Structural Study of the Complex Between DNA Polymerase Iota and Ub-PCNA

Harrison Taylor
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science
© Harrison Taylor 2020

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Biochemistry Commons, and the Structural Biology Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/6958

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.
Supervisor
Ling, Hong
*The University of Western Ontario*

This dissertation/thesis is available at Scholarship@Western: [https://ir.lib.uwo.ca/etd/6958](https://ir.lib.uwo.ca/etd/6958)
Abstract

DNA polymerase iota (polι) is a member of the Y-family, polymerases which are key components in translesion synthesis (TLS). As part of the DNA damage response, TLS allows cells to bypass damaged template DNA. Each member of the Y-family is capable of accurately replicating across from certain lesions. All Y-family polymerases are recruited by ubiquitination of the DNA sliding clamp, PCNA, by direct interaction with PCNA and ubiquitin. The mechanism of polymerase choice is not well understood, nor are the interactions between Ub-PCNA and the TLS polymerases. We studied the structure of the complex between the interacting region of polι and Ub-PCNA. Polι appears to be unable to bind all three monomers of homotrimeric Ub-PCNA simultaneously, even in a heavily truncated form. The maximum complex ratio observed was two polymerases per Ub-PCNA ring. This assembly ratio limit may give insight into switching of multiple polymerases at the PCNA platform in DNA damage response.

Keywords: DNA damage, translesion synthesis, PCNA, ubiquitin, UBM
Summary for Lay Audience

Accurate replication of DNA is a key biological process. Organisms have evolved complementary maintenance processes to maintain the integrity of genetic information, such as low-error replication machinery, damage repair mechanisms, and replacement mechanisms using templates. These all work together to keep the mutation rate low. Translesion synthesis (TLS) is one component of the damage-response process, but its role is different. TLS replication tolerates damaged DNA and allows replication to continue past lesions when repair processes cannot respond in time, at the risk of mutation. TLS must be kept inactive to avoid unnecessary mutations, but quickly activated at damaged sites. This is accomplished by modifying the major hub protein for replication processes, a ring-shaped trimeric protein called PCNA. A small marker protein called ubiquitin (Ub) is added to a specific site, which displaces the regular replication machinery and acts as a binding site for the TLS proteins. The structure of the complex between Ub-PCNA and TLS proteins is poorly understood. Each of the TLS polymerases handles different types of damage well, but all of them are called by adding ubiquitin to PCNA. DNA polymerase iota (polι), one of the TLS proteins, was chosen an example to investigate. After initial attempts at structure determination by crystallography failed, indirect methods of probing the complex were used. The stoichiometry of the complex was found to be limited to two small polι constructs per UbPCNA ring by multiple independent methods. Coarse structural determination was also attempted by small-angle X-ray scattering, with ambiguous results. Here we argue that the limited stoichiometry observed and the unusual organization of polι suggests that its role is that of a final attempt at lesion bypass when other TLS polymerases have failed. This is in line with the minimal effect of having knocked out polι in model organisms, and a reasonably safe role considering its high tendency to mutate and relatively small number of lesions it appears to handle compared to the major TLS polymerase.
Acknowledgements

I would like to thank Dr. Hong Ling for giving me the opportunity to pursue my studies under her supervision. I will always remember her advice and guidance through the experience. I would also like to thank the members of my advisory committee, Drs. Murray Junop and Gary Shaw, for their guidance and their ability to give me well-timed pushes in a new and usually rewarding direction.

I would like to thank Dr. Guangxin Xing for making me a far better bench scientist, his brilliant help in working out the nuts and bolts of experiments, and for being my sounding board when I needed to work out a problem. I’d like to thank Dr. Vikash Jha for his encouragement, his help in sorting through difficult data, and for his well-timed gifts of coffee. I’d like to thank my labmates and fellow graduate students: Sam Chu, Daniel Sze, and Derek Tse. Thank you for being my friends at the bench and for being the best coworkers I’ll ever have. I’d like to thank Lee-Ann Briere for teaching a total rookie to run his most important experiments and her helpful words when they were most needed.

I would also like to thank my parents, my friends, and my partner for all their support throughout my studies, for hearing my complaints and for sharing in my successes.

This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. “This project was supported by grant 9 P41 GM103622 from the National Institute of General Medical Sciences of the National Institutes of Health.”
## Contents

Abstract ................................................................................................................................. ii

Summary for Lay Audience ................................................................................................... iii

Acknowledgements ............................................................................................................... iv

List of Figures ....................................................................................................................... vii

List of Tables ......................................................................................................................... viii

List of Abbreviations ............................................................................................................ x

1 Introduction ......................................................................................................................... 1

1.1 DNA Damage .................................................................................................................... 1

1.2 Translesion Synthesis ..................................................................................................... 1

1.3 Y-family polymerases and Polι ..................................................................................... 3

1.4 PCNA and the Regulation of TLS .................................................................................. 6

1.5 Outline of Investigation .................................................................................................. 8

2 Materials and Methods ...................................................................................................... 9

2.1 Bacterial strains and plasmids ...................................................................................... 9

2.2 List of Standard Buffers .................................................................................................. 11

2.4 Expression and purification ............................................................................................ 12

2.4.1 PIPUBM1-PCNA Fusion Protein Purification ............................................................ 12

2.4.2 Polι C-terminal Protein Purifications ......................................................................... 15

2.4.3 PCNA Purification – Primary Method using His-tagged protein ................................ 16

2.4.4 PCNA purification – Method for untagged PCNA ....................................................... 17

2.4.5 Ubiquitin Purification .................................................................................................. 18

2.4.6 hUba1 Purification ...................................................................................................... 18

2.4.7 hUbC S22R Purification .............................................................................................. 19

2.5 Ubiquitination of PCNA .................................................................................................. 20
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1 Ubiquitination Reaction Conditions</td>
<td>20</td>
</tr>
<tr>
<td>2.5.2 Confirmation of ubiquitination specificity by Mass Spectrometry</td>
<td>21</td>
</tr>
<tr>
<td>2.5.3 Confirmation of ubiquitination specificity using mutant PCNA</td>
<td>22</td>
</tr>
<tr>
<td>2.6 Crystallization Trials</td>
<td>23</td>
</tr>
<tr>
<td>2.7 Size-Exclusion Chromatography – Multi-Angle Light Scattering (SEC-MALS)</td>
<td>24</td>
</tr>
<tr>
<td>2.7.1 Sample Preparation for SEC-MALS</td>
<td>24</td>
</tr>
<tr>
<td>2.7.2 Experimental Protocol</td>
<td>24</td>
</tr>
<tr>
<td>2.8 Analytical Ultracentrifugation</td>
<td>24</td>
</tr>
<tr>
<td>2.8.1 Sample Preparation for Analytical Ultracentrifugation</td>
<td>24</td>
</tr>
<tr>
<td>2.8.2 Analytical Ultracentrifugation Sedimentation Velocity Experiments</td>
<td>25</td>
</tr>
<tr>
<td>2.8.3 Analytical Ultracentrifugation Sedimentation Equilibrium Experiments</td>
<td>26</td>
</tr>
<tr>
<td>2.9 Small-Angle X-ray Scattering Experiments (SAXS)</td>
<td>28</td>
</tr>
<tr>
<td>2.9.1 Size-Exclusion Chromatography – Small Angle X-ray Scattering</td>
<td>28</td>
</tr>
<tr>
<td>3 Results</td>
<td>30</td>
</tr>
<tr>
<td>3.1 Purification of polt and (Ub)-PCNA proteins</td>
<td>30</td>
</tr>
<tr>
<td>3.1.1 PIP-UBM1-PCNA fusion proteins can be separated from contaminants</td>
<td>30</td>
</tr>
<tr>
<td>3.1.2 Short polt C-terminal proteins are stable, longer forms are more delicate</td>
<td>32</td>
</tr>
<tr>
<td>3.1.3 PCNA purification is highly effective</td>
<td>33</td>
</tr>
<tr>
<td>3.1.4 Ubiquitin purification is high yield</td>
<td>35</td>
</tr>
<tr>
<td>3.1.5 Ubiquitination reaction is high yield, scalable</td>
<td>35</td>
</tr>
<tr>
<td>3.2 Difficulties in crystallizing PIP-UBM1/Ub-PCNA complex</td>
<td>36</td>
</tr>
<tr>
<td>3.3 The ubiquitination reaction is site-specific and monoubiquitination-specific</td>
<td>37</td>
</tr>
<tr>
<td>3.3.1 K164R mutant PCNA abolishes ubiquitination</td>
<td>37</td>
</tr>
<tr>
<td>3.3.2 Mass spectrometry experiments detect ubiquitination at K164 only</td>
<td>38</td>
</tr>
<tr>
<td>3.4 SEC-MALS experiments demonstrate stability of 1:1 complex</td>
<td>40</td>
</tr>
</tbody>
</table>
3.5 Complex investigation using analytical ultracentrifugation......................... 43

3.5.1 Sedimentation velocity experiments observe a maximum complex ratio of 1:2 ................................................................. 43

3.5.2 AUC-SE experiments suggest that the first and second additions are of similar affinity and in equilibrium ................................................................. 46

3.5.3 Sedimentation velocity experiments with PIP-UBM1-UBM2 show similar 1:2 binding limit. ......................................................................................... 49

3.5.4 AUC-SE experiments with PIP-UBM1-UBM2 show similar binding pattern to PIP-UBM1 ......................................................................................... 51

3.6 Reprocessing of ITC data............................................................................. 53

3.7 Solution structure study by small-Angle X-ray scattering ....................... 55

3.7.1 SAXS measurements of Ub-PCNA are in line with literature values, the 1:1 complex with PIP-UBM1 can be detected after SEC .................................................................................................................. 55

3.7.2 SAXS envelope modelling can reconstruct Ub-PCNA, reconstruction of the complex is ambiguous ......................................................................................... 57

4 Discussion.................................................................................................. 60

4.1 The observed complex ratio between Ub-PCNA and pol is limited to two pol for each Ub-PCNA ring ......................................................................................... 60

4.2 Comparison of methods to investigate binding behaviour ..................... 61

4.3 Complex structure predictions .................................................................. 62

4.4 Implications for TLS mechanism and polymerase selection .................. 63

References....................................................................................................... 64

Appendices..................................................................................................... 69

Appendix A: Purification of Ubiquitination Enzymes .................................. 69

hUba1 Purification.......................................................................................... 69

hUbC S22R Purification.................................................................................... 70

Curriculum Vitae ............................................................................................ 72
List of Figures

Figure 1.1 Translesion synthesis activation .......................................................... 2
Figure 1.2 Diagram of the classic polymerase domain ........................................ 3
Figure 1.3 Comparison of Y-family polymerases ................................................. 4
Figure 1.4 Diagram of the structure of PCNA ....................................................... 6
Figure 1.5 Regulation of DNA damage responses by modification of human PCNA. 7
Figure 2.1 Schematic of polI proteins used in study. ........................................... 12
Figure 2.2 Workflow for ubiquitination site determination by mass spectrometry. 21
Figure 3.1 Fusion Protein Production .................................................................. 31
Figure 3.2 C-terminal protein production ......................................................... 33
Figure 3.3 PCNA Production ............................................................................... 34
Figure 3.4 Ubiquitination Products ..................................................................... 36
Figure 3.5 Ubiquitination Assay by Point Mutation ............................................. 38
Figure 3.6 Coverage map of PCNA by MALDI-TOF fragments. ......................... 40
Figure 3.7 Multi-angle light scattering refractive index vs elution volume and chart showing measured molecular weight ......................................................... 42
Figure 3.8 Absorbance spectrum of PIP-UBM1 .................................................. 44
Figure 3.9 AUC-SV of PIP-UBM1/UbPCNA complex ........................................ 45
Figure 3.10 Sedimentation equilibrium fit for PIP-UBM1/UbPCNA ..................... 48
Figure 3.11 AUC-SV of PIP-UBM1-UBM2/UbPCNA complex ............................. 50
Figure 3.12 Sedimentation equilibrium fit for PIP-UBM1-UBM2/UbPCNA ........... 52
Figure 3.13 Original ITC data analysis, against Ub-PCNA protomer .................... 53
Figure 3.14 ITC isotherm for trimeric Ub-PCNA vs PIP-UBM1 ............................ 54
Figure 3.15 Ub-PCNA and Complex SAXS comparison ....................................... 57
Figure 3.16 Reconstructed model of Ub-PCNA ................................................... 58
List of Tables

Table 2.1 List of bacterial strains used in the study ......................................................... 9
Table 2.2 Standard buffers used in the study ................................................................. 11
Table 2.3 Poli-PCNA fusion proteins used in this study .................................................. 13
Table 2.4 Poli C-terminal proteins used in the study ...................................................... 15
Table 2.5 Untagged PCNA proteins ................................................................................. 17
Table 2.6 Ubiquitination reaction conditions ................................................................. 20
Table 2.7 Crystal screening kits ...................................................................................... 23
Table 3.1 Tryptic digest products of Ub-PCNA identified by MALDI-TOF ....................... 39
Table 3.2 Table of theoretical molecular weights ........................................................... 41
Table 3.3 Sedimentation coefficients and estimated molecular weights for PIP-UBM1/Ub- PCNA complexes ................................................................................................. 45
Table 3.4 Speed/time settings for PIP-UBM1/Ub-PCNA AUC-SE experiment ............... 47
Table 3.5 Dissociation constants for PIPUBM1/Ub-PCNA complex ............................... 48
Table 3.6 Speed/time settings for Ub-PCNA/PIP-UBM1-UBM2 AUC-SE experiment .... 51
Table 3.7 Dissociation constants for PIP-UBM1-UBM2/Ub-PCNA complex ................. 52
Table 3.8 Sample injection conditions for SEC-SAXS experiments .............................. 55
Table 3.9 Basic SAXS measurements for Ub-PCNA, PIP-UBM1, and complex .......... 56
List of Abbreviations

° C  degrees Celsius  
µ-  micro prefix  
M  molar (mol/L)  
Tris  tris-(hydroxymethyl)aminomethane  
DTT  dithiothreitol  
BME  β-mercaptoethanol  
EDTA  ethylenediaminetetraacetic acid  
MBP  maltose binding protein  
Mocr  monomeric Ocr protein  
PIP  PCNA-interacting peptide  
UBM  ubiquitin-binding motif  
PCNA  proliferating cell nuclear antigen  
polt  DNA polymerase iota  
DNA  deoxyribonucleic acid  
aa  amino acid  
Ub  ubiquitin  
Ub-PCNA  ubiquitinated PCNA (usually linked to K164 in this context)  
StrpII  StrepII affinity tag, originally developed as an avidin binding tag  
TEV  tobacco etch virus, usually refers to the sequence-specific TEV protease  
SEC-MALS  size-exclusion chromatography multiangle light scattering  
SEC-SAXS  size-exclusion chromatography small-angle X-ray scattering  
MALDI-TOF  matrix-assisted laser desorption ionization time-of-flight mass spectrometry  
ITC  isothermal titration calorimetry  
AUC  analytical ultracentrifugation  
AUC-SV  AUC sedimentation velocity  
AUC-SE  AUC sedimentation equilibrium  
s  sedimentation coefficient  
$\text{s}_w$  sedimentation coefficient in water  
CV  Column volume  
OD  optical density
His-  His-tagged protein, confers nickel affinity
SDS-PAGE  sodium dodecyl sulfate polyacrylamide gel electrophoresis
mS  millisiemens
NaAc  sodium acetate
rpm  rotations per minute
g  standard gravity
MMT  malic acid/MES/Tris buffer
TCEP  tris(2-carboxyethyl)phosphine
ATP  adenosine triphosphate
PEG  polyethylene glycol, number indicates average molecular weight
PCR  polymerase chain reaction
PIPE cloning  Polymerase Incomplete Primer Extension cloning
IDCL  inter-domain connecting loop
IPTG  isopropyl β-D-1-thiogalactopyranoside
1 Introduction

1.1 DNA Damage

In a living cell, DNA is constantly exposed to diverse forms of damage from both external and internal sources. It is estimated that some form of DNA damage occurs over 10,000 times per day, per cell. Further complicating matters, DNA damage may take many forms. These range in severity from the minor (abasic sites, adducts) to the potentially catastrophic (double stranded breaks). Abasic sites are points where the nitrogenous base has been removed while leaving the sugar-phosphate backbone intact, leaving an empty space across from an unpaired base. Adducts are chemical additions to a correct base. Some adducts are harmless and are added as part of genetic regulation, such as DNA methylation, but other adducts are random occurrences. These random occurrences, especially if the adducts are large or heavily distorting (such as the TT cyclobutane dimer caused by UV radiation) are harmful to the cell, potentially producing a permanent mutation. Even minor DNA damage can create distortions in the DNA double strand: the strand must twist and adjust to accommodate the altered bases.

Cells have evolved a range of responses to DNA damage to maintain genome stability, most famously the myriad repair mechanisms and pathways that lead to apoptosis when the damage is too great. Many of these pathways are error-free. These responses are not, however, appropriate in all situations the cell may encounter.

1.2 Translesion Synthesis

Translesion synthesis (TLS), a DNA damage tolerance mechanism, is an important tool in the DNA damage response kit (Goodman and Woodgate, 2013). TLS fills a niche in between damage repair and apoptosis, occurring during DNA replication. Ordinary replicative polymerases (pol α, δ, and ε in humans) are characterized by high processivity, restrictive active sites, and 3’-5’ exonuclease activity. The narrow active sites reduce the probability of incorporating mismatched bases, and the exonuclease activity acts as proofreading by removing mismatched bases. These features of replicative polymerases are key to the high fidelity of replication but may also be a point of vulnerability for the cell with DNA damage.
DNA replication can be stalled at lesioned sites, which would be deadly to the cell if not resolved (Durando et al., 2013). As noted above, lesioned sites disrupt the normal shape of DNA bases and may distort the shape of the backbone strand around its site. These changes make it difficult for replicative polymerases to replicate across lesions. It becomes difficult to add a base across the lesion and increases the probability that the exonuclease activity will remove an added base. The polymerase stalls at this site, but the helicase at the replication fork decouples and continues to unwind the DNA. If the stalling is not resolved, the increasing tracts of ssDNA will lead to a stress response that will eventually activate apoptosis (Rodriguez et al., 2008). Error-free repair pathways tend to be slow and may not be able to rescue replication in time. For this reason, evolution has selected for the TLS pathway: a damage response that can act fast enough to rescue replication. TLS is initiated by monoubiquitination of the proliferating cell nuclear antigen (PCNA) and involves the displacement of the replicative polymerase and its replacement by a TLS polymerase.

**Figure 1.1** Translesion synthesis activation. Upon encountering a lesion, the replicative complex stalls at the damage site (a). Stalling activates a stress response culminating in ubiquitination of PCNA (b). Ubiquitination displaces the replicative polymerase and recruits the TLS polymerase (c) which can then replicate past the damaged site (d).

TLS does come with risk: it is inherently error-prone and thus mutagenic (Yang et al., 2018). Because of this danger, TLS must be tightly regulated (Ghosal and Chen, 2013). TLS has also been found to be improperly active in cancer cells (Gao et al., 2016; Mutter-Rottmayer et al., 2016). The most prominent family of TLS polymerases is the Y-family (Vaisman and Woodgate, 2017).
1.3 Y-family polymerases and Polτ

Y-family polymerases resemble replicative polymerases but have key differences to permit lesion bypass. These enzymes have low fidelity, low processivity, and low catalytic efficiency when replicating across normal template DNA (Yang, 2014). The basic Y-family organization contains a structured N-terminal catalytic domain and an unstructured C-terminal regulatory domain. The catalytic domain is similar to the classical polymerase structure, with the right-hand organization of a palm, finger, and thumb domain.

![Diagram of the classic polymerase domain](image)

**Figure 1.2** Diagram of the classic polymerase domain (left) and the same superimposed over a right hand for context (right). The finger domain (green) binds the incoming nucleotide while the thumb domain helps to position the DNA strand. The palm domain contains the active site.

The domain is differentiated by a looser active site, lack of exonuclease activity, and the addition of a further domain known as the little finger (LF) domain. These differences lead to an improved ability to accommodate and replicate across damaged template DNA at the cost of fidelity. The C-terminal domain regulates the protein’s activity. Though largely unstructured, they contain interacting motifs to that localize the enzyme to ubiquitinated proliferating cell nuclear antigen (PCNA), the eukaryotic. The two most important types of interacting motifs are the PCNA-interacting peptide (PIP or PIP box) (Haracska et al., 2005) and motifs for binding ubiquitin. The PIP-box of Y-family polymerases are non-canonical (Haracska et al., 2005) and have less affinity for PCNA.
relative to replicative polymerases. There are two classes of motifs for ubiquitin binding: the ubiquitin binding zinc finger (UBZ) and the ubiquitin-binding motif (UBM). Together, these binding motifs lead to strong affinity for Ub-PCNA but poor affinity to PCNA in general. This keeps Y-family activity at a minimum under normal conditions but rapidly recruits them to stalled replication forks after PCNA ubiquitination. Y-family polymerases are themselves able to be ubiquitinated, and it is hypothesized that this ubiquitination prevents them from activating by keeping them in a self-binding conformation to preclude them from binding to Ub-PCNA (McIntyre et al., 2013).

There are four members of the Y-family polymerases, each with different activities.

![Diagram of Y-family polymerases](image)

**Figure 1.3** Comparison of Y-family polymerases. N-terminal catalytic domain organization is conserved across all members of the family, with classic polymerase core in red and the little finger domain in red. The C-terminal regulatory domain shows significant differences in composition and organization (adapted from Yang and Woodgate, 2007).

Each polymerase displays different catalytic properties and play unique roles. Rev1 is an unusual case: its catalytic activity is limited to adding dC across from any template, and it itself serves as a scaffold protein for further responses. Polη is the best studied member of the family. Humans lacking polη activity are afflicted with a variant of xeroderma pigmentosum (XPV), a skin disease which is characterized by high sensitivity to UV light. This is caused by losing polη’s ability to bypass TT cyclobutane dimers.
quickly and accurately (Masutani et al., 1999), which is a function that other TLS polymerases lack. Polη also has a role in promoting monoubiquitination of PCNA at K164 (Durando et al., 2013). Polκ’s primary function is strand extension following lesion bypass (Jha and Ling, 2018; Lone et al., 2007), but it is also capable of bypassing lesions on its own. Polη’s role in TLS is poorly understood. Its activity largely overlaps with that of polη, and likely functions as a backup or complement polymerase, such as its activity in XPV organisms (Wang et al., 2007) and its accurate bypass of 8-oxo-guanine lesions (Kirouac and Ling, 2011). One of its notable qualities is its unusual mutation spectrum: polη replicates inaccurately over a template T and favours misincorporation of dC over dA by a factor of 3:1. Despite its inaccuracy, polη has been shown to play a role in maintaining genetic integrity (Iguchi et al., 2014). The structure of its catalytic domain (PDB code: 1T3N) has been solved in complex with various substrates (Nair et al., 2004).

Polη follows the basic Y-family organization, and its regulatory domain contains a PIP-box and two UBMs. The UBM binds to ubiquitin in an unorthodox manner and does not bind to the typical isoleucine-44, rather the nearby isoleucine-8 (Bomar et al., 2010). The binding is also stabilized by electrostatic contacts (Burschowsky et al., 2011).

Polη is a useful model for studying the Y-family polymerase family’s behaviour in vitro. Full-length Y-family proteins tend to be difficult to overexpress and purify in significant quantities, but the catalytic domain and regulatory domain may be individually soluble. The PIP-box of polη is located between the two domains, so that proteins of each domain can be created which contain the PIP-box in its natural point in the structure. This permits independent study of the N-terminal and C-terminal domains in proper complex with its key binding partner and regulator, PCNA.
1.4 PCNA and the Regulation of TLS

Human proliferating cell nuclear antigen (PCNA) is a key hub that physically bridges DNA and a multitude of proteins that act on DNA. PCNAs are the eukaryotic analogue to the bacterial β clamp. In humans, the protein naturally forms a ring-shaped homotrimer which can be loaded onto DNA and slide along the strand. The angle of the basic patches matches that of the backbone and contributes to the “cogwheel”-type sliding (De March et al., 2017). PCNA does not have enzymatic activity; rather it is a platform for a wide range of proteins.

![Diagram of the structure of PCNA](image)

Figure 1.4 Diagram of the structure of PCNA. PCNA is a ring-shaped homotrimer. Each unit is composed of two similar domains, between which is a hydrophobic pocket and an unstructured region called the inter-domain connecting loop (IDCL, shown in black). The overall appearance is a ring with pseudo-sixfold rotational symmetry. The central pore is lined with basic residues which form weak interactions with the DNA backbone, permitting its sliding action (March and Biasio, 2017).

The hydrophobic pocket and IDCL features form a binding site for the PIP box, which is a common binding motif in proteins which interact with DNA. The crystal structure of the polt PIP-box in complex with Ub-PCNA is known (Hishiki et al., 2009). The PIP binding site sits on the forward face of the ring, in the direction of the sliding. The angle of the basic patches matches that of the backbone and contributes to the “cogwheel”-type sliding (De March et al., 2017). PCNA does not have enzymatic activity; rather it is a platform for a wide range of proteins.
In addition to the PIP-binding site, PCNA is subject to varied modifications to recruit specific proteins under specific circumstances. PCNA has been found to be SUMOylated, ISGylated, and ubiquitinated, all at multiple possible sites (Chen et al., 2011). Each modification signals for different responses, so PCNA acts as both the interface between DNA and its proteins and the central point of regulation for many DNA-related processes.

![Diagram of PCNA modifications](image)

**Figure 1.5** Regulation of DNA damage responses by modification of human PCNA. Monoubiquitination of K164 is the signal that recruits translesion polymerases, including Y-family polymerases, to stalled replication forks to bypass lesions (adapted from Chen et al., 2011).

Stalled replication leads to monoubiquitination of PCNA at lysine-164 by RAD18, which activates the TLS response. This dislodges the replicative polymerase and recruits TLS polymerases, including polη. The exact mechanism is unknown, as is the structure of
the active TLS complex. All TLS enzymes are recruited by this same signal, which introduces a polymerase-selection problem: how can a cell ensure that a given lesion is bypassed by the most suitable polymerase? There are two competing models for this complex: the tool-belt model (Boehm et al., 2016) and the polymerase-switch model (Dovrat et al., 2014).

The tool-belt model is based on the homotrimeric structure of PCNA. Since PCNA (and Ub-PCNA) has threefold rotational symmetry, the system effectively contains 3 sets of identical binding sites. The active complex is hypothesized to be able to host a set of TLS polymerases (for example, pols η, ι, and κ) and rotate around the DNA to allow each polymerase to interrogate the lesion. In this way, the most suitable polymerase is the most likely to bypass the lesion and resolve the stalling.

The polymerase-switch model is not mutually exclusive with the tool-belt model, but it may occur without the creation of tool belts. In this model, only a single polymerase binds to Ub-PCNA at any time, requiring a complete detachment and new binding event to change the polymerase that is acting on DNA. Polymerases that bind to Ub-PCNA are themselves ubiquitinated while bound. Should the polymerase detach (a common occurrence since Y-family polymerases are distributive enzymes, with low processivity), it reverts to its self-binding closed conformation and is prevented from re-binding to Ub-PCNA. Like the tool-belt model, this model resolves the polymerase-selection problem by decreasing the probability that the lesion will be bypassed by a polymerase which does not easily act upon the lesion. Though this is perfectly compatible with a tool-belt system (since there are more than three TLS polymerases, it would be necessary for that model to function), this model does not require in vivo tool belts to function and is compatible with a system using one bound polymerase at a time.

1.5 Outline of Investigation

This project set out to investigate the structure of the complex between poli’s regulatory domain and Ub-PCNA. Determination of the crystal structure of the complex between Ub-PCNA and poli may offer clues about the mechanism of TLS and the behaviour of the lesion bypass complex.
Separate proteins of polI and (Ub-)PCNA and fusion proteins of two proteins had been developed for crystallization trials. After crystallization trials were exhausted, the unfused complex was examined to determine whether the complex was forming as hypothesized. Initial findings in multiangle light scattering experiments, followed by analytical ultracentrifugation and small-angle X-ray scattering studies, showed that the complex composition was not the expected 3:1 polymerase:trimer. Instead, the highest stoichiometry ever observed was 2:1 polymerase:trimer. Small-angle X-ray scattering experiments were performed to create a low-resolution structure in solution.

2 Materials and Methods

The following lists of materials were used in experiments. In certain procedures the standard buffer may not be used; in this case the modification may be noted in the section describing the method in question.

2.1 Bacterial strains and plasmids

*Escherichia coli* was the organism of choice for cloning procedures and protein production. The following strains were used in this project.

<table>
<thead>
<tr>
<th>Strain Name</th>
<th>Use</th>
<th>Key Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>BL21-DE3</td>
<td>Protein Production</td>
<td>Protease deficient, primary protein production strain (Novagen)</td>
</tr>
<tr>
<td>BL21-DE3 (pRARE)</td>
<td>Protein Production</td>
<td>Protease deficient, pRARE plasmid requires chloramphenicol selection (Novagen)</td>
</tr>
<tr>
<td>DH5α</td>
<td>Cloning</td>
<td>High transformation efficiency (Thermo Fisher Scientific)</td>
</tr>
</tbody>
</table>

The following plasmids were used as vectors for protein expression. Each plasmid is IPTG-inducible, encoded an N-terminal solubility tag and contained an antibiotic resistance for selection. The solubility tag is cleavable by TEV protease. The exception to the tagging system is the pET3a plasmid, which encodes a C-terminal 6xHis tag which is not cleavable. This vector was used for a single ubiquitination enzyme.
The pRARE plasmid was not transformed into bacteria at the same time as the vector plasmid; rather it was included in the competent cells. The plasmid encodes tRNAs that are rarely used in *E. coli* but common in other organisms. The extra tRNAs produced help to increase protein yield by removing the scarcity as a bottleneck. It is included on this list for the sake of completeness and to note the antibiotic required to select for its inclusion.
2.2 List of Standard Buffers

Each buffer listed with its formulation and its main use. In cases where an experimental protocol uses a modified buffer, the modification is listed in its Methods section.

**Table 2.2** Standard buffers used in the study.

<table>
<thead>
<tr>
<th>Buffer Name</th>
<th>Use</th>
<th>Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>His-Binding Buffer</td>
<td>Ni Affinity Chromatography</td>
<td>50 mM Tris pH 7.5, 500 mM NaCl, 5% glycerol (v/v)</td>
</tr>
<tr>
<td>MBP-Binding Buffer</td>
<td>Amylose Affinity Chromatography</td>
<td>20 mM Tris pH 7.5, 200 mM NaCl, 1 mM EDTA</td>
</tr>
<tr>
<td>Dialysis Buffer</td>
<td>Overnight removal of imidazole</td>
<td>25 mM Tris pH 7.5, 200 mM NaCl, 5% glycerol (v/v)</td>
</tr>
<tr>
<td>Tandem Loading Buffer</td>
<td>1-step 2\textsuperscript{nd} Ni affinity and Q column loading</td>
<td>25 mM Tris pH 7.5, 150 mM NaCl, 5% glycerol (v/v)</td>
</tr>
<tr>
<td>Q Buffer A</td>
<td>Anion Exchange Chromatography</td>
<td>25 mM Tris pH 7.5, 2.5% Glycerol, 1 mM EDTA</td>
</tr>
<tr>
<td>Q Buffer B</td>
<td>Anion Exchange Chromatography</td>
<td>25 mM Tris pH 7.5, 1 M NaCl, 2.5% glycerol (v/v), 1 mM EDTA</td>
</tr>
<tr>
<td>Heparin Buffer C</td>
<td>Heparin Affinity Chromatography</td>
<td>50 mM sodium acetate (NaAc) pH 5.5, 5% glycerol</td>
</tr>
<tr>
<td>Heparin Buffer D</td>
<td>Heparin Affinity Chromatography</td>
<td>50 mM NaAc pH 5.5, 1 M NaCl, 5% glycerol (v/v)</td>
</tr>
<tr>
<td>Sizing Buffer</td>
<td>Standard Experimental Buffer</td>
<td>20 mM Tris pH 7.5, 150 mM NaCl</td>
</tr>
<tr>
<td>SAXS Buffer</td>
<td>SEC-SAXS running buffer</td>
<td>20 mM Tris pH 7.5, 150 mM NaCl, 2.5% glycerol (v/v)</td>
</tr>
</tbody>
</table>
2.4 Expression and purification

The protein complex of interest is composed of UbPCNA and polt. UbPCNA is created by enzymatically ubiquitinating purified PCNA in vitro. Full-length polt cannot be expressed in soluble form, so truncated forms of the C-terminal were created. Fusion proteins featuring linked PCNA and polt were also created for crystallization purposes.

![Diagram showing various fusion proteins involving PCNA and polt](image)

**Figure 2.1** Schematic of polt proteins used in study. Segments of the polt C-terminal region (green) were expressed and purified, both as individual proteins and as fusion constructs with PCNA (blue). Fusion constructs are connected by flexible linkers. The PIP-UBM1 protein is the narrowest segment of polt used, including the PIP box (yellow) and UBM1 (orange). The LF constructs include the little finger (LF) domain at the N-terminal. PIP-UBM1-UBM2 constructs include the full C-terminal domain and all interacting regions of interest.

2.4.1 PIP-UBM1-PCNA Fusion Protein Purification

Fusion proteins were developed to express the polt-PCNA complex as a single unit to eliminate the need for excess PIP-UBM1 to saturate the PCNA ring and to simplify the production process. The C-terminus of the PIP-UBM1 protein was fused to the N-terminus of PCNA by a flexible linker. The linker must be flexible and long enough to permit the desired interactions. It must also be polar to stabilize solvent interactions (Chen et al., 2013).
Table 2.3 Poli-PCNA fusion proteins used in this study.

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PIP-UBM1-PCNA</td>
<td>Fusion between poli 415-530 and PCNA N-terminus, linked by a 24aa flexible linker. Purified using TEV-cleavable N-terminal His-Mocr tag.</td>
</tr>
<tr>
<td>StrpII-PIP-UBM1-PCNA</td>
<td>Fusion between poli 415-530 and PCNA N-terminus, linked by a 24aa flexible linker. Purified using TEV-cleavable N-terminal His-Mocr tag. Strep-II tag is inserted after the tag such that it is the N-terminal of the cleaved protein</td>
</tr>
<tr>
<td>StrpII- PIP-del26-PIP-PCNA</td>
<td>Fusion between poli 415-530 ΔK439-464 and PCNA N-terminus, linked by a 24aa flexible linker. Purified using TEV-cleavable N-terminal His-Mocr tag. Strep-II tag is inserted after the tag such that it is the N-terminal of the cleaved protein</td>
</tr>
<tr>
<td>StrpII- PIP-del36-PIP-PCNA</td>
<td>Fusion between poli 415-530 ΔK439-474 and PCNA N-terminus, linked by a 24aa flexible linker. Purified using TEV-cleavable N-terminal His-Mocr tag. Strep-II tag is inserted after the tag such that it is the N-terminal of the cleaved protein</td>
</tr>
<tr>
<td>StrpII- PIP-del46-PIP-PCNA</td>
<td>Fusion between poli 415-530 ΔK439-474, ΔK484-493 and PCNA N-terminus, linked by a 24aa flexible linker. Purified using TEV-cleavable N-terminal His-Mocr tag. Strep-II tag is inserted after the tag such that it is the N-terminal of the cleaved protein</td>
</tr>
</tbody>
</table>
Fusion proteins were developed to express the poli-PCNA complex as a single unit to eliminate the need for excess PIP-UBM1 to saturate the PCNA ring and to simplify the production process. The C-terminus of the PIP-UBM1 protein was fused to the N-terminus of PCNA by a flexible linker (see Figure 2.1). The linker must be flexible and long enough to permit the desired interactions. It must also be polar to stabilize solvent interactions (Chen et al., 2013).

The StrpII tag (Schmidt and Skerra, 2007), originally designed as an affinity tag, was added to the N-terminal of the cleaved protein. The tag was originally considered to stabilize the N-terminus of the protein and eliminate a possible degradation product, but it was ultimately included to alter the crystallization behaviour and open a new round of screening (Lukat et al., 2008).

All fusion proteins were expressed in E. coli BL21 (DE3) cells. Cells are grown in lysogeny broth (LB) + Ampicillin (100 mg/L) at 37 ºC to OD₆₀₀= 0.6 and then induced with 0.5 mM IPTG. PIP-UBM1 fusions are expressed at 37 ºC and harvested 4 hours after induction. LF-PIP-UBM1 fusions are expressed at 16 ºC and grown overnight before harvest. Cells are pelleted and stored frozen at -80ºC.

Cell pellet from 1 L culture was thawed in a room temperature water bath at use and resuspended in a lysis buffer made from His-binding buffer with 1 mM benzamidine, 1 mM PMSF, and 5 mM BME. Cells were lysed using a homogenizer and the crude lysate was centrifuged at 20,000 rpm (approximately 40,000 g) for 30 minutes. The supernatant was poured off and loaded on a 5mL Roche His cOmplete Ni column. The column was washed with 50mL His-binding buffer, 50mL of His-binding buffer adjusted to 1 M NaCl, 50 mL His-binding buffer, and finally 50 mL of His-binding buffer with 10 mM imidazole. The column was then eluted using His-binding buffer with 500 mM imidazole into 12x1.5 mL fractions. Fractions were pooled based on spot test and TEV protease was added at a ratio of 1 mg for every 30 mg of protein in the pooled fractions. The pooled sample was dialyzed against 150mM NaCl, 25 mM Tris pH 7.5, and 5% glycerol at 4 ºC overnight.

The digested sample was retrieved the next morning and loaded onto a tandemly connected Roche nickel column flowing into a 5 mL GE HiTrap Q column equilibrated to 150 mM NaCl. The digested sample was pushed through the column with additional buffer
and the Q column was removed, then connected to an AKTA Purifier for elution. The protein was eluted using a gradient of 15-70% buffer B over 7.5 CV and collected in 1.5 mL fractions based on UV absorbance. A contaminant band produced a pronounced shoulder in the peak, but the contaminant could be reasonably eliminated by pooling early fractions only. The early fractions were pooled and concentrated to 10mg/mL and were used to set up crystallization trays.

2.4.2 polt C-terminal Protein Purifications

Note on buffers: PIP-UBM1 proteins are purified in standard buffers at pH 8.7 to allow binding to Q column. Q buffer B contains 500 mM NaCl to make low-salt buffers more accurately.

*E. coli* BL21 (DE3) cells expressing 8xHis-PCNA were grown in LB + Ampicillin (100 mg/L) to OD$_{600}$ = 0.7 at 37 ºC and induced with 0.5 mM imidazole. Cells were grown a further four hours at 37 ºC before harvest. Cell pellets were stored at -80 ºC.

<table>
<thead>
<tr>
<th>Table 2.4 Polt C-terminal proteins used in the study.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Protein Name</strong></td>
</tr>
<tr>
<td>PIP-UBM1</td>
</tr>
<tr>
<td>PIP-UBM1 del 26</td>
</tr>
<tr>
<td>PIP-UBM1 del 36</td>
</tr>
<tr>
<td>LF-PIP-UBM1</td>
</tr>
<tr>
<td>PIP-UBM1-UBM2</td>
</tr>
</tbody>
</table>

Cell pellet from 1 L culture was thawed in a room temperature water bath at use and resuspended in a lysis buffer made from His-binding buffer with 1 mM benzamidine, 1 mM PMSF, and 5 mM BME. Cells were lysed using a homogenizer and the crude lysate
was centrifuged at 20,000rpm for 30 minutes. The supernatant was then loaded on a 5mL Roche His cOmplete Ni column.

The Ni column was washed with 150 mL of His-binding buffer and then eluted into 12x1.5 mL fractions using His-binding buffer with 500 mM imidazole. Fractions were pooled based on spot test and TEV protease was added at a ratio of 1 mg for every 20 mg of protein and dialyzed overnight against 20 mM Tris-HCl pH 8.7, 200 mM imidazole, 2.5% glycerol, and 5mM BME.

The digested protein was retrieved the next morning and diluted 6x in buffer A. The diluted protein was loaded on a 5mL HiTrap Q column equilibrated in 6% buffer B (30 mM NaCl). The protein was eluted on an AKTA Purifier from the column using a gradient of 6%-50% over 6 column volumes and collected in 1.5 mL fractions based on UV absorbance. Protein was concentrated to 15 mg/mL, spiked to 20% glycerol (v/v), and flash frozen before storage at -80 °C.

2.4.3 PCNA Purification – Primary Method using His-tagged protein

E. coli BL21 (DE3) cells expressing 8xHis-PCNA were grown in LB + Ampicillin (100 mg/L) to OD$_{600} = 0.7$ at 37 °C and induced with 0.5mM imidazole. Cells were grown a further three hours at 37 °C before harvest. Cell pellets were stored at -80 °C.

Cell pellet from 1 L culture was thawed in a room temperature water bath at use and resuspended in a lysis buffer made from His-binding buffer with 1 mM benzamidine, 1 mM PMSF, and 5 mM BME. Cells were lysed using a homogenizer and the crude lysate was centrifuged at 20,000rpm for 30 minutes. The supernatant was then loaded on a 10mL Roche His cOmplete Ni column.

The column was then eluted using His-binding buffer with 500 mM imidazole into 16x1.5 mL fractions. Fractions were pooled based on spot test and TEV protease was added at a ratio of 1 mg for every 40 mg of protein in the pooled fractions. The pooled sample was dialyzed against 200 mM NaCl, 25 mM Tris pH 7.5, and 5% glycerol at 4°C overnight.

Protein sample was retrieved the next morning and spiked to 500mM NaCl. The sample was passed through His cOmplete column again and the flowthrough collected.
Flowthrough was diluted with Q buffer A to 150 mM NaCl and loaded on 10 mL Q column
pre-equilibrated in 15% B. Column was eluted using a gradient of 15-70% B over 7.5 CV
and collected in 1.5 mL fractions based on UV absorbance. Peak fractions were pooled and
concentrated to roughly 10 mg/mL, spiked to 20% glycerol and flash-frozen before storage
at -80 °C.

2.4.4 PCNA purification – Method for untagged PCNA

Untagged PCNA was at first the only PCNA protein available, before HisPCNA
that was TEV-cleavable was designed. A small number of standard PCNA purifications
were performed using this method. The main use of this purification was to produce lysine
to arginine (K/R) mutant PCNA for ubiquitination tests (outlined below).

Table 2.5 Untagged PCNA proteins purified by ion exchange/heparin affinity method.

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCNA</td>
<td>Human PCNA, untagged</td>
</tr>
<tr>
<td>PCNA K164R</td>
<td>Untagged human PCNA with K164R point mutation</td>
</tr>
<tr>
<td>PCNA K168R</td>
<td>Untagged human PCNA with K168R point mutation</td>
</tr>
<tr>
<td>PCNA K164/168R</td>
<td>Untagged human PCNA with K164/168R point mutations</td>
</tr>
</tbody>
</table>

_E. coli_ BL21 (DE3) cells expressing human ubiquitin (Ub) were grown in LB +
Ampicillin (100 mg/L) at 37 °C to OD_{600} = 0.7 and then induced with 0.5mM IPTG. After
induction cells were incubated at 37°C for 3 hours and harvested (normal yield 3.5-4 g/L),
and the cell pellets were stored at -80°C.

Purification began by thawing 1 L worth of cells (3.5-4 g) and suspending in 15%
Buffer B with 5 mM BME, 0.01% (v/v) NP-40, and 1 mM each benzamidine and PMSF.
The suspended cells were then lysed using a homogenizer. The lysate was spun down at
20,000 rpm for 30 minutes at 4 °C. The supernatant was poured off and then loaded on a
pre-equilibrated Q column. The Q column was then connected to the purifier and eluted
using a 15-25% gradient over 10 mL, followed by 25-70% over 40mL. The initial slow
gradient removes a strong band of protein and DNA immediately before the elution of the
PCNA. Fractions with a conductivity between 20 mS and 40 mS were examined by SDS-PAGE and relatively clean fractions were pooled and dialyzed overnight against 10% buffer D to prepare for heparin chromatography the next day.

Protein was loaded on a pre-equilibrated heparin column and then connected to the purifier. The column was eluted using a 10-50% gradient over 40 mL. Peak fractions were pooled and diluted back to 10% D for reloading. The reloaded protein was then eluted using a 10-70% gradient and peak fractions were concentrated for storage, typically 10 mg/mL. Protein was flash frozen and stored at -80 °C.

2.4.5 Ubiquitin Purification

*E. coli* BL21 (DE3) cells expressing human ubiquitin (Ub) were grown in LB + Ampicillin (100 mg/L) at 37 °C to OD$_{600}$ = 0.7 and then induced with 0.5mM IPTG. After induction cells were incubated at 30°C overnight and harvested, and the cell pellets were stored at -80°C.

Full scale purification requires 4 L worth of cells to begin. Cells were thawed using a room-temperature water bath and resuspended in Buffer A (50mM NaAc pH 4.5, 1mM EDTA) and lysed using a homogenizer. The lysate was spun down at 20,000 rpm for 30 min at 4 °C. The supernatant was collected and heated in a water bath set to 75°C for 30 minutes. The sample was shaken every 3 minutes during this period to heat evenly. The heat-treated sample was then centrifuged as above, and the supernatant was collected. The supernatant was collected, and its pH was brought to 7.5 using 1M Tris-HCl pH 8 and loaded on a 10mL Q column equilibrated in 5% buffer A and the flowthrough collected. The Q column was cleared using 2M NaCl and the collected flowthrough was reloaded onto the column. This second flowthrough was collected and concentrated using a pressure-powered concentrator to 10mL and then concentrated to 19mg/mL using a centrifugal concentrator. The protein was aliquoted and flash-frozen before storage at -80 °C.

2.4.6 hUba1 Purification

*E. coli* BL21 (DE3) cells expressing His-hUba1 were grown in LB + Ampicillin (100 mg/L) at 37 °C to OD$_{600}$ = 0.7, and then induced with 0.25 mM IPTG. Cells were then grown overnight at 16 °C and then harvested. Cell pellets were stored at -80 °C.
Cells from 2 L of culture were thawed and resuspended in His-binding buffer with benzamidine and PMSF. Cells were lysed on homogenizer and spun down at 20,000 rpm. The supernatant was loaded on a 10 mL Roche His cOmplete Ni column and washed with 200 mL buffer and 100 mL buffer with 10 mM imidazole. The protein was eluted into 16x1.5 mL fractions with His-binding buffer + 500 mM imidazole. The fractions containing protein were pooled and diluted 10x in Q buffer A to 50 mM NaCl and loaded onto a 10 mL Q column. The Q column was eluted on an AKTA Purifier using a gradient of 5%-50% over 10 CV and collected in 1.5 mL fractions based on UV absorbance. Peak fractions were pooled and concentrated to 3.8 mg/mL. Aliquoted protein and flash-froze before storing at -80 °C.

2.4.7 hUbC S22R Purification

*E. coli* BL21 (DE3-pRARE) cells expressing His-hUba1 S22R were grown in LB + Kanamycin (50 mg/L) + Chloromphenicol (34 mg/L) at 37 °C to OD600 = 0.7 and then induced with 0.5 mM IPTG. Cells were grown at 16°C overnight and harvested. Cell pellets were stored at -80°C.

Cells from 2 L of culture were thawed and resuspended in His-binding buffer with benzamidine and PMSF and then lysed on a homogenizer. The lysed cells were centrifuged at 20,000 rpm for 30 minutes and the supernatant was then loaded onto a 10 mL Roche His cOmplete Ni column. The column was then washed with 100 mL His-binding buffer, followed by 50 mL of His-binding buffer with 10 mM imidazole. The protein was eluted into 16x1.5 mL fractions with His-binding buffer with 300 mM imidazole. Fractions with protein were pooled and digested with 1 mg TEV protease for every 20 mg protein while dialyzing against 25 mM Tris pH 7.5, 200 mM NaCl and 5% glycerol overnight. The next morning the digested protein was spiked to 500 mM NaCl and 10 mM MgCl2 before being loaded on a 10 mL Ni column. The flowthrough was collected and concentrated to 17.2 mg/mL, flash frozen, and stored at -80 °C.
2.5 Ubiquitination of PCNA

2.5.1 Ubiquitination Reaction Conditions

Reliable, site-specific monoubiquitination of PCNA at K164 was necessary for the study. The PCNA ubiquitination reaction was performed under the following conditions in aqueous buffer, modified from a previously published condition (Hibbert and Sixma, 2012). A typical ubiquitination reaction has a PCNA concentration of ~0.5 mg/mL. As a matter of good practice, protein concentrations were re-evaluated after new stocks of Uba1, UbC, and/or ubiquitin were prepared.

**Table 2.6** Ubiquitination reaction conditions.

<table>
<thead>
<tr>
<th>Component</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMT Buffer 10x pH 9</td>
<td>1x</td>
</tr>
<tr>
<td>NaCl</td>
<td>25 mM</td>
</tr>
<tr>
<td>MgCl₂</td>
<td>3 mM</td>
</tr>
<tr>
<td>TCEP</td>
<td>0.5 mM</td>
</tr>
<tr>
<td>ATP pH 7</td>
<td>3 mM</td>
</tr>
<tr>
<td>PCNA</td>
<td>16 μM</td>
</tr>
<tr>
<td>Ubiquitin</td>
<td>22 μM</td>
</tr>
<tr>
<td>Uba1</td>
<td>40 nM</td>
</tr>
<tr>
<td>UbC S22R</td>
<td>16 μM</td>
</tr>
</tbody>
</table>

Note: MMT 10x buffer is a broad-range buffer mixture that is 200 mM malic acid, 400 mM MES, and 400 mM Tris.

The final reaction buffer volume was calculated from the total mass of PCNA to be ubiquitinated. All non-protein components were mixed first at room temperature, and then each protein was added individually with mixings in between. The final component added was hUbiquitin. The reaction mixture was incubated in a water bath at 37 °C for 80 minutes, with gentle mixing performed every 20 minutes. At the end of the reaction time, the reaction mixture was loaded on an appropriately sized HiTrap Q column pre-equilibrated with 15% Buffer B and washed with 2 CV of the equilibration buffer. The column was then
eluted on an AKTA Purifier using a gradient of 15%-70% over 4 CV and collected in 1.5 mL fractions based on UV absorbance.

2.5.2 Confirmation of ubiquitination specificity by Mass Spectrometry

Pol1 specifically binds to PCNA ubiquitinated at lysine-164 (K164), so it was important to confirm that the ubiquitination reaction used was only ubiquitinating at this specific residue. Mass spectrometry was used to determine the ubiquitination site(s) present (Parker et al., 2009).

Freshly ubiquitinated Ub-PCNA from a small-scale ubiquitination was run on an SDS-PAGE gel. The ubiquitination reaction was run for two hours rather than the usual 80 minutes to improve the chances of multiple ubiquitinations. A gel was run using multiple concentrations of the same sample across all the lanes to provide options for resolution vs. sample size. Spots were cut from the gel using a robotic sampler and submitted for trypsin digestion. The peptide products of the digest were submitted for MALDI-TOF to measure the masses of the fragments.

**Figure 2.2** Workflow for ubiquitination site determination by mass spectrometry. Ubiquitinated protein (a) is digested in-gel to peptides (b), including a small GG isopeptide scar on the ubiquitination site(s). The masses of the peptides are determined by MALDI-TOF mass spectrometry (c).

Observed peptides were compared to a generated library of possible tryptic digestion products (including possible missed cutting sites, oxidation products, and non-
specific modifications) from Ub-PCNA to identify fragments containing a mass shift of +129 Da from the normal mass, which corresponds to the GG scar from the ubiquitin-lysine isopeptide bond. Fragments marked with this scar are identified as ubiquitination sites.

2.5.3 Confirmation of ubiquitination specificity using mutant PCNA

The specificity of the ubiquitination reaction was also tested by performing the normal ubiquitination reaction (see above) using K/R mutants of PCNA. Arginine is a positively charged amino acid with a long side chain like lysine, but it cannot be ubiquitinated like lysine (Hibbert and Sixma, 2012). Such mutations, properly placed, should abolish ubiquitination without affecting other properties of the protein.

Three PCNA mutants were generated by point mutation and purified using the untagged PCNA purification procedure: K164R, K168R, and K164/168R. K164 is the target site, and K168 was believed to be the most probable off-target site based on its proximity to the target site. A ubiquitination reaction was performed on each mutant and a wild-type control in parallel. The start and end points of the reaction were run on an SDS-PAGE gel to determine which of the mutants were able to be ubiquitinated under these conditions.
2.6 Crystallization Trials

PIP-UBM1-PCNA fusion proteins were screened for possible crystallization conditions using the following kits:

Table 2.7 Crystal screening kits used.

<table>
<thead>
<tr>
<th>Crystal Kit</th>
<th># of Wells</th>
<th>Manufacturer</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEG Smear</td>
<td>96</td>
<td>Mixed in house, based on (Chaikud et al., 2015)</td>
<td>PEG mixtures, salts</td>
</tr>
<tr>
<td>Wizard 1-4</td>
<td>192</td>
<td>Molecular Dimensions</td>
<td>Broad Screen</td>
</tr>
<tr>
<td>Natrix 1+2</td>
<td>96</td>
<td>Hampton (Scott et al., 1995)</td>
<td>Salt mixture</td>
</tr>
<tr>
<td>PEG 400 Ion</td>
<td>48</td>
<td>Hampton (Mcperson, 2001)</td>
<td>PEG 400 and salts</td>
</tr>
<tr>
<td>PEG 550 MME Ion</td>
<td>48</td>
<td>Hampton</td>
<td>PEG 550 monomethyl ether and salts</td>
</tr>
<tr>
<td>PEG 1000 Ion</td>
<td>48</td>
<td>Hampton</td>
<td>PEG 1000 and salts</td>
</tr>
<tr>
<td>PEG 1500 Ion</td>
<td>48</td>
<td>Hampton</td>
<td>PEG 1500 and salts</td>
</tr>
<tr>
<td>PEG 2000 Ion</td>
<td>48</td>
<td>Hampton</td>
<td>PEG 2000 and salts</td>
</tr>
<tr>
<td>PEG 2000 MME Ion</td>
<td>48</td>
<td>Hampton</td>
<td>PEG 2000 monomethyl ether and salts</td>
</tr>
<tr>
<td>PEG 4000 Ion</td>
<td>48</td>
<td>Hampton</td>
<td>PEG 4000 and salts</td>
</tr>
<tr>
<td>PEG 5000 Ion</td>
<td>48</td>
<td>Hampton</td>
<td>PEG 5000 and salts</td>
</tr>
<tr>
<td>PEG 6000 Ion</td>
<td>48</td>
<td>Hampton</td>
<td>PEG 6000 and salts</td>
</tr>
<tr>
<td>Morpheus</td>
<td>96</td>
<td>Molecular Dimensions (Gorrec, 2009)</td>
<td>Broad, includes ligands, cryoprotective</td>
</tr>
<tr>
<td>Index</td>
<td>96</td>
<td>Hampton (D’Arcy et al., 2003)</td>
<td>Broad screen, diagnostic</td>
</tr>
<tr>
<td>JCSG+</td>
<td>96</td>
<td>Molecular Dimensions (Newman et al., 2005)</td>
<td>Broad</td>
</tr>
<tr>
<td>MIDAS</td>
<td>96</td>
<td>Molecular Dimensions</td>
<td>Narrow salt/pH, unusual polymeric precipitants</td>
</tr>
<tr>
<td>Protein Complex</td>
<td>96</td>
<td>Qiagen</td>
<td>PEG and salt screen</td>
</tr>
<tr>
<td>PACT Premier</td>
<td>96</td>
<td>Molecular Dimensions (Newman et al., 2005)</td>
<td>Screens pH, cations, and anions</td>
</tr>
<tr>
<td>Crystal Screen 1+2</td>
<td>96</td>
<td>Hampton (Jancarik and Kim, 1991)</td>
<td>Broad</td>
</tr>
</tbody>
</table>
Fusion proteins were screened after concentrating ion exchange fractions, either from purification or after ion exchange refresh from storage at -80 ºC. Protein was concentrated to 10-15 mg/mL and diluted as needed before setting up trays. Trays were placed in incubators set to 18 ºC and observed over a few months for crystals using a microscope.

Screening proceeded using 1+1 µL drops using protein concentrations of 5 or 10 mg/mL. Two basic screening setups were used. First; hanging drop vapour diffusion experiments were set up in 24-well plates using a precipitant solution mixed in the drop and a well condition of 0.5 mL of 1M (NH₄)₂SO₄ or 1.5M NaCl. Second; sitting drop experiments were performed using a well condition of 80 µL of the precipitating condition.

2.7 Size-Exclusion Chromatography – Multi-Angle Light Scattering (SEC-MALS)

2.7.1 Sample Preparation for SEC-MALS

Ub-PCNA and PIP-UBM1 were taken out of -80 ºC storage and thawed. Both proteins were refreshed by injecting on S200 equilibrated in sizing buffer + 0.5 mM DTT. Peak fractions of each were pooled to create high quality samples of sufficient concentration directly from the column. Mixtures were prepared a minimum of 30 minutes before injection to ensure equilibration of the mixture.

2.7.2 Experimental Protocol

The SEC-MALS system was equilibrated in sizing buffer + 0.5 mM DTT (matching the buffer used to prepare samples). Once the sizing column was fully equilibrated, the flow into the MALS reference cell was stopped. 125 µL of sample was loaded into the 100 µL loop to ensure that it was fully loaded. Injection was programmed to occur after 10 mL of buffer had flowed through the column since starting the program. MALS data were collected and processed in ASTRA (Wyatt Technologies).

2.8 Analytical Ultracentrifugation

2.8.1 Sample Preparation for Analytical Ultracentrifugation

Frozen samples of proteins to be run PIP-UBM1(-UBM2) and Ub-PCNA were thawed and individually injected into a GE S200 size exclusion column pre-equilibrated in sizing buffer with 0.5 mM DTT. Peak fractions were collected and pooled in order to
produce a sample of sufficient concentration directly from the sizing column. An additional aliquot of the sizing buffer was kept on ice for use in the reference cell.

2.8.2 Analytical Ultracentrifugation Sedimentation Velocity Experiments

Sedimentation velocity experiments were performed to determine the stoichiometry of the complex between polI and Ub-PCNA by measuring the molecular weight of species in solution. UV absorbance optics were used to track sedimentation velocity.

Stored samples of Ub-PCNA and of PIP-UBM1 (or PIP-UBM1-UBM2) were thawed and refreshed (as noted above) by size-exclusion chromatography using an GE Superdex S200 column pre-equilibrated in Sizing Buffer with 0.5 mM DTT, which also served to match the sample buffers. A 10 mL sample of the Sizing Buffer was kept separate for sample dilutions and reference cells. 0.5 mL fractions were collected based on UV absorbance. Peak fractions of each protein were pooled for a working concentration (Roughly 5 mg/mL each for PIP-UBM1 experiments, and roughly 2.5 mg/mL each for PIP-UBM1-UBM2 experiments).

Experiments using PIP-UBM1 were performed using a wavelength of 293 nm to track Ub-PCNA only while leaving PIPUBM1 virtually invisible. Trimeric Ub-PCNA was kept to a constant concentration 18 μM while PIP-UBM1 ratios were varied between cells. Each individual experimental run included a Ub-PCNA standard cell to ensure reliability since the expected mass change is so small.

The sedimentation velocity experiments were run at 20 °C over a period of 7.5 hours, at a radially mean force of 89,180 g. Scans of absorbance vs radial length were conducted at 10-minute intervals for a total of 45 scans. Cells depleted over the course of the experiment.

Experiments using PIP-UBM1-UBM2 were performed using a wavelength of 280 nm to match the standard protein UV absorbance peak. The aromatic amino acid content, and thus absorbance spectra, of Ub-PCNA and PIP-UBM1-UBM2 are similar, thus the strategy employed in PIP-UBM1 experiments could not be applied here. Ub-PCNA trimer concentration was kept constant in each cell, but at 4.5 μM to keep the absorbance low enough to also track PIP-UBM1-UBM2. Ub-PCNA-only absorbance under these
conditions was 0.5, which left room for higher ratios of PIP-UBM1-UBM2. The binding partner ratio was varied to determine the possible complex stoichiometries.

Sedimentation velocity data were analyzed in the SEDFIT program (Schuck, 2000) to produce a continuous c(s) distribution, with peaks corresponding to the observed sedimentation coefficient ($s_w$) of the species in solution. This sedimentation coefficient can be used to estimate the molecular weight of the species in solution by the Svedberg relationship:

\[
\frac{M}{N_A} = \frac{sf}{1 - \bar{v}\rho}
\]

**Equation 1.** The Svedberg relationship, where $M$ is the molecular weight, $N_A$ is Avogadro’s number, $s$ is the sedimentation coefficient, $f$ is the frictional coefficient, $\bar{v}$ is the partial specific volume, and $\rho$ is the solvent density. The solvent density and partial specific volume were calculated using SEDNTRP, a SEDFIT-suite program.

Analysis proceeded using a single frictional coefficient for all runs to try to avoid biasing results. This coefficient was initially determined by fitting the Ub-PCNA standard estimate to its known molecular weight and applying the same coefficient to each of the remaining cells. The frictional coefficient is the relationship between the apparent size of the molecule and its resistance to sedimentation, i.e. between two molecules of identical molecular weight, the molecule with the smaller frictional ratio will sediment faster. This value may be floated during analysis but changing its value will cause species with identical sedimentation coefficient to have different apparent molecular weights. The frictional ratio is determined by the shape and flexibility of the species. Since the Ub-PCNA trimer forms the bulk of the PIPUMB1-Ub-PCNA complex, the frictional ratio of each complex should differ little from that of Ub-PCNA on its own.

2.8.3 Analytical Ultracentrifugation Sedimentation Equilibrium Experiments

Sedimentation equilibrium experiments were performed to determine the dissociation constants of the poli-Ub-PCNA complexes. Sample preparation was performed in the same manner as the sedimentation velocity experiments. UV absorbance
optics were again used to track sedimentation. Experiments were performed using 2-cell centrepieces.

Experiments using PIP-UBM1 complexes were again performed using a wavelength of 293 nm to simplify the system and track Ub-PCNA alone. Ub-PCNA trimer concentration was held constant at 8.9 μM across all cells and three complexes were run: Ub-PCNA alone, 1:2 trimer:PIP and 1:4 trimer:PIP. Samples were mixed in a 175 μL volume, from which 150 μL were loaded into the sample cell. 175 μL buffer was loaded into the reference cell.

Equilibrium experiments with PIP-UBM1 and Ub-PCNA were performed at 20 °C for a total of 87 hours. The PIP-UBM1 experiment was performed with the following speeds and timepoints. The remaining mixed sample was mixed with SDS sample and frozen so that an SDS-PAGE gel could be run to determine whether any degradation occurred.

Experiments using PIP-UBM1-UBM2 were performed using a wavelength of 280 nm to maximize sensitivity. The absorbance spectra of PIP-UBM1-UBM2 and Ub-PCNA are similar and there is not a convenient wavelength to observe only Ub-PCNA. Ub-PCNA trimer concentration was held constant at 2.7 μM and three complexes were run: 1:0.8, 1:1, and 1:2 trimer:PIP.

Equilibrium experiments with PIP-UBM1-UBM2 and Ub-PCNA were performed at 20 °C for a total of 68 hours. The PIP-UBM1-UBM2 complex equilibrium experiment was performed with the following speeds and timepoints.

Based on the results of the PIP-UBM1 complex experiments, the speeds and number of scans were reduced to improve the quality of the data. The experimental setup used does not accurately measure the concentration at A > 1.1, thus at higher speeds larger segments of each scan would be excluded from analysis. Since the most important part of a scan is the most concentrated section, excluding this end may lead to unreliable data.

Data analysis of all SE experiments were performed using SEDPHAT’s multispeed equilibrium processing (Vistica et al., 2004). The analysis used a 3-site symmetric binding model and treated the entire Ub-PCNA trimer as the central binding partner with
three binding sites. This model was chosen because the observed particle is the Ub-PCNA trimer in addition to its bound partners. Unfortunately, this model is much more complicated than a standard heteroassociation A+B-type model.

\[
K_D^I = \frac{(A + AB_{II} + AB_{III} + AB_{II}B_{III}) \cdot B}{(AB_I + AB_I B_{II} + AB_{II} B_{III} + AB_{II} B_{II} B_{III})}
\]

\[
K_D^{II} = \frac{(A + AB_{I} + AB_{III} + AB_{II} B_{III}) \cdot B}{(AB_{II} + AB_{II} B_{II} + AB_{II} B_{III} + AB_{II} B_{II} B_{III})}
\]

\[
K_D^{III} = \frac{(A + AB_{II} + AB_{III} + AB_{II} B_{II}) \cdot B}{(AB_{III} + AB_{II} B_{III} + AB_{II} B_{III} + AB_{II} B_{II} B_{III})}
\]

**Equation 2.** \(K_D\) equations in a three-site binding model. Each binding site is considered separately (\(AB_I\), \(AB_{II}\), and \(AB_{III}\)) and each individual \(K_D\) accounts for all possible populations in equilibrium (9 in total). This model keeps all binding sites in saturation with one another, rather than saturating each binding site stepwise.

### 2.9 Small-Angle X-ray Scattering Experiments (SAXS)

#### 2.9.1 Size-Exclusion Chromatography – Small Angle X-ray Scattering

SEC-SAXS experiments (Malaby et al., 2015) were performed at the Advanced Photon Source (APS) beamline 18D. Samples of Ub-PCNA and PIP-UBM1 were purified and concentrated in lab and shipped to APS. 250 µL samples of complex were prepared 30 minutes prior to injection to allow for equilibration and then injected into a fast-flow S200 gel filtration column pre-equilibrated in SAXS buffer at room temperature inside the beamline hutch. The beam wavelength was 1.54 Å. X-ray scattering images were collected before, across, and after the peak through an inline capillary. This experimental setup uses size exclusion chromatography to ensure that the sample is monodisperse (assuming good separation on the column) and the inline capillary combined with low exposure per frame reduces the effect of radiation damage on the sample collected. The capillary system is thoroughly cleaned using hydrophobic liquid to prevent buildup of damaged proteins and thus improve the quality of the data collected.

Basic SAXS processing was performed at the beamline using the RAW software suite. Molecular weight (MW) and radius of gyration (\(R_g\)) measurements are easily
processed and were used to assess the data quality. SAXS envelope modelling was performed using DAMMIN (Svergun, 1999), DAMMIF (Franke and Svergun, 2009), and GASBOR (Svergun et al., 2001). Visualization of the envelope was performed in PyMOL.
3 Results

3.1 Purification of pol and (Ub)-PCNA proteins

3.1.1 PIP-UBM1-PCNA fusion proteins can be separated from contaminants.

Fusion proteins were developed for crystallization experiments. Protein complexes crystallize best when one partner is saturated, but the excess of the other partner required may inhibit crystallization. Fusion proteins create a saturated “complex” without any excess binding partner.

Purifications proceeded according to the procedure outlined in Methods. A gel would be run following the Ni column on the first day to assess quality. Fractions containing few contaminating proteins and reasonable concentration would then be pooled. The concentration of protein was then measured (typical total protein between 30-60 mg per litre of culture used) and TEV protease was added.
The next and final check occurs after the final ion exchange column.

**Figure 3.1** Fusion Protein Production  
**a)** Expression and solubility test for StrpII-PIP-UBM1-PCNA protein, expressed with TEV-cleavable His-Mocr solubility and affinity tag.  
1) uninduced cell lysate. 2) induced cell lysate after 4h growth at 37 °C. 3) soluble fraction of 37 °C expression. 4) induced cell lysate after overnight growth at 16 °C. 5) soluble fraction of 16 °C expression. The expression and solubility are typical of the fusion proteins, save the little finger fusion proteins, which tend to be less soluble even after overnight low-temperature growth.  
**b)** TEV digestion and final ion exchange fractions for StrpII-PIP-36del-UBM1-PCNA fusion protein, an example of its type chosen because the contaminant band is easily seen. A low molecular weight (Mw) array is used as a reference standard with bands at 97.4, 66.2, 45, 31, 21.5, and 14.4 kDa. After successful TEV protease (31 kDa) digestion of the tagged protein at 66 kDa (lanes 1-2), the digestion product at 45 kDa was loaded on an anion exchange column (flowthrough in lane 3) and eluted (lanes 4-14). Early peak fractions (5-10) can be pooled to exclude the major contaminant at 40kDa.  

PCNA alone tends to run higher than expected from its molecular weight (see PCNA purification below), so it is not surprising that the fusion proteins do too. In this case, the 40 kDa StrpII-PIP-36del-UBM1-PCNA fusion appears to run the same as the 45 kDa marker. The contaminant just below the main protein band is present in all fusion proteins and can be removed by only pooling early peak fractions with minimal contamination. It is not a degradation product; the pooled early fractions can be kept at
room temperature for two weeks without any sign of degradation. PIP-UBM1-PCNA proteins can be concentrated to approximately 15mg/mL safely.

The fusion proteins were not able to be ubiquitinated at K164 alone. Since site-specific ubiquitination could not be achieved, the UBM had no binding partner and could not be anchored to a specific region. This may have been a factor in the failure of the crystallization trials.

3.1.2 Short polt C-terminal proteins are stable, longer forms are more delicate

Polt C-terminal proteins were the model proteins of choice for studying all aspects of the interaction between polt and Ub-PCNA. PIP-UBM1 (aa415-530) is the workhorse protein; it is stable, easy to purify, contains both major interacting motifs, and has no internal deletions. It has been used for every experiment during this project. PIP-UBM1-UBM2 (aa415-715) contains the entire C-terminal domain. It has been used for analytical ultracentrifugation experiments to determine whether the full-length protein behaves in the same manner as the truncated protein. PIP-UBM1-UBM2 more prone to aggregation than the PIP-UBM1 protein and cannot be concentrated to the same concentration (m/v). Its thermal stability is also poorer, but it does not have any significant effect at either 4 °C or 20 °C.

Purification of polt C-terminal proteins is straightforward and requires only two columns. For the first day, the Ni column only requires a spot test check to pool fractions and a Bradford assay to measure protein for TEV digestion (typical values are roughly 50-100 mg per litre of cell culture).
**Figure 3.2** C-terminal protein production

**a**) Expression and solubility test for PIP-UBM1 protein, expressed with TEV-cleavable His-Mocr solubility and affinity tag. 1) uninduced cell lysate. 2) induced cell lysate after 4h growth at 37 °C. 3) soluble fraction of 37 °C expression.  

**b**) Example SDS-PAGE gel outlining purification of PIP-UBM1 C-terminal fragment as an example of this class of protein. After enrichment and purification of the target protein by Ni-affinity chromatography (lane 5), the final anion exchange step eliminates the bulk of the contaminants. The remaining weak contaminants are eliminated by size-exclusion chromatography.

PIP-UBM1 proteins may be safely concentrated to 20 mg/mL but this is typically unnecessary; 10-15 mg/mL is a typical target concentration. The low molecular weight of the protein means that even this lower concentration has a high molarity. PIP-UBM1-UBM2 proteins may be safely concentrated to 10 mg/mL.

### 3.1.3 PCNA purification is highly effective

PCNA is the basis for the other major component of the complex being studied. It or Ub-PCNA has been used in all experiments in this investigation. Purification followed the procedure outlined in Methods. Each PCNA monomer is tagged with a TEV-cleavable 8xHis tag. The protein forms a strong trimer, so the protein being purified may be treated as a single unit with 3 such tags. The protein strongly binds the first Ni column and may
be washed with high imidazole before elution and is pure after this first stage. Subsequent purification steps are largely to separate the cleaved tag and the TEV protease.

| a | kDa | Mw | L | C | FT | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | b | Mw | Load | Elution Fractions |
| 97.4 | 66.2 | 45.0 | 31.0 | 21.5 | 14.4 |

**Figure 3.3** PCNA Production  
**a)** SDS-PAGE gel demonstrating Ni-affinity chromatography of His-PCNA. The 8xHis tag confers powerful Ni affinity to PCNA, which permits stringent washing (lane 1) and elution (lane 2) of pure protein at the first step of the purification. After tag removal, the 2nd Ni column traps nearly all the contaminants, leaving pure PCNA in the second flowthrough fraction (lane 5).  
**b)** SDS-PAGE gel of the final anion exchange step. Purity has not been improved between the load and elution fractions, but the protein is now in a more appropriate buffer without leached Ni and is contained in a smaller volume. The anion exchange chromatogram may also be used as a quality check.

The final ion exchange column removes any nickel or imidazole that may still be in the protein buffer and concentrates the protein before the final centrifugal concentration.

PCNA may be concentrated to 20 mg/mL, but typically 12-15mg/mL was used since it must be diluted anyway during its ubiquitination reaction. Typical yield is 10-15 mg per litre of culture used. The protein was dispensed into 5 mg aliquots for ease of use. Protein not immediately ubiquitinated was flash frozen and stored at -80 °C. Frozen PCNA may be ubiquitinated immediately after thawing; refreshing is achieved during the ubiquitination protocol and the customary size-exclusion before the final use.
3.1.4 Ubiquitin purification is high yield

Ubiquitin has been purified for the purpose of ubiquitinating PCNA. The ubiquitin purification follows an atypical protocol, and at no point does the target protein bind to a chromatography column. Anion exchange columns are used twice, but these only remove contaminants and the ubiquitin remains in the flowthrough fraction. This causes the volume of the buffer containing the protein to increase with each step. To accommodate the large final volume, a pressure-fed concentrator with a 40 mL capacity and a 5000 Da MWCO was used to concentrate the sample. The concentrator also had an internal magnetic stir bar to prevent overconcentration near the membrane. Even with this concentrator the flow rate was roughly 20 mL per hour. Once the total volume reached 10mL, the sample was transferred to a 15mL centrifugal concentrator to finish concentrating. hUb was concentrated to its working concentration of 18.6 mg/mL and had a final yield of 74.3 mg from 4 L of cell culture. 150 µL aliquots were made and flash frozen for later use.

3.1.5 Ubiquitination reaction is high yield, scalable

PCNA is ubiquitinated in 50 mL conical tubes, in the reaction buffer outlined above. The reaction buffer is then loaded on a Q column as outlined in Methods to isolate the Ub-PCNA. The excess ubiquitin and enzymes do not bind the column at the loading buffer’s ionic strength, and the Ub-PCNA that elutes from the column is pure.
Figure 3.4 Ubiquitination Products SDS-PAGE gel demonstrating the final state of Ub-PCNA after anion exchange chromatography to remove the enzymes involved in the reaction. Each numbered lane represents the sequential 1.4 mL peak fraction collected from the Q column. The faint band below the strong Ub-PCNA band is the remaining PCNA. Ub-PCNA is effectively pure at this stage.

Ub-PCNA can be concentrated to 20 mg/mL but is usually only concentrated to 15mg/mL. The same mass of Ub-PCNA can be recovered as the mass of PCNA used in the reaction (~90% yield).

3.2 Difficulties in crystallizing PIP-UBM1/Ub-PCNA complex

PIP-UBM1-PCNA fusion proteins were screened for crystals using the screens and experimental setups outlined in Materials and Methods. No protein crystals were observed nor any promising phase separations. Fusion proteins appeared to be more susceptible to aggregation compared to PIP-UBM1 + Ub-PCNA complexes of similar concentration, which suggested that the fusion proteins were less stable than the natural complex. Certain conditions (namely divalent cations, acidic pH, PEG range 4000-6000) tended to cause immediate heavy aggregations; screens which were biased toward these conditions were screened at lower concentrations initially then generally excluded from second-round screenings.
Natural complex (PIP-UBM1 and Ub-PCNA) screening had been attempted by the laboratory before the beginning of this investigation, without success. After the completion of fusion protein trials, the focus returned to the natural complex. Before returning to crystallization trials, the natural complex was studied to challenge the assumptions that had been made about it, particularly the specificity of the ubiquitination reaction and the excess of PIP-UBM1 required to saturate the complex.

3.3 The ubiquitination reaction is site-specific and monoubiquitination-specific

A possible reason for the failure of a crystallization trial would be a heterogeneity in the sample. In the case of the PIP-UBM1/Ub-PCNA complex, the most probable heterogeneity would be the presence of multiple ubiquitination sites. The SDS-PAGE gel of the ubiquitination product showed a single, strong band. Time-course experiments had never shown an intermediate product that would suggest that the product was polyubiquitinated, and a ubiquitination reaction that was oversupplied with ubiquitin and reacted for twice the normal time did not produce any bands higher than the normal product. These results demonstrated that the product of the ubiquitination reaction is exclusively mono-ubiquitinated PCNA. The literature on the reaction used (Hibbert and Sixma, 2012) had demonstrated that K164 was necessary for the reaction, but it did not conclusively prove that there were not multiple mutually-exclusive possible mono-ubiquitination sites.

3.3.1 K164R mutant PCNA abolishes ubiquitination

The specificity test using K/R point mutations that was performed by Hibbert and Sixma was extended to also investigate K168, which is the most likely off-target ubiquitination site (Hibbert and Sixma, 2012). PCNA of wild-type, K164R, K168R, and the K164/168R double mutant were purified for the experiment and then ubiquitination reactions for each were run simultaneously. The SDS-PAGE gel of the ubiquitination reactions is displayed below.
Figure 3.5 Ubiquitination Assay by Point Mutation SDS-PAGE gel of ubiquitination reaction for wild-type (WT) PCNA and three K/R PCNA mutants. Wild type PCNA is ubiquitinated as a control and the reaction ends in a single band. The K164R mutation abolishes the ubiquitination, leaving the PCNA band unchanged and no ubiquitinated products. K168R has little effect, with the majority of the PCNA being ubiquitinated in the same manner as wild type, suggesting that K168 is not a target of this reaction. The K164/168R mutation also abolishes ubiquitination, as expected after the results of the K164R mutation. These results strongly suggest that the reaction used is site-specific to K164.

The results of these ubiquitination reactions show that K168R can be ubiquitinated like wild-type PCNA, while K164R and K164/168R are unable to be ubiquitinated under the reaction conditions used. This suggests that K164 is the sole lysine ubiquitinated by the reaction. K168 may have a role in the reaction, since the K168R reaction did not go to completion, but it does not appear to be a ubiquitination target. These results suggest that the ubiquitination reaction is site-specific and that the products are homogenous. However, it was prudent to continue to confirm the homogeneity of the reaction products.

3.3.2 Mass spectrometry experiments detect ubiquitination at K164 only

Identification of other possible ubiquitination sites was performed using in-gel digestion of Ub-PCNA followed by mass spectrometry (Parker et al., 2009). After
digestion, peptide fragments of Ub-PCNA were identified by MALDI-TOF mass spectrometry.

The following peptide fragments of PCNA were detected by mass spectrometry. The underlined peptide is the sole fragment found with the GG scar from ubiquitination.

**Table 3.1** Tryptic digest products of Ub-PCNA identified by MALDI-TOF.

<table>
<thead>
<tr>
<th>Mass (Da)</th>
<th>Peptide from Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>653.3</td>
<td>MFEAR</td>
</tr>
<tr>
<td>754.4</td>
<td>YYLAPK</td>
</tr>
<tr>
<td>778.3</td>
<td>IEDEEGS</td>
</tr>
<tr>
<td>857.5</td>
<td>LVQGSILK</td>
</tr>
<tr>
<td>871.4</td>
<td>VSDYEMK</td>
</tr>
<tr>
<td>884.5</td>
<td>IADMGHLK</td>
</tr>
<tr>
<td>894.4</td>
<td>MPSGEFAR</td>
</tr>
<tr>
<td>932.5</td>
<td>YLNFFTK</td>
</tr>
<tr>
<td>974.4</td>
<td>SEGFDTYR</td>
</tr>
<tr>
<td>1261.6</td>
<td>CAGNEDIITLR (with carbamidomethyl adduct)</td>
</tr>
<tr>
<td>1293.6</td>
<td>FSASGELGNGNIK</td>
</tr>
<tr>
<td>1381.7</td>
<td>NLAMGVNLTSMSK (with oxidation)</td>
</tr>
<tr>
<td>1615.9</td>
<td>ILKCAGNEDIITLR (with carbamidomethyl adduct)</td>
</tr>
<tr>
<td>1984.0</td>
<td>DLSHIGDAVVISCAKDGVK (with GG scar)</td>
</tr>
<tr>
<td>2075.0</td>
<td>AEDNADTLALVFEAPNQEKK (with carbamidomethyl adduct)</td>
</tr>
<tr>
<td>2424.3</td>
<td>ATPLSSTVTLSMSADVPLVVYEYK (with oxidation)</td>
</tr>
<tr>
<td>2665.3</td>
<td>DGVKFSASGELGNGNIKLQSQTSNVDK</td>
</tr>
<tr>
<td>3307.6</td>
<td>LSQTSNVDKEEEAVTIEMNDEPVQLTFALR (with oxidation)</td>
</tr>
</tbody>
</table>
Figure 3.6 Coverage map of PCNA by MALDI-TOF fragments. Highlighted regions are those covered by the fragments detected by mass spectrometry. The red highlighted K residues are the three lysine residues not covered by the peptides.

These data show a single peptide containing a GG shift. This peptide contains two lysine residues: K164 and K168. It is unlikely that the K168 residue is the site of modification; the GG scar is expected to cause a missed cleavage at scarred residue and the cut site of the scarred fragment is at K168. It is also unlikely that there are ubiquitinations at the sites not detected. The sample was taken from a gel band containing solely monoubiquitinated PCNA, and there is no evidence of any unmodified K164. Furthermore, two of the unmapped lysines are on the internal pore of the PCNA trimer and are difficult to access for ubiquitination.

The results of this experiment confirm that the ubiquitination reaction created mono-Ub-PCNA site-specific to K164 within the detection range of the mass spectrometry experiment. Heterogeneity of the Ub-PCNA sample is not a probable cause of the failure of the crystallization trials.

3.4 SEC-MALS experiments demonstrate stability of 1:1 complex

Once crystallization trials were abandoned, other avenues of studying the structure of the complex were considered. SEC-MALS was considered as a starting point to characterize the complex ratio and the stability of the complex over time. A Ub-PCNA
trimer has binding sites to completely accommodate three PIP-UBM1 binding sites. Prior ITC studies have demonstrated that Ub-PCNA has good affinity for PIP-UBM1, with micromolar $K_d$. Size-exclusion chromatography experiments were able to separate the complex from unbound Ub-PCNA and PIP-UBM1. However, it has not been possible to distinguish between different ratios using the sizing column based on elution position, and the PIP-UBM1 protein has poor UV absorption which make it difficult to estimate the excess separated by the column. The combination of these factors may have contributed to an erroneous assumption that the PIP-UBM1/Ub-PCNA complex was reasonably saturated with a small excess of PIP-UBM1.

**Table 3.2** Table of theoretical molecular weights of PIP-UBM1 and Ub-PCNA complexes.

<table>
<thead>
<tr>
<th>Species</th>
<th>PIP-UBM1 (trimer)</th>
<th>1:1 Complex</th>
<th>1:2 Complex</th>
<th>1:3 Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Theoretical MW (kDa)</strong></td>
<td>13.3</td>
<td>111.9</td>
<td>125.2</td>
<td>138.5</td>
</tr>
</tbody>
</table>

SEC-MALS experiments were performed to determine whether a complete PIP-UBM1 complex could be formed and determine whether the complex is stable enough to survive size-exclusion chromatography.
Figure 3.7 Multi-angle light scattering refractive index vs elution volume and chart showing measured molecular weight. The red peak has Ub-PCNA Alone, the yellow peak has 9:2 PIPUBM1:Ub-PCNA (trimer), and blue has: 9:1 PIPUBM1:Ub-PCNA (trimer). The 9:2 and 9:1 peaks have identical elution volumes and estimated molecular weight.

Good single peaks were observed for all complexes, well separated from the excess PIP-UBM1 peak further down. Complexes elute earlier than the Ub-PCNA alone; the shift is small but definite. The two complex ratios do not have a notable difference in the elution volume.

The mass estimation and complex ratio results were unexpected. Even at a Ub-PCNA concentration of 7.7 µM and a 2:9 ratio of Ub-PCNA:PIP-UBM1, the largest stable complex that can be measured from the column is only a 1:1 Ub-PCNA(trimer):PIP.
is no change in the observed weight even after doubling the ratio of ligand to trimer. Given that ITC experiments had determined the $K_d$ of the interaction to be on the order of 5 µM, this was a great surprise and suggested that the assumed complex ratio may have been incorrect.

3.5 Complex investigation using analytical ultracentrifugation

The complex ratio observed in the SEC-MALS experiments did not match the prior expectations, so the next step was to characterize the complex in solution by a different method. Sedimentation velocity experiments were performed as outlined in Materials and Methods to estimate the molecular weight of the complexes present at varying ratios of PIP-UBM1 and Ub-PCNA. Once the range of possible complexes was determined, sedimentation equilibrium (AUC-SE) experiments were performed to determine the binding constants of the complex.

3.5.1 Sedimentation velocity experiments observe a maximum complex ratio of 1:2

The first set of AUC-SV experiments were performed using PIP-UBM1 and Ub-PCNA. PIP-UBM1 was the protein used to compare the results to those of the initial SEC-MALS experiments that showed unexpected complex behaviour.

Each experimental cell contained 2 mg/mL Ub-PCNA (18 µM trimer). Cells were set up with the following ratios of Ub-PCNA:PIP-UBM1: 1:0, 1:1, 1:3, 1:6, and 1:9. This concentration produced an absorbance value of 1 in the setup used, which gave a reliable signal with a small noise component. At the wavelength chosen (293 nm) PIP-UBM1 is nearly invisible (see Figure 3.8) but Ub-PCNA absorbance is still reasonably strong.
**Figure 3.8** Absorbance spectrum of PIP-UBM1. 280nm is a standard UV wavelength for measuring the concentration of protein. This absorbance is primarily from aromatic residues (Phe, Tyr, Trp). Tryptophan has a larger aromatic system and permits the absorbance of longer wavelengths. PIP-UBM1 lacks tryptophan residues, and the absorbance spectrum above shows weak absorbance in the 290-300 nm region.

The samples were mixed to 400 µL. The experiments were performed using 2-sector cells. 380 µL of sample were loaded into the sample cell and 400 µL of Sizing Buffer was loaded into the corresponding reference cell.
Figure 3.9 AUC-SV of PIP-UBM1/UbPCNA complex a) full range of sedimentation coefficients observed in solution during Ub-PCNA-PIPUBM1 AUC-SV experiments. $c(s)$ curves are normalized to the peak value. No significant impurities or additional species are observed. PIP-UBM1 is virtually invisible at the chosen wavelength by design, save the small peak at 1.5 s during the maximum concentration run (9:1 ratio, red curve). b) section of the data focusing on Ub-PCNA and complexes to better illustrate the shift of sedimentation coefficient as the ratio of Ub-PCNA:PIP-UBM1 increases. The single peak in each run suggests that multiple states are in equilibrium in each case and the interchange is much faster than the timescale of the experiment.

Table 3.3 Sedimentation coefficients and estimated molecular weights for PIP-UBM1/UbPCNA complexes, with inferred complex ratio.

<table>
<thead>
<tr>
<th>Solution Ratio</th>
<th>S\textsubscript{w}</th>
<th>MW (kDa)</th>
<th>Complex Ratio</th>
<th>Theoretical Mw</th>
</tr>
</thead>
<tbody>
<tr>
<td>1:0</td>
<td>5.0</td>
<td>115</td>
<td>1:0</td>
<td>111.9</td>
</tr>
<tr>
<td>1:1</td>
<td>5.4</td>
<td>125</td>
<td>1:1</td>
<td>125.2</td>
</tr>
<tr>
<td>1:3</td>
<td>5.5</td>
<td>130</td>
<td>1:1.5</td>
<td></td>
</tr>
<tr>
<td>1:6</td>
<td>5.6</td>
<td>136</td>
<td>1:2</td>
<td>138.5</td>
</tr>
<tr>
<td>1:9</td>
<td>5.7</td>
<td>137</td>
<td>1:2</td>
<td>138.5</td>
</tr>
</tbody>
</table>

Sedimentation coefficients are listed as $S\textsubscript{w}$ after conversion to the theoretical sedimentation coefficient in water. The probable complex ratio of each sample was determined based on theoretical molecular weights for PIP-UBM1, Ub-PCNA, and their possible complexes. Refer to Table 3.2 for theoretical molecular weights of the component proteins and possible complexes.
These molecular weight estimates suggested that each Ub-PCNA trimer can bind two PIP-UBM1 partners, not the three that had been originally assumed. There was virtually zero shift between the mass estimation of the 1:6 and the 1:9 runs, which suggests that the complex is the upper limit on the complex ratio. The 1:1 peak is broad, which suggests that it is not a homogenous complex but rather a mixture of 1:0, 1:1, and 1:2 complexes, with an average molecular weight corresponding to 1:1. This heterogeneity suggests that the two $K_d$ values are close enough to be in equilibrium with one another.

One incidental finding is the excess of PIP-UBM1 required to saturate the Ub-PCNA ring may have had an impact on crystallization experiments. Most crystallization trials had been performed with a ratio between 1:3 and 1:4.5 (trimer:PIP), but the results of this experiment suggest that the complex is not saturated in this condition. Crystallization experiments are most likely to succeed when performed using a homogeneous protein sample, and these AUC-SV experiments may have provided insight into the failure to crystallize the complex.

The main possible source of error in this experiment is the estimation of the molecular weight. The sedimentation values ($s$) are the direct product of the experiment. The molecular weight is estimated from this based on factors like partial specific density and the estimated shape of the particle. The molecular weight of Ub-PCNA was known and used to estimate the mass of the complexes. However, the shift in $s$ is so small between complexes that the estimated molecular weights may have been incorrect and so the estimated complex ratio would be incorrect as well.

3.5.2 AUC-SE experiments suggest that the first and second additions are of similar affinity and in equilibrium

Having determined the range of possible complexes by AUC-SV, further experiments were needed to measure the dissociation constant. Sedimentation equilibrium experiments were performed to determine the dissociation constants of the PIPUBM1/Ub-PCNA complex. Complex stoichiometry and sedimentation coefficients were determined by sedimentation velocity experiments and used to inform the experimental design of the SE experiments. Once again, a wavelength of 293 nm was used to keep the PIP-UBM1 nearly invisible and only track the position of Ub-PCNA and its complexes. A Ub-PCNA
A concentration of 8.9 µM was chosen to achieve an absorbance of 0.4 AU when the cell is at rest. This is to balance the need for a high enough concentration for a reliable signal with the tendency for the absorbance at the bottom of the cell exceed the detector range. Prior ITC experiments had determined a $K_d$ value near 5 µM, so the final concentration of 8.9 µM is appropriate. Cells were run with ratios of 1:2 and 1:4 Ub-PCNA:PIP-UBM1, which were chosen based on the qualitative assessment of the AUC-SV $c(s)$ distribution. The multi-speed equilibrium experiment was conducted using the following cell conditions and speed settings.

**Table 3.4** Speed/time settings for PIP-UBM1/Ub-PCNA AUC-SE experiment.

<table>
<thead>
<tr>
<th>Scan #</th>
<th>Mean RCF (g)</th>
<th>Hours between Scans</th>
<th>Total Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4,660</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>4,660</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>7,280</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>7,280</td>
<td>4</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>14,270</td>
<td>15</td>
<td>64</td>
</tr>
<tr>
<td>6</td>
<td>14,270</td>
<td>4</td>
<td>68</td>
</tr>
<tr>
<td>7</td>
<td>23,580</td>
<td>15</td>
<td>83</td>
</tr>
<tr>
<td>8</td>
<td>23,580</td>
<td>4</td>
<td>87</td>
</tr>
</tbody>
</table>

When setting a speed, a long period of time was left to allow for equilibration in the cell. A second scan was performed 4 hours later at each speed to confirm that equilibration had completed by overlaying scans. Each cell came to equilibrium successfully, but it was found that at the two highest speeds the proteins had sedimented too much and were not able to provide any useful data. Similar issues were found with the 4:1 cell. For that reason, they were excluded from analysis and the speeds chosen were adjusted for the PIP-UBM1-UBM2 experiments.

Data analysis was performed in SEDPHAT using the 3-site symmetric model, using the Ub-PCNA trimer as component A and PIP-UBM1 as component B. The sedimentation values of Ub-PCNA and the two observed complexes were given to the program, as was an extrapolated value for the hypothetical “saturated” complex.
**Figure 3.10** Sedimentation equilibrium fit for PIP-UBM1/UbPCNA. Upper graph: absorbance vs radial length for 4,660 g (green) and 7,280 g (red). Lower graph: residuals of the fit. Data is cut off before the end of the cell as absorbance exceeds A=1. Overall fit is good, the residuals are not perfectly randomly distributed as might be hoped.

**Table 3.5** Dissociation constants for PIPUBM1/Ub-PCNA complex determined by sedimentation equilibrium.

<table>
<thead>
<tr>
<th>$K_d1$</th>
<th>$K_d2$</th>
<th>$K_d3$</th>
<th>$\chi^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.01 ± 0.36 μM</td>
<td>2.85 ± 1.22 μM</td>
<td>$&gt;&gt;K_d1$</td>
<td>0.312</td>
</tr>
</tbody>
</table>

Dissociation constants were determined with a 95% confidence interval. These values agree with the qualitative observations of the sedimentation velocity experiment. Both experiments suggest that Ub-PCNA can host only two PIP-UBM1 binding partners despite containing 3 sets of binding sites. In this case, the $K_d3$ is so large that the program has effectively not observed any contribution and offered an absurdly large value as a placeholder. The two real $K_d$ estimates are close enough that they are in equilibrium with one another, rather than sequential addition. This agrees with the peak broadening seen in the SV experiments. Finally, the real $K_d$ estimates are both in the magnitude of the prior ITC experiments.
Despite best efforts, it is possible that the shift in the $s$ value is so small and the 3-site binding model is so complex that these values represent an overfitting of the data. Sedimentation values from the AUC-SV experiment were used as a guide for the fit, possibly introducing an improper bias into the analysis.

3.5.3 Sedimentation velocity experiments with PIP-UBM1-UBM2 show similar 1:2 binding limit.

Sedimentation velocity experiments were then run using the larger PIP-UBM1-UBM2 protein. PIP-UBM1 is a convenient protein of the polt C-terminal domain, but it is a small representative part of that larger domain. Since the ratio of complex using the small PIP-UBM1 fragment had been determined, the next step was to determine the complex size limit using the larger PIP-UBM1-UBM2 protein and whether the larger domain would face similar constraints. PIP-UBM1-UBM2 contains the entire suite of interacting motifs in the C-terminal domain; it would be more representative of the biologically relevant complex than the small fragment used in most experiments. Ub-PCNA was run alone as a reference standard at a concentration of 4.5 µM, and mixtures of Ub-PCNA:PIP-UBM1-UBM2 were run at ratios of 1:1 and 1:4.
**Figure 3.11** AUC-SV of PIP-UBM1-UBM2/Ub-PCNA complex. Observed c(s) distribution of PIP-UBM1-UBM2 and Ub-PCNA mixtures, estimated molecular weights, and estimated stoichiometry of the observed complexes. As with the PIP-UBM1 AUC-SV experiment in figure 3.8, single intermediate peaks are observed as a result of rapid interchange at equilibrium.

The molecular weight estimates were not as accurate as the PIP-UBM1 results, but they were enough to interpret the data. The higher molecular weight estimate suggested that the complex was sedimenting faster than its actual molecular weight would suggest. This was unexpected since unstructured proteins are generally assumed to sediment more slowly than a structured globular protein. PIP-UBM1-UBM2 appeared to adopt a maximum 1:2 complex with Ub-PCNA. If the first addition binds ubiquitin moieties with both UBMs, the second addition can only bind with one of its two UBMs, which would make the second addition less stable. The 1:2 peak has not sharpened to match the Ub-PCNA, suggesting that the complex has not fully saturated and may explain why the
molecular weight estimate is below the theoretical 1:2 weight. This also suggests that the second addition may be weaker than the second addition in PIP-UBM1.

3.5.4 AUC-SE experiments with PIP-UBM1-UBM2 show similar binding pattern to PIP-UBM1

Sedimentation equilibrium experiments were performed to determine the dissociation constants of the PIPUBM2-Ub-PCNA complex, again using the multi-speed equilibrium setup. The wavelength chosen for absorbance measurements was 280 nm, which was chosen to maximize sensitivity given there is not an available wavelength to observe Ub-PCNA but not PIP-UBM1-UBM2. The concentration of Ub-PCNA (2.7 µM) was again chosen to balance minimum and maximum absorbance. Ratios of 1:1 and 1:2 Ub-PCNA:PIP-UBM1-UBM2 were run in two cells.

**Table 3.6** Speed/time settings for Ub-PCNA/PIP-UBM1-UBM2 AUC-SE experiment.

<table>
<thead>
<tr>
<th>Scan #</th>
<th>Mean RCF (g)</th>
<th>Hours Before Scan</th>
<th>Total Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4,660</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>2</td>
<td>4,660</td>
<td>4</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>7,280</td>
<td>15</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>7,280</td>
<td>4</td>
<td>49</td>
</tr>
<tr>
<td>5</td>
<td>10,480</td>
<td>15</td>
<td>64</td>
</tr>
<tr>
<td>6</td>
<td>10,480</td>
<td>4</td>
<td>68</td>
</tr>
</tbody>
</table>

The scanning method was adjusted based on experience with the PIP-UBM1 complex. The speed of the third scan was reduced to improve the data range and the fourth scan was omitted. Data processing was performed in the same manner as the PIP-UBM1 complexes above.
Figure 3.12 Sedimentation equilibrium fit for PIP-UBM1-UBM2/UbPCNA. Upper graph: absorbance vs radial length for 4,660 g (blue), 7,280 g (yellow) and 10,480 g (red). Lower graph: residuals of the fit. Data is cut off before the end of the cell as absorbance exceeds A=1. Overall fit is good, the residuals are not perfectly randomly distributed as might be hoped, and the goodness of fit breaks down toward the bottom of the cell.

Table 3.7 Dissociation constants for PIP-UBM1-UBM2/Ub-PCNA complex determined by sedimentation equilibrium.

<table>
<thead>
<tr>
<th>$K_d$</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_d1$</td>
<td>0.38±0.04 µM</td>
</tr>
<tr>
<td>$K_d2$</td>
<td>23.0±1.32 µM</td>
</tr>
<tr>
<td>$K_d3$</td>
<td>$&gt;&gt;K_d1$</td>
</tr>
<tr>
<td>$\chi^2$</td>
<td>0.671</td>
</tr>
</tbody>
</table>

Dissociation constants were again determined with a 95% confidence interval. These $K_d$ values for the PIP-UBM1-UBM2/Ub-PCNA are consistent with those found for PIP-UBM1. The overall pattern is similar, but they show higher affinity for $K_d1$, lower affinity for $K_d2$, and $K_d3$ is so large that the addition has not been observed. The decrease in $K_d2$ may be caused by increased crowding effects, but the difference between the $K_d2$ values is so small that it may not reflect any actual difference.
3.6 Reprocessing of ITC data

Isothermal titration calorimetry experiments had been performed by the lab prior to this study. The experiment had assumed that each Ub-PCNA subunit would be able to host a PIP-UBM1 partner simultaneously, so the analysis considered the protomer Ub-PCNA as the binding partner of interest.

![Figure 3.13 Original ITC data analysis, against Ub-PCNA protomer. The n value is very nearly 1, and there is no reason to suspect that the maximum stoichiometry is anything but 1:1 (equivalent to 1:3 Ub-PCNA (trimer):PIP-UBM1).](image)

This result appears to rest on incorrect assumptions, based on the AUC experiments, therefore it was prudent to reinterpret the data considering the new observation. Ub-PCNA concentration was reset to consider the trimer as the particle of interest, and the $n$ value was set to float between 2.5 and 3.5 to represent the total number of available binding sites.
Figure 3.14 ITC isotherm for trimeric Ub-PCNA vs PIP-UBM1. The \( n \) value is nearly 3 and does not show any indications that the Ub-PCNA is not being fully saturated by PIP-UBM1. The approach to complete binding occurs between the mole ratio of 5-6, which is also the saturation point observed in the AUC-SV results.

ITC reprocessing does not show any signs that the Ub-PCNA trimer cannot be saturated by PIP-UBM1. The \( n \) value of 3.021 is a perfect match for complete saturation of the Ub-PCNA ring. It is possible that ITC cannot be used to determine the ratio of the complex. The \( K_d \) value 10 \( \mu \)M is a fair match for the AUC values (\( K_d = 1.01 \) and 2.85 \( \mu \)M). Considering that the ITC analysis is attempting to fit 3 identical binding sites, the underestimation of the \( K_d \) is unsurprising. Overall this experiment indicates that the AUC-SV experiment is likely measuring the strength of the binding well.

ITC has the fewest sources of error for investigating binding, so long as the protein concentrations are accurately measured. Its main shortcoming is that the software used to analyze it cannot separate sequential bindings. Rather than \( K_d1 \), \( K_d2 \), and \( K_d3 \), it returns a single \( K_d \) value and an \( n \) value for the binding ratio.
3.7 Solution structure study by small-Angle X-ray scattering

Small-angle X-ray scattering (SAXS) experiments were a good complement to the MALS and AUC experiments, and an opportunity to get structural information about the complex. A single set of experiments were able to provide both types of information.

3.7.1 SAXS measurements of Ub-PCNA are in line with literature values, the 1:1 complex with PIP-UBM1 can be detected after SEC.

Experimental runs were performed using Ub-PCNA alone, PIP-UBM1 alone, 3:1 PIP:Trimer, and 2:1 PIP:Trimer. In mixtures containing Ub-PCNA, its concentration was 5.3 mg/mL (47 µM) and PIPUBM1 was added to achieve the ratio required. PIP-UBM1 alone was run at a concentration of 11.6 mg/mL (872 µM) to maximize scattering of the small protein.

Table 3.8 Sample injection conditions for SEC-SAXS experiments.

<table>
<thead>
<tr>
<th>Injection Number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ub-PCNA Trimer (µM)</td>
<td>47.4</td>
<td>47.4</td>
<td>0</td>
<td>47.4</td>
</tr>
<tr>
<td>PIP-UBM1 (µM)</td>
<td>0</td>
<td>142.8</td>
<td>872.2</td>
<td>97.7</td>
</tr>
<tr>
<td>Injection Size (µL)</td>
<td>250</td>
<td>250</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>Ratio Trimer:PIPUBM1</td>
<td>1:0</td>
<td>3:1</td>
<td>0:1</td>
<td>2:1</td>
</tr>
</tbody>
</table>

Data were collected as a series of images and initial data processing was performed using the RAW analysis suite, followed by *ab initio* modelling using DAMMIN, DAMMIF, and GASBOR.

Basic SAXS processing was performed at the beamline using the RAW suite. Molecular weight (MW) and radius of gyration (R_g) measurements are easily processed and were used to assess the data quality. Ub-PCNA and the complex were both of good quality: each had higher molecular weights than theoretical but within the acceptable range for the experiment, and each had consistent R_g across the peak. These suggest that the sample quality was good and that the sample was monodisperse across the peak. PIP-UBM1 alone did not scatter well enough to measure R_g but its molecular weight was consistent with the theoretical.
The following molecular weights and $R_g$ values were measured during each run. Also included is the $D_{\text{max}}$ estimate. $D_{\text{max}}$ is a user-generated parameter and represents the longest linear dimension of the particle. $R_g$ and $D_{\text{max}}$ values for Ub-PCNA are in line with values previously reported (Hibbert and Sixma, 2012).

**Table 3.9** Basic SAXS measurements for Ub-PCNA, PIP-UBM1, and complex.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Mw (kDa)</th>
<th>Theoretical Mw</th>
<th>$R_g$</th>
<th>$D_{\text{max}}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ub-PCNA</td>
<td>118</td>
<td>111.9</td>
<td>39.9</td>
<td>127</td>
</tr>
<tr>
<td>PIP-UBM1</td>
<td>13.7</td>
<td>13.3</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>1:1 Complex</td>
<td>138</td>
<td>125.2</td>
<td>44.5</td>
<td>157</td>
</tr>
</tbody>
</table>
Figure 3.15 Ub-PCNA and Complex SAXS comparison (a) Pair distance distribution functions (PDDF) and (b) plotted I(q) vs q (intensity vs scattering angle) for Ub-PCNA alone and PIP-UBM1/Ub-PCNA complexes. Ub-PCNA: red and blue curves. Complex: purple and green. These demonstrate that there is a measurable difference between the Ub-PCNA and the complex. The PDDF graph shows that the complex is a physically larger particle, with a larger D$_{max}$ and larger average dimensions. The PDDF is also of approximately the correct shape. PCNA itself is ring shaped, while the ubiquitin moieties are placed at the outer edge and all face towards one of the faces. The overall shape of each plot is intermediate between the disc pattern and the sphere pattern, which is consistent with the structure of the molecule. The I(q) plot shows that the proteins are in good shape, and again has an observable difference between the complex and the Ub-PCNA alone.

3.7.2 SAXS envelope modelling can reconstruct Ub-PCNA, reconstruction of the complex is ambiguous

*Ab initio* structure modelling of Ub-PCNA and the complex was performed to create a model of the shape of the Ub-PCNA alone and the complex. DAMMIF/N and GASBOR were both tested before GASBOR was chosen. DAMMIF/N’s dummy atom modelling had difficulty generating a pore to locate the PCNA ring. Averaging and
clustering also showed that the reconstruction was ambiguous and many possible models could fit the data. GASBOR’s dummy residue model created more interpretable envelopes and produced particularly good Ub-PCNA envelopes.

![Ubiquitin and PCNA model](image)

**Figure 3.16** Reconstructed model of Ub-PCNA using GASBOR dummy residue modelling. This model is looking toward Ub-PCNA’s rear face, with the plane of the ring parallel with the page, and the ubiquitin domains pointing toward the reader.

This is a good reconstruction of Ub-PCNA in solution from low-resolution data. The PCNA pore is apparent, the pseudo-P6 symmetry of the PCNA ring looks to be in place, and the ubiquitin domains are placed in a semi-extended position. This reconstruction was achieved by enforcing P3 symmetry on the model. Unfortunately, this strategy cannot be applied to the complex reconstruction.
Figure 3.17 Reconstructed model of PIP-UBM1/Ub-PCNA 1:1 complex using GASBOR dummy residue modelling. Left image shows the PCNA ring in plane of the page, right image shows the PCNA ring perpendicular to the plane. PIP-UBM1 is believed to appear as the bulge on the right in image A and above on image B.

The complex is only expected to have one PIP-UBM1 based on the observed molecular weight, so the symmetry of the complex is P1. This leaves the modelling program with fewer constraints. The reconstructions of the complex did not average or cluster well, which suggests that the data may be ambiguous, and the complex reconstruction is a best guess. Tentatively, the reconstruction appears to be in a trans-arrangement, with a wide, distributed mass of PIP-UBM1 along a side of Ub-PCNA. The cis-arrangement would be expected to be a narrow, clustered mass. This trans-arrangement may be a reason for the limited complex ratio observed in sedimentation velocity experiments.
4 Discussion

4.1 The observed complex ratio between Ub-PCNA and polt is limited to two polt for each Ub-PCNA ring

At the beginning of this research project, the operating assumption that the complex between PIP-UBM1 and Ub-PCNA was straightforward: each Ub-PCNA protomer could host a single PIP-UBM1. Each PIP-UBM1 had a pair of binding motifs which bound to the complementary pair of binding sites on Ub-PCNA. We assumed that PIP-UBM1 was small enough (117aa) to be accommodated in multiple bindings on Ub-PCNA.

The initial SEC-MALS findings, with its hard cap of a 1:1 complex, were not expected but inconclusive. Even a large excess of PIP-UBM1 was not able to increase the measured molecular weight of the complex. The similar SEC-SAXS experiment confirmed the observation, but both shared the SEC component of the experimental setup. This system had the advantage of separating the complex from the excess binding partner. It was therefore possible that the 2nd and 3rd bindings were simply being shorn off as it flowed through the column. The difference in the speeds of the columns and the temperature during the columns may have contributed to the different MW estimates. To prove that the observed complex ratio was not an experimental artifact, another type of experiment had to be conducted.

Sedimentation velocity experiments were a perfect complement to the previous experiments. Like SEC-MALS and SEC-SAXS it provided a good estimate of molecular weight, but it did so with a completely different experimental method (sedimentation velocity vs. light scattering). This method also keeps the complex in question surrounded by the excess binding partner. Rather than stripping away excess protein, the complex sediments through a sea of excess protein and should maintain its state through the entire experiment. These experiments broadly agreed with the scattering experiments in that a 1:1 complex forms readily but differed in demonstrating that a 2:1 complex is achievable with enough excess PIP-UBM1. The second addition is likely present in some proportion of the population in the “1:1” experiment, along with an equal proportion of free Ub-PCNA. This heterogeneity would be a hindrance to further crystallization experiments.
Polt has moderately strong affinity for Ub-PCNA, as measured by ITC experiments. In both ITC and AUC experiments using PIP-UBM1 and Ub-PCNA, the complex appears to reach its final state at a ratio between 1:4 and 1:6. The apparent limit may be caused by the weak, dynamic binding, but the ease with which the first two additions occur suggests there may be a crowding factor that could be confirmed by structural studies.

4.2 Comparison of methods to investigate binding behaviour

Analytical ultracentrifugation and isothermal titration calorimetry were used to investigate the binding of the complex. Sedimentation velocity experiments measured the molecular weight of the complex to determine the complex ratio, followed by equilibrium experiments to determine the dissociation constants. This method was chosen to observe the complex in solution with a large excess of its binding partners, unlike experiments that relied on a size exclusion column. Sedimentation equilibrium experiments were also able to determine each $K_d$ separately and use those to make inferences about cooperativity in the binding.

Isothermal titration calorimetry experiments were used primarily for the determination of a dissociation constant. The $n$ value was also used as an indicator of the complex ratio. ITC returns a single $K_d$, which represents an average of the three separate bindings.

ITC is a much simpler method, with fewer sources of error than the AUC experiments. Fitting a binding curve to the data has fewer degrees of freedom. Determining the dissociation constants by AUC in a three-site system has more degrees of freedom and therefore it can be difficult to conclude that the determined $K_d$ values are accurate. Adding to this complexity, the data analysis requires determination of sedimentation and absorbance values for participating species, which introduces numerous possible sources of error. The sedimentation coefficients themselves are a possible source of error. Even if AUC-SV experiments determine accurate sedimentation coefficients, they must be converted to a molecular weight to be intelligible. Since the shift in mass is so small (each PIP-UBM1 added is only 11% of the mass of Ub-PCNA) it is possible that a small error in the mass calculation may lead to a misidentification of the complex observed. These errors combined make the AUC results less conclusive, so the disagreement of the $n$ value from
ITC and the complex ratio found in AUC is troubling. It may be wise to take the AUC results as provisional at best, but it would be useful to consider their implications if they are borne out by further experiments.

4.3 Complex structure predictions

Explicit structure determination by crystallographic and SAXS methods may have failed, but some predictions of the structure may be made. Ub-PCNA appears to be able to accommodate a maximum of two pol\(\iota\) based on AUC experiments. Although both additions are in equilibrium with one another, SEC-MALS data suggests that the single-binding arrangement is the most stable over long time courses. Combined with the unobserved third binding, this suggests that there is a crowding effect that affects each subsequent binding. This crowding is suggestive of a \textit{trans}-binding arrangement. A \textit{cis}-binding arrangement is not necessarily immune to crowding effects, but the \textit{trans} arrangements is virtually certain to cause this effect.

A definitive reconstruction of the SAXS envelope of the complex was not possible, but the best reconstructions tended to have broad regions assigned to the PIP-UBM1 section. This may indicate that the complex is in dynamic equilibrium during the SEC-SAXS experiment. Based on AUC-SV experiments, the apparent 1:1 complex may be an equilibrium between the 1:0, 1:1, and 1:2 species. The SEC-SAXS method is normally employed to isolate species and capture scattering from a homogenous sample. In this case, a “1:1” sample may have quickly reached a new equilibrium state that could not be further isolated by the size exclusion column.

The pol\(\iota\)/Ub-PCNA complex studied here may not be entirely biologically relevant. The TLS complex also includes Rev1 as a scaffold protein, and ubiquitinated pol\(\iota\) is necessary for pol\(\iota\) recruitment (Vaisman and Woodgate, 2017). It is possible that one or both UBMs may prefer to bind to other ubiquitinated proteins in the complex \textit{in vivo}. Pol\(\iota\)’s direct interaction with Ub-PCNA is still a matter of interest, as no structure of any Y-family C-terminal domain in complex has been solved. Models for TLS function rely on inferences from function experiments and molecular simulations. Solving the pol\(\iota\)/Ub-PCNA complex would be an important part of modelling the overall TLS complex.
4.4 Implications for TLS mechanism and polymerase selection

The observation that truncated polI proteins cannot form a 1:3 complex with trimeric Ub-PCNA suggests that it is a poor candidate to participate in tool belts. The tool-belt model of lesion bypass requires Ub-PCNA to host several TLS polymerases, each of which may probe the lesion. This increases the probability that the most efficient, and therefore likely the most suitable, polymerase performs the bypass. Tool-belts have been formed in vitro using PIP-containing proteins and a two-part tool belt of polη and Rev1 has been observed in yeast in vivo (Boehm et al., 2016). Y-family polymerases, which are the primary TLS polymerases, all have large, unstructured C-terminal domains that bind to Ub-PCNA.

There is variation in C-terminal organizations between the Y-family polymerases (Figure 1.1). PolI cannot easily form tool belts even as the only available polymerase, where all three hypothetical proteins would be able to adopt the same arrangement to minimize crowding effects. It therefore seems unlikely that it would readily form tool belts alongside differently organized proteins like other members of the Y-family.

While this finding does not exclude the possibility of tool belts within TLS, it demonstrates that PolI is not suited to participating in them. If the tool-belt model is validated in humans and the other TLS polymerases form functional tool-belts in vivo, then that suggests that PolI acts as a last-resort option. Polymerase ubiquitination (as outlined in the polymerase switch model) would also be at work in a tool-belt system. A lesion for which none of the tool-belt polymerases are suited would continue to stall replication. The first-response tool-belt polymerases would then continue to bind, probe, and be deactivated by ubiquitination. Eventually the local pool of these polymerases would deplete until their local concentrations dropped to the point that tool-belt formation could not continue, at which point PolI would be able to bind Ub-PCNA and engage the lesion. PolI’s role in this model would be that of a last-line lesion specialist. Since it would only be activated rarely, this model would account for PolI’s apparent low importance to organism survival; PolI knockout mice being apparently healthy unlike XPV patients (Kanao et al., 2015).
References


Appendices

Appendix A: Purification of Ubiquitination Enzymes

hUba1 Purification

hUba1 purification followed the procedure outlined in Methods. As a ubiquitination enzyme, small quantities are needed, so doing a large scale (2 litre culture) purification can produce enough to ubiquitinate < 100 mg of PCNA.

![Figure A1 SDS-PAGE gel outlining Ni-affinity and anion exchange steps in Uba1 purification. Strong contaminants remain.](image)

Purity is not crucial for this protein, but it is still unacceptable at this stage. Protein was concentrated to 1mL and run on S200 to clean it up.
Figure A2 Final S200 fractions of Uba1. Acceptable purity has been achieved at this stage.

Peak fractions containing Uba1 were pooled and concentrated to 3.8 mg/mL (32 
µM). The protein was distributed in 30 µL aliquots, flash frozen, and stored at -80 ºC until
used in ubiquitination reactions.

hUbC S22R Purification

hUbc S22R purification followed the procedure outlined in Methods. As a
ubiquitination enzyme, small quantities are needed, so doing a large scale (2 litre culture)
purification can produce enough to last a long time. The purification only requires two Ni
columns, one after lysis and one after tag cleavage.
Since the protein is to be frozen and then only used in small quantities for a single preparative reaction, the small amount of Ni which leached from the column is acceptable. FT2 and FT3 were pooled and concentrated to 17 mg/mL (1 mM) with a yield of 51 mg. The concentrated protein was distributed into 200 µL aliquots and frozen down for later use.

Figure A3 SDS-PAGE gel demonstrating the purification of UbC. High purity is achieved after the second Ni-affinity step that follows tag cleavage.
Curriculum Vitae

Name: Harrison Taylor

Post-Secondary Education and Degrees:

Queen’s University
Kingston, Ontario, Canada
2010-2014
Bachelor of Science in Biochemistry
Honours Specialization

University of Western Ontario
London, Ontario, Canada
2016-2019
Master of Science