1-1-2014

Structural Basis Of Epigenetic Regulation And Protein Scaffolding In Development And Diseases

Yuanyuan Jiang
Wayne State University,

STRUCTURAL BASIS OF EPIGENETIC REGULATION AND PROTEIN SCAFFOLDING IN DEVELOPMENT AND DISEASES

by

YUANYUAN JIANG

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

2014

MAJOR: BIOCHEMISTRY AND MOLECULAR BIOLOGY

Approved by:

_____

Advisor Date

_____

_____

_____

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the help of so many people in so many ways. First, I would like to express my deepest appreciation to my advisor, Dr. Zhe Yang for all his guidance and support during all these years. His passion, dedication and preciseness for science have allowed me to see what a real scientist should be and have put him as a role model for my future research. I am very fortunate to have him as my PhD mentor. My gratitude also goes to my committee members, Drs. Brian Edwards, Timothy Stemmler and Ashok Bhagwat for their guidance and suggestions during the course of my projects. It is such an honor for me to have these wonderful people on my committee. I also thank to Dr. Chunying Li and his lab members for their generous help with our joint NHERF1 project. I would also like to thank Joseph Brunzelle for helping us collect all the diffraction data at APS, Argonne National Lab. Special thanks to Dr. Chuan-Pu Lee for all the heartwarming encouragement and useful advices.

I thank all our former and current lab members, who have made contribution to these projects, for their help, suggestions and encouragement throughout the years. My deep thanks go to Dr. Nualpun Sirinupong for training me, and always being supportive from all aspects. She was the one who always found time to help me even when she was very busy. I also thank Guorong Lu who has been really helpful with my NHERF1 project. I thank the Department of Biochemistry and Molecular Biology for offering me the great learning experience as a PhD student.

Finally, I would like to thank my family and all my friends for their incredible love and encouragement. Without their unconditional support, I could not have accomplished what I got today.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PART I

Structural Insights into Epigenetic Regulation Revealed by Crystal Structures of The Histone Methyltransferase SMYD2

CHAPTER 1 General Introduction

1.1 Epigenetic regulation

Epigenetics can be described as the study of heritable changes in gene activity without changing the primary DNA sequence, such as DNA methylation and histone post-translational modification[5]. Epigenetic regulation is emerging as one of the most important research areas since it is greatly involved in a variety of biological processes such as signal transduction, cell cycle control, and stress response[6]. Abnormal epigenetic regulation has been linked with various diseases such as diabetes, heart disease, cancer and Alzheimer's disease[7-9].

1.1.1 Direct vs. indirect epigenetic mechanisms

Epigenetic regulation can be divided into two groups according to the mode of epigenetic modifiers involved in the regulation process. The most common type is the direct regulation, in which epigenetic regulators such as histone acetyltransferase and DNA methyltransferase act directly on chromatin to alter chromatin structure and subsequent gene expression. However, some epigenetic regulators can have dual

functions so that besides directly affecting chromatin, they can interact with proteins that are involved in the regulation of epigenetic gene, such as sequence-specific DNA transcription factors. In addition, some histone-modifying enzymes can directly modify transcription factors involved in gene regulation. For example, the histone lysine methyltransferase SET7/9 regulates the expression of p53 target genes through direct methylation on p53[10]. Similarly, the histone lysine methyltransferase SMYD2 represses the expression of the estrogen receptor target genes via ERα methylation[11].

1.1.2 Epigenetic signaling

Growing evidence indicates a coplay between epigenetic mechanisms and signaling pathways that establish transcriptional programs. Signaling pathways affect critical components of the epigenetic machinery; on the other hand, the epigenetic mechanisms are involved in signaling transduction regulation[6]. However little is known about molecular mechanisms regulating the crosstalk between signal transduction and epigenetic regulation, or the relationship between chromatin-associated proteins and essential signaling pathways. It is anticipated that studying the connection between cell signaling and epigenetic regulation will assist us in comprehending the intricate process of cellular transcriptional changes caused by the presence of external and internal signals.

1.1.3 Histone methylation

Posttranslational modifications of histones represent an essential epigenetic regulatory mechanism, which affects the chromatin structure and DNA accessibility, thereby controls gene transcription. There are many types of histone modifications including acetylation, methylation, phosphorylation, ubiquitination, glycosylation, sumoylation, ADP-ribosylation and carbonylation [12]. These modifications have been demonstrated to be essential for cell growth and development. Among these modifications, histone methylation regulates the fundamental processes such as heterochromatin formation, X chromosome inactivation, genomic imprinting, transcriptional regulation and DNA repair [13-16]. Disrupting the balance of histone methylation can lead to the altered expression of genes involved in tumorigenesis including proto-oncogenes and cell cycle regulators. Although the potential impact of histone methylation on health is self-evident, there are still many gaps in knowledge regarding to the enzymes that are responsible for histone methylation.

1.2 SMYD proteins

1.2.1 Overview

Members of the SET and MYND domain-containing (SMYD) family of proteins possess histone lysine methyltransferase capacity and are involved in the transcriptional control of cell differentiation and cell proliferation[17-21]. The SMYD protein family consists of five proteins (SMYD1–5) which are grouped based on the presence of two conserved domains (MYND and SET domains)[19](Figure 1). The MYND domain is a zinc finger motif that is involved in protein−protein interaction[22]. The SET domain is an

evolutionarily conserved motif responsible for adding methyl groups to lysine residues of protein using S-adenosylmethionine (AdoMet) as donor substrate. The function of the C-terminal domain (CTD) is still poorly understood. In all known SMYD structures, the CTD adopts a helix-turn-helix structure, which is similar to the architecture of the tetratricopeptide repeat (TPR) motifs. Since the TPR motifs mediate specific protein-protein interactions and the assembly of multi-protein complexes, the structural similarity between the CTD and the TPR motifs may suggest a potential function for the CTD as a protein-protein interaction module.

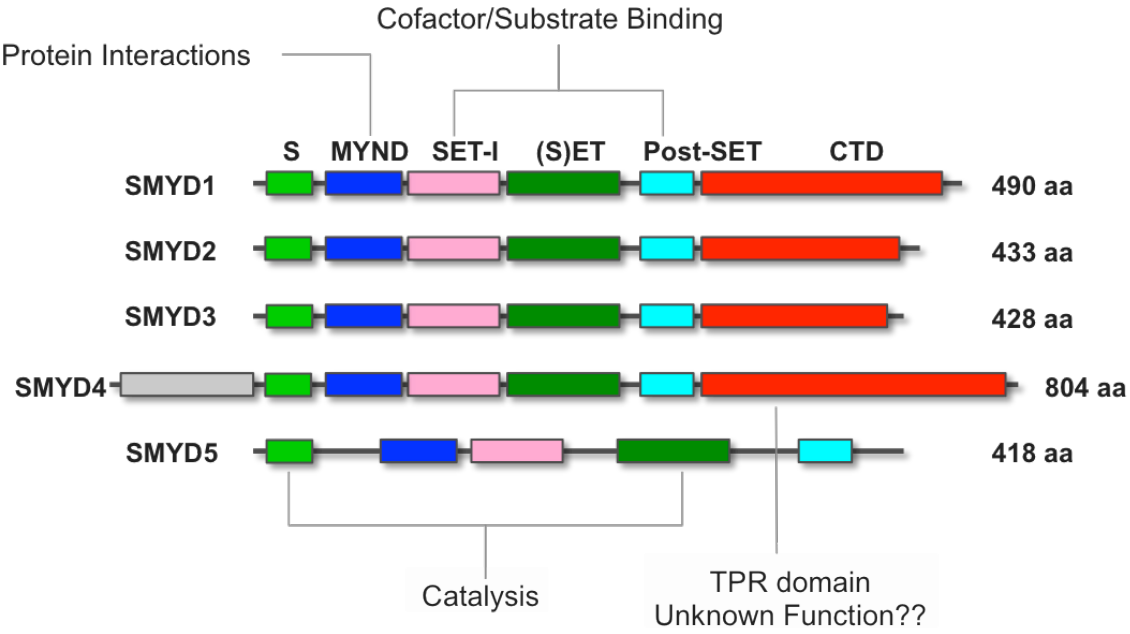

Figure 1. Schematic diagram of SMYD proteins domain structures and their related functions.

1.2.2 SMYD proteins in heart and muscle development

SMYD proteins are essential in the transcriptional regulation of cell differentiation and cell proliferation[17, 18, 20, 21, 23-26]. Evidence for a critical role of SMYD proteins during organ development was first shown by the constitutive knockout of SMYD1 in mice, resulting in early embryonic lethality due to disruption of cardiac differentiation and morphogenesis. Similar results have been observed in zebrafish that knockdown of *SMYD1* causes severe myofibrillar disorganization and malfunction of cardiac and skeletal muscles[23, 27]. Subsequent reports have indicated that SMYD proteins are indeed critical regulators of cardiac and skeletal muscle development[28-31]. In mice, SMYD2 methylates Hsp90 forming a complex with titin, a sarcomeric protein, to protect myocyte organization[32, 33].

1.2.3 SMYD proteins in cancers

Mounting evidence suggests that SMYD proteins play important roles in cancer development. Oncogenic activity of SMYD3 has been observed in many human cancers, such as breast cancer, hepatocellular cancer, and colorectal carcinomas[26]. SMYD4 has been identified as a potential tumor suppressor involved in breast cancer development. Expression levels of SMYD4 are significantly lower in breast cancer cells than healthy mammary cells. Disruption of one allele of the *SMYD4* gene through chromosome translocation resulted in tumorigenesis[34]. High level of SMYD2 expression have been observed in a number of human cancers[35, 36]. Overexpression of SMYD2 caused changes in expression of genes associated with cell cycle regulation and transcription regulation. A recent study has shown that SMYD2 assists in maintaining the self-renewal activity of

MLL-AF9-induced acute myeloid leukemia[37]. SMYD2 has also been shown to represses transcriptional p53 activity by lysine methylation (Lys370), exerting an oncogenic and drug resistance action through inhibition of p53-mediated cell death pathways[38]. In addition to p53 methylation, studies showed that the retinoblastoma tumor suppressor (RB), a central cell cycle regulator and tumor suppressor, can be methylated by SMYD2, which regulates the RB activity during cell cycle progression, cellular differentiation, and in response to DNA damage[39]. In agreement with these observations, *SMYD2* recently has been reported as a cancer-promoting gene through activation or overexpression in esophageal squamous cell carcinoma[36]. Together, these findings suggest the therapeutic potential of SMYD proteins in a range of human cancers.

1.3 Estrogen signaling

1.3.1 Overview

Estrogen signaling pathway controls cellular responses to estrogen and regulates gene transcription in diverse developmental processes[40]. Estrogen signaling is mediated by two estrogen receptors, ERα and ERβ, which are ligand-activated transcription factors and belong to the nuclear receptor superfamily[41]. Upon activation by estrogen, estrogen receptors form a dimer, which binds directly to the estrogen responsive elements (EREs) and then activates downstream gene expression [41-43] (Figure 2). Estrogen binding also induces a conformational change in the receptors, which allows the recruitment of a number of coregulators, including coactivators and corepressors, for specific regulation of gene activation and repression[41].

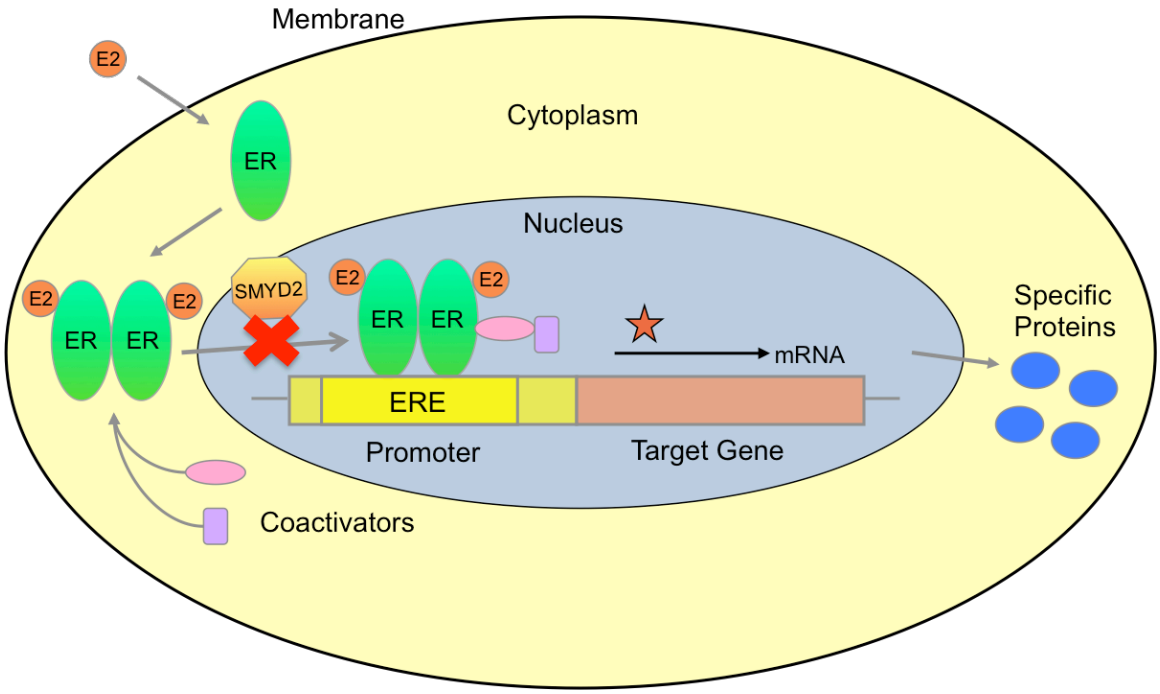

Figure 2. Classical estrogen signaling pathway.

Dysregulation of estrogen signaling can lead to a number of human diseases, including a variety of human cancers such as breast, ovarian, colorectal, prostate, and endometrial cancers, and also other diseases such as endometriosis, fibroids, and cardiovascular disease[44]. Thus, detailed understanding of the complex regulatory machinery underlying estrogen signaling has become extremely important.

1.3.2 Epigenetic regulation of estrogen signaling

Regulation of the transcriptional response of estrogen stimulation relies on both direct and indirect epigenetic signaling. Epigenetic regulators perform posttranslational modification either directly on histones or indirectly on estrogen receptors to alter ER target gene expression. For example, in response to hormone stimulation, the coactivator

complex p300/CBP gets recruited to the promoters of estrogen responsive genes where it acetylates local histones causing nucleosomal destabilization and recruitment of transcription factors[45]. In addition, the recruitment of the mixed lineage leukemia histone methylases (MLLs) is required for ERα transcriptional activity, and the knockdown of MLLs abolishes H3K4 trimethylation resulting in significant suppression of the estrogen-induced HOXC13 activation[46].

Besides the direct action on chromatin, some epigenetic enzymes regulate estrogen signaling via posttranslational modification of the estrogen receptors. Such modifications include acetylation, phosphorylation, methylation, sumoylation and ubiquitylation. These modifications are associated with distinct biological outcomes of ER-mediated signaling. Acetylation of ERα on K266/268 in the hinge region enhanced the DNA binding and transactivation activities of ERα, whereas acetylation on K302/303 represses target gene expression. In addition, ubiquitination at K302 and sumoylation at K266/268 have also been shown to affect estrogen receptor stability and activity [47-49]. Patients who have ERα S118 and/or S167 phosphorylated often have a better response to tamoxifen (Tam) therapy, while phosphorylation of S305 often results in tamoxifen resistance[50]. Recent identification of several ERα methylation sites has further expanded our knowledge in the role of the posttranslational modification in ER regulation. SET7/9 regulates ERα activity and stability via methylation on K302[51]. K260 of ERα was discovered as the target of protein arginine methyltransferase 1 (PRMT1) in the cytoplasm of normal and malignant epithelial breast cells[52]. We recently showed that the histone H3K4/H3K36 methyltransferase SMYD2 regulates ERα transactivation by K266

methylation, which attenuates ERα chromatin recruitment and prevents ERα target gene activation under an estrogen-depleted condition[11]. This finding identified a previously undefined inhibitory methylation event, contributing to a substantial body of evidence that posttranslational modifications of ERα provide complex and combinatorial regulation that assures the protein to be tightly regulated and coordinating the appropriate transcriptional response. Moreover, it highlights the importance of uncovering the structural basis of the SMYD2-mediated ERα methylation, as a necessary prerequisite of discovering small molecules that could fine-tune ERα activity or stimulate the restoration of normal ERα-dependent transcription programs. These findings, together with other epigenetic regulatory mechanisms[38], demonstrate an intricate relationship between epigenetics and estrogen signaling that function cooperatively to specify ER transcriptional consequences.

1.4 Concluding remarks and objectives

Epigenetic regulation plays an essential role in many biological processes, including cell-cycle regulation, development and differentiation. Since the dysregulation of SMYD proteins has been found in various diseases, these proteins have rapidly emerged as attractive therapeutic targets for drug discovery. There has been a growing interest in various mechanisms that involve epigenetic regulation, including modulation of SMYD proteins activities. However, little is known about the mechanisms underlying the function regulation of SMYD proteins. Structural studies have revealed new insights into the mechanisms behind the lysine methylation machinery and possible allosteric

regulation that may involve conformational flexibility. Future studies that can further uncover the molecular mechanisms of SMYD proteins will not only help us better understand the mechanism of the epigenetic inheritance, but also provide us great benefits in the development of alternative therapeutic strategies.

CHAPTER 2

Crystal Structures of Histone and p53 Methyltransferase SMYD2 Reveal A Conformational Flexibility of The Autoinhibitory C-terminal Domain

*Published in PLoS ONE 2011;6(6):e21640. doi:10.1371/journal.pone.0021640. All authors agreed with including their work in this dissertation.

Abstract

SMYD2 belongs to a new class of chromatin regulators that control gene expression in heart development and tumorigenesis. Besides methylation of histone H3 K4, SMYD2 can methylate non-histone targets including p53 and the retinoblastoma tumor suppressor. The methyltransferase activity of SMYD proteins has been proposed to be regulated by autoinhibition via the intra- and interdomain bending of the conserved C-terminal domain (CTD). However, there has been no direct evidence of a conformational change in the CTD. Here, we report two crystal structures of SMYD2 bound either to the cofactor product S-adenosylhomocysteine or to the inhibitor sinefungin. SMYD2 has a two-lobed structure with the active site located at the bottom of a deep crevice formed between the CTD and the catalytic domain. By extensive engagement with the methyltransferase domain, the CTD stabilizes the autoinhibited conformation of SMYD2 and restricts access to the catalytic site. Unexpectedly, despite that the two SMYD2 structures are highly superimposable, significant differences are observed in the first two helices of the CTDs: the two helices bend outwards and move away from the catalytic

domain to generate a less closed conformation in the sinefungin-bound structure. Although the overall fold of the individual domains is structurally conserved among SMYD proteins, SMYD2 appears to be a conformational "intermediate" between a closed form of SMYD3 and an open form of SMYD1. In addition, the structures reveal that the CTD is structurally similar to tetratricopeptide repeats (TPR), a motif through which many cochaperones bind to the heat shock protein Hsp90. Our results thus provide the first evidence for the intradomain flexibility of the TPR-like CTD, which may be important for the activation of SMYD proteins by Hsp90.

Introduction

Covalent histone modifications represent an important regulatory mechanism controlling gene transcription, essential for normal growth and development[16]. Disrupting the balance of histone modifications can lead to the altered expression of genes involved in tumorigenesis including proto-oncogenes and cell cycle regulators[53]; however, little is known about how the enzymes that control histone modifications are regulated posttranslationally. Members of the SET and MYND domain containing (SMYD) family of proteins possess histone lysine methyltransferase capacity and have been shown to be involved in the transcriptional control of cell differentiation and cell proliferation[53-57]. The SMYD protein family consists of five proteins (SMYD1–5) that share about 30% sequence identity with each other and are grouped based on the presence of two conserved domains (MYND and SET domains)[55]. The MYND domain is a zinc finger motif that is involved in protein–protein interaction[22]. The SET domain is an

evolutionarily conserved motif consisting of about 130 amino acids that is responsible for adding methyl groups to lysine residues of proteins using S-adenosylmethionine (AdoMet) as a donor substrate.

Evidence for a critical role of SMYD proteins during organ development was first shown by the constitutive knockout of SMYD1, resulting in early embryonic lethality due to disruption of cardiac differentiation and morphogenesis[54]. Subsequent reports have further indicated that SMYD proteins are indeed critical regulators of cardiac as well as skeletal muscle development[29, 56, 58-60]. Despite being highly expressed in heart and brain, a specific functional role for SMYD2 in these organs has not been well characterized[53, 59]. Overexpression of SMYD2 has been shown to cause changes in expression of genes associated with chromatin remodeling, cell cycle, and transcription regulation, indicating that this protein may function as a transcriptional regulator by methylating H3 K4 and participates in cell cycle regulation and cell growth[55]. Interest in SMYD2 has grown significantly because of recent reports indicating that SMYD2 repress transcriptional p53 activity by lysine methylation (Lys370), exerting an oncogenic and drug resistance action through inhibition of p53-mediated cell death pathways[38]. In addition to p53 methylation, a new study showed that the retinoblastoma tumor suppressor (RB), a central cell cycle regulator and tumor suppressor, can also be methylated by SMYD2 at lysine 860, which regulates the RB activity during cell cycle progression, cellular differentiation, and in response to DNA damage[61]. In agreement with these observations, SMYD2 recently has been shown to act as a cancer-promoting gene through activation or overexpression in

esophageal squamous cell carcinoma[36]. These studies thus support a role for SMYD2 in the regulation of proliferation and in tumor progression, which underscores the importance of elucidating the regulation of SMYD2 activity.

The molecular chaperone Hsp90 plays an important role in the folding, activation, intracellular transport, and assembly of a broad range of client proteins, specifically chaperoning molecules involved in signal transduction and cell cycle regulation[62]. Mounting evidence showed that Hsp90 is also involved in transcriptional regulation and epigenetic inheritance by interacting with epigenetic proteins that function in chromatin remodeling and histone modifications[63, 64]. Based on the ability of Hsp90 to stimulate the activity of SMYD proteins, recent studies have characterized SMYD proteins as new clients of Hsp90[55, 64]; however, the critical questions regarding how Hsp90 activates SMYD proteins remain poorly understood. Previous studies suggested that the methyltransferase activity of SMYD proteins is suppressed by an autoinhibited conformation maintained by the CTD, a helix bundle C-terminal to the catalytic SET domain that is conserved and unique in SMYD proteins[65, 66]. It has been proposed that the intra- and interdomain bending of the CTD may be central for the activation of SMYD proteins by Hsp90[66]. In this paper, we report two crystal structures of full-length SMYD2 in complex with the methyltransferase inhibitor sinefungin (SFG) and the cofactor product S-adenosylhomocysteine (AdoHcy). Our studies demonstrate for the first time the intradomain flexibility of the CTD and reveal the structural resemblance of the autoinhibitory CTD to tetratricopeptide repeat (TPR) motif, which suggest a mechanism

for the Hsp90-mediated activation of SMYD proteins. Our findings therefore contribute to the understanding of the mechanism that regulates the activity of SMYD proteins in early heart development and tumorigenesis.

Materials and Methods

Protein Preparation

Protein purification was performed essentially as described previously[65]. Briefly, mouse *SMYD2* was cloned into the pSUMO vector (LifeSensors), with an N-terminal His6-SUMO tag. Recombinant *SMYD2* was then transformed into *Escherichia coli* for protein expression. The transformants were grown to an OD_{600} (optical density at 600 nm) of 0.4 at 37°C in 2 L 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 a series of chromatography purification by an AKTA purifier system (GE healthcare), and the His6-SUMO tag was cleaved off with yeast SUMO Protease 1. SMYD2 proteins were finally purified to apparent homogeneity and concentrated to 10–20 mg/ml in 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1 mM β-mercaptoethanol, and 5% glycerol.

Crystallization and data collection

Prior to crystallization, SMYD2 (10 mg/ml) was incubated with 2 mM AdoHcy or sinefungin at 4 °C for 2 h. The binary complex of SMYD2–AdoHcy or SMYD2–sinefungin was then crystallized by hanging drop vapor diffusion at 20 °C, with

15% polyethylene glycol 8000, 50 mM NaCl, 100 mM Tris, pH 8.0. Crystals typically appeared within 1 day, achieved their full size in a week. X-ray diffraction data from single crystals were collected at beamline 21IDD at the Advanced Photon Source (Argonne, IL) and were then processed and scaled using the program HKL2000[67]. The crystals belong to the orthorhombic space group $P2_12_12_1$ and contain one molecule in the asymmetric unit (Table 1).

Structure determination and refinement

The crystal structure of SMYD2 in complex with AdoHcy was solved by the single-wavelength anomalous diffraction (SAD) method using three intrinsic zinc ions. Initial phases were obtained using the program SOLVE[68], which was able to identify all three zinc sites with a figure of merit of 0.329 in the resolution range 20–2.1 Å. After density modification with the program RESOLVE[68], the resulting electron density map is interpretable. With the modified phases, automated model building was carried out by RESOLVE, which built 80% of the protein residues including side chains. The model was then completed and improved by alternating cycles of manual model building and refinement using COOT[69] and BUSTER[70]. The final refined model is well ordered with the exception of the first two residues and the last residue. Because of isomorphism of crystals (Table 1), the crystal structure of SMYD2 in complex with sinefungin was solved by rigid-body fitting of the SMYD2–AdoHcy model followed by manual model building and refinement as described above. The final models were analyzed and validated with

PROCHECK[71]. All figures of 3D representations of the SMYD2 structures were made with PyMOL (www.pymol.org).

Protein Data Bank accession number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 3QWV and 3QWW for SMYD2−AdoHcy and SMYD2−SFG, respectively.

Results and Discussion

SMYD2 structure with the TPR-like CTD

Two crystal structures of full-length SMYD2 in complex with the cofactor product AdoHcy and the methyltransferase inhibitor sinefungin have been determined at 2.1 Å and 1.8 Å by zinc single-wavelength anomalous dispersion (Table 1). Similar to SMYD1 and SMYD3[65, 66], SMYD2 has a multidomain structure that folds into two lobes with overall dimensions of approximately 65 Å x 40 Å x 55 Å (Figure 3). Although the overall fold of their individual domains is structurally conserved, the SMYD family proteins differ dramatically in the relative orientation between the N- and C-terminal lobes. Detailed description of the structural differences will be addressed later in the article. The N-terminal lobe (residues 3−279) is composed of four domains: the catalytic SET domain, located in the middle of this lobe, is surrounded by the zinc finger MYND, insertion SET-I, and post-SET domains. Immediately C-terminal to the post-SET domain, the polypeptide forms a large domain of about 150 residues that constitutes the C-

Table 1. Crystallographic data and refinement statistics

	Sinefungin	**AdoHcy**
Space group	$P2_12_12_1$	$P2_12_12_1$
Cell parameters (Å)		
a	57.5	57.9
b	75.1	75.0
c	112.5	113.4
Wavelength (Å)	0.97872	1.28215
Resolution (Å)	30.0-1.8	30.0-2.03
R_{merge}[a]	0.083 (0.503)[b]	0.102 (0.512)
Redundancy	6.0 (5.8)	11.3 (10.3)
Unique reflections	45863	32539
Completeness (%)	99.7 (99.5)	99.9 (99.1)
$\langle I/\sigma \rangle$	8.9 (2.8)	9.5 (4.6)
Refinement		
Resolution (Å)	30-1.8	30-2.03
Molecules/AU	1	1
R_{work}[c]	0.186 (0.224)	0.173 (0.193)
R_{free}[d]	0.208 (0.275)	0.215 (0.226)
RMSD		
Bond lengths (Å)	0.010	0.010
Bond angels (°)	1.0	1.1
No. of atoms		
Protein	3465	3453
Sinefungin/AdoHcy	27	26
Water	429	392
Zinc	3	3
B-factor (Å^2)		
Protein	25.8	29.0
Sinefungin/AdoHcy	12.8	19.7
Water	33.4	35.5
Ramachandran plot		
Preferred regions (%)	97.12	96.92
Allowed regions (%)	2.88	3.08
Outliers (%)	0.0	0.0

[a] $R_{merge} = \Sigma |I-\langle I \rangle| / \Sigma I$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity of multiple observations of symmetry-related reflections.
[b] Numbers in parentheses refer to the highest resolution shell.
[c] $R_{work} = \Sigma |F_o-F_c| / \Sigma |F_o|$, where F_o is the observed structure factor, F_c is the calculated struture factor.
[d] R_{free} was calculated using a subset (5%) of the reflection not used in the refinement.

terminal lobe (residues 280–432). This domain is conserved in the SMYD proteins and was referred to as the CTD in our previous studies[65]. The CTD is composed of seven antiparallel α-helices (αH–αN) rotated relative to one another by an approximately 25°. This topology creates a right-handed superhelical structure generating a concave surface on one side with a convex surface on the other. Despite the absence of any significant sequence similarities, the overall fold of the CTD is reminiscent of that of TPR repeats that adopt a helix-turn-helix structure. Given that the TPR repeats mediate specific protein–protein interactions and the assembly of multiprotein complexes, the structural similarity of the CTD and the TPR repeats suggests a function for the CTD as a protein-protein interaction module.

The architecture of the catalytic SET domain of SMYD2 is essentially similar to that of SMYD1 and SMYD3[65, 66], which features a "split" domain defined by two separated segments, the S-sequence (residues 3–49) and the core SET domain (residues 183–246). Despite the split in the primary structure, the SET domain in SMYD2 has the similar overall fold to other SET domain containing proteins, characterized by one central 3_{10} helix ($3_{10\text{-}3}$) and 10 β-strands (β1–β5 and β8–β12) that are arranged into four antiparallel β-sheets (Figure 3). Of particular importance are the loop connecting $3_{10\text{-}3}$ and β10 that contributes conserved catalytic residues and functions to bind the cofactor at the bottom of the cofactor binding site, and the strand β8 and the loop following β10 that form a narrow cleft predicted to accommodate substrate H3 peptide (Figure 4). However, the SET domain alone is not sufficient for lysine methylation and it requires the

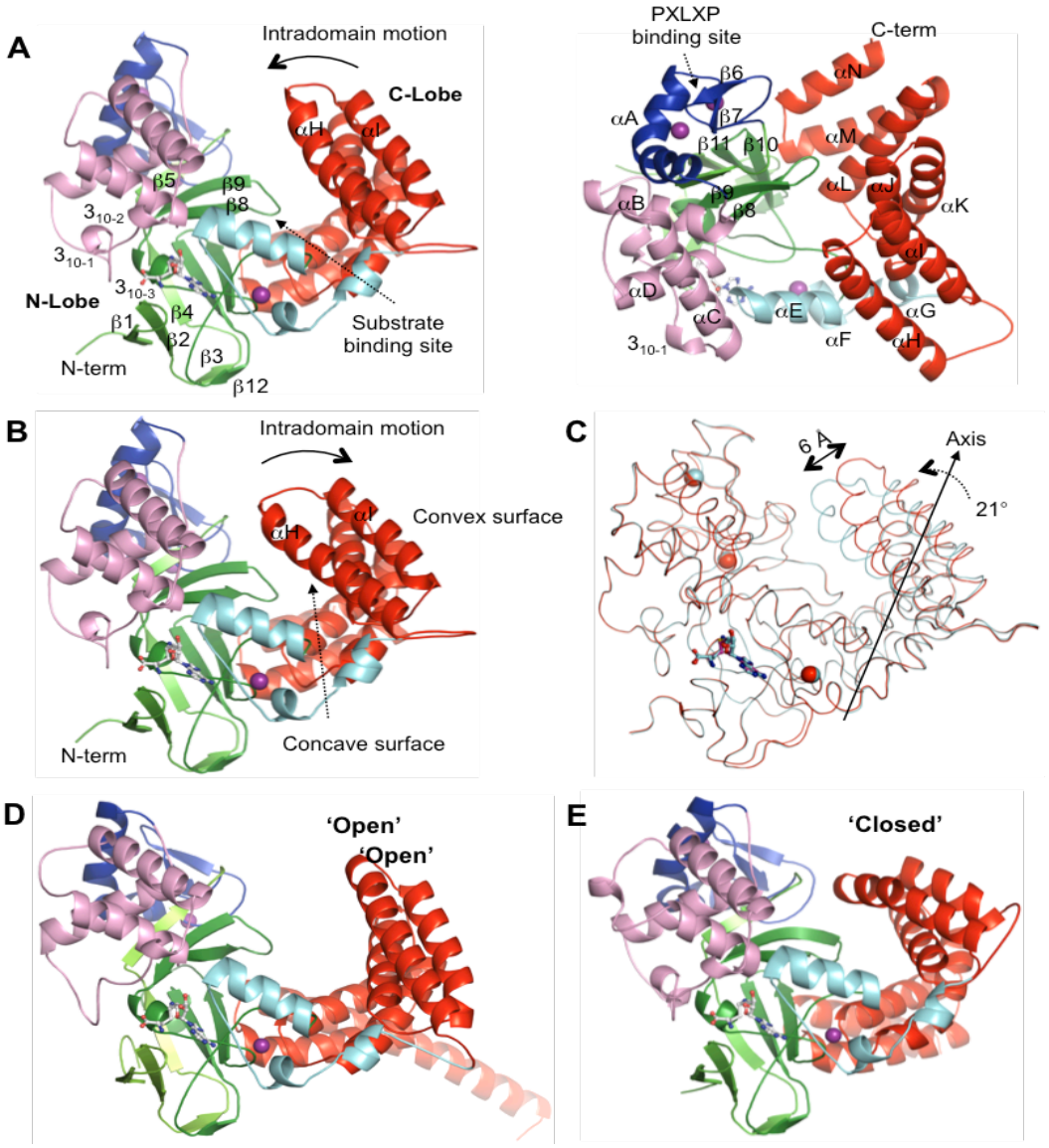

Figure 3. Ribbon diagram of the SMYD2 structures. (A) Side view (left) and top view (right) of the binary structure of SMYD2–sinefungin. (B) The structure of SMYD2–AdoHcy. Secondary structures of SMYD2, α-helices, 310-helices, and β-strands are labeled and numbered according to their position in the sequence. The S-sequence, MYND, SET-I, core SET, post-SET, and CTD are depicted in light green, blue, pink, green, cyan, and red, respectively, while sinefungin and AdoHcy are represented by balls-and-sticks and zinc ions are denoted by purple spheres. (C) Superposition of two SMYD2 structures in complex with sinefungin (red) and AdoHcy (cyan) based on their N-lobes. The maximum distance between the equivalent regions in the outer edge of their C-lobes is indicated. The intradomain motion is indicated by the straight arrow and the approximate rotation angle is given. (D) Ribbon diagram of the structure of SMYD1 and (E) SMYD3 with the domains colored the same as above.

cooperation with three other domains, including the N- and C-terminal flanking domains (pre-SET and post-SET) as well as the insertion SET-I domain[72-74]. The latter three domains are not conserved with highly variable structures in the known SET proteins but they occupy similar positions and play similar roles in these enzymes. Interestingly, SMYD2 does not contain the pre-SET domain, though this domain is required by other SET proteins to stabilize the SET domain fold or provide an extended histone binding site[73].

Both post-SET and SET-I domains are engaged in cofactor and substrate binding[72]. The post-SET domain, which is immediately downstream of the SET domain, is a small cysteine-rich region consisting of three short α-helices (αE, αF, and αG) that are organized around a single zinc ion (Figure 3). The zinc ion is coordinated by four highly conserved cysteine residues: Cys262, Cys264, and Cys267 from the post-SET domain and Cys209 from the SET domain. This zinc ion thus appears to be important for the folding of the post-SET domain and also tethers this domain to the SET domain. As a result of this tethering, the post-SET domain lies close to the active site, with the loop connecting αE and αF placed near the cofactor, and the C-terminal end of helix αE positioned to participate in the formation of the substrate binding cleft. Similar to other SMYD proteins[65, 66], SMYD2 has a large SET-I domain consisting of a helix bundle (αB, $3_{10\text{-}1}$, $3_{10\text{-}2}$, αC, and αD) of as many as 84 residues, together with the MYND inserted between the SET strands β5 and β8 (Figure 3). The equivalent region in Set7/9 or Dim-5, however, contains only one or two small helices of 15–20 residues[75, 76]. In contrast to the

MYND, the SET-I domain packs against the opposite face of the β-sheet containing β4, β10, and β11, contributing to the cofactor and substrate binding. Specifically, the last two helices (αC and αD) of the SET-I domain might be important for the recognition of the H3 N-terminal residues (Figure 4), while the loop between the $3_{10\text{-}1}$ and $3_{10\text{-}2}$ helices makes extensive contacts with the cofactor.

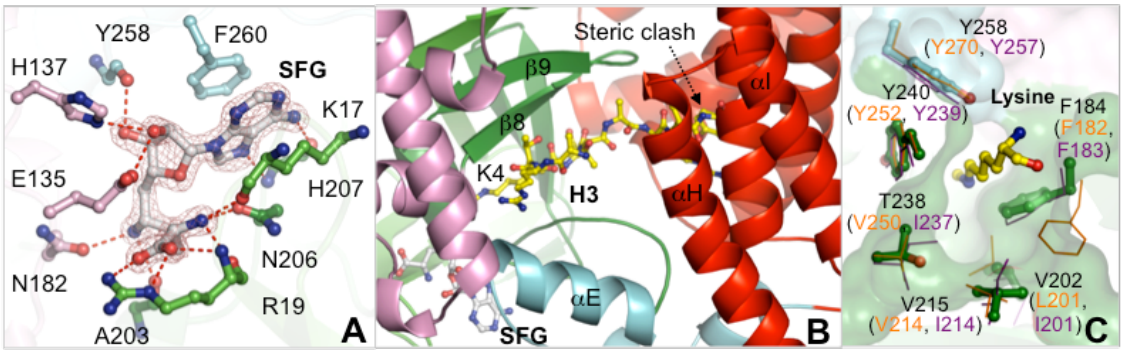

Figure 4. Cofactor binding pocket and substrate binding site.
(A) Interaction between SMYD2 and sinefungin. SMYD2 residues are represented by balls-and-sticks with their carbon atoms colored according to the scheme in Figure 1. Sinefungin is depicted by balls-and-sticks overlaid with 2Fo−Fc omit map calculated at 1.8 Å and contoured at 2.5 σ. Hydrogen bonds are illustrated as red broken lines. (B) Ribbon diagram of the putative substrate binding site, illustrating the interaction between SMYD2 and the modeled H3 peptide. The H3 peptide (1–10) from the Set7/9 structure (PDB code 1O9S) is displayed as balls-and-sticks with carbon atoms colored yellow. (C) Superposition of the target lysine-access channels of SMYD2, SMYD1, and SMYD3. The oval-shaped channel in SMYD2 is depicted by molecular surface. Residues in SMYD2 are represented by balls-and-sticks, while residues in SMYD1 and SMYD3 are displayed as sticks in purple and orange, respectively. Target lysine (H3K4) is colored in yellow.

MYND mediates protein–protein interactions by binding to a proline-rich sequence[22]. It has been demonstrated that the MYND present in SMYD2 interacts with proteins containing the PXLXP motif, such as EBP41L3, a functional suppressor of

epithelial ovarian cancers[55]. As shown in Figure 3A, the MYND domain consists of one kinked α helix (αA) and two antiparallel β-strands (β6, β7) that are organized around 2 zinc ions. Although it forms direct contacts with the catalytic SET domain, the MYND does not contribute residues to cofactor binding. In addition, this domain is more than 10 Å away from the putative substrate-binding pocket and may not be directly involved in substrate recognition (Figure 4). These observations are in agreement with previous findings that the MYND is dispensable for the histone methylation activity of SMYD2[55], implicating that the MYND may primarily function as a protein–protein interaction module and coordinate SMYD2 with other proteins to regulate tumor proliferation and progression. The structure of the MYND of SMYD2 is very similar to that of SMYD1, SMYD3 and AML1/ETO[22, 65, 66], with the following pairwise RMSDs for C_α atoms over 40 residues: 0.48 Å, 0.53 Å, and 0.81 Å, respectively. Superposition of the MYNDs of SMYD2 and AML1/ETO, which was solved in complex with a peptide containing the "PPPLI" motif[22], reveals that the proline-rich peptide is located in a shallow, fully exposed surface groove that is readily accessible by other proteins. One side of the groove is formed by a loop connecting β6 and β7, and the other side by the residues from the N-terminal half of helix αA. Three highly conserved residues (Trp80, Gln76, and Tyr70 in SMYD2), which are critical for AML1/ETO binding to the peptide are highly superimposable in the two structures. The high structural similarity suggests a similar mode of recognition of proline-rich sequences shared by these two MYNDs.

Active site characterized by a spacious target lysine access channel

We have determined two structures of SMYD2 bound either to the cofactor product AdoHcy or to a potent methyltransferase inhibitor sinefungin. The two structures are remarkably similar to each other in terms of cofactor binding. Therefore, the following discussion on the interaction between SMYD2 and the cofactor will be solely focused on the SFG-bound SMYD2 structure. Similar to that in other SMYD proteins[65, 66], the L-shaped sinefungin binds in a deep surface pocket formed by the SET-I, SET and post-SET domains (Fig 2A). In particular, the adenine moiety of sinefungin is sandwiched between the benzyl ring of Phe260 and the aliphatic side chain of Lys17, with its purine N6 and N7 atoms hydrogen-bonding to the backbone carbonyl and amide groups of His207, respectively. The ribose hydroxyls of the cofactor make three hydrogen bonds with the side chains of His137 and Glu135 and the carbonyl oxygen of Tyr258. At the opposite end of sinefungin, the positively charged α-amino group is recognized by a trigonal array of hydrogen bonds with the main chain carbonyl oxygens of Lys17 and Arg19 and the amide Oδ of Asn206, while the carboxylate moiety forms salt-bridge interactions with the guanidinium group of Arg19. The latter electrostatic interactions are present in most SET proteins including SMYD1 but are replaced by a hydrogen bond to a tyrosine residue in SMYD3, which represents an unusual variation[66]. In the middle of sinefungin, the C–NH3 amine group, which is in place of the S–CH3 sulfonium of AdoMet, engages in two hydrogen bonds with the backbone oxygen of Ala203 and the amide Oδ of Asn182. The similar interactions are expected in the case of AdoMet, which might contribute to enzymatic function by destabilizing the active methyl group. Collectively, the overall cofactor-binding mode of SMYD2 is structurally

conserved with SMYD1 and SMYD3 and other SET enzymes and serves to orient the methyl group of AdoMet into the methyltransfer pore during catalysis.

Although the SMYD2 structures were solved without substrate, superposition of SMYD2 with histone H3-bound Set7/9, a H3 lysine 4 methyltransferase, offers insights into substrate recognition. As shown in Figure 4B, the modeled H3 peptide binds in a deep, rectangle-shaped cleft formed by the SET, post-SET and SET-I domains. In the cleft, the β8 strand and the loop preceding the post-SET domain are predicted to interact with substrate histone in a hybrid β-sheet-binding mode as shown in other SET proteins[72]. Lys4 of the peptide is at the center of this β-sheet interaction with its side chain inserted into the target lysine access channel that leads to sinefungin that binds on the opposite face of the SET domain. Comparison of SMYD2, SMYD1, and SMYD3 reveals that the structures of the lysine access channel of these enzymes are similar to each other with a large oval-shaped opening (Figure 4C). The residues in SMYD2 involved in the formation of the channel including Tyr240, Tyr258, Val202, Val215, and Thr238, are highly structurally aligned with the equivalent residues in SMYD1 and SMYD3, except for Phe184. The spacious lysine access channel is a characteristic feature of SMYD proteins, which is mainly attributed to the replacement of some bulky aromatic residues in Set7/9 or other SET proteins by small hydrophobic ones in SMYD proteins[65, 66]. In SMYD1, substitution of Val214 by tyrosine, a mutation that would create a tighter active site pocket, results in a significant increase in H3 binding and also enhances SMYD1 methylation, indicating that this large channel made SMYD1 unable to

effectively interact with the target lysine during methyl transfer, affecting its enzymatic activity[65].

A unique feature of SMYD proteins is the presence of the conserved CTD, which is located near the substrate binding cleft and together with the SET domain forms a deep canyon that spans the entire molecule[65, 66]. Similar to SMYD3, the putative substrate-binding site of SMYD2 is located at the bottom of the 15-Å-deep crevice, with the CTD acting like a lid and partially covering the active site pocket (Figure 3). However, because of the location of the CTD, severe steric clashes are observed between the C-terminus of the peptide and the CTD inner surface in the SMYD2–H3 model (Figure 4B). The steric hindrance of the CTD suggests that the CTD prevents H3 binding and it may be required to move away to allow substrate entry and efficient catalysis. Alternatively, this might be an indication that the H3 peptide may adopt a different conformation when binding to SMYD2. Considering the potential motion of the CTD, it is also likely that the CTD conformation observed in the crystal structures represents a non-physiological state of the protein. Importantly, mutation or deletion of the CTD significantly increased both substrate binding and H3 methylation by SMYD1, demonstrating that this domain plays a negative role in the regulation of the protein's activity[65]. Together with previous functional studies[55, 65], these observations support the idea that the histone methyltransferase activity of SMYD proteins is regulated by autoinhibition that involves the conserved CTD.

Maintenance of SMYD2 autoinhibited conformation

SMYD2 methylates histone H3 to a very limited extent both *in vitro* and *in vivo*[55], and extensive interactions between the CTD and the SET domain appear to contribute to the maintenance of the autoinhibited state of SMYD2 (Figure 5). Specifically, the interactions involve contacts mediated by the turns connecting the CTD helices, which form a contiguous ridge that is anchored to the concave face of the β-sheet containing β4, β10, and β11. In addition, the residues within the antiparallel β-hairpin between β8 and β9 appear to play a central role in the interaction with the CTD. This hairpin protrudes deep into the middle of the concave face of the CTD, braced by the CTD helices and forming numerous direct interactions with αH, αL, αM, and αN. In contrast, the equivalent hairpin in SMYD1 interacts only with the last helix (αN) from the CTD, separated by a large crevice from the other CTD helices. In particular, the aliphatic side chain of Glu189 stacks with the aromatic ring of Tyr422, together with residues Leu191, Leu379, Leu386, Met412, and Ile426 forming a continuous hydrophobic core that stretches from the hairpin down to the bottom of the domain interface. Of particular importance, however, are hydrogen bonds formed between with the guanidinium group of Arg390, which projects from helix αL, and two acidic residues, Glu189 and Glu190 in the β8–β9 hairpin. A similar interaction between the β8–β9 hairpin and the CTD was also observed in SMYD3 but absent in SMYD1 that has an open conformation[65, 66]. Given that the β8–β9 hairpin makes extensive contacts with the CTD, this hairpin is likely to be important in holding the SET domain and the CTD together and maintaining the closed conformation of the substrate-binding cleft.

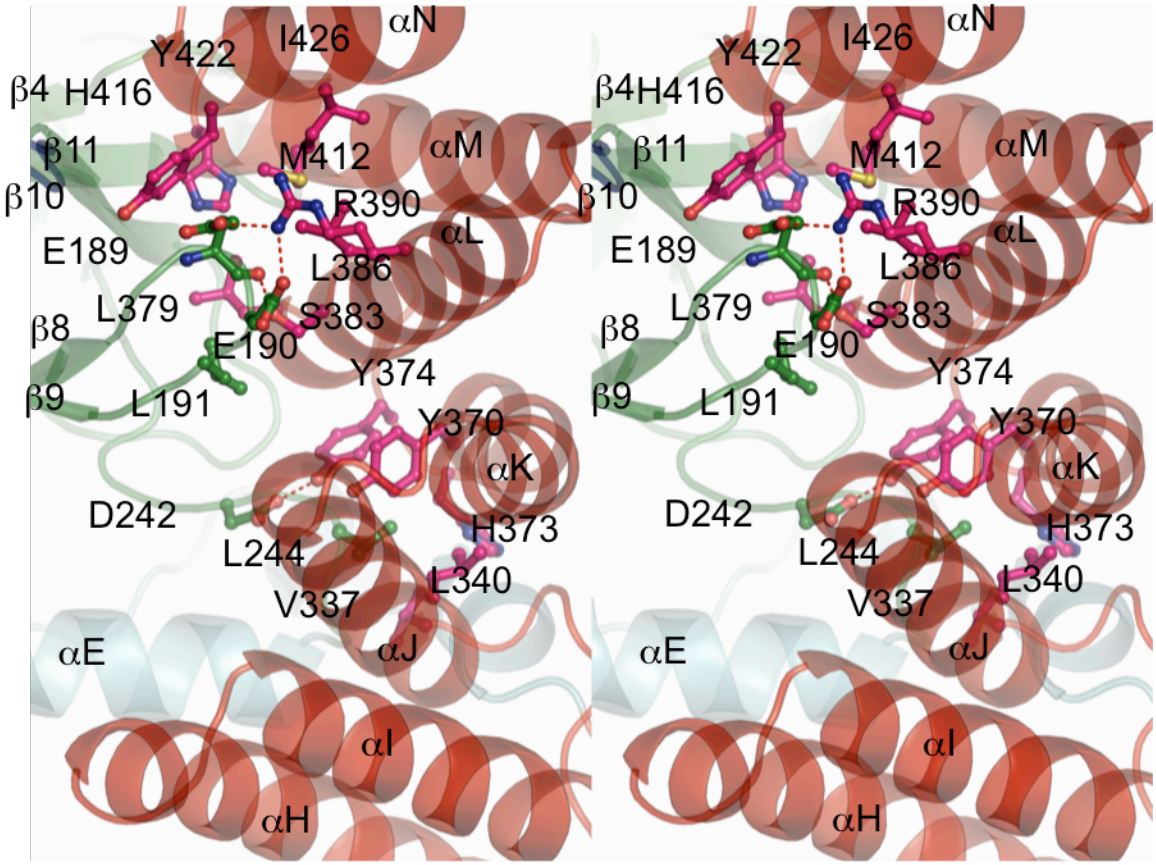

Figure 5. Stereo view ribbon diagram of the domain interface of N- and C-terminal lobes. Residues are colored according to domain in which they reside, and hydrogen bonds are indicated as red dashed lines.

Additional interactions that participate in stabilizing the closed conformation are made among the residues in the loop preceding the post-SET domain and residues in the third and fourth helices (αJ and αK) of the CTD (Figure 5). Specifically, Asp242 forms a hydrogen-bond interaction with Tyr374, while Leu243 participates in a hydrophobic cluster with Val337, Leu340, Tyr370, His373, and Tyr374 from the CTD. Most of these residues are well conserved in SMYD family proteins[65], suggesting that the interactions between them may also contribute to the maintenance of the autoinhibited state.

Interestingly, substitution of the corresponding Asp242 or Tyr370 by alanine is able to destabilize the autoinhibited state of SMYD1, leading to a significant increase in both H3 binding and the enzymatic activity[65].

Intradomain and interdomain flexibility of the conserved CTD

The intra- and interdomain bending of the CTD has been proposed to be central to the release of the autoinhibitory effect exerted by the CTD[66]. However, there has been no direct evidence to support this model. Significantly, the two SMYD2 structures in complex with the cofactor analogs sinefungin and AdoHcy differ dramatically in the conformation of the CTD (Figure 3C). Although the structures of SMYD2−SFG and SMYD2−AdoHcy are highly superimposable with RMSD of 0.36 Å over 400 residues, close examination reveals that the first two helices of the CTD (αH and αI) adopt different conformations. These two helices bend outwards with the loop between the two helices moving ~6 Å further away from the catalytic SET domain. This motion generates a less closed conformation in the SFG-bound SMYD2 structure and slightly tightens the cavity of the active site in SMYD2−AdoHcy. In agreement with the conformational changes, the flexible nature of the αH and αI helices is also indicated by their higher than average isotropic temperature factors of 41.9 $Å^2$ for SMYD2−AdoHcy and 39.5 $Å^2$ for SMYD2−SFG (Table 1). We use the program DynDom to further analyze this domain movement[77]. Two hinge bending motion regions are identified as containing residues 294−300 and 319−322, at which point the αH and αI helices pivot towards the SET

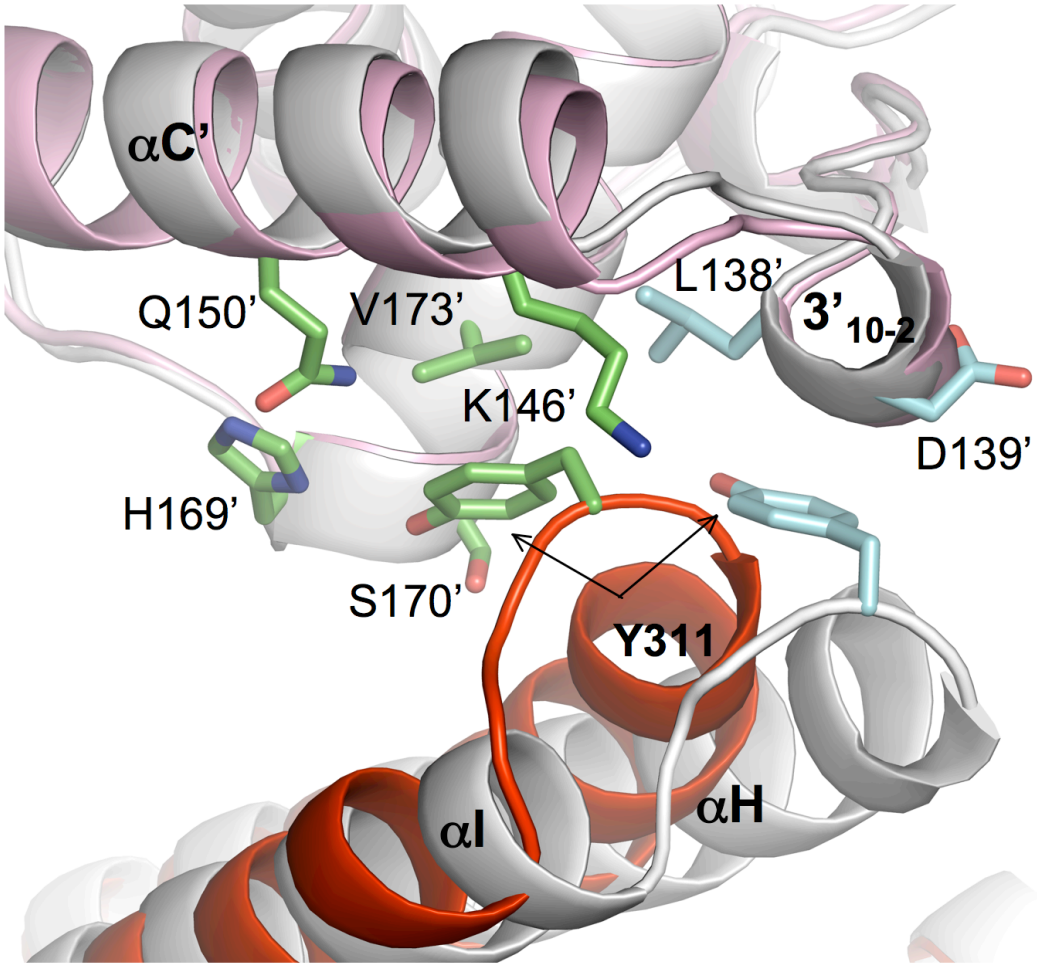

Figure 6. Comparison of the crystal contacts of SMYD2–SFG and SMYD2–SAH. SMYD2–SFG is depicted in red and its symmetry-related molecule in pink. Both SMYD2–SAH molecules are colored gray. Residues involved in crystal contacts are displayed as sticks and colored green and cyan in SMYD2–SFG and SMYD2–SAH, respectively. The prime symbol denotes residues and secondary structures in the symmetry-related molecules.

domain by the rotation. The hinge axis of the rotation runs approximately perpendicular to the axis of the CTD superhelix, intersects the helix αH and is located 1.4 Å from Cα of Arg299 and 1.3 Å from Cα of Asn300. The translation component of the screw operation describing the domain movement is 1.5 Å so that the movement is essentially a pure domain rotation. The crystal packing constraints appear to help to stabilize the

conformational diversity of SMYD2. The overall crystal packing is effectively identical in the SMYD2−AdoHcy and SMYD2−sinefungin complexes, except at the packing interfaces that involve αH and αI (Figure 6). The different orientations of these two helices are stabilized by the differences in crystal packing contacts that contain spatially close but distinct sets of residues. Collectively, these findings provide the first evidence of the intradomain flexibility of the CTD and the structural basis for the model of the conformational changes in the CTD that regulates the activity.

The CTD is located over 30 Å distant from the cofactor and does not contribute residues to cofactor binding. It is not apparent from the structure how such a long-range conformational change is triggered by the cofactors and propagated from the cofactor binding pocket, because of the highly superimposable cofactor binding sites and no significant structural changes in their immediate neighboring regions. There are, however, some differences caused by the CTD motion in the interaction networks between the CTD and post-SET domain, including a new hydrogen bond between the side chains of Arg299 and Glu248 and the potential salt-bridge interactions between Arg306 and Asp256 in SMYD2−AdoHcy. Nevertheless, the long-range conformational change triggered by the exchange of the cofactors could have at least one important functional implication. Sinefungin more resembles AdoMet than AdoHcy in structure, with the C–NH3 amine group in place of the S–CH3 sulfonium. Our findings may then suggest that the binding of the substrate AdoMet to SMYD2 may partially relieve the inhibition by the CTD by causing it to move away from the catalytic domain. The ability

of the conformation changes induced by cofactors appears to be specific to SMYD2. Several structures of SMYD3 have been recently been deposited in the protein data bank including SMYD3−AdoMet, SMYD3−sinefungin, and SMYD3−AdoHcy complexes[66, 78]. Despite marked differences in crystal packing, these SMYD3 complexes display essentially identical structures independent of the types of cofactor, suggesting some differences in allosteric properties among SMYD family members.

The exceptionally large differences in the domain-domain orientation or with respect to the distance separating the N- and C-terminal lobes have been observed between SMYD1 and SMYD3[65, 66]. As a result of the differences, the CTD in SMYD3 adopts a closed conformation that blocks the putative H3K4 binding cleft, whereas the SMYD1 CTD displays an open state with the active site completely exposed. Interestingly, the SMYD2 structures display substantial differences from both SMYD1 and SMYD3 in regard to the CTD orientation. The differences can be viewed when the N-terminal lobes from SMYD2, SMYD1, and SMYD3 are structurally aligned as shown in Figure 3. In this view, the N-terminal lobe remains essentially unchanged, but the CTDs move to either widen or narrow the deep crevice between the N- and C-terminal lobes, essentially mimicking how a clamshell opens and closes. In particular, the C_α atoms of some residues near the outer edge of the CTD move as much as 12 Å between SMYD3 and SMYD1, whereas two SMYD2 structures appear to be a conformational "intermediate" between the close form of SMYD3 and the open form of SMYD1. Although the active site pocket of both SMYD2 and SMYD3 is partially closed by the

CTD that leads to steric clash with the modeled H3 peptide, significant differences are observed in the first two helices of the CTD. The helices equivalent to αH and αI in SMYD3 form direct contact with the linker region between the SET and MYND domains[66], but the SMYD2 structures reveal that these two helices swing outwards and maintain a narrow gap with the SET domain on top of the active site pocket. This structural difference, however, does not cause a significant change in the contact area between the CTD and the rest of protein, with the total buried surface area in the domain interface of 3766 $Å^2$ in SMYD3 compared to 3796 $Å^2$ and 3682 $Å^2$ in SMYD2–AdoHcy and SMYD2–SFG, respectively. Taken together, the differences in the domain–domain orientation between SMYD2 and other SMYD proteins further suggest that the CTD is able to undergo a hinge bending-like motion, which could regulate access to the active site.

A model of SMYD2 activation by Hsp90

It has been reported that interaction between SMYD2 and Hsp90 is important for the histone methyltransferase activity of SMYD2, which is in agreement with results for SMYD1 and SMYD3[55-57]. The manner in which Hsp90 contributes as a cofactor of SMYD proteins is still unclear. Given the differences in the CTD conformations of SMYD proteins, it has been proposed that Hsp90 activates SMYD proteins through the displacement of the autoinhibitory effect of the CTD, which in turn leads to the exposure of the CTD-blocked active site[66]. However, the question regarding the mechanics of how Hsp90 causes the CTD motion remains elusive. Hsp90 is essential for maintaining the

activity of numerous signaling proteins and it plays a key role in cellular signal transduction networks[79]. In fulfilling its role, Hsp90 often operates by interacting with a variety of proteins that contain a TPR domain. At the very C-terminal end of Hsp90 is the TPR motif recognition site, a conserved MEEVD pentapeptide, that is responsible for the interaction with many TRP proteins such as the immunophilins FKBP51/52, the stress induced phosphoprotein Hop, cyclophilin Cyp40, and a protein phosphatase PP5[80].

Interestingly, a search using the Dali server reveals that the conserved CTD, which sterically blocks the substrate binding site, resembles the structure of TPR repeats[81]. The CTD is mainly comprised of three copies of 34-amino acid, helix-turn-helix TPR motifs, including αH–αI, αJ–αK, and αL–αM. As shown by superposition of the CTD of SMYD2 and the TPR2 domain of Hop, the overall configuration of these two domains are similar to each other with RMSD of 3.9 Å over 128 C_α atoms (Figure 7A). The only significant difference is the first two helices of the CTD (αH and αI), which have a different degree of superhelical twists. The structural similarity of the CTD and TPR repeats leads us to hypothesize that the CTD might interact with Hsp90 via the C-terminal MEEVD pentapeptide of the chaperone, which may be important for SMYD2 activation. This hypothesis is in agreement with previous studies showing that Hsp90 interaction with SMYD2 was mediated through a region other than the MYND and SET domains[55]. To assess potential interaction between the CTD and Hsp90, we performed a modeling study using the structure of the TPR2 domain of Hop in complex with a C-terminal pentapeptide MEEVD of Hsp90 (Figure 7A). In the structure of the

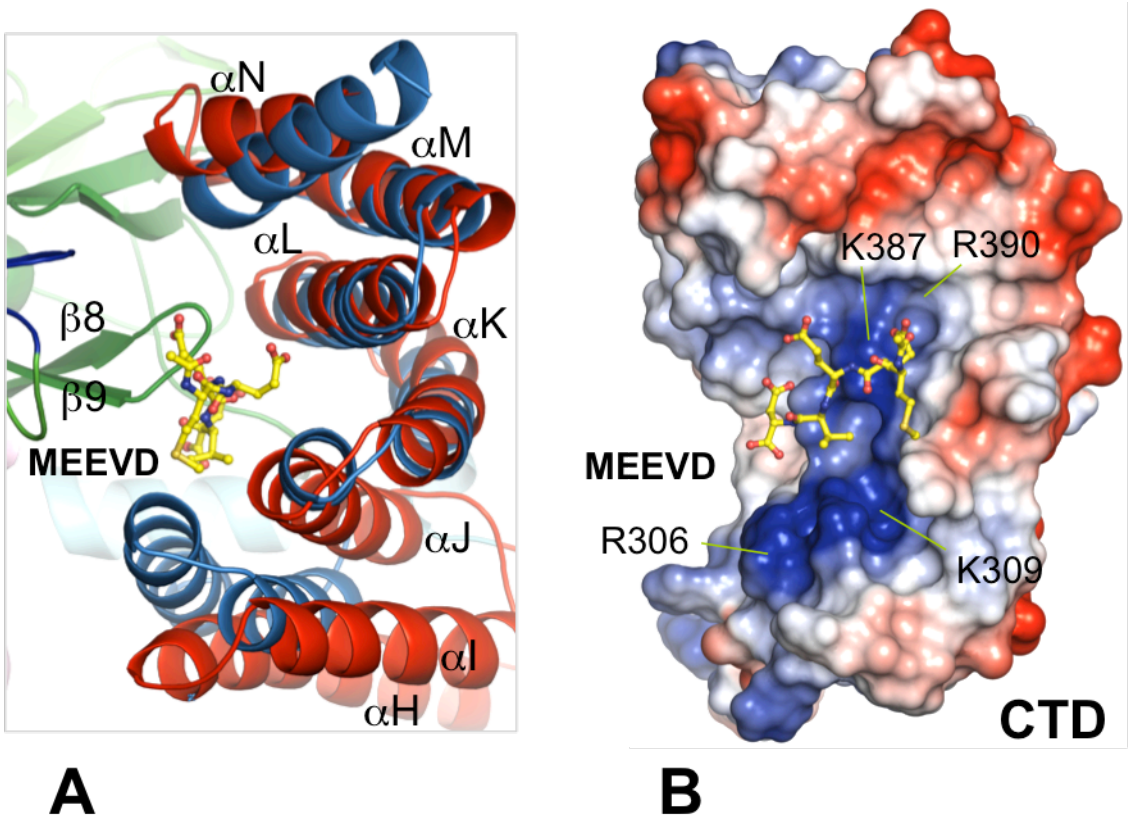

Figure 7. TPR-like CTD. (A) Superposition of the CTD of SMYD2–SFG (red) and the TPR2 domain of Hop (sky blue) (PDB code 1ELR). The Hsp90 MEEVD peptide in complex with the Hop TPR2 domain is displayed as balls-and-sticks with carbon atoms colored yellow. (B) Model of the Hsp90 MEEVD peptide bound in the SMYD2 CTD. The CTD is represented by molecular surface with color coding according to the electrostatic potential: red, white, and blue correspond to negative, neutral, and positive potential, respectively, whereas the peptide is shown as balls-and-sticks. Positively charged residues predicted to be essential for peptide binding are labeled.

Hop–MEEVD complex, the Hsp90 peptide interacts with the Hop TPR2 domain in an extended conformation, with the peptide sequence running parallel with the helices of the TPR motifs[82]. The peptide-protein interactions are primarily dominated by hydrogen bonds and salt-bridges involving the carboxylated groups of acidic residues and the C-terminus of the Hsp90 MEEVD motif interacting with conserved arginine and lysine residues lining the basic peptide-binding channel of Hop. Despite the low sequence

identities of 16% between the CTD and the TPR2 domain, most of the arginine, lysine and asparagine residues responsible for Hop–Hsp90 interactions are structurally conserved in SMYD2, including residues Arg306, Lys309, Gln345, Lys387, and Arg390. By lining up along the concave surface of the CTD, these residues create a continuous positively charged groove predicted for engagement of the Hsp90 acidic C-terminal region (Figure 7B). This putative MEEVD binding site, however, is partially buried and occupied by the loop between strands β8 and β9, the region that is involved in maintaining the autoinhibited state of the protein by interacting with the CTD (Figure 7A). The structural similarity of the CTD to the TPR2 domain together with the buried MEEVD binding site may suggest a mechanism of SMYD2 activation by Hsp90, which may resemble how PP5 is activated by Hsp90. The crystal structure of autoinhibited PP5 reveals that the TPR domain of PP5 engages with the catalytic channel of the phosphatase domain, restricting access to the catalytic site[83]. This autoinhibited conformation of PP5 is stabilized by the C-terminal helix that contacts a region of the Hsp90-binding groove on the TPR domain. Hsp90 activates PP5 by disrupting TPR–phosphatase domain interactions, permitting substrate access to the constitutively active phosphatase domain. Based on these analyses, we propose a model of SMYD2 activation by Hsp90, in which the Hsp90 MEEVD motif could compete with the β8–β9 hairpin for binding to the SMYD2 CTD, displacing the CTD from the substrate binding site and causing a conformational change in the CTD. This model is in agreement with the conformational flexibility of the CTD as revealed by the structural differences between SMYD2–AdoHcy and SMYD2–SFG (Figure 3). Additional research is required to

support this proposed mechanism and to determine whether Hsp90 interacts with the SMYD2 via the CTD and induces a conformational change in this domain.

CHAPTER 3

Structural Insights into Estrogen Receptor Alpha Methylation by Histone Methyltransferase SMYD2, a Cellular Event Implicated in Estrogen Signaling Regulation

*Published in Journal of Molecular Biology 2014 Mar 1. pii: S0022-2836(14)00101-6. doi: 10.1016/j.jmb.2014.02.019. All authors agreed with including their work in this dissertation.

Abstract

Estrogen receptor (ER) signaling plays a pivotal role in many developmental processes and has been implicated in numerous diseases including cancers. We recently showed that direct ERα methylation by the multi-specificity histone lysine methyltransferase SMYD2 regulates estrogen signaling through repressing ERα-dependent transactivation. However, the mechanism controlling the specificity of the SMYD2-ERα interaction and the structural basis of SMYD2 substrate binding diversity are unknown. Here we present the crystal structure of SMYD2 in complex with a target lysine (Lys266)-containing ERα peptide. The structure reveals that ERα binds SMYD2 in a U-shaped conformation with the binding specificity determined mainly by residues C-terminal to the target lysine. The structure also reveals numerous intrapeptide contacts that ensure shape complementarity between the substrate and the active site of the

enzyme, thereby likely serving as an additional structural determinant of substrate specificity. In addition, comparison of the SMYD2-ERα and SMYD2-p53 structures provides the first structural insight into the diverse nature of SMYD2 substrate recognition and suggests that the broad specificity of SMYD2 is achieved by multiple molecular mechanisms such as distinct peptide binding modes and the intrinsic dynamics of peptide ligands. Strikingly, a novel potentially SMYD2-specific PEG binding site is identified in the CTD domain, implicating possible functions in additional substrate binding or protein-protein interactions. Our study thus provides the structural basis for the SMYD2-mediated ERα methylation, and the resulting knowledge of SMYD2 substrate specificity and target-binding diversity could have important implications in selective drug design against a wide range of ERα-related diseases.

Introduction

Estrogen signaling regulates numerous developmental processes and plays important roles in cell growth and differentiation through influencing gene transcription[84]. Abnormal function of this hormonal signaling pathway can lead to many human diseases, including a variety of human cancers such as breast, ovarian, colorectal, prostate, and endometrial cancers, and also other diseases such as endometriosis, fibroids, osteoporosis, and cardiovascular disease[44]. Thus, detailed understanding of the molecular mechanisms underlying estrogen signaling is of clinical and therapeutic importance. In general, estrogen signaling is mediated by two estrogen receptors, ERα and ERβ, which are ligand-activated transcription factors and belong to the nuclear receptor

superfamily[41]. In the classical genomic pathway, the signaling mediated by these receptors begins with binding of estrogen to the receptors ERα or ERβ, and then the ligand-bound receptors dimerize and exert transcriptional control by binding to estrogen response elements in their target genes[41]. Estrogen binding also induces a conformational change in the receptors, which allows the recruitment of a number of coregulators, including coactivators and corepressors, for specific regulation of gene activation and repression[41].

Epigenetic mechanisms are known to play key roles in the regulation of estrogen signaling and contribute to ERα-mediated transcription[85, 86]. One such mechanism is covalent histone modifications including methylation, acetylation, phosphorylation, and ubiquitination on specific N-terminal residues of histones[85]. These modifications usually work together to regulate the functioning of the genome by altering the local structural dynamics of chromatin, primarily regulating its accessibility and compactness[87]. The interplay of these modifications creates an epigenetic landscape that defines distinct chromatin states compatible with either active or repressed gene transcription[5]. It was shown that the estrogen-induced ERα transcriptional outcome is regulated by dynamic interaction with various histone-modifying enzymes, which are generally associated with ER coactivators and corepressors[86]. For instance, optimal ERα-mediated transcription requires the recruitment of the coactivator complex p300/CBP, which acetylates local histones and causes nucleosomal destabilization, consequently facilitating the binding of transcription factors to promoter regions of estrogen responsive genes[45]. In addition, the

recruitment of the mixed lineage leukemia histone methylases (MLLs) is required for ERα transcriptional activity, and the knockdown of MLLs abolishes H3K4 trimethylation resulting in significant suppression of the estrogen-induced HOXC13 activation[46]. These findings, together with other epigenetic regulatory mechanisms[85], demonstrate an intricate relationship between epigenetics and estrogen signaling that function cooperatively to specify ER transcriptional consequences.

Remarkably, some of the histone modifying enzymes are capable of regulating estrogen signaling through direct ERα modification[11], constituting another layer of transcriptional regulation in signaling. For instance, the histone acetyltransferase p300 acetylates ERα at multiple lysine positions and regulates ERα transactivation in an estrogen-dependent fashion[88, 89]. Acetylation of K266/268 by p300 has been shown to promote ERα transactivation activity[88], while K302/303 acetylation has been reported to cause repression of ERα target gene expression[89]. Some of these residues are also subject to methylation catalyzed by histone lysine methyltransferases including SET7 and SMYD2[11, 51]. Methylation of K302 by SET7 has been associated with increased ERα activity and stability, and is essential for the efficient recruitment of ERα to its target genes[51]. We recently showed that the histone H3K4/H3K36 methyltransferase SMYD2 regulates ERα transactivation by K266 methylation, which attenuates ERα chromatin recruitment and prevents ERα target gene activation under an estrogen-depleted condition[11]. This recent finding identified a previously undefined inhibitory methylation event, contributing to a substantial body of evidence that posttranslational modifications

of ERα provide complex and combinatorial regulation that assures the protein to be tightly regulated and coordinating the appropriate transcriptional response. On the other hand, our finding highlights the importance of elucidating the structural basis of the SMYD2-mediated ERα methylation, as a necessary prerequisite of discovering small molecules that could fine-tune ERα activity or stimulate the restoration of normal ERα-dependent transcription programs.

In addition to ERα and histone H3, SMYD2 has been shown to be able to methylate several other histone and nonhistone proteins[90], which establishes SMYD2 as a multifunctional protein playing important roles in diverse cellular processes. Initial identification of SMYD2 as histone H3K4 and H3K36 methyltransferases suggests that SMYD2 functions as an epigenetic regulator involved in transcriptional control of cell proliferation and differentiation[55, 91]. Monomethylation of p53 by SMYD2 links SMYD2 to p53-mediated apoptosis and has been shown to contribute to tumorigenesis through inhibition of p53 transcriptional activity[38]. In addition, recent evidence shows that SMYD2-dependent RB methylation at K810 and K860 plays important roles in tumor progression and growth, capable of regulating RB tumor-suppressing activity during cell cycle progression and cellular differentiation in response to DNA damage[35, 39]. In agreement with these observations, overexpression of SMYD2 has been associated with multiple human malignancies, such as bladder cancer and esophageal carcinoma[35-37], indicating that it may act as a cancer-promoting protein regulating tumor progression via the protein methylation activity. Furthermore, methylation of the heat shock protein

Hsp90 by SMYD2 has been demonstrated as a key cytoplasmic event that stabilizes myofilament organization by promoting complex formation between Hsp90, SMYD2, and the sarcomeric protein titin[32, 33]. These findings together indicate that the diversity in SMYD2 substrate specificity dictates the diversification of its biological function. On the other hand, the complex picture of SMYD2-mediated methylation raises an intriguing problem regarding how SMYD2 achieves broad substrate specificity, the nature of which currently remains elusive.

Limited structural information on SMYD2-substrate complexes represents a major obstacle in understanding the molecular basis by which SMYD2 recognizes a diverse array of functionally different proteins. To date, the only available complex structure is the structure of SMYD2 in complex with the p53 peptide[92, 93]. This is also the only enzyme-substrate structure for the entire SMYD protein family[2, 65, 66]. Analysis of this structure reveals the SMYD2-p53 interaction involves both the catalytic SET domain and the tetratricopeptide repeat (TPR)-like CTD domain, suggesting the cooperative action of these conserved domains may be important for specific p53 recognition[92]. However, this structure alone offers little information on SMYD2 target binding diversity, and the structural principles governing its complex-specific interactions still remain unknown. Understanding the broad substrate specificity of enzymes generally requires structural determination of the proteins bound with various ligands, and comparative analysis of these liganded structures in turn will help to identify diversity determinants responsible for their substrate discrimination[94]. In this study, we present a

new SMYD2 structure in complex with a K266-containing ERα peptide. Comparison of the SMYD2-ERα and SMYD2-p53 structures reveals that the SMYD2 targeting diversity is facilitated by conformational flexibility in its substrate-binding pocket and also depends on intrinsic ligand dynamics of different sequences. Strikingly, our structure suggests the presence of an additional peptide-binding site that might confer an extended substrate-binding mode or allows SMYD2 binding of two different proteins. This study therefore provides important insights into the SMYD2-mediated ERα methylation and could be valuable in the development of novel therapeutic strategies against many ERα-dependent human diseases.

Materials and Methods

Protein Preparation

A DNA fragment encoding the full-length human SMYD2 was amplified by PCR and cloned into the pSUMO vector (LifeSensors). Recombinant SMYD2, which contains a N-terminal His6-SUMO tag, was then transformed into *Escherichia coli* BL21 Condon Plus (DE3) cells for protein expression. The transformants were grown to an OD_{600} (optical density at 600 nm) of 0.4 at 37°C in 2 L LB medium, and then induced with 0.1 mM isopropylthio-β-D-galactoside at 15°C overnight. The cells were harvested by centrifugation and lysed by French Press. The soluble fraction was then subjected to the Ni^{2+} affinity chromatography purification followed by the cleavage of the His6-SUMO tag with the yeast SUMO Protease Ulp1. SMYD2 proteins were separated from the cleaved tag by a second Ni^{2+} affinity chromatography and further purified by the size

exclusion chromatography. Finally, the proteins were concentrated to10–20 mg/ml in 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1 mM β-mercaptoethanol, and 5% glycerol.

Crystallization and Data Collection

Crystallization was performed using the hanging drop vapor diffusion method at 20 °C. Initial SMYD2 crystals were grown by mixing 1 μl of a protein solution containing 10 mg/ml SMYD2 and 2 mM AdoHcy with 1 μl of a well solution containing 15% PEG8000, 50 mM NaCl, and 100 mM Tris–HCl (pH 8.5). Crystals obtained under this condition were crushed and then used as a seed stock for growing SMYD2-AdoHcy-ERα cocrystals. An 11-residue synthetic peptide (GGRMLKHKRQR) corresponding to the ERα residues 261–271 (CPC Scientific) was used in cocrystallization. Prior to cocrystallization, SMYD2 (3 mg/ml) was incubated with 2 mM AdoHcy and 2 mM ERα peptide at 4 °C for 2 h. Complex crystals suitable for data collection were grown by mixing 1 μl of the protein-ERα solution, 0.3 μl of the seed stock, and 0.7 μl of a well solution containing 20% PEG 3350, 100 mM Tris–HCl (pH 8.7), and 5% ethanol. Crystals typically appeared within 1 day and achieved their full size in a week. X-ray data from single crystals were collected at beamline 21-ID-F at the Advanced Photon Source (Argonne, IL) and were processed and scaled using the program XDS[95]. The crystals belong to the tetragonal space group I4 and contain one molecule per asymmetric unit.

Structure Determination and Refinement

The crystal structure of SMYD2 in complex with AdoHcy and ERα peptide was solved by molecular replacement with the program PHASER[96] using the mouse SMYD2-sinfungin structure (PDB code: 3QWW) as a search model. Manual model building was carried out in COOT[69], and refinement was performed with PHENIX[97]. To reduce the effects of model bias, iterative-build OMIT maps have been used during model building and structure refinement. An elongated electron density clearly visible in the Fourier difference map during the last refinement cycles was modeled using coordinates for polyethylene glycol from the HIC-Up database[98]. The final model was analyzed and validated with Molprobity[99]. All figures of 3D representations of the SMYD2-ERα structure were made with PyMOL (www.pymol.org).

Protein Data Bank accession number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 4O6F.

Results and Discussion

Overall Structure of SMYD2-ERα Complex

The crystal structure of SMYD2 in complex with ERα peptide and cofactor product AdoHcy has been determined at 2.82 Å by molecular replacement (Table 2). The structure reveals that the overall fold of SMYD2 is bilobal with two lobes separated by a deep cleft (Figure. 8A). The ERα peptide binds at the bottom of the cleft that connects to the AdoHcy binding pocket located at the opposite face of the molecule (Figure. 8B).

Table 2. Crystallographic data and refinement statistics

Data	
Space group	*I*4
Cell parameters (Å)	
a=b	151.8
c	52.9
Wavelength (Å)	1.0781
Resolution (Å)	75.9-2.82 (3.15-2.82)[a]
R_{merge}[b]	0.110 (0.446)
R_{meas}[c]	0.136 (0.553)
$CC_{1/2}$[d]	0.993 (0.813)
Redundancy	3.8 (3.6)
Unique reflections	45284
Completeness (%)	83.6 (86.4)
$\langle I/\sigma \rangle$	10.8 (2.5)
Refinement	
Resolution (Å)	75.9-2.82
Molecules/AU	1
R_{work}[e]	0.176 (0.254)
R_{free}[f]	0.237 (0.298)
RMSD	
Bond lengths (Å)	0.006
Bond angels (°)	1.49
No. of atoms & B-factor (Å^2)	
Protein (3-432)	3460; 29.8
ERα Peptide (262-271)	90; 29.7
N-terminal residues (262-265)	31; 45.1
C-terminal residues (267-271)	50; 22.9
PEG	25; 27.2
AdoHcy	26; 16.4
Water	94; 21.5
Zinc	3; 23.5
Nickel	2; 72.4

[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]$R_{meas} = \Sigma [(n/n-1)]^{1/2} \Sigma |I-\langle I \rangle| / \Sigma I$, where n is the number of observations of a given reflection.

[d]Half-dataset correlation coefficient.

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

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

Similar to SMYD1[65], the N-lobe of SMYD2 (residues 3-276) is made up of four domains: SET, SET-I, post-SET, and MYND; and the C-lobe formed by the CTD domain (residues 280–425). The two lobes are connected by a nonconserved sequence (residues 277–279) of variable length, implicating a possible hinge motion between the CTD and

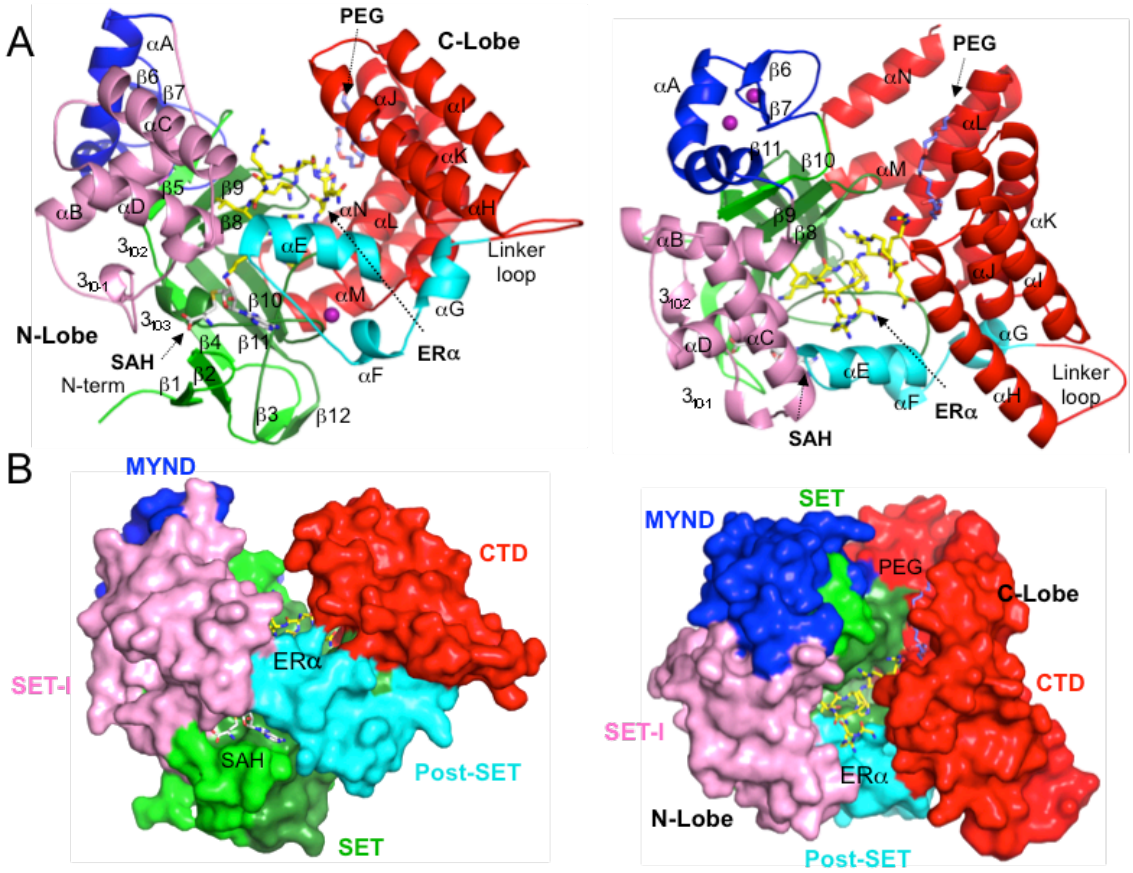

Figure 8. Overall structure of SMYD2-ERα complex. (A) Ribbon diagram of the SMYD2-ERα structure, side view (left) and top view (right). The S-sequence, MYND, SET-I, core SET, post-SET, and CTD are depicted in light green, blue, pink, green, cyan, and red. Secondary structures, α-helices, 3_{10}-helices, and β-strands, are labeled and numbered according to their position in the sequence. The ERα peptide, AdoHcy (SAH), and polyethylene glycol (PEG) are displayed as sticks with their carbon atoms colored in yellow, white, and light blue. Zinc ions are denoted by spheres and colored in purple. (B) Surface representation of the SMYD2-ERα structure, side view (left) and top view (right). SMYD2 is represented by the molecular surface colored according to domains. The ERα peptide, AdoHcy, and PEG are represented in the same way as in Figure 8A.

the rest of the protein[2]. Like other SMYD proteins[65, 66], SMYD2 has a characteristic split SET domain defined by two separate segments: the S-sequence (residues 3–49) and the core SET domain (residues 183–245). The spatial association of these segments creates an evolutionarily conserved fold comprised of one central 3_{10} helix ($3_{10\text{-}3}$) and 10 β-strands (β1–β5 and β8–β12) that are arranged into 4 antiparallel β-sheets (Figure. 8). The loop joining $3_{10\text{-}3}$ and β10 contributes conserved catalytic residues and binds the cofactor at the bottom of the cofactor-binding pocket (Figure. 9). The loop following β10 along with the strand β8 participates in the formation of the substrate-binding cleft responsible for accommodation of target ERα peptide (Figure. 10). Another functionally important region is the loop between β8 and β9, which interacts with the ERα peptide and has been shown to be required for SMYD2-mediated p53 methylation[92].

The flanking post-SET domain and the insertion SET-I domain both associate with the SET domain and assist the SET domain in lysine methylation[73, 100, 101]. The post-SET, which is immediately downstream of the SET domain, is a small cysteine-rich region comprised of 3 short α-helices (αE, αF, and αG) organized around a single zinc ion (Figure. 8). The SET-I domain is a helix bundle (αB, $3_{10\text{-}1}$, $3_{10\text{-}2}$, αC, and αD) that is inserted between the β5 and β8 strands of the SET domain (Figure. 8). Both domains (post-SET and SET-I) participate in cofactor and substrate binding (Figure. 9 and 10). The post-SET lies close to the active site, with the loop connecting αE and αF located near the cofactor and with the C-terminal end of αE positioned to stabilize the SMYD2-ERα interaction. For the SET-I domain, the functionally important structural elements are

the helix αC that is important for the recognition of the ERα N-terminal residues and the loop between $3_{10\text{-}1}$ and $3_{10\text{-}2}$ that makes extensive contacts with the cofactor. In contrast, the MYND, which also associates with the SET domain, does not contribute residues to cofactor and substrate binding. This observation is in agreement with previous findings that the MYND is dispensable for the histone methylation activity of SMYD2[55]. As shown in Figure. 8, the MYND consists of one kinked α helix (αA) and two antiparallel β-strands (β6 and β7) that are organized around two zinc ions. Evidence shows this domain is responsible for the interaction between SMYD2 and EBP41L3, a PXLXP motif-containing protein playing important roles in epithelial ovarian cancer suppression[55]. It is therefore conceivable that the MYND may primarily function as a protein-protein interaction module and coordinates SMYD2 with other proteins to regulate tumor proliferation and progression[2].

The exact function of the CTD domain in the SMYD protein family remains controversial[65, 92, 93]. In SMYD1, the CTD has been shown to play a key role in protein autoinhibition, and the deletion of the CTD increases the histone H3 binding and methyltransferase activity[65]. In contrast, the CTD in SMYD2 has been demonstrated to facilitate the formation of the substrate binding pocket and helps stabilizing p53 interaction[92]. It is not clear whether the functional differences of the CTDs may be related to the structural differences observed between SMYD proteins. We previously showed that there are exceptionally large differences in the domain-domain orientation between SMYD proteins and with respect to the distance separating the N- and C-lobes[2, 65, 66].

SMYD2 appears to be a conformational intermediate between a closed form of SMYD3 and an open form of SMYD1[2]. In addition, the CTD of SMYD2 has been demonstrated capable of adopting two distinct conformations when different cofactor analogues bind[2]. Nevertheless, the CTD itself is well conserved in the SMYD protein family comprised of 7 antiparallel α-helices (αH–αN) rotated relative to one another by approximately 25° (Figure. 8). Such helix-turn-helix topology has been noted resembling the structure of TPR repeats despite the absence of any significant sequence similarity[2, 93]. Given that the TPR repeats mediate specific protein interactions and the assembly of multiprotein complexes[102], the structural similarity between the CTD and the TPR repeats suggests a function for the CTD as a protein-protein interaction module.

Conserved Cofactor Binding Pocket

The L-shaped AdoHcy binds in a deep surface pocket formed by the SET, SET-I, and post-SET domains (Figure. 9). The bottom of the pocket is made by the SET domain signature motif NHXCXPN (residues 206–212), while the walls of the pocket are formed by three loops that are triangularly arranged (β1–β2, $3_{10\text{-}1}$–$3_{10\text{-}2}$, and αE–αF). In the pocket, the adenine moiety of AdoHcy is sandwiched between the side chains of Phe260 and Lys17, with its purine N6 and N7 atoms hydrogen-bonding to the backbone carbonyl and amide groups of His207. The ribose hydroxyls of AdoHcy make one hydrogen bond with the Tyr258 carbonyl and another hydrogen bond to the His137 imidazole ring. At the opposite end of AdoHcy, the carboxylate moiety is stabilized by salt-bridge interaction with the Arg19 guanido group, while the positively charged α-amino group is

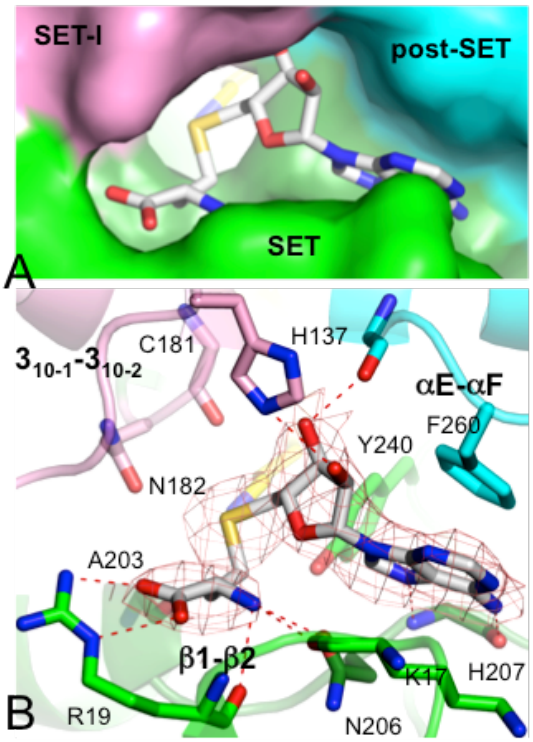

Figure 9. Conserved cofactor binding pocket. (A) Surface representation of SMYD2 cofactor binding pocket. The surface is colored according to SMYD2 domains. Bound AdoHcy is depicted by sticks with the carbon atoms colored in white. (B) Interaction between SMYD2 and AdoHcy. SmyD2 residues are represented by sticks with the carbon atoms colored according to the scheme in Figure 1. AdoHcy is depicted by sticks overlaid with $2F_o - F_c$ omit map calculated at 2.82 Å and contoured at 1.5 σ. Hydrogen bonds are illustrated as red dash lines.

recognized by triple hydrogen bonds with the carbonyls of Lys17 and Arg19 and the amide Oδ of Asn206. The former electrostatic interaction is present in most SET proteins including SMYD1 but is replaced by a hydrogen bond to a tyrosine residue in SMYD3[65, 66], which represents an unusual variation. In the middle of AdoHcy, the Sδ atom is surrounded by four oxygen atoms, of which two come from backbone carbonyls (Cys181 and Ala203), one from side chain hydroxyl (Tyr240), and one from side chain carbonyl (Asn182). Some of these oxygens have been shown to make atomic contacts with the AdoMet sulfonium and have been proposed to contribute to enzymatic function by destabilizing the active methyl group[65, 92]. Collectively, the overall cofactor-binding mode of SMYD2 is structurally conserved with SMYD1 and SMYD3 and other SET enzymes and serves to orient the methyl group of AdoMet into the methyltransfer pore during catalysis.

Specificity Determinants of SMYD2-ERα Interaction

The ERα peptide adopts a U-shaped structure and binds in the deep groove formed by the N- and C-lobes (Figure. 10A). The base of the U-structure is sandwiched between β8 and the loop preceding the post-SET domain, while the upward-protruding U-arms make interactions with the CTD and SET-I domains from each side of the ERα peptide (Figure. 10B). The stabilization of the SMYD2-ERα interaction is achieved by networks of atomic interactions, which are mainly attributed to the residue at position 0 and the residues C-terminal to the target lysine (position 0 referring to Lys266). In particular, the side chain of ERα Lys0 enters a deep hydrophobic channel composed of invariant residues Tyr258 from αE, Phe184 from β8, and Tyr240 from the loop connecting β10 and αE (Figure. 10C). These channel-forming residues are important for SMYD2 function; for example, we have shown that mutation of Tyr240 to Phe greatly diminishes ERα methylation in cells coexpressing ERα and the mutant, compared to the cells coexpressing ERα and the WT SMYD2[11]. In addition, the main chain of Lys0 makes one hydrogen bond to the Gly183 carbonyl group and two hydrogen bonds to the highly conserved residue Thr185. As a result of these interactions, the Lys0 is fully secured in the target lysine access channel with the side chain amine positioned 4.4 Å away from the sulfur atom of AdoHcy (Figure. 10D). Similar distance has been observed in many other SET methyltransferases and is deemed to be optimal for methyl transfer between AdoMet and target lysine[100].

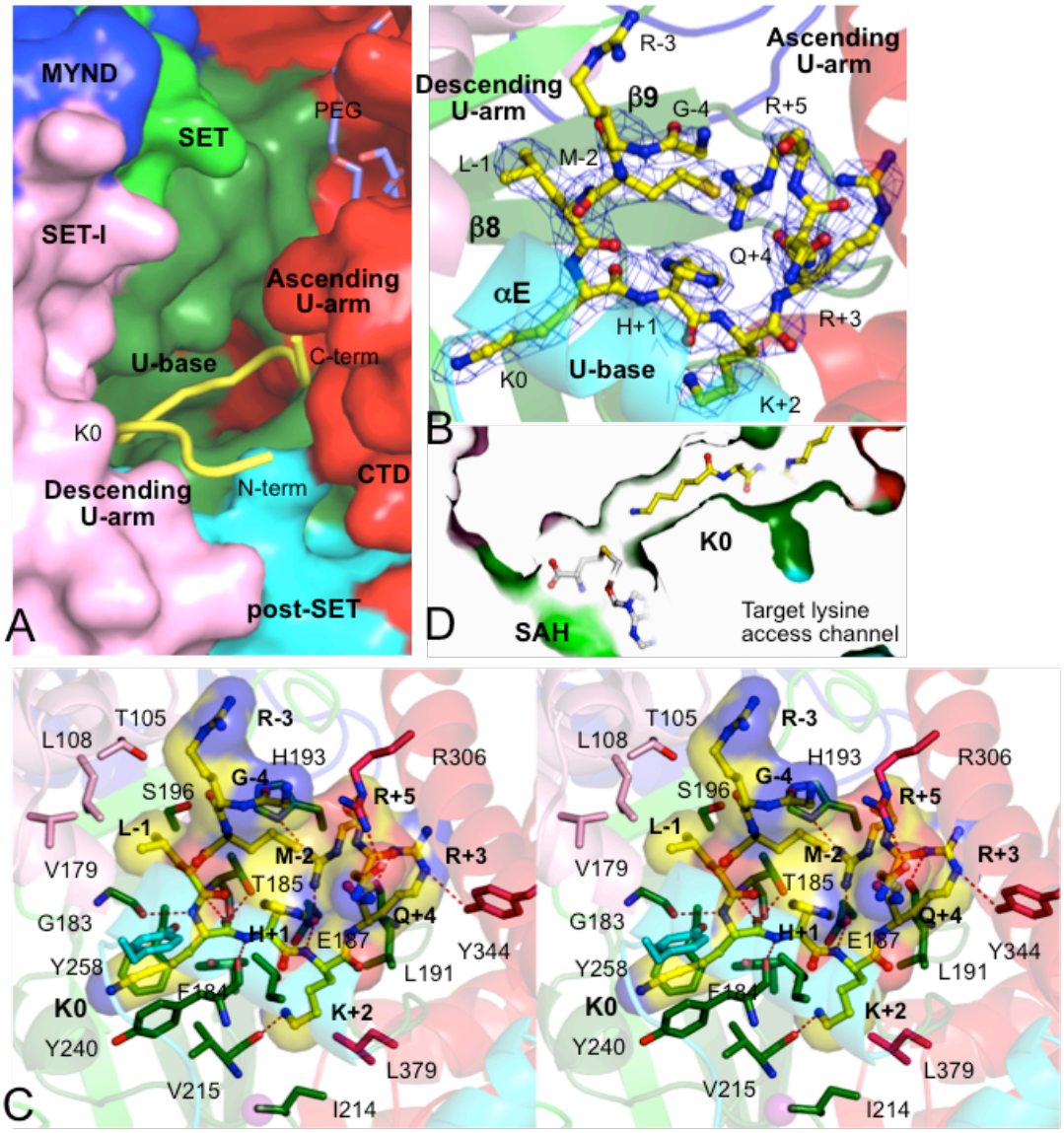

Figure 10. Interaction of SMYD2 and ERα. (A) Surface representation of SMYD2 substrate binding site. The surface is colored according to SMYD2 domains. The ERα peptide is depicted by ribbon and colored in yellow. (B) Overall view of ERα binding. SMYD2 is represented by ribbon and colored according to the scheme in Figure 1. ERα residues are shown as balls-and-sticks overlaid with $2F_o - F_c$ omit map calculated at 2.82 Å and contoured at 1.5 σ. (C) Stereo view of the detailed interactions between SMYD2 and ERα. SMYD2 backbone and residues are represented by ribbon and sticks. ERα residues are depicted by balls-and-sticks overlaid with translucent molecular surface. Hydrogen bonds are illustrated as red broken lines. (D) Target lysine access channel revealed by a slice section of the molecular surface. ERα and AdoHcy are represented by balls-and-sticks with their carbon atoms colored in yellow and white.

ERα recognition at the C-terminal side of the target lysine is mediated by numerous hydrogen bonds and hydrophobic interactions. This is in sharp contrast to the less significant N-terminal recognition (Figure. 10C). At +1 position, the interactions with His+1 include a hydrogen bond from its backbone amide to the main chain of Tyr240 and a hydrophobic contact between its imidazole ring and the side chain of Ile241. At the substrate +2 position, the side chain of Lys+2 inserts in a deep hydrophobic pocket, with the Nε atom hydrogen bonding to Val215 and the N atom hydrogen bonding to the Glu187 side chain. Mutation of Glu187 to Lys has been shown to significantly reduce the SMYD2-mediated p53 methylation[92], underlining the general importance of this residue in determining substrate binding specificity. Further C-terminal residues forming the ascending arm of the U-structure also contribute to specific interaction (Figure. 10C). Residues at +3 and +4 positions are both involved in direct contact with the CTD but to different substructures. The Nε atom of Arg+3 makes hydrogen bond with Tyr344 from the helix αJ, whereas the side chain of Gln+4 makes hydrogen bond with Arg306 located in αH. These observed interactions are consistent with recent evidence that the CTD plays an important role in SMYD2 substrate recognition, and the deletion of the CTD results in over 5-fold reduction in p53 methylation[92]. In addition, Arg+3 recognition involves the SET domain, and a hydrogen bond is formed between the Nη1 atom of Arg+3 and the backbone carbonyl of Leu191. It should be noted that the residue Leu191 and the aforementioned ERα-interacting residues Gly183, Thr185, and Glu187 are all situated within the antiparallel β8–β9 hairpin, delineating this hairpin as an important determinant of substrate specificity (Figure. 10B).

Consistent with this notion, the β8–β9 hairpin is also responsible for Arg+5 recognition, which involves bifurcated hydrogen bonding between the side chain of Arg+5 and the side chains of His193 and Glu187. In contrast, there is no significant interaction with the enzyme at the N-terminal side of the ERα peptide. This paucity of interaction is consistent with its poorly defined N-terminal structure and the higher than average B factors (Table 2). We observe no electron density for the first residue of the ERα peptide and the weak density for the following two N-terminal residues (Gly-4 and Arg-3). The specific interaction is observed only at position -1, where the side chain of Leu-1 is recognized by hydrophobic contacts with Ser196 from the SET domain and Thr105, Leu108, and Val179 from the SET-I domain. Together, our structural analysis indicates that the specificity of the SMYD2-ERα interactions is mediated by the coplay of both lobes, and the primary specificity-determining interactions appear to be C-terminal to the target lysine.

In addition to the intermolecular contacts, the ERα peptide makes a number of intrapeptide interactions that stabilize its U-shaped structure (Figure. 11). These interactions include stacking contact of the His+1 imidazole ring with the Met-2 side chain and the hydrogen bonding from the Nδ1 atom of His+1 to the O atom of Leu-1. In addition, the side chain of Arg+5 adopts a U-arm parallel rotamer that allows the formation of a hydrogen bond between its Nε atom and the main chain O of Arg+3. These interactions create a folded substrate that has a small hydrophobic core formed by the partially buried Met-2, His+1 and Arg+5 (Figure. 11A). Note that the overall

backbone structure of the folded ERα peptide differs substantially from that observed in other histone lysine methyltransferase-substrate complexes, where the bound peptide usually adopts a linear stretched conformation[100, 103, 104]. In particular, the structure of the ERα peptide is compactly asymmetric, having a horizontal cross section that is larger at one end than the other (Figure. 11B). This structural asymmetricity appears to facilitate substrate recognition by the funnel shaped-binding cleft between β8 and the β12–αE loop, as the surface of the binding pocket is highly complementary to the shape of the folded peptide (Figure. 11B). One would expect that the ordered ERα peptide structure may contribute to SMYD2 substrate binding specificity and may encode specificity-determining information additional to the peptide sequence. The structural order of peptide has been shown to be an important determinant of substrate specificity in a few other protein-peptide interaction systems. For example, experimental evidence shows that the degree and rate of modification of arginine residues to citrulline residues by PAD correlate not only with peptide sequence but also the structural order of the substrate[105]. Antibodies raised against cyclic citrullinated peptides have been shown to result in a more sensitive assay in RA diagnosis than one using linear citrullinated peptides, indicating the local structure of the peptides has important biological functions that rely on both sequence- and structure-dependent peptide recognition[106]. In addition, multiscale theoretical simulation shows that the intrapeptide interactions determine the secondary structure of amyloidogenic peptides and consequently the binding affinity to other molecules[107]. Therefore, the observed ERα peptide structure and its ability to make

intrapeptide contacts suggest an additional mechanism for determining SMYD2 substrate specificity, representing another layer of complexity in SMYD2 substrate recognition.

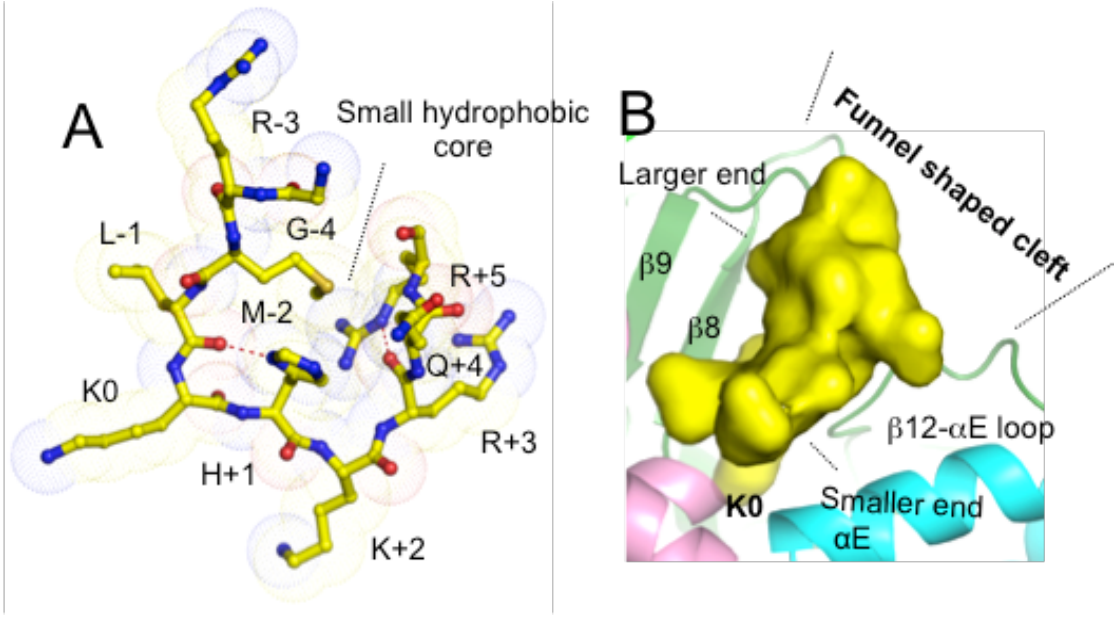

Figure 11. ERα intrapeptide interactions. (A) Van der Waals (VDW) dot representation of the folded ERα peptide. Hydrogen bonds are illustrated as red broken lines. (B) Shape complementarity between the ERα peptide and SMYD2 substrate binding pocket. ERα is represented by molecular surface (yellow) and SMYD2 is depicted by ribbon.

Mechanisms of SMYD2 Substrate Binding Diversity

To gain insights into SMYD2 substrate binding diversity, we compared the SMYD2-ERα structure with the structure of SMYD2 in complex with p53 peptide (Figure. 12). The two liganded SMYD2 structures are very similar with overall root mean-square deviations of 0.6 Å for 430 Cα atoms. The main chains of the two peptide ligands and the side chains of residues at positions -1, 0, +1, and +2 are superimposed well, whereas large deviations are observed at the ascending arms of the U-structures that

display completely different binding modes (Figure. 12A). In SMYD2-ERα, Arg+3 binds in the β8–β9 region of the SET domain, whereas in SMYD2-p53 the side chain of Lys+3 is stabilized by interaction with Tyr370 and Tyr374 from the CTD and Asp242 from the loop preceding the post-SET domain (Figure. 12B). Similarly, Arg+5 in SMYD2-ERα interacts with the β8–β9 hairpin, whereas Gln+5 in SMYD2-p53 inserts its side chain into a deep pocket formed by His341, Tyr344, Gln345, and Tyr370 from the CTD and Leu244 and Tyr245 from the post-SET-preceding loop (Figure. 12C). It is not clear whether the residues at position +4 also bind at different regions of the active site pocket, because a glycine residue is found in p53 at this position (Figure. 12D). However, the different backbone angles exhibited by Gln+4 and Gly+4 suggest any non-glycine residue replacing Gly+4 would point its side chain to the SET domain contrasting with the CTD-oriented conformation of the Gln+4 side chain. These findings indicate that SMYD2 has multiple distinct binding sites that allow the accommodation of the U-arm residues with different sequences, thereby explaining its broad specificity for these substrate positions (Figure. 12H).

Notably, the structural comparison reveals the mechanisms for the accommodation of diverse substrate residues at the U-base and the U-arms are different (Figure. 12E). This difference appears to be related to the fact that SMYD2 has a wide U-arm binding site compared to a narrow U-base binding cleft that may not allow peptide binding in different modes (Figure. 10A). Specifically, at the +1 position of the U-base, the side chain of ERα His+1 follows a path similar to that of the aliphatic portion of the

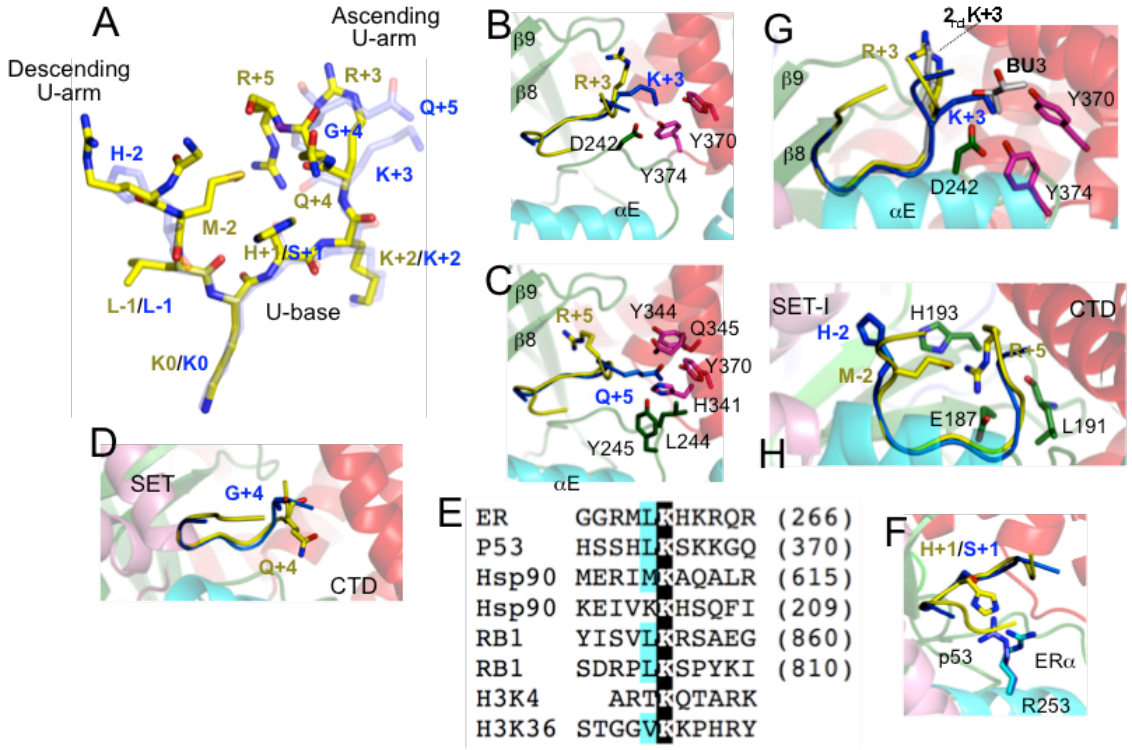

Figure 12**.** Structural comparison of SMYD2-ERα and SMYD2-p53. (A) Superposition of the structures of the ERα peptide (yellow) and the p53 peptide (light blue; PDB code: 3TG5). (B), (C), (D), (E), and (F) Structural and binding differences of ERα and p53 at position +3, +5, +4, +1, and -2. SMYD2 is colored according to domains while the ERα and p53 peptides are shown in yellow and light blue. (G) Structural superposition of +3 peptide residues of SMYD2-ERα (yellow), SMYD2-p53 (light blue), and the second SMYD2-p53 complex (gray; PDB code: 3S7D). A butanediol molecule (BU3) found in the second SMYD2-p53 structure is depicted by sticks with the carbon atoms colored in gray. (H) Sequence alignment of the target lysine and surrounding residues of SMYD2 methylation targets. Protein target names are shown at the left of the sequences. Target lysine residues are shown as white on black, and similar residues appear shaded in cyan. Position numbering is displayed above the alignment with position 0 referring to the target lysine. Sequence numbering of the target lysine is displayed to the right of the sequences.

p53 Ser+1 side chain, facing toward the SMYD2 residue Arg253. Note that the side chain of this SMYD2 residue exhibits the large conformational changes in the two complexes. In the SMYD2-p53 structure, the side chain of Arg23 is oriented toward the hydroxyl group of Ser+1, whereas in the SMYD2-ERα complex, the guanidino group is rotated

away from the bulkier side chain of His+1 because of steric effects (Figure. 12E). These observations indicate that the conformational changes of Arg253 underlie the SMYD2 flexibility to accommodate ligands with +1 side chains of different size and suggest the diversity in substrate recognition.

Comparison of the SMYD2-p53 structures solved by two independent groups[92, 93] indicates that the intrinsic dynamics of the peptide ligand may facilitate diverse substrate recognition by allowing considerably different peptide binding (Figure. 12F). In the two SMYD2-p53 structures, the side chains of the p53 Lys+3 adopt two different conformations. The second conformation is similar to the one observed for the ERα Arg+3, whereas the binding site for the first conformation (see above) is occupied by a butanediol molecule in the second SMYD2-p53 complex. This difference indicates that the bound peptide has significant flexibility capable of assuming different conformations in different solution environments. It also indicates that the intrinsic dynamics of the peptide ligand allows for the interaction with different substrate recognition residues, thereby likely contributing to the SMYD2 broad substrate specificity. It is of particular interest to note the recognition of a peptide-loaded MHC molecule by the cognate T-cell receptor depends on the dynamics properties of the peptides, and differential peptide flexibility resulting from MHC polymorphisms can broaden and expand T-cell receptor reactivity[108].

As discussed earlier, the intrapeptide interactions and the structural order of peptide substrates may play a role in determining substrate specificity. Interestingly, the structural comparison shows that the ERα and p53 peptides have completely different intrapeptide interactions. Unlike the ERα peptide (Figure. 11), only one hydrogen bond exists within the p53 peptide and there is no hydrophobic core present in the middle of the U-structure (Figure. 12A). For the most part, these observed differences are due to the conformational changes of the +2 residues in the two complexes. In SMYD2-p53, the side chain of His+2 points away from the U-structure and is oriented toward the SET-I domain. In SMYD2-ERα, the side chain of Met+2 is noted to participate in hydrophobic contacts with the Arg+5 side chain and assists in the completion of the Arg+5 binding pocket (Figure. 12G). Therefore, these ERα-specific interactions, together with the different binding modes of +5 residues, underline the potential role of the intrapeptide interactions in determining SMYD2 substrate binding diversity. Together, our comparison of the SMYD2-ERα and SMYD2-p53 structures provides the first structural information on how SMYD2 distinguishes and recognizes diverse substrates and suggests that the broad specificity of SMYD2 is achieved by manifold molecular mechanisms including multiple distinct binding sites, the conformational plasticity of the substrate-binding pocket, the intrinsic dynamics of peptide ligands, and the substrate-specific intrapeptide interactions. The interplay of these mechanisms would create sets of complex-specific states that may underlie the SMYD2 ability to methylate a broad spectrum of functionally and structurally different substrates.

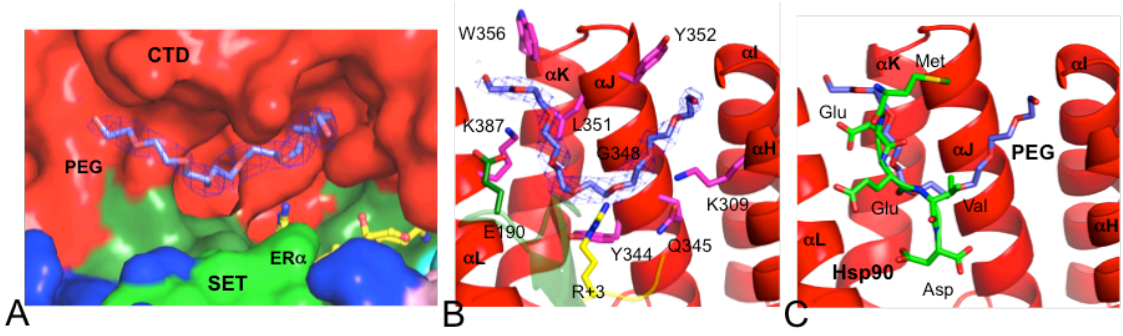

Figure 13. PEG binding site. (A) Surface representation of SMYD2 PEG binding site. The surface is colored according to SMYD2 domains. PEG is depicted by sticks overlaid with $2F_o - F_c$ omit map calculated at 2.82 Å and contoured at 1.5 σ. Nearby ERα residues are displayed as sticks with the carbon atoms colored in yellow. (B) Putative PEG interacting residues. SmyD2 residues are colored according to domains, ERα residues are shown in yellow, and PEG (light blue) is represented in the same way as in Figure 6A. (C) Structural superposition of the PEG binding site and the predicted Hsp90 binding site[2]. PEG is colored in light blue while the Hsp90 peptide (MEEVD) containing the TPR binding motif is colored in green.

Unexpected PEG Binding Site

The most striking finding in the SMYD2-ERα structure is an extra long density observed between the N- and C-lobes (Figure. 13). The entire stretch of this density is close to the inner surface of the CTD, with the central portion of the density adjacent to the ERα peptide-binding pocket. The density does not have a peptide characteristic density feature and is not interpretable as ERα residues, AdoHcy, or water. Based on the components in the crystal condition and the shape of the density, a polyethylene glycol (PEG) molecule (n=8) was assigned into the density (Figure. 13A). The modeled PEG molecule adopts an omega-turn conformation with one end anchored at a surface groove formed by αH, αI, and αJ and the other end extending between the αJ and αK helices (Figure. 13B). Residues likely contributing to PEG binding (within 4 Å of PEG) include

Lys309, Tyr344, Gln345, Gly348, Leu351, Tyr352, Trp356, and Lys387 from the CTD and Glu190 from the SET domain. The Arg+3 residue of the ERα peptide is in close proximity to the PEG molecule, indicating that ERα may also participate in interaction with PEG and stabilizes its position. It should be noted that all of the SMYD2 residues listed above are not conserved across the SMYD protein family except Lys309. This raises the interesting question whether the observed PEG binding in the present structure is SMYD2 specific.

Bound PEG molecules have been reported in several other protein structures[109-111]. In most cases, the PEG binding has important functional implications and mimics the ligand binding in the proteins, such as the putative odorant binding site of the odorant binding protein AgamOBP1 and the cephalosporin binding site of the putative methyltransferase CmcI[112]. While the functional significance of the PEG binding in SMYD2 requires future investigation, several possibilities can be envisioned. First, the PEG binding site might represent an additional substrate-binding pocket that could participate in ligand binding leading to increased affinity or novel specificity. This possibility is consistent with the observation that SMYD2 does not methylate the H3 or H4 peptides efficiently but can act on the full-length histones 10-fold more effectively[101]. Second, the PEG binding site might represent a protein interacting site responsible for modulating SMYD2 cellular localization and transcription activities. Proteomic analysis of the SMYD2 interactomes has shown that both CTD and MYND domains are important protein-protein interaction domains interacting with proteins involved in cell

cycle regulation and transcription regulation[113]. Third, the PEG binding site could play a regulatory role and regulates SMYD2 methyltransferase activities by binding activator proteins such as Hsp90. This suggestion is supported by the close proximity between the ERα binding site and the PEG binding site, and also by the observed PEG binding site overlapping with previously predicted Hsp90 binding pocket in the TPR-like CTD domain (Figure. 13C)[2]. Therefore, our finding of the PEG binding site in the SMYD2 protein will potentially open many new research directions that could improve functional understanding of this still poorly understood methyltransferase. On the other hand, the structural details of the PEG binding site may be valuable in developing new methods and strategies for selective drug design. For instance, strategies aiming at exploiting this novel, potentially SMYD2-specific binding site may represent a promising approach for the development of small molecules that could selectively inhibit this enzyme without cross-reacting with a wide range of other functionally important methyltransferases. Such strategy should have important implications in specific ERα signaling regulation and also in many ERα-dependent human diseases.

PART II

Structure Basis of Nherf1-Mediated CXCR2 Macromolecular Complex Assembly

CHAPTER 4

Structural Insights into Neutrophilic Migration Revealed by the Crystal Structure of the Chemokine Receptor CXCR2 in Complex with the First PDZ Domain of NHERF1

*Published in PLoS One 2013 Oct 2;8(10): e76219. doi: 10.1371/journal.pone.0076219. All authors agreed with including their work in this dissertation.

Abstract

Neutrophils play an essential role in host defense against infection, but uncontrolled neutrophilic infiltration can cause inflammation and severe epithelial damage. We recently showed that CXCR2 formed a signaling complex with NHERF1 and PLC-β2, and that the formation of this complex was required for intracellular calcium mobilization and neutrophilic transepithelial migration. To uncover the structural basis of the complex formation, we report here the crystal structure of the NHERF1 PDZ1 domain in complex with the C-terminal sequence of CXCR2 at 1.16 Å resolution. The structure reveals that the CXCR2 peptide binds to PDZ1 in an extended

conformation with the last four residues making specific side chain interactions. Remarkably, comparison of the structure to previously studied PDZ1 domains has allowed the identification of PDZ1 ligand-specific interactions and the mechanisms that govern PDZ1 target selection diversities. In addition, we show that CXCR2 can bind both NHERF1 PDZ1 and PDZ2 in pulldown experiments, consistent with the observation that the peptide binding pockets of these two PDZ domains are highly structurally conserved. The results of this study therefore provide structural basis for the CXCR2-mediated neutrophilic migration and could have important clinical applications in the prevention and treatment of numerous neutrophil-dependent inflammatory disorders.

Introduction

Interleukin 8 receptor, beta (CXCR2) is a G-protein-coupled receptor that mediates neutrophil migration to sites of inflammation and controls the positioning of oligodendrocyte precursors in developing spinal cord by arresting their migration[114, 115]. This receptor also functions in angiogenesis and wound healing, and plays an important role in both spontaneous and inflammation-driven tumorigenesis[114, 116, 117]. In almost all the cases, the ability of CXCR2 to direct cell trafficking and positioning depends on its ability to bind to a repertoire of structurally and functionally related chemokines[114]. For example, CXCR2 can bind all seven ELR-positive CXC chemokines, which include growth-related protein (Gro)-α, -β, and -γ, epithelial-derived neutrophil attractant-78 (ENA-78), granulocyte chemotactic protein-2 (GCP-2), interleukin-8 (IL-8) and neutrophil-activating peptide-2 (NAP-2)[118]. When binding to one of these chemokines,

CXCR2 is capable of initiating G-protein heterotrimeric dissociation, which in turn induces many downstream signaling events such as intracellular calcium mobilization and actin polymerization both required for the chemokine gradient-directed cell migration[114].

Although the general process of the CXCR2-mediated signaling is well established, the mechanisms regarding specific coupling of CXCR2 to its downstream signaling molecules still remain poorly understood. We recently showed that CXCR2 formed a complex with its downstream effector PLC-β2 (phospholipaseC) via the scaffold protein NHERF1 (Na+/H+ exchanger regulatory factor) in freshly isolated neutrophils and bone marrow-derived neutrophils[119]. We also showed that this complex played a critical role in the CXCR2-mediated signaling and was required for intracellular calcium mobilization and neutrophilic transepithelial migration[119]. Furthermore, we showed that the formation of this complex was mediated by the PDZ domains of NHERF1, which bridged CXCR2 and PLC-β2 by binding to their C-terminal PDZ-binding motifs[119]. Remarkably, the PDZ-mediated interaction of NHERF1 with the C-terminal sequence TSTTL of CXCR2 was essential for the functional assembly of the CXCR2/NHERF1/PLC-β2 complex, and disrupting the interaction with a cell permeable PDZ motif-containing peptide was sufficient to block the IL-8-induced CXCR2 neutrophilic signaling[119]. As neutrophil dysregulation is central to human immunopathology[120], the identification of this novel CXCR2 complex that contributed to neutrophil chemotactic regulation suggested that targeting this trimeric complex inside

the neutrophils might represent a new strategy for the treatment of numerous neutrophil-dependent inflammatory disorders[119]. This notion, in turn, highlights the importance of elucidating the structural basis of the PDZ domain-mediated CXCR2-NHERF1 interaction, as a necessary prerequisite of discovering small molecules that could fine-tune CXCR2 activity or suppress excessive, disease-causing neutrophilic infiltration.

In general, PDZ domains mediate protein interactions by recognizing short amino acid motifs at the C-termini of target proteins, through which PDZ domains play important roles in signal complex assembling and receptor recycling as well as in establishing cell polarity and directing protein trafficking[121]. Recent studies showed that individual PDZ motifs are capable of recognizing up to seven C-terminal ligand residues, with a vast potential to interact with a large number of biologically and functionally diverse ligands[122]. However, in many cases, the specificity of the PDZ-peptide interaction is determined mainly by the residues at positions 0 and -2 of the peptide (position 0 referring to the C-terminal residue), whereas other residues do not significantly contribute to the interaction[123]. Based on that, PDZ domains have been grouped into two major classes. Class I domains bind to peptides with the consensus sequence (S/T)X(V/I/L) (X denoting any amino acid), while class II domains recognize the motif (F/Y)X(F/V/A)[1, 124]. Corroborating this classification, structural studies revealed that PDZ domains share a similar peptide recognition mode, with the 0 residue of peptide occupying a hydrophobic pocket and the -2 residue participating in direct side chain interactions[123, 124].

In fact, the structural similarity in PDZ ligand recognition[123, 124], together with the fact that more than 500 PDZ domains in over 300 different proteins are present in the human genome[125], has led to years of intensive research regarding how PDZ domains, a structurally simple protein interaction module, can achieve robust and efficient ligand discrimination, the nature of which, however, still remains obscure. In this context, it is interesting to note that PDZ binding is also enormously promiscuous, with one domain capable of binding multiple targets[126]. For example, NHERF1 contains two PDZ domains (PDZ1 and PDZ2) that are known to interact with a variety of transmembrane proteins, such as the cystic fibrosis transmembrane conductance regulator (CFTR), the β2-adrenergic receptor (β2AR), the platelet-derived growth factor receptor (PDGFR) and the parathyroid hormone receptor (PTHR)[127, 128]. Moreover, PDZ promiscuity is exemplified by the fact that some PDZ domains have the ability to bind peptide sequences that belong to both class I and class II motifs[129]. Therefore, these examples have made it apparent that detailed analysis and comparison of many proteins will be required to establish and illuminate the full range of ligand discrimination operated by the PDZ domain fold[130]. A high-resolution structural interpretation of individual PDZ domain function should in turn provide considerable insights into the mechanisms regarding how the exquisite ligand discrimination dictates the diversification of biological functions. For this reason, we report here the high-resolution structure (1.1 Å) of the NHERF1 PDZ1 domain in complex with the CXCR2 C-terminal peptide TSTTL. The structure reveals PDZ1 ligand-specific interactions and new mechanisms that govern the PDZ1 target selection diversity. We also show that CXCR2 can bind both NHERF1 PDZ1 and PDZ2 in

pulldown experiments, consistent with the observation that the two domains share highly structurally-conserved peptide binding pockets. The results of this study therefore provide important insights into the CXCR2-mediated neutrophilic migration and could be valuable in the development of novel therapeutic strategies against many neutrophil-dependent inflammatory disorders.

Materials and Methods

Protein Expression and Purification.

For X-ray crystallography, a DNA fragment encoding the human NHERF PDZ1 (residues 11–94), and having the C-terminal extension TSTTL that corresponds to residues 356–360 of human CXCR2, was amplified using PCR and cloned in the pSUMO vector. The resulting clone that contains a N-terminal His6-SUMO tag was transformed into Escherichia coli BL21 Condon Plus (DE3) cells for protein expression. The transformants were grown to an OD600 (optical density at 600 nm) of 0.4 at 37 °C in LB medium, and then induced with 0.1 mM isopropylthio-β-D-galactoside and grown an additional 16 h at 15 °C. The cells were harvested by centrifugation and lysed by French Press. The soluble fraction was then subjected to Ni^{2+} affinity chromatography purification, followed by the cleavage of the His6-SUMO tag with yeast SUMO Protease 1. PDZ1 was separated from the cleaved tag by the second Ni^{2+} affinity chromatography and further purified by size-exclusion chromatography. Finally, the protein was concentrated to 10–20 mg/ml in a buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1 mM β-mercaptoethanol (BME), and 5% glycerol. For GST pulldown

experiments, full-length NHERF1, NHERF1 PDZ1 (residues 11-94) or NHERF1 PDZ2 was cloned into the BamHI/XhoI sites of pGEX6P-1 plasmid, which was then transformed into the Escherichia coli BL21 Gold (DE3) for protein expression. The proteins were expressed essentially similar as described earlier and purified by affinity chromatography using immobilized glutathione Sepharose 4B resin.

Crystallization, Data Collection and Structure Determination

Crystals were grown by the hanging-drop vapor-diffusion method by mixing the protein (~8 mg/ml) with an equal volume of reservoir solution containing 100 mM sodium acetate, pH 4.6, 2.7 M sodium chloride at 20 °C. Crystals typically appeared overnight and continued to grow to full size in 3-4 days. Before X-ray diffraction data collection, crystals were cryoprotected in a solution containing mother liquor and 25% glycerol and flash frozen in liquid nitrogen. The data were collected at 100 K at beamline 21-ID-F at the Advanced Photon Source (Argonne, IL) and processed and scaled using the program HKL2000[67]. Crystals belong to space group $P3_121$ with unit cell dimensions a = b = 45.1 Å, c = 63.6 Å, and one molecule in the asymmetric unit. The structure was solved by the molecular replacement method with program PHASER[96] using the PDZ1-CFTR structure (PDB code: 1I92) as a search model. The structure modeling was carried out in COOT[69], and refinement was performed with BUSTER[70]. The final models were analyzed and validated with PROCHECK[131]. All figures of 3D representations of the PDZ1-CXCR2 structures were made with PyMOL.

Protein Data Bank Accession Number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 4JL7.

Cell Culture and Transfection

HEK293 cells were obtained from the American Type Tissue Collection (Manassas, VA) and maintained as described previously[119]. Briefly, the cells were cultured in Dulbecco's modified Eagle's medium (DMEM) (Invitrogen) supplemented with 10% fetal bovine serum (FBS), 1% nonessential amino acids, 2 mM L-glutamine and 25 mg/mL penicillin or streptomycin. The cells were maintained at 37 °C in a 5% CO_2-95% air atmosphere and routinely passaged at a ratio of 1:4 when 70-80% confluent. Transfection was carried out with the Lipofectamine 2000 (Invitrogen) transfection kit according to the manufacturer's protocol. HEK293 cells were plated in 75-cm^2 flasks. After reaching of 80%–90% confluency, cells were provided with 12 ml of fresh medium and transfection was performed with pTriEx-4 vector encoding HA-tagged human CXCR2, murine CXCR2, and FLAG-tagged CXCR2 C-tail fragments.

Pulldown Assays

GST pulldown assay were preformed essentially similar as described in our previous studies[119]. Briefly, HEK293 cells overexpressing CXCR2 proteins were lysed with cell lysis buffer (PBS, 0.2% Triton X-100) supplemented with a mixture of protease inhibitors (1 mM phenylmethylsulfonyl fluoride, 1 g/ml of aprotinin, 1 g/ml of leupeptin,

and 1g/ml of pepstatin) and phosphatase inhibitor mixture (Sigma). The cell lysates were cleared by centrifugation at 16,000 g for 10 min, and then incubated with GST-NHERF1 fusion constructs (GST-NHERF1, GST-PDZ1, GST-PDZ2) or GST alone for 3 h at 4 ºC. After incubation, the complex was mixed with glutathione-agarose beads (BD Biosciences) and incubated for 1 h at 4 ºC with general shaking. The beads were then washed three times with 1 ml of lysis buffer, pelleted at 500g for 30 s, and boiled in Laemmli sample buffer. Finally, HA-tagged CXCR2 proteins, which bound to GST-NHERF1 proteins, were resolved by SDS-PAGE and detected by anti-HA antibodies. To verify direct CXCR2/NHERF1 interaction, purified GST-NHERF1 PDZ domains (0.5 uM; GST-NHERF1 PDZ1, GST-NHERF1 PDZ2, GST-NHERF1 PDZ1+2) or GST alone were mixed with a synthetic CXCR2 C-tail peptide (1.0 uM; last 13 a.a. with a biotin-conjugate at N-terminus) in binding buffer (PBS + 0.2% Triton X-100 + protease inhibitors) at 22-24ºC for 1 hr. The mixtures were incubated with Streptavidin beads (for binding to biotin-conjugate in the peptide) for 2 hrs. The beads were washed three times with binding buffer, and eluted with Laemmli sample buffer containing β-mercaptoethonal. The eluents were resolved by SDS-PAGE, and immunoblotted with anti-GST IgG.

Results and Discussion

Structure Determination

To facilitate NHERF1-CXCR2 cocrystallization and reveal the mechanism by which NHERF1 recognizes CXCR2, we generated a chimeric protein with the C-

terminus of the NHERF1 PDZ1 domain (residues 11–94) fused to five amino acids (TSTTL) corresponding to the CXCR2 residues 356–360. We reasoned that such design would take advantage of functional interaction between CXCR2 and NHERF1, allowing efficient crystal packing by promoting intermolecular contacts in a more site-specific manner. This strategy has previously been applied to several PDZ-target complexes[1, 4, 132], and indeed proved to be effective in obtaining diffraction-quality PDZ1-CXCR2 crystals in this study. The crystals diffracted to high resolution (1.1 Å), and the structure was determined by molecular replacement. The model was refined to R_{work} of 18.7% and R_{free} of 21.7%, and the evaluation of its stereochemistry using PROCHECK showed that 89.2% of the residues are in the most favored, 8.1% in the additional allowed, and 2.7% in the generously allowed regions; no residues are found in the disallowed regions (Table 3).

Overview of the Structure

The crystal structure reveals a polymeric PDZ1 arrangement with the carboxyl terminal region TSTTL of one PDZ1 molecule bound to a neighboring PDZ1, which leads to the formation of a linear, infinite PDZ1 filament throughout the crystals. The overall topology of NHERF1 PDZ1 is similar to other PDZ domains[124], consisting of a six-stranded β-barrel (β1–β6) that is capped top and bottom by two α-helices (αA and αB) (Figure. 14A). The β-barrel has a hydrophobic interior, lined up with highly conserved residues, including Leu59, Ile39, Phe26, Cys15, Val86, Leu88, Leu59, and

Table 3. Crystallographic data and refinement statistics

Space group	$P3_121$
Cell parameters (Å)	
a=b	50.4
c	66.0
Wavelength (Å)	0.97872
Resolution (Å)	20.0-1.16 (1.20-1.16)
R_{merge}[a]	0.063 (0.463)[b]
Redundancy	9.7 (7.0)
Unique reflections	33912
Completeness (%)	100 (100)
$\langle I/\sigma \rangle$	19.1 (3.3)
Refinement	
Resolution (Å)	20.0-1.16
Molecules/AU	1
R_{work}[c]	0.186 (0.217)
R_{free}[d]	0.208 (0.248)
RMSD	
Bond lengths (Å)	0.011
Bond angels (°)	1.2
No. of atoms	
Protein	655
Peptide	36
Water	102
Chloride	3
B-factor (Å^2)	
Protein	20.4
Peptide	15.1
Water	27.8
Chloride	16.4

[a] $R_{merge} = \Sigma |I-\langle I \rangle| / \Sigma I$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity of multiple observations of symmetry-related reflections.
[b] Numbers in parentheses refer to the highest resolution shell.
[c] $R_{work} = \Sigma |F_o-F_c| / \Sigma |F_o|$, where F_o is the observed structure factor, F_c is the calculated struture factor.
[d] R_{free} was calculated using a subset (5%) of the reflection not used in the refinement.

Val90 (Figure. 14B). These residues are also evolutionally conserved across other PDZ motifs, suggesting their universal role in stabilizing PDZ fold by forming a continuous hydrophobic core[123]. In contrast, the outside of the barrel is rather hydrophilic, with a region enriched with basic residues predicted to be involved in membrane association and direct interaction with acidic lipids[133]. These putative lipid-binding residues include Lys32, Lys34 and Arg40, which are located within β3 and its preceding loop with their side chains facing toward the solvent (Figure. 14A). The interaction of cholesterol with these surface residues was required for dynamic NHERF1-CFTR colocalization, and disruption of the NHERF1's cholesterol-binding activity resulted in aberrant CFTR channel activation[133].

In addition to its amphipathic nature, the PDZ1 β-barrel is structurally asymmetric, having a circular cross section that is larger at one end than the other (Figure. 14A). At the smaller end, the PDZ1 N- and C-termini curl close together and block the barrel opening. On the opposite end, a helix (αB) is positioned in a manner that still permits access to the barrel's interior core region. This helix (αB) is stabilized by VDW contacts with the residues from β3 and β4 but stays ~9 Å apart from β2. The nearly parallel arrangement of αB and β2 creates a shallow surface groove approximately 18 Å long, 8 Å wide, and 4 Å deep. The groove stretches deeply into the central cavity of the β-barrel, forming a peptide-binding pocket that is responsible for highly robust protein interactions[122]. Similar to other PDZ structures[1, 4], the CXCR2 C-terminal peptide TSTTL inserts into the PDZ1 binding pocket as an additional β-strand antiparallel to β2 (Figure.

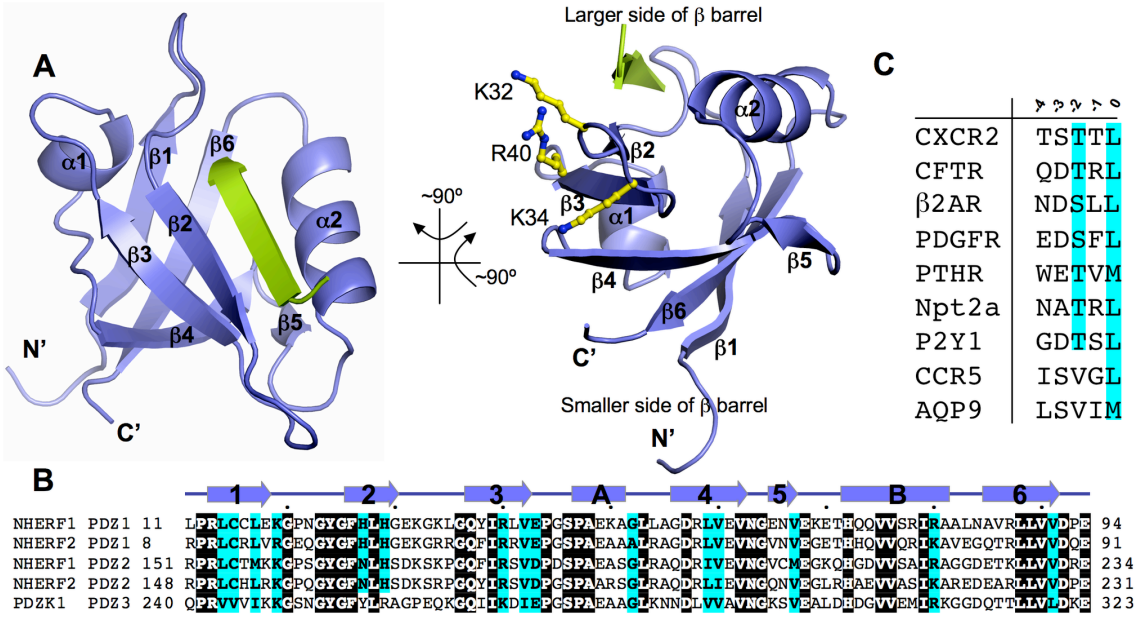

Figure 14. Structure of NHERF1 PDZ1 in complex with the CXCR2 C-terminal sequence TSTTL. (A) Ribbon diagram of the PDZ1-CXCR2 structure, front view on the left and side view on the right. PDZ1 is shown in purple and the CXCR2 peptide shown in green. Secondary structures of PDZ1, α-helices and β-strands, are labeled and numbered according to their position in the sequence. Side chains of putative PDZ1 lipid-binding residues are depicted by balls-and-sticks in the side view of the structure. (B) Sequence alignment of selected PDZ domains. The alignment was performed by ClustalW[42], including human NHERF1, human NHERF2 and mouse PDZK1. Identical residues are shown as white on black, and similar residues appear shaded in cyan. Secondary structure elements are displayed above the sequences and labeled according to the scheme in Fig. 14A. Sequence numbering is displayed to the left of the sequences, with every 10th residue marked by a dot shown above the alignment. (C) Sequence alignment of the last five residues of natural NHERF binding targets. The alignment includes CXCR2, CFTR, β2AR, PDGFR, PTHR, Npt2a (type 2 sodium-phosphate cotransporter), purinergic receptor P2Y1, CCR5 (C-C chemokine receptor type 5), and AQP9 (aquaporin 9). Protein names are shown at the left of the sequences. Position numbering is displayed above the alignment, with position 0 referring to the very C-terminal residue.

15). In this setting, the invading peptide is highly ordered as indicated by high quality electron density maps (Figure. 15A) and below average B factors (Table 3). It should be noted that the CXCR2-binding pocket is topologically distinct from the putative lipid binding sites (Figure. 14B), and that mutation of the cholesterol-binding residues did not

lead to significant changes in the NHERF1 ligand-binding activity[133]. Although the role of cholesterol in CXCR2 signaling is currently unknown, the PDZ topological asymmetry that places the CXCR2-binding sites opposite to the domain termini, along with direct cholesterol-NHERF interaction being important for cell signaling and protein networking[133], suggests a signaling platform with PDZ1 serving as a dual-specificity scaffold to bring together the membrane and juxtamembrane signaling complexes[133].

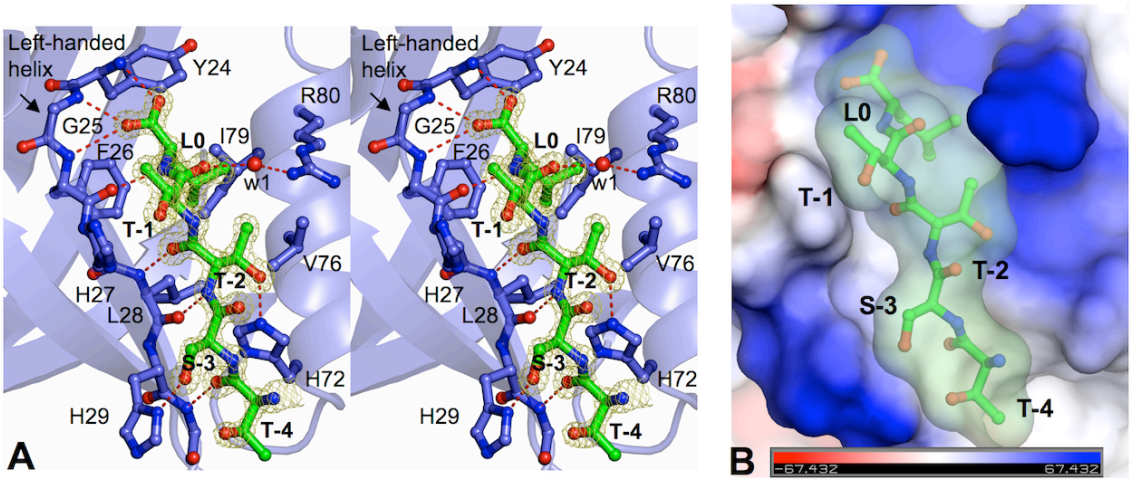

Figure 15. Interactions between PDZ1 and CXCR2. (A) Stereo view of the PDZ1 ligand-binding site bound to the CXCR2 C-terminal peptide. PDZ1 residues are represented by balls-and-sticks with their carbon atoms colored in purple. CXCR2 peptide is depicted by balls-and-sticks overlaid with 2Fo − Fc omit map calculated at 1.16 Å and contoured at 1.8 σ. Hydrogen bonds are illustrated as red broken lines. (B) Surface representation of the PDZ1 binding pocket with coloring according to the electrostatic potential: red, white, and blue correspond to negative, neutral and positive potential, respectively. The CXCR2 peptide is depicted by balls-and-sticks overlaid by its transparent molecular surface.

Specificity Determinants of Consensus PDZ1 Binding Motif

The CXCR2 pentapeptide (TSTTL) binds PDZ1 in an extended conformation, forming numerous contacts with β2 and αB and burying a total solvent-accessible surface

area of 360 $Å^2$ (Figure. 15). Only the last four residues of CXCR2 make specific contacts to PDZ1, whereas the first threonine adopts a well-defined conformation but is not directly involved in PDZ1 recognition, indicating that it does not contribute to the specificity of the interaction. Similar to other PDZ domains[123], the specificity and affinity of the PDZ1-CXCR2 interaction are achieved mainly by the residues at positions 0 and -2 of the peptide (position 0 referring to the very C-terminal residue), whereas residues -1 and -3 appear to be important for ligand-specific interactions (see below). Specifically, the side chain of CXCR2 Leu0 enters a deep hydrophobic pocket composed of invariant residues Tyr24, Phe26, and Leu28 from β2, and Val76 and Ile79 from αB. These pocket-forming residues are important for NHERF1 functions; for example, mutation of Tyr24 and Phe26 completely abolished the NHERF1-targets interaction and significantly altered cellular processes essential to tumor metastatic behaviors[134].

In the PDZ1 pocket, the position of Leu0 is fully secured by a hydrogen bond from its carbonyl oxygen to the Tyr24 amide nitrogen and by triple hydrogen bonding to the PDZ1 carboxylate-binding motif (Figure. 15A). The carboxylate-binding motif, located between β1 and β2, has a left-handed helical conformation that results in three amide nitrogens being directed toward the peptide, thereby allowing the hydrogen-bond formation between the Leu0 carboxylate and the amides of Phe24, Gly25, and Phe26. In addition, Leu0 fits tightly in the PDZ1 pocket, with the side chain directly abutting the benzene ring of Phe26 and the isobutyl group of Ile79. Remarkably, the surface of the pocket is highly complementary to the shape of leucine (Figure. 15B), which thus

provides a structural basis that governs the high affinity binding between CXCR2 and NHERF1[119]. This stereochemical complementarity also suggests that any model that substitutes Leu0 to larger hydrophobic residues would generate substantial steric clashes; to smaller ones would be energetically unfavorable. Interestingly, recent molecular dynamic simulation studies showed that replacement of Leu0 by Val or Ala of the CFTR ligand resulted in fewer interactions with NHERF1 PDZ1 and substantially lower binding energy[135]. Collectively, the present structure demonstrates, for the first time, the PDZ1 binding selectivity for the CXCR2 C-terminal leucine, which is mediated by the stereochemically complementary hydrophobic interaction in a fashion that is highly conserved in class I PDZ motif[123]. This conserved binding selectivity in turn provides structural rationalization for the importance of Leu0 in CXCR2 function. The competition experiments using the leucine-mutated peptides did not affect IL-8-induced CXCR2 signaling, but the treatment of bone marrow neutrophils with a CXCR2 peptide containing an intact PDZ motif, disrupting NHERF1-CXCR2 complex, resulted in a significant inhibition of intracellular calcium mobilization, chemotaxis, and transepithelial migration of neutrophils[119].

Another conserved feature of the PDZ1-CXCR2 interaction is that Thr-2 engages in numerous specific contacts with PDZ1 and plays an important role in determining the specificity and affinity of the interaction. Specifically, the amide nitrogen of Thr-2 hydrogen bonds to the carbonyl oxygen of Leu28, while the backbone carbonyl of Thr-2 hydrogen bonds to the main chain amide of the same residue. In addition, the side chain

hydroxyl of Thr-2 hydrogen bonds with the imidazole ring of His72, with its side chain aliphatic carbon making direct hydrophobic contacts to the conserved Val76. These observed interactions are consistent with biochemical studies showing that direct contacts between ligand -2 residue and the residues from PDZ αB helix are critical for the binding specificity of class I PDZ-ligand interaction[122, 123]. For example, mutation of the His72-equivalent residue in ERBB2IP-1 to Tyr, Asn, Gln or Lys, all capable of forming hydrogen bonds to threonine, did not alter specificity significantly, whereas substitution of the residue with Leu, Val or Met resulted in class II specificity profiles with preference for hydrophobic residues at -2 position[122]. Therefore, our structure, coupled with these previous results, indicates that the stabilization and specificity of PDZ1-CXCR2 interaction are dependent on both Leu0 and Thr-2 that possess the ability to form networks of hydrogen bonds and hydrophobic interactions with NHERF1.

Ligand-specific PDZ1-CXCR2 interactions

Compared to the motif residues (0 and -2), the peptide residues at positions -1 and -3 are largely exposed, with both side chains oriented upwards in the PDZ1-CXCR2 complex (Figure. 15). As a result of this orientation, the -1 and -3 residues make fewer direct contacts with PDZ1 and bury a much less extent of solvent-accessible surface areas than the motif residues (50 $Å^2$, -1; 40 $Å^2$, -3; 80 $Å^2$, -0; 75 $Å^2$, -2). These findings are consistent with previous evidence that both -1 and -3 residues in the peptide ligands were less stringently specified by individual PDZ domains than the residues at the 0 and -2 positions[123]. Specifically, the interactions with Thr-1 include a direct polar contact from

its side chain hydroxyl to the side chain of His27 and a water-mediated hydrogen bond between its carbonyl oxygen and the side chain of Arg80 (Figure. 15A). In these aspects, the PDZ1-CXCR2 structure differs significantly from the structures of other PDZ1-ligand complexes. In PDZ1-CFTR, the guanido group of Arg-1 forms two salt bridges to the Glu43 side chain and two hydrogen bonds with the carbonyl oxygen of Asn22[1], while in PDZ1-β2AR and PDZ1-PDGFR, the nonpolar residues at position -1 of the peptide ligands engage in direct hydrophobic interactions with the imidazole ring of His27[4]. These observed differences reveal that there is considerable diversity in PDZ1 interaction with -1 residue of different ligands, manifested by four chemically different amino acids (Asn22, His27, Glu43, and Arg80) combined in the discrete ways to discriminate the ligand residues of different hydrophobicity and polarity. This diversity reflects a high degree of selectivity in NHERF1 ligand recognition, consistent with a vast potential for PDZ domain family to bind different sequences[122].

The interactions between PDZ1 and CXCR2 at position -3 of the peptide are also very different compared to other PDZ1 complexes. In PDZ1-CXCR2, the hydroxyl group of Ser-3 forms a direct hydrogen bond with the His29 side chain and a water-mediated hydrogen bond to the imidazole ring of His27 (Figure. 15A). In contrast, the side chain of residue Asp-3, which is common in CFTR, β2AR, and PDGFR, is engaged in salt bridge interaction with the Arg40 guanidinium and direct hydrogen bonding to the side chain of His27[4, 124]. These structural differences appear to be important for PDZ1 ligand discrimination, as it was shown that highly specific contacts with different types of

contextual residues contributed significantly to the binding specificities of all peptide-mediated protein interactions[136]. In agreement with this conclusion, the structure of the NHERF2 PDZ2 in complex with the PSTRL sequence revealed the occurrence of similar interactions between Ser-3 and a histidine residue (His166) of the PDZ domain[132].Remarkably, the NHERF2 PDZ2 His166 residue corresponds to NHERF1 PDZ1 His29 (Figure. 14A), suggesting that the amino acid at this position may play a critical role in specific ligand recognition via interaction with the -3 residue of the peptide. Taken together, the present structure indicates that the peptide residues at positions -1 and -3 contribute to ligand specific PDZ1-CXCR2 interactions, suggesting that these positions may have been naturally selected to facilitate PDZ ligand selection within a complex network of NHERF-scaffolded interactions[122]. Interestingly, the residues at the -1 and -3 positions exhibit significant variability across natural NHERF1 binding targets, with the two-residue combination unique to each characterized ligand (Figure. 14C).

The considerable contacts between PDZ1 and the residues at positions -1 and 3 suggest that these residues may play an important role in the affinity of the PDZ1-CXCR2 interaction. Consistent with this suggestion, affinity selection experiments showed that NHERF PDZ1 almost exclusively selected ligands with arginine at position -1 from random peptides, and mutation of Arg to Ala, Phe, Leu, or Glu decreased the affinity of the PDZ1-ligand interaction by 2-10 fold[1, 137]. In addition, it has been shown that position -3 is also an important determinant of binding affinity, with PSD-95

preferring to bind peptides with acidic side chains at this position[123]. Furthermore, analysis of the binding specificities for nearly half of over 330 PDZ domains in human and worm revealed that there was a strong correlation between natural PDZ sequences and ligand specificities at both -1 and -3 positions of peptides[122]. Remarkably, the PDZ binding preferences at these positions can be influenced by multiple structural and chemical mechanisms involving both direct contacts and cooperative, long-range effects, suggesting that binding specificities can evolve rapidly, thus enabling PDZ for robust differentiation between biologically diverse ligands[122]. Therefore, our structure, together with these previous findings, suggests that the ligand specific contacts between PDZ1 and the CXCR2 -1 and -3 residues are important for the binding affinity and specificity of the PDZ1-CXCR2 interaction. In a broad term, the ligand specific interactions at these positions could lead to different PDZ-ligand complex stabilities, which, in conjunctions with an increasingly complex NHERF interaction network[138], may determine signaling orchestration and underlie the highly coordinated regulation of manifold NHERF-controlled signaling events[139]. In support of this idea, recent biochemical studies suggested that NHERF1, NHERF2, and CAL competed to regulate CFTR endocytic processing, and the differences in their CFTR binding affinities were required for CFTR to efficiently escape CAL-mediated degradation through repeated rounds of uptake and recycling[127].

Structural Comparison Reveals PDZ1 Target Selection Specificity

To uncover the structural details that govern the CXCR2-NHERF1 ligand specific

interactions, we compared the PDZ1-CXCR2 structure to the crystal structures of all available NHERF1 PDZ1-ligand complexes, including PDZ1-CFTR, PDZ1-β2AR, and PDZ1-PDGFR[1, 4]. The structural comparison reveals that the four PDZ1 structures are highly similar, with pairwise RMSDs (root-mean-square differences) for entire Cα atoms ranging from 0.48 to 0.83 Å (Figure. 16A). Larger differences in the PDZ1 backbone are found at two loop regions (β2-β3 and α2-β6 loops), but note that these loops made of non-conserved residues (Figure. 14B) are conformationally flexible, as indicated by poorly defined electron density and higher than average B factors (data not shown). Moreover, the backbone conformations of the bound peptides are also highly superimposed (RMSDs from 0.48 to 0.83 Å), as are their relative spatial positions to the conserved PDZ1 motifs (Figure. 16B). These findings therefore indicate that the binding of different peptides has little effect on the PDZ1 overall fold, consistent with previous studies showing that the localized changes at a few key positions within the PDZ fold were responsible for dramatically altered PDZ binding specificity[140]. Indeed, significant differences are observed only in the peptide-binding pocket, especially at PDZ residues that are involved in recognition of different side chains at position -1 and -3 of the peptide ligands. In particular, the structural alignments reveal that the Asn22 side chain has two different orientations, while the conformation of the Glu43 side chain differs among all four PDZ1 structures (Figure. 16B). Such structural differences have been noted before and led to the conclusion that the conformational changes of Asn22 and Glu43 underlay the PDZ1 flexibility to accommodate ligands with -1 side chains of different hydrophobicity and polarity[4].

The availability of the PDZ1-CXCR2 structure, however, not just confirms the above conclusion, but also has the potential to reveal differential -3 side chain recognition, i.e., how PDZ1 differentiates CXCR2 Ser-3 from Asp-3 of CFTR, β2AR and PDGFR. In this context, it is interesting to note that the most striking difference among the PDZ1 complexes is the His29 side chain, which adopts two different conformations. In PDZ1-CXCR2, the side chain of His29 is oriented toward the hydroxyl group of Ser-3, participating in specific ligand interaction; whereas in other three complexes, the imidazole ring of His29 points away from the bound ligands and does not engage in any peptide recognition (Figure. 16B). Strikingly, this conformational change is accompanied by large alteration in the Arg40 rotameric state, which rotates to make completely different PDZ1-peptide interactions. In PDZ1-CFTR, PDZ1-β2AR, and PDZ1-PDGFR, Arg40 is a key anchor residue for specific Asp-3 recognition and participates in direct ligand binding[1, 4]. In PDZ1-CXCR2, due to steric effects, the reorientation of His29 forces the Arg40 side chain to kink outwards and prevents it from interacting with the shorter side chain of Ser-3 (Figure. 16B). Therefore, these observed differences demonstrate that the structural variability surrounding the peptide-binding pocket is important for PDZ1 ligand specific interactions, and that the rotameric differences of a few key residues constitute the basis for PDZ1 robustness to bind a diverse array of functionally different proteins[122, 140].

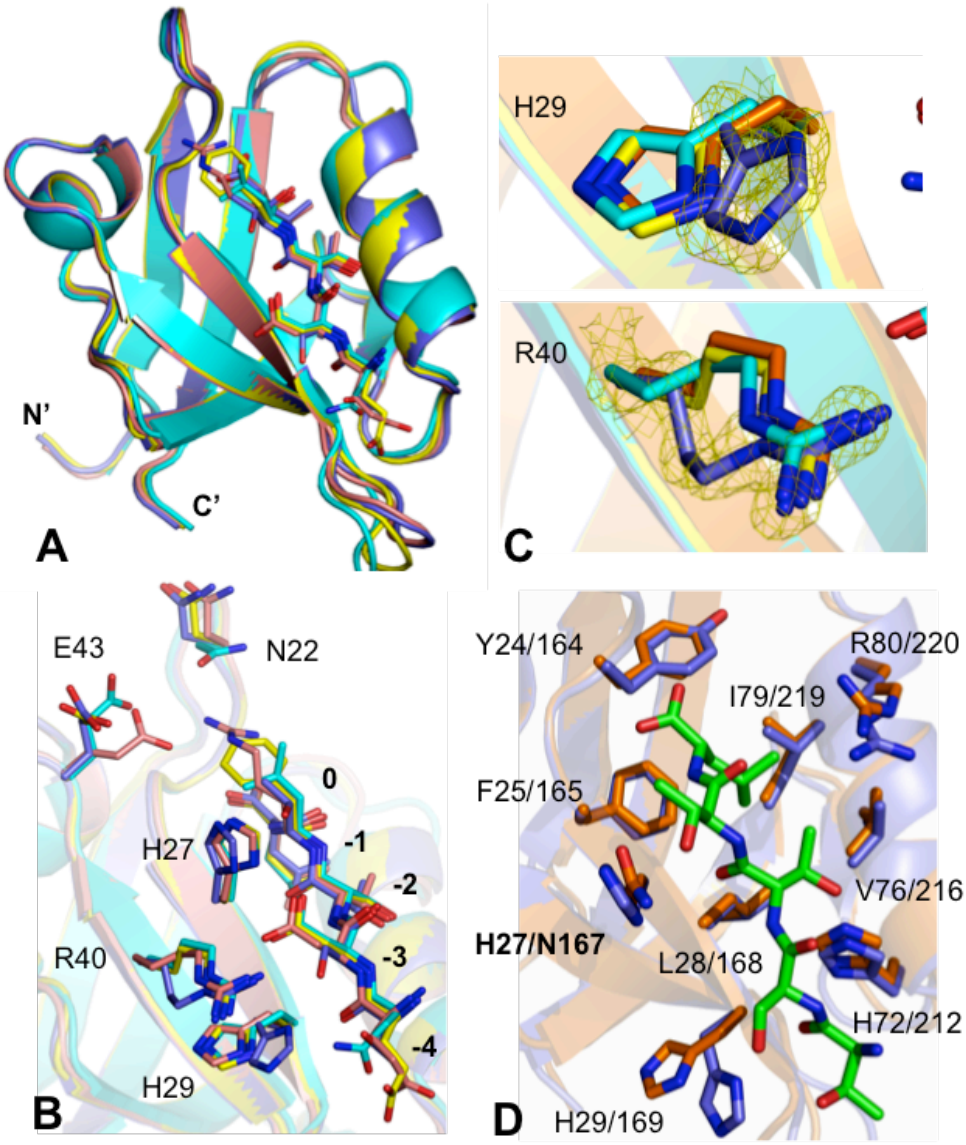

Figure 16. Structural comparison of PDZ domains. (A) Superposition of the structures of PDZ1-CXCR2 (purple; PDB code: 4JL7), PDZ1-CFTR (orange; PDB code: 1I92)[1], PDZ1-β2AR (cyan; PDB code: 1GQ4)[4], and PDZ1-PDGFR (yellow; PDB code: 1GQ5)[4]. PDZ domains are represented by ribbon, while residues in the ligands are displayed as sticks. (B) Superposition of the PDZ1 ligand binding pockets. Both PDZ1 and ligand residues are depicted by sticks and colored according to the scheme in Fig 3A. (C) Close-up views of structural differences of His29 (top) and Arg40 (bottom). The CXCR2 peptide is depicted by sticks overlaid with 2Fo − Fc omit map calculated at 1.16 Å and contoured at 2.0 σ. (D) Superposition of NHERF1 PDZ1 (purple) and PDZ2 (pink; PDB code: 2OZF) peptide binding pockets. CXCR2 peptide is shown in green and PDZ residues are depicted by balls-and-sticks.

CXCR2 Interacts with Both NHERF1 PDZ1 and PDZ2

The structural alignment reveals that NHERF1 PDZ1 and PDZ2 share highly similar overall structures and also highly conserved ligand binding pockets (Figure. 16C). The only notable difference in the ligand binding sites is residue 27, which is the His in PDZ1 and Asn (residue 164) in PDZ2. It should be noted that this conserved substitution maintains the amino functionality of the side chains, and thus, is not expected to disrupt the observed polar interactions between the CXCR2 peptide and PDZ1 (Figure. 15A). Based on that, we hypothesize that NHERF1 PDZ2 may also bind to CXCR2. Indeed, we showed that CXCR2 interacts with both PDZ1 and PDZ2 in the GST-pulldown experiments, with PDZ2 exhibiting higher binding affinities (Figure. 17). Specifically, we overexpressed CXCR2 in HEK293 cells and then performed pulldown assays from cell lysates using various GST-PDZ constructs. Whereas no CXCR2 was detected in the control lane containing GST alone, significant amounts of CXCR2 were found in the lanes containing PDZ1 domain (GST-PDZ1), PDZ2 domain (GST-PDZ2), and both PDZ domains together (GST-PDZ1-PDZ2). To test whether the PDZ-CXCR2 interactions are direct, we performed in vitro pulldown experiments with a biotinylated peptide corresponding to the last 13 amino acids of CXCR2. Similar binding results were observed in the experiments where CXCR2 interacts with both PDZ domains of NHERF1 (Figure. 17B).

Many other NHERF1 ligands, such as CFTR, PDGFR and PTH1R, were also known to bind both PDZ1 and PDZ2 in vitro[127, 141, 142], but in most cases, the biological

significance of such bivalent interactions remains unknown. It has been shown that bivalent binding was important for CFTR channel gating regulation, and disruption of the PDZ2-CFTR interaction, but keeping the interaction between PDZ1 and CFTR intact, was able to abolish the NHERF1 stimulatory effect on CFTR channel open probability[143, 144]. In addition, it has been suggested that a single NHERF1 molecule could assemble a PDGFR dimer and played a role in PDGFR signaling via stabilizing the ligand-induced receptor dimerization[145]. Later studies, however, revealed that PDGFR signaling was unexpectedly enhanced rather than impaired in NHERF1-null mouse embryonic fibroblasts, suggesting quite a different role of this bivalent molecule in PDGFR signaling regulation[146]. Remarkably, a recent article by Cardone et al. showed that NHERF1 PDZ1 and PDZ2 domains differently regulated invadopodia and podosome dynamics[134], and suggested that the differential functions of two PDZ domains might be dependent on their ability to interact with a unique array of functionally different signaling molecules[134]. Based on that, it is reasonable to speculate that the ability of CXCR2 to bind both NHERF1 PDZ domains may allow CXCR2 to operate in different signaling networks, which might be a key functional trait that has evolved to deal with the complexity of signaling transduction. While the biological impacts of this bivalent binding are currently unknown, future studies should be directed toward evaluation of its effects on CXCR2-mediated neutrophilic migration, receptor dimerization, CXCR2 internalization, and especially determining whether different NHERF1 PDZ domains could mediate the assembly of distinct and specific CXCR2 signal transduction complexes.

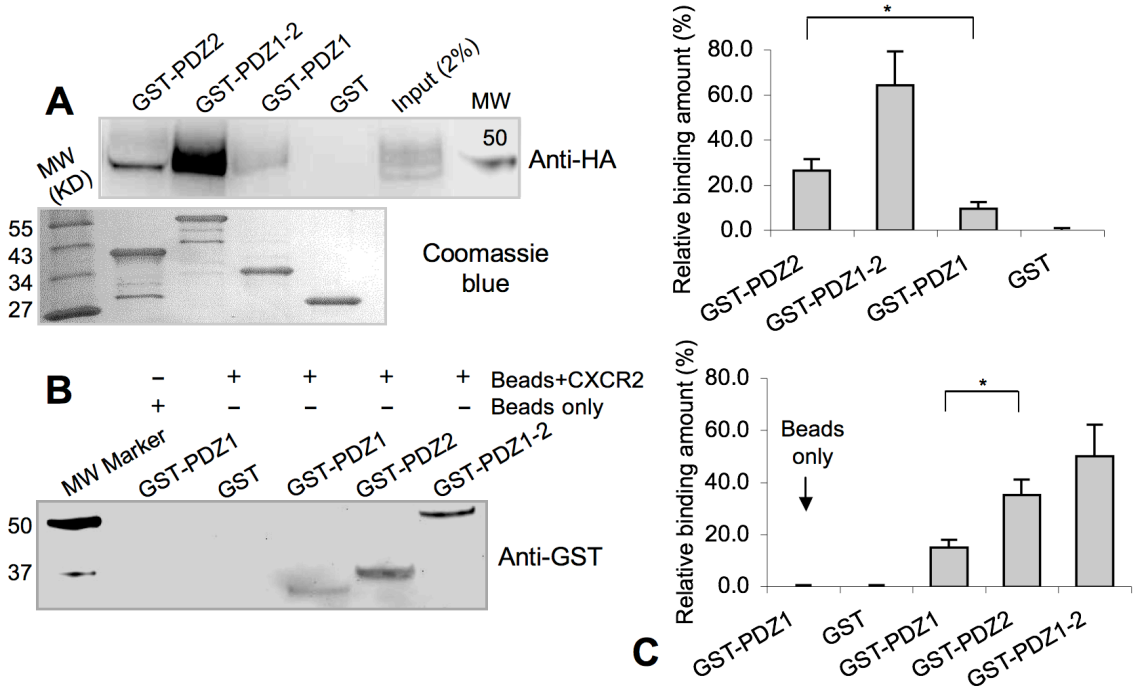

Figure 17. CXCR2 interacts with both PDZ1 and PDZ2 of NHERF1. (A) GST pull-down of CXCR2 with NHERF1. Lysates of HEK293 cells overexpressing HA-tagged CXCR2 were used as prey. GST fusion proteins of NHERF1 PDZ1, PDZ2, and PDZ1-PDZ2 were used as bait. GST alone served as a negative control. Binding experiments were analyzed by SDS-PAGE and visualized by immunoblot using anti-HA antibodies. The amount of beads-immobilized GST proteins in each reaction is shown in the lower panel. (B) Biotin pull-down assays to detect direct interaction between CXCR2 and NHERF1. A biotinylated peptide corresponding to the last 13 residues of CXCR2 was used as bait, while purified GST-PDZ1, GST-PDZ2, GST-PDZ1-PDZ2 and GST alone as prey. Binding was resolved by SDS-PAGE and immunoblotted with anti-GST antibodies. (C) All experiments performed in (A) and (B) were repeated three times. The results were quantified using the CCD gel imager (UVP Chemidoc) and presented as mean±standard deviation. Top: GST pull-down of CXCR2 with NHERF1, and bottom: Biotin pull-down of NHERF PDZ domains with the CXCR2 peptide.

Drug Design Perspective

Due to the exceptional importance of CXCR2 in inflammation and tumorigenesis[117], the structural determinants of the CXCR2-NHERF1 interaction may be valuable in developing new methods and strategies for targeted drug discovery. For

example, this information can be used to create new CXCR2 inhibitors that are potent and specific to block the CXCR2-NHERF1 interaction. Such inhibitors could in turn have a therapeutic potential in inhibiting neutrophil-driven inflammation by reducing neutrophil recruitment and restoring neutrophils to the tissue clearance pathway of apoptosis[147]. In this context, it is interesting to note that disruption of the CXCR2-NHERF1 complex was sufficient to inhibit the IL-8-induced neutrophilic chemotaxis and margination[119]. Therefore, small molecules and peptides that specifically block the CXCR2-NHERF1 interaction could act as CXCR2 antagonists and could be useful in attenuating the signaling activities of CXCR2 in various neutrophil-related inflammation disorders, such as inflammatory bowel diseases, chronic lung inflammation, and atherosclerosis[119].

However, the commonality of peptide recognition at position 0 and -2 by class I PDZ domains, together with NHERF1 capable of binding to a multitude of ligands (Figure. 14C), poses a challenge for designing CXCR2 inhibitors that are specific to the CXCR2-NHERF1 interface but do not cross-react with any of the other NHERF1-mediated interactions. NHERF1, through a network of PDZ domain-mediated interactions, regulates many cellular processes essential to normal physiological functions, such as testicular differentiation, signal transduction, endosomal recycling, membrane targeting, and hormone receptor desensitization[148, 149]. Therefore, it is conceivable that random targeting of NHERF1-ligand interactions by nonselective inhibitors could disrupt the NHERF1 interaction network and leads to considerable risks

with a diverse range of unwanted physiological and hormonal abnormalities. Regarding this possibility, it is particularly important to note that contextual specificity plays a key role in all peptide-mediated protein interactions[136], suggesting that the ability to achieve CXCR2 inhibitor selectivity is dependent on the identification and exploitation of structural features that differentiate CXCR2 from other NHERF1 binding partners, and on understanding how the peptide motif and context work in coordination to control the specificity and formation of each crucial NHERF-scaffolded signaling complex. This notion is consistent with accumulating evidence that the positions other than 0 and -2 make significant and variable contributions to both affinity and specificity of the PDZ-mediated interactions[4, 124]. For example, recent large-scale PDZ specificity mapping studies demonstrated that the PDZ domain family is surprisingly complex and diverse, recognizing up to 7 C-terminal ligand residues and forming at least 16 unique specificity classes across human and worm[122]. Furthermore, we recently showed that, although the motif-contacting residues involved in CXCR2 binding are all conserved in NHERF1 and PDZK1 (Figure. 14A), CXCR2 did not interact with PDZK1 in the *in vitro* GST pull-down assays[119], reciprocally suggesting that high affinity CXCR2 binding and selection by NHERF1 is also context dependent. Therefore, strategies aiming at exploiting CXCR2-NHERF1 contextual interactions may represent a promising approach for the development of small molecules that would selectively block this interaction and specifically inhibit the neutrophil-driven inflammation. In this context, it is particularly important that the ligand-specific structural principles that govern the NHERF1 target-selection diversity should be addressed in great detail.

CHAPTER 5

New Conformational State of NHERF1-CXCR2 Signaling Complex Captured by Crystal Lattice Trapping

*Published in PLoS One 2013 Dec 10;8(12): e81904. doi: 10.1371/journal.pone.0081904. All authors agreed with including their work in this dissertation.

Abstract

NHERF1 is a PDZ adaptor protein that provides a scaffold for the assembly of diverse signaling complexes and has been implicated in many cancers. However, little is known about the mechanism responsible for its ability to bind to multiple targets. Computational studies have indicated that PDZ promiscuity may be attributed to its conformational dynamics, but experimental evidence for this relationship remains very limited. Here we examine the conformational flexibility of the NHERF1 PDZ1 domain using crystal lattice trapping via solving PDZ1 structure of a new crystal form. The structure, together with prior PDZ1 structures of a different space group, reveals that 4 of 11 ligand-interacting residues undergo significant crystal packing-induced structural changes. Most of these residues correspond to the residues involved in allosteric transition when a peptide ligand binds. A subtle difference in peptide conformations causes the same ligand to bind in slightly different modes in different crystal forms. These findings indicate that substantial structural flexibility is present in the PDZ1

peptide-binding pocket, and the structural substrate trapped in the present crystal form can be utilized to represent the conformational space accessible to the protein. Such knowledge will be critical for drug design against the NHERF1 PDZ1 domain, highlighting the continued need for experimentally determined PDZ1-ligand complexes.

Introduction

The Na^+/H^+ exchanger regulatory factor 1 (NHERF1) is a multifunctional scaffold protein that plays a central role in diverse cellular events through recruiting receptors, transporters, and signaling molecules into specific functional complexes[150]. NHERF1 also plays a significant role in multiple cancers where its elevated expression correlates with aggressive stage and poor overall prognosis[134]. The functional diversity of NHERF1 in normal and pathological conditions depends largely on its two PDZ (PSD-95/Discs-large/ZO-1) domains, PDZ1 and PDZ2, which are highly promiscuous and capable of interacting with a large number of biologically different proteins[150]. To date, over 40 binding partners of NHERF1 have been identified; most of which are membrane receptors and transporters, such as the interleukin 8 receptor beta (CXCR2), the cystic fibrosis transmembrane conductance regulator (CFTR), the β2-adrenergic receptor (β2AR), the platelet-derived growth factor receptor (PDGFR), and the parathyroid hormone receptor (PTHR)[119, 127, 128, 151, 152]. Through these PDZ-interacting proteins, NHERF1 regulates many processes, including cell proliferation, invasion and migration, signal transduction, and protein trafficking[148, 149]. Our recent studies showed that the PDZ domains of NHERF1 bind CXCR2 in neutrophils, regulating neutrophil chemotaxis and

directing neutrophils to sites of inflammation[119]. A similar interaction was observed for pancreatic cancers, where disruption of this PDZ-mediated interaction was capable of suppressing human pancreatic tumor growth in vivo[152]. These recent evidence suggests targeting the PDZ-mediated NHERF1-CXCR2 interaction may represent a novel clinical strategy, which could be valuable in the development of new treatments against numerous neutrophil-dependent inflammatory diseases as well as pancreatic cancers[148, 149].

In general, PDZ domains mediate protein interactions by recognizing the C-terminal sequence of target proteins, and by binding to the targets through a canonically and structurally-conserved PDZ peptide binding pocket[121]. Based on the residues at positions 0 and -2 of the peptides (position 0 referring to the C-terminal residue), early studies grouped PDZ domains into two major specificity classes: class I, (S/T)X(V/I/L) (X denoting any amino acid); class II, (F/Y)X(F/V/A) [1, 123, 124]. However, growing evidence indicates that PDZ specificity is unexpectedly complex and diverse, with the PDZ domain family recognizing up to 7 C-terminal ligand residues and forming at least 16 unique specificity classes[122]. In addition, the complexity of PDZ-peptide interactions is exemplified by the facts that many PDZ domains can bind to multiple ligands of different peptide classes, and that single peptides are able to bind to distinct PDZ domains[122]. This complex picture of PDZ-peptide interactions raises a challenging problem regarding how PDZ domains, structurally simple protein interaction modules,

achieve binding promiscuity and specificity concomitantly, the nature of which remains obscure.

Because promiscuity and specificity have important implications in highly selective drug design[153], understanding the mechanism that determines PDZ interaction with specific peptide sequences is a subject of intensive research. Recent binding specificity studies of 157 mouse PDZ domains revealed that PDZ domains are evenly distributed throughout selectivity space, suggesting that they have been optimized across the proteome to minimize cross-reactivity[154]. The same study revealed a weak but significant correlation between the pairwise sequence divergence of PDZ domains and their divergence in ligand selectivities[154]. More recent specificity profiling studies with 91 point mutants of a model PDZ domain revealed that PDZ binding preference can be influenced by multiple structural and chemical mechanisms involving both direct contacts and cooperative, long-range effects, suggesting that PDZ specificity evolves rapidly, thus enabling PDZ for robust interaction with many biologically distinct ligands[122]. Using shotgun alanine scanning, another PDZ specificity study has yielded considerable insights into the relationships between primary sequence and specificity[140]. This study demonstrated that most of the alanine substitutions in HtrA1-PDZ are neutral with respect to peptide-binding selectivity and only a subset of mutations, mostly within the canonical PDZ binding pocket, affects its binding specificity[140]. Therefore, the results of these studies have offered considerable information about how the sequence composition

determines PDZ specificity, and a coherent picture of their relationships is beginning to emerge.

Despite the wealth of detail about PDZ specificity, the mechanism that determines PDZ promiscuity still remains poorly understood, partly because it has been difficult to explain PDZ promiscuity simply based on its sequence composition. It is important to note that a number of computational and experimental studies have suggested the conformational dynamics of PDZ domains may play a crucial role in ligand binding, especially in determining binding promiscuity[126, 155-157]. For example, molecular dynamics simulation of 12 PDZ domains revealed that binding dynamics and entropy are extremely variable not only across PDZ domains but also for the same PDZ domain bound with different ligands[155]. This indicates that complex-specific dynamical or entropic responses may form the basis for promiscuous binding and sustaining promiscuity in highly selective PDZ-peptide interactions[155]. Another computational study of five different PDZ domains came to similar conclusion. It revealed the existence of a close relationship between intrinsic dynamics and binding promiscuity and suggested the ability of PDZ domains to interact with multiple ligands requires the binding pocket to adopt significantly different conformations[126]. In addition, based on differential domain fluctuation profiles, the latter study also indicated that both induced fit and conformational selection play roles in PDZ ligand binding, but the extent to which these mechanisms are involved is highly variable across the PDZ domain family[126]. Remarkably, recent NMR dynamics studies demonstrated that the ligand-bound

conformation is already present in the conformational ensemble populated by unliganded protein, suggesting the intrinsicality of protein to fluctuate between multiple conformers, or conformational selection, might be the fundamental paradigm for promiscuous ligand binding[158-160]. These studies made it apparent that detailed and comparative analysis of PDZ conformational plasticity will be required to establish and illuminate the full range of ligand promiscuity specified by the PDZ domain fold. A high-resolution structural interpretation of individual conformational states should in turn provide considerable insights into the mechanisms whereby the exquisite ligand promiscuity dictates the diversification of biological functions.

In order to understand the promiscuity and specificity of the NHERF1 PDZ domains, we have previously reported a high-resolution PDZ1 crystal structure in complex with the CXCR2 C-terminal sequences[151]. We found that NHERF1 PDZ1 is capable of assuming distinct conformational states when the structure of PDZ1-CXCR2 was compared to the structures of three other PDZ1 complexes, including PDZ1-CFTR, PDZ1-PDGFR, and PDZ1-β2AR[1, 4]. Importantly, the complex-specific conformations were found to be closely associated with the various characteristics of peptide ligands[4, 151], suggesting that PDZ1 promiscuity is facilitated by protein flexibility that allows robust accommodation of peptides with distinct sequences. While these studies provided valuable insight about PDZ1 promiscuity and flexibility, the questions still remain concerning the dynamical features that control explicit binding of each of PDZ1 ligands and whether NHERF1 function relies on PDZ1 conformational diversity. Additionally,

we are still far from a complete description of PDZ1 conformational space, and the amount of available PDZ1 structures may represent only a tiny fraction of the entire ensemble[126, 155-157]. It is conceivable that limited numbers of PDZ1 structures could limit their usefulness in rational drug design owing to large unexplored conformational space that may compensate drug discovery efforts for potency and selectivity. Moreover, the lack of a complete picture of PDZ1 conformational space could lead to an incomplete understanding of the complex relationship between PDZ1 conformational dynamics and the promiscuous nature of its substrate specificity. In these regards, we here present a new conformational state of PDZ1 by solving the structure of the PDZ1-CXCR2 complex in a new crystal form. Multiple PDZ1 conformations observed in the present crystal form and another crystal form reported previously[151] provide an additional insight into PDZ1 conformational dynamics and a structural explanation for how PDZ1 is able to bind to different ligands. Alternatively, the variations in the structures of different crystal forms raise the challenge for selective drug design, emphasizing the need for obtaining X-ray crystal structures of various PDZ1 conformational states to inform the drug design process.

Materials and Methods

Protein Expression and Purification

A DNA fragment encoding the human NHERF1 PDZ1 (residues 11–94), and having the C-terminal extension TSTTL that corresponds to residues 356–360 of human CXCR2, was amplified using PCR and cloned in the pSUMO vector[151]. The resulting

clone containing a N-terminal His6-SUMO tag was transformed into *Escherichia coli* BL21 Condon Plus (DE3) cells for protein expression. The transformants were grown to an OD600 (optical density at 600 nm) of 0.4 at 37 °C in LB medium, and then induced with 0.1 mM isopropylthio-β-D-galactoside and grown an additional 16 hours at 15 °C. The cells were harvested by centrifugation and lysed by French Press. The soluble fraction was then subjected to Ni^{2+} affinity chromatography purification, followed by the cleavage of the His6-SUMO tag with yeast SUMO Protease 1. PDZ1 was separated from the cleaved tag by a second Ni^{2+} affinity chromatography and further purified by size-exclusion chromatography. Finally, the protein was concentrated to 40–50 mg/ml in a buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1 mM β-mercaptoethanol (BME), and 5% glycerol.

Crystallization, Data Collection and Structure Determination

Crystals were grown by the hanging-drop vapor-diffusion method by mixing the protein (~25 mg/ml) with an equal volume of reservoir solution containing 100 mM sodium acetate, pH 4.8, 0.2 M ammonium acetate, 25% PEG4000 at 20 °C. Crystals typically appeared overnight and continued to grow to full size in 3-4 days. Before X-ray diffraction data collection, crystals were cryoprotected in a solution containing mother liquor and 25% glycerol and flash cooled in liquid nitrogen. The data were collected at 100 K at beamline 21-ID-F at the Advanced Photon Source (Argonne, IL) and processed and scaled using the program HKL2000[67]. Crystals belong to space group $P2_1$ with unit cell dimensions a = 26.6 Å, b = 45.5 Å, c = 33.4 Å, β = 109.7°, and one molecule in the

asymmetric unit. The structure was solved by the molecular replacement method with program PHASER[96] using the $P3_121$-PDZ1 structure (PDB code: 4JL7) as a search model. The structure modeling was carried out in COOT[69], and refinement was performed with PHENIX[97]. The riding hydrogen and ADP features were included in the refinement, and no ADP restraint was employed. The final models were analyzed and validated with Molprobity[99]. The ADPs were analyzed using ANISOANL[161] and the PARVATI server[162]. All figures of 3D representations of the $P2_1$-PDZ1 structure were made with PyMOL.

Protein Data Bank Accession Number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 4MPA ($P2_1$-PDZ1) and 4N6X ($P3_121$-PDZ1).

Results

New Crystal Form of PDZ1-CXCR2 Complex

Alternative crystal forms can trap a protein in different conformational states, providing snapshots of the conformations accessible to the protein[163-165]. To reveal possible PDZ1 conformational states and how these may be important for PDZ1 promiscuity, we sought to use this well-recognized strategy to improve our understanding on PDZ1 conformational dynamics. Previously, we crystallized the PDZ1-CXCR2 complex in the $P3_121$ space group and have determined its structure at 1.16 Å resolution ($P3_121$-PDZ1)[151]. In the current study, by using different crystallizing precipitant under

Table 4. Crystallographic data and refinement statistics

Space group	$P2_1$
Cell parameters (Å)	
a	26.6
b	45.5
c	33.4
Wavelength (Å)	0.97856
Resolution (Å)	45.5-1.1 (1.16-1.10)[a]
R_{merge}[b]	0.024 (0.180)
R_{meas}[c]	0.034 (0.248)
$CC_{1/2}$[d]	0.999 (0.951)
Redundancy	3.7 (2.4)
Unique reflections	30032
Completeness (%)	97.7 (86.0)
$\langle I/\sigma \rangle$	24.8 (4.5)
Wilson B-factor	8.2
Refinement	
Resolution (Å)	25.9-1.10
Molecules/AU	1
R_{work}[e]	0.143 (0.158)
R_{free}[f]	0.156 (0.193)
Ramachandran plot by Molprobity	
Residues in favored region	97.9%
Residues in allowed region	2.1%
RMSD	
Bond lengths (Å)	0.010
Bond angels (°)	1.2
No. of atoms	
Protein (residues 9-94)	679
Peptide (residues 95-99)	39
Water	161
Chloride	1
Acetate	4
B-factor (Å^2)	
Protein	14.8
Peptide	9.7
Water	26.3
Chloride	16.2
Acetate	23.4

[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]$R_{meas} = \Sigma [(n/n-1)]^{1/2} \Sigma |I-\langle I \rangle| / \Sigma I$, where n is the number of observations of a given reflection.

[d]Half-dataset correlation coefficient.

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

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

similar pH, we obtained a new crystal form that diffracted to 1.10 Å resolution. The new crystal belongs to the $P2_1$ space group ($P2_1$-PDZ1), and the structure was solved by molecular replacement. The model was refined to R_{work} of 14.3% and R_{free} of 15.6%, and the validation of its stereochemistry using Molprobity[99] showed that 97.9% of the residues are in the most favored regions, 2.1% in the additional allowed regions, and 0.0% in the disallowed regions (Table 4).

Both crystal forms contain one molecule per asymmetric unit, but their crystal packing environments differ significantly. For $P3_121$ the distinctive packing pattern is manifested by linear stacking of PDZ1 complexes, hexagonal lateral association, and the existence of large solvent channels across the crystal (Figure. 18A). In the case of $P2_1$, the PDZ1 complexes are stacked in a staggered arrangement, displaying a densely packed, flattened configuration (Figure. 18B). Consistent with the packing environments, the solvent content in the $P3_121$ crystal form is higher than $P2_1$-PDZ1, ~50% compared to ~37%. However, analysis of crystal contacts reveals that there are more intimate packing interactions in the $P3_121$ crystal. For example, with distances of less than 3.5 Ådefined as contacts, $P3_121$ has 128 crystal contacts with symmetry-related molecules, whereas the $P2_1$ crystal has only 82 such contacts. Accessibility calculation with AREAIMOL[161, 166] shows that 2608 $Å^2$ of protein surface is buried by symmetry-related molecules in $P3_121$-PDZ1, compared to only 2382 $Å^2$ buried in $P2_1$-PDZ1. Thus, it appears that the protein molecules in the $P2_1$ crystal pack more loosely than in $P3_121$-PDZ1, though it has a relatively lower solvent content. Furthermore, their distinct

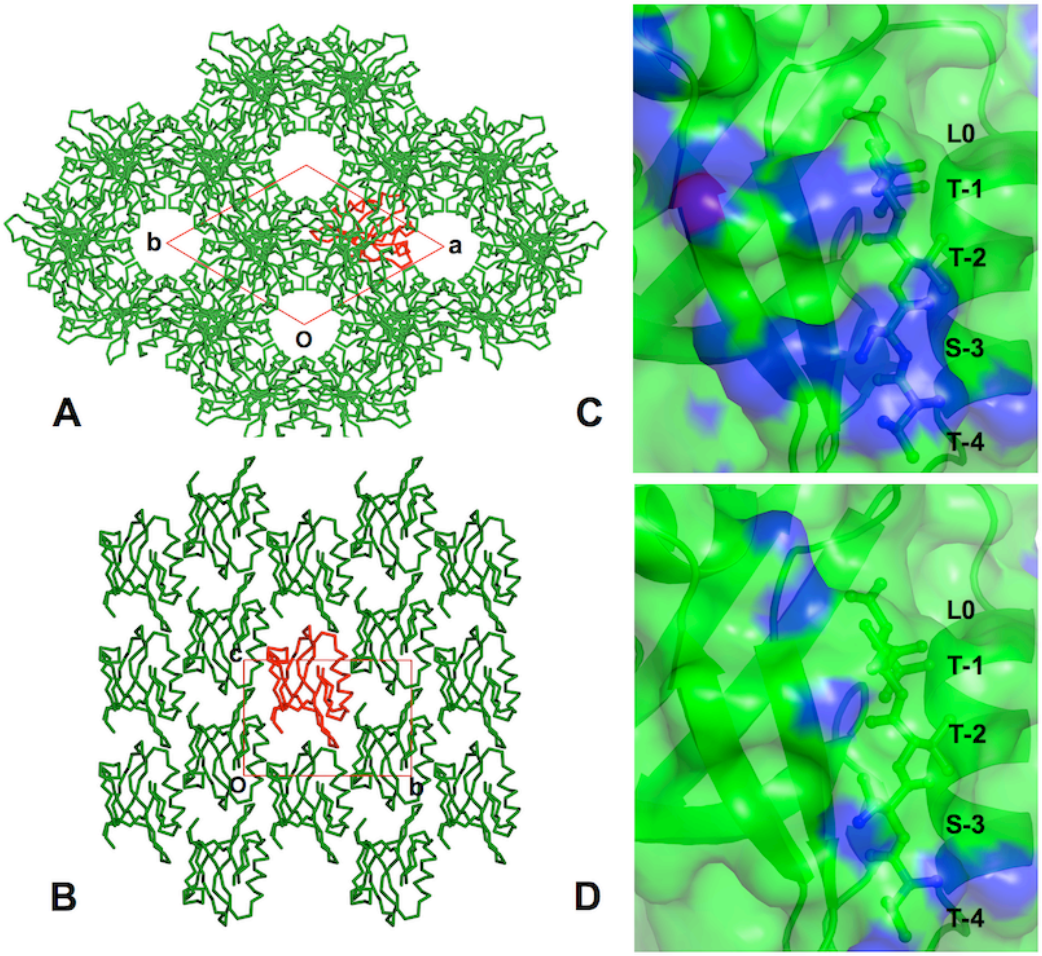

Figure 18. Crystal packing differences between two crystal forms. (A) Section of the crystal lattice of $P3_121$-PDZ1 and (B) $P2_1$-PDZ1. The unit cell is shown as a red box, with the origin and axes labeled. PDZ1 is shown as a Cα trace, with red standing for a reference molecule and green the symmetry-related molecules. (C) Surface representation of crystal contacts around the ligand-binding site of $P3_121$-PDZ1 and (D) $P2_1$-PDZ1. The surface is colored in blue if the distance to symmetry-related molecules is 3.5 Å or less and is colored in green otherwise. PDZ1 is depicted as ribbon and the bound CXCR2 peptide is labeled and represented by sticks.

packing environments are highlighted by strikingly large differences in their crystal contact surfaces. For all of the 128 contacts sites found in $P3_121$, there is no corresponding contact surface with equivalent residue composition in $P2_1$-PDZ1 (Figure.

18C and 18D). This difference provides the basis for us to utilize crystal packing in understanding PDZ1 conformational dynamics and should allow the capture of different conformational substrates.

Distinct PDZ1 Conformational States

Different crystal packing observed in $P2_1$-PDZ1 and $P3_121$-PDZ1 indeed causes significant differences in the ligand-binding pocket, but does not alter the overall fold of the protein (Figure. 19). In both crystal forms, PDZ1 adopts a conserved fold characterized by six β strands (β1–β6) and two α-helices (αA and αB). Superposition of the two structures reveals a high degree of overall structural similarity, with the rms (root-mean-square) differences of 0.91 Å for main chains and 1.46 Å for side chains. In addition, the crystal packing has little effect on the overall ligand interaction mode, as in both cases the CXCR2 peptide inserts between β2 and αB as an extra β-strand and the main-chain rms difference between the bound peptides is only 0.17 Å (Figure. 19B). Moreover, closer inspection of the PDZ1-CXCR2 interactions reveals that the specific ligand recognition modes at the peptide positions 0 and -2 are virtually indistinguishable. In both crystal forms, the side chain of Leu0 is nestled in a deep hydrophobic pocket formed by structurally identical residues, including Tyr24, Phe26, and Leu28 from β2, and Val76 and Ile79 from αB (Figure. 19D). At the ligand position -2, the side chain hydroxyl of Thr-2 in each structure hydrogen bonds to the imidazole ring of His72, with the side chain aliphatic carbon making contact to the structurally conserved residue Val76 (Figure. 19E). It should be noted that all these CXCR2 interacting residues are spared

from the crystal packing in both crystal forms, consistent with their spatially buried natures in the PDZ1-peptide complexes[151].

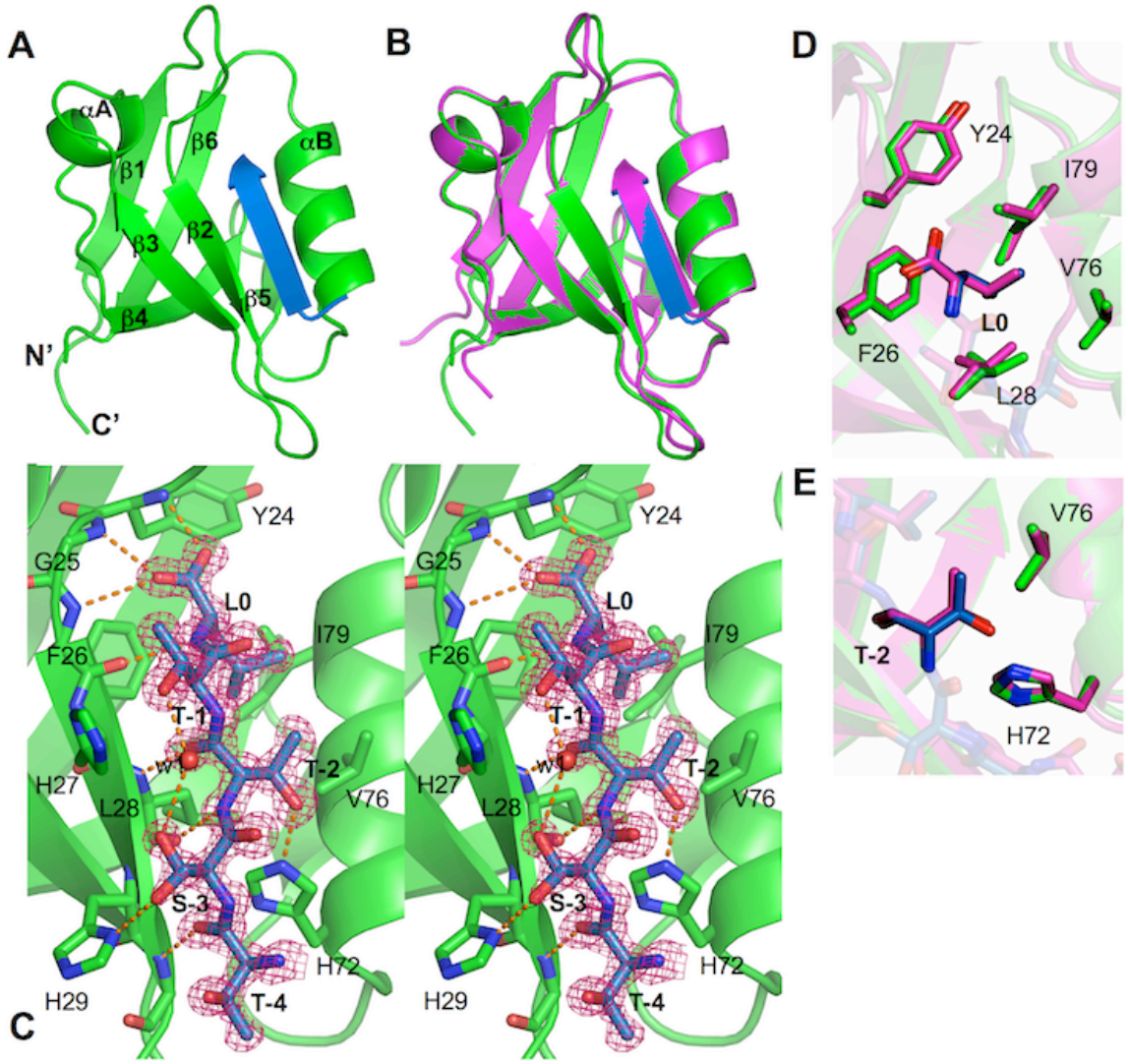

Figure 19. Structural similarities of two crystal forms. (A) Ribbon view of overall $P2_1$-PDZ1 structure. PDZ1 is shown in green and the CXCR2 peptide shown in blue. Secondary structures of PDZ1, α-helices and β-strands, are labeled and numbered according to their position in the sequence. (B) Superposition of $P2_1$-PDZ1 (green) and $P3_121$-PDZ1 (magenta). (C) Stereo view of the PDZ1-CXCR2 interaction in $P2_1$-PDZ1. The PDZ1 residues are represented by sticks with their carbon atoms colored in green. The CXCR2 peptide is depicted by sticks overlaid with $2F_o - F_c$ omit map calculated at 1.1 Å and contoured at 1.5 σ. Hydrogen bonds are illustrated as orange broken lines. (D) Superposition of the Leu0 and (E) Thr-1 recognition regions. Both $P2_1$-PDZ1 and $P3_121$-PDZ1 are depicted by sticks and colored according to the scheme in Figure B and C.

In contrast to the ligand recognition at positions 0 and -2, distinct conformations between the two forms of PDZ1 structures are observed in regions that are responsible for interactions with -1 and -3 residues. Notably, the residues at these two ligand positions are highly variable across natural PDZ1 binding targets, exemplifying its ability to bind multiple targets[151]. Thus, understanding the conformational dynamics that governs the specific interactions with residues -1 and -3 should be key to understanding the underlying mechanisms of PDZ1 promiscuity. Specifically, the PDZ1 residues exhibiting large conformational differences between the two crystal forms include His27 and His29 from β2, Arg40 from β3, and Glu43 from the loop following β3 (Figure. 20A). These residues are known to be important for -1 and -3 residue recognition, three of which (His29, Arg40, and Glu43) have been shown to undergo a large conformational change upon binding to different ligands[4, 151]. Remarkably, their differences in the conformations appear to be well correlated with differential crystal packing lattices, highlighting their adaptability to different environments that may be essential for PDZ1 promiscuity. In particular, the greatest difference between the ligand binding sites of the two PDZ1 forms is at residue Arg40, where the rms deviation of the main chain atoms is 0.02 Å and, for side chain atoms, 0.59 Å (Figure. 20B and 21). This large variation in Arg40 conformations is closely associated with the large differences in its crystal packing environments. In $P3_121$-PDZ1, the side chain atoms of Arg40 make a hydrogen bond to Met10 and van der Waals contacts with the Pro12 side chain of a neighboring molecule. In $P2_1$-PDZ1, only the Nη2 atom of Arg40 is within 4 Å distance to a neighboring Thr71,

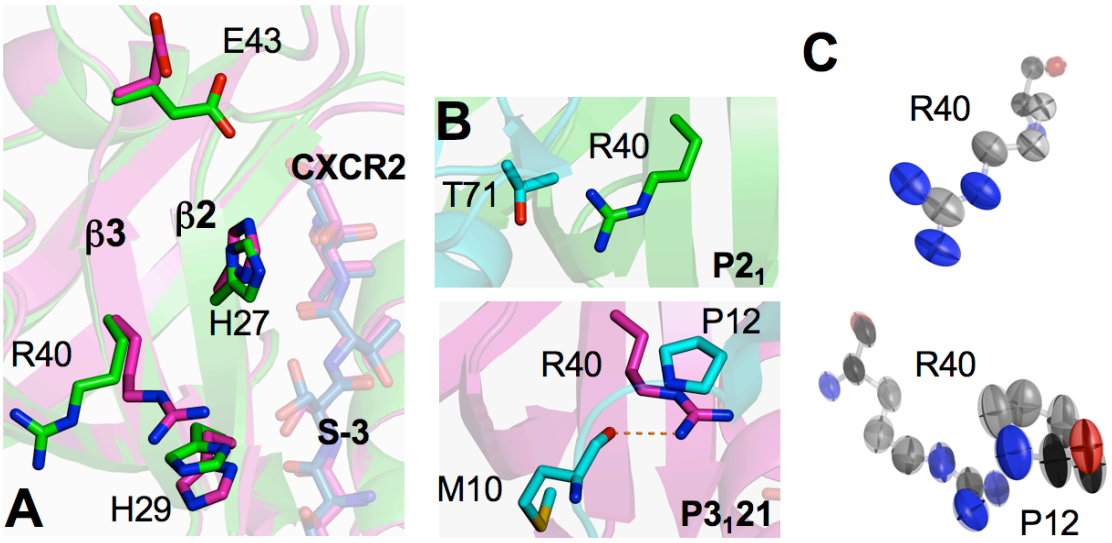

Figure 20**.** Different Arg40 conformations of two crystal forms. (A) Overall view of conformational differences in the peptide-binding pocket. $P2_1$-PDZ1 and $P3_121$-PDZ1 are superimposed and colored according to the scheme in Figure 2. (B) Arg40 crystal contacts in $P2_1$-PDZ1 (top) and in $P3_121$-PDZ1 (bottom). Symmetry-related molecules are represented by ribbons and sticks with their carbon atoms colored in cyan. (C) Thermal ellipsoid representation of Arg40 of $P2_1$-PDZ1 (top) and $P3_121$-PDZ1 (bottom). Carbon atoms are colored gray, nitrogen atoms blue, and oxygen atoms red. Thermal ellipsoids are contoured at the 50% probability level.

no any intermolecular contacts were observed below 3.5 Å. As a result of different crystal packing, Arg40 has very different rotameric conformations in the two crystal structures. In $P3_121$-PDZ1, the side chain of Arg40 is oriented toward the hydroxyl group of Ser-3, whereas in $P2_1$-PDZ1, its guanidinium points away from the bound ligand, adopting a conformation corresponding to ~ 90° rotation around the Cβ-Cγ bond of the side chain (Figure. 20B). In both crystal forms, Arg40 does not make contact with the CXCR2 peptide, whereas in other PDZ1 structures bound with different peptides[1, 4], it is a key anchor residue for specific Asp-3 recognition and engages in direct ligand binding.

Notably, a different rotameric state was assumed in the latter complexes, which allows Arg40 binding to the ligands with the longer side chain at the -3 position[151].

In addition to the conformational change, the intrinsic dynamics of Arg40 is different between the two crystal forms, and is significantly perturbed by the crystal packing. In $P3_121$-PDZ1, the mean anisotropy of Arg40 atoms is 0.337 (σ=0.070), which is considerably more anisotropic than that in $P2_1$-PDZ1 (A=0.489, σ=0.133) (Table 5). The majority of the largest anisotropic differences are located in the main chain region, which appear to correspond to the different crystal packing environments (Figure. 20C). In $P3_121$-PDZ1, the thermal ellipsoids for the main chain atoms of Arg40 are prolate, with the longest principal axis oriented roughly parallel with the side chain direction, indicating that displacements of the Arg40 backbone are least constrained along the side chain and most constrained in directions orthogonal to the side chain. This result contrasts sharply to the more isotropic displacements in $P2_1$-PDZ1, consistent with the extensive crystal packing and the fact that the orientations of the principal axes of the side chain fluctuations correspond closely to those of nearby neighboring atoms. Together, our crystallographic analysis demonstrates that residue Arg40 is intrinsically flexible, capable of exploring large conformational space, or visiting different conformations required for binding multiple partners.

Another large conformational difference occurs at His29, a residue that plays a key role in Ser-3 recognition[151]. This difference is not the direct result of crystal packing,

Table 5. Isotropic B-factor and anisotropy

Residue	**Arg40**		**His27**		**His29**		**Glu43**		**Ser-3:1[c]**		**Ser-3:2**	**Protein (aa 9–94)**		**Peptide (aa 95–99)**	
Space group	$P2_1$	$P3_121$	$P2_1$	$P3_121$	$P2_1$	$P3_121$	$P2_1$	$P3_121$	$P2_1$	$P3_121$	$P2_1$	$P2_1$	$P3_121$	$P2_1$	$P3_121$
Isotropic B-factor (Å²)															
All atoms	9.2	12.9	6.0	13.5	14.4	13.1	11.6	13.8	9.2	12.0	8.4	14.8	17.2	9.7	13.7
	(1.8)[a]	(1.5)	(0.6)	(2.5)	(2.4)	(1.7)	(1.5)	(3.6)	(0.7)	(1.2)	(0.6)				
Main-chain	5.8	9.5	5.1	9.3	11.0	9.9	10.4	10.2	8.5	11.8	8.4	13.7	15.3	8.3	12.1
	(0.0)	(0.1)	(0.0)	(0.0)	(0.2)	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)	(1.0)	(1.3)	(0.7)	(1.3)
Side-chain	11.2	14.9	6.6	16.4	16.7	15.2	12.5	16.7	10.6	12.2	8.6	16.0	19.4	11.3	15.6
	(0.2)	(0.1)	(0.1)	(0.2)	(0.1)	(0.1)	(0.1)	(0.3)	(0.1)	(0.1)	(0.0)	(1.4)	(2.6)	**(1.1)**	(1.6)
Anisotropy[b]															
All atoms	0.489	0.337	0.608	0.330	0.484	0.470	0.319	0.352	0.447	0.358	0.349	0.430	0.424	0.482	0.432
	(0.133)	(0.070)	(0.088)	(0.054)	(0.174)	(0.086)	(0.062)	(0.061)	(0.118)	(0.082)	(0.066)	(0.136)	(0.118)	(0.119)	(0.126)
Main chain	0.585	0.309	0.681	0.357	0.296	0.386	0.276	0.332	0.376	0.303	0.324	0.411	0.411	0.436	0.386
	(0.084)	(0.017)	(0.071)	(0.023)	(0.046)	(0.059)	(0.043)	(0.030)	(0.069)	(0.034)	(0.066)	(0.141)	(0.100)	(0.132)	(0.118)
Side chain	0.435	0.353	0.560	0.312	0.610	0.526	0.354	0.368	0.588	0.467	0.398	0.450	0.439	0.536	0.491
	(0.125)	(0.083)	(0.062)	(0.061)	(0.097)	(0.046)	(0.053)	(0.073)	(0.045)	(0.001)	(0.022)	(0.128)	(0.134)	(0.074)	(0.115)

[a]Numbers in parentheses refer to standard deviation.
[b]The mean anisotropy A is defined as the ratio of the smallest to the largest eigenvalue of U, where U is ADP tensor.
[c]Ser-3:1 conformation 1; Ser-3:2 conformation 2.

as neither crystal form involves His29 in the lattice interface, and the distance from His29 to the nearest neighboring atom is 4.6 Å for $P3_121$-PDZ1 and 6.2 Å for $P2_1$-PDZ1. However, through altering intramolecular interactions, the crystal packing has an indirect impact on His29 conformation. In particular, different His29 structures in different crystal forms are the result of altered local environments with altered Arg40 conformations (Figure. 21 and 22A). In $P3_121$-PDZ1, the imidazole ring of His29 is held down by the guanido group of Arg40 via parallel stacking interactions, whereas in $P2_1$-PDZ1, the packing-induced reorientation of the Arg40 side chain leads to breakage of such contacts, allowing His29 to adopt a more relaxed, ~25º upward-tilted rotamer. In both crystal forms, residue His29 maintains direct hydrogen bonding to the Ser-3 hydroxyl, but this is achieved with a concerted change in the peptide structure. Specifically, the movement of the His29 side chain induces a corresponding movement in the side chain of Ser-3, which preserves the His29/Ser-3 contacts and ligand specific recognition (Figure. 22A). Note the interaction of His29 with the -3 residue is dependent on the types of ligands; when binding to different ligands, the side chain of His29 can adopt very different conformers[151]. For example, in the PDZ1-CFTR complex[1], the side chain of His29 is completely oriented away from the CFTR peptide, adopting a conformation that is unable to interact with the ligand (Figure. 22A). A similar conformer has been observed in PDZ1-β2AR and PDZ1-PDGFR[4], where the -3 residue (Asp-3) of both complexes is common to the PDZ1-CFTR complex. In addition, the intrinsic dynamics of His29 is discernible from the anisotropic displacement parameters (ADPs) of the structures. The mean anisotropy of His29 in the two crystal forms are very similar,

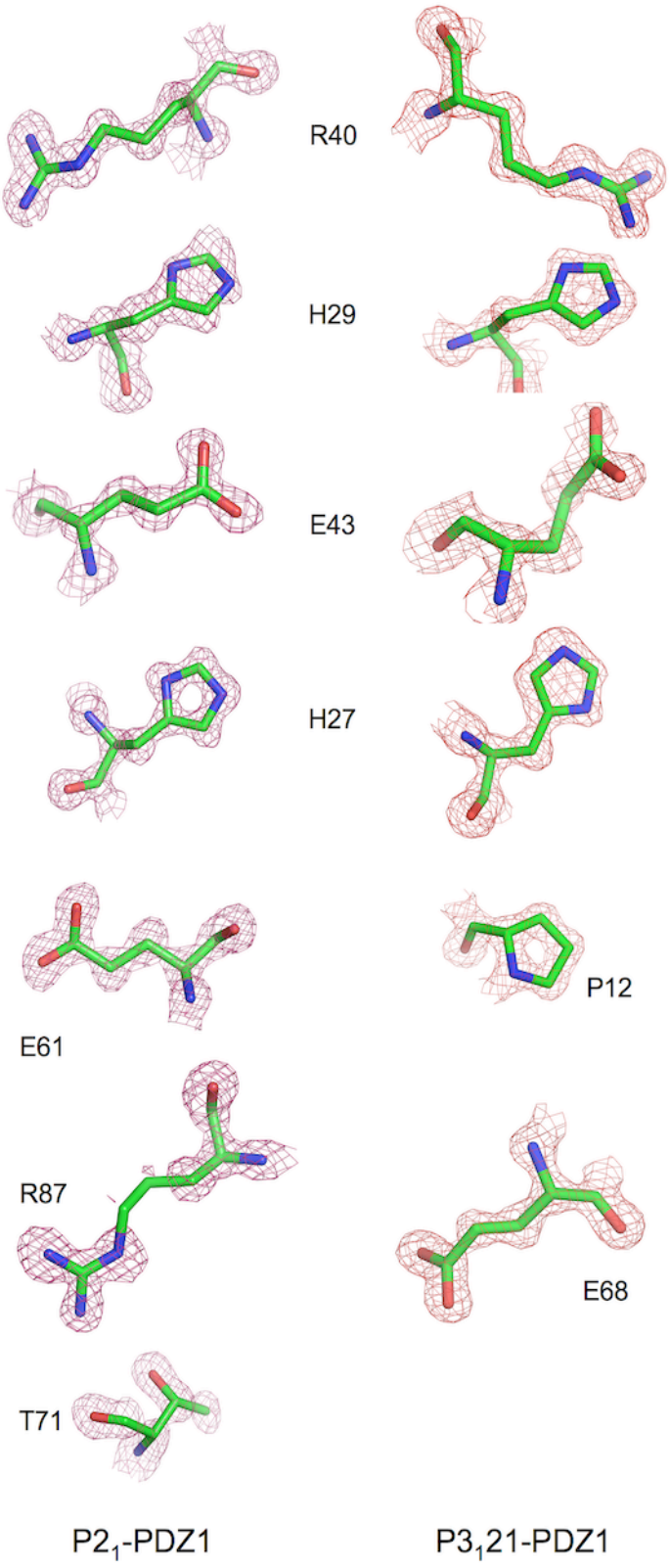

Figure 21. Electron density of selected residues. The left panel, P21-PDZ1; the right panel, P3121-PDZ1. Residues are depicted by sticks overlaid with 2Fo−Fc omit map calculated at 1.1 Å for P21-PDZ1 and 1.16 Å for P3121-PDZ1, and contoured at 1.5 σ.

0.484 (σ=0.174) for $P2_1$-PDZ1 and 0.470 (σ=0.086) for $P3_121$-PDZ1. However, the anisotropy of the main chain atoms and the side chain atoms is inversely different (Table 5). In $P2_1$-PDZ1, His29 exhibits the higher main chain anisotropy, the principle axes of which match excellently with the direction of the displacements deduced from the structural alignment (Figure. 22F). This indicates that His29 has an intrinsic propensity to undergo this movement. The higher side chain anisotropy of His29 in $P3_121$-PDZ1 appears to be related to the Arg40/His29 interaction, as the similar ADP magnitudes and orientations are observed for the Arg40 guanidinium and the His29 imidazole ring. These findings indicate that similar to Arg40, His29 is also intrinsically flexible and contributes to ligand specific binding and recognition.

Intriguingly, the peptide recognition by His29 and Arg40 appears to be mutually exclusive, as they occupy the overlapping space when binding to specific ligands. For instance, in PDZ1-CFTR, the interaction between Arg40 and Asp-3 induced steric effects that prevented His29 from ligand binding[1, 4]. In PDZ1-CXCR2, binding of His29 to Ser-3 caused a "kink" in Arg40's side chain resulting in the effective blockage of the Arg40-CXCR2 interaction[151]. This mutually exclusive peptide recognition may be advantageous, as the combined effects of individual recognition may increase PDZ1 robustness of ligand binding, or its capacity to interact with multiple ligands. The blend of His29 and Arg40 chemical properties, including hydrogen bonding, aromaticity, charge, and their intrinsic flexibility, may allow PDZ1 recognition of different -3 side chains. Interestingly, peptide affinity selection experiments showed that PDZ1 has no apparent amino acid

preference for -3 position of peptides, capable of binding the peptides with -3 side chains of different size and polarity[137]. These suggest PDZ1 promiscuity may be due to multiplicity of possible binding modes that use different structural elements for binding structurally different ligands. They also indicate that the functional interplay between different peptide recognition residues requires a flexible binding pocket and the ability of this pocket to adopt significantly different conformations.

The third notable conformational difference between the two crystal forms is at Glu43, which, together with His29 and Arg40, provides evidence that substantial structural flexibility is present in the PDZ1 peptide-binding pocket. Specifically, in $P3_121$-PDZ1, Glu43 is not engaged in any crystal contacts, having an upward-folded, solvent-pointing side-chain conformation (Figure. 22B and 21). In contrast, in $P2_1$-PDZ1, due to the contact with the Arg87 guanidinium of a neighboring molecule, the side chain of Glu43 adopts a distinct conformation that stretches out towards the bound ligand. This stretched conformation and its Arg-interacting ability, are reminiscent of PDZ1-CFTR interaction. Similar Glu43 conformation observed in the PDZ1-CFTR structure is required for specific binding with the Arg-1 of the ligand[1]. Intriguingly, comparing the structures of PDZ1-CFTR and $P2_1$-PDZ1 reveals that the positions of Arg-1 and Arg87 are completely different from one another, and they show no spatial overlap, approaching Glu43 from opposite directions (Figure. 22B). As a result, significant differences exist in the salt bridge interaction scheme between the two crystal forms. In PDZ1-CFTR, the Oε2 atom of Glu43 makes bifurcated hydrogen bonds with Nε and Nη2 of Arg-1,

whereas in $P2_1$-PDZ1, the carboxylate oxygens of Glu43 are involved in separate hydrogen bonding to Nη1 and Nη2 of Arg87. This difference indicates that the stretched Glu43 conformation adopted in $P2_1$-PDZ1 may be robust in Arg recognition, capable of binding Arg with different orientations. This conclusion is consistent with affinity selection experiments that showed NHERF1 PDZ1 prefers ligands with Arg at the -1 position, and the affinity of the PDZ1-ligand interaction can be reduced by mutation of Arg to Ala, Phe, Leu, or Glu[1, 137]. Given the recognized importance of Glu43 in peptide

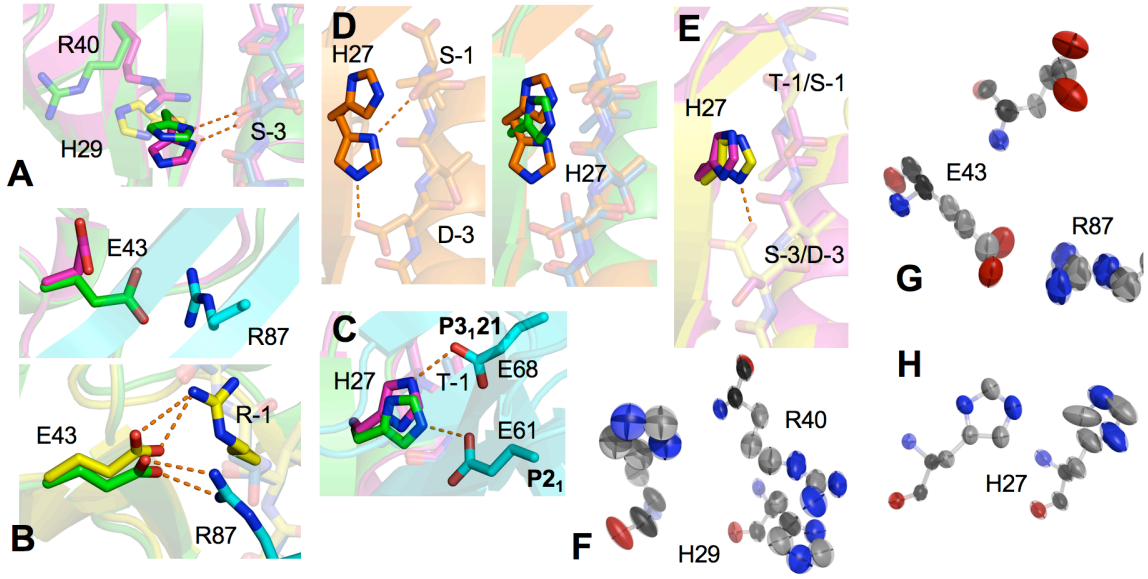

Figure 22**.** Conformational differences of His29, Glu43, and His27 between two crystal forms. (A) Superposition of His29 of $P2_1$-PDZ1 (green), $P3_121$-PDZ1 (magenta), and PDZ1-CFTR (yellow). (B) Comparative view of Glu43 of $P2_1$-PDZ1 and $P3_121$-PDZ1 shown together with $P2_1$-PDZ1 symmetry-related molecules (top); superposition of Glu43 of $P2_1$-PDZ1 and PDZ1-CFTR (bottom). (C) Superposition of His27 of $P2_1$-PDZ1 and $P3_121$-PDZ1. Hydrogen bonds are illustrated as orange broken lines, and residues from the symmetry-related molecules are colored in cyan. (D) Left: dual positioning of His27 of NHERF2 PDZ1 (PDB code: 2OCS) and right: superposition of His27 of $P2_1$-PDZ1 (green) and NHERF2 PDZ1 (orange). (E) Superposition of His27 of $P3_121$-PDZ1 and PDZ1-CFTR. (F), (G), and (H) Thermal ellipsoid representation of His29, Glu43, and His27 of $P2_1$-PDZ1 (top/left) and $P3_121$-PDZ1 (bottom/right).

recognition[1], together with its ability to adopt different conformations upon binding different ligands[151], the crystal packing-induced conformational change in Glu43 provides further evidence for its structural adaptability, consistent with general proposition that PDZ flexibility contributes to PDZ promiscuity. More evidence in favor of this interpretation is provided by the observation that the crystal packing has a significant impact on the Glu43 anisotropic displacement parameters. In both crystal forms Glu43 exhibits strong anisotropy ($P2_1$: A=0.319; $P3_121$: A=0.352), but the nature and orientations of their ADPs are discernibly different. In $P3_121$-PDZ1, the ADP orientations of the Glu43 atoms are not harmonized, whereas in $P2_1$-PDZ1, the principal axes of the uniformly oriented ellipsoids correlate with the direction of the neighboring Arg87 fluctuations, indicating the dynamic adaptation of Glu43 to different crystal packing environments (Figure. 22G).

Finally, the flexible nature of the PDZ1 peptide-binding pocket is evident in the observation that His27, which packs against -1 residue of the ligand, has different conformations in different crystal forms. The differences include a tilt of the side chain by 12° along the ligand and a 180° flip of the imidazole ring around the Cβ-Cγ bond (Figure. 22C and 23). As a result of this reorientation, the imidazole ring is 1.0 Å closer to Ser-3 in $P2_1$-PDZ1 than in $P3_121$-PDZ1, and there is an overall 2.3 Å displacement between its Nε2 atoms. Note that the flip of the His27 imidazole ring does not significantly affect the His27/Thr-1 interaction, as the plane of the imidazole ring in the two crystals is similarly oriented after flipping, and the σ-π stacking interaction between

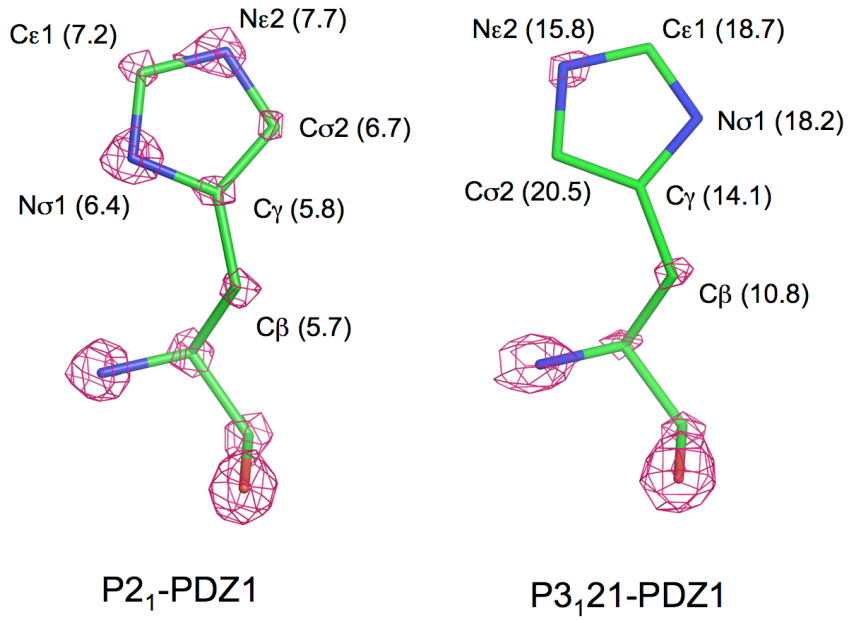

Figure 23. Electron density of His27 at high contour level. The left panel, P21-PDZ1; the right panel, P3121-PDZ1. His27 is depicted by sticks overlaid with 2Fo−Fc omit map calculated at 1.1 Å for P21-PDZ1 and 1.16 Å for P3121-PDZ1. The maps are contoured at 5.0 σ, which reveal the densities at the position of nitrogen atoms are stronger than the densities at the position of carbon atoms (Nε2 vs. Cε1; Nσ1 vs. Cσ2). The B factors of the side chain atoms are shown in parentheses after the atom names.

His27 and the Thr-1 hydroxyl is essentially independent of altered Nε2 positions. Nonetheless, the difference in Nε2 positioning is a manifestation of different crystal packing environments. In both crystal forms, the imidazole ring of His27 is involved in crystal contacts but interacts with different symmetry-related residues. In *P*3$_1$21-PDZ1, the Nε2 atom of His27 makes a hydrogen bond with the side chain Oε1 of Glu68, whereas in *P*2$_1$-PDZ1, it is hydrogen-bonded to the equivalent oxygen from Glu61. As shown in structure alignment, the side chain of Glu61 is similar in orientation to Glu68, but slides more than 5 Å along the peptide binding cleft (Figure. 22C). Remarkably, the direction of this shift corresponds to the direction of the His27 conformational change,

suggesting the intrinsic flexibility of His27 that has the ability to adapt to different environments. This conclusion is supported by the observed anisotropic displacement parameters of His27 that differ dramatically between the two crystal forms ($P2_1$: A=0.608; $P3_121$: A=0.330), and by the observation that the principle axes of the His27 ADPs correspond to the direction of the predicted structural changes (Figure. 22H). The intrinsic flexibility of His27 is also evident from prior findings that the His27 of NHERF2 PDZ1 shows dramatically double conformations; one conformer stacks with -1 residue and the other simultaneously interacts with both -1 and -3 residues of a ligand (Figure. 22D). Because NHERF2 PDZ1 His27 corresponds to NHERF1 PDZ1 His27, this implies that the conserved His27 is capable of exploring a large conformational space for promiscuous binding of various peptide sequences. It is intriguing to note the conformations of His27 are different when NHERF1 PDZ1 binds to different ligands. In PDZ1-CFTR, -β2AR, and -PDGFR, the conformations of His27 are highly superimposable, making a direct hydrogen bond to the common -3 residue (Asp-3) and a ligand-indiscriminative contact with the Cβ atom of the -1 side chain (Figure. 22E). In contrast, in PDZ1-CXCR2 ($P3_121$-PDZ1), the imidazole ring of His27 rotates 20°to accommodate the Thr-1 hydroxyl, and is positioned 0.5 Å further from -3 position of the ligand due to the lack of specific hydrogen binding with the shorter Ser-3 side chain. These differences reflect the relationship between His27 conformations and PDZ1 promiscuity as well as the importance of His27 flexibility in binding different ligands.

Different Modes of CXCR2 Peptide Interaction

One notable difference between the peptides is at the Ser-3 side chain, which adopts a double conformation in $P2_1$-PDZ1, but only one conformation in $P3_121$-PDZ1 (Figure. 24). This difference indicates that similar to PDZ1, the bound ligand also exhibits significant flexibility, capable of assuming different conformations in different environments. Specifically, a 130° rotation around the Cα-Cβ bond relates the two Ser-3 conformations present in the $P2_1$ crystal (Figure. 24). One conformation is similar to the one observed for $P3_121$-PDZ1 (conformation 1), while the other represents a new conformer with the side chain pointing to the opposite direction of the ligand (conformation 2). In $P2_1$-PDZ1, the two Ser-3 conformers are involved in completely different interaction networks resulting in two distinct modes of interaction with PDZ1. For conformation 1, the hydroxyl group of Ser-3 hydrogen bonds to the His29 imidazole ring, whereas in conformation 2, the Ser-3 side chain is stabilized by a van der Waals contact to the His27 Cδ2 atom and a hydrogen bond to a symmetry-related neighboring residue (Glu61) (Figure. 21 and 24A). Intriguingly, conformation 2 also engages in intrapeptide interaction and forms a water-mediated hydrogen bond with the Thr-1 side chain hydroxyl. It appears this dual positioning occurring in $P2_1$-PDZ1 but not in $P3_121$-PDZ1 is due to the dramatic differences in crystal packing between the two crystal forms. In $P2_1$-PDZ1, conformation 1 is not involved in any crystal contacts, whereas in $P3_121$-PDZ1, the tight packing between Ser-3 and Val91 may restrict the Ser-3 conformational flexibility and impedes the possible rotation of its side chain (Figure. 24C). This interpretation is supported by the fact that the axes of Ser-3 and Val91 fluctuations remarkably match with each other in $P3_121$-PDZ1, whereas the lack of the crystal contact

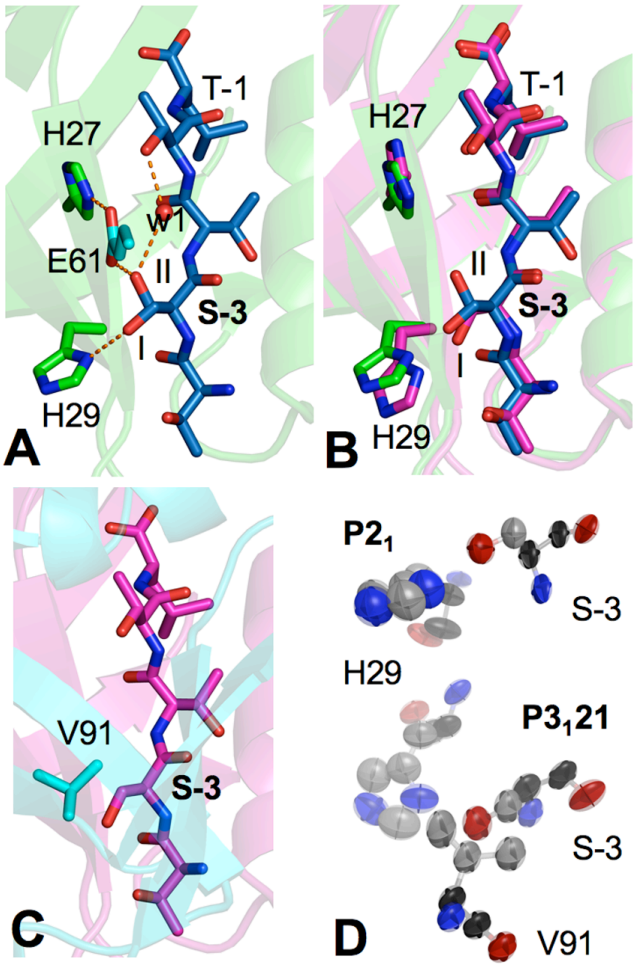

Figure 24. Distinct modes of CXCR2 peptide interaction. (A) Double conformation of Ser-3 in $P2_1$-PDZ1. Residues of the CXCR2 peptide are represented by sticks with their carbon atoms colored in blue. PDZ1 residues are shown in green, while its symmetry-related residue is shown in cyan. Orange broken lines depict hydrogen bonds, and the labels I and II indicate individual conformers of Ser-3 double conformation. (B) Superposition of the CXCR2 peptides of $P2_1$-PDZ1 (blue) and $P3_121$-PDZ1 (magenta). (C) Ser-3 crystal contacts in $P3_121$-PDZ1. (D) Thermal ellipsoid representation of Ser-3 interaction in $P2_1$-PDZ1 (top) and $P3_121$-PDZ1 (bottom).

in $P2_1$-PDZ1 results in apparently coupled motion between Ser-3 and His29, which is not found in $P3_121$-PDZ1 (Figure. 24D). Together, these observations provide some evidence that the intrinsic dynamics of the peptide ligand allows for interactions with different peptide recognition residues. While the functional significance of this dynamical response is unknown, future studies should be directed toward evaluation of its effects on PDZ specificity and promiscuity; especially to determine whether the peptide flexibility is important for single peptides to bind to distinct PDZ domains. It is of particular interest to note the recognition of a peptide-loaded MHC molecule (major histocompatibility complex) by the cognate T-cell receptor depends on the dynamics properties of the

peptides, and differential peptide flexibility resulting from MHC polymorphisms can broaden and expand T-cell receptor reactivity[108, 167].

Discussion

That different PDZ1 conformations are captured by different crystal forms is not surprising in itself, but to the extent it suggests the conformational space available to certain regions of the protein. The set of different crystal structures is thus particularly informative since it may represent different PDZ1 conformational states and reflects the protein's functional dynamics. One can argue that none of the PDZ1 crystal structures correspond exactly to native substrates because of the influence of crystal packing artifacts. Although this concern is likely to be somewhat valid, analysis of thermal factors of nonisomorphous lysozyme structures suggests that the crystal lattice does not just force some random conformational changes onto the molecule, but rather the molecule moves along essential eigenvectors to adapt to different lattice environments[168, 169]. Furthermore, the agreement of the residues undergoing the packing-induced conformational change with the residues involved in allosteric transition in response to ligand binding is in support of use of crystal forms for consolidating PDZ1 structural data, or gaining insights into potential ligand binding mechanisms (Figure. 3). Thus the present $P2_1$ crystal form plus the four original liganded structures in the $P3_121$ space group provide in total five independent views of PDZ1 bound to its targets. Although these structures likely account for only a tiny portion of the entire conformational space, they allow us to at least tentatively begin to sketch the mechanism that describes how the

protein works, and provide the basis for consideration of the PDZ1 structural dynamics and the mechanism by which PDZ1 flexibility contributes to PDZ1 promiscuity. On the other hand, given the exceptional importance of NHERF1 in tumorigenesis and inflammation[119, 134, 152], the knowledge of individual PDZ1 conformational states may be valuable in developing new methods and strategies for selective drug design. For instance, this information can be used to describe binding site flexibility that may allow for accurate modeling of PDZ1-inhibitor interactions. The information also allows for the use of ensemble docking in compound screening, and may contribute to druggable hot-spot identification, and the designing of highly selective compounds[170, 171]. Taken together, the collection of available PDZ1 structures provides insight into the PDZ1 conformational dynamics and the structural explanations of how PDZ1 is able to bind to different ligands. It is no doubt that further understanding of the rules that underlying the ligand-binding site dynamics will benefit from continued studies of PDZ1 liganded structures in different crystal forms.

CHAPTER 6

Crystallographic Analysis of NHERF1-PLCβ3 Interaction Provides Structural Basis for CXCR2 Signaling in Pancreatic Cancer

*Published in Biochemical and Biophysical Research Communication 2014 Apr 4;446(2):638-43. doi: 10.1016/j.bbrc.2014.03.028. All authors agreed with including their work in this dissertation.

Abstract

The formation of CXCR2-NHERF1-PLCβ3 macromolecular complex in pancreatic cancer cells regulates CXCR2 signaling activity and plays an important role in tumor proliferation and invasion. We previously have shown that disruption of the NHERF1-mediated CXCR2-PLCβ3 interaction abolishes the CXCR2 signaling cascade and inhibits pancreatic tumor growth in vitro and in vivo. Here we report the first crystal structure of the NHERF1 PDZ1 domain in complex with the C-terminal PLCβ3 sequence. The structure reveals that the PDZ1-PLCβ3 binding specificity is achieved by numerous hydrogen bonds and hydrophobic contacts with the last four PLCβ3 residues contributing to specific interactions. We also show that PLCβ3 can bind both NHERF1 PDZ1 and PDZ2 in pancreatic cancer cells, consistent with the observation that the peptide binding pockets of these PDZ domains are highly structurally conserved. This study provides an understanding of the structural basis for the PDZ-mediated NHERF1-

PLCβ3 interaction that could prove valuable in selective drug design against CXCR2-related cancers.

Introduction

CXC chemokine receptor 2 (CXCR2) is a G protein-coupled receptor that is activated by binding to the chemokine Gro-α, Gro-β, Gro-γ, ENA-78, GCP-2, IL-8, or NAP-2 [118]. CXCR2 mediates neutrophilic migration and plays critical roles in the positioning of oligodendrocyte precursors in developing spinal cord [114, 115, 119]. This receptor also functions in angiogenesis and wound healing and contributes to both spontaneous and inflammation-driven tumorigenesis [114, 116, 117]. Growing evidence suggests that CXCR2 signaling promotes pancreatic cancer progression where its elevated expression correlates with aggressive stages and poor overall prognosis in patients [172, 173]. More recent studies indicate that CXCR2 is expressed in various pancreatic ductal adenocarcinoma (PDAC) cell lines and enhances cell proliferation and survival via the autocrine or paracrine effect [152, 174, 175]. These findings imply that CXCR2 could be an attractive drug target for developing targeted treatment for pancreatic cancer.

Evidence suggests that CXCR2 interacts directly or indirectly with other receptors, ion channels, transporters, scaffolding proteins, effectors, and cytoskeletal elements to form macromolecular complexes at specialized subcellular domains [119, 152]. These dynamic protein-protein interactions regulate CXCR2 signaling function as well as its localization and processing within cells [176, 177]. We have shown that CXCR2,

phospholipase C-β3 (PLCβ3), and Na^+/H^+ exchanger regulatory factor-1 (NHERF1) form macromolecular complexes at the plasma membrane of pancreatic cancer cells, which functionally couple chemokine signaling to PLCβ3-mediated signaling cascade [152]. PLCβ3, a membrane bound enzyme, catalyzes the formation of inositol 1,4,5-trisphosphate and diacylglycerol from phosphatidylinositol 4,5-bisphosphate. This reaction uses calcium as a cofactor and plays an important role in the intracellular transduction of many extracellular signals [178]. NHERF1 is a PDZ domain-containing protein that typically functions as a scaffold to cluster transporters, receptors, and signaling molecules into supramolecular complexes [150]. We have demonstrated that the formation of the CXCR2-NHERF1-PLCβ3 complex is mediated by NHERF1 PDZ domains, which bridge CXCR2 and PLCβ3 through binding to their C-terminal PDZ-binding motif [152]. We also showed that disruption of this PDZ-mediated interaction abolishes CXC signaling and inhibits tumor growth in PNAC-1 cells and also in human PDAC xenograft animal model [152]. These findings imply that targeting the PDZ-mediated CXCR2-PLCβ3 interaction could provide new strategies for therapeutic interventions of CXCR2-related cancers.

In general, PDZ domains mediate protein interactions by recognizing the C-terminal sequence of target proteins and binding to the targets through a canonically and structurally conserved PDZ peptide-binding pocket [121]. The specificity of the interactions 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 [121]. This has led to the classification of PDZ domains into two major specificity classes: class I, (S/T)X(V/I/L) (X denoting any amino acid); class II, (F/Y)X(F/V/A) [1, 123, 124]. However, more recent evidence suggests that PDZ specificity is unexpectedly complex, with the PDZ domain family recognizing up to 7 C-terminal ligand residues and forming at least 16 unique specificity classes [122]. In addition, many PDZ domains can bind to multiple ligands of different peptide classes, and single peptides are capable of binding to distinct PDZ domains [122]. This complex picture raises a challenging problem of how PDZ domains, structurally simple protein-interaction modules, achieve the broad substrate specificity, the nature of which still remains obscure. In this context, we present the crystal structure of NHERF1 PDZ1 in complex with the PLCβ3 C-terminal peptide ENTQL. The structure reveals that the PLCβ3 peptide binds to PDZ1 in an extended conformation with the last four residues making specific side chain contacts. We also show that PLCβ3 can bind both NHERF1 PDZ1 and PDZ2 in PDAC tumor cells, consistent with the observation that the two domains share highly structurally conserved peptide-binding pockets. This study provides the structural basis of the PDZ-mediated NHERF1-PLCβ3 interaction and could be valuable in the development of novel therapeutic strategies against aggressive pancreatic cancers.

Materials and Methods

Protein Expression and Purification

For X-ray crystallography, a DNA fragment encoding the human NHERF1 PDZ1 (residues 11–94) was amplified by PCR using the full-length human NHERF1 cDNA as a

template. The C-terminal extension ENTQL that corresponds to residues 1230–1234 of human PLCβ3 was created by inclusion of 15 extra bases in the reverse primer. The PCR products were cloned in the pSUMO vector containing an N-terminal His6-SUMO tag. The resulting clone was transformed into *Escherichia coli* BL21 Condon Plus (DE3) cells for protein expression. The transformants were grown to an OD600 (optical density at 600 nm) 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 by centrifugation and lysed by French Press. The soluble fraction was then subjected to Ni^{2+} affinity chromatography purification, followed by the cleavage of the His6-SUMO tag with yeast SUMO Protease 1. PDZ1 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 30–40 mg/ml in a buffer containing 20 mM Tris–HCl (pH 8.0), 150 mM NaCl, 1 mM β-mercaptoethanol, and 5% glycerol. For pulldown experiments, glutathione S-transferase (GST) fusion proteins were generated by cloning NHERF1 PDZ1, PDZ2, or PDZ1-PDZ2 into the pGEX4T-1 vector [119]. His-S-tagged proteins were generated by cloning PLCβ3 C-terminal fragment (residues 1135–1234) into the pET30 vector [152]. GST-PDZ proteins were purified using glutathione agarose beads (BD Biosciences) and eluted with 50 mM glutathione. His-S-PLCβ3 was purified using Cobalt resins (Thermo Scientific) and eluted with 200 mM imidazole.

Crystallization, Data Collection and Structure Determination

Crystals were grown by the hanging-drop vapor-diffusion method by mixing the

protein (~25 mg/ml) with an equal volume of a reservoir solution containing 100 mM sodium acetate, pH 4.6, 2.5 M sodium chloride at 20 °C. Crystals typically appeared overnight and continued to grow to their full size in 3–4 days. Prior to X-ray diffraction data collection, crystals were cryoprotected in a solution containing the mother liquor and 25% glycerol and flash cooled in liquid nitrogen. The data were collected at 100 K at beamline 21-ID-F at the Advanced Photon Source (Argonne, IL) and processed and scaled using the program XDS [95]. Crystals belong to the space group $P3_121$ with unit cell dimensions a = b = 50.7 Å, c = 66.7 Å, and one molecule in the asymmetric unit (Table 6). The structure was solved by the molecular replacement method with the program PHASER [96] using the PDZ1-CXCR2 structure (PDB code: 4JL7) as a search model. Structure modeling was carried out in COOT [69], and refinement was performed with PHENIX [97]. To reduce the effects of model bias, iterative-build OMIT maps were used during model building and structure refinement. The final models were analyzed and validated with Molprobity [99]. All Figureures of 3D representations of the PDZ1-PLCβ3 structure were made with PyMOL (www.pymol.org).

Cell Culture

Human PDAC cell lines (PANC-1, AsPC-1, and BxPC-3) were obtained from American Type Culture Collection (Manassas, VA). Cells were cultured in Dulbecco's modified Eagle's medium (Thermo Scientific Hyclone) containing 4.5 mg/ml D-glucose and L-glutamine supplemented with 10% FBS, 100 units/ml penicillin, and 100 μg/ml streptomycin at 37°C in humidified air with 5% CO_2.

Pulldown Assays

GST pulldown assays were preformed as previously described [119]. Briefly, PDAC cells were lysed in a binding buffer containing phosphate-buffered saline (PBS), 0.2% Triton X-100, and a mixture of protease inhibitors. Supernatant was equally mixed with GST alone or various GST-PDZ fusion domains (GST-PDZ1, GST-PDZ2, or GST-PDZ1-PDZ2) at 4 °C for 2 hours. The mixture was pulled down by glutathione agarose beads at 4°C overnight, washed three times with the binding buffer, and then eluted in Laemmli sample buffer containing β-mercaptoethanol. The eluents were separated by SDS-PAGE and immunoblotted with anti-PLCβ3 antibody. To verify the direct NHERF1-PLCβ3 interaction, purified GST-NHERF1 PDZ domains or GST alone were mixed with a purified His-S-PLCβ3 C-terminal fragment (last 100 residues) in the binding buffer at 20 ºC for 1 hour. The mixtures were incubated with S-protein agarose beads for 2 hours. The beads were washed three times with the binding buffer and eluted with Laemmli sample buffer containing β-mercaptoethanol. The eluents were resolved by SDS-PAGE and immunoblotted with anti-His antibody. All experiments were repeated at least three times, and the results were consistent.

Protein Data Bank Accession Number

Coordinates and structure factors have been deposited in the Protein Data Bank with accession number 4PQW.

Table 6. Crystallographic data and refinement statistics

Data	
Space group	$P3_121$
Cell parameters (Å)	
a, b	50.7
c	66.7
Wavelength (Å)	1.1272
Resolution (Å)	33.3-1.47 (1.51-1.47)
R_{merge}[a]	0.041 (0.575)[b]
Redundancy	6.9 (6.5)
Unique reflections	17247
Completeness (%)	99.5 (100)
$\langle I/\sigma \rangle$	21.2 (2.7)
Refinement	
Resolution (Å)	26.6 - 1.47 (1.52- 1.47)
Molecules/AU	1
R_{work}[c]	0.181 (0.276)
R_{free}[d]	0.196 (0.325)
Ramachandran plot	
Residues in favored	98.9%
Residues in allowed	1.1%
RMSD	
Bond lengths (Å)	0.005
Bond angels (°)	0.95
No. of atoms	
Protein	1334
Peptide	80
Water	93
Chloride	4
Nickel	1
B-factor (Å^2)	
Protein	37.5
Peptide	41.2
Water	42.7
Chloride	37.7
Nickel	41.0

[a]$R_{merge} = \Sigma |I-\langle I \rangle| / \Sigma I$, where I is the observed intensity and $\langle I \rangle$ is the averaged intensity of multiple observations of symmetry-related reflections.
[b]Numbers in parentheses refer to the highest resolution shell.
[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.

Results and Discussion

Binding specificity of NHERF1-PLCβ3 Interaction

The overall structure of NHERF1 PDZ1 is similar to other PDZ domains [1, 140], consisting of six β strands (β1–β6) and two α-helices (αA and αB) (Figure. 25A and B). The PLCβ3 peptide binds in the cleft between β2 and αB, burying a total solvent-accessible surface area of 382 $Å^2$. The binding specificity of the PDZ1-PLCβ3 interaction is achieved through networks of hydrogen bonds and hydrophobic interactions (Figure. 25C). At the ligand position 0, the side chain of Leu0 is nestled in a deep hydrophobic pocket formed by invariant residues Tyr24, Phe26, and Leu28 from β2 and Val76 and Ile79 from αB (Figure. 25D). In the pocket, the position of Leu0 is further secured by both a hydrogen bond from its amide nitrogen to the Phe26 carbonyl oxygen and triplet hydrogen bonding between the Leu0 carboxylate and the amides of Tyr24, Gly25, and Phe26. Similar interactions have been observed in several other PDZ-mediated complexes [1, 140], which represent the most-conserved binding mode for terminal Leu recognition.

Residues at other peptide positions also contribute to the PDZ1-PLCβ3 complex formation (Figure. 25C). At the ligand position -1, the aliphatic portion of the Gln-1 side chain makes Van der Waals interaction with the imidazole ring of His27. At position -2, Thr-2 makes one hydrogen bond to the His72 imidazole group and two hydrogen bonds to the highly conserved residue Leu28. At the ligand position -3, the interactions with Asn-3 include a direct hydrogen bond from its side chain oxygen to the Nη1 atom of

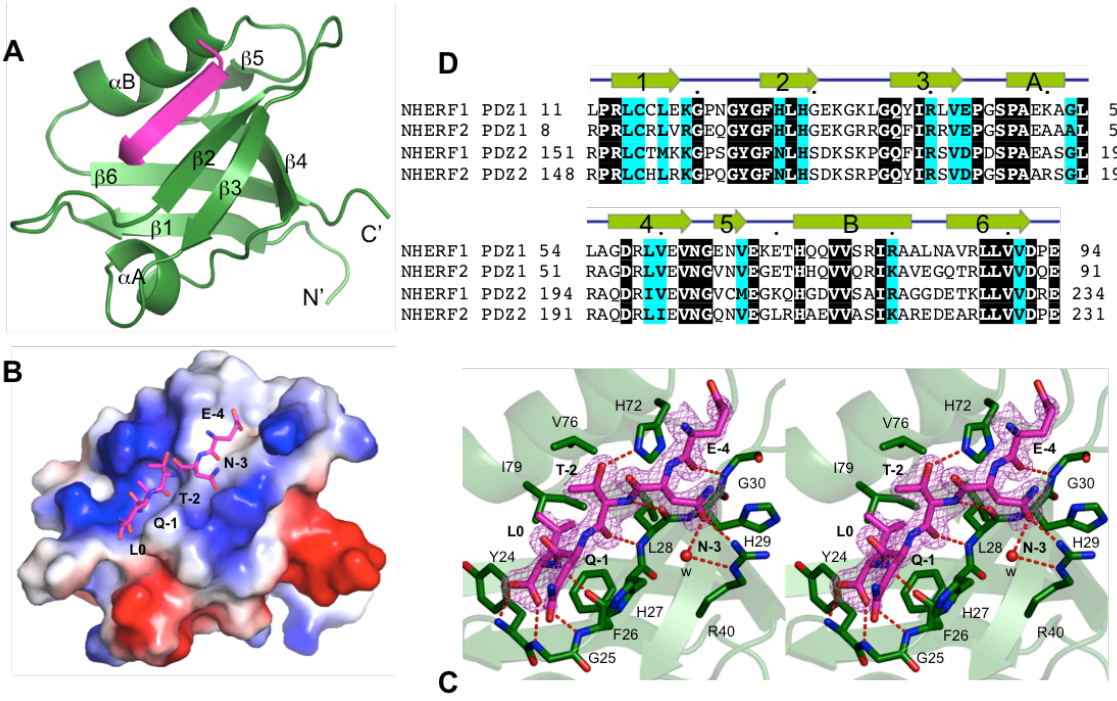

Figure 25. Structure of NHERF1 PDZ1 in complex with the PLCβ3 C-terminal sequence ENTQL. (A) Ribbon diagram of the PDZ1-PLCβ3 structure. PDZ1 is shown in green and the PLCβ3 peptide is shown in magenta. Secondary structures of PDZ1, α-helices, and β-strands are labeled and numbered according to their position in the sequence. (B) Surface representation of the PDZ1-PLCβ3 structure. Surface 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 PLCβ3 peptide is depicted by sticks. (C) Stereo view of the PDZ1 ligand-binding site bound to the PLCβ3 C-terminal peptide. PDZ1 residues are represented by sticks with their carbon atoms colored in green. The PLCβ3 peptide is depicted by sticks overlaid with 2Fo − Fc omit map calculated at 1.47 Å and contoured at 1.8 σ. Hydrogen bonds are illustrated as red broken lines. (D) Sequence alignment of selected PDZ domains. The alignment was performed by ClustalW [3], including human NHERF1 and NHERF2. Identical residues are shown as white on black, and similar residues appear shaded in cyan. Secondary structure elements are displayed above the sequences and labeled according to the scheme in Figure A. Sequence numbering is displayed to the left of the sequences, with every 10th residue marked by a dot shown above the alignment.

Arg40 and a water-mediated hydrogen bond to the Nε atom of Arg40. The latter two interactions represent ligand specific interactions, as the small side chain of Ser-3 is recognized by His29 in PDZ1-CXCR2 complex [179, 180]. Finally, the peptide residue Glu-4 engages in a main chain contact with Gly30, but does not participate in any specific side chain interactions. These observations indicate that the last four residues of PLCβ3 contribute to the binding specificity in the PDZ1-PLCβ3 complex formation.

Endogenous PLCβ3 Interacts with Both NHERF1 PDZ1 and PDZ2

To gain further insight into the NHERF1-PLCβ3 interaction, we performed GST pulldown assays to examine whether NHERF1 PDZ domains interact with endogenous PLCβ3 in PDAC cells. Lysates of various PDAC cells, PANC-1, AsPC-1, and BxPC-3 were used to interact with the GST fusion proteins GST-PDZ1, GST-PDZ2, and GST-PDZ1-PDZ2. As shown in Figure. 26A, no PLCβ3 was detected in the control lane containing GST alone, but significant amounts of PLCβ3 were found in the lanes containing PDZ1, PDZ2, and both PDZ domains together. Similar results were observed for all PDAC cells tested in our experiments. To check whether the PDZ-PLCβ3 interaction is direct, we performed GST pulldown assays with a purified peptide corresponding to the last 100 amino acids of PLCβ3. We observed similar binding results where PLCβ3 interacts with both PDZ domains of NHERF1 (Figure. 26B).

To understand the structural basis of the bivalent binding, we performed a

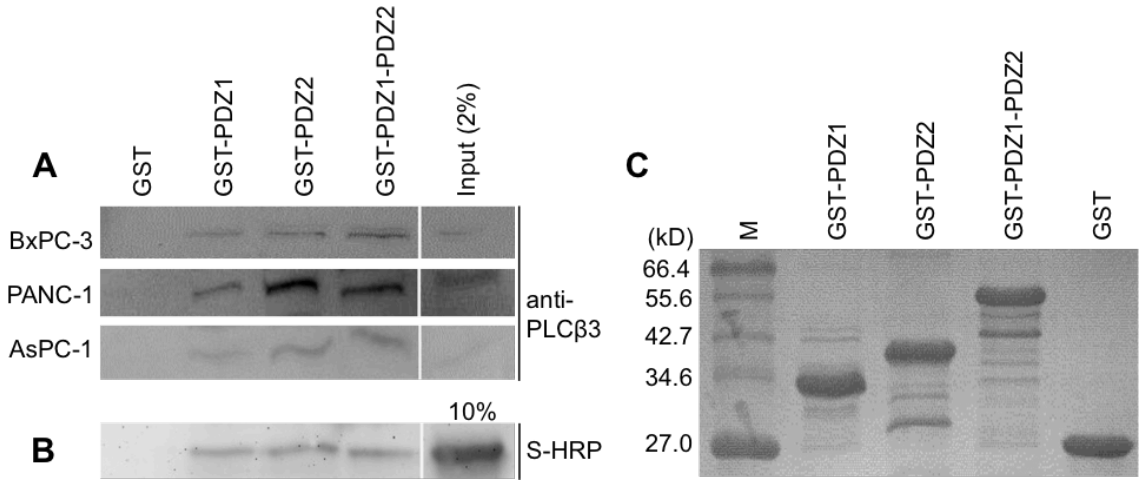

Figure 26. Endogenous PLCβ3 in human pancreatic cancer cells interacts with both NHERF1 PDZ1 and PDZ2. (A) GST pull-down of endogenous PLCβ3 with NHERF1 PDZ domains. Lysates of PDAC cells, BxPC-3, PANC-1, or AsPC-1 were used as prey. GST fusion proteins of NHERF1 PDZ1, PDZ2, or PDZ1-PDZ2 were used as bait. GST alone served as a negative control. Binding experiments were analyzed by SDS-PAGE and visualized by immunoblotting using anti-PLCβ3 antibody. (B) GST pull-down assays to detect direct interaction between purified PLCβ3 and NHERF1. A His-S-tagged peptide corresponding to the last 100 residues of PLCβ3 was used as prey. Purified GST-PDZ1, GST-PDZ2, GST-PDZ1-PDZ2, or GST alone was used as bait. Binding was resolved by SDS-PAGE and immunoblotted with anti-S antibody. (C) SDS-PAGE analysis of beads-immobilized GST proteins in each above reaction (loading control). Lane M is molecular weight markers. Molecular weights are indicated at the left of the gel. The gel is visualized by Coomassie blue staining.

structural alignment between the structure of NHERF1 PDZ1 and the structure of NHERF1 PDZ2. The alignment reveals that NHERF1 PDZ1 and PDZ2 share highly similar overall structures and highly conserved ligand-binding pockets (Figure. 27). The root mean square (rms) difference is 1.35 Å for the overall structure (86 Cα atoms), and for the ligand-interacting residues, 0.44 Å. The only notable difference in the ligand-binding sites is residue 27, which is His in PDZ1 and Asn (residue 164) in PDZ2. It should be noted that this conserved substitution maintains the amino functionality of the

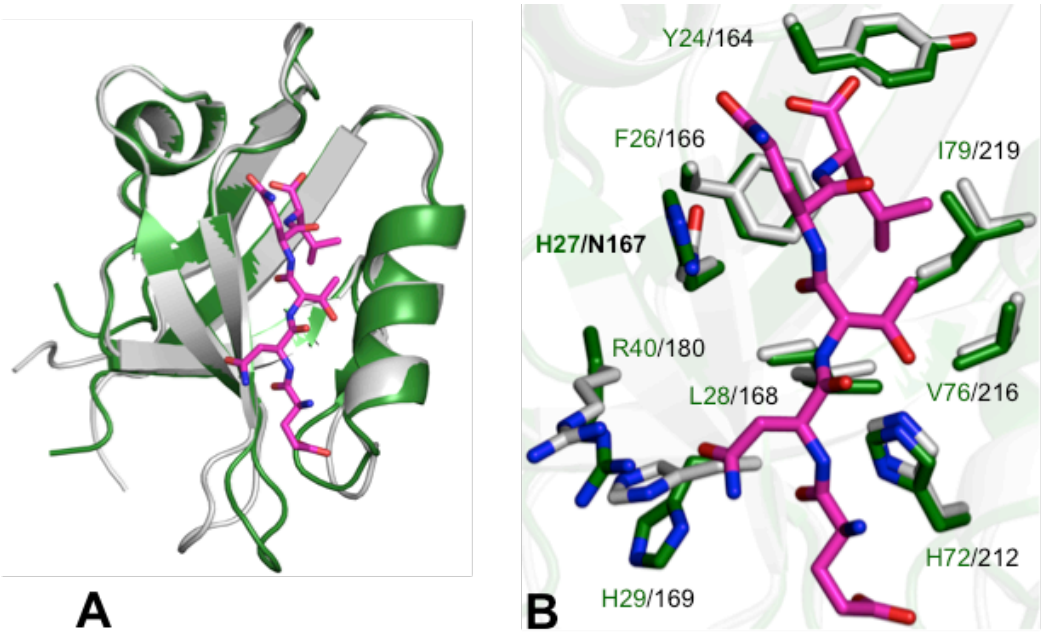

Figure 27. Structural comparison of NHERF1 PDZ1 and PDZ2. (A) Superposition of the structures of PDZ1-PLCβ3 (green; PDB code: 4PQW) and PDZ2 (gray; PDB code: 2OZF). PDZ domains are represented by ribbons. Residues in PLCβ3 are displayed as sticks with the carbon atoms shown in magenta. (B) Superposition of the PDZ ligand binding pockets. Both PDZ and ligand residues are depicted by sticks and colored according to the scheme in Figure A.

side chain, which is not expected to disrupt the observed Van der Waals contact between PDZ1 and PLCβ3 (Figure. 25C). Therefore, the comparison of PDZ1 and PDZ2 provides a structural explanation for the ability of PLCβ3 to bind to both PDZ domains.

Implication in Selective Drug Design

We previously have suggested that targeting the NHERF1-mediated CXCR2-PLCβ3 interaction may have a therapeutic potential in PDAC treatment, as inhibition of this interaction has been found to be sufficient to inhibit CXCR2 signaling activity both in vitro and in vivo [152]. These findings highlight the significance of our present structure

studies, and imply that structural details of the NHERF1-PLCβ3 interaction may be valuable in developing new methods and strategies for selective drug design. For instance, this information can be used to create new NHERF1 inhibitors that are potent and specific to block the NHERF1-PLCβ3 interaction. Such inhibitors have the potential to inhibit pancreatic tumor growth by suppressing CXCR2 signaling, and preventing tumor cell proliferation and invasion. In addition, the ability of PLCβ3 to bind both PDZ domains (Figure. 26), together with similar PDZ structures (Figure. 27), suggests NHERF1 inhibitors may be capable of targeting PDZ1 and PDZ2 simultaneously. Such inhibitors might be advantageous in cancer treatment, as PDZ1 and PDZ2 have been shown to have differential roles during metastasis. NHERF1 PDZ2 promotes visceral metastasis via invadopodia-dependent invasion and anchorage-independent growth, as well as by inhibition of apoptosis; while PDZ1 promotes bone metastasis by stimulating podosome nucleation, motility, angiogenesis, and osteoclastogenesis in the absence of increased growth or invasion [134]. It is conceivable that simultaneous targeting of the PDZ domains could lead to a combinatorially synergetic effect that would prevent metastatic behavior and inhibits mesenchymal-to-vasculogenic phenotypic transition in cancer patients. While the biological impact of the bivalent NHERF1-PLCβ3 interaction is currently unknown, future studies should be directed toward evaluation of its effect on CXCR2-mediated PDAC proliferation and invasion, and especially toward determining whether different PDZ domains could mediate the assembly of distinct CXCR2 signal transduction complexes. Such studies should have important implications in specific NHERF1 scaffolding regulation, and in many CXCR2-associated human cancers.

REFERENCES

[1] Karthikeyan, S., Leung, T., and Ladias, J. A. (2001) Structural basis of the Na+/H+ exchanger regulatory factor PDZ1 interaction with the carboxyl-terminal region of the cystic fibrosis transmembrane conductance regulator, *The Journal of biological chemistry 276*, 19683-19686.

[2] Jiang, Y., Sirinupong, N., Brunzelle, J., and Yang, Z. (2011) Crystal structures of histone and p53 methyltransferase SmyD2 reveal a conformational flexibility of the autoinhibitory C-terminal domain, *Plos One 6*, e21640.

[3] Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics 23*, 2947-2948.

[4] Karthikeyan, S., Leung, T., and Ladias, J. A. (2002) Structural determinants of the Na+/H+ exchanger regulatory factor interaction with the beta 2 adrenergic and platelet-derived growth factor receptors, *The Journal of biological chemistry 277*, 18973-18978.

[5] Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007) Epigenetics: a landscape takes shape, *Cell 128*, 635-638.

[6] Arzate-Mejía, R. G., Valle-García, D., and Recillas-Targa, F. (2011) Signaling epigenetics: Novel insights on cell signaling and epigenetic regulation, *IUBMB Life 63*, 881-895.

[7] Wang, J., Yu, J. T., Tan, M. S., Jiang, T., and Tan, L. (2013) Epigenetic mechanisms in Alzheimer's disease: Implications for pathogenesis and therapy, *Ageing research reviews 12*, 1024-1041.

[8] Vallaster, M., Vallaster, C. D., and Wu, S. M. (2012) Epigenetic mechanisms in cardiac development and disease, *Acta Bioch Bioph Sin 44*, 92-102.

[9] Wegner, M., Neddermann, D., Piorunska-Stolzmann, M., and Jagodzinski, P. P. (2014) Role of epigenetic mechanisms in the development of chronic complications of diabetes, *Diabetes research and clinical practice*.

[10] Chuikov, S., Kurash, J. K., Wilson, J. R., Xiao, B., Justin, N., Ivanov, G. S., McKinney, K., Tempst, P., Prives, C., Gamblin, S. J., Barlev, N. A., and Reinberg, D. (2004) Regulation of p53 activity through lysine methylation, *Nature 432*, 353-360.

[11] Zhang, X., Tanaka, K., Yan, J., Li, J., Peng, D., Jiang, Y., Yang, Z., Barton, M. C., Wen, H., and Shi, X. (2013) Regulation of estrogen receptor alpha by histone methyltransferase SMYD2-mediated protein methylation, *Proc Natl Acad Sci U S A 110*, 17284-17289.

[12] Strahl, B. D., and Allis, C. D. (2000) The language of covalent histone modifications, *Nature 403*, 41-45.

[13] Berger, S. L. (2007) The complex language of chromatin regulation during transcription, *Nature 447*, 407-412.

[14] Kouzarides, T. (2007) Chromatin Modifications and Their Function, *Cell 128*, 693-705.

[15] Rivenbark, A. G., and Strahl, B. D. (2007) MOLECULAR BIOLOGY: Unlocking Cell Fate, *Science 318*, 403-404.

[16] Berger, S. L. (2007) The complex language of chromatin regulation during transcription, *Nature 447*, 407-412.

[17] Brown, M. A., Sims, R. J., Gottlieb, P. D., and Tucker, P. W. (2006) 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 5*.

[18] Gottlieb, P. D., Pierce, S. A., Sims, R. J., Yamagishi, H., Weihe, E. K., Harriss, J. V., Maika, S. D., Kuziel, W. A., King, H. L., Olson, E. N., Nakagawa, O., and Srivastava, D. (2002) Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis, *Nat Genet 31*, 25-32.

[19] Abu-Farha, M., Lambert, J. P., Al-Madhoun, A. S., Elisma, F., Skerjanc, I. S., and Figeys, D. (2008) The tale of two domains - Proteomics and genomics analysis of SMYD2, a new histone methyltransferase, *Mol Cell Proteomics 7*, 560-572.

[20] Tan, X. G., Rotllant, J., Li, H. Q., DeDeyne, P., and Du, S. J. (2006) SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos (vol 103, pg 2713, 2006), *P Natl Acad Sci USA 103*, 7935-7935.

[21] Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F. P., Li, M. H., Yagyu, R., and Nakamura, Y. (2004) SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells, *Nat Cell Biol 6*, 731-740.

[22] Liu, Y., Chen, W., Gaudet, J., Cheney, M. D., Roudaia, L., Cierpicki, T., Klet, R. C., Hartman, K., Laue, T. M., Speck, N. A., and Bushweller, J. H. (2007) Structural basis for recognition of SMRT/N-CoR by the MYND domain and its contribution to AML1/ETO's activity, *Cancer Cell 11*, 483-497.

[23] Tan, X., Rotllant, J., Li, H., De Deyne, P., and Du, S. J. (2006) SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos, *Proc Natl Acad Sci U S A 103*, 2713-2718.

[24] Brown, M. A., Sims, R. J., 3rd, Gottlieb, P. D., and Tucker, P. W. (2006) 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 5*, 26.

[25] Gottlieb, P. D., Pierce, S. A., Sims, R. J., Yamagishi, H., Weihe, E. K., Harriss, J. V., Maika, S. D., Kuziel, W. A., King, H. L., Olson, E. N., Nakagawa, O., and Srivastava, D. (2002) Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis, *Nature genetics 31*, 25-32.

[26] Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F. P., Li, M., Yagyu, R., and Nakamura, Y. (2004) SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells, *Nat Cell Biol 6*, 731-740.

[27] Just, S., Meder, B., Berger, I. M., Etard, C., Trano, N., Patzel, E., Hassel, D., Marquart, S., Dahme, T., Vogel, B., Fishman, M. C., Katus, H. A., Strahle, U., and Rottbauer, W. (2011) The myosin-interacting protein SMYD1 is essential for sarcomere organization, *Journal of cell science 124*, 3127-3136.

[28] Li, D. L., Niu, Z. Y., Yu, W. S., Qian, Y., Wang, Q., Li, Q., Yi, Z. F., Luo, J., Wu, X. S., Wang, Y. Q., Schwartz, R. J., and Liu, M. Y. (2009) SMYD1, the myogenic activator, is a direct target of serum response factor and myogenin, *Nucleic Acids Res 37*, 7059-7071.

[29] Kawamura, S., Yoshigai, E., Kuhara, S., and Tashiro, K. (2008) smyd1 and smyd2 are expressed in muscle tissue in Xenopus laevis, *Cytotechnology 57*, 161-168.

[30] Diehl, F., Brown, M. A., van Amerongen, M. J., Novoyatleva, T., Wietelmann, A., Harriss, J., Ferrazzi, F., Bottger, T., Harvey, R. P., Tucker, P. W., and Engel, F. B. (2010) Cardiac Deletion of Smyd2 Is Dispensable for Mouse Heart Development, *Plos One 5*.

[31] Thompson, E. C., and Travers, A. A. (2008) A Drosophila Smyd4 Homologue Is a Muscle-Specific Transcriptional Modulator Involved in Development, *Plos One 3*.

[32] Donlin, L. T., Andresen, C., Just, S., Rudensky, E., Pappas, C. T., Kruger, M., Jacobs, E. Y., Unger, A., Zieseniss, A., Dobenecker, M. W., Voelkel, T., Chait, B. T., Gregorio, C. C., Rottbauer, W., Tarakhovsky, A., and Linke, W. A. (2012) Smyd2 controls cytoplasmic lysine methylation of Hsp90 and myofilament organization, *Genes Dev 26*, 114-119.

[33] Voelkel, T., Andresen, C., Unger, A., Just, S., Rottbauer, W., and Linke, W. A. (2013) Lysine methyltransferase Smyd2 regulates Hsp90-mediated protection of the sarcomeric titin springs and cardiac function, *Biochimica et biophysica acta 1833*, 812-822.

[34] Hu, L., Zhu, Y. T., Qi, C., and Zhu, Y. J. (2009) Identification of Smyd4 as a potential tumor suppressor gene involved in breast cancer development, *Cancer research 69*, 4067-4072.

[35] Cho, H. S., Hayami, S., Toyokawa, G., Maejima, K., Yamane, Y., Suzuki, T., Dohmae, N., Kogure, M., Kang, D., Neal, D. E., Ponder, B. A., Yamaue, H., Nakamura, Y., and Hamamoto, R. (2012) RB1 methylation by SMYD2 enhances cell cycle progression through an increase of RB1 phosphorylation, *Neoplasia (New York, N.Y.) 14*, 476-486.

[36] Komatsu, S., Imoto, I., Tsuda, H., Kozaki, K. I., Muramatsu, T., Shimada, Y., Aiko, S., Yoshizumi, Y., Ichikawa, D., Otsuji, E., and Inazawa, J. (2009) Overexpression of SMYD2 relates to tumor cell proliferation and malignant outcome of esophageal squamous cell carcinoma, *Carcinogenesis 30*, 1139-1146.

[37] Zuber, J., Rappaport, A. R., Luo, W., Wang, E., Chen, C., Vaseva, A. V., Shi, J., Weissmueller, S., Fellmann, C., Taylor, M. J., Weissenboeck, M., Graeber, T. G., Kogan, S. C., Vakoc, C. R., and Lowe, S. W. (2011) An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance, *Genes Dev 25*, 1628-1640.

[38] Huang, J., Perez-Burgos, L., Placek, B. J., Sengupta, R., Richter, M., Dorsey, J. A., Kubicek, S., Opravil, S., Jenuwein, T., and Berger, S. L. (2006) Repression of p53 activity by Smyd2-mediated methylation, *Nature 444*, 629-632.

[39] Saddic, L. A., West, L. E., Aslanian, A., Yates, J. R., 3rd, Rubin, S. M., Gozani, O., and Sage, J. (2010) Methylation of the retinoblastoma tumor suppressor by SMYD2, *The Journal of biological chemistry 285*, 37733-37740.

[40] Evans, R. M. (1988) The steroid and thyroid hormone receptor superfamily, *Science 240*, 889-895.

[41] Nilsson, S., Makela, S., Treuter, E., Tujague, M., Thomsen, J., Andersson, G., Enmark, E., Pettersson, K., Warner, M., and Gustafsson, J. A. (2001) Mechanisms of estrogen action, *Physiological reviews 81*, 1535-1565.

[42] Kumar, V., and Chambon, P. (1988) The Estrogen-Receptor Binds Tightly to Its Responsive Element as a Ligand-Induced Homodimer, *Cell 55*, 145-156.

[43] Berry, M., Nunez, A. M., and Chambon, P. (1989) Estrogen-Responsive Element of the Human Ps2 Gene Is an Imperfectly Palindromic Sequence, *P Natl Acad Sci USA 86*, 1218-1222.

[44] Burns, K. A., and Korach, K. S. (2012) Estrogen receptors and human disease: an update, *Archives of toxicology 86*, 1491-1504.

[45] Kim, M. Y., Hsiao, S. J., and Kraus, W. L. (2001) A role for coactivators and histone acetylation in estrogen receptor alpha-mediated transcription initiation, *The EMBO journal 20*, 6084-6094.

[46] Ansari, K. I., Kasiri, S., Hussain, I., and Mandal, S. S. (2009) Mixed lineage leukemia histone methylases play critical roles in estrogen-mediated regulation of HOXC13, *The FEBS journal 276*, 7400-7411.

[47] Berry, N. B., Fan, M., and Nephew, K. P. (2008) Estrogen receptor-alpha hinge-region lysines 302 and 303 regulate receptor degradation by the proteasome, *Molecular endocrinology (Baltimore, Md.) 22*, 1535-1551.

[48] Sentis, S. (2005) Sumoylation of the Estrogen Receptor Hinge Region Regulates Its Transcriptional Activity, *Molecular Endocrinology 19*, 2671-2684.

[49] Zhou, Q., Shaw, P. G., and Davidson, N. E. (2009) Epigenetics meets estrogen receptor: regulation of estrogen receptor by direct lysine methylation, *Endocrine Related Cancer 16*, 319-323.

[50] Anbalagan, M., Huderson, B., Murphy, L., and Rowan, B. G. (2012) Post-translational modifications of nuclear receptors and human disease, *Nucl Recept Signal 10*, e001.

[51] Subramanian, K., Jia, D., Kapoor-Vazirani, P., Powell, D. R., Collins, R. E., Sharma, D., Peng, J., Cheng, X., and Vertino, P. M. (2008) Regulation of estrogen receptor alpha by the SET7 lysine methyltransferase, *Molecular cell 30*, 336-347.

[52] Le Romancer, M., Treilleux, I., Leconte, N., Robin-Lespinasse, Y., Sentis, S., Bouchekioua-Bouzaghou, K., Goddard, S., Gobert-Gosse, S., and Corbo, L. (2008) Regulation of estrogen rapid signaling through arginine methylation by PRMT1, *Molecular cell 31*, 212-221.

[53] Brown, M. A., Sims, R. J., Gottlieb, P. D., and Tucker, P. W. (2006) 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 5*, 26.

[54] Gottlieb, P. D., Pierce, S. A., Sims, R. J., Yamagishi, H., Weihe, E. K., Harriss, J. V., Maika, S. D., Kuziel, W. A., King, H. L., Olson, E. N., Nakagawa, O., and Srivastava, D. (2002) Bop encodes a muscle-restricted protein containing MYND and SET domains and is essential for cardiac differentiation and morphogenesis, *Nat. Genet. 31*, 25-32.

[55] Abu-Farha, M., Lambert, J. P., Al-Madhoun, A. S., Elisma, F., Skerjanc, I. S., and Figeys, D. (2008) The tale of two domains: proteomics and genomics analysis of SMYD2, a new histone methyltransferase, *Mol Cell Proteomics 7*, 560-572.

[56] Tan, X., Rotllant, J., Li, H., De Deyne, P., and Du, S. J. (2006) SmyD1, a histone methyltransferase, is required for myofibril organization and muscle contraction in zebrafish embryos, *PNAS 103*, 2713-2718.

[57] Hamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F. P., Li, M., Yagyu, R., and Nakamura, Y. (2004) SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells, *Nat. Cell Biol. 6*, 731-740.

[58] Li, D., Niu, Z., Yu, W., Qian, Y., Wang, Q., Li, Q., Yi, Z., Luo, J., Wu, X., Wang, Y., Schwartz, R. J., and Liu, M. (2009) SMYD1, the myogenic activator, is a direct target of serum response factor and myogenin, *Nucleic Acids Res 37*, 7059-7071.

[59] Diehl, F., Brown, M. A., van Amerongen, M. J., Novoyatleva, T., Wietelmann, A., Harriss, J., Ferrazzi, F., Bottger, T., Harvey, R. P., Tucker, P. W., and Engel, F. B. (2010) Cardiac deletion of Smyd2 is dispensable for mouse heart development, *PLoS One 5*, e9748.

[60] Thompson, E. C., and Travers, A. A. (2008) A Drosophila Smyd4 homologue is a muscle-specific transcriptional modulator involved in development, *PLoS One 3*, e3008.

[61] Saddic, L. A., West, L. E., Aslanian, A., Yates, J. R., Rubin, S. M., Gozani, O., and Sage, J. (2010) Methylation of the retinoblastoma tumor suppressor by SMYD2, *J Biol Chem 24*, 24.

[62] Soti, C., Racz, A., and Csermely, P. (2002) A Nucleotide-dependent molecular switch controls ATP binding at the C-terminal domain of Hsp90. N-terminal nucleotide binding unmasks a C-terminal binding pocket, *The Journal of biological chemistry 277*, 7066-7075.

[63] Tariq, M., Nussbaumer, U., Chen, Y., Beisel, C., and Paro, R. (2009) Trithorax requires Hsp90 for maintenance of active chromatin at sites of gene expression, *Proc Natl Acad Sci U S A 106*, 1157-1162.

[64] Ruden, D. M., and Lu, X. (2008) Hsp90 affecting chromatin remodeling might explain transgenerational epigenetic inheritance in Drosophila, *Curr Genomics 9*, 500-508.

[65] Sirinupong, N., Brunzelle, J., Ye, J., Pirzada, A., Nico, L., and Yang, Z. (2010) Crystal structure of cardiac-specific histone methyltransferase SmyD1 reveals unusual active site architecture, *J Biol Chem 285*, 40635-40644.

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

[67] Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, *Methods in Enzymology 276*, 307-326.

[68] Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K., and Terwilliger, T. C. (2002) PHENIX: building new software for automated crystallographic structure determination, *Acta Crystallogr D Biol Crystallogr 58*, 1948-1954.

[69] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.

[70] Blanc, E., Roversi, P., Vonrhein, C., Flensburg, C., Lea, S. M., and Bricogne, G. (2004) Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT, *Acta Crystallogr D Biol Crystallogr 60*, 2210-2221.

[71] Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J. Appl. Cryst. 26*, 283-291.

[72] Couture, J. F., Collazo, E., Brunzelle, J. S., and Trievel, R. C. (2005) Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase, *Genes Dev 19*, 1455.

[73] Wilson, J. R., Jing, C., Walker, P. A., Martin, S. R., Howell, S. A., Blackburn, G. M., Gamblin, S. J., and Xiao, B. (2002) Crystal structure and functional analysis of the histone methyltransferase SET7/9, *Cell 111*, 105-115.

[74] Zhang, X., Tamaru, H., Khan, S. I., Horton, J. R., Keefe, L. J., Selker, E. U., and X., C. (2002) Structure of the Neurospora SET domain protein DIM-5, a histone H3 lysine methyltransferase., *Cell 111*, 117.

[75] Couture, J. F., Collazo, E., Hauk, G., and Trievel, R. C. (2006) Structural basis for the methylation site specificity of SET7/9, *Nat Struct Mol Biol 13*, 140-146.

[76] Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., and Cheng, X. (2003) Structural basis for the product specificity of histone lysine methyltransferases., *Mol. Cell 12*, 177-185.

[77] Hayward, S., and Berendsen, H. J. (1998) Systematic analysis of domain motions in proteins from conformational change: new results on citrate synthase and T4 lysozyme, *Proteins 30*, 144-154.

[78] Xu, S., Wu, J., Sun, B., Zhong, C., and Ding, J. (2011) 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 25*, 25.

[79] Pearl, L. H., and Prodromou, C. (2000) Structure and in vivo function of Hsp90, *Curr Opin Struct Biol 10*, 46-51.

[80] Young, J. C., Obermann, W. M., and Hartl, F. U. (1998) Specific binding of tetratricopeptide repeat proteins to the C-terminal 12-kDa domain of hsp90, *The Journal of biological chemistry 273*, 18007-18010.

[81] Holm, L., and Rosenstrom, P. (2010) Dali server: conservation mapping in 3D, *Nucleic Acids Res 38*, W545-549.

[82] Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000) Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine, *Cell 101*, 199-210.

[83] Yang, J., Roe, S. M., Cliff, M. J., Williams, M. A., Ladbury, J. E., Cohen, P. T., and Barford, D. (2005) Molecular basis for TPR domain-mediated regulation of protein phosphatase 5, *The EMBO journal 24*, 1-10.

[84] Kumar, V., and Chambon, P. (1988) The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer, *Cell 55*, 145-156.

[85] Mann, M., Cortez, V., and Vadlamudi, R. K. (2011) Epigenetics of estrogen receptor signaling: role in hormonal cancer progression and therapy, *Cancers 3*, 1691-1707.

[86] Zhou, Q., Shaw, P. G., and Davidson, N. E. (2009) Epigenetics meets estrogen receptor: regulation of estrogen receptor by direct lysine methylation, *Endocrine-related cancer 16*, 319-323.

[87] Sharma, S., Kelly, T. K., and Jones, P. A. (2010) Epigenetics in cancer, *Carcinogenesis 31*, 27-36.

[88] Kim, M. Y., Woo, E. M., Chong, Y. T., Homenko, D. R., and Kraus, W. L. (2006) Acetylation of estrogen receptor alpha by p300 at lysines 266 and 268 enhances the deoxyribonucleic acid binding and transactivation activities of the receptor, *Molecular endocrinology (Baltimore, Md.) 20*, 1479-1493.

[89] Wang, C., Fu, M., Angeletti, R. H., Siconolfi-Baez, L., Reutens, A. T., Albanese, C., Lisanti, M. P., Katzenellenbogen, B. S., Kato, S., Hopp, T., Fuqua, S. A., Lopez, G. N., Kushner, P. J., and Pestell, R. G. (2001) Direct acetylation of the estrogen receptor alpha hinge region by p300 regulates transactivation and hormone sensitivity, *The Journal of biological chemistry 276*, 18375-18383.

[90] Sese, B., Barrero, M. J., Fabregat, M. C., Sander, V., and Izpisua Belmonte, J. C. (2013) SMYD2 is induced during cell differentiation and participates in early development, *The International journal of developmental biology 57*, 357-364.

[91] Brown, M. A., Sims, R. J., 3rd, Gottlieb, P. D., and Tucker, P. W. (2006) 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 5*, 26.

[92] Wang, L., Li, L., Zhang, H., Luo, X., Dai, J., Zhou, S., Gu, J., Zhu, J., Atadja, P., Lu, C., Li, E., and Zhao, K. (2011) Structure of human SMYD2 protein reveals the basis of p53 tumor suppressor methylation, *The Journal of biological chemistry 286*, 38725-38737.

[93] Ferguson, A. D., Larsen, N. A., Howard, T., Pollard, H., Green, I., Grande, C., Cheung, T., Garcia-Arenas, R., Cowen, S., Wu, J., Godin, R., Chen, H., and Keen, N. (2011) Structural basis of substrate methylation and inhibition of SMYD2, *Structure 19*, 1262-1273.

[94] Zhang, Z. Y. (2002) Protein tyrosine phosphatases: structure and function, substrate specificity, and inhibitor development, *Annual review of pharmacology and toxicology 42*, 209-234.

[95] Kabsch, W. (2010) Xds, *Acta Crystallogr D Biol Crystallogr 66*, 125-132.

[96] McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.

[97] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

[98] Kleywegt, G. J. (2007) Crystallographic refinement of ligand complexes, *Acta Crystallogr D Biol Crystallogr 63*, 94-100.

[99] Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010)

MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr 66*, 12-21.

[100] Couture, J. F., Collazo, E., Brunzelle, J. S., and Trievel, R. C. (2005) Structural and functional analysis of SET8, a histone H4 Lys-20 methyltransferase, *Genes Dev 19*, 1455-1465.

[101] Wu, J., Cheung, T., Grande, C., Ferguson, A. D., Zhu, X., Theriault, K., Code, E., Birr, C., Keen, N., and Chen, H. (2011) Biochemical characterization of human SET and MYND domain-containing protein 2 methyltransferase, *Biochemistry 50*, 6488-6497.

[102] Blatch, G. L., and Lassle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions, *Bioessays 21*, 932-939.

[103] Xiao, B., Jing, C., Wilson, J. R., Walker, P. A., Vasisht, N., Kelly, G., Howell, S., Taylor, I. A., Blackburn, G. M., and Gamblin, S. J. (2003) Structure and catalytic mechanism of the human histone methyltransferase SET7/9, *Nature 421*, 652-656.

[104] Zhang, X., Yang, Z., Khan, S. I., Horton, J. R., Tamaru, H., Selker, E. U., and Cheng, X. (2003) Structural basis for the product specificity of histone lysine methyltransferases, *Mol Cell 12*, 177-185.

[105] Arita, K., Shimizu, T., Hashimoto, H., Hidaka, Y., Yamada, M., and Sato, M. (2006) Structural basis for histone N-terminal recognition by human peptidylarginine deiminase 4, *Proc Natl Acad Sci U S A 103*, 5291-5296.

[106] Schellekens, G. A., de Jong, B. A., van den Hoogen, F. H., van de Putte, L. B., and van Venrooij, W. J. (1998) Citrulline is an essential constituent of antigenic

determinants recognized by rheumatoid arthritis-specific autoantibodies, *J Clin Invest 101*, 273-281.

[107] Todorova, N., Makarucha, A. J., Hine, N. D., Mostofi, A. A., and Yarovsky, I. (2013) Dimensionality of carbon nanomaterials determines the binding and dynamics of amyloidogenic peptides: multiscale theoretical simulations, *PLoS Comput Biol 9*, e1003360.

[108] Borbulevych, O. Y., Piepenbrink, K. H., Gloor, B. E., Scott, D. R., Sommese, R. F., Cole, D. K., Sewell, A. K., and Baker, B. M. (2009) T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility, *Immunity 31*, 885-896.

[109] Deshpande, C. N., Harrop, S. J., Boucher, Y., Hassan, K. A., Di Leo, R., Xu, X., Cui, H., Savchenko, A., Chang, C., Labbate, M., Paulsen, I. T., Stokes, H. W., Curmi, P. M., and Mabbutt, B. C. (2011) Crystal structure of an integron gene cassette-associated protein from Vibrio cholerae identifies a cationic drug-binding module, *Plos One 6*, e16934.

[110] van den Hemel, D., Brige, A., Savvides, S. N., and Van Beeumen, J. (2006) 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, *The Journal of biological chemistry 281*, 28152-28161.

[111] Aufhammer, S. W., Warkentin, E., Ermler, U., Hagemeier, C. H., Thauer, R. K., and Shima, S. (2005) 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 14*, 1840-1849.

[112] McCleverty, C. J., Columbus, L., Kreusch, A., and Lesley, S. A. (2008) Structure and ligand binding of the soluble domain of a Thermotoga maritima membrane protein of unknown function TM1634, *Protein Sci 17*, 869-877.

[113] Abu-Farha, M., Lanouette, S., Elisma, F., Tremblay, V., Butson, J., Figeys, D., and Couture, J. F. (2011) Proteomic analyses of the SMYD family interactomes identify HSP90 as a novel target for SMYD2, *J Mol Cell Biol 3*, 301-308.

[114] Chapman, R. W., Phillips, J. E., Hipkin, R. W., Curran, A. K., Lundell, D., and Fine, J. S. (2009) CXCR2 antagonists for the treatment of pulmonary disease, *Pharmacology & therapeutics 121*, 55-68.

[115] Tsai, H. H., Frost, E., To, V., Robinson, S., Ffrench-Constant, C., Geertman, R., Ransohoff, R. M., and Miller, R. H. (2002) The chemokine receptor CXCR2 controls positioning of oligodendrocyte precursors in developing spinal cord by arresting their migration, *Cell 110*, 373-383.

[116] Vandercappellen, J., Van Damme, J., and Struyf, S. (2008) The role of CXC chemokines and their receptors in cancer, *Cancer Lett 267*, 226-244.

[117] Jamieson, T., Clarke, M., Steele, C. W., Samuel, M. S., Neumann, J., Jung, A., Huels, D., Olson, M. F., Das, S., Nibbs, R. J., and Sansom, O. J. (2012) Inhibition of CXCR2 profoundly suppresses inflammation-driven and spontaneous tumorigenesis, *J Clin Invest 122*, 3127-3144.

[118] Addison, C. L., Daniel, T. O., Burdick, M. D., Liu, H., Ehlert, J. E., Xue, Y. Y., Buechi, L., Walz, A., Richmond, A., and Strieter, R. M. (2000) The CXC chemokine receptor 2, CXCR2, is the putative receptor for ELR+ CXC chemokine-induced angiogenic activity, *J Immunol 165*, 5269-5277.

[119] Wu, Y., Wang, S., Farooq, S. M., Castelvetere, M. P., Hou, Y., Gao, J. L., Navarro, J. V., Oupicky, D., Sun, F., and Li, C. (2012) A chemokine receptor CXCR2 macromolecular complex regulates neutrophil functions in inflammatory diseases, *The Journal of biological chemistry 287*, 5744-5755.

[120] Geng, J. G. (2001) Directional migration of leukocytes: their pathological roles in inflammation and strategies for development of anti-inflammatory therapies, *Cell Res 11*, 85-88.

[121] Harris, B. Z., and Lim, W. A. (2001) Mechanism and role of PDZ domains in signaling complex assembly, *Journal of cell science 114*, 3219-3231.

[122] Tonikian, R., Zhang, Y., Sazinsky, S. L., Currell, B., Yeh, J. H., Reva, B., Held, H. A., Appleton, B. A., Evangelista, M., Wu, Y., Xin, X., Chan, A. C., Seshagiri, S., Lasky, L. A., Sander, C., Boone, C., Bader, G. D., and Sidhu, S. S. (2008) A specificity map for the PDZ domain family, *PLoS Biol 6*, e239.

[123] Sheng, M., and Sala, C. (2001) PDZ domains and the organization of supramolecular complexes, *Annu Rev Neurosci 24*, 1-29.

[124] Lee, H. J., and Zheng, J. J. (2010) PDZ domains and their binding partners: structure, specificity, and modification, *Cell Commun Signal 8*, 8.

[125] Schultz, J., Copley, R. R., Doerks, T., Ponting, C. P., and Bork, P. (2000) SMART: a web-based tool for the study of genetically mobile domains, *Nucleic Acids Res 28*, 231-234.

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

[127] Cushing, P. R., Fellows, A., Villone, D., Boisguerin, P., and Madden, D. R. (2008) The relative binding affinities of PDZ partners for CFTR: a biochemical basis for efficient endocytic recycling, *Biochemistry 47*, 10084-10098.

[128] Ardura, J. A., and Friedman, P. A. (2011) Regulation of G protein-coupled receptor function by Na+/H+ exchange regulatory factors, *Pharmacol Rev 63*, 882-900.

[129] Beuming, T., Skrabanek, L., Niv, M. Y., Mukherjee, P., and Weinstein, H. (2005) PDZBase: a protein-protein interaction database for PDZ-domains, *Bioinformatics 21*, 827-828.

[130] Zhang, Y., Yeh, S., Appleton, B. A., Held, H. A., Kausalya, P. J., Phua, D. C., Wong, W. L., Lasky, L. A., Wiesmann, C., Hunziker, W., and Sidhu, S. S. (2006) Convergent and divergent ligand specificity among PDZ domains of the LAP and zonula occludens (ZO) families, *The Journal of biological chemistry 281*, 22299-22311.

[131] Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) PROCHECK: A program to check the stereochemical quality of protein structures, *J Appl Crystallogr 26*, 283-291.

[132] Elkins, J. M., Papagrigoriou, E., Berridge, G., Yang, X., Phillips, C., Gileadi, C., Savitsky, P., and Doyle, D. A. (2007) Structure of PICK1 and other PDZ domains obtained with the help of self-binding C-terminal extensions, *Protein Sci 16*, 683-694.

[133] Sheng, R., Chen, Y., Yung Gee, H., Stec, E., Melowic, H. R., Blatner, N. R., Tun, M. P., Kim, Y., Kallberg, M., Fujiwara, T. K., Hye Hong, J., Pyo Kim, K., Lu, H., Kusumi, A., Goo Lee, M., and Cho, W. (2012) Cholesterol modulates cell signaling and protein networking by specifically interacting with PDZ domain-containing scaffold proteins, *Nat Commun 3*, 1249.

[134] Cardone, R. A., Greco, M. R., Capulli, M., Weinman, E. J., Busco, G., Bellizzi, A., Casavola, V., Antelmi, E., Ambruosi, B., Dell'Aquila, M. E., Paradiso, A., Teti, A., Rucci, N., and Reshkin, S. J. (2012) NHERF1 acts as a molecular switch to program metastatic behavior and organotropism via its PDZ domains, *Mol Biol Cell 23*, 2028-2040.

[135] Mamonova, T., Kurnikova, M., and Friedman, P. A. (2012) Structural basis for NHERF1 PDZ domain binding, *Biochemistry 51*, 3110-3120.

[136] Stein, A., and Aloy, P. (2008) Contextual specificity in peptide-mediated protein interactions, *Plos One 3*, e2524.

[137] Wang, S., Raab, R. W., Schatz, P. J., Guggino, W. B., and Li, M. (1998) Peptide binding consensus of the NHE-RF-PDZ1 domain matches the C-terminal sequence of cystic fibrosis transmembrane conductance regulator (CFTR), *FEBS Lett 427*, 103-108.

[138] Molina, J. R., Agarwal, N. K., Morales, F. C., Hayashi, Y., Aldape, K. D., Cote, G., and Georgescu, M. M. (2012) PTEN, NHERF1 and PHLPP form a tumor suppressor network that is disabled in glioblastoma, *Oncogene 31*, 1264-1274.

[139] Voltz, J. W., Weinman, E. J., and Shenolikar, S. (2001) Expanding the role of NHERF, a PDZ-domain containing protein adapter, to growth regulation, *Oncogene 20*, 6309-6314.

[140] Runyon, S. T., Zhang, Y., Appleton, B. A., Sazinsky, S. L., Wu, P., Pan, B., Wiesmann, C., Skelton, N. J., and Sidhu, S. S. (2007) Structural and functional analysis of the PDZ domains of human HtrA1 and HtrA3, *Protein Sci 16*, 2454-2471.

[141] Morales, F. C., Takahashi, Y., Momin, S., Adams, H., Chen, X., and Georgescu, M. M. (2007) NHERF1/EBP50 head-to-tail intramolecular interaction masks association with PDZ domain ligands, *Mol Cell Biol 27*, 2527-2537.

[142] Wang, B., Bisello, A., Yang, Y., Romero, G. G., and Friedman, P. A. (2007) NHERF1 regulates parathyroid hormone receptor membrane retention without affecting recycling, *The Journal of biological chemistry 282*, 36214-36222.

[143] Raghuram, V., Mak, D. O., and Foskett, J. K. (2001) Regulation of cystic fibrosis transmembrane conductance regulator single-channel gating by bivalent PDZ-domain-mediated interaction, *Proc Natl Acad Sci U S A 98*, 1300-1305.

[144] Raghuram, V., Hormuth, H., and Foskett, J. K. (2003) A kinase-regulated mechanism controls CFTR channel gating by disrupting bivalent PDZ domain interactions, *Proc Natl Acad Sci U S A 100*, 9620-9625.

[145] Voltz, J. W., Brush, M., Sikes, S., Steplock, D., Weinman, E. J., and Shenolikar, S. (2007) Phosphorylation of PDZ1 domain attenuates NHERF-1 binding to cellular targets, *The Journal of biological chemistry 282*, 33879-33887.

[146] Takahashi, Y., Morales, F. C., Kreimann, E. L., and Georgescu, M. M. (2006) PTEN tumor suppressor associates with NHERF proteins to attenuate PDGF receptor signaling, *The EMBO journal 25*, 910-920.

[147] Glynn, P. C., Henney, E., and Hall, I. P. (2002) The selective CXCR2 antagonist SB272844 blocks interleukin-8 and growth-related oncogene-alpha-mediated inhibition of spontaneous neutrophil apoptosis, *Pulm Pharmacol Ther 15*, 103-110.

[148] Weinman, E. J., Minkoff, C., and Shenolikar, S. (2000) Signal complex regulation of renal transport proteins: NHERF and regulation of NHE3 by PKA, *Am J Physiol Renal Physiol 279*, F393-399.

[149] Wang, B., Yang, Y., Abou-Samra, A. B., and Friedman, P. A. (2009) NHERF1 regulates parathyroid hormone receptor desensitization: interference with beta-arrestin binding, *Mol Pharmacol 75*, 1189-1197.

[150] Shenolikar, S., Voltz, J. W., Cunningham, R., and Weinman, E. J. (2004) Regulation of ion transport by the NHERF family of PDZ proteins, *Physiology (Bethesda) 19*, 362-369.

[151] Lu, G., Wu, Y., Jiang, Y., Wang, S., Hou, Y., Guan, X., Brunzelle, J., Sirinupong, N., Sheng, S., Li, C., and Yang, Z. (2013) 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*.

[152] Wang, S., Wu, Y., Hou, Y., Guan, X., Castelvetere, M. P., Oblak, J. J., Banerjee, S., Filtz, T. M., Sarkar, F. H., Chen, X., Jena, B. P., and Li, C. (2013) CXCR2 macromolecular complex in pancreatic cancer: a potential therapeutic target in tumor growth, *Transl Oncol 6*, 216-225.

[153] Gerek, Z. N., and Ozkan, S. B. (2010) A flexible docking scheme to explore the binding selectivity of PDZ domains, *Protein Sci 19*, 914-928.

[154] Stiffler, M. A., Chen, J. R., Grantcharova, V. P., Lei, Y., Fuchs, D., Allen, J. E., Zaslavskaia, L. A., and MacBeath, G. (2007) PDZ domain binding selectivity is optimized across the mouse proteome, *Science 317*, 364-369.

[155] Basdevant, N., Weinstein, H., and Ceruso, M. (2006) Thermodynamic basis for promiscuity and selectivity in protein-protein interactions: PDZ domains, a case study, *J Am Chem Soc 128*, 12766-12777.

[156] Gerek, Z. N., Keskin, O., and Ozkan, S. B. (2009) Identification of specificity and promiscuity of PDZ domain interactions through their dynamic behavior, *Proteins 77*, 796-811.

[157] Fuentes, E. J., Der, C. J., and Lee, A. L. (2004) Ligand-dependent dynamics and intramolecular signaling in a PDZ domain, *J Mol Biol 335*, 1105-1115.

[158] Tang, C., Schwieters, C. D., and Clore, G. M. (2007) Open-to-closed transition in apo maltose-binding protein observed by paramagnetic NMR, *Nature 449*, 1078-1082.

[159] Eisenmesser, E. Z., Millet, O., Labeikovsky, W., Korzhnev, D. M., Wolf-Watz, M., Bosco, D. A., Skalicky, J. J., Kay, L. E., and Kern, D. (2005) Intrinsic dynamics of an enzyme underlies catalysis, *Nature 438*, 117-121.

[160] Volkman, B. F., Lipson, D., Wemmer, D. E., and Kern, D. (2001) Two-state allosteric behavior in a single-domain signaling protein, *Science 291*, 2429-2433.

[161] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta Crystallogr D Biol Crystallogr 67*, 235-242.

[162] Merritt, E. A. (1999) Expanding the model: anisotropic displacement parameters in protein structure refinement, *Acta Crystallogr D Biol Crystallogr 55*, 1109-1117.

[163] Roujeinikova, A., Baldock, C., Simon, W. J., Gilroy, J., Baker, P. J., Stuitje, A. R., Rice, D. W., Slabas, A. R., and Rafferty, J. B. (2002) X-ray crystallographic studies on butyryl-ACP reveal flexibility of the structure around a putative acyl chain binding site, *Structure 10*, 825-835.

[164] Liebscher, M., and Roujeinikova, A. (2009) Allosteric coupling between the lid and interdomain linker in DnaK revealed by inhibitor binding studies, *J Bacteriol 191*, 1456-1462.

[165] Roujeinikova, A., Sedelnikova, S., de Boer, G. J., Stuitje, A. R., Slabas, A. R., Rafferty, J. B., and Rice, D. W. (1999) Inhibitor binding studies on enoyl

reductase reveal conformational changes related to substrate recognition, *The Journal of biological chemistry 274*, 30811-30817.

[166] Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility, *J Mol Biol 55*, 379-400.

[167] Pohlmann, T., Bockmann, R. A., Grubmuller, H., Uchanska-Ziegler, B., Ziegler, A., and Alexiev, U. (2004) Differential peptide dynamics is linked to major histocompatibility complex polymorphism, *The Journal of biological chemistry 279*, 28197-28201.

[168] Zhang, X. J., Wozniak, J. A., and Matthews, B. W. (1995) Protein flexibility and adaptability seen in 25 crystal forms of T4 lysozyme, *J Mol Biol 250*, 527-552.

[169] van Aalten, D. M., Crielaard, W., Hellingwerf, K. J., and Joshua-Tor, L. (2000) Conformational substates in different crystal forms of the photoactive yellow protein--correlation with theoretical and experimental flexibility, *Protein Sci 9*, 64-72.

[170] Boehr, D. D., Nussinov, R., and Wright, P. E. (2009) The role of dynamic conformational ensembles in biomolecular recognition, *Nat Chem Biol 5*, 789-796.

[171] Totrov, M., and Abagyan, R. (2008) Flexible ligand docking to multiple receptor conformations: a practical alternative, *Curr Opin Struct Biol 18*, 178-184.

[172] Hussain, F., Wang, J., Ahmed, R., Guest, S. K., Lam, E. W., Stamp, G., and El-Bahrawy, M. (2010) The expression of IL-8 and IL-8 receptors in pancreatic adenocarcinomas and pancreatic neuroendocrine tumours, *Cytokine 49*, 134-140.

[173] Li, A., King, J., Moro, A., Sugi, M. D., Dawson, D. W., Kaplan, J., Li, G., Lu, X., Strieter, R. M., Burdick, M., Go, V. L., Reber, H. A., Eibl, G., and Hines, O. J. (2011) Overexpression of CXCL5 is associated with poor survival in patients with pancreatic cancer, *Am J Pathol 178*, 1340-1349.

[174] Miyamoto, M., Shimizu, Y., Okada, K., Kashii, Y., Higuchi, K., and Watanabe, A. (1998) Effect of interleukin-8 on production of tumor-associated substances and autocrine growth of human liver and pancreatic cancer cells, *Cancer Immunol Immunother 47*, 47-57.

[175] Takamori, H., Oades, Z. G., Hoch, O. C., Burger, M., and Schraufstatter, I. U. (2000) Autocrine growth effect of IL-8 and GROalpha on a human pancreatic cancer cell line, Capan-1, *Pancreas 21*, 52-56.

[176] Magalhaes, A. C., Dunn, H., and Ferguson, S. S. (2012) Regulation of GPCR activity, trafficking and localization by GPCR-interacting proteins, *Br J Pharmacol 165*, 1717-1736.

[177] Neel, N. F., Barzik, M., Raman, D., Sobolik-Delmaire, T., Sai, J., Ham, A. J., Mernaugh, R. L., Gertler, F. B., and Richmond, A. (2009) VASP is a CXCR2-interacting protein that regulates CXCR2-mediated polarization and chemotaxis, *Journal of cell science 122*, 1882-1894.

[178] Kim, J. K., Lim, S., Kim, J., Kim, S., Kim, J. H., Ryu, S. H., and Suh, P. G. (2011) Subtype-specific roles of phospholipase C-beta via differential interactions with PDZ domain proteins, *Adv Enzyme Regul 51*, 138-151.

[179] Lu, G., Wu, Y., Jiang, Y., Wang, S., Hou, Y., Guan, X., Brunzelle, J., Sirinupong, N., Sheng, S., Li, C., and Yang, Z. (2013) 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 8*, e76219.

[180] Jiang, Y., Lu, G., Trescott, L. R., Hou, Y., Guan, X., Wang, S., Stamenkovich, A., Brunzelle, J., Sirinupong, N., Li, C., and Yang, Z. (2013) New Conformational State of NHERF1-CXCR2 Signaling Complex Captured by Crystal Lattice Trapping, *Plos One 8*, e81904.

ABSTRACT

STRUCTURAL BASIS OF EPIGENETIC REGULATION AND PROTEIN SCAFFOLDING IN DEVELOPMENT AND DISEASES

by

YUANYUAN JIANG

August 2014

Advisor: Dr. Zhe Yang

Major: Biochemistry and Molecular Biology

Degree: Doctor of Philosophy

Protein X-ray crystallography is a powerful approach for elucidating protein structure and function. The high-resolution data generated by X-ray allow us to visualize protein structures in a three-dimensional (3D) space, which is vital for our understanding of the protein intra- and intermolecular interactions that explain the mechanisms of various biological events. More importantly, such information can provide a structural basis for developing new methods and strategies of targeted drug discovery. In this dissertation, by using X-ray crystallography as the primary approach, we have performed the structural and functional studies of SMYD2 and NHERF1 and have determined their mechanisms of action in epigenetic regulation and protein scaffolding, respectively.

Primarily identified as a histone lysine methyltransferase, SMYD2 has been shown to be play important roles in muscle development and tumorigenesis. In addition to histone substrate, SMYD2 can also methylate non-histone proteins including p53,

retinoblastoma tumor suppressor and estrogen receptor alpha. However, there are still many gaps in knowledge regarding the mechanisms underlying the activity regulation and substrate recognition of SMYD2. In this dissertation, we solved the crystal structures of SMYD2 with two different cofactors. Both cofactor-bound SMYD2 structures have a two-lobed structure with the active site partially blocked by a domain at the C-terminus (CTD). Although the two structures are highly superimposable, detailed structural analysis revealed the significantly different CTD conformations, suggesting the CTD flexibility that may be involved in the regulation of SMYD2 histone methyltransferase activity. In addition, the structural similarity between the CTD and the tetratricopeptide repeats (TPR) suggests a possible mechanism for the Hsp90-mediated SMYD activity enhancement. Based on such knowledge, we then employed the co-crystallization approach to study the mechanisms for the substrate recognition. We have successfully co-crystallized SMYD2 with a non-histone substrate, estrogen receptor alpha (ERα). The complex structure revealed that ERα peptide binds SMYD2 in a U-shaped conformation with the binding specificity determined predominantly by residues C-terminal to the target lysine. The structure also showed that the broad specificity of SMYD2 is achieved by multiple molecular mechanisms such as distinct peptide binding modes and the intrinsic dynamics of peptide ligands. Interestingly, a novel potentially SMYD2-specific PEG binding site is identified in the CTD, implicating possible functions in additional substrate binding or protein-protein interactions.

The formation of CXCR2−NHERF1−PLCβ macromolecular complex plays vital roles in both inflammation and pancreatic cancers. In neutrophils, this NHERF1-mediated

macromolecular complex is essential in intracellular calcium mobilization and neutrophil migration. In pancreatic cancer cells, this complex regulates tumor proliferation and invasion. Therefore, targeting this NHERF1-mediated macromolecular complex will have great clinical importance. The second objective of this dissertation is to provide the structural basis for the formation of this NHERF1-mediated macromolecular complex. To achieve this, we first solved the complex structures of the NHERF1 PDZ1 domain with the C-terminal sequence of CXCR2 in two different crystal forms. Although the superposition revealed a high degree of overall structural similarity, distinct conformations were observed between the two forms in substrate-binding pocket and bound peptide. These conformational differences indicated that the flexibility of the ligand-binding pocket might be required for diverse peptide recognition. The structural comparison also reveals that the intrinsic dynamics of the peptide ligand may allow the PDZ1 domain for interactions with different peptide recognition residues.

The interactions between NHERF1 and the CXCR2 downstream effector PLCβ3 have been studied using the same strategy as mentioned above. The structural studies of the PDZ1−PLCβ3 complex allowed us to identify the determinants of the PDZ1 binding specificity. We also showed that PLCβ3 can bind PDZ2 in pancreatic cancer cells, consistent with the observation that the peptide binding pocket of these PDZ domains are highly structurally conserved.

In summary, the studies preformed in this dissertation have revealed new insights into the mechanisms behind the lysine methylation machinery and protein scaffolding

which are central to many biological processes and diseases. Such findings will be of great benefit in the development of alternative therapeutic strategies and drug design.

AUTOBIOGRAPHICAL STATEMENT

Education

2004-2008 B.S. in Biotechnology, Northwestern University of China, Xi'an, China,
2009-pres. Ph.D. in Biochemistry and Molecular Biology; Wayne State University, Detroit, MI

Rewards and Honors

2004-2005 Northwest University of China Scholarship
2013 Travel Grant for 2013 Annual Meeting of the American Crystallographic Association
2013-2014 Thomas C. Rumble University Graduate Fellowship, Wayne State University
2014 Graduate Student Professional Travel Awards, Wayne State University
2014 BMB C. P. Lee Travel Award, Wayne State University

Peer-reviewed Publications

1. Holcomb, J., **Jiang, Y.**, Guan, X., Trescott, L., Lu, G., Hou, Y., Wang, S., Brunzelle, J., Sirinupong, N., Li, C. and Yang. Z. Crystal Structure of the NHERF1 PDZ2 Domain in Complex with the Chemokine Receptor CXCR2 Reveals Probable Modes of PDZ2 Dimerization. *Biochem Biophys Res Commun,* 10.1016/j.bbrc.2014.04.085 (2014).
2. **Jiang, Y.**, Wang, S., Holcomb, J., Trescott, L., Guan, X., Hou, Y., Brunzelle, J., Sirinupong, N., Li, C. and Yang. Z. Crystallographic Analysis of NHERF1-PLCβ3 Interaction Provides Structural Basis for CXCR2 Signaling in Pancreatic Cancer. *Biochem Biophys Res Commun,* 10.1016/j.bbrc. 2014.03.028 (2014).
3. Holcomb, J., **Jiang, Y.,** Lu, G., Trescott, L., Brunzelle, J., Sirinupong, N., Li, C., Naren, A. and Yang, Z. Structural Insights into PDZ-mediated Interaction of NHERF2 and LPA2, a Cellular Event Implicated in CFTR Channel Regulation. *Biochem Biophys Res Commun*, 10.1016/j.bbrc. 2014.02.128 (2014).
4. **Jiang, Y.,** Trescott, L., Holcomb, J., Zhang, X., Brunzelle, J., Sirinupong, N., Shi, X. and Yang, Z. Structural Insights into Estrogen Receptor α Methylation by Histone Methyltransferase SMYD2, a Cellular Event Implicated in Estrogen Signaling Regulation. *Journal of Molecular Biology*, 10.1016/j.jmb.2014.02.019 (2014).
5. **Jiang, Y.,** Lu, G., Trescott, L., Hou, Y., Guan, X., Wang, S., Stamenkovich, A., Brunzelle, J., Sirinupong, N., Spaller, M., Li, C. & Yang, Z. New Conformational State of NHERF1-CXCR2 Signaling Complex Captured by Crystal Lattice Trapping. PLoS ONE, 8(12): e81904. doi:10.1371/journal.pone.0081904 (2013) [PMCID: PMC3858284].
6. Lu, G., Wu, Y., **Jiang, Y.,** Wang, S., Hou, Y., Guan, X., Brunzelle, J., Sirinupong, N., Sheng, S., Li, C. & Yang, Z. 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 8(10): e76219. doi:10.1371/journal.pone.0076219 (2013) [PMCID: PMC3788737]. (published as co-first author)
7. Zhang, X., Tanaka, K., Li, J., Yang, J., Peng, D., **Jiang, Y.,** Yang, Z., Barton, M., Wen, H., and Shi, X. Regulation of estrogen receptor α by SMYD2-mediated protein methylation. Proc. Natl. Acad. Sci. USA, 110:17284-9 (2013) [PMCID: PMC3808627].
8. **Jiang, Y.,** Sirinupong, N., Brunzelle, J. and Yang, Z. Crystal structures of histone and p53 methyltransferase SmyD2 reveal a conformational flexibility of the autoinhibitory C-terminal domain. PLoS ONE, 6:e21640 (2011) [PMCID: PMC3125274].