & Anderson, C. W. Two human 90-kDa heat shock proteins are phosphorylated in vivo at conserved serines that are phosphorylated in vitro by casein kinase II. *J. Biol. Chem.* **264**, 2431–2437 (1989).
48. Lees-Miller, S. P. & Anderson, C. W. The human double-stranded DNA-activated protein kinase phosphorylates the 90-kDa heat-shock protein, hsp90 alpha at two NH2-terminal threonine residues. *J. Biol. Chem.* **264**, 17275–17280 (1989).
49. Li, H. & Dong, S. Recent advances in the preparation of Fmoc-SPPS-based peptide thioester and its surrogates for NCL-type reactions. *Sci. China Chem.* **60**, 201–213 (2017).
50. Huang, H. & Rabenstein, D. L. A cleavage cocktail for methionine-containing peptides. *The Journal of Peptide Research* **53**, 548–553 (1999).
51. Zhang, H. *et al.* A dynamic view of ATP-coupled functioning cycle of Hsp90 N-terminal domain. *Sci Rep* **5**, 9542 (2015).
52. Karagöz, G. E. *et al.* N-terminal domain of human Hsp90 triggers binding to the cochaperone p23. *Proc Natl Acad Sci U S A* **108**, 580–585 (2011).
53. Mavridis, L. & Janes, R. W. PDB2CD: a web-based application for the generation of circular dichroism spectra from protein atomic coordinates. *Bioinformatics* **33**, 56–63 (2017).
54. Martinez-Yamout, M. A. *et al.* Localization of Sites of Interaction between p23 and Hsp90 in Solution. *Journal of Biological Chemistry* **281**, 14457–14464 (2006).
55. Bienkiewicz, E. A. & Lumb, K. J. Random-coil chemical shifts of phosphorylated amino acids. *J Biomol NMR* **15**, 203–206 (1999).
56. Dyson, H. J. & Wright, P. E. Defining Solution Conformations of Small Linear Peptides. *Annual Review of Biophysics and Biophysical Chemistry* **20**, 519–538 (1991).
57. Theillet, F.-X. *et al.* Cell signaling, post-translational protein modifications and NMR spectroscopy. *J Biomol NMR* **54**, 217–236 (2012).
58. Du, J.-T. *et al.* Low-barrier hydrogen bond between phosphate and the amide group in phosphopeptide. *J. Am. Chem. Soc.* **127**, 16350–16351 (2005).
59. Amaral, M. *et al.* Protein conformational flexibility modulates kinetics and thermodynamics of drug binding. *Nature Communications* **8**, 2276 (2017).
60. Chin, J. W., Martin, A. B., King, D. S., Wang, L. & Schultz, P. G. Addition of a photocrosslinking amino acid to the genetic code of Escherichiacoli. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 11020–11024 (2002).
61. Evans, T. C., Benner, J. & Xu, M. Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* **7**, 2256–2264 (1998).
62. Telenti, A. *et al.* The Mycobacterium xenopi GyrA protein splicing element: Characterization of a minimal intein. *J. Bacteriol.***179** (20):6378 (1997)

7 Supplementary Data

A

```
CAT ATG CAC CAT CAT CAC CAC CAT GAG AAC CTT TAT TTC CAA TGC TTC CAA
GCA GAA ATT GCT CAA CTG ATG AGC CTG ATC ATT AAC ACC TTC TAC TCG AAT
AAA GAA ATT TTC CTT CGC GAG CTT ATT AGT AAT TCT AGC GAT GCT TTA GAT
AAA ATC CGT TAT GAG TCG CTT ACG GAT CCG TCC AAG TTG GAC AGC GGC AAA
GAG CTT CAT ATT AAT CTG ATC CCC AAC AAA CAG GAT CGC ACG CTG ACA ATC
GTC GAT ACG GGT ATT GGG ATG ACG AAG GCG GAC CTG ATC AAC AAT CTT GGA
ACT ATT GCC AAA TCA GGA ACG AAA GCA TTC ATG GAA GCG TTG CAA GCT GGC
GCT GAC ATC AGC ATG ATC GGA CAA TTT GGG GTA GGG TTC TAT TCA GCC TAT
CTT GTG GCA GAA AAA GTC ACC GTA ATC ACC AAG CAT AAC GAT GAC GAA CAA
TAT GCT TGG GAG AGC AGC GCA GGT GGC AGT TTT ACT GTG CGC ACG GAC ACT
GGG GAG CCT ATG GGA CGC GGT ACG AAG GTA ATT TTG CAC TTG AAG GAG GAT
CAG ACT GAA TAC CTT GAA GAG CGT CGC ATT AAG GAA ATT GTA AAA AAG CAC
TCA CAG TTC ATC GGA TAC CCA ATT ACC TTA TTC GTC GAA AAG GAA CGT GAT
AAA GAG GTT TCT GAT GAC GAA GCA TGA CTC GAG
```

B

```
MHHHHHHENLYFQCFQAEIAQLM SLIINTFYSN KEIFLRELIS NSSDALDKIR YESLTDPSKL
DSGKELHINL IPNKQDRTLT IVDTGIGMTK ADLINNLGTI AKSGTKAFME ALQAGADISM
IGQFGVGFYS AYLVAEKVTV ITKHNDDEQY AWESSAGGSF TVRTDTGEPM GRGTKVILHL
KEDQTEYLEE RRIKEIVKKH SQFIGYPITL FVEKERDKEV SDDEA
```

S1: A) Codon sequence of Hsp90^{C1}. **B)** Amino acid sequence of Hsp90^{C1}.

A

```
CAT ATG CCG GAA GAA ACC CAG ACA CAA GAC CAA CCA ATG GAG GAA GAG GAG GTG GAA
ACA TTC GAG AAC TTG TAC TTT CAG TGC TTC CAG GCG GAA ATC GCG CAG TTA ATG AGT
TTA ATC ATC AAT ACA TTC TAC TCT AAC AAG GAA ATT TTT TTG CGT GAG CTG ATT TCC
AAT TCT TCC GAC GCA CTG GAC AAG ATT CGT TAC GAG TCC CTG ACG GAC CCA AGC AAA
TTG GAC AGC GGG AAA GAA TTA CAC ATC AAC CTT ATC CCG AAT AAA CAG GAC CGC ACA
CTG ACT ATT GTA GAT ACC GGA ATT GGC ATG ACA AAG GCT GAC CTG ATC AAC AAT TTG
GGG ACG ATT GCC AAG AGT GGC ACA AAG GCA TTC ATG GAG GCC CTT CAG GCA GGT GCT
GAC ATC AGT ATG ATT GGA CAA TTT GGC GTG GGC TTT TAT AGT GCG TAC TTA GTA GCG
GAG AAA GTG ACG GTA ATC ACT AAA CAT AAT GAT GAC GAG CAG TAC GCT TGG GAA TCA
TCC GCT GGG GGT AGC TTC ACT GTT CGC ACG GAC ACT GGA GAA CCT ATG GGT CGC GGA
ACT AAA GTT ATT TTG CAC CTT AAA GAG GAT CAG ACT GAA TAT CTT GAG GAA CGT CGC
ATT AAG GAA ATT GTG AAG AAG CAT AGT CAG TTT ATT GGC TAC CCT ATC ACA TTG TTT
GTT GAG AAA GAA CGC GAT AAA GAG GTT TCG GAC GAT GAA GCG CAT CAT CAC CAT CAT
CAT TGA CTC GAG
```

B

```
MPEETQTQDQ PMEEEEVETF ENLYFQ CFQAEIAQLM SLIINTFYSN KEIFLRELIS
NSSDALDKIR YESLTDPSKL DSGKELHINL IPNKQDRTLT IVDTGIGMTK
ADLINNLGTI AKSGTKAFME ALQAGADISM IGQFGVGFYS AYLVAEKVTV
ITKHNDDEQY AWESSAGGSF TVRTDTGEPM GRGTKVILHL KEDQTEYLEE
RRIKEIVKKH SQFIGYPITL FVEKERDKEV SDDEAHHHHHH
```

S2: A) Codon sequence of Hsp90^{C2}. **B)** Amino acid sequence of Hsp90^{C2}.

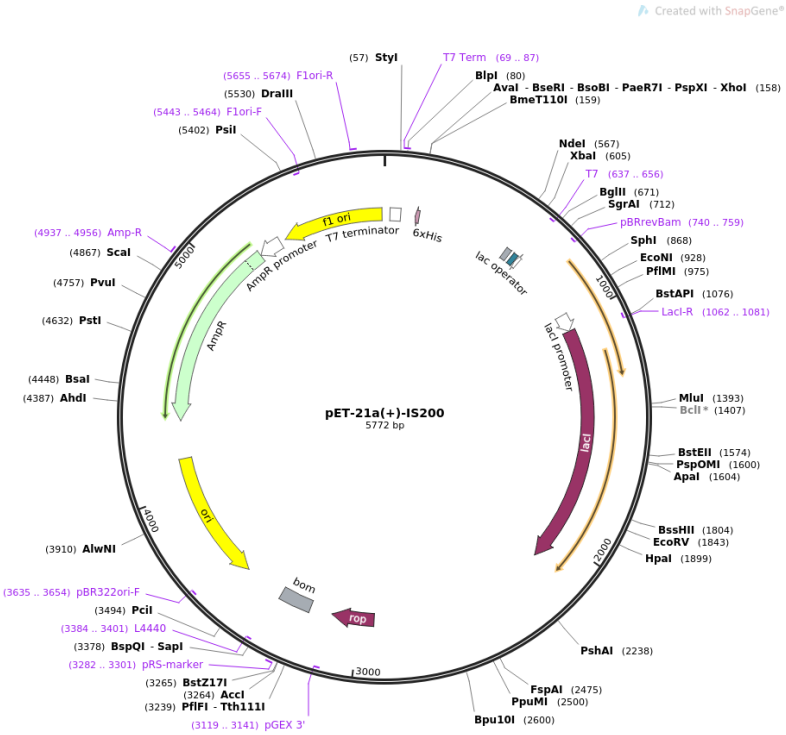
A

```
CAT ATG CCA GAG GAA ACA CAG ACA CAG GAT CAA CCA ATG GAG GAG GAG GAA
GTC GAG ACA TTC GCT TTC CAA GCC GAG ATT GCT CAA CTT ATG TCG CTT ATC
ATC AAT ACT TTC TAT TCC AAT AAA GAA ATT TTC CTT CGC GAA CTT ATT TCC
AAT AGT TCG GAT GCA TTA GAC AAA ATC CGC TAT GAA AGT TTA ACG GAC CCA
TCC AAA CTT GAT AGT GGC AAA GAA TTG CAC ATT AAT TTA ATT CCG AAC AAG
CAG GAC CGT ACT CTG ACC ATT GTT GAT ACT GGC ATT GGA ATG ACT AAA GCT
GAT CTT ATC AAT AAC TTG GGG ACA ATT GCT AAG TCT GGA ACA AAA GCC TTC
ATG GAG GCG CTT CAA GCT GGG GCG GAT ATT AGC ATG ATT GGG CAA TTT GGT
GTT GGT TTT TAT TCA GCG TAC TTG GTG GCC GAA AAG GTC ACT GTG ATC ACA
AAG CAC AAC GAC GAT GAG CAA TAT GCA TGG GAG TCT AGC GCG GGC GGT AGC
TTC ACT GTA CGC ACT GAT ACG GGT GAG CCT ATG GGT CGT GGA ACA AAA GTG
ATC TTG CAT TTG AAA GAG GAC CAA ACT GAA TAC CTT GAA GAA CGC CGC ATC
AAA GAA ATC GTA AAA AAA CAC AGT CAG TTT ATT GGC TAT CCT ATC ACG CTT
TTT GTT GAG AAG GAG CGC GAT AAA GAG GTT AGT GAC GAC GAA GCG CAT CAT
CAC CAC CAT CAT TGA CTC GAG
```

B

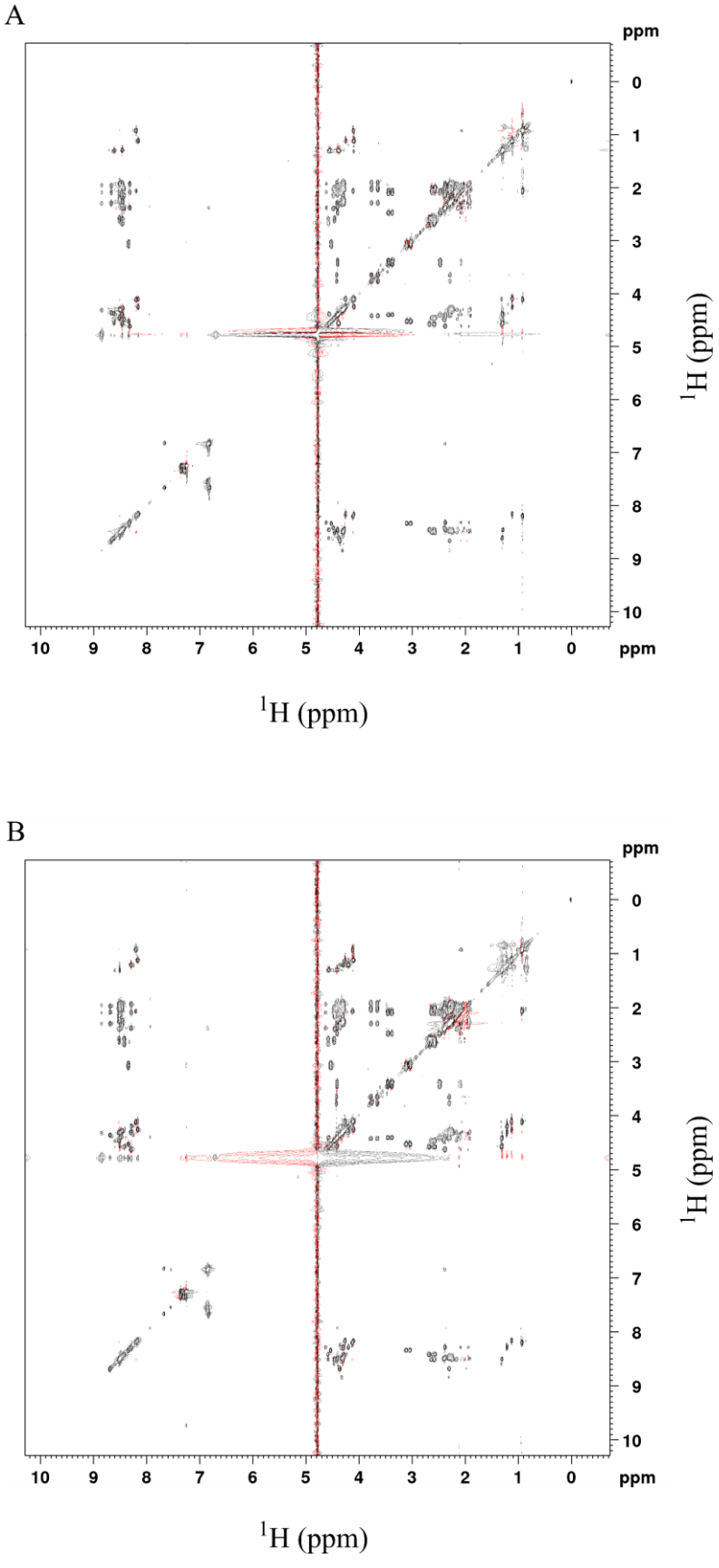
```
MPEETQTQDQ PMEEEEVETF AFQAEIAQLM SLIINTFYSN KEIFLRELIS
NSSDALDKIR YESLTDPSKL DSGKELHINL IPNKQDRTLT IVDTGIGMTK
ADLINNLGTI AKSGTKAFME ALQAGADISM IGQFGVGFYS AYLVAEKVTV
ITKHNDDEQY AWESSAGGSF TVRTDTGEPM GRGTKVILHL KEDQTEYLEE
RRIKEIVKKH SQFIGYPITL FVEKERDKEV SDDEAHHHHHH
```

S3: A) Codon sequence of Hsp90^{WT}. **B)** Amino acid sequence of Hsp90^{WT}.

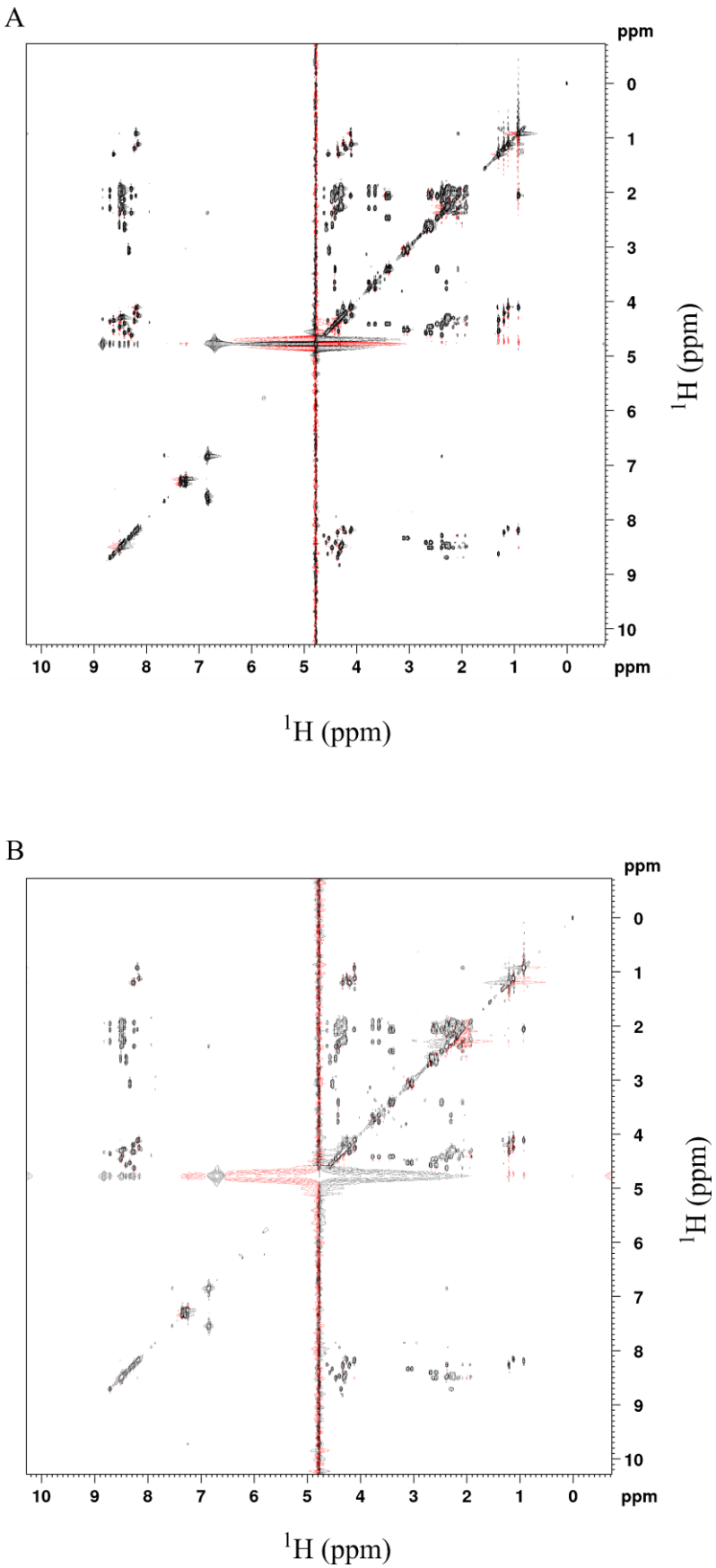


S4: Vector map of the bacterial expression vector pET-21a(+), which was used for the expression of Hsp90 constructs Hsp90^{C1}, Hsp90^{C2} and Hsp90^{WT}.

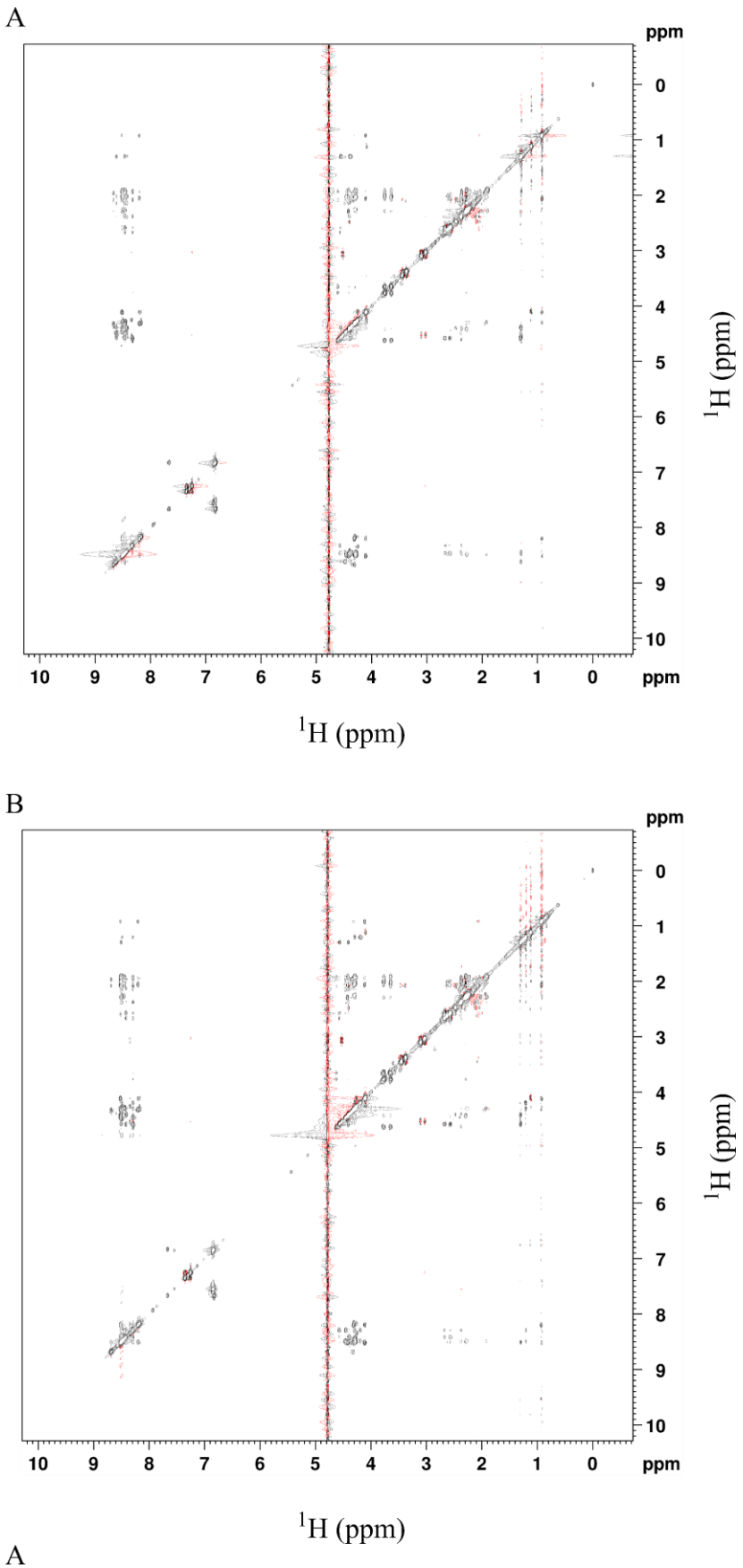
S5: Vector map of the bacterial expression vector TXB1, which was used for the expression of IgM CH4 1-92-intein fusion protein.



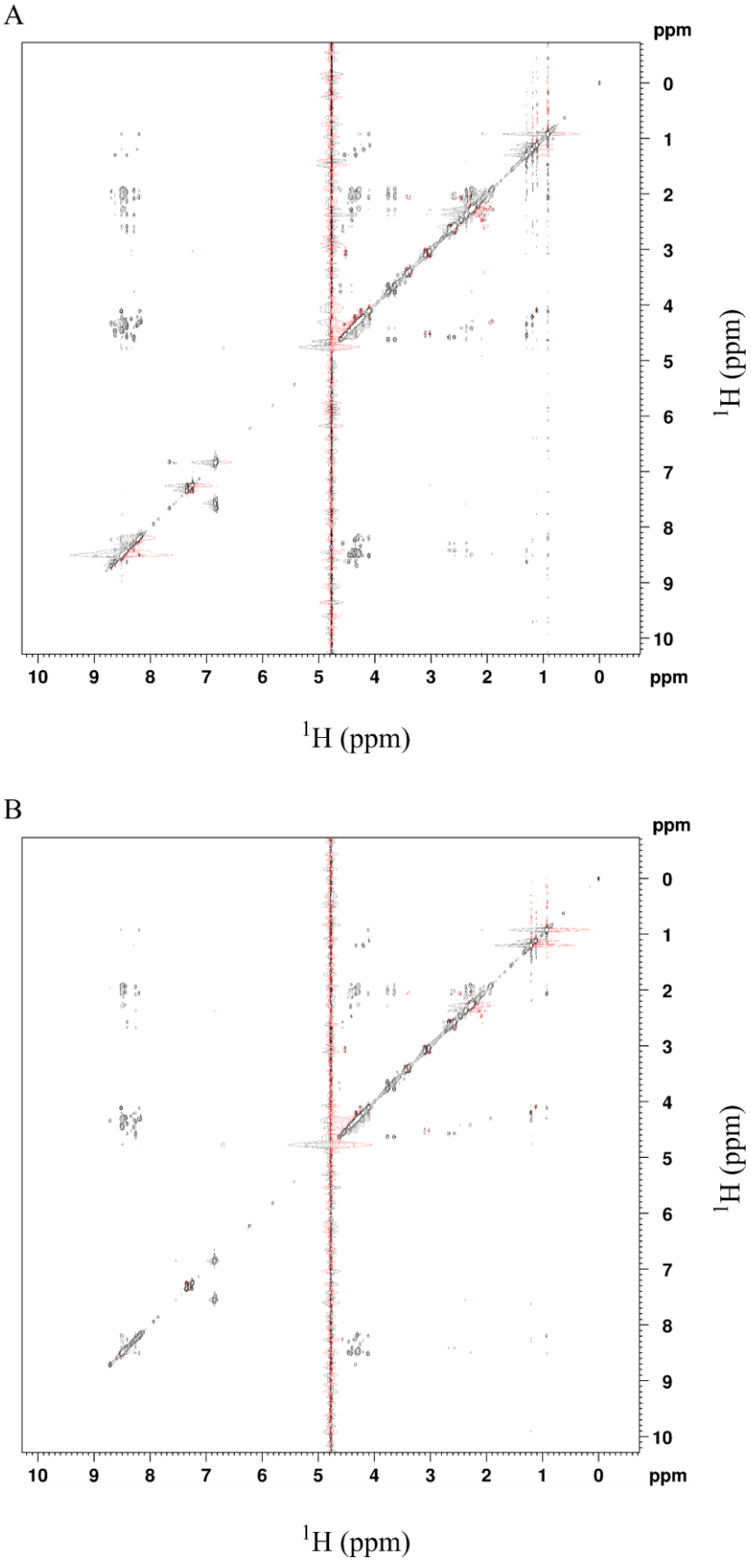
S6: ^{1}H-^{1}H TOCSY spectra of Hsp90 peptide hydrazides **H1** (**A**) and **H2** (**B**).



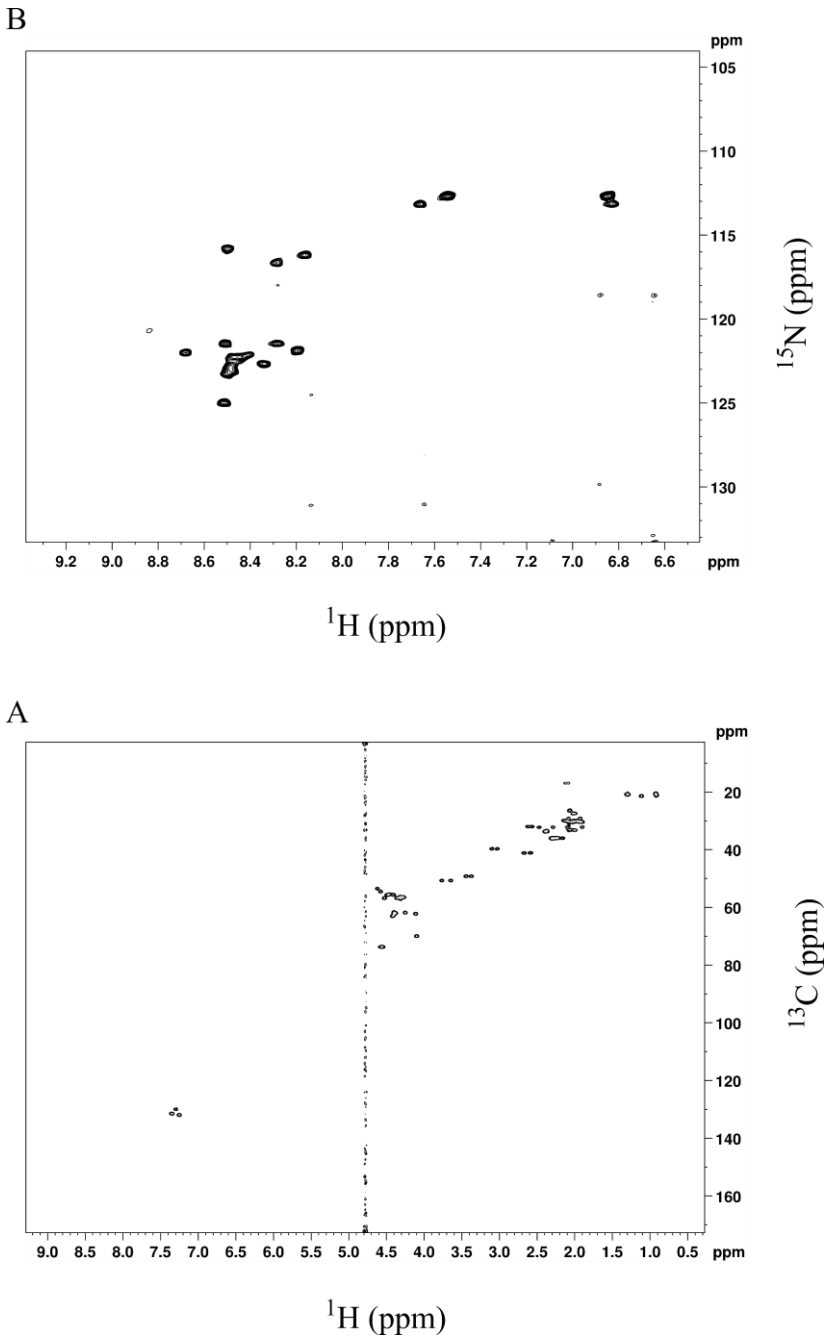
S7: ^{1}H-^{1}H TOCSY spectra of Hsp90 peptide hydrazides **H3** (**A**) and **H4** (**B**).



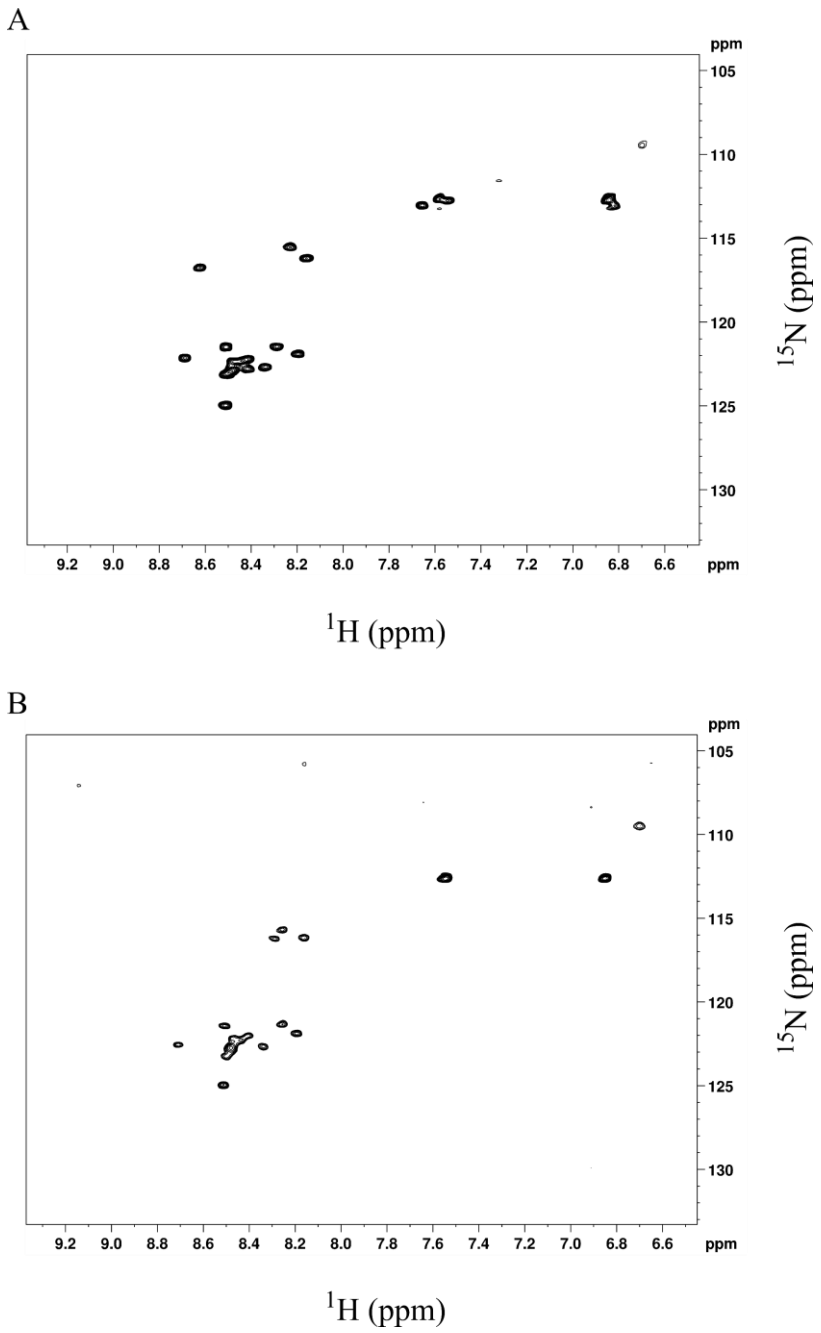
S8: ^{1}H-^{1}H NOESY spectra of Hsp90 peptide hydrazides **H1** (**A**) and **H2** (**B**).



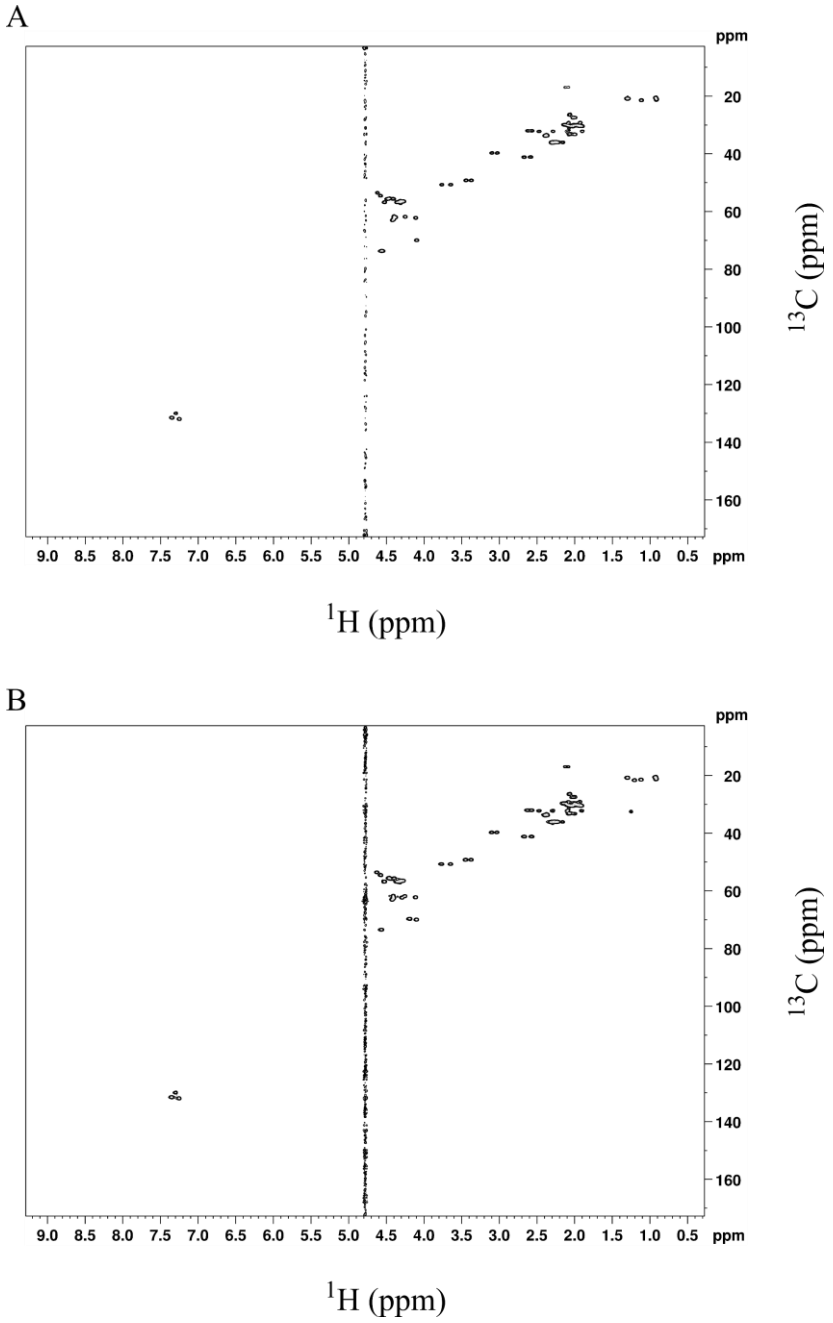
S9: ^{1}H-^{1}H NOESY spectra of Hsp90 peptide hydrazides **H3** (**A**) and **H4** (**B**).



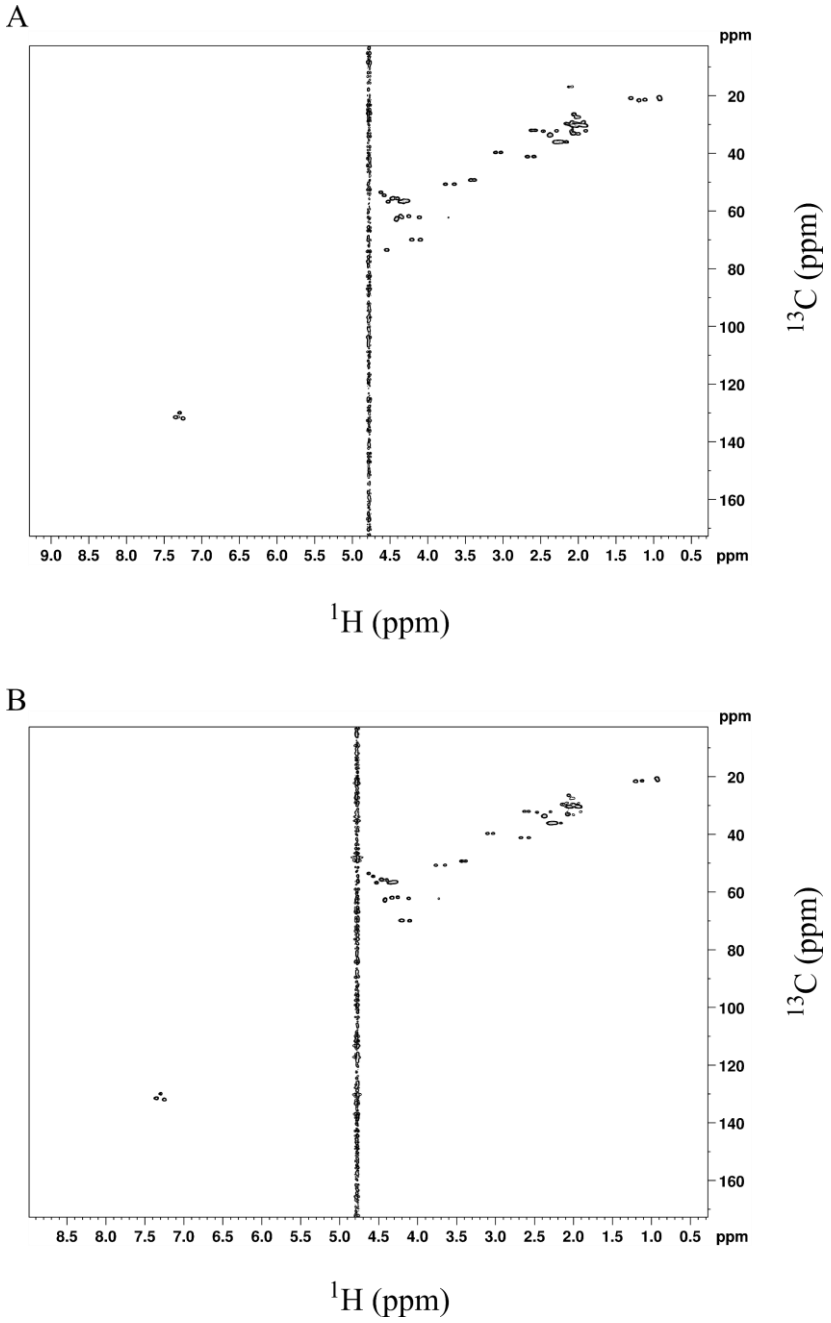
S10: ^{1}H-^{15}N HSQC spectra of Hsp90 peptide hydrazides **H1** (**A**) and **H2** (**B**).



S11: ^{1}H-^{15}N HSQC spectra of Hsp90 peptide hydrazides **H3** (**A**) and **H4** (**B**).



S12: ^{1}H-^{13}C HSQC spectra of Hsp90 peptide hydrazides **H1** (**A**) and **H2** (**B**).



S13: ^{1}H-^{13}C HSQC spectra of Hsp90 peptide hydrazides **H3** (**A**) and **H4** (**B**).

ST 1 Assigned chemical shifts (ppm) of Hsp90 peptide hydrazide **H1**.

Residue	NH	HN	Hα	Hβ	Cα	Cβ	Others
2 Pro		*	4.42	2.42/2.06	62.8	32.4	Hγ 2.06, Hδ1 3.44, Hδ2 3.38, Cδ 49.3
4 Glu	121.6	8.67	4.36	2.10/1.98	56.8	*	Hγ 2.30, Hδ 3.41
5 pThr	115.7	8.46	4.41	4.57	62.0	73.7	Hγ 1.29, Cγ 20.8
6 Gln	*	8.42	4.49	2.15/2.03	55.6	*	Hγ 2.40
7 pThr	116.8	8.62	4.38	4.56	62.0	73.7	Hγ 1.31, Cγ 20.8
8 Gln	122.7	8.47	4.40	2.13	55.6	*	Hγ 2.38, Cγ 33.6
9 Asp	122.0	8.46	4.41	2.67/2.60	54.6	41.2	
10 Gln	121.5	8.32	4.62	2.09/1.93	53.6	*	Hγ 2.38, Cγ 33.7
11 Pro		*	4.42	2.30/2.02	62.8	32.2	Hγ 1.91, Cδ 50.7
12 Met	*	8.51	4.46	*	55.7	*	Hγ1 2.63, Hγ2 2.58, Cγ 32.1
17 Val	121.9	8.20	4.12	2.07	62.2	*	Hγ1 0.92, Hγ2 0.92, Cγ 20.7
18 Glu	*	8.52	4.30	*	56.5	*	
19 Thr	116.2	8.16	4.25	4.10	61.8	69.9	Hγ 1.12, Cγ 21.5
20 Phe	122.6	8.34	4.52	3.10/3.03	56.8	39.7	

** Resonances could not be unambiguously assigned due to peak overlap*

ST 2 Assigned chemical shifts (ppm) of Hsp90 peptide hydrazide **H2**.

Residue	NH	HN	Hα	Hβ	Cα	Cβ	Others
2 Pro			4.42	2.42/2.06	62.8	32.4	Hγ 2.06, Hδ1 3.44, Hδ2 3.38, Cδ 49.3
4 Glu	122.0	8.69	4.36	2.08/1.97	56.8	*	Hγ 2.30, Hδ 3.41
5 pThr	115.8	8.50	4.41	4.56	62.0	73.4	Hγ 1.30, Cγ 20.8
6 Gln	*	8.51	4.46	*	55.8	*	
7 Thr	116.2	8.29	4.29	4.20	62.3	69.8	Hγ 1.20, Cγ 21.6
8 Gln	122.7	8.48	4.40	2.13(ba)	55.8	*	Hγ 2.36, Cγ 33.7
9 Asp	122.0	8.41	4.57	2.67/2.57	54.6	41.2	
10 Gln	121.4	8.29	4.63	2.09/1.94	53.6	*	Hγ 2.38, Cγ 33.7
11 Pro			4.42	2.30/2.02	62.8	32.2	Hγ 1.91, Hδ1 3.76, Hδ2 3.65, Cδ 50.7
12 Met	*	8.51	4.46	*	55.7	*	Hγ1 2.63, Hγ2 2.58, Cγ 32.1
17 Val	121.9	8.20	4.12	2.07	62.2	*	Hγ 0.92, Cγ 20.7
18 Glu	*	8.52	4.30	*	56.5	*	
19 Thr	116.2	8.16	4.25	4.10	61.8	69.9	Hγ 1.12, Cγ 21.5
20 Phe	122.6	8.34	4.52	3.10/3.03	56.8	39.7	

** Resonances could not be unambiguously assigned due to peak overlap.*

ST 3 Assigned chemical shifts (ppm) of Hsp90 peptide hydrazide **H3**.

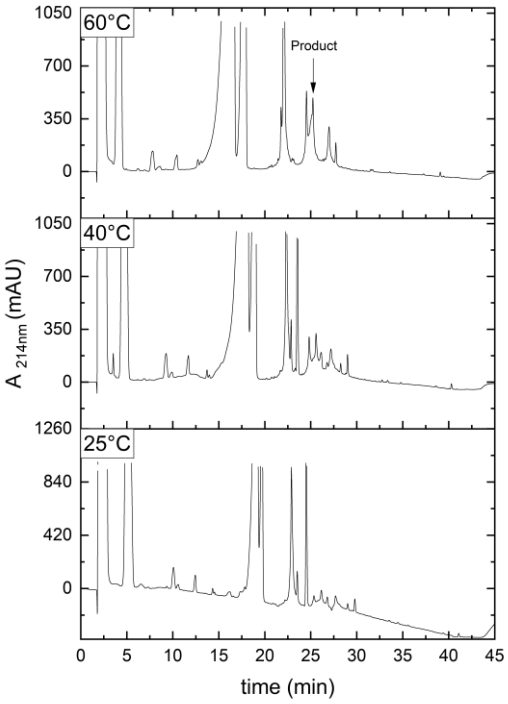
Residue	NH	HN	Hα	Hβ	Cα	Cβ	Others
2 Pro		*	4.42	2.42/2.06	62.8	32.4	Hγ 2.06, Hδ1 3.44, Hδ2 3.38, Cδ 49.3
4 Glu	122.1	8.70	4.37	2.08/1.96	56.8	*	Hγ 2.29, Hδ 3.41
5 Thr	115.5	8.34	4.36	4.21	62.0	73.4	Hγ 1.20, Cγ 20.8
6 Gln	*	8.51	4.47	*	55.8	*	
7 pThr	116.8	8.63	4.35	4.54	61.9	73.5	Hγ 1.30, Cγ 20.9
8 Gln	122.7	8.42	4.40	2.13/1.97	55.8	*	Hγ 2.37, Cγ 33.7
9 Asp	122.0	8.41	4.58	2.67/2.57	54.6	41.2	
10 Gln	121.4	8.29	4.62	2.08/1.93	53.6	*	Hγ 2.38, Cγ 33.7
11 Pro		*	4.42	2.30/2.02	62.8	32.2	Hγ 1.91, Hδ1 3.76, Hδ2 3.65, Cδ 50.7
12 Met	*	8.51	4.46	*	55.7	*	Hγ1 2.63, Hγ2 2.58, Cγ 32.1
17 Val	121.9	8.20	4.12	2.07	62.2	*	Hγ 0.92, Cγ 20.7
18 Glu	*	8.52	4.30	*	56.5	*	
19 Thr	116.2	8.16	4.25	4.10	61.8	69.9	Hγ 1.12, Cγ 21.5
20 Phe	122.6	8.34	4.52	3.10/3.03	56.8	39.7	

** Resonances could not be unambiguously assigned due to peak overlap.*

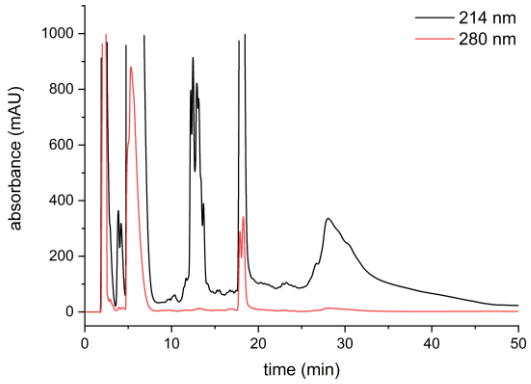
ST 4 Assigned chemical shifts (ppm) of Hsp90 peptide hydrazide **H4**.

Residue	NH	HN	Hα	Hβ	Cα	Cβ	Others
2 Pro		*	4.42	2.42/2.06	62.8	32.4	Hγ 2.06, Hδ1 3.44, Hδ2 3.38, Cδ 49.3
4 Glu	122.5	8.71	4.36	2.07/1.96	56.8	*	Hγ 2.29, Hδ 3.41
5 Thr	115.7	8.26	4.33	4.12	62.0	69.8	Hγ 1.20, Cγ 21.6
6 Gln	*	8.51	4.46	*	55.8	*	
7 Thr	116.2	8.29	4.33	4.21	62.0	69.8	Hγ 1.20, Cγ 21.6
8 Gln	122.7	8.48	4.40	2.13	55.8	*	Hγ 2.36, Cγ 33.7
9 Asp	122.0	8.41	4.57	2.67/2.57	54.6	41.2	
10 Gln	121.3	8.26	4.63	2.09 /1.94	53.6	*	Hγ 2.37, Cγ 33.7
11 Pro		*	4.42	2.30/2.02	62.8	32.2	Hγ 1.91, Hδ1 3.76, Hδ2 3.65, Cδ 50.7
12 Met	*	8.51	4.46	*	55.7	*	Hγ1 2.63, Hγ2 2.58, Cγ 32.1
17 Val	121.9	8.20	4.12	2.07	62.2	*	Hγ 0.92, Cγ 20.7
18 Glu	*	8.52	4.30	*	56.5	*	
19 Thr	116.2	8.16	4.25	4.10	61.8	69.9	Hγ 1.12, Cγ 21.5
20 Phe	122.6	8.34	4.52	3.10/3.03	56.8	39.7	

** Resonances could not be unambiguously assigned due to peak overlap.*



S14: Purification optimisation via RP-HPLC of the ligated wild-type IgM CH4 domain. The emergence of a distinct product peak at higher temperatures is labelled. Fractions were collected manually and analysed via ESI-MS.



S15: Attempted purification of the ligation reaction between recombinant Hsp90^{C2} and Hsp90 2-20 peptide hydrazide **H3**. Peaks were collected manually and analysed via ESI-MS. The observed masses can be assigned to different thiol adducts of the excess peptide **H3**. The mass of the ligation product (MW_{calc} 27247 Da) was not observed in the collected fractions.

8 Acknowledgement

First, I would like to thank Professor Christian Becker for the opportunity to complete my internship in his group and for his expert advice during meetings and talks.

I would like to acknowledge my supervisors Dr. Anne Conibear, Dr. Aleksandr Kravchuk and Mag. Gerhard Niederacher for their valuable guidance and encouragement throughout this difficult project. I greatly appreciate the chance to learn from such experienced and motivated scientists and I am thankful for their patience and support.

Furthermore, I would like to express gratitude to the whole group for their kind collaboration and for the great time inside and outside the laboratory.

In addition, I would like to thank my friends for their inspiration, motivation and broadening my horizons.

Finally, I would like to thank my family and my partner for their love, sympathetic ear and words of wisdom. I will be grateful forever.