The Dissertation Committee for Mengmeng Zhang Certifies that this is the approved version of the following dissertation:

PHOSPHATASES AND PROLYL-ISOMERASE IN THE REGULATION OF THE C-TERMINAL DOMAIN OF EUKARYOTIC RNA POLYMERASE II

Committee:

Yan Zhang, Supervisor

Jon D. Robertus

Dean R. Appling

Dionicio R. Siegel

Walter L. Fast

PHOSPHATASES AND PROLYL-ISOMERASE IN THE REGULATION OF THE C-TERMINAL DOMAIN OF EUKARYOTIC RNA POLYMERASE II

by

Mengmeng Zhang, B.E.; M.S.Bio.S.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

December 2012

Dedication

For my parents,

my husband, and my lovely baby girl.

Acknowledgements

The journey toward earning a Ph.D. degree is toilsome but enjoyable; it is a challenging adventure full of feelings of achievement! During this adventure, I was very fortunate to get help from exceptional mentors, colleagues, friends and my family.

My deepest gratitude is to my advisor, Dr. Yan Zhang, whose continuous support and encouragement promoted my maturity in scientific ability. She has been my tremendous mentor. Her advices on both my research as well as my career have been invaluable. Under her guidance, I found my scientific journey full of joy and surprises. I would like to express my sincere thanks to my committee members: Dr. Dean Appling, Dr. Walter Fast, Dr. Jon Robertus, and Dr. Dio Siegel, for their patience, dedication, and valuable comments on my research. My research journey would not have been so colorful without the guidance from my committee members. I am also grateful to Dr. Andrew Ellington, not only for teaching me how to evaluate ideas, but also for providing excellent comments on my research proposal.

I am happy to acknowledge my debt to Dr. David Hoffman, whose magic in the classroom for the first time made crystallography and other biophysical methods fascinating to me. I also would like to thank Dr. Ken Johnson, Dr. Marvin Hackert, Dr. Rick Russell, Dr. Ian Molineux, and Dr. Adrian Keatinge-Clay, whose truly inspiring teaching made me taste the wonderful flavor of biology.

During my scientific journey, I have been accompanied by my great lab-mates. I will never forget the days we stayed late in the lab, the days we discussed questions together, and days we trouble-shoot problems together. I would like to express my warm gratitude to them: Dr. Yonghua Luo, Dr. S.D. Yogesha, Wenzong Li, Wendy Matthews

PHOSPHATASES AND PROLYL-ISOMERASE IN THE REGULATION OF THE C-TERMINAL DOMAIN OF EUKARYOTIC RNA POLYMERASE II

Mengmeng Zhang, Ph.D.

The University of Texas at Austin, 2012

Supervisor: Yan Zhang

In eukaryotes, the first step of interpreting the genetic information is the transcription of DNA into RNA. For protein-coding genes, such transcription is carried out by RNA polymerase II. A special domain of RNA polymerase II, called the C-terminal domain (CTD), functions as a master controller for the transcription process by providing a platform to recruit regulatory proteins to nascent mRNA (Chapter 1-2). The modifications and conformational states of the CTD, termed the 'CTD code', represent a critical regulatory checkpoint for transcription. The CTD, found only in eukaryotes, consists of 26–52 tandem heptapeptide repeats with the consensus sequence, $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$. Phosphorylation of the serines and prolyl isomerization of the prolines represent two major regulatory mechanisms of the CTD. Interestingly, the phosphorylation sites are typically close to prolines, thus the conformation of the adjacent proline could impact the specificity of the corresponding kinases and phosphatases. Understanding how those modifying enzymes recognize and regulate the CTD is important for expanding our knowledge on the transcription regulation and deciphering the 'CTD code'.

During my PhD study, I studied the function of CTD phosphatases and prolyl isomerase in the CTD regulation using Scp1, Ssu72 and Pin1 as model regulators. Scp1 and Ssu72 are both Ser_5 phosphatases. However, Ssu72 is an essential protein and regulates the global transcription while Scp1 epigenetically silences the expression of specific neuronal genes. Pin1 is a highly conserved phosphorylation-specific prolyl isomerase that recognizes the phospho-Ser/Thr-Pro motif within the CTD as one of its primary substrates *in vivo*. Among these enzymes, Scp1 is the focal point of this dissertation, as it was studied from different angles, such as enzymatic mechanism (Chapter 3 describes the capture of phospho-aspartyl intermediate of Scp1 as a direct evidence for the proposed two-step mechanism), specific inhibition (Chapter 4 describes the identification and characterization of the first specific inhibitor of Scp1), and its non-active-site contact with the CTD (Chapter 5 describes the structural basis of this contact). These studies are of great importance towards understanding the molecular mechanism of the dephosphorylation process of the CTD by Scp1.

Table of Contents

List of Tables

List of Figures

xv

C-TERMINAL DOMAIN (CTD) OF EUKARYOTIC RNA POLYMERASE II

Chapter 1:   Introduction to the CTD of Eukaryotic RNA Polymerase II

Biological systems have long served as a source of inspiration for engineering, resulting in numerous inventions based on biomimetics. During the last few decades, as electronics-based information technology has matured and flourished, our understanding of the biological system has also proceeded to the molecular and informational level. Such understanding has enabled researchers to reprogram cells to undertake unnatural tasks, such as the production of proteins and metabolites for medical and industrial purposes. Recent development of systems and synthetic biology promises the development of cells with even more complex artificial functions that will require the collaboration of a large number of genes. With a similar philosophy, information-based molecular programming has also been established as a new path for nanotechnology. Central to all these new technologies, cellular or acellular, is the encoding and decoding of information at the molecular level, particularly using DNA as the information carrier. Therefore, from both a scientific and engineering point of view, it is important to understand how biological systems read information from DNA.

In biology, the major interpretation of this genetic information is through transcription regulation where eukaryotic RNA polymerase II plays the central role of transcribing the genetic information to the expressed protein. A special domain called the CTD of RNA polymerase II functions as a master controller for the transcription process by providing the template to recruit regulatory proteins to nascent mRNA (Corden 1990). The conformational states of the CTD, termed the CTD code, represent a critical regulatory check point for transcription (Dahmus 1996; Palancade et al. 2003; Meinhart

1

et al. 2005). The CTD, found only in eukaryotes, consists of 26–52 tandem heptapeptide repeats generally with the consensus sequence, $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$ (Corden 1990). Alterations of the sequence or the copy number of the heptapeptide may lead to distinguishable phenotypes or cell death (Chapman et al. 2007; Rogers et al. 2010). The CTD can spatially and temporally recruit different regulatory and processing factors to the transcriptional machinery, reviewed in Corden (Corden 1990) (**Figure 1-1**) but the domain is disordered in X-ray crystal structures. The CTD phosphorylation is a major mechanism by which cells regulate gene expression, with serines at position 2 and 5 as major phosphorylation sites (Phatnani et al. 2006). Recently, Ser_7 was also found to be phosphorylated *in vivo* although its function is still elusive (Chapman et al. 2007). A secondary mechanism for CTD regulation is prolyl-isomerization of the two prolines in the CTD heptapeptide sequence. By adjusting the cis–trans conformation of a proline adjacent to a phosphorylated serine, interaction of the CTD and binding partners it recruits can be modulated.

Coordinately regulated phosphorylation and dephosphorylation of the CTD plays an essential role not only in the recruitment and assembly of transcription complexes but also in temporal control of transcription and mRNA processing, reviewed in Refs (Buratowski 2009; Fuda et al. 2009). Evidence points to the phosphorylation state of Ser_2 and Ser_5 as the trigger for transcriptional process modulation (**Figure 1-2**). Ser_5 phosphorylation is required for assembly of the preinitiation complex (PIC) and facilitates mRNA capping via recruitment of capping enzymes (Cho et al. 1997; Komarnitsky et al. 2000). During the transition when the transcription complex moves away from the initiation site, Ser_5 gradually becomes dephosphorylated, whereas Ser_2 is phosphorylated. Ser_2 phosphorylation is the predominant CTD pattern on both elongating and terminating RNA Polymerase II, which ensures efficient 3'–RNA processing by

2

triggering the recruitment of 3'–RNA processing machinery (Fuda et al. 2009). At the end of transcription, CTDs are free of phosphate groups; non-phosphorylated CTDs are required for RNA polymerase II to recycle and bind a promoter for the next cycle of transcription (Fuda et al. 2009). Little is known about the timing of Ser_7 phosphorylation and how it affects the transcription but it appears to be an essential event specific for snRNA expression (Egloff et al. 2007).
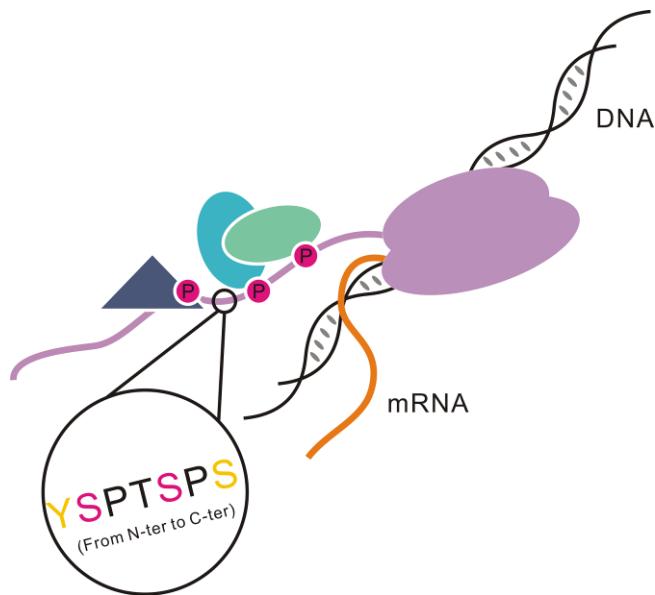
Figure 1-1:   Model of the CTD of RNA polymerase II.[1]

---

[1] The RNA polymerase II is colored with purple. Different shapes bound to the CTD indicate various proteins that are recruited by the CTD. Magenta circles labeled with 'P' indicate phosphorylation on the CTD. One repeat in the black circle is zoomed in to show its primary sequence 'YSPTSPS'. The Ser_2 and Ser_5 (colored with magenta) are always phosphorylated in each round of transcription, and Tyr_1 and Ser_7 (colored with yellow) are also detected as phosphorylation sites *in vivo*.

3
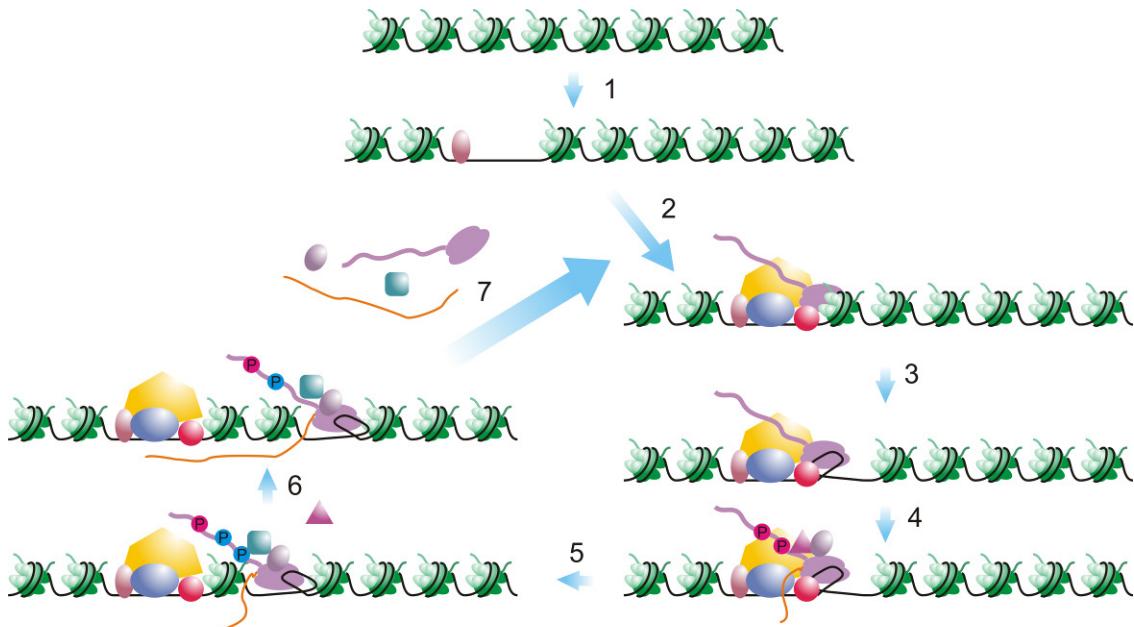
Figure 1-2:   Schematic diagram of RNA polymerase II-mediated transcription.[2]

One central question for CTD-directed transcription regulation is how high resolution recognition of different states of Ser_2 and Ser_5 are identified. Residues flanking Ser_2 and Ser_5 are highly similar, so it is puzzling how the transcription regulation is managed with such precision in location and timing. The most direct way to visualize and identify molecular elements in binding specificity is X-ray crystallography. A careful examination of the primary sequence of the CTD reveals the possibility that the CTD might have little secondary structure until its association with binding partners. In the next chapter, we will discuss our current understanding of the CTD structure and how the CTD is recognized by its binding partners (**Table 1-1**).

---

[2] The figure was adapted from (Fuda et al. 2009). Ser_5 phosphorylation (indicated by magenta blobs) occurs in promoter-proximal regions coincident with initiation and facilitates mRNA capping via recruiting capping enzymes (step 4). When the transcription complex moves away from the initiation site, Ser_5 gradually becomes hypophosphorylated (indicated by only one magenta blob), whereas Ser_2 gradually becomes hyperphosphorylated (indicated by blue blobs) (step 5). After termination, the CTD is dephosphorylated to reinitiate a new round of transcription (step 7). The colored-shapes indicate some essential accessory proteins and regulatory factors during transcription.

Species	Protein	Method	Preference	PDB code	References
Homo sapiens	Pin1	X-ray	Phospho.Ser$_5$ CTD	1f8a	(Verdecia et al. 2000)
Saccharomyces cerevisiae	Pcf11 CID	X-ray	Unphosphorylated CTD; phospho.Ser$_2$ CTD	1sza	(Meinhart et al. 2004)
Saccharomyces cerevisiae	Nrd CID	X-ray	Phospho.Ser$_5$ CTD	3clj	(Vasiljeva et al. 2008)
Homo sapiens	SCAF8 CID	X-ray	Phospho.Ser$_2$ CTD	3d9i	(Becker et al. 2008)
Homo sapiens	Scp1	X-ray	Phospho.Ser$_5$ CTD	2ght	(Zhang et al. 2006)
Saccharomyces cerevisiae	Fcp1	X-ray	Phospho.Ser$_2$ CTD	3ef0	(Ghosh et al. 2008)
Candida albicans	Cgt1	X-ray	Phospho.Ser$_5$ CTD	1p16	(Fabrega et al. 2003)
Homo sapiens	Set2 SRI	NMR	Phospho.Ser$_2$ CTD	2a7o	(Li et al. 2005)
Saccharomyces cerevisiae	Set2 SRI	NMR	Phospho.Ser$_2$ CTD	2c5z	(Vojnic et al. 2006)
Homo sapiens	CA150 FF	NMR	Unknown	2kis	(Murphy et al. 2009)
Saccharomyces cerevisiae	Prp40 FF	NMR	Unknown	2b7e	(Gasch et al. 2006)

Table 1-1:   Summary of CTD interacting proteins/domains.

REFERENCES

Becker, R., B. Loll and A. Meinhart (2008). "Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II." *J Biol Chem* **283**(33): 22659-22669.

Buratowski, S. (2009). "Progression through the RNA polymerase II CTD cycle." *Mol Cell* **36**(4): 541-546.

Chapman, R. D., M. Heidemann, T. K. Albert, R. Mailhammer, A. Flatley, M. Meisterernst, E. Kremmer and D. Eick (2007). "Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7." *Science* **318**(5857): 1780-1782.

Cho, E. J., T. Takagi, C. R. Moore and S. Buratowski (1997). "mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain." *Genes Dev* **11**(24): 3319-3326.

Corden, J. L. (1990). "Tails of RNA polymerase II." *Trends Biochem Sci* **15**(10): 383-387.

Dahmus, M. E. (1996). "Reversible phosphorylation of the C-terminal domain of RNA polymerase II." *J Biol Chem* **271**(32): 19009-19012.

Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick and S. Murphy (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." *Science* **318**(5857): 1777-1779.

Fabrega, C., V. Shen, S. Shuman and C. D. Lima (2003). "Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II." *Mol Cell* **11**(6): 1549-1561.

Fuda, N. J., M. B. Ardehali and J. T. Lis (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." *Nature* **461**(7261): 186-192.

Gasch, A., S. Wiesner, P. Martin-Malpartida, X. Ramirez-Espain, L. Ruiz and M. J. Macias (2006). "The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains." *J Biol Chem* **281**(1): 356-364.

Ghosh, A., S. Shuman and C. D. Lima (2008). "The structure of Fcp1, an essential RNA polymerase II CTD phosphatase." *Mol Cell* **32**(4): 478-490.

Komarnitsky, P., E. J. Cho and S. Buratowski (2000). "Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription." *Genes Dev* **14**(19): 2452-2460.

Li, M., H. P. Phatnani, Z. Guan, H. Sage, A. L. Greenleaf and P. Zhou (2005). "Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction

with the hyperphosphorylated C-terminal domain of Rpb1." *Proc Natl Acad Sci U S A* **102**(49): 17636-17641.

Meinhart, A. and P. Cramer (2004). "Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors." *Nature* **430**(6996): 223-226.

Meinhart, A., T. Kamenski, S. Hoeppner, S. Baumli and P. Cramer (2005). "A structural perspective of CTD function." *Genes Dev* **19**(12): 1401-1415.

Murphy, J. M., D. F. Hansen, S. Wiesner, D. R. Muhandiram, M. Borg, et al. (2009). "Structural studies of FF domains of the transcription factor CA150 provide insights into the organization of FF domain tandem arrays." *J Mol Biol* **393**(2): 409-424.

Palancade, B. and O. Bensaude (2003). "Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation." *Eur J Biochem* **270**(19): 3859-3870.

Phatnani, H. P. and A. L. Greenleaf (2006). "Phosphorylation and functions of the RNA polymerase II CTD." *Genes Dev* **20**(21): 2922-2936.

Rogers, C., Z. Guo and J. W. Stiller (2010). "Connecting mutations of the RNA polymerase II C-terminal domain to complex phenotypic changes using combined gene expression and network analyses." *PLoS One* **5**(6): e11386.

Vasiljeva, L., M. Kim, H. Mutschler, S. Buratowski and A. Meinhart (2008). "The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain." *Nat Struct Mol Biol* **15**(8): 795-804.

Verdecia, M. A., M. E. Bowman, K. P. Lu, T. Hunter and J. P. Noel (2000). "Structural basis for phosphoserine-proline recognition by group IV WW domains." *Nat Struct Biol* **7**(8): 639-643.

Vojnic, E., B. Simon, B. D. Strahl, M. Sattler and P. Cramer (2006). "Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription." *J Biol Chem* **281**(1): 13-15.

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." *Mol Cell* **24**(5): 759-770.

Chapter 2:   CTD Interacting Proteins

PIN1 AND THE CTD

The first glimpse of the CTD structure was through its interaction with human Pin1, a unique prolyl isomerase that catalyzes cis/trans isomerization of specific phospho.Ser/Thr-Pro motifs in signaling proteins (Lu et al. 1996; Ranganathan et al. 1997; Yaffe et al. 1997; Lu et al. 1999). Identification and characterization of this novel peptidyl-prolyl cis/trans isomerase (PPIase), Pin1, led to the discovery of a novel post-phosphorylation regulatory mechanism, in which regulation is achieved by conformational changes of a phosphorylated Ser/Thr-Pro peptide bond upon proline isomerization (**Figure 2-1**). This change in the configuration of the polypeptide has a profound effect on Pin1 targets and therefore modulates various signaling pathways at both transcriptional and post-translational levels. Specifically, prolyl-isomerization activity of Pin1 can interconvert the cis/trans conformation of the phospho.Ser/Thr-Pro motif of target proteins and make them better or worse substrates for conformation-specific signaling kinases (such as cyclin-dependent protein kinase, glycogen synthetase kinase 3 beta, and mitogen-activated protein kinase) and phosphatases (such as PP2A and Cdc25). Recent studies also provide substantial evidence implicating Pin1 in progression of malignant tumor cells (Lu 2004) and development of Alzheimer's disease (Etzkorn 2006). High affinity inhibitory unnatural peptides have been developed as a good template for chemical compounds targeting Pin1 for antineoplastic effects (Zhang et al. 2007).

Compelling data have implicated human Pin1 as a key modulator in the transcription mechanism. The yeast homologue of Pin1, Ess1, interacts physiologically and genetically with the CTD (Wu et al. 2000). Furthermore, hyperphosphorylated RNA

polymerase II appears to be the dominant binding target in yeast extracts (Morris et al. 1999). Considering the high local concentration of the phospho.Ser/Thr-Pro motif in the hyperphosphorylated CTD, it is plausible that the CTD is the major substrate of Pin1 *in vivo*. The K_d of a single CTD repeat that is phosphorylated at Ser_5 was reported to be 30 μM (Verdecia et al. 2000). Presumably, the CTD tail containing 26–52 such repeats localizes a substantial amount of Pin1.
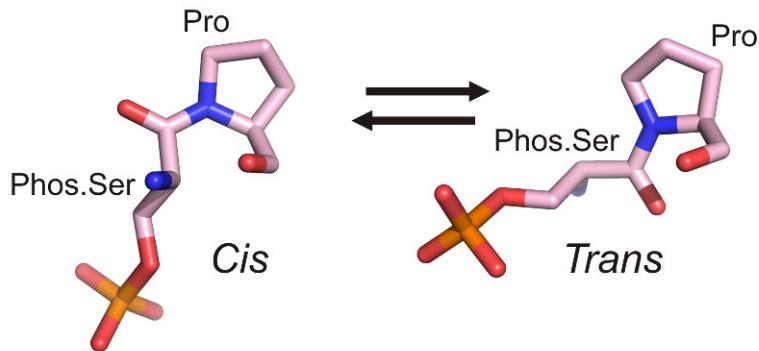
Figure 2-1:   Cis and trans conversion of proline in a phospho.Ser-Pro motif.

Pin1 is a 163 amino acid polypeptide that can be divided into two domains based on topology and function, a C-terminal PPIase domain and an N-terminal WW domain. Structure of the WW domain reveals three antiparallel β-strands forming a shallow interface with the PPIase domain for substrate peptide binding (Macias et al. 1996; Huang et al. 2000). It has long been realized that WW domains recognize proline-containing sequences but it was not clear until recently that WW domains join a group of modules that bind to protein ligands in a phosphorylation-dependent manner (Lu et al. 1999; Lu et al. 2002); these domains include SH2, PTB, 14-3-3, WD40, FHA, and FF domains. The modular nature of WW domain interactions leads to a classification into

four distinct groups based on binding specificity (Sudol et al. 2000). Group I WW domains, such as dystrophin and the Yes-associated protein YAP65, recognize 'PpxY' motifs (Chen et al. 1997). Group II, such as FE65 and formin binding proteins (FBPs), bind the 'PPLP' motif (Ermekova et al. 1997). A subset of FBPs interacts with 'PGM' motifs (Bedford et al. 1998). Group III WW domains select poly-proline motifs flanked by arginine or lysine (Komuro et al. 1999; Bedford et al. 2000). Group IV WW domains, including human Pin1 and Nedd4, specifically recognize a phospho.Ser/Thr-Pro motif (Lu et al. 1999; Verdecia et al. 2000; Myers et al. 2001).

The molecular detail of recognition of the CTD by Pin1 was elucidated by a structure of human Pin1 in complex with one doubly phosphorylated CTD repeat (Verdecia et al. 2000) (**Figure 2-2a**). So far, this is the only structure available for a full-length Pin1 binding to its target at the recognition module WW domain. The structure is consistent with the thermodynamic data, which showed that the phosphorylated Ser_5 in the CTD repeat is the major binding element in recognition. Loop 1 of the WW domain has been shown to be highly flexible in apo Pin1 (Kowalski et al. 2002) but this loop is essential for specificity recognition (Jager et al. 2006). In the complex structure, this loop warps toward the substrate peptide to ensure binding and results in an exaggerated twist in the triple-stranded β-sheet (Verdecia et al. 2000) (**Figure 2-2a**). This twist is coupled to a contraction of the WW domain ligand binding surface formed between the WW and PPIase domains. The two essential elements for peptide recognition include binding of phosphate by loop 1 residues and hydrophobic stacking of proline by Tyr23 and Trp34 (Verdecia et al. 2000). The binding of Pin1 to the CTD can modulate the regulatory effect of other CTD-binding proteins. *In vitro* experiments showed that Pin1 can influence the phosphorylation status of the CTD by inhibiting the transcription factor IIF-interacting CTD phosphatase 1 (Fcp1) and stimulating CTD phosphorylation by cdc2/cyclinB (Xu et

al. 2003; Palancade et al. 2004). The Rsp5, a ubiquitin ligase that binds to the CTD, also functions to oppose Pin1 effects on RNA polymerase II (Wu et al. 2001). The biological and structural results of Pin1 effects on RNA polymerase II function support a model that Pin1 works in a processive manner on the CTD with the WW domain acting as a binding element restricting movement to an efficient one dimensional walk and with the PPIase acting much like a reading head to processively isomerize the peptide bonds. The binding of Pin1 prolongs the phosphorylated state of the CTD by suppressing dephosphorylation, thereby enhancing the regulatory effect of CTD binding proteins.
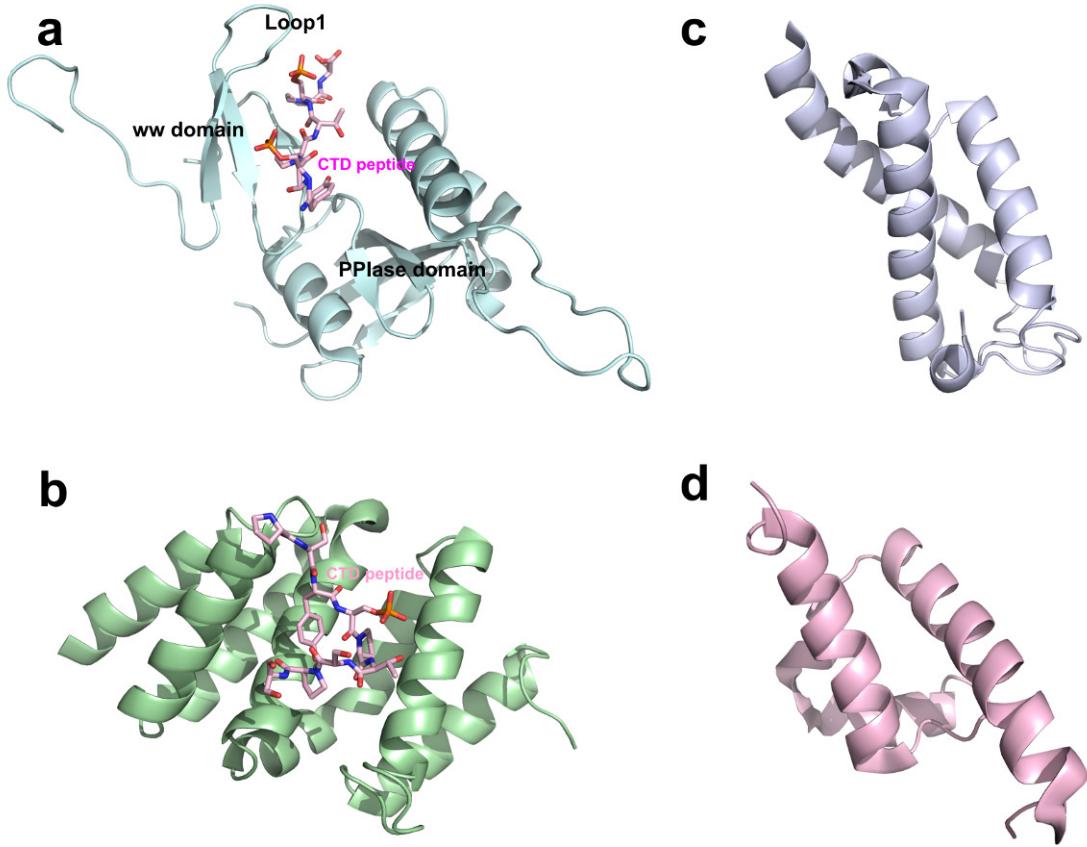
Figure 2-2:   Ribbon representation of four CTD recognition modules.[3]

CTD-INTERACTING DOMAIN (CID) AND THE CTD

Recognition of the phosphorylated CTD by Pin1 is mediated by its WW domain, a modular domain of around 40 residues that is essential for recognition of proline-rich motifs by the PPIase domain. However, the Pin1 WW domain also recognizes other substrates in addition to the CTD. A more specific recognition domain for the CTD is the CTD-interacting domain (CID) identified in multiple RNA processing and termination factors in eukaryotes (Meinhart et al. 2004; Noble et al. 2005) (**Figure 2-3**).

---

[3] (a) WW domain of Pin1 in complex with a short CTD peptide (1f8a); (b) CID domain of Pcf11 in complex with a short CTD peptide (1sza); (c) SRI domain of Set2 (2a7o); and (d) FF domain of CA150 (2kis).

In yeast, the effective termination of transcription relies on the recruitment of cleavage factors by Pcf11. The Pcf11 is a yeast protein of 70 kD with a CID domain directly targeted to the CTD of RNA polymerase II, which gives us the first glance of how a CID recognizes heptad repeats of the CTD. Interestingly, the CID domain of Pcf11 can bind to both unphosphorylated or phospho.Ser_2 CTD in biophysical binding assays (Meinhart et al. 2004). Consistent with the biochemical data for such preferences, the co-crystal structure of the Pcf11 CID and a CTD peptide shows no direct interaction between the phosphate group of phospho.Ser_2 and Pcf11, indicating no pre-requirement of phosphorylation of the CTD for protein binding (Meinhart et al. 2004) (**Figure 2-2b**). The CID domain, as an eight-helical bundle, recognizes a span of two heptad repeats in the CTD with a conserved groove, whereas the phosphate group of phospho.Ser_2 actually forms an intramolecular hydrogen bond with Thr_4 of the CTD and stabilizes the sharp β-turn formed by CTD, presenting the side chain to the binding groove (**Figure 2-2b**). Hydrogen bondings between the CID domain and the CTD peptide are distributed between the CID side chain and the main chain amide and carbonyl group of the CTD peptide. Importantly, hydrophobic interaction by Tyr_1 of the CTD to CID might contribute greatly for the specificity for phospho.Ser_2 over phospho.Ser_5. Meinhart and Cramer (Meinhart et al. 2004) conclude that Ser_2 phosphorylation stabilizes the CTD β-spiral that is incorporated into the transcription complex. Ser_5 phosphorylation would unwind the spiral resulting in an extended region that binds the capping enzyme. Overall, the binding of Pcf11 is not particularly strong but is consistent with its dynamic role involved in dismantling the elongation complex (Zhang et al. 2005; Hollingworth et al. 2006).

Another CID containing protein Nrd1 is an essential player in the termination pathway for RNA polymerase II-mediated transcription for snoRNA, snRNA as well as

cryptic unstable transcripts (CUT). The complex of Nrd1-Nab3-Sen1 is recruited to the transcription machinery through the CTD by the CID domain of Nrd1 of the complex. The CID domain derived from Nrd1 (**Figure 2-3**) has a very similar overall fold and a strong conservation of residues involved in CTD peptide binding with Pcf11, but Nrd1 shows a different specificity profile in which a much stronger preference for phospho.Ser_5 over phospho.Ser_2 in the CTD sequence is detected for the CID from Nrd1 by *in vitro* binding assays (Vasiljeva et al. 2008). Using yeast-two hybrid and fluorescence anisotropy, Vasiljeva showed a tenfold improvement in binding affinity for singly phosphorylated CTD double repeats at Ser_5 over Ser_2 (K_d = 40 µM for phospho.Ser_5 vs. 390 µM for phospho.Ser_2) and a slight improvement when the peptide is doubly phosphorylated (K_d = 16 µM upon double phosphorylation for both Ser_2 and Ser_5 sites). Specificity for phospho.Ser_5 on the protein is essential to understand how different termination pathways are selected. A potential different phosphate binding site was proposed but more insightful information about specificity will only be available with a structure of the Nrd1–CID complexed with the CTD peptide. Since the Nrd1-dependent termination pathway is usually associated with much shorter transcripts (a few hundred base-pairs), upon which point dephosphorylation of Ser_5 is not complete, a logical hypothesis is that phosphorylation at Ser_5 favors the selection of Nrd1 complex as a termination pathway in yeast. Indeed, Gudipati *et al.* (Gudipati et al. 2008) showed that reducing phosphorylation of Ser_5 using a mutant of kin28, a CTD Ser_5 kinase, will hamper Nrd-dependent termination. The selectivity for phospho.Ser_5 over phospho.Ser_2 might be the determining factor for the selection of termination pathways. A functional model for how the phosphorylation pattern of the CTD determines the transcriptional pathways suggests that termination by the Nrd1-Nab3-Sen1 complex is enforced when the CTD is still highly phosphorylated at Ser_5 (Rondon et al. 2008). The structure of a

complex of the CID of Nrd1 with the CTD domain will help provide a molecular explanation of transcription termination.
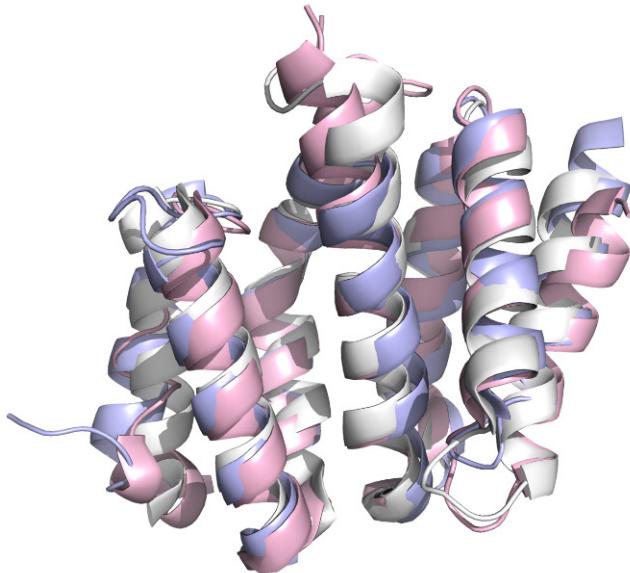
Figure 2-3:   Superimposition of CID domains from Pcf11, Nrd1 and SCAF8.[4]

Another CID containing protein human SCAF8 was crystallized with different phosphorylated forms of CTD peptides (**Figure 2-3**), providing more clues about how registration of a phosphate group on the CTD is encoded in target recognition for CID domains (Becker et al. 2008). The SCAF8 is implicated in splicing with a preference of phospho.Ser_2, similar to mRNA 3'-processing factor Pcf11. Indeed, a similar conformation of β-turn adapted by the CTD is observed in SCAF8 upon peptide binding but with one major distinction: phospho.Ser_2 is directly recognized by a basic residue, Arg112, through salt bridge interaction (Becker et al. 2008). At a similar position in Pcf11, a methionine residue was placed with no direct interaction with the CTD peptide.

---

[4] CID domain from Pcf11 is shown in light blue (PDB code: 1sza), the one from Nrd1 is shown in light pink (PDB code: 3clj), and the one from SCAF8 is shown in white (PDB code: 3d9i).

The replacement of methionine by arginine in SCAF8 distinguishes the phospho.Ser_2 CTD from the unphosphorylated form and might explain a tighter interaction for SCAF8 toward phosphoryl-peptide (K_d = 68 μM). In the study of Becker et $al.$ (Becker et al. 2008), an issue was raised whether phosphorylated Ser_7 contributes to recognition by the CID domain of SCAF8. Binding affinity measured by fluorescence anisotropy showed no advantage with additional Ser_7 phosphorylation (K_d = 68 μM for phospho.Ser_2 and K_d = 90 μM for doubly phosphorylated Ser_2 and Ser_7). Consistent with the binding interaction measurement, the complex structure of the SCAF8 CID and the peptide presents no interaction between the phosphate group of phospho.Ser_7 and the protein.

SMALL C-TERMINAL DOMAIN PHOSPHATASES (SCPS)/FCP PHOSPHATASES AND THE CTD

A wide range of enzymes participate in dynamic modifications of the CTD, including kinases and phosphatases responsible for addition and removal of phosphates. The CTD is principally phosphorylated by cyclin-dependent kinases (CDKs) with their associated cyclins. Specifically, Ser_5 phosphorylation is mainly catalyzed by Cdk7/cyclin H subunits of transcription factor IIH (TFIIH) (Lu et al. 1992; Hengartner et al. 1998; Fouillen et al. 2010); Ser_2 phosphorylation is mainly catalyzed by P-TEFb, which contains Cdk9/cyclin T subunits (Zhou et al. 2000; Shim et al. 2002). Intriguingly, it is suggested that Cdk9 also makes a contribution to Ser_5 phosphorylation and the relative contribution of TFIIH-associated Cdk7 varies between different genes based on experimental observations (Glover-Cutter et al. 2009). Moreover, the CTD can also be phosphorylated at both Ser_2 and Ser_5 by Cdk8 as part of the mediator complex, and preferentially phosphorylated at Ser_5 by MAPK2/Erk2 (Trigon et al. 1998). Recently, TFIIH-associated Cdk7 kinase has also been shown to phosphorylate Ser_7 in $vivo$ (Akhtar et al. 2009; Glover-Cutter et al. 2009). Even though structural information has been

17

obtained for Cdk7 (Lolli et al. 2004) and Cdk9/cyclin T (Baumli et al. 2008), how they recognize the CTD peptide and label the phosphorylation mark on the CTD is still elusive. It is assumed that specificity is achieved by other associated proteins in the multiprotein complexes they are involved (Cdk7 in TFIIH, Cdk8 in mediator, and Cdk9 in P-TEFb).

Dephosphorylation is essential for recycling RNA polymerase II, because after each round of transcription, the CTD has to be dephosphorylated in order to actively restart a new round of transcription. In humans, Fcp1, which is required for general transcription and cell viability, was the first discovered CTD-specific phosphatase with a catalytic preference for phospho.Ser$_2$. The Fcp1 is conserved among eukaryotes and was shown to be essential for cell survival in budding and fission yeast (Archambault et al. 1997; Kimura et al. 2002). The conserved region of Fcp1 is composed of two domains: an N-terminal FCP homology (FCPH) domain with phosphatase activity and a C-terminal breast cancer protein related C-terminal (BRCT) domain (Ghosh et al. 2008) (**Figure 2-4**).

Recently, a family of small CTD phosphatases (Scps) with activities preferential for phospho.Ser$_5$ was identified (Yeo et al. 2003; Yeo et al. 2005). This family includes three highly similar proteins designated Scp1, Scp2, and Scp3. The Scps also contain the FCPH catalytic domain that includes the DXDX(T/V) motif, the signature of a superfamily of phosphotransferases and phosphohydrolases called the haloacid dehalogenase (HAD) superfamily (Collet et al. 1998). Therefore, Fcp/Scp family members are classified as HAD superfamily enzymes. Interestingly, outside the signature motif, Scps share very little sequence similarity with the other enzymes in the HAD superfamily (Zhang et al. 2010). In humans, Scp1 has more than 20% sequence identity to Fcp1 in the FCPH domain but lacks the C-terminal BRCT domain that exists in Fcp1.

18

Moreover, Scp2 and Scp3, which also lack the BRCT domain, share more than 90% similarity with Scp1 in the FCPH domain (Yeo et al. 2003) (**Figure 2-4**).
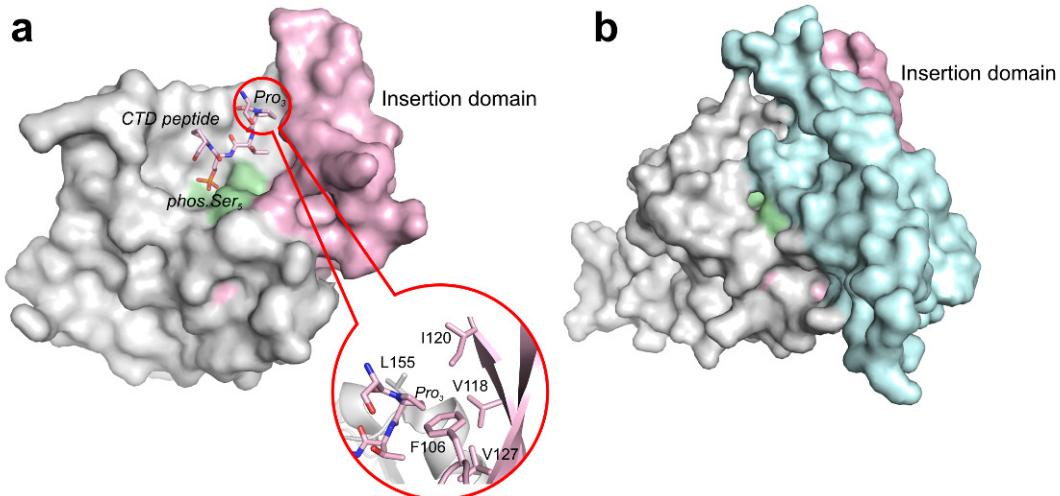
Figure 2-4:   Surface representation of Scp1 and Fcp1 FCPH domains.[5]

The apo structure of Scp1 solved by Kamenski *et al.* (Kamenski et al. 2004) showed a central parallel β sheet flanked by two α helices, a two-stranded β sheet, and a short 3_{10} helix. The conserved DXDX(T/V) signature motif lines part of a central crevice, which forms the active site and coordinates the Mg^{2+} ion that is essential to Scp1 phosphatase activity. The first aspartate in the signature motif is involved in Mg^{2+}-assisted phosphoryl transfer and acts as the phosphoryl acceptor. Mutation of this residue (Asp96 in Scp1) to alanine or asparagine abolished the activity of Scp1. The second aspartate (Asp98 in Scp1) also contributes to metal ion-binding and could possibly function as a general acid/base (Kamenski et al. 2004). The proposed phosphoryl transfer

[5] (a) Scp1 in complex with a short CTD peptide (2ght). The zoom-in picture shows the Pro_3 binds to the hydrophobic pocket with the aromatic residues shown in stick. (b) Fcp1 FCPH domain (3ef0). In both structures, the active site signature motif is colored with pale green, and the insertion domain is colored with light pink. Notably, the additional helical domain (light cyan) covers the 'insertion domain' in Fcp1 and makes it much less accessible for substrates.

mechanism for the Scp/Fcp family involves a phosphoryl-aspartate intermediate. Existence of this phosphoryl-enzyme intermediate was confirmed in a recent structural and functional study when we successfully trapped the phosphoryl-aspartate intermediate in the crystal structure of an Scp1D206A mutant soaked with *para*-nitrophenyl phosphate (*p*NPP) (see **Chapter 3**). The steady-state kinetic analysis of a variety of Scp1 mutants revealed the importance of Asp206 in Mg^{2+} coordination mediated by a water molecule. Moreover, snapshots of the phosphoryl transfer reaction at each stage of Scp1-mediated catalysis were also captured in this study (see **Chapter 3**).

In order to understand the discrimination of phospho.Ser_5 over phospho.Ser_2 as a substrate by Scp1, the complex structure of a dominant negative form of human Scp1 (Scp1D96N) bound with Ser_2/Ser_5-phosphorylated CTD peptide was obtained by crystal soaking (Zhang et al. 2006). The defined complex structure revealed a unique binding mode of the peptide in which $Ser_2Pro_3Thr_4$(phospho.Ser_5) forms a β-turn. The phospho.Ser_5 binds to the active site groove through Mg^{2+} coordination. The Pro_3 is recognized by an aromatic-rich hydrophobic pocket near the active site that further confers substrate specificity (Zhang et al. 2006). Notably, Scp1 shows remarkable specificity toward the trans peptide-bond configuration of the two prolines in the CTD repeat, which can adopt both cis and trans configurations. Such configuration switching is known to modulate the structure of the CTD and its accessibility to kinases or phosphatases (Morris et al. 1999; Verdecia et al. 2000). An insertion domain formed by a three-stranded β sheet directly follows the signature motif (**Figure 2-4**). This insertion domain is unique to Fcp1/Scp1 family phosphatases and may assist in substrate recognition. Consistent with the known specificity of Scp1 toward phospho.Ser_5 instead of phospho.Ser_2, in the complex structure the phospho.Ser_2 flips out of the active site, making no direct interaction with the protein.

20

Even though Scp and Fcp share similar phosphatase active sites, their strategies for substrate recognition might be different, as suggested by the recent crystal structure of apo *Schizosaccharomyces pombe* Fcp1 (SpFcp1) (Ghosh et al. 2008) (**Figure 2-4**). The minimal effective CTD substrate for SpFcp1 is a single heptad CTD peptide: $Ser_5Pro_6Ser_7Tyr_1(phospho.Ser_2)Pro_3Thr_4$, among which the Tyr_1 and Pro_3 flanking the phospho.Ser_2 are critical determinants of Fcp1 activity (Hausmann et al. 2004). The SpFcp1 structure revealed that it is a Y-shaped protein composed of three structural domains (Ghosh et al. 2008). The stem of the Y is the FCPH domain that contains a globular catalytic phosphatase core similar to that of Scp1. One major difference is that the three-stranded β sheet in Scp1 (insertion domain) is accessible for substrate recognition, whereas in Fcp1 it is buried by a helical insertion domain, suggesting a different binding interface between Fcp and the CTD with phospho.Ser_2 (**Figure 2-4**).

MRNA CAPPING ENZYME CGT1 AND THE CTD

Even though 26–52 heptad repeats exist in the CTD primary sequence, recognition of the CTD by proteins in all the complex structures discussed above only show a spanning of one or two repeats. This is understandable since a balance between favorable interactions of the CTD with its binding partners versus the entropy cost for binding a disordered peptide needs to be achieved. This leads to the proposed mechanism that a double repeat of the CTD sequence is the functional unit for transcription. To explore if such rule is consistent with the mRNA capping enzyme Cgt1, a four-heptad-repeat peptide with each Ser_5 phosphorylated was used in a cocrystallization experiment (Fabrega et al. 2003). Interestingly, a long span of CTD peptide consisting of 17 amino acids was modeled in the density of one of the two monomers with an extensive buried surface of 1,600 A^2 (Fabrega et al. 2003). Three of the four phosphorylated Ser_5 residues

were visible in this structure with both the first and third phosphate group recognized by positively charged patches on the Cgt1 surface. Consistent with a previous study of Cgt1, the Tyr_1 position in the CTD sequence is essential for recognition by the protein (West et al. 1995). On the other hand, a single mutation of the Cgt protein for the recognition interface did not show obvious deleterious effects during yeast mutation screening (Fabrega et al. 2003), possibly due to the extended binding surface without one interaction dominating binding. Since the other monomer in the asymmetric unit shows a much smaller interface, it suggests that multiple phosphorylation sites are not a prerequisite for ligand binding. Further affinity measurements with a different length and registration of the $phospho.Ser_5$ would elucidate whether interaction with multiple repeats of the CTD is essential for the effective binding by Cgt1.

NUCLEAR MAGNETIC RESONANCE (NMR) STRUCTURE OF SET2

The identification of the histone methyltransferase Set2 as a novel CTD binding partner bridges the CTD code to the histone code by implicating the regulatory effect of the CTD on histone modification and epigenetic control (Gerber et al. 2003; Hampsey et al. 2003) (**Figure 2-2c**). In eukaryotes, large genomes are efficiently organized into nucleosomes that are fundamental repeat units of chromatin. The nucleosome is composed of a histone octamer consisting of two copies of each of the core histones H2A, H2B, H3, and H4, around which 147 bp of DNA is wrapped. Such a structure is not only a strategy to compress large genomic DNA, but also provides a potential solution for tight regulation of DNA replication, repair, and transcription. During transcription, for example, the binding sites for a variety of transcription factors can be occluded by histones, leading to transcriptional silencing (Kornberg et al. 1999). Access to specific loci on the nucleosomal DNA is dynamically regulated by many factors including

chromatin modifiers and chromatin remodelers. Histones can be covalently modified by chromatin modifiers at particular loci, most of which are concentrated in the relatively unstructured N-terminal tails of histones. These modifications include acetylation, methylation, phosphorylation, ubiquitination, and sumoylation (Kouzarides 2007). The histone code, which is defined by covalent modifications, represents a fundamental regulatory mechanism of gene expression and repression (Strahl et al. 2000). Furthermore, the histone code can be interpreted by different modules in a modification-dependent manner to decide whether a gene is to be transcribed.

A novel Rpb1-binding domain of Set2, called Set2–Rpb1 interacting (SRI) domain, mediates the recognition and interaction with the phosphoryl CTD (**Figure 2-2c**). Lys4 of histone 3 is methylated by the proteins of the Set1 family, while Lys36 is methylated by proteins of the Set2 family (Gerber et al. 2003). Both Set1 and Set2 associate with the CTD but at different stages of transcription: Set1 binds to phosphorylated CTD enriched with Ser_5 at the promoter region through the mediation of the Paf complex (Krogan et al. 2003; Ng et al. 2003); whereas the recognition of the CTD by Set2 relies on the phosphorylation of Ser_2 in heptad repeats (Xiao et al. 2003). The SRI domains from both human (Li et al. 2005) and yeast (Vojnic et al. 2006) Set2 were determined by NMR and the binding interface mapped by phosphoryl-peptide titration. The study of affinity with peptides that have different length and phosphorylation registration showed that single repeats are not sufficient for SRI recognition and a strong preference for doubly phosphorylated peptide was observed. Five residues were identified as essential for the effective association between Set2 SRI and the CTD using NMR and mutagenesis (Krogan et al. 2003). One can assume the positively charged residues, Lys54, Arg58, and His62 of the Set2 SRI, are involved in phosphate binding

whereas Val31 and Phe53 are important for hydrophobic interaction with the side chain residues of the CTD, possibly tyrosine or proline.

In addition, the FF domain is a protein-protein interaction module existing in several CTD interaction proteins such as human transcription factor CA150 (Carty et al. 2000) and splicing factor Prp40 (Morris et al. 2000). Like the SRI domain, the FF domain exhibits a structure of three-helical bundle (**Figure 2-2d**). The FF domains usually are arranged as tandem arrays in proteins and such architecture might account for interaction with the CTD with each module contributing very weakly to binding. Identification of the specificity of phosphorylation states of the CTD recognized by FF domains has been challenging due to the weak binding. *In vitro* binding assays and NMR titration experiments did not detect interaction with a CTD peptide with the FF domain arrays derived from Prp40 (Gasch et al. 2006) or CA150 (Murphy et al. 2009). A binding constant of 50 μM was reported for the mammalian homologue of Prp40, FBP11, using isothermal titration calorimetry (Allen et al. 2002) but detailed information about the interaction between molecules is yet to be elucidated.

CONCLUSION

Recognition of the phosphorylation states of the CTD is essential for mediation of the expression of genetic information. Proteins that are highly selective toward specific phosphorylated forms of the CTD decipher the 'CTD code' and synchronize the transcriptional events accordingly. Communication of the CTD and histone codes provides an exquisite regulatory network for the precise control of transcription at multiple levels. Some CTD interacting proteins function as global transcriptional regulators non-discriminatingly. The inactivation of such molecules will lead to the shutdown of transcription machinery and eventually, cell death. For example, the deletion

of the CTD Ser_2 phosphatase Fcp1 is lethal for yeast due to the abolishment of RNA polymerase II recycling (Kobor et al. 1999; Kimura et al. 2002). Similar effects on yeast are observed when the CTD Ser_5 phosphatase Ssu72 is eliminated (Sun et al. 1996). On the other hand, evidence has indicated that CTD-interacting proteins can control gene expression at an epigenetic level and regulate expression of specific genes based on timing and the developmental needs. Human Scps are only found in higher eukaryotes and their expression is limited to neuronal stem cells or nonneural tissues. Consistent with their unique expression profile, Scps are identified as a component of the neuronal chromatin remodeling complex, REST (Yeo et al. 2005). The inactivation of Scp activity leads to the inappropriate differentiation of neuronal stem cells (Yeo et al. 2005). Another example of epigenetic regulation of a CTD-interacting protein is the Cdk8/cyclin C pair, which has been linked to transcriptional repression (Hengartner et al. 1998). Histone methyltransferases Set1 (Ng et al. 2003) and Set2 (Krogan et al. 2003) can identify the different phosphorylation stages of CTD. The phosphorylation of Ser_7 of CTD itself also occurs in a gene-specific fashion (Egloff et al. 2007). The apparent simple primary sequence of CTD integrates genetic and epigenetic information and plays a pivotal role in transcriptional activation or repression. Understanding how such coding is deciphered at the molecular level is essential to the interpretation of the central role of RNA polymerase II and its regulatory domains. The application of the CTD code provides a unique opportunity to engineer the transcriptional process in an epigenetic level in tissue-specificity manner.

REFERENCES

Akhtar, M. S., M. Heidemann, J. R. Tietjen, D. W. Zhang, R. D. Chapman, D. Eick and A. Z. Ansari (2009). "TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II." *Mol Cell* **34**(3): 387-393.

Allen, M., A. Friedler, O. Schon and M. Bycroft (2002). "The structure of an FF domain from human HYPA/FBP11." *J Mol Biol* **323**(3): 411-416.

Archambault, J., R. S. Chambers, M. S. Kobor, Y. Ho, M. Cartier, D. Bolotin, B. Andrews, C. M. Kane and J. Greenblatt (1997). "An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae." *Proc Natl Acad Sci U S A* **94**(26): 14300-14305.

Baumli, S., G. Lolli, E. D. Lowe, S. Troiani, L. Rusconi, A. N. Bullock, J. E. Debreczeni, S. Knapp and L. N. Johnson (2008). "The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation." *Embo J* **27**(13): 1907-1918.

Becker, R., B. Loll and A. Meinhart (2008). "Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II." *J Biol Chem* **283**(33): 22659-22669.

Bedford, M. T., R. Reed and P. Leder (1998). "WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif." *Proc Natl Acad Sci U S A* **95**(18): 10602-10607.

Bedford, M. T., D. Sarbassova, J. Xu, P. Leder and M. B. Yaffe (2000). "A novel pro-Arg motif recognized by WW domains." *J Biol Chem* **275**(14): 10359-10369.

Carty, S. M., A. C. Goldstrohm, C. Sune, M. A. Garcia-Blanco and A. L. Greenleaf (2000). "Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II." *Proc Natl Acad Sci U S A* **97**(16): 9015-9020.

Chen, H. I., A. Einbond, S. J. Kwak, H. Linn, E. Koepf, S. Peterson, J. W. Kelly and M. Sudol (1997). "Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands." *J Biol Chem* **272**(27): 17070-17077.

26

Collet, J. F., V. Stroobant, M. Pirard, G. Delpierre and E. Van Schaftingen (1998). "A new class of phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX(T/V) motif." *J Biol Chem* **273**(23): 14107-14112.

Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick and S. Murphy (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." *Science* **318**(5857): 1777-1779.

Ermekova, K. S., N. Zambrano, H. Linn, G. Minopoli, F. Gertler, T. Russo and M. Sudol (1997). "The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of Drosophila enabled." *J Biol Chem* **272**(52): 32869-32877.

Etzkorn, F. A. (2006). "Pin1 flips Alzheimer's switch." *ACS Chem Biol* **1**(4): 214-216.

Fabrega, C., V. Shen, S. Shuman and C. D. Lima (2003). "Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II." *Mol Cell* **11**(6): 1549-1561.

Fouillen, L., W. Abdulrahman, D. Moras, A. Van Dorsselaer, A. Poterszman and S. Sanglier-Cianferani (2010). "Analysis of recombinant phosphoprotein complexes with complementary mass spectrometry approaches." *Anal Biochem* **407**(1): 34-43.

Gasch, A., S. Wiesner, P. Martin-Malpartida, X. Ramirez-Espain, L. Ruiz and M. J. Macias (2006). "The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains." *J Biol Chem* **281**(1): 356-364.

Gerber, M. and A. Shilatifard (2003). "Transcriptional elongation by RNA polymerase II and histone methylation." *J Biol Chem* **278**(29): 26303-26306.

Ghosh, A., S. Shuman and C. D. Lima (2008). "The structure of Fcp1, an essential RNA polymerase II CTD phosphatase." *Mol Cell* **32**(4): 478-490.

Glover-Cutter, K., S. Larochelle, B. Erickson, C. Zhang, K. Shokat, R. P. Fisher and D. L. Bentley (2009). "TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II." *Mol Cell Biol* **29**(20): 5455-5464.

Gudipati, R. K., T. Villa, J. Boulay and D. Libri (2008). "Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice." *Nat Struct Mol Biol* **15**(8): 786-794.

Hampsey, M. and D. Reinberg (2003). "Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation." *Cell* **113**(4): 429-432.

Hausmann, S., H. Erdjument-Bromage and S. Shuman (2004). "Schizosaccharomyces pombe carboxyl-terminal domain (CTD) phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping." *J Biol Chem* **279**(12): 10892-10900.

Hengartner, C. J., V. E. Myer, S. M. Liao, C. J. Wilson, S. S. Koh and R. A. Young (1998). "Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases." *Mol Cell* **2**(1): 43-53.

Hollingworth, D., C. G. Noble, I. A. Taylor and A. Ramos (2006). "RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11." *RNA* **12**(4): 555-560.

Huang, K., K. D. Johnson, A. G. Petcherski, T. Vandergon, E. A. Mosser, N. G. Copeland, N. A. Jenkins, J. Kimble and E. H. Bresnick (2000). "A HECT domain ubiquitin ligase closely related to the mammalian protein WWP1 is essential for Caenorhabditis elegans embryogenesis." *Gene* **252**(1-2): 137-145.

Jager, M., Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele and J. W. Kelly (2006). "Structure-function-folding relationship in a WW domain." *Proc Natl Acad Sci U S A* **103**(28): 10648-10653.

Kamenski, T., S. Heilmeier, A. Meinhart and P. Cramer (2004). "Structure and mechanism of RNA polymerase II CTD phosphatases." *Mol Cell* **15**(3): 399-407.

Kimura, M., H. Suzuki and A. Ishihama (2002). "Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in Schizosaccharomyces pombe involves direct interaction between Fcp1 and the Rpb4 subunit of pol II." *Mol Cell Biol* **22**(5): 1577-1588.

Kobor, M. S., J. Archambault, W. Lester, F. C. Holstege, O. Gileadi, et al. (1999). "An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae." *Mol Cell* **4**(1): 55-62.

Komuro, A., M. Saeki and S. Kato (1999). "Npw38, a novel nuclear protein possessing a WW domain capable of activating basal transcription." *Nucleic Acids Res* **27**(9): 1957-1965.

Kornberg, R. D. and Y. Lorch (1999). "Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome." *Cell* **98**(3): 285-294.

Kouzarides, T. (2007). "Chromatin modifications and their function." *Cell* **128**(4): 693-705.

Kowalski, J. A., K. Liu and J. W. Kelly (2002). "NMR solution structure of the isolated Apo Pin1 WW domain: comparison to the x-ray crystal structures of Pin1." *Biopolymers* **63**(2): 111-121.

Krogan, N. J., M. Kim, A. Tong, A. Golshani, G. Cagney, et al. (2003). "Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II." *Mol Cell Biol* **23**(12): 4207-4218.

Li, M., H. P. Phatnani, Z. Guan, H. Sage, A. L. Greenleaf and P. Zhou (2005). "Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1." *Proc Natl Acad Sci U S A* **102**(49): 17636-17641.

Lolli, G., E. D. Lowe, N. R. Brown and L. N. Johnson (2004). "The crystal structure of human CDK7 and its protein recognition properties." *Structure* **12**(11): 2067-2079.

Lu, H., L. Zawel, L. Fisher, J. M. Egly and D. Reinberg (1992). "Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II." *Nature* **358**(6388): 641-645.

Lu, K. P. (2004). "Pinning down cell signaling, cancer and Alzheimer's disease." *Trends Biochem Sci* **29**(4): 200-209.

Lu, K. P., S. D. Hanes and T. Hunter (1996). "A human peptidyl-prolyl isomerase essential for regulation of mitosis." *Nature* **380**(6574): 544-547.

Lu, P. J., G. Wulf, X. Z. Zhou, P. Davies and K. P. Lu (1999). "The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein." *Nature* **399**(6738): 784-788.

Lu, P. J., X. Z. Zhou, Y. C. Liou, J. P. Noel and K. P. Lu (2002). "Critical role of WW domain phosphorylation in regulating phosphoserine binding activity and Pin1 function." *J Biol Chem* **277**(4): 2381-2384.

Lu, P. J., X. Z. Zhou, M. Shen and K. P. Lu (1999). "Function of WW domains as phosphoserine- or phosphothreonine-binding modules." *Science* **283**(5406): 1325-1328.

Macias, M. J., M. Hyvonen, E. Baraldi, J. Schultz, M. Sudol, M. Saraste and H. Oschkinat (1996). "Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide." *Nature* **382**(6592): 646-649.

Meinhart, A. and P. Cramer (2004). "Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors." *Nature* **430**(6996): 223-226.

Morris, D. P. and A. L. Greenleaf (2000). "The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II." *J Biol Chem* **275**(51): 39935-39943.

Morris, D. P., H. P. Phatnani and A. L. Greenleaf (1999). "Phospho-carboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3'-End formation." *J Biol Chem* **274**(44): 31583-31587.

Murphy, J. M., D. F. Hansen, S. Wiesner, D. R. Muhandiram, M. Borg, et al. (2009). "Structural studies of FF domains of the transcription factor CA150 provide insights into the organization of FF domain tandem arrays." *J Mol Biol* **393**(2): 409-424.

Myers, J. K., D. P. Morris, A. L. Greenleaf and T. G. Oas (2001). "Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1." *Biochemistry* **40**(29): 8479-8486.

Ng, H. H., F. Robert, R. A. Young and K. Struhl (2003). "Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity." *Mol Cell* **11**(3): 709-719.

Noble, C. G., D. Hollingworth, S. R. Martin, V. Ennis-Adeniran, S. J. Smerdon, G. Kelly, I. A. Taylor and A. Ramos (2005). "Key features of the interaction between Pcf11 CID and RNA polymerase II CTD." *Nat Struct Mol Biol* **12**(2): 144-151.

Palancade, B., N. F. Marshall, A. Tremeau-Bravard, O. Bensaude, M. E. Dahmus and M. F. Dubois (2004). "Dephosphorylation of RNA polymerase II by CTD-phosphatase FCP1 is inhibited by phospho-CTD associating proteins." *J Mol Biol* **335**(2): 415-424.

Ranganathan, R., K. P. Lu, T. Hunter and J. P. Noel (1997). "Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent." *Cell* **89**(6): 875-886.

Rondon, A. G., H. E. Mischo and N. J. Proudfoot (2008). "Terminating transcription in yeast: whether to be a 'nerd' or a 'rat'." *Nat Struct Mol Biol* **15**(8): 775-776.

Shim, E. Y., A. K. Walker, Y. Shi and T. K. Blackwell (2002). "CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo." *Genes Dev* **16**(16): 2135-2146.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." *Nature* **403**(6765): 41-45.

Sudol, M. and T. Hunter (2000). "NeW wrinkles for an old domain." *Cell* **103**(7): 1001-1004.

Sun, Z. W. and M. Hampsey (1996). "Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo." *Mol Cell Biol* **16**(4): 1557-1566.

Trigon, S., H. Serizawa, J. W. Conaway, R. C. Conaway, S. P. Jackson and M. Morange (1998). "Characterization of the residues phosphorylated in vitro by different C-terminal domain kinases." *J Biol Chem* **273**(12): 6769-6775.

Vasiljeva, L., M. Kim, H. Mutschler, S. Buratowski and A. Meinhart (2008). "The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain." *Nat Struct Mol Biol* **15**(8): 795-804.

Verdecia, M. A., M. E. Bowman, K. P. Lu, T. Hunter and J. P. Noel (2000). "Structural basis for phosphoserine-proline recognition by group IV WW domains." *Nat Struct Biol* **7**(8): 639-643.

Vojnic, E., B. Simon, B. D. Strahl, M. Sattler and P. Cramer (2006). "Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription." *J Biol Chem* **281**(1): 13-15.

West, M. L. and J. L. Corden (1995). "Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations." *Genetics* **140**(4): 1223-1233.

Wu, X., A. Chang, M. Sudol and S. D. Hanes (2001). "Genetic interactions between the ESS1 prolyl-isomerase and the RSP5 ubiquitin ligase reveal opposing effects on RNA polymerase II function." *Curr Genet* **40**(4): 234-242.

Wu, X., C. B. Wilcox, G. Devasahayam, R. L. Hackett, M. Arevalo-Rodriguez, M. E. Cardenas, J. Heitman and S. D. Hanes (2000). "The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery." *Embo J* **19**(14): 3727-3738.

Xiao, T., H. Hall, K. O. Kizer, Y. Shibata, M. C. Hall, C. H. Borchers and B. D. Strahl (2003). "Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast." *Genes Dev* **17**(5): 654-663.

Xu, Y. X., Y. Hirose, X. Z. Zhou, K. P. Lu and J. L. Manley (2003). "Pin1 modulates the structure and function of human RNA polymerase II." *Genes Dev* **17**(22): 2765-2776.

Yaffe, M. B., M. Schutkowski, M. Shen, X. Z. Zhou, P. T. Stukenberg, et al. (1997). "Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism." *Science* **278**(5345): 1957-1960.

Yeo, M., S. K. Lee, B. Lee, E. C. Ruiz, S. L. Pfaff and G. N. Gill (2005). "Small CTD phosphatases function in silencing neuronal gene expression." *Science* **307**(5709): 596-600.

Yeo, M., P. S. Lin, M. E. Dahmus and G. N. Gill (2003). "A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5." *J Biol Chem* **278**(28): 26078-26085.

Zhang, M., J. Liu, Y. Kim, J. E. Dixon, S. L. Pfaff, G. N. Gill, J. P. Noel and Y. Zhang (2010). "Structural and functional analysis of the phosphoryl transfer reaction mediated by the human small C-terminal domain phosphatase, Scp1." *Protein Sci* **19**(5): 974 - 986.

Zhang, Y., S. Daum, D. Wildemann, X. Z. Zhou, M. A. Verdecia, et al. (2007). "Structural basis for high-affinity peptide inhibition of human Pin1." *ACS Chem Biol* **2**(5): 320-328.

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." *Mol Cell* **24**(5): 759-770.

Zhang, Z., J. Fu and D. S. Gilmour (2005). "CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11." *Genes Dev* **19**(13): 1572-1580.

Zhou, M., M. A. Halanski, M. F. Radonovich, F. Kashanchi, J. Peng, D. H. Price and J. N. Brady (2000). "Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription." *Mol Cell Biol* **20**(14): 5077-5086.

SMALL C-TERMINAL DOMAIN PHOSPHATASE (SCP)

The unstructured C-terminal domain (CTD) of eukaryotic RNA polymerase II dynamically regulates the process of transcription by recruiting different factors to the nascent mRNA through its multiple phosphorylation patterns. A newly discovered class of phosphatases, the human small C-terminal domain phosphatases (Scps), specifically dephosphorylate phosphorylated Ser_5 (phospho.Ser_5) of the tandem heptad repeats of the CTD of RNA polymerase II. Scps also function as transcription regulators that epigenetically silence the expression of specific neuronal genes. Inactivation of Scps leads to neuronal stem cell differentiation. Therefore, Scps are a group of important enzymes not only as regulators of the dynamic transcription cycles, but also as potential targets for the treatment of neurodegenerative diseases. This section will reveal to the readers the detailed stories about this group of enzymes (focusing on Scp1), from their catalytic mechanism, to the selective inhibition of these enzymes, to the effort on unveiling how these enzymes read the 'CTD code'.

Chapter 3: Structural and Functional Analysis of the Phosphoryl Transfer Reaction Mediated by Scp1

INTRODUCTION

The CTD of RNA polymerase II is a unique structure that not only temporally and spatially modulates the progression of RNA polymerase II through the transcription cycle, but also associates mRNA transcription to mRNA processing, DNA repair and other cellular processes (Egloff et al. 2008). The biological functions of the CTD are performed through controlling the recruitment of regulatory factors in accordance with the dynamic modifications and conformation changes of the CTD (Meinhart et al. 2005), including phosphorylation, isomerization, and possibly glycosylation (Egloff et al. 2008).

The CTD largely consists of tandem heptad repeats with a consensus sequence $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$ where the phosphorylation of Ser_2, Ser_5 and Ser_7 has been shown to have important regulatory effects on the function of RNA polymerase II (Meinhart et al. 2005; Chapman et al. 2007; Egloff et al. 2007). The enzymes responsible for the phosphorylation and dephosphorylation of these residues therefore play essential roles in the regulation of transcription and related processes.

The transcription factor IIF (TFIIF)-interacting C-terminal domain phosphatase 1 (Fcp1) is the first discovered CTD-specific phosphatase that preferentially dephosphorylates Ser_2 (Chambers et al. 1994; Cho et al. 2001) in eukaryotes. Fcp1 has been shown to regulate the recycling of RNA polymerase II by dephosphorylating Ser_2 after each round of transcription (Cho et al. 1999). More recently, a family of small CTD phosphatases (Scps) with activities preferential for phosphoryl-Ser_5 (phospho.Ser_5) was identified (Yeo et al. 2003; Yeo et al. 2005). This family includes three highly similar proteins designated Scp1, Scp2 and Scp3. They share more than 90% similarity at the catalytic domain and have essentially the same activity to dephosphorylate phospho.Ser_5.

In humans, Scp1 has ~20% identity with the catalytic domain (FCPH domain) of Fcp1 but lacks the C-terminal breast cancer protein related C-terminal domain (BRCT) which is associated with CTD recognition by Fcp1. Scps show a strong preference towards Ser_5 over Ser_2 by 70-fold (Zhang et al. 2006). The first clue to the specificity of Scps was elucidated when the structures of Scp1 bound to the CTD peptides were determined (Zhang et al. 2006). These structures revealed a unique mode for CTD recognition and provided a structural explanation for the preferential binding and turnover of phospho.Ser_5 of the CTD. Consistent with the known specificity of Scp1 towards phospho.Ser_5 instead of phospho.Ser_2, in the complex structure the phospho.Ser_5 is bound to the active site whereas phospho.Ser_2 flips out of the active site, making no

36

direct interaction with the protein. Notably, Scp1 shows remarkable specificity toward the trans peptide-bond configuration of the two prolines in the CTD repeat which can adopt both cis and trans configurations. Such configuration switching is known to modulate the structure of the CTD and its accessibility to kinases and/or phosphatases (Morris et al. 1999; Verdecia et al. 2000). Computational docking shows that if a similar active site conformation is adopted, the peptide with cis Pro will result in an unfavorable steric clash with the protein. However, it has yet to be established if Scps can undergo a dramatic conformational change to accommodate cis Pro.

Unlike kinases that evolved from the same ancestor, the protein phosphatases derived from separate origins which results in different structures and reaction mechanisms (Tonks 2006). Protein tyrosine phosphatases mediate the phosphoryl transfer via a phosphoryl-protein intermediate at a conserved Cys residue at the active site (Zhang 2003). On the other hand, protein serine/threonine (Ser/Thr) phosphatases normally require metal ions for their associated reactions. The Ser/Thr phosphatase family is further divided into phosphoprotein phosphatase (PPP) and Mg^{2+}- or Mn^{2+}-dependent protein phosphatase (PPM) families with different metal requirements (Cohen 1997; Barford et al. 1998). Nonetheless, in both cases, the dephosphorylation reaction occurs in a single step in which a water molecule activated by metal ions targets the leaving phosphate group (Shi 2009). One of the major distinctions between tyrosine phosphatases and Ser/Thr phosphatases is the involvement of metal ions (Ser/Thr phosphatases) and the generation of phosphoryl intermediate (tyrosine phosphatases). Therefore, when a group of novel human CTD phosphatases (Fcp/Scp phosphatases) were identified, they were classified as the PP2C subfamily of the PPM Ser/Thr phosphatase family (Chambers et al. 1996; Yeo et al. 2003). Like other PP2C phosphatases, Fcp/Scp phosphatases are Mg^{2+}-dependent enzymes with multiple highly conserved aspartates.

However, when the Scp structure was determined (Kamenski et al. 2004), it exhibited several fundamental distinctions from other PP2C members. First, it lacks the di-metal center which is present in other PP2C Ser/Thr phosphatases but has a single Mg^{2+} ion bound at the active site (**Figure 3-1**). Moreover, the four strictly conserved aspartates are arranged in a different topology from PP2C active sites.
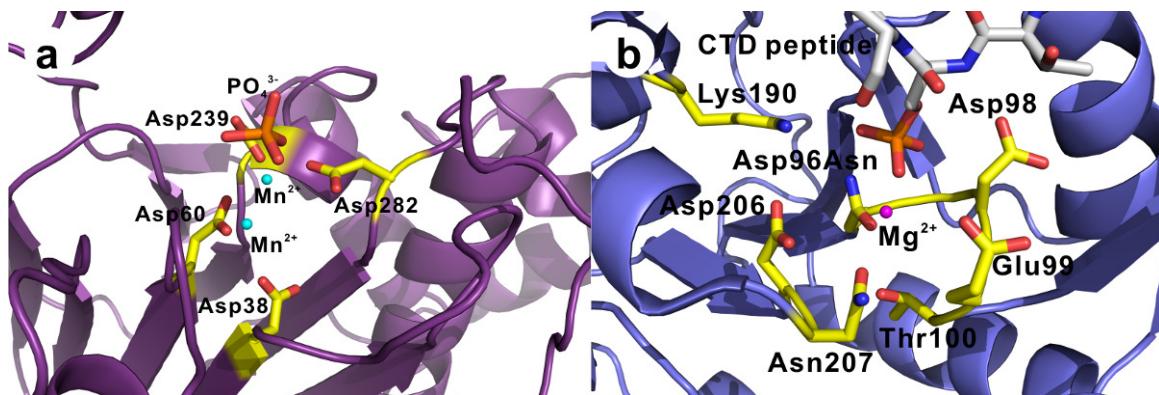
Figure 3-1:   Ribbon diagram of metal centers in human protein phosphatase 2C (1A6Q) and human Scp1 (2ghq).[6]

Based on the structure-based sequence alignment, Scp is similar to the HAD superfamily proteins, the majority of which are involved in phosphoryl transfer (Allen et al. 2004). Indeed, with little primary sequence identity, the active site of Scp superimposes perfectly with that of other HAD members (Wang et al. 2001; Lahiri et al. 2003; Kamenski et al. 2004). Even the Mg^{2+} ion position and its coordinating water molecules are conserved (Zhang et al. 2006). This suggests that Scp is a unique family of Ser/Thr phosphatases that mediates phosphoryl transfer through the generation of a phosphoryl-aspartate using its DXDX(T/V) signature motif, a mechanism conserved in

---

[6] (a) Di-metal ion center of PP2C. Invariant metal-coordinating residues (yellow) and phosphate group (red) are shown in stick, and the two catalytic essential Mn^{2+} ions are shown in cyan sphere. (b) Single-metal ion center of Scp1. Key residues at the active site are shown in yellow stick, and the essential Mg^{2+} ion is shown in purple sphere. The substrate, phosphorylated CTD peptide, is shown in white stick.

the HAD family. Using BeF$_3^-$ as a phosphoryl analogue, the first Asp of the motif was identified to be the nucleophile that undergoes phosphorylation leading to a phosphoryl-aspartate intermediate. This high-energy mixed anhydride is subsequently hydrolyzed to regenerate the enzyme (Allen et al. 2004). Although this proposed mechanism is well accepted, the phosphoryl intermediate was never directly visualized in Fcp/Scp family.

Considering the uniqueness of Fcp/Scp phosphatases in their phosphoryl reaction mechanism as a Ser/Thr phosphatase, we sought to isolate the phosphoryl intermediate combining mutagenesis and X-ray crystallography. We extend our previous structural and functional investigations to further probe the role of key residues in the dephosphorylation reaction. Steady-state kinetic analysis of a variety of Scp1 mutants verified the Mg^{2+}-coordinating function of Asp206, a conserved residue in the HAD superfamily whose role was not highlighted before. We have successfully captured the phosphoryl-aspartate intermediate in the crystal structure of Scp1D206A mutant soaked with *p*NPP, providing evidence for the proposed two-step mechanism. In this study, we were able to obtain the snapshots of Scp protein at each stage of the phosphoryl-transfer reaction which will be really valuable for the structure-function study of Fcp/Scp phosphatases.

RESULTS AND DISCUSSION

Design of mutants based on proposed phosphoryl transfer mechanism

Based on the structures of Scp1, a two-step reaction mechanism for the dephosphorylation reaction catalyzed by Scp1 was proposed (Zhang et al. 2006). In this proposed mechanism (**Figure 3-2a**), the Asp96 [the first Asp in the DXDX(T/V) signature motif] acts as the attacking nucleophile in the first step of the reaction where the phosphoryl group on phospho.Ser$_5$ is transferred to Asp96, forming a phosphoryl-

39

carboxyl mixed anhydride, which is subsequently hydrolyzed in the second step to release the phosphate and regenerate Asp96. Phosphoryl analogues have been used to understand the reaction mechanism (Kamenski et al. 2004). Most recently, a study conducted on *Schizosaccharomyces pombe* Fcp1 captured the mimic structures of the phosphoryl-aspartate intermediate and the transition state of the hydrolysis step using BeF_3^- and AlF_4^- respectively (Ghosh et al. 2008). In the Fcp1 structure, a water molecule occupies the position close to the Asp172 (corresponding to Asp98 in human Scp1) and can potentially be activated by Asp172 side chain for the breakdown of the phosphoryl-intermediate. This observation is consistent with the structural studies on phosphoserine phosphatase (PSP) (Wang et al. 2002). Based on the similarity between Fcp1 and Scp1, it is likely that Asp98 functions as the general acid to protonate the leaving group and as the general base to activate a water molecule in the first and second steps, respectively. However, structural or biochemical data that directly test this model in Scp1 are lacking.

One of the most efficient ways to understand the catalytic mechanism of an enzyme is to capture the intermediate of the reaction. It has been shown in an earlier study that a transient phosphoryl-cysteine intermediate was captured in the protein tyrosine phosphatase 1B crystal structure by generating an active site mutation at Gln262 to Ala (Pannifer et al. 1998). However, although the two-step mechanism of the dephosphorylation reaction (**Figure 3-2a**) by the HAD superfamily enzymes is generally accepted (Allen et al. 2004), the phosphoryl-aspartate intermediate has rarely been shown structurally (Lahiri et al. 2002). In theory, there are two scenarios where the phosphoryl-aspartate intermediate may be captured in crystal structure. First, if the general acid that protonates the product (in the first step, **Figure 3-2a**) is a different residue from the general base that catalyzes the hydrolysis of the phosphoryl-aspartate intermediate (the second step, **Figure 3-2a**) then mutation of the latter residue would result in the

accumulation of the intermediate. Second, if the general acid and the general base are the same residue, but the mutation of an auxiliary residue (e.g. a Mg^{2+}-coordinating residue) slows down the rate of both the first and the second reactions, the overall reaction rate might be low enough so that the added substrate is not completely consumed during the process of crystallization. In this case if the rate of the second step (k_2, **Figure 3-2a**) is not much higher than that of the first step (k_1, **Figure 3-2a**), a substantial fraction of the enzyme ($\dfrac{k_1}{k_1 + k_2}$) should exist as the phosphoryl-aspartate intermediate.

   With this reasoning in mind, we designed a series of Scp1 mutants where each of the acidic residues (namely, Asp98, Asp99 and Asp206) was mutated to either the isosteric neutral residue (i.e. Asn) or Ala, examined these mutants with biochemical assays and chose suitable ones for crystallographic investigation in order to capture the phosphoryl intermediate.
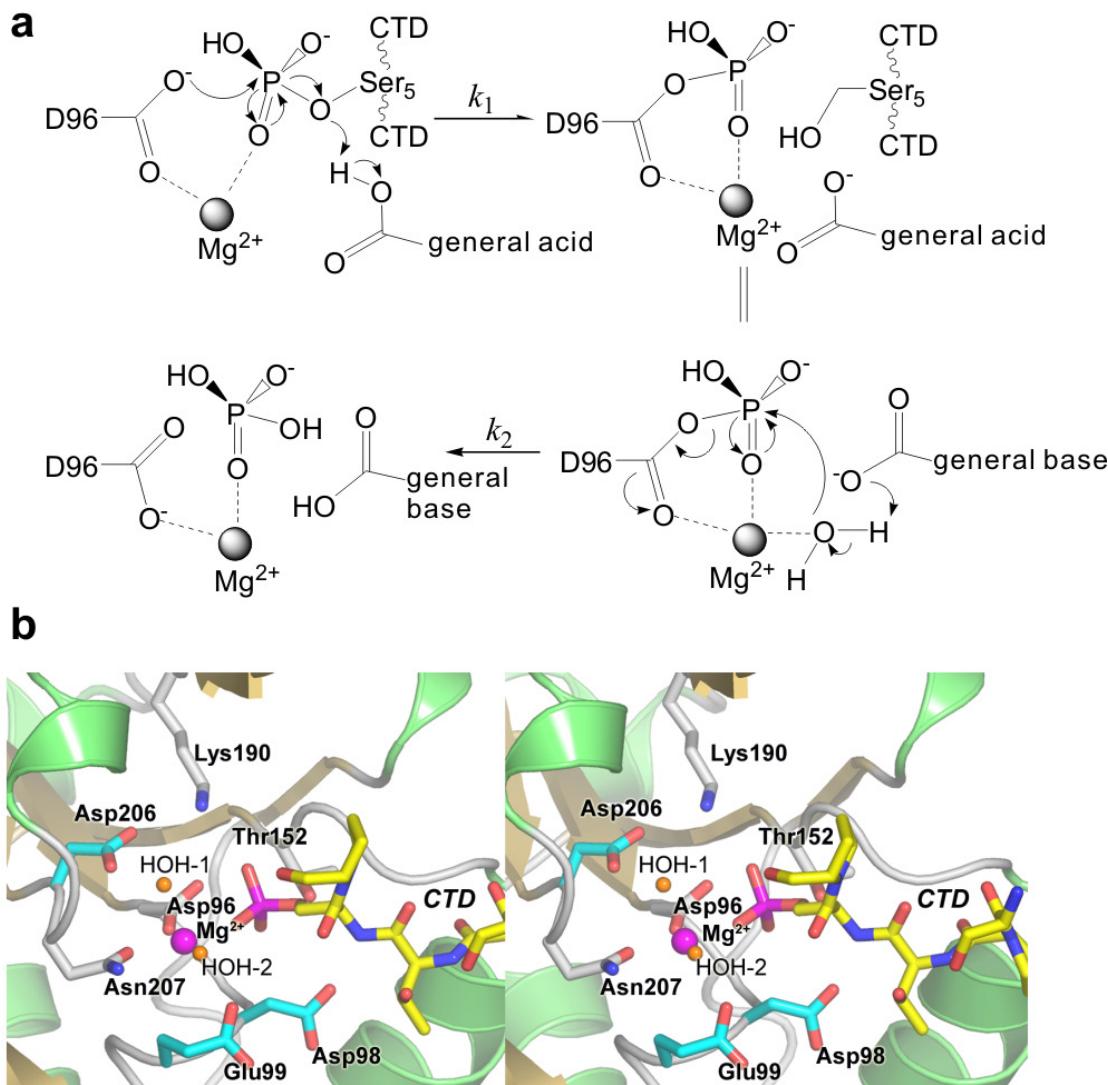
Figure 3-2: Proposed mechanism of phospho.Ser$_5$ dephosphorylation catalyzed by Scp1.[7]

---

[7] (a) Proposed mechanism of phospho.Ser$_5$ dephosphorylation catalyzed by Scp1. At early steady-state, the concentration of product has not accumulated to a significant level so that the reverse reaction of the first step can be ignored. (b) Stereo diagram of the CTD-bound active site of Scp1D96N. The structure of human Scp1D96N in complex with a CTD heptad repeat peptide phosphorylated at Ser$_5$ shown as a stereo pair of ribbon diagram. The active site residues are depicted as color-coded bonds. The three residues serving as candidate general bases for hydrolysis of the phosphoryl-Asp96 intermediate are highlighted with cyan colored bonds.

Phosphatase activity assays for Scp1 mutants

We used two complementary methods to measure the general dephosphorylation activity of Scp1 mutants: a colorimetric assay with pNPP as the substrate and a fluorescent assay using 6,8-difluoro-4-methylumbelliferyl phosphate (DiFMUP) as the substrate. The pNPP assay is highly robust but shows a strong background absorbance, resulting in insensitivity with an enzyme with low activity. On the other hand, the fluorescence-based DiFMUP assay is highly responsive towards dephosphorylation and can detect very low phosphatase activity. However, high concentrations of DiFMUP appeared to non-specifically inhibit the activities of all tested variants of Scp1. The inhibitory effect was evident at concentrations of DiFMUP around 500 μM and was nearly complete at higher concentrations of DiFMUP (>5 mM). For this reason, the highest concentration of DiFMUP used in our assay was 150 μM which is below the K_m of all active Scp1 variants, making the measurement of K_m and k_{cat} unfeasible. Nonetheless, the k_{cat}/K_m could still be measured in DiFMUP assays and were used to compare relative activities of Scp1 variants. Despite being less sensitive, the pNPP assay allowed high concentrations of substrate (up to 30 mM pNPP) and, when necessary, enabled us to obtain the k_{cat} and K_m for Scp1 variants toward pNPP. Using the two assays, significant information on the catalytic activity of these Scp1 variants was obtained.

As expected, the mutation of the active site nucleophile Asp96 to Asn (D96N) or Ala (D96A) abolished all measurable activity in both assays. Interestingly, although Asp98 was proposed to act as a general acid to protonate the leaving group of the first phosphoryl transfer and subsequently deprotonate a water to accelerate hydrolysis of the anhydride (Kamenski et al. 2004; Zhang et al. 2006), mutation of Asp98 to Asn (D98N) retained 30% activity of the wildtype Scp1 (DiFMUP assay data, **Table 3-1**). Even a D98A mutant showed 10% activity of the wildtype Scp1 (both assays, **Table 3-1**). The

k_{cat} of D98A using pNPP as the substrate was as high as 0.5 s^{-1}. This high k_{cat} is noteworthy because using the model described in Figure 3-2a, it can be deduced that

$$k_{cat} = \frac{k_1}{k_1 + k_2} \cdot k_2 \qquad [1]$$

Or in other words, k_2 must of necessity be greater than k_{cat}. Therefore, the first-order rate constant for the dephosphorylation of D98A mutant (k_2) must be greater than 0.5 s^{-1}, which is at least 4 orders of magnitude higher than the reported uncatalyzed hydrolysis of phosphoryl-aspartate mixed anhydride [10^{-4} s^{-1}, (Allen et al. 2004)]. These results argue against the proposed function of Asp98 as the only possible general acid/base in catalysis, suggesting that complementarity by another active site residue may be applied in Scp-mediated phosphate transfer.

Similarly, mutations of Glu99 to Gln or Ala did not abolish phosphatase activity, resulting in 10% and 5% activity compared to wildtype, respectively. The mutation of Asp206 to Asn resulted in greatly reduced stability and solubility of the protein. The corresponding mutation of Asp to Asn is also reported to be very disruptive for Fcp1 (Hausmann et al. 2003), possibly due to the misfolding of the protein induced by the disruption of the salt bridge between this Asp and active site Lys. Therefore, alanine mutation was tested to investigate the functional role of D206. Under the standard assay condition with 20 mM MgCl$_2$, the activity of D206A mutant was too low to be reliably detected in both assays. The crystal structure of Scp1 (**Figure 3-2b**) shows that an active site Mg^{2+} is coordinated by a water molecule which is further positioned by Asp206. Such secondary stabilization of metal ion binding via water molecules might be essential for the effective catalysis. Indeed, this role of Mg^{2+}-coordination by residues corresponding to D206 is also observed in many other HAD family proteins in three-dimensional structure (see structural alignment later). Consistent with these structural

observations, when the $MgCl_2$ concentration was increased to 50 mM, the D206A mutant displayed a well-detectable activity toward DiFMUP. The k_{cat}/K_m was 0.2 $mM^{-1}min^{-1}$ (~2% of the wildtype activity measured at 20 mM $MgCl_2$).

	DiFMUP assay	pNPP assay		
	k_{cat}/K_m ($mM^{-1}min^{-1}$)	k_{cat} (s^{-1})	K_m (mM)	k_{cat}/K_m ($mM^{-1}s^{-1}$)
Wildtype	10.3 ± 0.8	2.5 ± 0.1	3.6 ± 0.8	0.7 ± 0.2
D96N	N.A.[8]	N.A.[8]	N.A.[8]	N.A.[8]
D96A	N.A.[8]	N.A.[8]	N.A.[8]	N.A.[8]
D98N	3 ± 1	N.A.[9]	N.A.[9]	N.A.[9]
D98A	0.8 ± 0.2	0.50 ± 0.02	7.3 ± 0.8	0.068 ± 0.008
E99Q	1.14 ± 0.09	2.0 ± 0.3	8 ± 4	0.2 ± 0.1
E99A	0.5 ± 0.1	0.005 ± 0.001	N.A.[10]	N.A.[10]
D206A[11]	0.23 ± 0.04	0.022 ± 0.005	8 ± 4	0.002 ± 0.001

Table 3-1:   Steady-state parameters of Scp1 variants tested by DiFMUP and pNPP assays.

To further characterize the Mg^{2+}-dependence of D206A mutant, we tested the catalytic activity of wildtype Scp1 and the D206A mutant at various $MgCl_2$ concentrations (**Figure 3-3**). While the apparent K_d of wildtype Scp1 with Mg^{2+} was 1.3

---

[8] Signal below detection limit.

[9] Although D98N showed strong activity to pNPP, the activity of this mutant appeared unstable so that during the time course required for pNPP assay the enzyme did not maintain constant activity, leading to uninterpretable data.

[10] The K_m of E99A mutant is greatly increased, and thus can not be accurately determined from the rate-vs-pNPP concentration curve.

[11] Both DiFMUP and pNPP assay were performed with 20 mM Mg^{2+}, however, no signal can be detected for D206A mutant under this condition. Therefore, the data for D206A were obtained under 50 mM Mg^{2+} condition.

± 0.9 mM, that of the D206A mutant was at least 100-fold higher. These results indicate that Asp206 is critical during catalysis at least in part by coordinating the catalytic Mg^{2+}.
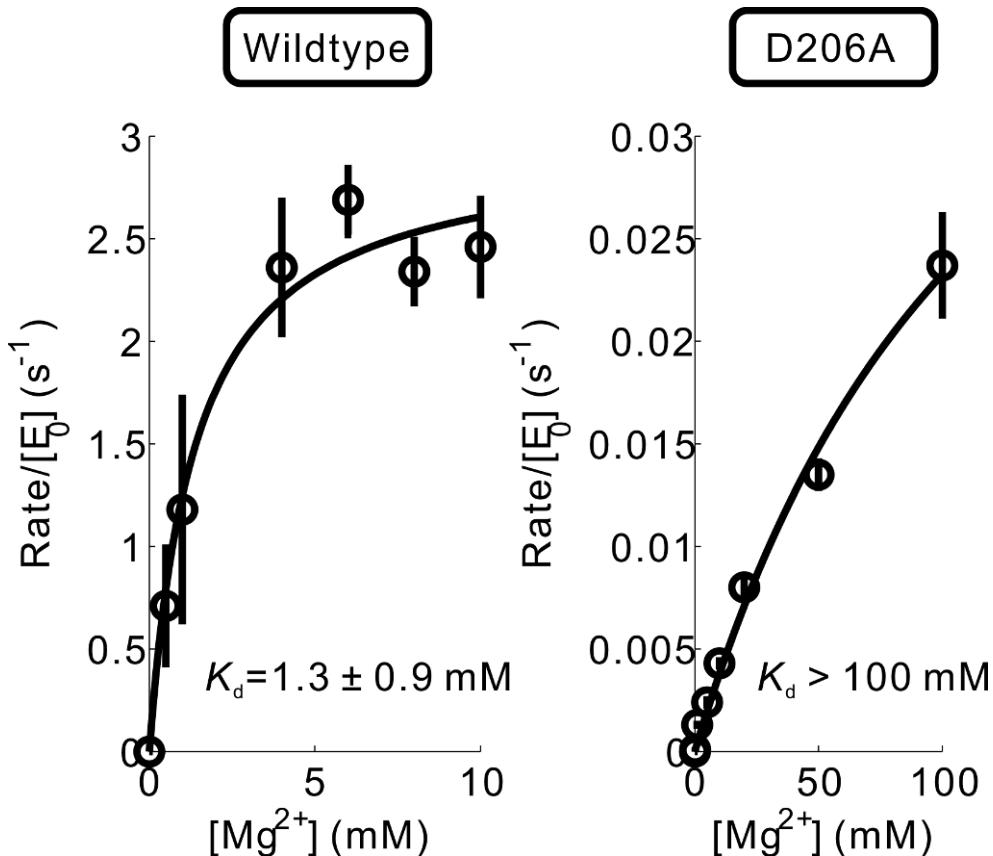
Figure 3-3: Mg^{2+}-dependence of enzyme activity of wildtype Scp1 and Scp1D206A mutant.

Phosphoryl intermediate of Scp1 obtained by incubating Scp1D206A mutant with *p*NPP at low Mg^{2+} concentration

Based on the kinetic characterization above, we crystallized D98A, E99A and D206A and soaked the crystals in the buffer containing a high concentration of *p*NPP. X-ray diffraction data were collected for each mutant at resolutions from 2.35-2.45 Å (**Table 3-2**). In the structure of the Scp1D98A mutant, a hydrogen bond originally observed between the side chain of Asp98 and that of Tyr158 found both in the apo

protein structure (**Figure 3-4**) and our previous complex structures (**Figure 3-5a**) is interrupted due to mutation of Asp98. This hydrogen bond appears to anchor Tyr158 to facilitate the formation of part of the CTD binding groove (Zhang et al. 2006). Without this hydrogen bond anchoring the side chain of Tyr158 by Asp98 side chain, an alternative conformation of Tyr is adopted (**Figure 3-5b**), which places the phenolic side chain of Tyr in the hydrophobic pocket previously observed to specifically sequester Pro_3 of the CTD repeat sequence (**Figure 3-5c**). This hydrophobic pocket prefers binding to hydrophobic molecules, and the adoption of Pro_3 of the CTD repeat is essential for CTD recognition by Scp1. The structure of Scp1E99A exhibits an identical metal center with wildtype Scp1 and Scp1D96N structures (**Figure 3-6**), with the only difference being the coexistence of both conformations of Tyr158 due to a minor change of the Asp98 position. Apparently, the orientation of the Tyr158 is directly affected by its hydrogen bonding to Asp98 carboxyl side chain.
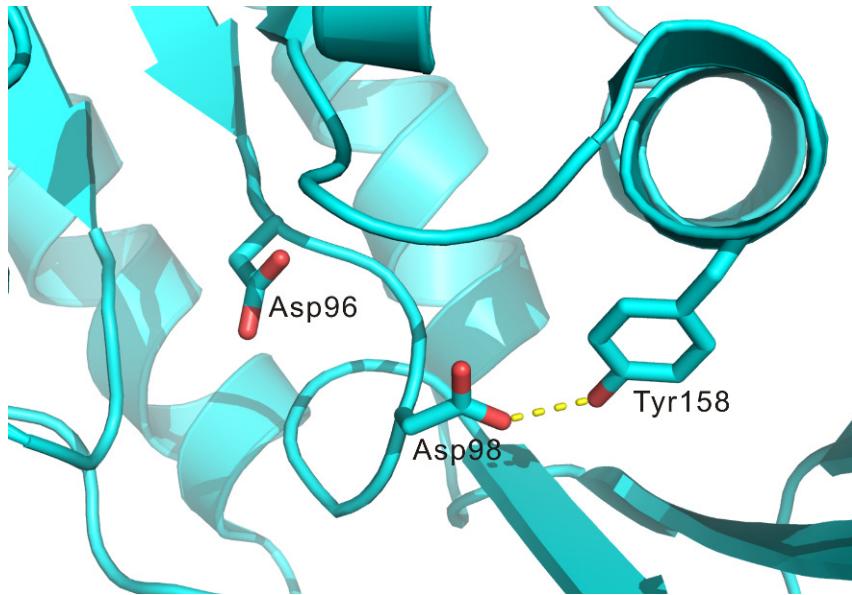
Figure 3-4:   Ribbon diagram showing the salt bridge between Asp98 and Tyr158 in Scp1 apo protein structure.

	Scp1D206A soaked in *p*NPP at low Mg^{2+}	Scp1D206A soaked in *p*NPP at high Mg^{2+}	Scp1D98A soaked in *p*NPP at low Mg^{2+}
Data collection			
Space group	C2	C2	C2
Cell dimensions			
a, *b*, *c* (Å)	124.7, 78.2, 62.6	125.5, 77.3, 63.1	124.8, 79.0, 62.9
α, β, γ (°)	90.0, 112.3, 90.0	90.0, 112.5, 90.0	90.0, 112.8, 90.0
Resolution (Å)	49.1–2.35 (2.43–2.35)	48.6–2.45(2.54–2.45)	48–2.30 (2.38–2.30)
No. of observation	39564	32699	39941
No. of unique reflections	23379	20421	24660
R_{sym} or R_{merge} (%)	7.1 (20.4)	7.5 (22.7)	7.2 (20.5)
I/σ(I)	16.4 (2.6)	13.4 (2.1)	17.5 (3.4)
Completeness (%)	91.3 (59.1)	90.7 (53.2)	90.2 (57.2)
Redundancy	2.4 (1.7)	3.3 (2.1)	3.4 (2.6)
Refinement			
Resolution (Å)	49.1–2.35	48.6–2.45	32.2–2.35
No. reflections (test set)	23286 (1109)	17693 (864)	21099 (1131)
R_{work} / R_{free} (%)	19.5(25.9)/24.5(34.5)	20.4(25.6)/25.4(37.7)	19.2(24.0)/23.1(30.6)
No. of atoms			
Protein	2930	2908	2928
ion	2	12	2
Water	95	32	99
B-factors (Å^2)			
Protein	36.2	37.8	37.4
Mg^{2+}	20.5	39.3	24
Water	37.2	38.2	39.1

Table 3-2:   Data collection and refinement statistics.[12]

---

[12] Highest resolution shell is shown in parenthesis. R$_{free}$ is calculated with 5% of the data randomly omitted from refinement.

(Table 3-2 cont.)

R.m.s deviations			
Bond lengths (Å)	0.013	0.014	0.011
Bond angles (°)	1.48	1.60	1.36
Ramachandran plot (%)			
Most favored	87.8	84.4	90.5
Additionally allowed	11.3	15.3	9.2
Generally allowed	0.9	0.3	0.3
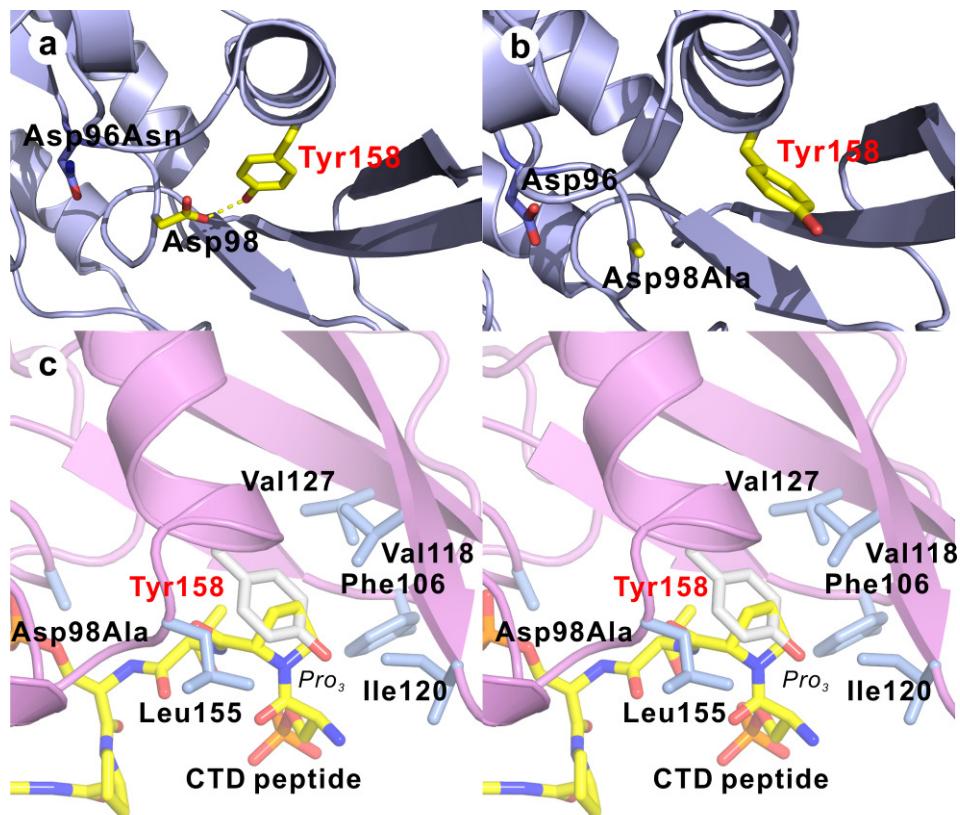
Figure 3-5: Two different conformations of Tyr158 observed in Scp1D96N structure (a) and Scp1D98A structure (b).[13]

---

[13] In (a), the side chain of Tyr158 forms a hydrogen bond (yellow dashed line) with the side chain of Asp98 which is lost in (b). (c) Stereo diagram of superimposition of Scp1D98A structure with Scp1D96N-CTD complex (PDB code: 2ghq). The CTD peptide is shown in yellow stick.
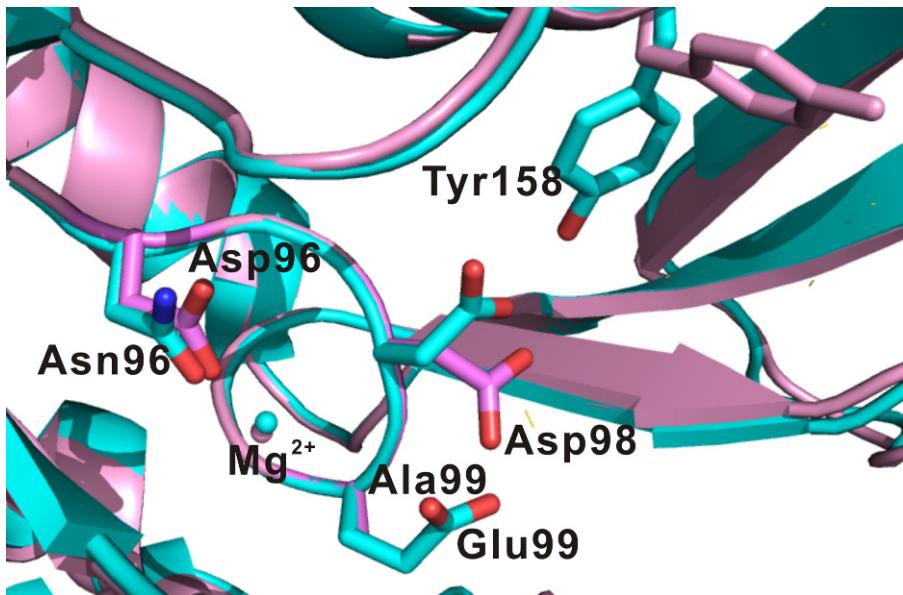
Figure 3-6:   Superimposition of Scp1E99A (magenta) and D96N (cyan) active site structures.[14]

Unlike D98A and E99A structures, the Scp1D206A mutant exhibited strong positive electron density directly contiguous with the electron density associated with the Asp96 residue. The shape and high signal to noise ratio of the contiguous density is consistent with a covalently trapped phosphate group, which was built and refined as a phosphoryl-Asp96 intermediate (**Figure 3-7a, b**). The remaining parts of the overall Scp1D206A structure are virtually identical to the wildtype and D96N Scp1 structures. Additional density resided in the Pro_3 binding pocket that was refined well as a part of PEG molecule in one of the two Scp1D206A monomers in the asymmetric unit. This is consistent with previous observation that this pocket highly favors hydrophobic molecules and might non-specifically bind to chemical molecules in crystallization solution (Kamenski et al. 2004).

---

[14] The structure of E99A mutant exhibits an identical metal center with that of D96N. Notably, the Tyr158 can adopt two conformations as discussed in the paper due to the minor change of Asp98 position.
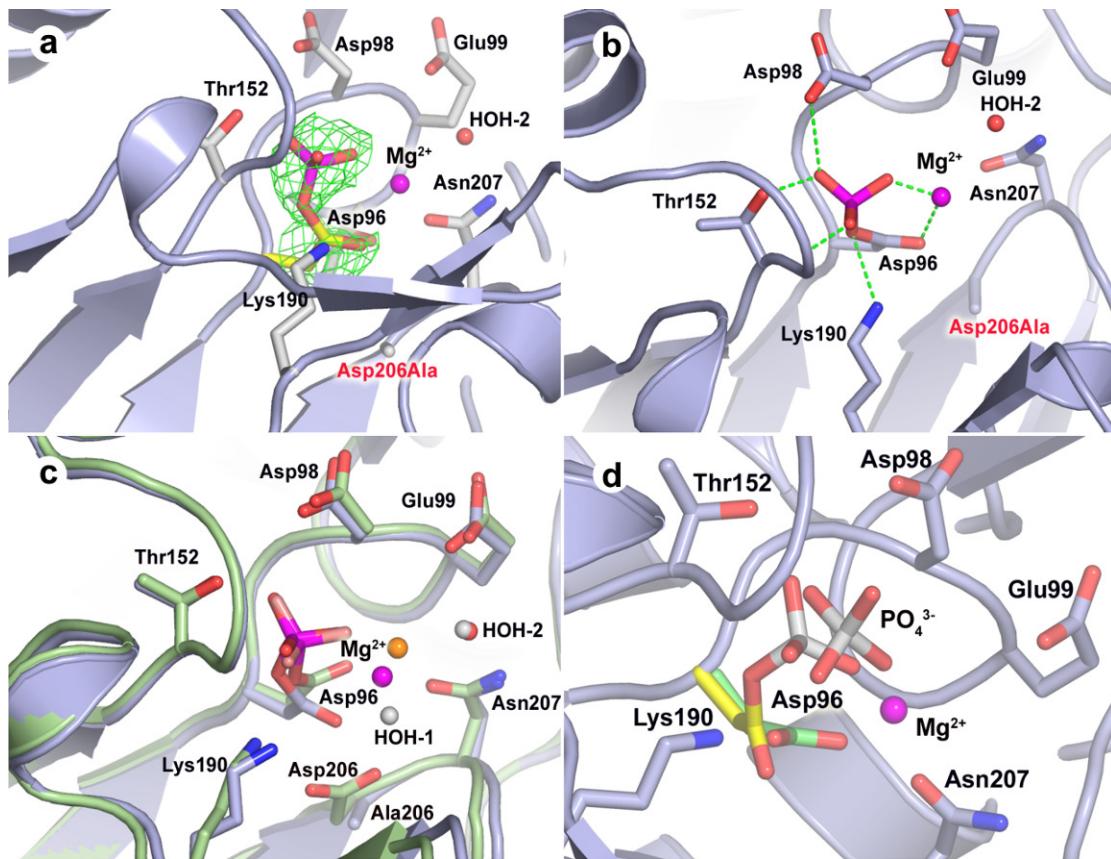
Figure 3-7: Structural studies of Scp1D206A.[15]

Apparently, in the crystalline environment of the Scp1D206A mutant, the high energy phosphoryl-enzyme intermediate bound to Asp96 reaches appreciable levels and remains stable over time to allow visualization by protein X-ray crystallography. Based on the peak height from initial F_o-F_c electron density maps as well as thermal factors from active sites, we estimated that the occupancy of the phosphate group covalently

---

[15] (a) Phosphorylation of Asp96. The mutation of Asp206 (labeled in red) stabilizes the trapped phosphoryl-aspartate intermediate of the enzyme. (b) The hydrogen bonding formed between phosphate group and the enzyme. (c) Superimposition of phosphate-trapping Scp1D206A (light blue) with Scp1 bound with BeF_3^- (1ta0) (lime). The water molecules are colored red and Mg^{2+} ion colored magenta for the former structure, whereas water as white and Mg^{2+} as orange for the latter. (d) Superimposition of Scp1D206A soaked in substrate pNPP at different Mg^{2+} concentrations. The active site residues are shown in stick. The Asp96 covalently linked to phosphate group is shown in yellow, and the phosphoryl-free Asp96 is shown in green.

attached to Asp96 reached approximately 50%. In the crystalline environment of the Scp1D206A mutant, the electron density of Mg^{2+} (**Figure 3-8**) is significantly weaker than that obtained in the Scp1D96N-CTD peptide complex structure, suggesting the Mg^{2+} occupancy is not 100% (refined as 50% occupancy with B factor around 20 $\mathring{A}^2$). This is again consistent with the kinetic study that the binding affinity between Mg^{2+} and Scp1D206A is much lower. Therefore, using lower Mg^{2+} concentration in the crystallization mother liquor could be one of the major reasons why we captured this phosphoryl-enzyme intermediate in Scp1D206A crystal. The crystallization state also sheltered the phosphoryl intermediate from bulk water in the D206A mutant and slowed down the hydrolysis of phosphoryl-Asp96.
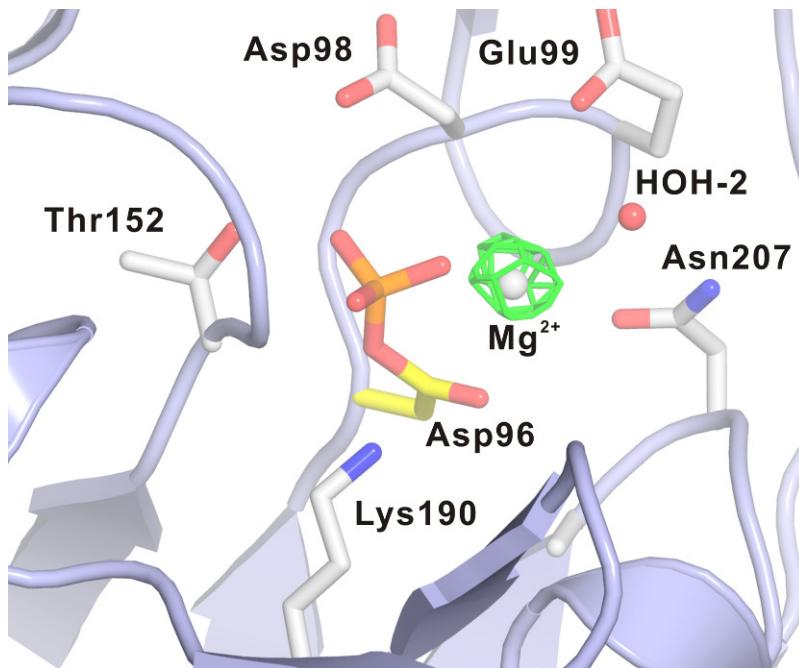
Figure 3-8: F_o-F_c electron-density map (green) of Mg^{2+} ion with a sigma cutoff of 8σ.

A previously reported structure of BeF_3^- in complex with Scp1 suggested the existence of a Scp1 phosphoryl-enzyme intermediate (Kamenski et al. 2004). Because of

the presumed chemical instability of phosphoryl-aspartate intermediates of HAD family proteins, BeF_3^- has been used as a phosphate mimic (Cho et al. 2001). By comparing the structure of our trapped phosphoryl-aspartate intermediate and that of BeF_3^- bound to Scp1 (PDB code: 1ta0), we conclude that the complex structure encompassing BeF_3^- bound to Asp96 is a reasonably acceptable mimic for the covalent phosphoryl-aspartate intermediate (**Figure 3-7c**).

Product-trapping structure of Scp1 obtained by incubating Scp1D206A mutant with *p*NPP at high Mg^{2+} concentration and high pH

Based on the proposed mechanism that a general base is involved in the second step and our kinetic data, we hypothesized that increased Mg^{2+} availability and higher pH would facilitate the hydrolysis of the phosphoryl-aspartate. To evaluate these effects on the activity of Scp1D206A crystallographically, we used the same Scp1D206A crystal form but soaked the samples in 25 mM *p*NPP with 0.2 M $MgCl_2$ at slightly higher pH (pH 6.5–7.0). Clear and unmistakable electron density indicates a bound phosphate group in the Scp1D206A active site. However, unlike the electron density associated with phosphate for the low Mg^{2+} soaked crystals, this electron density clearly showed a complete loss in covalent attachment to the enzyme (**Figure 3-7d**). The phosphate group in the structure was refined to 100% occupancy with the same thermal factors as the surrounding protein residues. The position of the phosphate group moves by approximately 1.2 Å away from its original position in the low Mg^{2+} complex while the side chain of the Asp96 nucleophile rotates away from the phosphate (**Figure 3-7d**). Together with previously reported structures, these newly determined Scp1 crystal structures provide static snapshots along the reaction pathway catalyzed by Scp1 (**Figure 3-9**). Our complex structure of Scp1D96N with CTD phosphoryl-peptide presents the pre-reaction binding mode. The D206A structure soaked with *p*NPP at low Mg^{2+}

53

concentration is the capture of phosphoryl intermediate, whereas the same mutant at high Mg^{2+} concentration is a product trapped image with inorganic phosphate group at the active site.
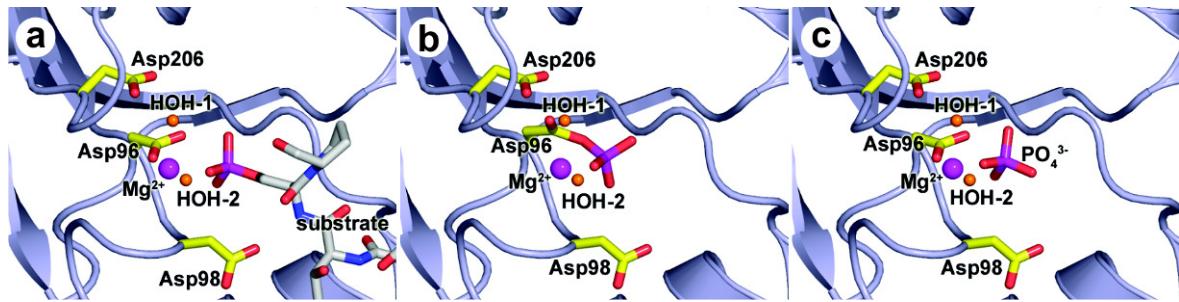
Figure 3-9: Snapshots of dephosphorylation reaction of human Scp1.[16]

Capture of phosphoryl intermediate for HAD superfamily

The formation of phosphoryl-enzyme intermediates is a crucial component of various enzymatic mechanisms involving phosphoryl transfer (Pannifer et al. 1998). Even though phosphoryl-aspartate is not very stable, it has been captured in other proteins containing a DXDX motif (Allen et al. 2004; Diaz et al. 2008), mostly using phosphoryl analogues. In addition to the direct visualization of phosphoryl intermediate in our X-ray structure for Scp1, β-phosphoglucomutase (β-PGM) is found to be phosphorylated at the active site aspartate (Asp8) in the crystalline state (PDB code: 1lvh) (Lahiri et al. 2002). β-PGM is a subclass I HAD family protein, which catalyzes the transfer of a phosphate group between the 1 and the 6 position of the cyclic glucose ring. As the enzyme needs to

---

[16] This is the presumed model for the reaction catalyzed by Scp1 in substrate recognition (a), phosphoryl transfer (b), and phosphate release (c) stages. In order to capture them in crystallization state, certain residues were mutated (see below). Scp1 is shown in ribbon diagram in light blue, and Asp96, Asp98 and Asp206 residues are shown in stick in yellow. Mg^{2+} ion and the two water molecules are shown in sphere in magenta and orange, respectively. (a) Scp1 bound to phosphorylated CTD (carbon chain shown in white). Asp96 was mutated to Asn to capture this complex structure. (b) Phosphate group of phospho.Ser_5 is transferred to Asp96 and generates the phosphoryl-aspartate intermediate. Asp206 was mutated to Ala in this structure. (c) Hydrolysis of the phosphoryl-aspartate intermediate and release of phosphate group from Asp96 side chain. Again, Asp206 was mutated to Ala in this structure.

transfer the phosphate group from position 1 on the hexose to position 6, the phosphoryl-enzyme intermediate is sufficiently stable to allow reorientation of the acceptor in the active site while preventing water-mediated loss of phosphate. Other phosphoryl-aspartate intermediate was investigated where various phosphate analogs (AlF_4^-, MgF_4^{2-} and BeF_3^-) were used to mimic the different stages of phosphoryl transfer. For instance, in Ca^{2+}-ATPase, the ATP-dependent phosphorylation of its catalytic Asp is a critical step in transducing chemical energy into conformational energy for Ca^{2+} ion transport (Toyoshima et al. 2004; Toyoshima et al. 2007). Subsequent to ATP-dependent phosphorylation of the ATPase, the dephosphorylation reaction of the phosphoryl-enzyme intermediate triggers a conformational change that closes the lumenal gate of the Ca^{2+} channel though the movement of a domain capping the transmembrane portal (Toyoshima et al. 2004; Toyoshima et al. 2007). Furthermore, the phosphoryl-aspartate intermediate was observed in *Schizosaccharomyces pombe* Fcp1 structure with BeF_3^- mimicking the phosphate group (PDB code: 3ef0) (Ghosh et al. 2008). In the crystal structure, a covalent bond was clearly observed between the BeF_3^- and Asp170 of Fcp1 (the first Asp in the DXDX motif). Human eyes absent (Eya) proteins, which also belong to the HAD family, play dual roles as a protein tyrosine phosphatase and a transcription factor in organ formation during development. Again, using phosphate analog BeF_3^- and AlF_3, a phosphoryl-aspartate intermediate and a transition sate of phosphoryl-enzyme were captured in the catalytic domain of Eya2 phosphatase, respectively (Jung et al. 2010). Additionally, phosphoryl-aspartate intermediates have been detected by radioactivity measurements in DNA 3' phosphatase Tpp1 (Deshpande et al. 2004), and in a recent study, Ca^{2+} ions were shown to have an inhibitory effect and induce the accumulation of a phosphoryl-aspartate intermediate detected by mass spectrometry in histidinol phosphate phosphatase (Rangarajan et al. 2006).

Conservation of D206 in HAD family

While most HAD family proteins vary considerably in primary sequence, they all share the DXDX(T/V) motif (motif I) (Asp96 and Asp98 in human Scp1) as well as two other motifs at similar positions in their 3D structures (**Figure 3-10a**). Motif II is a single residue, either Thr (Thr152 in human Scp1) or Ser, which hydrogen bonds through its side chain with the phosphate moiety of the substrate. These two motifs are highly conserved and easily identified using structure-based alignment. Motif III is a DD or DN pair with a Lys residue located a short distance away on the N-terminal side of the acidic pair. The first Asp of the DD (DN) pair (Asp206 in human Scp1) aligns a water molecule which coordinates the catalytic Mg^{2+} ion in the active site. This residue also forms a salt bridge to the side chain amino group of the upstream Lys (Lys190 in human Scp1) residue, which in turn forms a second intermolecular bridge to the phosphate moiety of the substrate. The second Asp or Asn (Asn207 in human Scp1) forms an axial coordination bond with the octahedrally coordinated Mg^{2+} ion (**Figure 3-2b**). Interestingly, the conservation of the second of the acidic stretch (Asn207) extends throughout HAD family whereas Asp206 is found to be replaced by Gly in a lot of HAD proteins (protein serine phosphatase, P-type ATPase, phosphomannomutase, histidinol phosphate phosphatase, sucrose phosphatase). This is very puzzling at first since our kinetic study strongly supports an essential structural function of Asp206 (or corresponding residues) in ion coordination and possibly a functional role in catalysis due to its mutation leads to capture of phosphoryl intermediate. A more careful examination of structure-based alignment shows that in HAD proteins where the first Asp of Motif III is replaced by Gly, the position corresponding to the side chain of Asp206 in Scp1 is always occupied by an Asp side chain extended from an amino acid 4~5 residue downstream of the second residue of Motif III (**Figure 3-10b**). Therefore, Asp206 side

chain function is conserved spatially with similar functions: Mg^{2+} coordination and salt bridging to Lys (**Table 3-3**).
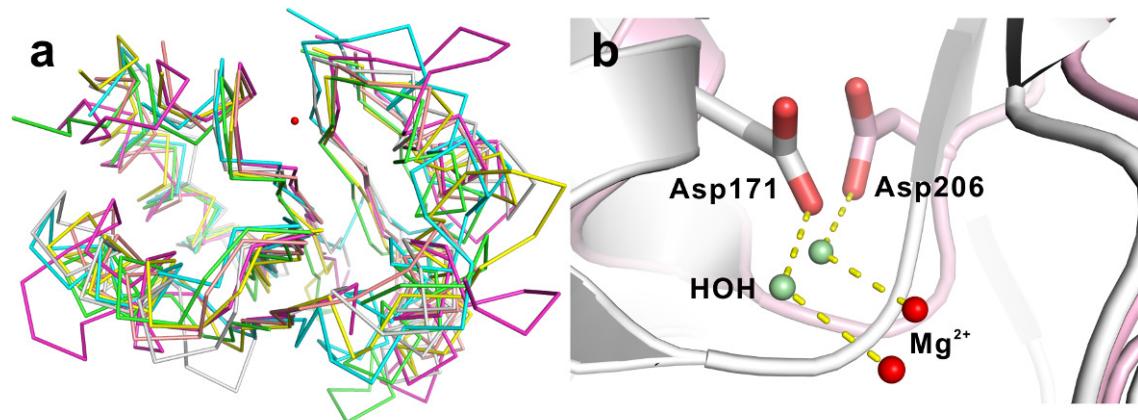
Figure 3-10: Structural alignment of HAD family members.[17]

[17] (a) Structural alignment of alpha-carbon backbone of Scp1 (yellow, 2ght), PSP (green, 1l7m), β-PGM (cyan, 1lvh), PMM (magenta, 2fuc), P-Type ATPase (pink, 2zbd) and Eya2 (gray, 3geb). Red sphere is the Mg^{2+} in Scp1, indicating the active site. For visual clarity, only the conserved portion is shown in each structure (Scp1: show residue 89-222 without residue 104-130; PSP: show residue 4-184 without residue 19-73 and 120-137; β-PGM: show residue 1-188 without residue 15-91; PMM: show residue 12-240 without residue 82-190; P-Type ATPase: show residue 346-719 without residue 359-602 and 641-673; Eya2: show residue 268-519 without 281-424). (b) Structural alignment of the active site of human Scp1 (light pink, 2ght) and MjPSP (white, 1l7m). The Asp206 in Scp1 and corresponding residue Asp171 in MjPSP are shown in stick. The water molecule hydrogen-bonding with Asp206 (or Asp171) is shown in pale green, and Mg^{2+} is shown in red. Hydrogen bonds are indicated by yellow dashed lines.

Protein name	PDB code	Seq. iden. (%)	Corres. a.a. for D96	Corres. a.a. for D98	Corres. a.a. for T152	Corres. a.a. for D206	Corres. a.a. for K190	Corres. a.a. for N207
			Motif I		Motif II	Motif III		
Pseudomonas sp. YL HAD	1jud	14	D10	Y12	S118	T175	K151	S176
Schizosaccharomyces pombe Fcp1	3ef0	24	D170	D172	T243	D297	K280	D298
Lactococcus lactis β-PGM	1lvh	14	D8	D10	S114	E169	K145	D170
T4 polynucleotide kinase	1ltq	16	D165	D167	S211	D277	K258	D278
Leishmania mexicana α-phosphomannomutase (PMM)	2i54	12	D10	D12	S46	D215	K188	D207
Human PMM	2fuc	16	D19	D21	G52	D226	K198	N218
Methanocaldococcus jannaschii Phosphoserine phosphatase (MjPSP)	1l7m	10	D11	D13	S99	D171	K144	D167
Human PSP	1l8l	15	D20	D22	S109	D183	K158	D179
Oryctolagus cuniculus P-Type ATPase	2zbd	13	D351	T353	T625	D707	K684	D703
Escherichia coli Histidinol phosphate phosphatase	2fpr	12	D10	D12	T55	D135	K106	D131
Synechocystis sp. PCC6803 Sucrose phosphatase	1tj3	15	D9	D11	T41	D190	K163	D186
Escherichia coli NagD	2c4n	14	D9	D11	T42	D206	K176	D201
Escherichia coli Class B acid phosphatase (AphA)	2b82	12	D44	D46	T112	D171	K152	D167
Human eyes absent phosphatase 2 (Eya2)	3geb	13	D274	D276	T448	E506	K480	D502

Table 3-3:   Conserved residues in the HAD family in comparison to Scp1.

58

Possible additional role of D206 in catalysis

Our studies have established that even though the coordination of Mg^{2+} by D206 is mediated through a water molecule, its mutation can reduce Mg^{2+} binding by at least 100-fold. When considering the reaction mechanism of phosphoryl transfer reaction, the most logical candidate for general acid/based is Asp98. As shown in a recent paper about a Scp family member, *Schizosaccharomyces pombe* Fcp1, a water molecule is occupying at an ideal position to be activated by Asp98 and nucleophically attacks phosphorus (Ghosh et al. 2008). Interestingly, our mutants of Asp98 still retain a significant percentage of phosphoryl activity, which prompted us to speculate if alternative general acid/base might exist especially when Asp98 is replaced. Considering the conservation of Asp206 throughout HAD family and our capture of phosphoryl intermediate upon mutation, it is possible that this residue may also play a catalytic role. With our current experimental method, Mg^{2+} concentration higher than 100 mM causes precipitation of the protein, therefore it is hard to determine whether saturating concentration of Mg^{2+} would fully rescue the activity of D206A mutant. However, it would be interesting to investigate the structural and functional roles of this residue in other HAD family. For example, it was reported that mutation at this residue in P-type ATPase (Asp707) totally abolishes its activity (McIntosh et al. 2004).

In summary, our current study provided definitive evidence for the existence of the phosphoryl-aspartate intermediate in the reaction catalyzed by Fcp/Scp phosphatase family. We identified Asp206 as a key residue for catalysis and metal binding, whose mutation can result in product trapping. Our results strongly suggest that Asp206 and its equivalent residues in other HAD family members play a structural and possible mechanistic role through divalent cation coordination. We believe these results will direct future studies on the mechanism of Scp1 and other HAD family enzymes.

MATERIALS AND METHODS

Cloning, mutagenesis and purification

Wildtype human Scp1, residues 77-256, was sub-cloned and expressed in *E.coli* BL21(DE3) strain using pHIS8 vector encoding a thrombin cleavable N-terminal octa-histidine tag (Jez et al. 2000). Mutant genes were generated using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, CA). The expression of Scp1 wildtype or mutants was induced in *E.coli* at low temperature (16°C) with 250 µM IPTG and allowed to accumulate to high levels overnight. Cells were harvested and re-suspended in 0.5 M NaCl, 15 mM imidazole, 50 mM Tris-HCl pH 8.0, 10% (v/v) glycerol and 0.1% (v/v) Triton X-100. After sonication, cell debris was removed by centrifugation. The supernatant was loaded onto a Ni^{2+}-NTA affinity column, washed extensively with sonication buffer and then with sonication buffer minus Triton X-100. Nearly homogeneous Scp1 was eluted with sonication buffer minus Triton X-100 but including 150 mM imidazole. Eluted protein was dialyzed into 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 5 mM β-mercaptoethanol and stored at a concentration of 2 mg/ml at 4 °C. Samples for protein x-ray crystallography were subjected to thrombin processing to remove the octa-histidine tag followed by size exclusion chromatography for final cleanup. These samples were stored in 20 mM HEPES-Na^+ pH 7.5, 100 mM NaCl, 3 mM dithiothreitol at a concentration of 12-16 mg/ml at -80 °C.

*p*NPP assay

The steady-state parameters of D98A/N, E99Q/A, and D206A were measured using *p*NPP as a substrate. The reaction mixtures contained 50 mM Tris-acetate pH 5.5, 20 or 50 mM $MgCl_2$, 0.5–50 mM of *p*NPP, and suitable amount of mutant proteins. The reactions were incubated at 37 °C for 15 mins and then quenched by adding 80 µl of 0.25

N NaOH. Release of pNP was determined by measuring absorbance at 410 nm. The absorbance was converted to product concentration by a pNP standard curve. The data were analyzed by nonlinear regression performed in Matlab (The MathWorks, Inc., MA). The Mg^{2+} dependence of wildtype Scp1 or Scp1D206A was measured by the same assay at constant substrate concentration with 0–100 mM Mg^{2+}.

DiFMUP assay

DiFMUP (Invitrogen, CA) was dissolved in 100% DMSO to make 100 mM stock. Before each measurement, the DiFMUP was freshly diluted in 1× reaction buffer containing 50 mM Tris-acetate pH 5.5, 20 mM $MgCl_2$ to desired concentrations. The substrate mixture and protein mixture were pre-incubated at 37 °C for 10 mins prior to mixing. The final reaction mixture contained 0–150 µM of DiFMUP and wildtype Scp1 or Scp1 mutants. The generation of the fluorescent DiFMU was monitored continuously at excitation/emission maxima ~358/450 nm for 2 hrs. The data were analyzed by the same method as described for pNPP assay.

Crystallization and structure determination

Scp1 mutants were crystallized in 0.5–0.8 M ammonium sulfate, 100 mM HEPES-Na$^+$ pH 7.0, with 0.2 M lithium sulfate. Crystals were then transferred to a stabilizer consisting of 100 mM sodium citrate, pH 5.5 (or MES-Na$^+$, pH 6.5), 30% (w/v) PEG 8000, 5-25 mM pNPP and 10 mM $MgCl_2$ or 0.2 M $MgCl_2$. The pH of the crystal stabilizer drop was retested and confirmed using pH paper. After soaking for 24 hrs, crystals were transferred to a cryo-protecting stabilizer containing 30% (v/v) glycerol, 30% (w/v) PEG8000 and 100 mM of the respective buffer of the original stabilizer. After a brief period of equilibration, crystals were frozen in nylon loops in liquid nitrogen and stored in liquid nitrogen prior to data collection. Data were collected at 100 K on beam-

line 8.2.2 of the Advanced Light Source (ALS). Diffraction data were processed with HKL2000 (Otwinowski et al. 1997). The data collection statistics are summarized in **Table 3-2**.

The crystal structures of Scp1 mutants were determined by molecular replacement (MR) using the Scp1D96N structure as a search model (PDB code: 2ghq) using the program AmoRe (Navaza 1994) available in the CCP4 software package (CCP4 1994). MR solutions were refined using CNS (Brunger et al. 1998) and REFMAC, reserving 5% of the measured and reduced structure factor amplitudes as an unbiased test set for cross validation (R_{free}) (Brunger 1992). SigmaA-weighted electron density maps ($2F_o$-F_c and F_o-F_c) were calculated after each cycle of refinement and carefully inspected to guide model rebuilding using O (Jones et al. 1991). The final models were evaluated by PROCHECK (Laskowski et al. 1993).

ACCESSION NUMBERS

Coordinates for structural studies in the paper have been deposited in the Protein Data Bank with human Scp1 D206A phosphoryl-Asp structure as 3L0B, D206A product trapped structure as 3L0C and D98A mutant as 3L0Y.

REFERENCES

Allen, K. N. and D. Dunaway-Mariano (2004). "Phosphoryl group transfer: evolution of a catalytic scaffold." *Trends Biochem Sci* **29**(9): 495-503.

Barford, D., A. K. Das and M. P. Egloff (1998). "The structure and mechanism of protein phosphatases: insights into catalysis and regulation." *Annu Rev Biophys Biomol Struct* **27**: 133-164.

Brunger, A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." *Nature* **355**(6359): 472-475.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5)**: 905-921.

CCP4 (1994). "Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography." *Acta Crystallogr.* **D50**: 760-763.

Chambers, R. S. and M. E. Dahmus (1994). "Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II." *J Biol Chem* **269**(42): 26243-26248.

Chambers, R. S. and C. M. Kane (1996). "Purification and characterization of an RNA polymerase II phosphatase from yeast." *J Biol Chem* **271**(40): 24498-24504.

Chapman, R. D., M. Heidemann, T. K. Albert, R. Mailhammer, A. Flatley, M. Meisterernst, E. Kremmer and D. Eick (2007). "Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7." *Science* **318**(5857): 1780-1782.

Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt and S. Buratowski (2001). "Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain." *Genes Dev* **15**(24): 3319-3329.

Cho, H., T. K. Kim, H. Mancebo, W. S. Lane, O. Flores and D. Reinberg (1999). "A protein phosphatase functions to recycle RNA polymerase II." *Genes Dev* **13**(12): 1540-1552.

Cho, H., W. Wang, R. Kim, H. Yokota, S. Damo, S. H. Kim, D. Wemmer, S. Kustu and D. Yan (2001). "BeF(3)(-) acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF(3)(-) complex with phosphoserine phosphatase." *Proc Natl Acad Sci U S A* **98**(15): 8525-8530.

Cohen, P. T. (1997). "Novel protein serine/threonine phosphatases: variety is the spice of life." *Trends Biochem Sci* **22**(7): 245-251.

Deshpande, R. A. and T. E. Wilson (2004). "Identification of DNA 3'-phosphatase active site residues and their differential role in DNA binding, Mg2+ coordination, and catalysis." *Biochemistry* **43**(26): 8579-8589.

Diaz, A. R., S. Stephenson, J. M. Green, V. M. Levdikov, A. J. Wilkinson and M. Perego (2008). "Functional role for a conserved aspartate in the Spo0E signature motif involved in the dephosphorylation of the Bacillus subtilis sporulation regulator Spo0A." *J Biol Chem* **283**(5): 2962-2972.

Egloff, S. and S. Murphy (2008). "Cracking the RNA polymerase II CTD code." *Trends Genet* **24**(6): 280-288.

Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick and S. Murphy (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." *Science* **318**(5857): 1777-1779.

Ghosh, A., S. Shuman and C. D. Lima (2008). "The structure of Fcp1, an essential RNA polymerase II CTD phosphatase." *Mol Cell* **32**(4): 478-490.

Hausmann, S. and S. Shuman (2003). "Defining the active site of Schizosaccharomyces pombe C-terminal domain phosphatase Fcp1." *J Biol Chem* **278**(16): 13627-13632.

Jez, J. M., J. L. Ferrer, M. E. Bowman, R. A. Dixon and J. P. Noel (2000). "Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase." *Biochemistry* **39**(5): 890-902.

Jones, T. A., J. Y. Zou, S. W. Cowan and M. Kjeldgaard (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." *Acta Crystallogr A* **47 (Pt 2)**: 110-119.

Jung, S. K., D. G. Jeong, S. J. Chung, J. H. Kim, B. C. Park, N. K. Tonks, S. E. Ryu and S. J. Kim (2010). "Crystal structure of ED-Eya2: insight into dual roles as a protein tyrosine phosphatase and a transcription factor." *Faseb J* **24**(2): 560-569.

Kamenski, T., S. Heilmeier, A. Meinhart and P. Cramer (2004). "Structure and mechanism of RNA polymerase II CTD phosphatases." *Mol Cell* **15**(3): 399-407.

Lahiri, S. D., G. Zhang, D. Dunaway-Mariano and K. N. Allen (2002). "Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis." *Biochemistry* **41**(26): 8351-8359.

Lahiri, S. D., G. Zhang, D. Dunaway-Mariano and K. N. Allen (2003). "The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction." *Science* **299**(5615): 2067-2071.

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

McIntosh, D. B., J. D. Clausen, D. G. Woolley, D. H. MacLennan, B. Vilsen and J. P. Andersen (2004). "Roles of conserved P domain residues and Mg2+ in ATP binding in the ground and Ca2+-activated states of sarcoplasmic reticulum Ca2+-ATPase." *J Biol Chem* **279**(31): 32515-32523.

Meinhart, A., T. Kamenski, S. Hoeppner, S. Baumli and P. Cramer (2005). "A structural perspective of CTD function." *Genes Dev* **19**(12): 1401-1415.

Morris, D. P., H. P. Phatnani and A. L. Greenleaf (1999). "Phospho-carboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3'-End formation." *J Biol Chem* **274**(44): 31583-31587.

Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." *Acta Crystallogr.* **A50**: 157-163.

Otwinowski, Z. and W. Minor (1997). "HKL: Processing of X-ray diffraction data collected in oscillation mode." *Methods Enzymol.* **276**: 307-326.

Pannifer, A. D., A. J. Flint, N. K. Tonks and D. Barford (1998). "Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by x-ray crystallography." *J Biol Chem* **273**(17): 10454-10462.

Rangarajan, E. S., A. Proteau, J. Wagner, M. N. Hung, A. Matte and M. Cygler (2006). "Structural Snapshots of Escherichia coli Histidinol Phosphate Phosphatase along the Reaction Pathway." *J Biol Chem* **281**(49): 37930-37941.

Shi, Y. (2009). "Serine/threonine phosphatases: mechanism through structure." *Cell* **139**(3): 468-484.

Tonks, N. K. (2006). "Protein tyrosine phosphatases: from genes, to function, to disease." *Nat Rev Mol Cell Biol* **7**(11): 833-846.

Toyoshima, C., H. Nomura and T. Tsuda (2004). "Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues." *Nature* **432**(7015): 361-368.

Toyoshima, C., Y. Norimatsu, S. Iwasawa, T. Tsuda and H. Ogawa (2007). "How processing of aspartylphosphate is coupled to lumenal gating of the ion pathway in the calcium pump." *Proc Natl Acad Sci U S A* **104**(50): 19831-19836.

Verdecia, M. A., M. E. Bowman, K. P. Lu, T. Hunter and J. P. Noel (2000). "Structural basis for phosphoserine-proline recognition by group IV WW domains." *Nat Struct Biol* **7**(8): 639-643.

Wang, W., H. S. Cho, R. Kim, J. Jancarik, H. Yokota, H. H. Nguyen, I. V. Grigoriev, D. E. Wemmer and S. H. Kim (2002). "Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states." *J Mol Biol* **319**(2): 421-431.

Wang, W., R. Kim, J. Jancarik, H. Yokota and S. H. Kim (2001). "Crystal structure of phosphoserine phosphatase from Methanococcus jannaschii, a hyperthermophile, at 1.8 A resolution." *Structure* **9**(1): 65-71.

Yeo, M., S. K. Lee, B. Lee, E. C. Ruiz, S. L. Pfaff and G. N. Gill (2005). "Small CTD phosphatases function in silencing neuronal gene expression." *Science* **307**(5709): 596-600.

Yeo, M., P. S. Lin, M. E. Dahmus and G. N. Gill (2003). "A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5." *J Biol Chem* **278**(28): 26078-26085.

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." *Mol Cell* **24**(5): 759-770.

Zhang, Z. Y. (2003). "Mechanistic studies on protein tyrosine phosphatases." *Prog Nucleic Acid Res Mol Biol* **73**: 171-220.

Chapter 4:   Selective Inactivation of Scp1 by a Small Molecule Inhibitor

INTRODUCTION

The CTD specific kinases and phosphatases function as house-keeping regulatory factors for global transcription (Majello et al. 2001). Recently, it has been shown that certain CTD regulatory factors can also epigenetically modulate the expression level of a specific group of genes (Li et al. 2005; Yeo et al. 2005; Zhang et al. 2010). As an example, a newly discovered class of phosphatases, the human <u>s</u>mall <u>C</u>-terminal domain <u>p</u>hosphatases (Scps), specifically dephosphorylates phosphorylated Ser_5 (phospho.Ser_5) of the tandem heptad repeats of the CTD (Yeo et al. 2003). Interestingly, Scps have also been shown to epigenetically silence the expression of a specific set of neuronal genes in neuronal stem cells and non-neuronal cells by acting as co-repressors in REST/NRSF complex (Yeo et al. 2005). Inhibition of Scps in P19 stem cells by the dominant negative mutants (which retain the overall structure but have abolished phosphatase activity) or microRNA miR-124 (which directly targets the untranslated region of Scp genes and suppresses their expression) allows neuronal gene expression and induces neuronal differentiation (Yeo et al. 2005; Visvanathan et al. 2007). Given the demonstrated role of Scps in limiting inappropriate expression of neuronal specific genes in pluripotent cells and the fact that their down-regulation leads to neuronal differentiation, Scps serve as promising new targets for small molecule inhibitors to regulate neuronal stem cell development and to promote neuronal differentiation.

However, one of the greatest challenges associated with phosphatase inhibitor identification is the cross inhibition of other phosphatases due to poor selectivity (Yu et al. 2007), which usually stems from small, uncharacteristic active sites of phosphatases.

68

For Scp inhibitors, selectivity is of great concern as two close family members in the Scp/Fcp family, Fcp1 (Cho et al. 1999) and Dullard (Kim et al. 2007), play essential roles in cell survival as well as proper development (Chambers et al. 1994; Archambault et al. 1997; Kim et al. 2007).

The crystal structures of Scp/Fcp family members Fcp1 and Scp1 were initially solved by Ghosh *et al.* (Ghosh et al. 2008) and Kamenski *et al.* (Kamenski et al. 2004). Unlike the traditional cysteine-based or dimetal-dependent phosphatases, the Scp/Fcp family members belong to a unique family of phosphatases that rely on the DXDX(T/V) motif and Mg^{2+} to catalyze the phosphoryl-transfer (Yeo et al. 2003; Kim et al. 2007), as found in the haloacid dehydrogenase (HAD) superfamily (Allen et al. 2004). In fact, the overall core fold of Scps resembles the core domains of other HAD family members, despite low sequence similarity (see **Chapter 3**) (Zhang et al. 2010).

The HAD superfamily can be further divided into three subfamilies (**Figure 4-1**) according to the presence and the location of a second domain known as the "cap domain" (Morais et al. 2000; Zhang et al. 2002). Both type I and II subfamily members utilize the cap domain to shield the active site from bulk solvent and to achieve substrate recognition (Allen et al. 2004). The type III HAD superfamily, including the Scp/Fcp family, do not have a cap domain, requiring specificity achieved through alternative strategies, such as recruitment of other regulatory proteins. This feature poses an additional challenge for inhibitor design. In fact, no specific inhibitors have been reported to date for type III HAD family members.

The complex structure of Scp1 bound to its substrate peptide has previously been solved (Zhang et al. 2006), and the snapshots of the phosphoryl-transfer reaction at each step and the formation of the phospho-aspartyl intermediate were captured using X-ray crystallography and the reaction mechanism of Scps was established (Zhang et al. 2010).

69

These crystallographic and biochemical studies not only provided insights into the catalytic mechanism of Scp/Fcp, but also hinted a novel strategy of specific recognition of substrates by Scps. The complex structure of Scp1 bound to its substrate peptide revealed a hydrophobic binding pocket which is specific for Scps, suggesting that a specific Scp inhibitor might be obtained through targeting this pocket (Zhang et al. 2006). In the present study, we exploited this possibility and identified rabeprazole as the first reported lead compound for Scp inhibition ($K_i = 5 \pm 1$ μM). This small molecule shows no inhibition towards Fcp1 or Dullard, nor towards bacteriophage lambda Ser/Thr phosphoprotein phosphatase (λPPase). This extraordinary selectivity can be explained through analysis of our high resolution structure of Scp1 complexed with the compound, which shows, as expected, the compound binds specifically to the unique hydrophobic binding pocket of Scps. The structure highlights the chemical functional groups that make essential contributions to binding. To the best of our knowledge, this is the first selective lead compound for the Scp/Fcp phosphatase family, as well as the type III HAD family.

a

b

c

Hydrophobic pocket
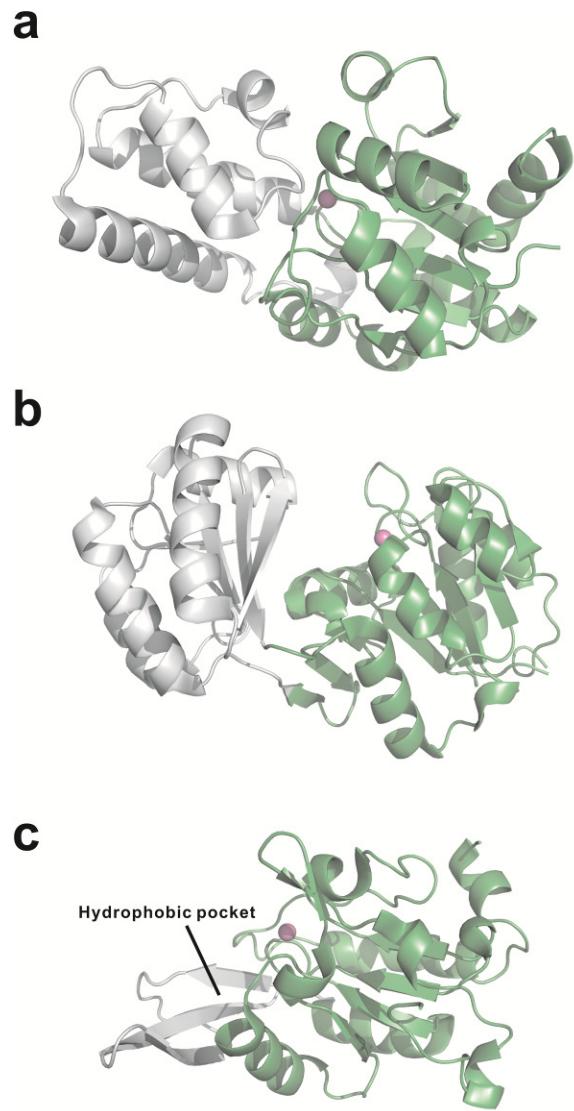
Figure 4-1:   Representative family members of HAD super family.[18]

[18] (a) Type I family member β-phosphoglucomutase (PDB code: 1o08), (b) Type II family member phosphatase *TM0651* (PDB code: 1nf2), (c) Type III family member Scp1 (PDB code: 2ght). The conserved core domain is colored in lime, and the portion that does not align structurally is colored in gray. Mg^{2+} is shown in violet to indicate the active site.

RESULTS AND DISCUSSION

Human Scps as target for inhibitor identification

Phosphatases have historically proven to be difficult targets for inhibitor design. The difficulty stems from three major challenges in inhibitor design: (1) affinity: the binding pockets of phosphatases tend to be small, and therefore limit potential molecular interactions; (2) specificity: the substrate-specificities of phosphatases tend to overlap, creating possibilities for cross inhibition; (3) permeability: the mimicry of the charged phosphate group, which is key to recognition by phosphatases, often prevents the compounds from penetrating cell membranes. Owing to these difficulties, one of the best phosphatases inhibitors, I-C11, inhibits its target lymphoid-specific tyrosine phosphatase (Lyp) with a K_i of 2.9 ± 0.5 μM and shows some degree of promiscuity (Yu et al. 2007). With hundreds of Ser/Thr PPases and protein tyrosine phosphatases identified, the specific inhibitors reported are very limited. The most potent inhibitor for any phosphatase is a small molecule reported recently which targets T cell protein tyrosine phosphatase with a K_i of 4.3 ± 0.2 nM and a minor degree of cross inhibition (Zhang et al. 2009).

The newly discovered Scps might prove to be viable targets for inhibitor design. First, the structure of Scp1 bound to its natural substrate phosphorylated CTD peptide reveals a spacious substrate-binding area compared to that of other phosphatases (**Figure 4-2a**). Most notably, there is a hydrophobic pocket unique to Scps that specifically recognizes the Pro_3 of the CTD and is about 7 Å away from the active site where the phosphate group binds (**Figure 4-2a**). Compounds that target this hydrophobic pocket are expected to have high binding affinities due to ample opportunities of making molecular contacts. Secondly, cross inhibition of close family members might be prevented by targeting this hydrophobic pocket since it is unique to Scps. The hydrophobic pocket is

located between the three-stranded β sheet insertion domain and the core domain. Many of the hydrophobic residues lining a part of the hydrophobic pocket are from the insertion domain which is unique to the Scp/Fcp family and is highly diversified within Scp/Fcp family members (**Figure 4-2b**). There is no sequence similarity between Scp, Fcp, Dullard and other family members in the insertion domain (**Figure 4-2c**). Finally, the affinity provided by binding to the hydrophobic pocket might mitigate the necessity of having a charged moiety to mimic phosphate, making the prospect compound more membrane-permeable. By considering these unique advantages of Scps, we reasoned that Scps might be feasible targets for selective and high affinity inhibitor design.

a

DLDET Motif

Mg²⁺

Hydrophobic pocket

phospho.Ser₅

Pro₃

b

Scp1	102-133	261 aa	
Scp2	113-144	271 aa	
Scp3	107-137	265 aa	
Dullard	73-103	244 aa	
Fcp1	75-108 (176-224 in *S. pombe*)	842 aa	

CTD phosphatase domain

Insertion motif

BRCT domain

TFIIF interacting helix

c

```
              β2              β3                              β4
HsScp1       →               →                               →
             1        10                     20          30
HsScp1    VHSSFKPVNNADFIIPVEI.................DGVVHQVYVLKRP
HsScp2    VHSSFKPINNADFIVPIEI.................EGTTHQVYVLKRP
HsScp3    VHSSFKPISNADFIVPVEI.................DGTIHQVYVLKRP
HsDullard IHSHHDGVLRPT.VRPGTP...............PDFILKVVIDKHP
HsFcp1    IHTTEQHCQQMSNKGIFHFQLG.............RGEP.MLHTRLRP
SpFcp1    IHATVDPTVGEWMSDPGNVNYDVLRDVRSFNLQEGPSGYTSCYYIKFRP
SpFcp1       →      llllll        lll         →           →
             β2       α2          η1          β3          β4
```

Figure 4-2: The hydrophobic binding pocket is unique to Scps.[19]

---

[19] (a) The complex structure of Scp1 and a CTD-derived peptide (PDB code: 2ght). The active site DXDXT motif is shaded with pink color and the hydrophobic pocket residues are shaded with blue color. (b) Domain construct of human CTD phosphatases Scp1 (NCBI accession number: AAH12977), Scp2 (AAH65920), Scp3 (NP_005799), Dullard (AAH09295), Fcp1 (AAC64549) and *S. pombe* Fcp1 (NP_594768). Each domain is represented by a colored block. The catalytic domain is colored in light blue. The insertion domain, which is partially conserved, is colored in yellow. (c) The zoom-in figure of the insertion domain alignment. The upper panel shows the secondary structure of human Scp1 insertion domain, and the lower panel shows the secondary structure of *S. pombe* Fcp1 insertion domain. α indicates α helix, β indicates β sheet, and η indicates 3_{10} helix.

Identification of Scp inhibitors

Scps include three highly similar homologous family members in the human: Scp1, Scp2, and Scp3 (Yeo et al. 2003). Since no differences are found in the catalytic activities or biological functions of Scp1–3, they are thought to be functionally redundant (Yeo et al. 2005). Our structural studies of Scp1, 2 and 3 show little overall structural difference and no observable difference at the active site (Zhang Y, unpublished data). Furthermore, they display identical kinetic characteristics against *para*-nitrophenyl phosphate (*p*NPP) and phosphorylated CTD peptide (Kim Y and Dixon JE, personal communication). Therefore, we chose the best-characterized family member, Scp1, in the following studies.

In order to identify the inhibitors for Scps, we screened a pilot library of the NIH clinical collection (~400 compounds) and spectrum collection (2000 compounds) for their ability to impede phosphatase activity of Scp1. The screening assay was performed using *p*NPP as the substrate at a concentration comparable to its K_m (~6 mM). It was confirmed that Scp1 was stable under the assay conditions and can tolerate up to 10% DMSO. A Z' factor of 0.87 was obtained using the optimized screening protocol. Thirty-nine compounds showed greater than 70% inhibition when screened at a 50 μM inhibitor concentration.

To eliminate the false positives, we compared our initial hits with the collaborative drug discovery database (CDD) and identified eight compounds that are sufficiently soluble and less likely to be false positives. To further confirm the inhibitory effect of the compounds, we used a secondary assay to monitor phosphatase activity, where the CTD-derived phosphorylated peptide ($S_{a5}P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}S_{b5}P_{b6}S_{b7}Y_{c1}S_{c2}P_{c3}T_{c4}\textbf{phospho.S}_{\textbf{c5}}P_{c6}S_{c7}$) was used as the substrate and malachite green reagent was used to capture the released phosphate, and in doing so

75

produce a colorimetric signal. Five of the eight compounds showed inhibition for Scp1 when using the natural substrate in the assay. Rabeprazole (**Figure 4-3a**), which showed the strongest inhibition, was further characterized to exhibit an IC_{50} of 4 ± 0.7 µM in the pNPP assay and 9 ± 3 µM in the malachite green assay, where the concentration of pNPP and phosphorylated peptide were 6 mM and 30 µM, respectively, close to their respective K_m values (**Figure 4-3b**). To eliminate the possibility that the inhibitory effect of the compound was caused by protein denaturation, the thermal stability of Scp1 in the absence or presence of rabeprazole was tested in differential scanning fluorimetry assay (Niesen et al. 2007). Under the measurement condition, the melting temperature of Scp1 is 57.8 °C and 59.1 °C in the absence and presence of the compound, respectively. This result excludes protein destabilization as the mechanism for the observed inhibitory effect of rabeprazole on Scp1 (**Figure 4-4**).

a

b

pNPP

peptide

$IC_{50} = 4.0 \pm 0.7$ μM
$H = 0.9 \pm 0.1$

$IC_{50} = 9 \pm 3$ μM
$H = 0.8 \pm 0.2$

Relative activity

Relative activity

[Rabeprazole] (μM)

[Rabeprazole] (μM)

c

Rate/[E$_0$] (s^{-1})

Without inhibitor:
$k_{cat} = 13 \pm 1$ s^{-1}
$K_m = 2.8 \pm 0.9$ mM

With inhibitor:
$k_{cat} = 6.6 \pm 0.6$ s^{-1}
$K_m = 5 \pm 1$ mM

[pNPP] (mM)

Rate/[E$_0$] (s^{-1})

Without inhibitor:
$k_{cat} = 10 \pm 0.4$ s^{-1}
$K_m = 27 \pm 5$ μM

With inhibitor:
$k_{cat} = 8.6 \pm 0.3$ s^{-1}
$K_m = 70 \pm 9$ μM

[peptide] (mM)
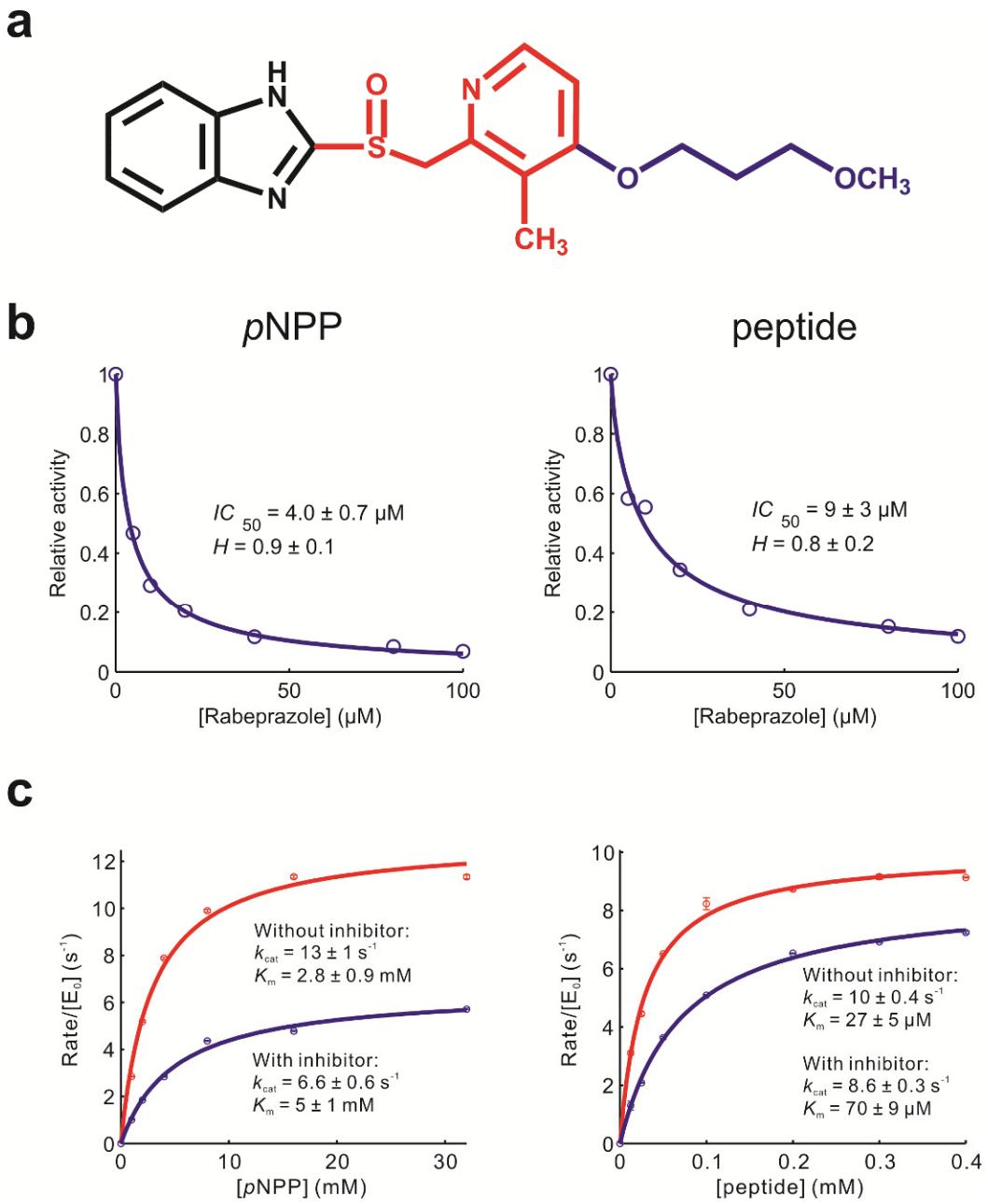
Figure 4-3:   Inhibition of Scp1 by rabeprazole.[20]

---

[20] (a) Structure of rabeprazole. The benzimidazole ring is in black. The methyl pyridine ring and methyl sulfinyl groups are in red. The methoxypropoxy tail is in blue. (b) Concentration-response of rabeprazole inhibition towards Scp1 tested by the pNPP assay and the malachite green assay. IC$_{50}$ values were derived by fitting the data to the equation (1) in Materials and Methods. (c) Steady-state kinetics of Scp1 in the presence (blue) and absence (red) of rabeprazole. Inhibition constants were derived by fitting the data to equation (2) in Materials and Methods.
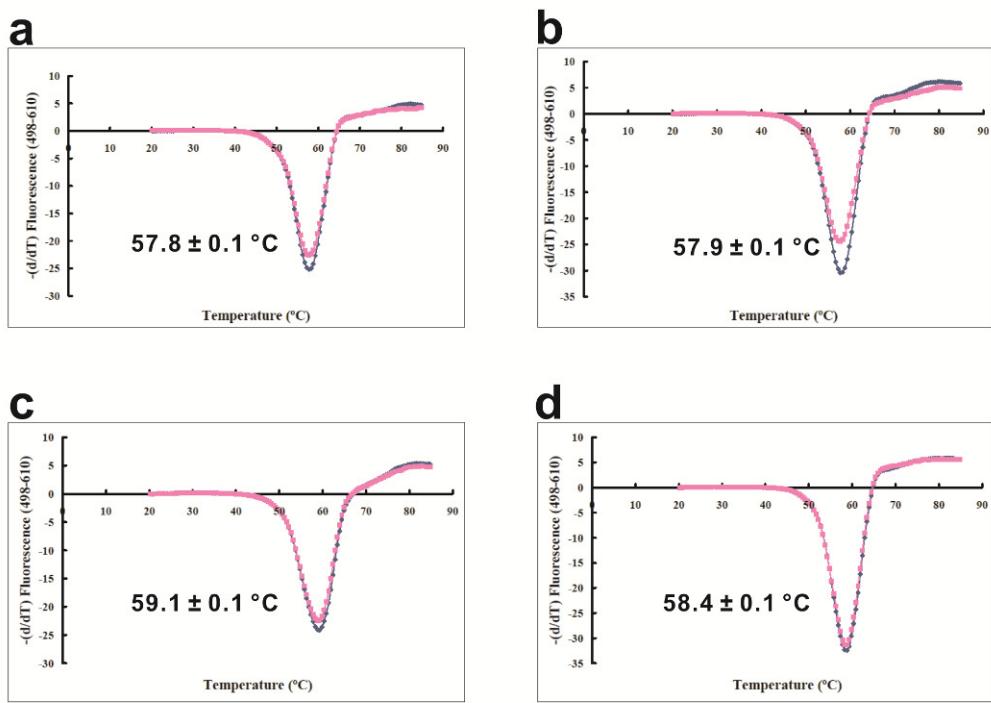
Figure 4-4: Melting curves measured by differential scanning fluorimetry (DSF).[21]

In order to probe the mechanism of inhibition, the steady-state kinetics using both the pNPP and malachite green assays were determined. Kinetic analysis revealed that the compound is a mixed inhibitor for Scp1 against pNPP. The inhibition constants K_i and K_i' of the compound was determined to be 5 ± 2 μM and 10 ± 2 μM using pNPP as the substrate. Interestingly, when the CTD-derived phosphorylated peptide was used as the substrate, rabeprazole exhibited characteristics of a competitive inhibitor: while the K_i was determined to be 5 ± 1 μM, the K_i' was at least ten times higher than K_i, and could not be precisely determined (**Figure 4-3c**). Rabeprazole very likely directly competes with the natural substrate CTD peptide in binding to Scp1.

---

[21] The samples are Scp1 only at 20 μM final concentration (a), Scp1 with 1% DMSO (b), Scp1 with rabeprazole at 100 μM final concentration plus 1% DMSO (c), and Scp1 with 17-mer peptide at 1mM final concentration (d). Each measurement was done in duplicates.

Complex structure of Scp1 and rabeprazole

To understand the inhibition mechanism of rabeprazole at the molecular level, we attempted to obtain the structure of Scp1 and rabeprazole by X-ray crystallography. Unfortunately, crystals derived from the published conditions proved to be too fragile upon compound soaking. We therefore identified new crystallization conditions for wildtype Scp1 which yielded Scp1 crystals in much higher quality (see **Materials and Methods**). Unlike the original conditions, with high concentration of ammonium sulfate as precipitant, the crystals grown from the new polyethylene glycol (PEG)-based conditions are much more durable upon compound soaking. The complex structures were obtained by soaking the Scp1 crystals in buffer containing 0.1–1 mM rabeprazole and incubating for various amounts of time at 25 ℃. After optimization of soaking conditions, the crystals soaked in ~0.5 mM rabeprazole for 2–3 hr were used in X-ray data collection with good diffraction quality (**Table 4-1**).

In all the data collected from multiple crystals, a strong area of electron density was observed at the hydrophobic pocket for recognition of the Pro_3 of CTD substrate (**Figure 4-5a**). The final structure of the Scp1-rabeprazole complex was refined to 2.35 Å with the protein portion of the structure highly identical to apo Scp1 (**Figure 4-6**). Between the two molecules in each asymmetric unit, molecule A shows strong and consistent positive density at the active site, indicating effective compound binding. In comparison to molecule A, the density at the active site of molecule B is much weaker. Our subsequent discussion of protein and ligand interaction focuses on molecule A in which a better model can be generated based on the diffraction data.

Data collection		Scp1-rabeprazole complex
Space group		C2
Cell dimensions:	a, b, c (Å)	125.5, 78.3, 62.7
	α, β, γ (°)	90.0, 111.9, 90.0
Resolution (Å)		50–2.35 (2.39–2.35)
No. of unique reflections		21285
R_{sym} or R_{merge} (%)		10.0 (39.7)
$I/\sigma(I)$		11.4 (1.6)
Completeness (%)		90.2 (47.3)
Redundancy		3.1 (2.2)
Refinement		
Resolution (Å)		64.95-2.35
No. of reflections (test set)		20188 (1097)
R_{work} / R_{free} (%)		20.7/26.5
No. of atoms:	Protein	2928
	Ion	2
	Ligand	25
	Water	112
B-factors (Å^2):	Protein	32.0
	Mg^{2+}	28.0
	Ligand	52.5
	Water	34.6
R.m.s deviations:	Bond lengths (Å)	0.020
	Bond angles (°)	1.977
Ramachandran plot (%): Most favored		85.7
Additionally allowed		13.7
Generally allowed		0.3

Table 4-1:   Crystallographic data statistics.[22]

---

[22] Highest resolution shell is shown in parenthesis. R_{free} is calculated with 5% of the data omitted.
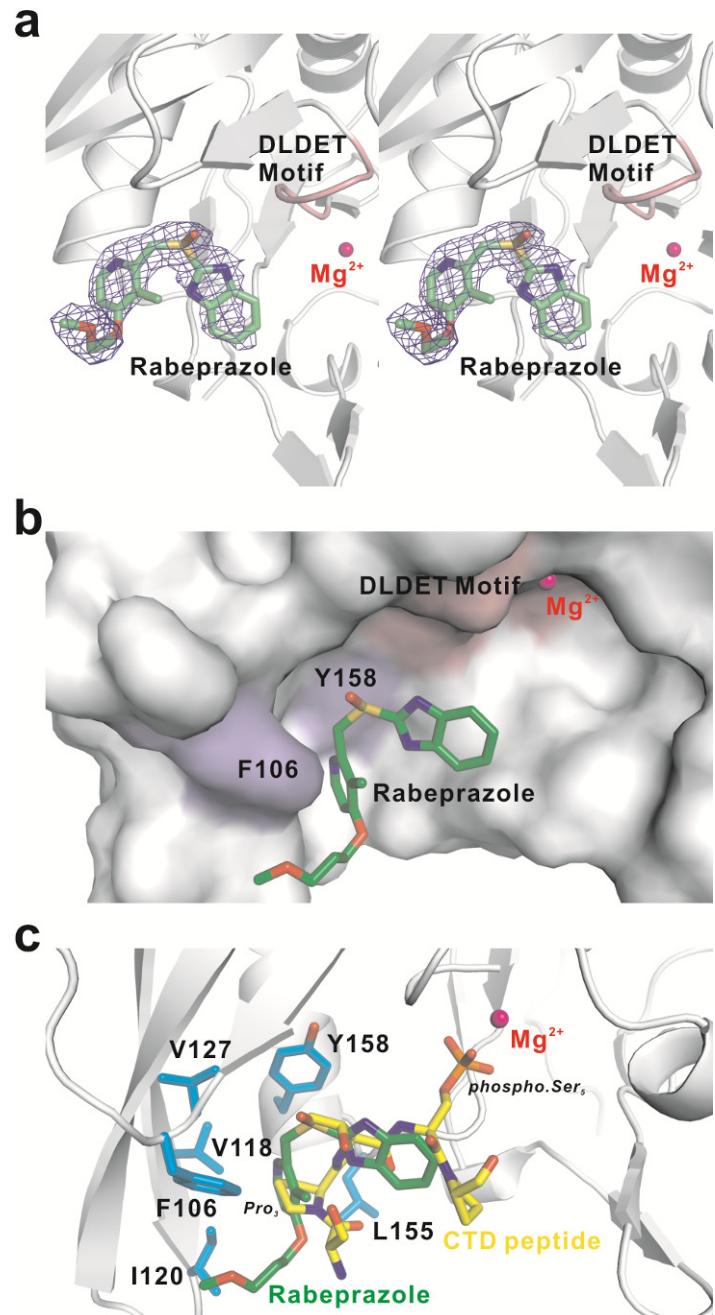
Figure 4-5: Complex structure of Scp1 and rabeprazole.[23]

---

[23] (a) Structure of rabeprazole bound to Scp1 with the blue SIGMAA-weighted $2F_o$-F_c electron-density map contoured at 1σ shown in stereo. (b) Surface representation of rabeprazole bound to the hydrophobic pocket of Scp1. (c) Superimposition of Scp1-rabperazole complex and Scp1-CTD peptide complex (PDB code: 2ght). The protein portion is identical. The hydrophobic pocket residues are shown in stick in cyan.
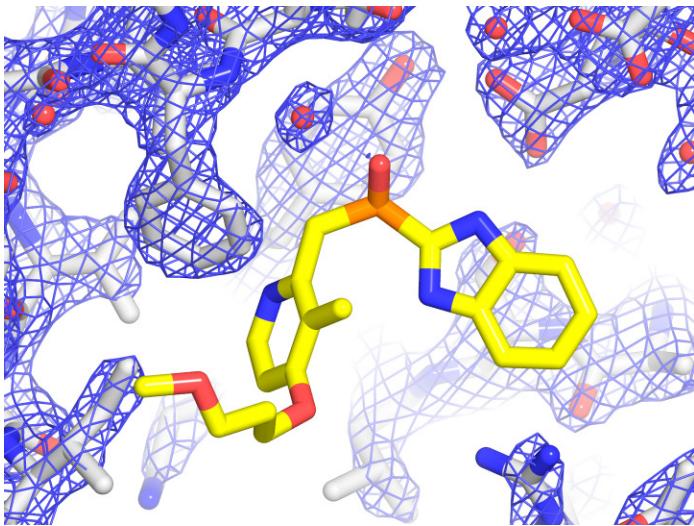
Figure 4-6: The SIGMAA-weighted F_o-F_c electron density map of Scp1 contoured at 1.7σ.

The total buried contact area of Scp1 upon rabeprazole binding is about 279 Å^2. The shape complementarily (Sc) value is calculated to be 0.54 which indicates a good shape complementarity between the protein and ligand surfaces (an Sc value of 1 reflects a perfect fit). The most distinguishing characteristic of the compound binding is the site of interaction of the methyl pyridine ring at the hydrophobic pocket surrounded by Tyr158, Phe106, Val118, Ile120, Val127 and Leu155 (**Figure 4-5b, c**). The distances between the methyl pyridine ring and surrounding key residues are within 3.5–4.5 Å. The binding to this hydrophobic pocket is noteworthy as this "insertion domain" is unique for Scps and is believed to be the major structural element for the specific recognition of Pro_3 of the CTD (Zhang et al. 2006). Since the compound binds to the hydrophobic pocket of Scp1, it is likely to prevent Pro_3 of the substrate from binding to the protein, explaining why the compound resembled a competitive inhibitor to Scp1 when the peptide was used as the substrate (**Figure 4-5c**). To further validate the binding mode of rabeprazole binding to Scp1, mutagenesis study around the hydrophobic binding pocket was

82

performed. Four mutants, including F106E, V127A, K157A and L155A, were tested by the pNPP assay. The IC_{50} of rabeprazole when tested against each mutant was derived from the concentration-response curve (**Figure 4-7**). It is clearly shown that F106E mutant increased IC_{50} by at least ten fold. The other mutants, V127A, K157A and L155A, all have moderately increased IC_{50}, indicating some degree of loss of inhibition. The mutagenesis results show that the hydrophobic residues in the insertion domain are very important for inhibition of Scp1 by rabeprazole.

The interaction of the sulfoxide group of rabeprazole and Tyr158 side chain resembles a cation-π interaction where the electron-rich π cloud of the phenyl ring of Tyr158 interacts with the partially positively charged sulfur of rabeprazole. Cation-π interactions are not uncommon in protein-inhibitor interactions and have been recognized as an important noncovalent interaction in structural biology (Dougherty 1996; Scrutton et al. 1996; Ma et al. 1997; Gallivan et al. 1999). For instance, the horse liver alcohol dehydrogenase is bound by its inhibitor sulfoxides through a cation-π interaction between the inhibitor and the benzene ring of Phe93 of the enzyme (Cho et al. 1997). In our structure, the sulfur atom locates directly over the phenyl ring of Tyr158. The distance between the sulfur atom and the center of phenyl ring is about 4 Å, which is within the range of typical distance for "amino-aromatic" interactions (Burley et al. 1986). Indeed, when we eliminated this cation-π interaction by mutating Tyr158 to Ala, a loss of inhibition by at least ten fold was observed when tested by the pNPP assay (**Figure 4-7**).
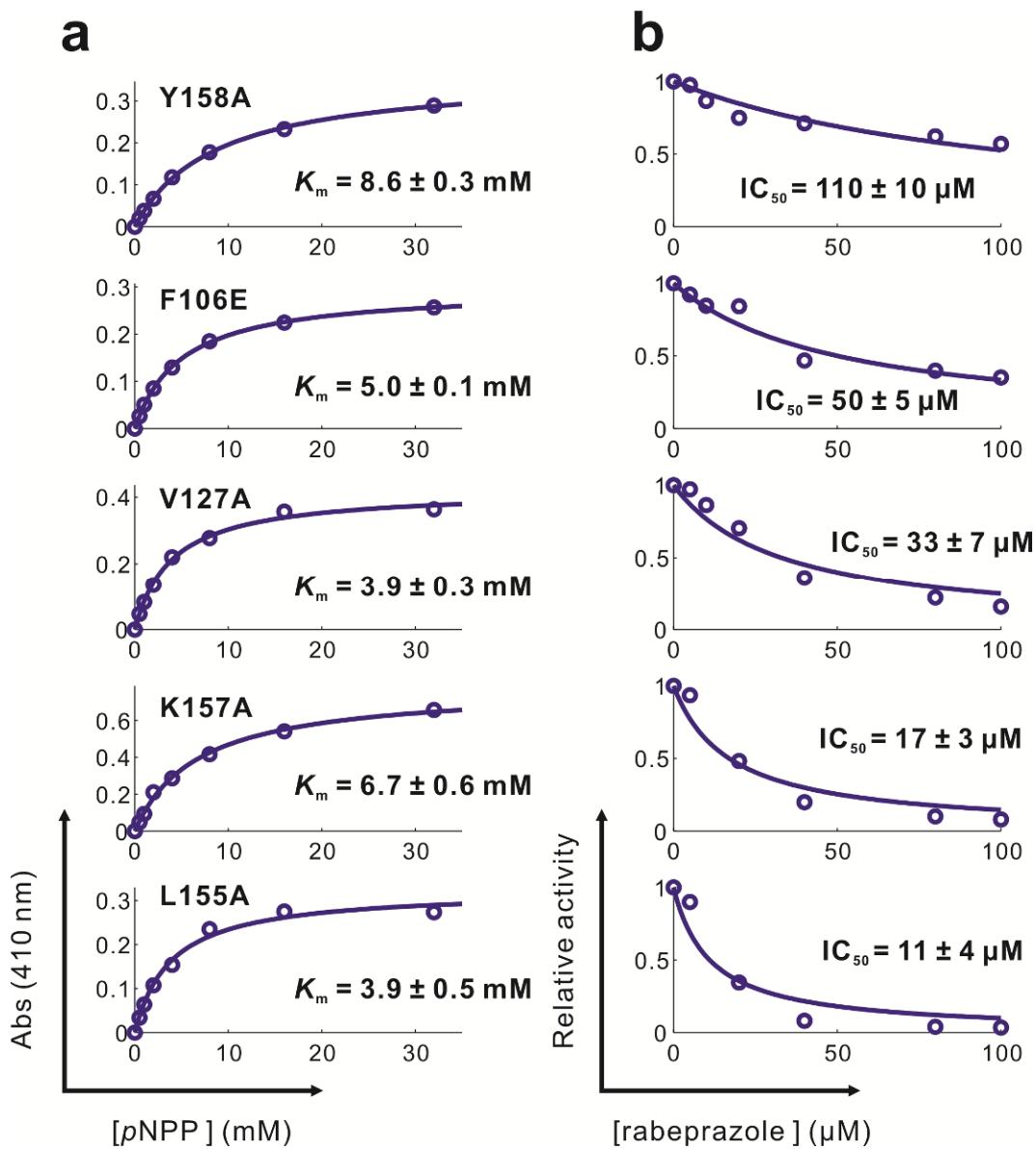
Figure 4-7:   Inhibition assay of Scp1 mutants.[24]

It is likely that not all functional groups of rabeprazole contribute to Scp1 binding.

Unlike pyridine ring and sulfoxide, the benzimidazole ring showed little if any electron

---

[24] (a) Steady-state kinetics of Y158A, F106E, V127A, K157A and L155A tested by the pNPP assay. (b) Concentration response of rabeprazole inhibition towards Y158A, F106E, V127A, K157A and L155A determined by the pNPP assay.

density which suggests that this portion is flexible in our structure. Accordingly, the refinement of this part of the structure presents a higher thermal factor than the active site residues, thus we rebuilt our model with the benzimidazole ring as 50% occupancy. The refinement puts the group in a position that extends it outward of the active site, contributing little to the recognition of the inhibitor. In addition, the methoxypropoxy "tail" portion also showed partial densities consistently in multiple crystals. Although there are two oxygen atoms on the "tail", no identifiable polar interactions were observed in our structure. It may also form hydrophobic interactions with the pocket, and the distances between the "tail" and surrounding hydrophobic residues are within 4–4.4 Å.

It should be noted that the rabeprazole used in our structural studies is a mixture of two enantiomers with the sulfur as the chiral center. However, only (R)-rabeprazole can be built into the electron density whereas the S enantiomer does not fit the density well (**Figure 4-8**). We predict that only the R enantiomer binds the protein and exerts inhibitory effect.
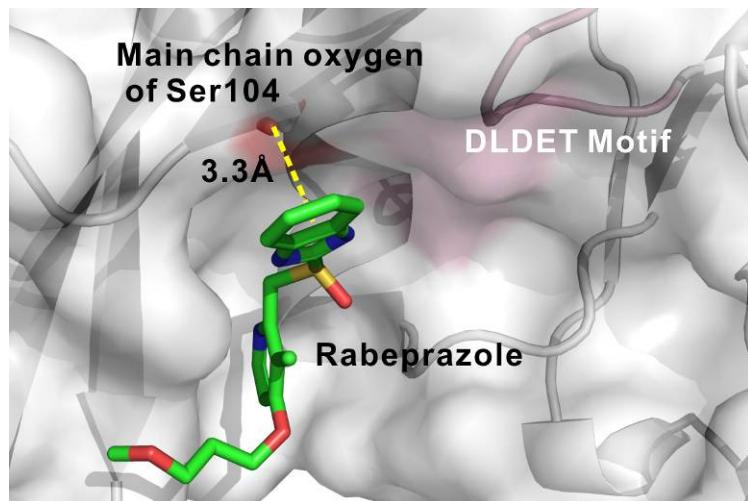
Figure 4-8:   Model of (S)-rabeprazole bound to Scp1.[25]

---

[25] The distance between the main chain oxygen of Ser104 and the benzoimidazole ring is 3.3 Å.

Selectivity of rabeprazole

To evaluate the specificity of rabeprazole as a Scps inhibitor, we tested its inhibition against other Scp/Fcp family phosphatases including Fcp1 and Dullard, whose inactivation usually lead to cell death or improper development (Chambers et al. 1994; Archambault et al. 1997; Kobor et al. 1999; Satow et al. 2002). Such tests are not only important to evaluate the potential value of rabeprazole in biological applications, but also crucial for confirming the mechanism of the inhibition. In the Scp/Fcp phosphatase family, all enzymes share the same signature motif DXDX(T/V) for phosphoryl transfer. They all belong to the HAD superfamily (also called DXDX(T/V) superfamily) type III subfamily which predominantly act on protein substrates (Allen et al. 2004).

The possible off-target inhibitory effect of rabeprazole was tested on the closest homologues of Scps: *Schizosaccharomyces pombe* (*S. pombe*) Fcp1 (Hausmann et al. 2003; Ghosh et al. 2008) and human Dullard. His-tagged *S. pombe* Fcp1 and His-GST-tagged Dullard were purified to homogeneity and exhibit stable phosphatase activity (**Figure 4-9**). The steady-state kinetic studies of these enzymes were performed. The K_m for Fcp1 and Dullard were determined to be 15 ± 3 mM and 10 ± 2 mM respectively, comparable with previously reported values (Hausmann et al. 2003; Kim et al. 2007). The concentration-response of inhibition was tested for both proteins using substrate concentrations comparable to the K_m. No inhibition was observed for either enzyme when up to 0.1 mM rabeprazole was used (**Figure 4-10a**).
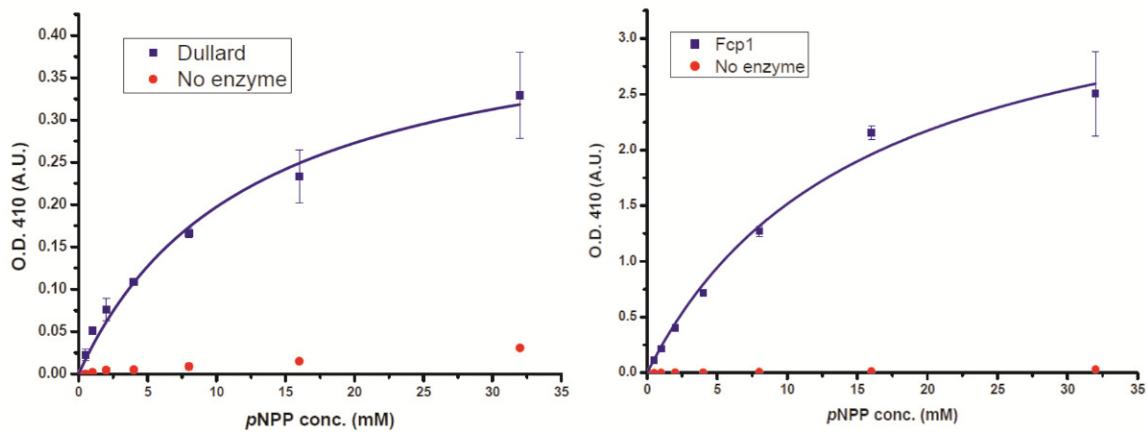
Figure 4-9: Steady-state kinetics of His-tagged Fcp1 and His-GST-tagged Dullard.[26]

The observed specificity of rabeprazole can be explained structurally. The major interactions of rabeprazole with Scp are attributed to the insertion domain which is unique to the Scp/Fcp family and poorly conserved among the family member (**Figure 4-2c**). For example, the three-stranded β sheet insertion domain in Scp1 is accessible for substrate recognition, whereas in Fcp1 it exhibits a different secondary structure and is buried by a helical insertion domain, suggesting a different binding interface between Fcp1 and the CTD with phospho.Ser$_2$ (**Figure 4-10b**). Even though the Dullard structure has yet to be determined, the secondary structure is predicted to be different from the three-stranded β sheet as observed in Scp1. It is highly likely that rabeprazole does not inhibit human Dullard due to a different insertion domain in Dullard.

Rabeprazole also exhibits no inhibition of λPPase, a Ser/Thr PPases outside the Scp/Fcp family, at up to 1 mM concentration (Data not shown).

---

[26] The red data points in both plots indicate the no enzyme control of the *p*NPP reaction.

a

Rabeprazole inhibition towards Fcp1

Rabeprazole inhibition towards Dullard

b

Scp1
insertion domain

Scp1

Fcp1 helical
insertion domain

Fcp1
insertion domain

Mg²⁺
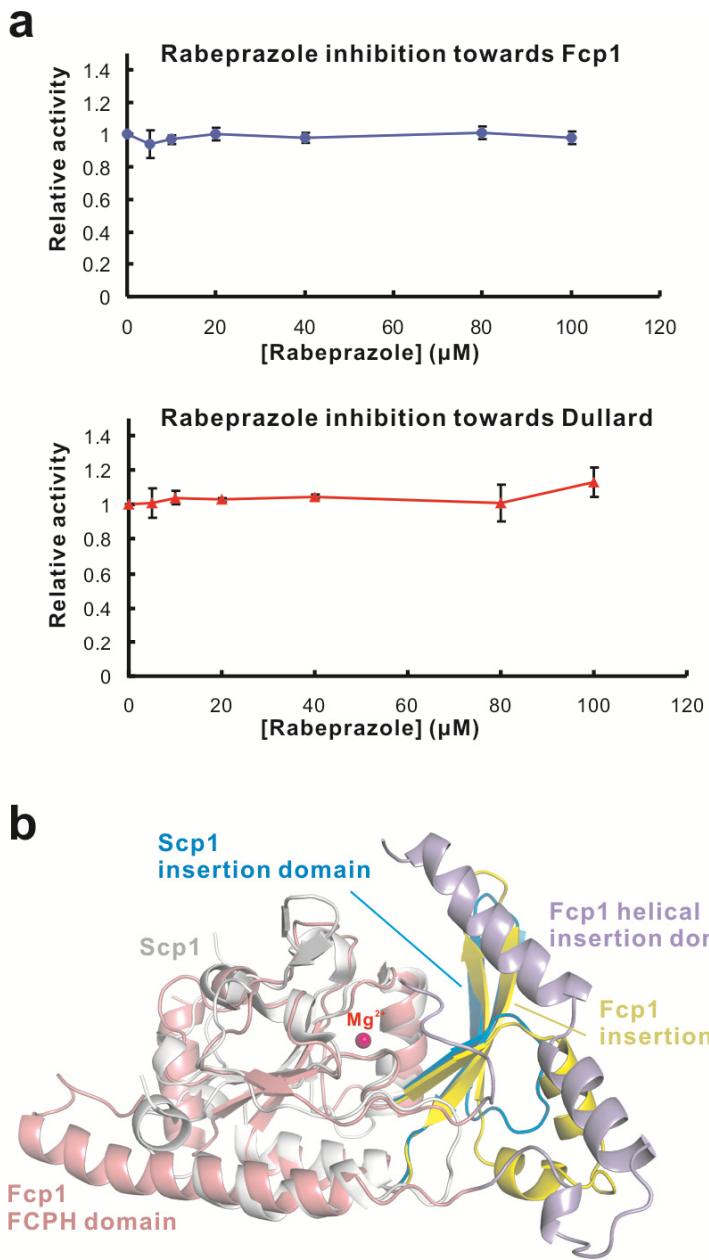
Fcp1
FCPH domain

Figure 4-10: Inhibition of Fcp1 and Dullard by rabeprazole.[27]

[27] (a) Concentration-response of rabeprazole inhibition towards *S. pombe* Fcp1 and human Dullard tested by the *p*NPP assay. (b) Superimposition of Scp1 (white) and *S. pombe* Fcp1 FCPH domain (pink, PDB code: 3ef0). The catalytic core folds of the two proteins are similar. The three-stranded β sheet insertion domain of Scp1 (cyan) is accessible for substrate recognition, whereas the insertion domain of Fcp1 (yellow) exhibits different secondary structure and is buried by an additional helical insertion domain (light blue).

We next determined if rabeprazole analogues also exhibit inhibitory effect on Scps (**Figure 4-11**). Rabeprazole sulfone, a metabolite of rabeprazole (Klotz 2000) did not show any inhibition in our concentration-response assay (data not shown). The absence of inhibition might stem from the loss of cation-π interaction between the sulfoxide group and Tyr158 with the extra oxygen in rabeprazole sulfone. The result further demonstrates that the sulfoxide group is important for the Scp1-rabeprazole interaction and should be kept intact.

Rabeprazole N-oxide (**Figure 4-11**), a contaminate during the rabeprazole synthesis process (Reddy et al. 2009), also exhibited no inhibition towards Scp1. It is likely that the change of the methyl pyridine ring interrupts or weakens the van del Waals interactions originally existing between the hydrophobic pocket residues and rabeprazole.

Likewise, lansoprazole (marketed as Prevacid in U.S., **Figure 4-11**) which only differs from rabeprazole at the tail portion, showed no inhibition, probably because the electrons of pyridine ring are dramatically influenced by converting the ether to the trifluoromethyl group. It is equally likely that the shorter tail and significant change in hydrophobicity disrupt important contacts. Since the pyridine ring is important in making hydrophobic interactions with the protein, the changed electron distribution or changed hydrophobicity will likely weaken the hydrophobic interactions.

Taken together, these results not only demonstrate the excellent specificity of rabeprazole, but also validate our strategy of targeting the hydrophobic binding pocket adjacent to the active site. To the best of our knowledge, rabeprazole is the first reported selective inhibitor for the Scp/Fcp family proteins, and also for the type III HAD subfamily proteins.

Lansoprazole

Rabeprazole sulfone

Rabeprazole N-oxide
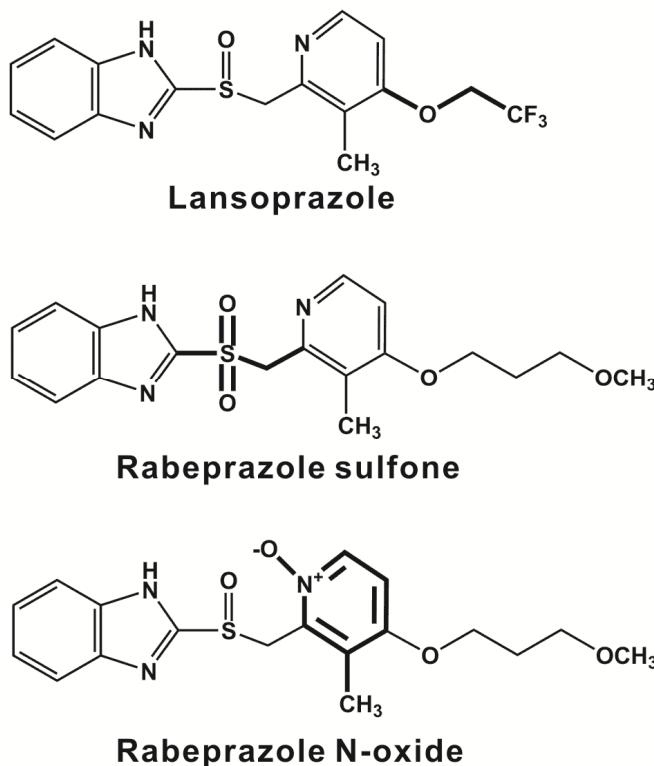
Figure 4-11: Structures of rabeprazole analogues.[28]

Considerations on the clinic application of rabeprazole

Rabeprazole is the active ingredient of an FDA-approved antiulcer drug AcipHex which is used to treat gastroesophageal reflux disease (GERD). In acidic environment such as lumen of gastric parietal cells, rabeprazole inhibits its target H^+/K^+ ATPase by forming a covalent bond with the active site cysteine under acidic condition (pH around 1). Rabeprazole behaves as a pro-drug for GERD and only becomes activated in highly acidic condition through two protonations and a subsequent spontaneous rearrangement to form the active sulfenamide (**Figure 4-12**) (Roche 2006). However, we reason that rabeprazole does not inhibit Scp1 through this mechanism, since no cysteines participate in the catalysis of Scp1 and no covalent bond between rabeprazole and Scp1 was

---

[28] The differences between the analogues and rabeprazole are indicated by thick lines.

observed in the complex structure (crystals obtained at neutral pH). This reasoning is corroborated by the fact that lansoprazole, another a proton pump inhibitor that shares the same mechanism with rabeprazole in treating GERD, does not inhibit Scp1.
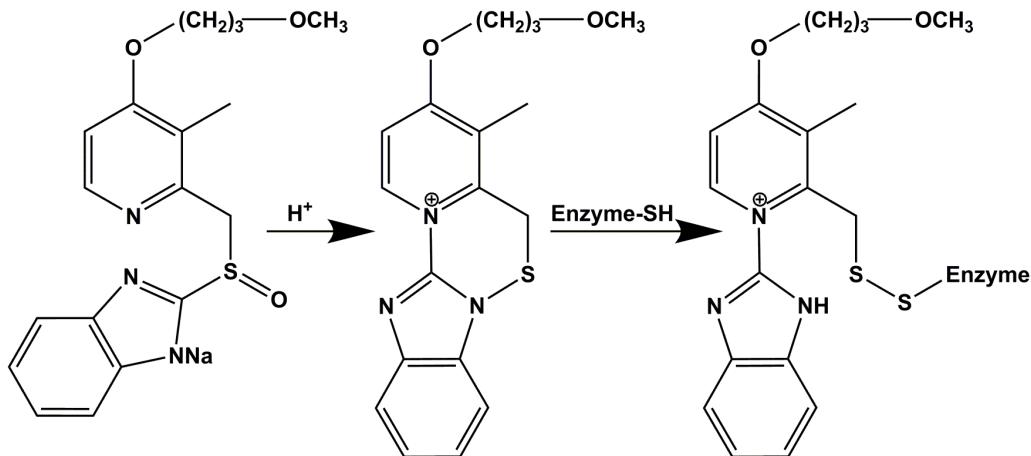
Figure 4-12: Mechanism of rabeprazole sodium inhibition of Cys-based enzymes.

Further development of inhibitors with higher binding affinity

Based on the complex structure of Scp1 and rabeprazole, further development of Scps specific inhibitors with higher affinity should be attainable. The important groups of this lead compound contributing to the binding are the methyl pyridine ring and methyl sulfinyl moieties. The benzimidazole ring seems to play little positive role if not hindering the compound binding, therefore, its replacement by other groups should be desirable. In our structure, no interaction of rabeprazole with the DXDX motif was observed. Further optimization can be explored by keeping the methyl pyridine ring and methyl sulfinyl groups of rabeprazole intact but changing other portions to explore interactions with the active site pocket.

In conclusion, we successfully identified the first selective lead compound for the inhibition of Scps in our high-throughput screening efforts. Our kinetic and structural

analyses further confirmed its inhibitory effect towards Scp1 but not its close family members, Fcp1 and Dullard. Thus, rabeprazole represents the first specific inhibitor for Scp/Fcp family phosphatases and also the first inhibitor for type III HAD family members. The complex structure of Scp1 bound with rabeprazole clearly shows the binding mode of the small molecule. As we expected, rabeprazole binds to the unique hydrophobic pocket of Scps which, in natural context, binds to Pro_3 of the CTD peptide. Since the hydrophobic pocket is located at the insertion domain which is unique to Scp/Fcp family members and shows diverse sequence among those family members, cross inhibition of other family members is prevented. The pyridine ring and sulfoxide groups are essential for this compound binding to Scp1. Other groups which may not substantially contribute to the binding can be optimized to make more potent inhibitors. The present study provides a starting point for the development of new inhibitors of Scps that can induce neuronal stem cell differentiation, which is of great interest in both basic and clinical research.

MATERIALS AND METHODS

Materials

*p*NPP was purchased from Fluka, Sigma-Aldrich. CTD peptides were purchased from Anaspec. SYPRO orange dye was purchased from Invitrogen. The *S. pombe* Fcp1 cDNA was purchased from National BioResource Project (NBRP)-Yeast Genetic Resource Center (Osaka City University, Osaka, Japan). Malachite green reagent was purchased from BIOMOL, Enzo Life Sciences. Rabeprazole and rabeprazole analogues were purchased from Toronto Research Chemicals, Inc., and the identity and purity of each compound was confirmed by our in-house mass spec facility.

Cloning, protein expression and purification

Wildtype or mutant human Scp1, residues 77–256, was sub-cloned and expressed in *E.coli* BL21(DE3) strain using pHIS8 vector encoding a thrombin cleavable N-terminal octa-histidine tag (Jez et al. 2000). The expression of Scp1 was induced in *E.coli* at low temperature (16 °C) with 250 µM IPTG and allowed to accumulate to high levels overnight (16 hr). Cells were harvested and re-suspended in lysis buffer (0.5 M NaCl, 15 mM imidazole, 50 mM Tris-HCl pH 8.0, 10% (v/v) glycerol and 0.1% (v/v) Triton X-100). Then the cells were sonicated. After sonication, cell debris was removed by centrifugation. The supernatant was incubated with Ni^{2+}-NTA affinity beads at 4 ℃ for 20 min, then loaded onto a column and washed extensively with the lysis buffer minus Triton X-100. Nearly homogeneous Scp1 was eluted with the lysis buffer minus Triton X-100 but including 150 mM imidazole. Eluted protein was dialyzed into 25 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM β-mercaptoethanol and stored at a concentration of 2 mg/ml at 4 °C. Samples for protein x-ray crystallography were subjected to thrombin processing to remove the octa-histidine tag followed by size exclusion chromatography for final cleanup. These samples were stored in 20 mM Tris-HCl pH 8.0, 50 mM NaCl, 10 mM β-mercaptoethanol at a concentration of 8–10 mg/ml at –80 °C.

Wildtype *S. pombe* Fcp1, residues 148–641, was also sub-cloned into pHIS8 vector. Wildtype human Dullard, residues 46–244, was sub-cloned into engineered pET28 vector encoding a 3C protease cleavable N-terminal His-GST tag. The expression and purification of human Dullard and yeast Fcp1 followed the same protocol as Scp1 with minor modifications. The protein samples were stored in the buffer containing 20 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM β-mercaptoethanol and 10% glycerol.

High-throughput screening

The high-throughput screening was performed in 384-well plates using pNPP as the substrate. We screened a library of NIH clinical collection (~400 compounds) and spectrum collection (2000 compounds) in our local high-throughput facility TI3D. The proteins were pre-incubated in the presence of 50 μM compounds in 0.8% DMSO at room temperature for 10 min. All the screening reactions (10 μl) were carried out at 37 °C in Master Buffer (50 mM Tris-acetate pH 5.5, 10 mM MgCl$_2$, 0.02% Triton X-100, 0.5% DMSO) with 25 ng of Scp1, and pNPP concentration at its K_m (6 mM). The reactions were quenched by addition of 40 μl of 0.25 N NaOH after 10 min of reaction. Release of pNP was determined by measuring absorbance at 410 nm.

***p*NPP assay**

The activity of each phosphatase toward pNPP in the absence or presence of inhibitor was measured in Assay Buffer (50 mM Tris-acetate pH 5.5, 10 mM MgCl$_2$, 0.02% Triton X-100, 1% DMSO) with appropriate amount of protein (50 ng of Scp1 or Scp1 mutants, 5 μg of Dullard or Fcp1) at 37 °C in 20 μl volume. When inhibitor was included, protein and inhibitor were pre-incubated in Assay Buffer at room temperature for 10 min in 15 μl volume. After 10–15 min of reaction, the reactions were quenched by adding 80 μl of 0.25 N NaOH. Released pNP was quantified by measuring absorbance at 410 nm.

Malachite green assay

The activity of Scp1 toward 17-mer Ser$_5$-phosphorylated peptide ($S_{a5}P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}S_{b5}P_{b6}S_{b7}Y_{c1}S_{c2}P_{c3}T_{c4}$**phospho.$S_{c5}$**$P_{c6}S_{c7}$) in the absence or presence of inhibitor was measured in Assay Buffer with 5 ng of Scp1 at 37 °C in 20 μl volume. When inhibitor was included, protein and inhibitor were pre-incubated in Assay

Buffer at room temperature for 10 min in 15 μl volume. After 3 min of reaction, the reactions were quenched by adding 40 μl of malachite green reagent. The release of free inorganic phosphate was determined by measuring the absorbance at 620 nm.

Determination of steady-state and inhibition constants

The raw absorbance readings were exported to Excel for further processing. For all reactions, a corresponding mock reaction containing everything but the protein was set up. The absorbance reading of a mock reaction was treated as background and used to subtract the raw absorbance readings of the corresponding true reaction, resulting in a processed absorbance reading caused by the product of the reaction. In the following text the term 'absorbance' is used to refer to 'processed absorbance reading'. When necessary, the absorbance was converted to concentration of product using a standard curve that was established beforehand.

To determine the IC_{50} of rabeprazole towards Scp1 (or Scp1 mutants), the relative activity of enzyme were determined by dividing the absorbance at each inhibitor concentration by the absorbance of positive control ($[I] = 0$). The relationship between the relative activity of enzyme and inhibitor concentration was fitted to the equation (1),

$$Y = \frac{1}{1 + (\frac{x}{IC_{50}})^{H}} \tag{1}$$

where Y is relative activity of enzyme, x is the inhibitor concentration, H is hill coefficient.

To determine steady-state constants (V_{max} and K_{m}) and inhibition constants (K_{i} and K_{i}'), the initial rate of reaction was obtained by dividing the product concentration by incubation time. The linearity of kinetics during such periods of time had been confirmed. The V_{max} and K_{m} values were obtained by fitting initial rates v.s. substrate

95

concentrations (in the absence of inhibitor) into Michaelis-Menten equation. Once the V_{max} and K_m are determined, the initial rates v.s. substrate concentrations in the presence of inhibitor at its IC_{50} were fitted into equation (2):

$$V = \frac{V_{max}[S]}{\alpha K_m + \alpha'[S]} \qquad (2)$$

where V and $[S]$ are initial rate and substrate concentration, respectively; V_{max} and K_m are pre-determined; α and α' are used to calculate K_i and K_i' using equations (3) and (4):

$$\alpha = 1 + \frac{[I]}{K_i} \qquad (3)$$

$$\alpha' = 1 + \frac{[I]}{K_i{}'} \qquad (4)$$

All non-linear fittings were carried out using the 'nlinfit' function of Matlab.

Differential scanning fluorimetry

The method was modified based on a published protocol (Niesen et al. 2007). Wildtype Scp1 was mixed with rabeprazole in the reaction containing 20 mM HEPES pH7.5, 50 mM NaCl, 1% DMSO, and 5× SYPRO orange dye. The final concentrations of protein and rabeprazole were 20 μM and 100 μM, respectively. The unfolding process of protein was monitored during the temperature increase from room temperature to 85 °C by LightCycler 480 machine (Roche). The protein only, protein with 1% DMSO and protein with 1 mM peptide were used as controls. The denature process of the protein was fitted into a monophase graph and the T_m was derived using the LightCycler 480 Protein Melting Software.

Crystallization and structure determination

Wildtype Scp1 was crystallized in 30% PEG 3350 and 0.2 M magnesium acetate. Crystals were transferred to a stabilizer consisting of the same buffer condition and 0.1–1 mM inhibitor. After soaking for 2–12 hrs, crystals were transferred to a cryo-protecting

stabilizer containing 30% (v/v) glycerol, 25% (w/v) PEG 3350, and 0.2 M magnesium acetate. After a brief period of equilibration, crystals were frozen in nylon loops in liquid nitrogen and stored in liquid nitrogen prior to data collection. Crystallographic data were collected at 100 K on beam-line 5.0.2 of the Advanced Light Source (ALS). Diffraction data were processed with HKL2000 (Otwinowski et al. 1997). The complex crystal structures of Scp1 were determined by molecular replacement (MR) using the Scp1D96N structure as a search model (PDB code: 2ghq) using the program PHASER (McCoy et al. 2007) available in the CCP4 software package (CCP4 1994). MR solutions were refined using REFMAC5 (Vagin et al. 2004), reserving 5% of the measured and reduced structure factor amplitudes as an unbiased test set for cross validation (R_{free}) (Brunger 1992). The inhibitor library was created by SKETCHER also available in the CCP4 software package. The model was built by COOT (Emsley et al. 2004) and refined by REFMAC5. The buried contact area of Scp1 by the compound and shape complementarity (Sc) were calculated by AREAIMOL (Lee et al. 1971) and SC (Lawrence et al. 1993) available in the CCP4 software package.

ACCESSION CODE

Coordinates of the Scp1-rabeprazole complex structure have been deposited in the Protein Data Bank with the accession number 3PGL.

REFERENCES

Allen, K. N. and D. Dunaway-Mariano (2004). "Phosphoryl group transfer: evolution of a catalytic scaffold." *Trends Biochem Sci* **29**(9): 495-503.

Archambault, J., R. S. Chambers, M. S. Kobor, Y. Ho, M. Cartier, D. Bolotin, B. Andrews, C. M. Kane and J. Greenblatt (1997). "An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae." *Proc Natl Acad Sci U S A* **94**(26): 14300-14305.

Brunger, A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." *Nature* **355**(6359): 472-475.

Burley, S. K. and G. A. Petsko (1986). "Amino-aromatic interactions in proteins." *FEBS Lett* **203**(2): 139-143.

CCP4 (1994). "Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography." *Acta Cryst.* **D50**: 760-763.

Chambers, R. S. and M. E. Dahmus (1994). "Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II." *J Biol Chem* **269**(42): 26243-26248.

Cho, H., T. K. Kim, H. Mancebo, W. S. Lane, O. Flores and D. Reinberg (1999). "A protein phosphatase functions to recycle RNA polymerase II." *Genes Dev* **13**(12): 1540-1552.

Cho, H., S. Ramaswamy and B. V. Plapp (1997). "Flexibility of liver alcohol dehydrogenase in stereoselective binding of 3-butylthiolane 1-oxides." *Biochemistry* **36**(2): 382-389.

Dougherty, D. A. (1996). "Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp." *Science* **271**(5246): 163-168.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132.

Gallivan, J. P. and D. A. Dougherty (1999). "Cation-pi interactions in structural biology." *Proc Natl Acad Sci U S A* **96**(17): 9459-9464.

Ghosh, A., S. Shuman and C. D. Lima (2008). "The structure of Fcp1, an essential RNA polymerase II CTD phosphatase." *Mol Cell* **32**(4): 478-490.

Hausmann, S. and S. Shuman (2003). "Defining the active site of Schizosaccharomyces pombe C-terminal domain phosphatase Fcp1." *J Biol Chem* **278**(16): 13627-13632.

Jez, J. M., J. L. Ferrer, M. E. Bowman, R. A. Dixon and J. P. Noel (2000). "Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase." *Biochemistry* **39**(5): 890-902.

Kamenski, T., S. Heilmeier, A. Meinhart and P. Cramer (2004). "Structure and mechanism of RNA polymerase II CTD phosphatases." *Mol Cell* **15**(3): 399-407.

Kim, Y., M. S. Gentry, T. E. Harris, S. E. Wiley, J. C. Lawrence, Jr. and J. E. Dixon (2007). "A conserved phosphatase cascade that regulates nuclear membrane biogenesis." *Proc Natl Acad Sci U S A* **104**(16): 6596-6601.

Klotz, U. (2000). "Pharmacokinetic considerations in the eradication of Helicobacter pylori." *Clin Pharmacokinet* **38**(3): 243-270.

Kobor, M. S., J. Archambault, W. Lester, F. C. Holstege, O. Gileadi, et al. (1999). "An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae." *Mol Cell* **4**(1): 55-62.

Lawrence, M. C. and P. M. Colman (1993). "Shape complementarity at protein/protein interfaces." *J Mol Biol* **234**(4): 946-950.

Lee, B. and F. M. Richards (1971). "The interpretation of protein structures: estimation of static accessibility." *J Mol Biol* **55**(3): 379-400.

Li, M., H. P. Phatnani, Z. Guan, H. Sage, A. L. Greenleaf and P. Zhou (2005). "Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1." *Proc Natl Acad Sci U S A* **102**(49): 17636-17641.

Ma, J. C. and D. A. Dougherty (1997). "The Cation-π Interaction." *Chem Rev* **97**(5): 1303-1324.

Majello, B. and G. Napolitano (2001). "Control of RNA polymerase II activity by dedicated CTD kinases and phosphatases." *Front Biosci* **6**: D1358-1368.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." *J Appl Crystallogr* **40**(Pt 4): 658-674.

Morais, M. C., W. Zhang, A. S. Baker, G. Zhang, D. Dunaway-Mariano and K. N. Allen (2000). "The crystal structure of bacillus cereus phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily." *Biochemistry* **39**(34): 10385-10396.

Niesen, F. H., H. Berglund and M. Vedadi (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." *Nat Protoc* **2**(9): 2212-2221.

Otwinowski, Z. and W. Minor (1997). "HKL: Processing of X-ray diffraction data collected in oscillation mode." *Methods Enzymol.* **276**: 307-326.

Reddy, G. M., K. Mukkanti, B. V. Bhaskar and P. P. Reddy (2009). "Synthesis of Metabolites and Related Substances of Rabeprazole, an Anti-Ulcerative Drug." *Synthetic Communications* **39**: 278-290.

Roche, V. F. (2006). "The chemically elegant proton pump inhibitors." *Am J Pharm Educ* **70**(5): 101.

Satow, R., T. C. Chan and M. Asashima (2002). "Molecular cloning and characterization of dullard: a novel gene required for neural development." *Biochem Biophys Res Commun* **295**(1): 85-91.

Scrutton, N. S. and A. R. Raine (1996). "Cation-pi bonding and amino-aromatic interactions in the biomolecular recognition of substituted ammonium ligands." *Biochem J* **319 (Pt 1)**: 1-8.

Vagin, A. A., R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long and G. N. Murshudov (2004). "REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use." *Acta Cryst.* **D60**(Pt 12 Pt 1): 2184-2195.

Visvanathan, J., S. Lee, B. Lee, J. W. Lee and S. K. Lee (2007). "The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development." *Genes Dev* **21**(7): 744-749.

Yeo, M., S. K. Lee, B. Lee, E. C. Ruiz, S. L. Pfaff and G. N. Gill (2005). "Small CTD phosphatases function in silencing neuronal gene expression." *Science* **307**(5709): 596-600.

Yeo, M., P. S. Lin, M. E. Dahmus and G. N. Gill (2003). "A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5." *J Biol Chem* **278**(28): 26078-26085.

Yu, X., J. P. Sun, Y. He, X. Guo, S. Liu, B. Zhou, A. Hudmon and Z. Y. Zhang (2007). "Structure, inhibitor, and regulatory mechanism of Lyp, a lymphoid-specific tyrosine phosphatase implicated in autoimmune diseases." *Proc Natl Acad Sci U S A* **104**(50): 19767-19772.

Zhang, G., A. S. Mazurkie, D. Dunaway-Mariano and K. N. Allen (2002). "Kinetic evidence for a substrate-induced fit in phosphonoacetaldehyde hydrolase catalysis." *Biochemistry* **41**(45): 13370-13377.

Zhang, M., G. N. Gill and Y. Zhang (2010). "Bio-molecular architects: a scaffold provided by the C-terminal domain of eukaryotic RNA polymerase II " *Nano Reviews* **1: 5502 - DOI: 10.3402/nano.v1i0.5502**.

Zhang, M., J. Liu, Y. Kim, J. E. Dixon, S. L. Pfaff, G. N. Gill, J. P. Noel and Y. Zhang (2010). "Structural and functional analysis of the phosphoryl transfer reaction mediated by the human small C-terminal domain phosphatase, Scp1." *Protein Sci* **19**(5): 974-986.

Zhang, S., L. Chen, Y. Luo, A. Gunawan, D. S. Lawrence and Z. Y. Zhang (2009). "Acquisition of a potent and selective TC-PTP inhibitor via a stepwise fluorophore-tagged combinatorial synthesis and screening strategy." *J Am Chem Soc* **131**(36): 13072-13079.

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." *Mol Cell* **24**(5): 759-770.

Chapter 5:   Potential Secondary Regulatory Site of Scp1

INTRODUCTION

Accurate transcription of genetic information is essential for cell survival. In eukaryotes, the major player of transcription is RNA polymerase II, responsible for the transcription of protein-coding genes as well as mammalian snRNA genes and yeast snoRNA genes (Egloff et al. 2008). The C-terminal domain (CTD) of RNA polymerase II forms a tail-like peptide chain extended from the catalytic core and resides closed to the RNA exit channel of the enzyme. In addition, a wealth of evidence reveals that CTD is an important mediator in coupling mRNA transcription, mRNA processing, DNA repair and other cellular processes (Egloff et al. 2008). This unusual structure serves as a docking platform for a variety of factors involved in transcriptional and co-transcriptional events.

Although the CTD undergoes very complicated modifications and regulates various cellular processes, the primary sequence of the CTD is surprisingly simple. The CTD consists of multiple tandem heptapeptide repeats of the consensus sequence: $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$. The number of repeats is different from species to species (Egloff et al. 2008). There is a minimum length of the CTD, which is required for normal cell growth and function in many species (Chapman et al. 2007). All the transcriptional events are highly related to the conformations and modifications, particularly phosphorylation patterns, of the CTD. Different phosphorylation patterns are the major component of the "CTD code" which orchestrates the function of RNA polymerase II (**Figure 5-1**). Moreover, different phosphorylation patterns could specifically recruit essential factors to the vicinity of genes, conferring success and possessiveness of transcriptional and co-transcriptional events.

103
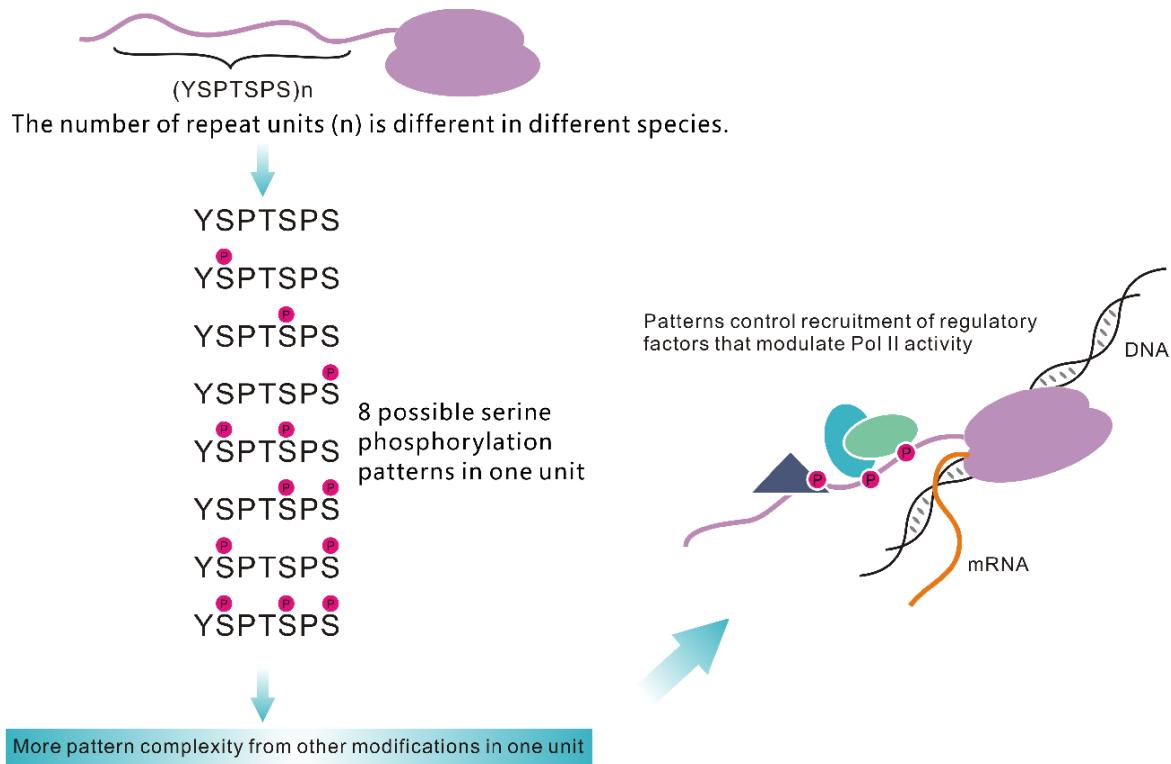
Figure 5-1:   Schematic diagram of "CTD code" and its important role in modulating RNA polymerase II activity.

Although there are five potential phosphorylation sites in the CTD consensus sequence, including Tyr_1, Ser_2, Thr_4, Ser_5 and Ser_7, CTD phosphorylation occurs mainly at residues Ser_2 and Ser_5 *in vivo* (**Figure 5-1**). In each round of transcription, Ser_5 phosphorylation occurs in promoter-proximal regions coincident with initiation and facilitates mRNA capping via recruiting capping enzymes (Cho et al. 1997; Komarnitsky et al. 2000). During the process in which the transcription complex moves away from the initiation site, Ser_5 gradually becomes hypophosphorylated, whereas Ser_2 gradually becomes hyperphosphorylated. Ser_2 phosporylation is the predominating pattern on both elongation and termination steps. Ser_2 phosphorylation ensures efficient 3'-RNA processing by triggering the recruitment of 3'-RNA processing machinery. Ser_2 and Ser_5

phosphorylation have been extensively studied; however, until recently the biological functions of phosphorylation at sites other than Ser_2 and Ser_5 have not been fully addressed. Only recently, Tyr_1, Thr_4 and Ser_7 have also been observed being phosphorylated *in vivo* in certain scenarios (Chapman et al. 2007; Hsin et al. 2011; Hintermair et al. 2012; Mayer et al. 2012). Two recent works suggest that Ser_7 phosphorylation occurs during transcription and is required for snRNA expression in humans. It is shown that Ser_7 phosphorylation occurs on both mRNA and snRNA genes and facilitates interaction with the snRNA gene-specific integrator complex (Chapman et al. 2007; Egloff et al. 2007).

A range of enzymes participate in the dynamic modifications of the CTD, including kinases and phosphatases responsible for the addition and removal of phosphates (Bataille et al. 2012). The CTD is principally phosphorylated by cyclin-dependent kinases (CDKs). Specifically, Ser_5 phosphorylation is mainly catalyzed by the general transcription factor IIH (TFIIH) which contains Cdk7/cyclin H subunits (Lu et al. 1992; Hengartner et al. 1998); Ser_2 phosphorylation is mainly catalyzed by the positive transcription elongation factor b (P-TEFb) which contains Cdk9/cyclin T subunits (Zhou et al. 2000; Shim et al. 2002). Intriguingly, it is suggested that Cdk9 also makes a contribution to Ser_5 phosphorylation (Garber et al. 2000; Gomes et al. 2006) and the relative contribution of TFIIH-associated Cdk7 varies between different genes based on experimental observations. Moreover, the CTD can also be phosphorylated at both Ser_2 and Ser_5 by Cdk8, and preferentially be phosphorylated at Ser_5 by mitogen-activated protein kinase 2 (MAPK2/Erk2). Recently, TFIIH-associated Cdk7 kinase has also been identified to phosphorylate Ser_7 *in vivo* (Glover-Cutter et al. 2009).

Dephosphorylation negatively controls the activity of RNA polymerase II and makes a significant contribution to the changes in the "CTD code". More importantly, it

is thought to be essential in recycling RNA polymerase II, because the CTD has to be dephosphorylated at the end of transcription in order to actively restart a new round of transcription. In humans, Transcription factor II F (TFIIF) interacting CTD phosphatase 1 (Fcp1), which is required for general transcription and cell viability, was the first discovered CTD-specific phosphatase with the catalytic preference for phospho.Ser_2. Later on, Ssu72 was identified to be a CTD-specific phosphatase, targeting on the phospho.Ser_5 (Krishnamurthy et al. 2004). Like Fcp1, Ssu72 is also considered to be a house-keeping gene whose deletion will result in cell death in yeast. On the contrary, a group of phosphatases called small CTD phosphatases (Scp's), which are also identified to dephosphorylate Ser_5, function at the epigenetic level and regulate the expression of a subset of genes. Scp's were identified to inhibit neuronal gene transcription in non-neuronal cells by acting as a co-repressor in REST/NRSF complex (Yeo et al. 2005). Scp's are expressed broadly in non-neuronal tissues but are largely excluded from the adult nervous system. The expression pattern of Scp's in human tissues parallels with that of REST/NRSF [which stands for repressor element 1 (RE-1)-silencing transcription factor/neuron-restrictive silencer factor] complex, the best-characterized transcription factors that regulate the expression of neuronal genes globally. Interestingly, dominant negative form of Scp1 (Scp1D96N) can de-repress the neuronal differentiation.

Although a large body of knowledge regarding the function of CTD has accumulated, there remains many questions to be answered in order to fully decipher the CTD code, for example, how the CTD enzymes are targeted to regulate general or gene-specific transcription, how Scps respond to the bivalent modification marks, etc. Previously, we have shown the complex structures of Scp1 and CTD-derived peptides, providing a detailed binding mode of the CTD. Although CTD peptides with different lengths and phosphorylation patterns were soaked into the Scp1 crystals, the same

binding mode in multiple crystals was observed. The $S_2P_3T_4(\textbf{p.S}_5)$ adopts a beta turn conformation which is recognized by the active site groove of Scp1. The *trans* Pro_3 at the N-terminal of the phospho.Ser_5, which is accommodated by the adjacent hydrophobic pocket, is crucial for the substrate specificity. However, one point worth mentioning is that the CTD is a very long tail and thus may have larger content area with Scp1. Therefore, in this study, we extended our studies to further investigate the binding mode of the CTD with Scp1 as well as how Scp1 functions on the CTD. Interesting observations were made based on X-ray crystallographic study. In addition, detailed kinetic studies were performed to further demonstrate the functional mode of Scp1 on the CTD.

RESULTS AND DISCUSSION

Complex structures of Scp1D96N mutant and the CTD-derived peptides

Obtaining the complex structure of Scp1 and its substrate CTD is of necessity to understand how Scp1 recognizes and regulates the CTD. In our previous study, we crystallized the dominant negative mutant of Scp1, Scp1D96N, and the CTD peptides through crystal soaking method (Zhang et al. 2006). These peptides were different in length and modification (**Table 5-1**). The structures of multiple crystals complexed with different peptides were solved. There are two interesting observations from these crystal structures. Firstly, consistent with the kinetic study that Scp1 is much more active toward phospho.Ser_5, only the phospho.Ser_5 is observed in the active site whereas the phospho.Ser_2 flips out of the active site, making no direct interaction with the protein (Zhang et al. 2006). Secondly, in both 9-mer and 14-mer phospho.Ser_5 peptides, only eight residues $[S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\textbf{p.S}_{\textbf{b5}})P_{b6}S_{b7}]$ showed strong electron density (Zhang et al. 2006), indicating that this segment is stably bound by the protein.

Number	Length	Sequence
		Previous Structural Study
1	9-mer	$P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
2	9-mer	$P_{a6}S_{a7}Y_{b1}(\mathbf{p.S_{b2}})P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
3	14-mer	$Y_{a1}S_{a2}P_{a3}T_{a4}(\mathbf{p.S_{a5}})P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
4	14-mer	$Y_{a1}S_{a2}P_{a3}T_{a4}S_{a5}P_{a6}S_{a7}Y_{b1}(\mathbf{p.S_{b2}})P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
5	14-mer	$Y_{a1}(\mathbf{p.S_{a2}})P_{a3}T_{a4}S_{a5}P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
		Current Structural Study
6	17-mer	$S_{z5}P_{z6}S_{z7}Y_{a1}S_{a2}P_{a3}T_{a4}S_{a5}P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
7	20-mer	$S_{z2}P_{z3}T_{z4}S_{z5}P_{z6}S_{z7}Y_{a1}S_{a2}P_{a3}T_{a4}S_{a5}P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
		Current Kinetic Study
8	14-mer	$Y_{a1}S_{a2}P_{a3}T_{a4}S_{a5}P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
6	17-mer	$S_{z5}P_{z6}S_{z7}Y_{a1}S_{a2}P_{a3}T_{a4}S_{a5}P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$
9	17-mer	$S_{z5}P_{z6}(\mathbf{p.S_{z7}})Y_{a1}S_{a2}P_{a3}T_{a4}S_{a5}P_{a6}S_{a7}Y_{b1}S_{b2}P_{b3}T_{b4}(\mathbf{p.S_{b5}})P_{b6}S_{b7}$

Table 5-1:   CTD peptides used in the previous and current studies.

But we reasoned that since the CTD is very long, there may be extended contact area between the CTD and Scp1 in the cells. To test this hypothesis, Scp1D96N mutant crystals were soaked in the buffer containing 17-mer or 20-mer synthetic peptide with a single phospho.Ser_5 (**Table 5-1**). These two peptides contain two full repeats and several more residues at the N-terminus. Expectantly yet still intriguingly, in both crystals, we observed strong and beautiful electron density of seventeen residues from Ser_{z5} to Ser_{b7} (**Figure 5-2**). The crystallographic statistics of the Scp1D96N and 17-mer peptide complex are summarized in **Table 5-2**. Compared with the previous complex structure of Scp1D96N and the 14-mer peptide, the complex structure of Scp1D96N and the 17-mer phospho.Ser_5 peptide has nine more residues clearly shown, although only three more residues ($S_{z5}P_{z6}S_{z7}$) were included in the 17-mer peptide. The newly obtained complex structures implied the existence of larger contact area and more stable binding formed between Scp1 and the CTD. It is surprising to see such large contact area by extending only three residues of the peptide. It is possible that these three additional residues make

108

important interactions with the protein; however, it is also likely that these three residues are not very important *per se* but facilitate the interaction between other residues and the protein. Besides the original interactions observed in the previous structures, the additional eight residues largely form hydrophobic interactions with the protein. The phenol ring of Tyr_{b1} of the peptide and the side chain of Ile120 of Scp1 forms amino-aromatic interaction (**Figure 5-3**). The Pro_{a6} and the imidazole ring of His125 form a pi-stacking interaction (**Figure 5-3**). There are also some hydrophilic interactions involved. For instance, the side chain of Ser_{a7} is 2.6 Å away from the carbonyl oxygen of Asp121, indicating the possibility of forming hydrogen bonds (**Figure 5-3**). The total buried surface of the protein by binding to the peptide is about 701.3 $Å^2$, and it has a surface complementarity (Sc) value of 0.814.
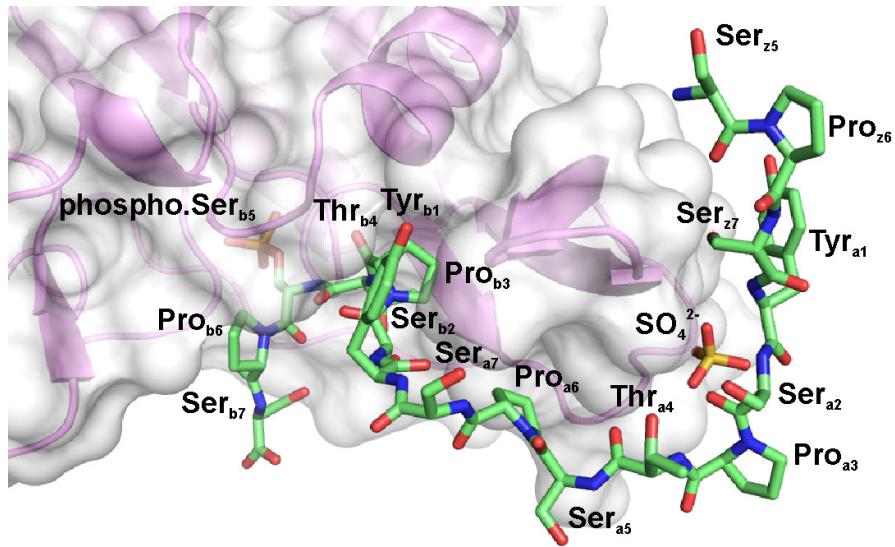
Figure 5-2:   Complex structure of Scp1 and the 17-mer peptide.

Interestingly, in the crystal soaked with 17-mer peptide, a tetrahedral shaped electron density was observed in between the protein and the peptide. In the crystallization condition, there was an excess amount of ammonium sulfate as the

precipitant, therefore we built a sulfate group into the density. However, when considering the environment in the cell, this sulfate-binding pocket might incorporate a phosphate group of the phosphorylated CTD. If so, this interaction might position Scp1 in the vicinity of other phosphorylation sites on the CTD for essentially higher local concentration of substrate. Because of the close proximity between the observed sulfate group and Ser_7, it is reasonable to hypothesize that phospho.Ser_7 might occupy this pocket and recruit Scp1 to the CTD. Moreover, it also possible that such phospho.Ser_7 might function as a docking site for Scp1 so that Scp1 can catalyze the dephosphorylation of Ser_5 on multiple sites of the CTD without dissociating from it (**Figure 5-4**). This notion is strongly fortified with the recent discovery that TFIIH can phosphorylate both Ser_5 and Ser_7, making the co-existence of phospho.Ser_5 and Ser_7 a very common species. Therefore, it is highly plausible that Ser_7 phosphorylation might play a role in the phosphorylation state of Ser_5.
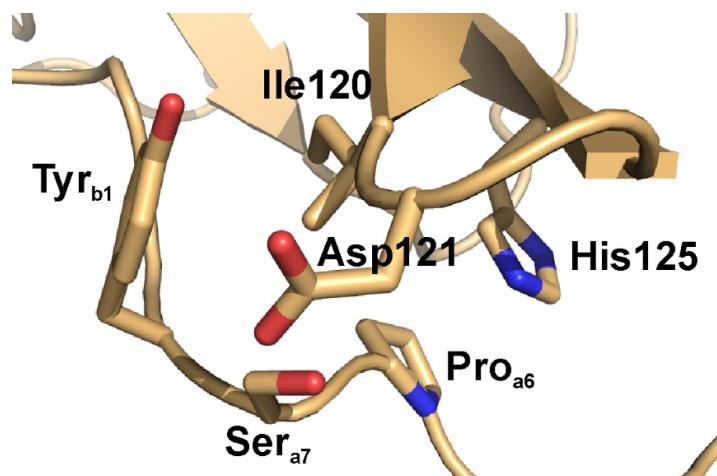
Figure 5-3:   Some potential interactions between the 17-mer peptide and Scp1.
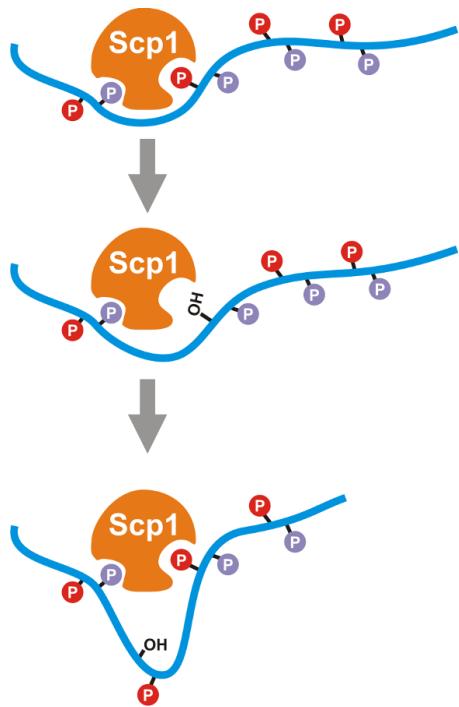
Figure 5-4:   Proposed processivity model of Scp1.

Kinetic analysis of the possible processivity model of Scp1 dephosphorylation

To initially examine whether the phospho.Ser_7 can contribute to the substrate binding, we tested the steady-state kinetics of Scp1 toward these two peptide substrates: one with both Ser_5 and Ser_7 phosphorylated, and the other with only Ser_5 phosphorylated (**Table 5-1**). If phospho.Ser_7 enhances the reactivity of CTD peptide towards Scp1, some increase of activity, most likely through decreased K_m, is expected. However, no difference was observed by comparing the kinetic parameters (**Figure 5-5**). There are two possibilities that can account for the data. One is that the Ser_7 phosphorylation has no effect on the substrate binding since the alternative binding site may not recognize phospho.Ser_7 at all. The other (unlikely but intriguing explanation) is that phospho.Ser_7 binds to the alternative binding site extremely tight so that the effect of this binding is not apparent in a multi-turnover reaction. If the second scenario is true, we would expect

dramatic processivity of Scp1 in true processivity assays. To perform these assays, phosphorylated substrates with multiple repeats were prepared.

GST-CTD was used as a model substrate in our processivity assays. In this system, yeast full length CTD was expressed as a fusion protein with GST to mimic the native Rpb1-CTD as well as ensure good solubility. Previous studies have shown that Erk2-treated GST-CTD is phosphorylated at Ser_5 in each repeat, and TFIIH-treated GST-CTD is phosphorylated at both Ser_5 and Ser_7. If the kinase treatment is done in the presence of radio-labeled ATP, $[\gamma\text{-}^{32}P]ATP$, we can then monitor the dephosphorylation of these two substrates by radioautography. The yeast GST-CTD was expressed in *E.coli*, purified to ≥90% homogeneity, and fully treated with TFIIH or Erk2 in the presence of $[\gamma\text{-}^{32}P]ATP$. The treated substrates were used in the following experiments.
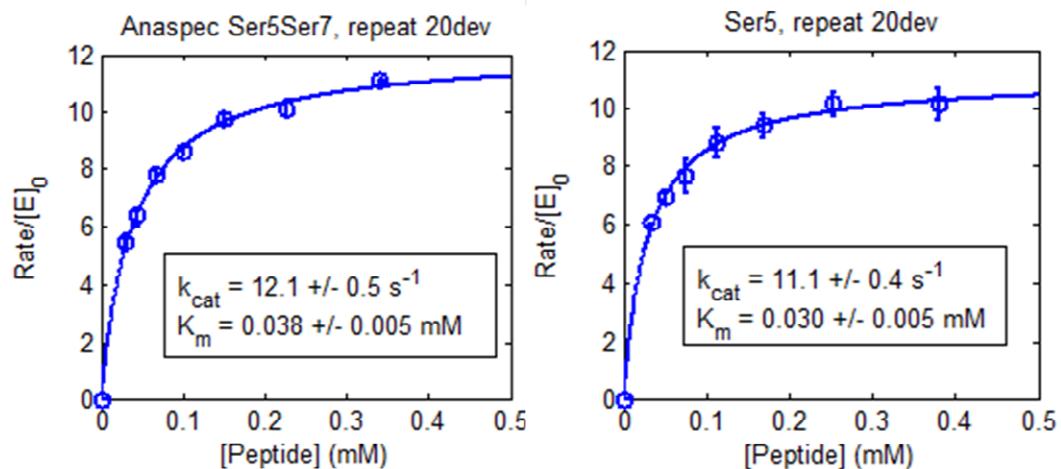
Figure 5-5: Steady-state kinetics of Scp1 toward 17-mer Ser_5Ser_7 phosphorylated peptide and Ser_5 phosphorylated peptide.

We first tested the activity of Scp1 toward these two GST-CTD substrates (phosphorylated by Erk2 and TFIIH) by varying the protein concentrations while keeping the substrate concentration constant. The result clearly showed that the TFIIH-treated substrate, which has both Ser_5 and Ser_7 phosphorylated, is a better substrate for Scp1

112

(**Figure 5-6**). But whether this observation is due to the binding between the alternative binding site and phospho.Ser$_7$ still remains a question. In order to test our processive model, a rapid dilution assay was performed. In this assay, Scp1 and TFIIH-treated GST-CTD were mixed at high concentration, incubated for a short period of time, and then the mixture was diluted 50-fold with the buffer. After dilution, a fraction of the mixture was taken out at each time point, quenched by protein gel loading dye, and resolved by SDS-PAGE. If our processive model is true, we would expect that the k_{off} rate between the GST-CTD and the protein is very low. And in this case, if we allow the protein and substrate to bind with each other at high concentration, they should form very stable complex. Even if the mixture is diluted, they should still associate with each other and the reaction can proceed in the diluted mixture. However, it is important to make sure that new complexes between the protein and substrate are not formed in the diluted mixture. To make sure this is the case, in the control group, diluted GST-CTD and Scp1 were directly mixed to the final volume.
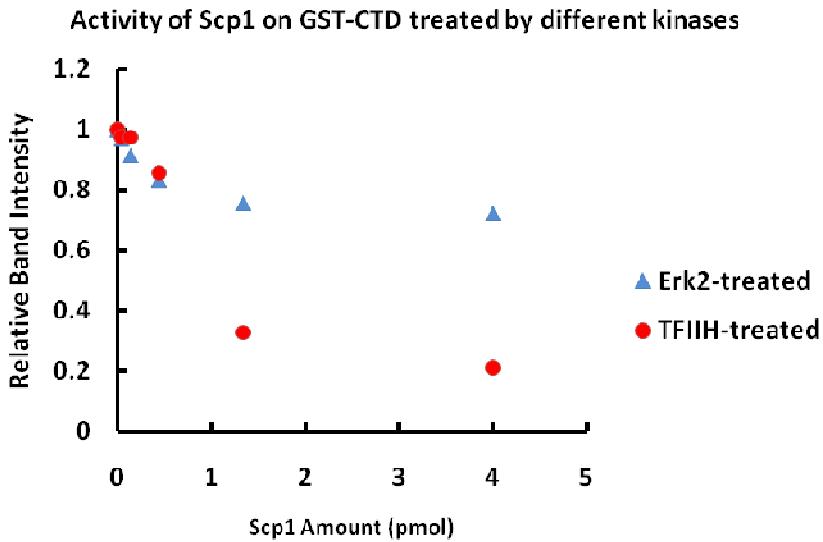
Figure 5-6:   Activity of Scp1 on the GST-CTD substrates treated with different kinases.

113

Our result showed that, after the dilution, the reaction no longer continued (**Figure 5-7**). If the binding pocket has a low k_{off}, we would expect to see decreasing of band intensity over time in the rapid dilution group. However, this result did not completely disprove our model as there are two explanations for this result. One possibility is that the enzyme is distributive. The other possibility is that: the enzyme is processive and is extremely fast. The second scenario means that either the enzyme and substrate do not meet with each other, or when they meet, the enzyme completely dephosphorylates the substrate rapidly.

To discriminate the above two possibilities, we performed a gel shift assay. The prerequisite of this assay is that GST-CTD substrates with different numbers of phosphorylations have different migration rates on SDS-PAGE and will appear at different positions on the gel. If our processive model is true, we should see two major species on the gel after dilution with one corresponding to fully phosphorylated GST-CTD, the other corresponding to hypophosphorylated GST-CTD. On the contrary, if the enzyme is distributive, then we would expect to see a smear on the gel, indicating various intermediates with different numbers of phosphorylation sites. In this assay, the density of each band on the gel was converted to a peak for easy observation and analysis. The result showed that there were different intermediates present in the mixture over time as indicated by the peak shift (**Figure 5-8**). This experiment, together with the previous supporting experiments, demonstrates that Scp1 is a distributive enzyme rather than a processive enzyme when dephosphorylating TFIIH-treated CTD substrate.
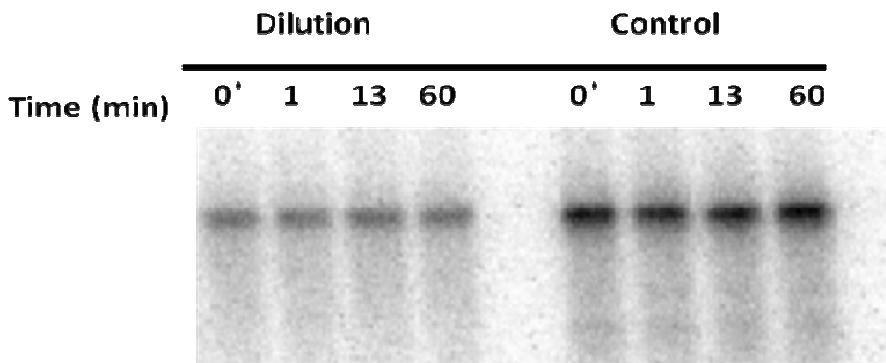
Figure 5-7: The result of rapid dilution assay showed that the reaction no longer continued after the dilution.[29]

The sulfate group may be a mimic of phosphate group from sites other than Ser$_7$

Though our kinetic analysis indicated that the sulfate group is not a mimic of phospho.Ser$_7$ to confer processivity of Scp1, it could be a mimic of phosphate group from other sites. Since more and more modifications on the CTD have been discovered, the combinatory effects between two or more modifications of the CTD on its binding proteins seem plausible. We then wanted to see if any of the combinatory effects were present that could potential affect the dephosphorylation by Scp1. The possible phosphorylation sites other than Ser$_7$ would be Tyr$_1$ and Ser$_2$. The synthetic peptides with both Ser$_5$ and Tyr$_1$ phosphorylated or Ser$_5$ and Ser$_2$ phosphorylated (**Table 5-1**) were used as substrates in our steady-state kinetic analysis. However, no substantial change of Scp1 activity was observed. Scp1 is not a processive enzyme in all of our assays, but it might be processive in cells in the presence of some binding partners.

---

[29] The protein and the GST-CTD peptide were pre-incubated for about 8 min.
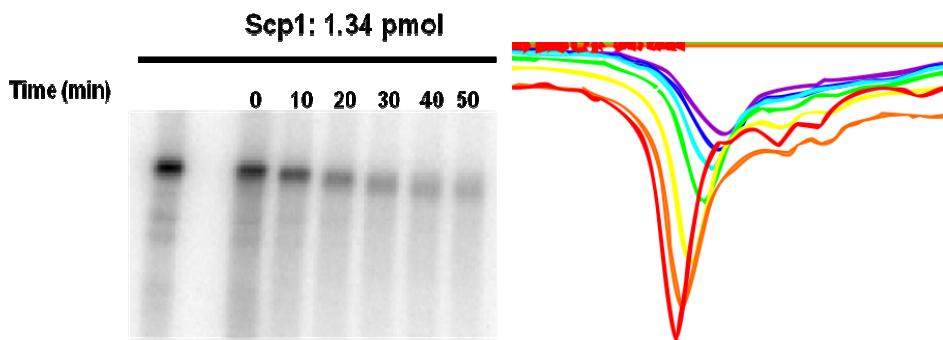
Figure 5-8: The result of the gel shift assay. The intensity of each band was converted to the peak on the right.

The extra three residues of the 17-mer peptide provide additional contact

Since the kinetic analysis demonstrated that the additional phosphorylation from any of the three residues (i.e. Ser_7, Tyr_1 and Ser_2) does not influence the activity of Scp1, we turned our focus to the extra three residues of the 17-mer peptide compared with the 14-mer peptide. Steady-state kinetics experiments were performed to probe whether this extra contact was biochemically relevant, and not just a crystallographic artifact. The activity of Scp1 toward the 14-mer and the 17-mer peptide was tested. As expected, the extra three residues makes the 17-mer peptide a much better substrate as indicated by decreased K_m and increased k_{cat}. The k_{cat}/K_m increased significantly to about 13-fold (**Figure 5-9**).

a

$k_{cat} = 3.9 \pm 0.4\ s^{-1}$
$K_m = 0.134 \pm 0.004\ mM$

Rate/[E]₀ (s⁻¹)

14-mer peptide conc. (mM)

b

Ser5, repeat 20dev

$k_{cat} = 11.1\ +/-\ 0.4\ s^{-1}$
$K_m = 0.030\ +/-\ 0.005\ mM$

Rate/[E]₀
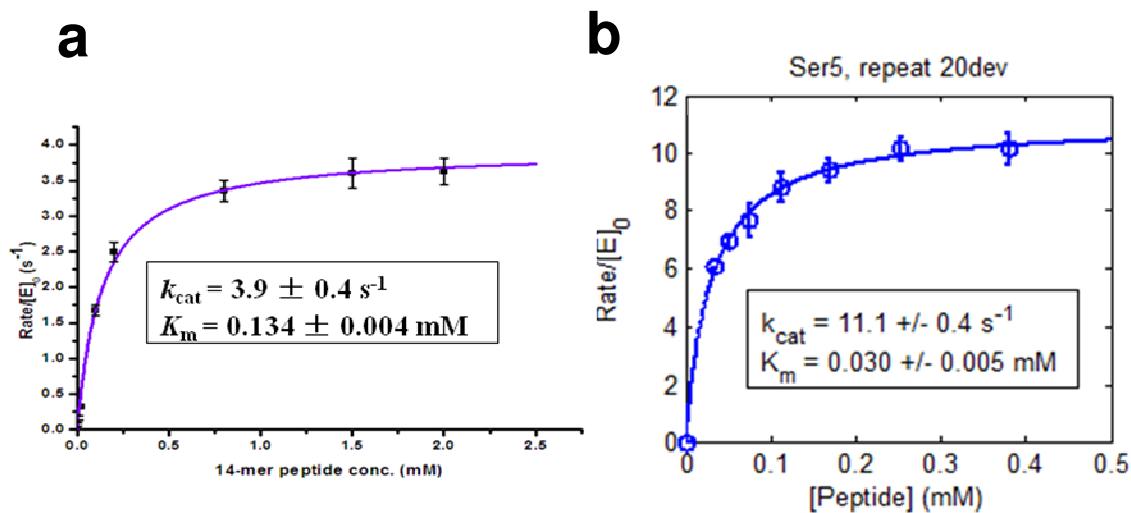
[Peptide] (mM)

Figure 5-9: Steady-state kinetics of Scp1 toward 14-mer peptide (a) and 17-mer peptide (b).

We next measured the binding affinity between Scp1 and the 14-mer or 17-mer peptide. While the K_d value between Scp1 and the 14-mer peptide was about 261 ± 21 µM, the K_d value between Scp1 and the 17-mer peptide was greatly reduced to 46 ± 8 µM. These results demonstrated that the contact between the extra three residues of the 17-mer peptide and the protein is indeed present in solution, at least in our *in vitro* system. The additional contact makes the 17-mer peptide bind to Scp1 much better, resulting in a faster dephosphorylation by the protein.

The repetitive nature of the CTD may require modifying enzymes, e.g. kinases and phosphatases, to function in a processive way to ensure efficient control of the transcription during the transcription cycle. Based on our crystallographic results, we hypothesized that Scp1 could be a processive enzyme by binding strongly to the phospho.Ser_7. However, in our *in vitro* system, we did not observe a strong processivity of Scp1 in the presence of phosphorylated Ser_7, Tyr_1 or Ser_2. On the other hand, the combinatory effect of the phosphorylation of the residues besides Ser_5 was not observed

for Scp1. All other potential phosphorylations on the CTD besides Ser_5 did not influence the activity of Scp1 *in vitro*. In cells, the activity of Scp1 may be regulated by other modifications on the CTD via regulation of its binding partners. Moreover, other phosphatases, for example Ssu72, may be processive enzymes or be significantly regulated by other modifications on the CTD. Our data for the first time demonstrate that the long CTD has additional contact area with the protein in addition to the interaction at the active site. The large contact area makes the CTD bind to the Scp1 much better and is more readily dephosphorylated by the Scp1. In addition, our results show that there is a secondary binding pocket on Scp1 which can be used to design more potent inhibitors of this protein.

Data collection	Scp1-17-mer peptide complex
Space group	C2
Cell dimensions: a, b, c (Å)	124.9, 78.3, 62.8
α, β, γ (°)	90.0, 112.6, 90.0
Resolution (Å)	50.0–2.29 (2.33–2.29)
No. of unique reflections	25291
R_{sym} or R_{merge} (%)	10.0 (48.9)
$I/\sigma(I)$	12.8 (2.2)
Completeness (%)	99.7 (99.9)
Redundancy	3.7 (3.7)
Refinement	
Resolution (Å)	64.8-2.29
No. of reflections (test set)	24000 (1288)
R_{work} / R_{free} (%)	18.0/23.3
No. of atoms: Protein	2928
Mg^{2+}	2
Ligand	181
Water	166
B-factors (Å^2): Protein	30.2
Mg^{2+}	24.1
Ligand	42.9
Water	36.4
R.m.s deviations: Bond lengths (Å)	0.022
Bond angles (°)	1.941
Ramachandran plot (%): Most favored	87.2
Additionally allowed	11.6
Generally allowed	1.2

Table 5-2:   Crystallographic data statistics.[30]

METHODS

Cloning, protein expression and purification

Wildtype and mutant Scp1 were purified using the method described previously in Chapter 3. Yeast CTD was sub-cloned from *Saccharomyces cerevisiae* (*S. cerevisiae*) genomic DNA into a pGTVL2 vector which contains a His-tag followed by a GST-tag at

---

[30] Highest resolution shell is shown in parenthesis. R_{free} is calculated with 5% of the data omitted.

the N-terminus. The GST-CTD fusion protein was expressed in *E.coli* BL21(DE3) strain and purified using the same method for Scp1.

Malachite green assay

The activity of Scp1 toward the CTD peptides was measured in Assay Buffer (see **Chapter 4**) with 5 ng of Scp1 at 37 ^{o}C in 20 µl volume. The reactions were quenched by adding 40 µl of malachite green reagent. The release of free inorganic phosphate was determined by measuring the absorbance at 620 nm.

X-ray crystallography

For crystal soaking, the Scp1D96N proteins were crystallized in previously identified conditions. The crystals soaked with 20-mer peptide were crystallized in the buffer containing 0.5–0.8 M ammonium sulfate, 100 mM HEPES-Na^{+} pH 7.0, and 0.2 M lithium sulfate. The crystals soaked with 17-mer peptide were crystallized in a different condition containing 25% polyethylene glycol (PEG) 3350 and 0.2 M magnesium acetate. Crystals were transferred to a cryo-protecting stabilizer containing 30% (v/v) glycerol, 25% (w/v) PEG 3350, and 0.2 M magnesium acetate. After a brief period of equilibration, crystals were frozen in nylon loops in liquid nitrogen and stored in liquid nitrogen prior to data collection. Crystallographic data were collected at 100 K on beam-line 5.0.2 of the Advanced Light Source (ALS). Diffraction data were processed with HKL2000. The complex crystal structures of Scp1 were determined by molecular replacement (MR) using the Scp1D96N structure as a search model (PDB ID: 2ghq) using the program PHASER available in the CCP4 software package. MR solutions were refined using REFMAC5, reserving 5% of the measured and reduced structure factor amplitudes as an unbiased test set for cross validation (R$_{free}$). The model was built in COOT and refined by REFMAC5. The buried contact area of Scp1 by the peptide and

shape complementarity (Sc) were calculated by AREAIMOL and SC available in the CCP4 software package.

Preparation of ^{32}P-labeled GST-CTD

GST-CTD was treated with either TFIIH (Millipore) or MAPK2/Erk2 (NEB) in the presence of [γ-^{32}P]ATP and 2 mM cold ATP. The labeling reactions were carried out at 30 °C overnight to ensure complete phosphorylation.

Phosphatase assay using labeled GST-CTD

All the dephosphorylation reactions (20 μl total volume) were performed in the Reaction Buffer containing 50 mM Tris-HCl pH 7.9, 10 mM MgCl$_2$, 20% glycerol, 0.02% triton X-100, 20 mM KCl, and 5 mM Dithiothreitol (DTT). Each reaction contained about 1 pmol ^{32}P-labeled GST-CTD and a certain amount of Scp1 (0-1.34 pmol). The reactions were carried out in 30 °C for different periods of time. At each time point, 5 μl of the reaction was taken out and quenched by adding 2x Laemmli buffer. The reaction products were resolved on a 12 % SDS-PAGE gel. The gel images were developed by radioautography. The gel shift assay was done using the similar procedure, but the SDS-PAGE was supplemented with 50 mM phosphate-binding compound (AAL-107) in the presence of Mn^{2+}.

Rapid dilution assay

Comparable amount of Scp1 and CAK-treated ^{32}P-labeled GST-CTD (~1 pmol) were pre-incubated in the 1x Reaction Buffer in 20 μl total volume either in the presence or absence of Mg^{2+} for 8 min. The reactions were rapidly diluted by 50-fold by the addition of 0.98 ml of 1x Reaction Buffer. After the rapid dilution, the reactions were incubated in 30 °C. As a control, the same amount of Scp1 in 0.5 ml of 1x Reaction Buffer was mixed with GST-CTD in 0.5 ml of 1x Reaction Buffer without further

dilution. A fraction of each reaction was taken out every 5-10 min and quenched by adding 2x Laemmli buffer. The products in each reaction were analyzed by SDS-PAGE.

Fluorescence polarization

The 14-mer and 17-mer Ser_5 phosphorylated CTD peptides labeled with 5-carboxylfluorescein (5-FAM) at the N-terminal were used in this assay. The peptide with a final concentration of 1 µM was mixed with various amount of the dominant negative form of Scp1 D96N mutant in 20 µl reactions. The final binding buffer was 20 mM Tris pH 8.0, 50 mM NaCl, 10 mM BME and 20 mM $MgCl_2$. The fluorescence signal was monitored at 485/535 nm (excitation/emission) wavelength using Tecan Infinite F200 microplate reader. The FP signal was calculated by the following equation:

$$FP = \frac{\Delta Signal_{parallel} \cdot G - \Delta Signal_{cross}}{\Delta Signal_{parallel} \cdot G + \Delta Signal_{cross}}$$

where $Signal_{parallel}$ and $Signal_{cross}$ are the signals parallel and cross to the incident light, respectively. G is the G-factor which equals to 1.138 on the instrument used.

The K_d value can be determined by fitting the data points to the following equation:

$$FP = A \frac{[E]_0}{K_d + [E]_0} + B$$

REFERENCES

Bataille, A. R., C. Jeronimo, P. E. Jacques, L. Laramee, M. E. Fortin, A. Forest, M. Bergeron, S. D. Hanes and F. Robert (2012). "A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes." Mol Cell 45(2): 158-170.

Chapman, R. D., M. Heidemann, T. K. Albert, R. Mailhammer, A. Flatley, M. Meisterernst, E. Kremmer and D. Eick (2007). "Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7." Science 318(5857): 1780-1782.

Cho, E. J., T. Takagi, C. R. Moore and S. Buratowski (1997). "mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain." Genes Dev 11(24): 3319-3326.

Egloff, S. and S. Murphy (2008). "Cracking the RNA polymerase II CTD code." Trends Genet 24(6): 280-288.

Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick and S. Murphy (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." Science 318(5857): 1777-1779.

Garber, M. E., T. P. Mayall, E. M. Suess, J. Meisenhelder, N. E. Thompson and K. A. Jones (2000). "CDK9 autophosphorylation regulates high-affinity binding of the human immunodeficiency virus type 1 tat-P-TEFb complex to TAR RNA." Mol Cell Biol 20(18): 6958-6969.

Glover-Cutter, K., S. Larochelle, B. Erickson, C. Zhang, K. Shokat, R. P. Fisher and D. L. Bentley (2009). "TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II." Mol Cell Biol 29(20): 5455-5464.

Gomes, N. P., G. Bjerke, B. Llorente, S. A. Szostek, B. M. Emerson and J. M. Espinosa (2006). "Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program." Genes Dev 20(5): 601-612.

Hengartner, C. J., V. E. Myer, S. M. Liao, C. J. Wilson, S. S. Koh and R. A. Young (1998). "Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases." Mol Cell 2(1): 43-53.

Hintermair, C., M. Heidemann, F. Koch, N. Descostes, M. Gut, et al. (2012). "Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation." Embo J 31(12): 2784-2797.

Hsin, J. P., A. Sheth and J. L. Manley (2011). "RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing." Science 334(6056): 683-686.

Komarnitsky, P., E. J. Cho and S. Buratowski (2000). "Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription." Genes Dev 14(19): 2452-2460.

Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore and M. Hampsey (2004). "Ssu72 Is an RNA polymerase II CTD phosphatase." Mol Cell 14(3): 387-394.

Lu, H., L. Zawel, L. Fisher, J. M. Egly and D. Reinberg (1992). "Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II." Nature 358(6388): 641-645.

Mayer, A., M. Heidemann, M. Lidschreiber, A. Schreieck, M. Sun, C. Hintermair, E. Kremmer, D. Eick and P. Cramer (2012). "CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II." Science 336(6089): 1723-1725.

Shim, E. Y., A. K. Walker, Y. Shi and T. K. Blackwell (2002). "CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo." Genes Dev 16(16): 2135-2146.

Yeo, M., S. K. Lee, B. Lee, E. C. Ruiz, S. L. Pfaff and G. N. Gill (2005). "Small CTD phosphatases function in silencing neuronal gene expression." Science 307(5709): 596-600.

Yeo, M., P. S. Lin, M. E. Dahmus and G. N. Gill (2003). "A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5." J Biol Chem 278(28): 26078-26085.

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." Mol Cell 24(5): 759-770.

Zhou, M., M. A. Halanski, M. F. Radonovich, F. Kashanchi, J. Peng, D. H. Price and J. N. Brady (2000). "Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription." Mol Cell Biol 20(14): 5077-5086.

124

SSU72 PHOSPHATASE

Reversible phosphorylation of the C-terminal domain (CTD) of eukaryotic RNA polymerase II largest subunit represents a critical regulatory mechanism during the transcription cycle and mRNA processing. Ssu72 is an essential phosphatase conserved in eukaryotes that dephosphorylates phosphorylated Ser_5 (phospho.Ser_5) of the CTD heptapeptide. Unlike Scps, Ssu72 is a house-keeping phosphatase for RNA polymerase II-mediated transcription. The function of Ssu72 is implicated in transcription initiation, elongation and termination as well as RNA processing. In this section, I will focus on Ssu72 phosphatase, introducing its characteristics both structurally and functionally.

Chapter 6:   Crystal Structure of Ssu72 in Complex with a Transition State Analogue

INTRODUCTION

Dynamic reversible phosphorylation of the CTD plays an essential role, not only in the recruitment and assembly of transcription complexes, but also in the temporal control of transcription and mRNA processing (reviewed in (Fuda et al. 2009)). Ser_5 phosphorylation is required for assembly of the pre-initiation complex (PIC), and facilitates mRNA capping via recruitment of capping enzymes. As the transcription complex moves away from the initiation site, Ser_5 gradually becomes dephosphorylated, whereas Ser_2 is phosphorylated. Ser_2 phosphorylation is the predominant CTD pattern on both elongating and terminating RNA polymerase II, which ensures efficient 3'-RNA processing by triggering the recruitment of 3'-RNA processing machinery. Non-phosphorylated CTDs are necessary for the recycling of RNA polymerase II and its subsequent binding to a promoter (see **Chapter 1**).

Phosphorylation/dephosphorylation of the CTD is catalyzed by cyclin-dependent kinases and CTD-specific phosphatases (Majello et al. 2001). The best characterized CTD-specific phosphatases are the Fcp/Scp family members. Fcp1 is essential for cell survival in budding and fission yeasts presumably due to its function in RNA polymerase II recycling (Archambault et al. 1997). Small CTD phosphatase (Scp) has been identified as a phospho.Ser_5 phosphatase (Yeo et al. 2003), but it only affects a subset of genes involved in neuronal differentiation (Yeo et al. 2005). It is a component of the neuronal silencing complex REST and prohibits the inappropriate differentiation of neuronal cells (Yeo et al. 2005). Scp1 and Fcp1 share the same active site topology and reaction mechanism (Zhang et al. 2006; Ghosh et al. 2008), but are distinct in preferences for Ser_2 or Ser_5 dephosphorylation and domain architecture (Zhang et al. 2010). Interestingly, Scps which exhibit a strong preference towards phospho.Ser_5 are conserved in higher eukaryotes but not in yeast. Inactivating Scp genes will promote neurogenesis but does not lead to cell death (Yeo et al. 2005). It is also observed that even though neuronal gene expression can be de-repressed upon Scp inactivation, general transcription in the cell is not eliminated (Pfaff, SL, personal communication). Therefore, eukaryotes must have other Ser_5 phosphatases which function as house-keeping proteins for RNA polymerase II recycling, while Scps function at an epigenetic level and only specifically affect the expression profile of a subset of genes.

A novel phosphatase, Ssu72, was identified as the Ser_5 phosphatase in yeast whose phosphatase activity is identified as essential for cell survival and transcription cycling (Sun et al. 1996; Krishnamurthy et al. 2004). Ssu72 belongs to a highly conserved protein family found in eukaryotes. Compromising Ssu72 phosphatase activity results in the accumulation of a hyperphosphorylated Ser_5 form of the CTD (Krishnamurthy et al. 2004). Ssu72 was first identified by its genetic interaction with the

126

transcription factor TFIIB, as a mutation in Ssu72 disrupted this interaction and affected the accuracy of start site selection (Sun et al. 1996). Further studies identified Ssu72 as a subunit of the cleaving and processing factor (CPF) complex, indicating its participation in mRNA processing regulation (He et al. 2003). Ssu72 may also be involved in transcription termination, as the mutations at Ssu72 directly alter the termination of the expression of snoRNA (Kim et al. 2006). Interestingly, the prolyl-isomerization regulation of the CTD, mediated by Ess1 in yeast, is also linked functionally to the phosphatase activity of Ssu72 (Krishnamurthy et al. 2009). It has been proposed that Ess1 can change the balance of the cis-trans conformation of proline by adjusting the conformation suitability of Ser_5-Pro_6 as Ssu72 substrate, which in turn determines the pathway selection for transcriptional termination (Krishnamurthy et al. 2009). Therefore the effectiveness of Ssu72 mediated dephosphorylation of Ser_5 of the CTD can be modulated by the prolyl-isomerization state of the CTD. Recently, an atypical phosphatase, Rtr1, has been defined as a CTD phosphatase, but its biological role is not fully understood (Mosley et al. 2009).

To gain insight into the molecular mechanism of phosphoryl transfer by the Ssu72 phosphatase family and the substrate recognition of Ssu72, we purified recombinant yeast, *Drosophila* and human Ssu72 to examine their catalytic activities and kinetic properties against a generic phosphatase substrate and their natural substrate CTD peptides. Furthermore, the X-ray crystal structures of *Drosophila* Ssu72 were obtained. The structure of Ssu72 clearly indicates it is a unique member of the low molecular weight tyrosine phosphatase (LMWPTP) superfamily, and further exhibits a deep groove that may potentially bind to the CTD of RNA polymerase II. A complex structure of Ssu72 with vanadate highlights the formation of a phosphoryl-cysteine intermediate and mimics the transition state of the phosphoryl transfer reaction. A unique "cap" domain

127

excludes direct access to the active site of Ssu72 by substrate and provides a possible explanation for the selectivity of Ssu72. We have also identified several residues in the CTD binding groove that play essential roles for substrate recognition.

RESULTS AND DISCUSSION

Phosphatase activity of Ssu72 from various organisms and the overall structure of *Drosophila* Ssu72

The sequence alignment of ten Ssu72 phosphatases from various organisms in the NCBI database (**Figure 6-1**) shows a high degree of conservation from yeast to human (43% identity yeast to human, 60% identity *Drosophila* to human), consistent with the proposed biological role of Ssu72 as an essential phosphatase for the C-terminal domain of RNA polymerase II. To study the phosphatase activity and structure of the Ssu72 family, the genes derived from *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Homo sapiens* were cloned and overexpressed in *E. coli*. The proteins were purified to homogeneity, and phosphatase activities were evaluated by pNPP assay which is a generic phosphatase assay. All three enzymes exhibit phosphatase activity with an optimal pH around 6.0–6.5 (data not shown). Yeast Ssu72 has the lowest enzymatic activity with a k_{cat} of 0.090 ± 0.002 s^{-1}, and K_m of 38 ± 3 mM (**Figure 6-2c**). For *Drosophila* Ssu72, much higher activity is observed with the k_{cat} and K_m determined to be 1.30 ± 0.04 s^{-1} and 11.5 ± 0.9 mM (**Table 6-1**, **Figure 6-2a**). The k_{cat} and K_m of human Ssu72 is comparable to *Drosophila* homologue, and was determined to be 0.47 ± 0.01 s^{-1} and 11.0 ± 0.8 mM (**Figure 6-2b**). Our kinetic data for Ssu72 are consistent with previously reported results on yeast and human Ssu72 (Meinhart et al. 2003). The results confirm that the proteins, which were subsequently used in structural studies, are catalytically active.

128

<pre>
          β1   α1   β2   β3   β4
Dro   →   ccccccccc   →   →   TT   →   TT   cccccc
          1   10   20   30   40   50   60

Dro   MTDP..SKLAVAVVCSSNMNRSMEAHNFLAKKGFNVRSYGTGERVKLPGMAFDKPNVYEFGTKYEDIYR
Hum   MP.S..SPLRVAVVCSSNQNRSMEAHNILSKRGFSVRSFGTGTHVKLPGPAPDKPNVYDFKTTYDQMYN
Sce   MPSHRNSNLKFCTVCASNNNRSMESHKVLQEAGYNVSSYGTGSAVRLPGLSIDKPNVYSFGTPYNDIYN
Spo   MAPK..TNLQISVICASNQNRSMEAHNVLKNAGYQVDSFGTGSAVRLPGPSIDKPNIYQFGYPYDEIYK
Bos   MS.F..SPLKVAVVCMSNMNRSMEAHSILRKKGFSVRSFGAGSRVRLPGTARNLPVVYDFSTTYEQMRK
Rat   MP.S..SPLRVAVVCSSNQNRSMEAHNILSKRGFSVRSFGTGTHVKLPGPAPDKPNVYDFKTTYDQMYN
Xen   MP.T..APLRVAVVCSSNQNRSMEAHNILSKRSFNVRSFGTGTHVKLPGPAPDKPNVYDFKTTYEQMYS
Gal   MP.S..SPLRVAVVCSSNQNRSMEAHNILSKRGFSVRSFGTGTHVKLPGPAPDKPNVYDFKTTYDQMYN
Sal   MP.S..HPLRVAVVCSSNQNRSMEAHNILSKRGFEVRSFGTGSHVKLPGPAPDKPNVYDFKTTYEQMYN
Ara   MR......FRYAMVCSSNQNRSMEAHALLKRQGLDVASYGTGSHVKLPGSLREPNVYDFGTPYKQMFD
            ▲   △ ▲▲   △   △   △

          α2   α3   η1   β5   α4
Dro   cccccccccccc   ccccccccccc   ccc   →   cccccccccccc   TT
          70   80   90   100   110   120   130

Dro   LESKDKEFYTQNGLLHMLDRNRRIKKCPERFQD..TKEQFDIIVTVEERVYDLVVMHMESMESVDNRPV
Hum   LLRKDKELYTQNGILHMLDRNKRIKPRPERFQN..CKDLFDLILTCEERVYDQVVEDLNSREQETCQPV
Sce   LLSQSADRYKSNGLLQMLDRNRRLKKAPEKWQE..STKVFDFVFTCEERCFDAVCEDLMNRGGKLNKIV
Spo   LEAQDSRLYTANGLLKMLDRNRRIKRAPCRWQD..QDSIYNIVITCEERCYDAICEDLYRRGETLNRPV
Bos   LVRIDRERYTSNGILHILGRNERIKPRPERFQE..CRDRFDVIFTCEESVYDRVIEELWVREQETCQPV
Rat   LLRKDKELYTQNGILHMLDRNKRIKPRPERFQN..CKDLFDLILTCEERVYDQVVEDLNSREQETCQPV
Xen   LLKKDKELYTQNGILHMLDRNRRIKPRPERFQN..CKDYFDLVITCEERVYDQVVEELNSREQETCQPV
Gal   LLRKDKELYTQNGILHMLDRNKRIKPRPERFQN..CKDVFDLILTCEERVYDQVVEDLNSREQETCQPV
Sal   LVRKDKELYTQNGILHMLDRNKRIKSRPERFQS..CKDQFDLVITCEERVYDQVLEDLSSREQETFTPV
Ara   LRRKDPELYKRNGILQMIKRNLSVKLAPQRWQDNAGDGVFDVVMTFEEKVFDSVLEDLNNREQSLTKTI
          △   △   △

          α5   α6   β7
Dro   →   ccccccccccccccccc   c.......ccccccccccccccc   →
          140   150   160   170   180   190

Dro   VLNVDVVDNAEDALMGAFVITDMINMMAKSTDL........DNDIDELIQEFEERRKRV.ILHSVLFY
Hum   VVNVDIQDNHEEATLGAFLICELCQCIQHTEDM........ENEIDELLQEFEEKSGRT.FLHTVCFY
Sce   VINVDIKDDDENAKIGSKAILELADMLNDKIEQCEKDDIPFEDCIMDILTEWQSSHSQLPSLYAPSYY
Spo   LINVDIKDNHEEASVGGKAILDLVNKLTEAQDK.......LEELFPSIMADFQSNHPKLPVLYTIHFF
Bos   VINVNMDDNTEDATLGSLIICELCERLQQAEDM........EGSLAELLLAVERKTGRS.FLHTVCFY
Rat   VVNVDIQDNHEEATLGAFLICELCQCIQHTEDM........ENEIDELLQEFEEKSGRA.FLHTVCFY
Xen   VINVDIQDNHEEATLGAFLICELCQCIQHTEDM........ENEIDELLQEFEEKSGRT.FLHTICFY
Gal   VINVDIQDNHEEATLGAFLICELCQCIQHTEDM........ENEIDELLQEFEEKSGRT.FLHTVCFY
Sal   VINVDIQDNHEEATLGAFLICELCQCIQHTDDM........ENEMDELLQEFEDKSNRP.FLHTVCFY
Ara   VMNLEVKDNHEEAAIGGRLALELCQEIEGNETW.......EDTIDDIVAGFEKQHRRK.LVYSISFY
          ▲
</pre>

Figure 6-1: Primary sequence alignment of Ssu72 from various species.[31]

---

[31] Sequence alignment of Ssu72 from *Drosophila* (Dro, NP_608342), *Homo sapiens* (Hom, NP_054907), *Saccharomyces cerevisiae*(Sce, NP_014177), *Schizosaccharomyces pombe* (Spo, NP_594076), *Rattus norvegicus* (Rat, NP_001020828), *Bos Taurus* (Bos, XP_595220), *Gallus gallus* (Gal, NP_001007876), *Salmo salar* (Sal, NP_001136192), *Xenopus laevis* (Xen, NP_001084864) and *Arabidopsis thaliana* (Ara, NP_177523). Helices and strands are indicated by coils and arrows, respectively. Active site residues are marked with filled triangles below the alignment. The residues (also included in the mutagenesis study) for the tentative substrate CTD proline binding are marked by hollow triangles.

a Drosophila Ssu72

k_{cat} 1.30 ± 0.04 S^{-1}
K_m 11.5 ± 0.9 mM

b Human Ssu72

k_{cat} 0.47 ± 0.01 S^{-1}
K_m 11.0 ± 0.8 mM

c Yest Ssu72

k_{cat} 0.090 ± 0.002 S^{-1}
K_m 38 ± 3 mM
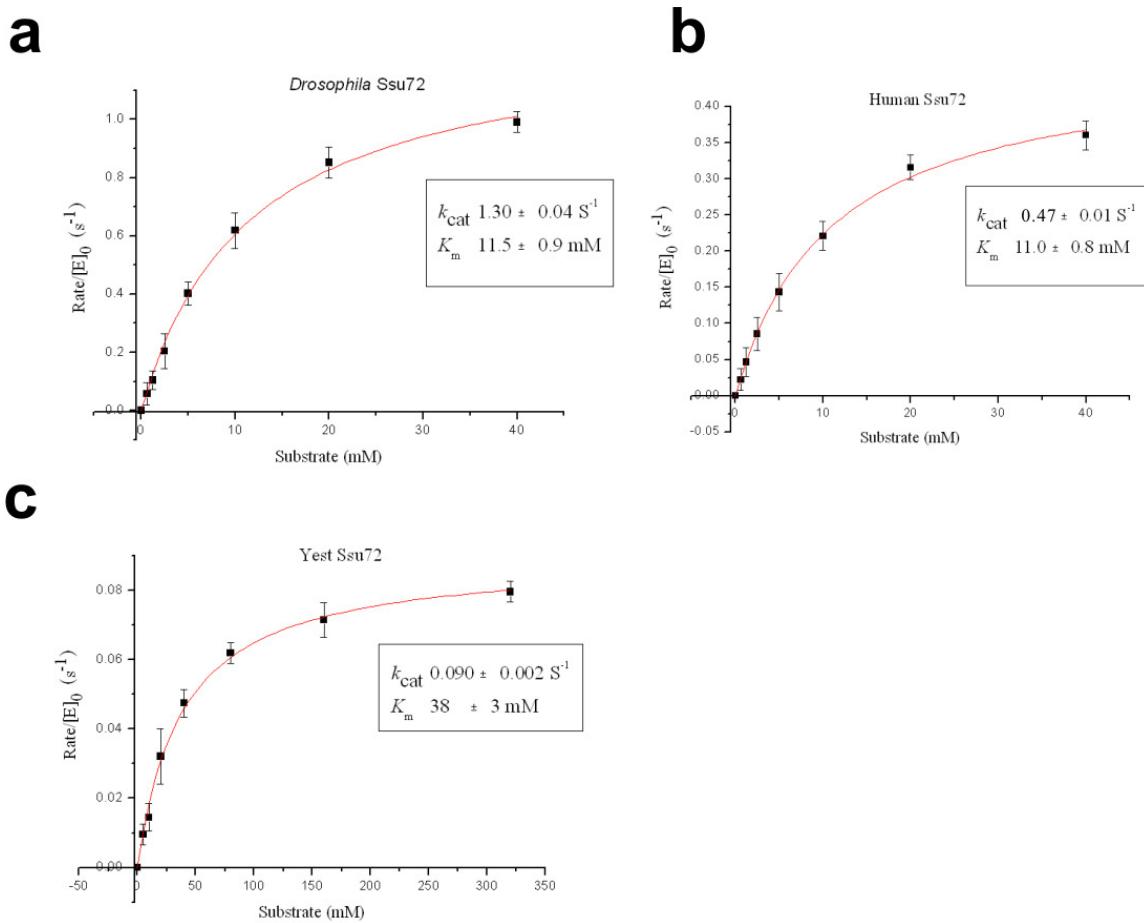
Figure 6-2:   Steady-state kinetics of Ssu72 measured by *p*NPP assay.[32]

_____

[32] The data points were fitted into Michaelis–Menten equation by Origin7.5.

Mutants	pNPP		phos-CTD peptide (10 mers) Y-S-P-T-**pS**-P-S-Y-S-P		phos-CTD peptide (14 mers) Y-S-P-T-**pS**-P-S-Y-S-P-T-**pS**-P-S	
	K_m (mM)	k_{cat} (s^{-1})	K_m (mM)	k_{cat} (s^{-1})	K_m (mM)	k_{cat} (s^{-1})
Wildtype	11.5 ± 0.9	1.30 ± 0.02	0.96 ± 0.03	0.42 ± 0.04	0.15 ± 0.01	0.397 ± 0.004
M17A	17 ± 2.6	0.28 ± 0.02	N.A.	N.A.	N.A.	N.A.
P46A	17.8 ± 1.7	0.69 ± 0.02	N.A.	N.A.	N.A.	N.A.
D51A	5.5 ± 0.7	1.34 ± 0.05	N.A.	N.A.	N.A.	N.A.
Y56A	144.2 ± 21.2	0.35 ± 0.04	N.A.	N.A.	N.A.	N.A.
Y77A	35.8 ± 1.9	1.39 ± 0.03	N.A.	N.A.	N.A.	N.A.
L82A	4.6±0.6	1.31±0.04	1.9±0.1	0.53±0.03	0.38±0.05	0.47 ± 0.04
M85A	34.9 ± 2.5	0.48 ± 0.02	N.A.	N.A.	N.A.	N.A.

Table 6-1:   Probing the residues that are important for the substrate CTD peptide binding.[33]

Drosophila Ssu72 was crystallized, and the diffraction of apo crystals was processed to resolution at 2.85 Å. The protein crystallizes in space group P2$_1$2$_1$2 (unit cell: a = 158.1 Å, b = 101.9 Å, c = 65.6 Å; α = β = γ = 90.0º) with four molecules per asymmetry unit. The protein structure consistently exhibits a high thermal factor (**Table 6-2**). This may also explain the difficulty in obtaining a high resolution structure for apo Ssu72 as well as the crystal's high sensitivity to environmental changes. The four molecules in each asymmetry unit are highly identical with the exception of two flexible loops (residues 47–53 and 127–134). The topology can be roughly divided into two portions with a deep groove cutting through the surface of the molecule, separating the protein into the "cap" and the "core" domains (**Figure 6-3a**). The connections between the "core" and "cap" domains are two flexible loops, suggesting that the individual

---

[33] N.A.: No activity detected.

domains can move in a hinge-like motion to desolvate the active site and impose selectivity toward substrates. The essential catalytic residue, Cys13, whose substitution results in loss of phosphatase activity and death of yeast strains (Sun et al. 1996), is located at the tip of the groove (**Figure 6-3a**). The "core" domain of the protein has a typical Rossmann fold with a series of twisted beta-strands sandwiched by alpha helices at each side (**Figure 6-3b**). The "cap" portion (residue 41–92), which is well conserved among Ssu72 from different species, is unique with no similar sequence or structure identified in other protein families (**Figure 6-1 and 6-3c**). When we color the protein according to residue conservation, the highly conserved residues are localized around the groove dividing the "cap" and "core" domains of the protein, indicating the functional importance of this groove (**Figure 6-3c**).

Data Statistics	Apo	Complex (Ssu72–vanadate)
Source	Advanced Light Source 5.0.1	Advanced Light Source 5.0.1
Wavelength (Å.)	0.9774	0.9765
Resolution (Å.)	48.3-2.85(2.90–2.85)	48.5- 2.35 (2.39–2.35)
Space group	$P2_12_12$	$P2_12_12$
Unit cell (Å.) a, b, c	158.1, 101.9, 65.6	157.6, 102.3, 65.8
Number of unique reflections	24877	45695
Redundancy	5.0 (4.6)	6.6 (6.1)
Completeness (%)	94.6 (96.2)	99.9 (98.4)
I/σ	21.8 (1.3)	36.5/1.5
R_{sym} (%)	8.7 (76.7)	7.0 (76.1)
Refinement Statistics		
Resolution limits (Å.)	48.3–2.85	48.65–2.35
Number of reflections (test)	23460 (1194)	45556(2295)
R_{work}/ R_{free} (%)	21.6 / 27.7	20.5 / 26.7
Number of atoms protein/water	6216/15	6236/171
B factors for protein atoms ($Å^2$)	83.7	62.5
B factors for ligand/water ($Å^2$)	-/66.7	54.0/60.8
Bond rmsd length(A)/angles (°)	0.008/1.2	0.018/2.1
Most favored	88.4%	87.7%
Additionally allowed	10.7%	11.1%
Generously allowed	0.6%	1.1%
Disallowed region	0.3%	0

Table 6-2:   Crystallographic data statistics.[34]

---

[34] Parentheses indicate statistics for highest resolution shell.
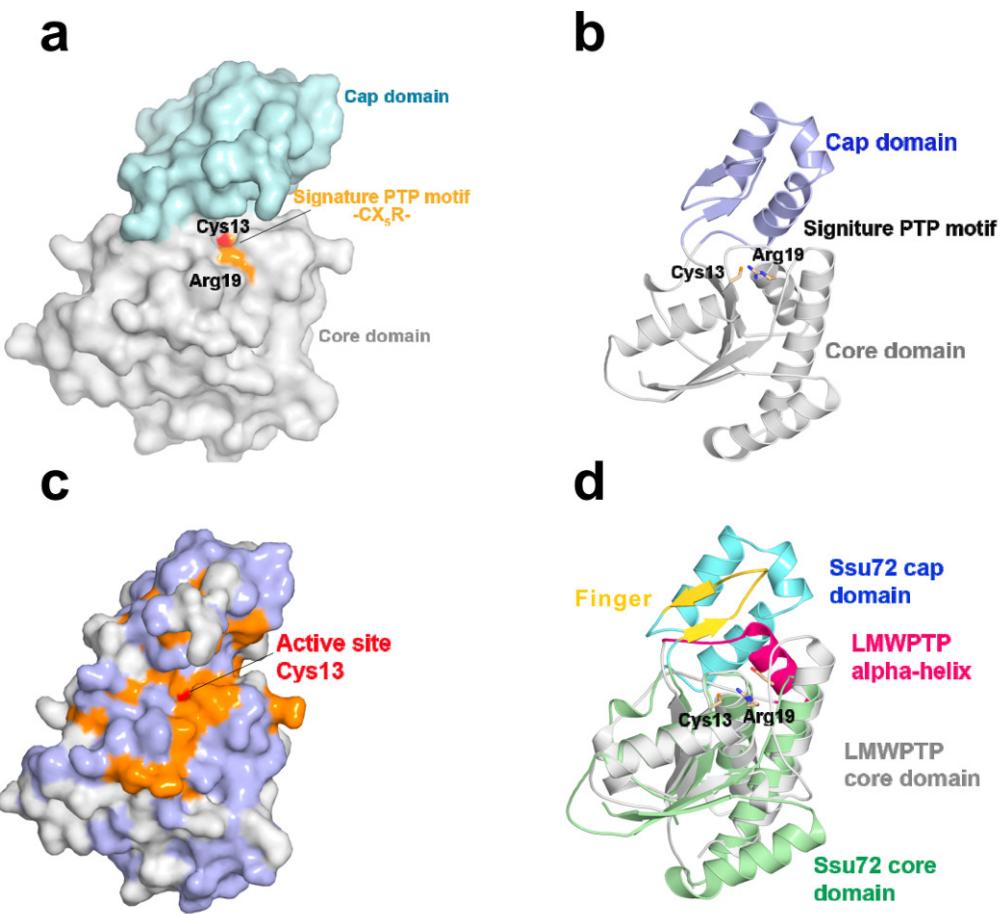
Figure 6-3:   Structures of Ssu72.[35]

Ssu72 is a low molecular weight tyrosine phosphatase

Unlike kinases which evolve from the same ancestor and therefore maintain the same three-dimensional topology, protein phosphatases adapt different catalytic mechanisms for the phosphoryl-transfer reaction. Three different catalytic mechanisms are identified among protein phosphatases (**Figure 6-4**). The first category, called

---

[35] (a) Surface representation of *Drosophila* Ssu72 apo protein structure. The signature motif residue arginine is highlighted in orange, and the nucleophilic cysteine is highlighted in red. (b) Ribbon representation of *Drosophila* Ssu72 structure. The active site essential residues are shown in stick representation. (c) Surface representation of the conserved residues (shown in orange) on *Drosophila* Ssu72 structure. Partially conserved residues are shown in light blue. (d) Superimposition of *Drosophila* Ssu72 (core domain in pale green, cap domain in cyan, finger region in yellow) and human LMWPTP 1 (PDB code: 1xww, core domain in white, a short helix that differs in different isoforms colored in red).

cysteine-based phosphatases, utilizes cysteine as the nucleophile at the active site (Denu et al. 1995; Tonks 2006). The well-studied mechanism establishes that the phosphate group from the substrate is transferred to the thiol group as a phosphoryl-intermediate, which in turn undergoes hydrolysis. On the other hand, Ser/Thr phosphatases such as PP1 and PP2 utilize a dramatically different strategy for the hydrolysis of phosphate monoesters by using di-metal ions and activating a water molecule to directly break the P–O bond without a phosphoryl-protein intermediate (reviewed in (Shi 2009)). The third category, HAD-like phosphatases, including the CTD phosphatase Scp/Fcp family, resembles the two-step reaction mechanism of cysteine-based phosphatases, but utilizes aspartic acid as a nucleophile (Allen et al. 2004).

The cysteine-based protein phosphatases are further divided into three sub-families based on the relative location of the signature motif, termed "PTP" loop (–CX_5R–), and their preferences for substrates. The first subfamily, classical tyrosine phosphatase, is usually around 30 kDa with the PTP loop close to the middle of the protein, and it specifically dephosphorylates phosphoryl tyrosine. Dual specificity protein phosphatases, on the other hand, recognize both tyrosine and serine/threonine phosphorylation, as indicated by the name, with the PTP loop situated closer to the C-terminal end of the protein. The third family, low molecular weight tyrosine phosphatases (LMWPTP), has a single catalytic phosphatase domain of around 18 kDa. The PTP loop is located at the N-terminus of the LMWPTP, with nucleophile cysteine usually in the vicinity of the 10–15th residues. LMWPTP favors phosphoryl-tyrosine as a substrate, even though high concentrations of phosphoryl-serine/threonine are also subjected to LMWPTP-catalyzed hydrolysis *in vitro* (Ramponi et al. 1997).

135
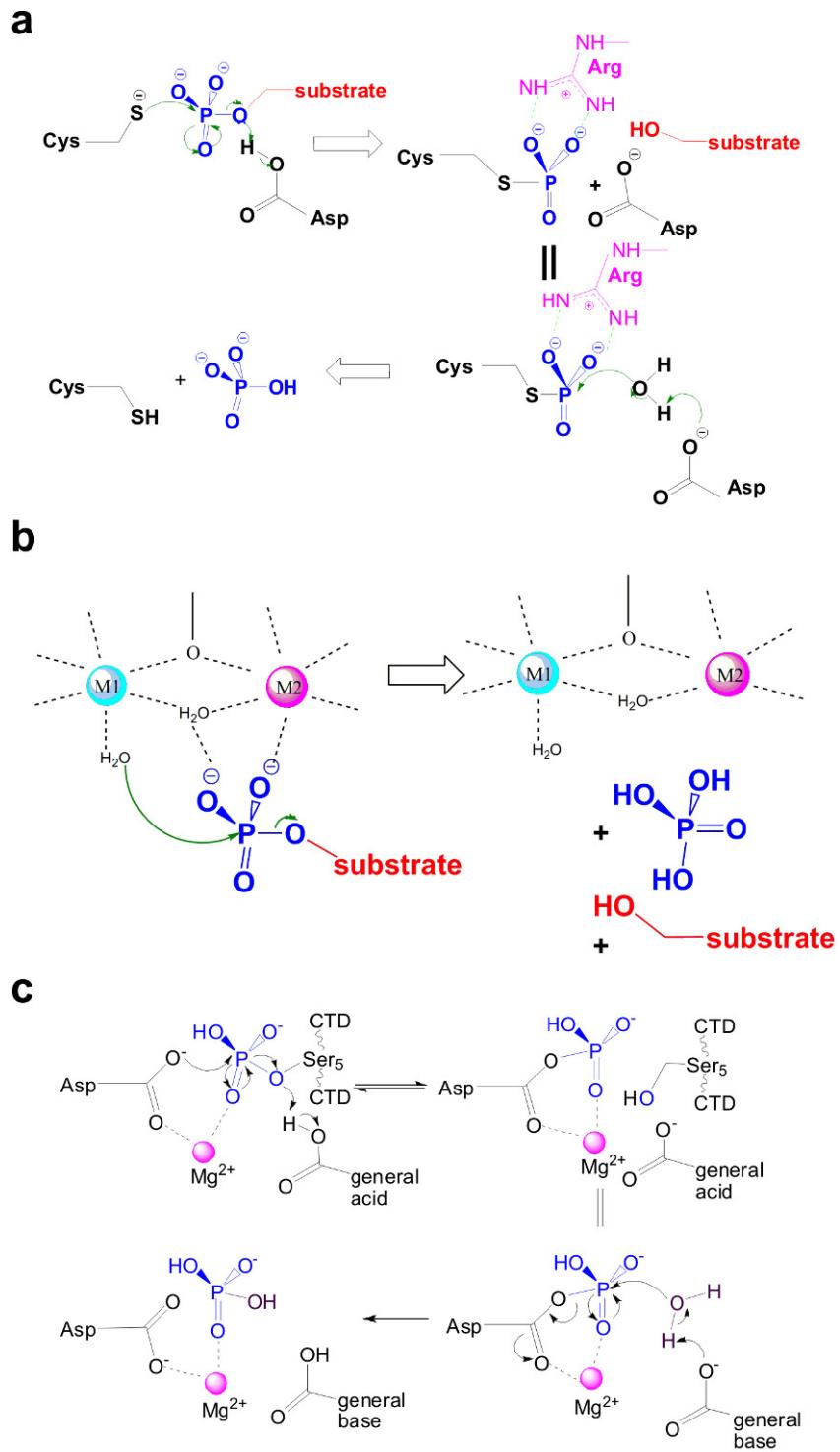
Figure 6-4: Summary of three different catalytic mechanisms identified among protein phosphatases.

Although Ssu72 exhibits very low primary sequence identity with proteins from any of the cysteine-based phosphatase families, the fold of the core domain is highly identical to that of LMWPTP (**Figure 6-3d**). A $-CX_5R-$ signature motif is located at the N-terminus of the protein. With the primary sequence consensus of only 15%, the overall fold of Ssu72 exhibits a Z score of 11.6 in DALI search (Holm et al. 2010) and a RMS deviation of 1.7 Å when superimposed with the main chain of LMWPTP (PDB code: 1xww) (Ssu72 6–195 excluding residue 41–96; LMWPTP 1–157 excluding residue 50–67).

Complex structure of Ssu72 with inhibitor vanadate

Another characteristic of cysteine-based phosphatases is their sensitivity to oxoanion compounds such as vanadate, tungsten and molybdate (Lindqvist et al. 1994). These compounds can inhibit cysteine-based phosphatases by forming transition states or product analogs with cysteine nucleophiles. In order to better understand the reaction mechanism and roles of catalytic residues, apo Ssu72 crystals were transferred to crystallization medium that contained 2mM Na_3VO_4. A structure of Ssu72 in which VO_4^{3-} was coordinated in the active site was then compared to the apo Ssu72 structure. Interestingly, the incorporation of the vanadate ion stabilizes the crystal, which exhibits a higher resolution and lower thermal factor than apo Ssu72 crystals (**Table 6-2**).

The structure of the Ssu72-vanadate complex was solved using apo Ssu72 as a search model in molecular replacement, and refined to a resolution of 2.35 Å (**Figure 6-5a**). The refined model is almost identical to apo Ssu72 with the signature motif ($-CX_5R-$) located at the crevice. Strong positive electron density can be observed close to the nucleophilic cysteine (Cys13), with a characteristic trigonal bipyramidal coordination (**Figure 6-5b**). A vanadate group can be built into the electron density with three oxygen

atoms at equatorial positions, and a cysteine sulfur atom, as well as another oxygen atom at each side of the apical position (**Figure 6-5c**). The distance of vanadium to sulfur is refined to 2.3–2.4 Å, and the other four oxygen atoms are all about 1.9 Å away, consistent with a previously published vanadate adduct with bovine LMWPTP (Zhang et al. 1997). The bond length of S–V suggests the existence of covalent bond formation between the compound and the active site residue of Ssu72, explaining the inhibitory effect of vanadate compounds on Ssu72 (**Figure 6-5c**).

The other conserved residue of the signature motif, Arg19, is located close to the nucleophilic cysteine (**Figure 6-5c**), and has dual functions in the phosphoryl transfer reaction which is conserved in all cysteine-based phosphatases (Zhang 2002). First, the positively charged arginine side chain can help recruit the phosphate group to the active site through electrostatic interactions. More importantly, the side-chain of arginine has the potential to stabilize the tetrahedral intermediate by forming hydrogen bonds with vanadate (Zhang et al. 1992) (**Figure 6-5d**). The stability of the coplanar bidentate is essential for the phosphoryl transfer reaction since lysine replacement of arginine cannot fully replace the arginine function in *Yersinia* PTPase (Zhang et al. 1992). Indeed, in the complex structure of Ssu72 with vanadate, the two nitrogen atoms of the guanidinium group from the highly conserved Arg19 pair with two equatorial oxygen atoms in the vanadate group (**Figure 6-5d**). The pairing of the vanadate oxygen and side chain of Arg19 constitutes a six-membered ring hydrogen bonding network that presumably reduces the free energy activation for the phosphoryl transfer reaction. The structure mimics the formation of the trigonal-pyramidal transition state in the phosphoryl transfer reaction, and explains how such a high-energy state is stabilized in Ssu72. Additional hydrogen bonding occurs between the backbone amides of the PTP loop and the equatorial oxygen atoms (**Figure 6-5d**). As found in other cysteine-based phosphatases, a

serine residue is located after this arginine and potentially stabilizes the thiolate anion of nucleophile cysteine by lowering its pK_a (Peters et al. 1998). Indeed, in our structure, the hydroxyl group of the Ser20 side chain forms a hydrogen bond with nucleophile Cys13, which can potentially promote the deprotonated state of cysteine and facilitate the nucleophilic attack on phosphorus (**Figure 6-5e**). The elimination of this hydrogen bond by replacing Ser20 with Ala abolished the phosphatase activity of the protein in our *p*NPP assay (data not shown).

Asp-containing flexible loop

In addition to the $-CX_5R-$ PTP motif, another highly conserved residue in all LMWPTP is an aspartic acid residue, usually 110 amino acids away from the nucleophile cysteine. The three-dimensional position of this aspartic acid is conserved in all cysteine based phosphatases. In *Yersinia* PTPase, this conserved aspartate accounts for the basic limb of the pH dependence curve (Zhang et al. 1994). This aspartate residue fills the role of a general acid to provide the proton for the substrate-leaving after phosphoryl transfer (Zhang 2002). In the primary sequence of *Drosophila* Ssu72, two aspartate residues, Asp141 and Asp144 (Asp140 and Asp143 in human), are highly conserved in this area. The mutation of either of these two aspartate residues in human Ssu72 reduced the k_{cat}/K_m of Ssu72 by about 20 fold (Meinhart et al. 2003). To identify which is the general acid for phosphoryl transfer reaction, we inspected the interaction of the transition state analogue vanadate with active site residues. In the complex structure of Ssu72 and vanadate, this highly conserved position is occupied by Asp144, which can make two potential hydrogen bonds with vanadate (**Figure 6-5c**). However, Asp141 is 12.7 Å away from the vanadate group (**Figure 6-5e**). Therefore, we conclude that the Asp144 is the general acid that contributes to the reaction, and the loss of activity of Asp141 mutation might be

caused by structural disruption. Indeed, when we mutated Asp141 to Ala, the proteins can be expressed (as detected by SDS-PAGE), but are not accumulated in soluble fraction, likely due to issue with protein folding.

In other cysteine-based phosphatases, this loop containing general acids/bases is presumably highly flexible, and can adopt the active conformation upon ligand binding or swing away when the active site is empty. For example, in classical protein tyrosine phosphatases, this aspartate is 10 Å away from the active site in the absence of substrate and only extends into the active site when occupied (Zhang 2002). In our crystal structure of *Drosophila* Ssu72 complexed with vanadate, the loop is in the "active" form with Asp144 extending into the active site, mimicking the conformation of an active site for phosphoryl transfer (**Figure 6-5c**). Without the vanadate compound, the loop still extends into the active site, forming a hydrogen bond with a highly ordered water molecule.
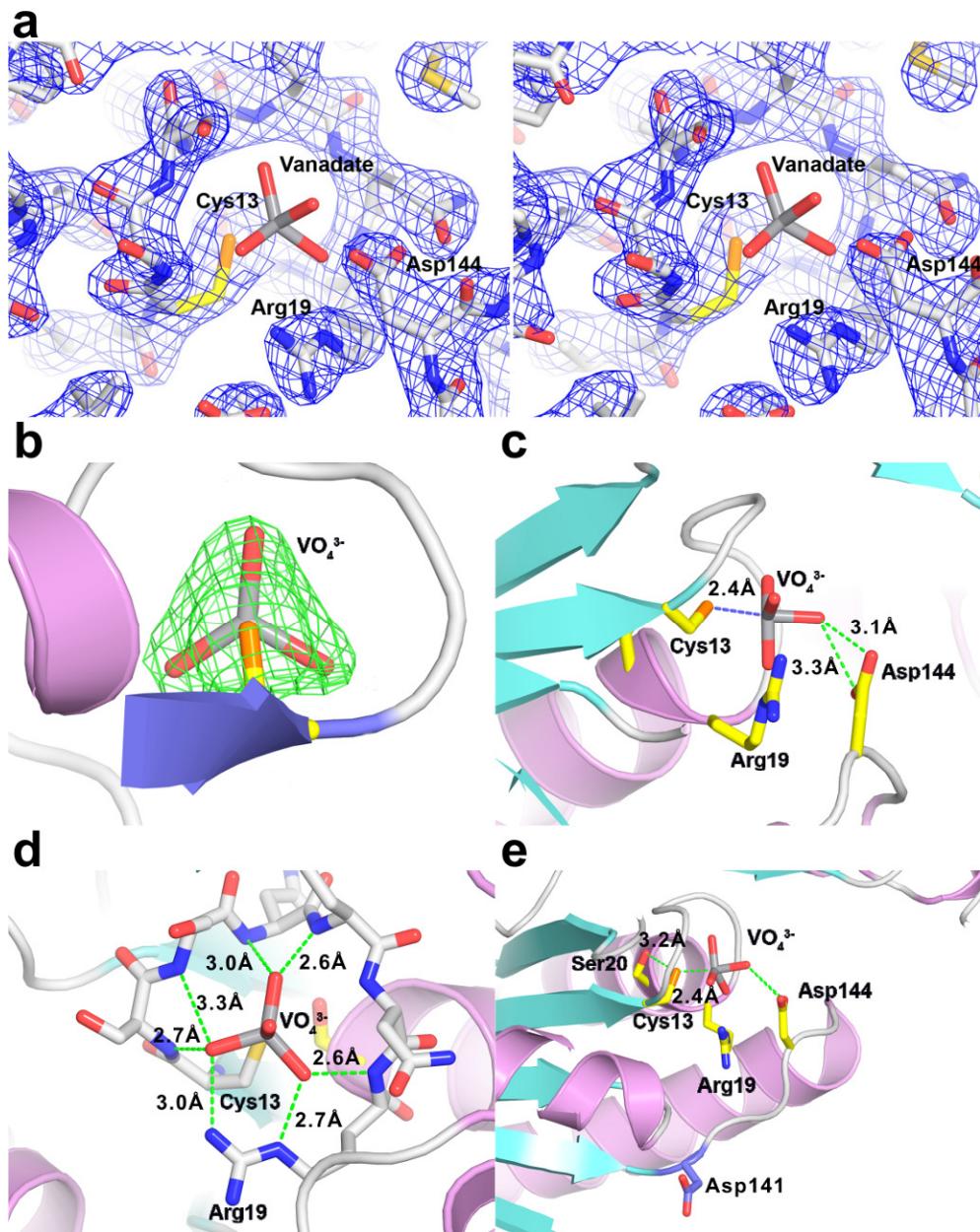
Figure 6-5:   Complex structure of Ssu72 and vanadate.[36]

---

[36] The green dashed lines indicate hydrogen bonds. The numbers are the distances of hydrogen bonds in Å. (a) Stereoview of $2F_o$-F_c electron density of active site residues contoured at 2σ. The nucleophile cysteine is in yellow. (b) $2F_o$-F_c electron density map of vanadate contoured at 1.6σ. (c) The interactions between vanadate and Cys13/Asp144. The S–V bond between Cys13 and vanadate (blue dashed line) is within covalent bond range. (d) The vanadate mimics the formation of the trigonal pyramidal transition state in phosphoryl transfer. (e) The hydroxyl group of the Ser20 side chain forms a hydrogen bond with the nucleophile Cys13. Asp141 is located distal to the active site with its carbon atom colored blue.

To confirm that the addition of transition state analog vanadate stabilizes the conformation of the loop containing the aspartate biochemically, we used DSF to assay the effect of an oxoanion compound on the folding of Ssu72 (Niesen et al. 2007). In this method, fluorescence from a dye with affinity for hydrophobic surface is monitored. The fluorescence increases upon the exposure of hydrophobic surfaces during protein unfolding. In our experiment, the incorporation of the vanadate compound consistently stabilized Ssu72 and increased the T_m. For yeast Ssu72, T_m increased from 37.8 ± 0.2 ℃ to 43.5 ± 0.3 ℃ upon vanadate addition. Similar trends were observed in human Ssu72 (T_m increased from 44.0 ± 0.3 ℃ to 49.2 ± 0.5 ℃) and *Drosophila* Ssu72 (T_m from 42.5 ± 0.2 ℃ to 45.6 ± 0.4 ℃) (**Figure 6-6a, b and c**). We also used CD to detect the effect of the vanadate compound and to correlate with our DSF results. Even though the buffer condition is slightly different, the stabilization effect of vanadate is consistent with a T_m of *Drosophila* Ssu72 increase from 44.0 ± 0.3 ℃ to 46.2 ± 0.2 ℃ (**Figure 6-6d**). We reason that the binding of vanadate stabilizes the flexible loop containing general acid Asp144 through hydrogen bonding, thereby enhancing protein folding.

The "cap" region

Unlike the bottom core portion of the Ssu72 protein whose topology is highly conserved in all cysteine-based phosphatase, the "cap" region (residue 41–92) of Ssu72 is unique and shows no structural similarity to any other protein (less than 4.0 in Z score) in DALI search (Holm et al. 2010). From residue Glu41 to Glu57, two short anti-parallel β-strands are connected by an extended flexible loop resembling a "finger", comprising the most dynamic portion of the Ssu72 structure (**Figure 6-3d**). This mobile tip excludes the active site pocket from the bulk solvent, and limits access to the binding pocket for Ssu72. When we mutated a residue at the tip of the finger, Asp51 to Ala, the binding to

CTD substrate was greatly reduced, even though the phosphatase activity was not affected when tested by the *p*NPP assay (**Table 6-1**). Consistent with our mutagenesis result, a complex structure of human Ssu72 with the substrate CTD and the scaffold protein symplekin was recently determined, showing the incorporation of the substrate CTD close to this flexible region (Xiang et al. 2010) (**Figure 6-7a**).
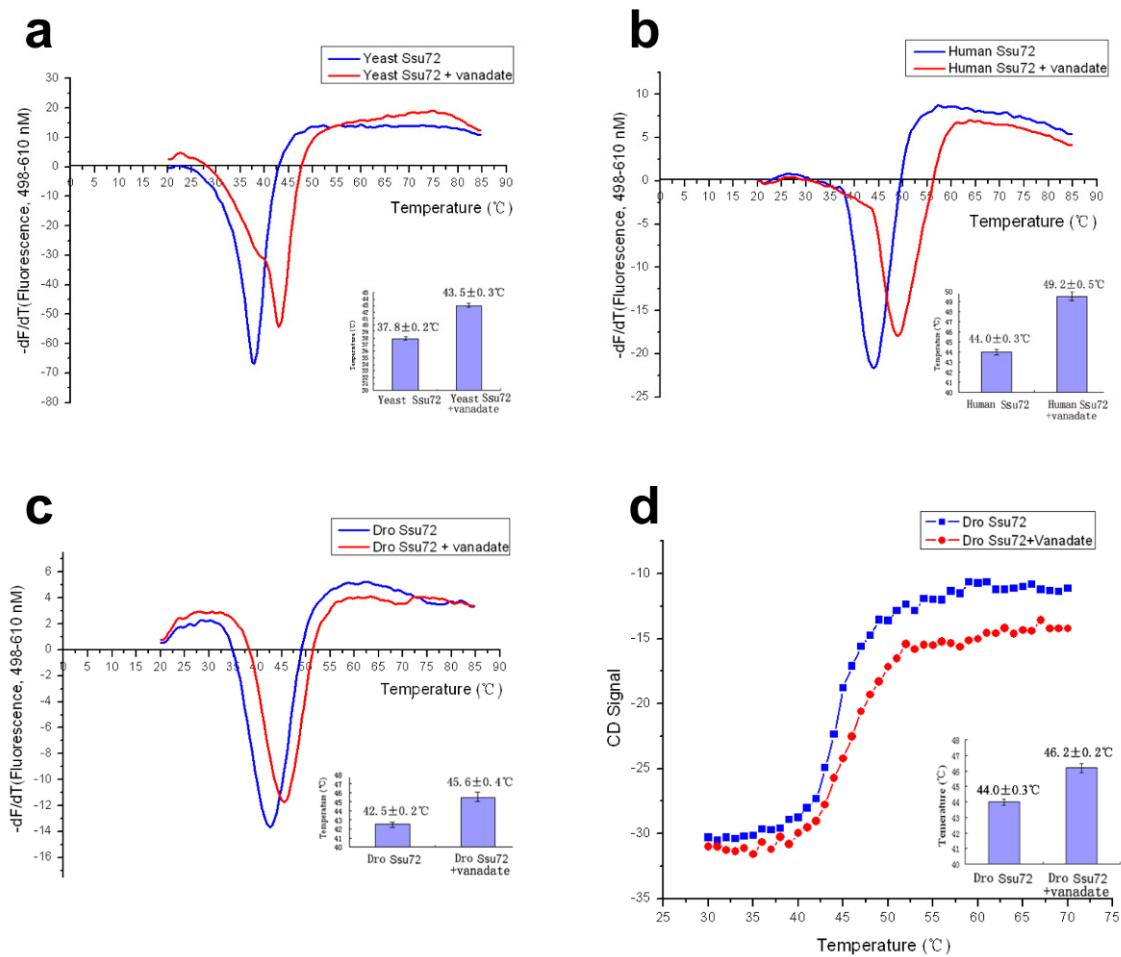
Figure 6-6:   Differential scanning fluorimetry and circular dichroism spectra of Ssu72 in the absence or presence of sodium vanadate.

Unlike Ssu72, the active site in LMWPTP (PDB code: 1xww) is highly exposed (**Figure 6-3d**). Such architectural design is consistent with the substrate specificity of

these phosphatases. LMWPTP target receptor tyrosine kinases such as platelet-derived growth factor receptor, insulin receptor (Chiarugi et al. 1997), ephrin receptor (Stein et al. 1998), and fibroblast growth factor receptor (Rigacci et al. 1999). Therefore, the active site of LMWPTP needs to be accessible and open to accommodate such bulky substrates. On the other hand, the only substrate Ssu72 recognizes *in vivo* is the CTD of RNA polymerase II, which is a highly disordered peptide and presumably adopts a long extended loop to fit into the more exclusive active site of Ssu72.

One important regulatory mechanism for LMWPTP activity is through the differential phosphorylation of Tyr131 and Tyr132 (Bucciantini et al. 1999). These two sites are proposed to increase the activity of the phosphatase or to recruit the SH2 domain of Grb2 to LMWPTP (Bucciantini et al. 1999). In contrast, no phosphorylation regulation has yet been observed for Ssu72. The flexible loop, containing two tyrosine residues (Tyr131 and Tyr132) in LMWPTP, is substantially shorter in Ssu72, with the two tyrosine residues omitted in the Ssu72 sequence (**Figure 6-7b**). We speculate that instead of using phosphorylation as regulatory mechanism, Ssu72 adapts the unique "cap" domain to modulate phosphatase activity and substrate selectivity. Use of divergent cap domains to limit the accessibility of the active site is a common strategy in biology. For example, the HAD super family, a large family of proteins for phosphoryl transfer, utilizes different "cap" domains to mediate different phosphoryl transfers of different substrates (Allen et al. 2004). A core domain with Rossmann fold is conserved among all 3000 HAD family members.
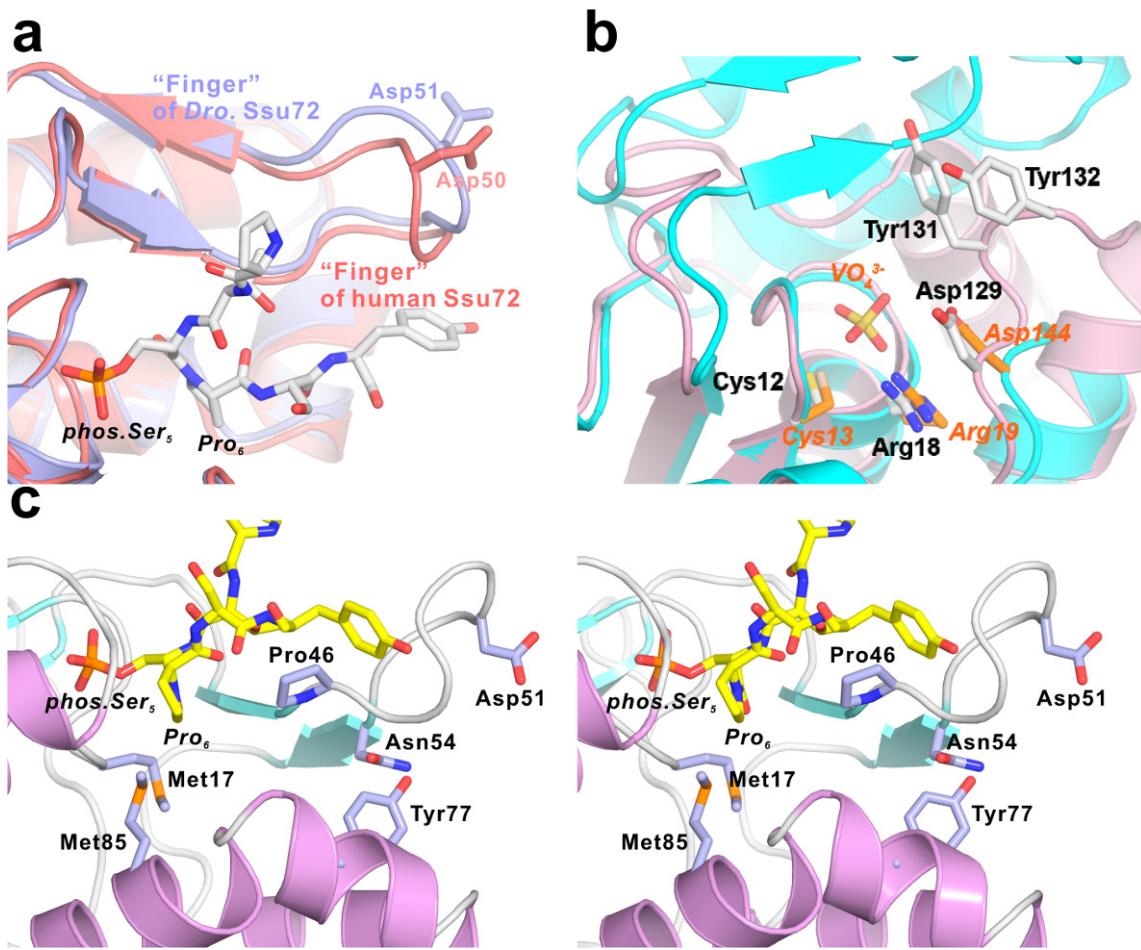
Figure 6-7: Important residues for Ssu72 function.[37]

Potential proline binding pocket

Ssu72 can dephosphorylate the hyperphosphorylated CTD *in vitro* (Krishnamurthy et al. 2004). To test the activity of *Drosophila* Ssu72 towards the CTD,

---

[37] (a) Superimposition of Drosophila (Dro) Ssu 72 and human Ssu72-CTD complex (PDB code: 3o2q). Dro. Ssu72 is in light blue color, and human Ssu72 is in salmon color. The CTD peptide is in white stick. The flexible loop region (indicated by "Finger") encloses the CTD binding site in the human Ssu72-CTD peptide complex structure. (b) Superimposition of the active sites of Ssu72 (cyan) and LMWPTP (PDB code: 1xww, pink). The key residues are shown in white stick in LMWPTP, and in orange stick in Ssu72, respectively. The vanadate is shown in yellow to indicate the active site. (c) Stereo figure of the CTD binding groove presented by superimposition of the CTD peptide from the symplekin-Ssu72-CTD complex structure (PDB code: 3o2q) into Drosophila Ssu72 structure. The residues in the mutagenesis study are shown in stick.

we used CTD derived peptides with different lengths and phosphorylation patterns to identify the optimal minimal length for the CTD recognition. It has been noticed that the binding of CTD by Ssu72 is highly dependent on the Tyr from the following repeat of phospho.Ser$_5$ (Hausmann et al. 2005). This is quite different from human Scp, in which the Pro$_3$ N-terminus from the phospho.Ser$_5$ is essential for the CTD recognition (Zhang et al. 2006). To evaluate this, we first tested against a 17-mer CTD peptide with phospho.Ser$_5$ site (SPSYSPTSPSYSPT**pS**PS), which is the optimal substrate for Scps. However, the activity is relatively low with a k_{cat} of 0.19 ± 0.01 s^{-1} and K_m of 0.85 ± 0.09 mM. Consistently, when using a singly phosphorylated double repeats peptide (YSPTSPSYSPT**pS**PS) as the substrate, the k_{cat} is measured to be 0.44 ± 0.03 s^{-1} and K_m is 1.2 ± 0.2 mM. The best activity is obtained when both Ser$_5$ are phosphorylated in the double repeats CTD peptide (YSPT**pS**PSYSPT**pS**PS) which results in a k_{cat} of 0.397 ± 0.004 s^{-1} and K_m of 0.152 ± 0.006 mM (**Table 6-1**, **Figure 6-8**). This turnover rate is 20 times better than the previously reported yeast Ssu72 (k_{cat} estimated to be 0.02 s^{-1}) (Hausmann et al. 2005), and comparable to the other Ser$_5$ phosphatases, Scps (K_m = 0.21 ± 0.05 mM and k_{cat} = 2.44 ± 0.04 s^{-1}) (Zhang et al. 2006). We further identified a minimal optimal peptide (10-mer, YSPT**pS**PSYSPT) which shows comparable activity with K_m = 0.96 ± 0.03 mM and k_{cat} = 0.42 ± 0.04 s^{-1} (**Table 6-1**). This result suggests different recognition element requirements to Scp and Ssu72 for substrate binding, even though they both dephosphorylate phospho.Ser$_5$ of the CTD. The residues that contribute greatly to the binding of CTD by Scp are primarily located at the N-terminal to the phospho.Ser$_5$ subject to dephosphorylation (Zhang et al. 2006). Conversely, Ssu72 requires certain C-terminal residues following the phospho.Ser$_5$ for CTD peptide recognition.

146
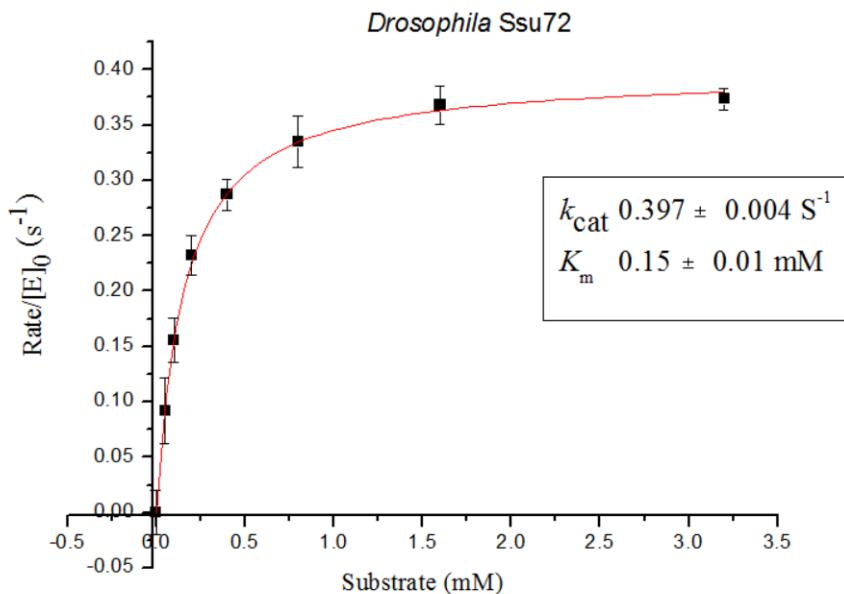
Figure 6-8: Steady-state kinetics of *Drosophila* Ssu72 towards CTD peptide.[38]

Most recently, the complex structure of human Ssu72 bound to CTD and scaffold protein symplekin was reported (Xiang et al. 2010). A particularly interesting question about Ssu72 is the identity of residues that play essential roles in the substrate recognition. We reasoned that such residues would not interfere with active site phosphatase activity, as characterized by the *p*NPP assay, but should exhibit much lower activity towards CTD peptides. To investigate which residues are important specifically for the substrate recognition of Ssu72, we made mutations of residues that are located in the substrate binding groove. We are particularly interested in the residues that might contribute to the binding of Pro_6 in the CTD sequence. Unlike any other CTD bound proteins such as Scp1 or Cgt1 (Zhang et al. 2010), Pro_6 exhibits a unique cis conformation upon Ssu72 binding (Xiang et al. 2010). Through mutagenesis and steady-

---

[38] The CTD peptide sequence is YSPT**p**SPSYSPT**p**SPS.

state kinetics, we identified five residues that play pivotal roles for the specific activity towards the CTD (**Table 6-1**). A pair of Met residues, Met17 and Met85, clamp on Pro_6 of the CTD, with Pro46 from the opposite side, through hydrophobic interactions (**Figure 6-7c**). The mutation of either of the two Met residues, or Pro46 to Ala results in the loss of Ssu72 activity towards the CTD peptide, even though the activity towards pNPP is unaffected or insignificantly reduced (**Table 6-1**). On the other hand, the loss of activity due to Tyr77 mutation is surprising, since Tyr77 does not directly contribute to the substrate binding. Instead, it orients the side chain of Asn54 with hydrogen bonding, and stabilizes the conformation of the highly flexible region (residues 41–57). The comparison of the human Ssu72-CTD complex structure and our *Drosophila* structures discloses that Ssu72 structure is highly conserved upon vanadate incorporation, with the only difference found at this highly flexible region (residues 41–57) enclosing the CTD substrate in the complex structure (**Figure 6-7a**). This region is formed by two anti-parallel beta-strands with a flexible loop, and is highly dynamic in apo structure (**Figure 6-3d**). The phosphate group of phospho.Ser_5 of the CTD is found at the location of the vanadate ion in our structure. Residues on this flexible loop are particularly important for the function of Ssu72. Asp51Ala mutant exhibits comparable phosphatase activity with the wildtype Ssu72 in the pNPP assay, but its activity is greatly compromised when phosphoryl CTD peptides are used as the substrate (**Table 6-1, Figure 6-7c**). Another mutation, Tyr56Ala, disrupts the protein activity in both the pNPP assay and malachite green assay (**Table 6-1**). Not all residues located in the groove contribute to Ssu72 activity. For example, the Ala mutation of Leu82, which is located very close to the substrate CTD, results in no loss of phosphatase or specificity activity (**Table 6-1**).

Ssu72 is an essential protein in eukaryotes whose phosphatase activity plays a key role in RNA polymerase II recycling and the selection of the transcription termination

pathways. In the present study, we successfully solved the high resolution X-ray crystal structures of *Drosophila* Ssu72 in complex with vanadate, mimicking the transition state of the phosphoryl transfer. This structure shows that Ssu72 is a unique subfamily of LMWPTP with a "cap" domain to confer its substrate specificity. Moreover, kinetic studies of *Drosophila* Ssu72 using Ser_5-phosphorlated peptides as the substrates demonstrated that Ssu72 can dephosphorylate phospho.Ser_5 of the CTD peptide. A deep groove engulfing the nucleophile Cys13 between the core domain and the "cap" domain was observed in the Ssu72 structures and is predicted to be the binding site for the CTD peptide. In addition, the complex structure of Ssu72 and vanadate mimics the formation of the trigonal pyramidal transition state in the phosphoryl transfer reaction, and explains how such a high-energy state is stabilized in Ssu72.

METHODS

Cloning and protein purification

The Ssu72 genes from various organisms were cloned from *Saccharomyces cerevisiae* genomic DNA, *Drosophila melanogaster* synthetic genes (*Drosophila* Genomics Resource Center) and *Homo sapiens* cDNA (Origene). The genes encoding Ssu72 were amplified by PCR using primers with convenient ligation-independent cloning sites and corresponding templates. The PCR products were directly inserted into an in-house-generated pET28b derivative vector, pETHis8-SUMO, which was linearized with T4 DNA polymerase. The resulting expression plasmid pETHis8-SUMO-ssu encodes a corresponding Ssu72 fusion protein with an N-terminal 124-amino-acid tag consisting of a His8 leader (MGSSHHHHHHHHSSGSDSEVNQEAKP) followed by an 86-amino-acid SUMO (small ubiquitin-related modifier) fragment and a PreScission protease recognition sequence (ALEVLFQGPGSG). *Escherichia coli* BL21 (DE3) were

149

transformed with the pETHis8-SUMO-ssu vectors and grown in Luria–Bertani medium containing 50 µg/ml kanamycin at 37 °C until the OD600 reached 0.6–1.0. The cultures were supplemented with 0.5 mM IPTG (isopropyl β-D-thiogalactopyranoside) and then grown at 16 °C for 16 h. The induced cells were harvested with centrifugation (5000 rev./min, Beckman Avanti J-26, JLA 8.1 rotor) and disrupted by sonication on ice (Misonix sonicator 4000, amplitude 90%) for 5 cycles (30 sec/cycle, with 1 sec on/off pulses, and a 4 min pause between cycles). The recombinant Ssu72 protein was affinity purified using a Ni-NTA (Ni^{2+}-nitrilotriacetate) column (Qiagen) and eluted with buffer A [250 mM imidazole, 500 mM NaCl, 100 mM Tris/HCl (pH 8.5) and 10 mM 2-mercaptoethanol]. A final concentration of 3 mM EDTA was added to the eluted protein to eliminate any residual nickel ion. The eluted protein was then dialysed against buffer B [20 mM Tris/HCl (pH 8.5), 100 mM NaCl and 10 mM 2-mercaptoethanol] and digested with PreScission protease (GE Healthcare) at 4 °C overnight. The truncated tag and untruncated protein were removed by reloading the digested sample on a second Ni-NTA column equilibrated with buffer C [30 mM imidazole, 20 mM Tris/HCl (pH 8.5), 100 mM NaCl and 10 mM 2-mercaptoethanol]. Flow-through fractions were pooled and applied on a MonoQ column, and proteins were eluted with a NaCl gradient. Ssu72 proteins were further purified by gel filtration using a Superdex-75 column (GE Healthcare) equilibrated against buffer D [25 mMHepes (pH 7.5), 80 mM NaCl and 1 mM DTT (dithiothreitol)]. Finally, the pure Ssu72 proteins were concentrated with a 10 kDa Vivaspin-20 concentrator (GE Healthcare) to ~10 mg/ml.

Crystallization and compound soaking

Crystallization trials were carried out using sitting-drop vapour diffusion at 4 °C. An initial Ssu72 crystal was identified from the Classics suite crystal screen (Qiagen) by

mixing equal amounts of protein solution (10 mg/ml) and reservoir solution containing 0.1 M Hepes sodium salt (pH 7.5), 10% (v/v) propan-2-ol and 20% (w/v) PEG [poly(ethylene) glycol] 4000. Subsequently, production of the crystal was optimized under the following conditions: 0.1 M Hepes sodium salt (pH 7.5), 10% (v/v) propan-2-ol and 14–18% (w/v) PEG 4000. Crystals appeared within 1 week and grew for 10 more days. A 20–25% (v/v) glycerol was supplemented with crystallization conditions as the cryoprotectant for all crystals used in the experiments. To obtain the complex structure of Ssu72 with vanadate, apo crystals were soaked in mother liquor containing 2 mM sodium orthovanadate solution for 2 h prior to cryoprotection. X-ray diffraction data were collected at beamline 5.0.1 (Advanced Light Source) using a 2×2 ADSC CCD (charge-coupled device) detector. Data were processed with HKL2000 and are summarized in **Table 6-2**.

Structure determination and refinement

The crystal structures of *Drosophila* Ssu72 were determined by molecular replacement using a low-resolution structure (PDB code 3FMV; 3.3 Å) from Northeast Structure Genomic Consortium as the initial search model using the program Phaser from the CCP4 package. Initial refinement was carried out using the Phenix refinement suite under NCS (non-crystallographic symmetry) restraints with a 5% test set (reflections) excluded for R_{free} cross-validation. Electron-density maps σA-weighted $2F_o - F_c$ and $F_o - F_c$ maps were calculated after each cycle of refinement and inspected to guide model rebuilding using Coot. For the complex structure of Ssu72 and vanadate, the locations of the vanadate group were clear in $F_o - F_c$ maps. The inhibitor model was built into the electron density using Coot. The final models were evaluated by PROCHECK.

151

Refinement statistics are summarized in Table 6-2. PyMOL was used to produce molecular graphics renditions.

Phosphatase activity assays

An assay for non-specific phosphatase activity was carried out using the general substrate *p*NPP (*p*-nitrophenyl phosphate) (Fluka, Sigma–Aldrich) in 0.2 ml of reaction mixture containing 0.1 M citrate buffer (pH 6.0), 1 mM DTT, various *p*NPP concentrations (0–40 mM for *Drosophila* and human Ssu72, 0–320 mM for yeast Ssu72), and 2.0 μg of the enzyme. After incubation at 28 °C (or 37 °C for human Ssu72) for 15 min, the reaction was terminated by adding an equal volume of 2.0 M NaOH, and the released *p*NPP was measured at 405 nm using a Tecan infinite 200 microplate reader. The optimum pH for enzyme activity was determined in 0.1 M citrate buffer, 0.1 M Mes buffer and 0.1 M Tris/HCl (adjusted in the range of pH 4–9) at 28 °C (or 37 °C for human Ssu72) with 2.4 mM *p*NPP as the substrate. Temperature effects on activity were measured in 0.1 Mcitrate buffer (pH 6.0) at different temperatures. Michaelis–Menten kinetic parameters for purified Ssu72 towards *p*NPP were determined by measuring initial reaction rates at various *p*NPP concentrations in the above reaction buffer. Data were fitted to the Michaelis–Menten equation with the program Origin7.5 (OriginLab).

The substrate specificity of the *Drosophila* Ssu72was examined using Ser5-phosphorylated CTD peptides (Anaspec) by malachite green colorimetric assay. The assay was performed in a 200 μl PCR tube in 20 μl of assay buffer [0.1 M Mes (pH 6.0) and 1 mM DTT] containing various concentrations of Ser5-phosphorylated CTD peptide in the absence and the presence of purified Ssu72 at 100 ng per reaction. The assay tubes were incubated in a PCR machine (Bio-Rad) at 28 °C for 10 min. It was stopped with 80 μl of malachite green reagent (Biomol Green), and the absorbance was read at 620 nm

according to the manufacturer's instructions. Kinetic data were analysed according to the Michaelis–Menten equation with the program Origin7.5.

Differential scanning fluorimetry

Yeast, *Drosophila* and human Ssu72s with various concentrations (1–50 μM) were mixed with 1 mM sodium vanadate in 96-well low-profile PCR plates (ABgene, catalogue number AB-0700) and incubated on ice for 30 min. SYPRO orange dye was added into each well immediately before placing the plate in LightCycler 480 (Roche). The protein unfolding experimentwas carried out with an increase of temperature from 20 °C to 85 °C. The melting temperature curves of Ssu72s were monophasic and T_m values were derived from the curves.

CD-monitored thermal denaturation

Drosophila Ssu72 (43 μM) was incubated with 1 mM sodium orthovanadate in 20 mM Hepes buffer (pH 7.5). CD spectra were monitored at 220 nm with an AVIV model 420 spectropolarimeter equipped with a thermoelectric temperature control unit. Data points were collected every min/°C as the sample temperature increased from 30 to 70 °C (1 °C per min). Melting temperature was obtained by fitting the data to a Boltzman sigmoidal function by Origin7.5.

ACCESSION CODE

The atomic coordinates and structure factors (codes 3OMW and 3OMX) have been deposited in the Protein Data Bank.

REFERENCES
Allen, K. N. and D. Dunaway-Mariano (2004). "Phosphoryl group transfer: evolution of a catalytic scaffold." *Trends Biochem Sci* **29**(9): 495-503.

Archambault, J., R. S. Chambers, M. S. Kobor, Y. Ho, M. Cartier, D. Bolotin, B. Andrews, C. M. Kane and J. Greenblatt (1997). "An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae." *Proc Natl Acad Sci U S A* **94**(26): 14300-14305.

Bucciantini, M., P. Chiarugi, P. Cirri, L. Taddei, M. Stefani, G. Raugei, P. Nordlund and G. Ramponi (1999). "The low Mr phosphotyrosine protein phosphatase behaves differently when phosphorylated at Tyr131 or Tyr132 by Src kinase." *FEBS Lett* **456**(1): 73-78.

Chiarugi, P., P. Cirri, F. Marra, G. Raugei, G. Camici, G. Manao and G. Ramponi (1997). "LMW-PTP is a negative regulator of insulin-mediated mitotic and metabolic signalling." *Biochem Biophys Res Commun* **238**(2): 676-682.

Denu, J. M. and J. E. Dixon (1995). "A catalytic mechanism for the dual-specific phosphatases." *Proc Natl Acad Sci U S A* **92**(13): 5910-5914.

Fuda, N. J., M. B. Ardehali and J. T. Lis (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." *Nature* **461**(7261): 186-192.

Ghosh, A., S. Shuman and C. D. Lima (2008). "The structure of Fcp1, an essential RNA polymerase II CTD phosphatase." *Mol Cell* **32**(4): 478-490.

Hausmann, S., H. Koiwa, S. Krishnamurthy, M. Hampsey and S. Shuman (2005). "Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases." *J Biol Chem* **280**(45): 37681-37688.

He, X., A. U. Khan, H. Cheng, D. L. Pappas, Jr., M. Hampsey and C. L. Moore (2003). "Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1." *Genes Dev* **17**(8): 1030-1042.

Holm, L. and P. Rosenstrom (2010). "Dali server: conservation mapping in 3D." *Nucl. Acids Res.* **38**(suppl_2): W545-549.

Kim, M., L. Vasiljeva, O. J. Rando, A. Zhelkovsky, C. Moore and S. Buratowski (2006). "Distinct pathways for snoRNA and mRNA termination." *Mol Cell* **24**(5): 723-734.

Krishnamurthy, S., M. A. Ghazy, C. Moore and M. Hampsey (2009). "Functional Interaction of the Ess1 Prolyl Isomerase with Components of the RNA Polymerase II Initiation and Termination Machineries." *Mol. Cell. Biol.* **29**(11): 2925-2934.

Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore and M. Hampsey (2004). "Ssu72 Is an RNA polymerase II CTD phosphatase." *Mol Cell* **14**(3): 387-394.

Lindqvist, Y., G. Schneider and P. Vihko (1994). "Crystal structures of rat acid phosphatase complexed with the transition-state analogs vanadate and molybdate. Implications for the reaction mechanism." *Eur J Biochem* **221**(1): 139-142.

Majello, B. and G. Napolitano (2001). "Control of RNA polymerase II activity by dedicated CTD kinases and phosphatases." *Front Biosci* **6**: D1358-1368.

Meinhart, A., T. Silberzahn and P. Cramer (2003). "The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase." *J Biol Chem* **278**(18): 15917-15921.

Mosley, A. L., S. G. Pattenden, M. Carey, S. Venkatesh, J. M. Gilmore, L. Florens, J. L. Workman and M. P. Washburn (2009). "Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation." *Mol Cell* **34**(2): 168-178.

Niesen, F. H., H. Berglund and M. Vedadi (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." *Nat Protoc* **2**(9): 2212-2221.

Peters, G. H., T. M. Frimurer and O. H. Olsen (1998). "Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases." *Biochemistry* **37**(16): 5383-5393.

Ramponi, G. and M. Stefani (1997). "Structure and function of the low Mr phosphotyrosine protein phosphatases." *Biochim Biophys Acta* **1341**(2): 137-156.

Rigacci, S., E. Rovida, S. Bagnoli, P. Dello Sbarba and A. Berti (1999). "Low Mr phosphotyrosine protein phosphatase activity on fibroblast growth factor receptor is not associated with enzyme translocation." *FEBS Lett.* **459**(2): 191-194.

Shi, Y. (2009). "Serine/Threonine Phosphatases: Mechanism through Structure." *Cell* **139**(3): 468-484.

Stein, E., A. A. Lane, D. P. Cerretti, H. O. Schoecklmann, A. D. Schroff, R. L. Van Etten and T. O. Daniel (1998). "Eph receptors discriminate specific ligand oligomers to determine alternative signaling complexes, attachment, and assembly responses." *Genes Dev* **12**(5): 667-678.

Sun, Z. W. and M. Hampsey (1996). "Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo." *Mol Cell Biol* **16**(4): 1557-1566.

Tonks, N. K. (2006). "Protein tyrosine phosphatases: from genes, to function, to disease." *Nat Rev Mol Cell Biol* **7**(11): 833-846.

Xiang, K., T. Nagaike, S. Xiang, T. Kilic, M. M. Beh, J. L. Manley and L. Tong (2010). "Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex." *Nature* **467**(7316): 729-733.

Yeo, M., S.-K. Lee, B. Lee, E. C. Ruiz, S. L. Pfaff and G. N. Gill (2005). "Small CTD Phosphatases Function in Silencing Neuronal Gene Expression." *Science* **307**(5709): 596-600.

Yeo, M., P. S. Lin, M. E. Dahmus and G. N. Gill (2003). "A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5." *J Biol Chem* **278**(28): 26078-26085.

Zhang, M., G. Gill and Y. Zhang (2010). "Bio-molecular Architects: A Scaffold Provided by the C-terminal Domain of Eukaryotic RNA Polymerase II." *Nano Reviews*.

Zhang, M., J. Liu, Y. Kim, J. E. Dixon, S. L. Pfaff, G. N. Gill, J. P. Noel and Y. Zhang (2010). "Structural and functional analysis of the phosphoryl transfer reaction mediated by the human small C-terminal domain phosphatase, Scp1." *Protein Sci* **19**(5): 974-986.

Zhang, M., M. Zhou, R. L. Van Etten and C. V. Stauffacher (1997). "Crystal structure of bovine low molecular weight phosphotyrosyl phosphatase complexed with the transition state analog vanadate." *Biochemistry* **36**(1): 15-23.

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." *Mol Cell* **24**(5): 759-770.

Zhang, Z. Y. (2002). "Protein tyrosine phosphatases: structure and function, substrate specificity, and inhibitor development." *Annu Rev Pharmacol Toxicol* **42**: 209-234.

Zhang, Z. Y., J. P. Davis and R. L. Van Etten (1992). "Covalent modification and active site-directed inactivation of a low molecular weight phosphotyrosyl protein phosphatase." *Biochemistry* **31**(6): 1701-1711.

Zhang, Z. Y., Y. Wang and J. E. Dixon (1994). "Dissecting the catalytic mechanism of protein-tyrosine phosphatases." *Proc Natl Acad Sci U S A* **91**(5): 1624-1627.

HUMAN PROLYL-ISOMERASE PIN1

The CTD of eukaryotic RNA polymerase II is an essential regulator for RNA polymerase II-mediated transcription. It is composed of multiple repeats of a consensus sequence $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$. CTD regulation of transcription is mediated both by phosphorylation of the serines and prolyl isomerization of the two prolines. Interestingly, the phosphorylation sites are typically close to prolines, thus the conformation of the adjacent proline could impact the specificity of the corresponding kinases and phosphatases. Experimental evidence of cross-talk between these two regulatory mechanisms has been elusive. Pin1 is a highly conserved phosphorylation-specific peptidyl-prolyl isomerase (PPIase) that recognizes the phospho-Ser/Thr (pSer/Thr)-Pro motif with CTD as one of its primary substrates *in vivo*. In this chapter, we provide structural snapshots and kinetic evidence that support the concept of cross-talk between prolyl isomerization and phosphorylation. We determined the structures of Pin1 bound with two substrate isosteres that mimic peptides containing pSer/Thr-Pro motifs in cis or trans conformations. The results unequivocally demonstrate the utility of both cis- and trans-locked alkene isosteres as close geometric mimics of peptides bound to a protein target. Building on this result, we identified a specific case in which Pin1 differentially affects the rate of dephosphorylation catalyzed by two phosphatases (Scp1 and Ssu72) that target the same serine residue in the CTD heptad repeat but that have different preferences for the isomerization state of the adjacent proline residue. These data exemplify for the first time how modulation of proline isomerization can kinetically impact signal transduction in transcription regulation.

Chapter 7:   Structural and kinetic analysis of prolyl-isomerization/phosphorylation cross-talk in the CTD code

INTRODUCTION

The unique chemical structure of the proline residue makes it the only amino acid enabling the Xaa-Pro peptide bond (where Xaa is any amino acid residue) to adopt a cis conformation to a significant extent (10-30%) (Brandts et al. 1975). The transition between cis and trans conformations of the prolyl peptide bond occurs at a slow rate, and phosphorylation of the serine or threonine preceding the proline (pSer/Thr-Pro) can further slow the transition (Schutkowski et al. 1998). Since protein kinases or phosphatases are specific for the proline isomeric state (Lu et al. 2002), it is possible that kinases and phosphatases could recognize the same Ser/Thr position but with different preferences for the isomerization state of the adjacent Pro. In other words, since proline isomerases can be dependent on the phosphorylated state, these enzymes may act as molecular 'switches' that govern the downstream recognition and kinetics of phosphatases.

To the extent this hypothesis is true, the conformation of the peptide bond should impact the substrate recognition of at least some modifying enzymes. Pin1 is a highly conserved peptidyl-prolyl isomerase (PPIase) that specifically recognizes the pSer/Thr-Pro motif and catalyzes faster transition between the two isomeric states, and thereby regulates protein functions. The isomerization of the pSer/Thr-Pro motif mediated by Pin1 is especially important for biological processes, e.g. cancer and neurodegenerative diseases such as Alzheimer's (Lu et al. 1996; Lu 2004). In humans, one of the most significant substrates of Pin1 is RNA polymerase II, the central molecule for eukaryotic transcription (Yaffe et al. 1997). The signature motif recognized by Pin1 is highly enriched in the C-terminal domain (CTD) of RNA polymerase II, which consists of 26-52

159

tandem heptapeptide repeats with the general consensus sequence from yeast to human, $Tyr_1Ser_2Pro_3Thr_4Ser_5Pro_6Ser_7$ (Corden 1990). CTD phosphorylation is a major mechanism by which cells regulate gene expression, with serines at positions 2 and 5 as the primary phosphorylation sites (Dahmus 1996). The conformational states of the prolines in the CTD also represent a critical regulatory checkpoint for transcription (Dahmus 1996; Palancade et al. 2003; Meinhart et al. 2005). By adjusting the cis-trans conformation of a proline adjacent to a phosphorylated serine, the interaction of the CTD and the binding partners it recruits can be modulated (Morris et al. 1999; Wu et al. 2000). These CTD binding proteins are involved in a variety of processes during the transcription. However, nearly all the complex structures of CTD-binding proteins and the CTD peptides solved so far contain the pSer-Pro motif in trans conformation (Zhang et al. 2010). Some examples of these CTD-binding proteins include Pcf11 (Meinhart et al. 2004), a subunit of yeast cleavage and polyadenylation factor I, and mRNA capping enzyme Cgt1 (Fabrega et al. 2003).

Substantial evidence has been accumulated that Pin1 modulates the dephosphorylation of the CTD of RNA polymerase II (Wu et al. 2000; Palancade et al. 2003; Xu et al. 2003; Xu et al. 2007). Congruent with the hypothesis set out above, in the absence of Pin1-catalyzed cis/trans isomerization, a phosphatase might not be able to 'undo' the phosphorylation catalyzed by the kinase, even though they recognize the same Ser/Thr. In other words, Pin1 might significantly affect the steady-state phosphorylation level of a protein even though it has neither kinase nor phosphatase activity. Notably, the function of Pin1 itself is tightly regulated in normal tissues on both expression level and post-translational modification level (Lu et al. 2007), indicating Pin1 as a regulatory switch of the isomeric states of pSer/Thr-Pro in signal transduction.

160

About a dozen CTD-specific kinases have been identified and characterized (Prelich 2002), but CTD phosphatases are understudied. Recently, our lab and others have structurally characterized two CTD-specific phosphatases: small CTD phosphatase 1 (Scp1) (Yeo et al. 2003; Yeo et al. 2005; Zhang et al. 2006) and Ssu72 (Zhang et al. 2011), and their interactions with phosphorylated CTD (Xiang et al. 2010; Werner-Allen et al. 2011). Scp1 was identified as a neuronal gene suppressor in non-neuronal cells as well as neuronal stem cells (Yeo et al. 2005), where it epigenetically regulates the expression of a subset of genes. Scp1 is the first structurally characterized CTD-specific phosphatase with its substrate bound in human (Yeo et al. 2003; Yeo et al. 2005; Zhang et al. 2006). In our high-resolution structure of the Scp1-CTD complex (PDB code 2ght), it is obvious that $pSer_5$ is the site that undergoes dephosphorylation, while the $pSer_2$ side chain extends outwards from the protein surface, even though it is also phosphorylated. The Pro_3 and Pro_6 are both in the trans conformation, and the conversion from trans to cis would make the substrate chain clash sterically with the protein unless dramatic conformational changes occurred (Zhang et al. 2006). Ssu72, on the other hand, has been recognized as a housekeeping gene that is pivotal to the general transcription cycle. It exhibits substrate specificity toward the cis conformation of its bound CTD peptide (Xiang et al. 2010; Werner-Allen et al. 2011). It recognizes $pSer_5$-Pro_6 motif only if Pro_6 adopts the cis conformation (Xiang et al. 2010). Based on the structure, the high energy cis-proline substrate can be stabilized in part by the intramolecular hydrogen bond between the hydroxyl side chain of Thr_4 and the carbonyl group of Pro_6. This is surprising since all previously identified CTD binding proteins are trans specific. Since the cis conformation only accounts for ~20% of the peptide, the unambiguous observation in crystal structures of the cis conformation binding in Ssu72 indicates selectivity, rather than an averaging effect of crystallography. The structures strongly

161

suggest that the cis or trans conformations of prolines upstream or downstream of a pSer site in the CTD can directly determine if the specific site can be subject to the dephosphorylation by the CTD phosphatases.

We therefore hypothesized that prolyl isomerization has a substantial impact on dephosphorylation rates by changing the suitability of CTD as the substrate for CTD phosphatases, and thus affects the phosphorylation patterns of the CTD (Etzkorn 2006). In particular, Scp1 and Ssu72 might respond differently to the Pin1-mediated prolyl isomerization. However, structural and kinetic evidence were lacking. In order to test this hypothesis, we determined the structures of two conformational-locked alkene isosteres bound to Pin1 that mimic the two endpoints of the isomerization reaction: pSer-*cis*-Pro and pSer-*trans*-Pro (Wang et al. 2004). These structures not only provide insight into how Pin1 recognizes its substrates, but also demonstrate unequivocally the utility of both cis- and trans-locked alkene isosteres as close geometric mimics of peptide bonds bound to a protein target. We further show that cis-specific phosphatase Ssu72 was highly activated by Pin1 activity, but that trans-specific phosphatase Scp1 was not affected substantially. These data illustrate how the ability of Pin1 to 'switch' the cis and trans conformation of its substrates may have significant implications for the regulation of RNA polymerase II-mediated transcription.

RESULTS AND DISCUSSION

The binding of cis and trans isosteric compounds to Pin1

In order to promote Pin1 crystals to endure prolonged chemical soaking, it was necessary to engineer Pin1 using an entropy reduction strategy (Zhang et al. 2007). R14A mutation of Pin1 has been shown to dramatically stabilize the protein crystal, yet it has little impact on the PPIase activity or WW domain binding on the substrates of Pin1 (Lu

et al. 1999). The R14A mutant also shows identical binding modes as wild-type protein when bound to high-affinity inhibitor Ac-L-Phe-D-pThr-L-Pipecolic acid-L-Naphthylalanine-L-Gln-NH$_2$ (D-PEPTIDE) (Zhang et al. 2007). Therefore, in our current investigation, crystals of this mutant were used to soak with cis and trans isosteric compounds. This strategy was extremely effective and the crystals diffracted to 2.1 and 2.3 Å on an in-house X-ray source for cis and trans complexes respectively.

The overall structure of Pin1 is highly consistent with previously reported structures. Briefly, human Pin1 has two distinctive domains, a WW domain that recognizes the signature motif pSer/Thr-Pro, and a PPIase domain that catalyzes the reaction of the prolyl-peptide. The linker between the two domains is highly flexible and disordered in all of the structures of Pin1 published thus far (**Figure 7-1**) (Ranganathan et al. 1997; Verdecia et al. 2000; Zhang et al. 2007). The flexible nature of the linker is inherited throughout the Pin1 family. The mobility of the linker and the inter-domain movement is proposed to be an integral regulatory mechanism for the communication between WW and PPIase domains and essential for the biological function of the protein (Li et al. 2005; Namanja et al. 2011). A PEG400 molecule, used as additive in the crystallization buffer, was found as usual in the groove between the two domains. This PEG400 molecule stabilizes the mobility between the two domains and enables the crystallization of Pin1 molecules. The only exception is the structure of Pin1 with the phosphoryl-peptide derived from the CTD of RNA polymerase II, in which case the peptide replaced the PEG molecule (PDB code: 1f8a) (Verdecia et al. 2000).
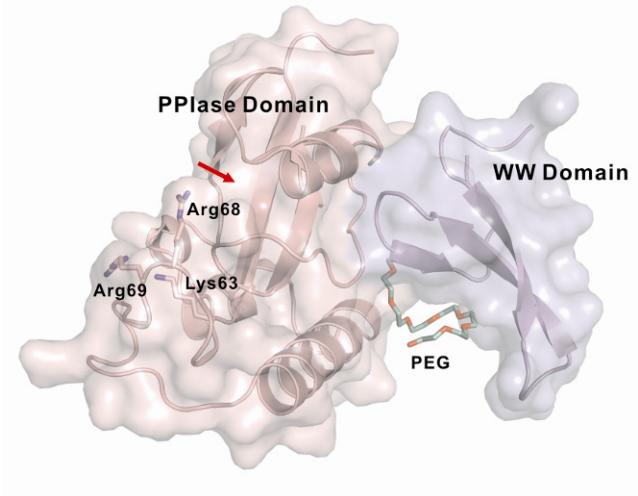
Figure 7-1:   Overall structure of Pin1 represented by cartoon and surface.[39]

The complex structures show that both cis and trans alkene compounds (**Figure 7-2**) bind to the PPIase domain of the Pin1. The structures indicate that Pin1 recognizes both cis and trans substrates in very similar conformations, even though the mode of the proline 5-membered ring analogue bound to the proline binding pocket is slightly different. The PPIase domain has two distinctive binding areas for each residue of the signature motif, pSer/Thr-Pro. Three essential residues of the PPIase domain, Lys63, Arg68 and Arg69, form a positive triad pocket that specifically binds to the phosphate group (**Figure 7-1**). The elimination of any of these residues greatly diminishes the activity of the enzyme but does not totally abolish the isomerization activity (Lu et al. 1999). However, eliminating two out of the three residues reduces the activity of the enzyme to an undetectable level (Lu et al. 1999). These residues, embracing the phosphate group of the substrate with electrostatic interactions, form a roomy and elastic

---

[39] The PPIase domain is colored light pink, and the WW domain is colored light blue. The linker in between of the two domains is missing due to its high flexibility. The residues, Lys63, Arg68 and Arg69, that bind phosphate group is shown as sticks. The red arrow indicates the pocket that recognizes proline. PEG molecule is shown in green and red as sticks.

pocket that can accommodate "rolling" of the phosphate. When not occupied, the positive triad loop preserves an open conformation that can close up upon inhibitor binding (**Figure 7-3**) (Verdecia et al. 2000). A hydrophobic pocket that binds to the proline residue of the substrate is also accountable for the unique selectivity of Pin1 isomerase. This greasy pocket highly prefers hydrophobic residues like proline. When a proline-containing compound is not provided in solution, density from additives sometimes can be found in the crystal structure, most likely due to non-specific binding. This implies a strong preference for hydrophobic interactions in this binding area and provides clues for inhibitor design (Zhang et al. 2007).
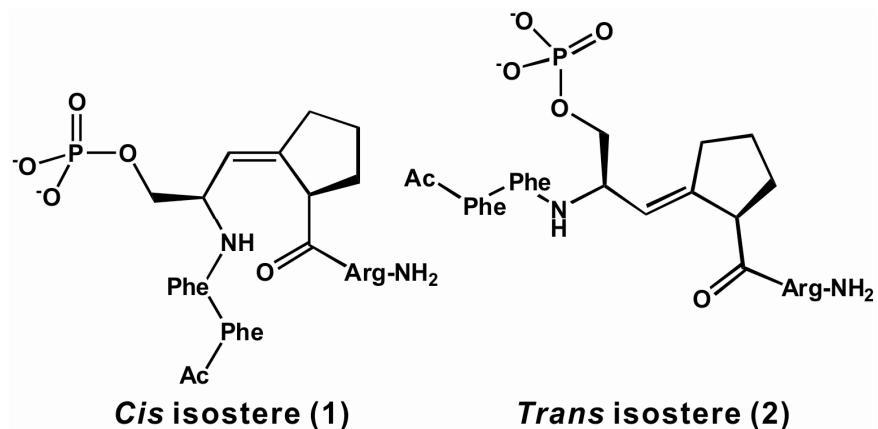
Figure 7-2:   Chemical structures of the cis and trans peptidomimetic inhibitors of Pin1: cis isostere (1) and trans isostere (2).
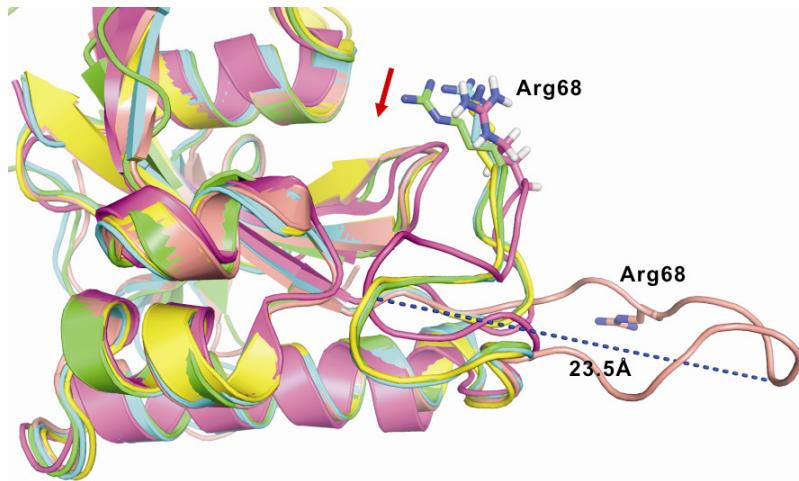
Figure 7-3:   Superimposition of multiple Pin1 structures published in PDB shows "open" and "closed" conformations of the loop containing Arg68.[40]

The Pin1 R14A-cis-isostere complex structure

The cis isosteric inhibitor binds to the PPIase domain, as predicted based on the kinetic results that the compound is a competitive PPIase inhibitor (Wang et al. 2004). Three residues of the inhibitor are ordered and modeled in the density at the active site (**Figure 7-4a**). Even though it has been proposed that the presence of hydrophobic residues N-terminal to the signature motif can enhance binding (Yaffe et al. 1997), the two phenylalanine residues in the cis peptidomimetics are not visible in our structure. This is similar to another complex structure of Pin1 (PDB code: 2itk) with high affinity peptide inhibitor (K_i of 19 nM) where N-terminal hydrophobic residue (Phe) immediately preceding the pSer/Thr-Pro motif (PDB code: 2itk) was also disordered (Zhang et al. 2007). The consistent lack of order in these amino acids suggests that the residue(s) N-terminal to the pSer/Thr-Pro motif might contribute very little structurally to the tight binding between Pin1 and those inhibitors. We took advantage of this observation in our

---

[40] The ones that have ligand bound at the phosphate-binding pocket (red arrow) include 1pin (green), 2itk (cyan), 1nmw (magenta), and 3jyj (yellow). The structure with PDB code 1f8a (pink) has no ligand at the phosphate-binding pocket, possessing the loop swung out about 23.5 Å from the "closed" conformation.

ensuing design of a reduced amide inhibitor of Pin1, Ac–pSer–Ψ[CH$_2$N]-Pro–tryptamine (Xu et al. 2011). In this case, no hydrophobic residues were placed in front of the pSer, the resultant compound exhibited much improved solubility that allowed structural determination.
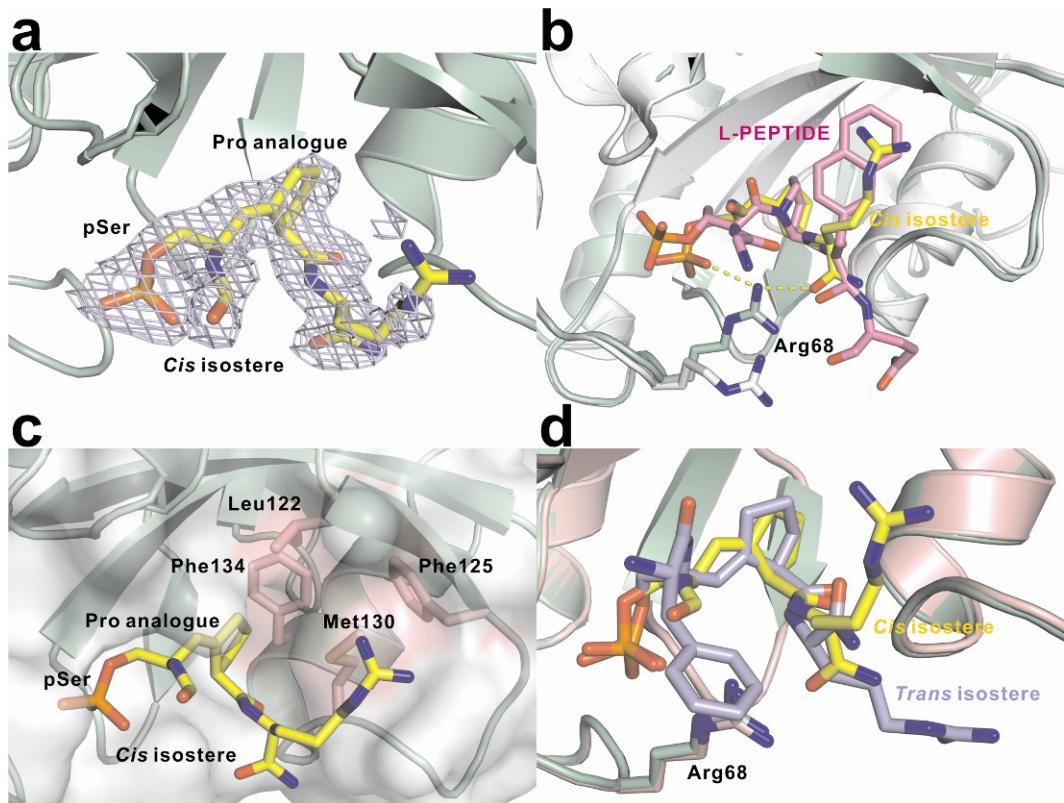
Figure 7-4:   Complex structures of Pin1 bound with cis or trans isosteres.[41]

One interesting aspect of our structure is the conformation of Arg68 of Pin1, whose side chain was disordered in the previous PPIase complex structure (Zhang et al.

---

[41] (a) Electron density map (2F$_o$-F$_c$) of cis isostere contoured at 1σ. (b) Superimposition of Pin1 bound with cis isostere (yellow) and L-PEPTIDE (magenta, PDB code: 2q5a). Pin1 bound with cis isostere is shown in pale green, and the Pin bound with L-PEPTIDE is shown in white. Arg68 in both structures is shown as sticks. The yellow dashed lines indicate the hydrogen bonds. (c) Hydrophobic pocket (pink surface) which recognizes Pro analogue of the peptidomimetic inhibitor. The key hydrophobic residues are shown in pink as sticks. (d) Superimposition of Pin1 bound with cis (yellow) and trans (light blue) isosteres.

2007). The loop containing the positively charged triad (Arg68, Arg69 and Lys63) forms favorable electrostatic interactions with phosphate group, stabilizing the interaction between protein and peptide inhibitor. The loop can adopt two dramatically different conformations, a closed conformation when a negatively charged group, such as inorganic sulfate or a phosphate group from substrate or substrate analogues, occupies the active site, and an open conformation when the site is unoccupied. Compared with the closed conformation, the tip of the loop swings 23.5 Å away when the pocket is empty (**Figure 7-3**). Furthermore, even in the closed conformation, the side chain of Arg68 is highly flexible in different structures, and it is usually totally disordered. However, in this pair of structures of Pin1 with substrate analogues, Arg68 covers the entrance of the active site cavity and provides a lid with hydrogen bonds to the amides of both inhibitors' C-terminal arginine residues and phosphates (**Figure 7-4b**). The position of the Arg68 side chain is also very close to the alkene bond that mimics the peptidyl-prolyl bond in the substrate. The highly mobile Arg68 side chain enables a very elastic binding pocket for phosphate, and permits the rotation of the phosphate group upon isomerization.

The proline-binding pocket, on the other hand, is less flexible and favors certain positioning of proline over others. This pocket is composed of Phe125, Phe134, Met130 and Leu122 (**Figure 7-4c**), which provides hydrophobic interactions with the proline-mimic moiety of the inhibitors. Hydrophobic interactions usually grant high-affinity binding for inhibitors. For example, rapamycin presents a 0.2 nM K_d toward another PPIase, FKBP12 of the FK506 binding protein (FKBP) family, yet all the strong interactions are driven by hydrophobic interactions with only one potential hydrogen bond between the ligand and the enzyme (PDB code: 2dg3). Exploitation of the proline-binding pocket in Pin1 will help us to design inhibitors for human Pin1 with stronger affinity and selectivity. Indeed, the high affinity of D- and L-PEPTIDE is at least partially

attributed to the exchange of proline residue by a 6-membered ring analogue of proline, pipecolic acid (Zhang et al. 2007). The natural substrate of Pin1 at this site, proline, actually has relatively weak binding to the PPIase domain, which is mimicked by the position of the proline analogue in our structure (**Figure 7-4c**). Instead, the WW domain shows a strong affinity for pSer/Thr-Pro sequences, and may function as a recruiter for natural substrates. Pin1 PPIase only binds substrate with K_d in millimolar range, therefore it is believed that the PPIase domain can only take the substrate after the WW domain targets the protein to the substrate (Ranganathan et al. 1997; Lu et al. 1999). This dual mode of action has been the center of investigation for human Pin1 function (Daum et al. 2007; Namanja et al. 2011), but how the two domains communicate and coordinate catalysis remains to be elucidated.

The Pin1 R14A-trans-isostere complex structure

The structure of Pin1 bound to the inhibitor mimicking trans-proline exhibits a very similar conformation as its cis counterpart (**Figure 7-4d**). Consistently, Arg68 is ordered in this structure and covers the active site entrance. However, one interesting distinction from the cis complexes is the much weaker density at the alkene bond that mimics the prolyl peptide bond even though the densities of the compound at the phosphate binding site and proline binding pocket are rather strong. Considering that both complexes were obtained with similar amounts of soaking time at similar resolutions, this suggests that the isosteric bond is more ordered in the cis compound compared to the trans. The K_i of the trans compound is 23-fold higher than that of the cis compound, possibly due to the different binding modes of the proline residues as restricted by the cis or trans conformation. Alternatively, it is possible that the binding of the trans conformation of carbocyclic proline analogue exerts strain upon binding to Pin1, as

evidenced by the protein dynamics measured by NMR (Namanja et al. 2011). In contrast, the cis conformation of the alkene bond introduces less distortion, resulting in more favorable binding.

Impacts on CTD dephosphorylation mediated by Scp1 and Ssu72

Even though both cis- and trans-proline are suitable substrates for Pin1, as mimicked in our structures, the impact of isomerase activity is not the same on different enzymes recognizing different prolines. Since cis-proline is only a minor component in naturally occurring proteins, enzymes recognizing cis-proline as substrate will have their substrate pool greatly affected by Pin1 activity. Phosphatases targeting the same substrate sequence motif, but requiring different proline isomers, represent the best system to test this. Scp1 and Ssu72 are eukaryotic phosphatases recognizing the Ser_5 position of the CTD. However, their complex structures suggest that Scp1 and Ssu72 prefer different proline conformations. We have examined how Pin1 activity affects the phosphatase activity of these enzymes.

In order to investigate whether Pin1 can regulate the activity of CTD-specific phosphatases, we tested the dephosphorylation of a CTD-derived peptide using the malachite green assay. This peptide includes four repeats, $(YSPTpSPS)_4$, with all four Ser_5 phosphorylated. Previously, Ssu72 has been shown to be more active upon the addition of yeast homologue of Pin1, Ess1 (Werner-Allen et al. 2011). We asked whether similar effects would be observed when human Pin1 was used. The isomerization effects of Pin1 on another CTD phosphatase, transcription factor IIF-interacting CTD-phosphatase 1 (Fcp1), have been controversial when both activation with human Pin1 (Kops et al. 2002) and inhibition with yeast homologue Ess1 (Xu et al. 2003) have been observed. We reasoned that the discrepancy comes from the different recognition sites for

Fcp1 and Pin1. Pin1 binds the Ser_5 of the CTD (Verdecia et al. 2000) whereas Fcp1 highly favors Ser_2, and only binds Ser_5 weakly (Hausmann et al. 2002). When the same recognition motif is being recognized, the effect of the PPIase towards isomer-specific phosphatases should be more consistent between the yeast and human versions of the PPIase.

In this study, we used *Drosophila* Ssu72 to test how Pin1 affects its activity. *Drosophila* Ssu72 shares 60% identity with human Ssu72, and structural conservation of 0.56 Å in the main chain (Zhang et al. 2011). The active site superimposes perfectly between the *Drosophila* and human counterparts (**Figure 7-5**). The *Drosophila* version of Ssu72 has much higher thermostability, making it a better version to use for the kinetic experiments. Consistent with prior reports (Xiang et al. 2010; Werner-Allen et al. 2011), Ssu72 is activated upon Pin1 addition (**Figure 7-6a**) by about 3-fold. This result is consistent with a scenario in which Pin1 quickly converts the trans-Pro to cis-Pro, and by doing so, makes the cis-trans ratio reach equilibrium much faster than the uncatalyzed auto-conversion. Therefore, the consumable substrate concentration for Ssu72 was increased in the presence of Pin1, and led to the apparent increased activity of Ssu72. Such effect is specifically caused by the prolyl isomerase activity of Pin1, because when we used a truncated version of Pin1, PPIase domain, which cannot target CTD substrate to PPIase active site (Verdecia et al. 2000), the activation effect is lost (**Figure 7-6a**).
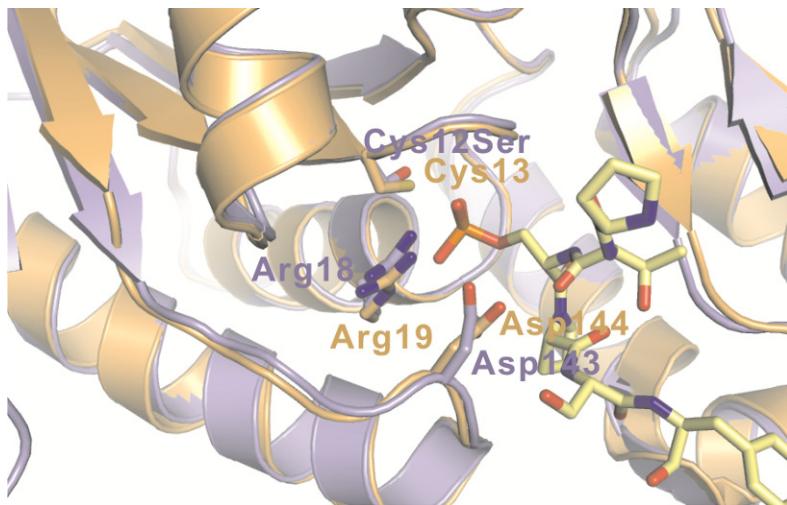
Figure 7-5: Superimposition of the active sites of the human (light blue, PDB code: 3o2q) and *Drosophila* (light orange, PDB code: 3omw) Ssu72.[42]

It should be noted that the fraction of peptide that was dephosphorylated by Ssu72 (~1/3) consists of both substrate that was in cis-conformation initially [estimated to be ~20%], and substrate that was auto-converted to cis-conformation during the process of the reaction. The difference between the reactions with and without Pin1 is caused by the effect of Pin1 'outracing' trans-to-cis auto-conversion.

Combined with a previous experiment that a catalytically impaired mutant of yeast homologue of Pin1, Ess1, cannot activate Ssu72 (Werner-Allen et al. 2011), our result shows that it is the isomerization of the CTD that promotes the enhanced phosphatase activity of Ssu72, rather than stabilization of Ssu72 protein or reducing the non-specific adsorption of Ssu72 protein to the test tube. Furthermore, we ruled out the possibility that Pin1 activates Ssu72 by physically interacting with Ssu72. Firstly, Pin1 specifically recognizes a Ser/Thr-Pro motif in its substrates only when the Ser/Thr is phosphorylated. However, Ssu72 contains no pSer/Thr-Pro motif in its primary sequence.

---

[42] The three key residues at the active site are shown as sticks. The nucleophile Cys12 was mutated to Ser in the human Ssu72 in order to obtain the complex structure with the peptide (shown as yellow sticks).

Secondly, we tested whether Pin1 and Ssu72 can directly interact with each other to form a stable complex using gel filtration chromatography. In this experiment, roughly equal amount of Pin1 and Ssu72 (~300 μM each) were mixed together and incubated at 4 °C for 6 hr. The mixture was then loaded on Superdex 75 column (GE Healthcare). No peak corresponding to a possible Pin1-Ssu72 complex was observed (**Figure 7-7**). This experiment shows that the enhancement of Ssu72 activity is not due to its physical interaction with Pin1.
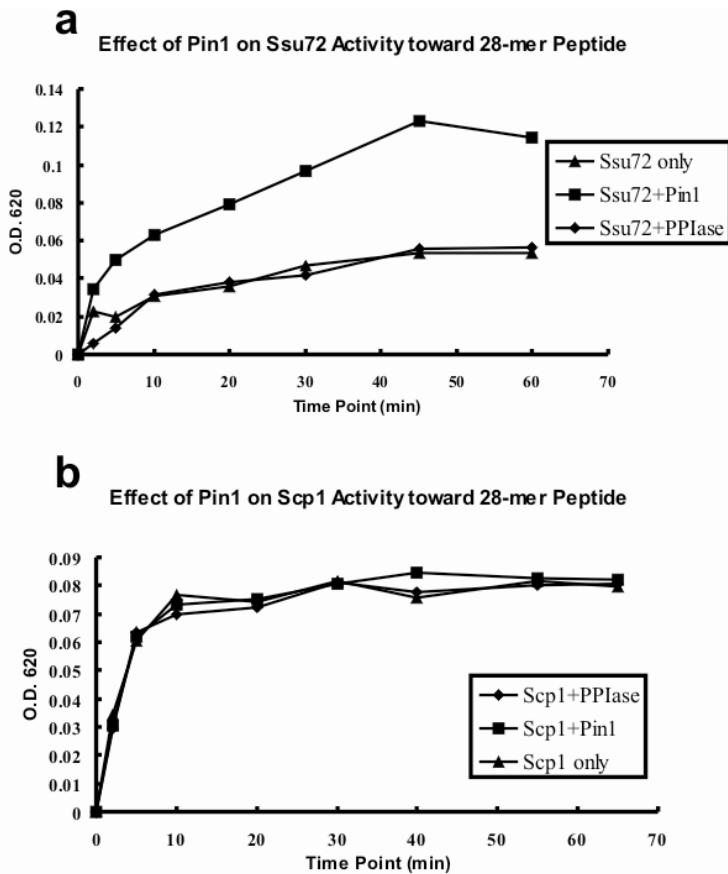
Figure 7-6: Effect of human Pin1 on the activities of *Drosophila* Ssu72 (a) and human Scp1 (b) phosphatases.[43]

In contrast, when we tested human Scp1 in the same assay, the phosphatase activity is not significantly affected by Pin1 (**Figure 7-6b**). The insensitivity of Scp1 to Pin1 is consistent with the structural observations for prolyl isomeric states. Both Pro3 and Pro6 of the CTD peptide exhibit only the trans conformation in the complex structure

---

[43] The activities of both phosphatases toward a 28-mer peptide [sequence: (YSPTpSPS)$_4$] were measured using malachite green assay. The Pin1 is wild-type full length protein and PPIase domain is a truncated version of Pin1 with residues 51-163. (a) The reaction (20 µL total volume) for Ssu72 was carried out in buffer containing 100 ng of Ssu72, 20 µM of peptide, 100 mM MES pH 6.5, and 10 ng of Pin1 or PPIase domain. (b) The reaction of Scp1 was performed in the buffer containing 5 ng of Scp1, 10 µM of peptide, 50 mM Tris-acetate pH 5.5, 10 mM MgCl$_2$ and 10 ng of Pin1 or the PPIase domain. The reactions were quenched by adding 40 µL of malachite green reagent at different time points. The release of inorganic phosphate was detected by measuring the absorbance at 620 nm.

of Scp1 and CTD peptide (Zhang et al. 2006). Unlike Ssu72, Pro_3 two residues upstream of $pSer_5$ is a recognition determinant for Scp1 binding. Since the majority of the peptide substrate [estimated to be 80%] has the proline in the trans conformation, Scp1 recognizes the substrate readily and dephosphorylates the substrate. In this case the addition of Pin1 only marginally improves substrate accessibility of Scp1. The slight improvement is hard to distinguish due to the sensitivity level of malachite green assay and is thus insignificant.
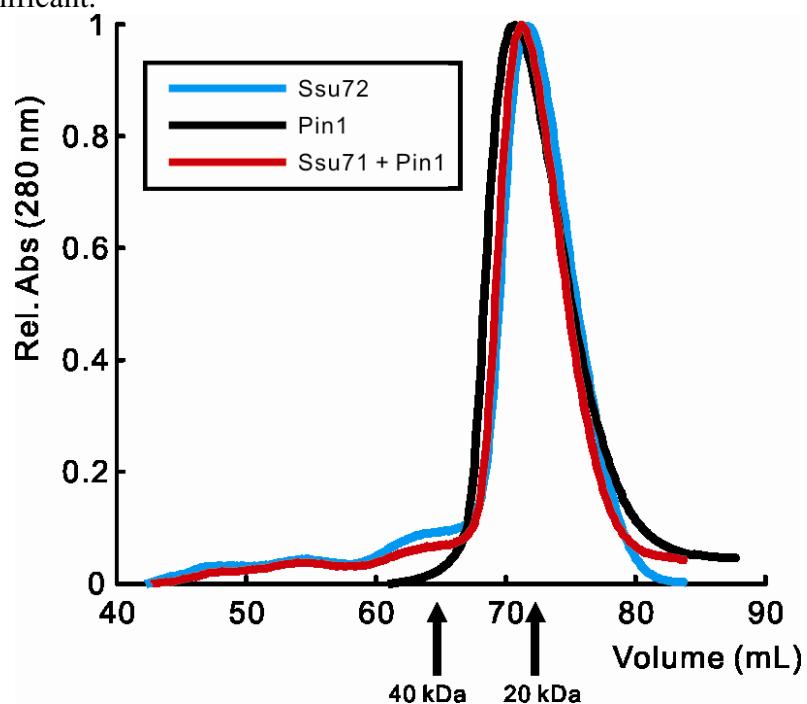
Figure 7-7:   Gel filtration traces of Ssu72 (~23 kDa), Pin1 (~18 kDa) and Ssu72-Pin1 mixture generated by Superdex 75 column (GE Healthcare).[44]

---

[44] The wild-type *Drosophila* Ssu72 and human Pin1 were purified by following the same procedure described in **Methods**. Ssu72-Pin1 mixture was prepared by mixing nearly equal molar concentration (~300 µM) of Ssu72 and Pin1 in 4 °C and incubated for 6 hr. The arrows indicate where the standard 40 kDa and 20 kDa protein should be.

Implication of Pin1 mechanism from the structures

Both cis- and trans-prolines are subject to isomerization by Pin1 with cis and trans alkene inhibitors mimicking substrate/product of the Pin1 isomerization reaction. The pSer-Pro dipeptide with either cis or trans conformation is modeled, based on our present complex structures, in the Pin1 structure to illustrate the real substrate binding (**Figure 7-8b, c**). The complex structures show that the same structural elements are used to recognize the substrate/product and interconvert the two species to reestablish the equilibrium cis:trans ratio.
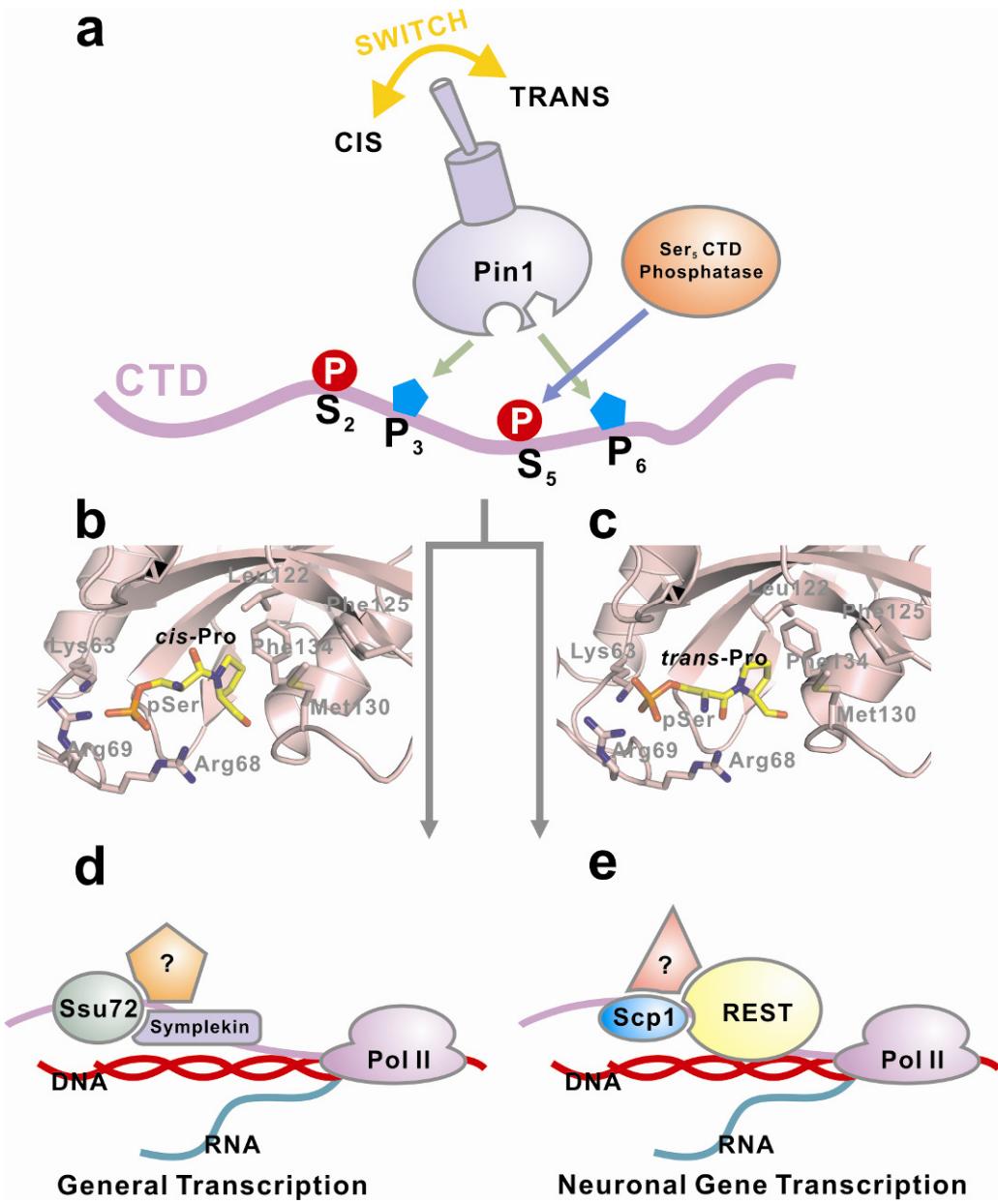
Figure 7-8: Model of the cross-talk between Ser$_5$ dephosphorylation and prolyl isomerization of the CTD.[45]

---

[45] (a) The Pin1 acts as molecular switch that changes the isomeric state of the two prolines, resulting in recruitment of different transcription complex and therefore, different outcomes of transcription. (b) pSer-*cis*-Pro dipeptide modeled in Pin1 structure. (c) pSer-*trans*-Pro dipeptide modeled in Pin1 structure. (d) Ssu72 is recruited with specific regulatory factors (colored shapes) to control general gene transcription in response to cis-proline. (e) Scp1 is recruited with a different set of regulatory factors (colored shapes) to control neuronal gene transcription in response to trans-proline.

A challenge of this field has been making substrates locked in only one conformation. Alkenes have a long history as peptide bond isosteres. The carbon-carbon double bond is close to the same length as the amide carbonyl-nitrogen bond, 1.40 Å vs 1.32 Å, and the distance between the α-carbons is identical, 3.8 Å (Shue et al. 1993). The dynamics of both cis- and trans-locked ligands are dramatically affected by which conformation is bound to Pin1; cis is more rigid than trans, and the rigidity of bound cis results in 23-fold tighter binding (Wang et al. 2004; Namanja et al. 2011). The protein dynamics of a conduit between the PPIase and WW domains of Pin1 are also differentially affected by binding of cis- or trans-locked substrate isosteres (Namanja et al. 2011).

Previously, we have obtained crystal structures of Pin1 complexed with two high affinity peptide inhibitors; Ac-Phe-(D/L)-pThr-Pip-Nal-Gln-NH$_2$, (D-PEPTIDE or L-PEPTIDE, respectively). Our new structures replace the prolyl-peptide which is subject to isomerization with non-rotatable carbon-carbon double bond, locking the two states of substrate-bound mode of Pin1. When we compared these two pairs of complex structures of Pin1, substrate-mimicking (cis and trans isosteres) versus high affinity inhibitors (D- and L-PEPTIDE), it has been observed that the phosphate positions of different Pin1 inhibitors are highly diversified. The architecture of the triad positive residues allows the rolling of the phosphate group to form electrostatic interactions with Lys63, Arg68 and Arg69. The side chains of these three residues also adopt different conformations with each isomer to accommodate different positions of the phosphate of the substrate, allowing the rotation of the phosphate group, yet still within the pocket. On the contrary, the C-terminus of the signature motif (pSer/Thr-Pro) provides a strong hold for the peptide. The proline pocket does not accommodate free rotation of the proline and the hydrogen bonding between the carbonyl of proline and the amide of Gln131 is highly

178

conserved among all of our Pin1 structures (**Figure 7-9**). These observations suggest that dynamic interactions of the phosphate group and the protein allows the rotation of the prolyl peptide at the N-terminus of the peptide subject to isomerization, which echoes the discovery in NMR studies on the issue (Labeikovsky et al. 2007; Namanja et al. 2010). In addition, the flexibility of the interactions between the phosphate group and the positive triad during the rotation permits transition-state stabilization (Xu et al. 2011).
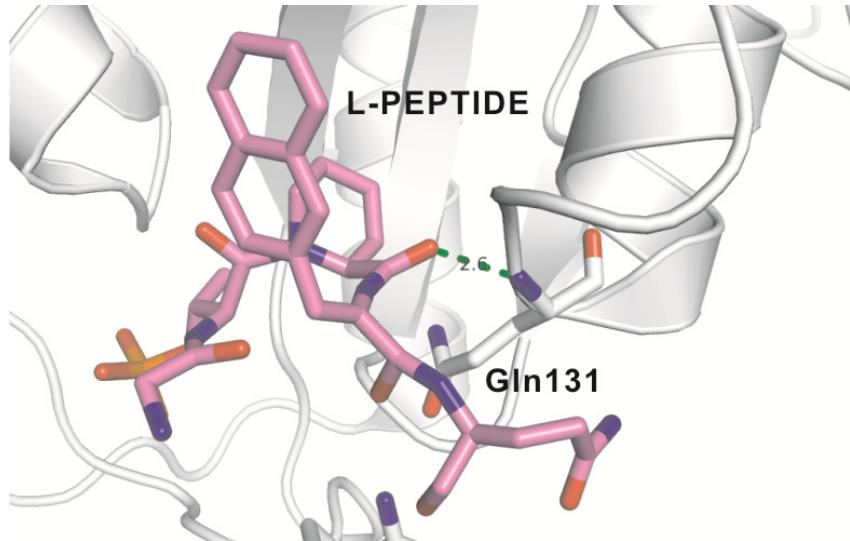
Figure 7-9:   Hydrogen bond (green dashed line) formed between the carbonyl of proline (or proline analogue) and the amide of Gln131.[46]

Implications for the regulatory mechanism of CTD

During the progression of the RNA polymerase II-mediated transcription cycle, many CTD-specific kinases and phosphatases are recruited to the CTD. The dynamic phosphorylation/dephosphorylation is a major regulatory mechanism of the CTD, greatly influencing transcription. However, various phosphatases and kinases may have different specificity toward the isomeric states of the prolines adjacent to the major

---

[46] Shown here is an example from complex structure of Pin1 and L-PEPTIDE (PDB code: 2q5a).

phosphorylation sites. Even though cis and trans forms of proline can reach equilibrium slowly under thermal isomerization called auto-conversion, the rate is too slow to allow efficient signal transduction in cells. Therefore, Pin1-mediated prolyl isomerization of the CTD is necessary to couple with the phosphorylation regulation to generate suitable substrates for both cis- and trans-specific kinases/phosphatases (**Figure 7-8a**). The dynamic nature of the CTD phosphorylation states during transcription determines that different outcomes can be reached for different phosphatases when their proline isomeric specificity is different (**Figure 7-8**).

In such a scenario, the prolyl isomerase activity can greatly affect the outcome when the pool of one species of the substrate is rapidly depleted. For a cis-specific enzyme, such as Ssu72, the isomerase activity of Pin1 guarantees the availability of cis-form substrate that is rapidly depleted. On the other hand, since Scp1 utilizes the trans-form CTD as substrate, which is the major species, the apparent dephosphorylation will not be affected much by Pin1. So even though both Scp1 and Ssu72 recognize the pSer$_5$ of CTD as substrate, their response towards Pin1 PPIase activity in cells will differ dramatically (**Figure 7-8d, e**). Their different responses will in turn affect the regulatory factors recruited to the vicinity of genes during transcription. Specifically, Ssu72 together with its binding partner, the scaffold protein symplekin (Xiang et al. 2010) and other regulatory factors from the cleavage/polyadenylation specificity factor (CPSF) complex (Krishnamurthy et al. 2004), will be recruited in response to pSer-*cis*-Pro, thus regulate general transcription. On the other hand, Scp1, together with REST complex (Yeo et al. 2005), will be recruited in response to pSer-*trans*-Pro to turn off neuronal gene expressions. To a certain extent, the phosphorylation state of the CTD is governed not only by the phosphorylation/dephosphorylation mechanism, but also the prolyl isomerization mechanism (**Figure 7-8**). The combination of the various post-translational

modifications on CTD can lead to different transcription outcome and therefore, various fates for the cell. The recognition of both phosphorylation and isomerization states of CTD by partner proteins are very likely to be a general mechanism adopted by other CTD-binding proteins in transcription regulation, indicating a "combinatorial" CTD code.

CONCLUSIONS

In this study, we determined the complex structures of human Pin1 with two isomer-locked peptidomimetics that mimic the substrates in the cis or trans form of a pSer-Pro peptide bond. The recognition by Pin1 has an impact on the downstream regulatory phosphatases of CTD to a different extent based on their specificity towards proline isomeric states. The existence of Pin1 isomerase activity can greatly stimulate the activity of a cis-proline specific phosphatase by increasing the potential substrate pool. However, the Pin1 effect is more limited on trans-proline specific phosphatases. Therefore, the up-regulation of Pin1 activity can alter the signal transduction pathway in CTD-mediated transcriptional regulation. The cross-talk between prolyl-isomerase and CTD phosphatases can differentially lead to various transcriptional outcomes in cells.

METHODS

Synthesis of the cis and trans peptide mimetic inhibitors

The cis and trans isosteres, Boc–Ser–Ψ[(*Z/E*)CH=C]-Pro–OH, where (*Z*) is the cis mimic, and (*E*) is the trans mimic, were synthesized as previously reported (Wang et al. 2003). Both peptidomimetics (**Figure 7-2**), Ac–Phe–Phe–pSer–Ψ[(*Z/E*)CH=C]-Pro–Arg–NH$_2$, were synthesized using solid-phase peptide synthesis with the Fmoc-protected, block-phosphorylated isosteres as described previously (Wang et al. 2004).

Purification of human Pin1 and human Scp1

The human Pin1 or Scp1 gene was sub-cloned in a pHIS8 vector, a derivative of pET28a vector (Novagene) (Jez et al. 2000). The Pin1 R14A mutant was produced using the QuikChange Site-Directed Mutagenesis Kit (Stratagene, CA). The purification of Pin1 R14A mutant or Pin1 PPIase domain (residue 51-163) was identical to the procedure previous reported (Zhang et al. 2007). Concisely, the protein was overexpressed using *E. coli* BL21(DE3) strain with isopropyl-β-D-thiogalactopyranoside (IPTG) induction at 16 °C overnight. The cells were pelleted and lysed with subsequent nickel affinity chromatography purification. After imidazole elution of the HIS-tagged recombinant protein, the N-terminal polyhistidine-tag was truncated with thrombin protease during dialysis at 4 °C, and subsequently purified on ion-exchange and size exclusion chromatography columns. The purified protein was homogenous on SDS-PAGE gel.

Purification of *Drosophila* Ssu72

A pET28b derivative vector, pETHIS8–SUMO, encoding *Drosophila* Ssu72 proceeded by an N-terminal 8xHIS-SUMO tag was constructed previously (Zhang et al. 2011). The protein was overexpressed in *E. coli* BL21 (DE3). The cells were grown at 37 °C in Luria-Bertani medium supplemented with 50 µg/mL kanamycin, then induced with 0.5 mM IPTG at 16 °C when the O.D. at 600nm reached 0.8. After overnight incubation, the cells were harvested by centrifugation and disrupted by sonication. Recombinant protein was initially purified with Ni-NTA column (Qiagen, Switzerland). The N-terminal 8xHIS-SUMO tag was then removed by PreScission protease (GE Healthcare). The protein was further purified by a size exclusion column Superdex-75 (GE Healthcare), equilibrated with 25 mM Tris-HCl (pH8.0) and 200 mM NaCl buffer. The collected Ssu72 protein was passed through the Ni-NTA column again to remove

any heterogeneous proteins, as evaluated by SDS-PAGE gels. The pure Ssu72 protein was finally flash frozen in liquid nitrogen and stored at –80 °C.

Crystallization, soaking and data collection

The R14A variant of human Pin1 was crystallized by vapor diffusion using a hanging drop of 1 μL protein plus 1 μL well solution. The crystals were obtained at 1.9-2.2 M ammonium sulfate, 1% PEG400 at pH 7.5 in 50 mM HEPES buffer. The crystals were then transferred to mother liquor containing 40% PEG400 and 50 mM HEPES pH 7.5 with 0.2 mM of peptide mimetic inhibitor. The crystals were soaked for 4 weeks with buffer exchange using fresh mother-liquor containing peptide mimetic inhibitor every week. The crystals were then frozen in liquid nitrogen and subjected to in-house X-ray beam using DIP100 imaging plate (MacScience, CO) with data collection. Diffraction data were processed using HKL2000. The statistics of the data are summarized in **Table 7-1**.

Structure determination and analysis

The complex structure of human Pin1 R14A with cis or trans peptide mimetic inhibitors were determined using molecular replacement with Pin1 complex structure with a high affinity inhibitor (PDB code: 2itk) as a search model. The solution of the structure was identified using AMoRe, a program in the CCP4 program suite (Navaza 1994). The refinement of the complex structures was performed using the program refmac in CCP4 (Vagin et al. 2004). Electron density maps (sigmaA weighted $2F_o$-F_c and F_o-F_c maps) were calculated after each cycle of refinement, and inspected to guide model rebuilding using Coot (Emsley et al. 2004). The quality of the final model was evaluated using Procheck (CCP4 1994). The statistics of the final model for both structures are summarized in **Table 7-1**.

Malachite green assay for Scp1 and Ssu72

The activity of the CTD phosphatases Scp1 and Ssu72 in the presence or absence of Pin1 toward 28-mer CTD peptide was measured in this assay (Martin et al. 1985). The 28-mer peptide contains 4 repeats of the consensus sequence with each Ser_5 phosphorylated: $(YSPTpSPS)_4$. The reaction (20 µL total volume) for human Scp1 was carried out in buffer containing 5 ng of Scp1, 10 µM of peptide, 50 mM Tris-acetate pH 5.5, 10 mM $MgCl_2$ and 10 ng of Pin1 or the PPIase domain, and was incubated at 37 °C. The reaction (20 µL total volume) of *Drosophila* Ssu72 was performed in the buffer containing 100 ng of Ssu72, 20 µM of peptide, 100 mM MES pH 6.5, and 10 ng of Pin1 or PPIase domain, and was incubated at 28 °C. The reactions were quenched by adding 40 µL of malachite green reagent at different time points. The release of inorganic phosphate was detected by measuring the absorbance at 620 nm.

ACCESSION CODE

Coordinates of the Pin1-cis compound and Pin1-trans compound complex structures have been deposited in the Protein Data Bank with the accession numbers 3tcz and 3tdb.

	Pin1 w/ cis compound	Pin1 w/ trans compound
Data collection		
Space group	P3$_1$21	P3$_1$21
Cell dimensions: *a*, *b*, *c* (Å)	69.4, 69.4, 79.6	69.3, 69.3, 79.7
α, β, γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 120.0
Resolution (Å)	50.00 – 2.10 (2.18 – 2.10) [*]	50.00– 2.26 (2.34 – 2.26) [*]
No. of unique reflections	12458 (1106)	10488 (908)
R_{sym} or R_{merge} (%)	5.4 (47.6)	4.3 (30.1)
I/σ(I)	26.4 (2.5)	33.2 (5.4)
Completeness (%)	93.2 (84.8)	97.5 (85.9)
Redundancy	5.5 (5.0)	4.6 (4.3)
Refinement		
Resolution (Å)	33.19 – 2.10	47.95 – 2.27
No. of reflections (test set)	10431 (1187)	9168 (1037)
R_{work} / R_{free} (%) [#]	22.3 / 26.5	21.7 / 25.6
No. of atoms: Protein	1164	1164
Ligand	30	39
PEG	24	24
Water	87	64
B-factors (Å^2): Protein	32.6	29.0
Ligand	47.0	55.6
PEG	31.7	30.2
Water	39.1	32.5
R.m.s deviations: Bond lengths (Å)	0.011	0.020
Bond angles (°)	1.411	1.977
Ramachandran plot (%): Most favored	92.8	92
Additionally allowed	6.4	7.2
Generally allowed	0.0	0.0
Disallowed [a]	0.8	0.8

Table 7-1:   Crystallographic data statistics.[47]

---

[47] [*] Highest resolution shell is shown in parenthesis. [#] R_{free} is calculated with 10% of the data randomly omitted from refinement. [a] Leu7 (chain A) in both structures is close to the N-terminus of Pin1.

REFERENCES

Brandts, J. F., H. R. Halvorson and M. Brennan (1975). "Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues." *Biochemistry* **14**(22): 4953-4963.

CCP4 (1994). "Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography." *Acta Cryst.* **D50**: 760-763.

Corden, J. L. (1990). "Tails of RNA polymerase II." *Trends Biochem Sci* **15**(10): 383-387.

Dahmus, M. E. (1996). "Phosphorylation of mammalian RNA polymerase II." *Methods Enzymol* **273**: 185-193.

Dahmus, M. E. (1996). "Reversible phosphorylation of the C-terminal domain of RNA polymerase II." *J Biol Chem* **271**(32): 19009-19012.

Daum, S., C. Lucke, D. Wildemann and C. Schiene-Fischer (2007). "On the benefit of bivalency in peptide ligand/pin1 interactions." *J Mol Biol* **374**(1): 147-161.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132.

Etzkorn, F. A. (2006). "Pin1 flips Alzheimer's switch." *ACS Chemical Biology* **1**(4): 214-216.

Fabrega, C., V. Shen, S. Shuman and C. D. Lima (2003). "Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II." *Molecular Cell* **11**(6): 1549-1561.

Hausmann, S. and S. Shuman (2002). "Characterization of the CTD phosphatase Fcp1 from fission yeast. Preferential dephosphorylation of serine 2 versus serine 5." *J Biol Chem* **277**(24): 21213-21220.

Jez, J. M., J. L. Ferrer, M. E. Bowman, R. A. Dixon and J. P. Noel (2000). "Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase." *Biochemistry* **39**(5): 890-902.

Kops, O., X. Z. Zhou and K. P. Lu (2002). "Pin1 modulates the dephosphorylation of the RNA polymerase II C-terminal domain by yeast Fcp1." *FEBS Lett* **513**(2-3): 305-311.

Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore and M. Hampsey (2004). "Ssu72 Is an RNA polymerase II CTD phosphatase." *Mol Cell* **14**(3): 387-394.

Labeikovsky, W., E. Z. Eisenmesser, D. A. Bosco and D. Kern (2007). "Structure and dynamics of pin1 during catalysis by NMR." *J Mol Biol* **367**(5): 1370-1381.

Li, Z., H. Li, G. Devasahayam, T. Gemmill, V. Chaturvedi, S. D. Hanes and P. Van Roey (2005). "The structure of the Candida albicans Ess1 prolyl isomerase reveals a well-ordered linker that restricts domain mobility." *Biochemistry* **44**(16): 6180-6189.

Lu, K. P. (2004). "Pinning down cell signaling, cancer and Alzheimer's disease." *Trends Biochem Sci* **29**(4): 200-209.

Lu, K. P., S. D. Hanes and T. Hunter (1996). "A human peptidyl-prolyl isomerase essential for regulation of mitosis." *Nature* **380**(6574): 544-547.

Lu, K. P., Y. C. Liou and X. Z. Zhou (2002). "Pinning down proline-directed phosphorylation signaling." *Trends Cell Biol* **12**(4): 164-172.

Lu, K. P. and X. Z. Zhou (2007). "The prolyl isomerase PIN1: a pivotal new twist in phosphorylation signalling and disease." *Nat Rev Mol Cell Biol* **8**(11): 904-916.

Lu, P. J., X. Z. Zhou, M. Shen and K. P. Lu (1999). "Function of WW domains as phosphoserine- or phosphothreonine-binding modules." *Science* **283**(5406): 1325-1328.

Martin, B., C. J. Pallen, J. H. Wang and D. J. Graves (1985). "Use of fluorinated tyrosine phosphates to probe the substrate specificity of the low molecular weight phosphatase activity of calcineurin." *J Biol Chem* **260**(28): 14932-14937.

Meinhart, A. and P. Cramer (2004). "Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors." *Nature* **430**(6996): 223-226.

Meinhart, A., T. Kamenski, S. Hoeppner, S. Baumli and P. Cramer (2005). "A structural perspective of CTD function." *Genes Dev* **19**(12): 1401-1415.

Morris, D. P., H. P. Phatnani and A. L. Greenleaf (1999). "Phospho-carboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3'-End formation." *J Biol Chem* **274**(44): 31583-31587.

Namanja, A. T., X. J. Wang, B. Xu, A. Y. Mercedes-Camacho, B. D. Wilson, K. A. Wilson, F. A. Etzkorn and J. W. Peng (2010). "Toward flexibility-activity relationships by NMR spectroscopy: dynamics of Pin1 ligands." *J Am Chem Soc* **132**(16): 5607-5609.

Namanja, A. T., X. J. Wang, B. Xu, A. Y. Mercedes-Camacho, K. A. Wilson, F. A. Etzkorn and J. W. Peng (2011). "Stereospecific gating of functional motions in Pin1." *Proc Natl Acad Sci U S A* **108**(30): 12289-12294.

Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." *Acta Cryst.* **A50**: 157-163.

Palancade, B. and O. Bensaude (2003). "Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation." *Eur J Biochem* **270**(19): 3859-3870.

Prelich, G. (2002). "RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function." *Eukaryotic Cell* **1**(2): 153-162.

Ranganathan, R., K. P. Lu, T. Hunter and J. P. Noel (1997). "Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent." *Cell* **89**(6): 875-886.

Schutkowski, M., A. Bernhardt, X. Z. Zhou, M. Shen, U. Reimer, J. U. Rahfeld, K. P. Lu and G. Fischer (1998). "Role of phosphorylation in determining the backbone dynamics of the serine/threonine-proline motif and Pin1 substrate recognition." *Biochemistry* **37**(16): 5566-5575.

Shue, Y. K., M. D. Tufano, G. M. Carrera, Jr., H. Kopecka, S. L. Kuyper, et al. (1993). "Double bond isosteres of the peptide bond: synthesis and biological activity of cholecystokinin (CCK) C-terminal hexapeptide analogs." *Bioorg Med Chem* **1**(3): 161-171.

Vagin, A. A., R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long and G. N. Murshudov (2004). "REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use." *Acta Cryst.* **D60**(Pt 12 Pt 1): 2184-2195.

Verdecia, M. A., M. E. Bowman, K. P. Lu, T. Hunter and J. P. Noel (2000). "Structural basis for phosphoserine-proline recognition by group IV WW domains." *Nat Struct Biol* **7**(8): 639-643.

Wang, X. J., S. A. Hart, B. Xu, M. D. Mason, J. R. Goodell and F. A. Etzkorn (2003). "Serine-cis-proline and serine-trans-proline isosteres: stereoselective synthesis of (Z)- and (E)-alkene mimics by Still-Wittig and Ireland-Claisen rearrangements." *J Org Chem* **68**(6): 2343-2349.

Wang, X. J., B. Xu, A. B. Mullins, F. K. Neiler and F. A. Etzkorn (2004). "Conformationally locked isostere of phosphoSer-cis-Pro inhibits Pin1 23-fold better than phosphoSer-trans-Pro isostere." *J Am Chem Soc* **126**(47): 15533-15542.

Werner-Allen, J. W., C. J. Lee, P. Liu, N. I. Nicely, S. Wang, A. L. Greenleaf and P. Zhou (2011). "cis-Proline-mediated Ser(P)5 dephosphorylation by the RNA polymerase II C-terminal domain phosphatase Ssu72." *J Biol Chem* **286**(7): 5717-5726.

Wu, X., C. B. Wilcox, G. Devasahayam, R. L. Hackett, M. Arevalo-Rodriguez, M. E. Cardenas, J. Heitman and S. D. Hanes (2000). "The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery." *Embo J* **19**(14): 3727-3738.

Xiang, K., T. Nagaike, S. Xiang, T. Kilic, M. M. Beh, J. L. Manley and L. Tong (2010). "Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex." *Nature* **467**(7316): 729-733.

Xu, G. G., Y. Zhang, A. Y. Mercedes-Camacho and F. A. Etzkorn (2011). "A reduced-amide inhibitor of pin1 binds in a conformation resembling a twisted-amide transition state." *Biochemistry* **50**(44): 9545-9550.

Xu, Y. X., Y. Hirose, X. Z. Zhou, K. P. Lu and J. L. Manley (2003). "Pin1 modulates the structure and function of human RNA polymerase II." *Genes Dev* **17**(22): 2765-2776.

Xu, Y. X. and J. L. Manley (2007). "Pin1 modulates RNA polymerase II activity during the transcription cycle." *Genes Dev* **21**(22): 2950-2962.

Yaffe, M. B., M. Schutkowski, M. Shen, X. Z. Zhou, P. T. Stukenberg, et al. (1997). "Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism." *Science* **278**(5345): 1957-1960.

Yeo, M., S. K. Lee, B. Lee, E. C. Ruiz, S. L. Pfaff and G. N. Gill (2005). "Small CTD phosphatases function in silencing neuronal gene expression." *Science* **307**(5709): 596-600.

Yeo, M., P. S. Lin, M. E. Dahmus and G. N. Gill (2003). "A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5." *J Biol Chem* **278**(28): 26078-26085.

Zhang, M., G. N. Gill and Y. Zhang (2010). "Bio-molecular architects: a scaffold provided by the C-terminal domain of eukaryotic RNA polymerase II." *Nano Rev* **1**.

Zhang, Y., S. Daum, D. Wildemann, X. Z. Zhou, M. A. Verdecia, et al. (2007). "Structural basis for high-affinity peptide inhibition of human Pin1." *ACS Chemical Biology* **2**(5): 320-328.

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." *Mol Cell* **24**(5): 759-770.

Zhang, Y., M. Zhang and Y. Zhang (2011). "Crystal structure of Ssu72, an essential eukaryotic phosphatase specific for the C-terminal domain of RNA polymerase II, in complex with a transition state analogue." *Biochem J* **434**(3): 435-444.

Perspective

The research described in this dissertation provided not only answers to several key questions regarding the function of Scps and the dynamic regulation of the CTD, but also a number of novel and powerful tools for the further study in these areas.

Firstly, our understanding of the enzymatic mechanism of Scps was significantly deepened by our capture of the phosphoryl-aspartate intermediate of Scp1. This was the first definitive evidence for the previously debated two-step mechanism of Scps. Secondly, the first selective inhibitor of Scp1, which represents the first selective inhibitor of the Fcp/Scp-family protein serine/threonine phosphatases, was identified. This discovery not only defied many stereotypes of phosphatase inhibition, but also provided a stepping-stone to develop more potent inhibitors of Scp1 for neuron regeneration. Thirdly, the potential secondary binding pocket for the CTD was identified on Scp1, which provided new insight of CTD regulation and may even facilitate the development of more potent Scp1 inhibitors. Finally, the communication between two different regulatory mechanisms of the CTD, i.e. phosphorylation and prolyl-isomerization, was established in our *in vitro* model system where Scp1, Ssu72 and Pin1 were effectors. This simplistic model represented an elegant example to delineate 'control at equilibrium' and 'control at steady-state', a concept that may underlie many regulatory processes including the dynamic regulation of transcription through the CTD.

These new findings and tools should enable us to ask deeper questions in these areas. With regard to the function of Scps, it would be interesting to identify all genes (other than neuronal genes) that are regulated by Scps. In addition, it is important to understand how Scp1 is recruited to the CTD and whether Scp1 displays processivity in the presence of binding partners. Moreover, the recent discovery of phosphorylation at

Tyr_1, Thr_4 and Ser_7 residues also prompts the question of whether additional modifications in conjunction with Ser_5 phosphorylation would influence the dephosphorylation of Ser_5 by Scp1.

Extraordinarily rapid progress has been made over the past few years in the field of CTD research; however, many important questions remain to be answered. One big hurdle is the inability of current technologies to identify the exact phosphorylation patterns across individual repeats at different stages during the transcription cycle. The ability to unambiguously identify (or 'sequence') the phosphorylation patterns at the resolution of individual repeats is likely to be essential to determine the existence and meaning of the postulated "CTD code". As the field stands now, this remains one of the most exciting and important challenges in the future of CTD research.

References

Akhtar, M. S., M. Heidemann, J. R. Tietjen, D. W. Zhang, R. D. Chapman, D. Eick and A. Z. Ansari (2009). "TFIIH kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II." *Mol Cell* **34**(3): 387-393.

Allen, K. N. and D. Dunaway-Mariano (2004). "Phosphoryl group transfer: evolution of a catalytic scaffold." *Trends Biochem Sci* **29**(9): 495-503.

Allen, M., A. Friedler, O. Schon and M. Bycroft (2002). "The structure of an FF domain from human HYPA/FBP11." *J Mol Biol* **323**(3): 411-416.

Archambault, J., R. S. Chambers, M. S. Kobor, Y. Ho, M. Cartier, D. Bolotin, B. Andrews, C. M. Kane and J. Greenblatt (1997). "An essential component of a C-terminal domain phosphatase that interacts with transcription factor IIF in Saccharomyces cerevisiae." *Proc Natl Acad Sci U S A* **94**(26): 14300-14305.

Barford, D., A. K. Das and M. P. Egloff (1998). "The structure and mechanism of protein phosphatases: insights into catalysis and regulation." *Annu Rev Biophys Biomol Struct* **27**: 133-164.

Bataille, A. R., C. Jeronimo, P. E. Jacques, L. Laramee, M. E. Fortin, A. Forest, M. Bergeron, S. D. Hanes and F. Robert (2012). "A universal RNA polymerase II CTD cycle is orchestrated by complex interplays between kinase, phosphatase, and isomerase enzymes along genes." *Mol Cell* **45**(2): 158-170.

Baumli, S., G. Lolli, E. D. Lowe, S. Troiani, L. Rusconi, A. N. Bullock, J. E. Debreczeni, S. Knapp and L. N. Johnson (2008). "The structure of P-TEFb (CDK9/cyclin T1), its complex with flavopiridol and regulation by phosphorylation." *Embo J* **27**(13): 1907-1918.

Becker, R., B. Loll and A. Meinhart (2008). "Snapshots of the RNA processing factor SCAF8 bound to different phosphorylated forms of the carboxyl-terminal domain of RNA polymerase II." *J Biol Chem* **283**(33): 22659-22669.

Bedford, M. T., R. Reed and P. Leder (1998). "WW domain-mediated interactions reveal a spliceosome-associated protein that binds a third class of proline-rich motif: the proline glycine and methionine-rich motif." *Proc Natl Acad Sci U S A* **95**(18): 10602-10607.

Bedford, M. T., D. Sarbassova, J. Xu, P. Leder and M. B. Yaffe (2000). "A novel pro-Arg motif recognized by WW domains." *J Biol Chem* **275**(14): 10359-10369.

Brunger, A. T. (1992). "Free R value: a novel statistical quantity for assessing the accuracy of crystal structures." *Nature* **355**(6359): 472-475.

Brunger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, et al. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5**): 905-921.

Bucciantini, M., P. Chiarugi, P. Cirri, L. Taddei, M. Stefani, G. Raugei, P. Nordlund and G. Ramponi (1999). "The low Mr phosphotyrosine protein phosphatase behaves differently when phosphorylated at Tyr131 or Tyr132 by Src kinase." *FEBS Lett* **456**(1): 73-78.

Buratowski, S. (2003). "The CTD code." *Nat Struct Mol Biol* **10**(9): 679-680.

Buratowski, S. (2009). "Progression through the RNA polymerase II CTD cycle." *Mol Cell* **36**(4): 541-546.

Burley, S. K. and G. A. Petsko (1986). "Amino-aromatic interactions in proteins." *FEBS Lett* **203**(2): 139-143.

Carty, S. M., A. C. Goldstrohm, C. Sune, M. A. Garcia-Blanco and A. L. Greenleaf (2000). "Protein-interaction modules that organize nuclear function: FF domains of CA150 bind the phosphoCTD of RNA polymerase II." *Proc Natl Acad Sci U S A* **97**(16): 9015-9020.

CCP4 (1994). "Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography." *Acta Crystallogr.* **D50**: 760-763.

195

CCP4 (1994). "Collaborative Computational Project, Number 4. The CCP4 Suite: Programs for Protein Crystallography." *Acta Cryst.* **D50**: 760-763.

Chambers, R. S. and M. E. Dahmus (1994). "Purification and characterization of a phosphatase from HeLa cells which dephosphorylates the C-terminal domain of RNA polymerase II." *J Biol Chem* **269**(42): 26243-26248.

Chambers, R. S. and C. M. Kane (1996). "Purification and characterization of an RNA polymerase II phosphatase from yeast." *J Biol Chem* **271**(40): 24498-24504.

Chapman, R. D., M. Heidemann, T. K. Albert, R. Mailhammer, A. Flatley, M. Meisterernst, E. Kremmer and D. Eick (2007). "Transcribing RNA polymerase II is phosphorylated at CTD residue serine-7." *Science* **318**(5857): 1780-1782.

Chen, H. I., A. Einbond, S. J. Kwak, H. Linn, E. Koepf, S. Peterson, J. W. Kelly and M. Sudol (1997). "Characterization of the WW domain of human yes-associated protein and its polyproline-containing ligands." *J Biol Chem* **272**(27): 17070-17077.

Chiarugi, P., P. Cirri, F. Marra, G. Raugei, G. Camici, G. Manao and G. Ramponi (1997). "LMW-PTP is a negative regulator of insulin-mediated mitotic and metabolic signalling." *Biochem Biophys Res Commun* **238**(2): 676-682.

Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt and S. Buratowski (2001). "Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain." *Genes Dev* **15**(24): 3319-3329.

Cho, E. J., T. Takagi, C. R. Moore and S. Buratowski (1997). "mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain." *Genes Dev* **11**(24): 3319-3326.

Cho, H., T. K. Kim, H. Mancebo, W. S. Lane, O. Flores and D. Reinberg (1999). "A protein phosphatase functions to recycle RNA polymerase II." *Genes Dev* **13**(12): 1540-1552.

Cho, H., S. Ramaswamy and B. V. Plapp (1997). "Flexibility of liver alcohol dehydrogenase in stereoselective binding of 3-butylthiolane 1-oxides." *Biochemistry* **36**(2): 382-389.

Cho, H., W. Wang, R. Kim, H. Yokota, S. Damo, S. H. Kim, D. Wemmer, S. Kustu and D. Yan (2001). "BeF(3)(-) acts as a phosphate analog in proteins phosphorylated on aspartate: structure of a BeF(3)(-) complex with phosphoserine phosphatase." *Proc Natl Acad Sci U S A* **98**(15): 8525-8530.

Cohen, P. T. (1997). "Novel protein serine/threonine phosphatases: variety is the spice of life." *Trends Biochem Sci* **22**(7): 245-251.

Collet, J. F., V. Stroobant, M. Pirard, G. Delpierre and E. Van Schaftingen (1998). "A new class of phosphotransferases phosphorylated on an aspartate residue in an amino-terminal DXDX(T/V) motif." *J Biol Chem* **273**(23): 14107-14112.

Corden, J. L. (1990). "Tails of RNA polymerase II." *Trends Biochem Sci* **15**(10): 383-387.

Dahmus, M. E. (1996). "Reversible phosphorylation of the C-terminal domain of RNA polymerase II." *J Biol Chem* **271**(32): 19009-19012.

Denu, J. M. and J. E. Dixon (1995). "A catalytic mechanism for the dual-specific phosphatases." *Proc Natl Acad Sci U S A* **92**(13): 5910-5914.

Deshpande, R. A. and T. E. Wilson (2004). "Identification of DNA 3'-phosphatase active site residues and their differential role in DNA binding, Mg2+ coordination, and catalysis." *Biochemistry* **43**(26): 8579-8589.

Diaz, A. R., S. Stephenson, J. M. Green, V. M. Levdikov, A. J. Wilkinson and M. Perego (2008). "Functional role for a conserved aspartate in the Spo0E signature motif involved in the dephosphorylation of the Bacillus subtilis sporulation regulator Spo0A." *J Biol Chem* **283**(5): 2962-2972.

Dougherty, D. A. (1996). "Cation-pi interactions in chemistry and biology: a new view of benzene, Phe, Tyr, and Trp." *Science* **271**(5246): 163-168.

Egloff, S. and S. Murphy (2008). "Cracking the RNA polymerase II CTD code." *Trends Genet* **24**(6): 280-288.

Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick and S. Murphy (2007). "Serine-7 of the RNA Polymerase II CTD Is Specifically Required for snRNA Gene Expression." *Science* **318**(5857): 1777-1779.

Egloff, S., D. O'Reilly, R. D. Chapman, A. Taylor, K. Tanzhaus, L. Pitts, D. Eick and S. Murphy (2007). "Serine-7 of the RNA polymerase II CTD is specifically required for snRNA gene expression." *Science* **318**(5857): 1777-1779.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." *Acta Crystallogr D Biol Crystallogr* **60**(Pt 12 Pt 1): 2126-2132.

Ermekova, K. S., N. Zambrano, H. Linn, G. Minopoli, F. Gertler, T. Russo and M. Sudol (1997). "The WW domain of neural protein FE65 interacts with proline-rich motifs in Mena, the mammalian homolog of Drosophila enabled." *J Biol Chem* **272**(52): 32869-32877.

Etzkorn, F. A. (2006). "Pin1 flips Alzheimer's switch." *ACS Chem Biol* **1**(4): 214-216.

Fabrega, C., V. Shen, S. Shuman and C. D. Lima (2003). "Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II." *Mol Cell* **11**(6): 1549-1561.

Fouillen, L., W. Abdulrahman, D. Moras, A. Van Dorsselaer, A. Poterszman and S. Sanglier-Cianferani (2010). "Analysis of recombinant phosphoprotein complexes with complementary mass spectrometry approaches." *Anal Biochem* **407**(1): 34-43.

Fuda, N. J., M. B. Ardehali and J. T. Lis (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." *Nature* **461**(7261): 186-192.

Fuda, N. J., M. B. Ardehali and J. T. Lis (2009). "Defining mechanisms that regulate RNA polymerase II transcription in vivo." *Nature* **461**(7261): 186-192.

198

Gallivan, J. P. and D. A. Dougherty (1999). "Cation-pi interactions in structural biology." *Proc Natl Acad Sci U S A* **96**(17): 9459-9464.

Garber, M. E., T. P. Mayall, E. M. Suess, J. Meisenhelder, N. E. Thompson and K. A. Jones (2000). "CDK9 autophosphorylation regulates high-affinity binding of the human immunodeficiency virus type 1 tat-P-TEFb complex to TAR RNA." *Mol Cell Biol* **20**(18): 6958-6969.

Gasch, A., S. Wiesner, P. Martin-Malpartida, X. Ramirez-Espain, L. Ruiz and M. J. Macias (2006). "The structure of Prp40 FF1 domain and its interaction with the crn-TPR1 motif of Clf1 gives a new insight into the binding mode of FF domains." *J Biol Chem* **281**(1): 356-364.

Gerber, M. and A. Shilatifard (2003). "Transcriptional elongation by RNA polymerase II and histone methylation." *J Biol Chem* **278**(29): 26303-26306.

Ghosh, A., S. Shuman and C. D. Lima (2008). "The structure of Fcp1, an essential RNA polymerase II CTD phosphatase." *Mol Cell* **32**(4): 478-490.

Glover-Cutter, K., S. Larochelle, B. Erickson, C. Zhang, K. Shokat, R. P. Fisher and D. L. Bentley (2009). "TFIIH-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II." *Mol Cell Biol* **29**(20): 5455-5464.

Gomes, N. P., G. Bjerke, B. Llorente, S. A. Szostek, B. M. Emerson and J. M. Espinosa (2006). "Gene-specific requirement for P-TEFb activity and RNA polymerase II phosphorylation within the p53 transcriptional program." *Genes Dev* **20**(5): 601-612.

Gudipati, R. K., T. Villa, J. Boulay and D. Libri (2008). "Phosphorylation of the RNA polymerase II C-terminal domain dictates transcription termination choice." *Nat Struct Mol Biol* **15**(8): 786-794.

Hampsey, M. and D. Reinberg (2003). "Tails of intrigue: phosphorylation of RNA polymerase II mediates histone methylation." *Cell* **113**(4): 429-432.

Hausmann, S., H. Erdjument-Bromage and S. Shuman (2004). "Schizosaccharomyces pombe carboxyl-terminal domain (CTD) phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping." *J Biol Chem* **279**(12): 10892-10900.

Hausmann, S., H. Koiwa, S. Krishnamurthy, M. Hampsey and S. Shuman (2005). "Different strategies for carboxyl-terminal domain (CTD) recognition by serine 5-specific CTD phosphatases." *J Biol Chem* **280**(45): 37681-37688.

Hausmann, S., B. Schwer and S. Shuman (2004). "An encephalitozoon cuniculi ortholog of the RNA polymerase II carboxyl-terminal domain (CTD) serine phosphatase Fcp1." *Biochemistry* **43**(22): 7111-7120.

Hausmann, S. and S. Shuman (2003). "Defining the active site of Schizosaccharomyces pombe C-terminal domain phosphatase Fcp1." *J Biol Chem* **278**(16): 13627-13632.

He, X., A. U. Khan, H. Cheng, D. L. Pappas, Jr., M. Hampsey and C. L. Moore (2003). "Functional interactions between the transcription and mRNA 3' end processing machineries mediated by Ssu72 and Sub1." *Genes Dev* **17**(8): 1030-1042.

Hengartner, C. J., V. E. Myer, S. M. Liao, C. J. Wilson, S. S. Koh and R. A. Young (1998). "Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases." *Mol Cell* **2**(1): 43-53.

Hintermair, C., M. Heidemann, F. Koch, N. Descostes, M. Gut, et al. (2012). "Threonine-4 of mammalian RNA polymerase II CTD is targeted by Polo-like kinase 3 and required for transcriptional elongation." *Embo J* **31**(12): 2784-2797.

Hollingworth, D., C. G. Noble, I. A. Taylor and A. Ramos (2006). "RNA polymerase II CTD phosphopeptides compete with RNA for the interaction with Pcf11." *RNA* **12**(4): 555-560.

Holm, L. and P. Rosenstrom (2010). "Dali server: conservation mapping in 3D." *Nucl. Acids Res.* **38**(suppl_2): W545-549.

Hsin, J. P., A. Sheth and J. L. Manley (2011). "RNAP II CTD phosphorylated on threonine-4 is required for histone mRNA 3' end processing." *Science* **334**(6056): 683-686.

Huang, K., K. D. Johnson, A. G. Petcherski, T. Vandergon, E. A. Mosser, N. G. Copeland, N. A. Jenkins, J. Kimble and E. H. Bresnick (2000). "A HECT domain ubiquitin ligase closely related to the mammalian protein WWP1 is essential for Caenorhabditis elegans embryogenesis." *Gene* **252**(1-2): 137-145.

Jager, M., Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele and J. W. Kelly (2006). "Structure-function-folding relationship in a WW domain." *Proc Natl Acad Sci U S A* **103**(28): 10648-10653.

Jez, J. M., J. L. Ferrer, M. E. Bowman, R. A. Dixon and J. P. Noel (2000). "Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase." *Biochemistry* **39**(5): 890-902.

Jones, T. A., J. Y. Zou, S. W. Cowan and M. Kjeldgaard (1991). "Improved methods for building protein models in electron density maps and the location of errors in these models." *Acta Crystallogr A* **47 (Pt 2)**: 110-119.

Jung, S. K., D. G. Jeong, S. J. Chung, J. H. Kim, B. C. Park, N. K. Tonks, S. E. Ryu and S. J. Kim (2010). "Crystal structure of ED-Eya2: insight into dual roles as a protein tyrosine phosphatase and a transcription factor." *Faseb J* **24**(2): 560-569.

Kamenski, T., S. Heilmeier, A. Meinhart and P. Cramer (2004). "Structure and mechanism of RNA polymerase II CTD phosphatases." *Mol Cell* **15**(3): 399-407.

Kim, M., L. Vasiljeva, O. J. Rando, A. Zhelkovsky, C. Moore and S. Buratowski (2006). "Distinct pathways for snoRNA and mRNA termination." *Mol Cell* **24**(5): 723-734.

Kim, Y., M. S. Gentry, T. E. Harris, S. E. Wiley, J. C. Lawrence, Jr. and J. E. Dixon (2007). "A conserved phosphatase cascade that regulates nuclear membrane biogenesis." *Proc Natl Acad Sci U S A* **104**(16): 6596-6601.

Kimura, M., H. Suzuki and A. Ishihama (2002). "Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in Schizosaccharomyces pombe involves direct interaction between Fcp1 and the Rpb4 subunit of pol II." *Mol Cell Biol* **22**(5): 1577-1588.

Klotz, U. (2000). "Pharmacokinetic considerations in the eradication of Helicobacter pylori." *Clin Pharmacokinet* **38**(3): 243-270.

Kobor, M. S., J. Archambault, W. Lester, F. C. Holstege, O. Gileadi, et al. (1999). "An unusual eukaryotic protein phosphatase required for transcription by RNA polymerase II and CTD dephosphorylation in S. cerevisiae." *Mol Cell* **4**(1): 55-62.

Komarnitsky, P., E. J. Cho and S. Buratowski (2000). "Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription." *Genes Dev* **14**(19): 2452-2460.

Komuro, A., M. Saeki and S. Kato (1999). "Npw38, a novel nuclear protein possessing a WW domain capable of activating basal transcription." *Nucleic Acids Res* **27**(9): 1957-1965.

Kornberg, R. D. and Y. Lorch (1999). "Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome." *Cell* **98**(3): 285-294.

Kouzarides, T. (2007). "Chromatin modifications and their function." *Cell* **128**(4): 693-705.

Kowalski, J. A., K. Liu and J. W. Kelly (2002). "NMR solution structure of the isolated Apo Pin1 WW domain: comparison to the x-ray crystal structures of Pin1." *Biopolymers* **63**(2): 111-121.

Krishnamurthy, S., M. A. Ghazy, C. Moore and M. Hampsey (2009). "Functional Interaction of the Ess1 Prolyl Isomerase with Components of the RNA Polymerase II Initiation and Termination Machineries." *Mol. Cell. Biol.* **29**(11): 2925-2934.

Krishnamurthy, S., X. He, M. Reyes-Reyes, C. Moore and M. Hampsey (2004). "Ssu72 Is an RNA polymerase II CTD phosphatase." *Mol Cell* **14**(3): 387-394.

Krogan, N. J., M. Kim, A. Tong, A. Golshani, G. Cagney, et al. (2003). "Methylation of histone H3 by Set2 in Saccharomyces cerevisiae is linked to transcriptional elongation by RNA polymerase II." *Mol Cell Biol* **23**(12): 4207-4218.

Lahiri, S. D., G. Zhang, D. Dunaway-Mariano and K. N. Allen (2002). "Caught in the act: the structure of phosphorylated beta-phosphoglucomutase from Lactococcus lactis." *Biochemistry* **41**(26): 8351-8359.

Lahiri, S. D., G. Zhang, D. Dunaway-Mariano and K. N. Allen (2003). "The pentacovalent phosphorus intermediate of a phosphoryl transfer reaction." *Science* **299**(5615): 2067-2071.

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

Lawrence, M. C. and P. M. Colman (1993). "Shape complementarity at protein/protein interfaces." *J Mol Biol* **234**(4): 946-950.

Lee, B. and F. M. Richards (1971). "The interpretation of protein structures: estimation of static accessibility." *J Mol Biol* **55**(3): 379-400.

Li, M., H. P. Phatnani, Z. Guan, H. Sage, A. L. Greenleaf and P. Zhou (2005). "Solution structure of the Set2-Rpb1 interacting domain of human Set2 and its interaction with the hyperphosphorylated C-terminal domain of Rpb1." *Proc Natl Acad Sci U S A* **102**(49): 17636-17641.

Lindqvist, Y., G. Schneider and P. Vihko (1994). "Crystal structures of rat acid phosphatase complexed with the transition-state analogs vanadate and molybdate. Implications for the reaction mechanism." *Eur J Biochem* **221**(1): 139-142.

Lolli, G., E. D. Lowe, N. R. Brown and L. N. Johnson (2004). "The crystal structure of human CDK7 and its protein recognition properties." *Structure* **12**(11): 2067-2079.

Lu, H., L. Zawel, L. Fisher, J. M. Egly and D. Reinberg (1992). "Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II." *Nature* **358**(6388): 641-645.

Lu, K. P. (2004). "Pinning down cell signaling, cancer and Alzheimer's disease." *Trends Biochem Sci* **29**(4): 200-209.

Lu, K. P., S. D. Hanes and T. Hunter (1996). "A human peptidyl-prolyl isomerase essential for regulation of mitosis." *Nature* **380**(6574): 544-547.

Lu, P. J., G. Wulf, X. Z. Zhou, P. Davies and K. P. Lu (1999). "The prolyl isomerase Pin1 restores the function of Alzheimer-associated phosphorylated tau protein." *Nature* **399**(6738): 784-788.

Lu, P. J., X. Z. Zhou, Y. C. Liou, J. P. Noel and K. P. Lu (2002). "Critical role of WW domain phosphorylation in regulating phosphoserine binding activity and Pin1 function." *J Biol Chem* **277**(4): 2381-2384.

Lu, P. J., X. Z. Zhou, M. Shen and K. P. Lu (1999). "Function of WW domains as phosphoserine- or phosphothreonine-binding modules." *Science* **283**(5406): 1325-1328.

Ma, J. C. and D. A. Dougherty (1997). "The Cation-π Interaction." *Chem Rev* **97**(5): 1303-1324.

Macias, M. J., M. Hyvonen, E. Baraldi, J. Schultz, M. Sudol, M. Saraste and H. Oschkinat (1996). "Structure of the WW domain of a kinase-associated protein complexed with a proline-rich peptide." *Nature* **382**(6592): 646-649.

Majello, B. and G. Napolitano (2001). "Control of RNA polymerase II activity by dedicated CTD kinases and phosphatases." *Front Biosci* **6**: D1358-1368.

Mayer, A., M. Heidemann, M. Lidschreiber, A. Schreieck, M. Sun, C. Hintermair, E. Kremmer, D. Eick and P. Cramer (2012). "CTD tyrosine phosphorylation impairs termination factor recruitment to RNA polymerase II." *Science* **336**(6089): 1723-1725.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." *J Appl Crystallogr* **40**(Pt 4): 658-674.

McIntosh, D. B., J. D. Clausen, D. G. Woolley, D. H. MacLennan, B. Vilsen and J. P. Andersen (2004). "Roles of conserved P domain residues and Mg2+ in ATP binding in the ground and Ca2+-activated states of sarcoplasmic reticulum Ca2+-ATPase." *J Biol Chem* **279**(31): 32515-32523.

Meinhart, A. and P. Cramer (2004). "Recognition of RNA polymerase II carboxy-terminal domain by 3'-RNA-processing factors." *Nature* **430**(6996): 223-226.

Meinhart, A., T. Kamenski, S. Hoeppner, S. Baumli and P. Cramer (2005). "A structural perspective of CTD function." *Genes Dev* **19**(12): 1401-1415.

Meinhart, A., T. Silberzahn and P. Cramer (2003). "The mRNA transcription/processing factor Ssu72 is a potential tyrosine phosphatase." *J Biol Chem* **278**(18): 15917-15921.

Morais, M. C., W. Zhang, A. S. Baker, G. Zhang, D. Dunaway-Mariano and K. N. Allen (2000). "The crystal structure of bacillus cereus phosphonoacetaldehyde hydrolase: insight into catalysis of phosphorus bond cleavage and catalytic diversification within the HAD enzyme superfamily." *Biochemistry* **39**(34): 10385-10396.

Morris, D. P. and A. L. Greenleaf (2000). "The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II." *J Biol Chem* **275**(51): 39935-39943.

Morris, D. P., H. P. Phatnani and A. L. Greenleaf (1999). "Phospho-carboxyl-terminal domain binding and the role of a prolyl isomerase in pre-mRNA 3'-End formation." *J Biol Chem* **274**(44): 31583-31587.

Mosley, A. L., S. G. Pattenden, M. Carey, S. Venkatesh, J. M. Gilmore, L. Florens, J. L. Workman and M. P. Washburn (2009). "Rtr1 is a CTD phosphatase that regulates RNA polymerase II during the transition from serine 5 to serine 2 phosphorylation." *Mol Cell* **34**(2): 168-178.

Murphy, J. M., D. F. Hansen, S. Wiesner, D. R. Muhandiram, M. Borg, et al. (2009). "Structural studies of FF domains of the transcription factor CA150 provide insights into the organization of FF domain tandem arrays." *J Mol Biol* **393**(2): 409-424.

Myers, J. K., D. P. Morris, A. L. Greenleaf and T. G. Oas (2001). "Phosphorylation of RNA polymerase II CTD fragments results in tight binding to the WW domain from the yeast prolyl isomerase Ess1." *Biochemistry* **40**(29): 8479-8486.

Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." *Acta Crystallogr.* **A50**: 157-163.

Ng, H. H., F. Robert, R. A. Young and K. Struhl (2003). "Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity." *Mol Cell* **11**(3): 709-719.

Niesen, F. H., H. Berglund and M. Vedadi (2007). "The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability." *Nat Protoc* **2**(9): 2212-2221.

Noble, C. G., D. Hollingworth, S. R. Martin, V. Ennis-Adeniran, S. J. Smerdon, G. Kelly, I. A. Taylor and A. Ramos (2005). "Key features of the interaction between Pcf11 CID and RNA polymerase II CTD." *Nat Struct Mol Biol* **12**(2): 144-151.

Otwinowski, Z. and W. Minor (1997). "HKL: Processing of X-ray diffraction data collected in oscillation mode." *Methods Enzymol.* **276**: 307-326.

Palancade, B. and O. Bensaude (2003). "Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation." *Eur J Biochem* **270**(19): 3859-3870.

Palancade, B., N. F. Marshall, A. Tremeau-Bravard, O. Bensaude, M. E. Dahmus and M. F. Dubois (2004). "Dephosphorylation of RNA polymerase II by CTD-phosphatase FCP1 is inhibited by phospho-CTD associating proteins." *J Mol Biol* **335**(2): 415-424.

Pannifer, A. D., A. J. Flint, N. K. Tonks and D. Barford (1998). "Visualization of the cysteinyl-phosphate intermediate of a protein-tyrosine phosphatase by x-ray crystallography." *J Biol Chem* **273**(17): 10454-10462.

Peters, G. H., T. M. Frimurer and O. H. Olsen (1998). "Electrostatic evaluation of the signature motif (H/V)CX5R(S/T) in protein-tyrosine phosphatases." *Biochemistry* **37**(16): 5383-5393.

Phatnani, H. P. and A. L. Greenleaf (2006). "Phosphorylation and functions of the RNA polymerase II CTD." *Genes Dev* **20**(21): 2922-2936.

Ramponi, G. and M. Stefani (1997). "Structure and function of the low Mr phosphotyrosine protein phosphatases." *Biochim Biophys Acta* **1341**(2): 137-156.

Ranganathan, R., K. P. Lu, T. Hunter and J. P. Noel (1997). "Structural and functional analysis of the mitotic rotamase Pin1 suggests substrate recognition is phosphorylation dependent." *Cell* **89**(6): 875-886.

Rangarajan, E. S., A. Proteau, J. Wagner, M. N. Hung, A. Matte and M. Cygler (2006). "Structural Snapshots of Escherichia coli Histidinol Phosphate Phosphatase along the Reaction Pathway." *J Biol Chem* **281**(49): 37930-37941.

Reddy, G. M., K. Mukkanti, B. V. Bhaskar and P. P. Reddy (2009). "Synthesis of Metabolites and Related Substances of Rabeprazole, an Anti-Ulcerative Drug." *Synthetic Communications* **39**: 278-290.

Rigacci, S., E. Rovida, S. Bagnoli, P. Dello Sbarba and A. Berti (1999). "Low Mr phosphotyrosine protein phosphatase activity on fibroblast growth factor receptor is not associated with enzyme translocation." *FEBS Lett.* **459**(2): 191-194.

Roche, V. F. (2006). "The chemically elegant proton pump inhibitors." *Am J Pharm Educ* **70**(5): 101.

Rogers, C., Z. Guo and J. W. Stiller (2010). "Connecting mutations of the RNA polymerase II C-terminal domain to complex phenotypic changes using combined gene expression and network analyses." *PLoS One* **5**(6): e11386.

Rondon, A. G., H. E. Mischo and N. J. Proudfoot (2008). "Terminating transcription in yeast: whether to be a 'nerd' or a 'rat'." *Nat Struct Mol Biol* **15**(8): 775-776.

Satow, R., T. C. Chan and M. Asashima (2002). "Molecular cloning and characterization of dullard: a novel gene required for neural development." *Biochem Biophys Res Commun* **295**(1): 85-91.

Scrutton, N. S. and A. R. Raine (1996). "Cation-pi bonding and amino-aromatic interactions in the biomolecular recognition of substituted ammonium ligands." *Biochem J* **319 (Pt 1)**: 1-8.

Shi, Y. (2009). "Serine/threonine phosphatases: mechanism through structure." *Cell* **139**(3): 468-484.

Shi, Y. (2009). "Serine/Threonine Phosphatases: Mechanism through Structure." *Cell* **139**(3): 468-484.

Shim, E. Y., A. K. Walker, Y. Shi and T. K. Blackwell (2002). "CDK-9/cyclin T (P-TEFb) is required in two postinitiation pathways for transcription in the C. elegans embryo." *Genes Dev* **16**(16): 2135-2146.

Stein, E., A. A. Lane, D. P. Cerretti, H. O. Schoecklmann, A. D. Schroff, R. L. Van Etten and T. O. Daniel (1998). "Eph receptors discriminate specific ligand oligomers to determine alternative signaling complexes, attachment, and assembly responses." *Genes Dev* **12**(5): 667-678.

Strahl, B. D. and C. D. Allis (2000). "The language of covalent histone modifications." *Nature* **403**(6765): 41-45.

Sudol, M. and T. Hunter (2000). "NeW wrinkles for an old domain." *Cell* **103**(7): 1001-1004.

Sun, Z. W. and M. Hampsey (1996). "Synthetic enhancement of a TFIIB defect by a mutation in SSU72, an essential yeast gene encoding a novel protein that affects transcription start site selection in vivo." *Mol Cell Biol* **16**(4): 1557-1566.

Tonks, N. K. (2006). "Protein tyrosine phosphatases: from genes, to function, to disease." *Nat Rev Mol Cell Biol* **7**(11): 833-846.

Tonks, N. K. (2006). "Protein tyrosine phosphatases: from genes, to function, to disease." *Nat Rev Mol Cell Biol* **7**(11): 833-846.

Toyoshima, C., H. Nomura and T. Tsuda (2004). "Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues." *Nature* **432**(7015): 361-368.

Toyoshima, C., Y. Norimatsu, S. Iwasawa, T. Tsuda and H. Ogawa (2007). "How processing of aspartylphosphate is coupled to lumenal gating of the ion pathway in the calcium pump." *Proc Natl Acad Sci U S A* **104**(50): 19831-19836.

Trigon, S., H. Serizawa, J. W. Conaway, R. C. Conaway, S. P. Jackson and M. Morange (1998). "Characterization of the residues phosphorylated in vitro by different C-terminal domain kinases." *J Biol Chem* **273**(12): 6769-6775.

Vagin, A. A., R. A. Steiner, A. A. Lebedev, L. Potterton, S. McNicholas, F. Long and G. N. Murshudov (2004). "REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use." *Acta Cryst.* **D60**(Pt 12 Pt 1): 2184-2195.

Vasiljeva, L., M. Kim, H. Mutschler, S. Buratowski and A. Meinhart (2008). "The Nrd1-Nab3-Sen1 termination complex interacts with the Ser5-phosphorylated RNA polymerase II C-terminal domain." *Nat Struct Mol Biol* **15**(8): 795-804.

Verdecia, M. A., M. E. Bowman, K. P. Lu, T. Hunter and J. P. Noel (2000). "Structural basis for phosphoserine-proline recognition by group IV WW domains." *Nat Struct Biol* **7**(8): 639-643.

Visvanathan, J., S. Lee, B. Lee, J. W. Lee and S. K. Lee (2007). "The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development." *Genes Dev* **21**(7): 744-749.

Vojnic, E., B. Simon, B. D. Strahl, M. Sattler and P. Cramer (2006). "Structure and carboxyl-terminal domain (CTD) binding of the Set2 SRI domain that couples histone H3 Lys36 methylation to transcription." *J Biol Chem* **281**(1): 13-15.

Wang, W., H. S. Cho, R. Kim, J. Jancarik, H. Yokota, H. H. Nguyen, I. V. Grigoriev, D. E. Wemmer and S. H. Kim (2002). "Structural characterization of the reaction pathway in phosphoserine phosphatase: crystallographic "snapshots" of intermediate states." *J Mol Biol* **319**(2): 421-431.

Wang, W., R. Kim, J. Jancarik, H. Yokota and S. H. Kim (2001). "Crystal structure of phosphoserine phosphatase from Methanococcus jannaschii, a hyperthermophile, at 1.8 A resolution." *Structure* **9**(1): 65-71.

West, M. L. and J. L. Corden (1995). "Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations." *Genetics* **140**(4): 1223-1233.

Wu, X., A. Chang, M. Sudol and S. D. Hanes (2001). "Genetic interactions between the ESS1 prolyl-isomerase and the RSP5 ubiquitin ligase reveal opposing effects on RNA polymerase II function." *Curr Genet* **40**(4): 234-242.

Wu, X., C. B. Wilcox, G. Devasahayam, R. L. Hackett, M. Arevalo-Rodriguez, M. E. Cardenas, J. Heitman and S. D. Hanes (2000). "The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery." *Embo J* **19**(14): 3727-3738.

Xiang, K., T. Nagaike, S. Xiang, T. Kilic, M. M. Beh, J. L. Manley and L. Tong (2010). "Crystal structure of the human symplekin-Ssu72-CTD phosphopeptide complex." *Nature* **467**(7316): 729-733.

Xiao, T., H. Hall, K. O. Kizer, Y. Shibata, M. C. Hall, C. H. Borchers and B. D. Strahl (2003). "Phosphorylation of RNA polymerase II CTD regulates H3 methylation in yeast." *Genes Dev* **17**(5): 654-663.

Xu, Y. X., Y. Hirose, X. Z. Zhou, K. P. Lu and J. L. Manley (2003). "Pin1 modulates the structure and function of human RNA polymerase II." *Genes Dev* **17**(22): 2765-2776.

Yaffe, M. B., M. Schutkowski, M. Shen, X. Z. Zhou, P. T. Stukenberg, et al. (1997). "Sequence-specific and phosphorylation-dependent proline isomerization: a potential mitotic regulatory mechanism." *Science* **278**(5345): 1957-1960.

Yeo, M., S.-K. Lee, B. Lee, E. C. Ruiz, S. L. Pfaff and G. N. Gill (2005). "Small CTD Phosphatases Function in Silencing Neuronal Gene Expression." *Science* **307**(5709): 596-600.

Yeo, M., S. K. Lee, B. Lee, E. C. Ruiz, S. L. Pfaff and G. N. Gill (2005). "Small CTD phosphatases function in silencing neuronal gene expression." *Science* **307**(5709): 596-600.

Yeo, M., P. S. Lin, M. E. Dahmus and G. N. Gill (2003). "A novel RNA polymerase II C-terminal domain phosphatase that preferentially dephosphorylates serine 5." *J Biol Chem* **278**(28): 26078-26085.

Yu, X., J. P. Sun, Y. He, X. Guo, S. Liu, B. Zhou, A. Hudmon and Z. Y. Zhang (2007). "Structure, inhibitor, and regulatory mechanism of Lyp, a lymphoid-specific tyrosine phosphatase implicated in autoimmune diseases." *Proc Natl Acad Sci U S A* **104**(50): 19767-19772.

Zhang, G., A. S. Mazurkie, D. Dunaway-Mariano and K. N. Allen (2002). "Kinetic evidence for a substrate-induced fit in phosphonoacetaldehyde hydrolase catalysis." *Biochemistry* **41**(45): 13370-13377.

Zhang, M., G. Gill and Y. Zhang (2010). "Bio-molecular Architects: A Scaffold Provided by the C-terminal Domain of Eukaryotic RNA Polymerase II." *Nano Reviews*.

Zhang, M., G. N. Gill and Y. Zhang (2010). "Bio-molecular architects: a scaffold provided by the C-terminal domain of eukaryotic RNA polymerase II " *Nano Reviews* **1: 5502 - DOI: 10.3402/nano.v1i0.5502**.

Zhang, M., J. Liu, Y. Kim, J. E. Dixon, S. L. Pfaff, G. N. Gill, J. P. Noel and Y. Zhang (2010). "Structural and functional analysis of the phosphoryl transfer reaction mediated by the human small C-terminal domain phosphatase, Scp1." *Protein Sci* **19**(5): 974 - 986.

Zhang, M., J. Liu, Y. Kim, J. E. Dixon, S. L. Pfaff, G. N. Gill, J. P. Noel and Y. Zhang (2010). "Structural and functional analysis of the phosphoryl transfer reaction mediated by the human small C-terminal domain phosphatase, Scp1." *Protein Sci* **19**(5): 974-986.

Zhang, M., M. Zhou, R. L. Van Etten and C. V. Stauffacher (1997). "Crystal structure of bovine low molecular weight phosphotyrosyl phosphatase complexed with the transition state analog vanadate." *Biochemistry* **36**(1): 15-23.

Zhang, S., L. Chen, Y. Luo, A. Gunawan, D. S. Lawrence and Z. Y. Zhang (2009). "Acquisition of a potent and selective TC-PTP inhibitor via a stepwise fluorophore-tagged combinatorial synthesis and screening strategy." *J Am Chem Soc* **131**(36): 13072-13079.

Zhang, Y., S. Daum, D. Wildemann, X. Z. Zhou, M. A. Verdecia, et al. (2007). "Structural basis for high-affinity peptide inhibition of human Pin1." *ACS Chem Biol* **2**(5): 320-328.

Zhang, Y., Y. Kim, N. Genoud, J. Gao, J. W. Kelly, S. L. Pfaff, G. N. Gill, J. E. Dixon and J. P. Noel (2006). "Determinants for dephosphorylation of the RNA polymerase II C-terminal domain by Scp1." *Mol Cell* **24**(5): 759-770.

Zhang, Z., J. Fu and D. S. Gilmour (2005). "CTD-dependent dismantling of the RNA polymerase II elongation complex by the pre-mRNA 3'-end processing factor, Pcf11." *Genes Dev* **19**(13): 1572-1580.

Zhang, Z. Y. (2002). "Protein tyrosine phosphatases: structure and function, substrate specificity, and inhibitor development." *Annu Rev Pharmacol Toxicol* **42**: 209-234.

Zhang, Z. Y. (2003). "Mechanistic studies on protein tyrosine phosphatases." *Prog Nucleic Acid Res Mol Biol* **73**: 171-220.

Zhang, Z. Y., J. P. Davis and R. L. Van Etten (1992). "Covalent modification and active site-directed inactivation of a low molecular weight phosphotyrosyl protein phosphatase." *Biochemistry* **31**(6): 1701-1711.

Zhang, Z. Y., Y. Wang and J. E. Dixon (1994). "Dissecting the catalytic mechanism of protein-tyrosine phosphatases." *Proc Natl Acad Sci U S A* **91**(5): 1624-1627.

Zhou, M., M. A. Halanski, M. F. Radonovich, F. Kashanchi, J. Peng, D. H. Price and J. N. Brady (2000). "Tat modifies the activity of CDK9 to phosphorylate serine 5 of the RNA polymerase II carboxyl-terminal domain during human immunodeficiency virus type 1 transcription." *Mol Cell Biol* **20**(14): 5077-5086.

Vita

Mengmeng Zhang was born in Xiangfan (now Xiangyang), China in 1984. She attended No. 3 Railway Middle School of Xiangfan and No. 1 Railway Middle School of Xiangfan (now Zhiyuan Middle School) before entering Shanghai Jiao Tong University in 2002 where she met her husband, Xi Chen. She studied bioengineering at the university. In 2005, she joined the Bio-X Center of Shanghai Jiao Tong University, led by Dr. Lin He, to study pharmacogenetics. After getting her Bachelor of Engineering degree in Bioengineering, she started her graduate study in the Department of Biological Sciences in Clemson University in 2006, under the supervision of Dr. George Xianzhong Yu. She got her Master of Science degree in 2008 and went to the University of Texas at Austin to continue her study as a Ph.D. student under the supervision of Dr. Yan Zhang in the Department of Chemistry and Biochemistry.

Email address: mzhang59@utexas.edu

This dissertation was typed by the author.