1-1-2018

Alternative Strategies To Inhibit Lysine Methyltransferases And Deubiquitinases In Human Cancers

Nicholas Spellmon
Wayne State University,

ALTERNATIVE STRATEGIES TO INHIBIT LYSINE METHYLTRANSFERASES AND DEUBIQUITINASES IN HUMAN CANCERS

by

NICHOLAS SPELLMON

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2018

MAJOR: BIOCHEMISTRY AND MOLECULAR BIOLOGY

Approved By:

_____

Advisor Date

_____

_____

_____

ACKNOWLEDGEMENTS

In the past six years, I have been fortunate to meet a handful of people who contributed to my graduate training and professional development. As I finalize my dissertation, I'd like to acknowledge those who made it possible for me to reach this point.

Joshua Holcomb. Josh was my first laboratory friend who taught me the basics in molecular cloning, protein purification, and X-ray crystallography. He and I worked diligently to together to solve problems and learn from each other's projects. Working with Josh has been a pleasure, and I look forward to his promising career endeavors.

Evan Cornett. Evan and I met at the ASBMB conference in 2017, and I was thrilled for the opportunity to work with Evan and his mentor, Dr. Scott Rothbart. He and I established an collaboration, and over time, we kept in touch to exchange ideas for our interrelated projects. I'm proud for the fruitful work we've accomplished, and I hope our collaboration continues beyond our current assignments.

April Wolak. I've known April since I entered the PhD program four years ago. She has helped me place orders for our lab and provided paperwork I needed throughout my graduate studies. I am thankful for April's remarkably steadfast and outstanding assistance.

Department Friends. More specifically, I'd like to thank my friends outside my lab: Kendall Muzzarelli, Tyler Peters and Chandni Patel. I'm much grateful for their help finding specific reagents, detailed protocols and proper instruments for some of my experiments. I couldn't have asked for better group of friends to work and laugh with.

Dissertation Committee. I'd like to extend my gratitude to my committee members: Drs. Brain Edwards, Ladislau Kovari and Young-Hoon Ahn. I sincerely appreciate our stimulating

discussions and your sagacious counsel. Your time and effort to aid in my training means more than I can express.

Dr. Zhe Yang. I remember four years ago walking to his office asking if he was accepting PhD students. At the time, I fervently wanted to learn X-ray crystallography, and I was thrilled he warmly accepted to train me in his lab. Since then, I've learned and accomplished more than I imagined. Thank you to Dr. Yang for setting the bar high enough for me to become nothing less than the best.

Family. To say, I couldn't complete my doctorate studies without my family's love and support, would be an understatement. My parents have ensured that I reached my fullest potential no matter the cost. They always said that you can't put a price on education; it's something that can't be taken away from you. Thank you to my parents, my siblings and my extended family for the sacrifices you made for me to reach this point.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PART I: STRUCTURAL INSIGHTS INTO THE BILOBAL REGULATION OF SMYD PROTEINS

GENERAL INTRODUCTION

Post-Translational Modifications

Post-translational modifications (PTM) are covalent modifications made to a protein after its translation by the ribosome. PTMs serve as central components in cell signaling, epigenetic regulation, proteasomal degradation, and protein structure. These small modifications may seem minor in comparison to a full size protein, but PTMs operate as vital mechanisms for organisms to expand and dynamically control a protein's function. Indeed, impaired post-translational machinery are linked to several human diseases. Understanding how PTMs are directed in human health are of strong interest among the research community.

PTMs can be divided into two categories: reversible and irreversible. Reversible modifications are usually small functional groups, such as phosphate, methyl and acetyl groups, attached to specific amino acid residues of a given protein, and these small chemical attachments can influence the protein's fold and function. Irreversible post-translational modifications include peptide bond cleavage and cofactor attachments. These type of modifications are also necessary for the protein maturation and function. While nature has many types of PTMs, my doctoral studies focus on the structure and function lysine methyltransferases, ubiquitin-specific proteases and disulfide-linked PDZ-domains.

SMYD proteins

SMYD (SET and MYND domain-containing proteins) proteins are histone and non-histone lysine methyltransferases which contribute various cellular roles including cell cycle control, chromatin remodeling, signal transduction and transcription [1-5]. Early reports demonstrated high SMYD expression in cardiac and skeletal muscles and suggested SMYD functioned in myogenesis

and cardiomyocyte differentiation [1, 6-12]; however, later reports found SMYD proteins to also be involved in immunity and inflammation [13-15]. SMYD proteins are of therapeutic interest due to a growing list of cancers associated with their overexpression and methylation activity, particularly SMYD2 and SMYD3 [4, 5, 16-22]. The lysine methyltransferase of interest for my project is SMYD2.

SMYD2 is overexpressed in many different cancers including but not limited to esophageal, gastric, breast, leukemia and pancreatic cancers [16-19, 23]. The growing number of reported cancers associated with SMYD2 expression and function highlight the importance of studying the molecular mechanism of SMYD2. Crystal structures of SMYD2 have shed some light into the mechanism of substrate and cofactor binding [24]; however, the mechanism at which substrates are channeled through the substrate channel remains unclear. Chapters 1, 2, 3 and 4 will explore the structure and function of SMYD proteins and examine how the TPR-like domain influences the accessibility of the substrate channel in SMYD proteins.

Ubiquitin-specific proteases

Ubiquitin-specific proteases (USPs) are a large family of deubiquitinases consisting of 56 different members in humans. Each USP is responsible for recognizing a specific protein substrate and ubiquitin branch type [25]. After deubiquitination, the protein of interest is recycled and protected against proteasomal degradation. Since proteasomal inhibitors were approved by the FDA for treating hematological malignancies, interest surrounding USPs as viable targets for treating cancers, autoimmune and neurodegenerative diseases is growing [26]. Since USPs are diverse class of enzymes upstream in the ubiquitin proteasome pathway, USPs serve as more selective targets potentially offering fewer side effects [27].

USP10 is one member of the USP family found to maintain p53 homeostasis, DNA repair and androgen receptor-dependent transcription. While USP10 appears to function multiple pathways, USP10's role in cancer is well established. In mutant p53 renal cancers, USP10 stabilizes oncogenic p53 promoting cellular growth [28]. Another study found USP10 creates a positive feedback loop in promoting androgen receptor transcription and G3BP2 expression resulting in repressed p53 function in prostate cancers [29]. More recently, USP10 was found to specifically deubiquitinate oncogenic FLT3-ITD, an oncogene commonly found in acute myeloid leukemias [30]. Altogether, USP10 appears to be an attractive therapeutic target for specific cancer treatments. Unfortunately, biochemical and structural studies of USP10 are currently lacking. In Chapter 5, I designed a soluble protein construct that incorporates the catalytic domain of USP10 (USP10-CA). This protein construct was used to study the binding and biochemistry of USP10-CA on reported inhibitors. Ultimately, our goal was to study the structure and function of USP10-CA in order to profile the mechanism of deubiquitination and inhibition.

PDZ-domains

PDZ domains are small, protein-protein interacting modules that anchor protein complex assembly. Humans possess roughly 260 proteins which possess either one or more PDZ domains into the open reading frame [31]. Each PDZ domain is capable of recognizing four amino acid motif, found at the carboxyl-end of target proteins. Disruption of PDZ-ligand interactions have been well-documented in impair cellular signaling [32]. For example, C-X-C chemokine receptor type 2 (CXCR2) is a membrane receptor capable of driving chemotaxis in neutrophils. Deleting the PDZ-binding motif in CXCR2 disrupted the PDZ-mediated complex with Na+/H+ exchanger regulatory factor-1 (NHERF1). Obstructing this complex attenuated CXCR2-dependent calcium mobilization and chemotaxis migration in murine neutrophils [33].

PDZ-domains are not necessarily noted for post-translational modifications; however, in chapter 6, our study was initially aimed to understand the binding promiscuity of the CXCR2 PDZ-interacting motif and PDZ-RhoGEF, but we identified a new disulfide bond link that appears to covalently scaffold two PDZ-RhoGEF molecules. Disulfide bonds are commonly used in nature to stabilize native and protein complex structures. In the case for PDZ-RhoGEF, disulfide bond formation creates an anti-parallel structure capable of recognizing two CXCR2 peptides. Interestingly, a separate study found disulfide bond formation of PDZ-RhoGEF and LARG, a conserved relative of PDZ-RhoGEF, were important for promoting ligand binding [34]. Here, we describe the structure and potential functional outcomes for this newly appreciated disulfide bond.

CHAPTER 1 STRUCTURE AND FUNCTION OF SET AND MYND DOMAIN-CONTAINING PROTEINS

*Published in Int J Mol Sci. 2015 Jan 8;16(1):1406-28. doi: 10.3390/ijms16011406. All authors agreed with including their work in this dissertation.

Introduction

SET and MYND domain-containing proteins (SMYD) are a special class of protein lysine methyltransferases involved in methylation of histones and non-histone targets [1-5]. To date, there are five members from the SMYD family, SMYD1–5 (Figure 1A) [6]. Each member contains a conserved SET (Suppressor of variegation, Enhancer of Zeste, Trithorax) domain that is "split" by a Myeloid-Nervy-DEAF1 (MYND) domain [35]. The SET domain is a conserved catalytic unit for lysine methylation found in nearly all histone methyltransferases (HMT) [36]. The MYND domain is a zinc finger motif that primarily functions as a protein–protein interaction module [37, 38]. Another feature is the C-terminal domain (CTD) found in SMYD1–4 but absent in SMYD5. Despite the lack of sequence similarity, this domain is structurally similar to tetratricopeptide repeats (TPR), which is a motif important for the binding of cochaperones with heat shock protein-90 (Hsp90) [24, 35, 39, 40].

SMYD proteins may regulate chromatin remodeling and gene accessibility by methylating histone targets and interacting with transcription mediators. SMYD1–3 methylate H3K4, which is a methylation site promoting active transcription [1, 4, 35, 41, 42]. However, SMYD does not have an effect on global H3K4 methylation but appears to impact selective promoter regions [43, 44]. SMYD1 binds directly to class I and class II histone deacetylases (HDAC) and represses transcription from an SV40-luciferase reporter [1]. SMYD2 was also found to dimethylate H3K36 in vitro and repress transcription through interaction with the Sin3A histone deacetylase complex [2]. However, it remains to be determined whether in vivo recruitment of Sin3A requires both

H3K36 methylation and the presence of SMYD2. SMYD3 plays an important role in transcriptional regulation as a member of an RNA polymerase complex [4]. SMYD3 interacts with RNA polymerase II and RNA helicase HELZ suggesting that it might regulate target gene expression by facilitating transcriptional elongation. In HEK293 cells, overexpression of SMYD3 was found to up-regulate a number of genes corresponding to oncogenes, homeobox genes, and genes of the cell cycle [4]. These genes are highly expressed in colorectal and hepatocellular carcinomas [45, 46]. SMYD4 was identified as a potential tumor suppressor involved in breast cancer [47]. Expression of SMYD4 partially inhibits the expression of PDGFα and the lack of SMYD4 promotes PDGFα production [47]. The Drosophila melanogaster homolog of SMYD4 was found to recruit the HDAC co-repressor complex and thereby aid in fly development [48]. Eri is another component of the HDAC co-repressor complex, which interacts with SMYD4 [48]. SMYD5 is known to associate with the NCoR co-repressor complex and regulate pro-inflammation genes through trimethylation of H4K20 [15]. In macrophages, the SMYD5–NCoR co-repressor complex was found to repress the expression of toll-like receptor 4 (TLR4) genes [15].

SMYD proteins methylate several non-histone targets. In the cell cycle, SMYD2 methylates p53 and retinoblastoma tumor suppressor (RB) [3, 49, 50]. p53 methylation by SMYD2 reduces the transactivation activity of p53 [3]. In esophageal squamous cell carcinoma (ESCC), p53 methylation and inactivation were associated with aberrant oncogenic expression of SMYD2 [16]. Additionally, SMYD2 has an anti-apoptotic effect when it methylates p53 in cardioblasts [51]. RB methylation at Lys860 is regulated during cell cycle progression and cellular differentiation [49, 50]. It has been shown that RB methylation binds to the transcriptional repressor L3MBTL1 causing repression of E2F target genes [49]. In response to DNA damage,

SMYD2 was also found to methylate PARP1 at lysine 528, and this methylation regulates the PARP1's poly(ADP-ribosyl)ation activity in HeLa cells [52]. In intracellular signaling, SMYD3 targets two important kinases for methylation: MAP3K2 and vascular endothelial growth factor receptor-1 (VEGFR1). Methylation of MAP3K2 prevents PP2A phosphatase, a key negative regulator of the MAP kinase pathway, from binding to MAP3K2 [5]. Methylated MAP3K2 links SMYD3 to Ras-driven cancer promoting cell proliferation and tumorigenesis [5]. VEGFR1 methylation by SMYD3 augments VEGRF1 kinase activity, which is thought to enhance carcinogenesis [53]. Since SMYD3 is primarily found in the cytoplasm during G0–G1 arrest, it is thought that SMYD3 enhances VEGFR1 signaling when cells are at the resting state [53].

Current data have shown that SMYD proteins methylate a variety of histone and non-histone targets which contribute to their various roles in cell regulation including chromatin remodeling, transcription, signal transduction, and cell cycle control. In order to better understand how SMYD proteins interact with such an extensive yet specific range of targets, structural examination of the SMYD family has provided significant insight to the diversity of SMYD binding and function. This review will provide a thorough description of SMYD structure and function and serve to inform rational drug design process targeting this cancer-related protein family.

SMYD Structure and Function

2.1. Overall SMYD Structure

Crystal structures of SMYD1, SMYD2, and SMYD3 with cofactors are currently available[24, 35, 39, 54-56]. Additionally, SMYD2 structures were solved with the estrogen receptor α (ERα) and p53 peptides enabling us to investigate the different interactions made between the two different substrates [54, 57, 58]. In all of the available SMYD structures, SMYD

proteins share a homologous bilobal structure separated by a non-conserved primary sequence of variable length (Figure 1B). The N-terminal lobe is divided into four domains: SET, MYND, SET-I, and post-SET. The catalytic SET domain is located in the middle of the N-terminal lobe in proximity to the C-terminal lobe. The C-terminal lobe is organized into helices that were found to be orientated in open or closed conformations [24, 35, 39]. SMYD1 has the most open structure and SMYD3 has the most closed one. SMYD2 is a conformational intermediate between SMYD1 and SMYD3 (Figure 1C). The difference in the relative positions of the N- and C-terminal lobes creates different shapes for substrate binding. The structure of SMYD1 resembles an open-ended "wrench" with two lobes separated by gap. Unlike SMYD1, SMYD2 and SMYD3 form a clamshell like structure due to the absence of the C-terminal protruding helix.

Currently, there are no structural data for SMYD4 and SMYD5. While SMYD1–3 share well aligned domains, SMYD4 and SMYD5 are vastly different in their primary sequence (Figure 1A). SMYD4 contains an additional TPR domain before the N-terminal lobe, and the CTD is far extended. SMYD5 completely lacks the CTD, yet the molecular size of SMYD5 is close to SMYD2 and SMYD3. Functional implications of the differences in SMYD4 and SMYD5 are unknown, but structural data of SMYD4 and SMYD5 are of interest to address these questions.

2.2. SET, the Evolutionary Conserved Methyltransferase Domain

The SET domain is split by the MYND domain into two sections: the S-sequence and the core SET domain (Figure 1A). The S-sequence is a small region that may aid in cofactor binding or protein–protein interaction along with its adjacent domain, MYND [7, 35]. The topology of the catalytic SET domain is well conserved between SMYD1–3, which is essentially similar to other traditional SET proteins despite the split in the primary sequence by the MYND domain [35]. SET domain often co-exists with post-SET, SET-I, and pre-SET, and together they contribute to

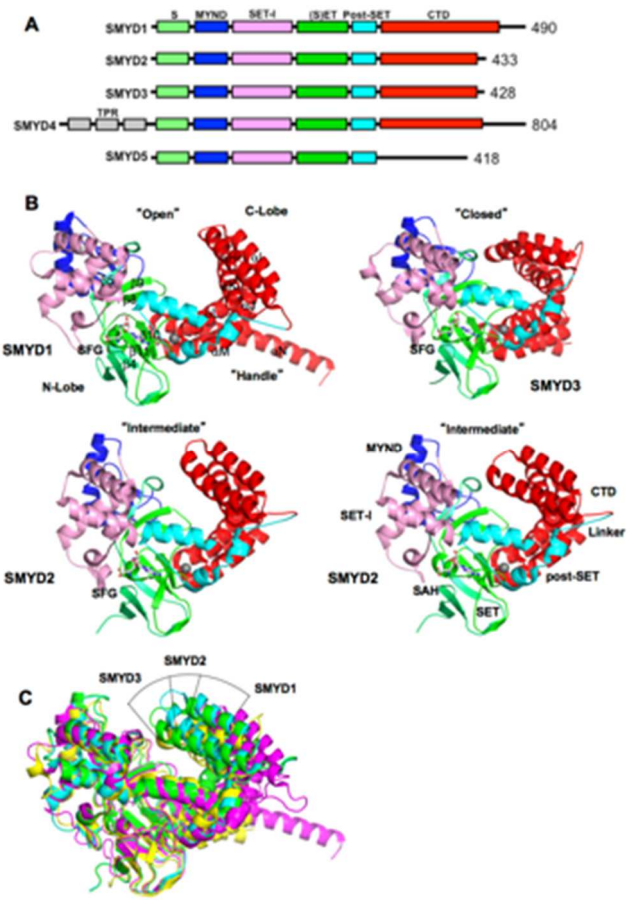

Figure 1. Overall structure of SET and MYND domain-containing proteins (SMYD)
(A) Domain diagram of SMYD protein family. S, S-sequence; MYND (Myeloid, Nervy, DEAF1); SET-I, insertion SET (Su(Var)3-9, Enhancer-of-zeste, Trithorax) domain; (S)ET, core SET domain; Post-SET, SET C-terminal flanking domain; CTD, C-terminal domain; (B) Ribbon diagram of SMYD1 (PDB code: 3N71), SMYD2 (PDB code: 3QWV and 3QWW), and SMYD3 (PDB code: 3PDN). The S-sequence, MYND, SET-I, core SET, post-SET, and CTD are depicted in light green, blue, pink, green, cyan, and red. Secondary structures, α-helices and β-strands are labeled and numbered according to their position in the sequence. Cofactors, sinefungin (SFG) and S-adenosyl-L-homocysteine (SAH), are depicted in balls and sticks; and (C) Structural superposition of SMYD proteins: SMYD1 (magenta), SMYD2 (SFG, cyan; SAH, green), and SMYD3 (yellow). The superposition is based on the N-terminal lobe.

cofactor binding, substrate binding, or the structural stability of the protein [59-61]. In SMYD proteins, the post-SET domain is made up of three α-helices bundled around a zinc atom coordinated by four cysteine residues. The SET-I domain along with MYND is an insertion region between the SET domain strands β5 and β8. Compared to other SET proteins like SET7 and Dim-

5, this insertion region is 6–10 times larger in size in SMYD proteins [35]. The pre-SET domain is often found in other SET containing proteins, but the pre-SET region is absent in SMYD1–3 proteins. Normally, this pre-SET region packs against an equivalent β-sheet made up of β4, β10, and β11, but in SMYD proteins, this β-sheet interacts with residues from the CTD in the αM–αN loop (Figure 1B). Interestingly, SMYD4 contains TPR repeats flanking the N-terminus of the N-terminal lobe (Figure 1A). This new region may introduce a pre-SET domain or add a third lobe to the overall structure.

2.3. MYND, the Zinc Finger Motif

MYND domain is a zinc finger motif identified to bind to proline-rich regions serving as a protein–protein interaction module [37, 38, 62]. In SMYD proteins, the MYND domain is part of the N-terminal lobe that interacts with the catalytic SET domain, but it does not participate in substrate or cofactor binding (Figure 1B) [35]. Consistently, deletion of the MYND domain does not affect the methyltransferase activity of SMYD2 in vitro, suggesting that the MYND domain is dispensable in methylation [41]. Despite the high sequence similarity to LIM (Lin11-Isl1-Mec3) domains, the MYND domain exhibits a different type of fold. The secondary structure of the MYND domain adopts a β–β–α topology, which is structurally similar to some PHD (Plant Homeo Domain) and RING motifs (Figure 2A). Although the MYND domains from AML1/ETO and SMYD proteins are only 30% identical in the primary sequence, the backbone and chelating zinc centers of the MYND are well superimposed (Figure 2A). The two structures share two anti-parallel β-strands (β6 and β7) and a small kinked α-helix (αA) that organize around two zinc atoms. Seven cysteine residues and one histidine are centered around the two zinc ions in a C4C2HC arrangement.

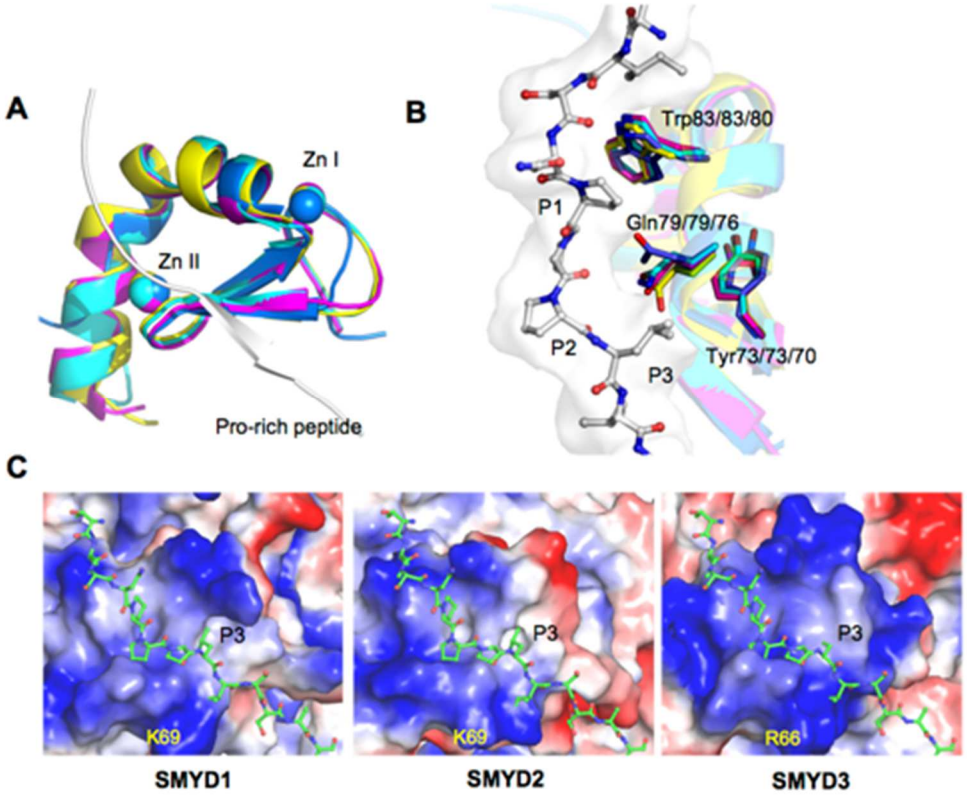

Figure 2. Structure of MYND domains
(A) Structural superposition of the MYND domains of SMYD and AML1/ETO (PDB code: 2ODD). MYND is represented by ribbon and colored in magenta (SMYD1), cyan (SMYD2), yellow (SMYD3), and blue (AML1/ETO). Proline-rich peptide bound to AML1/ETO is depicted by ribbon; (B) Superposition of the peptide binding pockets. Putative peptide interacting residues are colored according to the scheme in (A). The proline-rich peptide bound to AML1/ETO is depicted by balls-and-sticks; and (C) Surface representation of the MYND domains. Coloring is according to the electrostatic potential: red, white, and blue correspond to negative, neutral, and positive potential, respectively. The vacuum electrostatics/protein contact potential was generated by PyMOL. The proline-rich peptide, represented by balls-and-sticks, is modeled by superposition of the MYND domain of SMYD and AML1/ETO.

The MYND domain from AML1/ETO is known to bind to a PPPLI motif, and in SMYD1 and SMYD2, the MYND domain can interact with proteins with a similar proline-rich sequence [7, 37, 63]. SMYD1 binds to the muscle-specific transcription factor skNAC via a PPLIP motif [7]. In a previous yeast two-hybrid study, SMYD2 was found to interact with five different proteins possessing a PXLXP motif [63]. To date, a MYND-binding partner for SMYD3 has not been

identified, but the MYND structural similarities certainly suggest a proline-rich peptide-binding site for SMYD3. Three conserved and highly superimposed residues in SMYD3 (Trp80, Gln76, Tyr70) and in SMYD1–2 (Trp83, Gln79, Tyr73) may contribute to the binding of the proline-rich peptide (PXLXP) (Figure 2B). The tryptophan residue may pack against the first proline (P1), and the remaining glutamine and tyrosine residues may form a hydrophobic pocket for leucine (P3) to bind. The S-sequence in SMYD1 is also involved in binding to the skNAC proline-rich peptide [7], suggesting that regions other than the MYND domain may also play a role in determining binding specificity.

Electrostatic surface analysis shows the MYND domain is highly positively charged in SMYD1–3 (Figure 2C). This positively charged surface likely contributes to a protein–DNA interaction. SMYD binding to DNA was first identified in SMYD3, and binding to a specific DNA motif, 5'-CCCTCC-3', was found to regulate transcription of SMYD3 target genes such as Nkx2.8 [4]. Mutation of Arg66 within the MYND domain disrupted DNA binding of SMYD3 and abolished a DNA-induced increase in SMYD3 methyltransferase activity [64]. Interestingly, Arg66 appears to superimpose with similar conserved Lys69 in both SMYD1 and SMYD2, and the positively charged surface across the MYND domain is well observed across SMYD1–3 (Figure 2C). This certainly suggests that SMYD1 and SMYD2 are also involved in DNA binding. A recent study shows that SMYD2 binds to the promoter region of TACC2 and regulates TACC2 expression at a site different from the binding site for SMYD3 [41].

The exact nature of DNA binding in SMYD proteins is unknown, and how DNA binding affects the activity of SMYD proteins is yet to be identified. Structural studies of SMYD–DNA complexes are of interest to address these questions. Additionally, the overlap of the positively charged surface and proline-rich peptide-binding site in the MYND domain (Figure 2C) raises

intriguing questions regarding whether the peptide and DNA binding are mutually exclusive and what are the functional roles of such a scenario in the context of transcriptional regulation. In many cases, binding of SMYD3 to the promoter region of its target genes is associated with both H3K4 trimethylation and gene activation [44, 65-67]. Surprisingly, SMYD3 shows virtually no activity towards H3K4 in vitro compared to other targets such as H4K5 or MAP3K2 [5, 68]. This inconsistency suggests that DNA binding may induce a conformational change in SMYD3 that may subsequently affect substrate binding and specificity. Such a model remains to be determined, but the ability of SMYD2 to undergo a conformational change that alters the shape of the substrate-binding site provides a rationalization for this possibility [24].

2.4. Cofactor Binding Pocket

The SET-I, SET, and post-SET domains create a deep surface pocket allowing the L-shaped cofactors, S-adenosyl-L-homocysteine (SAH) and sinefungin (SFG) to bind (Figure 3A). Several cofactor and pocket interactions are shared among the SMYD family (Figure 3B). The adenine moiety of SAH or SFG is sandwiched between a conserved benzyl phenylalanine and aliphatic lysine or arginine side chain. The purine atoms N6 and N7 form a hydrogen bond to the carboxyl and amide groups of a conserved histidine residue, but the ribose hydroxyl groups form hydrogen bonds with somewhat similar neighboring residues among the SMYD family. At the positively charged amino group, a similar triangular array of hydrogen bonds is formed with the carbonyl oxygens from arginine and lysine (asparagine in SMYD3) and the amide Oδ from a separate asparagine. In the middle of cofactors, two backbone carbonyls and the side chain oxygens from conserved tyrosine and asparagine surround the Sδ atom in SAH or the C–NH2 amine group in sinefungin. The C–NH2 amine group of sinefungin corresponds to the S–CH3 sulfonium group in S-adenosyl-L-methionine (SAM). Some of the aforementioned surrounding oxygens are

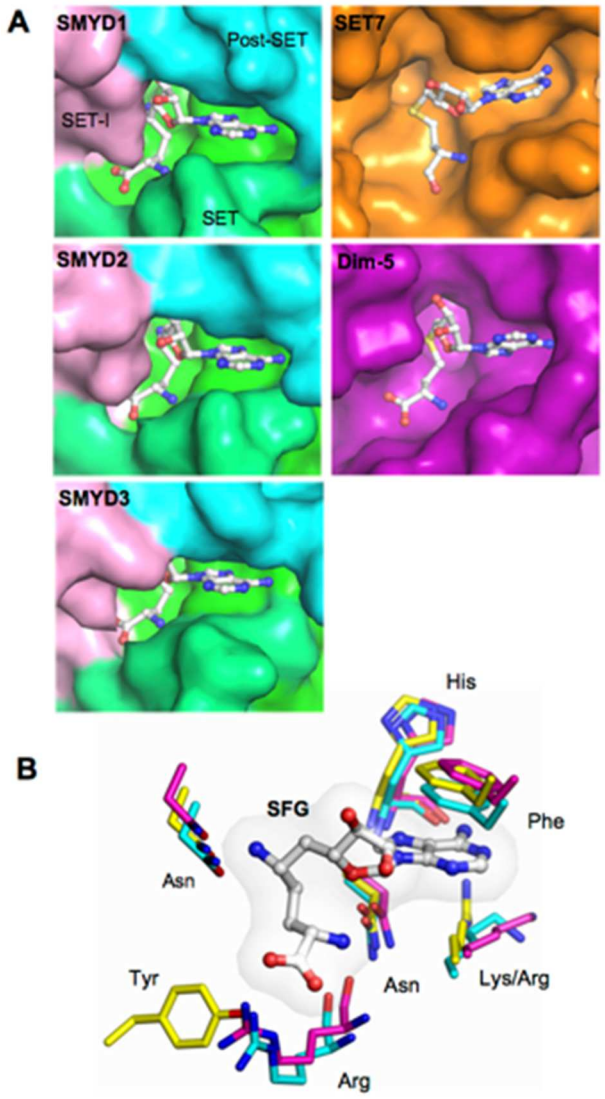

Figure 3. Cofactor binding pocket
(A) Surface representation of the cofactor-binding pocket of SMYD1–3, SET7 (PDB code: 1O9S), and Dim-5 (PDB code: 1PEG). The surface of SMYD proteins is colored according to domains. Bound SAH or SFG is depicted by sticks with the carbon atoms colored in white; (B) Superposition of the cofactor binding sites. SMYD residues are represented by sticks with the carbon atoms colored according to the scheme in Figure 1C. Cofactor is depicted by balls-and-sticks overlaid with translucent molecular surface.

responsible for hydrogen bonding with the C–NH2 amine group in sinefungin, and the same interaction is thought to contribute to destabilizing the active methyl group during enzymatic methylation [35, 58].

Large differences were observed at the carboxylate moiety of cofactors. In SMYD1–2, the carboxylate moiety is stabilized by a salt-bridge interaction with an arginine guanido group, but in SMYD3, this electrostatic interaction is substituted by a hydrogen bond to a tyrosine residue from a non-equivalent location (Figure 3B). This hydrogen bonding in SMYD3 represents an unusual variation, as the replaced electrostatic interaction is present in most SET containing proteins [24, 35, 69, 70]. As a result, the cofactor binding sites of SMYD1 and SMYD2 are more similar to one another than they are to the SMYD3 structure. However, all SMYD proteins have a nearly buried cofactor-binding site compared to SET7 and Dim-5. The bound cofactors share similar interactions in these proteins, but the large SET-I domain of SMYD proteins creates a nearly buried cofactor conformation (Figure 3A). This buried cofactor conformation, however, does not affect the enzymatic activity of SMYD1. Mutation of the SET-I residues responsible for the buried conformation only had modest effects on H3K4 methylation activity [35]. However, mutation of the S-sequence residues responsible for the adenine moiety binding completely abolished the enzymatic activity [35]. This suggests that the split S-sequence is an integral part of the SET domain contributing to cofactor binding.

2.5. Substrate Peptide Binding Site

To date, there are two SMYD–substrate complex structures available, SMYD2–ERα and SMYD2–p53 [54, 57, 58]. These structures have provided significant insight into the substrate binding and broad substrate specificity of the SMYD family. The ERα and p53 peptides bind to SMYD2 in a U-shape conformation (Figure 4A). The peptides are clamped between the N- and C-terminal lobes with the target lysine inserted into the lysine access channel. Residues contributing to peptide binding mainly come from the β8–β9 hairpin and a loop preceding the post-SET domain. The structure and residues at the turn of the β-hairpin vary significantly among the SMYD family,

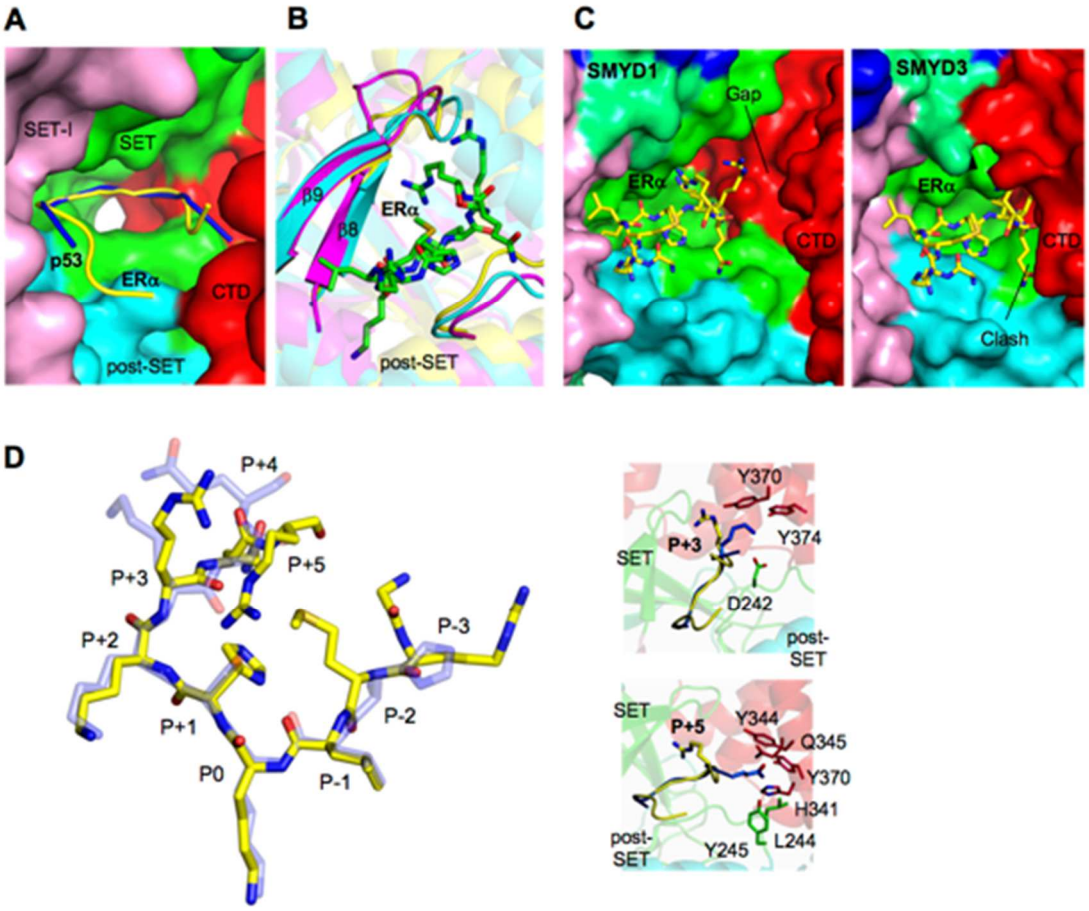

Figure 4. Substrate binding site

(A) Surface representation of SMYD2 substrate binding site. The surface is colored according to domains. ERα and p53 peptides are depicted by ribbons and colored in yellow and blue respectively; (B) Superposition of the substrate binding clefts. SMYD residues are represented by ribbons and colored according to the scheme in Figure 1C. ERα peptide is shown in balls-and-sticks colored in green; (C) Surface representation of the substrate-binding site of SMYD1 and SMYD3. The ERα peptide, represented by sticks, is modeled by superposition with the N-terminal lobe of SMYD2; and (D) Superposition of the SMYD2-bound ERα (yellow; PDB code: 4O6F) and p53 peptides (light blue; PDB code: 3TG5). Position 0 refers to the target lysine. Detailed structural and binding differences at position +3 and +5 are shown in callout boxes. Peptide-interacting SMYD2 residues are colored according to domains.

but in SMYD2 this region is important for substrate recognition (Figure 4B). This suggests that the structural features of the β8–β9 hairpin may contribute to the different substrate specificity of SMYD proteins. Additionally, the C-terminal domain of SMYD2 is directly involved in ERα and p53 binding, but the different CTD conformation in SMYD1 and SMYD3 implies a substantially

different substrate-binding mode in order to adapt to different substrate binding pockets (Figure 4C). The CTD conformation in SMYD proteins correlates with the size of the substrate-binding pocket, which may thereby provide another level of substrate specificity. Interestingly, SMYD3 appears to prefer smaller sized substrates and its methyltransferase activity is significantly higher for H4K5 and MAP3K2 than H4K5 [5, 68]. This is likely due to the small glycine residues neighboring the target lysine in H4K5 and MAP3K2. The smallest and most flexible amino acid, glycine, may facilitate substrate access to the closed active site of SMYD3 (Figure 4C).

Comparison of the SMYD2–ERα and SMYD2–p53 structures has provided insight into the broad substrate specificity of SMYD2 [57]. SMYD2 is able to methylate several targets and the structural basis of this broad substrate specificity lies in the presence of multiple substrate-binding sites in SMYD2 structure [54, 57, 58]. The structural comparison shows that the liganded SMYD2 structures are well superimposed with an RMSD value of 0.6 Å out of 430 Cα atoms [57]. The ERα and p53 peptides have a similar U-shape conformation and are well superimposed at positions −1, 0, +1, and +2 (position 0 referring to the target lysine) (Figure 4D). Large deviations are found in the peptides extending out of the U-base. For example, different interactions are seen at positions +3 and +5 between the ERα and p53 peptides (Figure, 4D). Arg+3 and Arg+5 in the ERα peptide binds to the β8–β9 region in the SET domain. In the p53 peptide, Lys+3 interacts with Tyr370, Tyr374, and Asp242, and the Gln+5 chain is inserted into a pocket formed by His341, Tyr344, Gln345, Tyr370, Leu244, and Tyr245. The two different peptide conformations and the broad substrate binding space inside SMYD2 present plausible explanations for the multiple accommodations for SMYD2 and substrate binding. Structural study of additional SMYD2–substrate complexes may be necessary to corroborate this model. Because of the sequence diversity

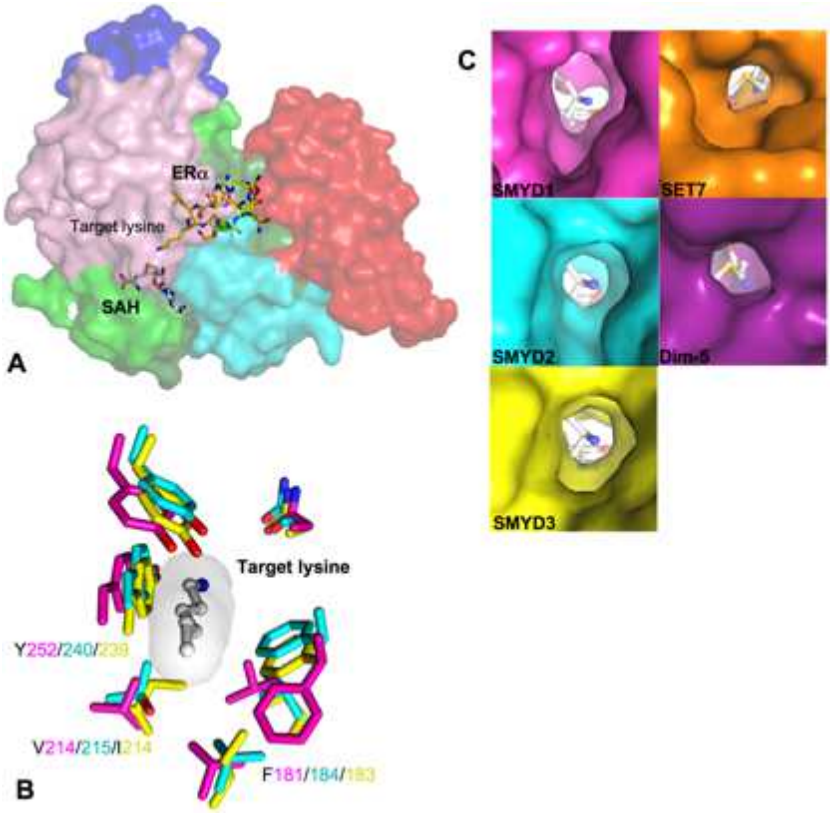

Figure 5. Target lysine access channel
(A) Surface representation of overall SMYD2–ERα structure. ERα peptide, SAH, and target lysine are indicated; (B) Superposition of the lysine access channels. SMYD residues are represented by sticks with the carbon atoms colored according to the scheme in Figure 1C. Target lysine is colored in white; and (C) Surface representation of the lysine access channel of SMYD1–3, SET7, and Dim-5. SAH or SFG is depicted by sticks with the carbon atoms colored in white.

of SMYD2 substrates [57], the new SMYD2–substrate complex structures could potentially lead to the identification of novel substrate binding modes.

2.6. Target Lysine Access Channel

In SET domain-containing enzymes, the lysine targeted for methylation fits into a hydrophobic pocket called the target lysine access channel (Figure 5A). The lysine access channel in SMYD proteins has a well-superimposed backbone and consists of a hydrophobic core surrounding the hydrophobic portion of the targeted lysine (Figure 5B). A highly conserved tyrosine residue preceding the post-SET domain is found in all SET classes including SMYD1

(Tyr252), SMYD2 (Tyr240), and SMYD3 (Tyr239). The tyrosine side chain appears to orient the target lysine into the channel, and substitution of this residue to phenylalanine completely abolished the enzymatic activity of SMYD2 and SMYD3 [2, 55, 71]. Seven residues are important for creating the lysine crevice, and of the seven, only three are aromatic residues. Two tyrosines and one phenylalanine are well conserved among SMYD1–3 with the exception of the orientation of phenylalanine in SMYD1. The aromatic ring is rotated about 110° around the Cα–Cβ bond axis which is pointed away from the lysine access channel; therefore, SMYD2 and SMYD3 have a well-defined and tighter channel than SMYD1 (Figure 5C). The other three small and non-aromatic residues are fairly conserved and responsible for creating a more open channel in comparison to SET7 and Dim-5 [35].

A more open channel may accommodate a larger substrate or is susceptible to ligand-induced conformational changes. It may also contribute to the broad and weak substrate binding especially for H3K4 methylation. This weak binding appears to be largely contributed to the substitution of large aromatic residues to smaller hydrophobic side chains such as valine, leucine, and isoleucine [35]. For example, substitution of Val214 to tyrosine in SMYD1 created a tighter access pocket, and thereby the binding of SMYD1 and H3 peptide significantly increased presumably due to a more compact channel and the additional hydrogen bond between the tyrosine hydroxyl group and an ε-amino from the neighboring lysine side chain [35]. This gain-of-function mutation indicates that a well-defined channel is important for securing the target lysine in the active site. A ligand-induced conformational change in the lysine access channel may be required for SMYD1 to efficiently methylate a substrate. In order to address these questions, it is of interest to determine a SMYD1 structure in complex with a substrate.

2.7. TPR-Like C-Terminal Domain

The CTD domain of SMYD proteins contains a series of antiparallel α-helices that display a similar structure to TPR domains despite the lack of sequence identity (Figure 1B). TPR domains are important for binding of cochaperones to Hsp90. For example, the Hop1 TPR domain binds to the very C-terminal end of Hsp90 mediating Hsp90 chaperone activity (Figure 6A) [72, 73]. The CTD structure is also well conserved in the SMYD family, and the only chief difference is the extended and protruded αN helix that resembles the "handle" of the wrench-shaped SMYD1 (Figure 6A). This feature is unique to SMYD1, as SMYD2–3 with the shorter αN helix resemble the shape of a clam-like shell. The unique portion of the αN helix in SMYD1 is well conserved from fish to human (data not shown). This region contains a patch of hydrophobic residues that mediate the crystal packing in SMYD1 crystals (Figure 6B). Interestingly, the CTD structure of SMYD1 is similar to the TPR structure of FKBP52 (Figure 6C). In FKBP52, the TPR domain also has a protruding C-terminal helix that contains a putative binding site for calmodulin [74, 75]. The function of the unique SMYD1 C-terminal helical tail is unknown, but the conserved sequence and involvement in the crystal packing suggests that it may serve as a site for protein–protein interaction.

The structural orientation of the CTD varies significantly among the SMYD family (Figure 1C). SMYD1 has an open CTD conformation with the substrate-binding cleft completely exposed in the protein. In SMYD3, the CTD conformation is closed as the significant contact between the N- and C-terminal lobes creates a narrower opening to the substrate-binding pocket [39]. SMYD2 is like a conformational intermediate between SMYD1 and SMYD3. Additionally, the CTD domain of SMYD2 is flexible and can undergo a conformational change when different cofactors bind [24]. The CTD conformational change results in two SMYD2 structures with a slight

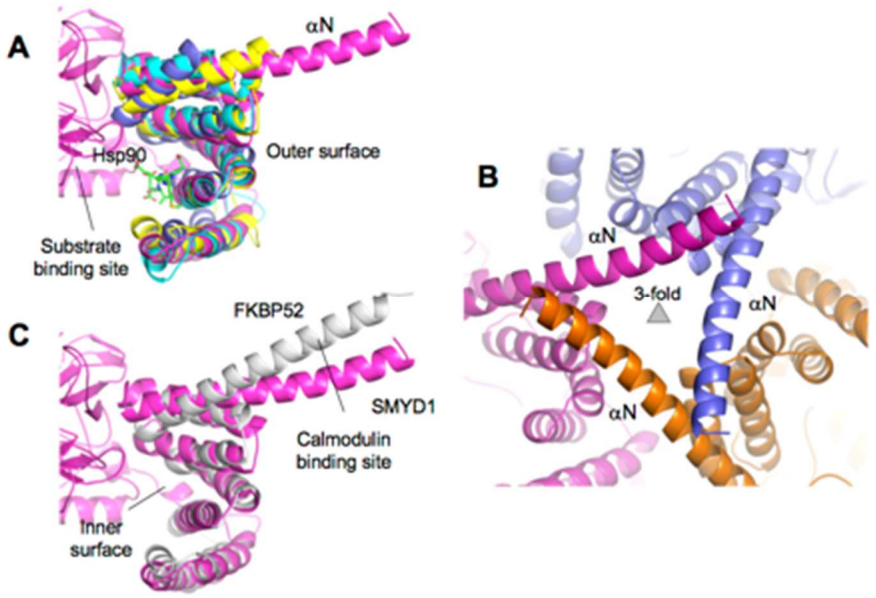

Figure 6. TPR-like C-terminal domain (CTD)
(A) Structural superposition of the CTD domains of SMYD and TPR domain of Hop1 (PDB code: 1ELR). SMYD proteins are represented by ribbons and colored according to the scheme in Figure 1C. The Hop1 TPR domain is shown in blue. Hsp90 peptide bound to Hop1 is depicted by balls-and-sticks; (B) Crystal lattice of SMYD1 shows the involvement of the protruding C-terminal α-helix in the crystal packing; and (C) Structural superposition of the CTD domain of SMYD1 and TPR domain of FKBP52 (PDB code: 1QZ2). SMYD1 is colored in purple and FKBP52 in white.

difference in the size and shape of the substrate-binding pocket [24]. Therefore, the orientation of the CTD domain may affect substrate specificity, and different pocket shapes and sizes may be involved in modulating the substrate preference of SMYD proteins. In addition, the CTD flexibility of SMYD2 suggests that SMYD2 may have the ability to adapt to substrates with different sizes, implying broad substrate specificity. Interestingly, SMYD2 so far has the broadest substrate specificity among SMYD proteins [57].

The CTD domain appears to play dual roles in substrate binding. Deletion of the CTD from SMYD1 results in increased binding and methylation on histone H3 suggesting that the CTD may have steric effects controlling substrate access to the active site [35]. In SMYD2, the CTD plays an important role in substrate recognition and binding, but the actual effect of the CTD appears to

be substrate-dependent. The CTD deletion has no effect on methylation of p53 peptide and histone H3 protein, but it results in a significant increase in H3K4 peptide methylation and significant decrease in p53 protein methylation [58]. Although these results seem paradoxical, the differential CTD effects have proved its complex roles in substrate recognition and binding. In addition, the TPR-like structure of the CTD suggests a potential role for the CTD in modulating protein–protein interaction. The predicted peptide-binding site in the CTD is located at the inner surface of the CTD in close proximity to the substrate-binding pocket (Figure 6A) [24, 57]. Binding to this location could therefore affect the substrate binding and enzymatic activity. Interestingly, the activity of SMYD proteins can be significantly increased in the presence of Hsp90 [4, 41, 42]. In the case of SMYD2, Hsp90 not only enhances the activity but also changes the substrate preference from H3K36 to H3K4 [41]. The questions remain whether Hsp90 regulates SMYD function via binding to the CTD domain and whether such binding has a reciprocal effect on Hsp90 chaperone activity. Nonetheless, the close proximity to the active site and flexibility and multi-orientations of the CTD suggest that it is a bona fide regulatory motif in SMYD proteins.

2.8. Additional Substrate Binding Site?

The polyethylene glycol (PEG)-binding site found in the SMYD2–ERα structure has suggested additional and extended substrate-binding pockets (Figure 7A) [57]. PEG binding was also found in other protein structures, and in most cases, PEG binding has important functional implications mimicking ligand binding in proteins [76-78]. In SMYD2, the PEG molecule primarily binds to the CTD domain with an omega-turn conformation with one end found near the surface groove shaped by αH, αI, and αJ and the other end extended between αK and αL helices (Figure 7B). The residues responsible for contributing PEG binding include Lys309, Tyr344,

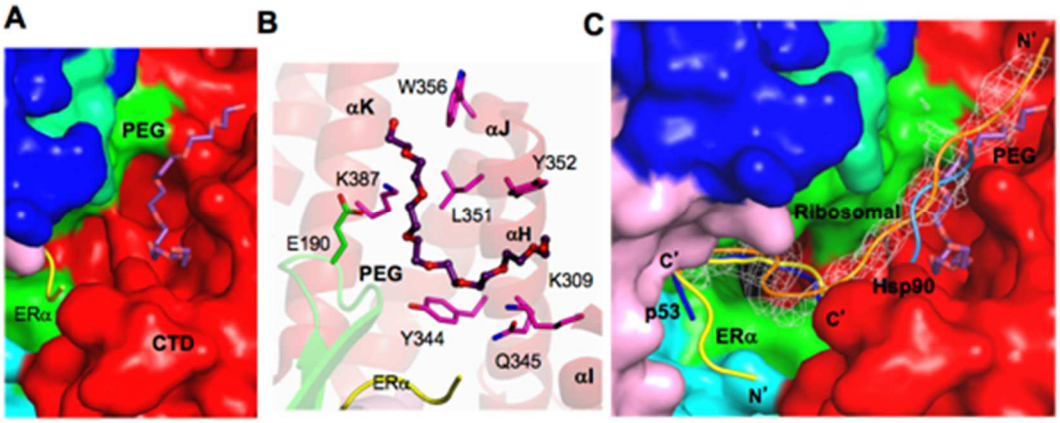

Figure 7. Additional substrate binding site
(A) Surface representation of the PEG binding site in SMYD2. PEG is depicted by sticks with the carbon atoms colored in purple. ERα peptide is displayed as ribbon and colored yellow; (B) Putative PEG interacting residues. SMYD2 residues are colored according to domains. ERα residues are shown in yellow. PEG is represented in the same way as in (A); and (C) Comparison of the binding sites of ERα (yellow), p53 (blue), PEG (purple), Hsp90 (light blue), and a ribosomal peptide (orange). The ribosomal peptide is overlaid with 2Fo−Fc omit map calculated at 2.8 Å and contoured at 1.5σ. The Hsp90 peptide is modeled by superposition of the SMYD2 CTD and Hop1 TPR.

Gln345, Gly348, Leu351, Tyr352, Trp356, and Lys387 from the CTD and Glu190 from the SET domain. The ERα peptide may also interact and stabilize PEG binding due to its close proximity to Arg+3. Note that all of the residues participating in PEG binding (except for Lys309) are not conserved in the SMYD family, which indicates a possible SMYD2-specific binding site.

The PEG binding site overlaps the predicted Hsp90 binding site (Figure 7C), suggesting that the PEG binding site might possess the peptide or substrate binding potential. This notion is supported by the SMYD2 structure binding to a ribosomal peptide (unpublished data). Instead of forming a U-shape formed by the ERα and p53 peptides, the ribosomal peptide density creates a partial U-shape at the ERα or p53-binding pocket, but the N-terminal end extends out through the αH/αI/αJ-binding groove of the PEG molecule (Figure 7C). This new mode of binding is vastly different to the ERα and p53 peptides and the orientations of the peptides are completely reversed.

This demonstrates the exceptional substrate adaptability of SMYD2, and the multiple binding sites and that some of these binding sites may be substrate-specific have provided explanations for its broad substrate specificity. It is of interest to reveal whether the αK/αL-binding groove of the PEG molecule in SMYD2 also indicate an additional substrate-binding pocket. Further investigation into different peptide binding and conformations are necessary to better characterize the diversity of SMYD structure and function.

Drug Design Perspective

SMYD proteins provide a new avenue for cancer and cardiovascular treatment. Overexpression of several of SMYD proteins is associated with nearly all cancer types[5, 16, 17, 50, 65]. Overexpression of SMYD1 represses transcription of genes necessary to produce ion channels in the heart, and repression of ion channel expression causes heart failure [79]. SMYD1 overexpression was also found in hypoplastic left heart syndrome (HLHS), a disorder characterized by the severely underdeveloped left ventricle [80]. SMYD2 is overexpressed in ESCC or p53-related cancers, and knockdown of SMYD2 inhibits tumor cell proliferation [16-18, 50]. SMYD3 is overexpressed in more than 14 types of cancers such as breast cancer, colon cancer, prostate cancer, lung cancer, and pancreatic cancer [4, 20, 21, 65, 81-88]. SMYD3 overexpression often correlates with poor prognosis, and knockdown of SMYD3 proved to inhibit tumor growth [4, 84-89]. Therefore, drug intervention of any of SMYD proteins may be beneficial to the fields of cardiovascular disease and cancer.

Efforts to create SMYD inhibitors are currently underway. AZ505 is a potent SMYD2 competitive inhibitor recently identified from a high throughput chemical screening [54]. In the crystal structure, AZ505 bound to the lysine access channel, and ITC analysis indicated inhibitor binding is primarily driven by hydrophobic interactions providing a low K_D ~0.5 μM (Figure

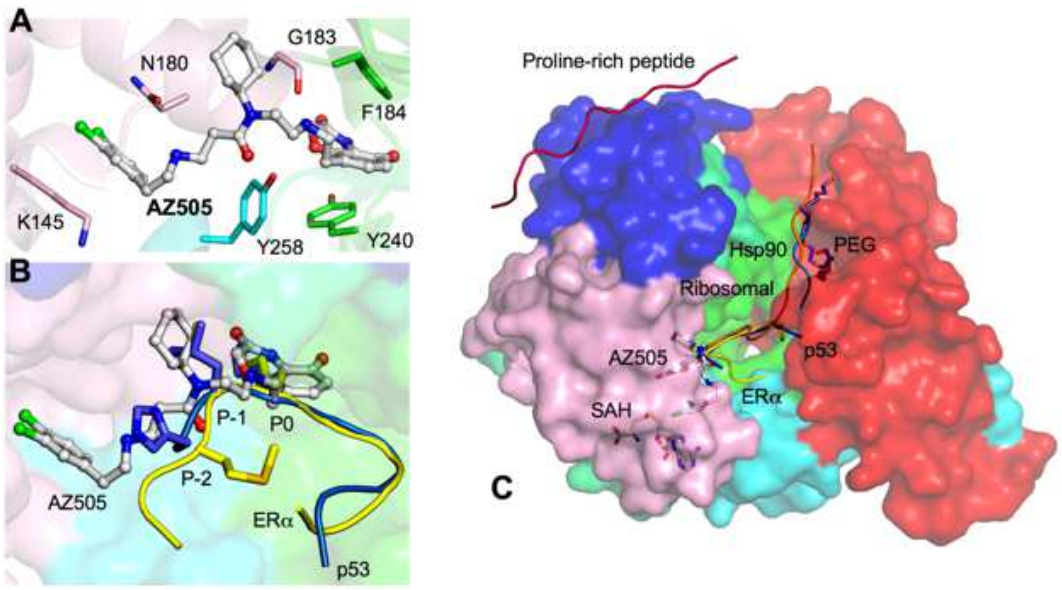

Figure 8. AZ505-bound SMYD2 structure
(A) Interactions between AZ505 and SMYD2. AZ505 is depicted by balls-and-sticks with the carbon atoms colored in white. SMYD2 residues are depicted by sticks colored according to domains; (B) Comparison of the binding sites of ERα, p53, and AZ505. The ERα and p53 peptides are depicted by sticks and colored in yellow and blue. AZ505 is represented in the same way as in (A); and (C) Surface representation of SMYD2–ERα structure illustrates potential drug targeting sites: ERα (yellow), p53 (blue), PEG (purple), Hsp90 (light blue), the ribosomal peptide (orange), and proline-rich peptide (hot pink). SAH and AZ505 are depicted by sticks.

8A)[54]. The three moieties of AZ505 have similar interactions found in the p53 and ERα peptides. The benzooxazinone group is packaged into the lysine channel where several hydrophobic and electrostatic interactions are made (Figure 8A). The cyclohexyl and dichlorophenethyl groups adopt the same −1 and −2 position, but they appear more compact to the surface than the p53 and ERα peptides (Figure 8B). In addition, the Gly183 carbonyl oxygen forms a similar hydrogen bond to the amide linker between the benzooxazinone and cyclohexyl groups. Therefore, the potency of AZ505 appears to be due to a complete blockage of the core region of the SMYD2 active site and preventing it from binding to the target lysine.

Therapeutic drug intervention of SMYD proteins may not be limited to inhibiting the lysine access channel. SMYD2 is necessary for methylating many targets and is involved in various

functionally independent cellular processes [57]. Complete knockdown of SMYD2 may not be a viable option since unselective SMYD2 inhibition may cause undesirable and perhaps lethal side effects. In order to selectively design a therapeutic drug to inhibit SMYD2 function in the context of cancer, one may consider targeting alternative binding sites for inhibition that will interfere with only a subset population of SMYD2. For example, designing a drug that will mimic the binding properties of the p53 peptide may provide specificity to oncogenic SMYD2 function in p53-related cancers. Targeting the binding properties of the ERα peptide may be beneficial to aggressive ERα-negative breast tumors by specifically restoring functional ERα expression [71]. The PEG or ribosomal binding site may also provide a genuine substrate-specific targeting option without the fear of interference with binding of the p53 and ERα peptides. Targeting MYND-mediated protein interactions may be another viable approach in cancer therapy as binding of the MYND to the proline-rich sequences of the tumor suppressor EBP41L3 links SMYD2 to meningiomas and lung cancer [41, 90, 91]. Finally, the CTD orientations related to the sizes and shapes of the SMYD substrate-binding pockets raise a possibility for selective drug design for different members of the SMYD family. Challenges remain because of the potential conformational flexibility of the CTD domain of SMYD proteins. Selective and potent drug design will require the consideration of conformational changes in SMYD proteins and understanding of the functional role of each conformational state. Further investigation into the structural differences between SMYD proteins will be necessary to distinguish specificity and efficacy into the drug design process.

Concluding Remarks

SMYD proteins are an exciting field of study as they are linked to many types of cancer-related pathways. Cardiac and skeletal muscle development and function also depend on SMYD proteins opening a possible avenue for cardiac-related treatment. The purpose of this review is to gather

current structural data to support the versatile roles of SMYD proteins. We provide a summary of the structures of the SMYD family focusing on their structural differences. The structures of the individual domains in SMYD1–3 are similar but the orientations of the CTD are substantially different resulting in open or closed conformations. Different CTD conformations suggest that SMYD proteins could undergo a conformational change that offers dynamics for regulation of substrate specificity [24, 35, 39]. SMYD2 conformations are sensitive to cofactor binding which alters the size of the substrate-binding pocket [24]. It is conceivable that the methyltransferase activity of SMYD proteins may be regulated by controlling lobe conformation and dynamics like some kinases [92, 93]. SMYD structures have a potential for efficacious drug intervention, but efforts to design a drug should not be limited to the target lysine access channel. Analysis of SMYD structures revealed many other binding sites with drug targeting potential, such as the broad substrate-binding pocket, PEG and ribosomal binding site, proline-rich peptide binding site, and a yet-unidentified DNA binding groove (Figure 8C). With the different binding sites and conformations, it is possible to effectively knockdown cancerous function of SMYD proteins such as SMYD2 and SMYD3 without disrupting the entire functional population of SMYD proteins. Additionally, analysis of current structures raised many new questions. The unique protruding C-terminal helix of SMYD1 may be involved in protein–protein interaction. The CTD orientation may determine substrate specificity. How DNA binding alters SMYD3 structure and function remains to be determined. Whether the possible binding of Hsp90 to the CTD provides a mechanism for SMYD activity enhancement and the potential role of this coplay in cancer and heart development are also unclear. In summary, SMYD proteins are of functional and therapeutic importance, and continued elucidation of their structural differences and substrate specificity will lead to additional functional implications.

CHAPTER 2 NEW OPEN CONFORMATION OF SMYD3 IMPLICATES CONFORMATIONAL SELECTION AND ALLOSTERY

*Published in AIMS Biophys. 2017;4(1):1-18. doi: 10.3934/biophy.2017.1.1. Epub 2016 Dec 20.
All authors agreed with including their work in this dissertation.

Introduction

SMYD3 belongs to a special class of protein lysine methyltransferases containing SET and MYND domains [94]. The SET is a catalytic motif responsible for lysine methylation. The MYND is a protein-protein interaction module involved in transcriptional cofactor recruitment. SMYD3 is overexpressed in more than 15 types of cancers such as breast cancer, colon cancer, prostate cancer, lung cancer and pancreatic cancer [4, 5, 94, 95]. Overexpression of SMYD3 often correlates with poor prognosis and its knockdown inhibits tumor growth [4, 5, 95]. Therefore, drug intervention of SMYD3 may be beneficial to the fields of cancer. SMYD3 is involved in tumorigenesis through methylation of histone and non- histone proteins. Histone methylation regulates gene expression and methylation of non- histone proteins can impact biochemical and cellular functions of the targets [5, 94, 95]. SMYD3 may directly or indirectly methylate histone H3K4, H4K20 and H4K5 [4, 68, 96]. Through these methylations, SMYD3 is involved in tumor cell viability, adhesion, migration and invasion. SMYD3 upregulates multiple cancer genes through H3K4 trimethylation. These include the telomerase reverse transcriptase (TERT), oncogenic c-Met, matrix metalloproteinase 9 (MMP-9), androgen receptor, myosin regulatory light chain 9 (MYL9) and retinoblastoma protein-interacting zinc finger gene 1 (RIZ1)[12, 44, 53, 65, 94, 97]. SMYD3 targets Cyclin D2 through H4K20 trimethylation and contributes to a more aggressive phenotype of prostate cancer [96]. H4K5 methylation by SMYD3 provides a potential new link between chromatin dynamics and neoplastic disease [68]. SMYD3 methylates three non-histone proteins: MAP3K2, vascular endothelial growth factor receptor-1 (VEGFR1) and AKT1.

Methylation of MAP3K2 prevents PP2A phosphatase, a key negative regulator of the MAP kinase pathway, from binding to MAP3K2 [5]. Methylated MAP3K2 links SMYD3 to Ras- driven cancer promoting cell proliferation and tumorigenesis [5]. VEGFR1 methylation by SMYD3 augments VEGRF1 kinase activity, which is thought to enhance carcinogenesis [53]. Methylation of AKT1 at lysine 14 is essential for AKT1 activation [98]. In addition, SMYD3 was found to promote formation of inducible regulatory T cells and may be involved in reducing autoimmunity [13, 99]. SMYD3 in vitro methyltransferase activity is not fully consistent with its cellular activity. SMYD3 only weakly methylates H3K4 in vitro but its cellular methyltransferase activity has been associated with H3K4 trimethylation at many genes [4, 5]. This functional inconsistency has hindered further understanding of the role of SMYD3 in epigenetic gene regulation [68, 100]. However, poor in vitro activity can be partly explained by the crystal structures [39]. SMYD3 has a closed conformation and a direct lobe-lobe interaction forms a cap over the substrate-binding site. Though this cap structure does not prevent substrate binding, the resulting narrow opening to the active site cavity could potentially affect the substrate binding competence of SMYD3 and thereby the catalytic activity [101]. SMYD3 in vitro activity can be enhanced by Hsp90 and DNA binding [4, 64]. The Hsp90 binding site has been mapped to a TPR-like C-terminal domain (CTD) [102]. Due to the closed conformation, the predicted Hsp90 binding site is half-buried and therefore the question remains how Hsp90 binds to SMYD3 and enhances its activity. The DNA binding site was predicted to be located within the zinc-finger MYND domain [64]. However, this domain is indispensable for SMYD enzymatic activities and how SMYD3 activity is regulated by DNA binding remains a puzzle [94]. Here we present an open SMYD3 conformation and both theoretical and experimental evidence that the conformational selection mechanism and allostery may be involved in SMYD3 functional control.

Materials and Methods

2.1 Molecular Dynamics Simulation

Molecular dynamics simulation was performed using NAMD [103]. Initial structure for simulation was the crystal structure of human SMYD3–sinefungin complex (PDB code: 3PDN). Prior to the simulation, this structure was modified by substituting the cofactor analog sinefungin with cofactor S-adenosyl methionine (SAM or AdoMet). The substitution was based on the structural comparison with the SMYD3–SAM complex (PDB code: 5CCL) and the two SMYD3 structures are very similar with a root-mean-squared-deviation (RMSD) of 0.6 Å. The resulting system including the cofactor SAM was parameterized using CHARMM force field (version 36). The net charge of the Zn ions in the structure was set to +2 and the chelating cysteine and histidine residues were deprotonated. The system was solvated inside an orthorhombic box of water molecules with a 13 Å padding in each direction. The system was then neutralized with NaCl at a concentration of 0.15 M. The final system contained 69,749 atoms. Simulation was performed with a 1 fs time step. Particle Mesh Ewald was used to treat long-range electrostatic interactions and a cutoff of 12 Å was used for non-bonded interactions. Periodic Boundary Conditions were applied during the simulation. The simulation was started with 2,000 steps of energy minimization. The first half of the minimization had harmonic restraints on the protein and the second half unrestrained minimization. The minimized structure was then slow heated from 0 to 300 K over 300 ps. At each integration step velocities were reassigned and the temperature was incremented by 0.001 K. The heated structure was then equilibrated for 300 ps and velocities were rescaled to 300 K at every integration step. The system was further equilibrated using Langevin dynamics for 300 ps at constant temperature (300 K) and pressure (1 bar). The production run was performed in

the NVE (microcanonical) ensemble at 300 K. The total simulation time was 50 ns and coordinates were recorded every 1 ps.

2.2 Principle Component Analysis

Principal component (PC) analysis was performed using Bio3D [104]. The entire 50 ns trajectory of 50,000 frames was used in the analysis. The overall translational and rotational motions in the trajectory were eliminated by least squares fitting to the first frame. A 3 N × 3 N covariance matrix was generated using Cartesian coordinates of Cα atoms. Diagonalization of the covariance matrix generated 3 N eigenvectors, each having a corresponding eigenvalue. The trajectory was projected onto a particular eigenvector to reveal concerted motions. Clustering of the trajectory in the PC space was performed using k-means algorithm. k-means partitions the observations into k clusters by minimizing the mean squared distance from each observation to its nearest cluster center. The number of clusters was chosen based on the "elbow criteria". At a cluster count of two the BSS/TSS (Between-group Sum of Squares/Total Sum of Squares) ratio is 79.8%. The PC analysis- based free energy landscapes were produced by Carma [105]. The domain motions along the PC axes were analyzed using the VMD plugin Hingefind [106].

2.3 Temporal analysis of structural attributes

Temporal changes of structural attributes including hydrogen bonds, salt-bridges, solvent accessible surface area (SASA), Phi and Psi were analyzed using the VMD plugin Timeline [107]. Hydrogen bonds were calculated with a distance cutoff of 3.2 Å and angle cutoff of 20°. Salt-bridges were calculated with a distance cutoff of 3.5 Å. SASA was calculated using a radius extension of 1.4 Å. The calculations were performed every 25 ps.

2.4 Running cross correlation

Residue-pair-wise cross-correlation coefficients were calculated with Bio3D. Running cross correlation (RCC) was calculated using an in-house code. The first element of RCC was obtained by taking the CC of the initial fixed subset of the trajectory. Then the subset was modified by shifting forward: excluding the first frame of the original subset and including the next frame following this subset in the trajectory. This created a new subset of frames, which was used to calculate the next CC. This process was repeated over the entire trajectory. RCC was a plot of the CC against the middle point of the CC time window. Inter- residue RCC deviation map was a heat-map of the standard deviation (σ) of residue-pair- wise RCC. σ was calculated for each RCC; the heat-map represents the magnitude of σ.

2.5 Dynamic network analysis

Dynamical network analysis was done in VMD according to previous protocols[107, 108]. Each amino acid in the network was represented by one node and SAM by three nodes. Amino acid nodes were centered on Cα atoms and SAM nodes were located at atoms Cα, C4' and N9. The edges between nodes were drawn if the residues were within a cutoff distance of 4.5 Å for at least 75% of the trajectory. The edge distances were derived from pairwise correlations which define the probability of information transfer across the edge. Correlations were calculated from the trajectory by the program Carma [105]. The community substructure of the network was obtained using the Girvan-Newman algorithm. Nodes in a community have more and stronger connections within that community than the nodes in other communities.

2.6 Targeted molecular dynamics

Targeted molecular dynamics (TMD) simulation was performed with NAMD. The initial and target structures used for simulation were the most dissimilar structures along the PC1 axis in the full-trajectory PCA (see above). During simulation, all heavy atoms in the CTD were guided

towards the final target structure by steering forces. The force on each atom was given by the gradient of the potential: UTMD = 1⁄2 * k * (RMS(t)−RMS0(t))2, where RMS(t) was the instantaneous best-fit RMS distance of the current coordinates from the target coordinates, RMS0(t) was the preset RMSD value for the current time step and the force constant k was 200 kcal·mol−1·Å−2. Other simulation parameters were the same as those used in the above conventional molecular dynamics simulation.

2.7 Protein Expression and purification

Human SMYD3 was essentially expressed and purified as previously described[39, 109]. In brief, SMYD3 was cloned with a His_6-SUMO tag in a pCDF-SUMO vector. Clones were inoculated in LB media and grew until an OD_{600} reached 0.4–0.6. Cells were induced with 0.1 mM isopropylthio-β-D-galactoside (IPTG) and grown overnight at 15 °C. Cells were harvested and lysed using a French Press. Lysate was spun down and the supernatant was collected for purification. The His_6-SUMO-SMYD3 was captured with a Ni^{2+}-affinity column and the His_6-SUMO tag was removed by yeast SUMO protease 1. Native protein was separated after running through a second Ni^{2+} column. Finally, SMYD3 was further purified by a size exclusion column in 20 mM Tris pH 8.0, 150 mM NaCl, 3% glycerol and 2 mM Tris(2-carboxyethyl) phosphine (TCEP).

2.8 Small angle X-ray scattering

Small angle X-ray scattering (SAXS) data were collected at BioCAT beamline at Argonne National Laboratory. Solution conditions were 20 mM Tris pH 8.0, 150 mM NaCl, 3% glycerol and 2 mM TCEP. All measurements were made at 25 °C using a 100 µL capillary flow-cell. Scattering data were collected at two SMYD3 concentrations: 1.2 and 7.7 mg/mL. Five frames with a 1s exposure were taken and data were averaged and subtracted from averaged buffer frames.

Low and high concentration data were merged based on an aligned middle q region to generate a single scattering curve with a q range of 0.0042–0.39 Å−1. Radius of gyration (Rg) values were calculated using the Guinier approximation [110]. The distribution function of interatomic distances within SMYD3, P(r), was estimated from the scattering data using the GNOM algorithm [110]. Ab initio dummy atom models were generated using DAMMIN [111]. Normal mode analysis was carried out by SREFLEX [112]. Theoretical scattering curves of SMYD3 structures were calculated with CRYSOL [110].

2.9 Statistical Analysis

Significance of mean differences for continuous data was evaluated by two-tailed t-test and circular data (Phi and Psi) by Watson-Williams high concentration F test. Association between continuous data was measured with Pearson correlation coefficient (r). Association of hydrogen bonds and salt-bridges with conformational states were evaluated by PHI coefficient and association of backbone angels or solvent accessible surface area by point- biserial correlation coefficient. For backbone angles, sine values of angles were used in correlation analysis.

Results

3.1 Conformational transition from the closed to open states

Molecular dynamics (MD) simulation reveals a striking conformational transition of SMYD3 from the closed to open states. The closed state is a crystal structure-like state characterized by a direct lobe-lobe interaction at top of the substrate-binding site (Figure 1A). The open state represents a previously-unidentified new conformational state which lacks the equivalent interaction between the two lobes (Figure 1B). In the closed state, the lobe-lobe interaction involves residues W300 from the C-lobe and S44, V47, V48, Q191 and V193 from the N-lobe (Figure S1A). The interaction includes a hydrogen bond from W300 to S44 and

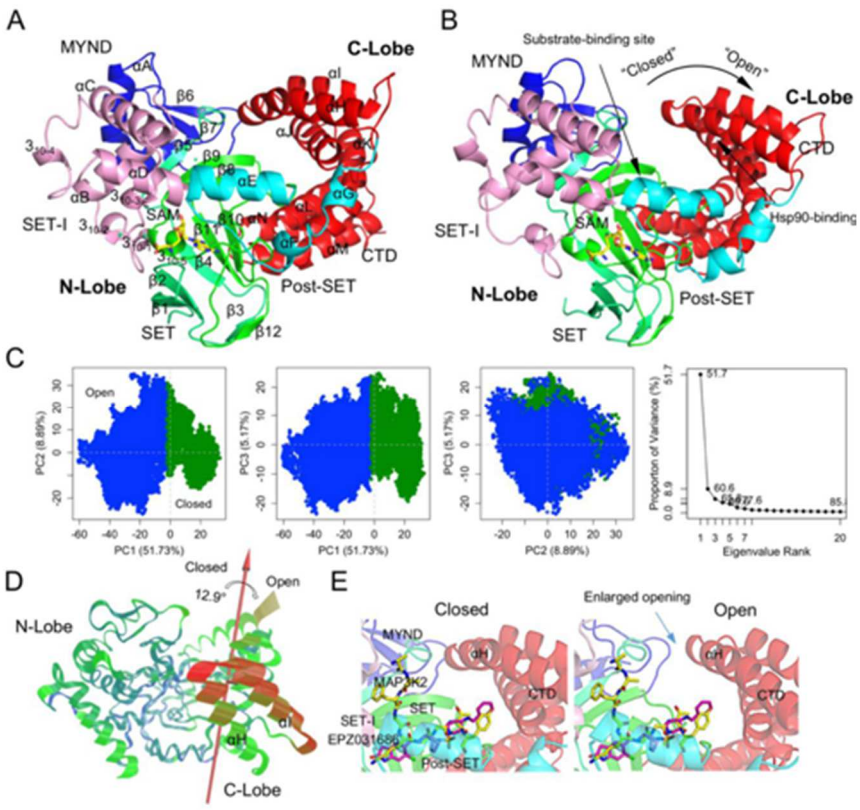

Figure 9. New open conformation of SMYD3
(A) A closed-state and (B) open-state structure. SMYD3 is colored according to domain. Secondary structures are labeled and numbered according to their position in the sequence. (C) Principle component analysis (PCA) of full 50-ns trajectory. Left three, projection of the trajectory onto the planes formed by the first three principle components. Conformers are colored according to the k-means clustering. Rightmost, scree plot showing the proportion of variance against its eigenvalue rank. (D) Visualization of the motions along PC1. Color scale from blue, green, to red depicts low to high atomic displacements. (E) Superimposition of the open and closed states with an SMYD3 bound peptide (MAP3K2, yellow) and inhibitor (EPZ031686, purple).

hydrophobic interaction of the W300 side chain with a pocket formed by the aforementioned N-lobe residues. The open state is characterized by the break of the direct lobe–lobe interaction. The W300–S44 hydrogen bond breaks and the side chain of W300 flips out from the small hydrophobic pocket. The substrate-binding site is widened and there is a clear gap between the N- and C-lobes. As a result, the open state shows larger structural difference from the crystal structure (Figure S1B).

The conformational transition can be illustrated by the change in W300–S44 distance. In the closed state W300–S44 maintains a hydrogen bonding distance for most of the time (Figure S2A). In the open state the hydrogen bond breaks and their distance fluctuates between 4.9 Å and 21.2 Å. The distance shows a steep rise during the transition phase and the transition happens in less than 0.3 ns. Therefore, the change in W300–S44 distance can clearly separate the two conformational states. Covariance-based principal component analysis (PCA) further demonstrates the presence of structure-distinct conformational states. The first PC axis alone is sufficient to define two major clusters, one corresponding to the closed state and the other the open state (Figure 1C). The two clusters are well separated along the PC1 axis and the boundary between them is marked by low population of conformers (Figure S2B). This statistically indicates a free-energy barrier for conformational transition. PC1 accounts for more than 50% of overall variance and the motion described by PC1 is a clamshell-like motion between the N- and C-lobes (Figure 1D and S2C). The rotation axis of this motion passes between the two lobes lying at the bottom of the gap between the two lobes. Therefore, this motion essentially depicts an open–closed dynamics and the conformational transition between the closed and open states.

3.2 New open ligand-binding-capable conformational state

The new open state may represent a conformational state that facilitates substrate or effector binding to SMYD3. The open state shows an enlarged opening to the substrate binding site which may make it more accessible to a substrate than the closed state (Figure 1E). There is over 35% increase in the accessible volume of the substrate binding cavity in the open state. The first α helix of the CTD (αH) is responsible for the widening and increased accessibility. This helix is involved in the direct lobe–lobe interaction and undergoes a large movement during the

transition from the closed to open states (Figure 1D). Because of this movement, the substrate-binding site is widened and more solvent-exposed in the open state.

The predicted Hsp90-binding site also becomes more solvent-exposed in the open state. The C-terminal MEEVD motif of Hsp90 was predicted to bind between αJ and αL at the inner surface of the CTD (Figure S2D) [94, 113]. This binding site is structurally similar to the putative TPR peptide-binding site[94]. However, in the closed state the Hsp90-binding site is half-buried due to the direct interaction between αH and the N-lobe. The binding site is further buried due to the lobe-bridging β8–β9 hairpin sitting in front of the binding site. In the open state the distance between the β8–β9 hairpin and Hsp90-binding site becomes significantly larger (Figure S2E) and the volume of the binding site cavity is three times more than that in the closed state. Therefore, the more exposed binding site in the open state may facilitate Hsp90 binding to SMYD3 and provide a mechanistic basis for Hsp90-induced activity enhancement.

3.3 Distinct structural characteristics of the closed and open states

The closed and open states show distinct structural characteristics. They are different in hydrogen bonding, salt-bridge, backbone angles and solvent accessible surface area. Hydrogen bonding is different in pattern but not total number (Figure 2A). The closed state has an average of 143.0 hydrogen bonds and open state 142.8. Their difference is not statistically significant ($p = 0.498$). However, there are 18 hydrogen bonds whose time- course pattern shows a significant correlation with the conformational states ($r > 0.5$). Six of them are strongly correlated with the open state and 12 with the closed state including the W300–S44 hydrogen bond. Residue D272 is involved in two conformational state-specific hydrogen bonds. One hydrogen bond (S246–D272) shows the strongest correlation with the closed state and the other (R249–D272) with the open state. D272 is located at the junction between the post-SET and CTD (Figure S3A). In the open

state, D272 moves slightly towards the substrate-binding site. The movement breaks its hydrogen bond to S246 and leads to the formation of the hydrogen bond with R249. This indicates that the hydrogen bonds S246–D272 and R249–D272 may be mutually exclusive. The time-course patterns of these two hydrogen bonds show a strong negative correlation ($r = -0.548$).

The numbers of salt-bridges in the closed and open states are significantly different ($p < 2.2 \times 10-16$). The closed state has 50.4 salt-bridges and open state 54.9. Nine salt-bridges show a significant correlation with the closed state and 16 with the open state ($r > 0.5$) (Figure 2A). The salt-bridge D332–K375 has the strongest correlation with the closed state ($r = 0.907$). This salt-bridge stabilizes the closed state by pulling together the helices αJ and αL of the CTD (Figure S3B). This also contributes to the buried state of the Hsp90-binding site. The salt-bridge D272–R249 shows the strongest correlation with the open state ($r = 0.838$). This correlation is consistent with the open-state-correlated hydrogen bonding between these two residues. The D272–R249 salt-bridge pulls αG towards the substrate-binding site and the pulling squeezes the bottom lobe–lobe interface. The salt-bridge D209–K271 also shows a significant correlation to the open state ($r = 0.795$). However, the direction of the force exerted by this salt-bridge is different from that by the D272–R249 salt-bridge. The D272– R249 exerts the force along the axis of the rotation describing the open-closed lobe–lobe motion. The D209–K271 exerts the force perpendicular to this axis at the opposite surface of the substrate-binding site. The D209–K271 stabilizes the open state by pulling the two lobes outwards.

Many residues show significant differences in the backbone torsion angles. Fifty-five percent of residues are significantly different in Psi and 51% in Phi ($p < 0.001$). Twelve and seven residues show more than 30° differences in Psi and Phi respectively (Table S1). There are 21

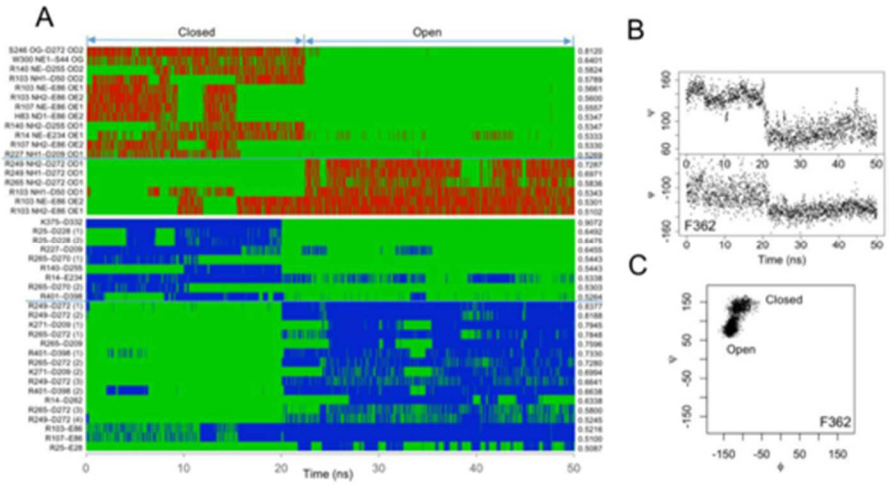

Figure 10. Distinct structural characteristics
(A) Conformational state-correlated hydrogen bonds (top) and salt-bridges (bottom). Red and blue lines indicate the presence of interactions and green lines absence. (B) Torsion angles of F362 as a function of time. (C) Ramachandran plot of F362 trajectory.

residues whose Psi changes show a significant correlation with the conformational states and 12 residues for Phi ($r > 0.5$). Both Psi and Phi of residue F362 show strong correlation with the conformational states (rpsi = 0.983, rphi = 0.839) (Figure 2B). There are clearly two populations in its Ramachandran plot, one corresponding to the closed state and the other open state (Figure 2C). The neighboring residues of F362 also show large changes in the backbone angles and significant correlation with the conformational states (residues 363– 366) (Table S1). These residues are located in a short loop connecting the fourth and fifth helices of the CTD. The changes in their backbone angles are correlated with a twisting motion between those two helices during the conformational transition (see below). Their backbone-angle changes are also correlated with a significant change in F362 interacting network. In the closed state, F362 forms a π-π stacking interaction with Y358 (Figure S3C). In the open state, this interaction is replaced by the stacking interaction with H366. As a result, F362 prevents M242 from interacting with H366. The loss of this interaction may weaken the interaction between the N- and C-lobes near the axis of the rotation describing the open-closed motion.

The SASA of the closed and open states are significantly different ($p < 2.2 \times 10{-}16$). Unexpectedly, the closed state is more solvent exposed. The average SASA of the closed state is 116,339.3 Å2 and open state 116,250.9 Å2. Sixty-eight percent of residues show a significant difference in SASA ($p < 0.001$). Fifty-six percent of these residues are more exposed in the closed state than open state. There are 24 residues whose SASA changes show a significant correlation with the conformational states ($r > 0.5$) (Figure S3D). Seventeen of them are located within the CTD. These include three residues (M335, L344 and Q372) lining the Hsp90-binding site, which are more exposed in the open state; and three residues (C309, A334 and C338) at the interface between the second and third helices of the CTD, which become more buried in the open state. The CTD is a key structural determinant of the closed and open states. The enrichment of residues with the conformational state-specific SASA values reflects the characteristic structural changes in the CTD defining the conformational states.

3.4 Different dynamical characteristics

The closed and open states have different dynamical characteristics. They are different in flexibility, cross correlation, interatomic distance fluctuation and dynamical network. The closed state is significantly less dynamical than open state ($p < 2.2 \times 10{-}16$). The average atomic displacement of the closed state is 0.81 Å and open state 1.24 Å (Figure S4A). The flexibility of the CTD increases to a larger extent than the N-lobe in the open state. The average ratio of root-mean-square fluctuation (RMSF) of the open to closed states is 1.36 for the N-lobe and 1.83 for CTD. However, the overall fluctuation pattern is not significantly different and the correlation

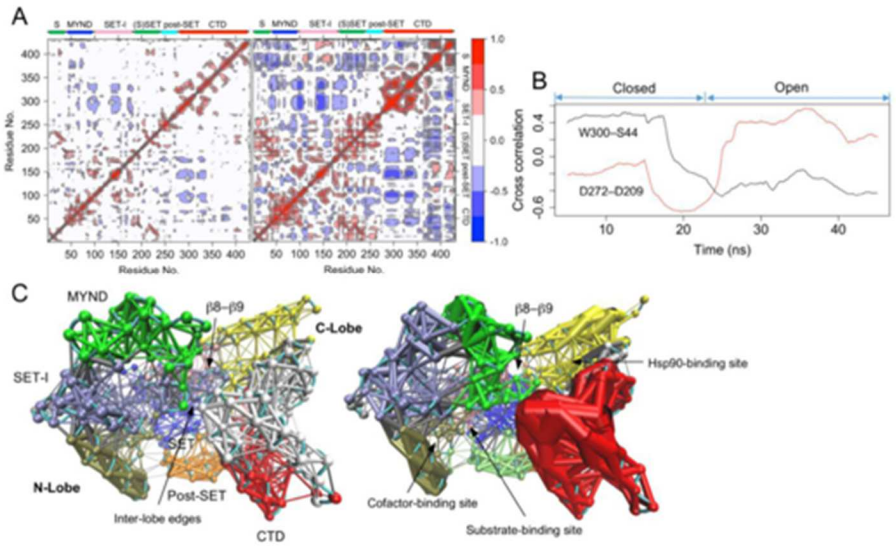

Figure 11. Different dynamical characteristics
(A) Cross-correlation map of the trajectory. Left, the closed state; right, open state. Blue and red indicate negative and positive correlation respectively. (B) Running cross correlation (RCC) of the residue pairs W300–S44 and D272–D209. (C) Dynamical network analysis of the closed (left) and open (right) states. Networks are colored according to communities. Points in the network are nodes and lines between the nodes represent edges. Thicker lines depict the stronger edges or stronger correlations.

between the two states is 0.753. Most of the residues in both states have a below 1 Å atomic displacement. The least dynamical region is the SET domain in both states. The SET is the catalytic domain responsible for cofactor and substrate binding. Several regions show a notable difference in flexibility. In the closed state the regions around residues W300 and S44 are less dynamical than those in the open state. The two regions interact with each other in the closed state and such interaction appears to restrain their flexibility.

Dynamic cross-correlation patterns of the closed and open states are different. The closed state shows a significantly lower level of correlated motions ($p < 2.2 \times 10{-}16$). The average correlation coefficients of the closed state and open state are 0.147 and 0.243 respectively. In both states, the SET-I and the first three helices of the CTD show strong negative correlated dynamics and the MYND motion is negatively correlated with the CTD motion (Figure 3A). Such negative

correlated dynamics is consistent with the open-closed motion between the N- and C-lobes. The open state shows many additional correlated motions. Among the most notable ones are those between the last three helices of the CTD and many regions across all domains. To quantitatively characterize the dynamical change in correlated motion, we developed the running cross correlation (RCC) method (see Methods). RCC shows a time- course change in cross correlation. It should smooth out short-term fluctuations and highlight longer-term trends or changes. RCC analysis shows that the cross-correlation profile of the residue pair W300–S44 evolves and changes during the simulation (Figure 3B). The motions of W300 and S44 are positively correlated in the closed state when they interact, but change to a negative correlated dynamics after the conformation is transited to the open state. Inter-residue RCC deviation analysis shows that W300–S44 is among the residues pairs with the largest RCC variations ($\sigma = 0.353$) (Figure S4B). The largest variation is found between the residue pair D272–D209 ($\sigma = 0.384$). These two residues are not in the close proximity but both involved in conformational state-specific hydrogen bonding and salt-bridges (Figure 2A).

The patterns of interatomic distance fluctuation are different between the closed and open states. The closed state shows a significantly lower level of fluctuation ($p < 2.2 \times 10{-}16$). The average fluctuation of the closed state and open state are 0.598 Å and 0.880 Å respectively. Both states show large distance variations between the lobes and the variations within the lobes are significantly lower (Figure S4C). The average level of the between-lobe variations of the open state is two times above that of the closed state. This indicates greater distance variability between the N- and C-lobes in the open state. All components of the N- lobe in the open state show significant distance variations with respect to the C-lobe, but only the MYND and SET-I shows large variations in the closed state. The W300–S44 distance deviates about 0.519 Å and 2.075 Å

in the closed and open states respectively. This difference is in agreement with the direct interaction of the two residues in the closed state and the break of this interaction in the open state. Dynamical network and communities are different between the closed and open states. There are ten communities in the closed state and 11 in open state (Figure 3C). The community assignment in both states is roughly correlated with the sequence- and structure- based domain assignment [94]. However, there are significant differences in the ways of partitioning the domains into communities. The most significant difference is found at the CTD. The CTD is split into three major communities in the closed state, whereas in the open state it is split into two. In both states the last three helices of the CTD form a separate community, but its first four helices form a single community in the open state and are split in half along the middle of the helices in the closed state.

This indicates that the residues in the first four helices of the CTD have stronger connections in the open state than they do in the closed state. Of note, the predicted Hsp90-binding site is located between the two open- state-CTD communities. Another notable difference in the dynamical networks is found at top of the substrate-binding site. Because of the direct lobe–lobe interaction, there are inter- lobe edges at this location in the closed state; but without the equivalent interaction, the open state has no edge. This indicates that the closed state may possess additional paths for dynamical inter-lobe communication.

3.5 Substrates

The conformers in the closed and open states can be further clustered into substates. PC analysis shows that both states consist of two major substates but the motions relating the substates are different (Figure S5A). For the closed state, PC1 accounts for nearly one fifth of the overall variance. The major motion along PC1 is a twisting motion of the N-lobe with respect to the C-lobe (Figure S5B). The axis of the twisting passes through the MYND, β8– β9-containing β sheet

and middle of the cofactor-binding site. For the open state, PC1 accounts for 38.4% of the overall variance. The major motions along PC1 include a clamshell-like motion between the N-lobe and first four helices of the CTD; and a twisting motion of the last three helices of the CTD with respect to the N-lobe (Figure S5C). The axis of the former rotation aligns with the axis of the motion depicting the conformational transition between the closed and open states (Figure 1B). In the closed state, the PC1- described twisting motion affects the funnel-shape substrate-binding site. The twisting pulls the β8–β9 hairpin and β12–αD loop together and apart. This alters the dimensions of the substrate-binding site. The funnel-shape substrate-binding site has been proposed to contribute to SMYD2 substrate recognition [113]. In the open state, the PC1-described motions affect the dimensions of the inter-lobe gap and the distance between the CTD and β8–β9 hairpin (Figure S5C). As a result, both substrate-binding site and Hsp90-binding site are exposed to different extents in the substrates.

3.6. Pathway of the conformational transition

Targeted molecular dynamics (TMD) simulation reveals the conformational transition pathway between the closed and open states (Figure S5D). The forward and reverse transitions follow similar structural conversion processes. The two conformational states are interconverted by a reversible CTD rotation. The axis of the rotation passes through the fifth helix (αL) of the CTD parallel to the helical axis. αL is relatively static during the conformational transition. The average RMSF of this helix is 1.7 Å compared to 4.2 Å for the first four helices of the CTD and 2.2 Å for the last two helices. The differences in these RMSFs are significant ($p < 9.0 \times 10{-}6$). αL is involved in direct interaction with the β8–β9 hairpin (Figure S5C). This interaction secures αL in position, appears to assist in rotating the CTD around this axis and thereby may contribute to a proper conformational transition between the closed and open states. In agreement with the

conventional molecular dynamics (Figure 1E and S2D), TMD also shows that the conformational transition regulates the degrees of exposure of the substrate-binding site and Hsp90-binding site.

3.7. Small angle X-ray scattering

To provide experimental support for the MD-sampled open state, the solution structure of SMYD3 was characterized using small angel X-ray scattering (SAXS) (Figure 4A). SAXS analysis shows that the radius of gyration (Rg) of SMYD3 is 24.5 Å in solution and Dmax (maximum particle dimension) 78.0 Å. These values are similar to the Rg (23.2 Å) and Dmax (77.8 Å) calculated from the crystal structure. The ab initio shape modeling shows that the dummy atom model visually matches the crystal structure (Figure 4B). The last three helices of the CTD fits into a slightly protruding envelope and there is a miniature groove between the N- and C-lobe-corresponded regions. However, this dummy model can also be fitted equally well with an open state structure (Figure 4B). This indicates that the low resolution of SAXS model is unable to distinguish between the closed and open states.

The theoretical scattering curve calculated from the crystal structure does not completely fit with the experimental data. The fitted $\chi 2$ is about 2.68. At the low q regions, the fitted curve is in a good agreement with the experimental data, but the high-q regions beyond 0.15 $Å^{-1}$ are not being well explained by the fitting (Figure 4A). This suggests that the crystal structure is somewhat different from the solution structure; more strictly, it is different from the average structure of the SMYD3 conformational space. However, the fitting statistics can be improved by normal mode analysis (NMA). NMA probes the large-scale motions in SMYD3 and estimates the structural flexibility to improve agreement with the SAXS data. The best model from the NMA has an

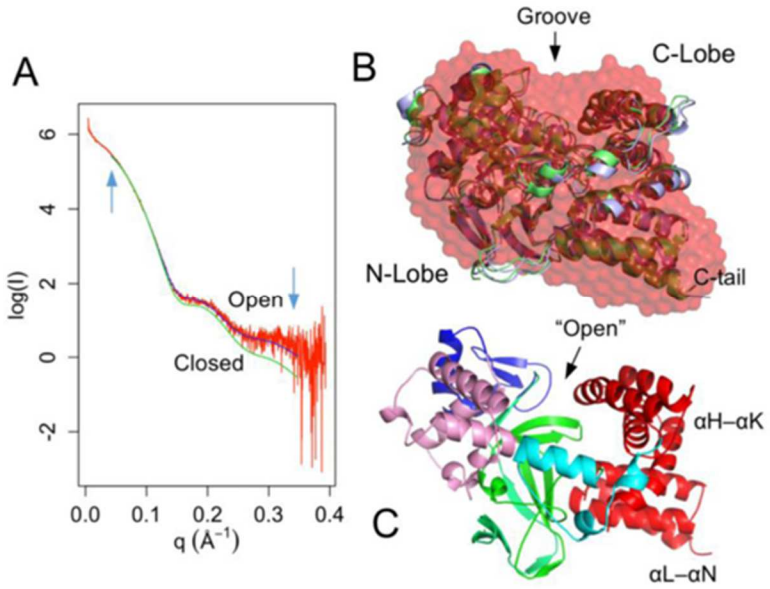

Figure 12. Small-angle X-ray scattering
(A) Experimental scattering curve (red) overlaid with theoretical scattering curves calculated from a closed (green) and open (blue) SMYD3 structure. The q range used for model fitting is indicated by arrows. (B) Ab initio dummy atom model (red) superimposed with a closed (green) and open (blue) structure. (C) An open structure derived from normal mode analysis.

improved χ2 of 1.72. The CTD in this model undergoes large conformational changes including a twisting motion of the first two helices and an outward-bending motion of its second half (Figure 4C). The lobe–lobe bridging interactions at the W300–S44 interface in this model break. Such a conformation resembles the open state structures sampled in the above MD analysis.

To correlate the MD simulation with the SAXS experiment, the entire MD trajectory was fitted to the experimental data. The average χ2 of the trajectory is 3.37 (Figure S6A). The closed state shows significantly lower χ2 than the open state ($p < 2.2 \times 10-16$). The average χ2 of the closed state is 2.71 and open state 3.90. This would indicate that the closed state fits better to the SAXS data than the open state. However, the best fitting conformer adopts an open structure. 32% of the open state has a χ2 less than the average value of the closed state. This is consistent with normal mode analysis where the open CTD structures show the best agreement with the experimental data (Figure 4C). This also indicates that the combination of all motions in a

conformational state determines the results of the experimental data fitting, rather than the open–closed motion alone (Figure 1D and S5). In the open state, the $\chi 2$ is widely spread with a σ value of 1.63 compared to 0.22 for the closed state (Figure S6A). This is consistent with highly dynamical nature of the open state (Figure S4A). The Rg of the trajectory shows a mixed negative/positive correlation with the $\chi 2$ (Figure S6B). The open state has larger Rg values than the close state ($p < 2.2 \times 10{-}16$). The average Rg for the closed state and open state are 23.2 Å and 23.6 Å respectively. This indicates that the closed-state structures are more compact than the open-state structures. The Rg is strongly negative correlated with the $\chi 2$ when it is less than 23.4 Å ($r = 0.612$) and changes to a positive correlation at the higher values ($r = 0.874$) (Figure S6B). The negative correlated region samples both closed and open conformers and the population of the closed state in this region is 2.5 times more than that of the open state. However, 75% of the top 1% best fitted conformers adopt an open conformation. This further indicates that some of the open state structures are closer to the average structure of the SMYD3 conformational space.

Conclusion

SMYD proteins are an exciting field of study as they are linked to many types of cancer-related pathways [4]. Cardiac and skeletal muscle development and function also depend on SMYD proteins opening a possible avenue for cardiac-related treatment [94]. Among SMYD proteins, SMYD3 has received the most attention because of its involvement in epigenetic and non-epigenetic regulation of numerous cancerous genes [4, 5, 94, 95]. Due to its tumor-growth-inducing role and association with poor prognosis SMYD3 has emerged as a key target for anti-cancer therapies [114]. However, the biochemical mechanism of SMYD3-mediated methylation remains elusive. The “closed” substrate-binding site and poor in vitro H3K4 methyltransferase activity have led to arguments that SMYD3 is not a histone lysine methyltransferase and the in

vivo-associated H3K4 tri-methylation might be catalyzed by other methyltransferases [5, 68]. Such arguments have obscured our understanding of the role of SMYD3 in epigenetic gene regulation, where a completely different interpretation of SMYD3 function could result from the arguments: SMYD3 functions as a histone code "writer" defining chromatin states, or only serves to anchor other chromatin-associated proteins through its sequence-specific DNA binding. Here we provide theoretical and experimental evidence that SMYD3 can adopt an open conformation. This new open conformational state is substantially different from the crystal structure-like closed state. The two states are related by a striking clamshell-like motion of the C-lobe with respect to the N- lobe and SMYD3 is transited by this large motion from a ligand binding-incapable state to a binding-capable state. A recent MD study revealed that the CTD can undergo a similar hinge-like motion resulting in expanded substrate binding crevice [115]. In the absence of the cofactor, the CTD samples more open configurations than it does in the presence of the cofactor [115]. It was postulated that the cofactor acts like a key and locks SMYD3 in a closed conformation [115]. However, the present MD study shows that SMYD3 can undergo a spontaneous conformational transition from the closed to open states in the presence of the cofactor. The conformational transition leads to the enlarged opening to the substrate binding site in the open state which could increase histone tail accessibility to the active site cavity and target lysine access channel. This would then provide the mechanism for SMYD3 activity on both H3K4 methylation and H3K4me3 binding. A recent study showed that SMYD3 interacts with H3K4me3 modified histone tails, which facilitates its recruitment to the core promoter regions of most active genes [95].

The conformational transition pathway involves a reversible twisting motion of the CTD and the transition from the closed to open states breaks the top lobe–lobe interface resulting in a more accessible substrate-binding site and Hsp90-binding site. Many structural and dynamical

changes are associated with this conformational transition and these changes may either contribute to the transitional process or stabilize the particular conformational states. While the exact portion of each conformational state in solution is unknown, the closed state statistically better fits the experimental data than the open state, but the best fitting conformers adopt an open structure. Nevertheless, the presence of both closed and open states in the conformational ensemble suggests two possible, mutually non-exclusive models for SMYD3 functional regulation. First, a conformational selection mechanism may regulate SMYD3's ligand binding. In the conformational selection model, the intrinsic dynamics of the protein lead it to spontaneously transition between a stable unbound and a less stable bound conformation. The apo-protein visits the bound state with significant probability and the ligand can bind directly to this conformation shifting the distribution of conformers towards the bound population [116]. Therefore, the open state with the exposed ligand- binding sites suggests that the ligand binding of SMYD3 may be regulated by the conformational selection mechanism. In addition, the highly correlated inter-lobe dynamics in the open state may facilitate SMYD3 promiscuity through the conformational selection mechanism, allowing the structural adaptation to different substrates. The conformational selection mechanism has been shown to be involved in promiscuous ligand binding and this assumes that the protein needs to visit multiple binding conformers capable of binding different ligands [116]. In SMYD3, the inter-lobe dynamics will alter the size of the substrate- binding site. The coupling of the two lobes by the correlated motion might thus offer the specificity and promiscuity for substrate recognition.

Second, our results provide a model for possible allosteric regulation and a population shift between the two conformational states may underlie the functional control of SMYD3. Recent data suggest that allostery can be mediated by transmitted changes in protein dynamics [117]. The

binding of an allosteric effector can result in the redistribution of protein conformational ensembles and cause changes in catalytic or ligand binding competence [117]. DNA binding to the N-lobe has been shown to enhance SMYD3 methyltransferase activity [64]. The interaction of SMYD3 with BRD4 mediates the recruitment of transcriptional cofactors at the myostatin gene and regulates skeletal muscle atrophy [12]. SMYD3 interacts with PC4 in tumor cells and such interaction stimulates oncogenic gene expression through deposition of H3K4 tri-methylation [97]. All these interactions are mediated via the MYND domain of SMYD3, but the structural and dynamical consequences of the interaction remain unknown. One possibility is that the interaction may affect the domain dynamics and inter- lobe dynamical correlation. Such an effect could be transduced to other parts of the protein through the edges bridging the dynamical communities and this might in turn cause a population shift between the existing conformational states, thereby modulating active site or binding site geometries. In summary, a detailed study of SMYD3 structure and dynamics is of functional and therapeutic importance. The identification of the open conformational state provides the basis for the conformational selection mechanism and allosteric regulation.

CHAPTER 3 MOLECULAR DYNAMICS SIMULATION REVEALS CORRELATED INTER-LOBE MOTION IN PROTEIN LYSINE METHYLTRANSFERASE SMYD2

*Published in PLoS One. 2015 Dec 30;10(12):e0145758. doi: 10.1371/journal.pone.0145758.
All authors agreed with including their work in this dissertation.

Introduction

SMYD is a special class of protein lysine methyltransferases involved in heart and muscle development[1, 94]. SMYD linked to tumorigenesis opens a possible avenue for cancer treatment [4, 94]. SMYD proteins contain five members, SMYD1–5 [24, 35, 39, 57, 94]. Each member contains a conserved SET domain that is "split" by a zinc-finger MYND domain [94]. The SET domain is a catalytic unit responsible for protein lysine methylation [36]. The MYND domain is a protein–protein interaction module and has also been shown to have a DNA binding ability in SMYD proteins[4, 37, 38]. Among SMYD proteins, SMYD2 has the broadest substrate specificity. In addition to histone proteins, SMYD2 is able to methylate p53, retinoblastoma tumor suppressor (RB), estrogen receptor α (ERα), poly(ADP-ribose) polymerase 1 (PARP1), and heat shock protein-90 (Hsp90)[3, 49, 50, 94]. Through these methylations, SMYD2 is involved in several cellular processes including cell cycle progression, apoptosis, cellular differentiation, DNA damage response, and epigenetic gene regulation [94].

The crystal structures revealed that SMYD proteins have a bilobal structure[24, 35, 39, 57, 94]. The N- lobe contains the SET, MYND, SET-I, and post-SET domains, and the C-lobe is made up of the CTD domain. The cofactor-binding site is located in a surface pocket in the N-lobe. The substrate-binding site is located between the N-lobe and CTD and situates at the bottom of a deep cleft. The orientation of the CTD is different among the SMYD family. This difference is reflected by the relative positions of the N- and C-lobes resulting in open and closed structures [39]. In SMYD2 the CTD is flexible and can undergo a conformational change upon binding to different

cofactors [24]. Such a conformational change results in two SMYD2 structures with a slight difference in the size and shape of the substrate-binding pocket. The functional significance of the SMYD2 conformational change is still unknown. One possible consequence is that the conformational change may affect substrate access to the active site, thereby regulate substrate binding [57]. Another possibility is that the conformational change may be important for SMYD2 promiscuity allowing the structural adaptation to different substrates [57]. Finally, the conformational change may provide an allosteric mechanism for the effector-induced activity enhancement and change in substrate specificity [24].

Current understanding of the SMYD conformational change is limited to the structural differences observed in the crystal structures. The dynamical nature of the SMYD proteins is still poorly understood. It remains unknown of the structure of dynamical networks and the pattern of correlated domain motions, both of which are fundamental in mediating substrate recognition and allostery [118, 119]. Using the molecular dynamics (MD) simulation, this study reveals that SMYD2 exhibits a negative correlated inter-lobe dynamics. Dynamical network analysis suggests optimal and suboptimal paths for such a correlation. This study provides insight into SMYD2 dynamics and could prove valuable in understanding SMYD2 substrate specificity.

Materials and Methods

Molecular Dynamics Simulation

Molecular dynamics simulation was performed using NAMD [103]. CHARMM force field was used to parameterize the simulation. Initial structure for simulation is the crystal structure of the SMYD2–SAH complex (PDB code: 3QWV; SAH: S-adenosyl-L-homocysteine or AdoHcy). The missing residues of the structure including two N-terminal residues and one C-terminal residue were filled using SWISS-MODEL [120]. The system was solvated inside an orthorhombic box of

water molecules with a 13 Å padding in each direction. The system was then neutralized with NaCl at a concentration of 0.15 M. The final system contained 78,008 atoms. Simulation was performed with a 1 fs time step. Particle Mesh Ewald was used to treat long- range electrostatic interactions and a cutoff of 12 Å was used for non-bonded interactions. Periodic Boundary Conditions were applied during the simulation. The simulation started with 2,000 steps of energy minimization. The first half of the minimization had harmonic restraints on the protein, and the second half unrestrained minimization. The minimized structure was then slow heated from 0 to 300 K over 300 ps. At each integration step velocities were reassigned and the temperature was incremented by 0.001 K. The heated structure was then equilibrated for 300 ps and velocities were rescaled to 300 K at every integration step. The production run was performed in the NVE (microcanonical) ensemble at 300 K. The total simulation time was 2 ns and coordinates were recorded every 1 ps.

Principal Component Analysis ad Clustering

Principal component (PC) analysis was performed using Bio3D [104]. The entire 2 ns trajectory of 2000 frames was used in the analysis. The overall translational and rotational motions in the trajectory were eliminated by least squares fitting to the first frame. A 3 N × 3 N covariance matrix were generated using Cartesian coordinates of Cα atoms. Diagonalization of the covariance matrix generated 3 N eigenvectors, each having a corresponding eigenvalue. The trajectory was projected onto a particular eigenvector to reveal concerted motions. Clustering of the trajectory in the PC space was performed using k-means and hierarchical clustering algorithms. k-means partitions the observations into k clusters by minimizing the mean squared distance from each observation to its nearest cluster center. Hierarchical clustering builds a hierarchy of clusters based on the distance between the observations.

Dynamical Network Analysis

Dynamical network analysis was done in VMD according to previous protocols [107, 108]. Each amino acid in the network was represented by one node and SAH by three nodes. Amino acid nodes were centered on Cα atoms and SAH nodes were located at atoms Cα, C4', and N9. The edges between nodes were drawn if the residues were within a cutoff distance of 4.5 Å for at least 75% of the trajectory. The edge distances were derived from pairwise correlations which define the probability of information transfer across the edge. Correlations were calculated from the trajectory by the program Carma [105]. The community substructure of the network was obtained using the Girvan-Newman algorithm. Nodes in a community have more and stronger connections within that community than the nodes in other communities. Critical nodes were defined based on the betweenness, which measures the importance of a node to the entire network. Critical nodes connect communities and lie at the interface between pairs of communities. Optimal and suboptimal paths were generated from the initial dynamical net- work matrix. The optimal path is the shortest path between two given nodes. Suboptimal paths are paths that are slightly longer than the optimal path.

Results

1. Dynamical Details of SMYD2-SAH Complex Structure

The simulation was performed using NAMD [103]. The starting structure is the crystal structure of SMYD2–SAH complex. The simulation was done in the NVE ensemble. The system was slow heated and equilibrated before 2 ns production simulation. The stability of the system during the production stage was evident by stable kinetic energy, potential energy, temperature, and RMSD (root mean square deviation) (data in S1 Fig and Fig 1A). The protein structure does

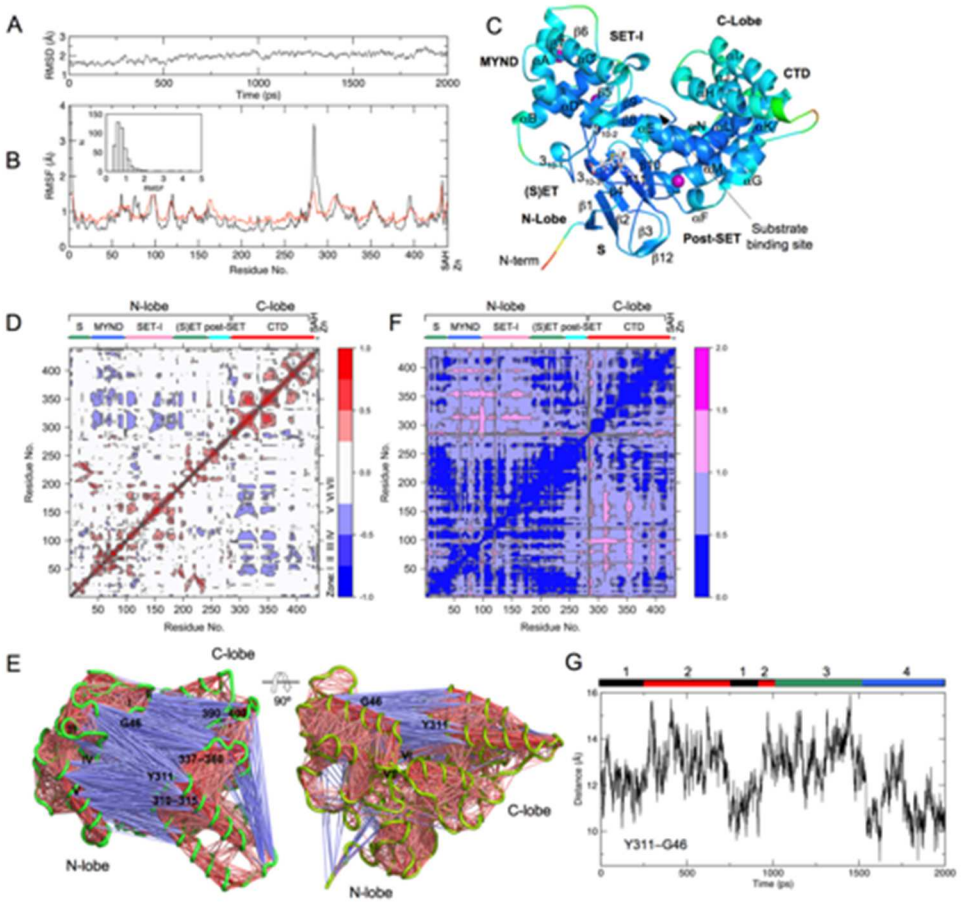

Figure 13. SMYD2 dynamics

(A) Backbone RMSD during the simulation. RMSD was calculated relative to the crystal structure. (B) Root mean square fluctuation (RMSF) of Cα atoms during the simulation (black line). Red line depicts the RMSF values converted from crystallographic B-factors. The inset depicts the distribution of the simulation RMSF. (C) Ribbon diagram of SMYD2 structure at 2 ns. The structure is colored according to the simulation RMSF. Color scale from blue to red depicts low to high atomic fluctuations. Secondary structures, α-helices and β-strands are labeled and numbered according to their position in the sequence. SAH is represented by sticks and zinc ions by purple spheres. (D) Cross-correlation map of the trajectory. Blue indicates a negative correlation between residue fluctuations, and red depicts a positive correlation. Lobe and domain structures of SMYD2 are indicated on the top of the map. (E) Visualization of residue–residue cross-correlations. SMYD2 is depicted by green coils. Blue and red lines indicate negative and positive correlated motions. (F) Inter-residue distance deviation map. Color scale from blue to magenta depicts small to large distance deviations. (G) Distance fluctuation of Y311–G46 during the simulation. Color bars depict the conformer clustering results obtained in Fig 2.

not significantly deviate from the crystal structure. The backbone RMSD fluctuates between 1.4 Å and 2.5 Å with an average value of 2.0 Å.

SMYD2 dynamics revealed by the MD is similar to that from the crystallographic B factor-based analysis (Fig 1B and 1C). Most of the residues have a below 1 Å atomic displacement. The least dynamical region is the SET domain. The SET is the catalytic domain responsible for cofactor binding and substrate binding. The post-SET which is tethered to the SET by a zinc ion also shows a less dynamical structure. The most dynamical region in SMYD2 is found in the N-termini. In the crystal structure the first two N-terminal residues were not observed. The second largest displacement is found at the linker region between the post-SET and CTD. This non-conserved region has a variable length in SMYD proteins. This region has been proposed to act as a hinge for inter-domain movement [24, 39]. Large motion is also observed for parts of the MYND and CTD. In CTD, the most dynamical regions are the linker regions between the pairs of up-down helices. In MYND, the variable regions are the N- and C-terminus of the kinked helix αA and a loop between β5 and β6. In SET-I, the most dynamical regions are the end of the first helix (αB) and the beginning and end of the second helix (αC), and the loop forming the cofactor-binding site is relatively static.

2. Correlated Inner-Lobe Motion

The CTD and N-lobe show strong negative correlated dynamics (Fig 1D). The regions in the CTD involved in such a correlation include residues 300–315, 337–360, and 390–400. These regions are among the most dynamical regions in the structure (Fig 1B) and located at the inner surface of the C-lobe (Fig 1E). The correlated regions in the N-lobe are divided into seven zones spanning from residues 40–200. The zones include the following regions: (I) residues 41–53; (II) 68–73; (III) 85–95; (IV) 100–115; (V) 150–160; (VI) 183–185; (VII) 195–200. These zones spread

across the MYND, SET-I and part of the SET domain but are clustered at the inner face of the N-lobe. As a result, the two correlated sets of residues are facing each other across the gap of the N- and C-lobes (Fig 1E). The residues in each set show positive correlated intra-lobe dynamics, whereas the two sets are related by the negative correlated inter- lobe dynamics (Fig 1D and 1E). During the simulation the contact distances between the residue-pairs of the two sets vary significantly (Fig 1F). The level of variation is two times above the average variation. This indicates a relatively large distance variability between the N- and C-lobes. This together with the negative correlated inter-lobe dynamics suggests a possible clamshell-like motion or open–closed motion between the lobes. The distances of the two representative residues, Y311 in the C-lobe and G46 in the N-lobe, range from 8.7 to 16.0 Å during the simulation. The fluctuated pattern of the distance indicates a slightly open and closed con- formation (Fig 1G).

3. Principal Component Analysis

To further understand SMYD2 correlated dynamics, principal component (PC) analysis was performed using Cα position covariance (Fig 2A). The first PC accounts for more than one fourth of the overall variance. The second PC accounts for 10%. The first three components together account for 45%. The individual component contributions afterward drop below 6%. The first PC describes a twisting motion of the CTD with respect to the N-lobe and a spring- bending motion within the MYND (Fig 2B). The second PC is dominated by a clamshell-like motion between the N- and C-lobes. It is therefore that the variance in the PC1–PC2 plane essentially dictates the negative correlated inter-lobe dynamics. Based on these variances, the conformers throughout the simulation were grouped into four clusters using k-means algorithm (Fig 2C). The number of clusters was chosen based on the “elbow criteria”. At a cluster count of four the BSS/TSS (Between-group Sum of Squares/Total Sum of Squares) ratio is 82.8%. Similar

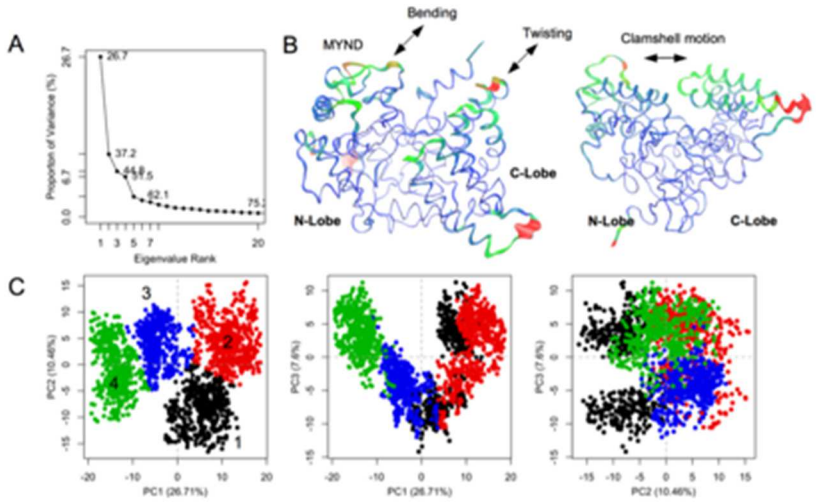

Figure 14. Principle component analysis

(A) Scree plot showing the proportion of variance against its eigenvalue rank. (B) Visualization of the motions along PC1 (left) and PC2 (right). The most dissimilar structures along a given PC are depicted by thicker coils. The interpolated structures produced by Bio3D [104] are shown by thinner coils. Color scale from blue, green, to red depicts low to high atomic displacements. (C) Projection of the trajectory onto the planes formed by the first three principle components. Conformers are colored according to the k-means clustering: cluster 1, black; 2, red; 3, blue; 4, green.

clustering was obtained using hierarchical clustering algorithm (data in S2 Fig). Cluster 1 populates in the first 0.25 ns and between 0.7 and 0.9 ns. Cluster 2 is intertwined with Cluster 1 (0.25–0.7 ns and 0.9–1.0 ns). Cluster 3 is sampled in a time window of 1.0–1.5 ns. Cluster 4 lasts for the remainder of the simulation. This PC1–PC2 plane-based clustering appears to correlate with the pattern of distance fluctuation between Y311 and G46 (Fig 1G). The Y311–G46 distance represents the distance between the N- and C-lobes or the open/closed state of the structure. Cluster 1 and 4 correspond to the closed state, while cluster 2 and 3 sample the open one.

4. Dynamical Network Analysis

Dynamical network analysis was performed to define the allosteric paths for SMYD2 correlated inter-lobe dynamics. This analysis revealed nine communities in the dynamical structural net- work (Fig 3A). The community assignment is roughly correlated with the sequence-

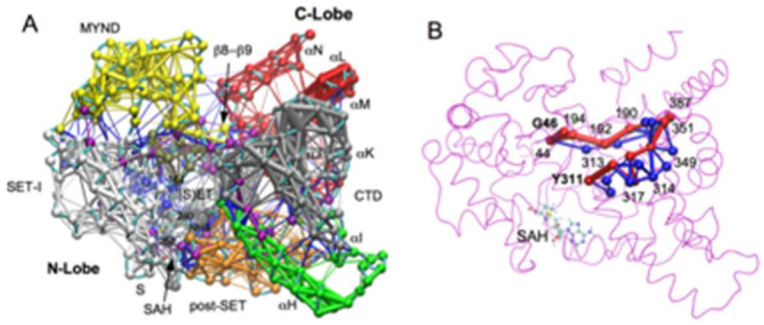

Figure 15. Dynamical network analysis
(A) SMYD2 dynamical network. The network is colored according to communities. Points in the network are nodes, and lines between the nodes represent edges. The thicker lines depict the stronger edges or stronger correlations. Critical nodes are colored in purple. (B) Optimal and suboptimal paths between Y311 and G46. The optimal path is colored in red and suboptimal paths in blue. The edge thickness is weighted by the number of suboptimal paths crossing the edge. Residues along the optimal path are labeled.

and structure-based domain assignment [35, 39]. The SET is split into two communities largely corresponding to the S-sequence and core SET. The cofactor product SAH is associated with the S-sequence community. This indicates a stronger correlated motion between SAH and the N-terminal S-sequence. The S-sequence has been shown to be involved in cofactor binding [24, 35, 39]. Mutation of two Gly residues in this sequence abolished SMYD1 enzymatic activity [35]. The SET-I, which is also involved in cofactor binding, forms a separate community. The MYND, a protein–protein interaction module, forms another community. There is a separate community formed at the interface of SET, MYND, and SET-I. This community contributes the residue Phe184 to the target lysine access channel. However, other two aromatic residues (Y240 and Y258) in this channel belong to the S-sequence-containing community. Most of residues in the post-SET belong to one community. This community also contains the residue H207 from the conserved active site motif NHS. H207 chelates the zinc atom of the post-SET, which may result in such an association. Other two residues (N206 and S208) in the NHS motif belong to the S-sequence-containing community. In CTD there are three communities formed by the first helix

(αH), αI–αJ–αK, and αL–αM–αN. The predicted Hsp90 binding site is located between the second and third communities, which is also the extended ribosomal binding site[24, 57, 94].

The communication between network communities is mediated through critical nodes [108]. Such nodes are important for allosteric signal transduction and dynamical correlation between the communities[108, 121]. Of note, the β8–β9 hairpin contains four critical nodes (residues 190–193). These nodes form a bridge connecting the N- and C-lobes (Fig 3A). Two of these nodes (residues 190 and 191) have direct interaction with the CTD. Disrupting this interaction has been found to reduce SMYD2 methyltransferase activity [58]. This suggests that the β8–β9 hairpin may represent an optimal path for dynamical inter-lobe communication. The optimal and suboptimal paths were generated between Y311 and G46. As mentioned earlier, these pair of residues move in concert toward the opposite direction. Their dynamical relationship can represent the open and closed state of the structure and correlated inter-lobe dynamics. The optimal path between the two residues passes through the β8–β9 hairpin (Fig 3B). All suboptimal paths also pass through the hairpin. The β8–β9 hairpin occurring in the highest number of suboptimal pathways may thus be necessary to guarantee an effective pathway for inter-lobe communication.

Discussion

The crystal structures revealed that SMYD proteins adopt different CTD conformations[24, 35, 39, 57, 94]. SMYD1 has an open CTD structure with the substrate-binding cleft completely exposed. SMYD3 has the narrowest substrate-binding cleft due to the direct CTD–N-lobe interaction [39]. SMYD2 is like a conformational intermediate, and when different cofactors bound, the CTD exhibits different conformations [24]. These data have suggested the dynamical nature of the CTD and a possible open–closed motion of the two lobes [94]. Our MD simulation of SMYD2 structure supports an open–closed motion. The simulation reveals that

SMYD2 exhibits a negative correlated inter-lobe dynamics, and this correlated dynamics is described by a twisting motion of the CTD with respect to the N-lobe and a clamshell-like motion between the lobes. Correlated inter-domain motions may mediate fundamental protein functions such as substrate recognition [119]. In SMYD2 the substrates bind to the protein in a U-shaped conformation[57, 58]. Both the N- and C-lobe contribute to the binding. The inter-lobe dynamics will alter the size of the substrate-binding site. The coupling of the two lobes by the correlated motion might offer the specificity and promiscuity for substrate recognition. Correlated inter-domain motions are also important for allostery [118]. In SMYD2 the cofactors exhibit allosteric effects. Binding of sinefungin and SAH to the cofactor-binding site in the N-lobe caused a structural difference in the CTD [24]. Such a long-range structural effect could not be explained by the crystallographic studies [24], but the correlated inter-lobe dynamics might provide a signal transduction pathway enabling a long-range domain–domain communication.

A complex mechanism regulates SMYD biochemical function. Binding of Hsp90 to the CTD significantly enhances the activity of SMYD proteins[4, 41, 42]. For SMYD2, Hsp90 binding not only increases the activity but also changes the substrate specificity [41]. Both SMYD2 and SMYD3 interact with DNA[4, 63]. DNA binding to the MYND has been shown to enhance SMYD3 enzymatic activity [64]. The mechanism of such an activity enhancement is unknown, but one possible mechanism is that the binding may affect the domain dynamics and inter-lobe dynamical correlation. Such an effect could be transduced to other parts of the protein through the critical nodes bridging the communities, which in turn might impact substrate binding and cofactor binding.

Studying SMYD2 conformational dynamics is of therapeutic interest. Dynamical information of SMYD2 structure would facilitate receptor-flexibility-enabled drug design. The

conformational states sampled by the MD simulation can be used in ensemble docking. In addition, the identification of the critical nodes and optimal path mediating the dynamical network communication could offer new strategies to manipulate SMYD2 function. Disrupting a specific network communication could represent a rational approach for the design of drugs with improved potency and selectivity. In summary, the MD simulation of SMYD2 structure has revealed that SMYD2 exhibits a negative correlated inter-lobe dynamics and provided additional insight into the structure of this multifunctional protein lysine methyltransferase.

PART II. ALTERNATIVE STRATEGIES TO INHIBIT POST-TRANSLATIONAL MODIFICATIONS

CHAPTER 4 CRYSTAL STRUCTURE OF SMYD2 – PARP1 COMPLEX, A NOVEL BINDING SITE IDENTIFIED.

Introduction

Poly (ADP-ribose) polymerase 1 (PARP1) is a mammalian, multifunctional enzyme involved in DNA repair, genome stabilization and chromatin restructuring [122-125]. Upon DNA damage, PARP1 serves as a first responder to recognize single and double stranded breaks (SSB and DSB), and recruits proteins necessary for repair and signaling the DNA damage response (DDR) pathway. Protein recruitment towards damaged DNA sites is made possible by the ADP-ribosyl transferase activity of PARP1, which serves to transfer linear or branched moieties of ADP-ribose upon itself and target nuclear proteins [126-129]. PARP's basal enzymatic activity is low, but activity can be enhanced through DNA binding, protein binding partners and post-translational modifications [127, 130-133]. Recently, a study found PARP1 is methylated at K528 by SMYD2, and methylation of K528 stimulated ADP-ribosyl activity and the DNA damage response [52].

SMYD2 (SET and MYND domain-containing protein 2) is a protein lysine methyltransferase reported to be involved in cardiomyocyte function, macrophage activation and cancer progression [14, 16-18, 51, 134]. The molecular mechanism of SMYD2 in relation to these responses have been fairly documented. One of the earliest studies found SMYD2 to methylate p53 at lysine 370 impairing p53's function to promote p21 and MDM2 expression [3]; however, SMYD2's anti-apoptotic role is not solely dependent on methylating p53 [16, 19]. Indeed, later studies found SMYD2 to methylate a library of protein substrates including Hsp90, estrogen receptor α (ERα), retinoblastoma (RB1), MAPKAPK3, all of which yield unique functional ends

[2, 19, 50, 52, 71]. Understanding how SMYD2 recognizes these reported substrates at the molecular level remains a challenge.

In order elucidate the mechanism of SMYD2 methylation and regulation, several efforts were made to understand the structure and biochemistry of SMYD2. Biochemical studies reported weak methyltransferase activity [61]. A possible explanation for low activity could be explained by the accessibility of the substrate channel. Crystal structures of SMYD2 demonstrate the TPR-like domain (C-lobe) appears to shade over the substrate channel and limit the space for peptide or protein entry . At the moment, two SMYD2-peptides (p53 and ERa) structures have been reported. Both peptides offer similar yet unique structures within the substrate channel offering insight into the multiple configurations a substrate can be recognized by SMYD2 [54, 57, 58].

In this study, we solved the crystal structure of SMYD2 in complex with PARP1 peptide. Unexpectedly, we were able to map only four amino acids in the primary site, including the target lysine and P-1 leucine. Since most of the residues in the peptide were unstructured, this may provide supporting evidence that the target lysine and P-1 leucine are necessarily and significant for SMYD2 substrate recognition and methylation [135]. However, we identified a second peptide bound in a separate site that appears to wedge between the N- and C-lobe of SMYD2. The function of this unexplored site may have implications in trafficking SMYD2 substrates into the substrate channel.

Materials and Methods

Protein Expression and Purification

The open reading frame of human SMYD2 was cloned into pCDF-SUMO vector and transformed into BL21 (DE3) cells for recombinant protein expression. Cells grew to an optical density of 0.4 – 0.6 and protein expression was induced with 0.1 mM IPTG. Growth continued

overnight at 15°C. Harvested cells were suspended in binding buffer (20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, 20 mM imidazole, 5 mM BME) and lysed using a French Press, and the supernatant was collected for purification. His_6-SUMO-SMYD2 was captured on a HisTrap column (GE Healthcare) and eluted out by applying an elution buffer (20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, 500 mM imidazole, 5 mM BME) gradient. Fractions containing His_6-SUMO-SMYD2 were pooled and dialyzed in binding buffer. Ulp1 was added to cleave off the His_6-SUMO tag. The native SMYD2 protein was collected by passing the protein pool through the HisTrap column. Finally, SMYD2 was further purified through a Superdex 200 column (GE Healthcare) into gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol, 5 mM BME). Fractions containing pure SMYD2 were pooled, concentrated to 20 mg/mL and stored in the -80°C.

Crystallization, Data Collection and Structure Determination

All crystals were prepared using the hanging drop method at 20°C. Crystal seeds were prepared by incubating 10 mg/mL SMYD2, 1 mM PARP1 peptide (RMKLTLKGGAAVD), and 600 μM SAH in buffer (20 mM HEPES 7.5, 50 mM NaCl) for 2 hours at 4°C and crystallized in a 1:1 volume of protein and crystallization solution (20% PEG 3350, 0.1 M Tris pH 7.5 and 5% ethanol). Seeds were crushed and crystallized again by supplementing 1.5 mg/mL SMYD2 and 1 mM PARP1 peptide into the crystallizing drop. Crystals typically achieved full size after one week. SMYD2-PARP1 crystals were harvested and flash-frozen in liquid nitrogen. X-ray diffraction data was collected at Advance Photon Source (beamline 21-ID-G) and images were processed using XDS and AIMLESS [136, 137]. Crystals were indexed to C121 space group containing 2 molecules per asymmetric unit. Phases were obtained through molecular replacement using human

SMYD2 structure (PDB ID: 5KJK) as a search model. Several rounds of model building and refinement were carried out in Coot and PHASER, respectively [138, 139].

Isothermal Titration Calorimetry

ITC titrations were carried out on a NanoITC SV calorimeter (TA Instruments) at 27°C whilst stirring at 375 rpm. All experiments were performed with 16 (15 μL) injections spaced 200 seconds apart. The first (3 μL) injection was excluded from modeling. Protein samples were dialyzed in the same buffer (25 mM Tris pH 7.5, 50 mM NaCl) and PARP1 and p53 peptides were dissolved in the buffer dialysate. The sample cell was loaded with approximately 1.2 mL of 40 μM SMYD2 solution, and the injection syringe was loaded with approximately 250 μL of 631 μM PARP1 peptide (RMKLTLKGGAAVD) solution. Protein and peptide concentrations were quantified using a Direct Detect (Millipore Sigma). Data was processed and fitted in NanoAnalyze (TA Instruments).

Results

Overall structure of the SMYD2-PARP1 complex

The SMYD2 structure in complex with PARP1 and s-adenosylhomocysteine (SAH) was solved at 2.1Å using molecular replacement. The overall fold of SMYD2 possesses a bilobal structure with a deep crevice separating the two lobes. The N-lobe (residues 5 – 278) is divided into four domains: SET, MYND, SET-I and post-SET. The SET, SET-I and post-SET domains associate to form an evolutionary conserved fold that make up the core SET domain structure. The MYND domain possesses a zinc-finger fold that sits outside of the SET domain. This domain doesn't contribute to substrate or cofactor binding but has been reported to contribute binding towards EBP41L3 through a PXLXP motif [41]. The cofactor, SAH sits inside a deep pocket

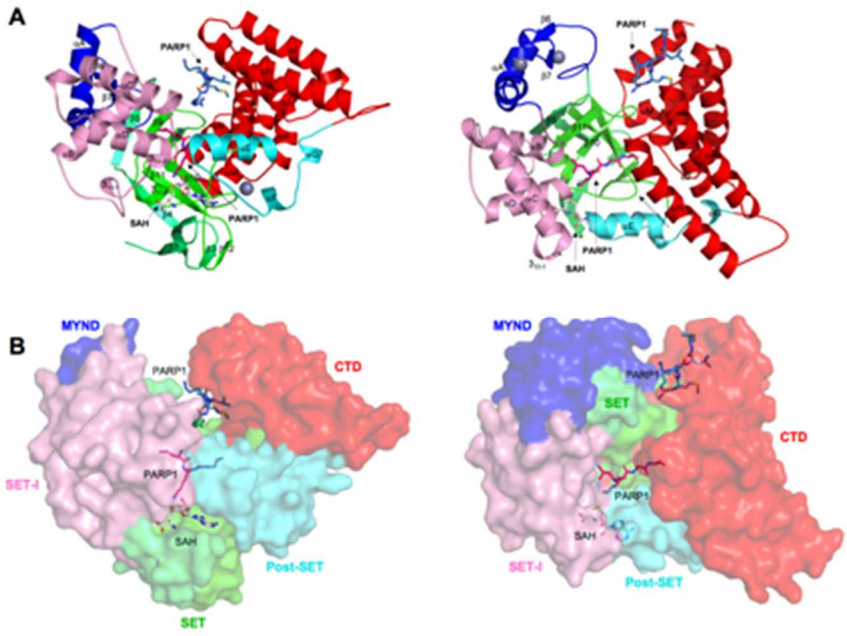

Figure 16 Overall structure of SMYD2-PARP1 complex
(A) Ribbon and (B) surface representation of SMYD2-PARP1 complex. The S-sequence, MYND, SET-I, core SET, post-SET, and CTD are colored in lime green, blue, pink, green, cyan, and red, respectively. Secondary structures are labeled based on their position in the sequence. PARP1 peptides inside substrate and secondary binding site are colored in magenta and sky blue, respectively.

formed the by SET (β1-β2) , SET-I ($3_{10\text{-}1}$–$3_{10\text{-}2}$) and post-SET domains (αE-αF) and essentially binds as previously described from another report [57]. The C-lobe or C-terminal domain (CTD) (residues 279-433) folds similarly to tetratricopeptide repeat (TPR) domains [35, 140]. The purpose of the CTD in SMYD proteins is unknown, but reports suggest the CTD contribute towards substrate and Hsp90 binding [54, 57, 58, 102]. Previous reports observed a significant increase in histone methylation from SMYD1, SMYD2 and SMYD3 in the presence of Hsp90; therefore, the CTD may play a role in modulating SMYD methylation [4, 41, 42]. Two binding sites of the PARP1 peptide were identified, one peptide bound in the active site and a second peptide bound between the N-lobe and the C-lobe. We named this area, “secondary binding site,” (SBS) since we found a second version of the PARP1 peptide.

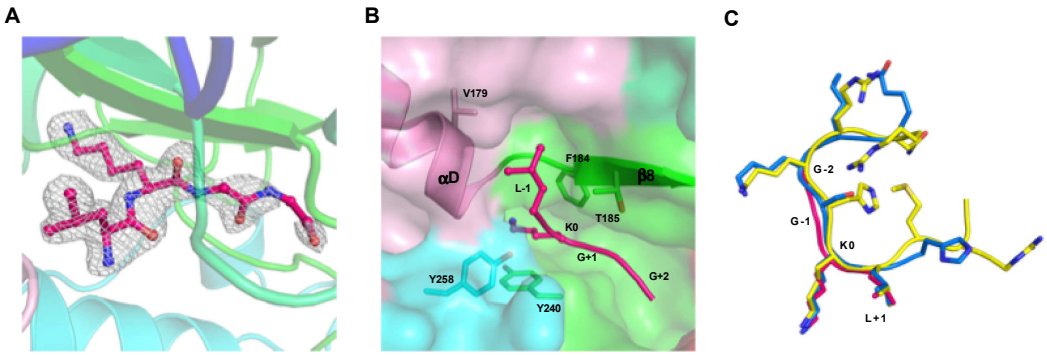

Figure 17. Substrate binding site.
A) 2Fo-Fc map of the PARP1 peptide in the substrate binding site was calculated at 2.1 Å and contoured at 1.0 s. PARP1 peptide is shown as ball and sticks. (B) Surface-stick representation of SMYD2-PARP1 complex. SMYD2 domains are colored according to Figure 16B. (C) Ribbon alignment of previously structured peptides. PARP1, p53 and ERα peptides are colored in magenta, blue and yellow respectively. K0 refers to the target lysine. L-1, G+1, and G+2 refer to the amino acid positions flanking the target lysine.

Structure of the PARP1 peptides in the substrate and secondary binding sites

The PARP1 peptide in the substrate binding site has a similar structure as previously reported p53 and ERα peptides [54, 57, 58]. The target lysine side chain sits inside the lysine access channel which is created by F184, Y240 and Y258, and the carbonyl oxygen of target lysine forms a hydrogen bond with T185 amide nitrogen. Leu-1 fits into a concave pocket formed by V179, T185 and the main chain loop between αD helix and β8 strand. Gly+1 and Gly+2 mimic the p53 and ERα peptide main chains. There is no electron density beyond Gly+2 and Leu-1 due to the lack of an electron density map. This may suggest residues beyond these positions do not significantly contribute towards binding to SMYD2; therefore, these resides were disordered in the crystal. This explanation can be highlighted by a previous report that SMYD2 substrate recognition is dependent on the P-1 and P+1 positions where P-1 must be leucine

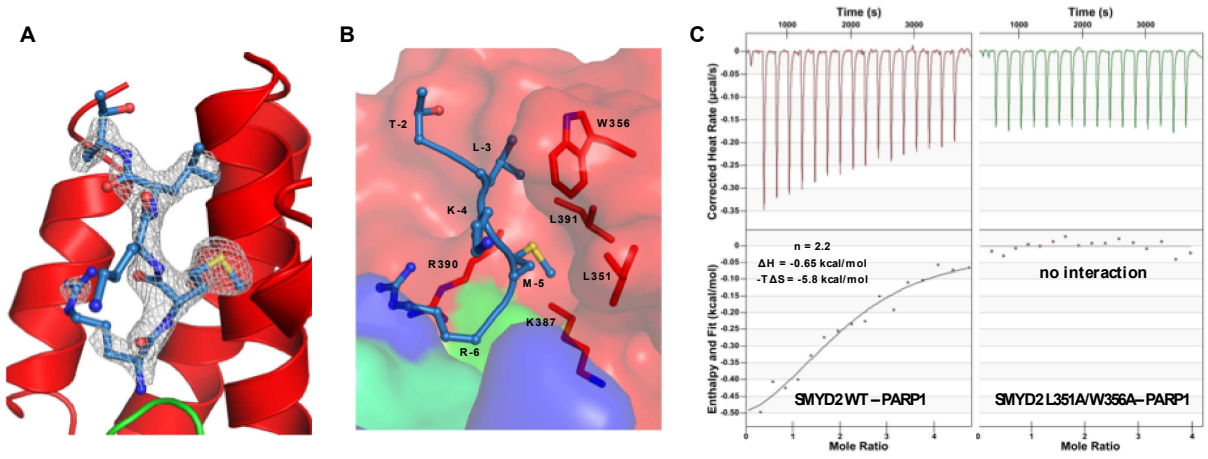

Figure 18. A second peptide identified in the C-terminal domain.
(A) Refined 2Fo-Fc map of the PARP1 peptide in the secondary binding site was calculated at 2.1 Å and contoured at 1.0 s. (B) Surface-stick representation of PARP1 (sky blue)and SMYD2 (red) interaction in the secondary binding site. Amino acid positions Thr-2 through Arg-6 refer to the positions flanking the target lysine in the PARP1 peptide (C) ITC of SMYD2 WT (red) and SMYD2 L351A/W356 mutant (green). Top panel displays corrected heat injections data. Bottom panel displays isotherm plot.

and the P+1 must be a small amino acid. Residues beyond these limits may aid in SMYD2 binding but are not necessary as hinted by the lack of an electron density map (Figure 17A).

The PARP1 peptide in the secondary binding site is nested into a small fissure in the CTD created by αJ, αK and αL. This interaction is hydrophobically driven by two residues in PARP1, Leu-3 and Met-5. Leu-3 packs against residues W356, L391 and G394. Met-5 fits into a hydrophobic crevice created by L351, K387, R390, and L391. An electron density map couldn't be mapped for the Lys-4 side chain which suggests Lys-4 adopts multiple orientations and doesn't significantly contribute towards SMYD2 binding (Figure 18A and B).

Two binding sites confirmed by Isothermal Titration Calorimetry.

In order to confirm two binding sites, ITC was performed to measure the stoichiometry with high accuracy. Titration of the PARP peptide (RMKLTLKGGAAVD) into full length human SMYD2 generated exothermic injections. Data was fit into an independent model with n = 2.21 ±

0.28. Measured enthalpy and dissociation constant were -0.65 kcal/mol and 28 μM, respectively. Calculated Gibb's energy and entropy (-TΔS) were -6.4 kcal/mol and-5.8 kcal/mol, respectively. The high entropic term, nearly nine fold higher than the enthalpic term, suggest peptide binding between SMYD2 and PARP1 is hydrophobically driven. This observation is consistent with the complex structure of SMYD2 and PARP1 peptides since both sides appear to be primarily driven by hydrophobic contacts (Figure 2B and 3B).

To confirm the PARP1 peptide was specifically binding in the secondary binding site, we mutated L351A and W356A in SMYD2 and purified it similarly as the wild type with no observable complications in mutant solubility. The same ITC experiment was performed on SMYD2 mutant (L351A and W356A.) Unexpectedly, no heat change was observed suggesting PARP1 didn't bind to the secondary binding site and substrate channel. This observation would suggest peptide binding to the substrate binding site requires a peptide bound to the secondary binding site. It's unclear whether peptide binding is sequential where the peptide from the secondary binding site is channeled to the active site or peptide binding in both sites are independent. Sequential binding is conceivable as this could influence the TRP-like domain to open the structure and tunnel the substrate towards the active site. Further work will be required to evaluate possible links between the substrate and the secondary binding sites.

Discussion

Crystal structures of SMYD proteins demonstrate the N- and C-lobe can potentially change [24, 35, 39]; however, the mechanism at which these lobes undergo bilobal conformational changes remains elusive. Additionally, the mechanism at which SMYD2 methylation clients are guided into the buried substrate channel is unclear. In this study, we found a peptide substrate predominantly bound to an new and unexplored area of the CTD in our SMYD2-PARP2 complex

crystal structure. Our ITC results demonstrate SMYD2 can bind to two PARP1 peptides; therefore, the second peptide bound to the secondary binding site doesn't appear to be a non-specific artifact. Interestingly, titration of the PARP1 peptide into SMYD2 L351A/W356A mutant didn't generate heat suggesting the PARP1 peptide binding into the substrate channel may be sequential, requiring PARP1 to bind to the secondary binding site first before PARP1 is guided into the substrate channel.

To our knowledge, this study presents the first SMYD structure that binds a second peptide outside of the substrate channel. We imagine this new peptide binding site may recognize a hydrophobic motif (MXL or ΦXΦ) within a substrate protein or perhaps protein chaperone that isn't a SMYD2 methylation client. Hydrophobic motifs binding to the secondary binding site could potentially change the dynamics of the C-lobe in SMYD2; thereby, allowing the substrate channel to become more accessible for methylation targets to enter. Since purified human SMYD2 methyltransferase activity is low [61], we predict the secondary binding site as an allosteric modulator that drives the C-lobe to open for substrate binding. To test this hypothesis, further work will require measuring the activity of SMYD2 L351A/W356A mutant to understand how the secondary binding site influences methylation of PARP1. At the moment, this study opens a new avenue for understanding how SMYD2 methylation clients are recognized and tunneled for lysine methylation.

Table 1. Crystallographic data and refinement statistics

Data	
Space group	*C*121
Cell parameters	
a, b, c (Å)	142.8, 52.2, 144.9
Wavelength (Å)	0.97856
Resolution (Å)	133.2-2.10 (2.21-2.1)[a]
R_{merge}[b]	0.105 (0.812)
$CC_{1/2}$	0.995 (0.772)
Redundancy	7.3 (7.4)
Unique reflections	51097
Completeness (%)	88.5 (96.7)
$\langle I/\sigma \rangle$	13.7 (2.5)
Refinement	
Resolution (Å)	66.6-2.10 (2.14-2.10)
Molecules/AU	2
R_{work}[c]	0.199 (0.246)
R_{free}[d]	0.234 (0.293)
Ramachandran plot	
Residues in favored	97.8%
Residues in allowed	2.2%
RMSD	
Bond lengths (Å)	0.003
Bond angels (°)	0.69
No. of atoms	
Protein	7290
Peptide	132
Water	154
B-factor (Å^2)	
Protein	43.8
Peptide	73.0
Water	33.9

[a]Numbers in parentheses refer to the highest resolution shell.

[b]$R_{merge} = \Sigma|I-\langle I\rangle| / \Sigma I$, where I is the observed intensity and $\langle I\rangle$ is the averaged intensity of multiple observations of symmetry-related reflections.

[c]$R_{work} = \Sigma|F_o-F_c| / \Sigma|F_o|$, where F_o is the observed structure factor, F_c is the calculated structure factor.

[d]R_{free} was calculated using a subset (5%) of the reflection not used in the refinement.

CHAPTER 5 EXPRESSION, PURIFICATION AND ACTIVITY STUDIES OF THE CATALYTIC DOMAIN OF USP10

Introduction

Ubiquitin-specific protease 10 (USP10) is an widely expressed deubiquitinase that plays a central role in maintaining p53 homeostasis, DNA repair and apoptosis [28, 141, 142]. USP10 was first identified in associate with G3BP which served to inhibit the deubiquitinase function of USP10. Later studies expanded USP10's role as a specific deubiquitinase for p53, histone variant H2A.Z, CFTR, AMPKα and FLT3 in various cellular pathways [28, 30, 142-145]. While the biological mechanism of USP10 appears to be diverse, one pathway of interest is USP10's involvement in maintaining p53 homeostasis. One study found that USP10 serves dual roles as a oncogene and tumor suppressor, depending on the oncogenic nature of p53. Under normal physiological conditions, USP10 serves as a tumor suppressor by preserving p53 homeostasis and counteracting p53 ubiquitination by Mdm2; however, in a p53 mutant environment, the role of USP10 reverses to an oncogene by stabilizing oncogenic p53 [28]. Inhibition of USP10 through RNA knockdown experiments greatly reduced cellular growth in mutant p53 renal cancer cells [28, 30].

Alternatively, USP10 was also found to inhibit p53 function though a positive feedback loop promoting androgen receptor-mediated transcription and G3BP expression. Since G3BP serves as an inhibitor of USP10 function, this resulted in repressing p53 function and promoting proliferation in G3BP2-overexpressed prostate cancer cells. Recently, USP10 was found to play a key role in stabilizing mutant FLT3 in acute myeloid leukemia (AML) cells. Mutagenesis of USP10's catalytic activity (C424S) resulted in reduced stabilization of mutant FLT3 in MOLM14 cells [30]. Therefore, designing small molecule inhibitors that specifically target USP10 may prove

to be a novel strategy for treating patients with mutant p53 cancers, mutant FTL AMLs and androgen receptor-dependent prostate cancers.

Alas, the atomic structure of USP10 and selective inhibitors for USP10 are currently lacking. One group identified two small molecule inhibitors that target USP10 activity *in vitro*; however, these compounds also target a related p53 deubiquitinase, USP7 [146, 147]. Another group identified spautin-1 as a selective inhibitor against USP10 and USP13 deubiquitinating function [148]. Recently, a mutant ubiquitin variant was reported to specifically inhibit USP10 *in vitro* which serves as an excellent aid to study USP10 for cellular and biochemical studies [149]. All of the USP10 inhibitors were reported by separate groups, and a collective study of these inhibitors is needed.

In this study, we explore the biochemistry of the catalytic domain of human USP10 and examine the potency of reported inhibitors: UbV10, spautin-1 and P22077. Our goal was to solve the crystal structure of the catalytic domain of USP10 which would serve as a template for virtual screening. In addition, comparing the inhibitory activity of reported inhibitors against USP10 can aid in the development of a fragment-based screen to improve more potent and soluble inhibitors. A selective inhibitor against USP10 would serve as a valuable tool to study USP10 overexpression in oncogenic p53 cancers and oncogenic FLT3 acute myeloid leukemias.

Materials and Methods

Protein Expression and Purification

The core deubiquitinase catalytic domain of USP10 (S374 – L798) was cloned into pCDF-SUMO vector and UbV10 construct was synthesized into pET151 vector (Invitrogen). Vectors were transformed into BL21 (DE3) cells for recombinant protein expression. Cells grew to an

optical density of 0.4 – 0.6 and protein expression was induced with 0.1 mM IPTG. Growth continued overnight at 15°C. Harvested cells were suspended in binding buffer (20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, 20 mM imidazole, 5 mM BME) and lysed using a French Press, and the supernatant was collected for purification. His_6-SUMO-USP10(374-798) was captured on a HisTrap column (GE Healthcare) and eluted out by applying a gradient of elution buffer (20 mM sodium phosphate, 500 mM NaCl, 5% glycerol, 500 mM imidazole, 5 mM BME). Fractions containing His_6-SUMO-USP10-CA were pooled and dialyzed in binding buffer. Ulp1 was added to cleave off the His_6-SUMO-USP10-CA tag. The native USP10-CA protein was collected by passing the protein pool through the HisTrap column. Finally, USP10-CA was further purified through a Superdex 200 column (GE Healthcare) into gel filtration buffer (20 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol, 5 mM BME). Fractions containing pure USP10-CA were pooled, concentrated to 20 mg/mL and stored in the -80°C. UbV10 was purified similarly, except the N-terminal His_6 tag was removed by TEV protease.

Isothermal Titration Calorimetry

ITC titrations were carried out on a NanoITC SV calorimeter (TA Instruments) at 28°C whilst stirring at 300 rpm. Experiments were performed with 23 (10 µL) injections spaced 300 seconds apart. The first (2 µL) injection was excluded from the data. USP10-CA and UbV10 were dialyzed in the same buffer (20 mM HEPES pH 7.5, 150 mM NaCl) overnight. The sample cell was loaded with approximately 1.2 mL of 25.5 µM USP10-CA, and the injection syringe was loaded with approximately 230 µL of 283 µM UbV10. Protein concentrations were verified using a Direct Detect (Millipore Sigma). Data was processed and fitted in NanoAnalyze (TA Instruments).

In vitro Deubiquitinase Activity Assay

Fluorescent activity assays were carried out on 384-well plates (Corning black flat/bottom) using a SpectraMax M5 (Molecular Devices). Ubiquitin-Rhodamine 110 (Ub-Rho) was used as a fluorescent substrate to monitor USP10-CA hydrolysis and liberation of Rhodamine 110. The final reaction buffer was 50 mM HEPES 7.5, 0.5 mM EDTA and 5 mM DTT. A typical experiment setup contained 200 nM USP10 and serial concentrations of Ub-Rho (0 – 30 μM) and UbV10 (0 – 10 μM) in a final volume of 25 (kinetic studies) or 30 μL (inhibitory IC_{50}/K_i studies). USP10-CA, UbV10, P22077 and Ub-Rho were diluted in reaction buffer and incubated at 30°C for at least 10 minutes. Fluorescence readings were recorded for 2 hours. Initial velocity was measured for the first 5 minutes. Data was fit in GraphPad.

Results

Cloning and Purification of the catalytic domain of USP10.

The full length USP10 protein was previously expressed and purified in *E.Coli* with no success [143]. Since we are specifically interested in the inhibition of USP10 deubiquitinase function, we sought to localize and purify the catalytic core of USP10 using *E.Coli* as a production host. In order to correctly define and clone USP10-CA, a sequence alignment was made across select, structured USP homologs and USP10 orthologs using COBALT and ESPript (Figure 19A and 19B) [150, 151].

In addition, a homology model of USP10-CA was made using SWISS-MODEL using USP8 (PDB ID: 5N3K) as a template with a sequence identity of 26% [120]. The overall structure and catalytic mechanism of USP10 should be expected to be similar to other USPs due to the homogenous spread of conserved residues within the model (Figure 19C). USP10 is highly conserved among orthologs; therefore, we designed the USP10-CA construct from S374-L798 that

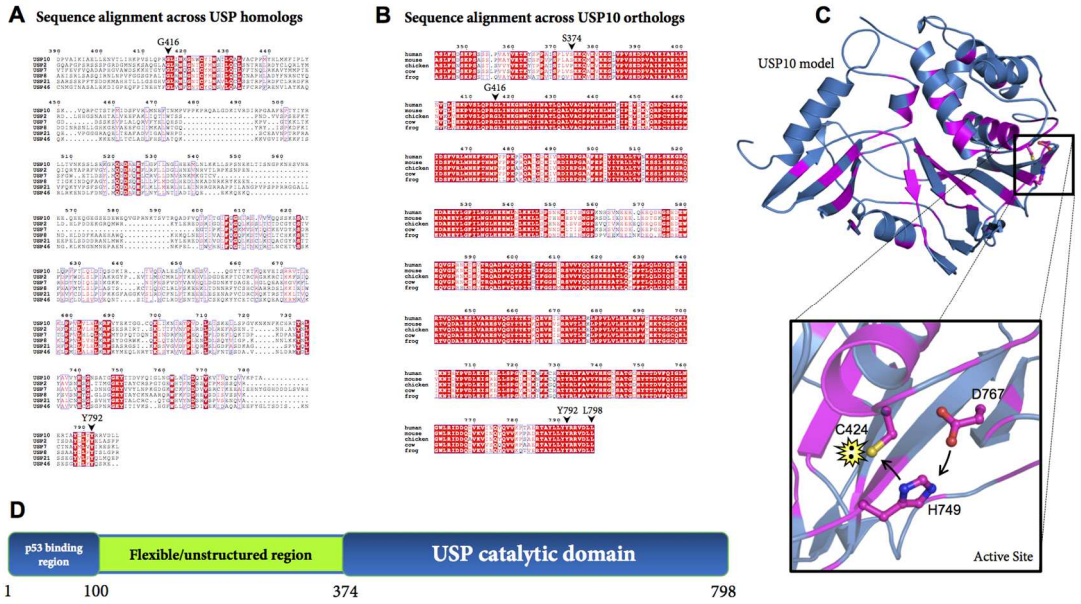

Figure 19. Sequence alignment of USP homologs and USP10 orthologs
(A) Sequence alignment of select USP homologs and (B) Sequence alignment of USP10 orthologs. Identical residues are highlighted in red in white font. Positive residues are boxed in red font. (C) SWISS model of USP10 using USP8 as a model base. Unstructured loops were excluded from the model. Conserved residues among select USP homologs are colored in magenta. Outlined box: Proposed mechanism for USP10 hydrolysis. (D) Open reading frame of human USP10.

incorporated most of the conserved regions within the active catalytic domain. After we successfully cloned and purified our catalytic domain construct, a separate group also reported purifying a similar USP10-CA construct, K376 – L798. [30].

The catalytic domain was expressed and purified using His_6-SUMO fusion tag adjacent to the N-terminal end of USP10. After capturing His_6-SUMO USP10-CA from Ni^{2+} column, the fusion tag was successfully removed from USP10-CA using Ulp1 protease. Several additional purification steps were needed to prepare the protein pool for a second round pass through the Ni^{2+} column, and we were able to separate the His_6-SUMO tag from USP10-CA (Figure 20A). USP10-CA was injected into a size exclusion column and eluted out homogeneously as a ~50 kDa monomer (Figure 20B) Altogether, soluble USP10-CA was obtained with high purity (Figure 20C).

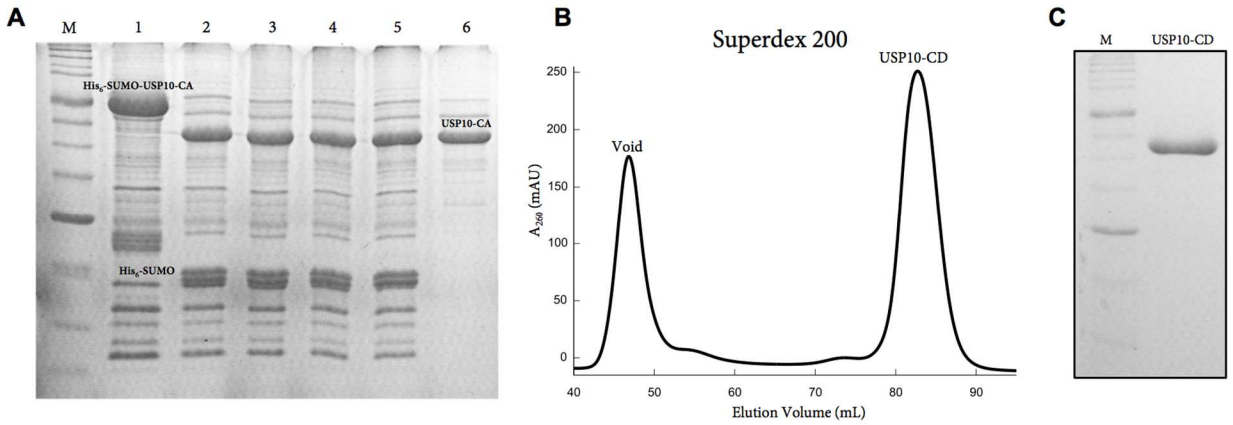

Figure 20 USP10 catalytic domain purification
(A) SDS-PAGE of SUMO-tag purification. 1: Pooled eluted fractions from containing His6-SUMO-USP10-CA. 2: Supplementing Ulp1 protease cleaved into USP10-CA and His_6-SUMO. 3| Sample after concentrating protein pool. 4: Sample after filtration. 5: Sample after dialysis to binding buffer. 6|:Pooled flowthrough and wash fractions after injecting the protein pool into a second HisTrap pass. (B) HiLoad 16/60 chromatogram of USP10-CA. USP10-CA elutes out as a homogenous monomer. (C) SDS-PAGE of the final purification product of USP10-CA.

In order to structure USP10-CA, we sought to crystallize USP10-CA, but our efforts to obtain well-ordered crystals were unsuccessful. Possible crystallization complications could arise due to large and distorted loop insertions (Figure 19A). To our knowledge, only four USP catalytic domains have been crystallized in the apo form, solved and deposited in the PDB. Most of the USP proteins are crystallized and structured as a complex with ubiquitin or a ubiquitin-derivative. This observation suggests the catalytic domains of USPs are challenging to crystallize. One possible explanation could be due to the conformational fluidity between active and inactive states as observed from the apo structures of USP7, USP8 and USP14 [152-154]. When USPs are bound to ubiquitin, this could limit the USP's conformational state and aid in protein crystallization. We sought to crystallize USP10-CA in complex with ubiquitin molecule in order to stabilize USP10 into a single conformation and promote our chances for USP10-CA crystallization.

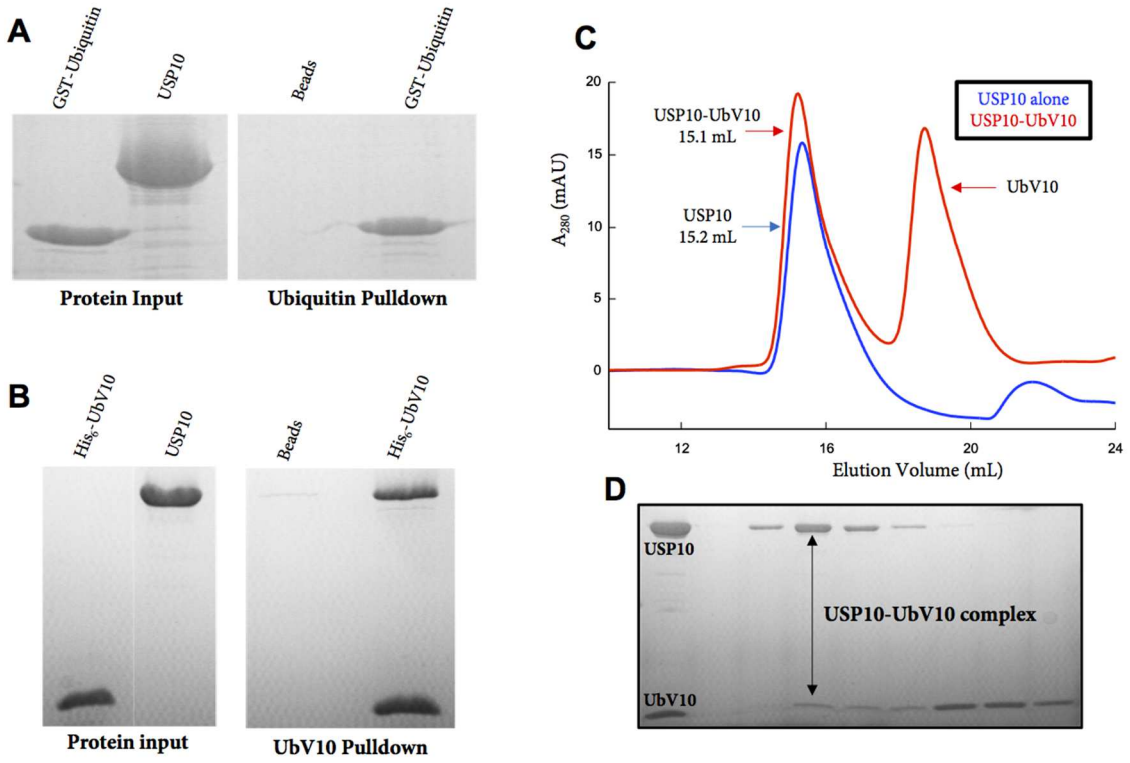

Figure 21. UbV10 complexes with USP10 catalytic domain
(A) GST-Ubiquitin pulldown of USP10-CA. (B) His-UbV10 pulldown of USP10-CA. (C) Gel filtration chromatogram of USP10-CA and UbV10. (D) SDS-PAGE of fractions under the chromatogram from C.

USP10-CA forms a complex with ubiquitin variant 10, not ubiquitin

Another challenge is complexing USP to ubiquitin since most USPs don't bind strongly to wild type ubiquitin. In order to structure USP and ubiquitin complexes, several groups covalently link the USP of interest with a covalent linker, attached to the carboxyl end of ubiquitin as a tactic to irreversibly link ubiquitin to the catalytic cysteine in USPs [152, 153, 155, 156]. An alternative strategy are the use of ubiquitin variants which are generated through phase-display libraries of mutated ubiquitin proteins [157]. The same group reported a mutated ubiquitin variant that specially binds and inhibits full length USP10 [149]. We obtained the reported protein sequence for UbV10, and purified UbV10 similarly as USP10 (data not shown). Consistently, we show USP10-CA binds strongly to UbV10 but not ubiquitin (Figure 21A and B). This suggests the

mutants created by phage display are important and necessary for USP10 binding. (Figure 21A and 21B). In addition USP10-CA and UbV10 appear to stability complex under gel filtration separation (Figure 21C and Figure 21D).

USP10 and UbV10 binding is strongly enthalpically driven

We next sought to measure the magnitude of binding (K_D) as well as additional thermodynamic information between USP10 and UbV10 by employing ITC. Measured enthalpy and dissociation constant were -13.4 kcal/mol and 984 nM, respectively. Calculated Gibb's free energy (ΔG) and entropy ($-T\Delta S$) were -8.3 kcal/mol and 5.1 kcal/mol, respectively. Overall, the binding between USP10 and UbV10 is a favorable process as indicated by enthalpy and Gibb's free energy; however, the interaction between USP10 and UbV10 is not entropically favorable. The high enthalpy and low entropic value suggest the interaction between USP10 and UbV10 are favorably driven by forming hydrogen bonds and resisted by unfavorable structuring of USP10 and UbV10.

USP10-CA activity is low, and Spautin-1 doesn't inhibit USP10-CA

To measure USP10-CA activity and inhibitory potency of UbV10, a fluorescence-based activity assay using Ubiquitin-Rhodamine 110 (Ub-Rho) as a substrate was employed. USP10 hydrolysis activity is weak compared to other DUBs [149] since we were unable to reach velocity saturation in our Michaelis-Menten curve. This is likely due to the high Michaelis constant (K_M) ~14.3 μM for substrate, which suggests USP10 doesn't bind strongly to the Ub-Rho. A separate group reported low activity for full length USP10 using Ub-Rho as a substrate (K_M >4 μM) [149]. Supplementation of UbV10 (1:20 molar protein:inhibitor) greatly diminished USP10-CA activity and greatly enhanced the modeled K_M. The modeled V_{max} between USP10-CA and USP10-CA-UbV10 were measured to 1.1 RFU/sec and 0.885 RFU/sec, respectively. The modeled V_{max} and

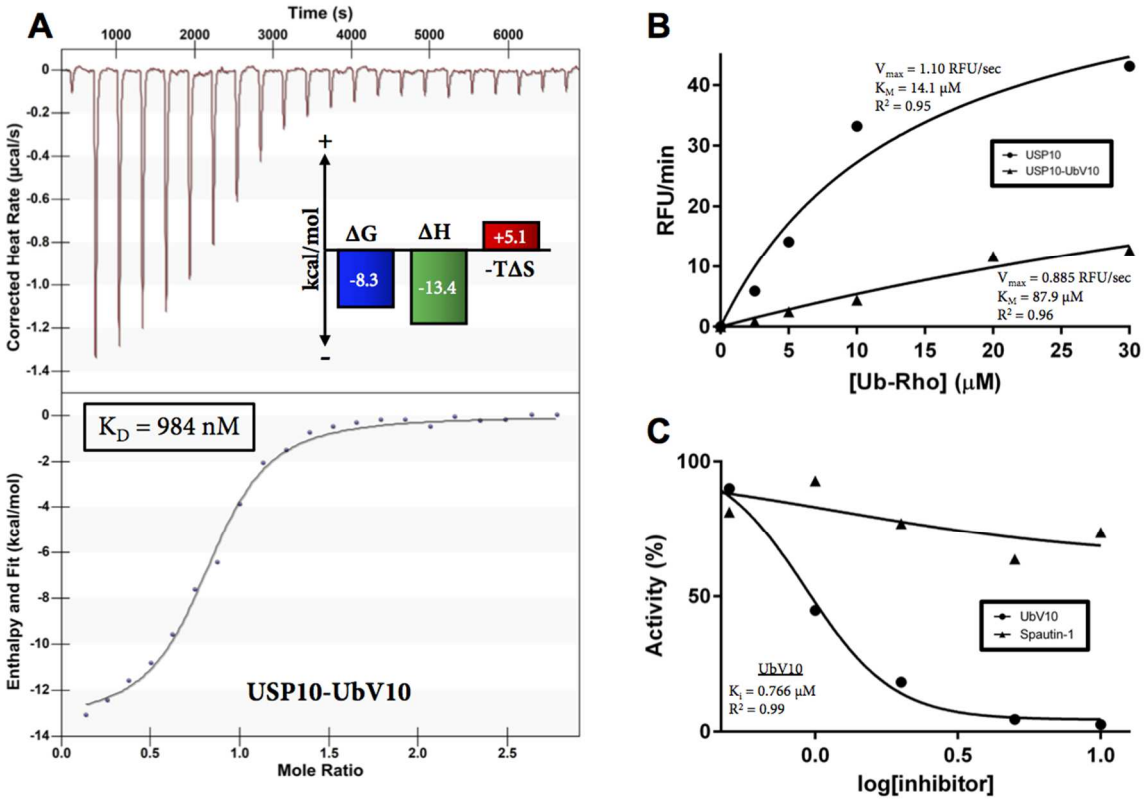

Figure 22. USP10-UbV10 binding is enthalpically driven and UbV10 competitively inhibits USP10.
(A) ITC data of USP10-CA and UbV10. Top panel displays corrected heat injections. Bottom panel displays isotherm plot. (B) Michaelis-Menten graph of USP10 activity. (C) Log-scale inhibition curve of UbV10 and spautin-1 against USP10-CA activity.

K_M parameters between USP10-CA alone and USP10-CA-UbV10 suggest UbV10 acts as a competitive inhibitor against USP10-CA. Additionally, we sought to compare how and how well UbV10 and spautin-1 inhibit USP10-CA. Our results indicate that spautin-1 doesn't directly inhibit USP10-CA. This finding is in conflict with another study which proposed spautin-1 selectively inhibits USP10 activity [148]. Further investigation will require either testing spautin-1 against full length USP10 or identifying an optimally active form of USP10. We were unable to detect P22077 inhibitory activity against USP10-CA (data not shown) since the compound was insoluble in our reaction buffer. Experiment will need to be repeated to accommodate the compound's poor solubility.

Discussion

In this study, we supply biochemical and structural studies for the catalytic domain of USP10 to pave the way for future drug development studies. We successfully purified the catalytic domain of USP10 using *E.Coli* as a production host eliminating the need to use other expensive expression systems; however, our *E.Coli* purified USP10-CA construct has low hydrolysis activity *in vitro*. Previous studies also report low deubiquitination activity for full length USP10 [149]. Therefore, USP10 hydrolysis activity is likely dependent on post-translational modifications and other protein chaperones like Beclin-1 [148]. Interestingly, USP10 is a phosphorylation target for ATM kinase signaling, AMPKα and CKII kinase [28, 145, 158]. Indeed, two separate studies found phosphorylation of USP10 S76 by AMPKα and USP10's association with Beclin-1 promoted deubiquitinase activity [145, 148]. Future studies may benefit exploring how phosphorylation and other protein chaperones influence USP10's deubiquitination activity.

We also provide supporting evidence that UbV10 is an effective competitive inhibitor against USP10-CA. The measured dissociation constant from our ITC results and inhibitory constant from our activity data are in fair agreement, 984 nM and 766 nM respectively. The thermodynamic binding profile of USP10-UbV10 appears to be similar to the thermodynamic profile reported for USP7 and Ub25 [159]. Despite the favorable enthalpic yield of USP10 and UbV10 binding, this complex outputs a disruptive entropic cost, which suggests the structuring of USP10 and UbV10 as a complex is an unfavorable process.

In summary, our study presents a soluble catalytic domain construct for human USP10 which is viable for further biochemical studies. Our data supports UbV10 as an active inhibitor against USP10-CA and may prove to be a useful as a tool to stabilize and potentially crystallize USP10. Unexpectedly, we did not observe significant inhibitory activity for spautin-1 which

suggests spautin-1 doesn't target the active site of USP10. Further studies towards developing small and selective molecule inhibitors for USP10-CA will require robust target-based chemical screening and understanding how USP10 achieves optimal biochemical activity. Attaining a high resolution structures of USP10 would also be of value since USP10 appears to be a fairly unique structure based on the lack of sequence conservation from other deposited USP structures. We look forward to future reports that dissect the structure and function USP10.

CHAPTER 6 STRUCTURAL BASIS OF PDZ-MEDIATED CHEMOKINE RECEPTOR CXCR2 SCAFFOLDING BY GUANINE NUCLEOTIDE EXCHANGE FACTOR PDZ-RHOGEF

*Published in Biochem Biophys Res Commun. 2017 Apr 1;485(2):529-534. doi:10.1016/ j.bbrc.2017.02.010. All authors agreed with including their work in this dissertation.

Introduction

CXCR2 is a G protein coupled receptor important for cellular mobility and chemotaxis through activation of calcium mobilization and actin polymerization [33]. CXCR2 is central to neutrophil migration to sites of inflammation and involved in wound healing and angiogenesis [33]. It has been shown that disrupting CXCR2 function plays a central role in multiple inflammatory diseases including rheumatoid arthritis, acute respiratory distress syndrome, septic shock, and chronic obstructive pulmonary dis- ease, likely due to the result of excessive release of neutrophils from the bone marrow [160]. CXCR2 also plays a critical role in many cancers such as lung cancer and pancreatic cancer by promoting tumor invasion and metastasis via autocrine and paracrine effects [161]. Elevated expression of CXCR2 enhances cancer cell proliferation and survival and often correlates with aggressive stages of cancer and poor overall prognosis [161].

CXCR2 directing cell trafficking depends on its ability to bind to ELR-positive CXC chemokines [162]. When binding to a chemokine, CXCR2 is capable of initiating G protein dissociation and inducing downstream signaling cascades that drive cell movement along chemokine concentration gradients. However, swift signaling requires direct and indirect interaction of CXCR2 with other membrane receptors, channels, intracellular scaffold proteins, effectors, and cytoskeletal elements, among which PDZ domain- containing proteins play a central role in efficient signaling by scaffolding the formation of macromolecular complexes at the plasma membrane and functionally coupling chemokine signaling to downstream signaling events [33]. In general, PDZ domains mediate protein interaction by recognizing the C-terminal sequence of

target proteins and binding to the targets through a canonically and structurally conserved PDZ peptide-binding pocket [163]. The specificity of the interaction is determined mainly by the residues at positions 0 and 2 of the peptides (position 0 referring to the C-terminal residue), whereas other residues do not significantly contribute to the interaction. As a result, PDZ domains are highly promiscuous capable of binding to multiple ligands; single peptides are capable of binding to distinct PDZ domains. Because of this promiscuity, PDZ-mediated interaction can generate complex and interconnected signaling networks that ensure precise and efficient signal transduction via protein-protein interaction.

However, the canonical ligand-binding of PDZ by itself has a limited capacity to scaffold multiprotein arrays within membrane microdomains, as PDZ domains can only bind to their ligands one at a time. Recent data suggest that PDZ dimerization plays an important role in increasing the scaffolding capacity [164]. PDZ dimerization with the same or different PDZ-containing proteins has been shown to amplify the complexity of interacting proteins in signal transduction networks and provide a mechanism to expand the scaffolding capacity in the assembly of multiprotein complexes. Of note, many PDZ domains can dimerize and 30% out of 150 PDZ domains in the mouse genome has been shown to participate in protein-protein interaction [165]. This suggests PDZ domains have evolved as a dual binding module in facilitating complex formation. Therefore, there has been a considerable interest in elucidating the structural basis of PDZ specificity, promiscuity and dimerization and how they can set up a specific interaction network for proper signaling, the nature of which still remains obscure.

Recently, we showed that the PDZ domains of NHERF1 play a pivotal role in CXCR2 signaling during the formation of macromolecular signaling complexes[33, 161]. NHERF1 scaffolds the interaction between CXCR2 and PLCβ2/3 by simultaneously binding to the C-

terminal tail of CXCR2 and PLCβ2/3 and physically connecting them through linked PDZ domains or PDZ dimerization. This resulted in a macromolecular complex essential for coupling CXCR2 activation to PLCβ2/3 signaling cascades in neutrophils and pancreatic cancer cells. Disruption of the interaction effectively abolished chemotaxis and transepithelial migration suggesting a functional importance of PDZ-mediated scaffolding. However, NHERF1 is the only PDZ domain-containing protein identified to date to interact with CXCR2. As a result, there is a limited understanding regarding the molecular mechanism of CXCR2 PDZ-binding promiscuity and how binding to different PDZ domains may interconnect different signaling pathways in CXCR2 signaling.

In this study, we sought to identify additional PDZ domains that could interact with the PDZ motif of CXCR2 and to better under- stand PDZ binding promiscuity and specificity. We identified several novel CXCR2 binding proteins using a PDZ binding array, among which PDZ-RhoGEF is of particular interest because it is also involved in signaling and cellular mobility. PDZ-RhoGEF is a PDZ and RGS-containing protein and belongs to the guanine nucleotide exchange factors family. It is a protein ubiquitously expressed in humans and involved in initiating the Rho signaling pathway for actin organization and cellular mobility [166]. To understand the molecular mechanism of CXCR2 PDZ motif binding to the PDZ domain of PDZ-RhoGEF, we solved the crystal structure of PDZ- RhoGEF PDZ domain in complex with the CXCR2 C-terminal PDZ binding motif. The structure reveals that the CXCR2 peptide binds to PDZ in an extended conformation with the last four residues making specific side chain contacts. Sequence alignment and structural comparison analyses suggest the sequence- and position-specific interactions determine CXCR2 PDZ-binding promiscuity and specificity. Unexpectedly, we identified a disulfide bond-linked PDZ dimer which enables parallel binding of CXCR2 peptides to the well-

separated ligand-binding pockets. This new mode of PDZ dimerization demonstrates structural diversity in PDZ-PDZ interaction and could prove valuable for understanding the complex-scaffolding function of PDZ-RhoGEF in CXCR2 signaling.

Materials and Methods

2.1 PDZ domain array screen

CXCR2-binding PDZ domains were screened using TranSignal PDZ Domain Array IV (Panomics) according to the manufacture's instruction. His-tagged C-terminal fragment of human CXCR2 (residues 316-360) was used in the assay screen, which was generated by PCR cloning into pET30 and purified using cobalt resins [33]. The purified CXCR2 was incubated with the PDZ Domain Arrays in blocking buffer for 1 h at room temperature, and washed thrice with wash buffer for 5 min. They were then incubated with Anti-histidine horseradish peroxidase (HRP) conjugate for 1 h at room temperature. Antibody complexes were detected by enhanced chemiluminescence and imaged using BioSpectrum 500 (UVP). The array was repeated twice and similar results were observed.

2.2 PDZ protein expression and purification

The cDNA fragment of human PDZ-RhoGEF PDZ (residues 41-123) was cloned into a pSUMO vector containing an N-terminal His_6-SUMO tag. The C-terminal extension TSTTL that corresponds to residues 356-360 of human CXCR2 was included in the reverse primer to create a chimeric clone. The clone was transformed into Escherichia coli BL21 Condon Plus (DE3) cells for protein expression. The transformants were grown to an OD_{600} of 0.4 at 37°C in LB medium, and then induced with 0.1 mM isopropylthio-β-D-galactoside at 15°C overnight. The cells were harvested and lysed by French Press. The soluble fraction was then subjected to Ni^{2+} affinity chromatography purification, followed by cleavage the His_6-SUMO tag with yeast SUMO

Protease 1. PDZ proteins were separated from the cleaved tag by a second Ni^{2+} affinity chromatography and further purified by size-exclusion chromatography. Finally, the proteins were concentrated to 10 mg/ml in a buffer containing 0.1 M sodium acetate pH 4.8., 150 mM NaCl, 5 mM b- mercaptoethanol (BME), and 5% glycerol.

2.3. Crystallization, data collection and structure determination

Crystals were grown by the hanging-drop vapor-diffusion method by mixing the protein (~10 mg/ml) with an equal volume of a reservoir solution containing 0.1 M sodium acetate pH 4.6, 0.1 M sodium citrate, 25% PEG8000 at 20°C. Crystals were cryoprotected in a solution containing 20% glycerol. Crystal data were collected at 100 K at the Advanced Photon Source (Argonne, IL) at beamline 21- ID-D and processed and scaled using XDS [136]. Crystals belong to the space group $C222_1$ with four molecules in the asymmetric unit (Supplementary data). The structure was solved by molecular replacement using PDZ-RhoGEF PDZ-PlexinB2 structure (PDB code: 5E6P) as a search model. Structure modeling was carried out in COOT [138] and refinement was performed with PHENIX [139]. The final model was analyzed and validated with Molprobity [139]. All figures of PDZ-CXCR2 structure were made with PyMOL.

2.4 Protein data bank accession number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 5TYT.

Results

3.1 CXCR2 C-terminus binds directly to PDZ-RhoGEF PDZ domain

Using a PDZ Array screen (Panomics), the CXCR2 C-terminus was identified to directly bind to the PDZ domains of PDZ-RhoGEF, leukemia associated RhoGEF (LARG), disks large homolog 3 (DLG3-D2), alpha-1-syntrophin (SNA1) and SH3/multiple ankyrin repeat domains

protein 1 (SHANK1) (SHK1) (Fig. 1A). DLG3-D2 is the second PDZ domain in synapse-associated protein 102 which serves as a post-synaptic scaffold for glutamate receptor signaling in developing cortical neurons. SNA1 is the only PDZ domain in alpha-1-syntrophin which serves as a scaffold for dystrophin protein complexes in Rac1 signaling in skeletal muscles. SHK1 is the first PDZ domain in SHANK1, a scaffolding protein that clusters neurotransmitter receptors necessary for synapse changes and development. PDZ-RhoGEF and LARG were of interest due to their imperative roles in cellular signaling and mobility [166], an analogous function of CXCR2. PDZ-RhoGEF and LARG possess one PDZ domain and were known to bind to PlexinB1/2, LPA1/2, and insulin-like growth factor-1 receptor in addition to CXCR2 [166]. These together raise an interesting question regarding the mechanism of how CXCR2 is recognized by different PDZ domains and how PDZ-RhoGEF PDZ binds to different substrates.

3.2 Binding specificity of the PDZ-CXCR2 interaction

To understand the interaction between PDZ-RhoGEF and CXCR2, we solved the crystal structure of PDZ-RhoGEF PDZ in complex with the C-terminal sequence (TTSTL) of CXCR2. The overall structure of PDZ-RhoGEF PDZ is similar to other PDZ domains [163], consisting of six β strands (β1-β6) and two α-helices (αA and αB) (Fig. 1B). The CXCR2 peptide binds in the cleft between β2 and βB, burying a total solvent-accessible surface area of 507.4 Å2. The binding specificity of the PDZ-CXCR2 interaction is achieved through networks of hydrogen bonds and hydrophobic interactions (Fig. 1C). At the ligand position 0, the side chain of Leu0 is nestled in a deep hydrophobic pocket formed by conserved residues Phe57, Phe59 and Val61 from β2 and Val106 and Ile109 from αB (Fig. 1D). In the pocket, the position of Leu0 is further secured by both a hydrogen bond from its amide nitrogen to the Phe59 carbonyl oxygen and bifurcated

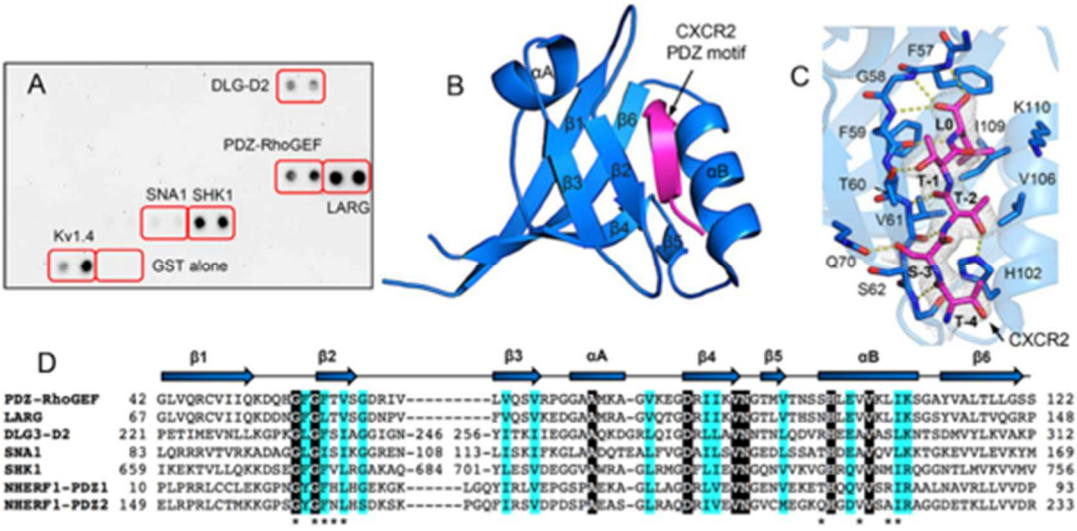

Figure 23. Structure of PDZ-RhoGEF PDZ (rPDZ) in complex with the CXCR2 C-terminal sequence TSTTL.
(A) PDZ Array screen of CXCR2-binding PDZ domains. Kv1.4 serves as a positive control for PDZ-peptide binding, and GST alone a negative control. (B) Ribbon diagram of rPDZ-CXCR2 structure. The PDZ is shown in blue and CXCR2 in magenta. Secondary structures are numbered and labeled based on their sequence position. (C) Detailed view of the PDZ ligand-binding site. 2Fo-Fc omit map of CXCR2 peptide was calculated at 2.4 Å and contoured at 1.0 s. Hydrogen bonds are illustrated as yellow broken lines. (D) Sequence alignment of CXCR2-binding PDZ domains. Identical residues are represented as white on black and similar residues are colored in cyan. Residues important for binding to CXCR2 are indicated by asterisks. The secondary structure elements are labeled according to 1B.

hydrogen bonding between the Leu0 carboxylate and the amides of Phe57 and Gly58. Similar interactions have been observed in several other PDZ-mediated complexes which represent the most-conserved binding mode for terminal leucine recognition [163]. Residues at other peptide positions also contribute to the PDZ-CXCR2 complex formation. At position 1, the side chain hydroxyl of Thr-1 forms a hydrogen bond with the Og1 atom of the Thr60 side chain. At position 2, Thr-2 makes one hydrogen bond to the His102 imidazole group and two hydrogen bonds to the highly conserved residue Val61. At the ligand position 3, the interactions with Ser-3 include one hydrogen bond from its side chain hydroxyl to the Oε1 atom of Gln70, and a VDW interaction with the side chain of Ser62. Finally, the peptide residue Thr-4 engages in a main-chain contact

with Gly63, but does not participate in any specific side-chain interactions. These observations indicate that the last four residues of CXCR2 contribute to the binding specificity in the PDZ-CXCR2 complex formation.

3.3 CXCR2 and PDZ binding promiscuity

To gain insight into promiscuous CXCR2 binding by different PDZ domains, we compared the structures of all available PDZ domains in complex with CXCR2, including NHERF1 PDZ1, NHERF1 PDZ2 and current PDZ-RhoGEF PDZ (rPDZ)[163, 167]. We also compared the rPDZ-CXCR2 structure to the structure of rPDZ in complex with a PlexinB2 peptide in order to understand PDZ binding promiscuity. These liganded PDZ structures are very similar with pairwise root-mean-square differences (RMSDs) ranging from 0.47 to 0.87 Å for entire C atoms (Fig. 2A). The main chains of the bound peptides superimpose well, as do their relative spatial positions to the conserved PDZ motifs. Additionally, the ligand recognition modes at the peptide positions 0 and 2 are virtually indistinguishable, characterized by structurally similar binding sites composed with highly conserved residues (Fig. 2B). This observation is consistent with previous evidence that the 0 and 2 residues of the ligand are critical for determining the binding specificity and affinity of PDZ-peptide interaction [163].

Large differences were found in ligand recognition at the peptide positions 1 and 3. In CXCR2-binding PDZ domains, residues that recognize these two positions are not conserved; in fact, the residues that recognize the 3 position are not even structurally equivalent (Figs. 1D and 2B). At position 1, the binding conformation of Thr-1 is nearly identical in different PDZ structures. The side chain of Thr-1 is oriented towards the same direction facing a residue equivalent to rPDZ Thr60. As a result, Thr-1 is recognized by different residues from the equivalent position. In NHERF1 PDZ1, the side chain hydroxyl of Thr-1 is stacked by the

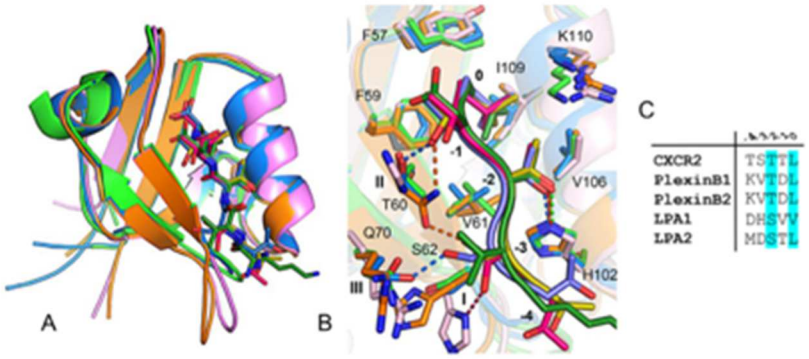

Figure 24. Structural comparison between PDZ-ligand complexes.
(A) Superposition of rPDZ-CXCR2 (blue), rPDZ-PlexinB2 (green), NHERF1 PDZ1-CXCR2 (pink) and NHERF1 PDZ2-CXCR2 (orange). PDZ domains are represented by ribbon, and ligand residues are displayed as sticks. (B) Superposition of PDZ ligand-binding sites. Colors are identical from 2A. (C) Sequence alignment of last five residues of known rPDZ binding targets.

imidazole ring of a histidine residue. In NHERF1 PDZ2, the side chain hydroxyl of Thr-1 forms a hydrogen bond with the side chain nitrogen of an asparagine. In current rPDZ structure, the equivalent hydroxyl group forms a hydrogen bond to the side chain of Thr60. In SHK1 and DLG3-D2, Thr-1 may interact with a valine and serine respectively (Fig. 1D). This demonstrates Thr-1 can be recognized by different residues via different interactions without the need for significant structural changes. This is also consistent with previous data that 1 residue in the peptide ligands is less stringently specified by individual PDZ domains than the residues at the 0 and 2 positions, thereby allowing binding promiscuity [163].

At position 3, the peptide binding is more PDZ specific, facilitated by the rotameric flexibility of Ser-3 (Fig. 2B). In the binding, Ser-3 adopts different rotamers and each rotamer is able to bind to a unique position in PDZ domains. In NHERF1 PDZ1, the side chain hydroxyl of Ser-3 points to the N-terminus of the peptide forming a hydrogen bond to a histidine residue at the receptor position 1. In NHERF1 PDZ2, the side chain hydroxyl of Ser-3 points towards the C-terminus forming a hydrogen bond to an asparagine at the receptor position 2. In current rPDZ, the pointing direction of the Ser- 3 side chain is perpendicular to the peptide direction which enables

hydrogen bonding with a glutamine residue at the receptor position 3. Of note, all these receptor positions have been suggested to contribute to a high degree of selectivity in PDZ ligand recognition and the ability for the PDZ domain family to bind to different sequences [168]. The present study is extending the role of these receptor positions in determining binding diversity, and the fact that one of such positions can be specifically selected for interacting with 3 residue makes it possible for different PDZ domains to recognize the ligand residues of same sequence, providing an explanation for promiscuous CXCR2 binding.

The structural comparison between rPDZ-CXCR2 and rPDZ-PlexinB2 provides some insight into PDZ binding promiscuity. The residues at the ligand positions 1 and 3 are highly variable across rPDZ binding targets (Fig. 2C) indicating an ability of rPDZ to bind to ligands with different 1 and 3 side chains. At position 1, Thr-1 of CXCR2 forms a hydrogen bond with Thr60, whereas most of the side chain atoms of Asp-1 in PlexinB2 are disordered suggesting no stable interaction between Asp-1 and rPDZ; and the 1 position of PlexinB2 may not significantly contribute to the binding specificity (Fig. 2B). At position 3, Val-3 of PlexinB2 forms less discriminative VDW interactions with Gln70 and Ser62, differing from Ser-3 of CXCR2 which forms a specific hydrogen bond with Gln70. This difference indicates that rPDZ is able to form different types of interactions with 3 residues, which may underlie its flexibility to accommodate ligands with different 3 side chains.

3.4. Disulfide bond linked PDZ dimer

The most intriguing finding in current rPDZ structure is an asymmetric disulfide bond-linked PDZ dimer found in the asymmetric unit of the crystal (Fig. 3A). PDZ dimerization has been well appreciated as an important mechanism for improving PDZ scaffolding capacity during

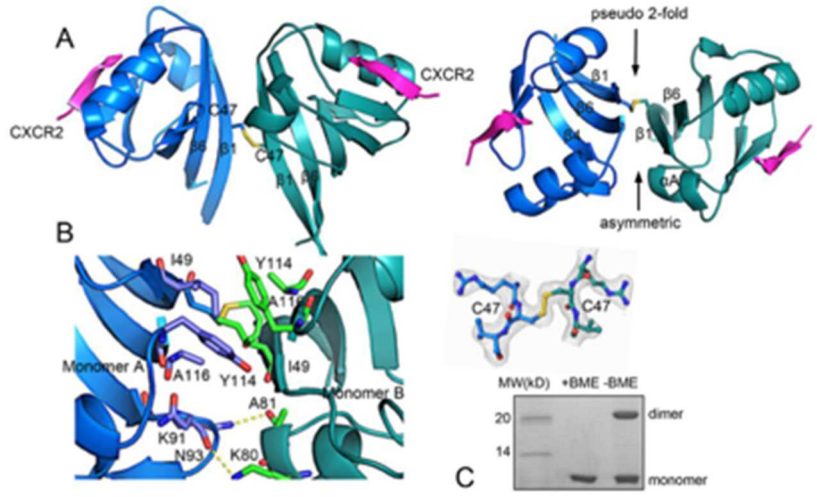

Figure 25. Disulfide bond linked PDZ dimer.
(A) Overall view of PDZ dimer (side view, left; top view, right). (B) Close-up view of the dimer interface. 2Fo-Fc omit map of the disulfide bond was calculated at 2.4 Å and contoured at 1.0 s. (C) SDS-PAGE of 5 μg of rPDZ-CXCR2 fusion protein with/without BME.

the formation of multiprotein complexes [165]. The dimerization is usually formed by non-covalent PDZ-PDZ interactions that put two canonical binding sites in a close proximity to facilitate a parallel or antiparallel nucleation of interacting proteins. However, the nature and significance of the less common disulfide bond-linked dimers remain largely elusive, though a few studies have suggested that the formation of disulfide bonds between proteins can be triggered by reactive oxygen species during cellular signaling [169]. In current rPDZ structure, two Cys47 on the outer surface of the b1 strand form an intermolecular disulfide bond responsible for PDZ dimerization (Fig. 3A). The dimer is an asymmetric dimer with the upper side having a pseudo 2-fold symmetry generated by parallel stacking of two copies of strands b1 and b6, and the lower side being asymmetric generated by the interaction between αA and β4 from different monomers. The buried surface area at the dimer interface is 777.6 Å2.

In addition to the disulfide bond, the dimer interface is further stabilized by several hydrogen bonds and hydrophobic interactions (Fig. 3B). At the upper side of the interface, there

is a continuous hydrophobic core formed by pairs of residues Ile49, Tyr114 and Ala116, which is located above the disulfide bond bridge. At the lower side of the interface, two asymmetric hydrogen bonds are formed: Lys91 from the β4-β5 loop in monomer A forms a hydrogen bond with Ala81 carbonyl oxygen from the αA helix in monomer B; Lys80 from the αA helix from monomer B forms a hydrogen bond with Asn93 carbonyl oxygen from the β4-β5 loop in monomer A. Of note, all these residues including the disulfide bond forming residue Cys47 are highly conserved in the PDZ and RGS-containing GEF protein family suggesting a conserved function of this disulfide bond-linked dimer (Fig. 1D). Reducing SDS-PAGE indicates that the disulfide bonds contribute to the formation of a dimer in solution (Fig. 3C). Previous studies have also demonstrated that rPDZ dimerization linked by Cys47 disulfide bond regulates the canonical ligand binding and enhances in vitro binding to a bivalent PlexinB2 PDZ motif [34].

Discussion

In this study, we identified several CXCR2 interacting PDZ domains including PDZ-RhoGEF PDZ. We solved the crystal structure of PDZ-RhoGEF PDZ in complex with the CXCR2 C-terminal tail that provides the molecular basis of the interaction. The crystal structure also reveals an unexpected asymmetric disulfide bond-linked PDZ dimer that allows simultaneous parallel binding of CXCR2 to two PDZ domains. While the functional link between CXCR2 and PDZ-RhoGEF in signaling and cellular mobility requires future investigation, the identification of new CXCR2-binding PDZ domains is enforcing the view that PDZ domains play important roles in CXCR2 signaling processes capable of scaffolding complex interaction networks and coupling CXCR2 signaling to specific signaling pathways, potentially Rho signaling and Rac1 signaling through interacting with PDZ-RhoGEF and SNA1 respectively [166]. The identified interactions also provide additional models that enable further understanding of PDZ and CXCR2 binding

promiscuity and specificity. The structural comparison is able to reveal the residues at the ligand positions 1 and 3 conferring PDZ- and ligand-specific recognition that may underlie the ability of CXCR2 to be bound by different PDZ domains and PDZ-RhoGEF PDZ to bind to different ligands. Additionally, the finding of the unexpected disulfide bond-linked PDZ dimer further demonstrates the structural diversity of PDZ dimerization. Diverse PDZ-PDZ interactions have been optimized as a mechanism in scaffolding the formation of distinct multiprotein complexes [165]. This non-canonical binding mode has been suggested to contribute more to defining the precise composition of protein complexes than does the canonical binding mode due to the structural diversity [165]. Therefore, there has been a continuous interest in revealing the specific nature of PDZ-PDZ interactions and their selectivity in precise scaffolding of temporal and spatial signaling networks. The current study provides an additional example of how PDZ domains may dimerize, and the asymmetric interface and rare disulfide bond linkage effectively define a new mode of PDZ dimerization, which is different from any reported structures [163]. Together with PDZ binding promiscuity, the new mode of dimerization could provide a reactive oxygen species-sensitive molecular scaffold for assembly of distinct CXCR2 signaling networks in actin polymerization and cell mobility.

Table 2. Crystallographic data and refinement statistics

Data	
Space group	$C222_1$
Cell parameters	
a, b, c (Å)	61.7, 66.6, 168.4
Wavelength (Å)	1.07822
Resolution (Å)	84.2-2.40 (2.44-2.40)[a]
R_{merge}[b]	0.094 (0.738)
Redundancy	6.5 (6.1)
Unique reflections	13997
Completeness (%)	99.9 (99.2)
$\langle I/\sigma \rangle$	11.2 (2.2)
Refinement	
Resolution (Å)	43.7-2.40 (2.44-2.40)
Molecules/AU	4
R_{work}[c]	0.206 (0.303)
R_{free}[d]	0.243 (0.356)
Ramachandran plot	
Residues in favored	98.0%
Residues in allowed	2.0%
RMSD	
Bond lengths (Å)	0.009
Bond angels (°)	1.19
No. of atoms	
Protein	2619
Peptide	144
Water	13
B-factor (Å^2)	
Protein	64.6
Peptide	70.7
Water	51.5

[a]Numbers in parentheses refer to the highest resolution shell.

[b]$R_{merge} = \Sigma|I-\langle I\rangle| / \Sigma I$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity of multiple observations of symmetry-related reflections.

[c]$R_{work} = \Sigma|F_o-F_c| / \Sigma|F_o|$, where F_o is the observed structure factor, F_c is the calculated structure factor.

[d]R_{free} was calculated using a subset (5%) of the reflection not used in the refinement.

REFERENCES

1. Gottlieb, P.D., et al., *Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis.* Nat Genet, 2002. **31**(1): p. 25-32.
2. Brown, M.A., et al., *Identification and characterization of Smyd2: a split SET/MYND domain-containing histone H3 lysine 36-specific methyltransferase that interacts with the Sin3 histone deacetylase complex.* Mol Cancer, 2006. **5**: p. 26.
3. Huang, J., et al., *Repression of p53 activity by Smyd2-mediated methylation.* Nature, 2006. **444**(7119): p. 629-32.
4. Hamamoto, R., et al., *SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells.* Nat Cell Biol, 2004. **6**(8): p. 731-40.
5. Mazur, P.K., et al., *SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer.* Nature, 2014. **510**(7504): p. 283-7.
6. Du, S.J., X. Tan, and J. Zhang, *SMYD proteins: key regulators in skeletal and cardiac muscle development and function.* Anat Rec (Hoboken), 2014. **297**(9): p. 1650-62.
7. Sims, R.J., 3rd, et al., *m-Bop, a repressor protein essential for cardiogenesis, interacts with skNAC, a heart- and muscle-specific transcription factor.* J Biol Chem, 2002. **277**(29): p. 26524-9.
8. Phan, D., et al., *BOP, a regulator of right ventricular heart development, is a direct transcriptional target of MEF2C in the developing heart.* Development, 2005. **132**(11): p. 2669-78.
9. Rasmussen, T.L., et al., *Smyd1 Facilitates Heart Development by Antagonizing Oxidative and ER Stress Responses.* PLoS One, 2015. **10**(3): p. e0121765.
10. Kawamura, S., et al., *smyd1 and smyd2 are expressed in muscle tissue in Xenopus laevis.* Cytotechnology, 2008. **57**(2): p. 161-8.
11. Fujii, T., et al., *Smyd3 is required for the development of cardiac and skeletal muscle in zebrafish.* PLoS One, 2011. **6**(8): p. e23491.
12. Proserpio, V., et al., *The methyltransferase SMYD3 mediates the recruitment of transcriptional cofactors at the myostatin and c-Met genes and regulates skeletal muscle atrophy.* Genes Dev, 2013. **27**(11): p. 1299-312.

13. Nagata, D.E., et al., *Epigenetic control of Foxp3 by SMYD3 H3K4 histone methyltransferase controls iTreg development and regulates pathogenic T-cell responses during pulmonary viral infection.* Mucosal Immunol, 2015. **8**(5): p. 1131-43.
14. Xu, G., et al., *The histone methyltransferase Smyd2 is a negative regulator of macrophage activation by suppressing interleukin 6 (IL-6) and tumor necrosis factor alpha (TNF-alpha) production.* J Biol Chem, 2015. **290**(9): p. 5414-23.
15. Stender, J.D., et al., *Control of proinflammatory gene programs by regulated trimethylation and demethylation of histone H4K20.* Mol Cell, 2012. **48**(1): p. 28-38.
16. Komatsu, S., et al., *Overexpression of SMYD2 relates to tumor cell proliferation and malignant outcome of esophageal squamous cell carcinoma.* Carcinogenesis, 2009. **30**(7): p. 1139-46.
17. Sakamoto, L.H., et al., *SMYD2 is highly expressed in pediatric acute lymphoblastic leukemia and constitutes a bad prognostic factor.* Leuk Res, 2014. **38**(4): p. 496-502.
18. Komatsu, S., et al., *Overexpression of SMYD2 contributes to malignant outcome in gastric cancer.* Br J Cancer, 2015. **112**(2): p. 357-64.
19. Reynoird, N., et al., *Coordination of stress signals by the lysine methyltransferase SMYD2 promotes pancreatic cancer.* Genes Dev, 2016. **30**(7): p. 772-85.
20. Hamamoto, R., et al., *Enhanced SMYD3 expression is essential for the growth of breast cancer cells.* Cancer Sci, 2006. **97**(2): p. 113-8.
21. Wang, S.Z., et al., *Knockdown of SMYD3 by RNA interference inhibits cervical carcinoma cell growth and invasion in vitro.* BMB Rep, 2008. **41**(4): p. 294-9.
22. *SMYD3 Is an Oncogene Required for Liver and Colon Tumorigenesis.* Cancer Discov, 2016. **6**(4): p. OF13.
23. Li, L.X., et al., *Lysine methyltransferase SMYD2 promotes triple negative breast cancer progression.* Cell Death Dis, 2018. **9**(3): p. 326.
24. Jiang, Y., et al., *Crystal structures of histone and p53 methyltransferase SmyD2 reveal a conformational flexibility of the autoinhibitory C-terminal domain.* PLoS One, 2011. **6**(6): p. e21640.
25. Komander, D. and M. Rape, *The ubiquitin code.* Annu Rev Biochem, 2012. **81**: p. 203-29.

26. Dou, Q.P. and J.A. Zonder, *Overview of proteasome inhibitor-based anti-cancer therapies: perspective on bortezomib and second generation proteasome inhibitors versus future generation inhibitors of ubiquitin-proteasome system.* Curr Cancer Drug Targets, 2014. **14**(6): p. 517-36.
27. Shi, D. and S.R. Grossman, *Ubiquitin becomes ubiquitous in cancer: emerging roles of ubiquitin ligases and deubiquitinases in tumorigenesis and as therapeutic targets.* Cancer Biol Ther, 2010. **10**(8): p. 737-47.
28. Yuan, J., et al., *USP10 regulates p53 localization and stability by deubiquitinating p53.* Cell, 2010. **140**(3): p. 384-96.
29. Takayama, K.I., et al., *Association of USP10 with G3BP2 Inhibits p53 Signaling and Contributes to Poor Outcome in Prostate Cancer.* Mol Cancer Res, 2018. **16**(5): p. 846-856.
30. Weisberg, E.L., et al., *Inhibition of USP10 induces degradation of oncogenic FLT3.* Nat Chem Biol, 2017. **13**(12): p. 1207-1215.
31. Schultz, J., et al., *SMART: a web-based tool for the study of genetically mobile domains.* Nucleic Acids Res, 2000. **28**(1): p. 231-4.
32. Fan, J.S. and M. Zhang, *Signaling complex organization by PDZ domain proteins.* Neurosignals, 2002. **11**(6): p. 315-21.
33. Wu, Y., et al., *A chemokine receptor CXCR2 macromolecular complex regulates neutrophil functions in inflammatory diseases.* J Biol Chem, 2012. **287**(8): p. 5744-55.
34. Paduch, M., et al., *Bivalent peptides as models for multimeric targets of PDZ domains.* Chembiochem, 2007. **8**(4): p. 443-52.
35. Sirinupong, N., et al., *Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture.* J Biol Chem, 2010. **285**(52): p. 40635-44.
36. Berger, S.L., *The complex language of chromatin regulation during transcription.* Nature, 2007. **447**(7143): p. 407-12.
37. Liu, Y., et al., *Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity.* Cancer Cell, 2007. **11**(6): p. 483-97.
38. Spadaccini, R., et al., *Structure and functional analysis of the MYND domain.* J Mol Biol, 2006. **358**(1): p. 498-508.

39. Sirinupong, N., et al., *Structural insights into the autoinhibition and posttranslational activation of histone methyltransferase SmyD3.* J Mol Biol, 2011. **406**(1): p. 149-59.
40. Allan, R.K. and T. Ratajczak, *Versatile TPR domains accommodate different modes of target protein recognition and function.* Cell Stress Chaperones, 2011. **16**(4): p. 353-367.
41. Abu-Farha, M., et al., *The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase.* Mol Cell Proteomics, 2008. **7**(3): p. 560-72.
42. Tan, X., et al., *SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos.* Proc Natl Acad Sci U S A, 2006. **103**(8): p. 2713-8.
43. Medjkane S Fau - Cock-Rada, A., J.B. Cock-Rada A Fau - Weitzman, and J.B. Weitzman, *Role of the SMYD3 histone methyltransferase in tumorigenesis: local or global effects?* Cell Cycle, 2012. **11**(10).
44. Cock-Rada, A.M., et al., *SMYD3 promotes cancer invasion by epigenetic upregulation of the metalloproteinase MMP-9.* Cancer Res, 2012. **72**(3): p. 810-20.
45. Lin, Y.M., et al., *Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas.* Oncogene, 2002. **21**(26): p. 4120-8.
46. Okabe, H., et al., *Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression.* Cancer Res, 2001. **61**(5): p. 2129-37.
47. Hu, L., et al., *Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development.* Cancer Res, 2009. **69**(9): p. 4067-72.
48. Thompson, E.C. and A.A. Travers, *A Drosophila Smyd4 homologue is a muscle-specific transcriptional modulator involved in development.* PLoS One, 2008. **3**(8): p. e3008.
49. Saddic, L.A., et al., *Methylation of the retinoblastoma tumor suppressor by SMYD2.* J Biol Chem, 2010. **285**(48): p. 37733-40.
50. Cho, H.S., et al., *RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation.* Neoplasia, 2012. **14**(6): p. 476-86.
51. Sajjad, A., et al., *Lysine methyltransferase Smyd2 suppresses p53-dependent cardiomyocyte apoptosis.* Biochim Biophys Acta, 2014. **1843**(11): p. 2556-62.

52. Piao, L., et al., *The Histone Methyltransferase SMYD2 Methylates PARP1 and Promotes Poly(ADP-ribosyl)ation Activity in Cancer Cells.* Neoplasia, 2014. **16**(3): p. 257-264.e2.
53. Kunizaki, M., et al., *The lysine 831 of vascular endothelial growth factor receptor 1 is a novel target of methylation by SMYD3.* Cancer Res, 2007. **67**(22): p. 10759-65.
54. Ferguson, A.D., et al., *Structural basis of substrate methylation and inhibition of SMYD2.* Structure, 2011. **19**(9): p. 1262-73.
55. Foreman, K.W., et al., *Structural and functional profiling of the human histone methyltransferase SMYD3.* PLoS One, 2011. **6**(7): p. e22290.
56. Xu, S., et al., *Structure of human lysine methyltransferase Smyd2 reveals insights into the substrate divergence in Smyd proteins.* J Mol Cell Biol, 2011. **3**(5): p. 293-300.
57. Jiang, Y., et al., *Structural Insights into Estrogen Receptor alpha Methylation by Histone Methyltransferase SMYD2, a Cellular Event Implicated in Estrogen Signaling Regulation.* J Mol Biol, 2014. **426**(20): p. 3413-25.
58. Wang, L., et al., *Structure of human SMYD2 protein reveals the basis of p53 tumor suppressor methylation.* J Biol Chem, 2011. **286**(44): p. 38725-37.
59. Couture, J.F., et al., *Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase.* Genes Dev, 2005. **19**(12): p. 1455-65.
60. Wilson, J.R., et al., *Crystal structure and functional analysis of the histone methyltransferase SET7/9.* Cell, 2002. **111**(1): p. 105-15.
61. Wu, J., et al., *Biochemical characterization of human SET and MYND domain-containing protein 2 methyltransferase.* Biochemistry, 2011. **50**(29): p. 6488-97.
62. Kateb, F., et al., *Structural and functional analysis of the DEAF-1 and BS69 MYND domains.* PLoS One, 2013. **8**(1): p. e54715.
63. Abu-Farha, M., et al., *Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2.* J Mol Cell Biol, 2011. **3**(5): p. 301-8.
64. Xu, S., et al., *Structural and biochemical studies of human lysine methyltransferase Smyd3 reveal the important functional roles of its post-SET and TPR domains and the regulation of its activity by DNA binding.* Nucleic Acids Res, 2011. **39**(10): p. 4438-49.
65. Luo, X.G., et al., *Histone methyltransferase SMYD3 promotes MRTF-A-mediated transactivation of MYL9 and migration of MCF-7 breast cancer cells.* Cancer Lett, 2014. **344**(1): p. 129-137.

66. Liu, C., et al., *Transcriptional regulation of 15-lipoxygenase expression by histone h3 lysine 4 methylation/demethylation.* PLoS One, 2012. **7**(12): p. e52703.
67. Kim, H., et al., *Requirement of histone methyltransferase SMYD3 for estrogen receptor-mediated transcription.* J Biol Chem, 2009. **284**(30): p. 19867-77.
68. Van Aller, G.S., et al., *Smyd3 regulates cancer cell phenotypes and catalyzes histone H4 lysine 5 methylation.* Epigenetics, 2012. **7**(4): p. 340-3.
69. Zhang, X., et al., *Structural basis for the product specificity of histone lysine methyltransferases.* Mol Cell, 2003. **12**(1): p. 177-85.
70. Xiao, B., et al., *Structure and catalytic mechanism of the human histone methyltransferase SET7/9.* Nature, 2003. **421**(6923): p. 652-6.
71. Zhang, X., et al., *Regulation of estrogen receptor alpha by histone methyltransferase SMYD2-mediated protein methylation.* Proc Natl Acad Sci U S A, 2013. **110**(43): p. 17284-9.
72. Onuoha, S.C., et al., *Structural studies on the co-chaperone Hop and its complexes with Hsp90.* J Mol Biol, 2008. **379**(4): p. 732-44.
73. Song, Y. and D.C. Masison, *Independent regulation of Hsp70 and Hsp90 chaperones by Hsp70/Hsp90-organizing protein Sti1 (Hop1).* J Biol Chem, 2005. **280**(40): p. 34178-85.
74. Wu, B., et al., *3D structure of human FK506-binding protein 52: implications for the assembly of the glucocorticoid receptor/Hsp90/immunophilin heterocomplex.* Proc Natl Acad Sci U S A, 2004. **101**(22): p. 8348-53.
75. Davies, T.H. and E.R. Sanchez, *Fkbp52.* Int J Biochem Cell Biol, 2005. **37**(1): p. 42-7.
76. Deshpande, C.N., et al., *Crystal structure of an integron gene cassette-associated protein from Vibrio cholerae identifies a cationic drug-binding module.* PLoS One, 2011. **6**(3): p. e16934.
77. van den Hemel, D., et al., *Ligand-induced conformational changes in the capping subdomain of a bacterial old yellow enzyme homologue and conserved sequence fingerprints provide new insights into substrate binding.* J Biol Chem, 2006. **281**(38): p. 28152-61.
78. Aufhammer, S.W., et al., *Crystal structure of methylenetetrahydromethanopterin reductase (Mer) in complex with coenzyme F420: Architecture of the F420/FMN binding*

site of enzymes within the nonprolyl cis-peptide containing bacterial luciferase family. Protein Sci, 2005. **14**(7): p. 1840-9.

79. Borlak, J. and T. Thum, *Hallmarks of ion channel gene expression in end-stage heart failure.* Faseb j, 2003. **17**(12): p. 1592-608.
80. Gambetta, K., et al., *Transcription repression and blocks in cell cycle progression in hypoplastic left heart syndrome.* Am J Physiol Heart Circ Physiol, 2008. **294**(5): p. H2268-75.
81. Liu, C., et al., *SMYD3 as an oncogenic driver in prostate cancer by stimulation of androgen receptor transcription.* J Natl Cancer Inst., 2013. **105**(22): p. 1719-1728.
82. O'Byrne, K.J., et al., *Epigenetic regulation of glucose transporters in non-small cell lung cancer.* Cancers (Basel), 2011. **3**(2): p. 1550-65.
83. Vieira, F.Q., et al., *Deregulated expression of selected histone methylases and demethylases in prostate carcinoma.* Endocr Relat Cancer, 2014. **21**(1): p. 51-61.
84. He, C., et al., *High expression of trimethylated histone H3 lysine 4 is associated with poor prognosis in hepatocellular carcinoma.* Hum. Pathol., 2012. **43**: p. 1425-1435.
85. Luo, X.G., et al., *Effects of SMYD3 overexpression on transformation, serum dependence, and apoptosis sensitivity in NIH3T3 cells.* IUBMB Life, 2009. **61**(6): p. 679-84.
86. Luo, X.G., et al., *Novobiocin decreases SMYD3 expression and inhibits the migration of MDA-MB-231 human breast cancer cells.* IUBMB Life, 2010. **62**(3): p. 194-9.
87. Sponziello, M., et al., *Epigenetic-related gene expression profile in medullary thyroid cancer revealed the overexpression of the histone methyltransferases EZH2 and SMYD3 in aggressive tumours.* Mol Cell Endocrinol, 2014. **392**(1-2): p. 8-13.
88. Dong, S.W., et al., *Effect of the downregulation of SMYD3 expression by RNAi on RIZ1 expression and proliferation of esophageal squamous cell carcinoma.* Oncol Rep, 2014. **32**(3): p. 1064-70.
89. Josse, R., et al., *Identification of early target genes of aflatoxin B1 in human hepatocytes, inter-individual variability and comparison with other genotoxic compounds.* Toxicol Appl Pharmacol, 2012. **258**(2): p. 176-87.

90. Gutmann, D.H., et al., *Loss of DAL-1, a protein 4.1-related tumor suppressor, is an important early event in the pathogenesis of meningiomas.* Hum. Mol. Genet., 2000. **9**: p. 1495-1500.

91. Tran, Y.K., et al., *A novel member of the NF2/ERM/4.1 superfamily with growth suppressing properties in lung cancer.* Cancer Res, 1999. **59**(1): p. 35-43.

92. Wiesner, S., et al., *A change in conformational dynamics underlies the activation of Eph receptor tyrosine kinases.* EMBO J, 2006. **25**(19): p. 4686-96.

93. Gosu, V. and S. Choi, *Structural dynamic analysis of apo and ATP-bound IRAK4 kinase.* Sci Rep, 2014. **4**: p. 5748.

94. Spellmon, N., et al., *Structure and function of SET and MYND domain-containing proteins.* Int J Mol Sci, 2015(1422-0067 (Electronic)).

95. Sarris, M.E., et al., *Smyd3 Is a Transcriptional Potentiator of Multiple Cancer-Promoting Genes and Required for Liver and Colon Cancer Development.* Cancer Cell, 2016. **29**(3): p. 354-66.

96. Vieira, F.Q., et al., *SMYD3 contributes to a more aggressive phenotype of prostate cancer and targets Cyclin D2 through H4K20me3.* Oncotarget, 2015. **6**(15): p. 13644-57.

97. Kim, J.M., et al., *Cooperation between SMYD3 and PC4 drives a distinct transcriptional program in cancer cells.* Nucleic Acids Res, 2015. **43**(18): p. 8868-83.

98. Yoshioka, Y., et al., *SMYD3-mediated lysine methylation in the PH domain is critical for activation of AKT1.* Oncotarget, 2016. **7**(46): p. 75023-75037.

99. Doughan, M., et al., *SMYD proteins in immunity: dawning of a new era.* AIMS Biophys, 2016. **3**(4): p. 450-455.

100. Zhou, Z., et al., *SMYD3-NY, a novel SMYD3 mRNA transcript variant, may have a role in human spermatogenesis.* Ann Clin Lab Sci, 2005. **35**(3): p. 270-7.

101. Van Aller, G.S., et al., *Structure-Based Design of a Novel SMYD3 Inhibitor that Bridges the SAM-and MEKK2-Binding Pockets.* Structure, 2016. **24**(5): p. 774-781.

102. Brown, M.A., et al., *C-terminal domain of SMYD3 serves as a unique HSP90-regulated motif in oncogenesis.* Oncotarget, 2015. **6**(6): p. 4005-19.

103. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-802.

104. Grant, B.J., et al., *Bio3d: an R package for the comparative analysis of protein structures.* Bioinformatics, 2006. **22**(21): p. 2695-6.

105. Glykos, N.M., *Software news and updates. Carma: a molecular dynamics analysis program.* J Comput Chem, 2006. **27**(14): p. 1765-8.

106. Wriggers, W. and K. Schulten, *Protein domain movements: detection of rigid domains and visualization of hinges in comparisons of atomic coordinates.* Proteins, 1997. **29**(1): p. 1-14.

107. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-8, 27-8.

108. Sethi, A., et al., *Dynamical networks in tRNA:protein complexes.* Proc Natl Acad Sci U S A, 2009. **106**(16): p. 6620-5.

109. Jiang, Y., et al., *Purification of Histone Lysine Methyltransferase SMYD2 and Co-Crystallization with a Target Peptide from Estrogen Receptor alpha.* Methods Mol Biol, 2016. **1366**: p. 207-217.

110. Petoukhov, M.V., et al., *New developments in the ATSAS program package for small-angle scattering data analysis.* J Appl Crystallogr, 2012. **45**(Pt 2): p. 342-350.

111. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.* Biophys J, 1999. **76**(6): p. 2879-86.

112. Panjkovich, A. and D.I. Svergun, *Deciphering conformational transitions of proteins by small angle X-ray scattering and normal mode analysis.* Phys Chem Chem Phys, 2016. **18**(8): p. 5707-19.

113. Holcomb, J., et al., *Structural insights into PDZ-mediated interaction of NHERF2 and LPA(2), a cellular event implicated in CFTR channel regulation.* Biochem Biophys Res Commun, 2014. **446**(1): p. 399-403.

114. Peserico, A., et al., *A SMYD3 Small-Molecule Inhibitor Impairing Cancer Cell Growth.* J Cell Physiol, 2015.

115. Chandramouli, B., et al., *Smyd3 open & closed lock mechanism for substrate recruitment: The hinge motion of C-terminal domain inferred from mu-second molecular dynamics simulations.* Biochim Biophys Acta, 2016. **1860**(7): p. 1466-74.

116. Munz, M., J. Hein, and P.C. Biggin, *The role of flexibility and conformational selection in the binding promiscuity of PDZ domains.* PLoS Comput Biol, 2012. **8**(11): p. e1002749.

117. Motlagh, H.N., et al., *The ensemble nature of allostery.* Nature, 2014. **508**(7496): p. 331-9.

118. Fenwick, R.B., S. Esteban-Martin, and X. Salvatella, *Understanding biomolecular motion, recognition, and allostery by use of conformational ensembles.* Eur Biophys J, 2011. **40**(12): p. 1339-55.

119. Esteban-Martin, S., et al., *Correlated inter-domain motions in adenylate kinase.* PLoS Comput Biol, 2014. **10**(7): p. e1003721.

120. Schwede, T., et al., *SWISS-MODEL: An automated protein homology-modeling server.* Nucleic Acids Res, 2003. **31**(13): p. 3381-5.

121. del Sol, A., et al., *Residues crucial for maintaining short paths in network communication mediate signaling in proteins.* Mol Syst Biol, 2006. **2**: p. 2006 0019.

122. Ray Chaudhuri, A. and A. Nussenzweig, *The multifaceted roles of PARP1 in DNA repair and chromatin remodelling.* Nat Rev Mol Cell Biol, 2017. **18**(10): p. 610-621.

123. Durkacz, B.W., et al., *(ADP-ribose)n participates in DNA excision repair.* Nature, 1980. **283**(5747): p. 593-6.

124. de Murcia, J.M., et al., *Requirement of poly(ADP-ribose) polymerase in recovery from DNA damage in mice and in cells.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7303-7.

125. Tulin, A. and A. Spradling, *Chromatin loosening by poly(ADP)-ribose polymerase (PARP) at Drosophila puff loci.* Science, 2003. **299**(5606): p. 560-2.

126. Ogata, N., et al., *Poly(ADP-ribose) synthetase, a main acceptor of poly(ADP-ribose) in isolated nuclei.* J Biol Chem, 1981. **256**(9): p. 4135-7.

127. D'Amours, D., et al., *Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions.* Biochem J, 1999. **342 (Pt 2)**: p. 249-68.

128. Huletsky, A., et al., *The effect of poly(ADP-ribosyl)ation on native and H1-depleted chromatin. A role of poly(ADP-ribosyl)ation on core nucleosome structure.* J Biol Chem, 1989. **264**(15): p. 8878-86.

129. Kraus, W.L. and J.T. Lis, *PARP goes transcription.* Cell, 2003. **113**(6): p. 677-83.

130. Piao, L., et al., *Regulation of poly(ADP-Ribose) polymerase 1 functions by post-translational modifications.* Front Biosci (Landmark Ed), 2018. **23**: p. 13-26.

131. Oei, S.L. and Y. Shi, *Transcription factor Yin Yang 1 stimulates poly(ADP-ribosyl)ation and DNA repair.* Biochem Biophys Res Commun, 2001. **284**(2): p. 450-4.

132. Kun, E., E. Kirsten, and C.P. Ordahl, *Coenzymatic activity of randomly broken or intact double-stranded DNAs in auto and histone H1 trans-poly(ADP-ribosylation), catalyzed by poly(ADP-ribose) polymerase (PARP I).* J Biol Chem, 2002. **277**(42): p. 39066-9.

133. Kim, M.Y., et al., *NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1.* Cell, 2004. **119**(6): p. 803-14.

134. Voelkel, T., et al., *Lysine methyltransferase Smyd2 regulates Hsp90-mediated protection of the sarcomeric titin springs and cardiac function.* Biochim Biophys Acta, 2013. **1833**(4): p. 812-22.

135. Lanouette, S., et al., *Discovery of substrates for a SET domain lysine methyltransferase predicted by multistate computational protein design.* Structure, 2015. **23**(1): p. 206-15.

136. Kabsch, W., *Xds.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 125-32.

137. Evans, P.R. and G.N. Murshudov, *How good are my data and what is the resolution?* Acta Crystallogr D Biol Crystallogr, 2013. **69**(Pt 7): p. 1204-14.

138. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.

139. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.

140. Scheufler, C., et al., *Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine.* (0092-8674 (Print)).

141. Zhang, M., et al., *Ubiquitin-specific Peptidase 10 (USP10) Deubiquitinates and Stabilizes MutS Homolog 2 (MSH2) to Regulate Cellular Sensitivity to DNA Damage.* J Biol Chem, 2016. **291**(20): p. 10783-91.

142. Bomberger, J.M., R.L. Barnaby, and B.A. Stanton, *The deubiquitinating enzyme USP10 regulates the post-endocytic sorting of cystic fibrosis transmembrane conductance regulator in airway epithelial cells.* J Biol Chem, 2009. **284**(28): p. 18778-89.

143. Soncini, C., I. Berdo, and G. Draetta, *Ras-GAP SH3 domain binding protein (G3BP) is a modulator of USP10, a novel human ubiquitin specific protease.* Oncogene, 2001. **20**(29): p. 3869-79.

144. Draker, R., E. Sarcinella, and P. Cheung, *USP10 deubiquitylates the histone variant H2A.Z and both are required for androgen receptor-mediated gene activation.* Nucleic Acids Res, 2011. **39**(9): p. 3529-42.

145. Deng, M., et al., *Deubiquitination and Activation of AMPK by USP10.* Mol Cell, 2016. **61**(4): p. 614-624.

146. Reverdy, C., et al., *Discovery of specific inhibitors of human USP7/HAUSP deubiquitinating enzyme.* Chem Biol, 2012. **19**(4): p. 467-77.

147. Altun, M., et al., *Activity-based chemical proteomics accelerates inhibitor development for deubiquitylating enzymes.* Chem Biol, 2011. **18**(11): p. 1401-12.

148. Liu, J., et al., *Beclin1 controls the levels of p53 by regulating the deubiquitination activity of USP10 and USP13.* Cell, 2011. **147**(1): p. 223-34.

149. Zhang, W., et al., *Generation and Validation of Intracellular Ubiquitin Variant Inhibitors for USP7 and USP10.* J Mol Biol, 2017. **429**(22): p. 3546-3560.

150. Papadopoulos, J.S. and R. Agarwala, *COBALT: constraint-based alignment tool for multiple protein sequences.* Bioinformatics, 2007. **23**(9): p. 1073-9.

151. Robert, X. and P. Gouet, *Deciphering key features in protein structures with the new ENDscript server.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W320-4.

152. Hu, M., et al., *Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde.* Cell, 2002. **111**(7): p. 1041-54.

153. Hu, M., et al., *Structure and mechanisms of the proteasome-associated deubiquitinating enzyme USP14.* EMBO J, 2005. **24**(21): p. 3747-56.

154. Avvakumov, G.V., et al., *Amino-terminal dimerization, NRDP1-rhodanese interaction, and inhibited catalytic domain conformation of the ubiquitin-specific protease 8 (USP8).* J Biol Chem, 2006. **281**(49): p. 38061-70.

155. Yin, J., et al., *Structural Insights into WD-Repeat 48 Activation of Ubiquitin-Specific Protease 46.* Structure, 2015. **23**(11): p. 2043-54.

156. Gersch, M., et al., *Mechanism and regulation of the Lys6-selective deubiquitinase USP30.* Nat Struct Mol Biol, 2017. **24**(11): p. 920-930.

157. Ernst, A., et al., *A strategy for modulation of enzymes in the ubiquitin system.* Science, 2013. **339**(6119): p. 590-5.

158. Bian, Y., et al., *Global screening of CK2 kinase substrates by an integrated phosphoproteomics workflow.* Sci Rep, 2013. **3**: p. 3460.

159. Zhang, Y., et al., *Conformational stabilization of ubiquitin yields potent and selective inhibitors of USP7.* Nat Chem Biol, 2013. **9**(1): p. 51-8.

160. Gonsiorek, W., et al., *Pharmacological characterization of Sch527123, a potent allosteric CXCR1/CXCR2 antagonist.* J Pharmacol Exp Ther, 2007. **322**(2): p. 477-85.

161. Wang, S., et al., *CXCR2 macromolecular complex in pancreatic cancer: a potential therapeutic target in tumor growth.* Transl Oncol, 2013. **6**(2): p. 216-25.

162. Addison, C.L., et al., *The CXC chemokine receptor 2, CXCR2, is the putative receptor for ELR+ CXC chemokine-induced angiogenic activity.* J Immunol, 2000. **165**(9): p. 5269-77.

163. Lu, G., et al., *Structural insights into neutrophilic migration revealed by the crystal structure of the chemokine receptor CXCR2 in complex with the first PDZ domain of NHERF1.* PLoS One, 2013. **8**(10): p. e76219.

164. Fouassier, L., et al., *Evidence for ezrin-radixin-moesin-binding phosphoprotein 50 (EBP50) self-association through PDZ-PDZ interactions.* J Biol Chem, 2000. **275**(32): p. 25039-45.

165. Chang, B.H., et al., *A systematic family-wide investigation reveals that ~30% of mammalian PDZ domains engage in PDZ-PDZ interactions.* Chem Biol, 2011. **18**(9): p. 1143-52.

166. Mikelis, C.M., et al., *PDZ-RhoGEF and LARG are essential for embryonic development and provide a link between thrombin and LPA receptors and Rho activation.* J Biol Chem, 2013. **288**(17): p. 12232-43.

167. Holcomb, J., et al., *Crystal structure of the NHERF1 PDZ2 domain in complex with the chemokine receptor CXCR2 reveals probable modes of PDZ2 dimerization.* Biochem Biophys Res Commun, 2014. **448**(2): p. 169-74.

168. Tonikian, R., et al., *A specificity map for the PDZ domain family.* PLoS Biol, 2008. **6**(9): p. e239.

169. Cumming, R.C., et al., *Protein disulfide bond formation in the cytoplasm during oxidative stress.* J Biol Chem, 2004. **279**(21): p. 21749-58.

ABSTRACT

ALTERNATIVE STRATEGIES TO INHIBIT LYSINE METHYLTRANSFERASES AND DEUBIQUITINASES IN HUMAN CANCERS

by

NICHOLAS SPELLMON

JUNE 2018

Advisor: Dr. Zhe Yang

Major: Biochemistry and Molecular Biology

Degree: Doctor of Philosophy

X-ray crystallography is the gold standard method for imagining macromolecules to atomic resolution. Three dimensional data is central to understanding the molecular mechanism how DNA, RNA and proteins function in biological events. Structural insights into these events provide a molecular window to visualize how biological molecules influence human health. Visualizing the architecture of these molecules set the stage for rational and selective drug design. The following dissertation utilizes biochemical and biophysical tools, including X-ray crystallography, to shed light on poorly understood mechanisms related to SMYD2 activity and regulation, USP10 architecture and function, and PDZ-RhoGEF dimerization.

SMYD2 is one member of the SET and MYND domain-containing protein (SMYD) family known to play key roles in cardiac function and development, innate immunity and tumorigenesis. While the molecular pathways involved in these events have been fairly described, the molecular mechanism of substrate recognition and bilobal changes have not. In this dissertation, I review the structure and function of SMYD protein family. In addition, I demonstrate SMYD2 and SMYD3 can exist in open and closed conformations based on X-ray crystallography, small angle X-ray

scattering, and molecular dynamic simulations data. Lastly, I revealed a novel binding site in SMYD2 that appears to be the first recognition site for SMYD methylation clients.

USP10 is one member of the ubiquitin-specific protease family important for DNA repair and apoptosis by recycling cytosolic p53. However, in the mutant p53 environment, USP10 serves as an oncogene; thereby promoting mutant p53-dependent cancer cell growth. Additional studies found related USP10 oncogene roles in other cancers. Unfortunately the biochemistry and structure of USP10 hasn't been thoroughly explored. My dissertation aims to understand the biochemistry and architecture of the catalytic domain of USP10 along with reported USP10 inhibitors which would be valuable for future studies to probe USP10 function and inhibition.

PDZ-RhoGEF is one member of the Rho guanine exchange factors (RhoGEF) family important for modulating Rho activity and actin-based cytoskeleton remodeling. PDZ-RhoGEF possesses a PDZ domain known for complexing with the cytoplasmic tail of Plexin B serving as modulator for downstream signaling factors. In our study, we found PDZ-RhoGEF complexes with the Interleukin-8 chemokine receptor, CXCR2. This novel interaction hasn't been reported before, and in my dissertation, I solved the crystal structure of PDZ-RhoGEF in complex with the PDZ motif of CXCR2. Unexpectedly, we identified a disulfide bond linking two PDZ-RhoGEF molecules. This disulfide bond was previously reported to be important for promoting PDZ-ligand binding between PDZ-RhoGEF and Plexin B2 peptides. Here, I describe the architecture of the disulfide-linked PDZ domain of PDZ-RhoGEF in complex with two CXCR2 PDZ-motifs.

AUTOBIOGRAPHICAL STATEMENT

Education

2008-2012 B.S. Biochemistry, Oakland University, Rochester, MI
2012-2014 M.Sc. Biochemistry and Molecular Biology; Wayne State University, Detroit, MI
2014-pres. Ph.D. Biochemistry and Molecular Biology; Wayne State University, Detroit, MI

Positions and Honors

2016-2017 Member, American Society for Biochemistry and Molecular Biology
2017 ASBMB Travel Award
2017 Experimental Biology 2017 poster presentation

Peer Reviewed Publications

Holcomb J, Doughan M, **Spellmon N***, Lewis L, Perry E, Zhang Y, Smith S, Nico L, Wan J, Chakravarthy S, Weifang S, Miao Q, Stemmler T, Yang Z. "SAXS analysis of a soluble cytosolic NgBR construct including extracellular and transmembrane domains" PLoS. 2018

Spellmon N*, Li C, Yang Z. "Allosterically targeting EGFR drug-resistance gatekeeper mutations" J Thorac Dis. 2017 Jul;9(7):1756-1758. doi: 10.21037/jtd.2017.06.43.

Kuiper BD, Slater K, **Spellmon N**, Holcomb J, Medapureddy P, Muzzarelli KM, Yang Z, Ovadia R, Amblard F, Kovari IA, Schinazi RF, Kovari LC. Increased activity of unlinked Zika virus NS2B/NS3 protease compared to linked Zika virus protease" Biochem Biophys Res Commun. 2017 Oct 28;492(4):668-673. doi: 10.1016/j.bbrc.2017.03.108. Epub 2017 Mar 22.

Spellmon N*, Holcomb J, Niu A, Choudhary V, Sun X, Zhang Y, Wan J, Doughan M, Hayden S, Hachem F, Brunzelle J, Li C, Yang Z. "Structural basis of PDZ-mediated chemokine receptor CXCR2 scaffolding by guanine nucleotide exchange factor PDZ-RhoGEF" Biochem Biophys Res Commun. 2017 Apr 1;485(2):529-534. doi: 10.1016/j.bbrc.2017.02.010. Epub 2017 Feb 6.

Spellmon N*, Sun X, Xue W, Holcomb J, Chakravarthy S, Shang W, Edwards B, Sirinupong N, Li C, Yang Z. "New open conformation of SMYD3 implicates conformational selection and allostery" AIMS Biophys. 2017;4(1):1-18. doi: 10.3934/biophy.2017.1.1. Epub 2016 Dec 20.

Spellmon N*, Sun X, Sirinupong N, Edwards B, Li C, Yang Z. "Molecular Dynamics Simulation Reveals Correlated Inter-Lobe Motion in Protein Lysine Methyltransferase SMYD2" PLoS One. 2015 Dec 30;10(12):e0145758. doi: 10.1371/journal.pone.0145758. eCollection 2015.

Spellmon N*, Holcomb J, Trescott L, Sirinupong N, Yang Z. "Structure and function of SET and MYND domain-containing proteins" Int J Mol Sci. 2015 Jan 8;16(1):1406-28. doi: 10.3390/ijms16011406. Review.