

BIOCHEMICAL CHARACTERIZATION OF THE ACTIVITY AND SPECIFICITY OF  
ULP1 FAMILY MEMBERS

APPROVED BY SUPERVISORY COMMITTEE

---

Kim Orth, Ph.D.

---

Diana Tomchick, Ph.D.

---

David Corey, Ph.D.

---

George DeMartino, Ph.D.

## DEDICATION

I would like to acknowledge a number of people whose insight and support made the past five years of my life such an enjoyable and life-changing experience.

Kim Orth has been more than just a boss to me. She has been my mentor. Her dedication and passion has shown me how to become a successful scientist. She has helped me become a biochemist and I will never be able to thank her enough for that.

I also want to thank every member of the Orth Lab, past and present, for pushing me when I needed to be pushed and supporting me when I needed support. Sohini and I were the first students to join Kim's lab and I have always viewed her dedication as an inspiration. Jen, Dara, and Mel made smile when I came into lab and we have shared many stories and baking competitions. I would like to thank Yong as my confidant in the lab. She has given me advice for anything and everything.

I would also like to thank the members of my Graduate Committee for their guidance, advice, and support. I would especially like to thank my chair, Diana Tomchick. She has taught me so much and was there during every step of my graduate career.

Finally, I would like to express my appreciation for my family. My sisters, Risa and Becky, have always encouraged me to achieve my dreams. My parents, Marilyn and Louis Chosed, have supported my dream of becoming a scientist since those days of middle school science fair projects. I would not have accomplished anything without their love and support. I also would have never made it through graduate school without that love, support, patience, and editing skills of my husband Damien.

BIOCHEMICAL CHARACTERIZATION OF THE ACTIVITY AND SPECIFICITY OF  
ULP1 FAMILY MEMBERS

by

RENEE JOANNA CHOSÉD

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

May, 2006

Copyright

by

RENEE JOANNA CHOSÉD, 2006

All Rights Reserved

BIOCHEMICAL CHARACTERIZATION OF THE ACTIVITY AND SPECIFICITY OF  
ULP1 FAMILY MEMBERS

Publication No. \_\_\_\_\_

RENEE JOANNA CHOSÉD

The University of Texas Southwestern Medical Center at Dallas, 2006

Supervising Professor: KIM ORTH, Ph.D.

The reversible posttranslational modification, SUMO, modulates the activity of a diverse set of target proteins. Conjugation machinery charges the processed-SUMO so that it can be linked via an isopeptide bond to a target protein. The removal of SUMO moieties from conjugated proteins by isopeptidases regenerates pools of processed-SUMOs and unmodified target proteins.

Yeast Ulp1 is the founding member of a growing family of these isopeptidases, or deSUMOylating enzymes. Another Ulp1, XopD from *Xanthomonas campestris* pv. *vesicatoria*, is expressed as a secreted virulence factor by this plant bacterial pathogen. *In*

*vitro* peptidase assays were performed to characterize the activity of XopD and showed that XopD can process plant SUMOs, but is unable to process mammalian SUMOs. Enzymatic studies further demonstrated that XopD functions as a cysteine protease. Isopeptide bond cleavage was demonstrated by the ability of XopD to cleave the plant SUMO moiety from *in vitro* SUMOylated RanGAP. These studies showed that XopD has a rigid SUMO substrate specificity in contrast to yeast Ulp1, which is a more promiscuous protease in its choice of SUMO substrate.

To further investigate the SUMO substrate specificity of XopD the crystal structures of the catalytic core of XopD and XopD C470A were determined. Although the catalytic core residues align between the XopD and Ulp1 structures, there are apparent structural differences that may account for the differences in choice of SUMO substrate. Mutational studies on tomato SUMO and yeast Smt3 identified some key residues in the enzyme-substrate recognition, but ultimately showed that there are many regions in the SUMO proteins that are accounting for the enzyme's ability to recognize certain substrates.

Experiments with the *Arabidopsis thaliana* family of SUMOs and ULP1s identified varying specificity of the AtULP1s. The evolutionarily conserved SUMO conjugating proteins, E1 and E2, recognize a diverse set of AtSUMO proteins using them to modify protein substrates. By contrast, the deSUMOylating enzymes differentially recognize the *Arabidopsis* SUMO proteins, resulting in specificity of the de-conjugating machinery. Therefore, the SUMO proteins in this signaling system have evolved to contain information that allows for both redundancy with the conjugation system and the diversity with deconjugating enzymes.

## TABLE OF CONTENTS

ABSTRACT.....	v
TABLE OF CONTENTS.....	vii
LIST OF PUBLICATIONS .....	x
LIST OF FIGURES .....	xi
LIST OF TABLES .....	xiii
LIST OF APENDICES .....	xiv
LIST OF ABBREVIATIONS .....	xv
CHAPTER 1: INTRODUCTION	
DeSUMOylating Enzymes.....	1
CHAPTER 2: REVIEW OF THE LITERATURE	
SUMO and deSUMOylating Enzymes.....	5
How it all began: The history of SUMO.....	5
SUMO: The present.....	7
DeSUMOylating Enzymes: What do we really know and what is next?.....	12
CHAPTER 3: METHODOLOGY	
Constructs.....	17
Protein expression and purification.....	21
<i>In vitro</i> peptidase assays.....	22

<i>In vitro</i> SUMOylation assays.....	23
Mass spectrometry analysis.....	23
Protein expression and purification for crystallography experiments.....	23
Protein crystallization.....	25
Data collection, structural determination, and refinement.....	26
Structure figures.....	27
Accession numbers.....	27

#### CHAPTER 4: RESULTS

<i>In vitro</i> characterization of <i>Xanthomonas</i> XopD.....	28
Introduction.....	28
Results.....	29
Discussion.....	39

#### CHAPTER 5: RESULTS

Structural analysis of XopD provides insight into its substrate specificity.....	44
Introduction.....	44
Results and Discussion.....	44
Conclusions.....	69

#### CHAPTER 6: RESULTS

Experimental Trials: <i>Xanthomonas</i> XopD in complex with Tomato SUMO.....	70
---	----



Introduction.....	70
Experimental Trials.....	70
CHAPTER 7: RESULTS	
Evolution of a signaling system that incorporates both redundancy and diversity: <i>Arabidopsis</i> SUMOylation.....	89
Introduction.....	89
Results and Discussion.....	90
Conclusions.....	109
CHAPTER 8: DISCUSSION AND CONCLUSIONS	
deSUMOylating Enzymes: What was known in the beginning.....	112
Discussion of research findings.....	114
Summary of main contributions to the field.....	118
Future directions.....	119
APPENDIX A: TABLE OF PRIMERS.....	120
APPENDIX B: XopD CRYSTALLOGRAPHIC DATA.....	126
BIBLIOGRAPHY.....	128
VITAE .....	132

## LIST OF PUBLICATIONS

Chosed, R., Mukherjee, S., Lois, M.L., and Orth, K. (2006) Evolution of a signaling system that incorporates both redundancy and diversity: *Arabidopsis* SUMOylation. *Biochemical Journal*. Submitted.

Hotson, A. \*, Chosed, R. \*, Shu, H., Orth, K., and Mudgett, M.B. (2003) *Xanthomonas* type III effector XopD targets SUMO-conjugated proteins *in planta*. *Molecular Microbiology* **50**: 377-389. (\* These authors contributed equally to this work.)

## LIST OF FIGURES

FIGURE 1 .....	8
FIGURE 2 .....	10
FIGURE 3 .....	13
FIGURE 4 .....	30
FIGURE 5 .....	33
FIGURE 6 .....	34
FIGURE 7 .....	39
FIGURE 8 .....	47
FIGURE 9 .....	48
FIGURE 10 .....	49
FIGURE 11 .....	50
FIGURE 12 .....	51
FIGURE 13 .....	52
FIGURE 14 .....	54
FIGURE 15 .....	56
FIGURE 16 .....	57
FIGURE 17 .....	58
FIGURE 18 .....	58
FIGURE 19 .....	60
FIGURE 20 .....	61
FIGURE 21 .....	62

FIGURE 22 .....	65
FIGURE 23 .....	66
FIGURE 24 .....	67
FIGURE 25 .....	68
FIGURE 26 .....	72
FIGURE 27 .....	75
FIGURE 28 .....	78
FIGURE 29 .....	79
FIGURE 30 .....	80
FIGURE 31 .....	81
FIGURE 32 .....	83
FIGURE 33 .....	84
FIGURE 34 .....	87
FIGURE 35 .....	93
FIGURE 36 .....	96
FIGURE 37 .....	100
FIGURE 38 .....	101
FIGURE 39 .....	103
FIGURE 40 .....	105
FIGURE 41 .....	108
FIGURE 42 .....	110

## LIST OF TABLES

TABLE ONE .....	120
TABLE TWO .....	122
TABLE THREE .....	126

## LIST OF APPENDICES

APPENDIX A: TABLES OF PRIMERS.....	120
APPENDIX B: XopD CRYSTALLOGRAPHIC DATA.....	126

## LIST OF ABBREVIATIONS

2x YT- two times yeast extract and tryptone

*At* – *Arabidopsis thaliana*

BLAST – Basic Local Alignment Search Tool; <http://ncbi.nih.gov/BLAST/>

βME – Beta-mercaptoethanol

CNS – Crystallography & NMR System

DTT – Dithiothreitol

*E. coli* – *Escherichia coli*

ESD4 – Early in short days<sup>4</sup>

FPLC – fast protein liquid chromatography

GST – glutathione S-transferase

HA – Influenza A virus hemagglutinin

Hrp –hypersensitive response and pathogenicity

IAA – iodoacetamide

IPTG – isopropyl-beta-D-thiogalactopyranoside

kDa - kilodalton

MEROPS – MEROPS peptidase database; <http://merops.sanger.ac.uk/>

NCBI - National Center for Biotechnology Information; <http://ncbi.nih.gov/>

Nedd8 – Neural precursor cell-expressed developmentally down-regulated

NEM – N-ethylmaleimide

Ni-NTA – nickel-nitriloacetic acid matrix (QIAGEN)

O.D. – Optical density

PBS – phosphate-buffered saline

PCR – polymerase chain reaction

PDB – protein data bank

PMSF – Phenylmethylsulphonylfluoride

RanGAP – Ran GTPase -activating protein

RMSD – root-mean-square deviation

RRL – rabbit reticulocyte lysate

SDS-PAGE – sodium dodecyl sulfate-polyacrylamide gel electrophoresis

SeMet – selenomethionyl

SENp – sentrin-specific protease

SUMO – small-ubiquitin related modifier

TEV – tobacco etch virus

TNT – transcription and translation

TTSS – type three secretion system

ULP1 – ubiquitin-like protein protease-1

*Xcv* – *Xanthomonas campestris* pathovar *vesicatoria*

XopD – *Xanthomonas* outer protein D

YopJ – *Yersinia* outer protein J



# CHAPTER ONE

## Introduction

### DeSUMOylating enzymes

In 2002, Noel *et al.* identified two novel type three secreted effector proteins in the plant pathogen, *Xanthomonas campestris* pathovar *vesicatoria* (1). One of these effectors, XopD (*X*anthomonas *O*uter *P*rotein *D*), showed sequence homology within its carboxy terminal domain to the C48 family of cysteine proteases of clan CE, specifically to the carboxy terminal domain of ubiquitin-like protein protease (Ulp1) family (2-4). Ulp1 is a cysteine protease from yeast that processes SUMO/Smt3-conjugated proteins (3). Ulp1 acts as a peptidase to process SUMO to its mature form by exposing its carboxy terminal GlyGly so that SUMO can be covalently added to its target protein via SUMOylation machinery (5). Ulp1 acts as an isopeptidase to deconjugate SUMO from SUMO-modified target proteins by cleaving the isopeptide bond formed between the carboxy terminal Gly of SUMO and the epsilon amino group of a Lys residue in the target protein (5).

*Xanthomonas campestris* pv. *vesicatoria* is the causal agent of bacterial spot disease in tomato and pepper plants (6). *Xanthomonas campestris* uses a *hrp*- (hypersensitive response and pathogenicity) type three encoded protein secretion system (TTSS) to inject effector proteins into host plant cells (7). The TTSS is found only in gram-negative bacteria and it functions to transport effector proteins across the inner bacterial membrane, the peptidoglycan layer, the outer bacterial membrane, and the host plasma membrane or cell wall to reach the host cell cytoplasm (8,9). The translocation of proteins into the host cell via

the TTSS is initiated by contact of the bacteria with the host cell (8). There are between 20-25 different proteins that make up a typical TTSS apparatus which forms a syringe like structure projecting from the surface of the bacteria (8). The effector proteins are transported through this syringe like channel in an unfolded state to reach the host cell interior (8). Once in the host plant cell, the effector proteins of *Xanthomonas campestris* have been reported to interact with specific host cell targets to cause plant cell death (7).

Since XopD showed high sequence similarity to Ulp1, it was hypothesized that XopD would share a similar protease activity in *Xanthomonas campestris* pv. *vesicatoria* pathogenesis. The hypothesized role of *Xanthomonas campestris* XopD, based on its sequence identity and that it is a type three secreted effector protein, was as a deSUMOylating protease that targets SUMO-conjugated proteins in host plant cells during *Xanthomonas campestris* infection. Characterization of the *in vitro* activity of XopD as a deSUMOylating protein is discussed in Chapter Four. These studies led to the hypothesis that XopD exhibits species specificity for the SUMO portion of SUMO modified target proteins. Tomato SUMO (T-SUMO) was the first SUMO substrate identified for which XopD exhibits peptidase activity. Tomato SUMO was first identified by Hanania *et al.* as playing a role in tomato plant cell defense responses (10). There has only been one SUMO protein identified in tomato plants, and there are no SUMO proteins identified in pepper plants to date. XopD is unable to process mammalian and yeast SUMO substrates, but is able to process plant SUMO conjugated targets proteins, suggesting that XopD is a species specific SUMO protease. The focus of Chapter Five is to determine how XopD recognizes its SUMO substrates and why XopD can only process certain SUMO substrates. To

understand how XopD recognizes its SUMO substrates I embarked on structural studies encouraged by the structure of yeast Ulp1 in complex with its SUMO substrate, Smt3 (11). The goal of solving the crystal structure of XopD was to understand how XopD is recognizing its SUMO substrates by understanding, which regions of XopD are in contact with certain regions of the SUMO substrate. To better understand how XopD is recognizing T-SUMO, I attempted numerous trials at purifying a covalent complex of XopD with T-SUMO to be used for crystallization trials. These trials were unsuccessful and are detailed in Chapter 6.

The XopD structural studies were complemented with mutational studies in SUMO substrates. SUMO proteins from *Arabidopsis thaliana* provided a clue as to which regions of the SUMO protein to focus mutational studies. Vierstra and colleagues had identified a family of eight genes in the *Arabidopsis thaliana* genome that encoded SUMO-like proteins (12). XopD was able to recognize two of the *Arabidopsis* SUMOs, which shared approximately 80% sequence identity with T-SUMO. The SUMO proteins for *Arabidopsis thaliana* not only provided insight into the specificity of XopD, but also provided a unique system to further study how other deSUMOylating enzymes recognize various SUMOs.

In addition to identifying SUMO genes in the *Arabidopsis thaliana* genome, Vierstra and colleagues identified a family of four *ULP1* genes (*AtULP1A*, *B*, *C* and *D*) encoded in the *Arabidopsis* genome and Coupland and colleagues have identified a fifth *ULP1* gene (*AtESD4*) (12,13). Since this plant system contained several *SUMO* and *ULP1* genes, I was able to ascertain which *ULP1* recognized which SUMO in *Arabidopsis thaliana*. Chapter Seven describes this study, which includes SUMO peptidase assays, as well as SUMO

isopeptidase assays using an *in vitro* SUMOylation system. The study also included four mammalian SUMOs and yeast SUMO (Smt3) along with yeast Ulp1. The results allowed me to group the various ULP1s based on their activity towards the various SUMO substrates. Some of the SUMO substrates tested were not recognized by any ULP1. I was able to use the *in vitro* SUMOylation system with the various SUMO substrates and showed that all the SUMOs tested were able to be utilized by the SUMO conjugation machinery. This led to the idea that the basic E1 and E2 of the SUMO conjugation machinery is quite promiscuous in its choice of SUMO substrate, yet the deSUMOylating machinery is rigid in its specificity for SUMO substrates.

Chapters Four, Five, and Seven all describe the use of biochemistry to characterize the activity and specificity of various ULP1s for their SUMO substrates. All of these studies have led to the hypothesis that ULP1 family members are not promiscuous enzymes but proteases that have specificity for SUMO substrates. This specificity of the deSUMOylating enzymes is predicted to play a role in regulating the activity of SUMO modifications in general. I have shown that the SUMOylation machinery recognizes any of the SUMO substrates that is presented. Undoubtedly, the growing family of E3 SUMO ligases contribute to the specificity of the SUMO conjugation machinery for diverse substrates. During the evolution of the SUMO signaling system, surface residues in SUMOs have been conserved so that the substrates are recognized by the conjugation machinery, while other surface residues have diversified thereby adding complexity to this type of posttranslational modification.

## **CHAPTER TWO**

### **Review of the Literature**

#### **SUMO and DeSUMOylating Enzymes**

#### **How it all began: The history of SUMO**

SUMO (small ubiquitin-like modifier), like ubiquitin, is an evolutionarily conserved, reversible posttranslational modification that is covalently attached to cellular proteins via an isopeptide bond. Unlike ubiquitin, modification by SUMO does not target proteins for degradation by the proteasome.

One of the first reports on SUMO demonstrated that RanGAP (Ran GTPase activating protein) is modified by mammalian SUMO-1 (M-SUMO-1) (14). Matunis *et al.* discovered that RanGAP (70 kDa) existed in a 90 kDa form when associated with the nuclear pore complex (14). The authors were able to use peptide sequencing to determine that this 90 kDa form of RanGAP contained an ubiquitin-like protein modification, later referred to as SUMO (14). Later, this modification of RanGAP by M-SUMO-1 was shown to be critical for the localization of RanGAP-SUMO-1 to the nuclear pore complex and its association with RanBP2 (15,16).

Around the same time, another report implicated SUMO (also referred to as Sentrin) as a molecule that blocked responses triggered by cell death receptors (17). Kamitani *et al.* saw the presence of high molecular weight bands in cell extracts when blotted with an anti-

SUMO antibody (18). The same group also determined that SUMO had two conserved glycine residues near its carboxy terminus, suggesting that SUMO may be able to modify or conjugate to proteins using a mechanism that is similar to that of ubiquitin conjugation (18). Upon mutation of the carboxy terminal Gly-Gly to Gly-Ala in SUMO, the high molecular weight species of SUMO were no longer detected with anti-SUMO antibody (18). Mahajan *et al.* were able to detect that SUMO modification of RanGAP requires ATP, just as ubiquitin modification is ATP-dependent (16).

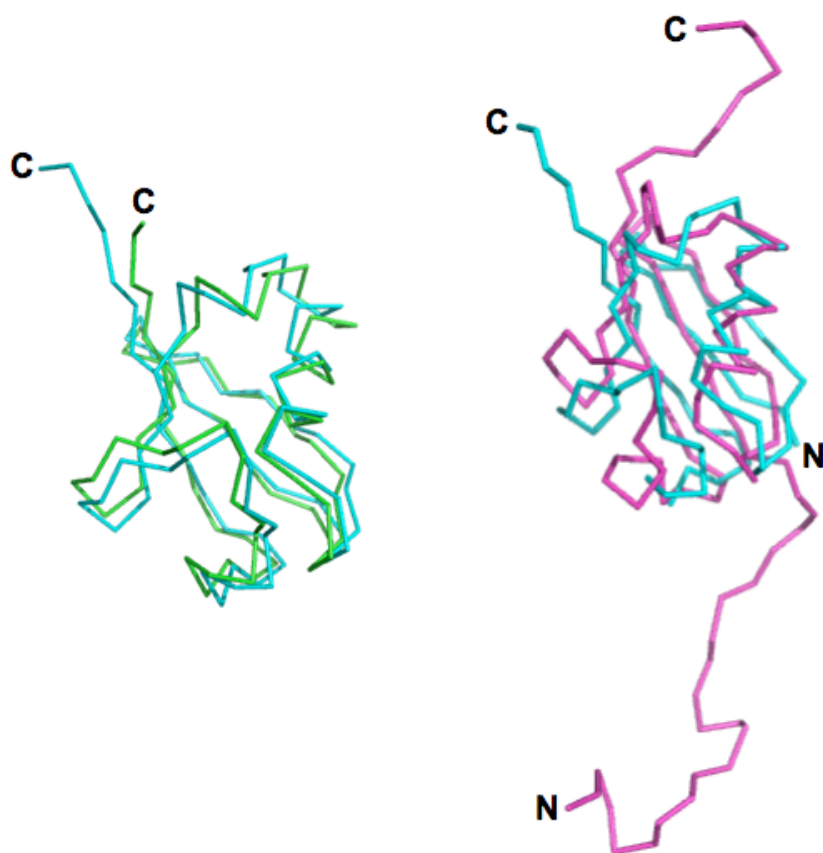
In the same year, Johnson *et al.* identified the yeast SUMO protein, Smt3 (19). This group showed that Smt3 had to be processed to its mature form, exposing its carboxy terminal Gly-Gly residues before it could be conjugated to target proteins (19). The same paper also identified Aos1/Uba2 as the potential E1 in the SUMO conjugation pathway (19). Shortly after this discovery, Ubc9 was identified as the E2 conjugating enzyme for the SUMO conjugation pathway by two labs (20,21). The next year brought the identification of a specific lysine (K526) residue in RanGAP that was conjugated to M-SUMO-1 via its carboxy terminal glycine 97 (22).

The structure of M-SUMO-1 was solved in 1998 and although M-SUMO-1 shares only 18% sequence identity with ubiquitin and does not target proteins for degradation, the tertiary structure of M-SUMO-1 was very similar to that of ubiquitin (23). Then in 1999, Li and Hochstrasser identified Ulp1 (Ubiquitin-like protein protease) as a SUMO deconjugating enzyme in yeast (3). The SUMO found in *Saccharomyces cerevisiae*, Smt3, is an essential gene in yeast and the deconjugation of Smt3 from target proteins has been shown to be involved in cell cycle progression (3). Later that same year, Suzuki *et al.* identified a de-

SUMOylating enzyme in bovine brain, and suggested that there could be more than one of these proteases in mammalian cells (24). Thus the field of deSUMOylating enzymes was born.

### **SUMO: the present**

As with ubiquitin, SUMO utilizes a conjugation machinery (E1, E2, E3) to modify specific target proteins. Similarly, another small ubiquitin-like protein, Nedd8, also has its own conjugation machinery, or Neddylation machinery. Ubiquitin and Nedd8 share approximately 60% sequence identity and exhibit very similar secondary structures as shown in Figure 1 (left side). While SUMO shares only 18% sequence identity with ubiquitin, their secondary structures are very similar (Figure 1, right) (23). All three protein structures contain a five-stranded beta sheet packed against two alpha helices, referred to as a ubiquitin fold (25). As depicted in Figure 1, the M-SUMO-1 crystal structure showed a very flexible N-terminal region unlike ubiquitin and Nedd8 (23). This extended charged amino-terminus may serve to aid in interactions with other proteins.

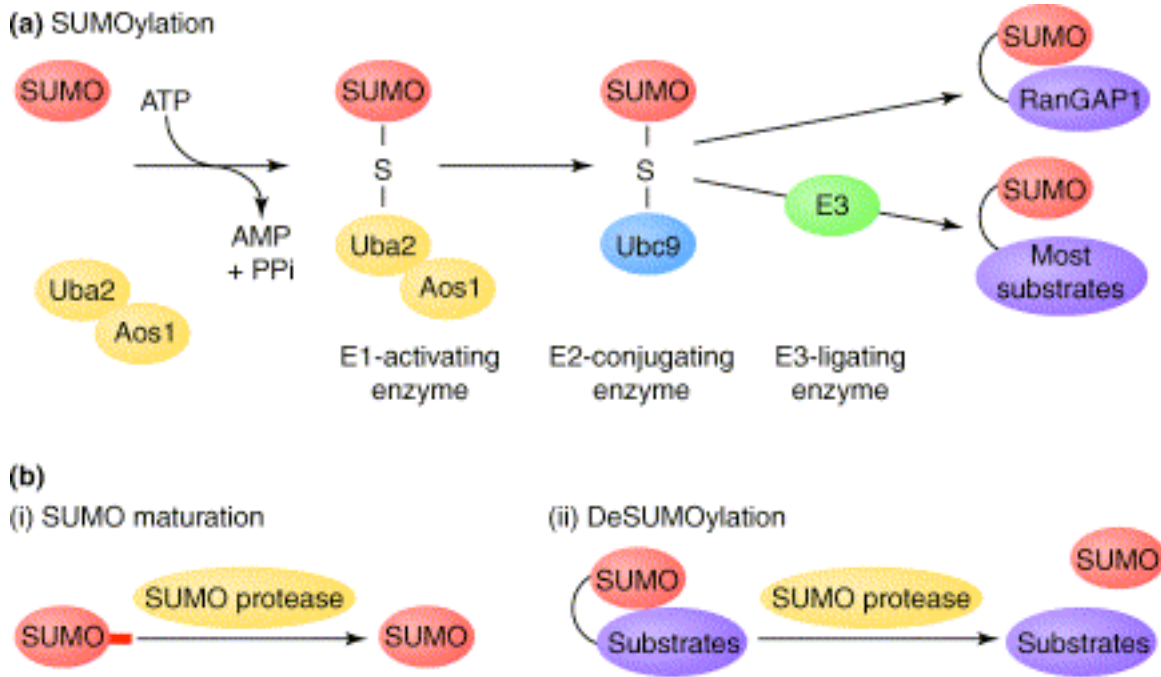


**Figure 1: The structures of M-SUMO-1, Ubiquitin, and Nedd8 show similar three dimensional structures.** Left, ribbon structure of Ubiquitin (cyan) (PDB ID: 1D3Z) aligned with Nedd8 (green) (PDB ID: 1NDD). Right, ribbon structure of M-SUMO-1 (magenta) (PDB ID: 1A5R) aligned with Ubiquitin (cyan).

SUMO, like ubiquitin, is attached to a target protein via an isopeptide bond between its carboxy terminal glycine and the  $\epsilon$ -amino group of a lysine residue on the target protein. SUMOylation, like ubiquitination, requires ATP for the first step of conjugation. The E1 activating enzyme for SUMOylation is Aos1/Uba2 (SAE1/SAE2) that forms a thioester bond with the carboxy terminal Gly of SUMO and Cys 173 on Uba2 (5). However, modification by SUMO is distinct from that by ubiquitin in a number of ways (26). First, SUMOylation machinery makes use of one universal E2 conjugating enzyme, Ubc9, which is able to



transfer SUMO directly to target proteins. In some cases, an E3 ligase is utilized to facilitate the transfer of SUMO to the target protein. The carboxylate group of the carboxy terminal Gly of SUMO forms an isopeptide bond with the  $\epsilon$ -amino group of a specific Lys in the target protein (Figure 2). Second, for SUMO, a “consensus” sequence modification motif in target proteins has been identified as  $\Psi$ KXE, where  $\Psi$  is a large hydrophobic residue, K is the lysine that SUMO is attached to, X is any residue, and E is glutamic acid (27). Ubiquitin can modify a protein either with a poly-ubiquitin chain or with a mono-ubiquitin (28). Modifications by SUMO are most commonly associated with mono-SUMOylation, although, recent studies have implicated the use of poly-SUMOylation *in vitro* (29). While both ubiquitin and SUMO are processed to a mature form, by exposing two glycine residues at their carboxy terminus, a distinct family of cysteine proteases, referred to as ubiquitin-like protein protease-1(s) (ULP1s), process the carboxy terminus of SUMO (26,30).



**Figure 2: The SUMO conjugation pathway makes use of an E1, E2, and E3 similar to the ubiquitin conjugation pathway (30).**

Since the original discovery of mammalian SUMO-1, three other mammalian SUMOs have been discovered. Mammalian SUMO-1 and SUMO-2 share 48% sequence identity, while SUMO-2 and SUMO-3 share 95% sequence identity with one another. Mammalian SUMO-2 and SUMO-3 are thought to be involved in how the cell responds to environmental stress (31). SUMO-2 and SUMO-3 (not SUMO-1) are known to form SUMO chains *in vitro* because they contain a Lys residue near the amino terminal region of the protein (29). M-SUMO-1 has also been shown to mainly modify RanGAP, thus M-SUMO-1 is localized at the nuclear envelope and nucleolus (14,15,32). In contrast, M-SUMO-2 and M-SUMO-3 are found in the nucleoplasm and not at the nuclear envelope (32). More recently, mammalian SUMO-4 was discovered and subsequently shown to be highly

expressed in the kidneys (33). Mammalian SUMO-4 shares 86% sequence identity with mammalian SUMO-2. A single polymorphism in Mammalian SUMO-4, Met55Val, was discovered. It has been proposed that this single amino acid change may link mammalian SUMO-4 to type I diabetes (33).

SUMO, which has been found in the nucleus, modifies a number of transcription factors and this modification regulates the activity of these transcription factors (34). Sp3, a transcription factor, is regulated by SUMOylation as well as by acetylation on the same Lys 539 (35). Mutation of this lysine to an arginine increases the transcriptional activity of Sp3 as does deSUMOylation of Sp3-SUMO by SENP2 (35). Another well studied transcription factor and tumor suppressor, p53, is SUMOylated on Lys 386 which activates its transcriptional activity (36). SUMOylation of K-bZIP, a transcriptional repressor from Kaposi's sarcoma-associated herpesvirus, increases its repressive activity, which is reversed by mutation of Lys to Arg (37). The SUMO modification of HDAC-1 causes activation of its deacetylase activity as well as stimulates its transcriptional corepressor activity (38). PML, a tumor suppressor, is modified by M-SUMO-1, which is required for its localization to PML nuclear bodies (39). PML nuclear bodies are also shown to contain several SUMOylated proteins including transcription factors (40). In 1998, Hay and colleagues discovered that I $\kappa$ B $\alpha$  is SUMOylated at Lys 21 which is also the site for ubiquitination on the protein (41). When I $\kappa$ B $\alpha$  is SUMOylated, it cannot be ubiquitinated, thus it becomes resistant to degradation by the proteasome and allows for inhibition of NF $\kappa$ B dependent transcription (41).

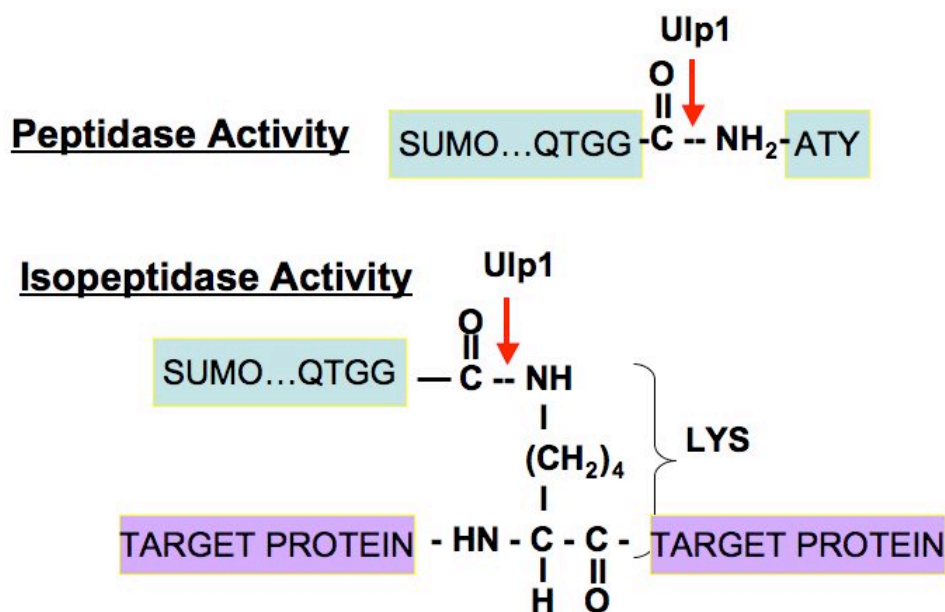
In early 2006, Melchior and colleagues observed that at low concentrations of ROS (reactive oxygen species) the levels of SUMO conjugates decreases, including those of SUMOylated transcription factors (42). They found that the mechanism of the reduced SUMO conjugates in response to ROS involves the formation of a disulfide bond between Cys 173 of the E1 subunit Uba2 and Cys 93 of Ubc9 (E2) (42). The disulfide bond formation inhibits the Ubc9 from forming a thioester bond with SUMO, thus the SUMO conjugation machinery is inhibited (42).

A SUMO protein from tomato plants was discovered in a yeast two-hybrid assay searching for proteins that interacted with ethylene inducing xylanase (10). More recently, Kurepa *et al.* identified eight SUMO encoding genes in *Arabidopsis thaliana* (12). Of these eight SUMO proteins, *Arabidopsis thaliana* SUMO-4, SUMO-6, and SUMO-7 do not contain the conserved GlyGly motif required for SUMO conjugation to target proteins. Kurepa *et al.* also did not detect mRNAs for *Arabidopsis thaliana* SUMO-4, SUMO-6, SUMO-7, and SUMO-8 (12). They were able to show that *Arabidopsis thaliana* SUMO-1 and SUMO-2 conjugated proteins are abundant in plants when exposed to stress. Another study showed that *Arabidopsis thaliana* SUMO-1 may be involved in abscisic acid signaling and the plant stress response (43). The SUMO conjugation machinery, E1, E2 and E3, in *Arabidopsis thaliana* has also been characterized (12).

### **deSUMOylating Enzymes: What do we really know, and what is next?**

ULP1s, like SUMOs, belong to a growing family of proteins that are evolutionarily conserved from yeast to man. Ulp1 from yeast plays a dual role, acting as a peptidase by

cleaving the full-length SUMO and as an isopeptidase by cleaving SUMO-conjugated target proteins (Figure 3) (3). Ulp1 from yeast has been shown to act as a peptidase by cleaving the full-length Smt3 to its mature form (carboxy terminal glycine exposed) to allow for Smt3 to be conjugated to target proteins (Figure 3) (3). Yeast Ulp1 also acts as an isopeptidase to cleave Smt3 that has been conjugated to target proteins. Yeast Ulp1 is localized to the nuclear pore complex (NPC) and is an essential gene in yeast, yet yeast that express only the carboxy terminal catalytic domain of Ulp1 are viable (3,30,44). Ulp2 is another member of the Ulp1 family found in yeast, although it is not an essential gene (30). Unlike Ulp1, Ulp2 is localized to the nucleoplasm and has been shown to be involved in chromosome segregation (30).



**Figure 3: ULP1s exhibit peptidase as well as isopeptidase activity for their SUMO substrates.**

Seven ULP1 proteases have been found in humans and are referred to as SENP1,2,3,5,6,7, and 8 (30,45). All of the SENPs share the conserved carboxy terminal catalytic domain with yeast Ulp1, yet differ in their size, amino terminal regulatory domain sequence, substrate preference, as well as cellular localization (30,45). SENP1 is localized to nuclear bodies, and is unable to process SUMO-1 modified RanGAP (46,47). SENP1 and SENP2 have been shown to process mammalian SUMO-1,-2, and -3 conjugates (46,48). SENP2 is better characterized than the other mammalian SENPs. SENP2 is localized to the nuclear face of the NPC and is associated with Nup153 (48,49). The amino terminal domain of SENP2 is required for its localization to the NPC (49). In one case, localization of human SENP2 appears to be regulated by differential splicing of its RNA message (50). Mouse SENP3 is localized to the nucleolus during interphase and is known to process mammalian SUMO-2 and -3 conjugates (51). SENP5 and SENP7 are not characterized in the literature to date. SENP6 is found in the cytoplasm and is considered a Ulp2 like enzyme whereby it is thought to exhibit mainly peptidase activity as opposed to isopeptidase activity (52). The amino terminal domain in SENP6 is required for its localization and SENP6 has been shown to process both mammalian SUMO-2 and -3 (52). However, Kim *et al.* reported to have only been able to detect peptidase activity for SENP6 since the protease was unable to remove GST-M-SUMO-1 from RanGAP *in vitro* (52). SENP6 has been shown to be highly expressed in the reproductive organs (52). SENP8, unlike the other characterized SENPs, is a deNeddylating enzyme, whereby it removes Nedd8 modifications from target proteins (53).

Tools for identification of SUMO, Ubiquitin, and Nedd8 conjugating and deconjugating enzymes has been aided by the development of selective inhibitors (54).

These inhibitors are ubiquitin-like proteins that have been modified, using intein based chemistry, so that their carboxy terminus acts as an electrophilic trap (55).

Differences in cellular localization of these seven SENPs have been proposed to account for the large number of SENPs and possibly their substrate specificity. While cellular localization accounts for the subset of SUMOylated proteins available to certain Ulp1 like proteins under certain cellular conditions, another layer of specificity intrinsic to these proteases may be responsible for substrate recognition and cleavage. Li and Hochstrasser performed extensive yeast genetic studies to decipher the function of the amino terminal domain of yeast Ulp1 *in vivo*. They concluded that the amino terminal domain of yeast Ulp1 was responsible for the NPC localization of Ulp1 in yeast. They also observed that expression of the full length Ulp1 processed Smt3-modified substrates more efficiently than the amino terminal domain deleted yeast Ulp1 (44).

Vierstra and colleagues have identified a family of four ULP1 genes (*AtULP1A*, *B*, *C* and *D*) encoded in the *Arabidopsis* genome and Coupland and colleagues have identified a fifth *ULP1* gene (*AtESD4*) (12,13). One *Arabidopsis thaliana* SUMO protease has been characterized. This protease, ESD4 (Early in Short Days 4), was shown to act as a SUMO protease in *Arabidopsis thaliana* and may help in regulating time of flowering (13).

Another ULP1, XopD (*Xanthomonas* outer protein *D*) from *Xanthomonas campestris* pv. *vesicatoria*, is expressed as a secreted virulence factor by this plant bacterial pathogen (56). *In planta* XopD reduces overall SUMO conjugates and exhibit peptidase as well as isopeptidase activity towards plant SUMOs *in vitro* (56). In 2004, a cytoplasmic protein from *Xanthomonas*, AvrXv4, was shown act as a deSUMOylating enzymes *in planta* (57).

As bacteria do not encode either the ubiquitin or the SUMO signaling machineries, *Xanthomonas* appears to have usurped the activity of eukaryotic ULP1s and uses this activity to aid in pathogenesis by disrupting host defense signaling in the infected plant cell (56). These and the aforementioned studies support the importance of SUMO modification in plant signaling (12,13,43,56).

The importance of SUMO modification of target proteins has been established in many cellular processes (30,58). However, the mechanisms that dictate the specificity, frequency and half-life of this modification are poorly understood. The basic machinery involved in the conjugation of SUMO (E1 and E2) is promiscuous, in that a diverse pool of SUMO molecules are used as substrates. The localization and diversity of the SUMO E3s also plays a major role in regulating the conjugation. As a consequence of subcellular localization, spatial restrictions on conjugation and deconjugation machinery are thought to contribute to the specificity of both the addition and removal of particular SUMO moieties. *In vitro* studies have supported the notion that the deSUMOylating enzymes are promiscuous and that their localization dictates their specificity (26,30).



## CHAPTER THREE

### Methodology

#### Constructs

#### *Chapter 4*

Polymerase chain reaction (PCR) was used to clone tomato SUMO (T-SUMO) and *Xcv xopD* and to construct gene fusions. PCR-generated DNA fragments were cloned into pCR-BluntII-TOPO or pCRII-TOPO (Invitrogen). The sequence of DNA constructs was verified by cycle sequencing. Tomato SUMO (1–131 bp) was amplified from a tomato cDNA library (a gift from W. Gruissem) and cloned to create pCRII (HA-T-SUMO) and pCRII (T-SUMO-HA) (10). For *in vitro* transcription and translation, tomato SUMO was cloned as a NdeI–XhoI fragment into pET15b (Novagen) creating pET15b (T-SUMO-HA). Tomato SUMO 1–131 bp was amplified from pET15b (T-SUMO-HA) and subcloned as a BamHI–HindIII fragment into pT7-LO (a gift from J. Clemens) to create T-SUMO (1–96) followed by Ala-Thr-Val-Asn-His<sub>6</sub>. Mammalian SUMO-1 was amplified from EST 2578604 (Research Genetics) and cloned in pET15b (Novagen) as an NdeI–XhoI fragment creating pET15b (M-SUMO-1-HA). RanGAP1 (a gift from N. R. Yaseen) was amplified and cloned into pcDNA3 (Invitrogen) as an XbaI–EcoRI fragment creating pcDNA3 (HA-RanGAP1).

The *xopD* open reading frame (ORF; 1–1638 bp) was amplified from *Xcv* strain 85-10 genomic DNA by PCR and cloned, creating pCRII(*xopD*<sub>1-546</sub>) (1). GST–XopD fusion protein was constructed by subcloning the BamHI–XhoI fragment from pCRII (*xopD*<sub>1-546</sub>)

into pGEX-5X-3 (Amersham Biosciences). Ulp1 was amplified by PCR from yeast genomic DNA and cloned into pGEX-KG (59) as a BamHI–EcoRI fragment creating pGEX-KG (GST-Ulp1). The catalytic mutant of XopD was generated using the Stratagene QuikChange™ site-directed mutagenesis kit by replacing cysteine codon 470 (TGC) with an alanine codon (GCC), and mutants were confirmed by DNA sequence analysis (Refer to Table 1 in Appendix A).

## *Chapter 5*

Polymerase chain reaction (PCR) was used to clone all SUMO and XopD constructs and to construct gene fusions. XopD, T-SUMO, yeast Ulp1, and RanGAP were cloned as described above in *Chapter 4*. For constructing GST-XopD truncations, XopD<sub>1-545</sub>, XopD<sub>285-545</sub>, XopD<sub>285-520</sub>, XopD<sub>305-520</sub>, XopD<sub>335-520</sub> were cloned in pGEX-rTEV (60). Refer to Table 2 of Appendix A for cloning details of the XopD constructs. These XopD constructs contained an N-terminal GST tag that could be cleaved with TEV protease (tobacco etch virus) or thrombin (GST-TEV protease expression vector was a gift from Yuh Min Chook). The catalytic mutant of XopD was generated using the Stratagene QuikChange™ site-directed mutagenesis kit by replacing cysteine codon 470 (TGC) with an alanine codon (GCC), and mutants were confirmed by DNA sequence analysis (Refer to Table 1 in Appendix A).

Smt3-HA was cloned from *Saccharomyces cerevisiae* cDNA. T-SUMO-His<sub>6</sub> was constructed by using pT7-LO (a gift from J. Clemens) and Smt3-His<sub>6</sub> was constructed by using pET15b (Novagen). These constructs contained an N-terminal His<sub>6</sub> tag directly following the Gly-Gly motif. The various mutants of T-SUMO-His<sub>6</sub> and Smt3-His<sub>6</sub> were

generated using the Stratagene QuikChange<sup>TM</sup> site-directed mutagenesis kit and mutants were confirmed by DNA sequence analysis (Refer to Table 1 in Appendix A).

For constructing His<sub>6</sub>-SUMO-GG-HA constructs to be used in *in vitro* peptidase assays, *At*SUMO (-1,-2,-3), T-SUMO, Smt3, and M-SUMO (-1, -2, -4) were cloned into pET15b (Novagen) and *At*SUMO-5 was cloned into pT7-LOH (gift from J. Clemens). These SUMOs were constructed with an N-terminal His<sub>6</sub> tag and their C-terminus contained an HA tag directly following the Gly-Gly motif. For constructing GST-SUMO-GG-STOP constructs to be used with *in vitro* SUMOylation assays, T-SUMO and M-SUMO-1 were cloned into pGEX-rTEV (60). These SUMOs were constructed with an N-terminal GST tag and their C-terminus contained a STOP codon after the Gly-Gly motif.

## Chapter 6

XopD and T-SUMO were cloned as described in *Chapter 4*. XopD<sub>285-545</sub>, XopD<sub>285-520</sub>, XopD<sub>305-520</sub>, XopD<sub>335-520</sub> were cloned in pGEX-rTEV as described in *Chapter 5*. These XopD constructs contained an N-terminal GST tag that could be cleaved with TEV protease or thrombin. The catalytic mutant of XopD and XopD<sub>305-520</sub> R334A, G495 were generated using the Stratagene QuikChange<sup>TM</sup> site-directed mutagenesis kit and mutants were confirmed by DNA sequence analysis. Refer to Table 2 of Appendix A for cloning details. *At*SUMO-His<sub>6</sub> construction is described in *Chapter 5*. The 17 amino acid N-terminal truncation of T-SUMO (NT-SUMO) was cloned using PCR. NT-SUMO-His<sub>6</sub> was constructed without an N-terminal tag using the pT7-LO vector. His<sub>6</sub>-NT-SUMO-GG-SCCT was constructed using the pT7-LO vector. This construct contained an N-terminal tag that

was not cleavable. His<sub>6</sub>-(TEV)-NT-SUMO-GG was constructed with an N-terminal tag followed by a TEV cleavage site using the pT7-LO vector. Refer to Table 2 of Appendix A for cloning details of each T-SUMO construct.

## Chapter 7

Polymerase chain reaction (PCR) was used to clone all SUMO and Ulp1 constructs and to construct gene fusions. *At*SUMOs (-1,-2,-3,-5), *At*ULP1A, *At*ULP1C, *At*ULP1D, and *At*ESD4 were cloned from *At* cDNA. Smt3-HA was cloned from *Saccharomyces cerevisiae* cDNA. M-SUMO-1 and M-SUMO-2 were cloned from mammalian cDNA. M-SUMO-4 was a gift from Cong-Yi Wang at Medical College of Georgia (61). T-SUMO, RanGAP, and yeast Ulp1 were cloned as described in *Chapter 4* methods above.

For constructing GST-SUMO-GG-STOP constructs to be used with *in vitro* SUMOylation assays, *At*SUMO (-1, -2, -3, -5), T-SUMO, SMT3, and M-SUMO (-1, -2, -4) were cloned into pGEX-rTEV (60). These SUMOs were constructed with an N-terminal GST tag and their C-terminus contained a STOP codon after the Gly-Gly motif. For constructing His<sub>6</sub>-SUMO-GG-HA constructs to be used with *in vitro* peptidase assays, *At*SUMO (-1,-2,-3), T-SUMO, SMT3, and M-SUMO (-1, -2, -4) were cloned into pET15b (Novagen) and *At*SUMO-5 was cloned into pT7-LOH (gift from J. Clemens). These SUMOs were constructed with an N-terminal His<sub>6</sub> tag and their C-terminus contained an HA tag directly following the Gly-Gly motif. For constructing GST-ULP1 constructs to be used with *in vitro* peptidase and isopeptidase assays, *At*ULP1 A, C, D, and ESD4 were cloned into pGEX-rTEV. These ULP1s were constructed with an N-terminal GST tag used for protein

purification purposes. The sequence of all DNA constructs was verified by cycle sequencing. Refer to Table 2 in Appendix A for cloning details.

The mutants of M-SUMO-1 and M-SUMO-4 were generated using the Stratagene QuikChange<sup>TM</sup> site-directed mutagenesis kit and mutants were confirmed by DNA sequence analysis. Refer to Table 1 in Appendix A for cloning details.

### **Protein expression and purification**

GST-tagged XopD constructs, GST-tagged SUMO constructs and all GST tagged *At* ULP1 family members were expressed in *E. coli* BL21/DE3 cells and then purified by standard GST affinity chromatography (62). Briefly, cells were grown to an O.D. (optical density) of 0.6-0.8 in 2x YT media and then induced with 400M IPTG (Roche) for 4 hours at 30C. The cells were lysed in PBS, pH8, 1% Triton X-100 (Fisher), 0.1% -mercaptoethanol (Bio-Rad) and 1mM phenylmethylsulphonyl fluoride (PMSF, Sigma) using a cell disrupter (Emulsiflex C5). The protein was bound to glutathione agarose beads and then eluted with 10mM reduced glutathione as previously described (62). Purified GST proteins were analyzed by SDS-PAGE and quantified using a modification of the Lowry procedure (63).

For purification of recombinant T-SUMO-His<sub>6</sub>, NT-SUMO-His<sub>6</sub>, His<sub>6</sub>-NT-SUMO-GG-SCCT, and His<sub>6</sub>-(TEV)-NT-SUMO-GG, cells were lysed as described above in 50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8, 300 mM NaCl, 10 mM imidazole (Sigma) or PBS, pH8, 1% Triton X-100 (Fisher) with 0.05% βME (Bio-Rad) and 1 mM PMSF (Sigma) and purified as specified by Qiagen using Ni-NTA resin chromatography. After dialysis into 50 mM Tris, pH 7.5,

150 mM NaCl, 0.1%  $\beta$ ME, recombinant T-SUMO was further purified using a HI-Trap MonoQ column and a Superdex 75 gel filtration column using a Pharmacia AKTA FPLC.

### ***In Vitro* Peptidase Assays**

*At*SUMO-1-HA, *At*SUMO-2-HA, *At*SUMO-3-HA, *At*SUMO-5-HA, M-SUMO-1-HA, M-SUMO-2-HA, M-SUMO-4-HA, Smt3-HA, T-SUMO-HA, and mammalian HA-RanGAP were *in vitro* translated in the TNT coupled rabbit reticulocyte lysate system (Promega) with L-[<sup>35</sup>S]-methionine (Amersham). For each *in vitro* peptidase assay, 2  $\mu$ L of the <sup>35</sup>S-labelled translation reaction mixture was added to 18  $\mu$ L of either glutathione elution buffer without glutathione or 0.5 mg/mL of purified enzyme for 1 hour at 30°C. The samples were then resolved by SDS-PAGE and then the gels were incubated with Amplify fluorographic reagent (Amersham) and analyzed by autoradiography.

Cysteine protease inhibitors, N-ethylmaleimide (NEM; Sigma) or iodoacetamide (IAA; Sigma), were used for inhibition assays. In these assays, purified protein was exchanged for 50 mM Hepes buffer, pH 7 that did not contain  $\beta$ ME or reduced glutathione. For the NEM inhibition assays, 5 mM NEM was added to the purified GST-XopD protein and incubated at room temperature for 20 min followed by the addition of 2  $\mu$ L of 100 mM l-cysteine and 2  $\mu$ L of 100 mM  $\beta$ ME to stop the reaction. For the IAA inhibition assays, 10 mM IAA was added to purified GST-XopD protein (0.2 mg/ml) in 50 mM Hepes buffer, pH 7, and incubated at 37°C for 40 min. The inhibited enzymes were tested for *in vitro* proteolytic activity as described above.

### ***In Vitro* SUMOylation Assays**

*In vitro* SUMOylation of  $^{35}\text{S}$ -labelled mammalian HA-RanGAP with GST-*At*SUMO-1, GST-*At*SUMO-2, GST-*At*SUMO-3, GST-*At*SUMO-5, GST-T-SUMO, GST-Smt3, GST-M-SUMO-1, GST-M-SUMO-2, and GST-M-SUMO-4 were performed as previously described (41). Assays using these SUMOylated RanGAP proteins were performed as described above in “*In vitro* Peptidase Assays”.

### **Mass spectrometry analysis**

Purified recombinant T-SUMO-His<sub>6</sub> was incubated with purified GST-XopD, and then XopD-cleaved T-SUMO and uncleaved T-SUMO-His<sub>6</sub> protein bands were excised from the Colloidal blue-stained SDS gels and digested with trypsin. The tryptic peptides were extracted and analyzed by nanoelectrospray tandem mass spectrometry using methods similar to those described previously (64). Proteins and peptides were identified by database searching of the MS and MS/MS spectra against the NCBI non-redundant databases using Knexus software.

### **Protein expression and purification for Crystallization Experiments**

GST-XopD and GST-XopD C470A constructs, were expressed in *E. coli* BL21/DE3 cells and then purified by standard GST affinity chromatography (62). Briefly, cells were grown to an OD of 0.6-0.8 in 2x YT media and then induced with 400  $\mu\text{M}$  IPTG (Roche) for

4 hours at 30°C. The cells were lysed in PBS, pH 8, 1% Triton X-100 (Fisher), 0.1% - $\beta$ mercaptoethanol ( $\beta$ ME) (Bio-Rad) and 1mM phenylmethylsulphonyl fluoride (PMSF, Sigma) using a cell disrupter (Emulsiflex C5). The protein was bound to glutathione agarose beads and then eluted with 10mM reduced glutathione as previously described (62). Protein buffer was exchanged with 50 mM Tris-HCl pH 8, 150 mM NaCl and 0.1%  $\beta$ ME by ultrafiltration using an Amicon (Millipore) device with a 10,000 molecular weight size exclusion. The GST portion of the proteins was cleaved from the XopD protein by overnight cleavage with thrombin (Novagen). The GST moiety was removed from the protein sample by running over glutathione agarose column (3mL of beads) three times. The protein was then concentrated to 2 mL using the Amicon ultrafiltration device and buffer was exchanged with 20mM Tris-HCl pH 7.5, 0.1 M KCl, and 1mM DTT. The concentrated protein was then charged onto a HiTrap MonoQ (1mL) ion exchange column and eluted with a linear gradient using 20mM Tris-HCl pH 7.5, 1.0 M KCl, and 1mM DTT using an Akta FPLC system (Amersham). Peak fractions were analyzed by SDS-PAGE and then concentrated by ultrafiltration. The protein at this stage was either used for initial crystallization screening or subjected to further purification. The concentrated protein (2 mL) was then loaded onto a HiLoad 16/60 Superdex 75 gel filtration columns that was pre-equilibrated with 20 mM Tris-HCl, 75 mM KCl, and 0.5 mM DTT. At this stage, the protein was pure by Coomassie gel staining of SDS-PAGE gels. Peak fractions were concentrated to 15 mg/mL for crystallization.

For the production of selenomethionine-labeled protein (SeMet), the XopD expression vector was transformed into B834 *E. coli*, which is a methionine auxotroph (Novagen). The



cells were grown in M9 medium supplemented with 125 mg/liter each of adenine, uracil, thymine, and guanosine nucleotide, 2.5 mg/liter thiamine, 4 mg/liter D-biotin, 20mM glucose, 2mM MgSO<sub>4</sub>, 50 mg/liter L-selenomethionine (SeMet) (Sigma). The SeMet XopD protein was purified as described above.

All purified proteins were analyzed by SDS-PAGE and quantified using a modified version of the Lowry procedure (63).

### **Protein Crystallization**

Wild-type XopD<sub>335-520</sub> and XopD<sub>335-520</sub> C470A crystals were grown by hanging drop vapor diffusion using 24-well plates (65). Index crystallization kits from Hampton Research were used to screen preliminary conditions. For this preliminary screening, wild-type XopD<sub>335-520</sub> protein that was only purified through the ion exchange step (not get filtration) was used. Only one condition of 96 total yielded crystals. Crystals were obtained using 1  $\mu$ L of purified protein at a concentration of 15mg/mL with 1  $\mu$ L of buffer containing 1.4 - 1.6 M sodium potassium phosphate at pH 7.5 - 7.75 at 20°C. The crystals were improved by purifying the protein on a gel filtration column after the ion exchange column prior to crystal growth. SeMet crystals were obtained under similar conditions but included the addition of 5 mM DTT or 10 mM  $\beta$ ME to the protein preparation. Crystals typically appeared in 2-7 days. The cryoprotectant solution used to freeze the crystals was 1.6 M sodium potassium phosphate pH 7.75 with 15% ethylene glycol. Crystals were flash-cooled in liquid propane and then stored under liquid nitrogen.

## Data Collection, Structure Determination, and Refinement

Diffraction data were collected at 100 K using synchrotron radiation at beamline 19-ID of the Structural Biology Center at the Advanced Photon Source (Argonne National Laboratory). The data sets were indexed, integrated, and scaled using the HKL2000 program package (66). The XopD<sub>335-520</sub> protein crystallized in space group P4<sub>1</sub>2<sub>1</sub>2 and had the following unit cell parameters:  $a = 91.618 \text{ \AA}$ ,  $b = 91.618 \text{ \AA}$ ,  $c = 44.820 \text{ \AA}$ ,  $\alpha = \beta = \gamma = 90^\circ$ . Data were collected from the selenomethionyl-substituted protein at the peak energy for anomalous absorption of selenium for structure solution via the single wavelength anomalous dispersion (SAD) method. Location of heavy atom sites and phase calculation and refinement was performed in the program Solve (67). Density modification was performed in the program dm (68). The program arp/Warp (69) automatically built 150 of 186 total residues, and assigned sequence to 122 of the 150 residues built. Further model building was performed manually in the program O (70). The model was refined using simulated-annealing, minimization, and individual B-factor refinement in the program CNS (65). Water molecules were added where stereochemically reasonable using the program CNS. The model of the wild-type XopD protein included residues 339-514 (missing 335-338, 406-407, 432-433, 465-468, and 512-520), 82 water molecules and 1 phosphate ion.

The XopD<sub>335-520</sub> C470A protein crystallized in space group P4<sub>1</sub>2<sub>1</sub>2 and had the following unit cell parameters:  $a = 91.953 \text{ \AA}$ ,  $b = 91.953 \text{ \AA}$ ,  $c = 45.013 \text{ \AA}$ ,  $\alpha = \beta = \gamma = 90^\circ$ . This structure was isomorphous to the wild type; refinement in the program

CNS followed a similar protocol to the wild type structure. The model of the XopD<sub>335-520</sub> C470A protein included residues 336-513 (missing 335, 466-467, and 514-520) and 10 water molecules. Alternate conformations were modeled using the O program. Data collection and refinement statistics are detailed in Table 3 of Appendix B.

### Structure Figures

All figures of XopD and yeast Ulp1 structures were constructed using the MacPyMOL program (71).

### Accession Numbers

*At*SUMO-1(At4g26840),      *At*SUMO-2(At5g55160),      *At*SUMO-3(At5g55170),  
*At*SUMO-5(At2g32765),      *At*ULP1A(At3g06910),      *At*ULP1C(At1g10570),  
*At*ULP1D(At1g60220), and *At*ESD4(At4g15880)

## CHAPTER FOUR

### Results

#### *IN VITRO* CHARACTERIZATION OF *XANTHOMONAS* XopD

##### Introduction

Recently, Noel *et al.* (2002) reported two novel TTSS effector proteins in *Xcv* (1). Upon further inspection, it was discovered that one of the effectors, XopD, is a homologue of the ubiquitin-like protein protease-1, Ulp1, substrates of which are SUMO-conjugated proteins (3). This striking homology supports the hypothesis that a suite of *Xcv* TTSS effectors may be working *in planta* to disrupt the steady state levels of SUMO-conjugated proteins during infection.

SUMO is a member of a family of ubiquitin-related proteins that are covalently attached to eukaryotic proteins by a conjugation system that operates similarly to the ubiquitin conjugation system (5,12). SUMO and SUMO proteases dynamically regulate protein conjugation and deconjugation of proteins and thus regulate a number of cellular processes. For example, SUMO modification controls nuclear transport, signal transduction, cell cycle progression and the stress response (5). In contrast to ubiquitin, SUMO stabilizes proteins and can be an antagonist of ubiquitin. It is thus becoming even more apparent that post-translational regulation of proteins by ubiquitin and ubiquitin-like proteins is analogous to the post-translational regulation of proteins by phosphorylation and is critical for maintaining cellular signaling and homeostasis (72).

This study shows that XopD, an *Xcv* type III effector protein, encodes an active cysteine protease with plant-specific SUMO substrate specificity. This protein is injected into plant cells by the Hrp TTSS during *Xcv* pathogenesis and is then translocated to the plant nucleus to subnuclear foci (56). XopD's *in planta* substrates are SUMO-conjugated proteins. These studies indicate that XopD mimics an endogenous plant SUMO isopeptidase(s), presumably to interfere with the regulation of host proteins during *Xcv* infection. We propose therefore that the pathogen uses SUMO protein deconjugation as a mechanism to alter plant signal transduction and, ultimately, plant physiology.

## Results

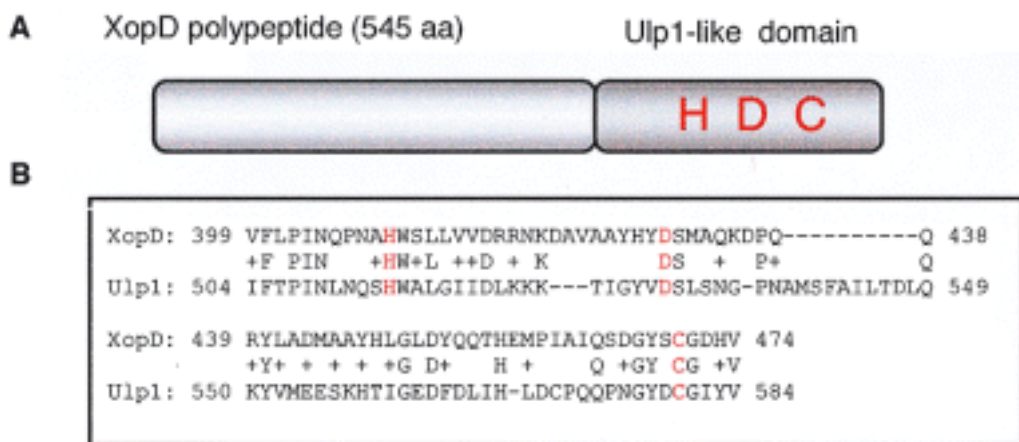
### XopD is an Ulp1 protein homologue

The XopD protein was recently identified as a novel type III secreted protein from *Xcv* (1). XopD shares sequence similarity with the virulence factor PsvA from *Pseudomonas syringae* pv. *eriobotryae*; however, homologous proteins with known function were not reported (1). Using BLAST and PROSITE analysis, it was found that the C-terminus of XopD (amino acids 322–520) shared primary sequence similarity with the C48 family of cysteine peptidases (Fig. 4A) (73,74). Moreover, amino acid positions 309–481 in the XopD polypeptide are most homologous to the C-terminal catalytic domain of the Ulp1 ubiquitin-like protease protein family (3).

Ulp1 is a cysteine protease that catalyzes two critical reactions in the SUMO/Smt3 pathway. In the first, Ulp1 processes the invariant C-terminal sequence (–GG XXX) of

SUMO to its mature form (–GG) so that the terminal glycine residue can be used as a substrate to modify target proteins covalently (5). In the second, Ulp1 deconjugates SUMO from its target protein by cleaving the isopeptide bond that links the C-terminus of SUMO to the epsilon-amine of a lysine residue on the target protein (5).

Alignment of the C-terminal domains of XopD and Ulp1 revealed that the residues comprising the putative catalytic core in the XopD cysteine protease are H409, D421 and C470 (Fig. 4B). Based on information from the co-crystal structure of Ulp1 and SUMO, we observed that residues important for substrate recognition are also conserved in XopD (amino acids 338–341) (11). Analysis of the N-terminus of XopD failed to reveal significant homology to any protein in existing databases possessing a known functional domain or structural motif.



**Figure 4: Features of the *Xcv* XopD polypeptide.** A. Schematic representation of XopD. XopD encodes a 545-amino-acid protein. Amino acids 309-481 encompass the Ulp1-like protease domain. The putative catalytic core residues are His-409, Asp-421 and Cys-470, depicted as H, D, and C. B. Sequence alignment of amino acids in the catalytic core of yeast Ulp1 and *Xcv* XopD.

Based on this striking sequence homology, it was hypothesized that XopD encodes an ubiquitin-like protein protease that functions *in planta* during *Xcv* pathogenesis. Previously, it was predicted that YopJ-like effector proteins in *Xcv* might similarly encode cysteine proteases with SUMO substrate specificity (75). Although the YopJ-like effectors, including AvrBsT, AvrRxv and AvrXv4, share limited sequence similarity with amino cysteine proteases including Ulp1 and adenovirus protease AVP, the catalytic domain of XopD possesses extensive sequence similarity with the catalytic core of Ulp1 (Figure 4) (2,4,75,76). Sequence and structural analyses categorize AVP, Ulp1 and XopD, and YopJ and YopJ-like effectors in the CE clan of cysteine proteases characterized by a cysteine nucleophile and a catalytic core with the ordered sequence: H, E or D, and C. Within the CE clan of cysteine proteases, these proteins have been assigned to the C5, C48 and C55 peptidase families, respectively (2,4). Thus, *Xcv* expresses two different families (C48 and C55) of cysteine proteases that possibly function as SUMO protease-like effectors.

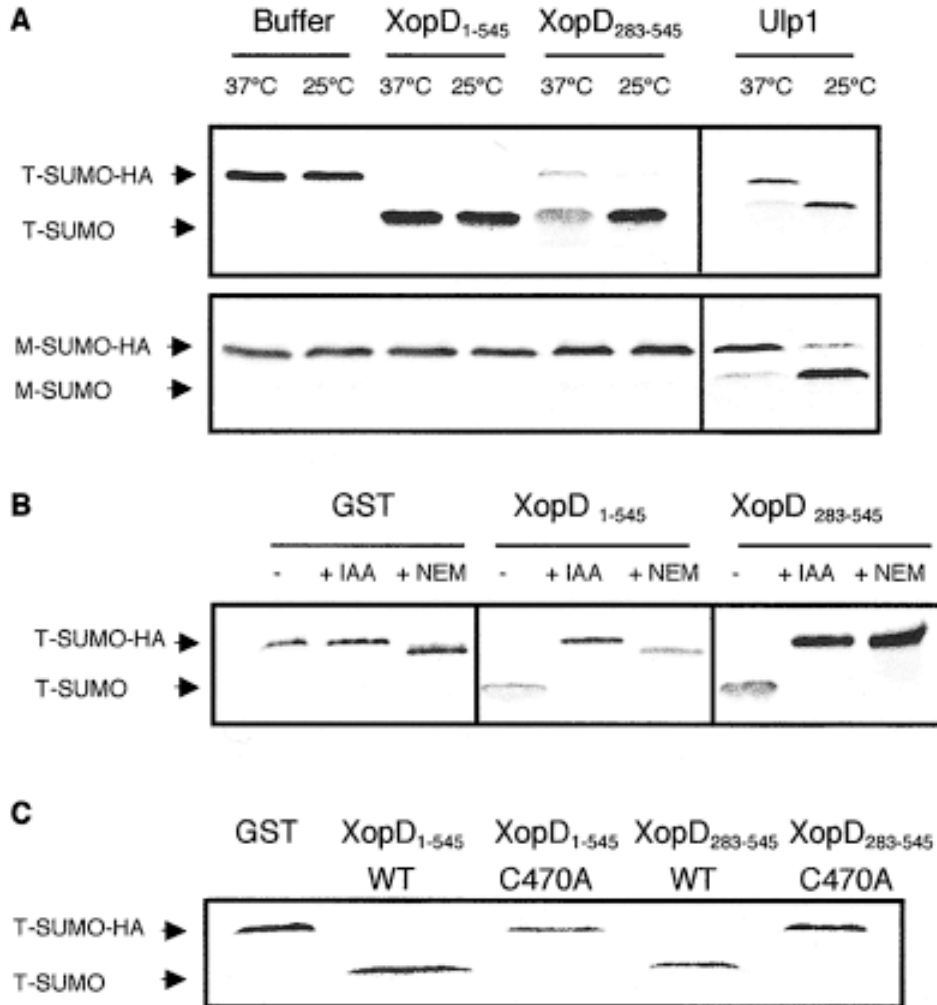
### **XopD possesses Ulp1-like protease activity**

To determine whether XopD possessed Ulp1-like protease activity, the *xopD* gene was cloned from *Xcv* strain 85-10 and constructed wild-type and mutant GST–XopD fusion proteins. GST–XopD and GST–Ulp1 fusion proteins were expressed in *Escherichia coli* BL21 cells, purified using standard GST–glutathione affinity chromatography and then analyzed with *in vitro* protease assays using as substrates <sup>35</sup>S-labelled tomato and mammalian SUMO with a carboxy-terminal HA-epitope tag. [<sup>35</sup>S]-T-SUMO-HA was incubated with buffer, GST–XopD<sub>1-545</sub>, GST–XopD<sub>285-545</sub> or GST–Ulp1 for 60 min and then analyzed by

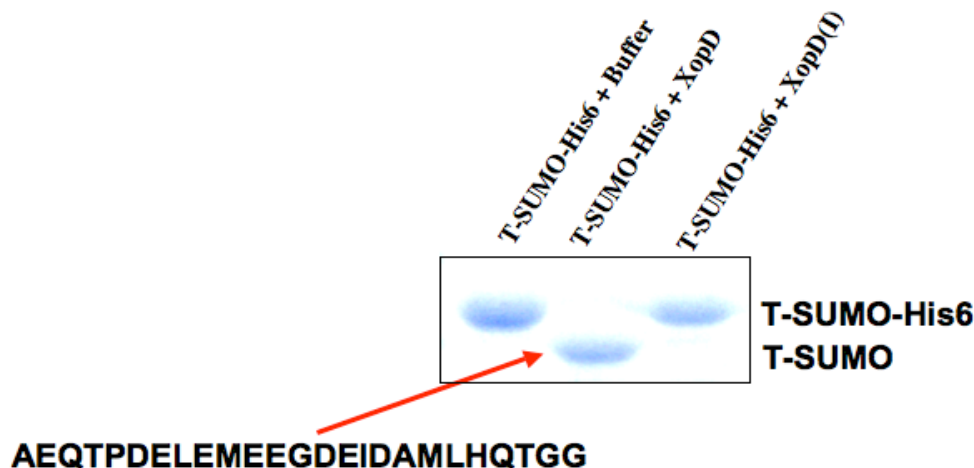
SDS-PAGE and autoradiography.

Wild-type *Xcv* GST–XopD<sub>1-545</sub> and yeast GST–Ulp1 cleaved the [<sup>35</sup>S]-T-SUMO-HA substrate (SUMO-GlyGly-HA) *in vitro* after the invariant C-terminal –GlyGly residues generating [<sup>35</sup>S]-T-SUMO (Fig. 5A). GST–XopD<sub>285-545</sub>, a fusion protein containing only the Ulp1-like domain, also cleaved the [<sup>35</sup>S]-T-SUMO-HA substrate. We confirmed that XopD cleaves the tomato SUMO polypeptide at the predicted cleavage site (after the invariant C-terminal –GlyGly residues) using tandem mass spectrometry. Recombinant T-SUMO-His<sub>6</sub> purified from *E. coli* was incubated with purified GST–XopD<sub>285-545</sub> *in vitro*, and then the amino acid composition of the XopD-cleaved T-SUMO and uncleaved T-SUMO-His<sub>6</sub> polypeptides was determined (Figure 6). Only T-SUMO-His<sub>6</sub> substrate incubated with purified XopD was cleaved after –GlyGly. These studies demonstrate that *Xcv* XopD possesses SUMO peptidase activity analogous to that observed for the yeast Ulp1 enzyme. SUMO peptidase activity is specifically localized to the predicted catalytic domain of XopD that shares homology with the catalytic domain of Ulp1. Moreover, the GST–XopD<sub>285-545</sub> enzyme is constitutively active and does not require any eukaryotic factors for its enzymatic activity.





**Figure 5: Xcv XopD is a plant-specific SUMO cysteine protease.** A. XopD exhibits peptidase activity similar to yeast Ulp1. [<sup>35</sup>S]-Tomato-SUMO-HA and [<sup>35</sup>S]-Mammalian-SUMO-HA were translated *in vitro* in a rabbit reticulocyte lysate and then incubated with buffer, 0.2 mg/ml GST-XopD<sub>1-545</sub>, 0.2 mg/ml GST-XopD<sub>283-545</sub> or 0.4 mg/ml GST-Ulp1 for 1 hour at the indicated temperatures. SUMO species were resolved on a 15% SDS gel and visualized by autoradiography. B. XopD enzyme activity is inactivated by cysteine protease inhibitors. [<sup>35</sup>S]-Tomato-SUMO-HA translated *in vitro* in a rabbit reticulocyte lysate was incubated for 1 h at 30°C with 0.2 mg/ml GST, 0.2 mg ml<sup>-1</sup> GST-XopD<sub>1-545</sub> or 0.2 mg/ml GST-XopD<sub>283-545</sub> treated with 10 mM IAA or 5 mM NEM. SUMO species were resolved on a 15% SDS gel and visualized by autoradiography. C. XopD C470A mutants are inactive proteases. [<sup>35</sup>S]-Tomato-SUMO-HA translated *in vitro* in a rabbit reticulocyte lysate was incubated with 0.2 mg/ml GST, 0.2 mg/ml GST-XopD<sub>1-545</sub>, 0.2 mg/ml GST-XopD<sub>1-545</sub> C470A, 0.2 mg/ml GST-XopD<sub>283-545</sub> or 0.2 mg/ml GST-XopD<sub>283-545</sub> C470A for 1 hour at 30°C. SUMO species were resolved on a 15% SDS gel and visualized by autoradiography.



**Figure 6: XopD cleaves T-SUMO-His<sub>6</sub> after its C-terminal Gly-Gly motif.** Recombinant GST-XopD<sub>285-545</sub> protein was incubated with recombinant T-SUMO-His<sub>6</sub> for one hour at 30°C. GST-XopD<sub>285-545</sub> was also treated with 10 mM IAA and then incubated with recombinant T-SUMO-His<sub>6</sub> for one hour at 30°C. Proteins samples were then run on a 15% SDS-PAGE gel. Protein bands were excised from the Colloidal blue-stained SDS gels and digested with trypsin. The tryptic peptides were extracted and analyzed by nanoelectrospray tandem mass spectrometry using methods similar to those described previously (64). Proteins and peptides were identified by database searching of the MS and MS/MS spectra against the NCBI non-redundant databases using KNEXUS software.

### **XopD is a plant-specific SUMO cysteine protease**

Further characterization of the activity of XopD was studied by varying the substrate and temperature. GST-XopD<sub>1-545</sub> and GST-XopD<sub>285-545</sub> were incubated *in vitro* with a mammalian SUMO-1 substrate, [<sup>35</sup>S]-M-SUMO-HA. Interestingly, [<sup>35</sup>S]-M-SUMO-HA was not cleaved by GST-XopD<sub>1-545</sub> or GST-XopD<sub>285-545</sub> (Fig. 5A), whereas yeast Ulp1 protease was able to cleave both the tomato and the mammalian SUMO-HA substrates. The specificity of XopD was not temperature dependent over physiological ranges, as XopD was able to cleave tomato SUMO-HA at 37°C and 25°C, growth temperatures for some animal

and plant hosts respectively (Fig. 5A). Although Ulp1 was able to cleave both tomato and mammalian SUMO-HA, the cleavage of tomato SUMO-HA by Ulp1 was more efficient at 25°C. These studies indicate that *Xcv* XopD has plant-specific SUMO substrate specificity, whereas yeast Ulp1 has broader host substrate specificity.

Ubiquitin-like proteases in the C48 family are cysteine peptidases with activity that depends on a functional catalytic triad (H, D, C) (2,4). Members of this family are inactivated by thiol protease inhibitors, including iodoacetamide (IAA) and N-ethylmaleimide (NEM), or by single amino acid substitutions within the catalytic core.

The next experiments explored whether XopD, like Ulp1, was acting as a cysteine protease. GST, GST-XopD<sub>1-545</sub> and GST-XopD<sub>285-545</sub> were incubated with [<sup>35</sup>S]-T-SUMO-HA at 25°C in the presence and absence of IAA and NEM. The GST-XopD<sub>1-545</sub> and GST-XopD<sub>285-545</sub> enzymes were specifically inhibited by IAA and NEM and were unable to cleave the T-SUMO-HA substrate (Fig. 5B). We consistently observed that NEM treatment alone slightly altered the migration of the T-SUMO-HA substrate whether or not enzyme was added. Furthermore, mutation of cysteine 470 to an alanine residue in the predicted XopD catalytic core completely inactivated the XopD peptidase activity (Fig. 5C). GST-XopD<sub>1-545</sub> C470A and GST-XopD<sub>285-545</sub> C470A were not able to cleave the T-SUMO-HA substrate. Collectively, these observations demonstrate that *Xcv* XopD encodes a plant-specific SUMO cysteine protease with peptidase activity.

### **XopD possesses plant SUMO isopeptidase activity**

To elucidate the role of XopD in pathogenesis, we next determined whether XopD

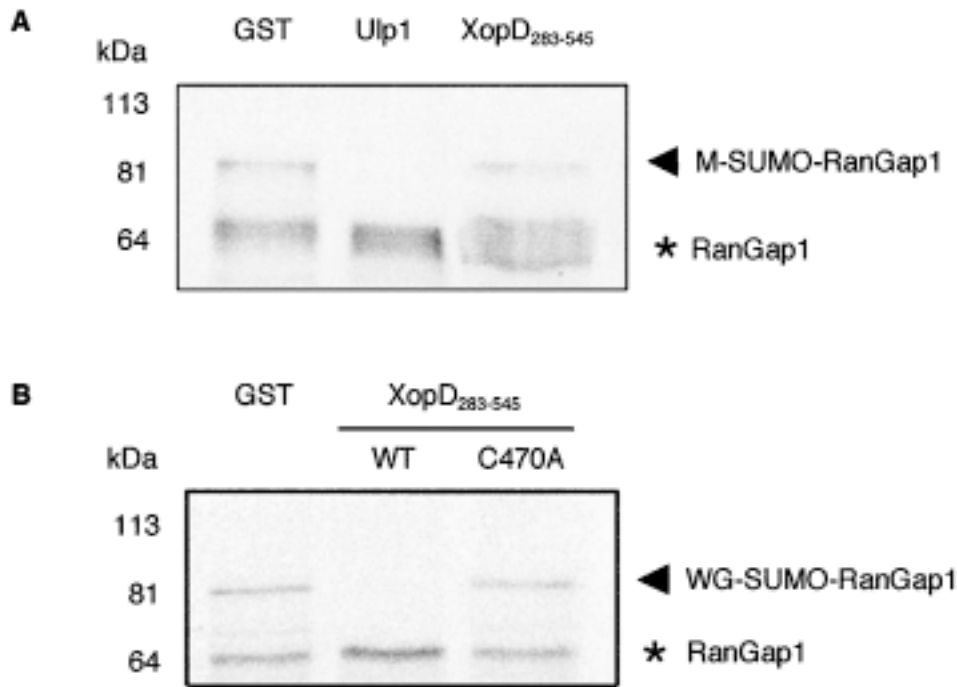
recognized plant proteins post-translationally modified by SUMO. In contrast to yeast and mammalian systems, little is known about the proteins that are modified by SUMO in plants. Thus, the extent of SUMO protein conjugation *in planta* using the *Agrobacteria* transient expression system to express tomato HA-SUMO in *N. benthamiana* was determined (56). Total protein was extracted from infected *N. benthamiana* leaves 48 h after inoculation and analyzed by immunoblot analysis using antisera specific for the HA epitope. Free tomato HA-SUMO and multiple HA-SUMO-conjugated proteins accumulated in the plant cells transiently expressing T-HA-SUMO, demonstrating that a number of plant proteins undergo SUMO post-translational modification in *N. benthamiana* (56).

The plant SUMO-conjugated proteins were used as substrates for XopD *in vitro*. Plant HA-SUMO-conjugated proteins were generated as described above and then used as substrates for GST-XopD<sub>1-545</sub>, GST-XopD<sub>1-545</sub> C470A and GST-Ulp1. GST-XopD<sub>1-545</sub> was able to hydrolyze the SUMO substrates resulting in a significant reduction in the HA-SUMO-conjugated proteins (56). The level of free HA-SUMO and total plant protein was not significantly altered, indicating that XopD action did not cause random protein degradation (56). GST-Ulp1 also recognized the plant SUMO substrates; however, GST-Ulp1 isopeptidase activity was less efficient than GST-XopD<sub>1-545</sub>. Mutant GST-XopD<sub>1-545</sub> C470A, as well as the GST control, exhibited no isopeptidase activity. Conversely, the mutant GST-XopD<sub>1-545</sub> C470A protein often stabilized SUMO-conjugated proteins relative to the GST alone control (56). These studies confirm that XopD possesses plant SUMO isopeptidase activity in addition to SUMO peptidase activity.

Considering that XopD functions as both a peptidase and an isopeptidase, the effect of XopD action on SUMO post-translational modification was assessed *in planta*. XopD-HA and T-HA-SUMO were co-expressed transiently in *N. benthamiana* leaves, and then HA-SUMO conjugates were analyzed by immunoblot analysis. Wild-type XopD-HA expression *in planta* led to a significant reduction in SUMO-modified proteins (56). The level of free HA-SUMO was not significantly altered indicating that XopD action *in planta* was not affecting the expression or accumulation of HA-SUMO. Not all potential SUMO substrates were hydrolyzed, indicating that XopD did not completely reduce protein SUMOylation. The mutant XopD-HA C470A was not able to hydrolyze the SUMO substrates (56). As with the *in vitro* analysis, we observed an increase in SUMO-conjugated proteins when XopD-HA C470A was co-expressed with T-HA-SUMO *in vivo* (56).

The *in planta* analysis revealed that XopD targets a number of SUMO-modified proteins and may therefore affect several plant signaling pathways during *Xcv* infection. Because individual plant proteins modified by SUMO have not yet been reported, we tested a major target of SUMO in higher eukaryotes, RanGAP1 (14). Initially, reticulocyte lysate was used to produce *in vitro*-translated, radiolabelled mammalian RanGAP1, which is known to produce both the [<sup>35</sup>S]-RanGAP1 protein and the [<sup>35</sup>S]-M-SUMO-RanGAP1 conjugate (15). When the translation reaction was incubated with purified Ulp1, the [<sup>35</sup>S]-M-SUMO-RanGAP1 band disappeared, indicative of isopeptidase activity by Ulp1 (Fig. 7A) (3). However, no isopeptidase activity was observed when the reticulocyte lysate translation reaction was incubated with GST or purified GST-XopD<sub>285-534</sub> (Fig. 7A). These results were not surprising based on our previous observations that XopD was able to cleave tomato

SUMO but unable to recognize mammalian SUMO as a substrate (Fig. 5A). Therefore, to test whether XopD has isopeptidase activity on a plant-derived SUMO-RanGAP1 substrate, we used a wheat germ lysate to translate mammalian RanGAP1 *in vitro*. We observed that the wheat germ extract, like the reticulocyte lysate, contains the enzymatic machinery that is sufficient to SUMOylate mammalian RanGAP1 (Fig. 7B). Wheat germ-generated [<sup>35</sup>S]-SUMO-RanGAP1 was then incubated with purified GST, GST-XopD XopD<sub>285-534</sub>, or GST-XopD XopD<sub>285-534</sub> C470A for 1 h, and we observed that incubation with GST-XopD XopD<sub>285-534</sub>, but not with GST or GST-XopD XopD<sub>285-534</sub> C470A, led to the disappearance of [<sup>35</sup>S]-WG-SUMO-RanGAP1. These *in vitro* results support our *in vivo* observations that XopD is an isopeptidase that can hydrolyze plant SUMO protein conjugates. Moreover, these results confirm that XopD is an isopeptidase, as well as a peptidase, that contains an activity that is plant SUMO specific.



**Figure 7: XopD hydrolyses WG-SUMO-RanGAP1 and not M-SUMO-RanGAP1 *in vitro*.** A. [ $^{35}\text{S}$ ]-mammalian-RanGap1 was translated *in vitro* in a rabbit reticulocyte lysate and then incubated with 0.2 mg/ml GST, 0.4 mg/ml GST-Ulp1 or 0.2 mg/ml GST-XopD<sub>283-545</sub> for 1 hour at 30°C. Products were resolved on a 10% SDS gel and visualized by autoradiography. M-SUMO-RanGAP1 (arrow) and RanGAP1 (asterisk) are indicated. B. [ $^{35}\text{S}$ ]-mammalian-RanGap1 was translated *in vitro* in a wheat germ lysate and then incubated with 0.2 mg/ml GST, 0.2 mg/ml GST-XopD<sub>283-545</sub> or 0.2 mg/ml GST-XopD<sub>283-545</sub> C470A for 1 h at 30°C. Products were resolved on a 10% SDS gel and visualized by autoradiography. WG-SUMO-RanGAP1 (arrow) and RanGAP1 (asterisk) are indicated.

## Discussion

These studies reveal that the *Xcv* XopD effector encodes an active Ulp1-like cysteine protease that functions inside plant cells to cleave SUMOylated proteins. Two other studies have linked SUMO and SUMO-like proteases to signal transduction events controlling plant defense responses revealing that the SUMO conjugation system may be a key target of plant

pathogens (10,75). Although SUMO protein targets in plant cells await discovery, a recent report revealed that the SUMO conjugation systems operate in plants (12). Characterization of the SUMO pathway in *Arabidopsis* shows that this conjugation system is more complex than any other system characterized to date. Bioinformatic studies predict that *Arabidopsis* uses eight SUMOs and 12 SUMO proteases, revealing that SUMOylation and deSUMOylation of proteins are pivotal regulatory steps in protein signal transduction in plants (12). A diverse array of SUMO-conjugated proteins exists *in planta*, and some of these are SUMO isoform specific. Moreover, SUMO protein conjugation is induced by heat, H<sub>2</sub>O<sub>2</sub>, ethanol and the amino acid analogue canavanine. Rapid and reversible modification by SUMO clearly reveals that plants use SUMOylation as an important regulator of the stress response (12).

Importantly, this work shows that *Xcv* can reduce SUMO protein conjugation *in planta*. Furthermore, the key substrate for XopD is plant SUMO and not mammalian SUMO, indicating that this enzyme is not a general eukaryotic SUMO protease. The co-crystal structure of Ulp1 and yeast SUMO reveals key residues that are important for enzyme:substrate recognition (11). Some of these contact residues in mammalian SUMO encode charge reversals in tomato SUMO, thereby supporting the proposal that the variability in these residues is playing a role in enzyme substrate specificity. SUMO cleavage by SUMO proteases may be distinct for different SUMO isoforms and/or for SUMOs from different organisms. Similarly, the enzyme structure for distinct SUMO proteases is expected to compensate for distinct SUMO structural features. We hypothesize that the target proteins affected by XopD proteolysis may be plant SUMO isoform specific.



The identification of SUMOylated plant targets will be paramount to understanding the complexity of cellular pathways modulated by these proteases during plant pathogenesis.

A survey of the MEROPS Protease Database found that proteins containing a protease domain similar to XopD (i.e. members of the C48 peptidase family) are found only in eukaryotic genomes and in bacteria known to associate with plant hosts (2,4). The plant pathogens *Xanthomonas campestris* pv. *campestris* and *Pseudomonas syringae* pv. *eriobotryae* and the plant symbiont *Mesorhizobium loti* each possess proteins with the invariant H/D/C residues in their putative protease domains, suggesting that these proteins are also cysteine proteases. HopPmaI from the plant pathogen *Pseudomonas syringae* pv. *maculicola* shares limited homology with the N-terminal domain of XopD but lacks the protease domain. Although HopPmaI is not predicted to be a protease, the N-terminal domain shared by HopPmaI and XopD may play a significant, yet to be identified, role in host-pathogen interactions.

The XopD effector does not appear to be essential for pathogenicity considering that an *Xcv* mutant strain lacking the *xopD* gene is still virulent on susceptible pepper hosts (1). It is not yet known whether *Xcv* has functionally redundant XopD-like proteins that can compensate for XopD action in *Xcv*. However, this strain of *Xcv* does contain the YopJ-like effector AvrRxv (77) that is predicted to function as a SUMO protease (76). XopD may play a role in the establishment of *Xcv* colonization in the plant apoplast by modulating SUMO-dependent signal transduction, possibly defense signaling, in plant cells. It is noteworthy that the aggressiveness of the plant pathogen *Ralstonia solanacearum* is modulated by the host

specificity factor PopP1, a protein with homology to YopJ-like effectors (78). The precise contribution of the XopD effector in *Xcv* pathogenesis remains to be established.

Leaves accumulating the wild-type EYFP–XopD active protease in subnuclear foci died by 7 days after transient infection (56). Conversely, tissue necrosis was never observed for leaves infected with the mutant EYFP–XopD C470A protein (56). Considering that *N. benthamiana* plants do not recognize the XopD effector (i.e. the plant is susceptible), the leaf necrosis observed is probably the consequence of XopD-specific proteolysis within the nucleus rather than a defense response triggered by the host (e.g. the hypersensitive cell death response). The deconjugation of SUMO targets in the nucleus is thus likely to play a central role in the control of a number of cellular processes. XopD's SUMO protease activity and its localization to subnuclear foci indicate that *Xcv*, and possibly other pathogens, can disrupt the regulation of several proteins modified by SUMO, interfering with the homeostasis of plant cellular signaling.

The identification of specific SUMO targets in the plant nucleus may permit the elucidation of the biological consequence of XopD proteolysis during *Xcv* pathogenesis. So far, only a few plant proteins have been shown to be localized to subnuclear foci. It is intriguing to speculate that one of these proteins is a XopD substrate. However, none of these proteins has been shown to be SUMOylated despite the fact that LAF1 and COP1 contain sequences similar to the minimal SUMO-specific consensus sequence (5). We speculate that some of these proteins may be affected by XopD action in the nucleus considering that light-induced signaling pathways control nuclear localization (79-81) and

interact with pathogen-dependent salicylic acid signal transduction pathways controlling defense responses in *Arabidopsis* (82).

Overall, these studies support a role for the XopD effector during *Xcv*–plant interactions. This study shows that XopD is an active cysteine protease with plant-specific SUMO specificity. This is the first evidence demonstrating a functional role for a phytopathogenic bacterial TTSS effector *in planta*. Moreover, these studies reveal a novel mechanism used by *Xanthomonas*, and possibly other plant-associated microbes, to modulate plant physiology during infection. Similar to what has been observed for other bacterial effector proteins (e.g. YopH, YopE, YopJ and YpkA), we hypothesize that *Xcv* has usurped the activity of a eukaryotic isopeptidase and uses this activity to alter directly SUMO protein targets in the plant nucleus that control plant susceptibility and/or plant defense (72,83). Based on these observations, we can explore candidate proteins that are regulated by plant SUMO modification and study their importance in plant–microbe interactions.

## **CHAPTER FIVE**

### **Results**

#### **STRUCTURAL ANALYSIS OF XopD PROVIDES INSIGHTS INTO ITS SUBSTRATE SPECIFICITY**

##### **Introduction**

The deSUMOylating enzyme *Xanthomonas* XopD, exhibits species specificity for its SUMO substrates based on the findings described in Chapter 4. XopD is unable to recognize mammalian SUMO substrates, but is able to recognize and process Tomato SUMO (T-SUMO). Unlike yeast Ulp1, which is promiscuous in its choice of SUMO substrate, XopD demonstrates specificity for the SUMO substrates it processes. By comparing the sequence alignments of several SUMOs from plants, human, and yeast, one can predict which residues in SUMO are dictating enzyme specificity. To further understand how XopD recognizes T-SUMO as a substrate, the crystal structure of XopD and the structure of the catalytically inactive mutant of XopD was solved. The goal of these studies is to use structural and mutational studies to better understand how XopD recognizes and processes its SUMO substrates.

##### **Results and Discussion**

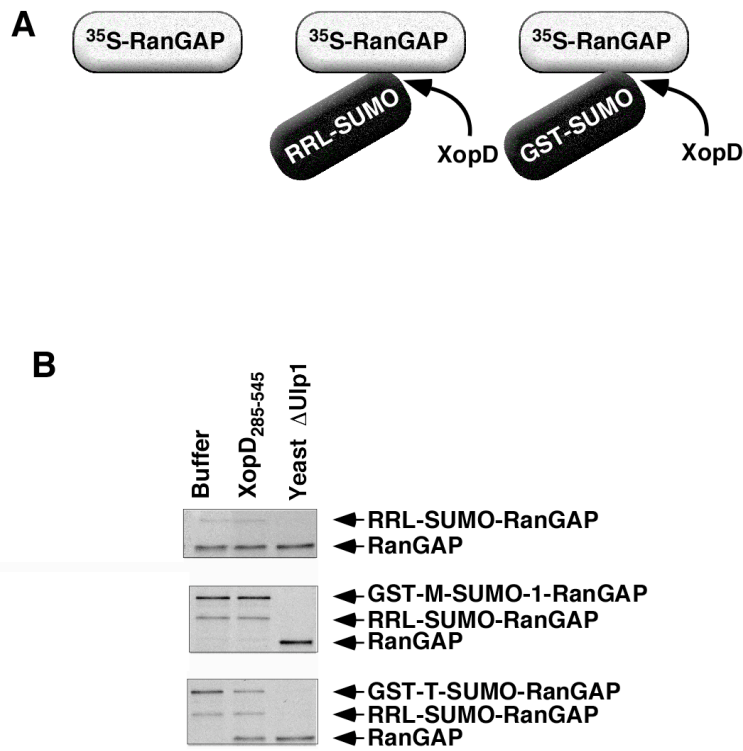
###### *Substrate Specificity of XopD*

As was shown in Chapter 4, XopD is able to process T-SUMO, but not M-SUMO-1 in peptidase assays, and is able to process wheat germ SUMO-modified RanGAP in isopeptidase assays. Using an *in vitro* SUMOylation system, the substrate specificity of XopD was further confirmed by showing that XopD can process RanGAP that has been modified by GST-T-SUMO, and not RanGAP that has been modified by GST-M-SUMO-1 (Figure 8). Thus XopD can process T-SUMO in peptidase assays as well as in isopeptidase assays when T-SUMO is conjugated to a target protein.

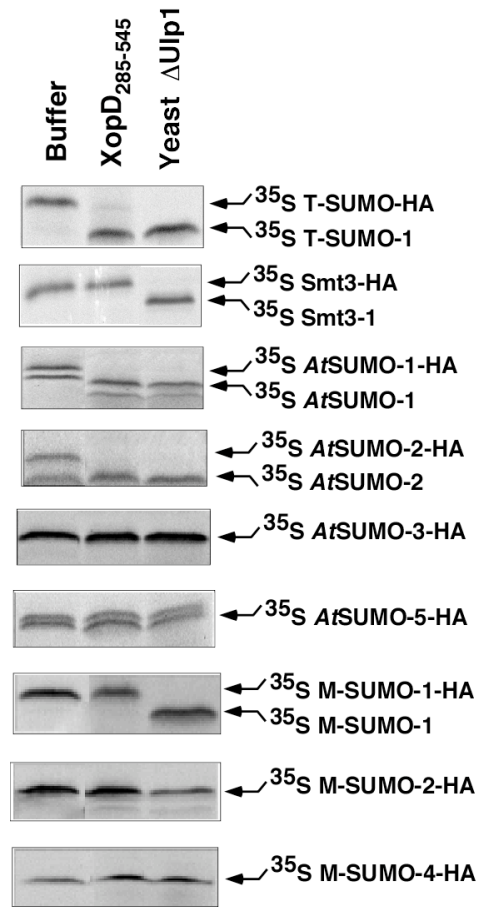
To further explore the SUMO substrate specificity of XopD, several *Arabidopsis thaliana* (*At*) SUMOs, mammalian and yeast SUMO were selected to determine if XopD and yeast Ulp1 would process these SUMOs using *in vitro* peptidase assays. XopD preferentially recognized T-SUMO, *At*SUMO-1, and *At*SUMO-2 from a panel of SUMO substrates (Figure 9).

By observing sequence alignments of the SUMOs that XopD and yeast Ulp1 can and cannot process, one can begin to determine which residues are important for XopD's SUMO specificity (Figure 10). Residues at the P<sub>5</sub>-P<sub>7</sub> positions in the SUMO substrates appear to be important for the recognition by XopD. T-SUMO, *At*SUMO-1, and *At*SUMO-2 share the same residues, Met-Leu-His, at the P<sub>5</sub>-P<sub>7</sub> positions. These three SUMOs are the only SUMOs processed by XopD in the peptidase assays. In contrast, Smt3 and M-SUMO-1, which are both processed by yeast Ulp1, contain a His-Arg-Glu and Tyr-Gln-Glu, respectively, at the P<sub>5</sub>-P<sub>7</sub> positions. The charge differences in this region may account for some of the specificity by which XopD recognizes its SUMO substrates. By looking at the co-crystal structure of yeast Ulp1 with Smt3 solved by Mossessova and Lima, the residues at

the P<sub>5</sub>-P<sub>7</sub> positions (H92, R63, and E94) of Smt3 appear to contact yeast Ulp1 (Figure 11) (11). Since the residues at P<sub>5</sub>-P<sub>7</sub> positions are part of the contact region between yeast Ulp1 and Smt3, it is likely that residues at the P<sub>5</sub>-P<sub>7</sub> positions are playing a role in how XopD is recognizing T-SUMO.



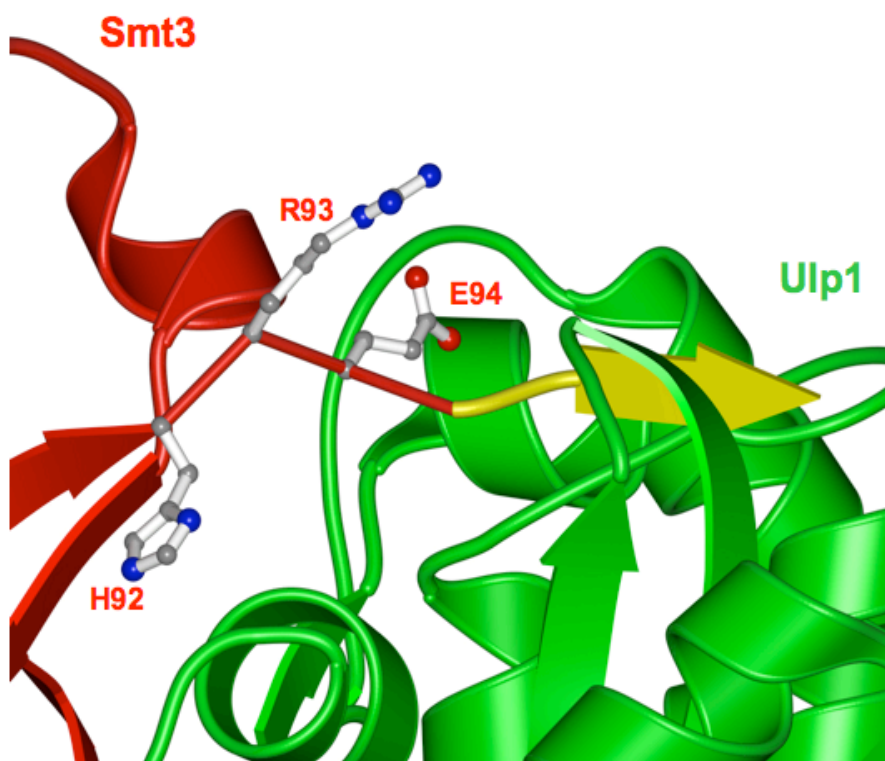
**Figure 8: XopD exhibits isopeptidase activity for T-SUMO modified RanGAP, but not M-SUMO-1 modified RanGAP.** A. Schematic of *in vitro* SUMO isopeptidase assay. B. XopD<sub>285-545</sub> shows specificity for the SUMO moiety of SUMOylated substrates. [<sup>35</sup>S]-mammalian RanGAP was *in vitro* translated in a rabbit reticulocyte lysate. In the rabbit reticulocyte lysate, some of the RanGAP is SUMOylated by endogenous SUMOylation machinery. The *in vitro* translated product was then used in an *in vitro* SUMOylation assay using recombinantly purified GST-M-SUMO-1 and GST-T-SUMO to produce GST-modified RanGAP. The SUMO-modified RanGAP was then incubated with buffer or 0.5 mg/mL of XopD<sub>285-545</sub> or yeast Ulp1 for 1 hour at 30°C. The samples were then resolved on 8% SDS gels and visualized by autoradiography.



**Figure 9: XopD is a plant-specific SUMO protease.** XopD exhibits peptidase activity similar to yeast Ulp1. [<sup>35</sup>S]-Tomato-SUMO-HA, Smt3-HA, AtSUMO-1-HA, AtSUMO-2-HA, AtSUMO-3-HA, AtSUMO-5-HA, M-SUMO-1-HA, M-SUMO-2-HA, and M-SUMO-4-HA were translated *in vitro* in a rabbit reticulocyte lysate and then incubated with buffer, 0.5 mg/ml GST-XopD<sub>285-545</sub>, or GST-Ulp1 for 1 h at 30°C. SUMO species were resolved on a 15% SDS gel and visualized by autoradiography.



**Figure 10: Sequence alignments of plant, yeast and mammalian SUMO proteins.** Sequence alignments of the C-terminal residues of T-SUMO, Smt3, M-SUMO (-1,-2,-4), and *At*SUMO (-1,-2,-3,-5). Yellow boxes represent the conserved QTGG region of the SUMO proteins. Cyan and magenta boxes represent similarities between SUMOs in the P<sub>5</sub>-P<sub>7</sub> positions. The “+++” symbol represents SUMO proteins that are processed by XopD or Ulp1.

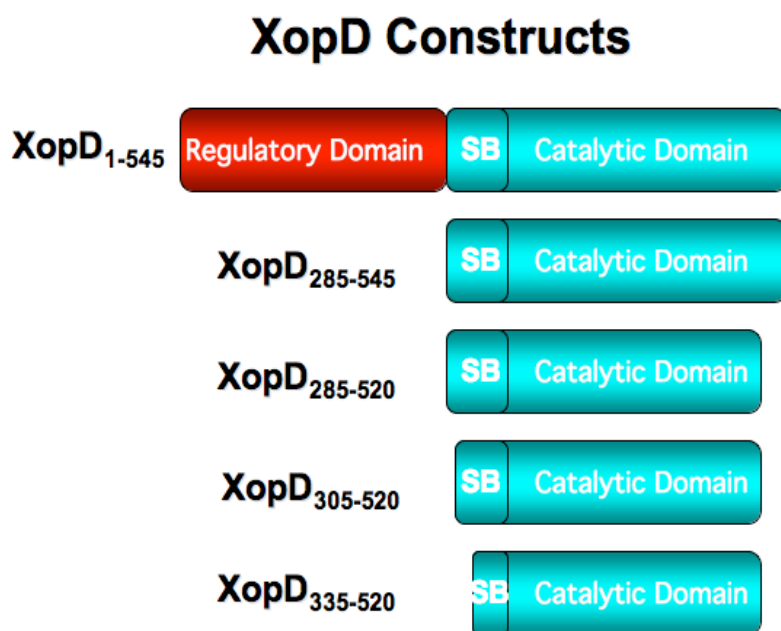


**Figure 11: Contact residues between yeast Ulp1 and Smt3 identified in the co-crystal structure (11).** In the Ulp1:Smt3 structure, Smt3 is drawn in red, Ulp1 in green, and the Gln-Thr-Gly-Gly C-terminus of Smt3 is shown in yellow. Residues H92, R63, and E94 of Smt3 (P<sub>5</sub>-P<sub>7</sub> positions) are in contact with Ulp1.

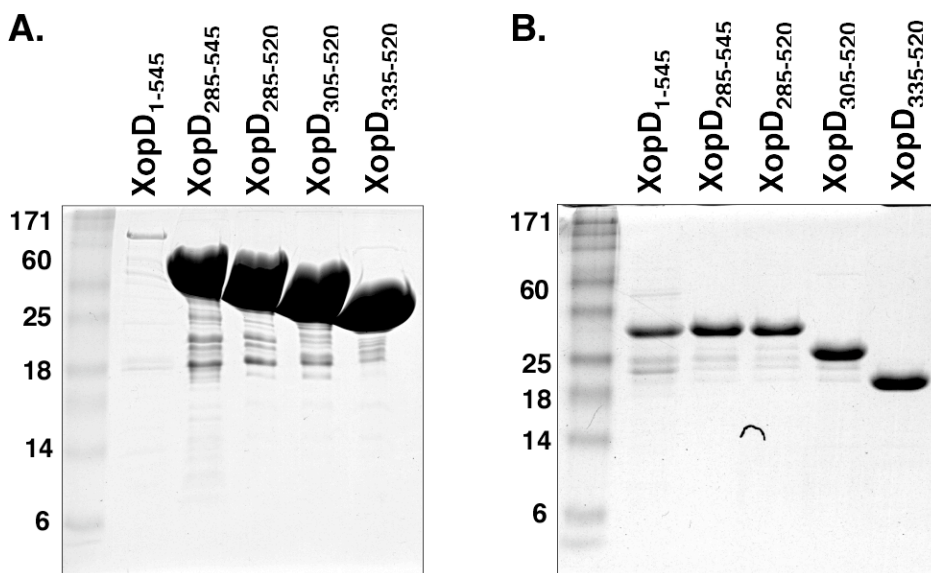
### *SUMO Mutational Studies*

To further understand which portions of SUMO are dictating how XopD is recognizing its substrates, mutation studies were undertaken. We predicted specific mutations in T-SUMO would inhibit XopD from processing T-SUMO. In a similar manner, specific residues mutated in Smt3 should enable XopD to process Smt3. Using the information gained from the SUMO peptidase assays above, site-directed mutagenesis was performed in T-SUMO and Smt3. Mutations were made in the P<sub>5</sub>-P<sub>7</sub> positions of T-SUMO

and Smt3 in order to make this region of T-SUMO look like Smt3, and vice versa. Also, by looking at the contact regions between yeast Ulp1 and Smt3 identified in the co-crystal structure, another mutation further upstream (Ala /Arg) was made to further alter the enzyme specificity (11). Several XopD constructs, XopD<sub>1-545</sub>, XopD<sub>285-545</sub>, XopD<sub>285-520</sub>, XopD<sub>305-520</sub>, XopD<sub>335-520</sub> (full-length XopD and XopD cleavage products) were used in the peptidase assays to determine which portions of XopD are important for SUMO substrate recognition (Figures 12 and 13). All of the XopD protein constructs are expressed with an amino-terminal GST tag. XopD<sub>1-545</sub> is self-cleaved to form XopD<sub>285-545</sub>, lacking the amino-terminal domain, and then to other smaller cleavage products of the carboxy-terminal domain (Figure 13).



**Figure 12: Diagram of all XopD protein constructs used for peptidase assays.** XopD<sub>1-545</sub>, XopD<sub>285-545</sub>, XopD<sub>285-520</sub>, XopD<sub>305-520</sub>, XopD<sub>335-520</sub> (full-length XopD and XopD cleavage products) were used in the peptidase assays to determine which portions of XopD are important for SUMO substrate recognition.

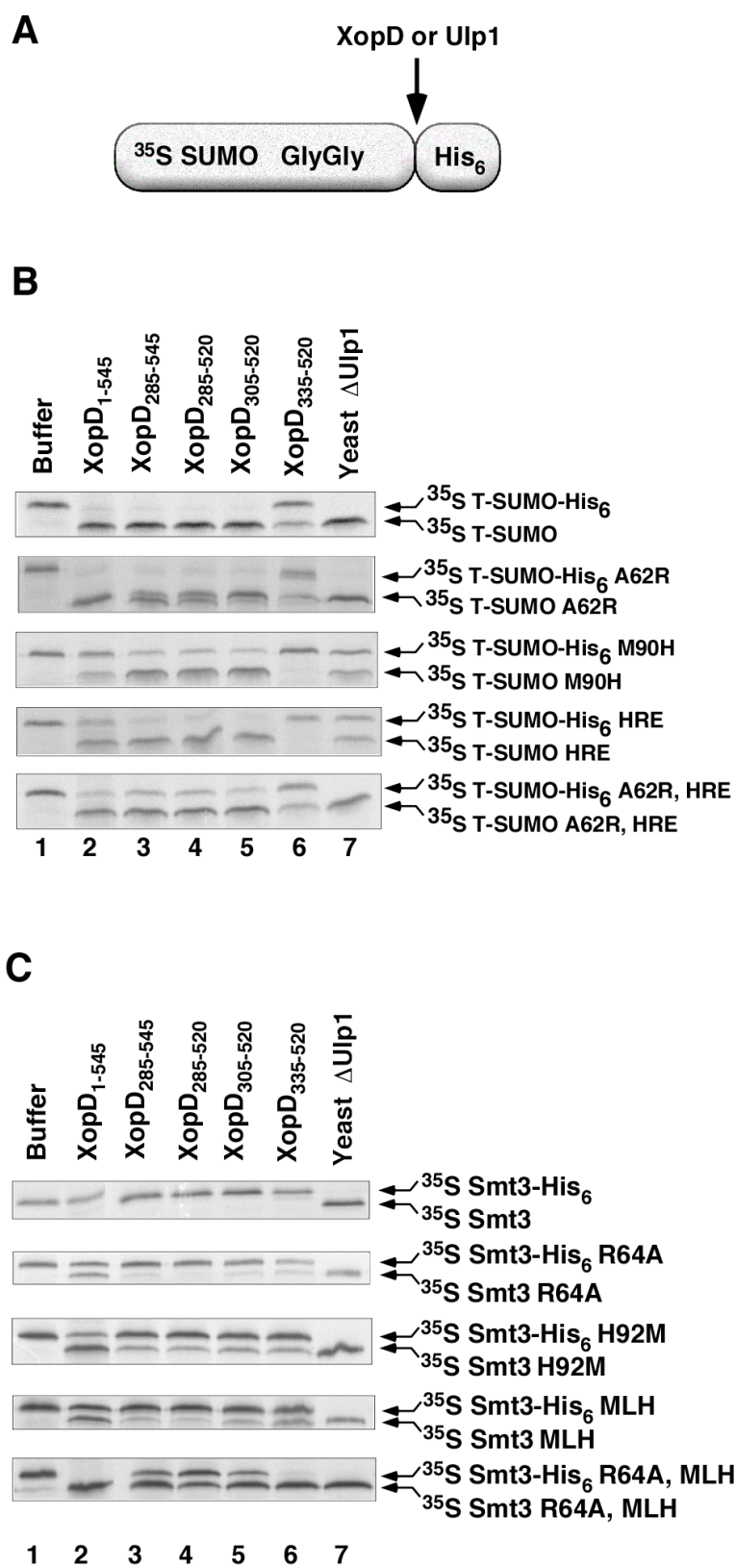


**Figure 13: Purified recombinant XopD protein expression .** Several XopD constructs, XopD<sub>1-545</sub>, XopD<sub>285-545</sub>, XopD<sub>285-520</sub>, XopD<sub>305-520</sub>, XopD<sub>335-520</sub> (full-length XopD and XopD cleavage products) were used in the peptidase assays to determine which portions of XopD are important for SUMO substrate recognition. A. GST-tagged XopD constructs were purified from bacteria and run on SDS gels and then stained with Coomassie Blue (overloaded lanes). B. Proteins from A were diluted and then run on SDS gels and stained with Coomassie Blue. (Note the change in size of the various XopD proteins).

As shown in Figure 14B, by mutating residues in the P<sub>5</sub>-P<sub>7</sub> positions of T-SUMO, the ability of XopD to process T-SUMO is reduced, but not abolished. Wild-type T-SUMO is processed well by all XopD constructs except XopD<sub>335-520</sub>. The lack of activity by XopD<sub>335-520</sub> is expected because it lacks a portion of the proposed substrate binding domain. Similar results to wild-type T-SUMO are observed with the A62R mutant. However, when the M90H mutation is made in T-SUMO, all of the XopD constructs exhibit a weakened ability to process T-SUMO. This trend is even more pronounced when all of the residues at the P<sub>5</sub>-P<sub>7</sub> positions are mutated to look like Smt3. This data suggests that the P<sub>5</sub>-P<sub>7</sub> region of T-

SUMO is important for XopD's ability to process T-SUMO, but that this region is not all that is necessary for recognition by XopD.

By mutating residues in the P<sub>5</sub>-P<sub>7</sub> positions of Smt3, XopD is able to process Smt3 to some extent (Figure 14C). Wild-type Smt3 cannot be processed by any of the XopD constructs, but Smt3 R64A is processed to a small extent by XopD<sub>1-545</sub> and XopD<sub>335-520</sub>. All of the XopD constructs are able to partially process Smt3 when residues in the P<sub>5</sub>-P<sub>7</sub> positions are mutated to look like those in T-SUMO. The results from this set of experiments further supports the idea that although the P<sub>5</sub>-P<sub>7</sub> region is dictating some of the specificity for XopD, there are other portions in SUMO that play a role in substrate recognition.



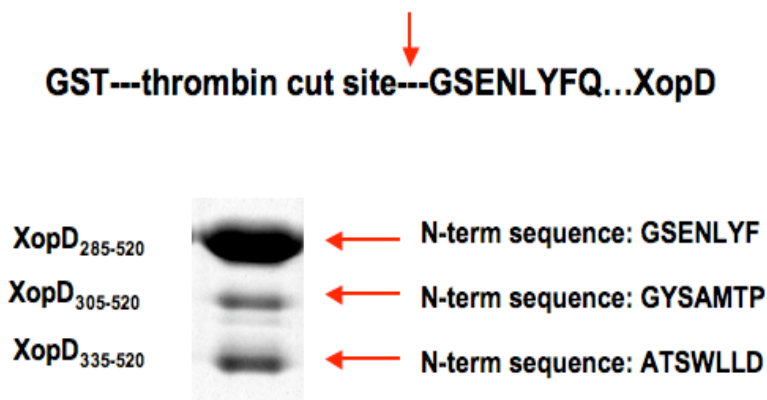
**Figure 14: Mutation of amino acids in T-SUMO and Smt3 alter substrate specificity.** A. Schematic of *in vitro* SUMO peptidase assay. B. [<sup>35</sup>S]-Tomato-SUMO-His<sub>6</sub> and [<sup>35</sup>S]-Tomato-SUMO-His<sub>6</sub> mutants were translated *in vitro* in a rabbit reticulocyte lysate and then incubated with buffer, 0.5 mg/ml GST-XopD<sub>1-545</sub>, GST-XopD<sub>285-545</sub>, GST-XopD<sub>285-520</sub>, GST-XopD<sub>305-520</sub>, GST-XopD<sub>335-520</sub> or GST-Ulp1 for 1 hour at 30°C. SUMO species were resolved on a 15% SDS gel and visualized by autoradiography. C. [<sup>35</sup>S]-Smt3-His<sub>6</sub> and [<sup>35</sup>S]-Smt3-His<sub>6</sub> mutants were translated *in vitro* in a rabbit reticulocyte lysate and then incubated with buffer, 0.5 mg/ml GST-XopD<sub>1-545</sub>, GST-XopD<sub>285-545</sub>, GST-XopD<sub>285-520</sub>, GST-XopD<sub>305-520</sub>, GST-XopD<sub>335-520</sub> or GST-Ulp1 for 1 hour at 30°C. SUMO species were resolved on a 15% SDS gel and visualized by autoradiography.

### *XopD structural studies*

The extent of contacts between XopD and T-SUMO appears to be very complex and difficult to fully understand with only mutational studies. Thus structural studies were undertaken in hopes of better understanding how XopD recognizes T-SUMO. Structural studies were encouraged based on two published X-ray crystal structures of yeast Ulp1 and human SENP2 (human Ulp1), and that XopD expressed in large and soluble quantities in *E. coli* (11,84).

The first step in beginning structural studies with XopD, was to determine the appropriate protein construct. Initial experiments were conducted using a XopD construct that was similar to that of yeast Ulp1, used in the structural studies of yeast Ulp1 with Smt3 (11). This construct, XopD<sub>285-520</sub>, was purified as an amino-terminal GST fusion protein by glutathione agarose affinity beads. The GST moiety was cleaved by thrombin, and the protein was further purified by a MonoQ anion exchange column. As depicted in Figure 13, XopD is self-cleaved into several truncated XopD proteins. The purified protein preparation of XopD<sub>285-520</sub> contained two cleaved forms of XopD that could not be purified away from one another (Figure 15). Thus a new XopD construct, XopD<sub>335-520</sub>, was used for

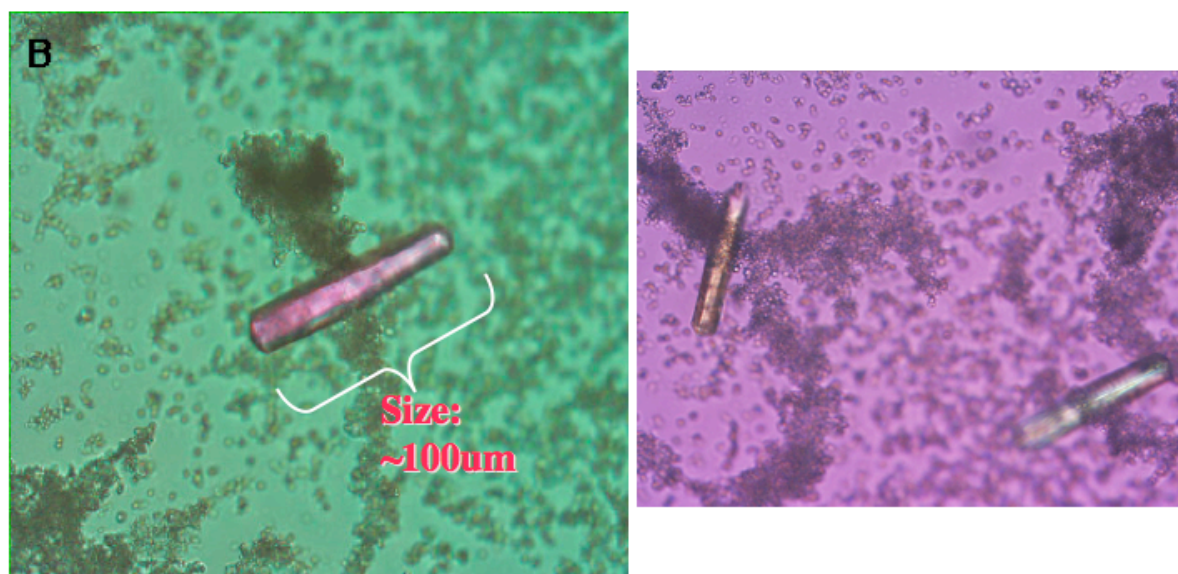
crystallization experiments since this was the smallest cleavage product of XopD. XopD<sub>335-520</sub> was purified as described for XopD<sub>285-520</sub>.



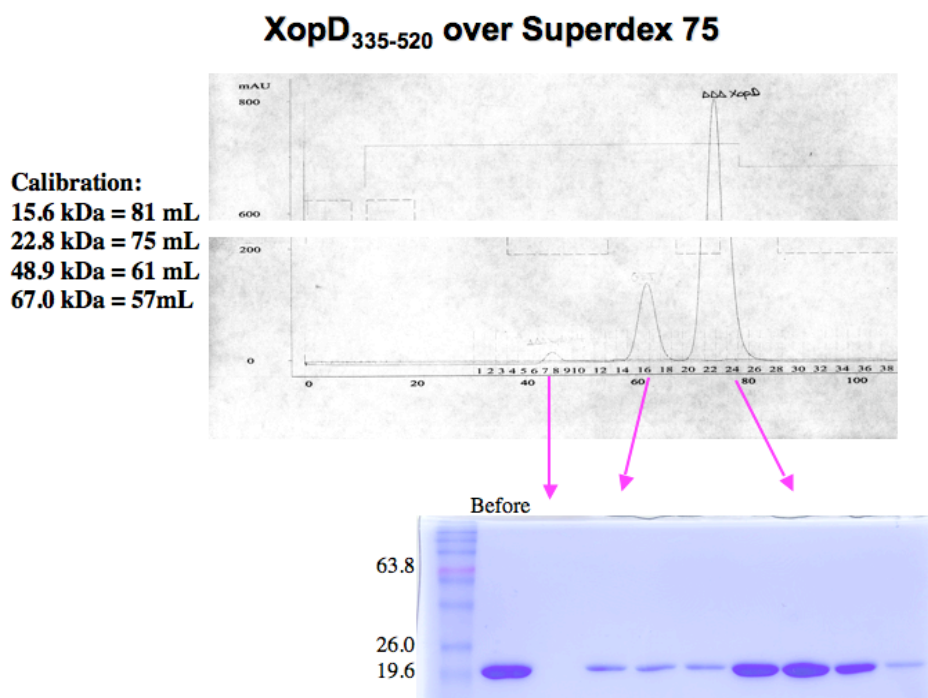
**Figure 15: XopD<sub>285-520</sub> purified protein preparation yields two XopD cleavage products.** The purified protein preparation of XopD<sub>285-520</sub> contained two cleaved forms of XopD, XopD<sub>305-520</sub> and XopD<sub>335-520</sub> that could not be purified away from one another. N-terminal sequencing confirmed the identity of the two XopD cleavage products.

The protein, XopD<sub>335-520</sub>, was concentrated and used for crystallization trials with commercially available screening kits from Hampton Research. Initial screening led to the identification of one set of buffer conditions that allowed for crystal growth. These conditions were further optimized to obtain more crystals for screening purposes. These crystals grew as small rods and were not single crystals (Figure 16). Thus further purification of the protein was required to grow single crystals. XopD<sub>335-520</sub> was further purified on a Superdex 75 gel filtration column after the anion exchange column to remove any XopD protein dimers or tetramers (Figure 17). This extra step of purification yielded large, single crystals as shown in Figure 18. The same purification protocol and crystallization conditions for the wild-type XopD<sub>335-520</sub> were used to obtain XopD<sub>335-520</sub> C470A protein crystals.

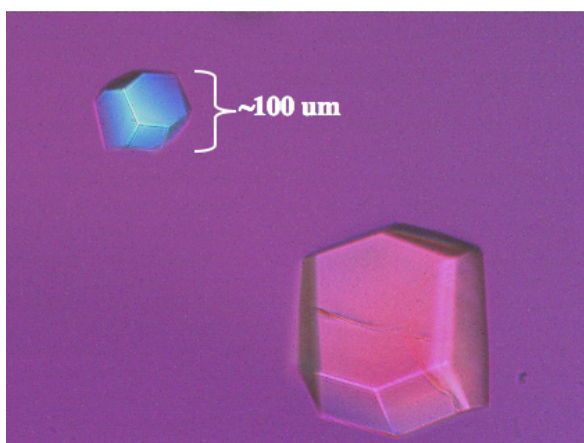




**Figure 16: Initial XopD proteins crystals are small rods and are not single crystals.** XopD<sub>335-520</sub> (purified with a Hi Trap MonoQ column) protein crystals were obtained using 1  $\mu$ L of protein at 15mg/mL with 1  $\mu$ L of buffer containing 1.4 - 1.6 M Sodium potassium phosphate at pH 7.5 - 7.75 at 20°C.

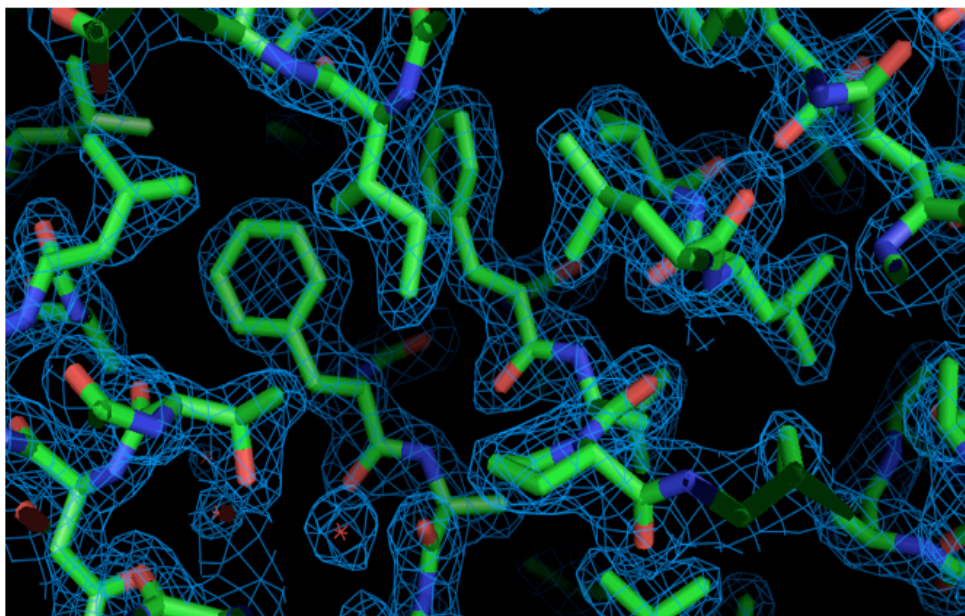


**Figure 17: XopD<sub>335-520</sub> protein purification using a Superdex 75 gel filtration column.** Further purification of the XopD<sub>335-520</sub> protein was required to grow single crystals. XopD<sub>335-520</sub> was further purified on a Superdex 75 gel filtration column after the anion exchange column to remove any XopD protein aggregates. Fractions were run on an 15% SDS gel and stained with Coomassie Blue.

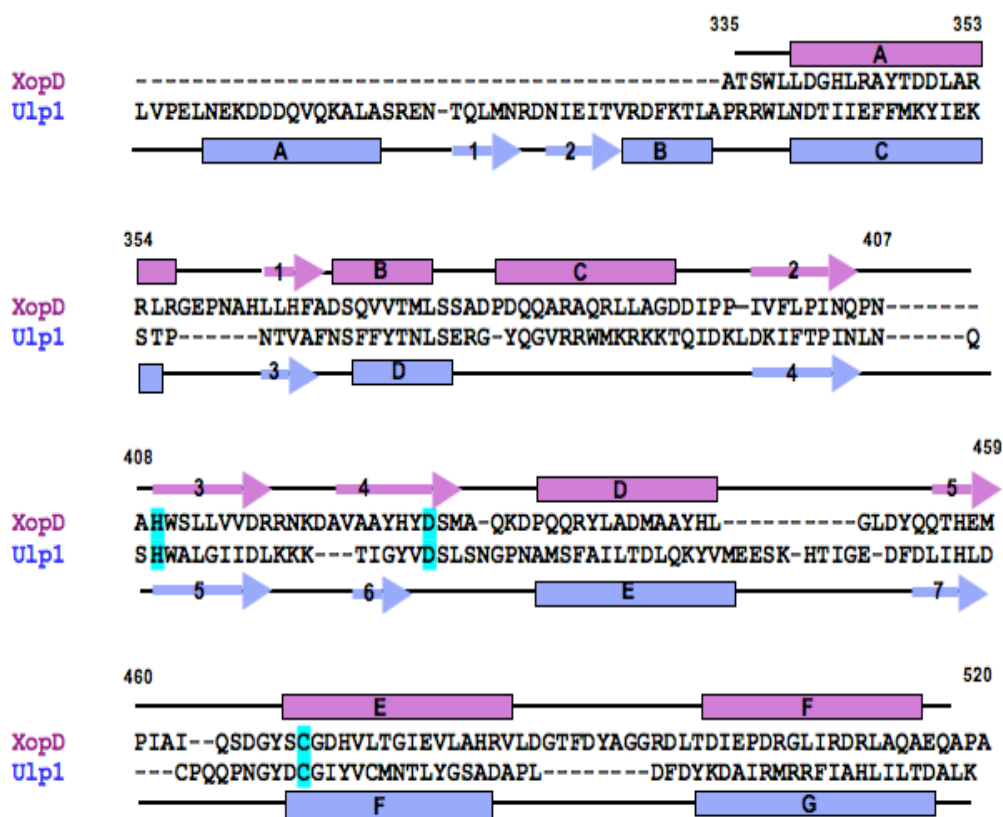


**Figure 18: XopD proteins crystals used for data collection.** XopD<sub>335-520</sub> (purified with a Hi Trap MonoQ column and Superdex 75) protein crystals were obtained using 1 μL of protein at 15mg/mL with 1 μL of buffer containing 1.4 - 1.6 M Sodium potassium phosphate at pH 7.5 - 7.75 at 20°C.

Due to low sequence similarity, it was not possible to use molecular replacement to solve the structure of XopD with the structure of yeast Ulp1. Therefore, selenomethionyl XopD<sub>335-520</sub> protein crystals were used to solve the structure of XopD<sub>335-520</sub>. We then used isomorphous replacement to solve the structure of XopD<sub>335-520</sub> C470A. The wild-type XopD<sub>335-520</sub> structure was solved to 1.95 Å resolution and the XopD<sub>335-520</sub> C470A structure was solved to 1.8 Å resolution (Figure 19). Figure 20 shows the structure-based sequence alignment of XopD<sub>335-520</sub> with yeast Ulp1. Ribbon diagrams of both wild-type XopD<sub>335-520</sub> and XopD<sub>335-520</sub> C470 show the catalytic triad of Cys/Ala 470, His 409, and Asp 429 (Figure 21A). The secondary structure of XopD<sub>335-520</sub> includes 6 alpha helices and 5 beta strands. The XopD structures show the beta sheets packed between the helical structural elements of the protein (Figure 21A and B).



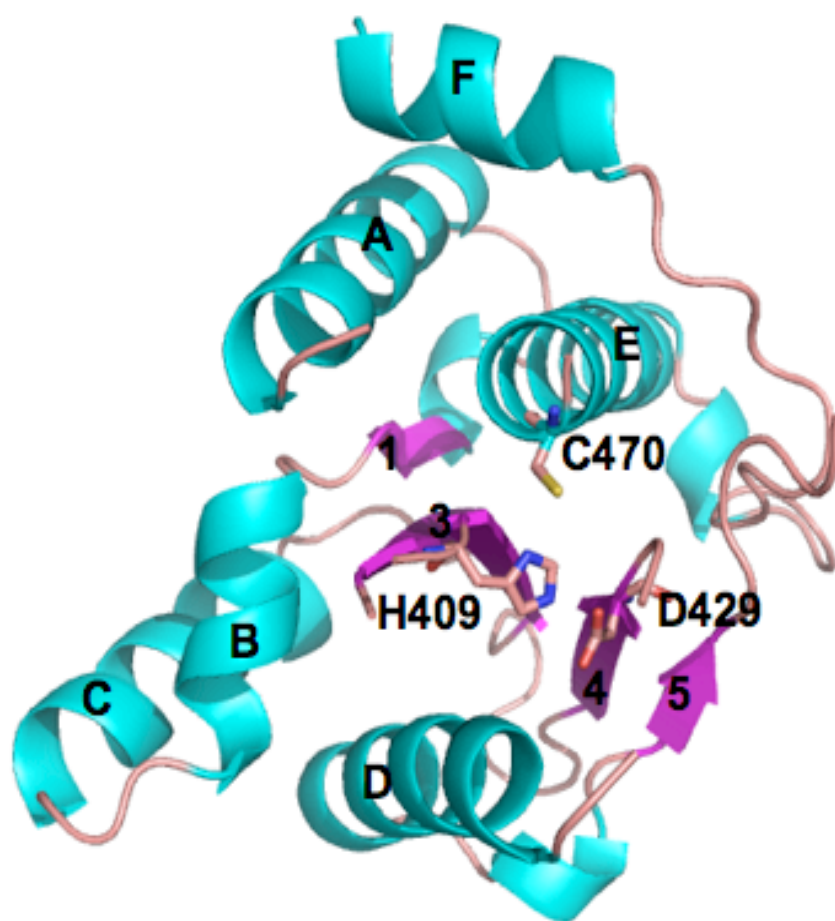
**Figure 19: Electron Density map of a portion of Wild-type XopD<sub>335-520</sub>.** Electron density map provides an example of the resolution at 1.95Å obtained from wild-type XopD<sub>335-520</sub>.



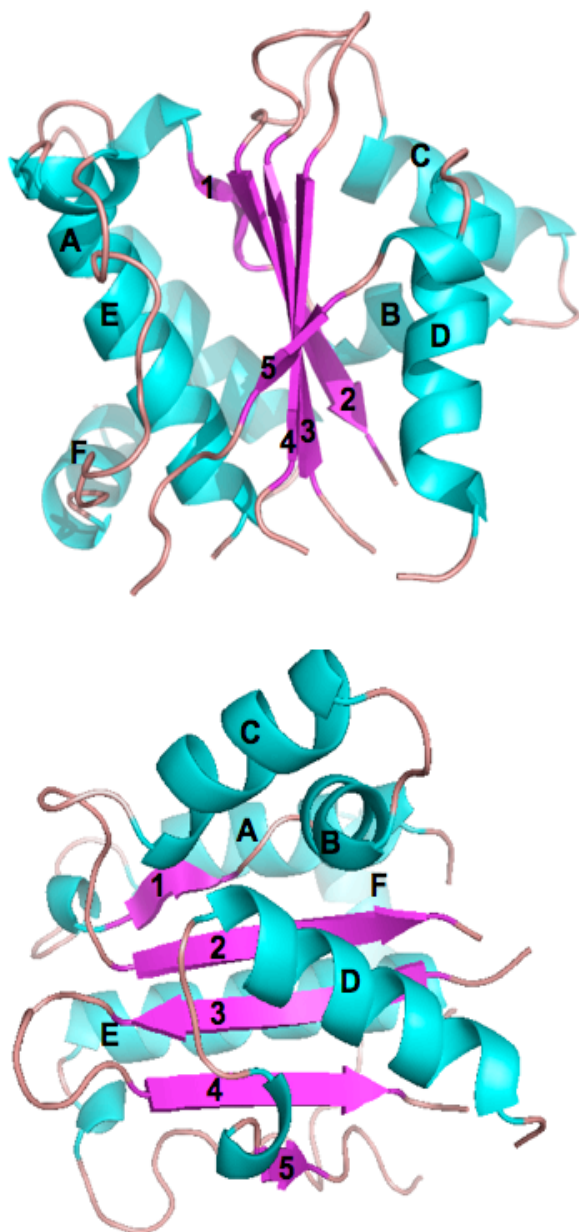
**Figure 20: Structure-based sequence alignment of XopD<sub>335-520</sub> with yeast Ulp1.**

Sequence alignment of XopD with yeast Ulp1. Gaps are denoted with dashed lines. Numbering is with respect to XopD. The XopD secondary structure is shown above the sequence in purple, and the yeast Ulp1 secondary structure is shown below the sequence in blue. Beta strands are numbered and helices are lettered. Catalytic residues are shown in cyan.

A.



B.

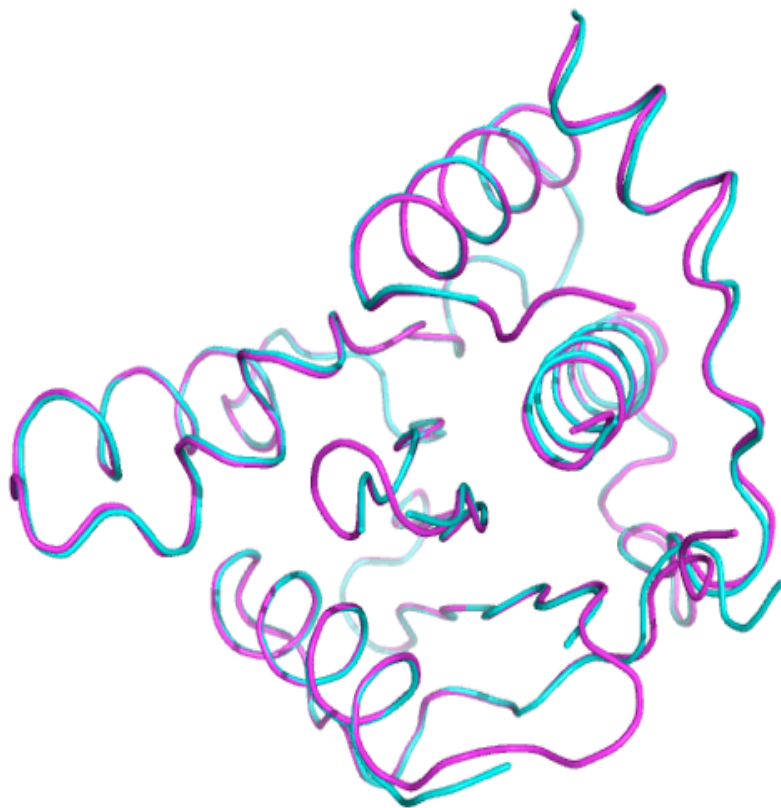


**Figure 21: Ribbon diagram of two views of wild-type XopD<sub>335-520</sub>.** A. The catalytic triad, C470, H409, and D429, shown in this ribbon diagram of wild-type XopD<sub>335-520</sub>. B. Other views of the XopD<sub>335-520</sub> structure shown in A.

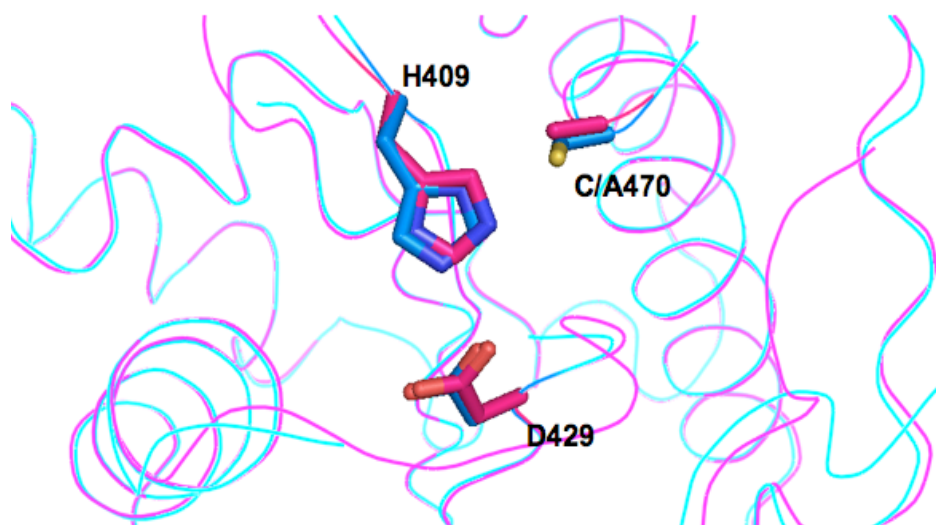


Superposition of wild-type XopD<sub>335-520</sub> and XopD<sub>335-520</sub> C470 C $\alpha$ -coordinates show very little difference in the two proteins (Figure 22). The two structures superimpose with an rmsd of 0.31 Å for 168 common C  $\alpha$  atoms of wild-type XopD<sub>335-520</sub> and XopD<sub>335-520</sub> C470. A close-up view of the catalytic core of wild-type XopD<sub>335-520</sub> and XopD<sub>335-520</sub> C470 show conservation between the two structures (Figure 23). Large conformational changes between the structures of wild-type XopD<sub>335-520</sub> and XopD<sub>335-520</sub> C470 were not predicted. The structure of the catalytically inactive XopD was solved to confirm the hypothesis that there would not be any large conformational changes between the mutant and wild-type protein and because there was not a structure of a catalytically inactive ULP1 family member reported in the literature. Superposition of either wild-type XopD<sub>335-520</sub> or XopD<sub>335-520</sub> C470A with yeast Ulp1:Smt3 show some similarity between secondary structural elements (Figure 20 and 24). The helices A, B, D, E and F of XopD share similarity with the Ulp1 structure (Figure 20 and 24). Interestingly, XopD contains an extra helix, helix C, that is not part of the Ulp1 structure (Figure 20). Beta strands 1, 2, 3, 4 and 5 of XopD also share similarity to Ulp1 secondary structure elements (Figure 20). The two structures superimpose with an rmsd of 1.95 Å for 101 common C  $\alpha$  atoms of XopD<sub>335-520</sub> and yeast Ulp1. While there are differences in the structures between wild-type XopD<sub>335-520</sub> and yeast Ulp1, the residues of the catalytic triad align between the two proteins (Figure 25).

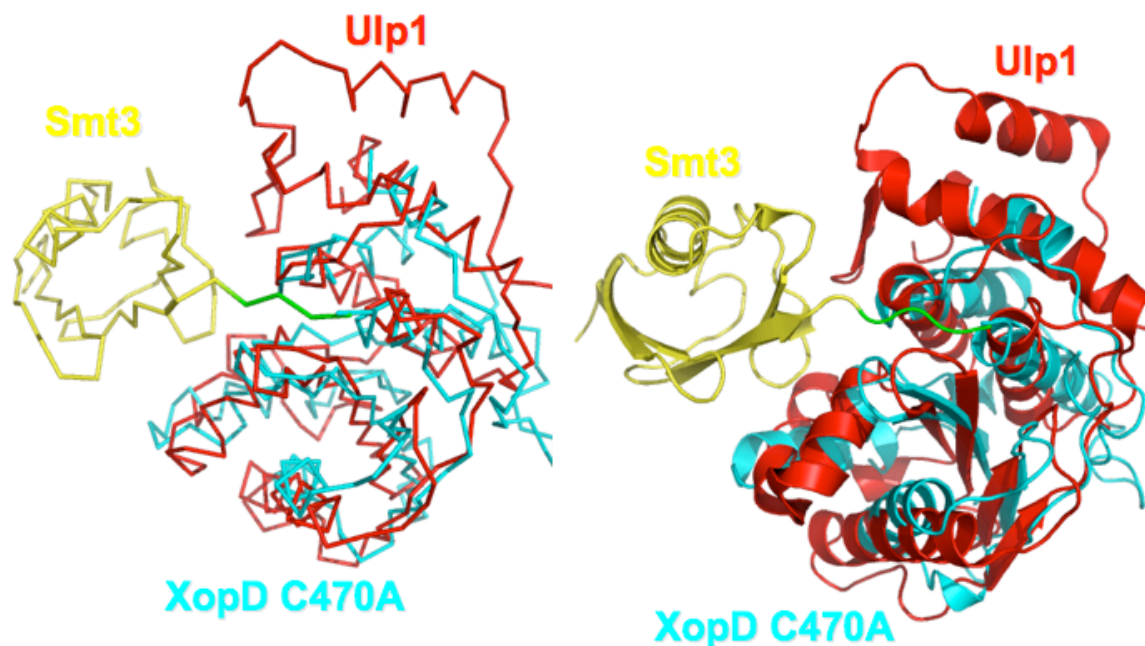




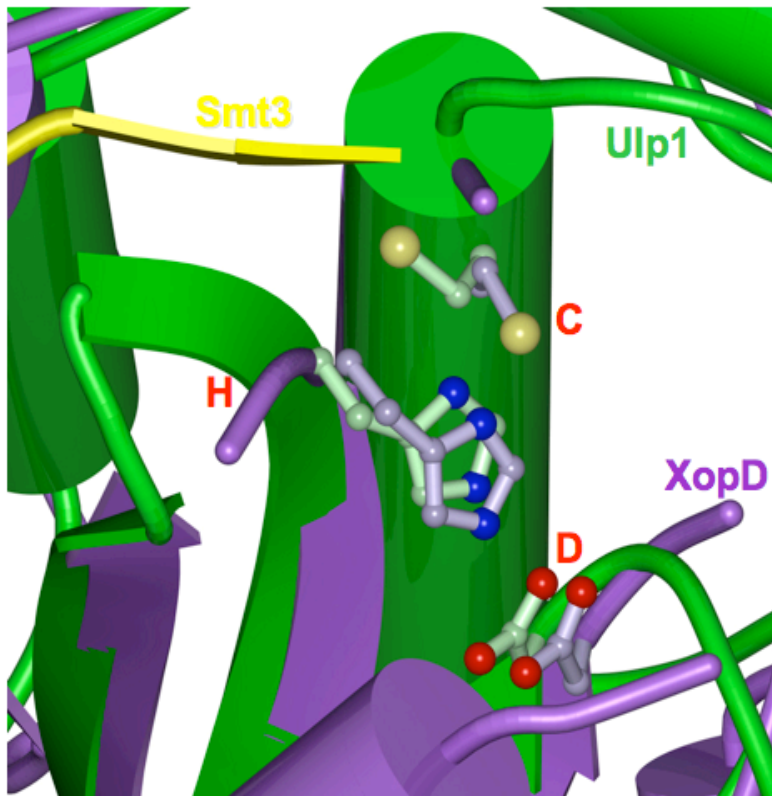
**Figure 22: Superposition of wild-type XopD<sub>335-520</sub> with XopD<sub>335-520</sub> C470.** The backbones of wild-type XopD<sub>335-520</sub> (cyan) and XopD<sub>335-520</sub> C470 (magenta) superimpose with an rmsd of 0.31 Å for 168 common C  $\alpha$  atoms of wild-type XopD<sub>335-520</sub> and XopD<sub>335-520</sub> C470.



**Figure 23: Superposition of wild-type XopD<sub>335-520</sub> with XopD<sub>335-520</sub> C470 catalytic triad.** The backbones of wild-type XopD<sub>335-520</sub> (cyan) and XopD<sub>335-520</sub> C470 (magenta) are also shown.



**Figure 24: Superposition of XopD<sub>335-520</sub> with the yeast Ulp1:Smt3 structure (11).** Yeast Ulp1 is shown in red, Smt3 is shown in yellow, the QTGG region of the Smt3 C-terminus is shown in green, and XopD is shown in cyan. The two structures superimpose with an rmsd of 1.95 Å for 101 common C  $\alpha$  atoms of XopD<sub>335-520</sub> and yeast Ulp1.



**Figure 25: Superposition of XopD<sub>335-520</sub> with yeast Ulp1 catalytic triad.** Yeast Ulp1 is shown in green, Smt3 is shown in yellow, and XopD is shown in purple. The catalytic triad of Cys, His and Asp of the XopD and Ulp1 align well.

## Conclusions

The results of the mutational and structural studies provide insight into the SUMO substrate specificity of XopD. The mutational studies with T-SUMO and Smt3 showed that the manner in which XopD is recognizing its substrate involves extensive contacts with SUMO that cannot be fully understood with mutational studies. The XopD structure shows differences in the secondary structural elements when compared to yeast Ulp1, while the catalytic triad is conserved. The structure of XopD revealed the presence of an additional helix, helix C, that is not present in yeast Ulp1. When the structure of XopD was aligned with the structure of yeast Ulp1 with regards to the catalytic residues, the helices and beta strands of each enzyme did not align well. Thus solving the crystal structure of XopD provides a clue as to the differences in the substrate specificity between XopD and yeast Ulp1. To fully understand how XopD is recognizing the SUMO substrate, the co-crystal structure of XopD with T-SUMO could be solved. Attempts at solving this structure are discussed in Chapter 6.

## **CHAPTER SIX**

### **Results**

#### **EXPERIMENTAL TRIALS:**

#### ***XANTHOMONAS* XopD IN COMPLEX WITH TOMATO SUMO**

##### **Introduction**

In order to understand how *Xanthomonas* XopD is recognizing its SUMO substrate, experiments were initiated with the goal of solving the structure of XopD in complex with its substrate, Tomato SUMO (T-SUMO). The idea for structural studies of this complex came from a paper from the lab of Chris Lima, where the structure of yeast Ulp1 in complex with its substrate, Smt3 (yeast SUMO) was solved (11). In these studies, the authors were able to determine specific residues in yeast Ulp1 that are in contact with specific residues in Smt3. Chapters 4 and 5 describe how XopD recognizes a distinct group of SUMO substrates (T-SUMO, *At*SUMO-1, and *At*SUMO-2) in contrast to Ulp1, which recognizes a diverse set of SUMO substrates. Since XopD exhibits differences in its SUMO substrate specificity as compared to Ulp1, solving the crystal structure of XopD in complex with T-SUMO would provide insight into this substrate specificity.

##### **Experimental Trials**

###### *Ulp1:Smt3 Complex Formation Protocol*

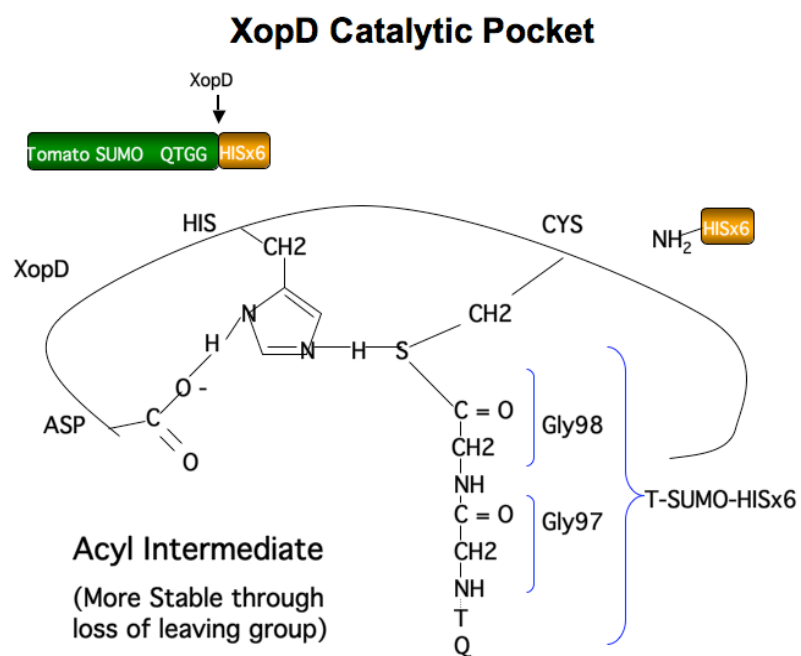
Before the structure of XopD in complex with T-SUMO can be solved, large quantities of protein complex must be purified from bacteria. Initially the method followed for protein complex purification were those of Chris Lima's lab as described in the Ulp1:Smt3 co-crystal

paper (11). This method involved purifying the enzyme, Ulp1, without a tag, and then purifying the substrate, Smt3, with a C-terminal His<sub>6</sub> tag. The C-terminal His<sub>6</sub> tag would be cleaved by the enzyme in the complex-forming reaction, leaving an untagged protein complex. The Ulp1:Smt3 protein complex is that of a stable transition state analog formed by the addition of sodium borohydride (NaBH<sub>4</sub>). The NaBH<sub>4</sub> reduces the acyl intermediate (formed between the thiol group of the cysteine in the Ulp1 active site with the carboxyl group of the C-terminal glycine in the Smt3) to a more stable thiohemiacetal transition state analogue (11,85) (Figure 26A-C). This reaction results in the formation of a thiohemiacetal bond between the active site cysteine of Ulp1 and the C-terminal glycine of Smt3 (Figure 26D). The protein complex formation protocol followed by the Lima lab involved using a 1:3 molar ratio of Ulp1 to Smt3 protein. The enzyme and substrate were incubated for 30 minutes with the addition of 5 aliquots of NaBH<sub>4</sub> (50mM final concentration). The percent of complex formation was reported to be between 20%-30% and was further purified on a MonoQ and Superdex 75 columns. The authors then concentrated the complex to 25 mg/mL for crystallization trials.

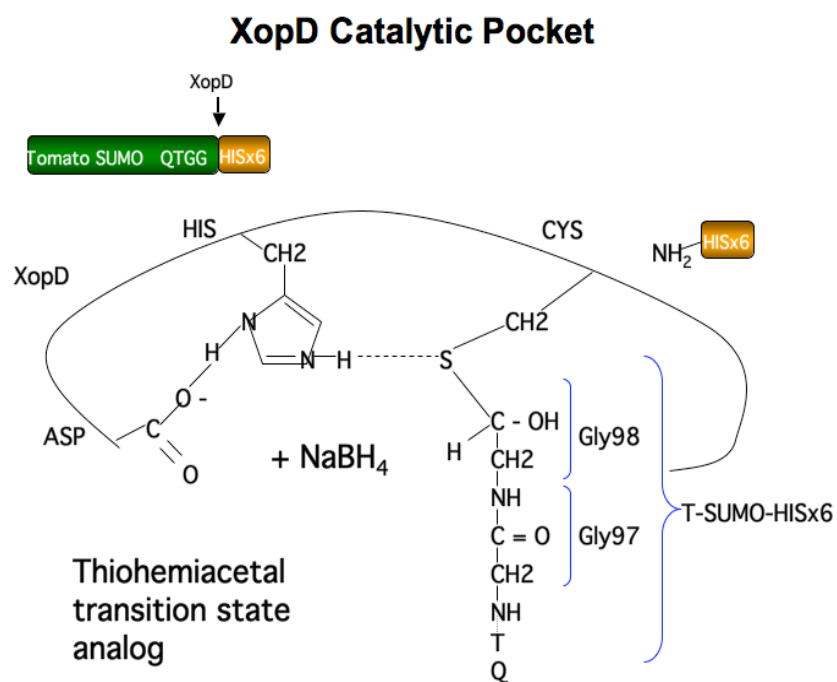




C.



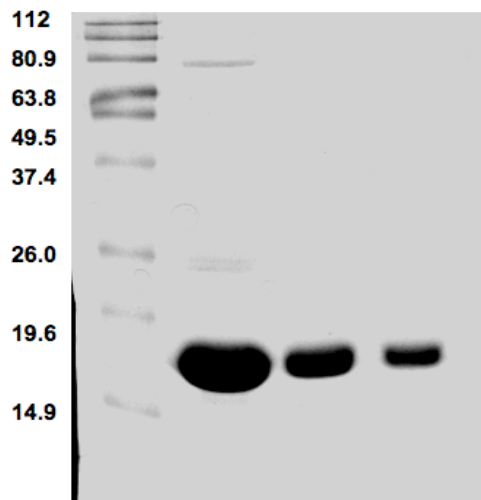
D.



**Figure 26: Model of the chemistry in the catalytic pocket of XopD during XopD:T-SUMO complex formation using NaBH<sub>4</sub>.** A. The Cys nucleophile in the catalytic pocket of XopD attacks the carbonyl carbon of the most C-terminal Gly of T-SUMO-His<sub>6</sub> leading to the formation of a tetrahedral intermediate (B.), which collapses to form a more stable acyl intermediate through the loss of the C-terminal leaving group (His<sub>6</sub> tag) (C.). D. NaBH<sub>4</sub> reduces the aldehyde to an alcohol, resulting in a thiohemiacetal transition state analogue.

#### *XopD:T-SUMO Complex Formation Trials*

Complex formation with XopD and T-SUMO using this Lima lab protocol was attempted. The enzyme, GST-XopD<sub>285-520</sub>, was used because it was similar to the Ulp1 construct used by the Lima lab (catalytic domain of Ulp1 with small portion of C-terminus removed). This protein was purified from bacteria and the GST tag was removed with thrombin without difficulty. The substrate, T-SUMO-GG-His<sub>6</sub>, was purified with the His<sub>6</sub> tag remaining on the protein (Figure 27). Several milligrams of each protein were purified in order to begin trials at forming the XopD:T-SUMO covalent complex.



**Figure 27: Purification of recombinant T-SUMO-His<sub>6</sub> protein for complex formation trials.** The T-SUMO-His<sub>6</sub> proteins was purified using Ni<sup>++</sup> beads, followed by a MonoQ column and a Superdex 75 gel filtration column. The purified protein fractions from the gel filtration column were run on a 15% SDS gel and stained with Coomassie Blue.

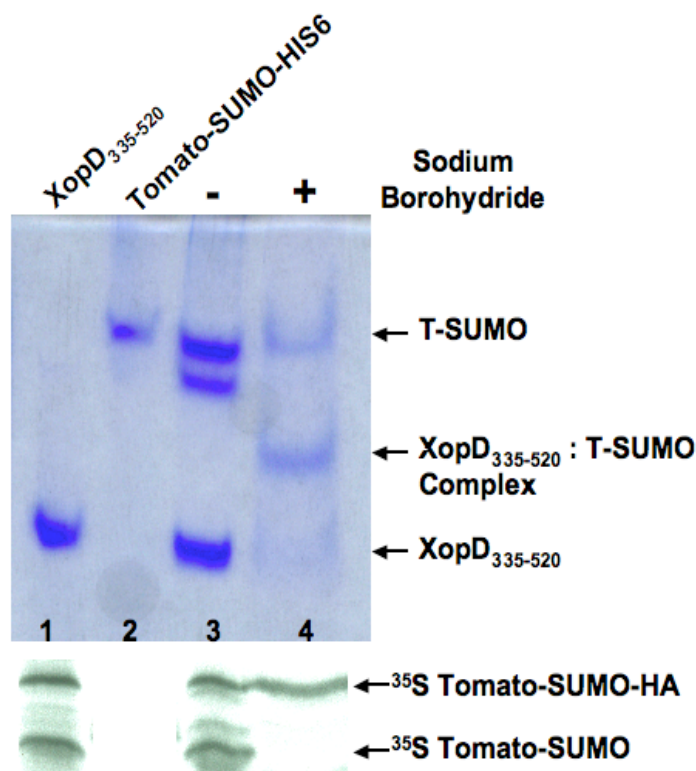
In trial #1, a small-scale reaction was setup using 120 µg of XopD<sub>285-520</sub> and 350 µg of T-SUMO-His<sub>6</sub>. Both the enzyme and substrate were added to a small 5mL tube and then small amounts (5 aliquots, totaling 50mM) of solid sodium cyanoborohydride were added over a 30 minute period at 37°C. The reaction formed foam upon each addition of the sodium cyanoborohydride. The reaction mixture was then extensively buffer exchanged using an Amicon ultrafiltration device to get rid of the sodium cyanoborohydride. The reaction mixture was then run on native gels, yet no protein complex was detected at this stage.

The protocol was altered in trial #2, whereby the addition of the first aliquot of sodium cyanoborohydride was added to the T-SUMO substrate at the same time that the enzyme was added. Again, there was no detectable complex formed as assessed by the inspection of native gels.

For trial #3 the reaction was scaled up by using 2.5 mg of XopD<sub>285-520</sub> and 2.5 mg of T-SUMO-His<sub>6</sub>, and the use of NaBH<sub>4</sub> for a one hour reaction at room temperature. The reaction was performed in the presence and absence of DTT. The final reaction mixture was run over a MonoQ column and a MonoS column. All of the protein loaded onto a MonoS column came out in the flow-through, and did not bind to the column. Protein bound to the MonoQ column was eluted with salt, yet there was no “complex” detected by inspection of native gels. This same set of experimental conditions was tested several times with no complex formation detected.

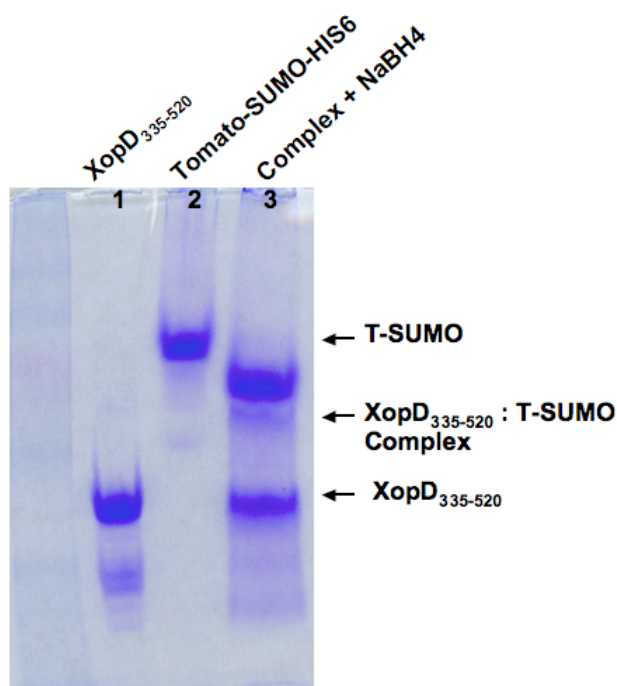
At this stage, a different XopD protein construct, XopD<sub>335-520</sub>, was used because the XopD<sub>285-520</sub> protein prep also contained the XopD<sub>335-520</sub> cleavage product. The XopD cleavage products were all active towards the SUMO substrate, and may have all been forming a complex with the substrate. Trial #4 included 1 mg of XopD<sub>335-520</sub> and 3 mg of T-SUMO-His<sub>6</sub>, and the use of NaBH<sub>4</sub> for a one hour reaction at room temperature. The native gels for this reaction showed a band migrating between the T-SUMO band and the XopD<sub>335-520</sub> band (Figure 28). This band was identified as the protein complex band due to its migration in the native gel. This protocol was repeated several times with different amounts of protein, and the “complex” band was consistently detected. To try and determine if this band was the covalent protein complex, an assay (similar to what was used in Chapters 4 and 5) using <sup>35</sup>S-labeled-T-SUMO-HA was attempted. In this crude assay, both XopD and T-SUMO were incubated for one hour at room temperature in the presence and absence of NaBH<sub>4</sub> (Figure 28). Thus if the NaBH<sub>4</sub> is trapping a covalent complex between enzyme and substrate, then the complex band should only be identified in the protein mixture that

contained NaBH<sub>4</sub>. Also, the protein mixture incubated in the presence of NaBH<sub>4</sub> should not be able to process <sup>35</sup>S-labeled-T-SUMO-HA since the XopD has formed a covalent complex with T-SUMO making its catalytic Cys inaccessible to additional SUMO substrate (Figure 28). The protein mixture that was not incubated in the presence of NaBH<sub>4</sub> is able to process <sup>35</sup>S-labeled-T-SUMO-HA because in this sample, the XopD is not covalently bound to its substrate, thus it can process additional substrate. This assay is crude and does not take into consideration that the NaBH<sub>4</sub> could somehow be inactivating the XopD by causing XopD dimers to form, or that there could be excess XopD in the reaction with NaBH<sub>4</sub> that did not form a complex with T-SUMO.



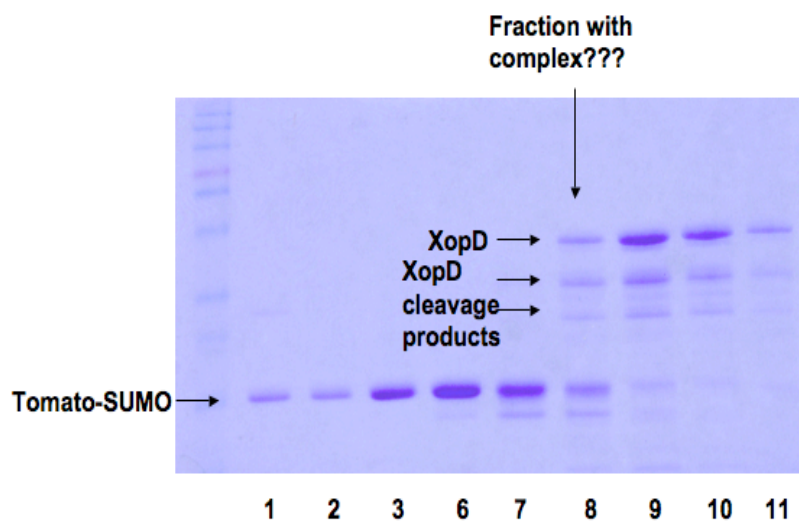
**Figure 28: Native gels were used to visualize formation of a covalent complex between XopD and T-SUMO.** Upper panel: In the assay, 1 mg of XopD<sub>335-520</sub> and 3 mg of T-SUMO-Hisx6 were incubated in the presence (lane 4) or absence (lane 3) of NaBH<sub>4</sub> for a one hour reaction at room temperature. The native gel shows a band migrating between the T-SUMO band and the XopD<sub>335-520</sub> band (lane 4). Lower panel: Protein samples from the complex forming assay were incubated with <sup>35</sup>S-labeled-T-SUMO-HA for 1 hour at 30°C. The samples were then run on a 15% SDS page gel and visualized by autoradiography. Lanes 1 and 3 show that when XopD is active (or not in a covalent complex) it can process <sup>35</sup>S-labeled-T-SUMO-HA to <sup>35</sup>S-labeled-T-SUMO. Lane 4 shows that when XopD forms the presumed covalent complex with T-SUMO, that XopD is no longer able to process <sup>35</sup>S-labeled-T-SUMO-HA.

Trial #5 involved using the same protocol as in trial #4, but 4 mg of XopD<sub>335-520</sub> with 12 mg of T-SUMO-His<sub>6</sub> were incubated for one hour at room temperature, and then run on a native gel. The same complex band as in trial #4 was observed (Figure 29). The complex was buffer exchanged to remove any NaBH<sub>4</sub> and then the complex was loaded onto a MonoQ column. The fractions from the MonoQ column were then run on native as well as SDS PAGE gels (Figure 30). This method of purification yielded very impure complex fractions.

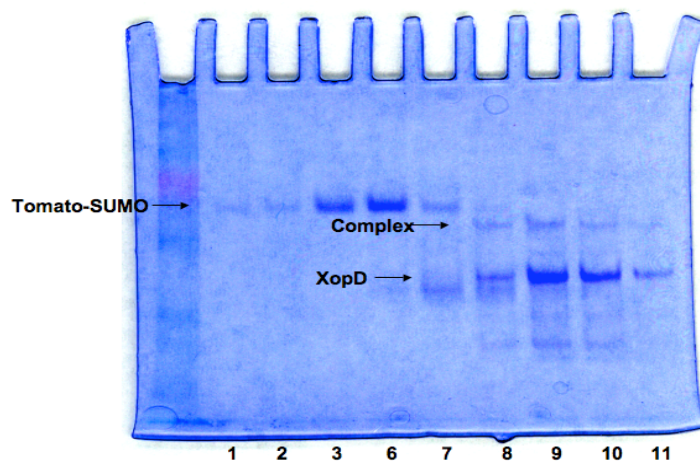


**Figure 29: Native gels were used to visualize formation of a covalent complex between XopD and T-SUMO.** In the assay, 4 mg of XopD<sub>335-520</sub> and 12 mg of T-SUMO-His<sub>6</sub> were incubated in the presence of NaBH<sub>4</sub> for a one hour reaction at room temperature (lane 3). The native gel shows a faint band migrating between the T-SUMO band and the XopD<sub>335-520</sub> band (lane 3).

A.



B.



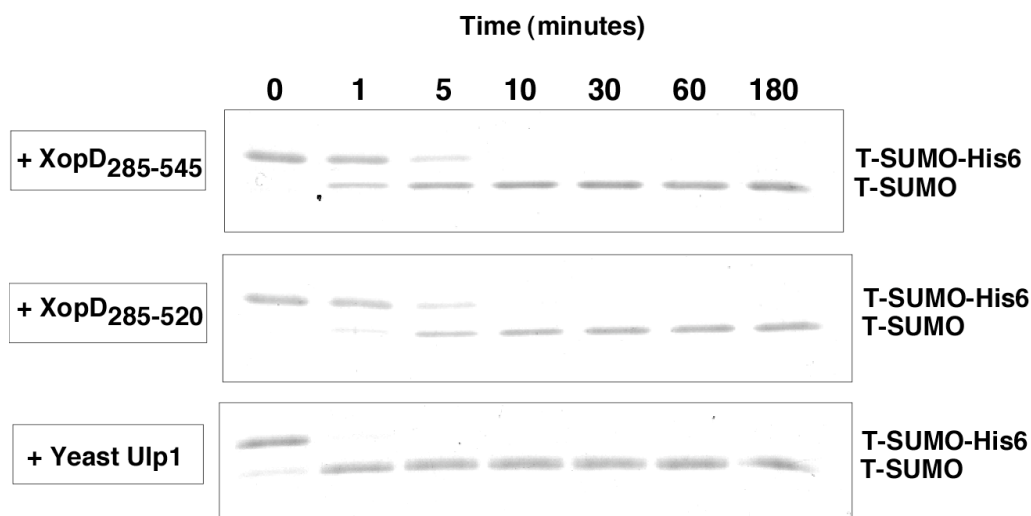
**Figure 30: XopD:T-SUMO complex purification using an anion exchange column.**

Protein sample (shown in lane 3 of Figure 30) was buffer exchanged to get rid of any  $\text{NaBH}_4$  and then the complex was loaded onto a Hi Trap MonoQ column. The fractions from the MonoQ column were then run on an SDS Page gel (A) and a native gel (B).



In Trial #6 a time course assay was performed to measure complex formation over a period of two hours. Complex formation was monitored by migration on native gels, and complex formation occurred by 20 minutes. Thus the formation of complex appeared quite fast, whereby all of the T-SUMO had either been cleaved by XopD and released, or cleaved by XopD and trapped by the NaBH<sub>4</sub>.

Time course peptidase assays with XopD and Ulp1 for T-SUMO-His<sub>6</sub> were performed to determine the efficiency of XopD (Figure 31). The time course assays shows that XopD, like Ulp1, rapidly processes all of the substrate within the first 5-10 minutes of the reaction.

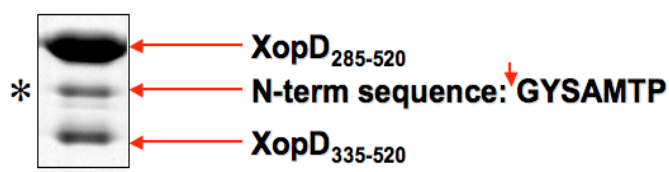


**Figure 31: Time course assay of XopD rapidly processing all of the T-SUMO-His<sub>6</sub> substrate in an *in vitro* peptidase assay.** Recombinant XopD<sub>285-545</sub>, XopD<sub>285-520</sub>, and yeast Ulp1 were incubated with recombinant T-SUMO-His<sub>6</sub> for the indicated times at 30°C. Samples were then run on 15% SDS gels and stained with Coomassie Blue.

For trial # 7 the reaction rate was slowed by running the reaction at 4°C for one hour. The complex formation did not appear to be enhanced by changing the temperature of the reaction. At this time, another paper came out from the Lima lab that reported the co-crystal structure of SENP2 with M-SUMO-1 (84). The formation of complex between SENP2 and M-SUMO-1 was done using the same methods as the Ulp1:Smt3 structure.

Closer inspection of the XopD protein construct that I had been using for complex formation yielded a new idea. By looking at the Ulp1:Smt3 structure, it was determined that XopD<sub>335-520</sub> was missing a portion of its substrate binding domain at its N-terminus which was shown to be a region of enzyme-substrate contacts from the Ulp1:Smt3 structure. The XopD<sub>335-520</sub> construct is also not as active in processing T-SUMO-His<sub>6</sub> as compared to the rest of the XopD constructs (Figure 14B). By looking at cleavage products from the full catalytic domain of XopD, another cleavage product that included more of this substrate binding domain than the XopD<sub>335-520</sub> construct was found (Figure 32A). This construct, XopD<sub>305-520</sub>, has one potential self cleavage site in the protein at G495 and one potential thrombin cleavage site at R334 (Figure 32B). These two sites were mutated to alanine residues in attempts to eliminate the cleavage of this protein construct.

A.



B.

### **XopD<sub>305-520</sub> R334A, G495A**

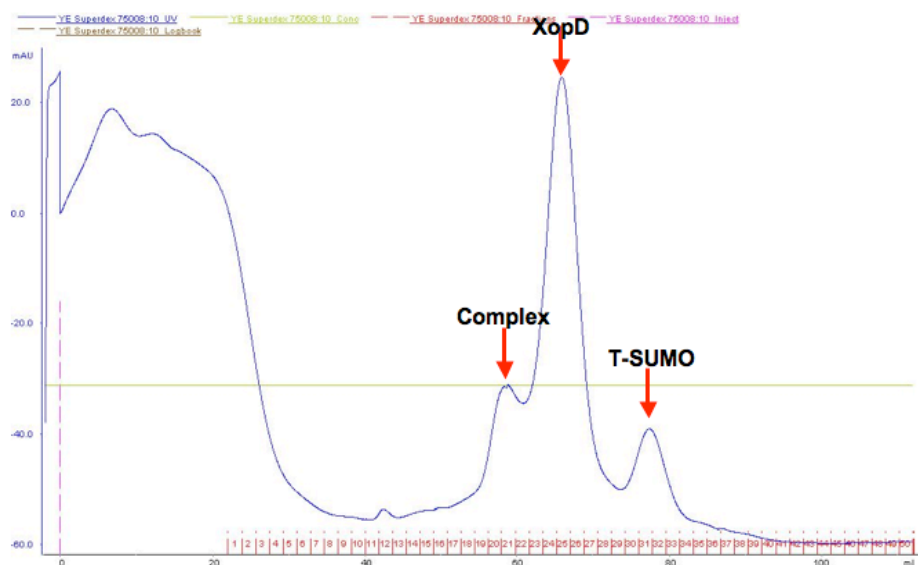
...DLNIPQQEEYPNNHGTQTPMGYSAMTPERIDVDN  
 LPSPQDVADPELPPVRATSWLLDGHLRAYTDDLARR  
 LRGEPNALLHFADSQVVTMLSSADPDQQARAQRL  
 LAGDDIPPVFLPINQPNAHWSLLVVDNRNKDAVAA  
 YHYDSMAQKDPQQRYLADMAAYHLGLDYQQTHE  
 MPIAIQSDGYSCGDHVLTGIEVLAHRVLDGTFDYA  
 GGRDLTDIEPDRGLIRDRLAQAEQAPA

(Red arrows point to the 'Y' in 'MGYSAMTP' and the 'R' in 'R334A', and the 'G' in 'G495A')

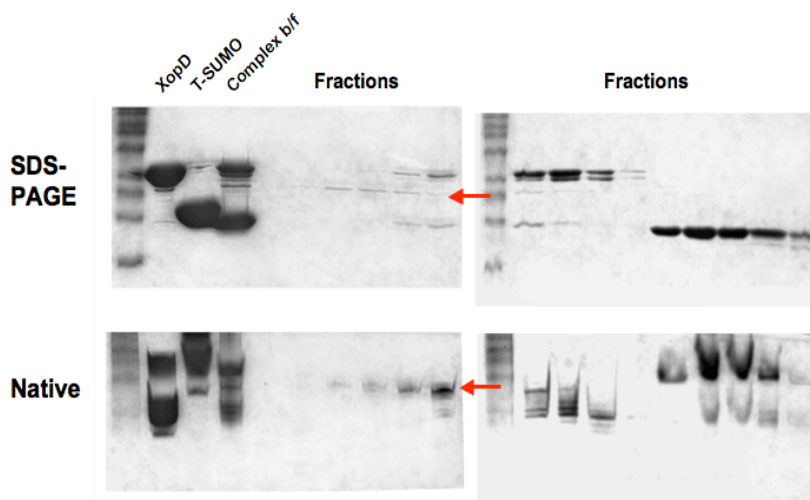
**Figure 32: A new XopD protein construct to be used in complex-forming assays by observing the cleavage products of the catalytic domain of XopD.** A. By looking at cleavage products from the full catalytic domain of XopD on an SDS gel stained with Coomassie Blue, another cleavage product that included more of this substrate binding domain than the XopD<sub>335-520</sub> construct was found (second band). B. Amino acid sequence of the XopD<sub>305-520</sub> protein construct. This construct, XopD<sub>305-520</sub>, has one potential self cleavage site in the protein at G495 and one potential thrombin cleavage site at R334 (noted by red arrows). These two sites were mutated to alanine residues in attempts to eliminate the cleavage of this protein construct.

Trial # 8 included using 10 mg of XopD<sub>305-520</sub> R334A, G495A with 30 mg of T-SUMO-His<sub>6</sub> in a one hour reaction at room temperature. The complex was then loaded onto a Superdex 75 gel filtration column and fractions were run on native as well as SDS gels (Figures 33). Figure 33 shows how the fractions that contained the complex also contained a presumed XopD<sub>305-520</sub> dimer. I attempted further purifications on a MonoQ column and was unable to separate the complex from the dimerized XopD<sub>305-520</sub>. Separation of the presumed complex from the dimerized XopD became a hindrance to future studies.

**A.**



B.



**Figure 33: Separation of XopD:T-SUMO complex from XopD protein was unsuccessful using gel filtration chromatography.** In the assay, 10 mg of XopD<sub>305-520</sub> R334A, G495A was incubated with 30 mg of T-SUMO-His<sub>6</sub> for 1 hour reaction at room temperature. The complex was then loaded onto a Superdex 75 gel filtration column (A) and fractions were run on native as well as SDS gels (B). The chromatogram from the Superdex 75 column in (A) shows how the fractions that contained the complex also contained a presumed XopD<sub>305-520</sub> dimer. This was confirmed with native and SDS PAGE gels in (B). Red arrows indicate the presumed protein complex band.

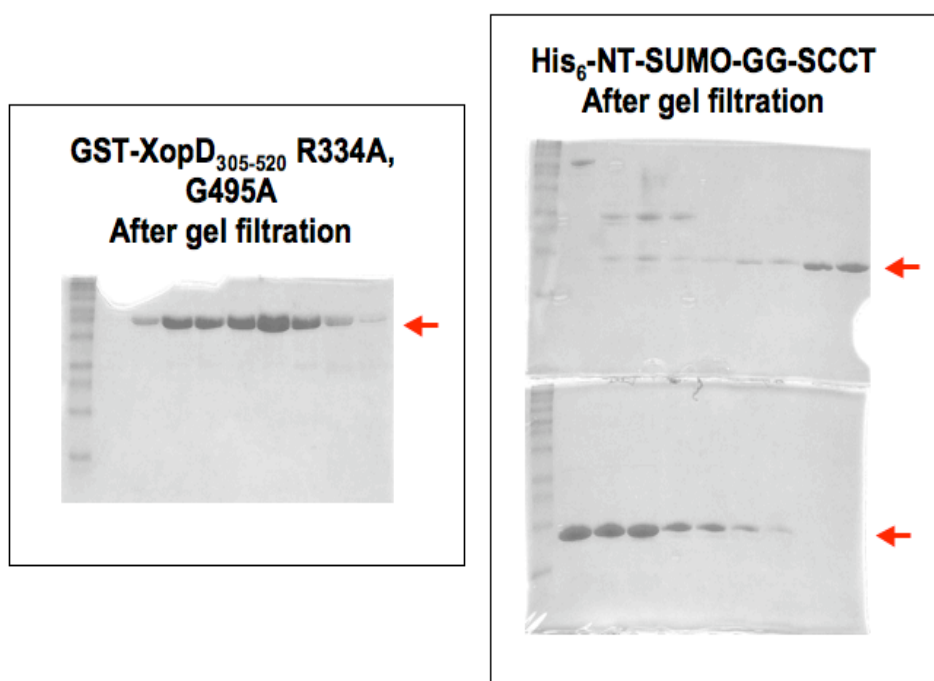
Another paper from the Lima lab reported the co-crystal structure of DEN1 with Nedd8-aldehyde (86). Further examination of the complex structures that had been solved revealed that the SUMO protein constructs used had a truncated N-terminus. The reason for this truncation is that original SUMO structure that was solved by NMR showed an unstructured or “floppy” N-terminus which would make crystallization difficult (Figure 1) (23). I decided to construct a T-SUMO construct with a 17 amino acid N-terminal truncation with a C-terminal His<sub>6</sub>tag.

In trial #9, this construct, NT-SUMO-His<sub>6</sub> (30mg), was used with 10mg of XopD<sub>305-520</sub> R334A, G495A in a complex forming assay. Unfortunately, there were no significant improvements in complex formation or in the ability to separate the complex from dimerized XopD.

Ron Hay and colleagues then reported the co-crystal structure of DEN1 with Nedd8 (similar to that of Reverter D. *et al.* (86)) (87). The method used for forming the complex was a modified version of that used in the Lima Lab. Hay and colleagues expressed and purified GST-tagged DEN1 and a His<sub>6</sub>-tagged Nedd8 (both proteins tagged on their N-terminus). The GST-tagged DEN1 had a TEV cleavage site after the GST moiety. The His<sub>6</sub>-Nedd8 construct contained four residues following its C-terminal GG motif. The SUMO and Smt3 constructs used by the Lima lab were His<sub>6</sub> tagged after the C-terminal GG motif. Both tagged GST-DEN1 and His<sub>6</sub>-Nedd8 were used for complex formation with sodium borohydride. This new method was then applied to experiments with XopD and T-SUMO.

For trial #10, as shown in Figure 34, His<sub>6</sub>-NT-SUMO-GG-SCCT and GST-XopD<sub>305-520</sub> R334A, G495A (with a TEV cleavage site after the GST) were constructed and purified. The tags on the proteins were used for purification of the complex as described by Hay and colleagues (87). In a similar manner, complex is formed with His<sub>6</sub>-NT-SUMO-GG-SCCT and GST-XopD<sub>305-520</sub> R334A, G495A with NaBH<sub>4</sub> as before. The complex is then run over a glutathione agarose column to remove any uncomplexed His<sub>6</sub>-NT-SUMO-GG-SCCT. The complex is then eluted from the glutathione agarose column, then buffer exchanged, and run over a nickel column to remove any unbound GST-XopD<sub>305-520</sub> R334A, G495A. The complex is then eluted with imidazole. The eluted complex is then cleaved with TEV to

remove the GST tag from GST-XopD<sub>305-520</sub> R334A, G495A. The final step is to run the complex (XopD<sub>305-520</sub> R334A, G495A : His<sub>6</sub>-NT-SUMO-GG) over a Superdex 75 gel filtration column to remove the free GST. This procedure was attempted with several large amounts of protein but the yield was very low using this multi-step purification. The order of the purification strategy was also altered to try and diminish the protein loss, but this method was also unsuccessful.



**Figure 34: GST-XopD<sub>305-520</sub> R334A, G495A and His<sub>6</sub>-NT-SUMO-GG-SCCT purified proteins samples.** GST-XopD<sub>305-520</sub> R334A, G495A was purified with the GST-tag intact using glutathione agarose, followed by a MonoQ column and then a Superdex 75 gel filtration column. Protein fractions from the gel filtration column are shown in the left panel. His<sub>6</sub>-NT-SUMO-GG-SCCT was purified using Ni<sup>++</sup> affinity chromatography followed by a MonoQ column and then a Superdex 75 gel filtration column. Protein fractions from the gel filtration column are shown in the right panel. Bands of the proteins of interest are noted by red arrows.

The co-crystal structure of DEN1:Nedd8 that was solved by the Lima lab made use of the Nedd8-Aldehyde to form their enzyme-substrate complex (86). The Nedd8-Aldehyde acts as an inhibitor of the DEN1 protein by forming a covalent bond with the catalytic cysteine of DEN1. The use of aldehyde derivatives of SUMO and ubiquitin proteins as covalent inhibitors of Ulp1s and DUBs is well characterized (54). The SUMO or ubiquitin aldehyde inhibitors form a thiohemiacetal intermediate with the cysteine in the active site of the DUB or Ulp1 (88). Attempts were made to form the T-SUMO-Aldehyde using the Carboxypeptidase Y catalyzed exchange for the C-terminal carboxylate with an aldehyde (88,89). The T-SUMO construct used to form the aldehyde was His<sub>6</sub>-TEV-NT-SUMO-GG, which included a TEV protease site for His<sub>6</sub> tag removal and the carboxy terminal -GG residues were exposed. The protocol outlined by Hu *et al.* involves using 3-amino-1,2-propanediol in an exchange reaction for the carboxylate group of the C-terminal glycine catalyzed by carboxypeptidase Y (88). The resulting diol is then purified with an ion exchange column. The diol is then oxidized to the aldehyde with NaIO<sub>4</sub> at an acidic pH (pH 4.55) for several hours. When this procedure was followed to make the T-SUMO-diol, it was determined that T-SUMO was unstable at the acidic pH, which is a step required for the reaction. After several attempts at this procedure, this method was abandoned.

Currently, trials for forming a complex between XopD C470S and NT-SUMO-GG-SCCT are in progress. Future studies are discussed in Chapter 8.



## CHAPTER SEVEN

### Results

#### EVOLUTION OF A SIGNALING SYSTEM THAT INCORPORATES BOTH REDUNDANCY AND DIVERSITY: *ARABIDOPSIS* SUMOYLATION

##### INTRODUCTION

As described in detail in Chapter 2, SUMO has been cloned from humans, yeast, and plants. To date, four mammalian SUMOs have been found and one yeast SUMO, Smt3. In plants, sequence mining of the *Arabidopsis thaliana* (*At*) genome by Vierstra and colleagues found eight genes encoding SUMO-like proteins (12,43).

SUMO utilizes a conjugation machinery (E1, E2, E3) to modify a target protein, similar to that of ubiquitin conjugation. SUMO is attached to a target protein via an isopeptide bond between its carboxy terminal glycine and the  $\epsilon$ -amino group of a lysine residue on the target protein. A family of cysteine proteases, referred to as ubiquitin-like protein protease-1(s) (ULP1s), are able to process SUMO to its mature form (exposing the conserved carboxy- terminal Gly-Gly residues) in order for SUMO to be utilized by the SUMOylation machinery (26,30). These proteases also act as isopeptidases by removing the SUMO moiety from SUMO conjugated target proteins. Seven ULP1 proteases have been found in humans and are referred to as SENP1, 2, 3, 5, 6, 7, and 8 (30). Vierstra and colleagues have identified four *ULP1* genes (*AtULP1A*, *B*, *C*, and *D*) encoded in the *Arabidopsis* genome and Coupland and colleagues have identified a fifth *ULP1* gene (*AtESD4*) (12,13).

To further our understanding on mechanisms utilized for the reversible post translational modification by SUMO, we investigated whether eukaryotes (plants and animals) encode such a large number of ULP1s and SUMOs to diversify this signaling system or to simply add redundancy to it. Our studies reveal that both of these mechanisms have been incorporated into the *Arabidopsis thaliana* SUMOylation signaling machinery. We have discovered that a great range of specificity for the different SUMO substrates is inherent in the catalytic core of the *Arabidopsis* ULP1s and that this system is further diversified by the addition of regulatory domains associated with the ULP1s. By contrast, we observed that the evolutionarily conserved E1 and E2 in the conjugation machinery do not discriminate amongst the various *Arabidopsis* SUMO proteins and are able to use all *Arabidopsis* SUMOs in conjugation reactions. Therefore, inherent in the *Arabidopsis* SUMO proteins is the information that dictates the specificity of their hydrolysis by isopeptidases information so that they can be uniformly recognized by the conjugation machinery.

## RESULTS AND DISCUSSION

### ***Arabidopsis* SUMOs are conjugated to RanGAP by evolutionarily conserved E1 and E2**

The sequences of the SUMO substrates used in this study can be aligned with high similarity, and include the four *At*SUMOs (*At*SUMO-1, *At*SUMO-2, *At*SUMO-3 and *At*SUMO-5), yeast Smt3, T-SUMO, M-SUMO-1, M-SUMO-2, and M-SUMO-4 (Figure 35A). Although mining of the *Arabidopsis* genome revealed eight SUMOs, only four are encoded by mRNA (12). *At*SUMO-1 and *At*SUMO-2 share 89% sequence identity with each

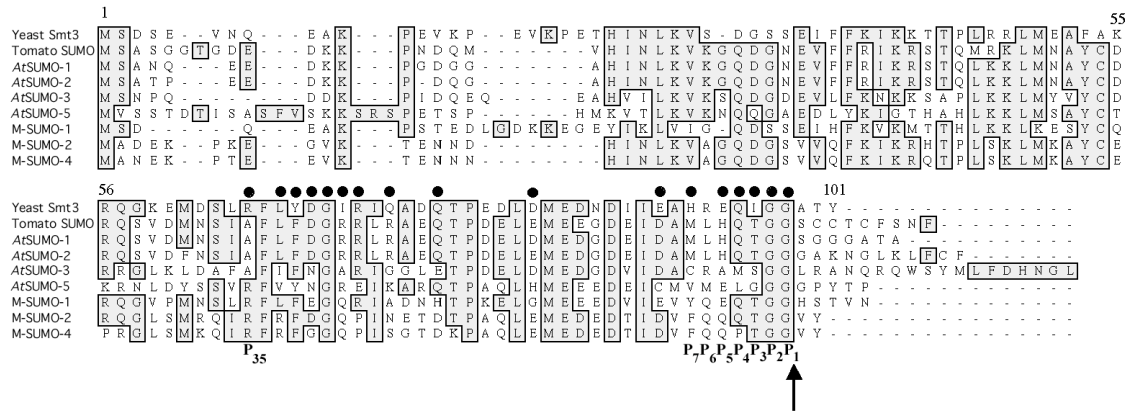
other and 83% with T-SUMO. For all of these proteins, the two conserved glycine residues in the carboxy-terminus mark the end of the processed form of the SUMOs.

To assess whether the various SUMOs can be used as substrates by the conjugation machinery, we constructed and purified recombinant N-terminally tagged GST-fusion proteins with the processed form of the *Arabidopsis* SUMOs, mammalian SUMOs, T-SUMO, and yeast Smt3 (GST-SUMO-GG-STOP). We utilized  $^{35}\text{S}$ -labeled RanGAP that was *in vitro* transcribed and translated in a rabbit reticulocyte lysate (RRL) as a target protein for the SUMOylation. As observed in Figure 1B (lane 1), a fraction of the translated  $^{35}\text{S}$ -labeled RanGAP was SUMOylated due to the presence of the endogenous SUMOylation machinery in RRL ( $^{35}\text{S}$ -RanGAP-RRL-SUMO) (Figure 35B, lane 1). The *in vitro* translated  $^{35}\text{S}$ -labeled RanGAP was used in an *in vitro* SUMOylation assay with recombinantly purified E1, E2 and the GST-SUMOs. The E1 and E2 conjugation machinery recognized all of the SUMOs as substrates and used them to modify RanGAP as observed by appearance of the GST-SUMO-RanGAP band (Figure 35B, lanes 2-9).

In crystallography studies, Reverter and Lima (2005), identified residues in Smt3 that interact with residues in its E2, Ubc9 (Figure 35A) (90). Smt3 uses Glu93 (P5) to make a salt bridge via its Arg 63 (P35) with UBC9 Glu122 (90). These residues are conserved in *At*SUMO-5, M-SUMO-1, M-SUMO-2 and M-SUMO-4. However, in T-SUMO, *At*SUMO-1 and *At*SUMO-2, the residue at P5 has been changed from a Glu (or Gln) to a His and the corresponding residue at P35 has been changed from an Arg to an Ala. We predict with this configuration that a His at P5 would directly interact with Ubc9 Glu122 to form a salt bridge. In the case of *At*SUMO-3, both P5 and P35 are replaced with an Ala and, therefore, these

changes do not hinder the conjugation process, as observed in Figure 35B. We propose that the residues that are required for recognition by Ubc9 alter in a manner that maintains the recognition of the SUMO molecules for Ubc9.

A



B



**Figure 35: SUMO family members can be utilized by SUMO conjugation machinery *in vitro*.** A. Sequence alignment of amino acids of yeast Smt3, Tomato-SUMO, AtSUMO-1, AtSUMO-2, AtSUMO-3, and AtSUMO-5. Numbering is shown with respect to yeast Smt3. Black dots above the alignment denote residues in yeast Smt3 that are in direct contact with yeast  $\Delta$ Ulp1 as described by Mossessova and Lima (11). The arrow indicates where SUMO substrates are cleaved by ULP1s. B. [<sup>35</sup>S]-mammalian-RanGAP was *in vitro* translated in a rabbit reticulocyte lysate. In the rabbit reticulocyte lysate, some of the RanGAP is SUMOylated by endogenous SUMOylation machinery (lane 1). The *in vitro* translated product was then used in an *in vitro* SUMOylation assay using recombinantly purified GST-AtSUMO-1, GST-AtSUMO-2, GST-AtSUMO-3, GST-AtSUMO-5, GST-T-SUMO, GST-Smt3, GST-M-SUMO-1, GST-M-SUMO-2, and GST-M-SUMO-4 to produce GST-SUMO-modified RanGAP. 5  $\mu$ L of the SUMO-modified RanGAP from each reaction was then added to 5x SDS sample buffer and the samples were resolved on 8% SDS gels and visualized by autoradiography.

### **The family of *Arabidopsis* ULP1s share limited sequence identity with yeast Ulp1**

Initially, we attempted to see whether SUMO specificity of *AtUlp1*-like enzymes could be predicted by alignment with various ULP1 family members and, as a guide, we used the co-crystal structure of the amino-terminally deleted yeast Ulp1 ( $\Delta$ Ulp1) with yeast Smt3 (11). In Figure 36A, we aligned the catalytic core ( $\Delta$ ) of the *AtULP1* proteases with yeast Ulp1 and designated the residues that are important for the interaction between yeast  $\Delta$ Ulp1 and its substrate, yeast SUMO, Smt3. Alignment of all of the enzymes was used to identify the amino terminal boundary of the catalytic domain for each *AtULP1* (Figure 36A). The four full-length *AtULP1*s share limited sequence identity with yeast Ulp1 (16-29% sequence identity), thereby making any predictions on substrate specificity difficult. By contrast, when the catalytic cores of *Arabidopsis* ULP1 family members are compared to one another, greater similarities are found to exist and the *AtULP1*s can be grouped into two pairs:  $\Delta$ *AtULP1A* and  $\Delta$ *AtESD4* with 65% sequence identity and  $\Delta$ *AtULP1C* and  $\Delta$ *AtULP1D* with 72% sequence identity (Figure 36B). On the basis of only contact residues identified in the yeast  $\Delta$ Ulp1:Smt3 structure, yeast  $\Delta$ Ulp1 is found to be most similar to  $\Delta$ *AtULP1A* (Figure 36A). Overall, it is difficult to determine which *AtULP1*-like enzymes either share a limited range of SUMO specificity as is observed with *Xanthomonas* XopD or exhibit more promiscuous activity as is seen with yeast Ulp1 (56).

To date, all enzymatic and structural studies have used only the carboxy terminal catalytic domain of ULP1s. However, recent studies have implicated the amino terminal regulatory domain in playing a role in substrate specificity (44). In this study, we have expressed and analyzed both the full-length (FL) and the catalytic core ( $\Delta$ ) of *AtULP1* family

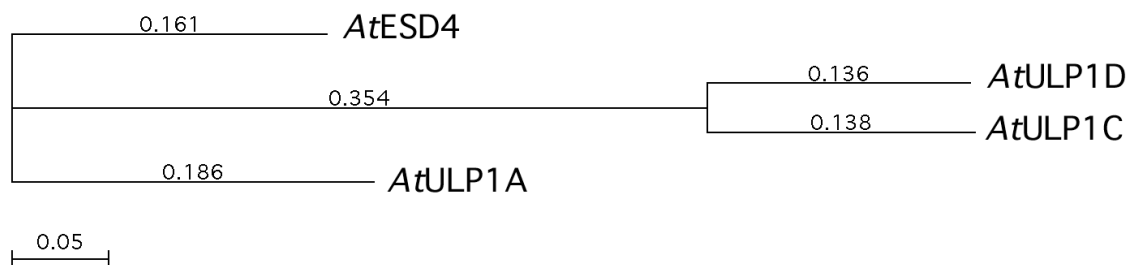
members to assess the impact of the regulatory domain on the specificity and activity of the catalytic domain.

# A

430	Yeast ΔUlp1	R D N I E I T V R D F K I T L A P R R W L N D T I I I E F F M K Y I E K - - - S T P N T V A - - - F N S F F Y T N L S E R G - - - G	483
	ΔAtULP1A	- - - I D I T G K I L R C L K P G K W L N D E V I N L Y M V L L K E R E A R E P K K F L K C H F F N T F F F T K L V N S A T - - - G	
	ΔAtULP1C	- - - V Q V S L K D L K C L S P P G E Y L T S P V I N F Y I R Y V Q H H V F S A D K I A A N C H F F N T F F Y K K L T E A V S Y K G	
	ΔAtULP1D	- - - V Q V C L K D L E C L A P R E Y L T S P V M N F Y M R F L Q Q I S S S N Q I S A D C H F F N T F F Y K K L S D A V T Y K G	
	ΔAtESD4	- - - I D I T G E V L Q C L T P S A W L N D E V I N V Y L E L L K E R E T R E P K K Y L K C H Y F N T F F Y K K L V S D S - - - G	
484	Yeast ΔUlp1	- - - Y Q G V R R W M K R K K T Q I D K L - - - D K I F I T P I N L N Q S H W A L G I I D L K K K T - - - I G Y V D S L S N - G P	537
	ΔAtULP1A	Y N Y G A V R R W T S M K R L G Y H L K D C D K I F I P I H M N - I H W T L A V I N I K D Q K - - - F Q Y L D S F K G R E P	
	ΔAtULP1C	N D R D A Y F V K F R R W W K G F D L F C K S Y I F I P I H E D - L H W S L V I I C I P D K E D E S G L T I I H L D S L G L H P R	
	ΔAtULP1D	N D K D A F F V R F R R W W K G I D L F R K A Y I F I P I H E D - L H W S L V I V C I P D K D E S G L T I L H L D S L G L H S R	
	ΔAtESD4	Y N F K A V R R W T T Q R K L G Y A L I D C D M I F V P I H R G - V H W T L A V I N N R E S K - - - L L Y L D S L N G V D P	
538	Yeast ΔUlp1	N - - A M S F A - I L T D L Q K Y V M E E - - - S K H T I G E D - - - F D L I H L D C P Q Q P N G Y D C G I Y V C M N T	588
	ΔAtULP1A	K - I L D A L A R Y F V D E V R D K S E V - - - D L D V S R W R - - - Q E F V Q D L P M Q R R N G F D C G M F M V K Y I	
	ΔAtULP1C	N L I F N N V K R F L R E E W N Y L N Q D - A P L D L P I S A K V W R D L P N M I N N E A E V Q V P Q K N D F D C G L F L F F I	
	ΔAtULP1D	K S I V E N V K R F L K D E W N Y L N Q D D Y S L D L P I S E K V W K N L P R R I S E A V V Q V P Q K N D F D C G P P V L F F I	
	ΔAtESD4	M - I L N A L A K Y M G D E A N E K S G K - - - K I D A N S W D - - - M E F V E D L P Q Q K N G Y D C G M F M L K Y I	
589	Yeast ΔUlp1	L Y G S A D A P L D F D Y K D - - - A I R M R R F I A H - - - L I L T D A L K - - - - -	621
	ΔAtULP1A	D F Y S R G L D L C F T Q E - - - - Q M P Y F R A R - - - T A K E I L Q L K A E - - -	
	ΔAtULP1C	R R F I E E A P Q R L T L Q Q D L K M I H K K W F K P E E A S A L R I K I W N I L V D L F R K G N Q T D	
	ΔAtULP1D	K R F I E E A P Q R L K R K D L G M F D K K W F R P D E A S A L R I K I R N T L I E L F R V S D Q T E	
	ΔAtESD4	D F F S R G L G L C F S Q E - - - - H M P Y F R L R - - - - T A K E I L R L R A D - - -	



B



**Figure 36: Sequence Alignments of the ULP1 Family Members.** A. Sequence alignment of amino acids in the catalytic core of yeast Ulp1, *AtULP1A*, *AtULP1C*, *AtULP1D*, and *AtESD4*. Numbering is shown with respect to yeast Ulp1. Catalytic Histidine, Aspartate, and Cysteine residues are shown in blue boxes. Black dots above the alignment denote residues in yeast Ulp1 that are in direct contact with yeast Smt3 as described by Mossessova and Lima (11). The catalytic domains of all *Arabidopsis* ULP1s are extended to their native carboxy terminus. B. Phylogenetic tree representation of amino acid sequence distance among *AtULP1*s used in this study. The tree was calculated using ClustalW (v1.4) algorithm in the MacVector program.

### The peptidase activity of ULP1 family members with *Arabidopsis* SUMOs reveals differences in substrate specificity

To analyze the peptidase activity of the ULP1s, we designed an assay whereby the SUMO substrates (SUMO-Gly-Gly-HA) are cleaved after the carboxy terminal Gly-Gly residues, resulting in a product that migrates faster than the substrate on SDS-PAGE gels (Figures 37 and 39). A recent report by Reverter *et al.* suggested that residues C-terminal of the conserved GlyGly motif of SUMOs might play a role in the specificity of ULP1 family members (86). We have observed only slight variability in the efficiency of the processing activity of  $\Delta$ XopD and yeast  $\Delta$ Ulp1 due to residues C-terminal to the GlyGly motif (R.C. and K.O., unpublished). To eliminate variability in the efficiency of cleavage due to the

differences at the carboxy terminus of SUMO substrates, thereby allowing the focus to be on differences in the mature SUMO proteins, we have utilized SUMO constructs that encode an Influenza A virus hemagglutinin (HA) tag directly following the GlyGly motif.

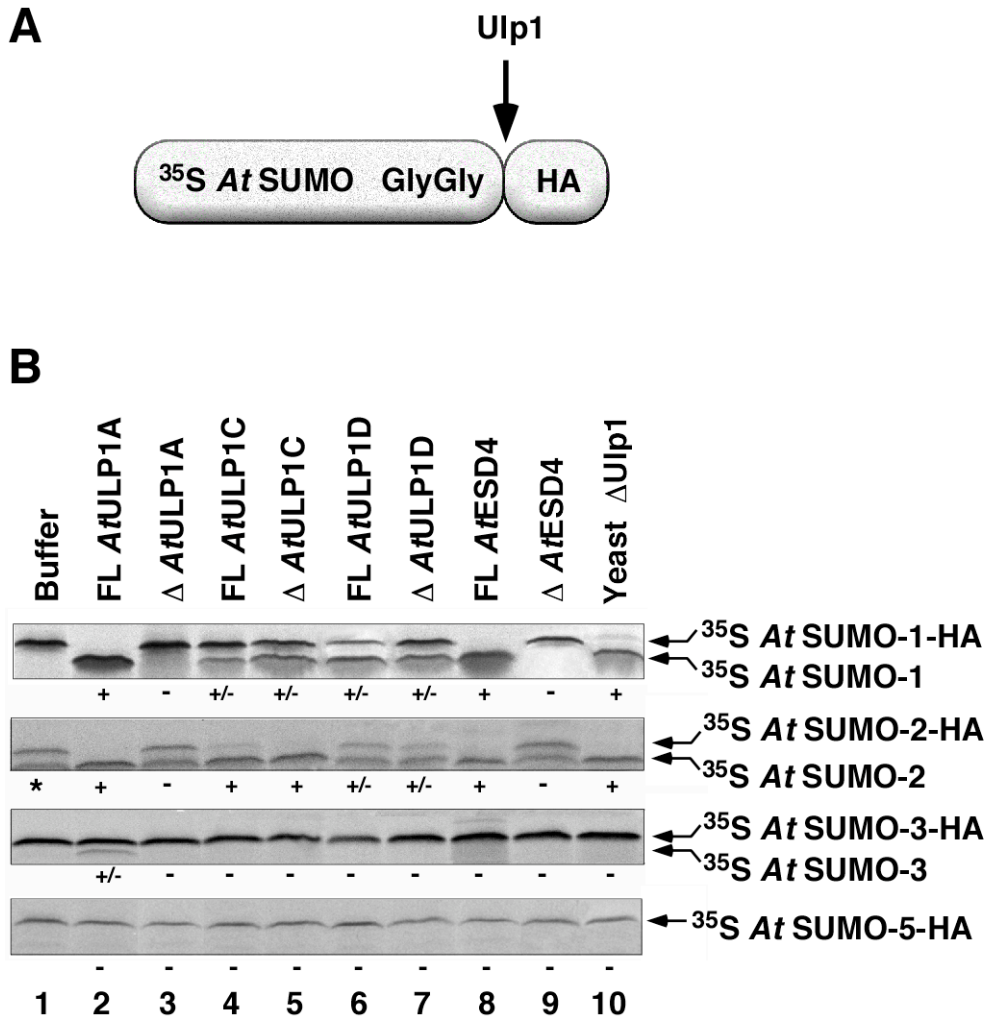
GST fusion proteins of yeast  $\Delta$ Ulp1 and the full-length and the catalytic core of *At*ULP1A, *At*ULP1C, *At*ULP1D, and *At*ESD4 were expressed in *E.coli* and purified using standard GST-glutathione affinity chromatography (Figure 38). All enzymes expressed as soluble proteins and the yield of the affinity purified proteins varied between 2 mg/L and 10 mg/L. Molar excess of each ULP1 is incubated with each  $^{35}$ S-labelled SUMO substrate for one hour at 30°C, to ensure cleavage of any potential substrate (Figures 37 and 39). Partial cleavage of specific SUMO substrates is denoted (Figures 37 and 39).

Within the first set of *in vitro* peptidase assays, we observed an unpredicted requirement for the regulatory domain of the *Arabidopsis* enzymes. Two of the four *Arabidopsis* ULP1s (*At*ULP1A and *At*ESD4) required their amino terminal regulatory domain for peptidase activity *in vitro* (Fig. 37B, lanes 2,3,8,9). Yet for the remaining two *Arabidopsis* enzymes, *At*ULP1C and *At*ULP1D, the activity of the catalytic core was unaffected by the presence of the regulatory domain (Fig. 37B, lanes 4,5,6,7,). Yeast  $\Delta$ Ulp1 is able to cleave both *At*SUMO-1 and *At*SUMO-2 regardless of whether or not its N-terminal domain is present (Fig. 37B, lane 10) (R.C. and K.O., unpublished). Only *At*ULP1A was able to cleave *At*SUMO-3, albeit weakly, and none of the ULP1-like enzymes cleaved *At*SUMO-5 (Fig. 37B, lane 2).

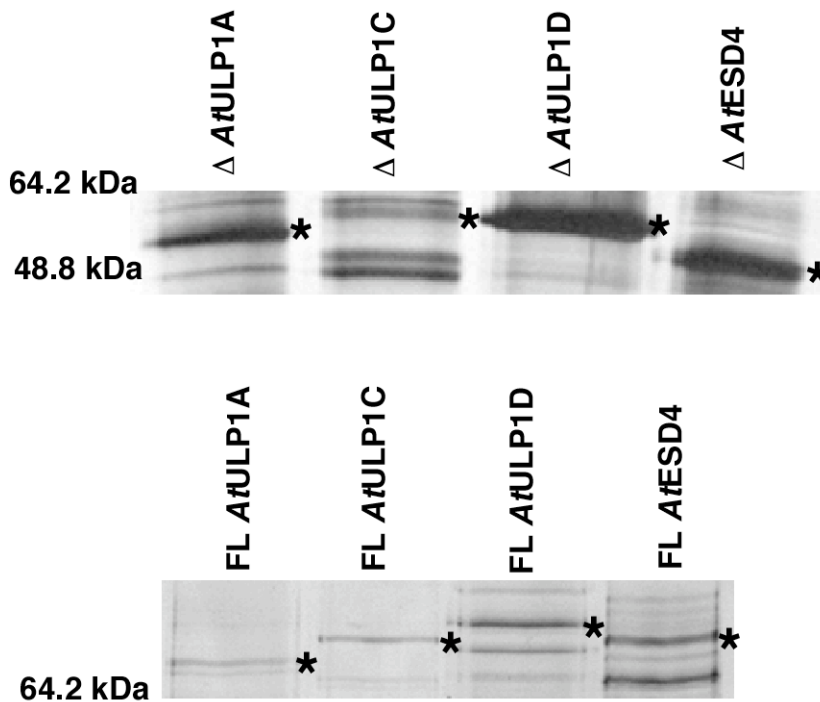
From these results, we observe a number of striking features on the specificity of the enzymes for various substrates. We observed similar activity profiles for the two *At*ULP1s

that encode catalytic cores that are most similar to one another (*AtULP1C* and *AtULP1D*; 72% sequence identity) (Fig. 35B). All of the active *AtULP1*s, and as predicted yeast  $\Delta$ Ulp1, were able to process *AtSUMO-1* and *AtSUMO-2*, which is not surprising, based on the substrate's 89% sequence identity. We observed a requirement for the N-terminal regulatory domain for the activity of *AtULP1A* and *AtESD4*, but not for *AtULP1C* and *AtULP1D*. These observations support the proposal that the N-terminal regulatory domains are able to affect the catalytic activity of the enzymes.

Insight was gained not only on the substrate specificity of the enzymes, but also on the ability of a substrate to be recognized as a substrate by a particular enzyme. Because none of the ULP1 family members tested can process the most distantly related *AtSUMO-5*, it is possible to predict those residues that may impede the processing of this SUMO. By scanning residues at the P<sub>1</sub>-P<sub>7</sub> positions, we observe that the conserved Gln residue at position P<sub>4</sub> is a Leu in *AtSUMO-5* and at position P<sub>7</sub> there is an uncharged Val residue (Fig. 35A). These residues may account for the lack of enzyme recognition of this SUMO substrate, since residues in these positions were observed to make hydrogen bonds between yeast  $\Delta$ Ulp1 and Smt3. In addition, many of the residues in Smt3 (R72E, D83H) that make contact with yeast  $\Delta$ Ulp1, encode charge reversals at the predicted residues in *AtSUMO-5* (Fig. 35A).



**Figure 37: *In vitro* peptidase activity of *Arabidopsis Thaliana* ULP1 family members.** A. Schematic of *in vitro* SUMO peptidase assay. B. *At*ULP1 family members show differing peptidase activity for *At*SUMO substrates. [<sup>35</sup>S]- *At*SUMO-1-HA, *At*SUMO-2-HA, *At*SUMO-3-HA, and *At*SUMO-5-HA were *in vitro* translated in a rabbit reticulocyte lysate and then incubated with buffer or 0.5 mg/mL of full-length (FL) *At*ULP1A, C, D and *At*ESD4, and the catalytic core (Δ) of *At*ULP1A, C, D and *At*ESD4, or yeast Ulp1 for 1 hour at 30°C. The samples were then resolved on 17% SDS gels and visualized by autoradiography. (+) symbol indicates total cleavage of the SUMO-HA, (+/-) symbol indicates partial cleavage of SUMO-HA, and (-) symbol indicates no detectable cleavage of SUMO-HA by each ULP1. The asterisk (\*) indicates that when *At*SUMO-2-HA is *in vitro* translated in RRL, background peptidase activity is detected (lane 1). This peptidase activity is probably due to a rabbit peptidase present in the RRL.



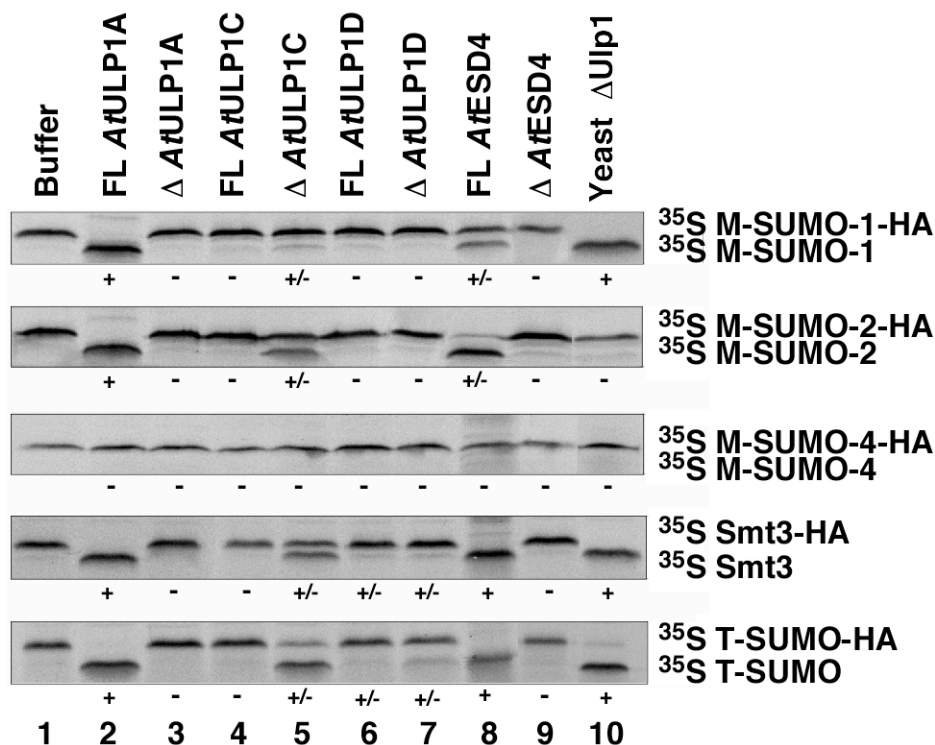
**Figure 38: Silver-stained gels of recombinant *Arabidopsis thaliana* Ulp1 proteins.** Full-length (FL) *AtULP1A*, *C*, *D* and *AtESD4* (bottom panel), and the catalytic core ( $\Delta$ ) of *AtULP1A*, *C*, *D* and *AtESD4* (lower panel) were purified as GST fusion proteins from bacteria. The proteins were purified using glutathione agarose, eluted, concentrated, and then run on SDS PAGE gels and visualized using silver staining protocol.

### Peptidase Activity of ULP1 family members with tomato, mammalian, and yeast SUMOs

To learn more about the substrate specificity of these enzymes for various SUMO substrates, the panel of GST-ULP1s was assayed with another group of SUMOs; *in vitro* translated  $^{35}\text{S}$ -labelled M-SUMO-1, M-SUMO-2, M-SUMO-4, yeast Smt3, and T-SUMO. As observed previously with *AtULP1A* and *AtESD4*, only the full-length forms of these

ULP1s exhibited protease activity (Fig. 37B). These two enzymes cleave all the aforementioned substrates with the exception of M-SUMO-4 (Fig. 39, lanes 2 and 8). In the case of *At*ULP1C, full length *At*ULP1C did not cleave any of the substrates (Fig. 39, lane 4), whereas  $\Delta$ *At*ULP1C cleaved M-SUMO-1, M-SUMO-2, Smt3 and T-SUMO, but not M-SUMO-4 (Fig. 39, lane 5). Both FL *At*ULP1D and  $\Delta$ *At*ULP1D are only able to partially cleave T-SUMO and Smt3 (Fig. 39, lane 6, 7). Yeast  $\Delta$ Ulp1 cleaved all substrates with the exception of M-SUMO-2 and M-SUMO-4 (Fig. 39, lane 10). The cleavage profile for yeast Ulp1 and this panel of substrates is the same, regardless of whether or not its N-terminal domain is present (R.C. and K.O., unpublished).

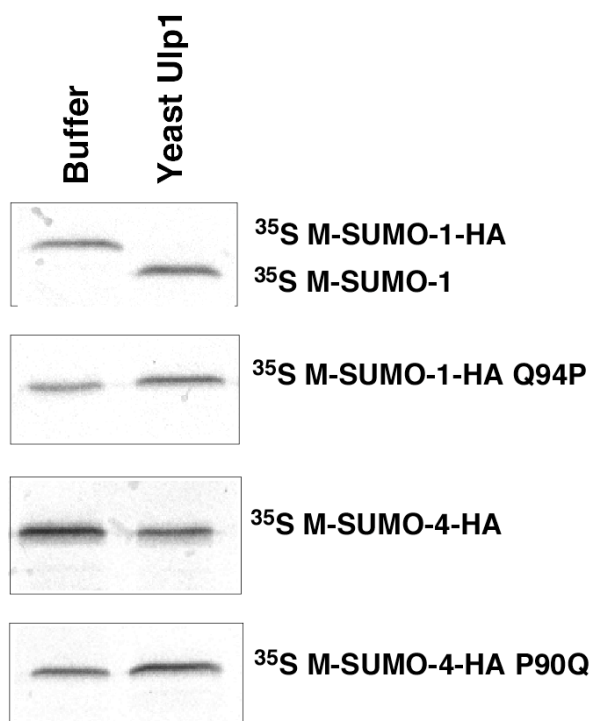
Using this panel of substrates, we can subdivide the ULP1s based on their substrate specificity. The most promiscuous enzyme is *At*ULP1A, albeit it requires its regulatory domain for its catalytic activity. In contrast, *At*ULP1C is active only when expressed without its regulatory domain. These observations support our hypothesis that the regulatory domain of the enzymes plays a key role in not only modulating the enzyme activity but also substrate specificity.



**Figure 39: Specificity of Plant, Animal, and Yeast ULP1 family members for their SUMO substrates.** *At* ULP1 family members show differing peptidase activity for mammalian, yeast, and tomato SUMO substrates. [ $^{35}\text{S}$ ]- M-SUMO-1-HA, M-SUMO-2-HA, M-SUMO-4-HA, yeast Smt3-HA, and Tomato-SUMO-HA (T-SUMO-HA) were *in vitro* translated in a rabbit reticulocyte lysate and then incubated with buffer or 0.5 mg/mL of full-length (FL) *At*ULP1A, C, D and *At*ESD4, and the catalytic core ( $\Delta$ ) of *At*ULP1A, C, D and *At*ESD4, or yeast Ulp1 for 1 hour at 30°C. The samples were then resolved on 17% SDS gels and visualized by autoradiography. (+) symbol indicates total cleavage of the SUMO-HA, (+/-) symbol indicates partial cleavage of SUMO-HA, and (-) symbol indicates no detectable cleavage of SUMO-HA by each ULP1.

None of the ULP1 family members can process M-SUMO-4. The most obvious difference distinguishing this SUMO from others is the proline (Pro) residue found at the P<sub>4</sub> position, as opposed to the conserved glutamine (Gln) residue (Fig. 35A). When M-SUMO-1 is mutated from a Gln residue to a Pro residue at the P<sub>4</sub> position, yeast  $\Delta$ Ulp1 is unable to cleave the mutated M-SUMO-1, indicating that the Pro at P<sub>4</sub> has caused a change in the structure of the M-SUMO-1 that is not compatible with substrate recognition (R.C. and K.O., unpublished). However, when M-SUMO-4 is changed from a Pro at P<sub>4</sub> to a Gln, yeast  $\Delta$ Ulp1 is still unable to cleave M-SUMO-4 (Figure 40). Thus this suggests the presence of other residues in M-SUMO-4 that are necessary for substrate recognition by yeast  $\Delta$ Ulp1.





**Figure 40: Mutation of Gln94 to a Pro in M-SUMO-1 abolished Ulp1 processing, but mutation of P90 to a Gln in M-SUMO-4 did not allow Ulp1 processing.** [ $^{35}\text{S}$ ]- M-SUMO-1-HA, M-SUMO-1-HA Q94P, M-SUMO-4-HA, and M-SUMO-4-HA P90Q were *in vitro* translated in a rabbit reticulocyte lysate and then incubated with buffer or 0.5 mg/mL yeast Ulp1 for 1 hour at 30°C. The samples were then resolved on 17% SDS gels and visualized by autoradiography.

### **Isopeptidase Activity of *Arabidopsis Thaliana* ULP1 family members**

All assays presented thus far have examined the SUMO specificity of ULP1s using peptidase assays. To further characterize the activity and specificity of this panel of proteases, we analyzed their specificity using isopeptidase assays. We utilized, as a substrate, RanGAP modified by various recombinant GST-SUMO proteins (Figure 35B). As observed previously (Figure 35B), a fraction of the translated  $^{35}\text{S}$ -labeled RanGAP is modified by the RRL endogenous SUMOylation machinery ( $^{35}\text{S}$ -RanGAP-RRL-SUMO).

Further modification by recombinantly purified GST-M-SUMO-1, GST-T-SUMO or GST-*At*SUMO-3 produced a population of GST-SUMOylated  $^{35}\text{S}$ -labeled RanGAP (Figure 35B and 41B). Therefore, each isopeptidase reaction includes radiolabelled  $^{35}\text{S}$ -RanGAP,  $^{35}\text{S}$ -RanGAP-RRL-SUMO,  $^{35}\text{S}$ -RanGAP-GST-SUMO and a test protease. Each protease is tested for activity with the same target protein, RanGAP, which is modified by either RRL-SUMO (as an internal control for each reaction) or recombinant GST-SUMO (Fig. 41A). We selected GST-M-SUMO-1, GST-T-SUMO, and GST-*At*SUMO-3 for *in vitro* SUMOylation assays for a number of reasons. First, GST-M-SUMO-1 serves as a GST control for the endogenous SUMO added to RanGAP during the *in vitro* translation reaction. Second, GST-M-SUMO-1 is variably recognized as a peptidase substrate. Third, GST-T-SUMO is recognized as a peptidase substrate by all full-length *At* proteins (except *At*ULP1C), and yeast  $\Delta\text{Ulp1}$ . Fourth, FL *At*ULP1A only processes GST-*At*SUMO-3. Thus, the three SUMOs provide a range of SUMO substrates in an internally controlled isopeptidase assay.

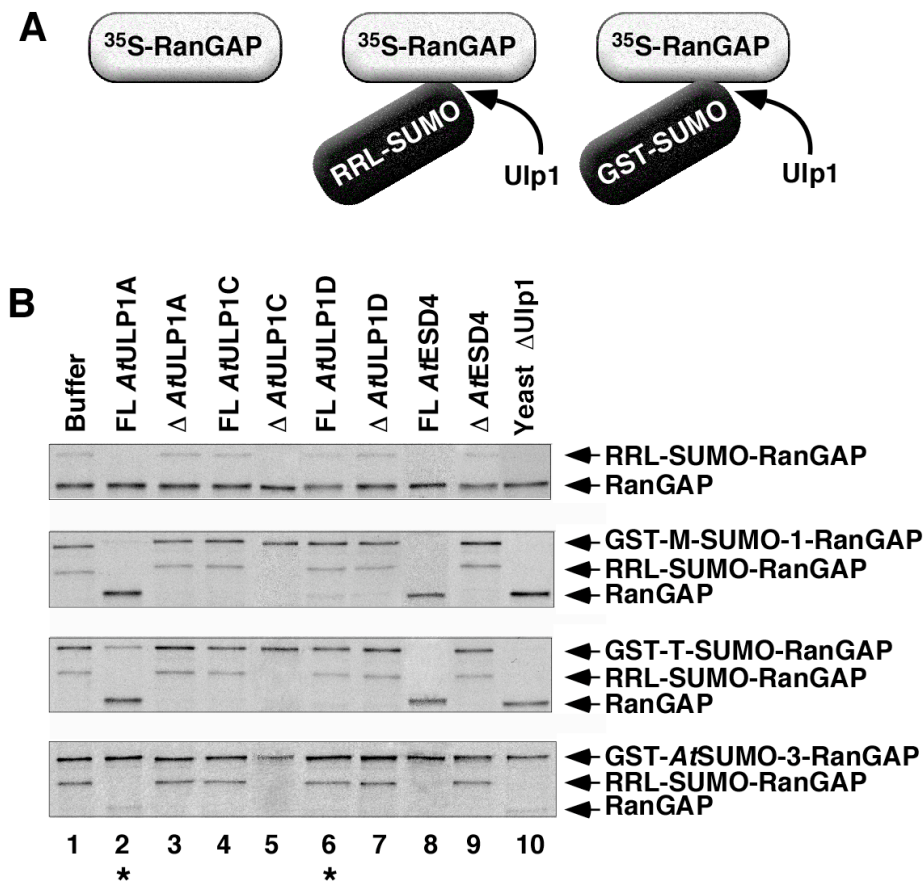
In the first isopeptidase assay,  $^{35}\text{S}$ -RanGAP-RRL-SUMO is used as substrate for the panel of ULP1s (Fig. 41B). Only full-length *At*ULP1A and *At*ESD4,  $\Delta\text{At}$ ULP1C, and yeast  $\Delta\text{Ulp1}$  process the endogenous SUMO from RanGAP (Fig. 41B, lanes 2, 5, 8, 10). This profile is similar to what was observed in the peptidase assays with M-SUMO-1.

In the second isopeptidase assay, both  $^{35}\text{S}$ -RanGAP-RRL-SUMO and  $^{35}\text{S}$ -RanGAP-GST-M-SUMO1 are used as substrates. The results are identical to the first isopeptidase assay with the exception of one enzyme,  $\Delta\text{At}$ ULP1C (Fig. 41B, lanes 2, 5, 8, 10).  $\Delta\text{At}$ ULP1C is able to cleave the RRL-SUMO but not the GST-M-SUMO1 (Fig. 41B, lane 5). Therefore,  $\Delta\text{At}$ ULP1C is able to recognize M-SUMO-1 as a substrate only in the peptidase

assay. Among the reasons that differences are observed for  $\Delta AtULP1C$  could be that the rabbit SUMO (RRL-SUMO) varies enough from the human SUMO (GST-M-SUMO-1) that it cannot be cleaved by  $\Delta AtULP1C$ . By contrast, FL  $AtULP1D$  is able to recognize and partially cleave GST-M-SUMO-1-RanGAP in the isopeptidase assay (Fig. 41B, lane 6), but neither FL nor  $\Delta AtULP1D$  are able to recognize M-SUMO-1 as a substrate in peptidase assays (Fig. 39, lanes 6,7).

In the third isopeptidase assay, the panel of ULP1s is tested using  $^{35}\text{S}$ -RanGAP-RRL-SUMO and  $^{35}\text{S}$ -RanGAP-GST-T-SUMO as substrates (Fig. 41B). All enzymes, with the exception of  $\Delta AtULP1C$ , that recognized T-SUMO as a peptidase substrate are able to cleave  $^{35}\text{S}$ -RanGAP-GST-T-SUMO. Finally, in the fourth isopeptidase assay, none of ULP1 family members tested are able to remove the GST- $At$ SUMO-3 modification from RanGAP (Fig. 41B). In all isopeptidase assays, the profile of cleavage of the internal control,  $^{35}\text{S}$ -RanGAP-RRL-SUMO, is consistent.

Some obvious differences and similarities are seen when comparing the *in vitro* isopeptidase and peptidase activities of these enzymes. FL  $AtULP1A$  is extremely efficient at cleaving peptide bonds, but less efficient with isopeptide bonds. In contrast,  $AtESD4$  cleaves SUMO conjugated proteins better than SUMO peptides. Both the peptidase and isopeptidase activity of the ULP1s are affected by the presence or absence of their amino terminal regulatory domain. The modulation of an enzymatic domain by a regulatory domain is a reoccurring theme in many signaling systems, including phosphatases and kinases (91,92).



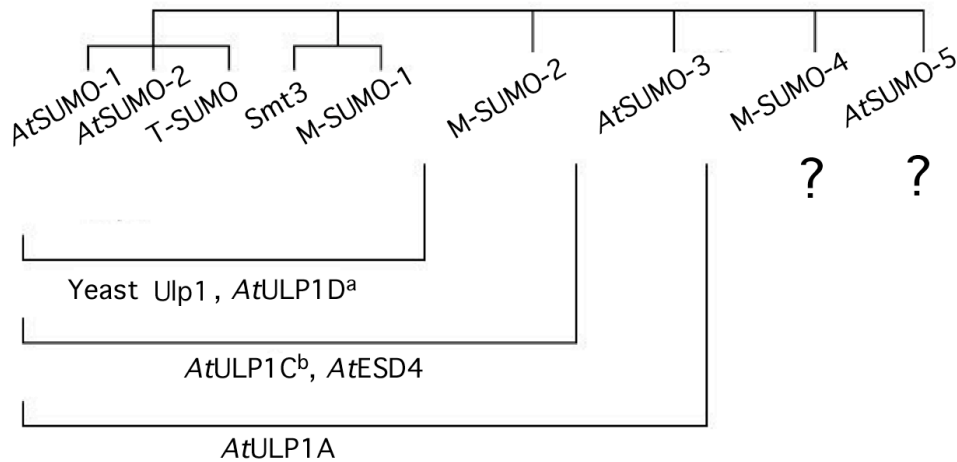
**Figure 41: *Arabidopsis thaliana* ULP1 family members exhibit isopeptidase activity *in vitro*.** A. Schematic of *in vitro* SUMO isopeptidase assay. B. The ULP1 family of proteases shows specificity for the SUMO moiety of SUMOylated substrates. [ $^{35}\text{S}$ ]-mammalian-RanGAP was *in vitro* translated in a rabbit reticulocyte lysate. In the rabbit reticulocyte lysate, some of the RanGAP is SUMOylated by endogenous SUMOylation machinery. The *in vitro* translated product was then used in an *in vitro* SUMOylation assay using recombinantly purified GST-M-SUMO-1, GST-T-SUMO or GST-*At*SUMO-3 to produce GST-SUMO-modified RanGAP. The SUMO-modified RanGAP reaction was then incubated with buffer or 0.5 mg/mL of full-length (FL) *At*ULP1A, *At*ULP1C, *At*ULP1D and *At*ESD4, and the catalytic core ( $\Delta$ ) of *At*ULP1A, *At*ULP1C, *At*ULP1D, *At*ESD4, or yeast Ulp1 for 1 hour at 30°C. The samples were then resolved on 8% SDS gels and visualized by autoradiography. Asterisks (\*) indicate partial cleavage of GST-T-SUMO-RanGAP by FL *At*ULP1A and GST-M-SUMO-1-RanGAP by FL *At*ULP1D.

## CONCLUSIONS

### **Both redundancy and diversity are exhibited by the *Arabidopsis* SUMOylation system**

The aforementioned observations demonstrate that the deSUMOylation of protein SUMO conjugates is a highly regulated and complex system. While cellular localization of proteases may account for part of the regulation of SUMOylated proteins, we observe, intrinsic to these proteases, is encoded another layer of specificity that dictates substrate recognition and cleavage. From our characterization of ULP1 family members expressed in *Arabidopsis*, we conclude that this family of proteases exhibits substrate specificity, both for the processing of SUMO and for the cleavage of SUMO conjugates. The substrate specificity of the enzymes used in this study is depicted in Figure 42. Assays with both the catalytic core and the full-length ULP1 proteins reveal variability in the requirement for the amino terminal regulatory domain.

The most striking examples of enzyme regulation are observed by the inhibitory and activating affects of the regulatory domain on the catalytic domain. In one case, the activity of *AtULP1C* was inhibited by the presence of its regulatory domain when processing certain substrates (Figure 39). These observations are reminiscent of the classic regulation of Src kinase activity by its amino terminal domain (92). In another case, the regulatory domain is required for both peptidase and isopeptidase activity for two of the *AtULP1*s (*AtULP1A* and *AtESD4*), and for another is necessary only for its peptidase activity (*AtULP1D*). Thus the role for the regulatory domain in these proteases varies from one protein to another, further demonstrating the diversity in the ULP1 family of enzymes.



**Figure 42: Classification of SUMOs and ULP1 family members based on activity.** SUMOs are grouped based on the ULP1 family members that cleave them as described in this paper. The classification is based on full-length and/or catalytic core ability to demonstrate peptidase and/or isopeptidase activity for SUMO substrates. <sup>a</sup>*AtULP1D* shows peptidase activity for *AtSUMO-1*, *AtSUMO-2*, T-SUMO, and Smt3, yet it only shows isopeptidase activity for GST-M-SUMO-1 conjugated to RanGAP. <sup>b</sup>*AtULP1C* show peptidase activity for *AtSUMO-1*, *AtSUMO-2*, T-SUMO, Smt3, M-SUMO-1, and M-SUMO-2, yet it only shows isopeptidase activity for RRL-SUMO conjugated to RanGAP.

In our studies, none of the *AtULP1*s tested were able to recognize *AtSUMO-5*, and only one enzyme was able to recognize *AtSUMO-3*. The *AtULP1*s used in our studies had been classified based on sequence identity with known ULP1s. Recently, a de-ubiquitinating enzyme (DUB) from Herpes Simplex Virus 1 (HSV-1) was discovered, and interestingly, this enzyme showed no sequence homology with any known de-ubiquitinating enzyme (93). Based on this finding, it is possible that other proteins in *Arabidopsis* have ULP1 activity, but do not share sequence homology to known ULP1s. Thus a ULP1 may exist in *Arabidopsis* that is able to recognize and hydrolyze *AtSUMO-5*.

Our results support the hypothesis that, not only is their sequence diversity, but also mechanistic diversity in this protein family. Our findings support the hypothesis that although genome duplication has occurred, these protein families have diversified and are not simply functionally redundant copies of one another. The four *Arabidopsis* ULP1s studied here have unique catalytic profiles with respect to their substrate specificity. Expression of the catalytic core is not always sufficient for observing the activity of an ULP1. In addition, in some cases, the regulatory domains can influence the specificity of these enzymes.

While we have observed specificity among the deSUMOylating enzymes for their SUMO substrates, we have also observed that the evolutionarily conserved SUMOylation machinery is promiscuous in its choice of SUMO substrate. We observed that all SUMOs can be used as substrates for conjugation, which supports the proposal that all SUMOs used in this study have conserved residues utilized for their interaction with Ubc9. Thus differences in the amino acid sequence of each SUMO determines which Ulp1 will process which SUMO, yet similarity is maintained to allow for their recognition by the SUMOylation machinery. Further more, we predict that the residues important for dictating specificity by the ULP1s, may also be important for the SUMO E3 ligases in specifying which SUMO is added to a target substrate.

Our studies contribute to the basic understanding of the complexity of this transient and reversible system of regulation and are reminiscent of many other systems involved in reversible posttranslational modifications.

## CHAPTER EIGHT

### Discussion and Conclusions

#### **deSUMOylating Enzymes: What was known in the beginning**

The field of SUMO, SUMOylation, and deSUMOylation began with the initial discovery of SUMO. One of the first reports on SUMO demonstrated that RanGAP is modified by mammalian SUMO-1 (M-SUMO-1) (14). The authors were able to use peptide sequencing to determine that this 90 kDa form of RanGAP contained ubiquitin-like protein modification, later referred to as SUMO (14). In 1997, Johnson *et al.* identified the yeast SUMO protein, Smt3 (19). This group showed that Smt3 had to be processed to its mature form, exposing its carboxy terminal Gly-Gly residues before it could be conjugated to target proteins (19). Shortly after the discovery of SUMO, the E1, E2, and E3s for the SUMO conjugation system were identified (19-21).

In 1999, Li and Hochstrasser identified Ulp1 (Ubiquitin-like protein protease), as a SUMO deconjugating enzyme in yeast (3). The authors show that Ulp1 is an essential gene in yeast and that Ulp1 could reduce the number of total Smt3 protein conjugates in yeast extracts (3). Later that same year, Suzuki *et al.* identified a de-SUMOylating enzyme in bovine brain, and suggested that there could be more than one of these proteases in mammalian cells (24). In 2000, Yeh and colleagues reported the cloning of Senp1, a human sentrin (SUMO)-specific protease that was able reduce the levels of SUMO-conjugated proteins in cell lysates (46). They showed that SENP1 was unable to remove the SUMO modification from RanGAP, and at the time, suggested that this was due to the localization of



the SENP1 (localized to nuclear bodies) relative to SUMO-modified RanGAP (localized to nuclear pore complex) (46). Seven ULP1 proteases have since been found in humans and are referred to as SENP1,2,3,5,6,7, and 8 (30,45). All of the SENPs share the conserved carboxy terminal catalytic domain with yeast Ulp1, yet differ in their size, amino terminal regulatory domain sequence, substrate preference, as well as cellular localization (30,45).

A family of five *ULP1* genes (*AtULP1A*, *B*, *C*, *D* and *ESD4*) were found to be encoded in the *Arabidopsis* genome (12,13). One *Arabidopsis thaliana* protease, ESD4, has been characterized as a SUMO protease in *Arabidopsis thaliana* (13).

In 2003, Li and Hochstrasser used yeast genetics in a study on the function of the amino terminal domain of yeast Ulp1 *in vivo* (44). They concluded that the amino terminal domain of yeast Ulp1 was responsible for the NPC localization of Ulp1 and it allowed for more efficient processing of Smt3-modified substrates based on genetic studies(44).

The cellular distribution of the human SENPs is thought to account for the large number of SENPs and possibly their substrate specificity. While cellular localization accounts for the subset of SUMOylated proteins available to certain ULP1s under certain cellular conditions, we demonstrated that another layer of specificity intrinsic to these proteases is responsible for substrate recognition and cleavage.

Another protein, *Yersinia* YopJ, was proposed to have deSUMOylating activity based on sequence similarity with the catalytic domain of yeast Ulp1 (72,75). YopJ is an effector protein secreted from *Yersinia spp.* In 2002, Noel *et al.* identified XopD (*Xanthomonas* Outer Protein D) as a secreted effector protein in the plant pathogen, *Xanthomonas campestris* pathovar *vesicatoria* (1). XopD showed sequence homology within its carboxy

terminal domain with yeast Ulp1 (4). Thus the hypothesized role of *Xanthomonas* XopD, based on its sequence similarity with Ulp1 and that it is a type III secreted effector protein, was as a deSUMOylating protease that targets SUMO-conjugated proteins in host plant cells during *Xanthomonas campestris* infection.

The objectives of this project were to characterize the *in vitro* activity (using biochemical and structural studies) of XopD, and extend this study to better understand other ULP1 family members and their SUMO substrate specificity.

### **Discussion of Research Findings**

*Xanthomonas* XopD is a cysteine protease, with deSUMOylating activity. XopD exhibits peptidase as well as isopeptidase activity demonstrated by its ability to process SUMO to its mature form as well as process SUMO protein conjugates. The activity of XopD can be inhibited by mutating its catalytic cysteine to an alanine (or serine) or by treatment with the chemical inhibitors, N-ethylmaleimide or iodoacetamide. XopD exhibits species specificity for its SUMO substrates by only processing T-SUMO, *At*SUMO-1, and *At*SUMO-2. Unlike yeast Ulp1, XopD is more rigid in its choice of SUMO substrate. The crystal structure of XopD (and XopD C470A) shows differences when compared to the crystal structures of yeast Ulp1 and human SENP2. Although the catalytic core residues align between the structures, there are other apparent structural differences that may account for the different choices of SUMO substrate. Mutational studies on T-SUMO and yeast Smt3 identified some key residues in the enzyme-substrate recognition, but ultimately showed that

there are many regions in the SUMO proteins that account for the enzyme's ability to recognize certain substrates.

Experiments with the *Arabidopsis thaliana* family of SUMOs and ULP1s identified varying specificity of the *At*ULP1s. These studies also showed that in some cases, the N-terminal regulatory domain of the ULP1s plays a role in specificity of the enzyme. All SUMOs (plant, yeast, and human) used in the experiments in Chapter 7 were used by the SUMOylation machinery, while the deSUMOylating enzymes were specific in their choice of SUMO to process. Thus the hypothesis was put forward that the basic E1 and E2 of the SUMOylation machinery is promiscuous in its choice for SUMO substrates, while the deSUMOylating enzymes exhibit great discretion for their choice of SUMO substrate.

*Xanthomonas XopD is a plant-specific deSUMOylating enzyme*

*Xanthomonas XopD* is an active cysteine protease that exhibits plant-specific SUMO specificity. The characterization of the activity of *XopD* represented the first evidence demonstrating a functional role for a phytopathogenic bacterial TTSS effector *in planta*. By uncovering the deSUMOylating function of *XopD*, this work has identified a novel mechanism used by *Xanthomonas*, and possibly other plant-associated microbes, to modulate plant physiology during infection. Similar to what has been observed for other bacterial effector proteins (e.g. YopH, YopE, YopJ and YpkA), we hypothesize that *Xcv* has usurped the activity of a eukaryotic isopeptidase (namely, Ulp1) and uses this activity to alter directly SUMO protein targets in the plant nucleus that control plant susceptibility and/or plant defense (72,83). After the study on *XopD*'s functional role was published, another study in

2004 reported that a cytoplasmic protein from *Xanthomonas*, AvrXv4, was shown act as a deSUMOylating enzyme *in planta* (57). As bacteria do not encode either the ubiquitin or the SUMO signaling machineries, *Xanthomonas* appears to have usurped the activity of eukaryotic ULP1s, as shown by XopD and AvrXv4, and used this activity to aid in pathogenesis by disrupting host defense signaling in the infected plant cell (56).

XopD exhibits species specificity for its SUMO substrate, whereby XopD will only process plant SUMOs. Yeast Ulp1 acts as a more promiscuous deSUMOylating enzyme as compared to XopD since yeast Ulp1 will process a variety of SUMO substrates. By performing structural studies to better understand the substrate specificity of XopD, it appears that XopD differs structurally from Ulp1, and this difference may account for the difference in SUMO specificity. Mutational studies with SUMOs confirmed previous data from structural studies that shows that extensive contacts are made between the Ulp1 family member and its SUMO substrate (11,84).

The studies regarding the SUMO specificity of XopD (described in Chapter 4 and 5) provide some answers as to the rigid specificity of XopD, yet a better understanding of the role of the amino-terminal domain and the actual SUMO-modified targets in plants could provide more answers.

#### *Evolution of a Signaling System that Incorporates Both Redundancy and Diversity:*

##### *Arabidopsis SUMOylation*

The studies described in Chapter 7 involving the *Arabidopsis thaliana* system demonstrate that the deSUMOylation of protein SUMO conjugates is a highly regulated and

complex system. While cellular localization of proteases may account for part of the regulation of SUMOylated proteins, we observe, intrinsic to these proteases, is encoded another layer of specificity that dictates substrate recognition and cleavage. From this characterization of *At*ULP1 family members, it was shown that this family of proteases exhibits substrate specificity, both for the processing of SUMO and for the cleavage of SUMO conjugates. Assays with both the catalytic core and the full-length ULP1 proteins reveal variability in the requirement for the amino terminal regulatory domain.

These results support the hypothesis that, not only is their sequence diversity, but also mechanistic diversity in this protein family. Although genome duplication may have occurred in *Arabidopsis*, these protein families have diversified and are not just functionally redundant copies of one another. The four *Arabidopsis* ULP1s studied here have unique catalytic profiles with respect to their substrate specificity. Expression of the catalytic core is not always sufficient for observing the activity of an ULP1. In addition, in some cases, the regulatory domains can influence the specificity of these enzymes.

We have observed that SUMOylation machinery is promiscuous in its choice of SUMO substrate while there is specificity among the deSUMOylating enzymes for their SUMO substrates. All SUMOs used in these studies can be used as substrates for conjugation, which supports the proposal that all SUMOs used in this study, have conserved residues utilized for their interaction with Ubc9. Thus differences in the amino acid sequence of each SUMO determines which Ulp1 will process which SUMO, yet similarity is maintained to allow for their recognition by the SUMOylation machinery. Further more, we

predict that the residues important for dictating specificity by the ULP1s, may also be important for the SUMO E3 ligases in specifying which SUMO is added to a target substrate.

### **Summary of main contributions to the field**

All of these studies involving XopD and *Arabidopsis thaliana* ULP1 family members have led to the hypothesis that ULP1 family members exhibit specificity for the SUMO substrate each will recognize and process. It is this rigid specificity of the deSUMOylating enzymes that may play a role in regulating that activity of SUMO modifications in general. These studies have shown that the SUMOylation machinery will recognize any SUMO it is challenged with, therefore the regulation of the SUMO modification may be due in part to the deSUMOylating enzyme. The literature is filled with numerous examples of proteins in mammalian cells that are modified by SUMO, ranging from transcription factors, histones, and signaling molecules. This diverse group of SUMOylated targets suggests that there must be regulation imposed on the SUMOylation system. From this body of work, I hypothesize that it the deSUMOylating enzymes, namely their specificity for SUMO proteins in combination with cellular location that imposes regulation on the SUMOylation system.

Our studies contribute to the basic understanding of the complexity of this transient and reversible system of regulation and are reminiscent of many other systems involved in reversible posttranslational modifications.

## Future Directions

Future studies to further characterize XopD would include identification of the actual SUMO modified target proteins in plant cells. Once the SUMO-modified protein targets are identified, assays using the various XopD truncations could be used to further understand the contribution that the amino-terminal regulatory domain plays in substrate recognition and processing. Solving the crystal structure of the amino-terminal regulatory domain of XopD or any ULP1 family member may provide information about the function of this region of the protein. Solving the crystal structure of a full-length ULP1 family member in complex with a SUMO-modified target proteins will provide further information as to the role of regulatory domain and would show if this domain is involved in recognizing the target protein as opposed to the SUMO moiety.

Solving the co-crystal structure of the XopD:T-SUMO complex would provide greater insight into how XopD is recognizing T-SUMO. Several attempts were made at forming large amounts of pure covalent complex between XopD and T-SUMO using sodium borohydride, which were unsuccessful. In future studies I plan to try and form a complex between XopD C470S and T-SUMO-GGSCCT to be used for crystallization trials.

To further understand how the cellular location may affect the activity of Ulp1 family members, the localization of the *Arabidopsis thaliana* ULP1 family members and SUMOs in plant cells could be determined. If SUMOylated target proteins in *At* were identified, this would provide a way to better understand the specificity of the *At*ULP1 family members.

## APPENDIX A

### TABLES OF PRIMERS

#### GST-XopD,

Construct Name	5' Primer	3'Primer
GST-XopD C470A	5'CGCGATACAGTCGGACGGTTA TTCCGCCGGCGATCATGTGCTG AC	5'GTCAGCACATGATCGCCGGCG GAATAACCGTCCGACTGTATCG CG
GST-XopD R334A	5'CCCCGAACTTCCTCCAGTGGC GGCCACTTCGTGGCTGCTGGAT GG	5'CCATCCAGCAGCCACGAAGTG GCCGCCACTGGAGGAAGTTCGG GG
GST-XopD G495A	5'GGCACCTTCGACTACGCAGGC GCAAGGGACCTGACTGATATC	5'GATATCAGTCAGGTCCCTTGC GCCTGCGTAGTCGAAGGTGCC

#### T-SUMO-His<sub>6</sub>

Construct Name	5' Primer	3'Primer
T-SUMO-His <sub>6</sub> A62R	5'CAGTGGACATGAACTCAATTA GATTCTTATTTGATGGGCGCAG GC	5'GCCTGCGCCCATCAAATAAGA ATCTAATTGAGTTCATGTCCACT G
T-SUMO-His <sub>6</sub> M90H	5'GGGTGATGAAATCGATGCAC ACCTACATCAAACCTGGAGGCG	5'CGCCTCCAGTTTGATGTAGGT GTGCATCGATTTTCATCACCC
T-SUMO-His <sub>6</sub> M90H, L91R, H92E (HRE)	5'GTGATGAAATCGATGCACACC GAGAACAACTGGAGGCGCTA CTG	5'CAGTAGCGCCTCCAGTTTGTT CTCGGTGTGCATCGATTTTCATC AC

#### Smt3-His<sub>6</sub>

Construct Name	5' Primer	3'Primer
Smt3-His <sub>6</sub> R64A	5'GGGTAAGGAAATGGACTCCTT AGCAATTCTTGTACGACGGTATT AG	5'CTAATACCGTCGTACAAGAAT GCTAAGGAGTCCATTTCCTTAC CC
Smt3-His <sub>6</sub> H92M	5'GAGGATAACGATATTATTGAG GCTATGAGAGAACAGATTGGTG GTGCTA	5'TAGCACCACCAATCTGTTCTC TCATAGCCTCAATAATATCGTT ATCCTC
Smt3-His <sub>6</sub> H92M, R93L, E94H (MLH)	5'GATATTATTGAGGCTATGTTA CATCAGATTGGTGGTGCTACGT ATC	5'GATACGTAGCACCACCAATCT GATGTAATACAGCCTCAATAAT ATC



## M-SUMO-1-HA

Construct Name	5' Primer	3'Primer
M-SUMO-1-HA Q94P	5'GTGATTGAAGTTTATCAGGAA CCAACGGGTGGTTACCCATACG AC	5'GTCGTATGGGTAAACCACCCGT TGGTTCCTGATAAACTTCAATC AC

## M-SUMO-4-HA

Construct Name	5' Primer	3'Primer
M-SUMO-4-HA P90Q	5'CAATTGATGTGTTTCAACAGC AGACGGGAGGTGTCTACCCATA CG	5'CGTATGGGTAGACACCTCCCG TCTGCTGTTGAAACACATCAAT TG

**Table 1. Table of primers used for site-directed mutagenesis of specified constructs.** In the table, the red bases represent the altered codon in each primer.

## pGEX-rTEV: GST-XopD

Construct Name	5' Restriction Site	3' Restriction Site	5' Primer	3'Primer
GST-XopD <sub>1-545</sub>	BamHI	XhoI	5'TACGGATCCGAAT ATATACCAAGATAT GAAG	5'GATCCTCGAGCTA GAACTTTTCCACC ACTT
GST-XopD <sub>285-545</sub>	EcoRI	XhoI	5'TGACGAATCCGA CCTCAACATCCCC AG	5'GATCCTCGAGCTA GAACTTTTCCACC ACTT
GST-XopD <sub>285-520</sub>	EcoRI	NotI	5'TGACGAATCCGA CCTCAACATCCCC AG	5'GATCCTCGAGCTA TGCTGGAGCTTGCT CCGC
GST-XopD <sub>305-520</sub>	EcoRI	XhoI	5'TGACGAATCCGG ATATTCGGCCATGA CTC	5'GATCCTCGAGCTA TGCTGGAGCTTGCT CCGC
GST-XopD <sub>335-520</sub>	EcoRI	XhoI	5'TGACGAATCCGC CACTTCGTGGCTGC TG	5'GATCCTCGAGCTA TGCTGGAGCTTGCT CCGC

## pT7-LO: T-SUMO constructs for complex formation

Construct Name	5' Restriction Site	3' Restriction Site	5' Primer	3'Primer
T-SUMO-GG-His <sub>6</sub>	BamHI	HindIII	5'TACGGATCCTCTG CTAGCGGCGGCAC	5'GATCAAGCTTAAT GATGATGATGATGA TGATGGTTTACAGT AGCGCCTCCAGTTT GATGTAGC
NT-SUMO-GG-His <sub>6</sub>	BamHI	HindIII	5'TACGGATCCATGG TTCATATCAATCTC AAGG	5'GATCAAGCTTAAT GATGATGATGATGA TGATGGTTTACAGT AGCGCCTCCAGTTT GATGTAGC
His <sub>6</sub> -NT-SUMO-GG-SCCT	BamHI	HindIII	5'TACGGATCCCATC ATCATCATCATCAT GGAGGAGGAATGGT TCATATCAATCTCA AGG	5'GATCAAGCTTTTA AGTGCAGCAACTGC CTCCAGTTTGATGT AGC
His <sub>6</sub> -TEV-NT-SUMO-GG	BamHI	HindIII	5'TACGGATCCCATC ATCATCATCATCAT GGCAGCGAAAACCT GTATTTTCAGGGCG TTCATATCAATCTC AAGGTTA	5'GATCAAGCTTTTA GCCTCCAGTTTGAT GTAGC

## pET15b: His6-SUMO-Gly-Gly-HA

Construct Name	5' Restriction Site	3' Restriction Site	5' Primer	3' Primer
<i>At</i> SUMO-1-HA	NdeI	XhoI	5'TACCATATGTCT GCAAACCAGGAGGA AG	5'GATCCTCGAGTTA AGCGTAGTCTGGGA CGTCGTATGGGTAAC CACCAGTCTGATGG AGC
<i>At</i> SUMO-2-HA	NdeI	XhoI	5'TACCATATGATGTC T GCTACTCCGGAAG	5'GATCCTCGAGTTA AGCGTAGTCTGGGA CGTCGTATGGGTAAC CACCAGTCTGATGA AGC
<i>At</i> SUMO-3-HA	NdeI	XhoI	5'TACCATATGATGTC T AACCCTCAAGATGA C	5'GATCCTCGAGTTA AGCGTAGTCTGGGA CGTCGTATGGGTAAC CACCCTCATCGCCC G
T-SUMO	NdeI	XhoI	5'GGAATTCCATATG TCTGCTAGCGGCGGC AC	5'GATCCTCGAGTTA AGCGTAGTCTGGGA CGTCGTATGGGTAGC CTCCAGTTTGATGTA G
Smt3-HA	NdeI	XhoI	5'GGAATTCCATATG AT GTCGGACTCAGAAG TCAA	5'GATCCTCGAGTTA AGCGTAGTCTGGGA CGTCGTATGGGTAAC CACCAATCTGTTCTC TGT
M-SUMO-1	NdeI	XhoI	5'GGAATTCCATATG ATGTCTGACCAGGA GGCAA	5'GATCCTCGAGTTA AGCGTAGTCTGGGA CGTCGTATGGGTAAC CCCCCGTTTGTCCT G
M-SUMO-2	NdeI	XhoI	5'GGAATTCCATATG ATGCCCGACGAAAA GCCC	5'GATCCTCGAGTTA AGCGTAGTCTGGGA CGTCGTATGGGTAG ACACCTCCCGTC
M-SUMO-4	NdeI	XhoI	5'GGAATTCCATATG ATGGCCAACGAAAA GCCC	5'GATCCTCGAGTTA AGCGTAGTCTGGGA CGTCGTATGGGTAG ACACCTCCCGTAGG

## pT7LOH: His6-SUMO-Gly-Gly-HA

Construct Name	5' Restriction Site	3' Restriction Site	5' Primer	3'Primer
<i>At</i> SUMO-5-HA	EcoRI	BamHI	5'TACGAATTCGCG TGAGTTCCACAGAC ACAT	5'GATCGGATCCCCTT AAGCGTAGTCTGGG ACGTCGTATGGGTA GCCACCACCAAGTTC CATGA

## pGEX-rTEV: GST-SUMO-Gly-Gly

Construct Name	5' Restriction Site	3' Restriction Site	5' Primer	3'Primer
GST- <i>At</i> SUMO-1	BamHI	XhoI	5'TGACGGATCCATG TCTGCAAACCAGGA GG	5'GATCCTCGAGTTA GCCACCAGTCTGATG GAG
GST- <i>At</i> SUMO-2	BamHI	XhoI	5'TGACGGATCCATG TCTGCTACTCCGGAA G	5'GATCCTCGAGTTA ACCACCAGTCTGATG AAGC
GST- <i>At</i> SUMO-3	EcoRI	XhoI	5'TGACGAATTCCAT G TCTAACCCTCAAGAT GAC	5'GATCCTCGAGTTA ACCACCACTCATCGC CCG
GST- <i>At</i> SUMO-5	BamHI	XhoI	5'TGACGGATCCATG GTGAGTTCCACAGA CAC	5'GATCCTCGAGTTA GCCACCACCAAGTTC CATGA
GST-T-SUMO	BamHI	XhoI	5'TACGGATCCTCTGC TAGCGGCGGCAC	5'GATCCTCGAGTTA GCCTCCAGTTTGATG TAGC
GST-Smt3	BamHI	XhoI	5'TGACGGATCCATG TCGGACTCAGAAGT CAA	5'GATCCTCGAGTTA ACCACCAATCTGTTC TCTGT
GST-M-SUMO-2	BamHI	XhoI	5'TGACGGATCCATG GCCGACGAAAAGCC C	5'GATCCTCGAGTTA ACCTCCCGTCTGCTG TTG
GST-M-SUMO-4	BamHI	XhoI	5'TGACGGATCCATG GCCAACGAAAAGCC CA	5'GATCCTCGAGTTA ACCTCCCGTAGGCTG TTG

## pGEX-rTEV: GST-ULP1

Construct Name	5' Restriction Site	3' Restriction Site	5' Primer	3' Primer
FL <i>At</i> ULP1A	EcoRI	XhoI	5'TGACGAATTCCAT GAAAAACCAATCTA GGGTTT	5'GATCCTCGAGTTA CTCGGCTTTCAGTTG CAGA
Cat. Core <i>At</i> ULP1A	EcoRI	XhoI	5'TGACGAATTCCATT GATATTACCGGGAA AATCT	5'GATCCTCGAGTTA CTCGGCTTTCAGTTG CAGA
FL <i>At</i> ULP1C	EcoRI	NotI	5'TGACGAATTCCAT GTTGAACATAGATTG GGAC	5'TTTTCCTTTTGCGG CCGCTTAATCTGTTT GGTTACCCTTGC
Cat. Core <i>At</i> ULP1C	EcoRI	XhoI	5'TGACGAATTCCGTT CAAGTATCTCTTAAA GATC	5'GATCCTCGAGTTA ATCTGTTTGGTTACC CTTGC
FL <i>At</i> ULP1D	BamHI	EcoRI	5'TGAGGATCCATGA CGAAGAGGAAGAAG GA	5'GATCGAATTTCGTT ACTCTGTCTGGTCAC TGACA
Cat. Core <i>At</i> ULP1D	BamHI	EcoRI	5'TGAGGATCCGTTT AAGTTTGTCTTAAAG ATC	5'GATCGAATTTCGTT ACTCTGTCTGGTCAC TGACA
FL <i>At</i> ESD4	BamHI	XhoI	5'TGACGGATCCATG GGTGCCGTAGCGAT C	5'GATCCTCGAGTTA ATCAGCTCGTAGCCT CAGA
Cat. Core <i>At</i> ESD4	EcoRI	XhoI	5'TGACGAATTCCATT GATATTACTGGAGA AGTTC	5'GATCCTCGAGTTA ATCAGCTCGTAGCCT CAGA

**Table 2. Table of primers used for gene amplification and subsequent cloning for the indicated constructs.**

## APPENDIX B

### XopD Crystallographic Data

#### Data collection

Crystal	Native	SeMet <sup>a</sup>	Mutant
Space group	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2	P4 <sub>1</sub> 2 <sub>1</sub> 2
<i>a</i>	91.618	91.790	91.953
<i>b</i>	91.618	91.790	91.953
<i>c</i>	44.820	45.487	45.013
Energy (eV)	12660.63	12648.92	12663.73
Resolution range (Å)	50.00-1.95 (2.0-1.95)	50.00-1.85 (1.89-1.85)	50.00-1.80 (1.84-1.80)
Unique reflections	14,282	17,009	17,346
Multiplicity	7.6 (7.5)	14.6 (10.6)	7.6 (3.1)
Data completeness (%)	98.6 (99.9)	98.9 (91.9)	94.2 (77.4)
<i>R</i> <sub>merge</sub> (%) <sup>b</sup>	6.1 (71.0)	9.0 (76.8)	7.5 (58.0)
<i>I</i> /σ( <i>I</i> )	31.2 (2.25)	36.2 (2.55)	30.6 (1.95)
Wilson B-value (Å <sup>2</sup> )	37.8	23.4	30.5

#### Phase determination

Anomalous scatterer	selenium (3 of 5 possible sites)
Figure of merit (64.6 – 1.85 Å)	0.218

#### Refinement statistics

	Native	Mutant
Resolution range (Å)	32.0-1.95 (2.00-1.95)	50.0-1.80 (1.86-1.80)
No. of reflections <i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	13,554/720 (982/49)	17,318/1,346 (514/37)
Atoms (non-H protein/solvent)	1,337/82	1396/100
<i>R</i> <sub>work</sub> (%)	23.8 (41.9)	23.6 (38.57)
<i>R</i> <sub>free</sub> (%)	28.7 (48.6)	25.8 (35.36)
R.m.s.d. bond length (Å)	0.017	0.012
R.m.s.d. bond angle (°)	1.46	1.570
Mean B-value (Å <sup>2</sup> )	36.6	38.2
Missing residues	335-338, 406-407, 432-433, 465-468, 515-520	335, 466-467, 514-520

**Table 3. Table showing data collection, structure determination, and refinement.**

Data for the outermost shell are given in parentheses.

<sup>a</sup>Bijvoet-pairs were kept separate for data processing

<sup>b</sup> $R_{\text{merge}} = 100 \sum_h \sum_i |I_{h,i} - \langle I_h \rangle| / \sum_h \sum_i I_{h,i}$ , where the outer sum (h) is over the unique reflections and the inner sum (i) is over the set of independent observations of each unique reflection.

## BIBLIOGRAPHY

1. Noel, L., Thieme, F., Nennstiel, D., and Bonas, U. (2002) *J Bacteriol* **184**(5), 1340-1348
2. Rawlings, N. D., Morton, F.R. & Barrett, A.J. (2006) *Nucleic Acids Res.* **34**, D270-D272
3. Li, S. J., and Hochstrasser, M. (1999) *Nature* **398**(6724), 246-251
4. MEROPS Protease Database, <http://merops.sanger.ac.uk/>. In.
5. Melchior, F. (2000) *Annu Rev Cell Dev Biol* **16**, 591-626
6. Jones, J. B., Stall, R. E., and Bouzar, H. (1998) *Annu Rev Phytopathol* **36**, 41-58
7. Alfano, J. R., and Collmer, A. (2004) *Annu Rev Phytopathol* **42**, 385-414
8. Ghosh, P. (2004) *Microbiol Mol Biol Rev* **68**(4), 771-795
9. Mudgett, M. B. (2004) *Annu Rev Plant Biol*
10. Hanania, U., Furman-Matarasso, N., Ron, M., and Avni, A. (1999) *Plant J* **19**(5), 533-541
11. Mossessova, E., and Lima, C. D. (2000) *Mol Cell* **5**(5), 865-876
12. Kurepa, J., Walker, J. M., Smalle, J., Gosink, M. M., Davis, S. J., Durham, T. L., Sung, D. Y., and Vierstra, R. D. (2003) *J Biol Chem* **278**(9), 6862-6872
13. Murtas, G., Reeves, P. H., Fu, Y. F., Bancroft, I., Dean, C., and Coupland, G. (2003) *Plant Cell* **15**(10), 2308-2319
14. Matunis, M. J., Coutavas, E., and Blobel, G. (1996) *J Cell Biol* **135**(6 Pt 1), 1457-1470
15. Matunis, M. J., Wu, J., and Blobel, G. (1998) *J Cell Biol* **140**(3), 499-509
16. Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997) *Cell* **88**(1), 97-107
17. Okura, T., Gong, L., Kamitani, T., Wada, T., Okura, I., Wei, C. F., Chang, H. M., and Yeh, E. T. (1996) *J Immunol* **157**(10), 4277-4281
18. Kamitani, T., Nguyen, H. P., and Yeh, E. T. (1997) *J Biol Chem* **272**(22), 14001-14004
19. Johnson, E. S., Schwienhorst, I., Dohmen, R. J., and Blobel, G. (1997) *Embo J* **16**(18), 5509-5519
20. Johnson, E. S., and Blobel, G. (1997) *J Biol Chem* **272**(43), 26799-26802
21. Desterro, J. M., Thomson, J., and Hay, R. T. (1997) *FEBS Lett* **417**(3), 297-300
22. Mahajan, R., Gerace, L., and Melchior, F. (1998) *J Cell Biol* **140**(2), 259-270
23. Bayer, P., Arndt, A., Metzger, S., Mahajan, R., Melchior, F., Jaenicke, R., and Becker, J. (1998) *J Mol Biol* **280**(2), 275-286
24. Suzuki, T., Ichiyama, A., Saitoh, H., Kawakami, T., Omata, M., Chung, C. H., Kimura, M., Shimbara, N., and Tanaka, K. (1999) *J Biol Chem* **274**(44), 31131-31134
25. Orengo, C. A., Jones, D. T., and Thornton, J. M. (1994) *Nature* **372**(6507), 631-634
26. Johnson, E. S. (2004) *Annu Rev Biochem* **73**, 355-382
27. Rodriguez, M. S., Dargemont, C., and Hay, R. T. (2001) *J Biol Chem* **276**(16), 12654-12659
28. Pickart, C. M., and Fushman, D. (2004) *Curr Opin Chem Biol* **8**(6), 610-616



29. Bylebyl, G. R., Belichenko, I., and Johnson, E. S. (2003) *J Biol Chem* **278**(45), 44113-44120
30. Melchior, F., Schergaut, M., and Pichler, A. (2003) *Trends Biochem Sci* **28**(11), 612-618
31. Saitoh, H., and Hinchey, J. (2000) *J Biol Chem* **275**(9), 6252-6258
32. Ayaydin, F., and Dasso, M. (2004) *Mol Biol Cell* **15**(12), 5208-5218
33. Bohren, K. M., Nadkarni, V., Song, J. H., Gabbay, K. H., and Owerbach, D. (2004) *J Biol Chem* **279**(26), 27233-27238
34. Gill, G. (2004) *Genes Dev* **18**(17), 2046-2059
35. Ross, S., Best, J. L., Zon, L. I., and Gill, G. (2002) *Mol Cell* **10**(4), 831-842
36. Rodriguez, M. S., Desterro, J. M., Lain, S., Midgley, C. A., Lane, D. P., and Hay, R. T. (1999) *Embo J* **18**(22), 6455-6461
37. Izumiya, Y., Ellison, T. J., Yeh, E. T., Jung, J. U., Luciw, P. A., and Kung, H. J. (2005) *J Virol* **79**(15), 9912-9925
38. David, G., Neptune, M. A., and DePinho, R. A. (2002) *J Biol Chem* **277**(26), 23658-23663
39. Zhong, S., Muller, S., Ronchetti, S., Freemont, P. S., Dejean, A., and Pandolfi, P. P. (2000) *Blood* **95**(9), 2748-2752
40. Zhong, S., Salomoni, P., and Pandolfi, P. P. (2000) *Nat Cell Biol* **2**(5), E85-90
41. Desterro, J. M., Rodriguez, M. S., and Hay, R. T. (1998) *Mol Cell* **2**(2), 233-239
42. Bossis, G., and Melchior, F. (2006) *Mol Cell* **21**(3), 349-357
43. Lois, L. M., Lima, C. D., and Chua, N. H. (2003) *Plant Cell* **15**(6), 1347-1359
44. Li, S. J., and Hochstrasser, M. (2003) *J Cell Biol* **160**(7), 1069-1081
45. Yeh, E. T., Gong, L., and Kamitani, T. (2000) *Gene* **248**(1-2), 1-14
46. Gong, L., Millas, S., Maul, G. G., and Yeh, E. T. (2000) *J Biol Chem* **275**(5), 3355-3359
47. Best, J. L., Ganiatsas, S., Agarwal, S., Changou, A., Salomoni, P., Shirihai, O., Meluh, P. B., Pandolfi, P. P., and Zon, L. I. (2002) *Mol Cell* **10**(4), 843-855
48. Zhang, H., Saitoh, H., and Matunis, M. J. (2002) *Mol Cell Biol* **22**(18), 6498-6508
49. Hang, J., and Dasso, M. (2002) *J Biol Chem* **277**(22), 19961-19966
50. Nishida, T., Kaneko, F., Kitagawa, M., and Yasuda, H. (2001) *J Biol Chem* **276**(42), 39060-39066
51. Nishida, T., Tanaka, H., and Yasuda, H. (2000) *Eur J Biochem* **267**(21), 6423-6427
52. Kim, K. I., Baek, S. H., Jeon, Y. J., Nishimori, S., Suzuki, T., Uchida, S., Shimbara, N., Saitoh, H., Tanaka, K., and Chung, C. H. (2000) *J Biol Chem* **275**(19), 14102-14106
53. Gan-Erdene, T., Nagamalleswari, K., Yin, L., Wu, K., Pan, Z. Q., and Wilkinson, K. D. (2003) *J Biol Chem* **278**(31), 28892-28900
54. Borodovsky, A., Ovaa, H., Meester, W. J., Venanzi, E. S., Bogoy, M. S., Hekking, B. G., Ploegh, H. L., Kessler, B. M., and Overkleeft, H. S. (2005) *Chembiochem* **6**(2), 287-291
55. Hemelaar, J., Borodovsky, A., Kessler, B. M., Reverter, D., Cook, J., Kolli, N., Gan-Erdene, T., Wilkinson, K. D., Gill, G., Lima, C. D., Ploegh, H. L., and Ovaa, H. (2004) *Mol Cell Biol* **24**(1), 84-95

56. Hotson, A., Chosed, R., Shu, H., Orth, K., and Mudgett, M. B. (2003) *Mol Microbiol* **50**(2), 377-389
57. Roden, J., Eardley, L., Hotson, A., Cao, Y., and Mudgett, M. B. (2004) *Mol Plant Microbe Interact* **17**(6), 633-643
58. Hay, R. T. (2005) *Mol Cell* **18**(1), 1-12
59. Guan, K. L., and Dixon, J. E. (1991) *Anal Biochem* **192**(2), 262-267
60. Chook, Y. M., and Blobel, G. (1999) *Nature* **399**(6733), 230-237
61. Guo, D., Li, M., Zhang, Y., Yang, P., Eckenrode, S., Hopkins, D., Zheng, W., Purohit, S., Podolsky, R. H., Muir, A., Wang, J., Dong, Z., Brusko, T., Atkinson, M., Pozzilli, P., Zeidler, A., Raffel, L. J., Jacob, C. O., Park, Y., Serrano-Rios, M., Larrad, M. T., Zhang, Z., Garchon, H. J., Bach, J. F., Rotter, J. I., She, J. X., and Wang, C. Y. (2004) *Nat Genet* **36**(8), 837-841
62. Zhou, G., Bao, Z. Q., and Dixon, J. E. (1995) *J Biol Chem* **270**(21), 12665-12669
63. Bailey, J. L. (1967) *Techniques in Protein Chemistry*, Elsevier, New York
64. Shu, H., Chen, S., Lyons, K., Hsueh, R., and Brekken, D. (2003) *AfCS Res Reports* **1**, 1-10
65. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54** (Pt 5), 905-921
66. Otwinowski, Z. a. M., W. (1997) *Methods Enzymol.* **276**, 307-326
67. Terwilliger, T. C., and Berendzen, J. (1999) *Acta Crystallogr D Biol Crystallogr* **55** (Pt 4), 849-861
68. Cowtan, K. D., and Main, P. (1996) *Acta Crystallogr D Biol Crystallogr* **52**(Pt 1), 43-48
69. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat Struct Biol* **6**(5), 458-463
70. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard. (1991) *Acta Crystallogr A* **47** (Pt 2), 110-119
71. DeLano, W. L. (2002) The PyMOL Molecular Graphics System. In.
72. Orth, K. (2002) *Curr Opin Microbiol* **5**(1), 38-43
73. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res* **25**(17), 3389-3402
74. Falquet, L., Pagni, M., Bucher, P., Hulo, N., Sigrist, C. J., Hofmann, K., and Bairoch, A. (2002) *Nucleic Acids Res* **30**(1), 235-238
75. Orth, K., Xu, Z., Mudgett, M. B., Bao, Z. Q., Palmer, L. E., Bliska, J. B., Mangel, W. F., Staskawicz, B., and Dixon, J. E. (2000) *Science* **290**(5496), 1594-1597
76. Staskawicz, B. J., Mudgett, M. B., Dangel, J. L., and Galan, J. E. (2001) *Science* **292**(5525), 2285-2289
77. Whalen, M. C., Stall, R.E., and Staskawicz, B.J. (1988) *Proc Natl Acad Sci U S A* **85**, 6743-6747
78. Lavie, M., Shillington, E., Eguiluz, C., Grimsley, N., and Boucher, C. (2002) *Mol Plant Microbe Interact* **15**(10), 1058-1068

79. Nagy, F., Kircher, S., and Schafer, E. (2001) *J Cell Sci* **114**(Pt 3), 475-480
80. Yamaguchi, R., Nakamura, M., Mochizuki, N., Kay, S. A., and Nagatani, A. (1999) *J Cell Biol* **145**(3), 437-445
81. Kircher, S., Gil, P., Kozma-Bognar, L., Fejes, E., Speth, V., Husselstein-Muller, T., Bauer, D., Adam, E., Schafer, E., and Nagy, F. (2002) *Plant Cell* **14**(7), 1541-1555
82. Genoud, T., Buchala, A. J., Chua, N. H., and Metraux, J. P. (2002) *Plant J* **31**(1), 87-95
83. Cornelis, G. R. (2002) *Nat Rev Mol Cell Biol* **3**(10), 742-752
84. Reverter, D., and Lima, C. D. (2004) *Structure (Camb)* **12**(8), 1519-1531
85. Pickart, C. M., and Rose, I. A. (1986) *J Biol Chem* **261**(22), 10210-10217
86. Reverter, D., Wu, K., Erdene, T. G., Pan, Z. Q., Wilkinson, K. D., and Lima, C. D. (2005) *J Mol Biol* **345**(1), 141-151
87. Shen, L. N., Liu, H., Dong, C., Xirodimas, D., Naismith, J. H., and Hay, R. T. (2005) *Embo J* **24**(7), 1341-1351
88. Hu, M., Li, P., Song, L., Jeffrey, P. D., Chenova, T. A., Wilkinson, K. D., Cohen, R. E., and Shi, Y. (2005) *Embo J* **24**(21), 3747-3756
89. Dunten, R. L., and Cohen, R. E. (1989) *J Biol Chem* **264**(28), 16739-16747
90. Reverter, D., and Lima, C. D. (2005) *Nature* **435**(7042), 687-692
91. Farooq, A., and Zhou, M. M. (2004) *Cell Signal* **16**(7), 769-779
92. Boggon, T. J., and Eck, M. J. (2004) *Oncogene* **23**(48), 7918-7927
93. Kattenhorn, L. M., Korbel, G. A., Kessler, B. M., Spooner, E., and Ploegh, H. L. (2005) *Mol Cell* **19**(4), 547-557

## VITAE

Renee Joanna Chosed has been pursuing her Ph.D. in biological chemistry in the Division of Basic Sciences at the University of Texas Southwestern Medical Center in Dallas since August of 2001. Prior to entering UT-Southwestern, Renee attended Trinity University in San Antonio, Texas where she was awarded the degree of Bachelor of Science in May, 2001. During her four years at Trinity, Renee majored in biochemistry and molecular biology with a minor in religion. Renee participated in summer undergraduate research at Trinity University as well as in the SMART Program at Baylor College of Medicine in Houston, Texas. Renee attended Trinity after being named valedictorian of her graduating class from Westbury High School in Houston, Texas in 1997. Renee, who was born in Houston, Texas on February 25, 1979, is the daughter of Marilyn Deutser Chosed and Louis Edward Chosed. She has been married to Geoffrey Damien Pierce since May of 2004. Renee will begin a post-doctoral fellowship at M.D. Anderson Cancer Center in Houston, Texas in the fall of 2006.

Permanent Address: 5610 Briarbend  
Houston, TX 77096