Spring 2015

Structural and biophysical analysis of the proteasomal deubiquitinase, UCH37

Marie Elizabeth Morrow
Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Part of the Biochemistry Commons, and the Biophysics Commons

Recommended Citation
https://docs.lib.purdue.edu/open_access_dissertations/522

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
This is to certify that the thesis/dissertation prepared

By Marie Elizabeth Morrow

Entitled STRUCTURAL AND BIOPHYSICAL ANALYSIS OF THE PROTEASOMAL DEUBIQUITINASE, UCH37

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Chittaranjan Das

Andrew D. Mesecar

Christine A. Hrycyna

Kavita Shah

To the best of my knowledge and as understood by the student in the Thesis/Dissertation Agreement, Publication Delay, and Certification/Disclaimer (Graduate School Form 32), this thesis/dissertation adheres to the provisions of Purdue University’s “Policy on Integrity in Research” and the use of copyrighted material.

Chittaranjan Das

Approved by Major Professor(s): ________________________________

Approved by: R. E. Wild ________________________________ 04/15/2015

Head of the Department Graduate Program  Date
ACKNOWLEDGEMENTS

To my parents, Jamie and Christine, thank you for everything and for supporting my dreams. To my siblings, Thomas and K-Kolbz, you kids are the best, I love you.

To Missy Ray (Kennedy) who made me love chemistry first. I want to be as intimidating as you were back then.

Thanks to everyone at the University of Florida (GO GATORS!), especially Carrie Haskell-Luevano who gave me the opportunity to do research in her lab, and Erica Haslach-Meckes and Anamika Singh who introduced me to research and inspired me to continue doing this, I’m so lucky to have had the both of you as mentors.

Thanks to everyone at Purdue who helped me in classes, trained me or showed me a new technique, let me borrow chemicals and instruments, or just made me laugh on bad days. First, to my advisor, Chitta Das, who convinced me to pursue protein crystallography even though I swore I never would get near it. I have been lucky to be able to learn from you and I appreciate every fight we’ve had. Seriously. Second, to my committee members who have provided me with tons of guidance and support, amazing letters of recommendation, and a few memorable parties: Andy Mesecar, Chris Hrycyna, and Kavita Shah. Thanks to
the entire Hrycyna group for all of their help over the years, especially the great Karen Olsen, Patty Wiley, and Kelsey Bohn. To Lake Paul for the biophysics help and the abuse. To my Purdue friends for all our adventures outside of lab: Sarah St. John, Chris Clark, Chris Collins, Renata Everett, Chad Keyes, Ian Klein, Vanessa Klein, Ruth Campbell, Karen Olsen and Walter Bosley, Patty Wiley, Kelsey Bohn, Judy Ronau, and Mike Sheedlo. Special thanks of course to all former and current members of the Das lab: David Boudreaux, Joe Chaney, Chris Davies, Myung-Il Kim, Judith Ronau, Mike Sheedlo, Rashmi Shrestha, Amy Bueno, Raj Babar, and Cameron Wade. Judy, you are the craziest person I’ve ever met, but grad school has been so much better because of that. I appreciate all your help in my project and your awesome friendship. Rashmi, you don’t deserve all the teasing we do, but you’ve grown up a lot since the first time we met and I’m so glad we’ve been labmates and friends. Mike, we’ve done this whole thing together, luckily for both of us. Now we finally get to have separate projects! Thanks for all of the help, BFF.

The last acknowledgment goes to Alex Riehm: I love you!!! The past 5 years haven’t been easy, and I couldn’t have done this without your support, every day.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES ................................................................. vii</td>
</tr>
<tr>
<td>LIST OF FIGURES ................................................................. viii</td>
</tr>
<tr>
<td>ABSTRACT ............................................................................. xi</td>
</tr>
<tr>
<td>CHAPTER 1: INTRODUCTION ..................................................... 1</td>
</tr>
<tr>
<td>1.1 Ubiquitination ..................................................................... 1</td>
</tr>
<tr>
<td>1.2 Deubiquitination ............................................................... 4</td>
</tr>
<tr>
<td>1.2.1 UCH Family ................................................................. 7</td>
</tr>
<tr>
<td>1.3 The 26S Proteasome ............................................................ 8</td>
</tr>
<tr>
<td>1.3.1 Deubiquitination at the 26S Proteasome ......................... 13</td>
</tr>
<tr>
<td>1.4 References ........................................................................ 17</td>
</tr>
<tr>
<td>CHAPTER 2: STRUCTURE OF TSUCH37\textsuperscript{cat}-UBVME ......... 26</td>
</tr>
<tr>
<td>2.1 Introduction ....................................................................... 26</td>
</tr>
<tr>
<td>2.2 Materials and Methods ...................................................... 27</td>
</tr>
<tr>
<td>2.2.1 Synthesis of Ubiquitin Vinyl Methyl Ester ....................... 27</td>
</tr>
<tr>
<td>2.2.2 Cloning, Expression, and Protein Purification of TsUCH37\textsuperscript{cat} .... 30</td>
</tr>
<tr>
<td>2.2.3 Complexation of TsUCH37\textsuperscript{cat} with UbVME ............ 31</td>
</tr>
<tr>
<td>2.2.4 Selenomethionine-labeled Protein Purification ................ 32</td>
</tr>
<tr>
<td>2.2.5 Crystallization and Structure Solution ............................ 33</td>
</tr>
<tr>
<td>2.3 Results .............................................................................. 37</td>
</tr>
<tr>
<td>2.3.1 Structure of TsUCH37\textsuperscript{cat}-UbVME .......................... 37</td>
</tr>
<tr>
<td>2.3.2 Active Site Binding ....................................................... 39</td>
</tr>
<tr>
<td>2.3.3 Distal Site Binding ......................................................... 40</td>
</tr>
<tr>
<td>2.3.4 Crossover Loop ............................................................ 41</td>
</tr>
<tr>
<td>2.4 Discussion ......................................................................... 43</td>
</tr>
<tr>
<td>2.5 References ........................................................................ 45</td>
</tr>
</tbody>
</table>
### CHAPTER 3: KINETIC AND BIOPHYSICAL CHARACTERIZATION OF TSUCH37

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Introduction</td>
<td>47</td>
</tr>
<tr>
<td>3.2 Materials and Methods</td>
<td>50</td>
</tr>
<tr>
<td>3.2.1 Cloning, Expression, and Protein Purification</td>
<td>50</td>
</tr>
<tr>
<td>3.2.2 Analytical Ultracentrifugation</td>
<td>51</td>
</tr>
<tr>
<td>3.2.3 Ubiquitin-AMC Hydrolysis</td>
<td>51</td>
</tr>
<tr>
<td>3.2.4 Isothermal Titration Calorimetry</td>
<td>52</td>
</tr>
<tr>
<td>3.3 Results</td>
<td>53</td>
</tr>
<tr>
<td>3.3.1 Analysis of Crystallographic Dimerization of TsUCH37cat-UbVME</td>
<td>53</td>
</tr>
<tr>
<td>3.3.2 Kinetic Characterization of TsUCH37cat and TsUCH37FL</td>
<td>54</td>
</tr>
<tr>
<td>3.3.3 Analysis of UCH37 Oligomeric State</td>
<td>57</td>
</tr>
<tr>
<td>3.3.4 Analysis of UCH37 Binding to Rpn13</td>
<td>58</td>
</tr>
<tr>
<td>3.4 Discussion</td>
<td>59</td>
</tr>
<tr>
<td>3.5 References</td>
<td>62</td>
</tr>
</tbody>
</table>

### CHAPTER 4: ANALYSIS OF THE TSUCH37ΔC46–UBVME STRUCTURE AND THE ROLE OF THE ULD

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Introduction</td>
<td>64</td>
</tr>
<tr>
<td>4.2 Materials and Methods</td>
<td>67</td>
</tr>
<tr>
<td>4.2.1 Molecular Dynamics Simulations</td>
<td>67</td>
</tr>
<tr>
<td>4.2.2 Site-directed Mutagenesis and Protein Purification</td>
<td>67</td>
</tr>
<tr>
<td>4.2.3 Ubiquitin-AMC Hydrolysis Assays</td>
<td>68</td>
</tr>
<tr>
<td>4.2.4 Synthesis of Asymmetric Triubiquitin Substrate and Assays</td>
<td>69</td>
</tr>
<tr>
<td>4.3 Results</td>
<td>71</td>
</tr>
<tr>
<td>4.3.1 Analysis of Molecular Dynamics Simulations</td>
<td>71</td>
</tr>
<tr>
<td>4.3.2 Ubiquitin-AMC Hydrolysis by UCH37 ULD Mutants</td>
<td>73</td>
</tr>
<tr>
<td>4.3.3 Triubiquitin Cleavage by ULD Mutants</td>
<td>75</td>
</tr>
<tr>
<td>4.4 Discussion</td>
<td>80</td>
</tr>
<tr>
<td>4.5 References</td>
<td>81</td>
</tr>
</tbody>
</table>

### CHAPTER 5: RECENT FINDINGS ON THE STRUCTURE AND ACTIVATION OF UCH37 AND OTHER DEUBIQUITINASES

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Introduction</td>
<td>82</td>
</tr>
<tr>
<td>5.2 Analysis of UCH37-Rpn13-Ub and UCH37-NFRKB-Ub structures</td>
<td>82</td>
</tr>
<tr>
<td>5.2.1 Crossover loop</td>
<td>84</td>
</tr>
<tr>
<td>5.2.2 NFRKB Mode of Inhibition</td>
<td>85</td>
</tr>
<tr>
<td>5.3 Analysis of Kinetic Findings in Vanderlinden et. al and Sahtoe et. al</td>
<td>87</td>
</tr>
<tr>
<td>5.4 Small Structures Effect Large Changes: A Review of Deubiquitinases</td>
<td>87</td>
</tr>
<tr>
<td>5.4.1 Unproductive Active Sites</td>
<td>88</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>2.1</td>
<td>Table of crystallographic statistics</td>
</tr>
<tr>
<td>3.1</td>
<td>Dissociation constants of proteasome-associated DUBs and their binding partners</td>
</tr>
<tr>
<td>4.1</td>
<td>Lys48-Glu51 distances in ubiquitin-bound PDB structures</td>
</tr>
<tr>
<td>4.2</td>
<td>ULD mutations introduced into human UCH37</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>1.1</td>
<td>Scheme of mechanisms of ubiquitination and deubiquitination</td>
</tr>
<tr>
<td>1.2</td>
<td>Scheme of the types of ubiquitination</td>
</tr>
<tr>
<td>1.3</td>
<td>Scheme of directionally specific deubiquitination</td>
</tr>
<tr>
<td>1.4</td>
<td>Domain diagrams of UCH family deubiquitinases</td>
</tr>
<tr>
<td>1.5</td>
<td>Structure of the 26S proteasome</td>
</tr>
<tr>
<td>1.6</td>
<td>Structure of the 20S and AAA ATPases</td>
</tr>
<tr>
<td>1.7</td>
<td>Scheme of deubiquitination/degradation at the 26S proteasome</td>
</tr>
<tr>
<td>2.1</td>
<td>Synthesis of UbVME</td>
</tr>
<tr>
<td>2.2</td>
<td>Generation of the TsUCH37\textsuperscript{cat}-UbVME complex</td>
</tr>
<tr>
<td>2.3</td>
<td>Crystals of the TsUCH37\textsuperscript{cat}-UbVME complex</td>
</tr>
<tr>
<td>2.4</td>
<td>Diffraction patterns</td>
</tr>
<tr>
<td>2.5</td>
<td>Active site of TsUCH37</td>
</tr>
<tr>
<td>2.6</td>
<td>Distal site of TsUCH37</td>
</tr>
<tr>
<td>2.7</td>
<td>Crossover loop of TsUCH37\textsuperscript{cat}-UbVME structure</td>
</tr>
<tr>
<td>3.1</td>
<td>Domain diagram of UCH37 and Rpn13</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>3.2 Structure of human UCH37</td>
<td>49</td>
</tr>
<tr>
<td>3.3 Analytical ultracentrifugation of the TsUCH37\textsuperscript{cat}-UbVME complex</td>
<td>53</td>
</tr>
<tr>
<td>3.4 Michaelis Menten kinetics for TsUCH37\textsuperscript{cat} UbAMC hydrolysis</td>
<td>55</td>
</tr>
<tr>
<td>3.5 Analytical ultracentrifugation of human UCH37 and Rpn13</td>
<td>56</td>
</tr>
<tr>
<td>3.6 Isothermal titration calorimetry of UCH37 and Rpn13 binding</td>
<td>58</td>
</tr>
<tr>
<td>4.1 Structure of TsUCH37\textsuperscript{ΔC46}-UbVME</td>
<td>65</td>
</tr>
<tr>
<td>4.2 Mutant triubiquitin synthesis</td>
<td>70</td>
</tr>
<tr>
<td>4.3 Molecular dynamics simulations of TsUCH37-UbVME</td>
<td>72</td>
</tr>
<tr>
<td>4.4 UbAMC hydrolysis by UCH37 ULD mutants</td>
<td>74</td>
</tr>
<tr>
<td>4.5 HPLC/MS separations of mutant ubiquitin monomers</td>
<td>76</td>
</tr>
<tr>
<td>4.6 Cleavage of mutant triubiquitin by UCH37 in the presence of Rpn13</td>
<td>78</td>
</tr>
<tr>
<td>4.7 Cleavage of varying length K48 polyubiquitin chains by UCH37</td>
<td>79</td>
</tr>
<tr>
<td>5.1 Conformational changes in the ULD domain</td>
<td>83</td>
</tr>
<tr>
<td>5.2 Activation of UCH37 by crossover loop binding to Rpn13</td>
<td>84</td>
</tr>
<tr>
<td>5.3 Mode of NFRKB inhibition of UCH37</td>
<td>86</td>
</tr>
<tr>
<td>5.4 Misaligned active sites of deubiquitinating enzymes</td>
<td>90</td>
</tr>
<tr>
<td>5.5 Insertion domains of JAMM metalloproteases</td>
<td>92</td>
</tr>
<tr>
<td>5.6 Crossover loops for ubiquitin-bound UCH DUBs</td>
<td>93</td>
</tr>
<tr>
<td>5.7 Occluding loops of USP14</td>
<td>95</td>
</tr>
<tr>
<td>Appendix Figure</td>
<td>Page</td>
</tr>
<tr>
<td>-----------------</td>
<td>------</td>
</tr>
<tr>
<td>A 6.1 Scheme of Ubl-UIM purification method of endogenous proteasomes</td>
<td>103</td>
</tr>
<tr>
<td>A 6.2 Polyubiquitination of GFP-titin-cyclin&lt;sup&gt;PY&lt;/sup&gt; substrate</td>
<td>107</td>
</tr>
<tr>
<td>A 6.3 Silver-stained SDS PAGE gel of rabbit 26S proteasome</td>
<td>109</td>
</tr>
<tr>
<td>A 6.4 Activity of UbVME-treated proteasome</td>
<td>110</td>
</tr>
<tr>
<td>A 6.5 Activity of UbVME- and MG132-treated proteasome</td>
<td>112</td>
</tr>
</tbody>
</table>
ABSTRACT

Morrow, Marie Elizabeth. Ph.D., Purdue University, May 2015. Structural and Biophysical Analysis of the Proteasomal Deubiquitinase, UCH37. Major Professor: Chittaranjan Das.

Ubiquitin carboxyl-terminal hydrolase 37, or UCH37, is a deubiquitinating enzyme associated with the 26S proteasome, the primary protein degradation machinery in eukaryotic cells. UCH37 is responsible for the disassembly of polymeric ubiquitin chains, or polyubiquitin, which have been ligated onto proteins in order to target them for degradation. The 26S utilizes two associated deubiquitinating enzymes, UCH37 and USP14, and one intrinsic, Rpn11, to remove polyubiquitin chains from substrate proteins as they are unfolded and translocated into the proteolytic core of the proteasome, where proteins are cleaved into small peptides and then released for recycling by the cell. UCH37 associates with the proteasome via binding of its C-terminal KEKE motif to the C-terminus of Rpn13, a proteasomal ubiquitin receptor which ensnares polyubiquitinated prey for degradation. UCH37 is known to be catalytically activated upon binding to Rpn13, allowing cleavage of Lys48-linked polyubiquitin chains from their distal end, an exo-specific deubiquitination. However, free UCH37 cleaves polyubiquitin poorly and is believed to be autoinhibited by its C-
terminal UCHL5-like domain, or ULD, which may also be responsible for its oligomerization in solution. This work examines the structural, biophysical, and catalytic characteristics of UCH37 in order to elucidate its mechanism of activation by Rpn13, assess its biophysical assembly with Rpn13 within the greater proteasomal context, and ascertain its mechanism of exo-specificity despite the proteasome’s processing of a variety of polyubiquitinated substrates.

To this end, a 1.7 Å resolution X-ray crystal structure was solved of the catalytic domain of a UCH37 homolog from *Trichinella spiralis* in complex with ubiquitin vinyl methyl ester (UbVME), a suicide inhibitor substrate. Our structure, in combination with another solved of a longer construct of TsUCH37 in complex with UbVME, provided structural insights into the ability of UCH37 to process polyubiquitin, namely that its C-terminal UCHL5-like domain (ULD) is responsible for its exo-specific activity due to a network of interactions with ubiquitin’s Lys48.

Through biophysical and kinetic characterization, we have affirmed the poor activity of UCH37 alone, but do not ascribe it to autoinhibition because it does not oligomerize as previously thought, rather we find that it sediments in a monomer-dimer equilibrium in analytical ultracentrifugation experiments. We have characterized its binding and activation by Rpn13, finding that UCH37 binds to Rpn13 with a 22 nM dissociation constant and that mutations to UCH37’s ULD render it unable to be activated by Rpn13. Interestingly, we have found that while Rpn13 activates UCH37 for ubiquitin-AMC cleavage, a monoubiquitin fluorogenic substrate, it appears to slow the enzyme’s processing of Lys48-linked polyubiquitin chains in our assays.
Altogether, we have confirmed that UCH37 exists primarily as a monomer which binds tightly to its proteasomal subunit, Rpn13, and can exo-specifically cleave Lys48-linked polyubiquitin chains. However, UCH37 may not be activated as was previously thought, by Rpn13 alone, and likely requires full association with the 26S proteasome.
CHAPTER 1: INTRODUCTION

1.1 Ubiquitination

Ubiquitination occurs through a coordinated enzymatic cascade ending in the attachment of ubiquitin’s C-terminal glycine (Gly76) to an acceptor lysine residue via an isopeptide bond. This is achieved through sequential ubiquitin activation (E1 enzymes), conjugation (E2 enzymes), and ligation (E3 enzymes). The E1 enzyme, of which there are only two in humans, binds both ubiquitin and ATP-Mg$^{2+}$, forms an adenylated ubiquitin intermediate, and then its catalytic cysteine attacks this adenylated ubiquitin to form a ubiquitin-charged E1, connected by a high energy thioester bond. Ubiquitin is then passed on to one of about 40 E2 enzymes by attack of their catalytic cysteine to form a charged E2. Subsequently, the charged E2 binds to one of hundreds of E3 enzymes, which then permits ubiquitin ligation onto a target protein either through direct transfer from the E2 onto the substrate, or by E2 hand-off to the E3 enzyme, which itself ligates the ubiquitin onto an acceptor lysine (Fig. 1.1). The determinant of either of these two mechanisms is inherent in the E3 enzyme; RING/U-box ligases mediate direct E2 transfer, while HECT ligases form a thioester with ubiquitin and transfer it themselves. RBR ligases (RING in-between RING) act by combining both
mechanisms, binding the E2~Ub complex in the same manner as RING ligases, but utilizing a catalytic cysteine to form a thioester and then transfer the ubiquitin themselves \(^3\). After deposition of an initiator ubiquitin, E2s and E3s work in tandem to build a polyubiquitin chain onto a substrate protein. Specific E2s are responsible for chain initiation and emphasize recognition of substrate, whereas others are chain elongators that recognize the initiator ubiquitin and facilitate chain extension using the initiator as a handle \(^2\). Chain specificity is generally set by the elongating E2.
As ubiquitin contains seven lysines itself (K6, K11, K27, K29, K33, K48, K63), its monomers can be ligated together to form polymeric chains, known as polyubiquitin (Fig 1.2). This host of acceptor lysines provides a wide variety of homotypic chain linkages, all responsible for a variety of cellular events. The most well-studied and classical polyubiquitin chain is K48-linked, the signal for protein degradation at the 26S proteasome. K11-linked chains are also primarily a protein degradation signal, but have also been found to be crucial for cell cycle progression through degradation of cell cycle regulators. K63-linked chains are attributed to lysosomal trafficking through the ESCRT pathway, NF-κB activation,
and DNA damage repair. Ubiquitin can also be linked through its start methionine to form linear ubiquitin chains, which are involved in NF-κB activation as well as cell death \(^{10}\). Additionally, monoubiquitination serves as a signal for a variety of cellular events, notably transcriptional regulation and degradation of membrane proteins \(^{11-13}\). Currently, little is known about the biological function of chains linked through K6, K27, K29, and K33 \(^{14}\). Adding further complexity to the system, ubiquitin chains can be heterotypic, either through mixed ubiquitin chain linkages that may be “branched” (mixed chain type) or “forked” (two ubiquitin chains stemming from one monomer) chains, or as mixed ubiquitin-SUMO chains, all of which are in their early stages of biological characterization \(^{9,15-17}\). The mechanisms by which E2s and E3s recognize, bind, initiate, and elongate ubiquitin chains of varying topologies is still under investigation, as well as identification of their specific substrates.

### 1.2 Deubiquitination

In opposition to ubiquitination lies deubiquitination, the hydrolysis of the isopeptide bond (or Met1-linked amide bonds in linear polyubiquitin) and subsequent release of ubiquitin from its substrate (Fig. 1.1). This is achieved by a \(~100\)-membered group of enzymes called deubiquitinases, or DUBs. They are further broken down into mechanistic families, the cysteine proteases and the metallo-proteases. Cysteine DUBs hydrolyze isopeptide bonds utilizing catalytic Cys, His, and Asp triads, as well as an oxyanion-stabilizing Gln residue. Metallo-
DUBs have a zinc ion responsible for catalysis by an active site-bound water molecule. This Zn is stabilized by coordination to two His and one Asp residue, with the catalytic water stabilized by hydrogen bonds to a nearby Glu residue. The cysteine proteases include four subfamilies: the ubiquitin C-terminal hydrolases (UCHs), ubiquitin-specific proteases (USPs), the ovarian tumor proteases (OTUs), and the Machado Josephin domain proteases (MJDs). The zinc metalloproteases have only one subfamily, the JAB1/MPN/MOV34 metalloenzymes (JAMMs). DUBs can hydrolyze both isopeptide bonds between ubiquitin and a substrate protein’s lysine residue, as well as between two Ub monomers. DUBs have additional specificities for particular polyubiquitin chain types, directional chain cleavage, and substrate preference. Some DUBs are highly specific for one chain type, for example the metallo-DUB AMSH (associated molecule of the SH3 domain of STAM) can only cleave Lys63-linked

Figure 1.3: Scheme of directionally specific deubiquitination.
polyubiquitin chains. Other DUBs have specificity for the substrate which has been ubiquitinated. DUBs that are responsible for chain cleavage have further specificity for the directionality of their cleavage activity: some remove whole chains from the site of attachment to a substrate, called en bloc cleavage; some cleave in the middle of a chain, or endo specificity; and the third group cleaves from the furthest end of the chain (distal monomer) and removes monomers sequentially, exo-specific cleavage (Fig. 1.3) \textsuperscript{18}.

In addition to their DUB domains, many deubiquitinases contain ubiquitin binding domains (UBDs) which either provide additional stabilization to ubiquitin binding or confer specificity. Typically, these domains bind monoubiquitin, sometimes polyubiquitin, with weak affinity in the high micromolar range. They are most efficient at improving ubiquitin binding when multiple UBDs are found in one DUB, or if a DUB within a larger complex binds to other proteins containing UBDs \textsuperscript{18,19}. Examples of some of the most frequently-occurring UBDs are UBAs (ubiquitin associated domains), UIMs (ubiquitin interacting motifs), and ZnFs (zinc finger ubiquitin binding domains) \textsuperscript{19}. UBDs are crucial for the activity of many deubiquitinases and are also critical regulators of ubiquitin binding across the entire proteome.
1.2.1 UCH Family

The smallest family of DUBs is the UCHs, composed of four family members in humans: UCHL1, known to be involved in Parkinson’s disease; UCHL3, of unknown function; UCH37, involved in deubiquitination at the 26S proteasome; and BAP1, which deubiquitinates histone H2A \(^{20-28}\). The first two family members, UCHL1 and UCHL3, are single-domain proteins which are only known to cleave small substrates from the C-terminus of ubiquitin, not polyubiquitin chains. This is believed to be due to the presence of a gating loop, referred to as the crossover loop, which sterically controls access to their catalytic clefts \(^{29,30}\). The other two family members, UCH37 and BAP1, contain an N-terminal UCH domain as well as extra C-terminal domains (Fig 1.4). UCH37’s

![Figure 1.4: Domain diagrams of UCH family deubiquitinases.](image)
C-terminal extension was named the ULD, or UCHL5-like domain, which is believed to autoinhibit the enzyme’s catalytic activity. At the end lies its KEKE motif, a region responsible for its binding to the 26S proteasome through the proteasomal subunit Rpn13, which has a complementary KEKE motif of its own. BAP1 has a putative ULD domain, by sequence similarity, which has yet to be characterized. BAP1 additionally has a nuclear localization signal at its far C-terminal end responsible for its cellular localization. Both UCH37 and BAP1 are known to process larger substrates than UCHL1 and UCHL3; UCH37 disassembles polyubiquitin chains at the 26S proteasome, while BAP1 deubiquitinates histone H2A as part of the Polycomb repressor DUB complex (PR-DUB). UCH37 has been found within the assembly of another macromolecular complex, the Ino80 chromatin remodeling complex, where it exists in a generally inactive form, the role of which has yet to be explored. This study focuses on the activity of UCH37, especially as it relates to its role at the 26S proteasome.

1.3 The 26S Proteasome

The 26S proteasome is a 2.5 MDa proteolytic machine responsible for degrading the majority of cellular proteins. It consists of a 20S core particle composed of proteolytic enzymes and a 19S regulatory particle responsible for capturing and feeding ubiquitinated proteins into the mouth of the 20S. The 20S is made up of 4 stacked heptameric rings of structurally similar, but not identical, subunits. The external rings contain seven α subunits while the internal rings
contain seven β subunits. The external α subunits act as the gated channel into the proteolytic chamber, where β subunits utilize their caspase-like (β1), trypsin-like (β2), and chymotrypsin-like (β5) activities to cleave proteins at a wide variety of sequence sites \(^{39}\). The three β-subunit proteases all rely on an N-terminal threonine for catalytic activity \(^{39,43}\). Passage of proteins through this chamber ensures cleavage into small peptides, provided they can enter. Entry is regulated by the α subunits; the “closed” form of the channel is only 9 Å wide, only allowing

![Figure 1.5: Structure of the 26S proteasome. The 19S regulatory particle is highlighted in blue, 20S core particle in yellow, AAA ATPases in purple, Rpn11 deubiquitinase in red, and ubiquitin receptors Rpn10 and Rpn13 in green. Adapted from PDB ID 4CR2.](image-url)
entry by smaller peptides, rather than entire globular proteins\textsuperscript{44}. When converted to the “open” form at \(~20\ \text{Å} \) wide, the core particle accommodates an unfolded polypeptide chain (Fig. 1.6)\textsuperscript{44}. This opening is facilitated by docking of the C-termini of Rpts 2, 3, and 5 of the 19S AAA ATPases into the binding pockets of the 20S \(\alpha\) subunits\textsuperscript{42,45-51}.

In order to translocate through the pore of the 20S core particle, proteins must be unfolded into a linear polypeptide chain; there is some debate as to whether unfolding and translocation are sequential, however evidence is mounting to support the theory that unfolding is coupled to translocation. This

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure16.png}
\caption{Structure of the 20S and AAA ATPases. The 20S has been crystallized in an open and closed gate (left) and the heterohexameric AAA ATPases, Rpts 1-6, are shown at right. Adapted from PDB ID 4CR2, 1G0U, and 1RYP.}
\end{figure}
event is achieved by the 19S regulatory particle’s AAA ATPase subunits, Rpts 1-6, a heterohexameric motor which utilizes ATP hydrolysis to pull polypeptide chains into the 20S (Fig. 1.6). These Rpts dock to the outer $\alpha$ rings of the 20S and serve as the base of the 19S RP. Studies of other AAA unfoldases, especially ClpXP, a bacterial unfoldase, has suggested that translocation and unfolding are simultaneously achieved through bursts of mechanical force. Both ClpXP and the $\varphi 29$ DNA packaging motor have been shown to exist 90% of the time in a dwell state, with only 10% of its time spent in a burst of activity. This has yet to be confirmed in the 26S Rpts, but cryoEM structures of the Rpts engaged and disengaged with substrate suggest this may be the case.

In addition to Rpts 1-6, the base of the 19S regulatory particle contains two scaffolding proteins, Rpn1 and Rpn2, as well as the two constitutive ubiquitin receptors, Rpn10 and Rpn13. Rpn1 and Rpn2 act to recruit associated proteins and shuttle factors to the 19S. Through interactions with Ubl (ubiquitin-like) domains, Rpn1 acts as a docking site for shuttle factors which bring polyubiquitinated proteins to the proteasome, such as Rad23B and Dsk2. Rpn1 is also responsible for recruitment of one of the proteasome’s associated deubiquitinating enzymes, USP14, discussed below. Thus far, Rpn2 is only known to anchor one of the intrinsic proteasomal ubiquitin receptors, Rpn13, to the proteasome, no other shuttle factors or associating ubiquitin receptors.

Rpn10 and Rpn13 are the intrinsic ubiquitin receptors at the proteasome, although shuttle factors and some temporarily-associating ubiquitin receptors (Rad23B, Dsk2, Dss1, Ddi1, AIRAP) also bind polyubiquitin and transport it to the
proteasome \(^{58,65,66}\). Interestingly, deletion of these receptors and shuttle factors (currently known ones) does not impair growth of yeast \(^{60,64}\). Rpn10 and Rpn13 bind tightly to the proteasome, whereas the other shuttle factors bind weakly and transiently \(^{61,63}\). It is possible that there are even more shuttle factors or receptors to be discovered that may rescue protein degradation upon deletion of this set. Rpn10, or S5a in humans, utilizes two ubiquitin interacting motifs (UIMs) to bind polyubiquitin avidly and can also recruit the shuttle factor Rad23B \(^{57,58,67-72}\). It has an additional N-terminal von Willebrand A (VWA) domain of unknown function. Rpn13 contains an N-terminal pleckstrin homology domain referred to as the pleckstrin-like receptor for ubiquitin (Pru) domain, which binds ubiquitin in a novel mode compared to other ubiquitin binding domains \(^{21-23,62-64}\). Rpn13’s C-terminal domain is responsible for binding UCH37, the second proteasome-associated deubiquitinase. Rpn10 and Rpn13 lie on the outer edge of the 19S, at opposite ends, affording polyubiquitin chains a broad surface area for binding as well as the flexibility of multiple conformations and chain branching (Fig. 1.5) \(^{50,56,57}\).

Wrapping around and above the 19S base complex lies its lid complex, one of the least understood components of the 26S. Functions have not been assigned for its 9 subunits except Rpn11, the proteasome’s constitutive deubiquitinase. Rpns 3, 5, 6, 7, 9, and 12 contain a proteasome cyclosome initiation factor (PCI) domain, but the function of these proteins is currently unknown, aside from acting as scaffolds for other components \(^{73}\). Rpn11, a JAMM metallo-DUB, requires dimerization with Rpn8, which contains an inactive MPN domain, to form its active deubiquitinating module \(^{55,74-80}\). The lid sits above
and around the opening pore of the AAA ATPases, with Rpn11 poised immediately adjacent to the access point of polyubiquitinated substrates.  

### 1.3.1 Deubiquitination at the 26S Proteasome

After polyubiquitinated proteins are brought to the 26S proteasome via shuttle factors and transient ubiquitin receptors, they bind to the proteasome’s intrinsic ubiquitin receptors, Rpn13 and Rpn10. As substrates are unfolded and translocated into the interior of the core particle, the metallo-deubiquitinase, Rpn11, cleaves off whole ubiquitin chains from the substrate protein, releasing them back into the cellular pool of ubiquitin. Rpn11 utilizes a catalytic zinc ion bound by two histidines and an aspartate to cleave polyubiquitin chains in an *en bloc* fashion, that is, the entire chain is removed from its acceptor lysine on a substrate protein. From cryo-EM structures of the 26S engaged and free of ubiquitinated substrates, it is known that Rpn11 initially exists in an occluded state that is misaligned with the central pore and ATPases, which subsequently undergoes a dramatic conformational change upon substrate binding and engagement. This conformational change aligns the active site of Rpn11 immediately above the central pore and ATPase ring opening, which then allows it to cleave entire polyubiquitin chains from an engaged substrate protein. Rpn11 is a highly promiscuous DUB capable of cleaving many different chain types and possibly having endopeptidase activity.
as well as *en bloc*; this promiscuity is necessary given the broad variety of polyubiquitinated substrates that must feed into the 26S proteasome.

![Diagram of proteasome](image)

**Figure 1.7:** Scheme of deubiquitination/degradation at the 26S proteasome. The top half of the 26S proteasome is shown, with the 19S regulatory particle highlighted in blue, 20S core particle in yellow, AAA ATPases in purple, deubiquitinases in red, and ubiquitin receptors in green. A ubiquitinated protein (grey with yellow Ub chain) is shown binding to Ub receptor Rpn10. Adapted from PDB ID 4CR2.

Either simultaneous with this activity, or prior to engagement (still currently unknown), the proteasome’s associated deubiquitinases, UCH37 and USP14, act to trim polyubiquitin chains that have bound to Rpn13 or Rpn10. Both
UCH37 and USP14 are cysteine protease DUBs which cleave polyubiquitin chains exo-specifically, that is from the furthest monomer (distal) from the substrate protein and working their way inwards. USP14 associates with the proteasome through its Ubl domain, which binds to Rpn1, a known docking point for other Ubl domain-containing proteins. UCH37, however, binds to an ubiquitin receptor, Rpn13, through matching KEKE motifs within both of their C-terminal domains. USP14 and UCH37 have poor basal levels of deubiquitinase activity alone, but become significantly activated upon recruitment to the 26S proteasome\textsuperscript{21,78,84,87,88}. They are generally thought to be present in substoichiometric amounts at the 26S, especially USP14 due to the fact that its binding partner, Rpn1, is known to bind to multiple proteins at that same site. Currently it is believed that UCH37 is specific for Lys48-linked chains and that USP14 may process other chain types, however, the variety of ubiquitinated species brought to the proteasome indicates that these DUBs are probably more promiscuous than first thought.

A few theories exist as to what role these associated DUBs play in proteasome degradation: (1) they recycle monoubiquitin, for further use by the cell \textsuperscript{89,90}; (2) they may allow dissociation of chains prior to substrate commitment for degradation, and in turn rescue a small portion of proteins slated for degradation that may be inappropriately labeled \textsuperscript{34,35}, or (3) after a polyubiquitin chain has been freed from its substrate by Rpn11, the two associated DUBs sequentially remove ubiquitin monomers until the affinity of the polyubiquitin chain for Rpn10 or Rpn13 is poor enough to dissociate from the 26S, allowing
“resetting” of the proteasome for another round of degradation \(^{39,91}\). Their inhibition has been shown to accelerate proteasomal degradation, however, further work is needed to clarify the biological role of proteasome-associated deubiquitination \(^{86,92}\).

Within this work, we present the X-ray crystal structure of a UCH37 homolog bound to ubiquitin, as well as biophysical and kinetic data, which provides a better structural understanding of the specificity and activation of this proteasome-bound DUB. Despite the broad spectrum of ubiquitinated proteasomal substrates, this DUB appears to maintain a limited specificity. We hope that these studies of UCH37 will obtain a better picture of how deubiquitinating enzymes in general balance a need for specificity in the face of a plethora of ubiquitinated proteins, as well as how these enzymes are kept in inactive/active states by cellular protein partners.
1.4 References


CHAPTER 2: STRUCTURE OF TSUCH37CAT-UBVME

2.1 Introduction

Structural approaches to studying ubiquitination/deubiquitination machinery has yielded extensive information about its detailed mechanisms, providing vital understanding of these proteins’ ability to recognize either highly specific chain types or to be grossly promiscuous for any ubiquitinated molecule available. This approach has given the field incredible insight into the biological significance of ubiquitination. The structures of many deubiquitinases have been solved alone and in complex with ubiquitin or a ubiquitin variant. Generally, DUBs bind monoubiquitin quite poorly, especially if they act as polyubiquitin chain trimmers in cells. Therefore, in order to capture a DUB-ubiquitin bound state, covalent linkage of monoubiquitin is required, to prevent dissociation during crystallography. For this end, a handful of suicide inhibitor ubiquitin variants are used in structural biology. One of these is ubiquitin vinyl methyl ester (UbVME), used in this study, which seems to have the highest reactivity with UCH family DUBs. Here, I have solved the X-ray crystal structure of a UCH37 homolog from *Trichinella spiralis* in complex with ubiquitin vinyl methyl ester. This structure highlights the similarities of UCH-family DUB binding to ubiquitin, as many contacts are conserved with UCHL1 and UCHL3. However, the active site
crossover loop, a structural feature common to UCH enzymes, is not resolved in the TsUCH37\textsuperscript{cat}-UbVME structure due to a high amount of flexibility that is not abrogated upon ubiquitin binding, an unexpected result that hints at UCH37’s mechanism of activation.

2.2 Materials and Methods

2.2.1 Synthesis of Ubiquitin Vinyl Methyl Ester

The synthesis of glycine vinyl methyl ester (GlyVME) has been previously published, but was modified in our hands (Fig 2.1)\textsuperscript{1-3}. For the Boc protection reaction, 8 grams (88 mmol) of 3-amino-1,2-propanediol was dissolved in 150 mL water, then cooled on ice in order to add 23 grams (105 mmol) of di-tert-butyl dicarbonate (Boc anhydride), after which the reaction was returned to room temperature. Then the reaction was brought to pH 10.5 by addition of sodium hydroxide and the reaction was allowed to run overnight at room temperature. The reaction was diluted with 100 mL of ethyl acetate, cooled on ice, and then brought to pH 2.5 with hydrochloric acid. The product was then extracted out with 8 x 50 mL ethyl acetate. The organic layer was washed with NaHSO\textsubscript{4} and brine, dried over sodium sulfate, and then rotovapped down and stored at -20 C. For the oxidation reaction, 7-8 g of Boc-propanediol was dissolved in 125 mL water, to which 1.4 molar equivalents of NaIO\textsubscript{4} were added. The reaction was stirred for 2-12 hours. The product was extracted out with 3 x 100 mL ethyl acetate, dried over sodium sulfate, and then rotovapped down. The aldehyde product was used
within a day and stored at -20 C. For the Horner Wadsworth Emmons reaction, 1 equivalent of sodium hydride (60% suspension in mineral oil) was added to a flame-dried round bottom and immediately suspended in 40 mL dry THF, then purged with N₂. The sodium hydride was washed 3 x 30 mL dry THF and then 1 equivalent of trimethyl phosphonoacetate was added over 1 hr on ice. Additional THF was added as needed to keep the reaction in solution. The Boc-aldehyde was dissolved in minimal THF and added to the reaction over 1 hr on ice. After addition, the reaction was allowed to warm to room temperature and run from 6-12 hrs. The reaction was quenched with 200 mL water and then THF was removed by rotovapping. The product, Boc-GlyVME, was extracted out with 3 x 50 mL chloroform and the organic layer was washed once with 50 mL of 2% hydrochloric acid and once with 50 mL saturated sodium carbonate. The organic layer was dried over sodium sulfate and then rotovapped down and stored until purification. Boc-GlyVME was purified by silica flash chromatography using a gradient of 0-20% ethyl acetate in hexanes, pooling only fractions containing the E isomer. Solvent was rotovapped off, the product was washed 2 x with DCM, and then rotovapped down again. For Boc deprotection and crystallization of the final product, 2 molar equivalents of p-toluenesulfonic acid was dissolved in 100-200 mL diethyl ether, dried over sodium sulfate, and decanted off. Boc-GlyVME was dissolved in minimal ether and added to the pTSA solution. GlyVME tosyl salt crystallized out overnight, was filtered out, and stored at -20 C until reaction with UbMESNa.
Ubiquitin\textsubscript{1-75} intein fused to a chitin-binding domain (CBD) was expressed in \textit{E. coli} Rosetta cells to an O.D. of 0.8 and cells were harvested after expression overnight at 18°C. Cells were lysed by French press and spun down at 100,000 x g for 1 hour. The supernatant was applied to a chitin resin and incubated, with rocking, at 4°C for 2-4 hours. Unbound protein was washed off with column buffer and then 2-mercaptoethane sulfonate (MESNa) was added to
the column to displace Ub1-75 from the intein group by incubating overnight at 37°C. The eluate was collected and concentrated down to 1.5 mL.

In order to generate UbVME, UbMESNa was incubated with 200 mg GlyVME and 125 mg NHS dissolved in 1 M NaHCO₃ at pH 8 overnight at room temperature. After incubation, UbVME was dialyzed into 50 mM NaOAc pH 4.5 for 4 hours, then applied to a Mono S cation exchange column for purification from UbMESNa or hydrolyzed Ub1-75. Fractions were tested for reactivity with UCHL3 and the most reactive fractions were pooled, concentrated down, and flash frozen and stored at -80°C.

2.2.2 Cloning, Expression, and Protein Purification of TsUCH37\textsuperscript{cat}

Full-length *Trichinella spiralis* (Ts) UCH37 in the pET28a vector was sent from the lab of Katerina Artavanis-Tsakonas, who had previously identified the enzyme as a UCH family member and confirmed it to be UCH37 by co-immunoprecipitations and pull-downs of proteasomal subunits \(^4\). Following standard cloning protocols, the catalytic domain of TsUCH37, residues 1-226, was subcloned into the pGEX 6P1 vector between BamHI and XhoI digestion sites. The protein was expressed in *E. coli* Rosetta DE3 cells to an O.D. of 1.0 and the cells were harvested after expression overnight at 18°C. Cells were lysed by French press and spun down at 100,000 x g for 1 hour. The supernatant was applied to glutathione sepharose beads and unbound proteins were washed off with column buffer (1 x PBS, 400 mM KCl). GST-fused TsUCH37\textsuperscript{cat} was eluted with reduced glutathione and incubated with PreScission Protease (GE
Biosciences) overnight at 4°C. TsUCH37\textsuperscript{cat} was run back over the glutathione beads to capture GST, and then the pure protein was concentrated down and run on a HiLoad Superdex 75 for further purification. Pure fractions were concentrated down, flash frozen, and stored at -80°C.

2.2.3 Complexation of TsUCH37\textsuperscript{cat} with UbVME

Test reactions to complex TsUCH37\textsuperscript{cat} with UbVME were set up in 12 uL scale to determine the ideal concentration to push complexation to completion. Three tests were done at 37°C for 3 hours at 29 mg/mL, 14.4 mg/mL, and 9.6 mg/mL TsUCH37\textsuperscript{cat} (Fig. 2.2). For the final scale up reaction, 14.4 mg/mL was chosen. The scale-up reaction was composed of 600 uL of 14.4 mg/mL TsUCH37\textsuperscript{cat}, 600 uL UbVME, and 70 uL 1M Tris pH 8.0, for a total volume of 1.9 mL (Fig. 2.2). After 3 hours at 37°C, the reaction was diluted to 4 mL and run on a Superdex 75 for further purification, but an unexpected higher molecular weight species was not purified, so all fractions from this step were pooled, buffer exchanged, and run on a MonoQ anion exchange column in 0-40% 50 mM Tris pH 7.6, 1 M NaCl, 1 mM DTT over 45 column volumes. The pure complex eluted at 17% 1 M NaCl (Fig. 2.2). Pure fractions were pooled, concentrated down to 3-5 mg/mL, flash frozen, and stored at -80°C.
Figure 2.2: Generation of the TsUCH37\textsuperscript{cat}-UbVME complex. A) SDS PAGE gel of test titration with varied concentrations of TsUCH\textsuperscript{37cat} at 37\textdegree C for 3 hrs. B) SDS PAGE gel of final reaction of TsUCH\textsuperscript{37cat} with UbVME at 14.4 mg/mL TsUCH\textsuperscript{37cat}, reacted for 3 hrs at 37\textdegree C. C) SDS PAGE gel of fractions from MonoQ anion exchange column. The chromatogram shown above is monitored at 280 nm.

2.2.4 Selenomethionine-labeled Protein Purification

In order to introduce heavy atoms into the TsUCH37\textsuperscript{cat}-UbVME complex for experimental phasing of protein crystals, a selenomethionine enriched TsUCH\textsuperscript{37cat} protein was purified and complexed with UbVME in addition to the native protein. TsUCH\textsuperscript{37cat} was expressed in M9 minimal media supplemented with amino acids (Lys, Phe, Thr, Ile, Leu) and selenomethionine after growing to an O.D. of 0.6. The protein was purified exactly as native TsUCH\textsuperscript{37cat}, except
that each buffer was supplemented with 2-5 mM DTT to keep selenomethionine in a reducing environment. Mass spectrometry of SeMet TsUCH37cat by protein MALDI confirmed that all four methionines in the protein were enriched with SeMet, an M+1 molecular weight of 26238.6 Da and M+2 of 13117.1 Da, with a calculated molecular weight of 26238 Da. SeMet TsUCH37cat was complexed with UbVME and purified by MonoQ anion exchange chromatography. Pure fractions were pooled, concentrated down, and flash frozen and stored at -80°C. Yields for the SeMet protein were reduced; therefore the SeMet complex was lower concentration than the original complex.

2.2.5 Crystallization and Structure Solution

Native TsUCH37cat-UbVME was screened at 3 mg/mL in ~700 crystallographic conditions by sitting drop vapor diffusion. A hit was identified in the Hampton Research Ammonium Sulfate grid screen, composed of 3 M ammonium sulfate, 0.1 M bicine pH 9 at room temperature after 2 days by hanging drop vapor diffusion. However, rather than single, 3D crystals, the initial hit appeared to be stacks of 2D plate crystals. In anticipation of poor data due to multi-latticed crystals, the initial hit was optimized by additive screening. Single 3D crystals appeared with the addition of 2 mM glutathione (mixture of oxidized and reduced). Crystallization attempts with the SeMet complex in the same mother liquor composition as the native hit did not yield any crystals, therefore microseeding with native crystals was done to induce SeMet complex
crystallization. Microseeds were obtained by crushing native crystals in ~40 µL mother liquor, vortexing for 5 minutes, and spinning down any large crystal fragments. A volume of 0.2 µL of microcrystals were added to a drop composed of 1 µL SeMet complex, 1 µL mother liquor, 0.4 µL glutathione additive. Crystals grew over 2 days at room temperature. Both native and SeMet complex crystals were looped and flash frozen using 2.5 M sodium malonate pH 7.0 as a cryoprotectant.

Figure 2.3: Crystals of the TsUCH37\textsuperscript{cat}-UbVME complex. A) Unoptimized crystals in 3 M ammonium sulfate, 0.1 M bicine pH 9.0. Crystals are generally layers of 2D plates. B) Optimized crystals with 2 mM glutathione additive. C) Small crystals of the SeMet TsUCH37\textsuperscript{cat}-UbVME complex. D) An optimized SeMet complex crystal looped and mounted at the beamline at APS GM/CA CAT.
Data was collected at the 23-ID-B beamline at Argonne National Laboratory on a Mar300 CCD detector (Mar USA) and processed with HKL2000. Native crystal data was collected up to 1.9 Å at 1.033 Å and 1.7 Å data was collected at the selenium peak, 0.979 Å for single-wavelength anomalous dispersion (SAD) phasing with an f' of -8.74 and an f'' of 7.12. Diffraction data was collected on some of the unoptimized crystals, lacking the glutathione additive. The diffraction pattern indicated a multi-lattice or multiple-crystal data, which was unable to be indexed (Fig. 2.4).

Figure 2.4: Diffraction patterns. From A) the unoptimized TsUCH37\textsuperscript{cat}-UbVME crystals and B) the SeMet TsUCH37\textsuperscript{cat}-UbVME crystals after optimization with additives. Collected at the ID-B beamline of GM/CA CAT at Argonne’s APS.
The initial model was obtained from the Phenix AutoSol wizard using selenium SAD phases with an input of 8 Se sites (from Matthews coefficient, determined to be a dimer in the asymmetric unit). The initial model was given a FOM (figure of merit) of 0.338, and initial $R_{\text{work}}$ of 0.3695 and $R_{\text{free}}$ of 0.3884. Its sequence was built in using the Phenix AutoBuild wizard with additional manual model building in Coot \textsuperscript{7,8}. Two copies of the complex were found in the asymmetric unit, having a space group of C2. Refinement of the structure was done in Phenix using some TLS refinement (entire asymmetric unit considered to be one TLS group) and optimized weighting for stereochemical restraints \textsuperscript{7}. Overall completeness of the data was poor, at 88.5%, but this can be credited to poor completeness in the highest resolution shells (42%), which did not prevent structure solution or refinement. The final model had an $R$ factor of 17.4% and an $R_{\text{free}}$ of 21% with <0.2% of residues in the disallowed region of the Ramachandran plot and scoring a 98% in assessment by Molprobity \textsuperscript{9}. The structure was deposited in the Protein Databank (PDB) under the entry 4I6N \textsuperscript{10}. 
2.3 Results

2.3.1 Structure of TsUCH37\textsuperscript{cat}-UbVME

TsUCH37\textsuperscript{cat}-UbVME crystallized in the C2 space group with two copies of the complex in the asymmetric unit. The final model had an $R$ factor of 17.4\% and an $R_\text{free}$ of 21\% (Table 2.1).

The first structural element to come to our attention was the presence of electron density for a disulfide bond between Cys71 of each TsUCH37\textsuperscript{cat} monomer, leading to disulfide-mediated dimerization in the asymmetric unit. Human UCH37 was previously thought to oligomerize in solution through its C-terminal domain, therefore this result was unexpected. In order to determine if this dimerization has biological relevance, the TsUCH37\textsuperscript{cat}-UbVME complex and TsUCH37\textsuperscript{cat} alone were both subjected to analytical ultracentrifugation, the results of which are discussed in Section 3.3. We concluded that this disulfide bond formation was a crystallographic artifact rather than a biologically significant event. It likely arose as a result of the introduction of glutathione as an additive and may have assisted crystal packing into a better form than the initial 2D plate crystals. The two copies of the complex have an RMSD of 0.39, indicating very few differences between them. Analysis of the complex, for the purpose of this document, will focus on Chains A and B in the PDB file rather than the copy, Chains C and D.
Table 2.1: Table of crystallographic statistics.

<table>
<thead>
<tr>
<th>Data collection</th>
<th>SeMet TsUCH37cat:UbVME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>C1 2 1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
<tr>
<td>a, b, c (Å)</td>
<td>171.2, 55.8, 73.9</td>
</tr>
<tr>
<td>α, β, γ (°)</td>
<td>90, 113.4, 90</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.979</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>50.00-1.70</td>
</tr>
<tr>
<td>R_sym or R_merge (%)</td>
<td>8.7 (50.0)</td>
</tr>
<tr>
<td>I / σI</td>
<td>15.9 (3.3)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>88.5 (42.0)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.8 (3.5)</td>
</tr>
</tbody>
</table>

| Refinement               |                        |
| Resolution (Å)           | 27.9-1.7               |
| No. unique reflections   | 62326, 3126            |
| R_work / R_free (%)      | 17.4 / 21.1            |
| No. atoms                |                        |
| Protein                  | 4650                   |
| Ligand                   | 24                     |
| Water                    | 437                    |
| Average B-factors (Å²)   |                        |
| Protein                  | 36.2                   |
| Ligand                   | 36.5                   |
| Water                    | 44.0                   |
| R.m.s deviations         |                        |
| Bond lengths (Å)         | 0.013                  |
| Bond angles (°)          | 1.48                   |
| Ramachandran plot        |                        |
| Favored (%)              | 98.1                   |
| Allowed (%)              | 1.6                    |
| Outliers (%)             | 0.4                    |

*Numbers in parentheses refer to data in the highest resolution shell.

*R_sym = Σ|Iₙ| - <Iₙ>/|Iₙ|, where Iₙ is the observed intensity and <Iₙ> is the average intensity.

*R_work = Σ||Fₜ₇| - k|Fₜ₇||/Σ|Fₜ₇|

*R_free is the same as R_obs for a selected subset (5% and 9%, respectively) of the reflections that was not included in prior refinement calculations.

*Ordered residues: Pro-3 to Gly141 and Lys153 to Asp224 in Chain C; Pro-3 to Gly141 and Gln152 to Gln225 in Chain A.
2.3.2 Active Site Binding

The structure of TsUCH37\textsuperscript{cat}-UbVME shows very similar features to the four other UCH family ubiquitin-bound structures in the PDB: Yuh1-Ub aldehyde (PDB ID 1CMX), UCHL3-UbVME (1XD3), UCHL1-UbVME (3KW5), and \textit{Plasmodium falciparum} UCHL3-UbVME (2WDT). Its active site and general catalytic cleft (Fig. 2.5) bind ubiquitin in a highly conserved manner\textsuperscript{11-13}. The C-terminal tail of ubiquitin, residues 70-75, is thoroughly stabilized by an extensive hydrogen bond network within the catalytic cleft, notably through contacts between the backbone.

Figure 2.5: Active site of TsUCH37. A) Active site architecture of TsUCH37 (teal) compared to human UCH37 (purple) upon ubiquitin binding. B) Interactions between the C-terminal tail of UbVME (orange) and TsUCH37\textsuperscript{cat} (teal), with human UCH37 residues superposed in purple. All contacts are 2.7 – 3.2 Å.
of the tail as well as Arg72 and Arg74’s side chains. The active site tetrad, composed of Cys85, Asp176, His161, and the oxyanion-stabilizing residue Gln79, is arranged in a canonical orientation for the UCH family, which is seen in papain-like cysteine proteases as well. The catalytic Cys85 of TsUCH37 has flipped about 90° compared to Cys88 of the unbound human enzyme (PDB 3IHR) upon binding to GlyVME, a mimic of the acyl-enzyme intermediate during catalysis. This phenomenon is also seen in the ubiquitin-bound and unbound structures of UCHL1, and is believed to be a conformational switch from an unproductive form of the enzyme that may exist as a protective mechanism 12. Some deubiquitinases operate within an oxidative environment, and this conformational change may protect the enzyme against cysteine oxidation 14.

### 2.3.3 Distal Site Binding

Stabilization of ubiquitin’s C-terminal tail is the primary mode of ubiquitin binding by UCH family enzymes, with the second-most important being its distal site interactions with ubiquitin’s Leu8, Thr7, and Thr9 as well as ubiquitin’s Ile44 patch. Ubiquitin-interacting residues from the Ts to human UCH37 are not highly conserved compared to its catalytic cleft residues. The distal site of TsUCH37 utilizes different hydrophobic groups than the human enzyme for ubiquitin binding, such as replacement of Ser37 with Leu36 and substitution of the large Trp36 with Val34 and Val35 (Fig 2.6). Additionally, comparing the distal pockets of the human unbound enzyme versus bound TsUCH37, there appears to be a
conformational change that encloses the distal residues tighter around ubiquitin’s L8-T9 hairpin turn.

Figure 2.6: Distal site of UCH37. A) Ubiquitin (orange) binding to the distal site of TsUCH37 (teal) compared to human UCH37 (purple) unbound. B) Interactions between ubiquitin’s Ile44 patch (orange) and TsUCH37cat (teal), with human UCH37 residues superposed in purple.

2.3.4 Crossover Loop

The crossover loop is a structural feature common to all UCH enzymes, which varies in length among the family members. It lies across the catalytic cleft of the UCH domain and binds to ubiquitin when it is bound to the enzyme, stabilizing its C-terminal tail further for catalysis. The single domain UCH family members crystallized in complex with an ubiquitin variant found in the PDB all make contacts with ubiquitin via their crossover loops, which in unbound structures of the enzymes, are not resolved due to high flexibility (Fig. 2.7). It is believed that the UCH crossover loop is locked into a specific conformation upon
ubiquitin binding, which likely contributes to the selectivity of this family of DUBs for cleavage of small moieties from the C-terminus of ubiquitin. However, it is known that the larger family members, UCH37 and BAP1, are capable of accommodating larger substrates, polyubiquitin chains in the case of UCH37 and ubiquitinated histone H2A for BAP1 \(^{15-17}\).

**Figure 2.7:** Crossover loop of TsUCH37\(^{cat}\)-UbVME structure. A) Crossover loop of TsUCH37 shown in teal compared to other UCH-Ub bound structures, UCHL1-UbVME in red (PDB ID 3KW5), UCHL3-UbVME in green (1XD3), YUH1-UbAl in purple (1CMX), PfUCHL3-UbVME in yellow (2WDT). B) Crossover loop (pink) of each monomer in the asymmetric unit of the TsUCH37\(^{cat}\)-UbVME structure.

Interestingly, the structure of TsUCH37\(^{cat}\)-UbVME has an unresolved crossover loop, despite being bound to ubiquitin. It has retained enough flexibility, despite crystal packing, that no electron density can be seen for residues 142 – 151 in either copy in the asymmetric unit (Fig. 2.7). It does not
make any contacts with ubiquitin in the situation of this complex: UCH37 – monoubiquitin. This leads one to believe that an additional protein binding event would be required to stabilize the crossover loop, that it may require a different minimal substrate (diubiquitin, triubiquitin, etc) or that the crossover loop binds to another protein regulator. We speculate that this other protein may be Rpn13, and that this binding event may be the source of activation of UCH37’s catalytic activity.

2.4 Discussion

Here we have presented the structure of TsUCH37\textsuperscript{cat} bound to ubiquitin vinyl methyl ester, which has provided some valuable insights into the mechanism of this UCH family deubiquitinase. The enzyme relies on a complex network of interactions around ubiquitin’s C-terminal tail for substrate stabilization, which is highly conserved between TsUCH37 and the other yeast and human homologs of UCH enzymes. Additionally, TsUCH37 utilizes distal site binding to recognize ubiquitin’s Ile44 patch and Leu8-Thr9 motif. However, the residues responsible for distal site binding are not as conserved as those in the catalytic cleft, compared to human UCH37. This lack of conservation may impact the affinity of ubiquitin binding, which will be explored in Part 2 through comparison of the enzyme’s catalytic activity compared to the human enzyme. This region of the enzyme may confer selectivity among UCH family enzymes, distinguishing each from one another, as their catalytic clefts are nearly identical.
The most significant structural difference between TsUCH37\textsuperscript{cat} and the other UCH family structures is that its crossover loop has not gained sufficient stabilization upon ubiquitin binding to be visualized in its X-ray crystal structure. The crossover loop is a structural element of UCH enzymes which is responsible for substrate filtering and binding, which appears to not play a role in ubiquitin binding for TsUCH37, and likely human UCH37 as well. We speculate that the crossover loop would be resolved in the structure if it was satisfying all its necessary contacts, which probably requires binding to an additional partner. We further hypothesize that this binding partner may be Rpn13, the proteasomal subunit which anchors UCH37 to the 26S proteasome. It seems therefore that the crossover loop in UCH37 may be a key element in the regulation of UCH37’s catalytic activity through protein-protein contacts. Further examination of the crossover loop in binding studies should confirm our hypothesis.
2.5 References


CHAPTER 3: KINETIC AND BIOPHYSICAL CHARACTERIZATION OF TSUCH37

3.1 Introduction

The biophysical characteristics of UCH37 keenly regulate its kinetic activity as well as biological association with its proteasomal binding partner, Rpn13. Its ULD, or UCHL5-like domain, has been shown to alter its activity and ability to bind to the 26S proteasome. Within this ULD lies the KEKE motif, a region spanning the final 20-30 amino acids of the protein, which is responsible for its binding to the proteasomal subunit Rpn13. Rpn13 harbors a complementary C-terminal KEKE motif, which binds to UCH37 (Fig 3.1). Interestingly, the ULD of UCH37 is also thought to play two additional roles within the enzyme: (1) regulation of its oligomeric state and (2) autoinhibition of the enzyme's catalysis\(^1\text{-}\text{5}\). The oligomerization of UCH37 was explained by tetramerization of the human enzyme in its X-ray crystal structure (PDB ID 3IHR) as well as in-solution higher order oligomers observed during size-exclusion chromatography (Fig 3.2)\(^5\). Autoinhibition has been seen by multiple groups in the context of purified protein, by deletion of the ULD and comparison of its activity versus that of the full-length enzyme against ubiquitin 7-amino-4-methylcoumarin, a fluorogenic monoubiquitin substrate standard in the DUB field, but limited in that it does not address the processing of polyubiquitin \(^1\text{,}2\text{,}5\text{-}9\).
UCH37 has never been shown to cleave polyubiquitin chains appreciably on its own, yet it has good activity against UbAMC. The concept of autoinhibition appears rather advanced, when the enzyme has not been definitively capable of dismantling its proteasomal substrate, polyubiquitin. To add to the confusion, when Rpn13 has been added to UCH37, it is slightly more active in polyubiquitin cleavage, which has led others to speculate that Rpn13 relieves the effects of ULD autoinhibition through KEKE motif binding. Ultimately, UCH37 is maximally activated upon binding to the entire 19S regulatory particle, where Rpn13 resides on the proteasome. Examination of the enzyme without the 19S and only in the presence of UbAMC is a limited approach, but was the system within our reach at the time of these experiments.

Figure 3.1: Domain diagram of UCH37 and Rpn13.
Our goals for this component of the project were to dissect the role of the ULD: does it mediate oligomerization in solution? If so, does this oligomerization confer its autoinhibitory effect, or is this a separate event? And how does all of this ultimately relate to binding to the proteasome, as this one region appears to be responsible for all three activities.

Figure 3.2: Structure of human UCH37 (PDB ID 3IHR) as a tetramer in the asymmetric unit. The catalytic Cys residue is shown in yellow spheres.

Herein, the kinetic and biophysical properties of TsUCH37 as well as human UCH37 are probed, in order to better understand the mechanisms of autoinhibition, activation, and proteasomal association still eluding the field. Through kinetic analysis of TsUCH37, we find that the enzyme is indeed autoinhibited in its full-length form, but despite investigations into the biophysical
characteristics of the event, we are still unclear about how the enzyme transitions between its basal and activated states.

3.2 Materials and Methods

3.2.1 Cloning, Expression, and Protein Purification

TsUCH37\textsuperscript{cat} and TsUCH37\textsuperscript{cat-UbVME} were purified as described previously in Part 2.2.2. TsUCH37\textsuperscript{FL} with an N-terminal 6xHis tag in pET28a+ was expressed in \textit{E. coli} by Dr. Myung-Il Kim as described in Morrow et. al, 2013\textsuperscript{10}. For isothermal titration calorimetry, human Rpn13 was expressed and purified by Dr. Judith Ronau from \textit{E. coli} on glutathione beads and subsequently by size exclusion chromatography. Human UCH37 proteins used for isothermal titration calorimetry were wild-type and an E284A mutant (discussed further in Section 4.3.2), both expressed from a pET28a+ plasmid in \textit{E. coli}, purified on Ni-NTA beads using a 50 – 500 mM imidazole gradient (purification buffer of 20 mM sodium phosphate pH 7.0, 300 mM NaCl, ). Due to an engineered HRV 3C protease site, Prescission Protease (GE Biosciences) was added to remove the 6xHis tag and linker, which was subsequently removed by incubation with glutathione beads. Second step purification was done on both wild-type and UCH37 E284A on a Sephadex S200 size exclusion column (GE Biosciences) and pure fractions were pooled, concentrated down, and flash frozen as aliquots.
3.2.2 Analytical Ultracentrifugation

TsUCH37\textsuperscript{cat} and TsUCH37\textsuperscript{cat}-UbVME were both dialyzed extensively against 50 mM Tris pH 7.4, 200 mM NaCl, and 1 mM DTT. Samples were run at concentrations a range of concentrations: 8, 16, and 32 µM for TsUCH37\textsuperscript{cat} and 10, 18, and 31 µM for TsUCH37\textsuperscript{cat}-UbVME to determine oligomeric states at high concentrations. Samples were run on a Beckman-Coulter XLA analytical ultracentrifuge at 50,000 rpm and monitored at 280 nm for 150 scans. Sedimentation coefficient distributions were analyzed by SEDFIT (v. 13.0b)\textsuperscript{11}.

For analysis of human UCH37 and Rpn13, proteins were both extensively dialyzed against 50 mM Tris pH 7.5, 50 mM NaCl, 1 mM TCEP. For analysis of individual oligomerization states, UCH37 was run at 8, 16, and 32 µM and Rpn13 was run at 13.5, 27, and 54 µM. Analysis of the UCH37-Rpn13 complex was run at concentrations of 4 and 8 µM, 4 and 16 µM, and 4 and 32 µM of UCH37 and Rpn13, respectively. Samples were run and analyzed by the same methods as TsUCH37\textsuperscript{cat} and TsUCH37\textsuperscript{cat}-UbVME, above.

3.2.3 Ubiquitin-AMC Hydrolysis

Cleavage of 7-amino-4-methylcoumarin from the C-terminus of monoubiquitin, or UbAMC cleavage, was monitored in a reaction buffer containing 50 mM Tris pH 7.6, 0.5 mM EDTA, 0.1% bovine serum albumin, and 5 mM DTT. TsUCH37\textsuperscript{cat} and TsUCH37\textsuperscript{FL} were diluted in reaction buffer to 7 nM final reaction concentration and preincubated at 30°C for 5 minutes prior to the reaction. Reactions were initiated by addition of UbAMC (Boston Biochem) and
were measured on a Tecan fluorescence plate reader (Männedorf, Switzerland) with 380 nm excitation wavelength and 465 nm emission wavelength at 30°C for 1 hr. Progress curves and Michaelis-Menten kinetics were plotted and fit in SigmaPlot (Systat Software).

3.2.4 Isothermal Titration Calorimetry

For isothermal titration calorimetry, wild-type UCH37, UCH37 E284A, and Rpn13 were dialyzed extensively together against 50 mM Tris pH 7.5, 50 mM NaCl, 1 mM TCEP. ITC experiments were done using a MicroCal ITC200 (GE Biosciences). For determination of the $K_d$ of UCH37 wild-type and Rpn13 binding, two experiments were averaged together: 20 μM UCH37 in the cell with 228 μM Rpn13 injected, and 10 μM UCH37 in the cell with 100 μM Rpn13 injected. For UCH37 E284A, 10 μM E284A was in the cell and 100 μM Rpn13 was injected. The data was analyzed and fit to a single binding site model in SEDPHAT $^{11}$. 
3.3 Results

3.3.1 Analysis of Crystallographic Dimerization of TsUCH37\textsuperscript{cat}-UbVME

As discussed in Section 2.3.1, the TsUCH37\textsuperscript{cat}-UbVME complex crystallized as a dimer, with dimerization mediated by a disulfide bond between Cys71 of each TsUCH37 protein. In order to determine if TsUCH37\textsuperscript{cat}-UbVME dimerization was a biologically relevant process or merely an artifact of crystallization, the oligomeric state of the complex as well as the catalytic domain by itself was determined by analytical ultracentrifugation (AUC). Taking aliquots from the same batch of the TsUCH37\textsuperscript{cat}-UbVME complex used for crystallization, AUC was run on the complex as well as the catalytic domain of TsUCH37 alone.

Figure 3.3: Analytical ultracentrifugation of the TsUCH37\textsuperscript{cat}-UbVME complex (B) compared to TsUCH37\textsuperscript{cat} alone (A). C) Table of sedimentation coefficients.
At concentrations higher than that in cells (8 – 32 µM), neither TsUCH37\textsuperscript{cat} nor the TsUCH37\textsuperscript{cat}-UbVME complex were found to exist in solution as dimers. Both are monomeric, with sedimentation coefficients ($S_{20,w}$) of 3.3 for the complex and 2.8 for the catalytic domain (Fig. 3.3). Therefore, the dimerization event observed in the crystal structure is an artifact of crystal packing, mediated by disulfide bond formation resulting from oxidative conditions prevailing in the crystallization buffer (glutathione additive).

3.3.2 Kinetic Characterization of TsUCH37\textsuperscript{cat} and TsUCH37\textsuperscript{FL}

In order to characterize the catalytic activity of TsUCH37, its activity against a standard DUB substrate, a fluorogenic monoubiquitin derivative called UbAMC, was assessed. The original goal of studying TsUCH37 previously was for drug targeting\textsuperscript{12}, therefore, it was of interest to examine its catalytic mechanism compared to that of human UCH37 and the other UCH family DUBs. Compared to the catalytic domain of human UCH37, TsUCH37\textsuperscript{cat} has about a 20-fold lower $K_M$, indicating an improvement in substrate binding, however, the $k_{cat}$ was 100-fold lower, yielding an overall 5-fold decrease in efficiency of the enzyme (Fig 3.4)\textsuperscript{9}. It would appear that TsUCH37’s catalytic domain binds substrate tighter, but that may also impair its ability to dissociate product for another round of catalysis. Not surprisingly, TsUCH37\textsuperscript{cat}’s $K_M$ is about 14-fold higher than UCHL3 and 23-fold higher than that of UCHL1, both of which bind
monoubiquitin well and are believed to only be capable of cleaving small substrates off the C-terminus of

ubiquitin. Since UCH37 is thought to cleave polyubiquitin, it is reasonable that its catalysis of a monoubiquitinated substrate is poor. The $k_{\text{cat}}$ for TsUCH37cat is 50-fold worse than that of UCHL3, one of the most efficient DUBs for UbAMC cleavage, but was 12-fold better than that of UCHL1. The $k_{\text{cat}}/K_M$ for TsUCH37

Figure 3.4: Michaelis Menten kinetics for TsUCH37cat UbAMC hydrolysis. A) TsUCH37cat cleavage of up to 12 µM UbAMC. Curve fit to Michaelis Menten parameters generated in Sigma Plot. B) Michaelis Menten parameters for TsUCH37cat compared to human UCH37cat, UCHL3, and UCHL1 from Boudreaux et. al, 2012. Curve fitting error is listed for experimental values of TsUCH37cat, generated by Sigma Plot.
and UCHL1 are only about 2-fold different, indicating that their efficiencies are similar. As nearly all residues in the UCH family active site are highly conserved, differences in $K_m$ and ubiquitin binding are generally due to differences in their distal binding site residues, which may confer some selectivity for this family. It is plausible that having a relatively lower $K_m$ compared to its single-domain cousins, UCHL1 and UCHL3, makes UCH37 better suited for regulatory control through protein-protein contacts.

Figure 3.5: Analytical ultracentrifugation of human UCH37 and Rpn13. Sedimentation of A) UCH37 alone, B) Rpn13 (ADRM1) alone, and C) the UCH37 – Rpn13 complex compared to the individual profiles. D) Sedimentation coefficients and apparent molecular weights of UCH37, Rpn13 (ADRM1), and the complex determined by AUC.
3.3.3 Analysis of UCH37 Oligomeric State

In order to probe whether the previously proposed model of UCH37 tetramerization or higher order oligomerization was possible, we examined its oligomerization by analytical ultracentrifugation (Fig 3.5)\(^5\). Additionally, the stoichiometry of the binding of UCH37 to Rpn13, its proteasomal binding partner, was determined. Analytical ultracentrifugation of UCH37 alone at 8, 16, and 32 µM yielded data indicating that the enzyme primarily exists as a monomer at lower concentrations, but is capable of a concentration-dependent rapid monomer-dimer equilibrium, which is seen most prominently in the 32 µM concentration sample. Higher order oligomers were not detected at those concentrations, which does not rule out the possibility, but indicates that at cellular concentrations, the enzyme is likely monomeric.

As for the UCH37-Rpn13 complex, first the solution state of Rpn13 was determined alone at 13.5, 27, and 54 µM. Rpn13 primarily exists as a monomer with a small population of higher order oligomers or aggregates, however this proportion is quite small. The UCH37-Rpn13 complex was run at three different concentrations of Rpn13 (8, 16, and 32 µM), but with UCH37 fixed at 4 µM. The complex exists in a 1:1 stoichiometry, which is not a surprise given that Rpn13 only has one recognition motif for UCH37 to bind. These results do not appear to support the theory that Rpn13 may relieve UCH37 of its autoinhibition through binding its ULD to change the oligomeric state of the enzyme. Both proteins are predominantly monomeric in solution and form a 1:1 complex at the proteasome.
3.3.4 Analysis of UCH37 Binding to Rpn13

Ultimately, the most important biophysical parameter missing in our knowledge of UCH37 is its binding to Rpn13. Previous studies have examined the binding affinities of the proteasome subunit Rpn1 for USP14, the other cysteine DUB that associates with the proteasome, as well as the affinity of the two ubiquitin receptors, Rpn10 and Rpn13, for ubiquitin\textsuperscript{13-15}. However, UCH37’s relationship with Rpn13 has been less explored. For this experiment, we used purified full-length human UCH37 and full-length human Rpn13; most studies have looked at the yeast homologs. For our experiments, two UCH37 proteins were studied for their binding to Rpn13: wild-type and a ULD mutant, E284A (Fig 3.6). The dissociation constant was determined by averaging two experiments for

![Figure 3.6](image-url)  

Figure 3.6: Isothermal titration calorimetry of UCH37 and Rpn13 binding. Heats of binding curves of A) wild-type UCH37 and Rpn13 and B) UCH37 E284A and Rpn13. C) Table of thermodynamic parameters.
wild-type, one with 20 µM UCH37 in the cell and 228 µM Rpn13 as the titrant, and a second with 10 µM UCH37 in the cell and 100 µM Rpn13 as the titrant. The average $K_d$ was 22 ± 6 nM. UCH37 E284A was only run as a single experiment, with 10 µM E284A in the cell and 100 µM Rpn13 injected, which yielded a $K_d$ of 18.5 ± 7 nM. Although UCH37 in cells is known to exist as a population of free enzyme, not bound to the 26S proteasome and can associate with the Ino80 chromatin remodeling complex, these dissociation constants suggest very tight binding between this DUB and its proteasomal anchor, Rpn13. This interaction is known to be abolished upon deletion of UCH37’s KEKE motif, and it is clear that even though the ULD mutation E284A impairs activation of the enzyme (Section 4.3.2), the ULD region likely does not contribute significantly to binding to Rpn13. Additionally, as the $K_d$ of UCH37-Rpn13 binding is so low, it further disproves the possibility that UCH37 oligomerizes in cells.

3.4 Discussion

Thorough characterization of UCH37’s kinetic and biophysical properties is necessary to dissect its cellular association with the 26S proteasome and potential autoinhibition. These studies have shed more light on the role of its ULD in catalysis and binding, but more work is still needed to understand its activation. We have confirmed that dimerization of the TsUCH37$^{\text{cat}}$-UbVME complex in its crystal structure is a crystallographic artifact of tight packing. Our studies of the kinetic activities of TsUCH37$^{\text{cat}}$ and TsUCH37$^{\text{FL}}$, and our analysis
of the oligomerization state of human UCH37, both support and conflict with previous literature, in that we do not see dimerization of the full-length enzyme at cellular concentrations, but we still observe the autoinhibition phenomenon when comparing activity to the catalytic domain alone \textsuperscript{5,9,16,17}. We have confirmed that UCH37 binds to Rpn13 in a 1:1 ratio, and with a 22 nM dissociation constant (Table 3.1). This data seems to indicate that UCH37, in contrast to USP14, may be a constitutive member of the 26S proteasome rather than a transiently associating subunit. Its binding is within range of the previously

Table 3.1: Dissociation constants of proteasome-associated DUBs and their binding partners. Values for * are from ref \textsuperscript{13} while those under # are from ref \textsuperscript{18}.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>$K_D$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rpn13-Ub*</td>
<td>8000 ± 500</td>
</tr>
<tr>
<td>Rpn2-Rpn13\textsuperscript{j}</td>
<td>12 ± 3</td>
</tr>
<tr>
<td>Rpn2-Rpn13-Lb*\textsuperscript{(10kRpn2)}</td>
<td>500 ± 40</td>
</tr>
<tr>
<td>Rpn13-UCH37</td>
<td>22 ± 6</td>
</tr>
<tr>
<td>Rpn1-Ubp6\textsuperscript{#}</td>
<td>62 ± 4</td>
</tr>
<tr>
<td></td>
<td>1920 ± 260</td>
</tr>
</tbody>
</table>

measured 12 nM dissociation constant for the Rpn2-Rpn13 interaction, two undoubtedly constitutive members\textsuperscript{18}. In comparison, the yeast homolog of
USP14, Ubp6, has two possible binding sites on Rpn1, one a tighter 62 nM $K_d$ site, the other much weaker at nearly 2 µM$^{13}$. Rpn1 is known to bind to shuttle factors and other Ubl domain-containing proteins, therefore, USP14 is not always bound to it. These numbers would suggest that UCH37 is more frequently found in a proteasomal context than USP14 and may play a more significant biological role.

However, these investigations still leave open the question of how UCH37 is activated at the proteasome, if it occurs merely through association with a conformationally-accessible Rpn13, or if another binding partner is required. A UCH37 mutant, E284A, which could not be activated by Rpn13 during ubiquitin-AMC hydrolysis (Section 4.3.2) bound to Rpn13 with nearly the same $K_d$ as the wild-type enzyme. This mutation isolates Rpn13’s activation of UCH37 to an event independent of simple binding. Further studies of this mutant in the presence of di- or tri-ubiquitin, as well as in the presence of Rpn2, the proteasomal subunit which binds Rpn13’s N-terminus, may provide the key to UCH37’s mode of activation.
3.5 References


CHAPTER 4: ANALYSIS OF THE TSUCH37$^{ΔC_{46}}$-UBVME STRUCTURE AND THE ROLE OF THE ULD

4.1 Introduction

Although the TsUCH37$^{cat}$-UbVME structure provided valuable insight into the mechanism of UCH37 and its ability to recognize and bind monoubiquitin, we are still lacking information about the role of the ULD in catalysis, binding, and activation. However, another group member, Dr. Myung-Ill Kim, was able to crystallize and solve the structure of a longer construct of TsUCH37 in complex with UbVME for us to glean information regarding the ULD, hereafter referred to as TsUCH37$^{ΔC_{46}}$-UbVME (Fig 4.1). Due to cleavage of the protein during purification, only a portion of the ULD was shown in the structure, but it provided important clues regarding ubiquitin recognition by the enzyme. Contacts between the catalytic domain of TsUCH37$^{ΔC_{46}}$ and ubiquitin are identical to that of TsUCH37$^{cat}$, including a lack of ordered density for the crossover loop residues. However, contacts between ubiquitin and the ULD are seen, making this an additional ubiquitin binding interface. These contacts involve hydrogen bonds and salt bridge interactions between TsUCH37’s Arg261 and Tyr262 with ubiquitin’s Gln49 and Lys48, which forces a salt bridge interaction between
Figure 4.1: Structure of TsUCH37ΔC46-UbVME with inset of Lys48 interactions with the ULD.

Lys48 and Glu51 of ubiquitin, an event heretofore unseen in any other ubiquitin-bound structures in the PDB (Table 4.1). It has been previously shown that UCH37 behaves exospecifically, that it only cleaves from the distal monomer and works its way inward to the proximal monomer of a polyubiquitin chain. We believe that the TsUCH37ΔΔC46-UbVME accounts for the structural basis of this specificity; UCH37 is only capable of cleaving the distal monomer because its Lys48 will not be engaged in an isopeptide bond and will be free for interacting with the ULD. This recognition of Lys48 would only permit exospecific cleavage. This mechanism explains the enzyme’s specificity in cleaving Lys48-linked polyubiquitin, however, it does not provide structural clues as to its ability to
cleave other chain topologies. This chapter explores the possibility that UCH37 utilizes additional interactions by the ULD to confer specificity and bind ubiquitin.

Table 4.1: Lys48-Glu51 distances in ubiquitin-bound PDB structures. Multiple entries indicate multiple copies of Ub in the crystallographic asymmetric unit.

<table>
<thead>
<tr>
<th>UCH Family</th>
<th>PDB ID</th>
<th>DU8-ubiquitin complex</th>
<th>Lys48-Glu51 distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1KOC</td>
<td>UCHL3-UbVME</td>
<td></td>
<td>9.1, 11.5</td>
</tr>
<tr>
<td>1CMX</td>
<td>YUL1-Ub</td>
<td></td>
<td>10.9</td>
</tr>
<tr>
<td>2KBDT</td>
<td>PIUHL3-UbVME</td>
<td></td>
<td>7.3, 9.6</td>
</tr>
<tr>
<td>3FWM</td>
<td>UCHL1 S18T-UbVME</td>
<td></td>
<td>8.7</td>
</tr>
<tr>
<td>3KVP</td>
<td>UKHL1 I93M-UbVME</td>
<td></td>
<td>12.3</td>
</tr>
<tr>
<td>3KW5</td>
<td>UKHL1-UbVME</td>
<td></td>
<td>13.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>USP Family</th>
<th>PDB ID</th>
<th>DU8-ubiquitin complex</th>
<th>Lys48-Glu51 distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3TIDP</td>
<td>DU84-Ub</td>
<td></td>
<td>8.3</td>
</tr>
<tr>
<td>2Y5B</td>
<td>USP12-ub</td>
<td></td>
<td>7.8, 10.6, 6.6</td>
</tr>
<tr>
<td>1NBF</td>
<td>HAUSF-Ub</td>
<td></td>
<td>10.0, 10.7</td>
</tr>
<tr>
<td>2Y5O</td>
<td>USP14-Ub</td>
<td></td>
<td>10.9</td>
</tr>
<tr>
<td>36HSH</td>
<td>SA6A complex (UBP8)-Ub</td>
<td></td>
<td>9.2</td>
</tr>
<tr>
<td>2H65</td>
<td>USP2, Ub</td>
<td></td>
<td>9.0</td>
</tr>
<tr>
<td>2G45</td>
<td>IA70, Ub</td>
<td></td>
<td>8.6, 8.9</td>
</tr>
<tr>
<td>2T7Q</td>
<td>M4:USP-UbVME</td>
<td></td>
<td>9.9, 10.8</td>
</tr>
<tr>
<td>3V5E</td>
<td>USP2, Ub variant</td>
<td></td>
<td>7.0</td>
</tr>
<tr>
<td>3V5C</td>
<td>USP2, Ub variant</td>
<td></td>
<td>7.4</td>
</tr>
<tr>
<td>3HHP</td>
<td>USP5, Ub covalent</td>
<td></td>
<td>9.6, 6.3</td>
</tr>
<tr>
<td>3MTT</td>
<td>USP21, ubiquitin-based JSP21-specific inhibitor</td>
<td></td>
<td>8.5</td>
</tr>
<tr>
<td>3T3</td>
<td>USP21, Ub covalent</td>
<td></td>
<td>7.4, 7.5, 7.4, 7.1</td>
</tr>
<tr>
<td>3N3K</td>
<td>USP8, covalent Ub-like variant</td>
<td></td>
<td>10.4</td>
</tr>
<tr>
<td>3N3E</td>
<td>USP2x, Ub</td>
<td></td>
<td>7.4</td>
</tr>
<tr>
<td>2T1I</td>
<td>USP2, Ub covalent</td>
<td></td>
<td>7.6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>OTU Family</th>
<th>PDB ID</th>
<th>DU8-ubiquitin complex</th>
<th>Lys48-Glu51 distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4OMM</td>
<td>alfavirus papain-like protease 2, Ub</td>
<td></td>
<td>7.4</td>
</tr>
<tr>
<td>324H</td>
<td>CCHF viral, ub propargyl</td>
<td></td>
<td>9.5</td>
</tr>
<tr>
<td>4P4L</td>
<td>OTU1, Ub</td>
<td></td>
<td>9.1</td>
</tr>
<tr>
<td>3PT2</td>
<td>Viral OTU, Ub</td>
<td></td>
<td>10.7</td>
</tr>
<tr>
<td>3H3Q</td>
<td>Nakovirus viral OTU, Ub</td>
<td></td>
<td>8.3, 10.1</td>
</tr>
<tr>
<td>3By4</td>
<td>OTU, Ub</td>
<td></td>
<td>6.0</td>
</tr>
<tr>
<td>3PRM</td>
<td>CCHF viral OTU, Ub</td>
<td></td>
<td>9.5, 10.0</td>
</tr>
<tr>
<td>3PRP</td>
<td>CCHF viral OTU, Ub</td>
<td></td>
<td>10.3, 10.9</td>
</tr>
<tr>
<td>3ORP</td>
<td>OTU, Ub</td>
<td></td>
<td>8.4</td>
</tr>
<tr>
<td>4DH2</td>
<td>Nococcus1,ubiquitin adddehyde-UbC133-Ub</td>
<td></td>
<td>9.2, 8.8</td>
</tr>
<tr>
<td>4DH3</td>
<td>cecUTB1 ubiquitin adddehyde UBC133-Ub complex</td>
<td></td>
<td>11.4, 6.5, 6.9, 9.8, 10.1</td>
</tr>
<tr>
<td>4DD1</td>
<td>OTU1/UbCH5b-UbUb</td>
<td></td>
<td>7.5, 12.6, 12.6, 7.5, 12.6, 7.5</td>
</tr>
<tr>
<td>4DDG</td>
<td>OTU1/UbCH5b-UbUb</td>
<td></td>
<td>11.3, 7.1, 8.0</td>
</tr>
<tr>
<td>3FH6V</td>
<td>OTU Domain of CCHF Virus, Ub</td>
<td></td>
<td>7.6, 7.2, 9.0, 5.3</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MJD Family</th>
<th>PDB ID</th>
<th>DU8-ubiquitin complex</th>
<th>Lys48-Glu51 distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3C65</td>
<td>Atrasin-3-like, Ub</td>
<td></td>
<td>9.6, 11.4, 12.8, 11.5</td>
</tr>
<tr>
<td>2J8R</td>
<td>Atrasin 3, Ub</td>
<td></td>
<td>12.3, 12.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>JAMM Family</th>
<th>PDB ID</th>
<th>DU8-ubiquitin complex</th>
<th>Lys48-Glu51 distance (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22NV</td>
<td>AM3NLP, Lys63-linked Ub ubiquitin</td>
<td></td>
<td>9.0, 8.3, 11.9, 13</td>
</tr>
</tbody>
</table>
4.2 Materials and Methods

4.2.1 Molecular Dynamics Simulations

A starting model of human UCH37-UbVME was made by modeling of full-length TsUCH37 using the human UCH37 structure (PDB ID 3IHR) as the search model in the SwissModel homology modeling server. The final 46 residues missing in the TsUCH37$^{ΔC46}$-UbVME were appended from the homology model in Coot, which then underwent one round of refinement in Phenix. Professor Markus Lill (Purdue University) then utilized this model for molecular dynamics simulations, methods described in Morrow et. al, 2013. From the 2 ns simulation, snapshots were examined for specific residues’ proximities to ubiquitin. The majority of potential interactions could be seen at the 1.3 ns snapshot.

4.2.2 Site-directed Mutagenesis and Protein Purification

Based on interactions seen in the MD simulations described above, a list of Ub-interacting ULD residues were generated from the Ts enzyme and corresponding residues in the human enzymes were mutated to Ala. Site-directed mutagenesis was performed using the AccuPower PCR PreMix (Bioneer) and mutations were confirmed by sequencing. Proteins were expressed in Rosetta2 DE3 E. coli expression cells and purified by Ni NTA beads. After cleavage of the 6xHis tag by Prescision Protease (GE Biosciences), proteins were passed over GSH beads to remove the tag and
protease. Proteins were further purified by size-exclusion chromatography on a Superdex 200 HiLoad column (GE Biosciences). Pure fractions were pooled, concentrated down, and flash frozen as aliquots. Concentrations were determined by UV/Vis. Human Rpn13\textsuperscript{FL} was provided by Dr. Judith Ronau.

4.2.3 Ubiquitin-AMC Hydrolysis Assays

Cleavage of 7-amino-4-methylcoumarin from the C-terminus of monoubiquitin, or UbAMC cleavage, was monitored in a reaction buffer containing 50 mM Tris pH 7.6, 0.5 mM EDTA, 0.1% bovine serum albumin, and 5 mM DTT. UCH37 wild-type and mutants were pre-incubated with Rpn13 on ice for 1 hr, and then diluted in reaction buffer to final reaction concentrations of 0.5 nM and 15 nM, respectively, and warmed to 30°C for 5 minutes prior to the reaction. Reactions were initiated by addition of UbAMC (Boston Biochem) and were measured on a Tecan fluorescence plate reader (Männedorf, Switzerland) with 380 nm excitation wavelength and 465 nm emission wavelength at 30°C for 1 hr. Progress curves were plotted in Kaleidagraph.
4.2.4 Synthesis of Asymmetric Triubiquitin Substrate and Assays

Site-directed mutagenesis was used to introduce a Gly76Val mutation into a Ub^{W77} construct in the pGEX-6P-1, which was subsequently confirmed by sequencing. The double mutant Ub^{G76V W77} was expressed in Rosetta2 DE3 E. coli cells and purified on glutathione beads. After treatment with Prescision Protease (GE Biosciences) to remove its N-terminal GST tag, the protein was run back over glutathione beads to remove the tag and protease. Ub^{G76V W77} was further purified by size exclusion chromatography on a Sephadex 75 HiLoad column (GE Biosciences) and pure fractions were pooled, concentrated down, and flash frozen as aliquots.

Untagged ubiquitin in the pRSET vector was expressed in Rosetta2 DE3 E. coli cells, spun down, resuspended in purification buffer A (50 mM sodium acetate pH 4.5, 2 mM DTT), lysed by French press, heated to 80°C for 10 minutes, then spun down at 30,000xg for 30 minutes. The supernatant was brought to pH 4.5 by 1N HCl and then was purified by cation exchange chromatography on SP sepharose beads (GE Biosciences) by gradient elution with purification buffer B (same as A, but with 1 M NaCl). Pure fractions were pooled and concentrated down, then further buffer exchanged and purified by size exclusion chromatography on a Sephadex 75 HiLoad column (GE Biosciences) into 50 mM Tris pH 7.6, 50 mM NaCl, 1 mM DTT. Pure fractions were pooled, concentrated down, and flash frozen as aliquots.
Figure 4.2: Mutant triubiquitin synthesis. A) Scheme of asymmetric triubiquitin showing native distal monomer and proximal Ub\textsuperscript{G76V W77}. B) Chromatogram (left) and SDS PAGE gel (right) of MonoS purification of asymmetric triUb.

In order to generate the substrate shown in Figure 4.2, first native diubiquitin was biosynthetically made using human Uba1, Cdc34, wild-type Ub, and ATP-Mg\textsuperscript{2+}. Native diubiquitin was purified by cation exchange chromatography on a MonoS column (GE Biosciences). Pure fractions were pooled and saved at -80°C for further reactions. Mutant triubiquitin was generated by incubating human Uba1, Cdc34, wild-type diubiquitin, Ub\textsuperscript{G76V W77}, and ATP-Mg\textsuperscript{2+} for 12 hrs at 37°C. Triubiquitin was purified from excess Ub\textsuperscript{G76V W77} on a MonoS column using a 200 – 600 mM NaCl salt gradient over 50 column volumes. Pure triubiquitin fractions were pooled, concentrated down, and flash frozen as aliquots.
For polyubiquitin cleavage assays, wild-type K48-linked di-, tri-, and tetra-
ubiquitin were generated biosynthetically in the same manner as mutant tri
Ub$_{G76V \ W77}$. 1.5 µM wild-type or E284A UCH37 was incubated for 1 hour on ice
with 50 µM GST-Rpn13 (for triUb$_{G76V \ W77}$ assays) or 5 µM untagged Rpn13 (for
wild-type polyubiquitin assays) in buffer containing 50 mM Tris pH 7.6, 0.5 mM
EDTA, 0.1% bovine serum albumin, and 5 mM DTT. Reactions were started with
the addition of 15 µM triUb$_{G76V \ W77}$, tetraubiquitin, wild-type triubiquitin, or
diubiquitin and time points were quenched with SDS PAGE buffer. Reactions
were run on 15% SDS PAGE gels and stained with Coomassie.

4.3 Results

4.3.1 Analysis of Molecular Dynamics Simulations

After Professor Markus Lill (Purdue University) generated a 2 ns molecular
dynamics simulation for full-length TsUCH37-UbVME, each frame was analyzed
for potential interactions between the ULD of TsUCH37, the mobile element, and
ubiquitin, which was held stationary. The majority of interactions were seen at a
1.3 ns snapshot, showing potential interactions between many ubiquitin-facing
ULD residues and ubiquitin (Fig. 4.3). Although some interactions are only within
van der Waals or salt bridge distances, some possible hydrogen bonds were also
observed. The majority of the residues at this interface are also highly conserved
from yeast up to human UCH37, therefore, the most conserved residues were
Figure 4.3: Molecular dynamics simulations of full-length TsUCH37-UbVME (in cyan, grey, and orange). TsUCH37ΔC46-UbVME is shown in olive. (i) and (ii) highlight interactions from a 1.3 ns snapshot, with highly conserved residues highlighted in red.

...deemed to be candidates for solution studies. Single point mutations to Ala were introduced in a handful of these highly-conserved residues (Table 4.2) in order to determine if these residues are responsible for UCH37’s activation by Rpn13 or, more likely, its exospecificity.
Table 4.2. ULD mutations introduced into human UCH37. Mutations listed with * were unable to be purified, due to poor solubility or instability.

<table>
<thead>
<tr>
<th>ULD mutants</th>
<th>TsUCH37</th>
<th>hUCH37</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>R261</td>
<td>R280A</td>
</tr>
<tr>
<td></td>
<td>Y262</td>
<td>Y281A</td>
</tr>
<tr>
<td></td>
<td>E265</td>
<td>E284A</td>
</tr>
<tr>
<td></td>
<td>R268</td>
<td>R287A*</td>
</tr>
<tr>
<td></td>
<td>H269</td>
<td>R288A</td>
</tr>
<tr>
<td></td>
<td>H271</td>
<td>H290A</td>
</tr>
<tr>
<td></td>
<td>N272</td>
<td>N291A*</td>
</tr>
<tr>
<td></td>
<td>Y273</td>
<td>Y292A</td>
</tr>
</tbody>
</table>

4.3.2 Ubiquitin-AMC Hydrolysis by UCH37 ULD Mutants

To assess the effects of ULD mutations on the ability of UCH37 to bind and cleave ubiquitin chains within the proteasome, we utilized a more minimal system examining the efficiency of these mutants in cleaving UbAMC in the presence and absence of Rpn13. If ULD interactions are necessary for Rpn13 binding and activation of the enzyme, presumably mutations within this region would abrogate activation. For the UbAMC assay, ULD mutations (Table 4.2) were pre-incubated with 30-fold excess Rpn13 at 4°C for 1 hour to allow association of the two proteins. Immediately prior to addition of substrate, the complexes were warmed to 30°C in reaction buffer. UbAMC hydrolysis was monitored and plotted as progress curves, shown in Figure 4.4.
Figure 4.4: UbAMC hydrolysis by UCH37 ULD mutants. Progress curves of UCH37 with and without Rpn13.

All ULD mutants assayed were activated in the presence of Rpn13 for UbAMC hydrolysis approximately 2-fold, except E284A. These results conflict with previously published activation of wild-type UCH37 in the range of a 10-fold change\textsuperscript{6,7}. The lack of activation of E284A was intriguing, though, because this mutation has somehow abolished either Rpn13’s ability to bind to UCH37 or merely catalytic activation of the enzyme. E284 is highly conserved from yeast up to human UCH37, which would indicate that it must have functional significance either in binding and/or catalysis of polyubiquitin, binding and/or activation by Rpn13, or binding and/or deactivation to partners within the Ino80 complex. To dissect its role within the context of Rpn13 activation and binding, we pursued
isothermal titration calorimetry (ITC) to determine the binding affinity of E284A with Rpn13 compared to wild-type UCH37. The results, shown in Section 3.3.4, indicate that binding to Rpn13 is not impaired by this mutation as its $K_d$ is close to that of wild-type UCH37. Therefore, this mutation may specifically inhibit the mechanism of activation of UCH37 by Rpn13, specifically.

Interestingly, ULD mutations near E284 do not impair activation within the context of the UbAMC assay, such as R280 and Y281. These two residues are not as conserved as E284; R280 is substituted with Met, Lys, or Leu, and Y281 is replaced by Trp in lower organisms. Perhaps these residues are more important for ubiquitin recognition, rather than activation.

### 4.3.3 Triubiquitin Cleavage by ULD Mutants

In order to assess directional cleavage by UCH37, an asymmetric polyubiquitin substrate was needed. To this end, Ub$^{W77}$, a construct utilized for studies of the activity of UCHL1, was given an additional mutation, Gly76Val, by site-directed mutagenesis in order to render its Trp77 non-cleavable by UCH37.

This double mutant monomer can be detected by an HPLC/MS assay due to changes in its biochemical properties: 1) increased hydrophobicity and 2) increased molecular weight. If ULD mutations abrogated exospecificity, equal proportions of Ub$^{wt}$ and Ub$^{G76VW77}$ would be cleaved from either end of the
Figure 4.5: HPLC/MS separations of ubiquitin monomers done by Zhen Wou, Wirth group, Purdue University. Panels show HPLC chromatograms (left) and MS/MS spectra (right) for Ub\textsuperscript{G76V,W77} (A) and Ub\textsuperscript{wt} (B) monomers, and C) HPLC chromatogram of a mixture of the two. D) Table of Ub monomer masses calculated and observed.

The triubiquitin chain, however, if they maintained exospecificity, initial time points would only detect the distal or middle monomers, Ub\textsuperscript{wt}. To confirm these properties, HPLC/MS experiments were done by collaborators in Mary Wirth’s lab, Purdue University, demonstrating that the mutant and wild-type monomers could be separated (Fig 4.5).

After generation of the triUb\textsuperscript{G76V,W77} substrate (Methods in Section 4.2.4), a gel-based cleavage assay was done to show activity on the substrate by wild-type UCH37 and the E284A mutant in the presence and absence of Rpn13.
Interestingly, Rpn13 appears to slow processing of triUb for both the wild-type and E284A UCH37 (Fig 4.6). In the presence of Rpn13, almost no monoubiquitin is generated by the two hour timepoint but the monoubiquitin band at the two hour timepoint for UCH37 alone is about eight times more intense. Additionally, after two hours, UCH37 + Rpn13 still has a significant amount of triubiquitin to cleave, whereas UCH37 alone has cleaved almost all of its triubiquitin down to
Figure 4.6: Cleavage of mutant triubiquitin by UCH37 in the presence of Rpn13. SDS PAGE gel of triUb\textsuperscript{G76V W77} cleavage by wild-type or E284A UCH37 in the presence and absence of GST-Rpn13.

di- or mono-Ub. The UCH37 E284A mutant appears to be mildly impaired at polyubiquitin processing compared to the wild-type enzyme because it still has triubiquitin left over after two hours. E284A is similarly inhibited by Rpn13.

In order to determine if these same kinds of rates of polyubiquitin cleavage can be seen using a variety of polyubiquitin chain types, the activity of UCH37 was assessed against tetraubiquitin in parallel with triubiquitin and diubiquitin substrates. Tetraubiquitin cleavage was also done in the presence or absence of Rpn13 to determine if inhibition persists with a longer polyubiquitin chain, one which may bind better to the ubiquitin-binding PRU domain in Rpn13’s N-
Figure 4.7: Cleavage of varying length K48 polyubiquitin chains by UCH37. SDS PAGE gel of tetra-, tri-, and di-ubiquitin cleavage by wild-type UCH37 in the presence and absence of Rpn13.

terminus. However, Rpn13 continued to inhibit polyubiquitin cleavage even with a tetraubiquitin substrate (Fig 4.7). From this experiment, it appears that UCH37 cleaves tetra-, tri-, and di-ubiquitin equally well and does not have a preference for ubiquitin chain length (Fig 4.7). From qualitative analysis of these gels, it would also appear that UCH37, while primarily an exo-specific DUB, is still capable of endo-cleavage. The rate of monoubiquitin accumulation is faster than that of diubiquitin species (especially noticeable with tetraubiquitin cleavage), however, there is still a small population of diubiquitin that is building up in initial time points. These polyubiquitin cleavage assay results are from single-trials however and results will need to be verified by repeat experiments accompanied by ImageJ quantification.
4.4 Discussion

We have analyzed the contribution of UCH37’s ULD and found it to provide 1) exo-specificity through binding to ubiquitin’s Lys48 and 2) a means of activation of the enzyme through interactions with Rpn13. After analysis of ubiquitin-AMC cleavage by ULD mutants in the presence and absence of Rpn13, we have identified Glu284 as a critical regulator of Rpn13’s activation, in that when it is mutated to Ala, activation is lost.

We have generated a novel polyubiquitin for the study of directional-specific cleavage, triUb\textsubscript{G76V W77}, which allows detection of a monoubiquitin variant by differences in molecular weight, hydrophobicity, and molar absorptivity. We have not utilized this triubiquitin for exo-specificity assays yet, but have analyzed UCH37’s ability to process tetra-, tri-, and di-ubiquitin in the presence and absence of Rpn13. It initially appears that Rpn13 slows polyubiquitin cleavage by UCH37, which has been noted by others but not fully explored\textsuperscript{9}. The investigation into the mechanism of activation of UCH37 has yet to be completely exhausted, especially as it pertains to polyubiquitin cleavage. However, our novel substrate has broader uses for other directional-specific deubiquitinating enzymes that have sufficient rates of polyubiquitin cleavage and may be a useful tool within this field.
4.5 References


5.1 Introduction

As of the writing of this document, recent work by two groups has uncovered two novel structures of UCH37: (1) bound to its activator, Rpn13, and ubiquitin and (2) bound to a fragment of NFRKB, its deactivator and a component of the Ino80 complex, both done by Vanderlinden et. al, 2015 and Sahtoe et. al, 2015\(^1,2\). These papers confirm some of the findings of this work, as well as leave some questions open still open about how UCH37 is regulated. This chapter will encompass an analysis of the new structures of UCH37 followed by a review of deubiquitinating enzymes whose specificity and activation rely on small structural elements, such as loops, in the same manner as UCH37.

5.2 Analysis of UCH37-Rpn13-Ub and UCH37-NFRKB-Ub Structures

Both structures reveal dramatic conformational changes on the part of UCH37's ULD domain (Fig 5.1) The ULD of apo UCH37 involves a helix-turn-helix (\(\alpha 9\) and \(\alpha 10\)) followed by a shorter helix, \(\alpha 11\), and ending with a short unstructured loop, as the structure lacks density for the final 18 amino acids of
Figure 5.1: Conformational changes in the ULD domain. From seven structures, as labeled.

UCH37’s C-terminus. However, this region dramatically remolds itself upon binding to both Rpn13 and NFRKB, in which helix α11 of UCH37 rotates back about 120° towards helix α10 and a new helix, α12, is resolved, which is bent approximately 90° towards the catalytic domain (Fig 5.1). In the NFRKB-bound structure without ubiquitin, even greater conformational changes are seen in the ULD, with helix α10 kinking at its center, positioning the rest of α10, α11, and α12 above the catalytic domain. These structures all show a direct role of the ULD in binding to Rpn13 and NFRKB as well as provide a structural basis for UCH37’s activation.
5.2.1 Crossover Loop

Surprisingly, our prediction that the crossover loop would be completely resolved in x-ray crystal structures upon UCH37 binding to ubiquitin and Rpn13 was proven incorrect by these recent structures. Despite the presence of activator and substrate, the crossover loop (residues 143-163) still maintains enough flexibility that it could not be entirely resolved (residues 154-159 are still unresolved)\(^1,2\). However, important interactions between it and Rpn13 are seen in these structures (Fig 5.2). Both Met148 and Phe149 make direct contact with Rpn13, a structural validation for Rpn13’s activation mechanism. Rpn13’s

![Figure 5.2: Activation of UCH37 by crossover loop binding to Rpn13. Crossover loop is in pink, with key residues highlighted. From PDB ID 4UEL.](image-url)
stabilization of an open conformation of the crossover loop likely allows improved substrate binding and catalysis, especially for the proximal ubiquitin monomer not seen in these structures. Mutations to Met148 and Phe149 render the enzyme unable to be activated by Rpn13 in UbAMC assays\textsuperscript{1,2}. It can be predicted that the rest of the crossover loop would be visualized in a UCH37-diubiquitin-Rpn13 structure, and that more of the residues in this region would contribute to substrate stabilization at the active site, especially the isopeptide bond. Alternatively, it is possible that the cross over loop maintains its dynamic character throughout catalytic steps of the enzyme, independent of substrate binding.

5.2.2 NFRKB Mode of Inhibition

The structures of UCH37 bound to the Ino80 component, NFRKB, illuminate the way in which the DUB is inhibited catalytically through both its active and distal sites. NFRKB hijacks the distal region of UCH37 that binds to the Leu8-Thr9 hairpin of ubiquitin, a key motif within ubiquitin for binding. NFRKB buries its own Phe100 and Arg101 within the distal pocket, occluding ubiquitin binding (Fig 5.3). Additionally, the large helix of NFRKB that lies across the active site face of UCH37 induces small conformational strains that lead to an unproductive orientation of the active site. The catalytic His has rotated and now lies an unproductive 6.3 Å from the catalytic cysteine. All of these small changes can be utilized for small-molecule targeting of UCH37, as they directly occupy binding sites of ubiquitin.
Figure 5.3: Mode of NFRKB inhibition of UCH37. NFRKB inhibition of UCH37 by B) active site rearrangement and C) occlusion of the UCH37 distal ubiquitin binding site. NFRKB is shown in yellow (PDB ID 4WLP), ubiquitin in orange, UCH37apo in purple (PDB ID 3IHR), UCH37 bound to NFRKB and Ub in green (PDB ID 4WLP), and UCH37 bound to Rpn13 and Ub (4WLR). Rpn13 not shown.
5.3 Analysis of Kinetic Findings in Vanderlinden et. al and Sahtoe et. al

Interestingly, these groups studied the activation of the enzyme in UbAMC hydrolysis assays using one of the ULD mutants discussed earlier, E284A in my studies, but numbered E283A in the isoform these groups used. They found that Rpn13 lowers the $K_M$ of UCH37 for ubiquitin 5-fold, but that the $K_M$ of the E283A mutant in the presence of Rpn13 is only 1.5-fold improved compared to UCH37 alone, indicating that this residue may be essential to its activation mechanism, similar to the results presented previously (Section 4.3.2)$^2$. These results validate our earlier findings, that E284 is essential to the activation mechanism of UCH37.

5.4 Small Structures Effect Large Changes: A Review of Deubiquitinases

Among the ~100 deubiquitinating enzymes in the human genome, a little less than half of these do not contain auxiliary ubiquitin binding domains or ubiquitin-like domains beyond their active sites$^3$. These deubiquitinases must rely on conformational movement and binding pockets inherent in their active sites alone, or utilize non-canonical interactions with ubiquitin to bind substrate. There have been many thorough reviews of the various ubiquitin binding and ubiquitin like domains; however, little focus has been drawn to the smaller dynamic movements that significantly contribute to deubiquitinating enzyme catalysis$^4$-$^8$. 
5.4.1 Unproductive Active Sites

One of the simplest, but integral, conformational changes within deubiquitinases (DUBs) is the reorientation of catalytic residues from an unproductive form in the apo enzyme to a productive conformation upon ubiquitin binding. This has been seen in structures of the cysteine protease DUBs and frequently involves misaligned catalytic cysteines or histidines within their papain-like active sites, less often their catalytic aspartic acid or stabilizing oxyanion glutamine residues (Fig 5.4).

Catalytic rearrangement occurs upon ubiquitin binding for the UCH family members UCHL1\(^9,10\) and UCH37\(^1,2,11-14\). In the apo UCHL1 active site, the catalytic histidine resides 8.2 Å away from the catalytic cysteine and is turned 90° from the typical orientation of a papain-like active site, an unproductive distance for catalysis\(^9\). Upon ubiquitin binding, the histidine swings 90° to lay in-plane with the catalytic Cys and has moved to a productive 4 Å distance\(^10\). In the apo active site of UCH37, the catalytic His is in a productive orientation, but its catalytic cysteine is rotated inwards, toward the His residue rather than towards the rest of the catalytic cleft where ubiquitin will bind\(^11-13\). Upon ubiquitin binding, the Cys rotates 70° to face Gly76 of ubiquitin, an appropriate conformation\(^1,2,14\).

Unproductive active sites have been found in the active sites of OTU-domain containing DUBs as well, both OTU1B, K48-specific, and OTULIN, a linear polyubiquitin-cleaving DUB. Upon binding to ubiquitin, OTUB1 has a similar conformational change to UCHL1 and UCH37; its His rotates down 90° to lock in plane with the catalytic Cys, and its Cys flips inward to face Gly76 of
ubiquitin, altogether moving the two residues closer by 3 Å into a catalytically-productive conformation\textsuperscript{15,16} (Fig 5.4). The structure of OTULIN’s active site has partial occupancy for two distinct orientations: one in which the catalytic His and Cys are in appropriate conformations, and one in which the catalytic His is flipped to occupy the space which Gly76 resides in the linear diubiquitin-bound structure\textsuperscript{17}.

HAUSP/USP7, one of the most well-characterized USP-family DUBs also contains a misaligned active site, wherein its catalytic Cys is positioned 10 Å away from the catalytic His (Fig 5.4). Upon ubiquitin binding, its catalytic Cys, His, and Asp move closer together, to a 3.7 Å distance between the Cys and His and a 2.7 Å distance between the His and Asp.

The current theory as to why these DUBs prefer a catalytically-unproductive active site orientation in absence of ubiquitin is that it may protect against oxidation\textsuperscript{18}. Some deubiquitinases have been found to be highly susceptible to oxidation, which can lead to irreversible modification (sulphinic or sulphonic acid) of their catalytic cysteines, rendering the enzyme catalytically dead. In the seminal work describing DUB oxidation, the OTU A20 was capable of an initial state of reversible oxidation, which would attain irreversibility upon continued exposure to the oxidant\textsuperscript{18}. A20 does not have a misaligned active site, however it is believed that DUBs with misoriented cysteines may induce protective interactions with nearby residues, keeping the cysteine shielded from oxidants.
Figure 5.4: Misaligned active sites of deubiquitinating enzymes. Compared to ubiquitin-bound states for A) UCHL1 (yellow) and UCH37 (blue), B) HAUSP/USP7 unbound (wheat) and bound to Ub (dark red), and C) unbound yeast OTU1 (green) and Ub-bound OTUB1 (cyan).
5.4.2 Insertions and the JAMM Domain

Regulation of JAMM domain containing deubiquitinating enzymes is mainly held by their insertion domains, numbered Ins-1 and Ins-2\textsuperscript{19}. The insertions act as substrate stabilizers and confer specificity, as seen in the structure of AMSH-LP bound to diubiquitin\textsuperscript{19} (Fig 5.5). AMSH-LP uses the sheets of Ins-1 to bind the distal ubiquitin monomer and stabilize the isopeptide bond for cleavage, and Ins-2 for additional isopeptide stabilization and binding of the proximal ubiquitin monomer. Superposition of the structures of human AMSH, another JAMM family DUB, and the AMSH-like \textit{S. pombe} orthologue Sst2 on the AMSH-LP structure shows similar modes of binding and stabilization by their highly-conserved insertion domains. These small domains provide significant contribution to catalysis; when mutated, they can render the enzyme catalytically impaired. However, mutations to Ins-2 do not contribute to substrate binding as they only affect $k_{\text{cat}}$, not $K_m$\textsuperscript{20,21}. The isopeptide contacts by the insertions also maintain specificity, in that AMSH-LP, AMSH, and Sst2 are all highly specific for K63-linked polyubiquitin chains.

In contrast, two structures were recently solved for Rpn11, the JAMM DUB resident in the 26S proteasome responsible for \textit{en bloc} cleavage of ubiquitin chains from proteasomal substrates, in which only Ins-1 was utilized for substrate recognition and catalysis\textsuperscript{22-24}. Ins-2 does not contribute to substrate catalysis; rather, it assists docking Rpn11 to the proteasomal subunit Rpn2, as predicted
Figure 5.5: Insertion domains of JAMM metalloproteases. Contributions of Ins-1 and Ins-2 to ubiquitin binding in A) the structure of AMSH-LP bound to K63-diUb and the models bound to diUb of B) AMSH, C) the AMSH-like *S. pombe* orthologue Sst2, and D) Rpn11. JAMM domains are in dark grey with insertions in pink, teal, light grey, and green, respectively.

by the recent cryoEM maps of the 26S proteasome\textsuperscript{24-26} (Fig 5.5). This non-catalytic engagement of Ins-2 provides a structural basis for Rpn11’s lack of specificity\textsuperscript{27}. It is known to cleave entire chains in an *en bloc* fashion, that is, complete removal of polyubiquitin from the lysine of attachment onto a substrate protein\textsuperscript{22,28-30}. 
5.4.3 Substrate-filtering Loops

Members of the UCH family of DUBs utilize gating loops to restrict the types of ubiquitin substrates they act upon. Structures of all the UCH family members (except BAP1) contain a flexible loop spanning their catalytic clefts, termed the crossover loop, to filter out larger substrates for cleavage\textsuperscript{31-33}. In the absence of ubiquitin, these DUBs' crossover loops are flexible and their electron density is not resolved. Upon ubiquitin binding, nearly every family member

![Figure 5.6: Crossover loops for ubiquitin-bound UCH DUBs. The UCH domain is in grey surface, ubiquitin in orange, and crossover loops highlighted for UCHL1 (red), yeast YUH1 (purple), UCHL3 (green), and UCH37 (teal).]
displays an ordered crossover loop that contributes some interactions to ubiquitin’s C-terminal tail, buried within the active site (Fig 5.6). Only UCH37 still lacks an ordered crossover loop, even in the presence of ubiquitin and its activator, the proteasomal subunit Rpn13\textsuperscript{1,2}.

Through diubiquitin cleavage assays and protein chimeras altering the length of the crossover loop, it is apparent that this loop is responsible for substrate specificity by steric filtering\textsuperscript{32,33}. Generally, the UCH family is believed to only cleave small moieties from the C-terminus of ubiquitin, not processing of polyubiquitin. The smaller family members, human UCHL1, human UCHL3, and yeast YUH1, which contain solely a UCH domain, have the shortest crossover loops and are not capable of polyubiquitin cleavage, only cleavage of ubiquitin-AMC, a fluorogenic ubiquitin substrate with only the small AMC fluorescent group attached to its C-terminus. The two larger UCH family members, UCH37 and BAP1, have C-terminal extensions beyond their catalytic domains and contain longer crossover loops than the other family members, which is believed to allow them to cleave polyubiquitin. A Drosophila homolog of BAP1, Calypso, has been shown to deubiquitinate histone H2B\textsuperscript{34} and UCH37 is proteasome-bound, and therefore must process the variety of polyubiquitinated prey captured by the 26S. The substrate-filtering theory was proven by an elegant experiment in which the crossover loop of UCH37 was expanded by a poly-glycine insertion\textsuperscript{32}. This chimeric DUB was capable of both K48- and K63-polyubiquitin chain cleavage\textsuperscript{32}. Within that same study, the crossover loop of wild-type UCHL3 was biochemically nicked, but this damage did not inhibit the DUB’s cleavage of
UbAMC\textsuperscript{32}. This would indicate that although the crossover loop does make contacts with ubiquitin's C-terminal tail, it does not significantly contribute to tail stabilization and catalysis.

5.4.4 Substrate-occluding Loops

In contrast to the insertion domains of the JAMM DUBs or the crossover loop of the UCH DUBs, some loops occlude substrate from active sites. The best

Figure 5.7: Occluding loops of USP14. USP14 substrate-occluding loops BL1 and BL2 in the apo form (pink) and bound to ubiquitin-aldehyde (blue). Zoomed in views in B) Ub-bound and C) unbound. From PDB ID 2AYN and 2AYO.
known case of this is seen in the structures of USP14 alone and bound to ubiquitin-aldehyde. In its apo form, USP14’s active site is blocked by two loops, BL1 and BL2, which occupy a portion of the space where ubiquitin’s C-terminal tail would bind in order to access its catalytic cysteine (Fig 5.7). Upon binding to ubiquitin, the entirety of these loops, as well as individual side chains, open up to allow ubiquitin binding. Oddly, many of the residues within these loops are also responsible for ubiquitin binding, such as Tyr333 and Phe331. These loops are attributed to USP14’s poor reactivity with ubiquitin probes, namely ubiquitin-vinyl sulfone, but that it becomes activated upon association with the 26S proteasome, potentially through conformational restraints of these loops into a more open form. Although many other USP family DUBs contain loops within this region, such as HAUSP/USP7, they do not sterically block those USPs active sites. It is believed that the length and conformation of these loops confer USP14 with unique reactivity compared to other USP family DUBs.

5.5 Conclusions

Through examination of the recent UCH37 activating and deactivating x-ray crystal structures as well as an assessment of other small conformational contributions to the regulation of deubiquitinating enzymes, we have a newfound appreciation for the layers of specificity and mechanisms of action of DUBs, a class of enzymes which process an incredible variety of cellular substrates. Only ~100 DUBs specifically recognize monoubiquitinated substrates, 8 homotypic
chain types, and a startling number of mixed linkage chain possibilities. Even more shocking is that only half of them require additional ubiquitin binding domains beyond their catalytic core in order to attain specificity and improve substrate binding. DUBs containing only a catalytic domain rely on small structures within themselves to restrict absolute specificity, as in the case of the JAMM DUBs AMSH and AMSH-LP, or to allow processing of a broad spectrum of substrates at a highly specific location and under certain conditions, as in the case of the proteasomal DUBs Rpn11, USP14, and UCH37.

Although we have uncovered significant information regarding the binding and catalysis of ubiquitin, as well as the binding and activation of Rpn13, to UCH37, many questions still remain regarding its ability to process polyubiquitin at the 26S proteasome. Despite our structural findings regarding its exo-specific recognition of Lys48-linked chains, it is highly implausible that UCH37 would have such limited polyubiquitin processing skills at the proteasome, with increasing reports of proteasome processing of non-canonical ubiquitin signals\textsuperscript{36,37}. Further structural and biochemical studies of UCH37 within the context of the 26S proteasome are needed to understand its role in the coupling of deubiquitination and degradation.
5.6 References


APPENDIX
Inhibition of Uch37 and Usp14 at the 26s Proteasome and Its Effects on Degradation

6.1 Introduction

Although UCH37 is capable of activation by Rpn13 alone, its activity is further enhanced within the context of the entire 26S proteasome, a mechanism of activation which remains a mystery\textsuperscript{1-3}. Due to UCH37’s poor cleavage of di- or polyubiquitin substrates alone and in the presence of Rpn13, it became clear that activity assays would not succeed without the entire 26S proteasome present as an activator. To this end, we pursued purification of endogenous 26S from rabbit tissue using an affinity-tag method developed by the Goldberg group\textsuperscript{4,5}. Using this proteasome, we hoped to study the roles of the deubiquitinases UCH37 and USP14 within overall protein degradation and whether their activities are coupled to the rate of degradation of polyubiquitinated substrates.

This purification method relies on the affinity of the Rpn1 and Rpn10 subunits of the 26S proteasome for the Ubl domain of Rad23B, a shuttle factor
capable of binding to either of those subunits\textsuperscript{6,7}. Its Ubl binds to either the Ubl-binding domain of Rpn1 or to one of the UIM domains of Rpn10, a ubiquitin receptor. For purification of the proteasome, the Ubl of Rad23B is fused to a GST-tag, allowing immobilization on glutathione beads. Proteasomes are captured by this GST-Ubl, non-specifically binding proteins are washed away, and then the proteasome is eluted by addition of a 6xHis-tagged UIM of Rpn10, which binds to the GST-Ubl and releases the proteasome. Excess His-UIM is removed by a subtraction step over Ni NTA beads. This method is considered a more gentle purification method than the traditional sucrose or glycerol gradient

Figure A 6.1: Scheme of Ubl-UIM purification method of endogenous proteasomes.
centrifugation method or the TAP- or FLAG-tagged Rpn11 method, allowing association of transient factors and ensuring the presence of the associated deubiquitinases, UCH37 and USP14\(^4,5\). This purification method is outlined in Figure 6.1.

The goal of this purification method is to investigate the contribution of UCH37 and USP14 to proteasomal degradation. There is evidence for the role of USP14 from purified *S. cerevisiae* proteasomes, however, these lack UCH37\(^8,9\). The goal of the following work is to understand how inhibition of UCH37 and USP14, achieved by incubation with the suicide inhibitor UbVME, affects substrate degradation and deubiquitination.

### 6.2 Methods

#### 6.2.1 Purification of Rabbit 26S Proteasomes

Following the method established by Besche et. al, we purified endogenous levels of 26S proteasome from rabbit tissue (Pel-freez)\(^4,5\). This affinity-tag method relied on the affinity of the proteasomal shuttle factor, Rad23B, specifically its Ubl domain, for either Rpn10’s UIM domain or Rpn1’s Ubl-binding site\(^6,7\). 2-4 grams of rabbit muscle tissue were homogenized on ice in proteasome purification buffer (25 mM HEPES pH 7.5, 50 mM NaCl, 5 mM MgCl\(_2\), 10% glycerol), PB, to which 1 mM ATP and 1 mM DTT were added, and then were spun down at 100,000xg for 1 hour. 2 mg of GST-tagged Rad23B Ubl, which had been previously purified recombinantly from *E. coli*, was immobilized
on glutathione beads and any excess washed off with PB. Rabbit lysate was rocked with immobilized GST-Ubl for 2 hours at 4°C. Beads were collected in an empty glass column and unbound proteins were allowed to flow through. 20 column volumes of PB with added ATP and DTT were used to wash the beads. 2 mg of 6xHis-tagged Rpn10 UIM was added to the beads and was rocked overnight at 4°C to induce elution of pure proteasomes. Proteasomes were then run over Ni NTA beads to re-capture the His-UIM. Pure proteasomes were concentrated down and flash frozen as aliquots.

6.2.2 20S Activity Assays

To confirm the presence of the 20S core particle and test its activity, the hydrolysis of succinate-Leu-Leu-Val-Tyr-AMC, a known 20S substrate, was measured in the presence of rabbit 26S proteasomes. Suc-LLVY-AMC was dissolved in DMSO. 5 nM 26S proteasome was diluted in AMC assay buffer containing 50 mM Tris pH 7.6, 0.5 mM EDTA, 0.1% bovine serum albumin, and 5 mM DTT and allowed to reach 30°C. 100 µM Suc-LLVY-AMC was added to 5 nM proteasome in assay buffer and AMC hydrolysis was measured on a Tecan plate reader at 380 nm excitation wavelength and 465 nm emission wavelength at 30°C for 1 hr. Progress curves were plotted in Kaleidagraph.
6.2.3 Ubiquitin-AMC Hydrolysis Assays

To test the deubiquitinating activity of endogenous UCH37 and USP14, hydrolysis of UbAMC was measured. Proteasomes were diluted to 5 nM in reaction buffer, 50 mM Tris pH 7.6, 0.5 mM EDTA, 0.1% bovine serum albumin, and 5 mM DTT. Reactions were initiated by addition of UbAMC (Boston Biochem) and were measured on a Tecan fluorescence plate reader (Männedorf, Switzerland) with 380 nm excitation wavelength and 465 nm emission wavelength at 30°C for 1 hr. Progress curves were plotted in Kaleidagraph.

6.2.4 Synthesis and Degradation of GFP-Titin-Cyclin<sup>Py</sup> Substrate

In order to measure rates of degradation by the rabbit 26S, a polyubiquitinated proteasomal substrate was needed. In the literature, there are a handful of substrates, however, each is limited in scope and/or synthetic simplicity. We utilized a substrate developed by the Martin group at UC Berkeley, a GFP-tagged unstructured protein, titin, fused to cyclin, a known proteasomal substrate, with an engineered PY motif, a degron which signals E3 ligases for polyubiquitination<sup>10,11</sup>. After expression and purification from <i>E. coli</i>, the GFP-titin-cyclin<sup>Py</sup> was polyubiquitinated by the E3 ligase Rsp5, also bacterially expressed.
A mixture of 2.3 µM E1 enzyme, 5 µM Ubc4 (E2 enzyme), 11 µM Rsp5 (E3 enzyme), 0.5 mM ubiquitin, 40 µM GFP-titin-cyclin\textsuperscript{PY}, 5 mM ATP, and 1 mM DTT in buffer containing 50 mM HEPES pH 7.5, 50 mM NaCl, 10% glycerol was incubated for 3-6 hours at 37°C (Fig 6.2). After the final time point, the substrate mixture was aliquoted out, flash frozen, and stored at -80°C until further use, at which point the substrate was used directly with an assumed stock concentration of 40 µM GFP-titin-cyclin\textsuperscript{PY}-Ubn.

For degradation reactions, rabbit 26S proteasome was incubated with either 1 µM UbVME or buffer control for 2 hours on ice. Degradation reactions were run with one time point per tube, and with 5xSDS PAGE buffer already added to the t=0 time point to prevent any proteasomal degradation. Reactions contained 100 nM 26S proteasome (+/- UbVME), 2 µM GFP-titin-cyclin\textsuperscript{PY}-Ubn, 1
mM ATP, 1 mM DTT, and an ATP recycling system (creatine phosphokinase, inorganic pyrophosphate, creatine phosphate) in 50 mM HEPES pH 7.5, 50 mM NaCl, 10% glycerol. The reactions were run at 25 °C and were quenched with 5x SDS PAGE buffer.

6.3 Results

6.3.1 Impact of Deubiquitinase Inhibition on Proteasome Degradation

The 26S proteasome was purified from rabbit muscle tissue using the Ubl-UIM method developed by the Goldberg group and can be seen in Fig 6.34,5. This method did not purify a large amount of proteasomes and retained the GST-Ubl and His-UIM proteins despite subtraction over Ni-NTA beads. However, there was sufficient to study the effects of inhibition of the associated deubiquitinases, UCH37 and USP14, by UbVME treatment. To this end, proteasomes were treated with and without 1 µM UbVME for 2 hours on ice to catalytically inactivate the endogenous associated deubiquitinases. Then each sample was assessed for its deubiquitinating activity by UbAMC hydrolysis assays and for its 20S core particle activity using a fluorogenic peptide substrate, succinate-Leu-Leu-Val-Tyr-AMC, or SucLLVY-AMC. Not surprisingly, in the presence or absence of UbVME, the proteasome’s 20S activity was not changed (Fig 6.4). However,
Figure A 6.3: Silver-stained SDS PAGE gel of rabbit 26S proteasome. Purified using the Ubl-UIM method.

deubiquitinating activity was completely abolished upon treatment with UbVME (Fig 6.4). UbVME-inhibited proteasomes were next assessed for their ability to degrade a proteasomal substrate, a GFP-titin-cyclinPY fusion protein ubiquitinated by Rsp5, a system developed by the Martin group\textsuperscript{10,11}. This substrate contains cyclin, a known proteasomal substrate, with an engineered degron fused to titin, a largely unstructured protein ideal for degradation by the 26S. Degradation rates in the presence and absence of UbVME were determined by ImageJ analysis of SDS PAGE gels showing disappearance of the highly polyubiquitinated substrate. UbVME inhibition of the proteasome resulted in about half as slow
Figure A 6.4: Activity of UbVME-treated proteasome. A) 20S activity of UbVME-treated proteasome and B) Deubiquitinase activity of UbVME-treated proteasome. C) SDS PAGE gel of GFP-Ubn substrate degradation used in quantitative ImageJ analysis. D) Quantitative results of UbVME inhibition from SDS PAGE gels.
degradation compared to uninhibited proteasomes, despite activity of the 20S core particle being unchanged in the presence of UbVME (Fig 6.4). This would suggest that deubiquitination by UCH37 and USP14 are coupled to degradation, a theory still under investigation.

Due to this result, we were curious about the effect of simultaneous deubiquitinase and 20S inhibition, achieved by the addition of UbVME and the 20S inhibitor MG132. MG132 inhibits the β5 subunit of the 20S core particle, thereby slowing its proteolytic activities by inhibiting one of the proteolytic subunits12,13. We were curious as to the contribution of UbVME in slowing proteasomal degradation compared to MG132, a well-characterized inhibitor. Incubation of the 26S with inhibitors was done on ice for 2 hours in the presence of either MG132 or a DMSO control, or UbVME or a buffer control. First, we tested 20S and deubiquitinase activity by SucLLVY-AMC and UbAMC hydrolysis (Fig 6.5). As expected, MG132 inhibits 20S activity but not deubiquitinating activity, and UbVME inhibits deubiquitinating activity but not 20S activity. 20S activity was slightly enhanced in the presence of UbVME, a phenomenon previously observed by the Goldberg group (using ubiquitin-aldehyde), but explained as an increase in AAA ATPase activity upstream14. Interestingly, MG132 seems to slightly enhance deubiquitinating activity, but is within error, therefore was not further investigated. These results indicate appropriate levels of inhibition of the 20S and deubiquitinases by their respective inhibitors, so
Figure A 6.5: Activity of UbVME- and MG132-treated proteasome. A) 20S activity of UbVME- or MG132-treated proteasomes and B) Deubiquitinase activity of UbVME- or MG132-treated proteasomes. C) SDS PAGE gel of GFP-Ubn substrate degradation used in quantitative ImageJ analysis. D) Quantitative results of UbVME or MG132 inhibition from SDS PAGE gels.
these samples were used to test degradation of the GFP-titin-cyclinPY-Ubn substrate (Fig 6.5).

The rates of degradation of the GFP-Ubn substrate were measured by running time points on SDS PAGE gels to show disappearance of the heavily polyubiquitinated band, which were subsequently quantified by ImageJ analysis. MG132 alone appears to only decrease degradation by about 25% compared to the uninhibited sample, which is understandable because it only inhibits one of the catalytic subunits of the 20S, rather than all three. UbVME alone shows a decrease of 50% in degradation, similar to the results shown above in Figure 6.5. However, the combination of MG132 and UbAMC slows degradation by >90% compared to the uninhibited sample. As this amount is even greater than the additive 75% of the two inhibitors alone, this result indicates that significant inhibition of degradation is occurring in a coupled deubiquitination-degradation mechanism.

6.4 Further Directions

Investigation into the coupling of UCH37/USP14 deubiquitination and degradation by the 26S proteasome could prove vital to our understanding of all protein degradation, but especially how these deubiquitinases may be the first step in regulation of this cellular machine. The experiments addressed here indicate that deubiquitination may be coupled to degradation, however, it is necessary to separate out the effects of deubiquitination by Rpn11 before any conclusions can be made. This could be achieved by incubation of proteasomes
with 1,10-phenanthroline, a known inhibitor of Rpn11\textsuperscript{12,15-17}. It is necessary to determine if deubiquitination by the cysteine protease deubiquitinases has a separate function from that of Rpn11 and which level of deubiquitination contributes most to slowing the rate of proteasomal degradation.

Additionally, the assays described here rely on SDS PAGE gel analysis of the disappearance of a band of highly polyubiquitinated GFP substrates, however, disappearance does not isolate deubiquitination from degradation. These experiments are currently being pursued by a labmate, Michael Sheedlo, using a T7 probe and polyubiquitinated Sic1, another proteasome substrate, to determine the contributions of deubiquitination vs degradation. More specific answers to these questions are needed before we can definitively say that deubiquitination by UCH37 and USP14 are indeed coupled to and a contributing factor during proteasomal degradation.
6.5 References


VITA
Marie Morrow was born in Winston-Salem, North Carolina to Jamie and Christine Morrow. She grew up in Jacksonville, Florida and went to Stanton College Preparatory School where she first came to love chemistry in Missy Ray’s AP/IB Chemistry class. She went on to study chemistry at the University of Florida where she did undergraduate research in the lab of Carrie Haskell-Luevano, developing peptide inhibitors for melanocortin receptors. After graduating with her Bachelor’s degree, she went to Purdue University to pursue her Ph.D in chemistry. In Chitta Das’s group, she has studied the structure and biophysical/biochemical properties of the proteasomal deubiquitinase UCH37. After graduating, she will start a post-doctoral research position in the lab of Cynthia Wolberger, Johns Hopkins University School of Medicine, studying the structure and function of ubiquitination/deubiquitination machinery.
Stabilization of an Unusual Salt Bridge in Ubiquitin by the Extra C-Terminal Domain of the Proteasome-Associated Deubiquitinase UCH37 as a Mechanism of Its Exo Specificity

Marie E. Morrow,† Myung-Il Kim,† Judith A. Ronau,† Michael J. Sheedlo,† Rhiannon R. White,‡ Joseph Chaney,† Lake N. Paul,§ Markus A. Lill,∥ Katerina Artavanis-Tsakonas,‡ and Chittaranjan Das*,†

†Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, Indiana 47907, United States
‡Division of Cell and Molecular Biology, Imperial College London, Sir Alexander Fleming Building, Imperial College Road, London SW7 2AZ, U.K.
§Bindley Biosciences Center, Purdue University, West Lafayette, Indiana 47907, United States
∥Department of Medicinal Chemistry and Molecular Pharmacology, Purdue University, 575 Stadium Mall Drive, West Lafayette, Indiana 47907, United States

Supporting Information

ABSTRACT: Ubiquitination is countered by a group of enzymes collectively called deubiquitinases (DUBs); ~100 of them can be found in the human genome. One of the most interesting aspects of these enzymes is the ability of some members to selectively recognize specific linkage types between ubiquitin in polyubiquitin chains and their endo and exo specificity. The structural basis of exo-specific deubiquitination catalyzed by a DUB is poorly understood. UCH37, a cysteine DUB conserved from fungi to humans, is a proteasome-associated factor that regulates the proteasome by sequentially cleaving polyubiquitin chains from their distal ends, i.e., by exo-specific deubiquitination. In addition to the catalytic domain, the DUB features a functionally uncharacterized UCH37-like domain (ULD), presumed to keep the enzyme in an inhibited state in its proteasome-free form. Herein we report the crystal structure of two constructs of UCH37 from Trichinella spiralis in complex with a ubiquitin-based suicide inhibitor, ubiquitin vinyl methyl ester (UbVME). These structures show that the ULD makes direct contact with ubiquitin stabilizing a highly unusual intramolecular salt bridge between Lys48 and Glu51 of ubiquitin, an interaction that would be favored only with the distal ubiquitin but not with the internal ones in a Lys48-linked polyubiquitin chain. An inspection of 39 DUB−ubiquitin structures in the Protein Data Bank reveals the uniqueness of the salt bridge in ubiquitin bound to UCH37, an interaction that disappears when the ULD is deleted, as revealed in the structure of the catalytic domain alone bound to UbVME. The structural data are consistent with previously reported mutational data on the mammalian enzyme, which, together with the fact that the ULD residues that bind to ubiquitin are conserved, points to a similar mechanism behind the exo specificity of the human enzyme. To the best of our knowledge, these data provide the only structural example so far of how the exo specificity of a DUB can be determined by its noncatalytic domain. Thus, our data show that, contrary to its proposed inhibitory role, the ULD actually contributes to substrate recognition and could be a major determinant of the proteasome-associated function of UCH37. Moreover, our structures show that the unproductively oriented catalytic cysteine in the free enzyme is aligned correctly when ubiquitin binds, suggesting a mechanism for ubiquitin selectivity.

The ubiquitin proteasome system (UPS), present in all eukaryotes, is responsible for the majority of controlled degradation and recycling of proteins within the cell.1−5 Polyubiquitinated, and to some extent monoubiquitinated, proteins are recognized and degraded by the 26S proteasome, a 2.5 MDa self-compartmentalizing proteolytic complex.6−13 It is composed of two major units: the 20S core particle (CP) consisting of 28 subunits and the 19S regulatory particle (RP) containing 19 subunits in yeast. The proteolytic active sites are housed within the luminal chamber of the barrel-shaped CP, capped on both ends by the RP, which contains ubiquitin receptors and enzymes that prepare substrates for degradation. Entry of substrates into the CP is regulated by the RP, primarily by opening and closing of the substrate translocation channel. Before the substrate is translocated into the narrow channel leading to the lumen of the CP, it is obligatorily deubiquitinated with the help of the RP-resident JAMM metalloprotease Rpn1114−16 and unfolded by Rpt subunits that sit within the base subcomplex of the RP.7,9,14 However, additional regulation...
is performed by proteasome-associated deubiquitinating enzymes, whose underlying mechanism is still poorly understood.16,17 Attachment of ubiquitin to a lysine residue(s) on target proteins is catalyzed by the sequential action of three enzymatic systems: E1 (ubiquitin-activating), E2 (ubiquitin-conjugating), and E3 (ubiquitin-ligating) enzymes.18,19 Usually, ubiquitination of a target protein results in the attachment of a polyubiquitin chain in which successive ubiquitin moieties are attached to one of the seven lysines, or the N-terminal amino group of the preceding monomer, to generate a homopolymeric polyubiquitin chain in which successive ubiquitin moieties are thus generated depending on which amino group of ubiquitin is used for chain extension (lysines 6, 11, 27, 29, 33, 48, and 63 or the amino group of Met1). A polyubiquitin chain of a specific topology is meant for a specific type of functional outcome.20–22 For example, a Lys48 (K48)-linked chain usually serves as the signal for proteasomal degradation, whereas K63 chains signal other types of functions such as endocytosis, DNA repair, and NF-κB signaling.23–26

Ubiquitination works as a reversible post-translational modification, like phosphorylation. Deubiquitinating enzymes, or DUBs, can hydrolytically remove ubiquitin from protein adducts, thereby opposing the action of ubiquitin conjugating machinery.27–35 Consequently, DUBs have been found to play important regulatory roles in numerous ubiquitin-dependent cellular processes.32–35 In mechanistic terms, these enzymes can be categorized into two main groups: cysteine proteases and zinc metalloproteases. The zinc metalloproteases consist of only cysteine proteases (USPs), ovarian tumor proteases (OTUs), and Machado-Joseph domain proteases (MJDs).32

UCH37 (also known as UCHL5) is a 37 kDa DUB of the UCH family and is one of the two proteasome-associated DUBs, the other being USP14 (Ubpep6 in yeast), known to regulate protein degradation by the mammalian proteasome.36–40 These associated DUBs, along with Rpn11, a constitutive member of the RP, conduct deubiquitination at the proteasome. However, the activities of the three enzymes are distinct. Rpn11 is responsible for en-bloc removal of polyubiquitin chains prior to (or concurrent with) unfolding and translocation of the substrate into the CP, an activity that appears to be coupled to substrate degradation.15–17 USP14 and UCH37 on the other hand are known to have chain-trimming functions.37,38,41 The importance of these associated DUBs to proteasome function was revealed through pharmacological inhibition of these enzymes. A small-molecule inhibitor of USP14 appears to accelerate proteasomal degradation of certain substrates, whereas UCH37 inhibition can still proteolysis, consistent with distinct functional roles played by the two enzymes.42–44

UCH37 was first identified as the PA700 isopeptidase, the cysteine DUB tightly associated with the RP, also known as PA700.39,43,46 Like other UCH family members, it contains a conserved catalytic triad of a cysteine, a histidine, and an aspartate. UCH37 has a canonical UCH domain that is 45% similar to UCHL1 and 49% similar to UCH3, its single-domain family members.47–50 It also has an additional C-terminal tail domain responsible for its interaction with the Rpn13 subunit of the RP.51–54 Proteasome-bound UCH37 is thought to behave as an “editor”, relieving poorly ubiquitinated substrates from degradation by sequentially dismantling their K48-linked polyubiquitin chains from the very distal end, removing one ubiquitin at a time.37,38,42 Such a type of chain disassembling activity can be termed as an exo cleavage activity in contrast to the endo activity, which leads to dismantling of chains by cleavage between internal ubiquitins. Although it has respectable UbAMC (ubiquitin aminomethyl coumarin) hydrolysis activity in its unbound form, UCH37 has been shown to require association with the proteasome to cleave ubiquitin (and polyubiquitin) chains.37 Additionally, its UbAMC hydrolysis activity is enhanced upon binding with Rpn13.37,54 Interestingly, UCH37 also associates in the nucleus with the human Ino80 chromatin remodeling complex, where it is held in an inactive state compared to the free enzyme.55 It thus serves as an example of a DUB whose catalytic activity is both positively and negatively regulated by binding to specific protein partners, making it an attractive target for structural studies. Crystal structures have been determined for both the catalytic domain and full-length human UCH37;38,39–41 however, the mechanism of its catalytic regulation upon binding to associated protein factors is not known. Any mechanistic understanding of its regulation must require structural information about UCH37 and its catalytic domain bound to ubiquitin, which has yet to be reported.

TsUCH37 is a recently characterized lower-organism homologue of UCH37 from *Trichinella spiralis* (Ts), an infectious helminth found nearly worldwide. TsUCH37 was identified by White et al. by incubation of the whole-cell lysate of *Ts* larvae with the HA-UbVME probe (HA, the hemagglutinin epitope, fused with the N-terminus of ubiquitin vinyl methyl ester), an epitope-tagged irreversible inhibitor of cysteine DUBs.39 Its structural and functional homology with human UCH37 was then confirmed by sequence analysis, co-immunoprecipitation with proteasomal subunits, and UbAMC hydrolysis assays. TsUCH37 is 45% identical to its human homologue and was shown to pull down TsADRM1, the corresponding Rpn13 homologue, by co-immunoprecipitation.52 The sequence and functional conservation between the Ts and human enzymes implies a similar chain-editing role of the former at the proteasome. To understand the mechanisms associated with UCH37, we have crystallized two constructs of TsUCH37 bound to ubiquitin vinyl methyl ester. The structures illuminate the mode of action in the enzyme by revealing binding interactions with the catalytic domain, which are conserved among UCH enzymes, and interactions unique to UCH37, notably ubiquitin binding by the ULD, providing further explanation of the proteasome-associated exo-specific deubiquitination activity of the DUB.

### MATERIALS AND METHODS

**Cloning, Expression, and Purification.** TsUCH37**<sub>226** (residues 1–226) was subcloned from the full-length construct (residues 1–309) in PET28a(+) into pGEX-6P-1 (GE Biosciences) using BamHI and Xhol restriction sites. The protein was expressed in *Escherichia coli* Rosetta cells (Novagen) grown at 37 °C in LB medium containing 100 μg/L ampicillin to an OD<sub>600</sub> of 1.0 and then induced with 0.5 mM isopropyl β-D-thiogalactoside (IPTG) at 18 °C for 16 h. Harvested cells were resuspended in lysis buffer (10 mM phosphate buffered saline and 400 mM KCl) and lysed with a French press. The lysate was then purified on a glutathione S-transferase (GST) column (GE Biosciences) followed by
cleavage of the GST tag by PreScission Protease (GE Biosciences) per the manufacturer’s instructions. It was further purified by size exclusion chromatography on a Superdex 75 column (GE Biosciences). Intein-fused ubiquitin–ΔC46 in pTXR1 was expressed in E. coli Rosetta cells and purified on chitin beads (New England BioLabs). Ubiquitin vinyl methyl ester (UbVME) was synthesized by overnight incubation of Ub−ΔC46−MESNa (MESNa, sodium mercaptoethanesulfonate) with glycine vinyl methyl ester and then purified on a MonoS cation exchange column (GE Biosciences). Glycine vinyl methyl ester was synthesized by a modified, previously published procedure.60 TsUCH37ΔC46 was reacted with UbVME for 4 h, followed by purification on a MonoQ anion exchange column (GE Biosciences) to separate any unreacted TsUCH37ΔC46. Selenomethionine TsUCH37ΔC46 protein (SeMet TsUCH37ΔC46) on a Superdex 75 column (GE Biosciences). Crystals were cryoprotected in 2.5 M sodium malonate and published procedure.60 TsUCH37 cat was reacted with a cation exchange column (GE Biosciences). Glycine vinyl methyl ester was synthesized by a modified, previously published procedure.60 TsUCH37ΔC46 was grown in M9 minimal medium supplemented with selenomethionine, reacted with UbVME, and purified as described above.

TsUCH37ΔC46, TsUCH37ΔC, was cloned previously into pET28(++) with an N-terminally fused His tag (Novagen). TsUCH37ΔC46 was expressed in E. coli Rosetta cells, grown at 37 °C in LB medium containing 10 μg/L kanamycin to an OD600 of 0.8, and then induced with 0.5 mM IPTG at 18 °C for 16 h. Harvested cells were resuspended in lysis buffer [50 mM Tris-HCl (pH 7.6), 200 mM NaCl, and 3 mM β-mercaptoethanol] and lysed with a French press. His-tagged TsUCH37ΔC was purified by immobilized metal affinity chromatography (IMAC) and eluted with 500 mM imidazole. TsUCH37ΔC proteomes were further purified by size exclusion chromatography (SEC) on a Superdex 75 column (GE Biosciences) in 50 mM HEPES (pH 7.6) and 3 mM dithiothreitol (DTT). SDS−PAGE on the fractions indicated a cleavage of the full-length protein, so the construct described is actually a proteolytic cleavage product of the full-length protein. The crystal structure (described below) lacks density for the last 46 amino acids from the C-terminus; therefore, this construct will hereafter be described as TsUCH37ΔC46−ΔC. Fractions containing the target protein were pooled, concentrated, and reacted with UbVME. UbVME was synthesized and reacted with purified TsUCH37ΔC46−ΔC as was done with TsUCH37ΔC46. To separate unreacted TsUCH37ΔC46−ΔC, the complex was further purified by SEC on a Superdex 75 column (GE Biosciences).

Crystallization and Structure Determination. TsUCH37ΔC46−UbVME Complex. The TsUCH37ΔC46−UbVME complex was concentrated to 3 mg/mL in 50 mM Tris (pH 7.6), 200 mM NaCl, and 1 mM DTT. Crystals were grown in 2 days at room temperature by hanging vapor diffusion in 3 M ammonium sulfate and 0.1 M bicine (pH 9.0) with 2 mM l-glutathione (mixture of oxidized and reduced) additive. Crystals were cryoprotected in 2.5 M sodium malonate and flash-frozen in liquid nitrogen.61 Diffraction data were collected on a Mar300 CCD detector (Mar USA) at the 23-ID-B beamline at Argonne National Laboratory. Data up to 1.7 Å were collected on SeMet TsUCH37ΔC46−UbVME crystals at the selenium peak (0.979 Å) for SAD (single-wavelength anomalous dispersion) phasing. Data were processed with HKL2000.62

The initial model was obtained by molecular replacement using the Phenix AutoMR wizard, with a monomer of the TsUCH37ΔC46−UbVME complex as the search model.63 The initial model building was conducted in Coot, and a single round of refinement was conducted in Refmac using TLS restraints.63,64 Structural refinement was conducted in Phenix using TLS group), as well as optimized weighting for stereochemical restraints.63 The data were run through Phenix Xtriage, which confirmed the chosen space group, C2, and did not detect evidence of crystal twinning.65 The completeness of the crystallographic data for the TsUCH37ΔC46−UbVME complex was less than ideal (see Table 2), however, this did not hinder the determination of the structure or the generation of the structural model presented herein and can be ascribed to poor completeness in the highest-resolution shells.

TsUCH37ΔC46−UbVME Complex. The TsUCH37ΔC46−UbVME complex was concentrated to 5 mg/mL in 50 mM HEPES (pH 7.6) and 2 mM DTT. Crystals were grown in 60 days at room temperature in 0.2 M ammonium chloride (pH 5.8) and 18% PEG3350. Crystals were cryoprotected in ethylene glycol and flash-frozen in liquid nitrogen. Diffraction data were collected on a Mar300 CCD detector (Mar USA) at the 23-ID-B beamline at Argonne National Laboratory. Data up to 2.0 Å were collected on TsUCH37ΔC46−UbVME crystals at 1.033 Å. Data were processed with HKL2000.62

The initial model was obtained by molecular replacement using the Phenix AutoMR wizard, with a monomer of the TsUCH37ΔC46−UbVME complex as the search model.63 The initial model building was conducted in Coot, and structural refinement was conducted initially in Refmac using TLS refinement and then using simulated annealing and individual B factor refinement in Phenix.63,64 The data were run through Phenix Xtriage, which confirmed the chosen space group, P3.5, and did not detect any evidence of crystal twinning.

UbAMC Hydrolysis Assay. TsUCH37ΔC46−ΔC was diluted in reaction buffer [50 mM Tris (pH 7.6), 0.5 mM EDTA, 0.1% bovine serum albumin, and 5 mM DTT] to a final reaction concentration of 7 nM and preincubated at 30 °C for 5 min prior to the addition of the UbAMC substrate (Boston Biochem). UbAMC cleavage was measured on a Tecan (Männedorf, Switzerland) fluorescence plate reader with 380 nm excitation and 465 nm emission wavelengths at 30 °C. Data were fit to Michaelis–Menten kinetics in SigmaPlot (Systat Software, San Jose, CA).

Analytical Ultracentrifugation. Sedimentation velocity experiments were conducted with the Beckman-Coulter XLA analytical ultracentrifuge. The sample was extensively dialyzed against 50 mM Tris-HCl, 200 mM NaCl, and 1 mM DTT (pH 7.4). The TsUCH37ΔC46 and TsUCH37ΔC46−UbVME complex concentration ranged from 10 to 32 μM. The samples were centrifuged at 50000 rpm using a two-sector 1.2 cm path-length carbon-filled Epon centerpiece. The experiments were conducted on an An-50 Ti rotor at 20 °C. The density and relative viscosity of the buffers were calculated using SEDNTERP version 1.09 (http://www.rasmol.org/sedntrp/windows/ sednterp-philo). 1.0079 g/mL and 0.01036 P, respectively. The partial specific volume (v̅) of the protein was also calculated from the protein sequence using SEDNTERP (0.7340 mL/g for TsUCH37ΔC46 and 0.7317 mL/g for the TsUCH37ΔC46−UbVME complex). The samples were monitored at 280 nm with a 4 min delay and 150 scans. The c(s) distributions were analyzed using SEDFIT version 13.0b.65

Molecular Dynamics Simulations. A model of full-length TsUCH37 was generated by the SwissModel homology modeling server using the structure of the full-length human protein as a template.66 Missing ULD residues produced by the model were appended to the TsUCH37ΔC46−UbVME structure in Coot, and a single round of refinement was conducted in Phenix, to produce a final model hereafter termed “the system”.63,64 The system was solvated in a box of TIP3P
water with the minimal distance between any solute atom and the boundary of the box set to 10 Å. The system was neutralized with 15 Na\(^+\) ions, which were automatically positioned by the tleap program. Molecular dynamics (MD) simulations were performed using Amber 10 with Amber force field ff03.67. Periodic boundary conditions were applied, and the full electrostatic energy was calculated using the particle mesh Ewald (PME) method.\(^{68}\) The simulation consisted of three sequential steps: energy minimization for 5000 steps (2500 steps of steepest-descent followed by 2500 steps of conjugate gradient minimization), equilibration for 100 ps of solvent with the protein restraint with a force constant of 5 kcal mol\(^{-1}\) Å\(^{-1}\), and a final MD simulation for 2 ns. All simulations were conducted at 300 K with a constant volume. A time step of 2 fs was used, and the SHAKE algorithm was applied to constrain the bonds involving hydrogen atoms.\(^{69}\)

## RESULTS

TsUCH37, like its mammalian counterpart, contains a catalytic UCH domain, and an additional polypeptide chain following it called the C-terminal tail comprising the conserved UCH37-like domain (ULD) followed by a putative KEKE motif (Figure 1b).\(^{37,51,70,71}\) The ULD in human UCH37 is thought to have an inhibitory role, presumably by folding onto the catalytic domain thereby occluding ubiquitin binding.\(^{37}\) However, how ubiquitin binds to UCH37 has not been structurally characterized. To gain insight into how ubiquitin is recognized by TsUCH37, we aimed to crystallize both the catalytic domain of TsUCH37 bound to ubiquitin vinyl methyl ester (UbVME) (the TsUCH37\(^{cat}\)−UbVME complex) and the UbVME complex of the full-length protein. UbVME is a suicide substrate of cysteine DUBs, which react with the former via nucleophilic attack of the catalytic cysteine at the vinyl group of the VME moiety, resulting in an irreversible modification whereby a covalent bond is formed between the catalytic cysteine and the VME portion of the inhibitor (Figure 1a).\(^{36,48,60}\) This covalent adduct is thought to mimic the acyl-enzyme intermediate formed during deubiquitination reactions catalyzed by the DUB (II and IV in Figure 1a). If diubiquitin is used as the substrate, the distal ubiquitin moiety is the acyl component of the acyl-enzyme intermediate, with the proximal ubiquitin acting as the leaving group during isopeptide bond hydrolysis (in diubiquitin, a lysine residue of one ubiquitin, called the proximal ubiquitin, is linked via an isopeptide bond to the C-terminal carboxylate group of another ubiquitin, called the distal ubiquitin) (III in Figure 1a).

The TsUCH37\(^{cat}\)−UbVME complex crystallized in the \(C_2\) space group with two molecules of the complex in the asymmetric unit. Our attempts to crystallize the full-length version, however, were met with limited success, the full-length protein being susceptible to proteolysis as indicated by at least two closely migrating bands in an SDS–PAGE gel (data not shown). While attempting to purify the full-length construct, we managed to retrieve a truncated version of the protein lacking 46 amino acids from the C-terminal end of the protein (see Materials and Methods). This truncated protein was
purified by Ni affinity chromatography and reacted with UbVME, and the complex was purified using ion-exchange chromatography. This complex, hereafter termed the TsUCH37\textsuperscript{cat}–UbVME complex (TsUCH37 missing the last 46 residues), crystallized in the R3 space group with one complex in the asymmetric unit.

The catalytic activity of TsUCH37\textsuperscript{cat} was measured with a Ub\textsubscript{aMC} hydrolysis assay (Figure 1c), which yielded Michaelis–Menten parameters as shown in Table 1. Compared to the catalytic domain of human UCH37, TsUCH37\textsuperscript{cat} has an approximately 20-fold lower $k_{cat}$ indicating a higher affinity for this substrate compared to that of the human protein, but a 100-fold lower $k_{cat}$, a substantially lower turnover number. Consequently, TsUCH37\textsuperscript{cat} is nearly 5-fold less efficient than the UCH domain of human UCH37.

Crystals of the TsUCH37\textsuperscript{cat}–UbVME complex diffracted to 1.7 Å. The structure was determined by single-wavelength anomalous dispersion (SAD) using anomalous scattering from seleniun (TsUCH37\textsuperscript{cat} was labeled with seleniun). Manual model building using Coot, followed by multiple rounds of refinement using Phenix, produced a final model with an R factor of 17.4% and an $R_{free}$ of 21% (see Table 2 for crystallographic and refinement parameters).\textsuperscript{63} The final refined model corresponding to the asymmetric unit consists of two copies of the TsUCH37\textsuperscript{cat}–UbVME complex, composed of TsUCH37\textsuperscript{cat}, residues 1–226, covalently connected via a thioether bond linking the catalytic cysteine with the VME group of UbVME (residues 1–75 of ubiquitin attached to GlyVME as the 76th residue, which is modeled as methyl 4-TsCH\textsubscript{3}7).

Table 1. Kinetic Parameters for TsUCH37\textsuperscript{cat}

<table>
<thead>
<tr>
<th>enzyme</th>
<th>$K_{M}$ (nM)</th>
<th>$k_{cat}$ (s\textsuperscript{-1})</th>
<th>$k_{cat}/K_{M}$ (x10\textsuperscript{5} M\textsuperscript{-1} s\textsuperscript{-1})</th>
</tr>
</thead>
<tbody>
<tr>
<td>TsUCH37\textsuperscript{cat}</td>
<td>1085</td>
<td>0.37</td>
<td>3.4</td>
</tr>
<tr>
<td>UCH37\textsubscript{N466}</td>
<td>21493</td>
<td>34</td>
<td>16</td>
</tr>
<tr>
<td>UCH3L\textsuperscript{a}</td>
<td>77.1</td>
<td>19</td>
<td>2414</td>
</tr>
<tr>
<td>UCHL1\textsuperscript{a}</td>
<td>47.0</td>
<td>0.03</td>
<td>7.4</td>
</tr>
</tbody>
</table>

\textsuperscript{a}Kinetic parameters previously determined, from ref 75.

The TsUCH37\textsuperscript{cat}–UbVME complex was used to determine the oligomerization state of complexed and the apo protein exist as monomers in solution by analytical ultracentrifugation (AUC). We found that both the complex and Gly141 and Glu157, Ala263 of the TsUCH37\textsuperscript{cat} chains (Figure 2a). It is possible that the disulfide bond forms because the protein exists as a dimer in solution, bringing the cysteines into proximity of each other, or is a result of crystallographic packing. To determine if this disulfide is a crystallographic artifact or a biologically relevant association, we determined the oligomerization state of complexed and the apo protein. The TsUCH37\textsuperscript{cat}–UbVME structure and Gly4–Lys57, Thr72–Gly141, and Glu57–Ala63 of the TsUCH37\textsuperscript{cat}–UbVME structure.
The overall structure of the TsUCH37 catalytic domain is similar to that of other structurally characterized UCH enzymes. It has the classical αβα fold, in which a central six-stranded β-sheet is surrounded by six α-helices, five on one side (α1−α5) and one on the other (α6) (Figure 3). The overall architecture of TsUCH37cat can be seen as bilobal, with one of the lobes comprising helices α1−α5 and the other comprising the β-sheets and helix α6. The active site is located at the interface of the two lobes, with Cys85 from helix α2 in one lobe and His161 from β3 in the other forming the catalytic Cys-His pair. An adjacent loop provides the third member of the triad, Asp176. Most of the secondary structural elements seen in TsUCH37cat are conserved in UCHL1 and UCHL3, with the only noticeable difference being the conformation of a segment following β2, residues 57−71. This segment is a helix in UCHL1 and UCHL3 and is in somewhat of an extended looplike conformation in human UCH37 (hUCH37) but is fairly ordered; in the structure of the various constructs of human UCH37 determined so far, this loop has been found to be in a similar conformation regardless of crystallographic packing (Figure S1 of the Supporting Information). In contrast, this segment appears to be flexible in TsUCH37 and is visualized only in the TsUCH37cat−UbVME structure, in which it forms the dimer interface between the two subunits in the asymmetric unit. In the TsUCH37cat−UbVME complex, a crystallographic monomer, this loop is disordered (Figure 3). Although the possibility that its binding can influence the loop dynamics cannot be ruled out, it is unlikely that UbVME has anything to do with the dynamic behavior of the loop because it does not bind to it. We therefore propose that the loop is intrinsically flexible in TsUCH37 but can become ordered under certain circumstances, such as under the constraints of crystallographic packing.

It is possible that the corresponding loop segment in hUCH37 is somewhat dynamic as well, but it appears to be significantly more flexible in TsUCH37. The significance of this difference in dynamics between the two proteins is not clear at the moment. Intriguingly, the loop’s dynamic behavior appears to have an effect on the conformation of a tryptophan residue (Trp55) adjacent to the active site (Figure S2 of the Supporting Information). This tryptophan is conserved among Schizosaccharomyces pombe (Sp), Ts, and human UCH37 (Figure S3 of the Supporting Information). This tryptophan is conserved among Schizosaccharomyces pombe (Sp), Ts, and human UCH37 (Figure S3 of the Supporting Information).
of the Supporting Information). In the TsUCH37ΔC46–UbVME complex, Trp55 makes contact with the OMe group of the suicide inhibitor, which in the actual ubiquitinated protein (a ubiquitinated protein or the diubiquitin motif of a polyubiquitin chain) would be replaced by the hydrocarbon portion of the isopeptide-linked lysine side chain (Figure 1a). The same residue in the TsUCH37ΔC46–UbVME complex shows a different orientation with respect to the OMe group and appears to have adopted a more open position for interaction with the isopeptide unit (Figure S2 of the Supporting Information). Therefore, Trp55 not only may provide important contacts with the isopeptide link to hold it in place near the active site but also may confer a certain plasticity to the active site of UCH37, which may be useful for an induced-fit type of engagement with the substrate.

As stated before, in the TsUCH37ΔC46–UbVME structure, we are able to visualize 40 additional amino acids after the UCH domain, the first 41 amino acids (residues 223–263) of the ULD in TsUCH37. The polypeptide chain, after emerging from the C-terminus of the UCH domain, adopts a helical structure of six turns (α7), takes a U-turn, and then continues as a helix (α8). α7 and α8 are arranged as a helix–turn–helix motif with a number of interhelical contacts, and this motif adopts a similar orientation with respect to the UCH domain as observed in hUCH37 (Figures 3 and 4b). The only difference in this motif between TsUCH37 and hUCH37 is that it is somewhat shorter in the former. The ULD in TsUCH37 appears to have a proteolytically susceptible region after Ala263, perhaps immediately following it, producing a C-terminal truncation we are observing here. When we model the missing part of the ULD, using the structure of hUCH37 as a template (see Materials and Methods), it is apparent that α8 could have continued on after the cleavage site (Figure 4b) almost as a long helix all the way up to residue 285, except for an interruption at Arg268 where four successive residues, including the arginine, adopt nonhelical dihedral angles producing a kink (a kink featuring equivalent residues is also seen in the template structure). As expected from the hUCH37 structure, the model shows that after the interruption, the helix would terminate at or near amino acid 285 (Figure 4), where the polypeptide chain reverses its direction as a turn segment that appears to cap the C-terminus of the helix. The putative KEKE motif was not modeled because it is absent in the template structure. Interestingly, the structure of the TsUCH37ΔC46–UbVME complex reveals side chains from α8 making contact with ubiquitin, specifically with its Lys48 residue, an interaction that may explain the distal end specificity displayed by UCH37 (discussed in more detail below). Also, the side chains from the modeled part of the ULD, missing in our structure, appear to present themselves for additional contacts with ubiquitin. Indeed, the two most conserved residues in the ULD, Glu265 and Asn272, are facing ubiquitin and lie within contact distances (Figure 4c). Thus, it is possible that they may actually bind to ubiquitin. Alternatively, in contrast to what is predicted by the model, these residues may be used for making contact with Rpn13, explaining why they are conserved.

Active-Site Geometry. The catalytic triad in this cysteine protease assumes a canonical arrangement in the ubiquitin-bound complex. The distance between the catalytic cysteine and histidine is 3.9 Å (Nε–Sδ distance) in both structures, and that between the histidine and aspartate is 2.8 Å (Nε–Oδ) in the TsUCH37ΔC46–UbVME complex and 2.9 Å in the TsUCH37ΔC46–UbVME complex. The distance between the CεH group of the catalytic histidine and the side chain carbonyl oxygen of the oxynonan stabilizing glutamine (Gln79) is 3.3 Å in the TsUCH37ΔC46–UbVME complex and 3.1 Å in the TsUCH37ΔC46–UbVME complex, suggesting a significant CH–O interaction between them, an interaction seen in other cysteine proteases as well.23 We were unable to crystallize

Figure 4. ULD–ubiquitin interactions. (a) Sequence alignment of the ULD of UCH37 highlighting conserved residues in UCH37 homologues. Glu265 and Asn272 (according to Ts numbering) are absolutely conserved (highlighted in red). (b) Superposition of the TsUCH37ΔC46–UbVME complex (ULD colored olive and UbVME orange), human UCH37 (ULD colored purple, PDB entry 3IHR), and TsUCH37 with the entire ULD modeled (cyan) based on the structure of the ULD in human UCH37. The model was generated using SwissModel and MD simulation (please see Materials and Methods). This model is taken from a snapshot collected at 1.3 ns during a 2 ns MD simulation run. (c) Structure of the TsUCH37–ubiquitin complex with the entire ULD modeled as shown in panel b, showing that the conserved residues of the ULD could make additional contacts with ubiquitin. The regions marked i and ii are expanded in the panels below. The UCH domain is colored gray.
the apo form of either TsUCH37 or TsUCH37ΔC46. In its place, we use the structure of apo human UCH37 to gain insight into structural changes in the active-site region that may occur upon ubiquitin binding. Comparison with the structures of human UCH37 reveals that the catalytic cysteine has changed its orientation, going from the apo form to the ubiquitin-bound form, adopting a more productive orientation in the latter, an orientation in which the catalytic cysteine’s side chain faces the catalytic cleft (Figure 5g). This analysis suggests that UCH37 exists in an unproductive form in the absence of ubiquitin, with the catalytic thiol facing the interior of the protein rather than the open space in the catalytic cleft, but is induced to adopt a more productive form upon its binding. Thus, UCH37 may offer yet another example of a UCH DUB that undergoes substrate-induced reorganization to a more productive form.

Crossover Loop Flexibility. A common structural feature present in all UCH enzymes is the crossover loop, which in TsUCH37 spans residues 141−157 (connecting α5 with β3). It straddles the active-site cleft as a flexible loop and is known to
provide steric constraint, limiting the size of the leaving group at the C-terminus of ubiquitin.36,75 Accordingly, UCH enzymes, such as UCHL1 and UCHL3, can cleave only small leaving groups from the C-terminus of ubiquitin, not large proteins or another ubiquitin.75 However, UCH37 is known to cleave diubiquitin (and polyubiquitin chains), but only when it is associated with the RP, being activated upon binding to its protein cofactor, Rpn13.37 All previously determined structures of UCH enzymes bound to ubiquitin have shown a resolved crossover loop, which makes contact with at least one residue from the C-terminal tail of ubiquitin. In the apo form of UCHL3, the closest homologue of UCH37, the loop is disordered but becomes ordered when ubiquitin is bound.38,50 The ubiquitin-bound structures of PiUCHL3 and the yeast ubiquitin hydrolase Yuh1 show an ordered crossover loop making contacts with side chains on the C-terminal tail of ubiquitin.38,78 In contrast, the structures of the TsUCH37–UbVME constructs present the only examples so far of a UCH DUB in which the crossover loop is still disordered even after ubiquitin is bound, indicating that the loop is flexible and does not contribute to ubiquitin binding. A small network of van der Waals interactions and hydrogen bonds seem to stabilize part of the crossover loop (residues 152–157) in a short helical conformation in the structure of the TsUCH37ΔC–UbVME complex, but the same segment in the TsUCH37ΔAC–UbVME structure is disordered and hence not visible, supporting dynamic sampling of conformations by this loop. The observation that the crossover loop is flexible despite the bound ubiquitin may be related to its activation by its proteasome cofactor Rpn13.37 By not engaging with ubiquitin, the loop is available to freely interact with the cofactor, which may stabilize it in a conformation that leaves the active site maximally open to accommodate the isopeptide bond between two ubiquitins or between ubiquitin and an acceptor protein.

**Interactions with Ubiquitin.** The interaction of UbVME with the TsUCH37ΔC–UCH domain buries a total of 2355 Å² of solvent accessible surface area, a value comparable to the amount buried in other UCH domain ubiquitin complexes (the buried accessible surface area in the TsUCH37ΔAC–UbVME complex is 2479 Å²).46,76 The interaction is predominantly localized at two areas on TsUCH37, the active-site cleft and the distal site (Figure 5a,b). The active-site cleft engages the C-terminal hexapeptide segment, Leu1–ArgLeuGly-Gly–ValVME, of UbVME with numerous intermolecular contacts that include van der Waals, hydrogen bonding, electrostatic, and water-mediated interactions (Figure 5c). This segment sits in the active-site cleft with an extended conformation to maximize interactions with both backbone and side chain atoms of nearby residues of the enzyme. As seen in other UCH structures, the narrowest part of the active-site cleft surrounds the terminal Gly-Gly motif, with the last Gly (GlyVME in this case) being placed immediately adjacent to the β7 atom of the catalytic cysteine, precisely located for nucleophilic attack on the isopeptide bond (Figure 5a,b). It is interesting to note that Arg172 of UbVME is engaged in at least three major interactions (Figure 5d), suggesting that it contributes significantly to stabilizing the enzyme–substrate complex. The interactions with Arg72 imply that TsUCH37 will find NEDD8 (neural precursor cell expressed, developmentally downregulated 8, a structurally similar ubiquitin-like protein modifier with a sequence that is 60% identical with that of ubiquitin) as a poorer substrate because this arginine is replaced with alanine in NEDD8. Indeed, TsUCH37 does not cleave NEDD8–AMC (see Figure S4 of the Supporting Information). Many of the active-site interactions observed in the ubiquitin-bound structures of UCHL1, UCHL3, PiUCHL3, and Yuh1 are conserved in both TsUCH37 structures. Additionally, those residues surrounding the C-terminal hexapeptide tail of ubiquitin are strongly conserved between the Ts and human protein (Figure 5c).

The interactions at the active-site cleft appear to be necessary for precise cleavage at the terminal glycine residue of ubiquitin, while the distal site provides additional interaction to stabilize the enzyme–substrate complex (Figure 5e,f). The distal site engages the N-terminal β-hairpin of ubiquitin, which docks by utilizing interactions primarily involving the two-residue β-turn segment, Leu8 and Thr9 of ubiquitin. These interactions are mostly hydrophobic in nature, involving van der Waals contact of Leu8 and Thr9 with Val35, Leu36, Ile206, Phe216, and Leu218, residues that constitute the surface-exposed hydrophobic crevace that is the distal site. Leu36, Ile206, Phe216, and Leu218 are conserved among Sp, Ts, Pf, and human UCH37 (Figure S3 of the Supporting Information), suggesting the importance of distal-site binding in enzyme–substrate recognition.

Ile44 of ubiquitin, a residue widely used in recognition by ubiquitin-binding proteins, including DUBs, is seen making van der Waals contacts with Val34 on a greasy loop in TsUCH37, residues 34–36 (residues Val35 and Leu36 extend into the distal-site pocket) (Figure 5f). A similar motif is used in other UCH enzymes to bind to Ile44 of ubiquitin. Val34 of TsUCH37 also makes contacts with His68 and Val70, which, together with Ile44 and Leu8 from the N-terminal β-hairpin turn, form the so-called Ile44 patch on ubiquitin. Thus, the binding potential of the Ile44 patch on ubiquitin appears to be fully satisfied in structures of the two complexes presented here, with each residue in the patch making at least one contact with the enzyme. The structural data presented here are supported by previously reported mutational analysis of the PA700 isopeptidase. Replacing Ile44 and Leu8 from the Ile44 patch with alanine in the distal ubiquitin of a diubiquitin substrate results in significantly impaired catalysis with no detectable hydrolysis product.45 Val34 and Val35 are replaced with tryptophan and serine, respectively, going from Ts to human UCH37 (Figure 5f) (Val34 provides additional contacts with Val70 of UbVME). These residues also show variability among other UCH family members. Subtle differences in the Ile44 patch-binding residues could be one of the contributing factors in the difference in $K_a$ between human and Ts UCH37, especially as most of the residues at the active site are conserved between the two.

There appear to be no striking conformational changes between the ubiquitin-bound form of TsUCH37 and apo hUCH37 except for the aforementioned reorientation of the catalytic cysteine. However, we cannot rule out the possibility that significant conformational changes might have occurred as a result of ubiquitin binding in the Ts enzyme because we could not crystallize its apo form.

**Ubiquitin Binding by the ULD.** As mentioned earlier, the ULD of hUCH37 was thought to have an inhibitory role, presumably by folding onto the catalytic domain and obstructing substrate binding.37 In contrast, the structure of the TsUCH37ΔAC–UbVME complex provides crystallographic evidence that the ULD can actually contribute to ubiquitin binding and therefore can play a productive role in catalysis, Arg261 and Tyr262 on α8 of the ULD approach ubiquitin to
engage in van der Waals contact with three of its side chains, Lys48 (with Arg261) and Gln49 and Arg72 (both with Tyr262) (Figure 6). Most notably, Arg261 is oriented in such a way to engage in close van der Waals contact with the hydrocarbon portion of the Lys48 side chain, forcing it to adopt an unusual conformation that allows an intramolecular salt bridge interaction with Glu51. This interaction is not observed in any of the 39 other ubiquitin-bound DUB structures currently found in the PDB, catalogued in Table 3; the Lys48−Glu51 distance is greater than 5.8 Å in all. Figure 6b shows the orientation of the same lysine in the TsUCH37cat−UbVME complex. Clearly, the orientation is different in this structure, and the intramolecular salt bridge in ubiquitin is absent, suggesting that Arg261 of the ULD plays a role in inducing the unusual orientation of Lys48 of ubiquitin. Arg261 is conserved among Sp, Ts, and human UCH37 (Figure 7) but is replaced with leucine in PIUCH37 (also known as PIUCH54). Tyr262 is conserved in human and Ts forms but is substituted with tryptophan in Sp and PIUCH37. Inspection of the structure reveals that the van der Waals contact with Lys48 is still feasible with leucine in place of arginine and tryptophan can conservatively replace tyrosine as well. Thus, it is likely that ULD binding with Lys48 and subsequent formation of the intramolecular salt bridge we are observing here are conserved features of UCH37 in general.

UCH37, as a part of PA700, is known to selectively cleave polyubiquitin chains from the very distal end, sequentially removing one ubiquitin at a time. The structural basis of this exo cleavage specificity is not yet known. The unique orientation of Lys48 stabilized by Arg261 leading to the intramolecular salt bridge may explain this selectivity. We propose that although a similar type of interaction between Arg261 and ubiquitin’s Lys48 is possible with an internal ubiquitin, the intramolecular salt bridge will be absent in this case because the amino group of the lysine is acylated and hence not charged. Thus, it is the lack of an additional interaction with an internal ubiquitin that makes binding to Lys48 of the terminal ubiquitin more favored, hence the exo selectivity.

**DISCUSSION**

UCH37 is a proteasome-associated UCH DUB known to have polyubiquitin chain-editing function. It preferentially cleaves the chain from its very distal tip. Such a function might rescue...
certain substrates from being committed to further downstream action of the proteasome. It is also possible that certain substrates carry inappropriate polyubiquitin tags that are not optimal for their degradation. The chain-editing function might be essential for releasing these substrates to clear up ubiquitin receptors for binding to productive substrates. A regulator of proteasome function, it is itself regulated by binding to the proteasome: UCH37 is activated upon binding to Rpn13, a subunit of PA700 (the 19S proteasome or RP), the mechanism of which is not understood. We report here the structure of two constructs of UCH37 from the infectious helminth T. spiralis (Ts) bound to the suicide inhibitor UbVME. This work constitutes the first structural analysis of a ubiquitin pathway protein in the organism showing how ubiquitin is recognized by this UCH family DUB in Ts. The structures reveal striking conservation of the ubiquitin binding mode among UCH DUBs, from lower eukaryotes to human (Figure $S$ of the Supporting Information). It also shows important structural differences between other UCH DUBs, such as UCHL1 and UCHL3, some of which could be used for the specialized function of UCH37. While revealing interesting differences, the Ts structures provide a number of details that may also hold true for the human enzyme, advancing our understanding of UCH37 in general.

The active-site cysteine may undergo ubiquitin-mediated reorientation to a more productive form (Figure 5g), making UCH37 yet another example of a UCH DUB that shows regulation of activity by ubiquitin, a feature that may provide selectivity to this group of cytosine proteases. Structures of the two constructs reveal invariant parts of the enzyme, likely less dynamic parts, while also revealing parts that are more dynamic in nature, such as the segment of residues 57–71 and Trp55. Future studies should reveal the role of such dynamic parts in catalysis or regulation thereof.

Importantly, the structure of the construct with the additional 40 amino acids after the UCH domain reveals that the ULD could contribute to ubiquitin binding (Figures 4 and 6), an unexpected finding because it was thought to be inhibitory in the human enzyme. The interaction of Arg261 on the ULD appears to engage Lys48 of the distal ubiquitin in a way that would be energetically most favored with the very terminal ubiquitin in a polyubiquitin chain, possibly explaining the exo specificity displayed by mammalian UCH37. These structural data are consistent with previously reported mutational analysis probing substrate specificity of the PA700 isopeptidase: mutation of Lys48 to cysteine on the distal ubiquitin of a diubiquitin substrate results in severely impaired catalysis. Apart from the broad agreement with the aforementioned experimental work, this observation of the intramolecular Lys48–Glu51 salt bridge in the distal ubiquitin, apparently induced by Arg261, is purely crystallographic at this point, although it seems unlikely that lattice forces have anything to do with it. Even if the opposite is true, the fact that such interactions are physiologically relevant cannot be ignored. The lack of an intramolecular Lys48–Glu51 salt bridge in any other ubiquitin-bound DUB structures to date (Table 3) makes this unusual interaction more intriguing, and worth additional study. This observation therefore lays the structural groundwork for future mutational analysis aimed at validating their existence in solution and their role in substrate specificity.

It is interesting to note that a salt bridge interaction, albeit an intermolecular one, involving Lys48 of ubiquitin and an acidic side chain of the enzyme is also seen in the structure of USP7 bound to ubiquitin aldehyde (the Lys48 side chain of the distal ubiquitin is interacting with Asp305 and Glu308 of USP7). Such bifurcated salt bridges will perhaps contribute substantially to the binding of the enzyme to distal ubiquitin in a K48-linked chain, based on which one may predict that USP7 will also exhibit exo specificity. This needs to be examined.

| Table 3. Lys48–Glu51 Distances for All DUB–Ubiquitin Complexes |
|-------------------|-----------------|----------------|
| PDB entry | DUB–ubiquitin complex | Lys48–Glu51 distance* (Å) |
| 1X3D | UCHL3–UbVME | 9.1, 11.5 |
| 1CMX | YUH1–Ub | 10.8 |
| 2WDT | PUCHL3–UbVME | 7.3, 9.6 |
| 3FW | UCHL1 S18Y–UbVME | 8.7 |
| 5KV | UCHL1 H3M–UbVME | 12.3 |
| 5KW | UCHL1–UbVME | 13.1 |
| 3TM | DUBA–Ub | 8.3 |
| 2YSB | USP21–linear diUb | 7.8, 10.8, 6.6 |
| 1NBF | HAUSP–Ub | 10.0, 10.7 |
| 2AYO | USP14–Ub | 10.9 |
| 3MHS | SAGA complex (UBP8)–Ub | 9.2 |
| 2HDS | USP2, Ub | 9.0 |
| 2G4S | Iso7, Ub | 8.6, 8.9 |
| 2TQ | M48 USP–UbVME | 9.9, 10.8 |
| 3VE | USP2, Ub variant | 7.0 |
| 3VC | USP2, Ub variant | 7.4 |
| 3HP | USP5, Ub covalent | 9.6, 6.3 |
| 3MTN | USP21, ubiquitin-based USP21-specific inhibitor | 8.5 |
| 3TF | USP21, Ub covalent | 7.4, 7.5, 7.4, 7.4 |
| 3N1K | USP9, covalent Ub-like variant | 10.4 |
| 3NHE | USP2a, Ub | 7.4 |
| 3BI | USP2, Ub covalent | 7.6 |
| 4UM | arterivirus papain-like protease 2, Ub | 7.4 |
| 32XH | CCHF viral, Ub-propargyl | 9.5 |
| 46L | OTUB1, Ub | 9.1 |
| 3PT2 | viral OTU, Ub | 10.7 |
| 4HDX | Nairovirus viral OTU, Ub | 8.3, 10.1 |
| 3BY | OTU, Ub | 6.0 |
| 3PRM | CCHF viral OTU, Ub | 9.5, 10.0 |
| 3PR | CCHF viral OTU, Ub | 10.3, 10.9 |
| 3CR | OTU, Ub | 8.4 |
| 4DHZ | h/cceOTUB1-ubiquitin aldehyde-UBC13–Ub | 9.2, 8.8 |
| 4DHJ | cceOTUB1 ubiquitin aldehyde UBC13–Ub complex | 11.4, 8.5, 6.9, 11.8, 9.8, 10.1 |
| 4DOI | OTUB1/UbcH5–Ub | 7.5, 12.6, 12.6, 7.5, 12.6, 7.5 |
| 4DOG | OTUB1/UbcH5–Ub | 11.3, 7.1, 8.0 |
| 3PHW | OTU domain of CCHF virus, Ub | 7.6, 7.2, 9.0, 5.8 |
| 3OS6 | Atatins-3-like, Ub | 9.6, 11.4, 12.8, 11.5 |
| 2JRI | Atatin 3, Ub | 12.3, 12.9 |

"Multiple distance entries refer to those in the other subunits of the crystallographic asymmetric unit.

Supporting Information. It also shows important structural conservation of the ubiquitin binding mode among UCH DUBs, from lower eukaryotes to human (Figure $S$ of the Supporting Information). It also shows important structural differences between other UCH DUBs, such as UCHL1 and UCHL3, some of which could be used for the specialized function of UCH37. While revealing interesting differences, the Ts structures provide a number of details that may also hold true for the human enzyme, advancing our understanding of UCH37 in general.

The active-site cysteine may undergo ubiquitin-mediated reorientation to a more productive form (Figure 5g), making UCH37 yet another example of a UCH DUB that shows regulation of activity by ubiquitin, a feature that may provide selectivity to this group of cytosine proteases. Structures of the two constructs reveal invariant parts of the enzyme, likely less dynamic parts, while also revealing parts that are more dynamic in nature, such as the segment of residues 57–71 and Trp55. Future studies should reveal the role of such dynamic parts in catalysis or regulation thereof.

Importantly, the structure of the construct with the additional 40 amino acids after the UCH domain reveals that the ULD could contribute to ubiquitin binding (Figures 4 and 6), an unexpected finding because it was thought to be inhibitory in the human enzyme. The interaction of Arg261 on the ULD appears to engage Lys48 of the distal ubiquitin in a way that would be energetically most favored with the very terminal ubiquitin in a polyubiquitin chain, possibly explaining the exo specificity displayed by mammalian UCH37. These structural data are consistent with previously reported mutational analysis probing substrate specificity of the PA700 isopeptidase: mutation of Lys48 to cysteine on the distal ubiquitin of a diubiquitin substrate results in severely impaired catalysis. Apart from the broad agreement with the aforementioned experimental work, this observation of the intramolecular Lys48–Glu51 salt bridge in the distal ubiquitin, apparently induced by Arg261, is purely crystallographic at this point, although it seems unlikely that lattice forces have anything to do with it. Even if the opposite is true, the fact that such interactions are physiologically relevant cannot be ignored. The lack of an intramolecular Lys48–Glu51 salt bridge in any other ubiquitin-bound DUB structures to date (Table 3) makes this unusual interaction more intriguing, and worth additional study. This observation therefore lays the structural groundwork for future mutational analysis aimed at validating their existence in solution and their role in substrate specificity.

It is interesting to note that a salt bridge interaction, albeit an intermolecular one, involving Lys48 of ubiquitin and an acidic side chain of the enzyme is also seen in the structure of USP7 bound to ubiquitin aldehyde (the Lys48 side chain of the distal ubiquitin is interacting with Asp305 and Glu308 of USP7). Such bifurcated salt bridges will perhaps contribute substantially to the binding of the enzyme to distal ubiquitin in a K48-linked chain, based on which one may predict that USP7 will also exhibit exo specificity. This needs to be examined. Preferential cleavage from the very distal tip of a Lys48-linked polyubiquitin chain may be a feature common to DUBs that work on chains of this topology. Lys48-linked chains are known...
to adopt a compact structure.\textsuperscript{80} However, the terminal ubiquitin, being less packed than the internal ones (packed from both sides), is more likely to fray and be susceptible to DUB cleavage for stereochemical reasons. Certain DUBs may have evolved a mechanism for grabbing onto those fraying ends and start disassembling chains from there. There may be other DUBs that prefer internal ubiquitins, or the terminal ones on the other extreme end of the chain, such as isopeptidase-T (USP5),\textsuperscript{81} and there may be some with no preference at all. The structure of AMSH-LP (a Lys63-linked chain-specific DUB) in complex with Lys63-linked diubiquitin shows that Lys63 on the distal ubiquitin is not engaged by the enzyme, suggesting it is unlikely to show any preference between the terminal and internal cleavage sites.\textsuperscript{82} This is consistent with the structure of a Lys63-linked chain, which adopts a more extended conformation in crystals and perhaps in solution as well.\textsuperscript{83} Future structural studies should reveal more details explaining exo and endo specificity seen in certain DUBs.

The structural analysis, combined with MD simulation, shows the contribution of the ULD in ubiquitin binding. In theory, certain residues in TsUCH37’s ULD, missing in our structure, also appear to be correctly positioned for contacting ubiquitin. Notably, the modeling study provides a possible explanation of why Glu265 and Asn272 are so strictly conserved in UCH37 from different organisms, with virtually no exception. Contributing to ubiquitin binding, as suggested by our modeling studies, may be one of the functional constraints underlying the conservation of the amino acids, although one cannot rule out whether binding to other proteins such as Rpn13 may be involved. It should be noted that Bap1, a UCH DUB mutated in several cancers, also features a ULD.\textsuperscript{70,71,86} Like UCH37, Bap1 becomes activated upon binding to a larger protein complex, demonstrated with the Drosophila orthologue, Calypso, binding to the polycomb repressor DUB complex.\textsuperscript{87} Interestingly, the putative ubiquitin-binding residues of the ULD of UCH37 are also conserved in Bap1 (data not shown), suggesting a role in ubiquitin binding for Bap1’s ULD as well (in some Bap1 orthologues, the glutamate corresponding to TsUCH37’s Glu265 is replaced with an aspartate). However, human Bap1 has a linker of approximately 300 amino acids separating the UCH domain and its ULD. It will be interesting to see how the ULD positions itself to bind ubiquitin, if it does. Of more interest is...
knowing whether the ULD has independent ability to bind to ubiquitin.

ASSOCIATED CONTENT

Supporting Information

Supporting figures and NEDD8 hydrolysis data (Figure S4). This material is available free of charge via the Internet at http://pubs.acs.org.

Accession Codes

Coordinates and structure factors have been deposited in the Protein Data Bank as entries 46N and 4IG7.

AUTHOR INFORMATION

Corresponding Author

*Brown Laboratory of Chemistry, 560 Oval Dr., West Lafayette, IN 47907. E-mail: cdao@purdue.edu. Phone: (765) 494-5478. Fax: (765) 494-0239.

Author Contributions

M.E.M. and M.L.K. contributed equally to this work.

Funding

Financial support from the National Institutes of Health (1R01RR026273, C.D.) is gratefully acknowledged. The General Medicine and Cancer Institutes Collaborative Access Team (GM/CA CAT) of the Advanced Photon Source at Argonne National Laboratory has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-GM-110438) and the National Institute of General Medical Sciences (Y1-GM-1104).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge Venugopalan Nagarajan, Ruslan Sanishvili, and Craig Ogata at beamline 23-ID-B of the Advanced Photon Source for assistance with data collection. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract DE-AC02-06CH11357. We thank Emma DeWalt and Garth Simpson from the Department of Chemistry, Purdue University, for their assistance with SONICC imaging of initial crystals.

ABBREVIATIONS

SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; DTT, dithiothreitol; IPTG, isopropyl β-D-thiogalactopyranoside; UbVME, ubiquitin vinyl methyl ester; UbAMC, ubiquitin aminomethylcoumarin; UCH37, ubiquitin carboxy-terminal hydrolase 37; DUB, deubiquitinating enzyme or deubiquitinase; PDB, Protein Data Bank.

REFERENCES


Integrated nonlinear optical imaging microscope for on-axis crystal detection and centering at a synchrotron beamline

Jeremy T. Madden,* Scott J. Toth,* Christopher M. Detttmar,* Justin A. Newman,* Robert A. Oglesbee,* Hartmut G. Hedderich,* R. Michael Everly,† Michael Becker,* Judith A. Ronau,* Susan K. Buchanan,c Vadim Cherezov,d Marie E. Morrow,* Shenglan Xu,b Dale Ferguson,b Oleg Makarov,b Chittaranjan Das,* Robert Fischetti b and Garth J. Simpsona*

aDepartment of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47906, USA, bGM/CA@APS, Advanced Photon Source, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, IL 60439, USA, cNIDDK, National Institutes of Health, Building 50, Room 4503, 50 South Drive, Bethesda, MD 20814, USA, and dDepartment of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA. E-mail: gsimpson@purdue.edu

Nonlinear optical (NLO) instrumentation has been integrated with synchrotron X-ray diffraction (XRD) for combined single-platform analysis, initially targeting applications for automated crystal centering. Second-harmonic-generation microscopy and two-photon-excited ultraviolet fluorescence microscopy were evaluated for crystal detection and assessed by X-ray raster scanning. Two optical designs were constructed and characterized; one positioned downstream of the sample and one integrated into the upstream optical path of the diffractometer. Both instruments enabled protein crystal identification with integration times between 80 and 150 ms per pixel, representing a $10^3$–$10^4$-fold reduction in the per-pixel exposure time relative to X-ray raster scanning. Quantitative centering and analysis of phenylalanine hydroxylase from Chromobacterium violaceum cPAH, Trichinella spiralis deubiquitinating enzyme TsUCH37, human $\mu$-opioid receptor complex kOR-T4L produced in lipidic cubic phase (LCP), intimin prepared in LCP, and $\alpha$-cellulose samples were performed by collecting multiple NLO images. The crystalline samples were characterized by single-crystal diffraction patterns, while $\alpha$-cellulose was characterized by fiber diffraction. Good agreement was observed between the sample positions identified by NLO and XRD raster measurements for all samples studied.

Keywords: XRD; NLO; SHG; SONICC; centering; protein; TPE-UVF; microscopy; LCP; two-photon.

1. Introduction

The high photon flux and energy tunability of synchrotron radiation sources have made them indispensable tools for X-ray analysis, with applications spanning protein structure determination through materials science and nanotechnology (Rasmussen et al., 2011; Moukhametzianov et al., 2008; Bates et al., 2006; Berger et al., 2010; Dauter, 2006; Ihee et al., 2010; le Maire et al., 2011; Parker et al., 2006; Riekel et al., 2005). The increasing drive toward tighter focusing has enabled structure determination on ever-smaller crystals and sub-domains within materials, but presents growing challenges for reliable crystal centering. These challenges are particularly relevant for protein crystal diffraction, in which the drive toward fully automated X-ray diffraction analysis at synchrotron sources has introduced bottlenecks in sample positioning (Andrey et al., 2004; Moukhametzianov et al., 2008; Pothineni et al., 2006; Aishima et al., 2010; Cherezov et al., 2009; Stepanov et al., 2011a). Diffraction-quality protein crystals are typically obtained through crystallization screenings, followed by optimization, and then are placed into cryoloops, which are flash-cooled in liquid nitrogen to reduce X-ray damage and aid in sample handling (Dobrianov et al., 1999; Karain et al., 2002). High-throughput methods for automated crystal positioning are frustrated by complications of reliable centering of smaller and smaller protein crystals within more complex and turbid matrices. The current most reliable methods for crystal centering involve rastering the
sample using a focused X-ray beam (Accardo et al., 2010; Hilgart et al., 2011; Cherezov et al., 2009; Stepanov et al., 2011a; Aishima et al., 2010; Song et al., 2007). From the resulting X-ray diffraction images recorded as a function of sample position in the beam, protein crystals are centered based on the locations of strongest Bragg-like diffraction. X-ray fluorescence raster is also relatively fast, but it requires a convenient X-ray fluorescent element to be present in the crystal (Stepanov et al., 2011a). While generally successful, X-ray raster scanning suffers from several limitations. First, the method is relatively slow, often utilizing >2 s per pixel (raster cell), corresponding to analysis times from several minutes up to an hour depending on the number of cells in the raster grid and on the exposure time (Aishima et al., 2010). Rastering is commonly performed first with a coarse grid, and then a finer grid, to minimize the number of cells, and to increase speed. The total pixel number is in turn dependent on the size of the X-ray beam, the speed of the detector and analysis, as well as the scanned size of the cryo-loop and the crystal itself (Cherezov et al., 2009; Song et al., 2007). Recent advances in diffraction image read times using single-photon-counting arrays (pixel array detectors) (Broennimann et al., 2006), allowing integration times as low as 2 ms per image (Aishima et al., 2010), can significantly reduce the time frame for raster scanning measurements. However, the time required for raster scanning will still ultimately be limited by the collective times required to obtain sufficient signal to noise (S/N) in a given pixel, to translate the sample through the X-ray source, and to reconstruct the crystal positions based on automated analysis of the compiled diffraction images. Diffraction is a relatively inefficient process with far more X-ray photons absorbed or inelastically scattered than detected for diffraction analysis, contributing to sample damage, even under the cryogenic conditions typically utilized. With small crystals or beams, incident X-ray intensities must be increased accordingly to achieve diffracted intensities equivalent to those for large crystals, thereby increasing absorbed X-ray dose and exacerbating damage. Alternative methods for automated loop centering based on optical imaging include bright-field image analysis and ultra-violet fluorescence (UVF) microscopy, which takes advantage of intrinsic fluorescent properties of protein crystals (Kain & Stojanoff, 2007; Vernede et al., 2006; Pohl et al., 2004; Andrey et al., 2004; Pothineni et al., 2006). However, algorithms for protein crystal centering (e.g. based on crystal edge-finding algorithms) are error-prone for microcrystals and turbid matrices, such as lipidic cubic phase (LCP). Methods optimized for analysis within the mother liquor often prove unreliable for a loop-mounted crystal, in part because algorithms often cannot easily distinguish between the loop, features in the cryo-cooled mother-liquor and the crystal. Furthermore, both bright-field and UVF imaging are challenging to reliably implement in turbid matrices, where optical scattering frustrates reliable crystal imaging. UVF also has a potential disadvantage of inducing UV photodamage to samples from long exposures, or in highly labile proteins, but the exposure times required for imaging are typically short enough to minimize such effects (Vernede et al., 2006; Chen et al., 2009; Nanao & Ravelli, 2006).

More recently, nonlinear optical imaging (NLO) methods such as second-harmonic generation (SHG) and two-photon-excited UV fluorescence (TPE-UVF) have emerged as viable alternatives for high-contrast crystal visualization (Kissick et al., 2010; Madden et al., 2011). SHG, or the frequency doubling of light, is symmetry forbidden in disordered media (e.g. amorphous protein aggregates or proteins in solution) but is allowed for certain classes of crystals (Haupert et al., 2011). Fortuitously, the chirality intrinsic to proteins typically results in the adoption of SHG-active crystal classes. Recent quantum chemical calculations suggest an SHG coverage of approximately 84% of protein crystals in the Protein Crystal Database using an optimized instrument (Haupert et al., 2012). TPE-UVF provides a complimentary method to SHG for protein crystal detection, with contrast dependent on the presence of aromatic side-chains (primarily tryptophan), independent of crystallinity. Crystals that are weakly active to SHG imaging but contain fluorescent amino acid residues can be detected (Madden et al., 2011). Furthermore, TPE-UVF can aid in distinguishing SHG-active small-molecule and salt crystals from protein crystals. The high selectivity for crystals and negligible background from disordered protein aggregates typically produces high-contrast SHG images, which are highly compatible with automated image analysis algorithms designed for protein crystal detection and centering (Haupert & Simpson, 2011). SHG measurements have recently enabled crystal detection for diffraction centering using off-line instrumentation (Kissick et al., 2013), in which protein crystals were first imaged under cryogenic conditions with an SHG microscope, and then manually compared with diffraction images obtained by X-ray raster scanning with good agreement. A major benefit of NLO instruments is the reduction in time required to determine crystal locations with high contrast, as measurements for an entire loop can be obtained in as little as a few seconds, compared with tens of minutes routinely required for X-ray raster imaging. The spatial resolution of NLO instruments is also high (~1–2 μm), whereas X-ray diffraction (XRD) rastering with this type of resolution would take substantially longer to scan an area equivalent to that of the entire NLO image (>72 h at 1 s per pixel for a 512 × 512 pixel image). Furthermore, reducing the reliance on X-ray raster imaging would minimize X-ray-induced sample damage (Hilgart et al., 2011; Ravelli & Garman, 2006).

By integrating SHG and TPE-UVF imaging directly into a synchrotron X-ray diffraction beamline, the robotic controls, automated positioning capabilities, cryogenics and other beamline utilities of high-throughput synchrotron facilities can be leveraged. However, the spatial constraints of a typical synchrotron X-ray experimental hutch represent a nontrivial hurdle for development of compatible NLO instrumentation. Typical research NLO instruments occupy a large footprint (an optical table approximately 120 cm × 300 cm), far greater than the space available on a typical beamline. In this work, two complementary prototypes for an on-line compa-
A versatile instrument combining synchrotron XRD and NLO imaging are described. Assessment of these systems was performed by direct comparisons between NLO images and those obtained by X-ray diffraction rastering.

2. Experimental methods

Two separate instruments were designed and constructed for integrating XRD and NLO imaging, each with its own advantages and limitations. The upstream version introduced the incident light coaxial and parallel with the direction of the X-ray beam path, while the downstream system was coaxial and anti-parallel. The upstream version was designed to fully integrate with the existing optical path, while the downstream version was optimized for high flexibility and compatibility with diverse beamline configurations. Both systems were rated as Class I laser systems on-site, with enclosed beam paths, shutters and interlocks to ensure no exposed collimated optical radiation. The integrated NLO microscopes were installed at beamlines 23-ID-B and 23-ID-D at the Advanced Photon Source (APS) at Argonne National Laboratory in Argonne, IL, USA. A basic schematic of the instruments and beam paths as they were installed on the synchrotron beamline can be seen in Fig. 1. Detailed descriptions and photographs are provided.

2.1. Integrated nonlinear optical microscope designs

The upstream illumination NLO system was designed to sit above the existing instrumentation at GM/CA beamline 23-ID-B at the APS, and couple directly into the existing optical path. A Fianium FemtoPower 1060 ultrafast fiber laser was utilized, producing $\sim 160$ fs pulses centered around 1060 nm, with a 50 MHz repetition rate, maximum power of 1.5 W, allowing for a maximum power of $\sim 140$ mW at the sample, with 80% of the overall loss arising from the objective. The Fianium source was composed of an oscillator coupled via a 1.5 m fiber to a dispersion compensator and free-space coupler unit, with dimensions of approximately 15 cm $\times$ 13 cm $\times$ 8 cm. A heated doubling crystal (Newlight Photonics Inc., SHG1663-IM, HTS 85141000) was permanently assembled in the beam path, with the fundamental beam focused into the crystal with a plano-convex lens ($f = 35$ mm) and collimated with another plano-convex lens ($f = 100$ mm) after the doubling crystal. The efficiency of SHG from the doubling crystal was controlled by either introducing or removing a 1064 nm zero-order half-wave plate using a flip mount (New Focus, 8892-K). The scanning assembly consisted of a galvanometer mirror (Cambridge Technology, 6210H) and resonant scanning mirror (Cambridge Technology, L-003-3002509), controlling the beam position on the horizontal slow-scan and vertical fast-scan axes, respectively. The beam was directed into a telescopic lens pair consisting of two plano-convex lenses ($f = 75$ mm and $f = 250$ mm) leading to an additional 3.3 beam expansion after the scan head. The incident light then reflected off a dichroic mirror stack (Semrock, PBP01-529/23-25x36 and Chroma, 900dcsp) designed to reflect 1060 nm and s-polarized 530 nm incident light. The p-polarized component of the returning 530 nm light was transmitted by this same dichroic for epi-detected SHG (i.e. SHG detected in the backward direction through the same objective as the incident light). High-reflectivity dichroic mirrors for both 1060 nm and 530 nm light (Semrock, FF550-Di01-25x36) delivered both wavelengths to the back aperture of the 10x objective (Optem, 28-21-10), which was modified with a $\sim 1.2$ mm hole bored through the center to allow X-ray access. In epi, the p-polarized SHG returning through the dichroic mirror was passed through a bandpass filter set (Chroma, HQ530/30m and CVI, 03FCG567/KG3) and into a compact photomultiplier tube (PMT) module (Hamamatsu, H10722-10). SHG and TPE-UVF were collected in the transmission direction by a plano-convex lens ($f = 25.4$ mm)
affixed to a right-angle prism using optical epoxy (Norland Optical Adhesive 63). Another plano-convex lens \( (f = 25.4 \text{ mm}) \) coupled the detected light into a near-UV-comparable liquid light guide (Oriel Instruments, 77554) collimated with a plano-convex lens \( (f = 25.4 \text{ mm}) \) into the detection assembly. Both the SHG and TPE-UVF were then reflected off a primary dichroic beam splitter (Semrock, FF555-Di03-25x36), then separated at a second dichroic beam splitter (Chroma, z1064dc-sp) for selective detection of SHG (through Chroma, HQ530/30m and CVI, 03FCG567/KG3 filters) and TPE-UVF (through Semrock, SP01-532RS-25 and FF01-440SP-25 filters). Both the SHG and TPE-UVF were focused onto the faces of the PMT modules (Hamamatsu, H10722-10) by a plano-convex lens \( (f = 60 \text{ mm}) \) positioned between the primary and secondary dichroic beam splitters. Backlight illumination was achieved using an LED (ThorLabs, MCWHL2) passing through the primary dichroic beam splitter and into the liquid light guide. The illumination light was then focused through the trans-SHG/TPE-UVF collection optics and onto the sample.

The downstream NLO system was also designed with the optical axis of the objective co-axial with the axis of X-ray propagation [Figs. 1(a) and 1(b)], using a similar laser source. The size constraints associated with this beamline, specifically the restrictions imposed by the support structure of the beamline and the area and instruments surrounding the sample, limited the available footprint of the NLO system to 39 cm \( \times \) 19 cm. The scanning assembly was composed of dual galvanometers (Cambridge Technologies, 6210HSM40B), mounted in a two-dimensional galvo 30 mm cage cube (Thorlabs, GCM002), with each scanning mirror rotating along either the x or y axis. With the scan head inducing a 90° turn into the beam path, the incident light was directed through a telecentric lens pair, mounted in a 30 mm cage cube, and composed of an aspheric lens \( (f = 10 \text{ mm}) \) and a plano-convex lens \( (f = 50 \text{ mm}) \), leading to a 5 x beam expansion. The incident light was then focused onto the sample by a long-working-distance IR \( 10 \times \) objective (Mitutoyo, NT46-403) generating SHG at 530 nm. Up to 650 mW of 1064 nm light could be delivered to the sample with this system with the use of the IR objective (compared with 140 mW with the upstream system). The SHG was detected in the epi-direction, collected through the incident objective and reflected through a filter set and onto a compact PMT module (Hamamatsu, H10722-10) by a dichroic mirror (Omega Optical, 580DCLP) centered around 532 nm and mounted in a rotatable kinematically controlled cage cube platform. The SHG signal was detected through a filter set composed of a KG3 (Thorlabs, FGS900) and 530 nm filter (Chroma, z532/10x). Bright-field images were also collected in the epi-direction using a module composed of an aspheric lens \( (f = 20 \text{ mm}) \) and a CMOS camera (Thorlabs, DC1C1645C), manually inserted when bright-field images were desired. Including the laser source, the total footprint of the microscope was 25 cm \( \times \) 15 cm \( \times \) 15 cm. The microscope was translated to the sample, at a height of 1.4 m, to perform SHG detection and centering measurements. The foundation of the microscope was a high-resolution long-travel translation stage (Newport, M-IMS300V), and its electronics box (Newport, ESP 300, three-axis motion controller), capable of translating the laser pulse-compressor/output coupler, the microscope and the support structure to and from the sample between X-ray measurements, corresponding to approximately 20 cm of travel, with an absolute accuracy of 2 \( \mu \)m.

The electronics package was designed and constructed in collaboration with the Jonathan Amy Facility for Chemical Instrumentation at Purdue University (JAFCI). The electronics package integrated the electronics associated with the microscope, including the power supplies, control boards and data acquisition card (National Instruments), into a compact housing for easy mounting and transport, with a footprint of 46 cm \( \times \) 61 cm \( \times \) 31 cm. Data were acquired as photon counts using a gated multi-scaler card (Becker & Hickl, PMS-400a), controlled using a custom-designed Labview program, which was also written in collaboration with JAFCI. Data reconstruction and imaging were completed through ImageJ (NIH, 2011).

### 2.2. X-ray raster scan scheme

XRD analysis and NLO images were acquired on all samples studied on 23-ID-B. Diffraction of kOR-T4L was acquired with a 5 \( \mu \)m-diameter X-ray beam, 5 \( \times \) 5 \( \mu \)m cell size, 12.0 keV X-ray beam, with 1 s exposure times, a photon flux of \( 2.7 \times 10^{10} \) photons s\(^{-1}\) (full unattenuated beam) and a detector distance of 300 mm. Diffraction of TsUCH37 was acquired with a 10 \( \mu \)m-diameter X-ray beam, a 10 \( \times \) 10 \( \mu \)m cell, a photon flux of 1.3 \( \times \) 10\(^9\) photons s\(^{-1}\) (10-fold attenuation) and detector distance of 300 mm. Diffraction of \( \alpha \)-cellulose was acquired with a 10 \( \mu \)m-diameter X-ray beam, a 10 \( \times \) 10 \( \mu \)m X-ray beam with a photon flux of 2.8 \( \times \) 10\(^9\) photons s\(^{-1}\) (50-fold attenuation) and detector distance of 300 mm. The resulting NLO images and XRD raster measurements were compared using ImageJ and JBluIce (Hilgart et al., 2011), which employs DISTL (Zhang et al., 2006), to assess the degree of correlation of the sample position within the loop. The boundaries of the raster grids and raster cell sizes were defined using the software GUI JBluIce (Stepanov et al., 2011b). Bragg candidates, which estimate the number of well-ordered reflections, were generated for each X-ray diffraction image; they are shown color-coded in the figures as unsmoothed XRD raster images. The X-ray beam size was adjusted using a mini-beam collimator (Fischetti et al., 2009).

### 3. Sample materials

Phenylalanine hydroxylase from Chromobacterium violaceum (cPAH) was purified as a glutathione s-transferase (GST) fusion protein. The GST tag was cleaved with PreScission protease (GE Biosciences). For crystallization, cPAH was concentrated to 10 mg ml\(^{-1}\) in a solution of 5 mM HEPES, pH 7.4. Crystals of cPAH were obtained at ambient temperature utilizing hanging-drop vapor diffusion from solution 43 of

A Mosquito LCP robot (TTP Labtech) was used to dispense 100 nL protein-lipid droplets, overlaid with 750 nL well solutions. Intimin crystals grew from 100 mM sodium citrate, pH 4.5–5.5, 50–100 mM NaCl, 100–150 mM MgCl₂, and 30–34% PEG 400. Crystals were mounted directly from the LCP mixture and flash-cooled in liquid nitrogen.

4. Results and discussion

Data were acquired with both downstream and upstream versions of the NLO instrument, and schematic representations along with photographs of the beam paths are shown in Fig. 1.

Fig. 2 (acquired via the upstream system) shows a large TsUCH37-UbVME crystal. Both the presence and position of the crystal can be independently confirmed with bright-field imaging (a), NLO microscopy and XRD measurements. Signal intensities of the corresponding epi-SHG (b), transmission-SHG (c) and TPE-UVF (d) were measured and processed in ImageJ. Although the crystal is visible using conventional optical imaging approaches, NLO microscopy produced substantial improvements in contrast compared with bright-field imaging. An X-ray diffraction raster was acquired (e) and a representative diffraction image is shown (f).

Intimin protein crystals in LCP were examined using the upstream NLO system. In Fig. 3 the bright-field image is shown in (a), with the corresponding trans-SHG image (b), and X-ray raster acquired with a 5 × 5 μm beam, confirming the presence of a protein crystal (c), with the spot having greatest protein-like diffraction circled and the resulting diffraction pattern provided (d). All protein crystals identified by SHG and XRD were accurate for absolute position within the resolution of the 5 μm X-ray beam.

In Fig. 4 (acquired via the upstream system) a bright-field image of a kOR-T4L crystal within frozen lipidic cubic phase is shown (a). As often arises with lipidic mesophase crystallizations, the looped droplets exhibited high optical scattering upon freezing that frustrated conventional bright-field imaging approaches for crystal positioning. Transmission SHG (b) and TPE-UVF (c) images were acquired, exhibiting localized areas (~2–5 μm) of signal within the loop, suggesting the presence of a crystal. Crystals were confirmed via a 5 μm diameter X-ray beam and 5 × 5 μm cell X-ray raster scan (d), in which several pixels exhibit weak, but detectable, diffraction with Bragg analysis consistent with the presence of a protein crystal. Diffraction patterns for the brightest spot are shown in Fig. 4(e). However, signal is observed in the trans-SHG and TPE-UVF images that does not correspond to areas of protein-like diffraction in the X-ray raster image. This signal discrepancy is tentatively attributed to protein crystals that are too small to produce Bragg peaks by XRD, or to the presence of other ordered materials arising in a false positive. False negatives for particular focal planes were also observed, in which analysis of the diffraction patterns obtained from the raster image indicates the presence of protein-like diffraction located in areas that did not exhibit substantial SHG or TPE-UVF due to the finite depth of field (~25 μm). However,
false positives. False positives can arise using TPE-UVF if there is protein aggregate located within the loop because TPE-UVF probes the presence of aromatic residues and is not crystal specific. Salt crystals and protein aggregates are common occurrences with protein crystal growth, generating false positives for SHG and TPE-UVF measurements, respectively. Fortunately, most simple salts adopt SHG-inactive centrosymmetric structures. Complementary use of these two techniques can significantly reduce the likelihood of false positives and false negatives.

Combined NLO imaging and XRD was also applied to studies of α-cellulose, which exhibits fiber-like diffraction. NLO measurements performed on loop-mounted cellulose generated moderate S/N for multiple fibers within the sample loop (Fig. 5, acquired via the upstream system). Although fiber diffraction was evident from the cellulose samples, the DISTL algorithm used in raster scanning, which searches for discrete Bragg reflections or spots and not fiber diffraction, does not indicate these areas, but rather seems to show that no measurable sample is present. Manual inspection of the individual acquisition of multiple focal planes through samples has been observed to recover crystal locations more quantitatively (not shown).

In SHG measurements the possibility of false positives exists from other SHG-active structures. Most notably, some salts commonly used in crystallization screening can adopt non-centrosymmetric SHG-active lattices and produce bright SHG. Alternatively, noncrystalline structures exhibiting molecular ordering over distances significantly greater than the wavelength of light can also potentially produce false positives for SHG. An example of a false positive, from a noncentrosymmetric vanadate salt crystal, is shown in Fig. S1 of the supplementary information4 in which a cryo-loop containing a crystal grown in LCP was examined with the upstream NLO instrument, and yielded substantial signal in the epi- and transmission-SHG directions. X-ray raster scans suggested the presence of salt-like diffraction, in addition to ice diffraction, as there was ice present on the sample loop. Key signatures for an SHG-active salt were found to be bright epi-SHG and little to no detectable TPE-UVF. These salt crystal signatures can be exploited to reduce the likelihood of

Figure 2

(a) Bright-field image of a T. spiralis UCH37 1-226/UbVME complex crystal (~100 μm thick) and the corresponding (b) epi-SHG, (c) trans-SHG, (d) TPE-UVF and (e) X-ray raster scan within the 300 × 300 μm box. (f) X-ray diffraction of a representative 10 μm-diameter area from (c). X-ray energy: 12 keV; exposure time: 1 s; photon flux: 2.7 × 109 photons s⁻¹ (10-fold attenuation); detector distance: 300 mm; maximum theoretical resolution: 2.25 Å. The large difference in the epi- and trans-SHG signals is expected for thick samples owing to the difference in the forward and backward coherence length. The intensities of the two directions will approach equality as the sample thickness approaches the backwards coherence length (~100 nm). Scale bars are 100 μm. (Three darkened spots, apparent in this figure, arose from separate X-ray ‘burn tests’ to assess X-ray damage, the results of which will be published in a future study.)

4 Supplementary data for this paper are available from the IUCr electronic archives (Reference: WAS051). Services for accessing these data are described at the back of the journal.
signals above the background correlated with the areas of the crystal generating a detectable protein-like diffraction, providing preliminary confirmation of the ability of the downstream instrument to rapidly generate information for crystal position as a complement to X-ray raster scanning.

The polyimide loops (MiTeGen) were found to undergo noticeable deformation with less than 100 mW incident power using the downstream system, whereas the nylon loops were more robust, and were not damaged at these powers. No noticeable damage could be induced in either loop types using the upstream system during either SHG or TPE-UVF measurements (120 mW and 90 mW, respectively). Several mechanisms were considered for the observed laser-induced damage to the polyimide loops when measured with the downstream system. Previous studies suggest that damage from multi-photon absorption and plasma formation was found to be an important, if not dominant, mechanism for damage in biological NLO imaging (Sacconi et al., 2006). However, those measurements were performed under conditions of tight focusing [high numerical aperture (NA)] and on live cells/tissues. However, alternative mechanisms may dominate in the present low-NA studies of purified protein crystals maintained under cryogenic conditions. Local heating was also considered as a possible damage mechanism, arising from either one- or two-photon absorption of the incident beam. The marked difference in damage susceptibilities between the upstream and downstream systems is consistent with this mechanism, differing notably in the use of a resonant

Figure 3
(a) Bright-field for an intimin protein crystal generated in LCP with corresponding (b) trans-SHG and (c) X-ray raster summary overlay showing corrected Bragg-like reflection counts. (d) X-ray diffraction of the 5 μm-diameter area corresponding to the red circles in each image, with X-ray energy 12.0 keV, exposure time 1 s, photon flux $2.7 \times 10^{10}$ photons s$^{-1}$ (unattenuated beam), sample-to-detector distance of 300 mm, resulting in a maximum theoretical resolution of 2.25 Å. Scale bars are 50 μm. Cross-hairs were added to (a) and (b) to assist in orienting the field of view with respect to the diffraction raster images.

Figure 4
(a) Bright-field image of a membrane protein (human κ-opioid receptor complex) crystal in lipidic cubic phase and the corresponding (b) trans-SHG and (c) TPE-UVF with (d) an X-ray raster summary overlay showing corrected Bragg-like reflection counts. (e) X-ray diffraction of the 5 μm-diameter area corresponding to the red circles in each image, X-ray energy: 12.0 keV; exposure time: 1 s; photon flux: $2.7 \times 10^{10}$ photons s$^{-1}$ (unattenuated beam); sample-to-detector distance: 300 mm; maximum theoretical resolution: 2.25 Å. Scale bars are 20 μm. Cross-hairs were added to (b) and (c) to assist in orienting the fields of view with respect to the bright-field and diffraction raster images.
By positioning the loop to avoid the outer turning points of the fast-scan mirror or blocking the beam at those locations, no noticeable damage could be induced in the polyimide loops during TPE-UVF imaging.

Both of the NLO imaging systems presented in this paper have strengths and limitations, and either could be utilized as a method for locating and centering protein crystals on a synchrotron beamline. With a small footprint and the ability to insert and remove the instrument, there is potential for a single design of the downstream instrument to be utilized on a variety of different beamlines. However, the time required for translating the entire microscope to and from the sample increases the total time for collecting SHG images and XRD of the protein. Indeed, the microscope positioning required substantially more time (~2 min) than the sample imaging (~40 s). Furthermore, the absolute accuracy of the translation stage (in this case, ±2 μm) can ultimately dictate the precision in crystal positioning. In addition, the downstream instrument did not have transmission-SHG detection capabilities. For protein crystals, detection in transmission provides substantial improvements in detection limits for weakly SHG-active proteins, as thickness greater than the crystals’ coherence lengths can decrease the overall SHG intensity in the epi direction (Boyd, 2009; Kestur et al., 2012). The absence of transmission detection could potentially be remedied by introducing additional optics or integrating into existing optical paths.

The direct integration of the upstream system eliminated the need for a translation stage for inserting the microscope, as was used with the downstream system. This significantly reduced the time between imaging and XRD, which allowed for a marked improvement on throughput of data collection. The upstream system did still require the transmission detection optics to translate in and out for XRD collection in transmission, but epi-detected SHG can be performed concurrently with X-ray diffraction, with only a factor of three reduction in signal intensity with the mini-beam collimator in place. The positioning of the collection optics does not, however, require precise realignment allowing for a significant improvement on the translation time, as compared with the downstream instrument, where the entire microscope requires translation with high precision. The upstream system had some design trade-offs to accommodate the existing optical path, which in part accounted for the lower infrared (IR) throughput and available power in the upstream system. The biggest losses came from the incident objective in which 80% of the IR power was lost from reflections because it was not designed for IR incident light. Choosing optics with a more broadband anti-reflective coating (ARC) will significantly improve the power throughput. Testing performed in-house, with an IR-ARC objective, resulted in a doubling of the IR transmittance, corresponding to an anticipated four-fold improvement in signal at the sample (unpublished). The multiple imaging modes (SHG and TPE-UVF), as well as both epi and transmission detection, improves the ability of the upstream system to detect protein crystals that could otherwise be missed on the downstream system.
Based on these combined results, integrating a NLO microscope with a synchrotron XRD instrument complements stand-alone X-ray raster scanning for crystal centering in three key respects. First, it is expected to minimize radiation-induced sample damage compared with X-ray raster techniques for X-ray labile crystals or small crystals difficult to quickly detect at low X-ray flux (Kissick et al., 2013). Second, NLO microscopy significantly increases the spatial resolution and reduces the total acquisition time for the determination of crystal location. For a large sample area (150 × 150 μm) scanned with a small beam size (5 × 5 μm), X-ray raster images for the protein crystals typically required approximately 30 min to acquire with a 1 s X-ray exposure time. For NLO measurements on identical samples, the acquisition time for the collection of each image was typically <10 s. The downstream NLO system allows 512 × 512 pixel images with 40 s acquisitions, and the upstream system allows 150 × 150 pixel images with 1 s acquisitions, which is roughly a >10^4-fold reduction in the per-pixel acquisition time compared with the X-ray raster acquisition per cell (~3 s per pixel, corresponding to a 1 s exposure, with 2 s of dead-time between pixel acquisitions). The theoretical resolution of the objective was 1.6 μm with 2 μm measured spatial resolution. The downstream NLO system required a total time of 2.5 min for translation of the microscope from its resting position to the sample and then back to the resting position following NLO measurements, resulting in a total acquisition time for each sample of the order of 3 min, which is still significantly faster and of higher resolution compared with X-ray raster scan measurements performed on the same sample. In the upstream system, no dead-time was required for epi-detection (in fact, SHG imaging can be performed while acquiring diffraction measurements), and only a few seconds of translation time were required to raise and lower the collection optics in transmission. Third, for weakly diffracting systems where rapid automated diffraction scoring is challenging, NLO measurements may significantly increase the ability to locate protein crystals.

5. Conclusion

Two different designs of integrated NLO instruments were constructed and characterized targeting applications for automated sample positioning. The systems were evaluated using protein crystals (TsUCH37-UbVME, kOR-T4L, cPAH, Intimin) and fibers (α-cellulose). Both NLO and XRD exhibited good agreement for crystal positioning, consistent with previous off-line measurements specifically targeting protein crystals (Kissick et al., 2013). The integrated NLO and synchrotron XRD instrument was found to enable precise centering of α-cellulose samples for fiber diffraction without requiring the development of an application-specific analysis algorithm. The NLO instrument produced images with <10 s image acquisition times, compared with 3–60 min for X-ray rastering performed at much lower spatial resolution. By nature of the higher resolution of NLO image acquisition, the per-pixel raw data acquisition time was approximately five orders of magnitude faster than X-ray raster scanning. Once fully developed, NLO imaging may serve to identify regions of interest for targeted X-ray scanning, or ultimately serve as the sole or primary method for precise automated crystal positioning, such that all of the X-rays striking the crystal are dedicated to structure elucidation.

Despite these successes, a relatively small variety of crystals were used to characterize the instruments in this initial study. Further studies on a greater diversity of protein crystals will help define the scope of use for NLO methods in automated centering. Additionally, the present study focused exclusively on the hardware for visualization, and not on subsequent algorithms for image analysis and automated crystal positioning. Higher contrast afforded by NLO imaging has the potential to significantly improve the reliability of such algorithms if the combined techniques of SHG and TPE-UVF provide sufficient protein crystal coverage for general-purpose use.

These studies provided a foundation for future efforts combining NLO measurements with synchrotron X-ray diffraction. The data presented here support the use of the NLO microscopy for automated or manual crystal centering prior to or in lieu of raster scanning. Potential scope of use where all optical crystal positioning would be preferred includes the analysis of smaller crystals (<5 μm), where the low crystal volume may present challenges for rapid crystal positioning by X-ray raster scanning. SHG also enables positioning of fibrous material exhibiting fiber diffraction, such as cellulose, collagen, chitin etc. Further potential applications include defect studies, X-ray damage studies and studies of active pharmaceutical ingredients.

The authors acknowledge Huixian Wu, Victoria J. Hall, Emma L. DeWalt, Valerie Pye, Martin Caffrey, Nicholas Noijaj and James W. Fairman for aiding in sample preparation. Instrumentation development was supported in part by the Center for Direct Catalytic Conversion of Biomass to Biofuels (C3Bio), an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Award No. DE-SC0000997, and by the NIH-NIGMS through the R01GM-103401. GM/CA@APS has been funded in whole or in part with federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). Use of the Advanced Photon Source was supported by the US Department of Energy, Basic Energy Sciences, Office of Science, under contract No. DE-AC02-06CH11357. Support is also acknowledged from the NIH Common Fund in Structural Biology, grant P50 GM073197. SKB is supported by the Intramural Research Program of the NIH, National Institute of Diabetes and Digestive and Kidney Diseases.

References
