CHALMERS
UNIVERSITY OF TECHNOLOGY

Biophysical approaches in a structure-guided SMYD3 ligand discovery

Master's thesis in Biotechnology

MARTIN JOHANNES TALU

Department of Biology and Biological Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2019

Master's thesis 2019

Biophysical approaches in a structure-guided SMYD3 ligand discovery

MARTIN JOHANNES TALU

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Biology and Biological Engineering
Biotechnology
Chalmers University of Technology
Gothenburg, Sweden 2019

Biophysical approaches in a structure-guided SMYD3 ligand discovery
MARTIN JOHANNES TALU

Supervisor: Vladimir O. Talibov, Uppsala University, Department of Chemistry, Uppsala Biomedical Centre - BMC, Danielson group
Examiner: Joakim Norbeck, Chalmers University of Technology, Molecular Biotechnology, Biology and Biological Engineering, Systems and Synthetic Biology

Master's Thesis 2019
Department of Biology and Biological Engineering
Chalmers University of Technology
SE- 412 96 Gothenburg

Gothenburg, Sweden 2019

Biophysical approaches in a structure-guided SMYD3 ligand discovery
MARTIN JOHANNES TALU
Department of Biology and Biological Engineering
Chalmers University of Technology

Abstract

In eukaryotic cells, DNA is wrapped around nucleosomal cores formed of protein heterooctamers, which consist of core histones. Nucleosomes are the main units of chromatin organization. Chromatin exist in two states - either as eu- or heterochromatin, which either promotes or silences gene expression as a consequence of its packing states. Eukaryotic cells have developed epigenetic regulation to control the chromatin state, and to guarantee a high level of differentiation.

The basis of epigenetic regulation are patterns of post-translational modifications of the nucleosomal proteins. These modifications are performed by epigenetic enzymes. This thesis focuses on one of these enzymes - the human lysine methyltransferase SMYD3. SMYD3 is also capable of interacting with certain cytosolic proteins, such as the molecular chaperone HSP90 - the human Heat Shock Protein 90. Both proteins are of high interest in the drug research and development landscape, as a drastic change in their activity and expression levels have both been shown to be related to several cancers or neurodevelopmental diseases.

In this work, various truncated forms of HSP90 were produced and probed for their interactions with SMYD3 using surface plasmon resonance-based biosensor technology. A previously reported interaction of SMYD3 with the C-terminal domain of HSP90 was confirmed, with an affinity discovered to be $K_D = 1.3 \times 10^{-5}$ M. Additionally, the biosensor-based assay was used to test potential ligands of SMYD3, including low affinity fragment-like organic molecules. To complement the study, extensive crystallization and co-crystallization trials were carried out with SMYD3. As a result, conditions for the formation of various crystal forms of SMYD3 were mapped, with the best crystal form found to have high stability and good diffraction properties.

A set of experiments presented herein develops expertise in the tools one can use for an efficient and rational ligand discovery campaign targeting SMYD3 histone methyltransferase.

Keywords: SMYD3, HSP90, SPR, XRD, MST, TSA, drug discovery

Acknowledgements

First and foremost, I am grateful to the Alfred Ots Scholarship foundation, for funding my studies at Chalmers University of Technology. I have met a lot of new and amazing people through the events hosted by the foundation, and even had the honour of hosting some of them myself. The members of the foundation have created a positive and enforcing environment, which helped me settle down easily in Sweden.

I would also like to thank Professor U. Helena Danielson, for welcoming me into her group, and my supervisor, Vladimir Talibov, for his friendship and all his guidance. In the past year I have learned so much more than I could have possibly imagined. I have had the chance of working with and bettering myself in several new key methods, which will help me a lot in my coming years. I could not have asked for a better supervisor. I would also like to extend my thanks to all the other members of the Danielson group, for their continued support and friendliness at the lab.

Last but certainly not least, I would like to thank my mother for urging me to continue with my studies, and for supporting me through them. I would not be the person I am today without her.

Martin Johannes Talu, Gothenburg, June 2019

Contents

List of Figures

List of Tables

1
Introduction

In order to facilitate efficient lead discovery campaigns towards challenging targets, exemplified by epigenetic enzymes, a better biophysical expertise has to be developed. SET- and MYND- domain containing histone methyltransferase (SMYD3) is a perspective drug target which has attracted the interest of the medicinal chemistry community in recent years. The aim of this Master's thesis was to develop novel and to optimize existing biophysical approaches which could assist selection of SMYD3-interacting ligands in a direct manner with the efficiency and robustness required of a successful early stage drug discovery campaign.

1.1 Epigenetic enzymes

Eukaryotic cells regulate gene expression on multiple levels. Unlike in prokaryotes, the genomic DNA of eukaryotes is organized in complex structures, forming chromatin - a supramolecular assembly of nucleic acids with various proteins, with well defined regions differing in density and accessibility for the cellular transcription machinery. On a basic descriptive level, DNA strands are wrapped around nucleosomal cores - protein heterooctamers, formed by histones. Nucleosomes are the main units of chromatin organization, which in turn exist in two states. To promote gene expression and to render the genomic DNA transcription-competent, the chromatin adopts a relaxed state, so-called euchromatin, with promoter regions of the genes exposed to various RNA polymerases. Gene expression can be efficiently silenced by increasing the tightness of packing of the DNA, turning the chromatin in a denser state - hetero-chromatin. This allows changes in gene expression levels and patterns, and corresponds to changes in the cell's phenotype and its biological functions simply by varying the balance of euchromatin and hetero-chromatin in different areas of the genome.

The chromatin state is controlled by specific pathways, summed up as epigenetic regulation. On a biochemical level, epigenetic regulation originates from patterns of post-translational modifications of the nuclear proteins, post-replicational modifications of the genomic DNA, and partially, post-transcriptional modifications of the RNA [1], including mRNA. As most chemical transformations in a biological system, these modifications are performed by specific types of enzymes, denoted nowadays as epigenetic enzymes. In case of proteinogenic substrates, exemplified by the main nucleosomal proteins - histones, the following types of post-translational modifications are involved in the epigenetic regulation: methylation, acetylation, phosphorylation and ubiquination. Epigenetic enzymes are categorized mainly as

writers or erasers, by their ability to add (*to write*) or to remove (*to erase*) the aforementioned modifications onto nucleosomal proteins. These modifications have a dramatic effect on the chromatin state, as can be exemplified by the correlation between the presence or absence of methyl-or acetyl marks on the side chains of lysine or arginine amino acid residues of histones, and the local density of chromatin, together with the downstream activity of genes [2].

One of the most commonly studied epigenetic modifications is cytosine methylation, which occurs in CpG motifs, which cluster in CpG island regions of the genome. These regions harbour about 60% of human gene promoters, and are unmethylated under normal circumstances, allowing for gene transcription. An alteration in the DNA methylation pattern leads to pathogenic cellular conditions, *e.g.* hypermethylation of CpG islands leads to transcriptional inactivation. At the organism level, the outcomes of this impairment are cancer and neurodevelopmental disorders. Cancer cells have 20-60% less 5-methylcytosine, causing a loss in DNA methylation. Hypomethylation promotes chromosomal instability, gene disruptions, and an aberrant expression of certain oncogenes: S100P in a pancreatic cancer, SNCG in breast and ovarian cancers, DPP6 in melanomas. Following a similar logic, a global reduction of acetylation by an overexpression or mutation of the Histone Deacetylases (HDACs) of monoacetylated H4K16 is also a prominent cause of cancer. A key step of development for the cells of the neural system is the mitotic exit, when the cells lose their multipotency. This step requires fine tuning of their transcriptional patterns, depending on the cellular tissue localization [3]. Mutations in genes that are responsible for one's epigenetic profile can lead to somatically acquired or heritable neurodevelopmental disorders: a mutation in MeCP2 causes the Rett syndrome, and a deregulation of one or more genes at 15q11-13 leads to Angelman or Prader-Willis Syndrome [4].

Epigenetic enzymes are thus attractive drug targets in oncology, as the change in gene expression levels can hinder cancer cell growth, exemplified by activating tumour suppressor genes, or repressing cell growth in cancer cells. [2]

1.2 SMYD3

SMYD3 is a lysine methyltransferase that expresses predominantly in the testis and skeletal muscles of humans. The protein requires a co-factor, S-adenosyl-L-methionine (SAM), which serves as a methyl donor [5]. Having protein-based substrates, SMYD3 demonstrates some degree of substrate promiscuity. Originally discovered to bear an enzymatic activity towards histone 2A, it was demonstrated to methylate the whole range of core nucleosomal proteins, except of histone 1. SMYD3 methylates histones H2A to H4 with the following specificity: H4>H2A>H3>H2B, having the highest activity towards the dimethylated lysine 20 of histone H4 (H4K20me2) [6].

As most epigenetic enzymes, SMYD3 does not exclusively bear catalytic functions. The protein was shown to be a part of certain supramolecular complexes, including the RNA polymerase II transcription complex. By interacting with the RNA polymerase II and RNA helicase HELZ, it acts as a transcriptional regulator [7]. The protein was also demonstrated to interact with DNA in a sequence-specific

manner [5]. Additionally, it was shown that the protein is able to interact with certain cytosolic proteins, including a molecular chaperone HSP90 [8].

SMYD3 belongs to the SMYD protein family, characterized by the presence of a conserved SET (Supperssor of variegation, Ehancer of Zeste, Trithorax) domain, which is split by a MYND (Myeloid-Nervy-DEAF1) domain [7]. SMYD3 has a two-lobed structure (Figure 1.1) with a substrate protein binding cleft at the bottom of a 15 Å deep crevice, which is located between the SET and C-terminal domains. The C- and N-terminal lobes are connected by a non-conserved region across SMYD family paralogues (277-279), which might imply a possible hinge motion between the two domains. The SET domain (3-245) is responsible for the enzymatic activity, and the zinc-finger MYND domain is a protein-protein interaction site which mediates contacts with other proteins via proline-rich sequences [9], as well as likely interacts with DNA. The SET domain composes of the pre-SET (3-46), SET-I (86-181), and core-SET (182-245) regions, and is divided by the MYND (47-85) domain insertion. The two fragments come together to form a conserved SET-domain fold. The pre-SET domain is said to stabilize the SET domain, and is essential for interaction with the methylation co-factor SAM. Residues 280-425 make up the C-terminal domain, hypothesized to be a regulatory motif to modulate the substrate methylation activity and substrate selectivity.

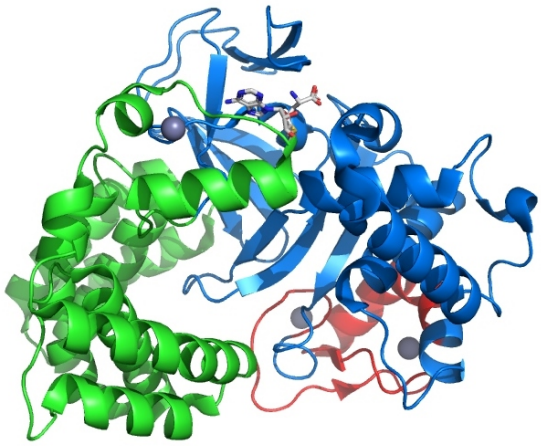

Figure 1.1: Ribbon representation of the SMYD3 crystallographic structure. The SET-domain is depicted in blue, the zinc finger containing MYND-domain in red, and the C-terminal domain in green. The protein was co-crystallized with SAM bound to its catalytic site. The model was made using PyMOL, PDB ID:3MEK.

SMYD3 is shown to be a potentiator of a large variety of cancer-related oncogenes. For example, it is required to chemically induce liver or colon cancer in mice disease models. Overexpression of SMYD3 in humans has shown a positive correlation in cancer progression of human colorectal carcinoma, hepatocellular carcinoma, breast cancer, lung- and pancreatic ductal adenocarcinomas [10]. Therefore, inhibitors of SMYD3 could be promising candidates for cancer therapy, possibly with reduced toxicity risks and limited adverse effects. However, the precise biology of the protein is still unclear and on-going research would benefit from novel tool compounds with various mechanisms of modulating SMYD3 functions.

1.3 Heat shock protein 90

Heat shock protein 90, abbreviated as HSP90, is a protein folding chaperone, which interacts with host proteins to protect them from misfolding and aggregation. HSP90 was discovered because of its elevated expression over the heat shock response, and was found to be important for various cellular processes. It has multiple roles: protein folding, DNA repair, mediation of the immune response. Its known clients are kinases, transcription factors, steroid hormone receptors, E3 ubiquitin ligases, and many more [11]. Being a part of numerous signalling pathways, HSP90 has become a drug target for the treatment of various diseases, including oncology-related.

The protein in its active form is a dimer. An HSP90 monomer consists of three domains: an amino-terminal domain, which mediates adenosine triphosphate (ATP) binding; a middle substrate binding domain, which is necessary for ATP hydrolysis and binding of most of the client proteins, and a carboxy-terminal domain, which is responsible for HSP90 dimerization and interaction with co-chaperones via its MEEVD-motif [11] (Figure 1.2). HSP90 has two main isoforms in humans, the alpha (HSP90α, or HSP90AA1) and the beta (HSP90β, or HSP90AB1), which appeared ca. 500 million years ago following a gene duplication mechanism. The isoforms are usually studied in unison, since it is relatively complicated to distinguish between them biochemically in cellular and animal models. The two different isoforms may form both homo-and heterodimers. Some of the differences in their functions are that the HSP90AA1 isoform dimerizes more efficiently than HSP90AB1, and therefore is more active in a cellular environment [12]. The expression of HSP90AA1 is induced primarily by a heat shock, whereas HSP90AB1 is expressed at a constant rate. HSP90AA1 promotes cell cycle regulation and cell growth, whereas HSP90AB1 is important for cellular transformation, signal transduction cascades, and long-term adaptation. Both of them however contribute significantly to the regulation of cellular proliferation and differentiation [13].

HSP90 interacts with substrate and client proteins in different conformations, with the main conformational changes occurring in a region spanning the middle domain. The chaperone was shown to make protein-protein contacts utilizing its N- and C- terminal domains. Both isoforms have a Met-Glu-Glu-Val-Asp (MEEVD) conserved motif at the C-terminus of the C-terminal domain, which is a common recognition motif for tetratricopeptide domains [11] - domains, which facilitate specific protein-protein interactions and are important for the functioning of, for example, transcription complexes [14].

Inhibitors of HSP90 interact with the N-terminal domain, blocking its ability to bind ATP, and therefore inhibit the stabilization of substrates. This causes the client proteins to fold incorrectly and to undergo degradation within the ubiquitin-proteosome pathway. The build-up of misfolded proteins may lead to malignant states, Alzheimer's, Parkinson's, and multiple sclerosis. The inhibition of HSP90 induces the expression of other Heat Shock Proteins, translating into resolubilization and disaggregation of the misfolded species [15].

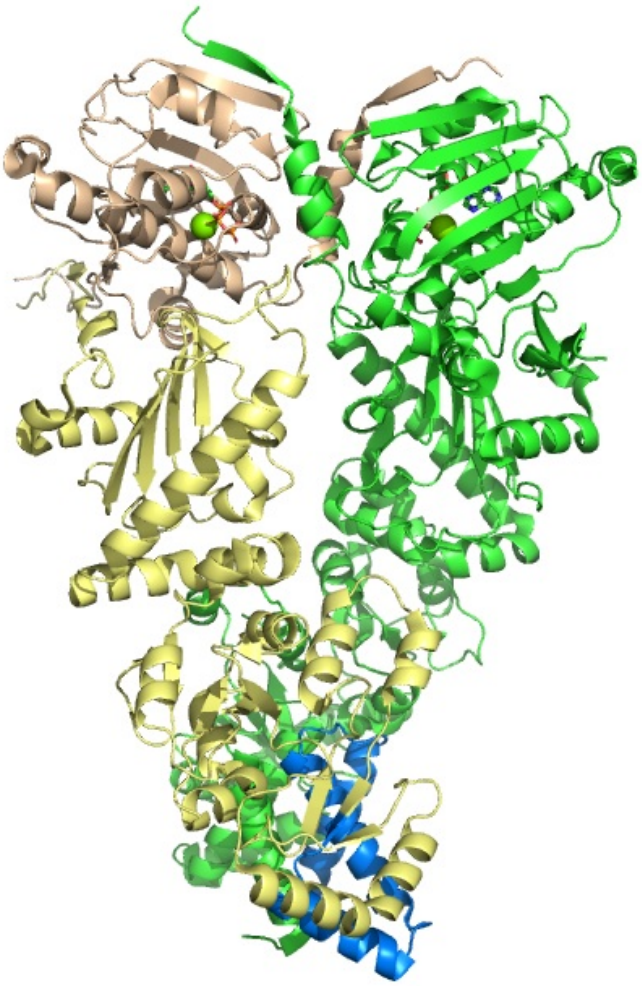

Figure 1.2: Human heat shock protein 90. Ribbon diagram of the HSP90AB1 homodimer. Monomer 1 depicted in green. N-terminal domain in salmon, middle domain in yellow, C-terminal domain in blue, with ATP and Mg^2+ cations bound to its corresponding sites. The model was made using PyMOL, PDB ID: 5FWK.

1.4 Early-stage drug discovery

The five stages of drug discovery are basic research (hypothesis generation and the validation of a target), early stage drug discovery (drug-like molecule search), pre-clinical development (basic laboratory testing to verify if the drug candidate is safe for a set of test subjects, as well as an accumulation of preliminary pharmacokinetic data), clinical development (large scale test on humans with statistical verification to see if the desired effect is achievable, and if it can surpass the already existing medications), and a final confirmation of the results by governmental authorities (*e.g.* US Food and Drug Administration, FDA), which will decide whether or not a drug is safe and effective to be released to the market [16]. The development of a new drug is a challenging process which can take up to 15 years and cost about 1 billion USD (2019).

Once the function of a target is known, it is also necessary to understand in

which cellular compartments, types and tissues it resides, which pathways the protein influences, and what eventual outcome it has on a physiological level [17]. As a first step in protein studies, one has to assign the protein to a specific process to eventually understand its individual role. Basic structural information (open reading frames and cDNA sequences) is acquired from transcriptional studies and can be used to produce the protein of interest in its pure form, and *in vitro* conditions can be mimicked to monitor the function of the protein [18]. Often enough, after optimization and scaling of the functional analysis, similar assays with an activity readout (exemplified by monitoring the enzyme kinetics) are used for screening of drug-like compound libraries, a lengthy experiment involving testing of thousands of unique chemical structures. However, functional assays have certain limitations. Sometimes the nature of the target does not allow to monitor its activity in such a reductionistic manner, employing simple *in vitro* conditions. Alternatively, the assays able to do it are either impractically expensive or have a complex set up. In such cases biophysical methods, *i.e.* methods that employ various physical phenomena to monitor direct interactions between the search molecules and the target protein, are hugely beneficial for both initial screens, and for the lead compound optimization in a rational manner.

1.4.1 Biophysical methods

Biophysical methods aim to measure interaction parameters between two different biomolecules, or biomolecules and synthetic compounds, to characterize the binding process, its mode, and unique physico-chemical signatures [19]. The most common methods used in the early-stage drug discovery are the Surface Plasmon Resonance (SPR)-based biosensor technology, Isothermal Titration Calorimetry (ITC), Nuclear Magnetic Resonance (NMR), Mass Spectrometery (MS), X-ray based protein crystallography (XRD), and thermal unfolding methods (thermal shift assays, TSA). These methods aim to answer if the protein structure is maintained correctly, does it interact with a molecule of interest, how and where does the interaction occur, and what are the thermodynamics and kinetics behind the interaction. These methods often require the compound to be soluble at levels higher than the expected affinity constant, and non-aggregating in the experimental conditions [20]. Additionally, biophysical methods put an extra limitation on the quality of materials to be used for the assays - high purity of the chemicals and high quality of protein preparations are required, especially for the structural biology applications.

1.4.2 Surface Plasmon Resonance

Surface Plasmon Resonance is a method commonly used in preliminary pharmacokinetic profiling of drugs, medium through-put screening (up to few thousand compounds with the contemporary instrument), fragment-based drug discovery [21] and lead compound characterization. The method is based on the phenomena of plasmon resonance, which is induced by an incident light oscillation of planar electrons in a conductive media. At certain angles, light energy couples through the gold metal plating, and the electrons in the metal surface layer move due to the

excitation. The plasmon waves, or the electron movements, propagate parallel to the metal surface, creating an approximately 300 nm electric field from the boundary between the phases - a so-called evanescent field [22]. The field propagation is sensitive to the optical properties of the surface layer, for example to its refractive index. Therefore, the changes in the surface refractivity translate to changes of the angle of the resonance induction, and can be measured optically [23]. The optical unit, coupled to a fluidic system and a pumping unit allows to monitor such changes in real time. SPR biosensors allow for extraction of kinetic parameters (for example the reaction rate constant), the interaction mechanism and thermodynamic parameters [23].

In an experiment, a ligand is immobilized on the sensor surface, and an analyte is addressed over the sensing element. The assay readout is a sensorgram, or a progress curve, which can be split in two phases - an association (a to b) and a dissociation (c) phase, figure 1.3. During the first phase, the analyte association with the ligand gives a rise in signal until the ligand's occupancy reaches a steady-state equilibrium (b) at a given analyte concentration. At the end of the analyte injection, the formed complex dissociates to the flow of the buffer, creating the dissociation phase. The surface is thereafter regenerated (d) by removing the excess of injected analyte. Analysis of the progress curve employing either global curve analysis with the corresponding rate equations (exemplified on Scheme 1.1 and Eq. 1.1) or the relationships between the signal levels and analyte concentration allows to calculate kinetic and thermodynamic parameters of the interaction [24].

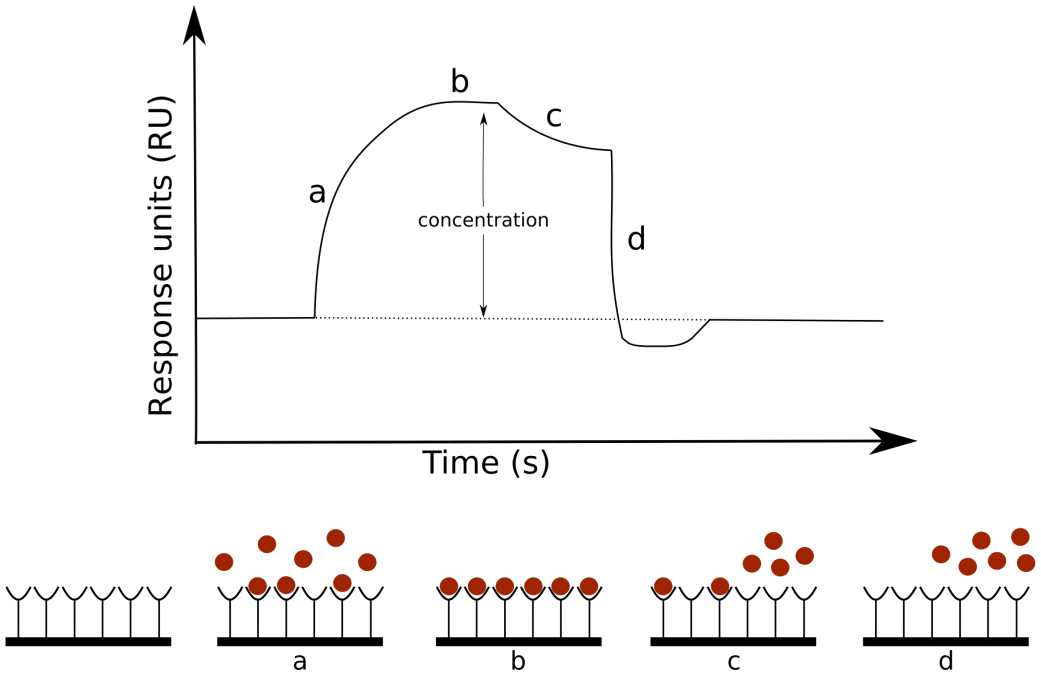

Figure 1.3: Interaction kinetic curves showing association (a), steady-state (b), dissociation (c) and regeneration (d). The change in the curvature can be used to analyze the kinetics of the proper interaction. The bottom figure shows the binding of analyte (red) to the protein (black) in different experimental stages.

$$\mathrm{L} + \mathrm{A} \underset{\mathrm{k_{off}}}{\overset{\mathrm{k_{on}}}{\rightleftharpoons}} \mathrm{LA}$$

Scheme 1.1: Second order reversible association between a ligand (L) and analyte (A) molecules, characterized by association (k_{on}) and dissociation (k_{off}) rate constants. The corresponding reaction rate equation is given in Eq. 1.1

$$\frac{d[LA]}{dt} = k_{on}[L][A] - k_{off}[LA] \tag{1.1}$$

1.4.3 Fluoroscopic methods

Fluorescence is the ability of a molecule to absorb light and to emit it back with a lower energy (longer wavelength). It can be used to trace natively fluorescent proteins, which have aromatic amino acid residues, such as tyrosine and tryptophan. Fluorescent molecules can also be bound to proteins as fluorophoric groups, which is useful for studying the distribution of proteins in tissue samples or for ligand-binding assays [25].

Thermal shift assay (TSA), or differential scanning fluorimetry, is a common method to asses protein stability and to detect recognition events between a protein and a ligand. Protein denaturation is a two-state transition between folded and unfolded states. Unfolding will occur rapidly above a certain critical temperature upon gradually heating the protein sample. The melting temperature corresponds to a transition point in which half of the protein is unfolded [26]. This can be measured by coupling a gradually heating thermocycler to a spectrofluorimeter, detecting either changes in intrinsic fluorescence of certain amino acid residues (*e.g.* tryptophan [27]), or fluorescent properties of the reporter solvatochromic dyes - dyes that attach to the unfolded hydrophobic parts of the protein, and are highly fluorescent in a non-polar environment. If a ligand binds to the protein, or an environment (*e.g.* ionic strength, buffer pH) stabilises or destabilises the protein fold, a change in its melting temperature can be recorded. TSA is useful for screening for optimal protein storage conditions and further downstream applications, allowing to find parameters crucial for the protein's stability and its structural integrity.

Microscale Thermophoresis (MST) is an emerging technology useful for addressing the ligand-binding properties of biomolecules. The method monitors the fluorescence of proteins along a temperature gradient [20]. A laser is focused on a sample, which creates a spatial temperature distribution of about 25 µm. After heating, the system reaches a steady state of heat dissipation equilibrium. The change in protein concentration of initial vs steady state in the localized distribution is measured. This distribution may be influenced by multiple factors, including the hydrodynamic radius of a molecule - which, in turn, changes upon interaction with a ligand. Thus, the method allows for direct monitoring of intermolecular interactions [28].

1.4.4 X-ray crystallography

One can usually predict some secondary structure elements from the amino acid sequence of a protein. But it is still not possible to deduce the whole structure, even

despite recent progress in *in silico* methods. Thus, structural methods are used to gain an insight into the fine atomic details of the macromolecular structure, as well as to identify the chemistry of an interaction between a protein and its ligand, with X-Ray based protein crystallography being the main method of choice [17].

X-Rays are a type of electromagnetic radiation with a wavelength of about 0.1 nm. There are different sources of X-Rays, however, contemporary diffraction experiments with protein crystals are performed in synchrotron facilities [29]. Upon irradiation of a crystalline matter, most of the rays will pass through it, whereas some of them will be scattered by the electron clouds of the sample atoms. These scattered waves can then be detected and recorded as a diffraction pattern. A set of patterns obtained under various angles indirectly represent a distribution of electron density inside the crystal, and thus allow to build an atomic model of a molecule [17].

One of the major challenges in macromolecular crystallography is to obtain crystalline material of suitable quality. The process of crystallization can best be explained by a phase diagram (Figure 1.4), as an equilibrium between protein concentration and other adjustable parameters (*e.g.* *p*H, concentration of salts or precipitants). Crystallization occurs when a crystal nucleus appears, and then spontaneously grows. When the protein mixture is undersaturated, the protein is fully dissolved and spontaneous nucleation will not occur. When the protein is oversaturated, precipitation of amorphous material composed of disordered protein aggregates occurs. The nucleation zone is where conditions are right for nucleation to occur, but the crystal will most likely not grow to the required dimensions. The metastable zone is where nucleation takes place, and conditions are stable enough for crystals to grow. Following the phase diagram shown in Figure 1.4, nucleation will lead to a decrease in protein concentration in the solution, which will move the balance towards the metastable zone. This is usually not the case, as excess nucleation sites occur, which lead to numerous low quality crystals. To overcome this, one can influence the kinetics of the crystallization process, by bypassing the nucleation zone, or by limiting the amount of nuclei. The kinetics can be influenced by altering the nucleation path using a different crystallization process, or by utilizing an oil barrier to slow down diffusion of the solvent. The easiest way to bypass the nucleation zone, is to use pre-grown nuclei as a seed for bigger monocrystals [30].

The main requirement of protein crystallography is a pure (>95 %) and structurally homogeneous protein. The production of protein with a concentration high enough (>10 mg mL^{-1}) is time consuming, constituting of multiple steps of cloning, generation of constructs, expression, purification, assessment of activity and stability. From then on, crystallography can be divided to two steps: screening for the optimal crystallization conditions, and optimization of the conditions to obtain well-ordered monocrystals big enough to sustain the radiation damage. Initial screening can be automated by use of commercially available kits, which for example target a range of different additives, *p*H, precipitant types and their concentrations [31].

Crystallization is usually set up in a microbatch, or by utilizing the vapour diffusion or dialysis methods. In microbatch experiments, protein solution is put under a low density (0.87 mg mL^{-1}) paraffin oil. The protein drop is denser than the oil, so it will remain under the organic layer, and is thus protected from rapid evaporation or airborne contaminants. In vapour diffusion methods, a drop of liquid

containing the protein is equilibrated against a reservoir solution. The experiment can be set in two ways, commonly referred to as the sitting or hanging drop (Figure 1.5), in which a protein drop is either sitting on a pedestal or hanging from a cover slide. In dialysis-based crystallization, the protein mixture is separated from the reservoir/precipitant solution by a semi-permeable membrane, that allows for a slow passage of low molecular weight precipitants to the protein mixture [30]. In microbatch, the sample is at its end concentration, therefore no changes in the *p*H or drop volume will occur. The downside to this is that it can not be used when dealing with volatile or lyophobic substances, since they will dissolve into the oil. Vapour diffusion is by comparison a dynamic system, with conditions such as drop volume and *p*H (due to volatile ions) changing throughout the crystallization process. Dialysis provides an alternative method to cross the phase diagram, but requires high expertise and a relatively complicated set-up [30].

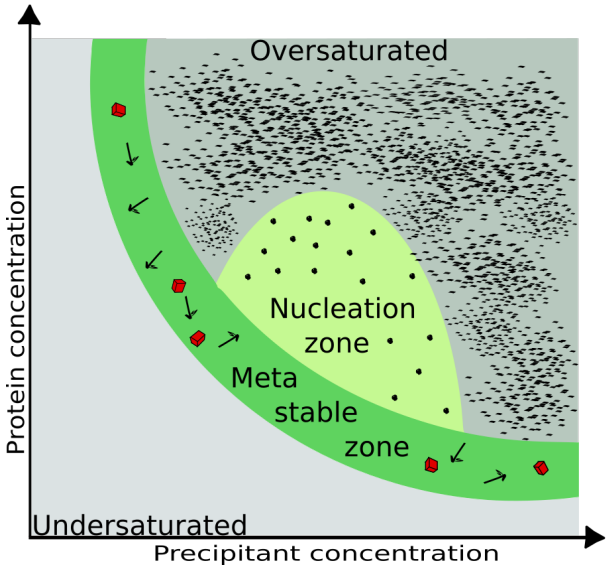

Figure 1.4: Crystallization phase diagram. The saturation curve of protein crystallization from liquid to crystal to precipitate.

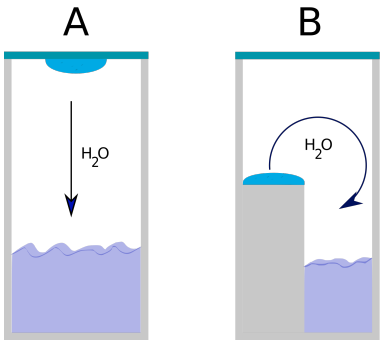

Figure 1.5: The hanging (A) and sitting (B) drop methods of crystallization utilizing vapour diffusion.

1.5 Present work

Prior to the current study, SMYD3-related projects have been evolving in Prof. U. Helena Danielson's group (Uppsala University, Sweden). A reliable SPR biosensor-based interaction assays, as well as a protein production platform were established (E. Fabini, V. Talibov, et al. *Submitted, 2019*). The medicinal chemistry application of the assay was validated, employing a small library of drug-like compounds, and the assay hits were confirmed with structural studies. One of the hit compounds (denoted further in this work as **DIP2**) was co-crystallized with SMYD3, revealing an allosteric binding site (V. Talibov, E. Fabini, et al. *Manuscript, 2019*). The current study evolved on the basis of the latter work, with aims to link the discovered allosteric site to a protein function and to develop a better X-Ray crystallographic expertise for SMYD3-related ligand discovery campaigns.

The discovered allosteric site was found to overlap with a proposed SMYD3-HSP90 protein-protein interaction interface [8], Figure 1.6. It was claimed that HSP90 interacts with SMYD3 C-terminal domain specifically, employing its C-terminal moiety composed of amino acid residues Met-Glu-Glu-Val-Asp (MEEVD). Several peptides with varying length and charge were synthesized to verify if this motif is crucial for SMYD3 recognition by HSP90. Additionally, the current work was focused on an orthogonal approach to verify if the HSP90-SMYD3 interaction exists, with a more direct manner than previously reported by Brown et al. [8]. Lastly, the current study aimed to investigate if the screened compound acts as a protein-protein interface inhibitor, diminishing potential recognition between SMYD3 and HSP90 C-terminal domain. After setting up an SPR assay to screen for interactions between SMYD3 and HSP90, it was viable to study whether or not **DIP2** could compete with HSP90 for SMYD3 binding.

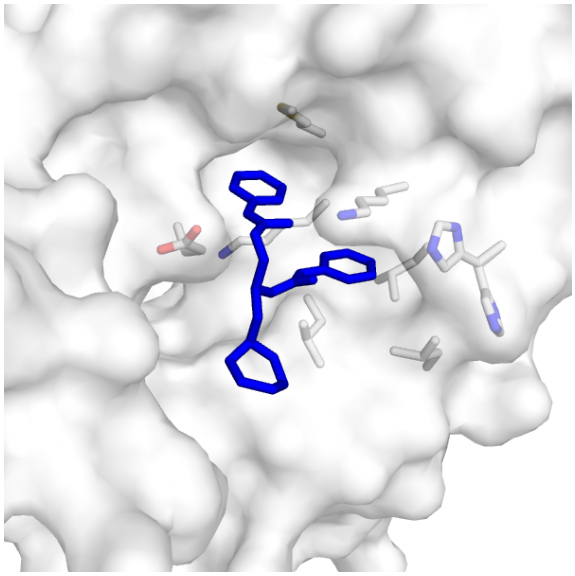

Figure 1.6: Structure of co-crystallized SMYD3-DIP2 complex. The ligand (blue sticks) revealed a novel allosteric binding site, located within SMYD3 C-terminal domain (white surface).

At the same time, work was carried out to optimize the crystallization platform for SMYD3, in hopes of increasing the diffraction properties of the previously obtained protein crystals. Potential binders were also tested by co-crystallizing SMYD3 with a set of *in silico* predicted perspective small molecular ligands, HSP90 C-terminal domain, and several HSP90 recognition motif-based peptides.

2
Methods

2.1 Design of expression constructs

The donor and recipient plasmids used in this thesis are shown in the appendix (Figures E.1, E.2, E.3, E.4). The selections and propagation of plasmids were performed in *E. coli* DH5α strains. All proteins were produced with heterologous expression, employing the *E. coli* Rosetta 2 expression strain.

2.1.1 Preparation of chemically competent cells

Competent cells were prepared following a slightly modified Inoue method [32]. The cells were grown in SOB medium (per 1 L: 5 g NaCl, 20 g trypton, 5 g yeast extract, 2.5 mM KCl, *p*H 7.5) to $OD_{600\,nm} = 0.5$ at 18 °C. All subsequent manipulations were performed at 4 °C or on ice. The cells were centrifuged at 2500×g for 10 minutes, washed in cold 0.22 µm filter-sterilized Inoue solution (86.5 mM $MnCl_2$, 20 mM $CaCl_2$, 250 mM KCl, 10 mM HEPES, *p*H 6.7), and re-suspended in the same solution. At the final step, the cell suspension was supplemented with warm dimethylsulfoxide (DMSO) to 7 % (v/v). The suspension was incubated for 10 min on ice, aliquoted and snap-frozen in liquid nitrogen.

After each preparation, the cells were spread on LB-Agar plates with or without antibiotics to guarantee an absence of cross-contamination of the competent cells with the wrong selection markers.

2.1.2 Preparation of pET15b-HSP90 C-terminal domain construct

The expression plasmid encoding GST-$HSP90_{626-732}$ was a gift from William Sessa (Addgene plasmid nr 22483; http://n2t.net/addgene:22483; RRID: Addgene_22483). The plasmid-bearing culture (agar slab) was titrated on an ampicillin LB-agar plate and grown overnight at 37 °C. The next day, a single colony was taken from the plate and grown overnight at 37 °C in ampicillin-supplemented rich media (LB-Amp, antibiotic to 100 µg mL^{-1}).

The plasmid was purified using an alkaline lysis procedure with a GeneJET Plasmid Miniprep Kit (ThermoFisher Scientific, Netherlands); the purified product was sequenced using single-strand Sanger sequencing (EuroFins Genomics, Netherlands).

cDNA was amplified with PCR employing Phusion DNA polymerase (ThermoFisher Scientific, Lithuania), using ACTGAACATATGGGTTACATGGCAGC

as the forward and ATATAGGATCCCCTAGTCTACTTCTTCC as the reverse primers. The program and the reaction mixture composition are shown in the appendix (appendix C.1). The resulting PCR products were run on a 1.6% agarose gel, and purified using a QIAquick PCR Purification Kit (QIAGEN, Germany).

The resulting cDNA fragment with NdeI and BamHI flanking sites was dual digested using FastDigest restriction endonucleases (ThermoFisher Scientific, Lithuania). The recipient plasmid pET-15b (Novagen, Merck KGaA, Germany) was digested under similar conditions, but with *in situ* dephosphorylation with 1 U of alkaline phosphotase (ThermoFisher Scientific, Lithuania). The restriction reactions were stopped by heating the samples to 85 °C for 15 min; ligation was performed using T4 DNA ligase (ThermoFisher Scientific, Lithuania) at RT for 3 h. The ligation reaction was used to transform *E. coli* DH5α cells.

Colony PCR using T7 promoter/terminator-specific primers was performed to identify positive clones. The program and reaction mixture composition are shown in the appendix (appendix C.6). Further on, the positive colonies were inoculated in LB-Amp and the purified plasmids were sequenced using single-strand Sanger sequencing (EuroFinns Genomics, Germany).

The construct was transformed into chemically competent Rosetta 2 *E. coli* cells, and grown on chloramphenicol (35 µg mL^{-1}) and ampicillin (100 µg mL^{-1}) LB-agar plates overnight at 37 °C.

2.1.3 Preparation of pET15b-HSP90 N-terminal domain construct

The expression plasmid encoding GST-HSP90$_{9-236}$ was a gift from William Sessa (Addgene plasmid nr 22481; http://n2t.net/addgene:22481; RRID: Addgene_22481). The construct re-design steps were similar to the aforementioned pET15b-HSP90 C-terminal domain. Briefly, PCR was carried out using AACTGACATATGGACCAACCGATGGAGG as the forward and AATAGCCTCGAGTTCAGCCTCATCATCGCTTAC as the reverse primer, to amplify the HSP90 N-terminal domain (HSP90 NTD) HSP90$_{9-236}$ encoding sequence (table C.2). The insert was ligated into a pET-15b plasmid. The ligation mixture was propagated in *E. coli* DH5α cells, positive ligation products were sequenced, and the correct construct was transformed into chemically competent Rosetta 2 *E. coli* cells.

2.1.4 Preparation of pET15-HSP90$_{293-732}$ construct

The cDNA encoding full-length HSP90$_{1-732}$ was a gift from William Sessa (Addgene plasmid nr 22487 ; http://n2t.net/addgene:22487 ; RRID: Addgene_22487). HSP90$_{293-732}$ cDNA was produced using the following primers - FWD: (1) GCGAGTCATATGACCAAGCCTATTTGG , REV: (2) GACGGAGGATCCCTAATCGACTTCTTC, G2A-FWD: (3) GGCTTTTCCCTTGAAGATCCCCAG, G2A-REV: (4) GGGTCTGGG-GATCTTCAAGGG, which annealed as shown in Figure 2.1. A guanine to adenine (G2A) mutation was introduced at main ORF codon position 679, to remove an endogenic BamHI site. The first PCR reaction (table C.3) resulted in two products using primers (1, 3) and (2, 4), both bearing overhangs for

downstream hetero-annealing. The second PCR (table C.4) was carried out using a 1:5 ratio of the longer (1, 3) to shorter (2, 4) fragments, creating $HSP90_{293-732}$ with unique NdeI and BamHI sites. This DNA was purified using a PCR cleanup kit. A third PCR reaction (table C.5) was used to amplify the desired product.

Figure 2.1: The primer annealing sites on $HSP90_{1-732}$ DNA. All positions are given related to the main ORF codons. Primer (1) annealed at position 293, (4) and (3) around 679, (2) on 732. G2A overhang and mutation *mismatch* shown as black and red lines on figure.

The re-designed cDNA was sequenced as a PCR product and inserted in the multiple cloning site of the pET-15b plasmid. The plasmid was propagated in *E. coli* DH5α cells, positive colonies were sequenced on both strands and the correct expression construct was transformed into a chemically competent Rosetta 2 *E. coli* cells.

2.2 Purification

All proteins, unless stated otherwise, were expressed utilizing the *E. coli* Rosetta 2 strain in Lysogeny Broth medium, supplemented with $100\,\mu g\,mL^{-1}$ of ampicillin and $35\,\mu g\,mL^{-1}$ chloramphenicol, employing pre-cultures from an overnight single colony growth, which were expanded with a ratio of 1:100.

2.2.1 Purification of HSP90 C-terminal domain

N-terminally His-tagged $HSP90_{626-732}$ expression was induced with 0.5 mM IPTG at $OD_{600\,nm} = 0.8$ for 4 h at 37 °C. Cells were harvested with centrifuging (5000×g, 4 °C), washed using cold TBS buffer (50 mM Tris, 150 mM NaCl, *p*H 8.0), harvested again, and re-suspended in a lysis buffer consisting of 50 mM Tris, 300 mM NaCl, 2 mM $MgCl_2$, 0.5 mM $CaCl_2$, $10\,\mu g\,mL^{-1}$ of both DNAse I and RNAse I, 5 mM phenylmethylsulfonyl fluoride, *p*H 8.0. Cells were lysed using a French press at 1.7 kbar.

$HSP90_{626-732}$ (HSP90CTD) was purified on ÄKTA Explorer FPLC system (Phadia, Sweden), by means of immobilized nickel affinity chromatography on a Ni-NTA IMAC column (in-house packing, media - Chelating Sepharose FF, GE Healthcare, Sweden). The lysate was applied in IMAC buffer A (50 mM TRIS, 300 mM NaCl, 10 mM imidazole, *p*H 8.0), washed with 60 mM imidazole, and eluted at 100 % of buffer B (same as A, but 300 mM imidazole). The protein was desalted in TBS *p*H 8.0 with a Sephadex G25 gel-filtration column (GE Healthcare, Sweden). The hexahistidine tag was removed from the protein by incubating the desalted fraction with 0.5 catalytic units of human thrombin (Merck KmgA, Germany) per apparent mg of protein overnight on ice. The proteolytic reaction was supplemented to

300 mM NaCl and 40 mM imidazole, and passed through a Ni-NTA IMAC column. The collected flow-through was desalted in a citric acid buffer (50 mM citric acid, 50 mM NaCl, *p*H 6.0). The sample was loaded on a HiTrap Q HP anion exchange column (GE Healthcare, Sweden), washed with 240 mM NaCl, and the fraction of interest was eluted with 335 mM NaCl. The eluate was desalted in TBS *p*H 8.0. The protein was concentrated using a 3 kDa cut-off centrifugal microconcentrator, aliquoted, and snap-frozen in liquid nitrogen.

To purify the GST-fused $HSP90_{626-732}$ (GST-HSP90CTD), the original donor plasmid (Addgene plasmid nr 22481 ; http://n2t.net/addgene:22481 ; RRID: Addgene_22481) was employed. Cell growth, harvesting and lysis conditions were identical to the his-tagged version, except of the lysis buffer supplementation with dithiotreitol (DTT) to 5 mM. Glutathione agarose CL-4B (GSH-Sepharose, in-house preparation) beads were equilibriated with the purification buffer (300 mM NaCl, 50 mM Tris, 5 mM DTT, *p*H 8.0), mixed with the lysate, and agitated on ice for 3 hours. The beads were washed with the same buffer, and the protein was eluted with 30 mM of reduced GSH. The eluate was collected, desalted into TBS *p*H 8.0, concentrated, aliquoted and frozen.

2.2.2 Purification of HSP90 N-domain

E. coli Rosetta 2 cells bearing pET15b-$HSP90_{9-236}$ were inoculated and grown in 500 mL LB-Amp-Cam to $OD_{600\,nm} = 0.6$ at 37 °C. The cells were harvested by centrifugation (5000×g, 15 min, 4 °C) and washed with 1 × M9 salts (67.5 mM Na_2HPO_4, 44 mM KH_2PO_4, 17.1 mM NaCl, *p*H 7.3). The cells were then transferred to 100 mL M9 mineral medium (Appendix B.2), and let to regenerate at room temperature for an hour under extensive aeration. Expression was induced with 0.4 mM IPTG, and the cells were grown overnight at 20 °C. Cells were harvested, washed using 1×M9 salts, and lysed using a French press at 1.7 kbar. The lysis buffer consisted of 10 mM NaH_2PO_4, 300 mM NaCl, 2 mM $MgCl_2$, 0.5 mM $CaCl_2$, 10 µg mL^{-1} of both DNAse I and RNAse I, 5 mM phenylmethylsulfonyl fluoride, *p*H 7.5.

$HSP90_{9-236}$ was purified in a batch manner with Nickel-NTA beads. The beads were equilibrated using IMAC buffer C (10 mM NaH_2PO_4, 300 mM NaCl, *p*H 7.5), and the lysate was gently agitated with the affinity beads on ice for 1 h. The beads were washed with the IMAC C buffer supplemented to 50 mM imidazole, the protein-containing fractions were eluted with 300 mM imidazole. The fractions were desalted into PBS *p*H 7.4 (Appendix A.2). The 6×His tag was removed from the protein by incubating with human thrombin overnight at 4 °C. At the final step, the purification was done on an anion exchange column using a citric acid buffer (50 mM citric acid, 50 mM NaCl, *p*H 6.0), the fraction of interest eluted in 1 M NaCl, and desalted into TBS *p*H 8.0. The purified protein was concentrated using a 3 kDa cut-off centrifugal microconcentrator, aliquoted and frozen.

2.2.3 Purification of SMYD3

The expression construct pET15b-SMYD3 was prepared by Vladimir Talibov (Uppsala University, Sweden). The expression was induced in LB-Amp-Cam broth

at $OD_{600\ nm}$ = 0.6 with 0.4 mM IPTG, with an additional supplementation of the growth media with $ZnSO_4$ to 50 µM, and the cells were grown overnight at 22 °C. The cells were lysed using a French Press at 1.7 kBars. The lysis buffer consisted of 50 mM Tris, 300 mM NaCl, 10 mM imidazole, 2 mM $MgCl_2$, 0.5 mM $CaCl_2$, 10 µg mL^{-1} of both DNAse I and RNAse I, 5 mM phenylmethylsulfonyl fluoride, 5 mM 2-mercaptoethanol, *p*H 8.0.

A crude SMYD3-containing fraction was obtained using IMAC as follows: loading with Buffer D - 50 mM Tris, 300 mM NaCl, 5 mM 2-mercaptoethanol (2-ME); an intermediate wash with 50 mM imidazole, and elution at 300 mM imidazole. The obtained fraction was desalted into TBS *p*H 8.0, supplemented with 2 mM DTT. The 6×His tag was removed from the protein by incubating the fraction with human thrombin on ice overnight (0.5 U thrombin per apparent mg of protein); the next day, the proteolytic reaction was supplemented to 40 mM imidazole and 300 mM NaCl and passed through reverse IMAC. The flow through was desalted to 50 mM TRIS, 50 mM NaCl, 5 mM 2-ME, *p*H 7.4, and the residual impurities were removed using anion-exchange chromatography, and the fraction of interest eluted at 200 mM NaCl. The SMYD3-containing fraction was desalted in a storage buffer (TBS with 2 mM dithiothreitol (DTT), *p*H 8.0), and the protein was concentrated to more than 10 mg mL^{-1} employing centrifugal microconcentrators with a 10 or 30 kDa Mw cut-off. The protein samples were aliquoted and snap-frozen in liquid nitrogen.

2.2.4 Gels for electrophoresis

DNA Agarose gels were run at 250V at room temperature. The gel composed of varying concentrations of agarose (w/w) in LAB buffer (10 mM lithium acetate, 10 mM boric acid, *p*H 6.9). GelRed Nucleic Acid Gel Stain (Biotinum, USA) was used to stain the gels, and a DNA Gel Loading Dye (ThermoFisher Scientific, Netherlands) was used for the sample preparation.

Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis (SDS-PAGE) was performed at 4 °C or room temperature. The gel and buffer compositions either followed the Laemmli protocol [33], or were made employing the commercial MES-based discontinuous system NuPAGE (ThermoFisher Scientific, Netherlands). After complete separation, the gels were washed twice with distilled water, and stained using in-house made colloidal Coomassie G-250 staining solution (60 mg L^{-1} G-250, 35 mM HCl) [34]). For the in-house prepared gels, the stacking gel composed of 6 % acrylamide, 125 mM Tris *p*H 6.6, 0.1 % SDS, 0.1 % APS, 0.01 % TEMED. The separating gel composition was 375 mM Tris *p*H 8.8, 0.1 % SDS, 0.1 % APS, 0.01 % TEMED, and the acrylamide percentage was varied depending on the desired separation range.

PageRuler Unstained Protein Ladder was used as the reference sample, and a 4× Laemmli protein sample buffer (BioRad, USA) as the sample preparation buffer.

Gels were imaged using a ChemiDoc imaging system (BioRad, USA).

2.3 Surface plasmon resonance

2.3.1 SMYD3-functionalized sensing surfaces

Interaction kinetic experiments were carried out using SPR-based biosensors Biacore 3000 and Biacore S51 (GE Healthcare, Sweden). SMYD3 (250 µg mL^{-1} in 10 mM bis-Tris, *p*H 7.0) was immobilized on the surface using an amine coupling protocol at a surface density of 12000 response units, employing HBS-T (10 mM HEPES, 150 mM NaCl, 0.05 % Tween-20, *p*H 7.5) as a running buffer. Depending on the experimental setup, denatured SMYD3 was immobilized as a reference. The experiments were carried out at 15 °C. Unless stated otherwise, running buffer for the interaction experiments was TBS-T (TBS, supplemented with 0.05 % Tween-20 and 2 mM DTT), supplemented with DMSO to 2-4 % wherever work with synthetic molecules was carried out. Ligands were prepared in a 2-fold concentration series, with 2 blank injections of buffer between each series. The injection time as well as the dissociation phase were varied depending on the ligand. Solvent correction cycles were included in the assay scheme to correct possible mismatches in the refractive index of the samples whenever the running buffer was supplemented with DMSO.

2.3.2 HSP90-functionalized sensing surfaces

GST-HSP90$_{626-732}$ (25 µg mL^{-1} in 10 mM sodium acetate, *p*H 5.0 buffer) was immobilized using amine coupling at a surface density of 200-1000 response units using HBS-T running buffer. The experiments were carried out at 15 or 25 °C, employing TBS-T as the running buffer. After each cycle, the surface was regenerated with a 60 s injection of 1 M ethanolamine, *p*H 8.5, followed by an additional baseline stabilization for 1 min.

2.4 MST experiments

Label-free microscale thermophoresis experiments (MST) were carried out using the Monolith NT.Automated instrument (NanoTemper Technologies, Germany). The capillaries were filled with a solution containing 2 µM protein, various concentrations of the peptides, in 2 mM DTT, 1.6 % DMSO, 0.05 % Tween-20, 50 mM Tris, 150 mM NaCl, *p*H 8.0 buffer. The concentration of the protein was kept a constant, with each consecutive capillary having a three fold dilution of the peptides **1**, **2**, **3** or **4**, spanning a concentration range from 400 to 0.1 µM. Measurements (figure 3.7) were carried out at 20 % UV power, 80 % infrared laser power, measuring 5 s for the baseline, 10 s for the gradient, 5 s for the back-diffusion.

2.5 TSA experiments

Thermal Shift Assay experiments were carried out using the NanoTemper Tycho NT.6 instrument (NanoTemper Technologies, Germany). The capillaries were filled

with a solution containing 1.5 mg mL^{-1} SMYD3, 200 mM NaCl, and 100 mM of the buffer component. The buffer used varied in *p*H and buffering agent: *p*H 5.0-5.5 Citrate, 6.0-6.5 MES, *p*H 6.85-7.0 bis-Tris or *p*H 7.0 MOPS, *p*H 7.25-7.5 HEPES, *p*H 7.75-8.5 Tris, *p*H 8.5-9 bicine, *p*H 9.5 Ethanolamine. Further experiments on the quality of the protein between batches were carried out with 1.5 mg mL^{-1} SMYD3 in a storage buffer (TBS with 2 mM DTT, *p*H 8.0).

2.6 Crystallization

2.6.1 Crystallization of SMYD3

Crystallization trials were performed at RT with a hanging drop vapour diffusion method, with a total drop volume of 2 µL. SMYD3 in a storage buffer (TBS, 2 mM DTT, *p*H 8.0) at concentrations 9-17 mg mL^{-1} was used in grid screenings, varying PEG 3550 or PEG8000 concentrations (9-17%, v/v), DMSO (0-10%), 100 mM buffer component (*p*H 6.85 bis-Tris, *p*H 7.0 MOPS, *p*H 7.25-7.5 HEPES, *p*H 7.75-8.5 Tris, *p*H 8.75-9 bicine), and magnesium or calcium acetate 50-100 mM. The obtained crystals were cryoprotected in the reservoir solution, containing either 20 % (v/v) 1,5-pentanediol or 10 % glycerol, mounted on cryoloops and snap-frozen in liquid nitrogen.

2.6.2 Co-crystallization of SMYD3-ligand complexes

Co-crystallization experiments were performed similar to SMYD3 crystallization, but with minor changes: reservoir and protein solutions were supplemented with DMSO to 10-20 % (v/v), and ligands of interest were pre-incubated with the protein at room temperature or 4 °C at concentrations 0.5-5 mM for 1 to 12 h. Obtained crystals were cryoprotected with 10 % (v/v) glycerol.

2.6.3 Data collection

Diffraction data was collected at the European Synchrotron Radiation Facility beamlines ID24 and ID30B (ESRF, Grenoble, France) or MAXIV beamline BioMAX (Lund, Sweden). The diffraction data was indexed and integrated on site ([35]), scaled, merged and reduced employing aimless package from CCP4 suite ([36], [37]). The phase problem was solved using molecular replacement by the Phaser software ([38]). Model building was performed with Coot ([39]), refinement with refmac5 ([40]). Restraints for the synthetic ligands were generated using elBOW ([41]).

3
Results

3.1 Protein expression and purification

3.1.1 Expression and purification

In this thesis, three unique expression constructs were generated by means of molecular biology: pET15b-HSP90$_{9-236}$, pET15b-HSP90$_{626-732}$ and pET15b-HSP90$_{293-732}$.

Molecular cloning was performed using the standard approach of plasmid ligation. For the designed constructs, cDNA 5'-and 3' extension was carried out to flank the fragments of interest with the restriction sites, and to add an extra overhang to promote the binding of endonucleases. For HSP90$_{9-236}$ and HSP90$_{626-732}$, this process required one round of DNA polymerase-based amplification. However, the source cDNA used for the generation of pET15b-HSP90$_{293-732}$ revealed the presence of a unique endogenic BamHI site (sequence GGATCC), spanning nucleotides 2034-2039. In order to subclone the construct, the region was mutated. The mutation was carried out at position 2034, with a G2A substitution. The sequence at 2032-2040 in the wild-type cDNA is GAGGATCCC, translating to Glu-Asp-Pro at HSP90 amino acid residues 678-680. The G2A substitution at position 2034 altered the first codon in this region from GAG to GAA, retaining Glu678 in its corresponding reading frame.

Expression and purification of the HSP90 domains and SMYD3 was carried out. HSP90$_{626-732}$ was purified as a GST-fuse, and as a 6×His tagged and tag-free form, with a yield of more than 8 mg mL^{-1} of pure protein from 1.5 L of culture. Similarly, HSP90$_{9-236}$ was purified as a 6×His tagged or tag-free form, with a yield of more than 7 mg mL^{-1} of pure protein from 1.5 L of culture.

The N-terminal domain of HSP90, HSP90$_{9-236}$, was also probed for expression in M9 mineral medium. Surprisingly, the yields were higher than for LB-based standard expression protocol, with a strong over-expression seen already in the soluble fraction of the lysate.

HSP90$_{293-732}$ was expressed only in a series of trial experiments. The protein was shown to express in the probed conditions in a soluble form, but optimization of the system and large-scale purification was not performed. This form of HSP90 was not used for downstream applications.

The expression conditions of SMYD3 were developed prior to this work, but they were optimized in order to increase the yield of the protein. Optimization included varying the temperature, the composition of the expression medium, and the concentration of inducer. The best conditions were identified as induction with 0.4 mM IPTG at 22 °C for 16-20 h in LB media supplemented with $ZnSO_4$ to 50 µM. Sur-

prisingly, expression in richer media, *e.g.* Terrific Broth or 2×YT, did not increase the protein yields. The enzyme was purified in two forms, as a 6×His tagged- or tag-free form, with an end yield of more than 10 mg mL^{-1} of pure protein from 3 L of LB-based culture. Protein preparations of crystallographic quality were obtained regularly following a refined purification protocol, Figure 3.1,

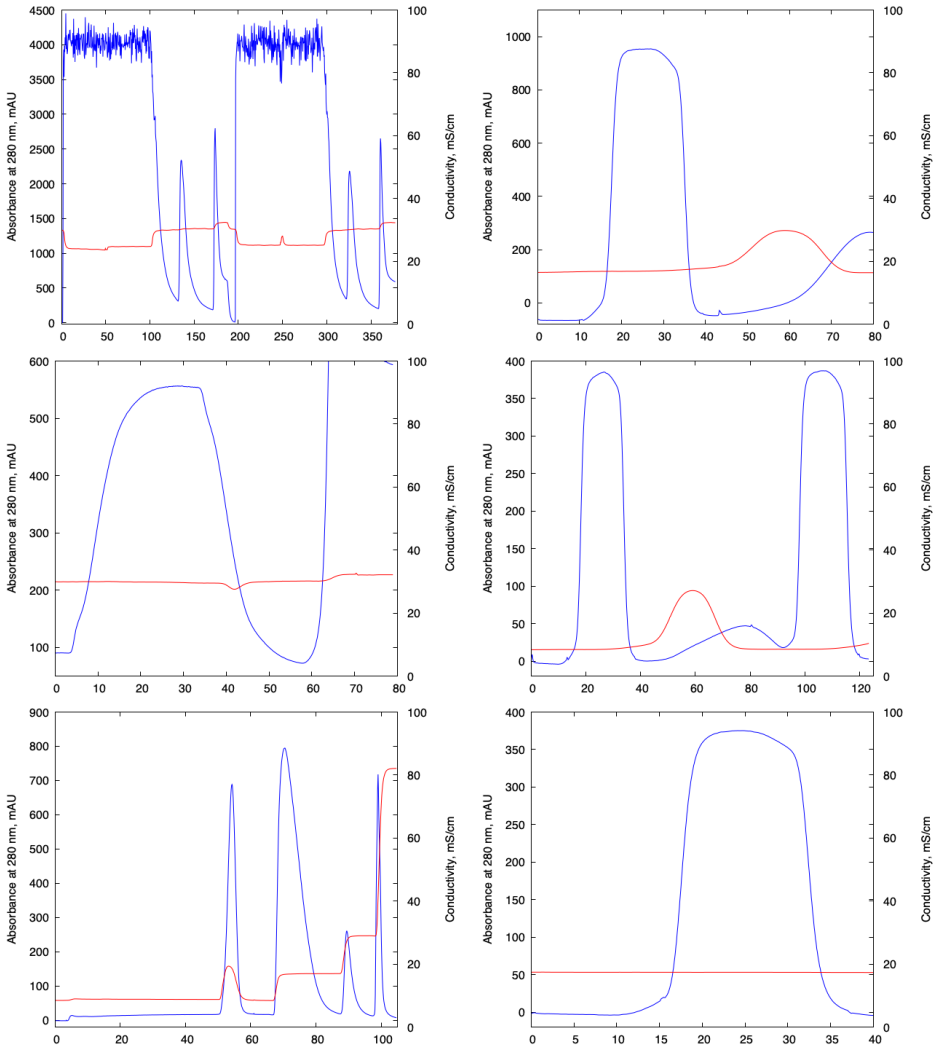

Figure 3.1: Purification of SMYD3 on the ÄKTA FPLC system. The chromatograms follow the purification steps of SMYD3. Top left: IMAC, top middle: desalting to TBS *p*H 8.0, top right: reverse IMAC, bottom left: desalting to TBS *p*H 7.4, bottom middle: anion exchange, bottom right: desalting to the storage buffer. The blue line responds to $A_{280\ nm}$ and the red one to the conductivity of the mobile phase.

The homogeneity of the produced proteins was verified via SDS-PAGE analysis and analytical size-exclusion FPLC on Superose 12 media. All the produced proteins had the correct size and high homogeneity according to the SDS-PAGE analysis: HSP90CTD around 12 kDa, HSP90NTD 25 kDa, SMYD3 49 kDa. Figure 3.2 shows a comparison between HSP90 C-terminal domain, purified with or without a 6×His- or GST tag.

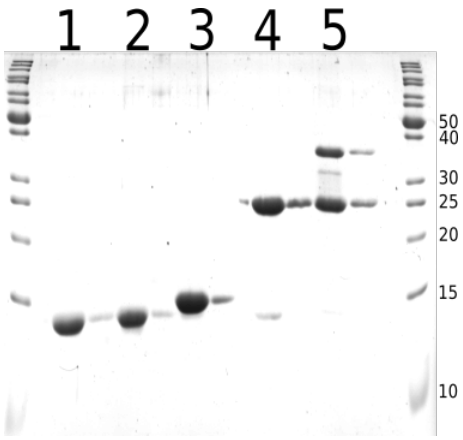

Figure 3.2: 15% SDS-PAGE, 250 µg mL^{-1} and 50 µg mL^{-1} protein per lane. 1: March HSP90CTD; 2: July HSP90CTD; 3: March HSP90CTD 6×His; 4: HSP90CTD GST cleaved; 5: HSP90CTD GST-fuse. Molecular weight in kDa on the right: PageRuler Unstained Protein ladder.

3.1.2 Thermal shift assays

To guarantee high quality of protein batches and to find suitable conditions for downstream applications, the effects of *p*H and certain additives to SMYD3 thermal stability was also evaluated, as seen in Table 3.1 and Figure 3.3. A Tris-based buffer with a *p*H of 8.0 yielded the highest inflection temperature for SMYD3, and was thus chosen as the buffering agent for SMYD3 storage. Thermal shift assays were also used to compare the quality of tryptophan-containing purified proteins in and between protein preparations (Figure 3.4), to guarantee batch-to-batch consistency.

Table 3.1: SMYD3 melting temperature as a function of buffer composition

*p*H	DMSO (%)	buffer	T_m, °C
5.0	0	Citrate	ni
5.5	0	Citrate	ni
6.0	0	MES	ni
6.5	0	MES	45
7.0	0	bis-Tris	49
7.5	0	HEPES	51
8.0	0	Tris	51
8.0	2	Tris	50
8.0	4	Tris	50
8.0	6	Tris	49
8.0	8	Tris	48
8.0	10	Tris	48
8.5	0	Bicine	49
9.0	0	Bicine	49
9.5	0	Ethanolamine	ni

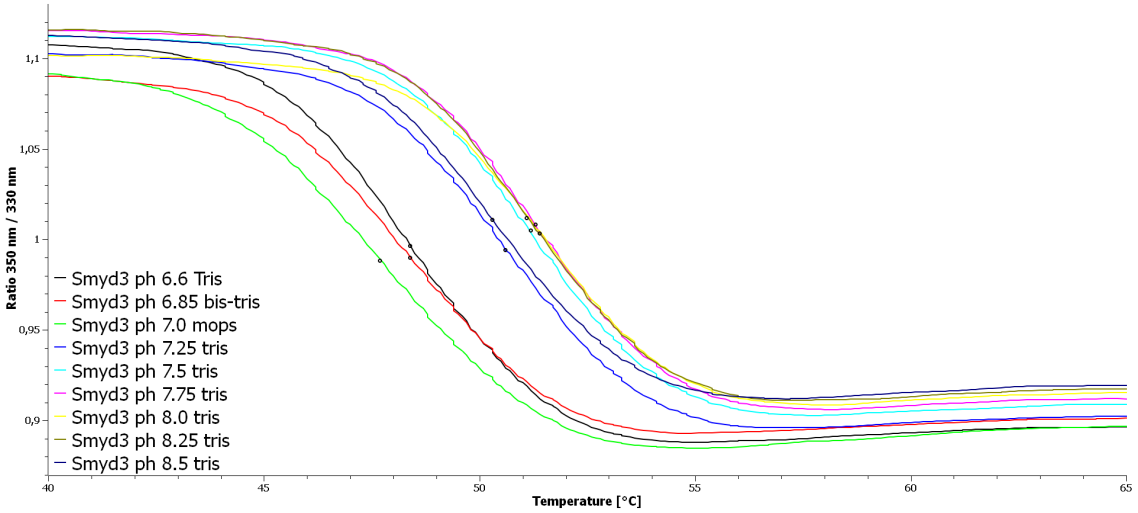

Figure 3.3: TSA analysis of the effects of storage buffer *p*H on the stability of 6×His-tagged SMYD3. T_m *p*H 7.0 mops 47.7 °C; T_m *p*H 8.0 tris 51.4 °C

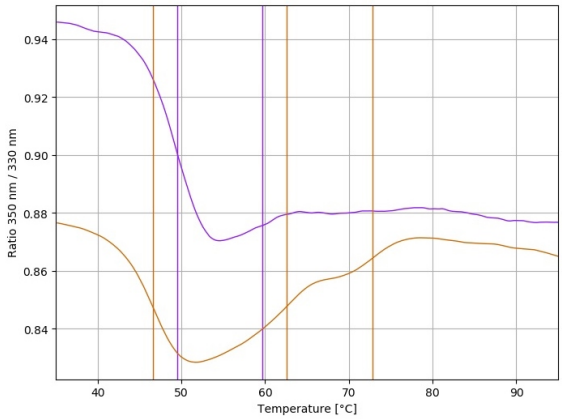

Figure 3.4: TSA analysis of two different batches of 6×His-tagged SMYD3. A higher T_m and a more distinctive differential melting profile corresponds to a higher percentage of stable protein in a batch, and therefore a higher quality of the preparation. Ratio 350/330 nm on the Y-axis, temperature in °C on the X-axis.

3.2 Interaction analysis

3.2.1 SMYD3 interactions with natural ligands

In order to establish an interaction kinetic assay and to monitor the activity of immobilized SMYD3 during the biosensor-based experiments, interactions with the methylation reaction by-product S-adenosyl-L-homocysteine (SAH) were established. The assay was performed at 15 °C and was based on a design developed prior to this work. SAH interacted with immobilized SMYD3 in a specific manner, with a 1:1 stochiometry and with resolvable kinetics, which was attributed to a 1:1 Langmuir-alike interaction model, Figure 3.5. The quantified interaction parameters were

as followed: $K_D = (9.8 \pm 0.9) \times 10^{-8}\,\mathrm{M}$, $k_{on} = (6.3 \pm 3.2) \times 10^{5}\,\mathrm{M^{-1}\,s^{-1}}$, $k_{off} = (6.4 \pm 3.4) \times 10^{-1}\,\mathrm{s^{-1}}$ ($n = 3$). The quantified parameters were similar to the ones observed before (*E. Fabini, V. Talibov et al., Submitted, 2019*).

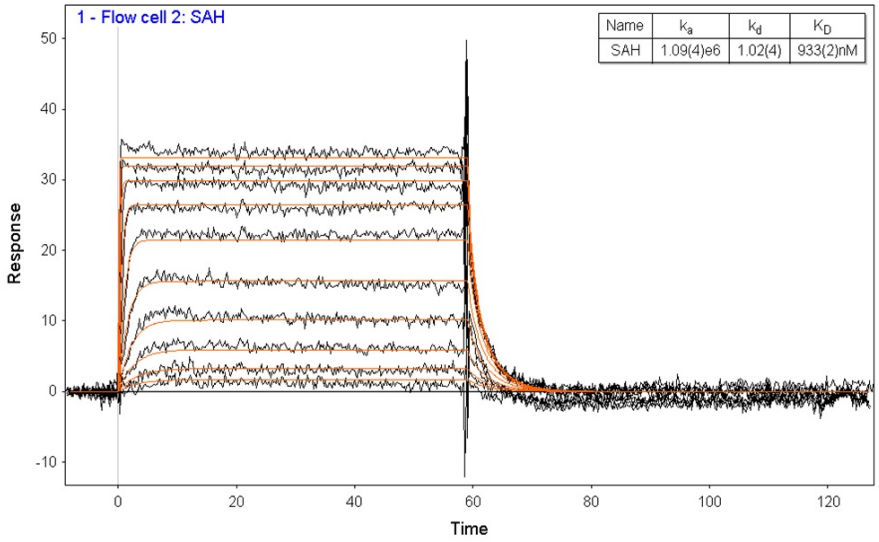

Figure 3.5: Biosensor analysis of SMYD3 interactions with SAH. The interaction of SMYD3 with SAH validates the assay. Interaction kinetic curves were fitted using a reversible second order rate equation; the inset corresponds to extracted interaction parameters. Response units on the Y-axis, time (seconds) on the X-axis.

3.2.2 SMYD3 interactions with peptides

SPR-based interaction kinetic assays of the HSP90 C-terminal domain-derived peptides (Table 3.2) with surface-immobilized SMYD3 showed an interaction with peptides 2, 3, and 4, but no interaction with peptide 1 (Figure 3.6). The interactions appeared to be complex and could not be analysed by any simple kinetic model, however, the specific nature of the interactions could be seen from an absence of a significant response of structurally-related peptide 1.

Table 3.2: Primary structures of HSP90 C-terminal domain-based peptides 1, 2, 3 and 4.

Peptide	Structure	Net charge
Peptide 1	NH_2-MEEVD-COOH	-3
Peptide 2	AcNH-MEEVD-COOH	-4
Peptide 3	NH_2-TSRMEEVD-COOH	-2
Peptide 4	NH_2-DDDTSRMEEVD-COOH	-5

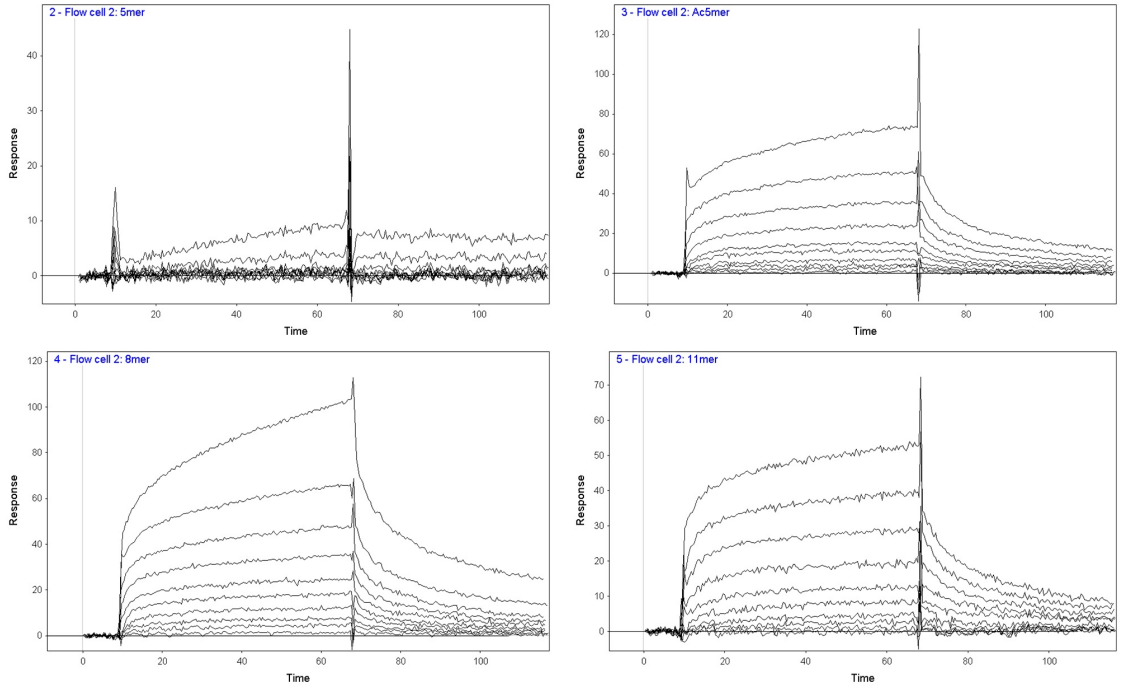

Figure 3.6: Biosensor analysis of SMYD3 interactions with peptides 1,2,3 and 4. SAH (top left), peptides 1 (top right), 2 (bottom left), 3 (bottom middle) and 4 (bottom right). No interaction was observed for peptide 1, unlike from peptides 2, 3 and 4. Response units on the Y-axis, time (seconds) on the X-axis.

To confirm this observation with an orthogonal method, a microscale thermophoresis analysis with the given peptides was executed in a label-free mode (Figure 3.7). The qualitative nature of interactions remained the same, with an observable change in the thermophoretic mobility of SMYD3 in complexes with peptides 2-4, but not with peptide 1.

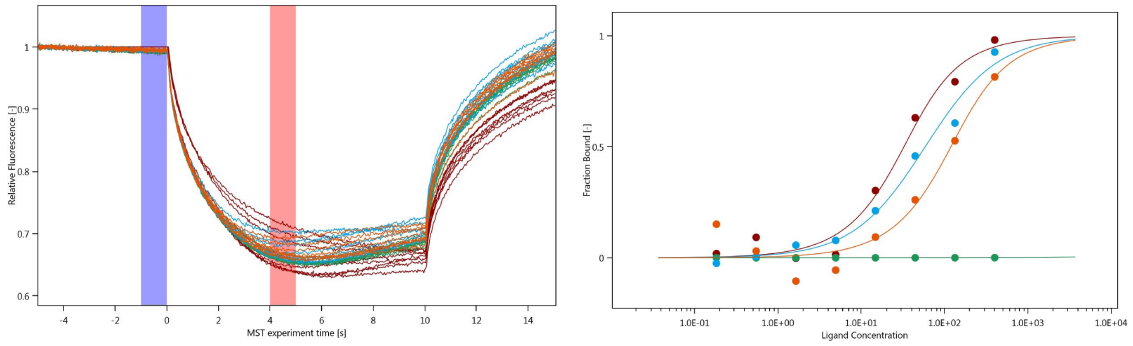

Figure 3.7: Thermophoretic mobility of SMYD3 was measured in a label-free manner over a range of 400 µM to 91 nM of HSP90 C-terminal domain-based peptides. The results were plotted as a function of fraction of SMYD3 bound to the peptides (right panel). Peptide 1 - green, peptide 2 - red, peptide 3 - blue, peptide 4 - orange. Left: Y - relative fluorescence [-], X - MST experiment time. Right: Y - Fraction bound [-], X - Ligand concentration (Molar)

3.2.2.1 SMYD3 interactions with HSP90 domains

Surface plasmon resonance biosensor-based interaction assays were carried out at 25 °C with GST-fused $HSP90_{626-732}$ immobilized on the surface, with injections of SMYD3 from 30 to 0.23 µM (Figure 3.8). The results confirmed a specific interaction with the C-terminal domain of HSP90. The interaction was characterized by a fast on/off rates, with a corresponding steady-state derived K_D value of $(13 \pm 1) \times 10^{-6}$ M.

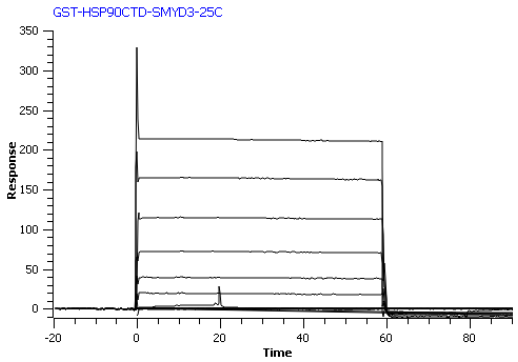

Figure 3.8: Biosensor-based analysis on GST-fused HSP90 C-terminal domain (surface) with 30 to 0.23 µM SMYD3 at 25 °C. Steady-state derived $K_D = (13 \pm 1) \times 10^{-6}$ M.

The experiment was repeated at 15 °C with a serial dilution of SMYD3 from 4 to 0.06 µM (figure 3.9). Surprisingly, at lower temperature the interaction gained in complexity, with an apparent association-and dissociation kinetics. Steady-state analysis was performed on an apparent level, with the resulting apparent $K_D = (1.7 \pm 0.2) \times 10^{-6}$ M. At the end of each injection, GST-HSP90 C-terminal domain surface was regenerated with an injection of 1 M ethanolamine, *p*H 8.5.

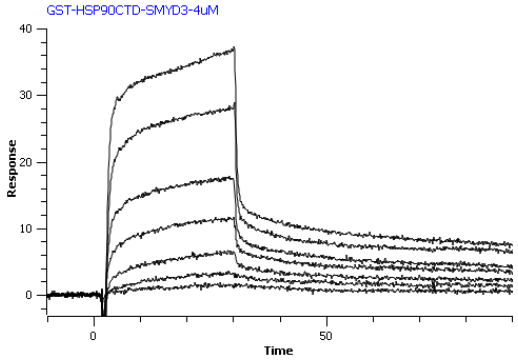

Figure 3.9: Biosensor-based analysis on GST-fused HSP90 CTD (surface) with serial dilutions of 4 to 0.06 µM SMYD3 at 15 °C. Response units on the Y-axis, time (seconds) on the X-axis.

To confirm specificity of the observed interactions and to exclude a possible cross-reactivity of SMYD3 with the GST tag itself, the assay was reversed at 15 °C. SMYD3 was immobilized on the surface, and a dilution series of tag-free

$HSP90_{626-732}$ from 50 to 0.4 µM was injected over the functionalized surface (Figure 3.10). In the reverse setup, the relative binding activity of SMYD3 towards HSP90 did not exceed 10 %, in contrast to its activity against S-adenosyl-L-homocysteine. However, interaction kinetic curves were similar to the one obtained for HSP90-fictionalized surface at 15 °C (Figure 3.9).

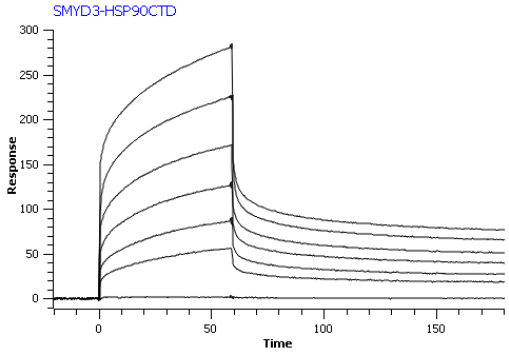

Figure 3.10: Biosensor-based analysis on SMYD3 (surface) interactions with serial dilutions of 50 to 0.4 µM HSP90 CTD at 15 °C. Response units on the Y-axis, time (seconds) on the X-axis.

As a hypothetical negative control, SMYD3 was immobilized on the surface, and serial injections of $HSP90_{9-236}$ (50 to 0.4 µM) were performed at 15 °C (Figure 3.11). Surprisingly, an interaction was observed, with extremely slow apparent association/dissociation kinetics. However, the amplitude of the signal was much lower than the one observed for $HSP90_{626-732}$, a polypeptide with a twice lower molecular weight.

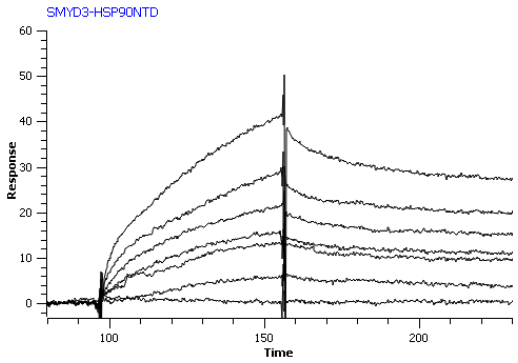

Figure 3.11: SPR analysis on SMYD3 with serial dilutions of 50 to 0.4 µM HSP90 NTD at 15 °C. Response units on the Y-axis, time (seconds) on the X-axis.

3.2.2.2 SMYD3 interactions with small synthetic molecules

Biosensor-based interaction assays were set up with SMYD3, immobilized at surface densities sensitive enough to detect recognition events with fragment-sized organic molecules, fragments **1**, **2** and **3**. Surface plasmon resonance experiments were also carried out on benzamidine (Figure 3.12), a potential bioisostere of N-phenylcarbamate with a positive charge, and four additional compounds **5**, **6**, **7**

and **8**, screened *in silico* (Figures 3.12, 3.13). For the smallest ligands, *i.e.* **1-3** and benzamidine, no interaction was observed. However, signatures of a low-affinity interactions were observed for compounds **6-8**, while no interaction with SMYD3 was detected for **5**.

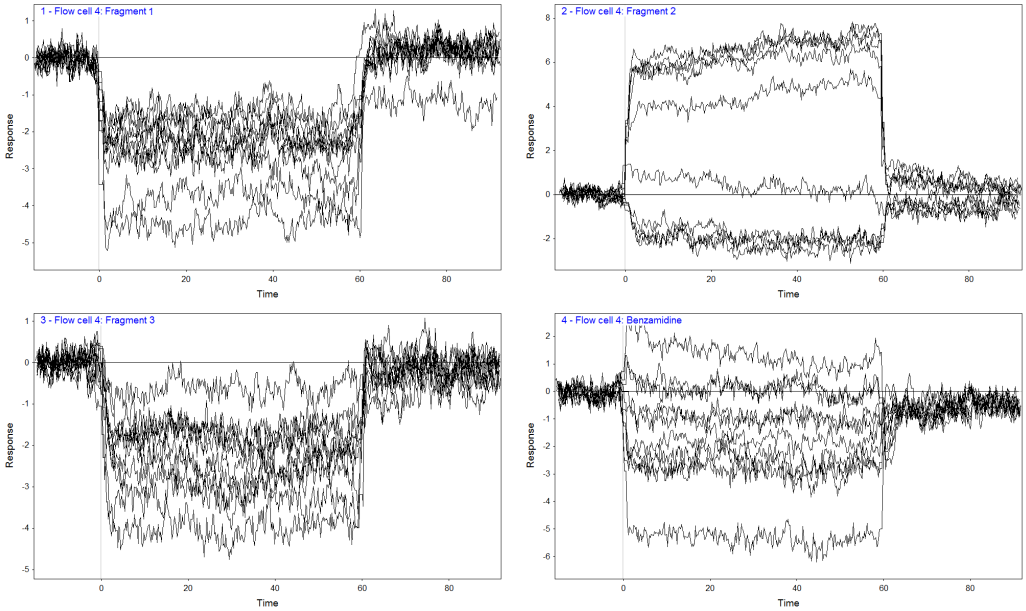

Figure 3.12: Biosensor analysis on fragments **1**, **2**, **3** and benzamidine. No recognition events were detected for titration series starting from 400 µM. Response units on the Y-axis, time (seconds) on the X-axis.

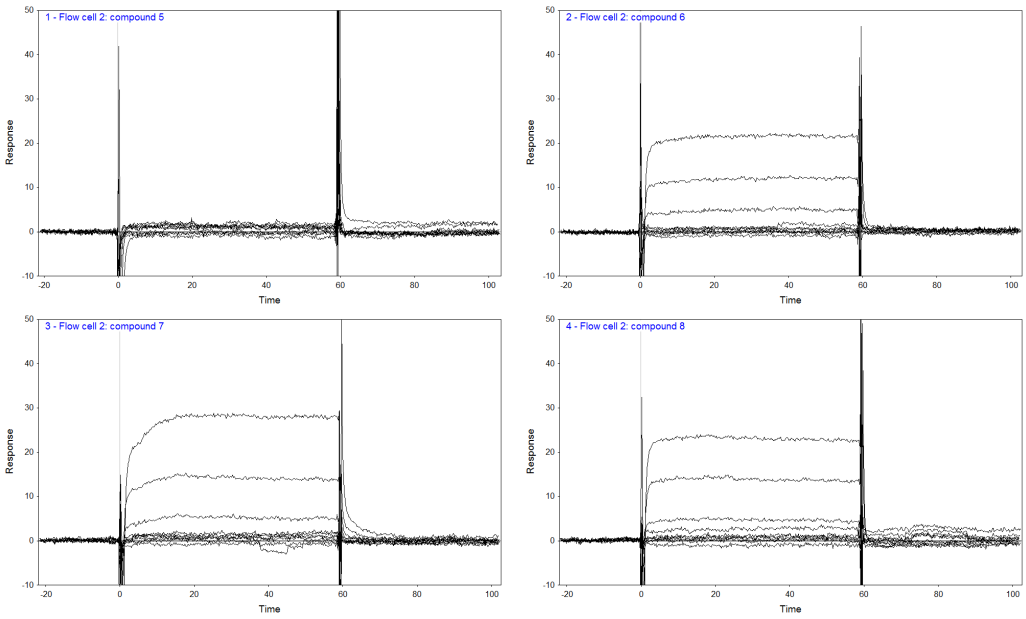

Figure 3.13: Biosensor analysis on compounds **5**, **6**, **7** and **8**. The interaction kinetic curves highlighted potential interactions with **6-8**, but no interaction with compound **5**. Response units on the Y-axis, time (seconds) on the X-axis.

3.2.3 X-Ray crystallography

3.2.3.1 SMYD3 crystallization

The initial crystallization conditions for SMYD3 were found prior to this work: 100 mM Tris *p*H 8.0, 13 % PEG8000, 50 mM Mg(OAc)2, 12 mg mL^{-1} SMYD3; P_{21}, two macromolecular chains per asymmetric unit. However, they were modified extensively, employing multiple manual and robotic grid screens, and an alternative crystal form with better diffraction and packing parameters was obtained.

Initially, crystallization conditions were probed at RT with both His-tagged SMYD3 and the untagged form of the enzyme. No hits were detected with the tagged protein. Trials were conducted in a hanging drop vapour diffusion set up with a total drop volume of 2 µL, with two-dimensional multidrop screening varying the protein-to-reservoir solution ratios (0.5:1.5, 1:1, 1.5:0.5), precipitant concentration (PEG8000 or PEG3350), additives ($Mg(OAc)_2$, $Ca(OAc)_2$, Li/Na/K alkaline mix, DMSO) and buffer *p*H (7.0-9.0).

The best crystals were obtained in 13% PEG3550, 100 mM $Mg(OAc)_2$, 100 mM Tris *p*H 7.75-8.25, 0.5 mM SAM, with or without DMSO supplementation. The crystals nucleated within a few hours after the set up, and grew to their maximum dimensions within one week. A crystal loaded onto a cryoloop and it's diffratction pattern are shown in figures 3.14 and 3.15

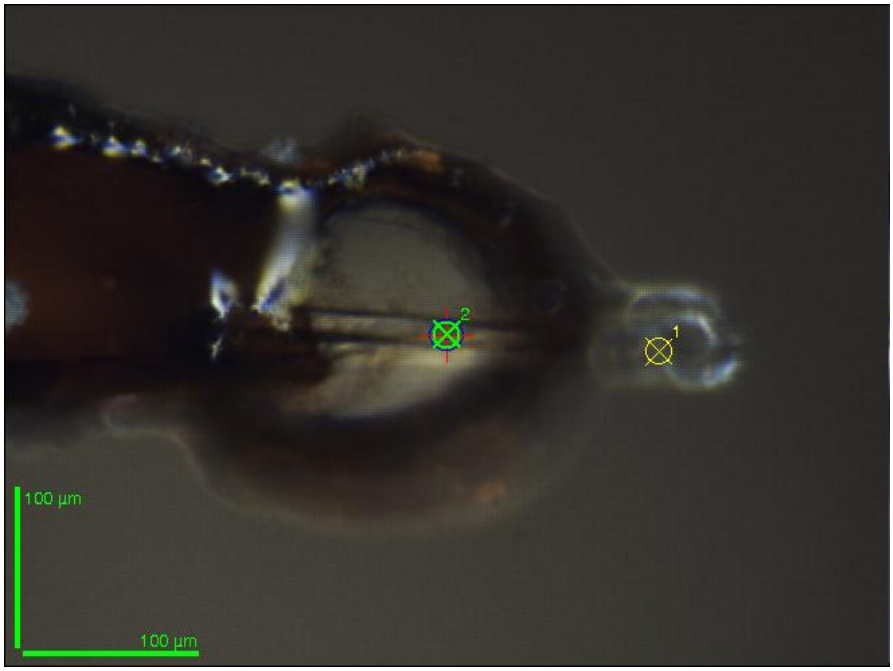

Figure 3.14: A form 2 monocrystal of SMYD3, mounted on a cryoloop and installed at the beamline

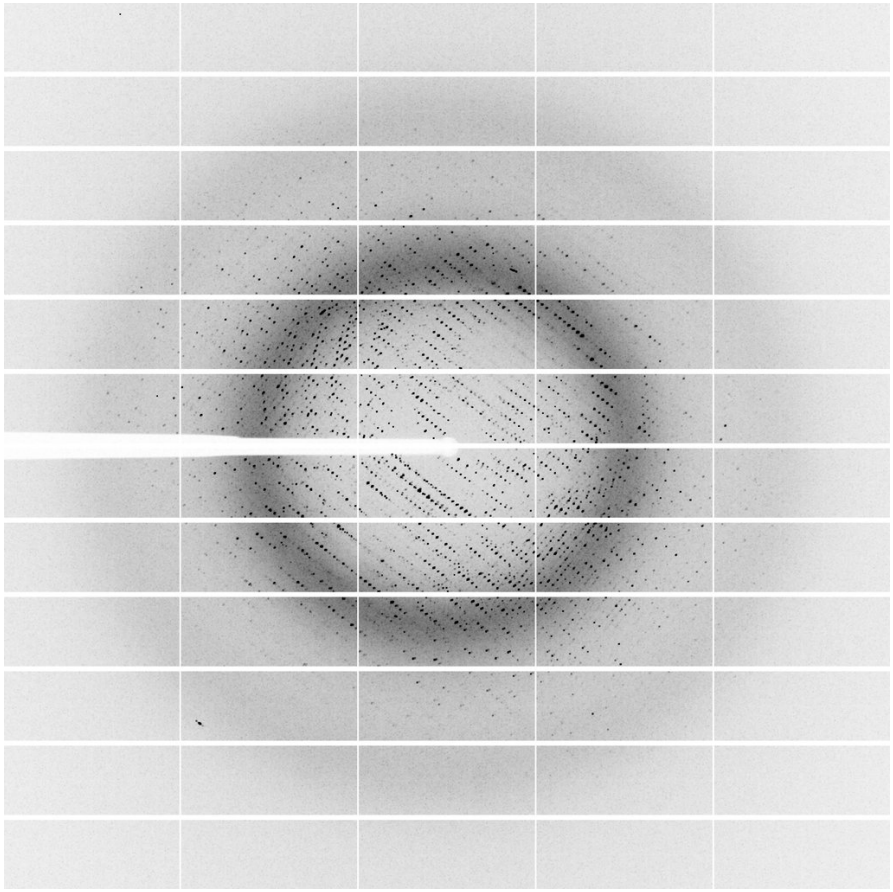

Figure 3.15: A diffraction pattern of a SMYD3 crystal, form 2

Curious enough, DMSO-free conditions resulted in a simultaneous growth of two crystal forms: a prism-like or pinacoidal crystals (form 1) and columnar crystals, grouped in clusters (form 2). Supplementation of the reservoir solutions with DMSO promoted the nucleation and growth of form 2 exclusively. Also, form 2 was often observed whenever PEG3350 was used as a precipitant, while it did not appear in PEG8000-equilibrated systems.

Attempts to co-crystallize SMYD3 with peptides 2-4 were performed, but did not succeeded. Planar microcrystals appeared in commercial crystallization screens, but did not diffract.

Co-crystallization and soaking experiments were performed with a set of small organic molecules, derived computationally from a structure of SMYD3-bound DIP2 compound (compounds **1-3** and benzamidine). The crystals diffracted, but solving the structure did not yield promising positioning for the small molecules. It correlated well with an absence of SMYD3 recognition of the very same molecules in biosensor-based experiments, as mentioned above.

3.2.3.2 Diffraction data analysis

Diffraction data from the form 2 crystals was collected on a regular basis, employing synchrotron radiation facilities in ESRF (Grenoble, France) and MAXIV (Lund, Sweden). In comparison to the priorly discovered crystal form 1, form 2 led to a

better diffraction, with routinely observed resolution significantly below 2 Å. An example of the reduction statistics for crystal form 2 diffraction data set is given in Table 3.3.

Table 3.3: An example of diffraction data collection and reduction statistics, obtained from SMYD3 crystal form 2. Data was collected at BioMAX beamline (MAXIV, Lund, Sweden) with wavelength 0.976 Å.

Parameter	All shells	High resolution shell
Space group	P_{212121}	
a, b, c, Å	60.94, 65.99, 107.32	
α, β, γ, °	90, 90, 90	
Average mosaicity	0.07	
Resolution, Å	107.32-1.6	1.63-1.6
$R_{p.i.m}$	0.023	0.224
Observations (unique)	420180 (57796)	14217 (2752)
$I/\sigma(I)$	19.2	3.2
$CC_{1/2}$	0.99	0.853
Completeness, %	99.9	98.1
Multiplicity	7.3	5.2

Analysis of the unit cell composition of crystal form 2 revealed differences in comparison to form 1. The phase problem was solved with molecular replacement, utilizing ligand-free PDB 5CCM [42] as a search model. In comparison to the previously developed form 1, form 2's asymmetric unit consisted of one macromolecular chain. Additionally, form 2 diffracted better than form 1, in particular with a higher signal-to-noise ratio ($I/\sigma(I)$) and better reduction statistics.

4

Discussion

4.1 SMYD3 interactions with small molecules

The assay presented here was originally designed to score interactions between surface-immobilized SMYD3 and small molecular weight analytes, including SMYD3 natural ligands and synthetic molecules. As shown in Figure 3.5, the assay is sensitive enough to probe potential interaction partners with a Mw below 400 Da. In fact, the surface binding capacity towards the co-factor site regularly exceeded 60 %. Judging from the repeatable injections of the control compound SAH, the assay was stable for multiple days when carried out at 15 °C.

Besides the reaction by-product, a set of synthetic ligands were probed in interaction kinetic assays. Ligands **1-3** were designed rationally by Edward Fitzgerald (Uppsala University, Sweden) on the basis of N-phenyl-carbamate binding moieties of the compound DIP2. However, they did not demonstrate any appreciable binding, as seen in Figure 3.12. It's worth to mention that these ligands were designed as fragments and were expected to have small affinities towards SMYD3. Nevertheless, no unambiguous electronic densities that could be attributed to ligands **1-4** was found in co-crystallization experiments, suggesting that they are unable to recognize SMYD3.

DIP2 was also evolved into a set of bigger compounds - ligands **5-8**, with Mw approaching 400, having more of a drug-like structure. Signs of interaction were detected for ligands **6**, **7** and **8**, while ligand **5** did not demonstrate any response with SMYD3 even at a concentration of 400 µM, Figure 3.13. Ligands **6-8** were scored as weak binders, due to an absence of saturation signs over the tested concentration range. However, their kinetic features (fast on-and off-rates, absence of secondary effects) are expected from specific binders of high µM affinity. Currently, these ligands are tested in co-crystallization trials to verify the interactions, possibly localized in the DIP2 binding site.

4.2 SMYD3 interactions with HSP90

Supporting the interest in the biology of the discovered allosteric binding site in SMYD3, a set of experiments was performed to study SMYD3-HSP90 interactions. Experimental design was based on attempts to replicate the work presented by Brown et al. [8], where the C-terminal domain of HSP90 was claimed to recognize the C-terminal domain of SMYD3. In the original work, authors used a modified Ni-NTA-beads based extraction assay, with a binding analysis via SDS-PAGE. Herein,

a direct approach utilizing SPR-based biosensors is proposed.

At the initial stage, a set of peptides was probed against surface-immobilized SMYD3. The peptides were design to contain the main recognition motif, C-terminal sequence MEEVD, as proposed by Brown et al. [8]. It is worth mentioning that in the original work the pentapeptide was probed as a fuse to GST. The set of peptides used in the current study were based on the minimal sequence itself, Table 3.2. Surprisingly, peptide **1** did not serve as a minimal recognition motif. In contrast, the addition of extra amino acid residues from its N-terminus rendered peptides **3** and **4** binding-competent, providing saturable interaction kinetic curves of a complex interaction mode, Figure 3.6. Attempts to quantify the parameters of interaction were skipped, as the nature of secondary effects remained unclear. An exciting finding was the discovery of the role of the N-terminal residues. Peptide **2** was designed as an alpha-amino-acetylated analog of peptide **1**, and also interacted with the SMYD3 surface in an appreciable manner. The dramatic difference between the recognition of peptide **1** and peptide **2** is a clear structure-activity relationship, which suggests a specific recognition of the extended MEEVD motif. This observation was confirmed with an alternative method in a qualitative manner, employing microscale thermophoresis in a label-free mode (figure 3.7).

To elaborate the study, various domains of HSP90 were expressed and purified. The design of a protein-protein biosensor-based interaction assay proved to be challenging, possibly due to the low affinity of the studied interaction. In general, reported experiments utilized two types of surfaces: either a SMYD3-modified surface, using tag-free HSP90 C-terminal domain as an analyte, or a reversed assay with GST-HSP90 C-terminal domain fuse being immobilized and SMYD3 injected. For the latter case, a fused protein was used for the immobilization from practical issues - the different amino acid compositions of a Glutathione-S-transferase-based tag (GSH-tag) and the HSP90 C-terminal domain (Appendix D) could potentially provide a site-directed immobilization, with most amide bonds forming with the purification tag.

The immobilized SMYD3 recognized HSP90, Figure 3.10, but the apparent surface activity was way below the expected value. It might be hypothesized that the presence of multiple lysine residues in the C-terminal domain of SMYD3 could render it interaction-incompetent. Nevertheless, when the assay was reversed and GST-HSP90 was immobilized on the surface, the kinetic curves demonstrated a similar shape and behaviour, while the expected response values approached the theoretical ones. For the inverted assay, experiments were carried out both at 15 and 25 °C, Figures 3.9 & 3.8. At room temperature, the interaction was found to have fast on-and off-rates, with a K_D value of 13 µM. The quantified K_D is relatively close to the value reported before [8]. Unexpected secondary effects appeared at 15 °C, mostly envisioned in a complex mechanism of the interaction, similar to the one observed earlier for a direct SMYD3-HSP90 C-terminal domain interaction assay (Figure 3.10) or for the interactions with the MEEVD-based peptides 1-4 (Figure 3.6). An apparent K_D value was quantified despite an absence of a clear steady-state equilibrium, and was found to be 1.7 µM. The role of the temperature factor for the given interaction remains a question.

Attempts to observe a competition between HSP90 C-terminal domain and SMYD3

in the presence of **DIP2** compound were performed, but are rendered incomplete. A competition assay for a weak binder (SMYD3-DIP2 interaction $K_D = 40\,\mu M$) requires a direct set up, with SMYD3 being immobilized on the surface to assure its saturation by the small molecule. As mentioned above, amine coupling-immobilized SMYD3 had surprisingly low relative activity, and the experiments in a reversed assay did not show any significant influence of DIP2 on the SMYD3-HSP90 interaction. Either HSP90 CTD recognizes SMYD3 at a different site, or the potency of DIP2 is not high enough to disrupt the protein-protein interaction interface. A better assay is required to further verify this. One of the possible alternatives is to design an affinity-based capturing system for SMYD3, employing for example a purification tag. It can be achieved with anti-His antibody-based target capturing and will be performed for this system in the future.

4.3 Crystallographic studies

The previously established SMYD3 crystallization conditions were optimized. This resulted in a discovery of conditions translated into a consistent nucleation of form 2 - crystal form with exceptionally high diffraction abilities. Form 2, reported herein, was found to be isomorphic to several SMYD3 crystal systems reported before. However, conditions mapped in the current study supported the nucleation and growth of SMYD3 crystals within a useful chemical space, exemplified by the tolerance to high DMSO concentrations. High quality crystals that can tolerate organic co-solvent (up to 20 % in the current work) are of high interest for crystallographic screening campaigns.

The crystals obtained diffracted with a resolution below 2 Å. Upon processing, the resulting electronic density maps were of high quality, exemplified in Figure 4.1. It is worth noting that the diffraction ability of the crystals was significantly underestimated - as can be seen from Table 3.3; even for the high resolution shell the value of $I/\sigma(I)$ is relatively high. The factual resolution of potential diffraction experiments can be significantly improved. Therefore, crystallization experiments reported herein expand the toolbox of methods which are able to study interactions of SMYD3 with weak, fragment-like binders.

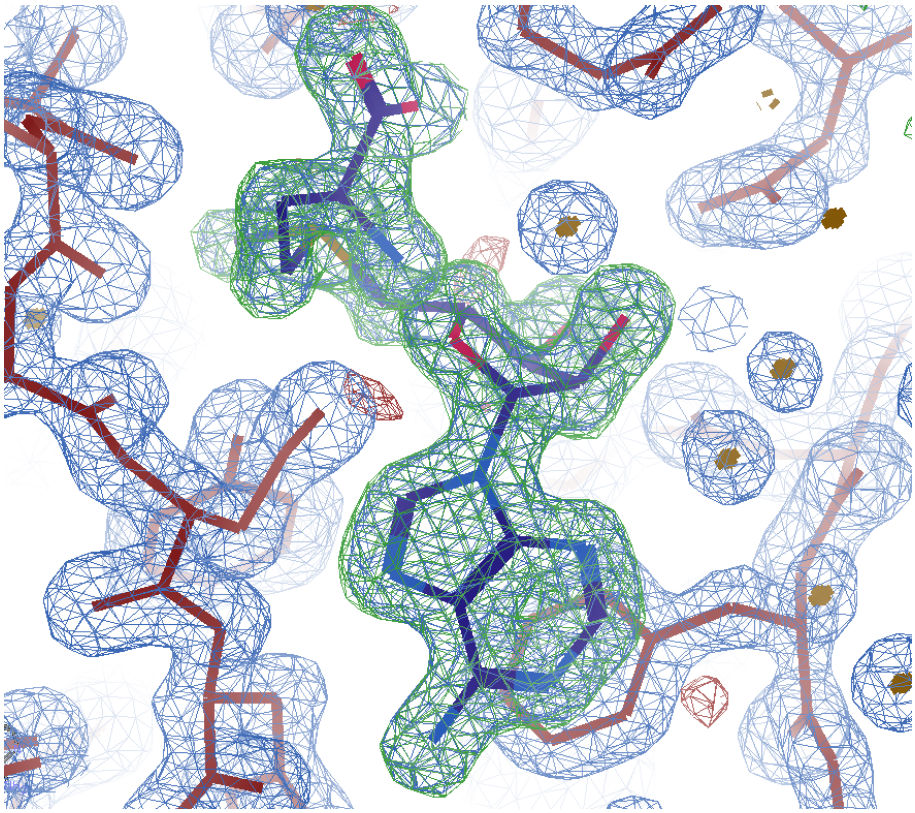

Figure 4.1: An example of a map of electronic densities, derived from the diffraction data sets of SMYD3 crystal form 2. The protein is depicted in red, bound small molecular ligand S-adenosyl-L-methionine (SAM) in dark blue. Blue isomesh - $(2 \times mF_0 - DF_c)$ electronic density map, countered at $1.5 \times \sigma$. Green isomesh - SAM omit $(mF_0 - DF_c)$ difference map, countered at $4 \times \sigma$. The quality of the density maps provide a near-atomic resolution.

Bibliography

[1] Perri F, Longo F, Giuliano M, Sabbatino F, Favia G, Ionna F, et al. Epigenetic control of gene expression: potential implications for cancer treatment. Critical reviews in oncology/hematology. 2017;111:166–172.

[2] Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nature Reviews Genetics. 2016;17(8):487.

[3] Portela A, Esteller M. Epigenetic modifications and human disease. Nature biotechnology. 2010;28(10):1057.

[4] Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429(6990):457.

[5] Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, et al. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nature cell biology. 2004;6(8):731.

[6] Foreman KW, Brown M, Park F, Emtage S, Harriss J, Das C, et al. Structural and functional profiling of the human histone methyltransferase SMYD3. PloS one. 2011;6(7):e22290.

[7] Spellmon N, Holcomb J, Trescott L, Sirinupong N, Yang Z. Structure and function of SET and MYND domain-containing proteins. International journal of molecular sciences. 2015;16(1):1406–1428.

[8] Brown MA, Foreman K, Harriss J, Das C, Zhu L, Edwards M, et al. C-terminal domain of SMYD3 serves as a unique HSP90-regulated motif in oncogenesis. Oncotarget. 2015;6(6):4005.

[9] Sirinupong N, Brunzelle J, Doko E, Yang Z. Structural insights into the autoinhibition and posttranslational activation of histone methyltransferase SmyD3. Journal of molecular biology. 2011;406(1):149–159.

[10] Sarris ME, Moulos P, Haroniti A, Giakountis A, Talianidis I. Smyd3 is a transcriptional potentiator of multiple cancer-promoting genes and required for liver and colon cancer development. Cancer Cell. 2016;29(3):354–366.

[11] Schopf FH, Biebl MM, Buchner J. The HSP90 chaperone machinery. Nature Reviews Molecular Cell Biology. 2017;18(6):345.

[12] Subbarao Sreedhar A, Kalmár É, Csermely P, Shen YF. Hsp90 isoforms: functions, expression and clinical importance. FEBS letters. 2004;562(1-3):11–15.

[13] Prince TL, Kijima T, Tatokoro M, Lee S, Tsutsumi S, Yim K, et al. Client proteins and small molecule inhibitors display distinct binding preferences for constitutive and stress-induced HSP90 isoforms and their conformationally restricted mutants. PloS one. 2015;10(10):e0141786.

[14] Blatch GL, Lässle M. The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. Bioessays. 1999;21(11):932–939.

[15] Chaudhury S, Welch TR, Blagg BS. Hsp90 as a target for drug development. ChemMedChem: Chemistry Enabling Drug Discovery. 2006;1(12):1331–1340.

[16] Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. British journal of pharmacology. 2011;162(6):1239–1249.

[17] Raff M, Alberts B, Lewis J, Johnson A, Roberts K. Molecular Biology of the Cell 4th edition. National Center for Biotechnology InformationÕs Bookshelf; 2002.

[18] Simpson JC. In: Functional Assays. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006. p. 617–620. Available from: https://doi.org/10.1007/3-540-29623-9_3700.

[19] Holdgate G, Geschwindner S, Breeze A, Davies G, Colclough N, Temesi D, et al. Biophysical methods in drug discovery from small molecule to pharmaceutical. In: Protein-Ligand Interactions. Springer; 2013. p. 327–355.

[20] Renaud JP, Chung Cw, Danielson UH, Egner U, Hennig M, Hubbard RE, et al. Biophysics in drug discovery: impact, challenges and opportunities. Nature Reviews Drug Discovery. 2016;15(10):679.

[21] Nordstrom H, Gossas T, Hamalainen M, Kallblad P, Nystrom S, Wallberg H, et al. Identification of MMP-12 inhibitors by using biosensor-based screening of a fragment library. Journal of medicinal chemistry. 2008;51(12):3449–3459.

[22] Nguyen H, Park J, Kang S, Kim M. Surface plasmon resonance: a versatile technique for biosensor applications. Sensors. 2015;15(5):10481–10510.

[23] Olaru A, Bala C, Jaffrezic-Renault N, Aboul-Enein HY. Surface plasmon resonance (SPR) biosensors in pharmaceutical analysis. Critical reviews in analytical chemistry. 2015;45(2):97–105.

[24] Patching SG. Surface plasmon resonance spectroscopy for characterisation of membrane protein–ligand interactions and its potential for drug discovery. Biochimica et Biophysica Acta (BBA)-Biomembranes. 2014;1838(1):43–55.

[25] Williams R, Bridges J. Fluorescence of solutions: A review. Journal of clinical pathology. 1964;17(4):371.

[26] Fedorov O, Niesen FH, Knapp S. Kinase inhibitor selectivity profiling using differential scanning fluorimetry. In: Kinase Inhibitors. Springer; 2012. p. 109–118.

[27] Labant M. Protein–Protein Interactions in Plain View. Genetic Engineering & Biotechnology News. 2018;38(15):1–10.

[28] Wienken CJ, Baaske P, Rothbauer U, Braun D, Duhr S. Protein-binding assays in biological liquids using microscale thermophoresis. Nature communications. 2010;1:100.

[29] Schoenlein R, Chattopadhyay S, Chong H, Glover T, Heimann P, Shank C, et al. Generation of femtosecond pulses of synchrotron radiation. Science. 2000;287(5461):2237–2240.

[30] Chayen NE, Saridakis E. Protein crystallization: from purified protein to diffraction-quality crystal. Nature methods. 2008;5(2):147.

[31] Tickle I, Sharff A, Vinković M, Yon J, Jhoti H. High-throughput protein crystallography and drug discovery. Chemical Society Reviews. 2004;33(8):558–565.

[32] Inoue H, Nojima H, Okayama H. High efficiency transformation of Escherichia coli with plasmids. Gene. 1990;96(1):23–28.

[33] Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. nature. 1970;227(5259):680.

[34] Lawrence AM, Besir H. Staining of proteins in gels with Coomassie G-250 without organic solvent and acetic acid. JoVE (Journal of Visualized Experiments). 2009;(30):e1350.

[35] Kabsch W. Integration, scaling, space-group assignment and post-refinement. Acta Crystallographica Section D: Biological Crystallography. 2010;66(2):133–144.

[36] Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, et al. Overview of the CCP4 suite and current developments. Acta Crystallographica Section D: Biological Crystallography. 2011;67(4):235–242.

[37] Evans PR, Murshudov GN. How good are my data and what is the resolution? Acta Crystallographica Section D: Biological Crystallography. 2013;69(7):1204–1214.

[38] McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ. Phaser crystallographic software. Journal of applied crystallography. 2007;40(4):658–674.

[39] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D: Biological Crystallography. 2004;60(12):2126–2132.

[40] Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, et al. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallographica Section D: Biological Crystallography. 2011;67(4):355–367.

[41] Moriarty NW, Grosse-Kunstleve RW, Adams PD. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. Acta Crystallographica Section D: Biological Crystallography. 2009;65(10):1074–1080.

[42] Mitchell LH, Boriack-Sjodin PA, Smith S, Thomenius M, Rioux N, Munchhof M, et al. Novel oxindole sulfonamides and sulfamides: EPZ031686, the first orally bioavailable small molecule SMYD3 inhibitor. ACS medicinal chemistry letters. 2015;7(2):134–138.

[43] Frisk D. A Chalmers University of Technology Master's thesis template for LaTeX. Unpublished;.

A

Used buffers and stock solutions

A.1 Chemicals & molecular biology reagents

All chemicals used in this study were purchased from Sigma-Aldrich (USA), Alfa Aesar (USA), Acros Organics (Belgium), VWR Sverige (Sweden) or Merck KGaA (Germany). All the chemicals were of at least ACS grade or higher, and were used without any additional purification.

Enzymes used for molecular biology applications were from ThermoFisher Scientific (Netherlands & Lithuania). Unless stated otherwise, enzymatic reactions were carried out according to the manufacturer recommendations.

Peptides 1-4 were a kind gift from Prof. Jan Kihlberg (Uppsala University, Sweden). The peptides were synthesized on a solid phase, employing Fmoc-based chemistry, purified with HPLC and analyzed with MALDI-TOF MS.

Compounds **1**, **2** and **3** originated from SciLifeLab DDD fragments library.

A.2 Stock solutions

Primer solutions
Ordered primers (ThermoFisher Scientific, Sweden) were diluted to 100 µM with TE buffer (10 mM TRIS, 10 mM EDTA, *p*H 8.0, filtered) to create a master stock. Said stock was later diluted to 25 micromolar concentrations for a *working solution*.
Chloramfenicol 1000× stock
For a 1000× Chloramfenicol antibiotic stock solution, 350 mg-s of Chloramfenicol was measured, which was dissolved in 7 mL-s of 98% ethanol, and the volume was adjusted to 10 mL-s.
Ampicillin 1000× stock
For a 1000× Ampicillin antibiotic stock solution, 1 g of Ampicillin was measured, which was dissolved in 4 mL-s of 98 % ethanol, and the volume was adjusted to 10 mL-s.
Phosphate buffered saline (PBS) 5× stock
For a 5× PBS stock solution, 10 PBS tablets from Fisher Scientific (BP2944-100) were dissolved in 400 mL-s of MilliQ and the solution filtered. A 1x PBS has 10 mM phosphate,2.7 mM KCl, 137 mM NaCl, *p*H 7.4.

B

Composition of broths

B.1 Lysogeny broth

For 1 liter of Lysogeny broth, 10 grams of tryptone, 5 grams of yeast extract, and 10 grams of NaCl was measured, and 950 ml of distilled water was added. The *p*H was adjusted to 7.5, and the total volume was adjusted to 1 liter. The broth was autoclaved at 121 °C. Lysogeny Broth always included 100 µg mL^{-1} Ampicillin and 35 µg mL^{-1} Chloramfenicol, unless stated otherwise.

B.2 M9 mineral medium

The end concentration of M9 mineral medium was 67.5 mM Na_2HPO_4, 44 mM KH_2PO_4, 17.1 mM NaCl, 0.1 % NH_4Cl, 0.4 % D-glucose, 4 mM $MgSO_4$, 0.2 mM $CaCl_2$, 1x BME vitamin solution (Sigma-Aldrich),100 µg mL^{-1} Ampicillin, 1x Studier's Salts.

Studier's salts were prepared as a 1000x stock, composing of 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $Cocl_2$, 2 mM $Cucl_2$, 2 mM $NiSo_4$, 2 mM Na_2MoO_4, 2 mM Na_2SeO_3, 2 mM H_3BO_3, in 60 mM HCl.

B.3 Agar plates

For agar plates, 5 grams of tryptone, 2.5 grams of yeast extract, and 5 grams of NaCl was measured, and 400 ml of distilled water was added. The *p*H was adjusted to 7.5. 7.5 grams of agar was added to the lysogeny broth mixture, and the volume adjusted to 500 milliliters. It was then autoclaved at 121 °C. To make antibiotic agar plates, 125 µL-s of the 1000x stock solution was added to 125 mL-s of the agar plate mixture, and poured on to petri dishes.

C

Polymerase chain reactions

C.1 HSP90$_{626-732}$ PCR

The program used for amplification of HSP90$_{626-732}$ is shown in table C.1. The PCR mixture final composition was 6.35 pg µL^{-1} HSP90$_{626-732}$ DNA, 1 × *Phusion* buffer (with $MgSO_4$), 1 µM forward and reverse primer, 200 µM NTP, 0.025 Units of *Phusion* DNA polymerase. The total volume of the PCR reaction mixture was 50 µL-s. Steps 2-4 (denaturation, annealing, extention) were repeated 40 times.

Table C.1: Program used for HSP90$_{626-732}$ PCR amplification

	degrees (°C)	time (s)
Initiation	95 °C	120
Denaturation	95 °C	60
Annealing	50 °C	30
Extention	74 °C	60
Final extention	74 °C	300

C.2 HSP90$_{9-236}$ PCR

The program used for amplification of HSP90$_{9-236}$ is shown in table C.2. The PCR mixture final composition was 6.35 pg µL^{-1} HSP90$_{9-236}$ DNA, 1 × *Phusion* buffer (with $MgSO_4$), 1 µM forward and reverse primer, 200 µM NTP, 0.025 Units of *Phusion* DNA polymerase. The total volume of the PCR reaction mixture was 50 µL-s. Steps 2-4 (denaturation, annealing, extention) were repeated 30 times.

Table C.2: Program used for HSP90$_{9-236}$ PCR amplification

	degrees (°C)	time (s)
Initiation	95 °C	120
Denaturation	95 °C	60
Annealing	56 °C	30
Extention	74 °C	60
Final extention	74 °C	300

C.3 $HSP90_{293-732}$ PCR

The programs used for amplification of $HSP90_{293-732}$ are shown in tables C.3,C.4 and C.5. All PCR mixtures had a final volume of 50 µL-s, and were made of 1 ×*Phusion* buffer (with $MgSO_4$), 1 µM forward and reverse primer, 200 µM NTP, 0.025 Units of *Phusion* DNA polymerase, varying the template DNA and concentration, and the number of cycles.

The first PCR reaction (table C.3) had 6.5 pg µL^{-1} $HSP90_{1-732}$ DNA, and repeated steps 2 to 4 (denaturation, annealing and extention) for 40 cycles. The second PCR (table C.4) had 672 pg µL^{-1} $HSP90_{283-683}$ and 3780 pg µL^{-1} $HSP90_{675-732}$ PRC product, and did 10 cycles. The third PCR (table C.5) had 280 pg µL^{-1} $HSP90_{283-732}$ PCR-2 annealed product, and did 15 cycles.

Table C.3: $HSP90_{293-732}$ PCR 1

	degrees (°C)	time (s)
Initiation	98 °C	120
Denaturation	98 °C	45
Annealing	53 °C	30
Extention	72 °C	168
Final extention	74 °C	900

Table C.4: $HSP90_{293-732}$ PCR 2

	degrees (°C)	time (s)
Initiation	98 °C	30
Denaturation	98 °C	10
Annealing	62 °C	30
Extention	72 °C	90
Final extention	74 °C	600

Table C.5: $HSP90_{293-732}$ PCR 3

	degrees (°C)	time (s)
Initiation	98 °C	120
Denaturation	98 °C	60
Annealing	47 °C	42
Extention	72 °C	120
Final extention	74 °C	900

C.4 Colony PCR

The program used for Colony PCR is shown in table C.2. A colony from an Agar plate was smeared to the bottom of the PCR tube. The PCR mixture final compo-

sition was 1 ×*DreamTAQ* buffer (with $CaCl_2$), 1 µM T7 forward and reverse primer, 200 µM NTP, 0.03 Units of *DreamTAQ* DNA polymerase. The total volume of the PCR reaction mixture was 50 µL-s. Steps 2-4 (denaturation, annealing, extention) were repeated 25 times.

Table C.6: Program used for colony PCR amplification

	degrees (°C)	time (s)
Initiation	95 °C	300
Denaturation	95 °C	30
Annealing	42 °C	30
Extention	72 °C	60
Final extention	72 °C	300

D

Amino acid analysis of HSP90 CTD and GST-fused HSP90 CTD

Amino acid composition of HSP90 CTD

The amino acid sequence of HSP90 CTD is as follows:

```
        10         20         30         40         50         60
GYMAAKKHLE INPDHSIIET LRQKAEADKN DKSVKDLVIL LYETALLSSG FSLEDPQTHA

        70         80         90        100
NRIYRMIKLG LGIDEDDPTA DDTSAAVTEE MPPLEGDDDT SRMEEVD
```

with a theoretical isoelectric point of 4.34 (ExPASy, ProtParam tool). The sequence has 7 lysine residues.

Amino acid composition of GST-fused HSP90 CTD

The amino acid sequence of HSP90 CTD as a GST-fuse is as follows:

```
        10         20         30         40         50         60
MSPILGYWKI KGLVQPTRLL LEYLEEKYEE HLYERDEGDK WRNKKFELGL EFPNLPYYID

        70         80         90        100        110        120
GDVKLTQSMA IIRYIADKHN MLGGCPKERA EISMLEGAVL DIRYGVSRIA YSKDFETLKV

       130        140        150        160        170        180
DFLSKLPEML KMFEDRLCHK TYLNGDHVTH PDFMLYDALD VVLYMDPMCL DAFPKLVCFK

       190        200        210        220        230        240
KRIEAIPQID KYLKSSKYIA WPLQGWQATF GGGDHPPKSD LVPRGSGYMA AKKHLEINPD

       250        260        270        280        290        300
HSIIETLRQK AEADKNDKSV KDLVILLYET ALLSSGFSLE DPQTHANRIY RMIKLGLGID

       310        320        330
EDDPTADDTS AAVTEEMPPL EGDDDTSRME EVD
```

with a theoretical isoelectric point of 5.02 (ExPASy, ProtParam tool). The sequence has 28 lysine residues.

E

Construct maps

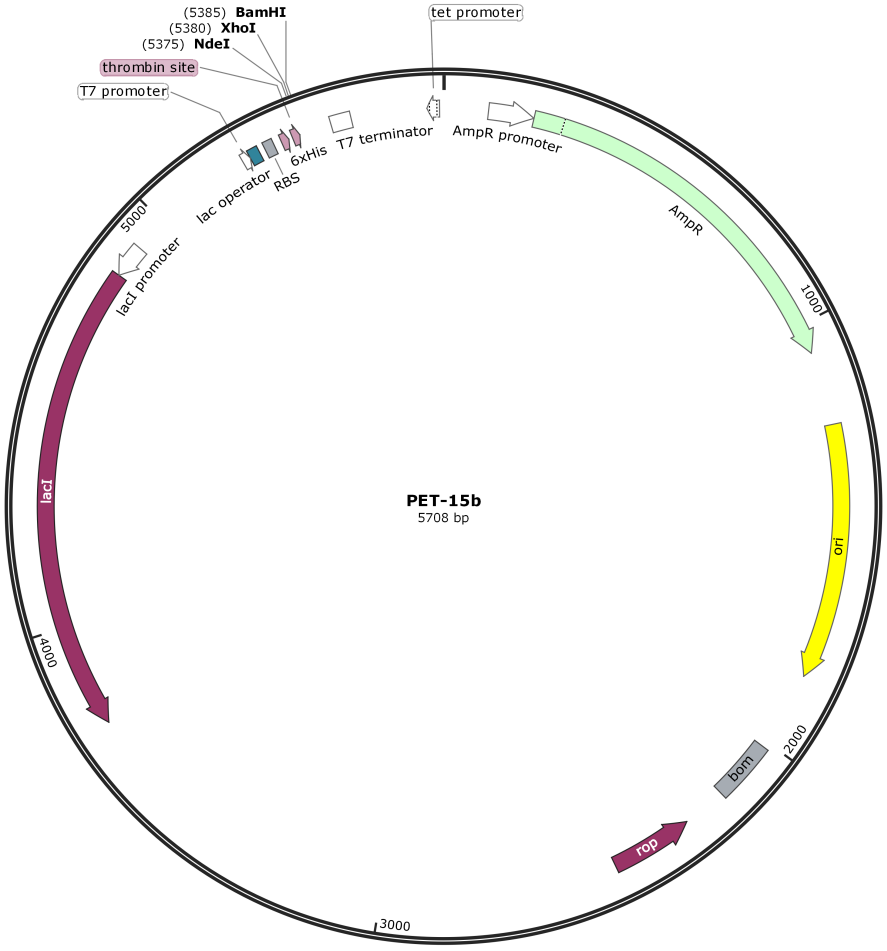

Figure E.1: A map of the pET-15b plasmid, with NdeI, XhoI and BamHI recognition sequences shown near the 6×His and thrombin cleavage site.

Figure E.2: Donor plasmid for $\text{HSP90}_{626-732}$. Primer annealing sites shown in purple for NdeI and BamHI.

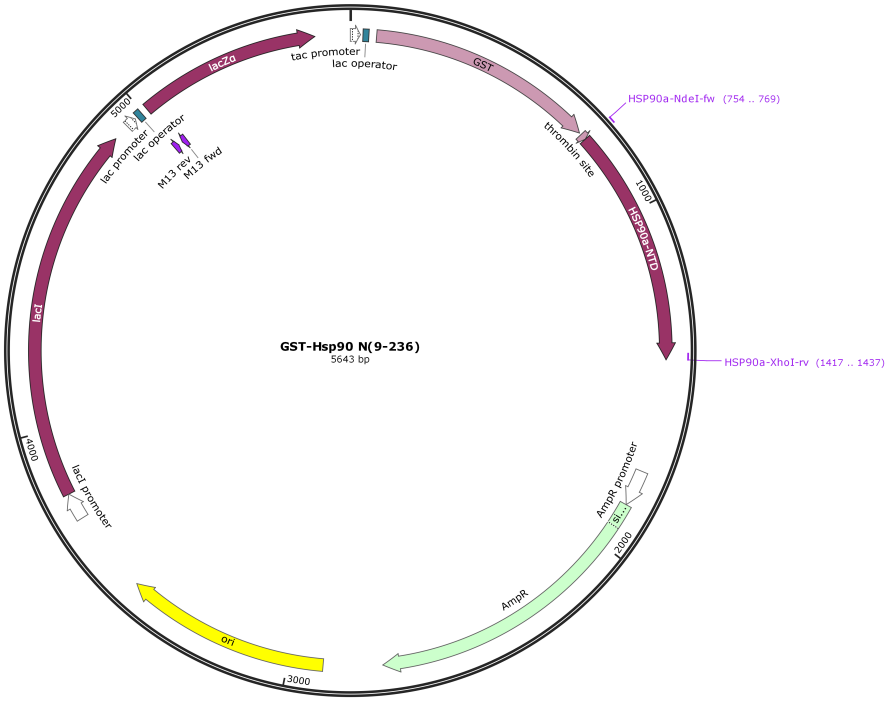

Figure E.3: Donor plasmid for HSP90_{9-236}. Primer annealing sites shown in purple for NdeI and XhoI.

Figure E.4: Donor map for $HSP90_{1-732}$. Primer annealing sites shown in purple for NdeI and BamHI, and for the primers for G2A mutation.