NOTE TO USERS

This reproduction is the best copy available.

UMI®
RICE UNIVERSITY

Structural and thermodynamic analysis of human PCNA bound to peptides derived from DNA polymerase-delta p66-subunit and flap endonuclease-1 (FEN1) proteins

by

John Bruning

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Yousif Shamoo, Assistant Professor, Thesis Advisor
Biochemistry and Cell Biology

John S. Olson, Ralph and Dorothy Looney Professor
Biochemistry and Cell Biology

Seiichi Matsuda, Professor
Biochemistry and Cell Biology
Chemistry

Yizhi Jane Tao, Assistant Professor
Biochemistry and Cell Biology

Ronald Parry, Professor
Chemistry

Houston, Texas
November, 2004
INFORMATION TO USERS

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleed-through, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.
ABSTRACT

Structural and thermodynamic analysis of human PCNA bound to peptides derived from DNA polymerase-δ p66-subunit and flap endonuclease-1 (FEN1) proteins

by

John Bruning

DNA replication and repair requires the coordination of multiple proteins at the replication fork. Among the proteins found at the replisome, hPCNA (human Proliferating Cell Nuclear Antigen) is essential for many of the protein-protein interactions required for replication and repair processes. hPCNA is a member of the sliding clamp family of DNA replication processivity factors. Recent studies have shown that hPCNA is able to make specific protein-protein interactions to a wide range of DNA replication, recombination and repair proteins through a small peptide motif termed the PCNA-Interacting-Protein or “PIP-Box.” The structures have been solved of hPCNA bound to PIP-Box containing peptides derived from the p66-subunit of the human replicative DNA polymerase δ at 2.6 Å resolution and of the human flap endonuclease (FEN1) at 1.85 Å resolution. Both structures indicate that the pol-δ p66 subunit and
FEN1 endonuclease peptides bind at a site previously shown to bind a PIP-Box containing peptide from the cdk-inhibitor p21 (CIP1). Both interacting peptides adopt a similar elongated overall conformation with a central $3_{10}$ helix that plugs into a hydrophobic cavity of hPCNA. Isothermal titration calorimetry studies indicate that PIP-box containing peptides from the p66-subunit of the pol-$\delta$ holoenzyme and FEN1 bind hPCNA less tightly, 15.6 $\mu$M and 59.9 $\mu$M respectively, than p21-derived peptide (82.6 nM). Although it is likely that hPCNA may make additional interactions to the p66 and FEN1 proteins, these structural and calorimetry studies indicate that interactions made through the PIP-box motif can account for much of the hPCNA specificity for these proteins observed in vivo. These studies support a model where hPCNA uses a conserved set of protein-protein contacts made through the PIP-Box to attenuate binding to the proteins of DNA replication, recombination and repair, and that these proteins compete for binding through the same site on hPCNA. The significance of structural studies within the human PCNA system is underscored by the potential antibiotic and cancer drug targets on the PIP-box binding surface.
ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Shamoo. He always kept me challenged scientifically and intellectually, allowing the proper advancement of my scientific expertise. He also never let me quit when I wanted to; when I had convinced myself of absolute failure he believed in me and moved me forward. He also always made it a lab standard to have the latest equipment and materials necessary for lab members, no easy feat when grant money is hard to come by. Additionally, he was always available for help when I had problems in lab. Finally, I would also like to thank him for all the time he spent with me at the computer helping me rebuild my models.

I am grateful to my committee: Dr. Olson, Dr. Tao, Dr. Matsuda, and Dr. Singleton. They were very supportive of me and very understanding and flexible. As I am a very independent person, they gave me the freedom I needed to explore my project in a manner that fit my personality. I would also especially like to thank Dr. Tao for looking at my data whenever I asked, and helping me overcome my twinning problems.

I would like to thank my lab members: Jeff, Raf, Siyang, Danny, Marc, Jim, and Megan. They have tolerated a lot from me, and always been there.

I would like to thank Kevin and the entire MacKenzie lab. Kevin always made himself available for help, and often presented alternative ways of looking at things scientifically that often helped foster a broader perspective.

I would like to thank Dr. Shamoo, Siyang, Jeff, and M. Szebenyi for help at the synchrotron.
Todd Yeates was also very helpful, and gave me advice to help deal with my twinning problem.

I am also grateful for friendships with Sean Moran, Philip, Junhua, and Corey.

I would also like to thank the financial supporters: Houston Livestock Show and Rodeo, the HAMBP training grant given by the NIH training grant (GM08280), a grant to Dr. Shamoo from the American Cancer Society (RSG GMC-104925) and a fellowship to J.B. from The Robert A. Welch Foundation (C-1584).
# TABLE OF CONTENTS

Chapter 1: Introduction ........................................... 1  
  1.1 The Replisome ............................................. 1  
  1.2 Sliding Clamps ............................................ 6  
  1.3 hPCNA's Interactions and the PIP(PCNA Interacting Protein)-Box ...... 9  
  1.4 DNA polymerases ........................................ 12  
  1.5 Flap Endonuclease-1 (FEN1) ................................ 15  
  1.6 The Werner protein ....................................... 19  
  1.7 RecQ5 ..................................................... 22  
  1.8 p21 and Cell Cycle Control ................................ 24  
  1.9 Structure of p21(139-160) bound to hPCNA ....................... 26  
  1.10 Mutational and biochemical analysis of PIP-box protein's interaction with hPCNA .............................................. 31  

Chapter 2: Preparation of hPCNA and PIP-box peptides for crystallization .............................................. 33  
  2.1 Cloning of the hPCNA expression plasmid ...................... 33  
  2.2 Cell growth and protein expression .......................... 34  
  2.3 Previously published purification of hPCNA .................... 35  
  2.4 Purification of hPCNA ...................................... 37  
  2.5 Peptide production, purification, and quality analysis .......... 46  

Chapter 3: Crystallization of hPCNA with interacting peptides ...... 57  
  3.1 General strategy ........................................... 57  
  3.2 Proof of concept: recreating hPCNA-p21(139:160) co-crystals ...... 59  
  3.3 Crystallization of hPCNA with the p66 PIP-box peptide .......... 64  
  3.4 Crystallization of hPCNA with the FEN1 PIP-box peptide .......... 67  
  3.5 Crystallization of hPCNA with the WRN PIP-box peptide .......... 70  
  3.6 Crystallization of hPCNA with the RecQ5 PIP-box peptide .......... 72  
  3.7 Cryo-crystallography ........................................ 74  

Chapter 4: Preliminary characterization of crystal diffraction and
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Screening of crystals by means of diffraction</td>
<td>77</td>
</tr>
<tr>
<td>4.2</td>
<td>X-ray diffractometers and collection apparatus</td>
<td>79</td>
</tr>
<tr>
<td>4.3</td>
<td>hPCNA-p21(139:160) co-crystals</td>
<td>79</td>
</tr>
<tr>
<td>4.4</td>
<td>hPCNA-p66(452:466) co-crystals</td>
<td>81</td>
</tr>
<tr>
<td>4.5</td>
<td>hPCNA-FEN1(331:350) co-crystals</td>
<td>83</td>
</tr>
<tr>
<td>4.6</td>
<td>hPCNA-WRN co-crystals</td>
<td>85</td>
</tr>
<tr>
<td>4.7</td>
<td>hPCNA-RecQ5 co-crystals</td>
<td>86</td>
</tr>
<tr>
<td>5.1</td>
<td>Indexing, scaling, and merging of hPCNA-p66(452:466) diffraction data</td>
<td>89</td>
</tr>
<tr>
<td>5.2</td>
<td>Indexing, Scaling, and merging of hPCNA-FEN1(331:350) diffraction data</td>
<td>89</td>
</tr>
<tr>
<td>6.1</td>
<td>Background on crystal twinning</td>
<td>92</td>
</tr>
<tr>
<td>6.2</td>
<td>Detwinning</td>
<td>95</td>
</tr>
<tr>
<td>6.3</td>
<td>Warning signs of twinning and twinning detection</td>
<td>96</td>
</tr>
<tr>
<td>6.4</td>
<td>Detection of hPCNA-FEN1(331:350) peptide crystal twinning</td>
<td>98</td>
</tr>
<tr>
<td>7.1</td>
<td>hPCNA-p66(452:466) structure solution</td>
<td>103</td>
</tr>
<tr>
<td>7.2</td>
<td>hPCNA-FEN1(331:350) structure solution</td>
<td>108</td>
</tr>
<tr>
<td>7.3</td>
<td>Comparison of hPCNA structure quality</td>
<td>111</td>
</tr>
<tr>
<td>8.1</td>
<td>Background of p21 PIP-box peptide ITC</td>
<td>113</td>
</tr>
<tr>
<td>8.2</td>
<td>ITC methodology</td>
<td>114</td>
</tr>
<tr>
<td>8.3</td>
<td>Comparison of PIP-box peptide association constants and thermodynamic properties</td>
<td>116</td>
</tr>
<tr>
<td>9.1</td>
<td>Overall structure comparison</td>
<td>120</td>
</tr>
</tbody>
</table>
9.2 p66 and FEN1 PIP-box Interactions ........................................ 123
  9.2a Comparison of PIP-box N-terminal regions ....................... 123
  9.2b Comparison of PIP-box 3_{10} helical regions ..................... 127
  9.2c Comparison of PIP-box C-terminal regions ....................... 128
  9.2d Conclusions .................................................................... 129

Chapter 10: p21 is a model for a high affinity PIP-box ................. 130
  10.1 3_{10} helix packing volumes of p66, FEN1, and p21 PIP-box peptides .. 130
  10.2 Hydrogen bonding and ionic interactions ............................ 132

Chapter 11: PIP-boxes have a small contact surface with PCNA ...... 136

Chapter 12: Comparison of DNA polymerase-sliding clamp
  interactions ........................................................................... 139
  12.1 The Pol IV/DinB co-structure ........................................... 139
  12.2 Comparison of the prokaryotic and human clamp interactions .. 140

Chapter 13: Biological significance of hPCNA-p66(452:466),
  hPCNA-FEN1(331:350), and hPCNA-p21(139:160) ...................... 144

Chapter 14: PCNA: a putative anti-cancer drug target ............... 147

References ............................................................................... 149
LIST OF TABLES

Table 1.1  Alignment of known human PIP-box sequences .......................... 11
Table 5.1  Data processing statistics ......................................................... 91
Table 7.1  Cross-rotation solutions using the hPCNA-p66(452:466) data ........ 103
Table 7.2  Translation solutions using the hPCNA-p66(452:466) data .......... 104
Table 7.3  Cross-rotation solutions using the hPCNA-FEN1(331:350) data .... 109
Table 7.4  Translation solutions using the hPCNA-FEN1(331:350) data ....... 109
Table 7.5  Refinement statistics for hPCNA-p66(452-466) and
           hPCNA-FEN1(331:350) .................................................................. 112
Table 8.1  Peptides used in ITC experiments ............................................. 115
Table 8.2  hPCNA isothermal titration calorimetry with p21, p66a, p66b,
           and FEN1 ............................................................................. 120
LIST OF FIGURES

Figure 1.1 Events prior to replication initiation _________________2
Figure 1.2 Schematic diagram of the eukaryotic replication bubble ______4
Figure 1.3 Crystal structures of sliding clamps from across kingdoms ______7
Figure 1.4 The human PCNA monomer bound to the p21-derived peptide ____8
Figure 1.5 Various DNA secondary structure elements ______________________16
Figure 1.6 Crystal structure of A. fulgidus FEN1 bound to a partial DNA
    substrate ________________________________________20
Figure 1.7 Human RecQ family proteins aligned by their RecQ helicase domains
    compared to the E. coli RecQ to which they are related _________22
Figure 1.8 Interactions of the 3_10 helix of the p21-derived peptide with hPCNA __28
Figure 1.9 N-terminal interactions of the p21-derived peptide with hPCNA ___29
Figure 1.9 C-terminal interactions of the p21-derived peptide with hPCNA ___30
Figure 2.1 SDS-PAGE of DE52 elutions ___________________________________38
Figure 2.2 Phenyl sepharose FPLC chromatogram __________________________40
Figure 2.3 Q sepharose chromatography ____________________________41
Figure 2.4 Heparin sepharose chromatography ____________________________43
Figure 2.5 Mono Q chromatography ______________________________________45
Figure 2.6 p21 peptide purification used for crystallography and ITC ______48
Figure 2.7 Purification of the p66a peptide used for ITC __________________49
Figure 2.8 Purification of the p66b peptide used for ITC __________________50
Figure 2.9 Purification of the p66 peptide used for crystallography _________51
Figure 2.10 Purification of the FEN1 peptide used for crystallography ______52
Figure 2.11 Purification of the FEN1 peptide used for ITC __________________53
Figure 2.12 Purification of the WRN peptide used for crystallography and ITC _54
Figure 2.13 Purification of the RecQ5 peptide used for ITC ________________55
Figure 2.14 Purification of the RecQ5 peptide used for crystallography_____56
Figure 3.1 Published crystal conditions for hPCNA-p21 peptide __________62
Figure 3.2 Hexagonal shaped co-crystals of hPCNA-p21peptide __________64
Figure 3.5 Co-crystals of hPCNA and p66 peptide ________________________67
Figure 3.4  PCNA-FEN1 peptide co-crystals ................................. 69
Figure 3.5  hPCNA-WRN peptide co-crystals ................................ 72
Figure 3.6  hPCNA-RecQ5 peptide co-crystals ................................ 74
Figure 4.1  Diffraction pattern of hPCNA complexed with p21 peptide ... 81
Figure 4.2  Diffraction pattern of hPCNA complexed with p66 peptide ... 83
Figure 4.3  Diffraction pattern of hPCNA-FEN1 peptide co-crystals ....... 85
Figure 4.4  Diffraction pattern of hPCNA-WRN peptide co-crystals ....... 86
Figure 4.5  Diffraction pattern of hPCNA-RecQ5 peptide co-crystals ...... 88
Figure 6.1  Schematic representation of hemihedral twinning within theoretical crystal lattices ................................................................. 95
Figure 6.2  Intensity statistics (<H>) distribution for various twin fractions ... 100
Figure 8.1  ITC measurements of binding for peptides derived from the PIP-box of p66, FEN1, and p21 to hPCNA .............................................. 117
Figure 9.1  Overall structure comparison of hPCNA bound to peptides derived from human DNA polymerase-δ p66-subunit, FEN1, and p21 ........ 121
Figure 9.2  Stereo view of electron density from 2Fo-Fc maps (contoured at 1σ) of peptides derived from p66 and FEN1 overlayed on a ribbons diagram of the hPCNA monomer .............................................. 122
Figure 9.3  Molecular recognition surfaces of hPCNA in complex with peptides derived from the C-terminus of PCNA-interacting proteins p66, FEN1, and p21 ................................................................. 124
Figure 9.4  The p66 peptide N-terminus wraps around C-terminus of hPCNA .... 126
Figure 9.5  Superimposition of conserved PIP-box Gln from p66, FEN1, and p21 ................................................................. 127
Figure 10.1  Packing of the conserved 310 helix of p66 (452-466) and FEN1 (331-350) with hPCNA ................................................................. 131
Figure 10.2  Alignment of PIP-box structures in hydrogen bonding networks ... 134
Figure 11.1  Superimposition of the PIP-box regions from six PIP-Box to sliding clamp co-crystal structures ............................................. 138
Figure 12.1  Crystal structure of the E. coli beta-clamp with the pol IV/DinB(LF domain) ................................................................. 142
Figure 12.2  Contact surface of polIV/DinB with the beta-clamp 143
LIST OF ABBREVIATIONS

α, twinning fraction

aPCNA, Archae proliferating cell nuclear antigen

AP, apurinic or abasic

BER, base excision repair

β, E. coli beta clamp

Cdc, cell division cycle

CDK, cyclin dependent kinase

Cdt1, Cdc 10-dependent transcript 1

CIP, CDK interacting protein

DDK, Dbf4-dependent kinase

ΔG, free energy

ΔH, enthalpy

ΔS, entropy

DMSO, dimethylsulfoxide

DTT, dithiothreitol,

FEN1, flap endonuclease-1

FPLC, Fast Performance Liquid Chromatography

hPCNA, human proliferating cell nuclear antigen

HPLC, high performance liquid chromatography

HRDC, helicase and RNase D C-terminal

IPTG, Isopropyl-β-d-thiogalactopyranoside
ITC, isothermal titration calorimetry
K, affinity constant
K_d, equilibrium dissociation constant
K_m, the Michaelis constant
LB, Luria-Bertani Media
MALDI-TOF, matrix assisted laser desorption/ionization, time of flight mass spectroscopy technique
MES, 2-[N-Morpholino]ethanesulfonic acid
MCM, minichromosome maintenance protein
MIR, multiple isomorphous replacement
MPD, 2-methyl-2, 4-pentanediol
N, stoichiometry
ORC, origin recognition complex
p, protein
PCNA, proliferating cell nuclear antigen
PCR, polymerase chain reaction
PDB, protein data bank
PEG, polyethylene glycol
pfuPCNA, *Pyrococcus furiosus* proliferating cell nuclear antigen
PIP, PCNA-interacting-protein
PMSF, phenyl methyl sulfonyl fluoride
pol, polymerase
RecQ-Ct, RecQ C-terminal
RFC, replication factor C

RPA, replication protein A

SDS-PAGE, sodium dodecyl sulphate polyacrylamide gel electrophoresis

WRN, Werner

yPCNA, yeast proliferating cell nuclear antigen
Chapter 1: Introduction

1.1 The replisome

DNA replication requires the cooperation of many proteins, collectively known as the replisome, at the replication fork. Stages required for genome duplication, via the replisome, include pre-replication, pre-initiation, initiation and replication, and post-replication (Reviewed by Davey et al., 2000). DNA replication requires a host of protein machinery to ensure proper fidelity, processivity, and timing. In addition to the machinery directly involved in DNA replication, several proteins are required at the replisome for processes such as DNA repair, DNA modification, and proper cell cycle control.

Proteins involved in pre-replication and pre-initiation are known as initiators. Eukaryotic DNA replication starts from many points on the chromosome known as origins of replication and can only be "fired" once during each cell division. Initiators bind to these origins of replication and recruit the replication machinery as well as contribute to the control of replication timing. Initiators recognize specific sequences or structural elements of DNA and are capable of helicase activity or helicase loading. During the cell cycle, origins of replication must become activated and competent for DNA replication. During late mitosis and early G1 phase of the cell cycle, initiators bind to the origin of replication and "license" the origin for replication. Once "licensed," the ensemble at the origin of replication is known as the pre-replication complex. Once the cell is prepared to enter S-phase, CDK's (Cyclin Dependent Kinases) become active and
allow replication to proceed. This is accompanied by the addition of more initiators to the origin of replication which are capable of recruiting polymerases and other replication machinery.

![Diagram](image)

**Figure 1.1** Events prior to replication initiation. A The origin replication complex (ORC) binds to DNA at origin B The ORC recruits Cdc6 and Cdt1 to the origin C Mcm2-7 complexes (shown as M) are recruited and this is known as licensing D Cdt1 and Cdc6 dissociate after licensing E Cdc7 is recruited and phosphorylates the Mcm2-7 complexes allowing replication fork initiation to begin under control of Cdc45 (Blow, 2001).
protein (Maiorano et al., 1994), DDK (Cdc7-Dbf4 kinase), and Cdc45 (Zou et al., 1997). How initiators assemble at the origin of replication is represented in Figure 1.1. The ORC contains six subunits which are the first proteins to bind the origin of replication and do so in an ATP-dependent manner (Bell and Stillman, 1992). Key proteins of the pre-replication complex include Cdc6 and Cdt1 which interact with the ORC (Mizushima et al., 2000). Cdc6 and Cdt1 facilitate the loading of the two hexameric MCM complexes to the origin of replication (Mendez and Stillman, 2003). DDK phosphorylates the MCM complexes during S-phase, activating its helicase activity (Lei and Tye, 2001). The helicase activity of the MCM complex creates the replication "bubble" by unwinding the DNA and allowing Cdc45 to bind at each fork of the bubble. Once the replication bubble is formed, the replication machinery is recruited and DNA replication is initiated.

DNA replication is characterized by moving forks of a replication bubble composed of leading and lagging strand DNA synthesis in the 5' to 3' direction as depicted in Figure 1.2. DNA is replicated by a host of DNA polymerases with a variety of specialized functions. Efficient and processive DNA replication requires pol α, pol δ, and pol ε (Johnson et al., 1985; Sitney et al., 1989; Morrison et al., 1990). DNA synthesis is primed by the formation of RNA primers. DNA replication is initiated by Pol α (primase) which synthesizes the RNA primer and remains attached to the primer as well as to the single-stranded DNA binding protein RPA (Hubscher et al., 2002). RPAs bind to single-stranded DNA denaturing adventitious secondary structure elements disruptive to replication and protect DNA from nuclease activity (Wold, 1997). Pol α synthesizes an RNA/DNA hybrid of about 10 nucleotides followed by 20-30 more of strictly DNA. After synthesis of the initial 30-40 bases, a pol switch occurs and synthesis
Figure 1.2 Schematic diagram of the eukaryotic replication bubble. Major constituents include the replicative polymerases dimeric Pol δ (subunits not shown for simplicity) and Pol ε on the lagging strand, Pol α required for primer formation, PCNA clamping the polymerases to the DNA, a hexameric helicase unwinding the DNA, and RP-A, the eukaryotic single stranded binding protein (Burgers 1998).
is continued by pol δ (Stucki et al., 2001). Pol δ is the main polymerase responsible for DNA synthesis at the leading and lagging strand during elongation phase (Waga et al., 1994). Four subunits of pol δ have been identified: p125, p66, p50, and p12 (Liu et al., 2000). DNA polymerase δ requires a processivity factor known as the sliding clamp (known as PCNA in eukaryotes) (Prelich et al., 1987). PCNA encircles DNA and tethers pol δ to the DNA via a C-terminal tail located on p66 (Ducoux et al., 2001). PCNA is loaded onto DNA by means of the clamp loader which is known as the RFC (Replication Factor C) (Yoder and Burgers, 1991; Burgers and Yoder, 1993). The clamp loader acts as a molecular wrench, requiring ATP, opening the circular clamp and closing it around DNA (Jeruzalmi et al., 2001a; Jeruzalmi et al., 2001b). Pol α is displaced by binding of the RFC to RPA, allowing for placement of the sliding clamp and polymerase at the replication fork. In a similar manner, pol δ competes with RFC for placement on the sliding clamp and displaces the RFC once the clamp has been loaded. Other DNA polymerases, such as pol ε, are capable of catalytic replication but are essential in cell cycle control and DNA damage repair (Lee et al., 1991; Turchi and Bambarra, 1993; Mossi et al., 1998). Lagging strand DNA synthesis creates a logistical challenge in which several newly synthesized DNA strands, known as Okazaki fragments, must be processed and joined together. FEN1 (Flap endonuclease 1) is a nuclease responsible for proper processing of Okazaki fragments (Bambarra et al., 1997). DNA ligases are also necessary to join these newly synthesized strands of DNA.

In addition to replication machinery, the replisome acts as a docking station for proteins of varied functions requiring access to DNA. Proteins that are involved in controlling the cell cycle, including cyclins and CDKs, have been localized to the
replisome and more specifically have been found to interact with PCNA (Koundrioukoff et al., 2000). Enzymes required for the alteration of DNA have also been localized to the replisome. Examples include methyltransferases and uracil DNA glycosylases. DNA repair proteins are required at the replisome including proteins involved in mismatch repair as well as nucleotide excision repair.

1.2 Sliding Clamps

All species require a processivity factor, usually in the form of a sliding clamp, for highly processive DNA synthesis and the proper recruitment of DNA processing enzymes involved in DNA modification, Okazaki strand resolution, nucleotide excision repair, and mismatch repair. Although there is only minimal sequence identity among species, sliding clamp quaternary structures among species is strikingly similar. Most sliding clamps have a quaternary structure in the shape of a toroid (Figure 1.3), although some viruses such as Herpes Simplex virus UL42 (Randell and Coen, 2004) do not. The central channel of the toroid allows the movement of double-stranded DNA movement through the center. Crystal structures are available for the T4 sliding clamp (gp45) (Moarefi et al., 2000), RB69 sliding clamp (Shamoo et al., 1999), Archaeoglobus fulgidus sliding clamp (aPCNA), Pyrococcus furiosus sliding clamp (pfuPCNA) (Matsumiya et al., 2001), E. coli sliding clamp (β) (Kong et al., 1992), yeast sliding clamp (yPCNA) (Krishna et al., 1994), and human sliding clamp (hPCNA) (Gulbis et al., 1996). All species have a trimeric sliding clamp except for E. coli which has a dimeric sliding clamp. Also, all known sliding clamps are proportionate in size with an inner diameter of approximately 30 Å.
Figure 1.3 Crystal structures of sliding clamps from across kingdoms. All clamps shown are trimers with monomers color-coded (green, red and blue) with the exception of *E. coli* which is a dimer with monomers color-coded (red and blue) A hPCNA:p66(452:466) (PDB:1U76) B yeast PCNA (PDB:1PLQ) C *E. coli* beta clamp (PDB:1MMI) D Archaeal sliding clamp, (PDB:1GE8) E Phage RB69 Gp45 sliding clamp (PDB:1B77)
Figure 1.4 The human PCNA monomer bound to the p21-derived peptide. The p21-derived peptide is shown in red. Residues 119-134, the interdomain connector loop, are colored green. For clarity the C-terminus of hPCNA is colored blue. The 18 beta sheets as well as the 4 alpha helices of the hPCNA monomer are labeled using the original author's nomenclature (Glubis et al., 1996).

In addition, the oligomeric states of the sliding clamps have differential intrinsic stabilities. hPCNA can maintain a trimeric state at concentrations down to 100nM (K_d=21nM) (Yao et al., 1996). The β clamp of E. coli is much more stable and has a K_d of less than 60pM for its dimer, possibly due to its simpler oligomeric state (Yao et al., 1996). Interestingly, T4 gp45 is comparatively less stable as a trimer and has a K_d of
only 250nM (Yao et al., 1996). Without polymerase, T4 gp45 spontaneously dissociates whereas hPCNA and the β clamp must be removed from DNA by clamp loaders, which may explain why the association rate of the trimeric state of T4 gp45 is lower than the other species (Yao et al., 1996).

The structural interface of hPCNA monomers are held together by anti-parallel interactions, as shown in Figures 1.3 and 1.4, beginning between strands βD₂ and βl₁ (for nomenclature definitions see Figure 1.4). All together, these intermolecular interactions lead to a total of 9 three-stranded β-sheets which form a surface across the monomer-monomer interface. Each monomer also contains similar anti-parallel β-sheets near the interdomain boundaries that form a scaffolding for 12 α-helices that line the center of hPCNA’s ring in a manner that puts them almost perpendicular to the ring face. Also of importance is the long loop which spans from the N-terminus down to almost the C-terminus of each monomer known as the interdomain connector loop. Although no specific interactions have been discovered as yet by DNA with the center ring of PCNA, the inner lining of the central ring has an electropositive surface which complements that of the DNA which slides through it.

1.3 hPCNA’s Interactions and the PIP(PCNA Interacting Protein)-Box

hPCNA was originally discovered due to its role as the processivity factor in DNA replication (Burgers, 1988). Interactions with other proteins at the replisome have also been identified by many means including yeast two-hybrid screens, peptide scanning, deletion mutants, and random as well as site-directed mutagenesis. As shown
in Table 1.1, proteins found to interact with hPCNA include polymerases, DNA repair proteins, DNA modification enzymes, the clamp loader, cell cycle control proteins, DNA ligases, transposases, helicases, and apoptosis proteins (Reviewed in Maga and Hubscher, 2003). The recruitment of these PCNA-interacting proteins is mediated largely through a conserved motif usually located within the C-terminal region of the protein (Warbrick, 1996). The conserved protein-protein interaction motif has been coined the "PCNA Interacting Protein" or "PIP"-Box. The PIP-box consensus sequence, $Q^{\text{xx(M/L/I)}^{\text{xxF(Y/F)}}}$, is well conserved among most kingdoms of life, as well as some viruses, which supports an important role for these interactions early in evolution (Hingorani and O'Donnell, 2000; Klemperer et al., 2000). The PIP-box is located near the C-terminus of many proteins, including flap endonuclease (FEN1), the p66 subunit of pol-$\delta$ (p66), cdk-inhibitor p21$^{\text{CIP1}}$ (p21), and DNA polymerase $\eta$ (Harper et al., 1993; Jonsson et al., 1998; Kannouch et al., 2004; Warbrick, 2000). Although the PIP-box is commonly found near the C-terminus, other proteins, including the clamp loader RFC, DNA-cytosine-5-methyltransferase and WRN DNA helicase, have internal PIP-boxes. Deletion or mutation of the conserved residues within the PIP-box can abrogate PCNA interactions in vivo and in vitro (Eissenberg et al., 1997; Jonsson et al., 1998; Stucki et al., 2001; Zheleva et al., 2000).
| Table 1.1 Alignment of known human PIP-box sequences¹ |
|-----------------|------------------|
| p21             | QTSMTDFY         |
| pol β           | QLQKV-HF         |
| pol δ           | QVSITGFF         |
| pol ι           | SRGVLSFF         |
| pol η           | Q-TLESFF         |
| pol κ           | KHTLDIFF         |
| pol λ           | SVPVLELF         |
| WRN             | QWKLLRDF         |
| RecQ            | QNLIRHFF         |
| XPG             | QLRIDSFF         |
| MSH6            | QSTLYSFF         |
| MSH3            | QAVLSRFF         |
| MCMT            | QTTITSHF         |
| RF-C            | MDIRKFF          |
| LigI            | QRSIMSFF         |
| Topo-IIα        | QTTLAFKP         |
| FEN1            | QGRLDFFF         |
| UNG2            | QKTLYSFF         |
| ING1            | QLHLVNYV         |
| Tigger2         | QTSLLSYF         |

¹Conserved PIP-box residues are highlighted in gray.
1.4 DNA polymerases

Most knowledge pertaining to polymerase function was elucidated in the *E.coli* system. Hence, human polymerases are classified by their catalytic domain homology to *E.coli* polymerases (Ito and Braithwaite, 1991; Aravind and Koonin, 1999; Friedberg et al., 2000). The major *E. coli* polymerases are polymerases I-V. Human polymerases specific to the mitochondrion, polymerases γ and θ, are classified as family A, due to their similarity to *E. coli* polymerase I (Sharief et al., 1999). The major human replicative polymerases α, δ, and ε as well as translesion polymerase ζ are classified as family B due to their homology to *E. coli* polymerase II (Nelson et al., 1996). The polymerase domains of family B polymerases are highly conserved and some members share well conserved exonuclease and zinc-finger domains (Ito and Braithwaite, 1991). No human homologs to *E. coli* polymerase III exist and therefore there is no family C. Human polymerases with homology to *E. coli* polymerases IV and V are classified in the super-family Y. Human translesion polymerases η, ι, and κ belong to superfamily Y (Ohmori, 2001). Historically, eukaryotic polymerases with no homology to *E. coli* were classified into family X. Human polymerases in family X include polymerase β, polymerase λ, and polymerase μ.

To date, there is only one atomic level structure of a human polymerase, pol β (Saway et al., 1997). Several B family polymerase crystal structures exist for RB69 and several Archaeal polymerases (Wang et al., 1997; Zhao et al., 1999; Hopfner et al., 2001; Rodriguez et al., 2000). In addition, a crystal structure exists for the Y family *E. coli* polymerase IV/Dinb (Bunting et al., 2003). The crystal structures of polymerases
show that they have similar domains that can be likened to a right hand: the thumb, palm, and fingers domains (for review see Brautigan and Steitz, 1998). The palm domain is well conserved as is the catalytic core of the enzyme. The palm contains two conserved aspartates important for catalysis. DNA is cradled in the palm and it is the site of nucleotidyl transfer. The fingers domain interacts with and positions the template DNA, as well as the incoming dNTP. The thumb domain binds the duplex DNA in a non-sequence dependent manner along the minor groove. Polymerases have been shown to use a two metal cation reaction mechanism using two aspartates to coordinate two magnesium ions. One of the Mg ions activates the 3'-hydroxyl group of the primer to nucelophilically attack the α-phosphate of the dNTP leading to pyrophosphate cleavage. The second metal ion stabilizes the negative charge on the leaving oxygen and chelates the β and γ phosphates which are the leaving group. This one nucleotide strand elongation is followed by translocation of the polymerase.

The functions of the human replicative enzymes ε and δ have been well studied in recent years but many questions still remain. Although the number and molecular weights of the subunits of pol δ show much degeneracy among species, their functions are well characterized (Gerik et al., 1998). Pol δ contains an active 3'-5' exonuclease proofreading domain and 5'-3' polymerization activity (Johnson, 1993). In addition, pol δ contains a PIP-box within the C-terminal region of its p66 domain and another putative non-PIP-box region in the N-terminus of the p125 subunit which may contact hPCNA (Zhang et al., 1999; Reynolds et al., 2000). hPCNA in the presence of pol δ increases it's processivity up to 100-fold over that in the absence of hPCNA (Prelitch et al., 1987). In contrast, pol ε does not receive any increase in processivity in the presence of hPCNA.
(Syvaoja and Linn, 1989). Interestingly though, hPCNA increases the association rate of pol ε to the DNA primer, and increases the nucleotide incorporation rate, but does not affect the dissociation rate from DNA (Maga and Hubscher, 1995). Pol ε has a higher inherent processivity and fidelity than does pol δ (Chui and Linn, 1995). Although pol ε has been shown to interact with hPCNA and its contact points have been mapped, pol ε contains no conserved PIP-box. It is still unclear as to the exact role that pol δ and pol ε play together at the replisome. Several models have been proposed, but no clear evidence has been presented which explains how pol δ and pol ε coordinate replication of the leading and lagging strands (Burgers, 1991; Morrison et al., 1990; Zlotkin et al., 1996).

Recent studies have discovered and characterized many polymerases required for DNA synthesis in the presence of DNA damage. Polymerase η can polymerize through several DNA lesions including cis-syn thymine-thymine dimers and 5'-TC-3' and 5'-CC-3' cyclobutane dimers (Washington et al., 2000; Yu et al., 2001). In contrast, polymerase τ cannot replicate through such lesions, but can incorporate nucleotides opposite the 3'-T of the (6-4) T-T photoproduct and opposite of abasic sites (Johnson et al., 2000). Polymerase κ cannot replicate through any of the aforementioned lesions, but is capable of replicating through a guanine-AAF (N-2-acetyl-aminofluorine) adduct as well as several poorly characterized lesions (Johnson et al., 2000; Ohashi et al., 2000).

Polymerase β has also been implicated as mainly functioning as a DNA repair polymerase. Pol β is involved in short-gap filling DNA synthesis which has been shown to be important in base excision repair (Sobol et al., 1996). Of all the translesion DNA polymerases, polymerases β, τ, η, and κ have been shown to have PIP-boxes (Maga et al., 2002; Haracska et al., 2002; Haracska et al., 2001).
1.5 Flap endonuclease-1 (FEN1)

One of the key PCNA interacting proteins is FEN1. FEN1 is required for proper DNA replication and repair. In mammals FEN1 is an essential gene and in lower organisms such as yeast, FEN1 deletions cause defects in replication and repair (Kucherlapati et al., 2002; Larsen et al., 2003). FEN1 is a structure specific 5' endo/exonuclease. The endonuclease activity of FEN1 recognizes double-stranded DNA that is branched with a single stranded flap and removes the flap endonucleolitically (Harrington et al., 1994). Schematic definitions of flaps can be seen in Figure 1.5. Both DNA and RNA flaps can be removed by FEN1. FEN1 can also act as an exonuclease whereby it recognizes nicks and gaps in DNA and can degrade 5' nucleotides one by one (Murante et al., 1994). On a cellular level, the nuclease activity of FEN1 is essential in two processes: Okazaki fragment processing and flap removal during base excision repair (Liu et al., 2004).

DNA synthesis is semi-discontinuous with one strand being synthesized in fragments of 100-150 nucleotides in length, known as Okazaki fragments, and must be removed of their RNA primers and joined by DNA ligase I. DNA synthesis of one Okazaki fragment meets the next Okazaki fragment displacing its 5' end region. This single stranded 5' region is known as a flap (defined in Figure 1.5) and contains the original RNA primer. FEN1 removes this primer with its nuclease activity. RNase H removes most of the RNA primer but leaves one residual ribonucleotide which is cleaved by the exonuclease activity of FEN1 (Turchi et al., 1994). The nicked DNA is then joined by DNA ligase I (Turchi et al., 1994). In the absence of RNase H,
Figure 1.5 Various DNA secondary structure elements

A Blunt duplex  B 5' single-stranded tail  C 3' single-stranded tail  D 3' flap  E 5' flap  F Bubble  G Fork  H 3-way junction  I Holiday junction  J G4 tetraplex
FEN1 is still able to process the entire flap containing the RNA primer by means of its endonuclease activity.

DNA base damage, caused by processes such as alkylation or oxidation, is predominantly corrected by base excision repair (BER) (Lindahl, 1993). BER is a process in which a single base is removed and corrected. DNA glycosylase cleaves the glycosidic bond of the damaged base to the sugar phosphate backbone. This creates what is known as a apurinic or abasic (AP) site. Consequently, AP endonuclease cleaves the AP site on the other 5' end, cleaving the DNA backbone. The resulting gap is filled in by polymerase β creating a flap of 2 to 12 nucleotides (Matsumoto and Kim, 1995). FEN1 removes the DNA flap endonucleolytically and the nick is joined by DNA ligase I.

FEN1 is localized at the replisome by its PIP-box which interacts with hPCNA (Chen et al., 1996; Gomes and Burgers, 2000; Jonsson et al., 1998; Wu et al., 1996). hPCNA stimulates FEN1 endonuclease and exonuclease activity by 5-10-fold. The mechanism of hPCNA's FEN1 stimulatory property is similar to that of increasing polymerase δ activity. FEN1 is held onto its substrate by PCNA rather than changing any of its catalytic characteristics. hPCNA reduces the $K_m$ value of the FEN1 nuclease activity by 11-12-fold (Samson et al., 2000).

FEN1 has also been implicated in less well characterized processes. Some of these processes include maintaining telomere stability, nonhomologous DNA end joining, maintaining genome stability, and resolving Holiday junctions (defined schematically in Figure 1.5) by its interaction with the Werner protein (for a review see Liu et al., 2004).
Figure 1.6  Crystal structure of *A. fulgidus* FEN1 (green) bound to a partial DNA substrate (red). The DNA substrate lacks the 5' flap and additional double-stranded DNA region of a true FEN1 5' flap substrate. The DNA is ~25 Å from the active site; the double-stranded DNA is thought to position the 5' flap into the active site for cleavage with the flexible helical clamp region clamping the flap into the active site (Chapados et al., 2004).

Although no human FEN1 crystal structure exists, several crystal structures of FEN1 homologs exist, as well as one crystal structure with FEN1 bound to a partial, double-stranded DNA substrate. FEN1 consists of 3 domains: the N-terminal domain (N), the intermediate domain (I), and the C-terminal domain (C). The N and I domains of FEN1 contain the elements responsible for substrate binding and catalysis while that of
the C domain contains the sequences responsible for interacting with other proteins including the PIP-box and the WRN protein interaction sequence. The FEN1-partial DNA substrate is shown in Figure 1.6. The 5' flap inserts onto the surface of FEN1, a conformational change occurs, and a flexible helical clamp closes on top of the flap initiating the nuclease reaction. The FEN1 nuclease reaction is a 2 metal ion phosphoryl transferase reaction similar to that of polymerases, phosphatases, and nucleases (Tock et al., 2003).

1.6 The Werner protein

Werner syndrome (WS) is a disease characterized by symptoms of premature aging resulting from mutation in a single gene product known as the Werner protein or WRNp. Werner syndrome was originally classified by its clinical phenotypes including skin atrophy, graying hair, the subcutaneous deposition of fat on the trunk, osteoporosis, bilateral cataracts, type II diabetes mellitus, as well as malignancy and myocardial infarction (Epstein et al., 1966). Cells from patients with WS have been found to have many cellular defects such as deletions or rearrangements in their chromosomes, an extended S phase, and defects in homologous recombination (Chen and Oshima, 2002).

The WRN gene product is a 1432 amino acid protein with several domains including a 3'-5' and 5'-3' exonuclease domain, a transcriptional activation domain, an ATP-dependent 3'-5' RecQ helicase domain, a nuclear localization domain as well as a PIP-box (Gray et al., 1997; Huang et al., 1998; Rodriguez-Lopez et al., 2003). The helicase action of WRNp starts by the binding of WRNp to the single stranded portion of
a DNA overhang and then moves to unwind the double-stranded DNA. The helicase action moves in the 3' to 5' direction in an ATP-dependent manner. WRNp can also unwind DNA/RNA duplexes in a similar manner. The helicase activity of WRNp is enhanced by single-stranded binding proteins such as the human single-stranded DNA binding protein RPA (Brash et al., 1999). The WRNp has also been shown to destabilize G quartets (defined schematically in Figure 1.5) as well as promote branch migration of Holiday junctions. The WRNp is different from other proteins in the RecQ family in that it has an exonuclease function. No ATP is required for the exonuclease function, but ATP does stimulate nuclease activity (Shen and Loeb, 2000). The exonuclease activity has been shown to start from the 3' end of blunt ended DNA as well as the 3' overhang of double-stranded DNA. In addition, Holiday junctions are even more susceptible to the exonuclease activity (Shen and Loeb 2000).

The exact cellular functions of the WRNp are still somewhat uncertain, but its primary function appears to be in a variety of DNA metabolic processes. WRNp has been shown to interact with PCNA, topoisomerase I, and pol δ (Chen and Oshima, 2002). The WRNp stimulates the activity of pol δ, but pol δ does not stimulate the activity of WRNp (Kamath-Loeb et al., 2000). The WRNp helps to enable pol δ to polymerize through secondary elements such as hairpins and G2 bimolecular tetraplexes (Kamath-Loeb et al., 2001). The large complex Ku-DNA-PK, responsible for non-homologous end joining of double-stranded DNA breaks, has been shown to interact with the WRNp. The role of the WRNp in non-homologous end joining may be to unwind broken ends of double stranded DNA as well as to process the ends with its exonuclease activity in order for the ends to be rejoined. In addition, the WRNp has been localized to telomeres and
cells with mutations or deletions of the WRNp lack telomeres with proper lengths (Johnson et al., 2001). Telomere repeats have been shown to be unwound by the WRNp and the WRNp has been shown to interact with telomere processing proteins TRFI and TRFII. Finally, the WRNp has been shown to have a role in cell cycle control by means of its interaction with p53. p53 has been shown to physically interact with the WRNp and cells lacking the WRNp have more attenuated p53 mediated apoptosis (Spillare et al., 1999; Blander et al., 1999).
There are five human helicases discovered so far in the RecQ family: BLMp, WRNp, RecQ1, RecQ4, and RecQ5. Although it appears that the RecQ helicases function predominantly in preventing excessive recombination in maintaining genome stability, they also participate in a wide host of functions in DNA metabolism (Opresko et al., 2004). Several of the human RecQ helicases when mutated result in clinical disorders...
such as BLMp resulting in Bloom syndrome, WRNp resulting in Werner syndrome, and RecQ4 resulting in Rothmund-Thomson syndrome: these diseases are characterized by premature aging and predisposition to cancer (Hickson, 2003). The helicases share a conserved helicase domain and have received the term "RecQ" from the first discovered RecQ helicase of the family in *E. coli*, RecQ. The helicase activity of the RecQ helicases works in a 3' to 5' direction and requires the hydrolysis of ATP. Helicases in this family are capable of resolving blunt or forked termini, duplexes with single-stranded overhangs, D-loops, and Holiday junctions, with each of the particular helicases resolving particular structures (see Figure 1.5 for structure schematic definitions) (Harman and Kowalczykowski, 1998). Most of the RecQ helicases share a low processivity, usually less than 100 base pairs (Bennet and Keck, 2004). RecQ family helicases not only share the conserved RecQ helicase domain but also a RecQ-Ct (RecQ C-terminal) domain and a HRDC (Helicase and RNase D C-terminal) domain. The RecQ-Ct domain has been shown to be responsible for protein-protein interactions as well as DNA interactions while the HRDC domain has been shown to be responsible for DNA binding interactions (Marozov, 1997; Bennet and Keck, 2004). The HRDC receives its name by means of its homology to the RNase D family of nucleases. The RecQ-Ct and HRDC domains are not necessary for catalysis.

RecQ helicases have been shown to be important in homologous recombination. Homologous recombination often occurs when DNA is damaged or broken, stalling or breaking replication forks. The RecQ helicases ensure proper strand cross-over and are therefore required for replication restart (Opresko et al., 2004). In addition, RecQ proteins make interactions with proteins such as polymerases which allow for the
relaxing of the secondary structure elements necessary for replication and recombination (Kamath-Loeb et al., 2001). In a like manner, RecQ helicases resolve G-quadruplex DNA, often found in telomeres, which allows for the proper maintenance of telomere length.

Consistent with most RecQ helicases, RecQ5 has recently been shown to unwind specific DNA structures. RecQ5 was shown to preferentially unwind a 3' flap, a three-strand junction and a three way junction whereas it could not unwind a blunt duplex, a Holiday junction, a 5' flap, or a 12 nucleotide bubble which have been shown to be unwound by other RecQ proteins such as BLMp and WRNp (Ozsoy et al., 2003). The types of structures unwound by RecQ5 are very similar to those found at the lagging strand of stalled replication forks. Therefore, it has been proposed that RecQ5 may be involved in the recombinatorial repair of stalled replication forks of DNA substrates which BLMp and WRNp cannot act upon. In addition, although RecQ5 has not been implicated in any human disease as of yet, it has been shown that RecQ5 deficiency in C. elegans results in reduced lifespan and increased sensitivity to radiation; RecQ5 appears to play an important role in recombination and repair which is just now beginning to be understood (Jeong et al., 2003).

Because RecQ5 has been relatively uncharacterized as yet, no information regarding interactions with hPCNA have been identified. The BLMp and WRNp have been shown to interact with hPCNA and contain putative PIP-boxes. Based on sequence homology, we have identified a putative PIP-box in the RecQ5 sequence (Table 1.1).

1.8 p21 and cell cycle control
Cell cycle control, regulation, and progression are driven by cyclin-dependent kinases (CDKs). CDKs are heterodimeric serine-threonine kinases whose activity promotes progression through particular cell cycle checkpoints (Pavletich et al., 1999). Different stages of the cell cycle are usually controlled by specific CDKs which act as on and off switches for the different stages of the cell cycle (Morgan, 1997). CDKs are activated by their interactions with cyclins and by phosphorylation, and are inactivated by CDK inhibitors. When cells have undergone extensive DNA damage, DNA replication is slowed in order to prevent the transmission of mutations. CDKs and CDK inhibitors control this passage by means of a G₁ to S phase check point (Peter and Herskowitz, 1994). When DNA damage reaches a threshold point, a tumor suppressor protein (p53) becomes elevated in expression and activates a signaling pathway to halt DNA replication (Cox, 1997). p21, a CDK inhibitor, is transcriptionally activated by p53; p21 halts initiation of the S phase of the cell cycle. p21 is a member of a family of CDK inhibitors whose structures share homology. Two classes of CDK inhibitors exist: the Cips such as p21 and p27 and the INK4s such as p15 and p16 (Pavletich, 1999). Cips act by binding to, and inhibiting, the entire active cyclin-CDK complex, whereas INK4s act by binding to and inhibiting only the nonactive isolated CDK. INK4s only participate in control of the G₁ cell cycle checkpoints, whereas the Cips participate in a broader range of cell cycle checkpoints. CDKs all share a common and conserved cAMP-dependent protein kinase active site. The CDKs also contain a more variable site for particular cyclin binding and activation. The cyclins usually bind to one side of the active site cleft causing a conformational change activating the kinase. Furthermore, phosphorylation of
a conserved threonine on a loop, causes further conformational changes activating the kinase.

Several CDK-cyclin complexes have been shown to interact with hPCNA (Zhang et al., 1993; Koundrioukoff et al., 2000). The significance of these interactions is still unclear. However, the p21 CDK inhibitor has been shown to have a clear function in its interaction with hCPNA. Not only does the p21 CDK inhibit DNA synthesis by inactivating cyclin-CDK complexes, but it also does so directly by binding to and competing for the same spot on hPCNA as the replicative polymerase δ (Chen et al., 1995; Gulbis et al., 1996). p21 contains a C-terminal PIP-box, whereas other CDK inhibitors do not, and mediates its interactions with hPCNA through the PIP-box. A short C-terminal peptide containing the PIP-box has been shown to be sufficient for binding hPCNA and blocking DNA replication (Chen et al., 1996).

1.9 Structure of p21(139-160) bound to hPCNA

Prior to our studies, there existed only one example of a crystal structure of the human sliding clamp: that of hPCNA bound to a peptide derived from the C-terminus of p21 containing the PIP-box (residues 139-160) (Gulbis et al., 1996). The overall shape of the hPCNA trimer is toroidal, resembling that of other sliding clamp crystal structures from other species (as seen in Figure 1.3). The interactions of p21 with hPCNA can be divided into three major regions: a C-terminal region that makes an anti-parallel beta-sheet with the interdomain connector loop as well as ionic interactions, a central 3₁₀ helix region that plugs into a hydrophobic cleft, and an N-terminal region that makes limited
hydrogen bond contacts with the C-terminus of hPCNA as well as containing several disordered residues.

Three of the four conserved residues of p21's PIP-box are located within the 3$_{10}$ helix that inserts into the hydrophobic core (Figure 1.8). Tyr-151 plugs deepest into the core, while at the same time making two hydrogen bond contacts: one with hPCNA residue Gln-131 and one with a water which also coordinates to residue Tyr-133 of hPCNA (Figure 1.8b). The other conserved aromatic residue of the PIP-box is a phenylalanine (Phe-150) and does not insert into the hydrophobic cleft but rather provides a barrier between the solvent and the hydrophobic cleft. The non-aromatic conserved hydrophobic residue of the PIP-box is a methionine (Met-147), which also inserts into the hydrophobic cleft.
Figure 1.8 Interactions of the $3_{10}$ helix of the p21-derived peptide with hPCNA. Hydrogen bonding interactions are shown with red dots. A hPCNA is shown with a surface representation and the p21-derived peptide shown as a stick representation. Conserved PIP-box residues Tyr 151 and Met 147 plug into a hydrophobic core of hPCNA. B Conserved PIP-box residue Tyr 151 hydrogen bonds with hPCNA Gln 131 as well as a water molecule (Gulbis et al., 1996).

The N-terminus of p21 was not well resolved in the crystal structure: the first five residues of the p21 peptide were disordered to the point that they could not be modeled. The only non-hydrophobic residue of the PIP-box, Gln-144, is also located in the N-terminal region and makes two sets of hydrogen bonds from its side-chain: one to residue Ala-252 of hPCNA and one to a water which in turn coordinates to residues Ala-252 and Ala-208 of hPCNA (Figure 1.9A). The only other interaction of the N-terminal region of p21 of note is that of Thr-145 which hydrogen bonds to hPCNA Pro-253 (Figure 1.9B).
Figure 1.9 N-terminal interactions of the p21-derived peptide with hPCNA. Hydrogen bond interactions are shown with red dots. A Conserved PIP-box residue Gln 144 of p21 makes one hydrogen bond with hPCNA main-chain carbonyl Ala 252 and one hydrogen bond with a water B Thr 145 of p21 makes a hydrogen bond interaction from its side-chain oxygen with hPCNA main-chain carbonyl Pro 253 (Gulbis et al., 1996).

The C-terminus of p21 makes an extensive anti-parallel beta-sheet of 9 residues (residues 152-160) with the interdomain connector loop of hPCNA (residues 119-127) (Figure 1.9A). There are also two arginine residues of p21 making ionic interactions with hPCNA: Arg 155 making contacts with Asp-122 of hPCNA and Arg-156 making contacts with hPCNA Gln-125 and Asp-29 (Figure 1.9B). In addition, the side-chain of Ile-158 is making a hydrophobic contact with a hydrophobic pocket within hPCNA. Finally, residue Ser-160 is involved in a hydrogen bond from its side-chain to hPCNA Gly-69 (Figure 1.9E).
Figure 1.9 C-terminal interactions of the p21-derived peptide with hPCNA. Hydrogen bonds are shown as red dots and ionic interactions are indicated with + and - symbols. A The p21-derived peptide is shown in red and the interdomain connector loop of hPCNA is shown in blue. The C-terminal region of p21 is making an antiparallel beta-sheet with hPCNA. B The side-chain of Ile 158 plugs into a hydrophobic core of hPCNA (shown as a surface representation). C Arg 156 side-chain of p21 makes an ionic contact with hPCNA Asp 29 as well as a hydrogen bond with hPCNA Gln 125 side-chain. D p21 Arg 156 side-chain makes an ionic contact with hPCNA Glu 124 side-chain as well as a hydrogen bond with hPCNA Asp 122 side-chain. E Side-chain of p21 Ser 160 makes a hydrogen bond with hPCNA main-chain atoms of Met 119 and Gly 69 (Gulbis et al., 1996).
1.10 Mutational and biochemical analysis of PIP-box protein's interaction with hPCNA

Biochemical data on human p66 has been historically lacking due to its difficulty of purification. Not until 2001, when Ducoux et al., used a refolding technique for p66 purification, was purified p66 shown to directly interact with PCNA in a pull-down assay (Ducoux et al., 2001). In the same pull-down assay, when the PIP-box was deleted from p66, the interaction was abrogated, showing that the p66 PIP-box is necessary for binding. This was further supported by studies that showed a p66-derived peptide containing the PIP-box is able to inhibit PCNA activation of pol δ in activity assays (Zhang et al., 1998). In the yeast system, which is well conserved with the human system, pull-down assays were used to show that the PIP-box of pol δ not only to be necessary, but also sufficient for binding to PCNA (Reynolds et al., 2000). Although there are no direct binding studies with p66 and PCNA, the human sliding clamp stimulates processive DNA replication by approximately 10-100-fold (Burgers 1988).

The interaction between FEN1 and PCNA was initially shown in a yeast two-hybrid system as well as a far Western experiment (Warbrick et al., 1997). In addition, a p21-derived peptide containing the PIP-box was shown to disrupt the FEN1-PCNA association (Warbrick et al., 1997) with concentrations that indicated the p21 interaction is stronger. PCNA has been shown to stimulate FEN1 cleavage 5-50 fold (Wu et al., 1996; Li et al., 1995). In the absence of DNA, purified yeast FEN1 was shown, by means
of surface plasmon resonance, to have a $K_D$ of 0.02 $\mu$M, although accurate equilibrium values using surface plasmon resonance can often be difficult to quantitate due to the chip surface effects (Gomes et al., 2000).

In addition, the PIP-box of p21 has been shown to be necessary and sufficient for \textit{in vivo} function. A p21-derived peptide containing the PIP-box (residues 141-160) binding to hPCNA was shown to have a $K_d$ of 87.7 nM which is only slightly higher than that of the full-length p21 which was shown to have a $K_d$ of 15 nM (Zheleva et al., 2000). Zheleva et al., also showed that mutation of PIP-box residues to alanine greatly reduces or completely inhibits p21 peptide interaction with hPCNA (Zheleva et al., 2000).

To date, biochemical characterization of the hPCNA:p21 interaction is comprehensive, while in contrast hPCNA interactions with proteins that require DNA for function are not. Biochemical binding studies are complicated and haven't been thoroughly investigated due to many factors: FEN1 and p66 also interact with DNA, the replicative polymerase is composed of several subunits, and p66 is difficult to purify. PCNA requires a clamp loader and, once loaded, slides along the DNA. Proper binding studies of FEN1 and p66 would first require that PCNA is loaded onto DNA with a clamp loader, allowing for more than a simple one step equilibrium interaction.
Chapter 2: Preparation of hPCNA and PIP-box peptides for crystallization

2.1 Cloning of the hPCNA expression plasmid.

cDNA encoding hPCNA (Accession NM_002592) was obtained from the NCBI I.M.A.G.E. consortium and used as a template for PCR amplification. Oligonucleotide primers (5’ATTGAGCTCAGATGTGTTGAGGCGC-3’) and (5’ATGGGATCCCTAAGATCCCTTCTCATCCT-3’) were used in the PCR reaction. The PCR reaction was a 100 µL reaction consisting of 20 mM Tris-HCl pH 8.8, 10 mM KCl, 10 mM (NH)₂SO₄, 2 mM MgSO₄, 0.1% Triton X-100, 1.85 mM MgOAc, 250 µM dNTPs, and 5 units total of Taq polymerase. The reaction began with a 5 minute 94°C melt followed by 25 cycles of a 94°C melt, a 52°C annealing, and a 74°C DNA polymerization; the reaction was completed with a 7 minute 74°C DNA polymerization to ensure complete extension of duplicated DNA strands. The DNA was cleaned with PCR Quickclean (Qiagen). 1 µg of the PCR product was digested with 20 units BamHI and 10 units BspHI in a reaction containing 50 mM NaCl, 10 mM Tris-HCl, 10 mM MgCl₂, and 1 mM DTT; the reaction buffer was at pH 7.9. The reaction was stopped after 2 hours at 37°C and the DNA was cleaned with the PCR Quickclean (Qiagen). Plasmid pET-28a (Novagen) was digested with the same enzymes using the same conditions. The digested DNA insert was ligated into the digested pET-28a using the following conditions: 50 mM Tris-HCl (7.5), 10 mM MgCl₂, 10 mM DTT, 1 mM ATP, and 400 units T4 DNA
ligase; the reaction was carried out at room temperature for 16 hours. The ligated DNA was cleaned and the clone is here-to-after referred to as pET-28a-hPCNA. pET-28a-hPCNA was transformed into *Escherichia coli* BL21 (DE3) using competent cells (Biorad) by means of a calcium chloride heat shock protocol. The clone was sequenced by Lone Star Labs to verify insertion of the correct hPCNA DNA sequence.

2.2 Cell growth and protein expression

Because no protocols existed for expression of pET-28a-hPCNA, optimal conditions for cell growth and protein expression had to be explored. Several small scale expression experiments of pET-28a-hPCNA in BL21 (DE3) were carried out in order to optimize such variables as IPTG concentration, time duration of expression, and type of media. Small scale cell growths of 50 mL were used, inoculating from a -80°C glycerol stock. Cells were induced with IPTG when the optical density reached an absorbance of 0.9 units at a wavelength of 590 nm. IPTG concentrations tested were 0.5 mM, 0.75 mM, and 1 mM. Induction times tested were 2 hours, 4 hours, 6 hours, and 16 hours. Media types tested were LB, 2xYT, and terrific broth. The antibiotic was kept constant at 50 μg/mL kanamycin. All cell growths were grown at 37°C and shaken at 250 r.p.m. 0.5 mM IPTG was decided to be the optimal concentration for induction as determined by hPCNA band size on SDS-PAGE. Important points for consideration when preparing protein for crystallography are protein yield and the presence of proteolysis products. These points were assayed by means of the hPCNA bands on SDS-PAGE gels of varying time points during protein expression. Optimal yield was achieved with longer
expression time. Even after 16 hours of protein expression, no visible proteolysis products could be seen on the SDS-PAGE gels. Therefore, the optimal condition for cell growth and expression was determined to be 0.5mM IPTG induction for 16 hours at 37°C in 2xYT media being shaken at 250 r.p.m.

Small scale protein expression conditions do not always coincide with large scale protein expression conditions. To this end, large scale growth conditions were tested using small scale growth conditions mentioned above as a starting point. 50 mL of 2xYT were inoculated from a frozen glycerol stock and grown 16 hours at 37°C and 50 μg/mL kanamycin; this was used as a starter culture. The large scale growth culture consisted of 5 2L cultures grown in 4L baffle flasks inoculated with 5mL starter culture into each flask. Cells were grown at 37°C shaken at 250 r.p.m. until reaching an optical density of 0.9 at 590 nm absorbance at which time they were induced with 0.5 mM IPTG. Cells were induced 16-18 hours. Cells were then pelleted by centrifugation at 5g. The cell yield was on the order of 80g. Cell yields were variable when large scale growths were reproduced and cell yields varied from 40-80g. Cell pellets were aliquoted into 15-20g aliquots and stored at -80°C.

2.3 Previously published purification of hPCNA

Prior to our work, hPCNA had been crystallized with a PIP-box peptide (Gulbis et al., 1996). Our initial aim was to use the Gulbis et al purification protocol for hPCNA in our work to ensure crystallizability. Possibly due to publication space restrictions, the published purification protocol was limited in description, "human PCNA was purified
by standard methodologies, using chromatography on Q Sepharose Fast Flow and Heparin Sepharose (Pharmacia)." In our hands, following this limited protocol did not produce pure or crystallizable protein. Upon personal communication with the author, we were able to obtain a more detailed purification protocol (Kelman, 1996). Throughout purification, hPCNA buffer contained 10% glycerol, 0.5mM EDTA, and 2mM DTT. After cell lysis, hPCNA was loaded onto Q sepharose resin at pH 7.5 and eluted with a gradient from 100 mM to 800 mM NaCl. Fractions containing hPCNA were then dialyzed to low salt buffer with sodium acetate at pH 7.5. This was loaded onto heparin sepharose and hPCNA flowed through. The protein solution was then brought to pH 5.5 by adding acetic acid directly to the solution. Finally, the sample was reapplied to the heparin column and washed with 150 mM NaCl and eluted with no gradient at 300 mM NaCl. The protein was then concentrated by applying it to a Q Sepharose column and bumping it off with 1.5 M NaCl. Unfortunately, this protocol did not produce hPCNA that would crystallize. After purification using this protocol, the protein appeared over 90% pure by SDS-PAGE but was contaminated by nucleic acid as determined by 280/260 absorbance ratios well below 1.

Due to the inability to produce purified hPCNA which could be crystallized using published protocols, we searched for columns and procedures of our own that would produce crystal quality hPCNA. One issue of importance was the ability to remove DNA from the preparation. hPCNA has an affinity for nucleic acid that allowed nucleic acid to associate with hPCNA throughout purification. Thus, a large portion of our experimental design in hPCNA purification was directed to finding methods to separate hPCNA from nucleic acid. Several columns were experimented with in hPCNA purification as well as
salt concentrations, pH's, gradient lengths, additives, dialysis method, and protein
concentration methods. Columns tried include phosphocellulose P11 (Whatman), DEAE
DE52 (Whatman), Source 15 Q (Pharmacia), Q sepharose (Pharmacia), Mono Q
(Pharmaica), Heparin sepharose (Pharmacia), Superdex 200 (Pharmacia), Hydroxy
apetite, Phenyl sepharose (Pharmaica), and SP sepharose (Pharmacia). Only the columns
and time schedule arrived at as the best purification method will be described in the next
section.

2.4 Purification of hPCNA

15g of cells were removed from the -80° freezer and resuspended in 100 mL of
20mM Tris-HCl 7.5, 50 mM NaCl, 10% sucrose, 2 mM DTT, 0.5 mM EDTA, 0.2 mM
PMSF, and 30 mM spermidine. All purification buffers from this point forward
contained 2mM DTT, 0.5mM EDTA, and 0.2mM PMSF. Spermidine was added for its
polycaionic nature and its ability to bind and pack nucleic acid; spermidine may help
sequester nucleic acid from hPCNA. Cells were resuspended on ice. Cell lysis was
carried out by sonication using a Ultrasonics Heat Systems sonicator model w-375. Cells
were sonicated using a 50% duty cycle for 1 minute carried out 4 times with manual
stirring in between. Following sonication, insoluble material was pelleted by
centrifugation at 15,000 g for 45 minutes.

Soluble material was applied to 10 mL of phosphocellulose P11 resin (Whatman)
loaded within a column (23 cm in length and 3 cm in diameter). The column was
equilibrated in 20 mM Tris-HCl 7.5 and 0.1M NaCl. The P11 resin was washed with 5
column volumes (50 mL) equilibration buffer. Solution containing hPCNA flowed through the column and was collected.

Material containing hPCNA which flowed through the P11 resin was applied to a DEAE DE52 (Whatman) column (23 cm in length by 3 cm in diameter) equilibrated in 20 mM Tris-HCl 7.5 and 0.1M NaCl. The column was washed with equilibration buffer until the 280 nm absorbance was below 0.5. A linear NaCl gradient was applied (0.1-0.5M NaCl) and 3-5 mL fractions were collected. Fractions were analyzed by absorbance at 280 and 260 nm as well as SDS-PAGE. Protein fractions contained large amounts of nucleic acid contamination as judged by the 280/260 ratio (well below 1) and the SDS-PAGE showed numerous contaminants. Fractions containing hPCNA were pooled.

![SDS-PAGE of DE52 elutions](image)

**Figure 2.1** SDS-PAGE of DE52 elutions (indicated with arrows). Red arrows indicate fractions containing hPCNA pooled for further purification while black arrows indicate elution fractions not pooled for further purification. Molecular weight standards are shown on the right (kDa)

Analytical experiments showed that hPCNA solutions below 1mg/mL protein could tolerate ammonium sulfate concentrations over 1M without precipitation of hPCNA. After DE52 anion exchange chromatography, the hPCNA solution was brought
up to 1M ammonium sulfate by adding 4M ammonium sulfate solution slowly and drop-wise to the hPCNA solution. The hPCNA solution was then stirred gently at 4°C for 45 minutes. Next, the solution was pelleted by centrifugation for 10 minutes at 15,000 g. Little if any precipitate was visible. Finally, the supernatant containing hPCNA was filtered through a 0.22 micron syringe filter.

The filtered ammonium sulfate hPCNA solution was applied to a 10 mLHiTrap Phenyl Sepharose column (Pharmacia) equilibrated in 20 mM Tris-HCl 7.5, 1M ammonium sulfate, 0.5 M NaCl, and 10% glycerol (all buffers after this point in the purification protocol contained 10% glycerol). A 10 column volume linear gradient was applied extending from the equilibration buffer to 20 mM Tris-HCl 7.5 and 50 mM NaCl. Fractions containing hPCNA eluted between 0.75 M and 0.15 M ammonium sulfate. 2.5 mL fractions were collected and the flow rate was set at 2.0 mL per minute. Flow through fractions contained large amounts of nucleic acid and little or no hPCNA. Fractions containing hPCNA were pooled and dialyzed to 20 mM Tris-HCl 7.5, and 0.1M NaCl. Dialysis was carried out using a 6-8 kD molecular weight cut off dialysis bag. Dialysis consisted of three steps changing the buffer at each step: 1L for 3 hours, 1 L for 1 hour, and 1 L for 1 hour.
Figure 2.2 Phenyl sepharose FPLC chromatogram. Conductivity (brown line), %B (green line), and absorbance at 280 nm (blue) are shown. Fractions pooled for further purification are located between the 2 red arrows.

Dialyzed material was next filtered through a 0.22 micron syringe filter. hPCNA was loaded onto a 10 mL HiTrap Q Sepharose column (Pharmacia) equilibrated in 20 mM Tris-HCl 7.5 and 0.2 M NaCl. The flow rate used was 2.5 mL/min and 2.5 mL fractions were collected. A salt gradient was applied using buffer B (20 mM Tris-HCl 7.5, 0.8 M NaCl). The gradient went form 0-10% B in 1 column volume, 10-55% B in 10 column volumes, and 55-100% B in 3 column volumes. hPCNA eluted from the column at 0.24-0.4 M NaCl. Fractions containing hPCNA were found using SDS-PAGE and pooled. hPCNA was dialyzed to 20 mM Tris-HCl 7.5 and 0.1 M NaCl. Dialysis was carried out in three steps changing the buffer for each step: 1L for 1 hour, 1L for 45 minutes, and 1L for 45 minutes.
Figure 2.3 Q sepharose chromatography  

A Chromatogram of Q sepharose column with conductivity (brown line), %B (green line), and absorbance at 280 nM (blue) shown. Fractions pooled for further purification are located between the 2 red arrows. Molecular weight standards are located on the right (kDa).  

B SDS-PAGE of elution fractions (arrows). Red arrows indicate fractions containing hPCNA pooled for further purification while black arrows indicate elution fractions not pooled for further purification.
After filtration through a 0.22 micron filter, the material dialyzed from the preceding step was applied to a 5 mL HiTrap Heparin Sepharose column equilibrated in 20 mM Tris-HCl 7.5 and 0.1M NaCl. hPCNA flows through the column at this pH and was pumped over the column at 1 mL/minute. One large fraction was collected based on the 280 nm absorbance profile and SDS-PAGE verification. hPCNA was then dialyzed to 50 mM sodium acetate pH 5.5 and 0.1 M NaCl. The dialysis was carried out in three steps changing the buffer at each step: 1L for 1 hour, 1L for 45 minutes, and 1L for 45 minutes.

Next, hPCNA was applied to a 10 mL HiTrap Heparin Sepharose column (Pharmacia) equilibrated in 50 mM sodium acetate and 0.2 M NaCl. A 10 column volume salt gradient (0.2-0.4 M NaCl) was carried out at 2.5 mL/minute collecting 2.5 mL fractions. hPCNA was eluted at 0.35-0.4 M NaCl. Fractions were analyzed by SDS-PAGE and those containing hPCNA were pooled. hPCNA was then dialyzed to 20 mM Tris 7.5 and 50 mM NaCl. Dialysis was carried out in three steps changing the dialysis buffer at each step: 1L for 1 hour, 1L for 45 minutes, and 1L for 45 minutes.
The final column step in the purification was a Mono Q HR 10/10 column (~7-8 mL) (Pharmacia). Dialyzed hPCNA was filtered through a 0.22 micron syringe filter and loaded onto a Mono Q HR 10/10 column (Pharmacia) equilibrated in 20 mM Tris 7.5 and
50 mM NaCl. Protein was eluted with a salt gradient using buffer B (20 mM Tris 7.5 and 1M NaCl). The gradient was 0-30% B in 2 column volumes, 30-60% B in 10 column volumes, and 60-100% B in 1 column volume. hPCNA eluted at 0.4M NaCl. Fractions were analyzed by SDS-PAGE and those containing hPCNA were pooled. hPCNA was then dialyzed extensively to its storage buffer: 20 mM Tris-HCl 7.5, 10% glycerol, 0.5mM EDTA, and 2 mM DTT. Dialysis consisted of two steps changing the dialysis buffer at each step: 2L for 16 hours, and 1L for 4 hours.
Figure 2.5 Mono Q chromatography. A Chromatogram of Mono Q column with conductivity (brown line), %B (green line), and absorbance at 280 nM (blue) shown. Fractions pooled for further purification are located between the 2 red arrows. Molecular weight markers are shown on the right (kDa). B SDS-PAGE of elution fractions (arrows). Red arrows indicate fractions containing hPCNA pooled for further purification and black arrows indicate fractions not pooled for further purification.

As with cell growth, protein yield was also variable. Protein yield was on the order of at least 1.5-2 mg per gram of cells lysed, which is not surprising given the
number of purification steps. Protein purity was of good quality with only 1 or 2 trace contaminants as judged by SDS-PAGE with overloaded amounts of protein on the gels to highlight contaminants. Nucleic acid contaminants were also present in trace amounts as judged by 280/260 absorbance ratios usually on the order of 1.5.

The published purification protocol concentrated the hPCNA protein to 20 mg/mL for crystallography by bumping the protein off of a Q column with high salt (1.5 M NaCl) and then dialyzing. This proved challenging due to the inability to accurately adjust the concentration with column elution. This method also added dialysis time to the preparation. For these reasons, we chose to concentrate hPCNA by centrifugation using a 10,000 molecular weight cut off Vivaspin 20 mL concentrator (Vivascience). This allowed us to accurately concentrate the hPCNA sample to 20 mg/mL as judged by absorbance at 280 nm. hPCNA concentration was also verified later by amino acid analysis using a ABI 420A Analyzer at Baylor College of Medicine Department of Immunology Protein Chemistry Core Laboratory. Protein was aliquoted into 50 μL aliquots and flash frozen with liquid nitrogen and stored at -80° C. A small amount of sample from each preparation was not frozen, but kept at 4° C and used immediately for crystallography to ensure that oxidation and freeze/thaw would not be variables in crystal quality. In total, 14 separate preparations of hPCNA used solely for crystallography were carried out as well as 3 additional preparations of hPCNA used for crystallography as well as isothermal titration calorimetry.

2.5 Peptide production, purification, and quality analysis
All peptides used for our crystallographic and isothermal titration calorimetry studies were obtained from Tufts University Core Facility. Peptides were made with an ABI 431 peptide synthesizer which employs FastMoc chemistry. All peptides were produced on a 0.1 mmol scale. After synthesis peptides were purified by reverse-phase HPLC. For purification, 40 mg crude peptide was used and resulted in yields on the order of 10-25 mg purified peptide. Peptides were generally a minimum of 90% pure. Peptide purity was monitored by means of mass spec; mass spec was done with an Applied Biosystems Voyager DE pro using MALDI-TOF. Peptide purity was also monitored by means of analytical HPLC. After purification, peptides were lyophilized and shipped to us. Once in our hands, all peptides were resuspended in two separate samples: one just in water and the other in 10 mM Tris 7.5. Samples were brought up to a 10 mM concentration based on peptide powder weight. To verify the correct concentration of our peptide solutions, we sent them to Baylor College of Medicine Department of Immunology Protein Chemistry Core Laboratory for amino acid analysis using an ABI 420A Analyzer.
Figure 2.6  p21 peptide purification used for crystallography and ITC
(NH₂GRKRRQTSMTDFYHSKRRLIFS₇COOH) (Tufts University Core Facility)  A
Purified peptide HPLC chromatogram  B  Purified peptide mass spec analysis.
Figure 2.7 Purification of the p66a peptide \((\text{NH}_2\text{GKANRQVSITGFFQRK}_{\text{COO}OH})\) used for ITC (Tufts University Core Facility)  

A  HPLC chromatogram of peptide purification with horizontal dashes indicating the portion collected  

B  Purified peptide mass spec analysis.
Figure 2.8 Purification of the p66b peptide (NH$_2$ GKANRQVSITGFQKRKLIFS$_{COOH}$) used for ITC (Tufts University Core Facility). A HPLC chromatogram of peptide purification with horizontal dashes indicating the portion collected. B Purified peptide mass spec analysis.
Figure 2.9  Purification of the p66 peptide (NH$_2$KANRQVSITGFFQRKCOOH) used for crystallography (Tufts University Core Facility)  A  HPLC chromatogram of peptide purification with horizontal dashes indicating the portion collected  B  Purified peptide mass spec analysis.
Figure 2.10 Purification of the FEN1 peptide (\(\text{NHE}_{2}\text{SRQGSTQGRLDDFFKVTGSL}_{\text{COOH}}\)) used for crystallography (Tufts University Core Facility) A HPLC chromatogram of peptide purification with horizontal dashes indicating the portion collected  B Purified peptide mass spec analysis.
Figure 2.11 Purification of the FEN1 peptide (NH2RQGSTQGRLDFFKVTGLSALSACOOH) used for ITC (Tufts University Core Facility)  A HPLC chromatogram of peptide purification with horizontal dashes indicating the portion collected  B Purified peptide mass spec analysis.
Figure 2.12 Purification of the WRN peptide (NH2EGDQWKLLRDIFKLNFCOOH) used for crystallography and ITC (Tufts University Core Facility) A HPLC chromatogram of peptide purification with horizontal dashes indicating the portion collected B Purified peptide mass spec analysis.
Figure 2.13 Purification of the RecQ5 peptide (NH2VKEEAQNLIRHFFHGRCESECOOH) used for ITC (Tufts University Core Facility) A HPLC chromatogram of peptide purification with horizontal dashes indicating the portion collected B Purified peptide mass spec analysis.
Figure 2.14 Purification of the RecQ5 peptide (NH$_2$KEAAQNLIRHFFHGRCOOH) used for crystallography (Tufts University Core Facility) A HPLC chromatogram of peptide purification with horizontal dashes indicating the portion collected B Purified peptide mass spec analysis.
Chapter 3: Crystallization of hPCNA with interacting peptides

3.1 General strategy

Solutions of hPCNA and interacting peptides were made by mixing the respective purified solutions (described in Chapter 2) in a microcentrifuge tube. hPCNA-peptide solutions were equilibrated at either 4° C or 20° C, depending upon the temperature of the crystallographic experiment, for 15 to 30 minutes prior to the initiation of the crystallographic experiment. With certain exceptions, such as the Hampton Salt Rx screen (Hampton Research) using robot technology, all crystallization experiments were done by means of the hanging drop vapor diffusion method using a 24 well Limbro plate. Experiments used 0.5 mL of well solution and a variable amount of protein solution hanging on a silanized glass cover slip sealed with vacuum grease.

Several general strategies were employed in the design of crystallographic experiments. First, conditions were tried that were published for previously crystallized hPCNA. In addition, new crystal conditions were searched for. Searching for new crystal conditions consisted of two stages: broad screening of crystallization conditions in which the goal is an initial "hit" such as crystalline precipitate, spherulites, or crystals and a second stage which consists of more narrow optimization of such conditions. New crystal conditions were initially searched for by means of commercially available sparse matrix screens such as Hampton's Crystal Screen I, Crystal Screen II, Index screen, as well as the Salt Rx screen (Hampton Research). These screens employ screening conditions developed from historically successful conditions devised from a wide variety
of protein crystallization experiments from hundreds of previously crystallized proteins. In addition, initial searches for new crystal conditions consisted of using broad footprinting screens and pH matrixes using variable concentrations of commonly used precipitants such as ammonium sulfate, various molecular weight polyethylene glycols (PEGs), and methypentanediol (MPD) combined with various buffers at different pHs. Once initial crystalline precipitate, spherulites, or crystals were found, the conditions were optimized using additives (such as salts, divalent metals, small organic molecules, or sugars), detergents, different temperatures, fine pH grids, fine precipitant grids, different buffers, different peptide to protein ratios, more intense hPCNA purification, different drop sizes, different drop ratios (protein to precipitant volume ratios in the hanging drop), different crystallization apparatus (such as sitting drops or micro batch crystallization), streak-seeding, and macro-seeding.

Seeding is useful in increasing the size or quality of already obtained crystals: crystals are transferred to a new, already equilibrated drop to feed the crystals with more protein to continue crystal growth. Streak-seeding consists of touching a cat whisker to an old crystal and streaking a newly equilibrated drop: when the crystal is touched it picks up micro-crystals that when transferred to a new drop are seeds for crystal growth. Macro-seeding simply consists of washing already obtained crystals and transferring them to a new drop to feed more crystal growth. Because crystallization consists of two stages, nucleation (formation of the initial micro-crystals from which the crystal forms) and crystal growth, seeding is often successful when nucleation is very slow or difficult to reproduce. The type and number of strategies employed directly correlated to the ease
of crystallization of each particular hPCNA-peptide complex and will be described in more detail for each complex.

3.2 Proof of concept: recreating hPCNA-p21(139:160) co-crystals

Prior to our work, there existed only one published case of hPCNA crystals: that of hPCNA co-crystallized with a peptide derived from p21 (residues 139:160) (Gulbis et al., 1996). It is often the case, but not always, that proteins can be crystallized with different ligands or mutations under similar conditions. We thought this would likely be the case with our system due to the high homology of our PIP-box peptides with the p21 peptide used in the published structure. However, one troubling feature of the published hPCNA-p21 crystal structure was that the p21 peptide was involved in crystal contacts, allowing for the possibility that other PIP-box peptides would not make the same crystal contacts and, thus, not crystallize. Regardless, our first goal was to recreate the published hPCNA-p21(139:160) co-crystals as a proof of concept that our crystallization materials and methods were capable of crystallization. Protein crystallization conditions often differ from lab to lab due to variables in personal technique while handling protein samples during purification and crystallization. This proof of concept would also allow us to use this crystallization protocol as a template protocol for other PIP-box peptides.

The published crystallization protocol for the hPCNA-p21(139:160) co-crystals consisted of mixing a 1.5 molar excess of p21 peptide (\textsuperscript{139}GRKRRQTSMTDFYHSKRRLIFS\textsuperscript{160}) with hPCNA at 20 mg/mL. The crystallization experiment used the hanging drop vapor diffusion method at 4° C. 1 µL of hPCNA-
peptide complex was mixed with 0.5 μL 1 M spermine and 1 μL well solution. The well solution consisted of 100 mM sodium citrate (pH 5.6), 200 mM sodium/potassium phosphate, and 30% (v/v) 2,5-methylpentanediol (MPD). A second crystal form with perfect hemihedral twinning was also reported using the same protocol but using a well solution consisting of 100 mM MES (pH 6.0), 200 mM magnesium acetate, and 30% (v/v) MPD. Several aspects of the published crystallization protocol presented themselves as potentially troublesome to us. First, hPCNA was documented as having a propensity to form twinned crystals (discussed in detail in Chapter 6). Also, the non-twinned crystals report using 100 mM sodium citrate as a buffer and 200 mM sodium/potassium phosphate as a salt additive. The sodium/potassium phosphate is at a concentration twice that of the sodium acetate; because sodium/potassium phosphate also has buffering capacity it cannot possibly be buffered by the sodium citrate. The pH of the sodium/potassium phosphate was not reported nor was the final pH of the well solution. In addition, whether the phosphate was monobasic or dibasic was not reported nor was the ratio of sodium to potassium phosphate reported. Finally, the use of 1 M spermine seemed very high as it is usually used as an additive in crystallographic experiments at a concentration of 1 to 100 mM. Additionally, spermine has a buffering capacity and the pH of the spermine solution was also not reported.

Due to the strange nature of the crystallization conditions and the lack of published information on the physical morphology of the crystals, we contacted the original author. Communication with the author, Jackie Gulbis, revealed that the large concentration of spermine in the drop lead to crystallization of spermine. Spermine
crystallized as long spears. hPCNA crystallized as hexagonal plates with the spermine spears running through the center of the hexagonal hPCNA plates.

Obtaining hPCNA-p21(139:160) co-crystals proved challenging for many reasons. Quality crystals could not be obtained with the first six preparations of hPCNA. Altering the published purification protocol was necessary to obtaining crystals. No crystals formed until nucleic acid contamination was successfully removed from the hPCNA preparation. Spermine crystals formed in almost all drops making it difficult to identify hPCNA from salt crystals readily. Spermine crystals also made seeding difficult because spermine seeds would always be transferred to new drops. Finally, the published crystallization conditions were never recreated and new crystallization conditions had to be found in which the complex crystallized. In total, 1462 conditions were searched to find hPCNA-p21(139:160) co-crystals of similar morphology and quality as the published crystals.
The first crystallization trials consisted of trying to recreate the published conditions and the experiments were all carried out at 4°C. Trials were based on varying the conditions of MPD, the pH of sodium citrate (2.5-5.5), the buffer types, the concentration of buffer, the concentration and pH of sodium/potassium phosphate, and the concentration and pH of the spermine additive. In addition, Hampton additive screen 2 and detergent screen 1 and 3 were also tried (Hampton Research). Several other
additives were also tried such as sugars, divalent metals, small organic molecules, and spermidine in place of spermine. Crystals of spermine were readily obtainable but no hexagonal hPCNA crystals were obtained. Small and thin plates formed that stained blue with Izit dye (commercially available dye that stains protein crystals blue but not salt crystals such as spermine) (Hampton Research) but were too small for diffraction. Streak seeding also failed to form quality crystals.

Next, the search for novel crystallization conditions was carried out. The search began with sparse matrix screens such as the Hampton crystal screen I, crystal screen II, and Index screen (Hampton Research). Foot-printing screens using ammonium sulfate, PEG 350, PEG 4000, PEG 8000, MPD, and MPD with ammonium sulfate at pH's in the range of 2.5 to 9 were also employed. The screens were carried out at 4° C, 10° C, and 20° C. Spherulites were found in several ammonium sulfate conditions. Spherulites were optimized using combinations and permutations of the following variables: temperature (4° C, 10°, and 20° C), Hampton additive screen 2 (Hampton Research), Hampton detergent screen 1 and 3 (Hampton Research), other additives (alcohols, organics, sugars, and salts), drop size (1 to 5 μL total), protein solution to well solution size (1 μL protein solution with 0.5, 1, 1.5, 2, 2.5 or 3 μL well solution), pH (2.5 to 9), buffer type, protein concentration (10 to 25 mg/mL), ammonium sulfate concentration (1 to 3 M), and the Hampton heavy metal screens (Hampton Research). Large diamond shaped crystals were found by mixing 1 μL hPCNA-p21(139:160) complex with 1 μL well solution consisting of 10% MPD, 2M ammonium sulfate, and 0.1 M sodium acetate (pH 4.6) at 4° C. Hexagonal plate crystals similar in morphology to the published crystals were finally obtained using 1 μL hPCNA-p21(139:160) complex mixed with 0.2 μL Hampton
detergent screen detergent #2 (C_{12}E_8) and 1 μL well solution consisting of 2M ammonium sulfate and 0.1M sodium acetate (pH 4.6) at 20°C. The detergent was absolutely necessary for optimal crystal size, shape, and diffraction quality.

**Figure 3.2** Hexagonal shaped co-crystals of hPCNA-p21peptide. Crystals grew in 10% MPD, 2 M ammonium sulfate, and 0.1 M sodium acetate. 1μL complex was mixed with 1 μL well solution and 0.2 μL detergent. Crystals had the physical morphology of previously published crystals.

3.3 Crystallization of hPCNA with the p66 PIP-box peptide
The first step in crystallizing hPCNA with the p66 PIP-box peptide was the design of the peptide length. Our goal was to include the important conserved residues of the PIP-box while also choosing a peptide length that supported ease of crystallization. Unlike p21, the p66 native protein terminates earlier in its homologous PIP-box region than does p21, so the construction of the C-terminus of our peptide was dictated by its native C-terminus. We designed the N-terminus of our p66 peptide to be several residues shorter than that of the N-terminus of the p21-derived peptide. Our N-terminal design of the p66-derived peptide was based on the hPCNA-p21 crystal structure in which the N-terminal region of the p21 peptide was disordered. We hypothesized that decreasing the disorder at the N-terminus might aid in ease of crystallization. Our p66-derived peptide was 15 amino acids in length (\(^{452}\text{KANRQVSITGFFQRK}\)\(^{466}\), with the conserved residues of the PIP-box highlighted in red).

Crystallization experiments began by exploring the published hPCNA-p21(139:160) conditions as well as the new conditions we found which crystallized the complex. In addition, sparse matrix screens such as the Hampton Salt Rx, Index, and Crystal Screen I (Hampton research) were used. Foot-printing screens using ammonium sulfate, MPD, and MPD with ammonium sulfate were also carried out. Small, irregular cubic crystals formed in almost every condition containing ammonium sulfate (1.5-3 M) as a precipitant at pH 3-5.5 within 3 days. The best crystals as judged by size and shape grew at 20°C.

Because the hPCNA-p66(452:466) crystals were small, 2M ammonium sulfate and 0.1M sodium acetate (pH 4.6) was used as a starting point for optimization. Crystals were optimized by trying these strategies: fine ammonium sulfate grids (0.05 M
ammonium sulfate steps), temperature (4°C, 10°C, and 20°C), Hampton additive screen 2 (Hampton Research), Hampton detergent screen 1 and 3 (Hampton research), other additives (alcohols, organics, sugars, and salts), drop size (1 to 5 µL total), protein solution to well solution size (1 µL protein solution with 0.5, 1, 1.5, 2, 2.5 or 3 µL well solution), pH (2.5 to 9), using other sulfate containing salts (i.e., lithium sulfate), sodium acetate concentration grids (0.25M steps), peptide to hPCNA ratio (0.5 to 7 peptide:hPCNA ratio), and the Hampton heavy metal screens (Hampton Research). Micro batch, using Al's oil, and sitting drop methods were also tried. Large, nicely shaped cubic crystals suitable for data collection were found by mixing 5 µL hPCNA-p66(452-466) (at a 0.8 peptide:hPCNA ratio) with 5 µL well solution consisting of 2 M ammonium sulfate and 0.1M sodium acetate (pH 4.6) at 20°C. Crystals formed within 3 days and reached maximum size within one week. The large drop size was absolutely necessary to obtaining large crystals. In total, 1255 conditions were tried before obtaining diffraction quality crystals.
3.4 Crystallization of hPCNA with the FEN1 PIP-box peptide

The FEN1-derived peptide length was designed based on a length similar to that of the p21-derived peptide as well as on biochemical studies which showed the residues necessary to compete with other proteins for hPCNA binding (Warbrick et al., 1997). The FEN1-derived peptide was 20 amino acids in length.
\(^{(331)SRQGSTQGRDLDFKVTGSL^{350}}\), with conserved PIP-box residues highlighted in red).

Crystallization experiments began by searching conditions used in previous hPCNA crystallization. In addition, sparse matrix screens were carried out using the Hampton Salt Rx screen (Hampton Research). Small diamond shaped crystals formed in 2.6 M ammonium sulfate and 0.1 M sodium acetate (pH 4.6), but required at least 2 weeks to form and had rough edges. This condition was used as a starting point for optimization. Crystals were optimized by trying these strategies: fine ammonium sulfate grids (0.05 M ammonium sulfate steps), temperature (4\(^{\circ}\) C, 10\(^{\circ}\) C, and 20\(^{\circ}\) C), Hampton additive screen 2 (Hampton Research), Hampton detergent screen 1 and 3 (Hampton Research), other additives (alcohols, organics, sugars, and salts), drop size (1 to 5 \(\mu\)L total), protein solution to well solution size (1 \(\mu\)L protein solution with 0.5, 1, 1.5, 2, 2.5 or 3 \(\mu\)L well solution), pH (2.5 to 9), sodium acetate concentration grids (0.25M steps), and peptide to hPCNA ratios (0.5 to 7 peptide:hPCNA ratio). Micro batch, using Al's oil, and sitting drop methods were also tried. Streak-seeding was also employed but produced no crystals. Unlike the hPCNA-p66(452:466) the range of ammonium sulfate concentration in which quality crystals grew was very narrow, often within \(\pm\) 0.05 M ammonium sulfate. Reproducible diamond shaped crystals with well-shaped edges that grew within 3 days were obtained by mixing 1 \(\mu\)L hPCNA-FEN1(331-350) complex (using a 0.8:1 peptide:hPCNA ratio) with 1 \(\mu\)L well solution consisting of 2.35 M ammonium sulfate and 0.1 M sodium acetate (pH 4.6) at 20\(^{\circ}\) C.
Figure 3.4 PCNA-FEN1 peptide co-crystals. Crystals grow in 2.35 M ammonium sulfate and 0.1 M sodium acetate pH 4.6 at room temperature in a hanging drop. The quality diamond-shaped crystal is highlighted with a red circle.

Because hPCNA-FEN1(331-350) co-crystals exhibited partial hemihedral twinning (described further in chapter 6), several conditions were tried to eliminate or decrease the twinning. Small concentrations of small organic molecules and small PEGs have been shown to decrease or stop twinning. These include DMSO, dioxane, isopropanol, ethanol, butanol, MPD, and PEGs less than 1000 molecular weight. Using 2.35 M ammonium sulfate and 0.1 M sodium acetate (pH 4.6) as a starting point, 0.2-5% DMSO, dioxane, butanol, MPD, PEG 350, PEG 550, and PEG 200 were screened in order to decrease or eliminate the twinning. Although, the twinning was not eliminated, DMSO, at a final drop concentration of 0.2%, was used in order to help decrease the amount of twinning in the crystals grown for synchrotron data collection.
In order to obtain quality hPCNA-FEN1(331:350) co-crystals suitable for data collection, 920 different conditions were tried.

3.5 Crystallization of hPCNA with the WRN PIP-box peptide

The WRN-derived peptide was designed largely on sequence homology as little biochemical data on the PIP-box exists. Furthermore, because the p66-derived peptide crystallized readily and was short in length (15 amino acids), we designed the WRN-derived peptide to be closer in length to the p66-derived peptide rather than the p21-derived peptide (22 amino acids). The WRN-derived peptide was 18 amino acids in length: 141EGDQWKLLRDFDIKLKNF158 (with conserved PIP-box residues highlighted in red).

The hPCNA-WRN complex crystallization began by trying the other hPCNA crystal conditions as well as sparse matrix screens such as the Hampton Salt Rx and Crystal Screen I screens (Hampton Research). Coarse foot-printing screens using ammonium sulfate, ammonium formate, sodium malonate, di-ammonium phosphate, and MPD versus various pHs were also tried. Large diamond shaped crystals, very similar in morphology to the hPCNA-FEN1(331-350) co-crystals, with rough surfaces were found in conditions around 2 M sodium malonate and 0.1 M sodium acetate (pH 4.6) at 20° C. These crystals were used for further optimization. Crystals were optimized by trying these strategies: fine sodium malonate grids (0.05 M ammonium sulfate steps), temperature (4° C, 10°, and 20° C), Hampton additive screen 2, Hampton detergent screen 1 and 3, other additives (alcohols, organics, sugars, and salts), drop size (1 to 5 μL total),
protein solution to well solution size (1 μL protein solution with 0.5, 1, 1.5, 2, 2.5 or 3 μL well solution), pH (2.5 to 9), sodium acetate concentration grids (0.25M steps), and peptide to hPCNA ratios (0.5 to 7 peptide:hPCNA ratio). Micro batch, using Al's oil, and sitting drop methods were also tried. Streak-seeding was also employed and produced similar crystals of no higher quality as judged by morphology and size. The best crystals obtained were obtained by mixing 1 μL hPCNA-WRN complex (using a 1:1 peptide:hPCNA ratio) with 1 μL well solution consisting of 2 M sodium malonate and 0.1 M sodium acetate (pH 4.6) at 20° C. Crystals grew to full size within 3 days. To find the best crystal conditions, 601 different crystal trials were conducted.
3.6 Crystallization of hPCNA with the RecQ5 PIP-box peptide

Similar to the WRN-derived peptide, the RecQ5-derived peptide was designed largely on sequence homology as little biochemical data on the PIP-box exists. As with the p66 and WRN derived peptides, the RecQ5-derived peptide was designed to be shorter in length to aid in crystallization. The RecQ5-derived peptide was designed to be exactly the same length as well as the homologous residues of p66-derived peptide. The
RecQ5-derived peptide was 15 amino acids in length: $^{960}\text{KEEAQNLIRHFFHGR}^{974}$ (with the conserved residues of the PIP-box highlighted in red).

hPCNA-RecQ5 crystallization began by searching conditions which produced crystals for the other hPCNA complexes. In addition, course foot-printing screens using ammonium sulfate and MPD versus pH were carried out. Clusters of thin, rectangular plates were found at 20°C at 2 M ammonium sulfate and 0.1 M sodium acetate (pH 4.6). This condition was used for optimization. Optimization trials included: Hampton additive screen 2 (Hampton Research), Hampton detergent screen 1 (Hampton Research), drop size (1 to 5 μL total), protein solution to well solution size (1 μL protein solution with 0.5, 1, 1.5, 2, 2.5 or 3 μL well solution), hCPNA to peptide ratios (0.5 to 7 peptide:hPCNA ratios), and course ammonium sulfate and MPD grids with varying pH. Only 187 conditions have been tried as yet, and better crystals may be found with more optimization. To date, the best rectangular plate crystals were made by mixing 1 μL hPCNA-RecQ5 complex (using a 1:1 hPCNA:peptide ratio) with 1 μL well solution consisting of 2 M ammonium sulfate and 0.1 M sodium acetate (pH 4.6) at 20°C.
Figure 3.6 hPCNA-RecQ5 peptide co-crystals. Crystals grew in 2 M ammonium sulfate and 0.1 M sodium acetate pH 4.6 at 20 °C. More optimization experiments may lead to improved crystal quality.

3.7 Cryo-crystallography

Crystallography at 100 K has been shown to protect crystals from radiation damage suffered from prolonged exposure to x-rays; cryo-crystallography also prevents protein oxidation and is employed for ease of storage of crystals. All crystals in our studies were flash cooled with liquid nitrogen, stored at cryo-temperatures, and exposed to x-rays at 100 K (Henderson, 1990). Cryo-protectants are required to protect crystals upon flash cooling. Crystals without proper cryo-protectants can suffer cracking, ice
formation, and overall poor diffraction. Proper cryo-protectants must be found on a case by case basis by means of experimentation. Poor cryo-protectants can be identified by ice rings on the diffraction pattern, physical cracking, disappearance of the crystal (the protein going back into solution), or poor diffraction as compared to a capillary mounted crystal. A successful cryo-protectant has been found when the diffraction quality of the cryo-cooled crystal matches that of the capillary mounted crystal exposed to x-rays at room temperature as well as the lack of ice-rings on the diffraction pattern.

Several cryo-protectants were tested for the hPCNA-p21(139:160) and hPCNA-p66(452:466) co-crystals. Crystals were harvested from the drop with a loop, transferred to a drop containing the cryo-protectant, and then flash cooled and exposed to x-rays. Cryo-protectants tried were: 2 M Li₂SO₄, 30% glycerol with 2 M ammonium sulfate and 0.1M sodium acetate (pH4.6), 20% glycerol with 2 M ammonium sulfate and 0.1 M sodium acetate (pH4.6), 10% glycerol with 2 M ammonium sulfate and 0.1 M sodium acetate (pH4.6), 30% glycerol, 20% glycerol, 10% glycerol, 50% sucrose with 2 M ammonium sulfate, 2 M LiCl, and Al's oil (50% paraffin oil, 50% silicon oil). In addition, a gradual step-wise transfer to 30% glycerol and 2 M ammonium sulfate was tried in which the crystal was first transferred to 5% glycerol and 2 M ammonium sulfate and equilibrated for 10 minutes, and likewise increasing the glycerol concentration by 5% at each step until 30% glycerol was reached. The optimal cryo-protectant for both crystals was determined to be 30% glycerol, 2 M ammonium sulfate, and 0.1M sodium acetate (pH4.6) as determined by the lack of ice-rings in the diffraction pattern as well as diffraction quality which was consistent with that of the capillary mounted crystals. 2 M ammonium sulfate and 0.1M sodium acetate (pH4.6) were found to be essential to the
cryo-protectant as in their absence crystals cracked and exhibited poor diffraction. This
cryo-protectant was also used for all subsequent hPCNA co-crystals and showed
consistent cryo-protectant quality as judged by diffraction quality. All harvested crystals
were stored under liquid nitrogen.
Chapter 4: Preliminary characterization of crystal diffraction and data collection

4.1 Screening of crystals by means of diffraction

Although crystal quality is initially screened by morphology and size, crystals must next be screened by diffraction quality. Although crystals may look perfectly formed, they may not diffract well. In addition, crystals that look similar or crystals from very similar conditions may not diffract in a similar manner. Even crystals formed in the same drop may have different diffraction qualities. When screening crystals several factors must be considered when determining whether a crystal is suitable for full data set collection. These factors are described below. Crystals were screened on our home source by exposures taken at four angles: 0, 90, 180, 270 degrees.

Resolution. Resolution is perhaps the most important factor in determining crystal diffraction quality. For our studies, all diffraction resolution had to be higher than 3 Å in order to be of significant scientific merit. For many of our crystals, this standard could not be met. Longer exposure times can often be used to determine the absolute resolution limit of the crystal as well as whether the crystal resolution limit is inherently limited or limited by the x-ray intensity. This is important when determining if crystal diffraction resolution could be significantly improved by diffraction at a synchrotron. Crystal screening at our home source was usually carried out with a minimum of 10 to 20 minute exposures.
Flash cooling. Even after proper cryo-protectants have been found, a "bad freeze" can still occur. A "bad freeze" occurs when crystals still become cracked or form ice when flash cooled. Often this occurs because the flash cooling is not quick enough or the loop experiences solvent loss while being transferred to the cryo-stream.

Anisotropy and completeness. Some crystals may exhibit diffraction patterns which lack reflections in some quadrants of each individual frame. These diffraction patterns appear "lop-sided". Some crystals, usually due to morphology, may have high quality diffraction at one angle and not at another angle. In both of these cases, data sets may be incomplete and better crystals should be searched for.

Crystal variability. Although crystals grown in the same conditions may appear identical under a light microscope, their diffraction characteristics can vary dramatically. In our studies, when a drop contained promising crystals, a minimum of 3-5 crystals from that drop were screened. In addition, crystals from different but similar drops, i.e., a different pH or with the addition of a detergent, may diffract completely different. In our studies, all promising crystals from different conditions were screened even if their morphology and size were similar.

Crystal cracking and multiple crystals. Sometimes crystals become cracked upon flash cooling, but the cracking cannot be seen by physical inspection. Also, sometimes crystals appear to be single crystals but are in reality multiple crystals attached to each other. In both of these cases, the diffraction pattern can be easily identified. The diffraction pattern will appear to have overlapping or irregular lattice lines; the diffraction pattern will appear as two diffraction patterns superimposed. Crystals with these problems are not suitable for further data collection.
**Reflection intensity.** Sometimes crystals may diffract to high resolution, but may not have a high intensity over background. These reflections may not be very well measured, leading to poor data processing and an overall poor structure solution. This can often be overcome by increasing exposure time or by using more intense synchrotron radiation.

**Synchrotron.** It is nearly certain that using the synchrotron x-ray source will improve resolution and reflection quality compared to a home source. There are often long waits to get time at a synchrotron and synchrotron trips are expensive. Due to the possibility of human error and time constraints at the synchrotron, all crystals were pre-screened at the home source and a minimum of 10 crystals of each complex were taken to the synchrotron. Crystals taken to the synchrotron were stored and shipped under liquid nitrogen.

### 4.2 X-ray diffractometers and collection apparatus

Data was collected on our home source using a Rigaku R-axis IV++ detector, a rotating copper anode at 1.54 Å wavelength, and Osmic Mirrors™. Data was collected at 100 K using a nitrogen cryo-stream.

All synchrotron data was collected at the CHESS A1 beamline at 0.9764 Å wavelength. Data was collected at 100 K using a nitrogen cryo-stream. Data was collected using a CCD ADSC QUANTUM 4 detector.

### 4.3 hPCNA-p21(139:160) co-crystals
Several hPCNA-p21(139:160) co-crystal forms were screened. Co-crystals which were long, thin, and grown in 30% MPD did not diffract at all (Figure 3.1). These crystals were most likely limited by their size; larger, thicker crystals of this form could not be obtained. The diamond shaped co-crystals grown in conditions around 10% MPD and 2 M ammonium sulfate diffracted to only ~7-8 Å resolution. These crystals also suffered from a low number of reflections and appeared incomplete. The octagonal plates (seen in Figure 3.2) grown in conditions around 2 M ammonium sulfate diffracted to ~3-3.4 Å resolution and appeared to have quality reflections based on shape, pattern, and intensity (Figure 4.1). The unit cell for these crystals was calculated as 83.5 x 83.5 x 200 Å in a trigonal space group; this was very similar to the published space group which was also trigonal and 83.5 x 83.5 x 233.9 Å. No data set was collected on the hPCNA-p21(139:160) co-crystals as they were solely a proof of concept; all efforts after this point were directed at other PIP-box peptides. In all, about 25 separate hPCNA-p21(139:160) co-crystals were screened by means of diffraction.
Figure 4.1 Diffraction pattern of hPCNA complexed with p21 peptide. Diffraction came from hexagonal crystals shown in Figure 3.2 with a 10 minute exposure. The resolution limit was below 3.2 Å. The unit cell dimensions were calculated as 83.5 by 83.5 by 200 Å and the crystal was classified with a trigonal space group. A full data set was not collected.

4.4 hPCNA-p66(452:466) co-crystals

Two crystal forms of hPCNA-p66(452:466) were screened by diffraction: the football shaped crystals formed in conditions around 2 M ammonium sulfate and 0.1 M citrate (pH 2.7) and the irregular cubic crystals formed in 2 M ammonium sulfate and 0.1 M sodium acetate pH 4.6 (Figure 3.3). Only ~10 of the football shaped crystals were screened. These crystals were characterized by poor diffraction: the resolution only extended to 7-12 Å and very few reflections were observed. No improvement in the diffraction of this crystal form could be achieved.
The major crystal form of the hPCNA-p66(452:466) complex was screened: the irregular cubic crystals grown in conditions around 2 M ammonium sulfate and 0.1 M sodium acetate (pH4.6). In total, 25-50 separate crystals were screened by means of diffraction. Diffraction resolution was quite variable and was significantly dependent on crystal size. Most crystals diffracted to 3 to 4 Å (Figure 4.2). The best crystal diffracted to 2.8 to 3 Å at our home source and was the largest crystal which was screened. At the home source, low resolution reflections were large and very intense over background while high resolution reflections were much smaller and weaker in intensity. Our goal was to increase high resolution diffraction as well as intensity over background for high resolution reflections at the synchrotron. Two complete data sets were measured with the same crystal: one at our home source and one at a synchrotron source. Data was over-collected in both cases (more data than theoretically necessary for a complete data set) and 90° of data was taken at 0.5° oscillations. At the home source 20 minute exposures were taken while at the synchrotron 20 second exposures were taken. The synchrotron improved the resolution limit by almost 0.5 Å: the resolution limit at home was 2.8 to 3 Å while that at the synchrotron appeared to be 2.4 Å. Although reflections were visible at 2.4 Å, they were not well measured as dictated by R factors in refinement (discussed in chapter 7) and data had to be later cut to 2.6 Å. Data processing and statistics are discussed in chapter 5.
Only one crystal form of hPCNA-FEN1(331:350) was screened by means of diffraction: the diamond shaped crystals grown in conditions around 2.35 M ammonium sulfate and 0.1 M sodium acetate pH 4.6 (Figure 3.4). In total, 25-50 separate crystals
were screened by diffraction. All crystals had similar resolution limits: at our home source they were 2.2 to 2.4 Å, while at the synchrotron they diffracted to ~1.8 Å. These crystals were not limited by resolution. In addition, the reflections were well shaped, appearing nearly perfectly round. The reflections measured at high resolution were of high quality as judged by their intensity and later by data processing and refinement.

Because these crystals were twinned (discussed in Chapter 6), additives were used to try to alleviate the twinning. These crystals were screened by means of diffraction to observe whether twinning had been alleviated. The measure of this was seen by the unit cell size: unit cell sizes with implausible Matthew's coefficients were judged as twinned. Of all the additives screened, all crystals screened by diffraction still showed the same unit cell size. Twinning could not be alleviated. Although twinning could not be completely alleviated, it is possible that additives could have lessened the twinning fraction. The only method to measure the twinning fraction is by measuring a complete data set, and measuring several data sets for different additives was not plausible due to time constraints at the synchrotron.

Two complete data sets were measured: one set at our home source and one set at a synchrotron source. Data was over-collected, and in both cases 180° of data was collected using 0.5° oscillations. Data at the home source was collected using 20 minute exposures while data at the synchrotron was collected using 10 second exposures. Data processing statistics are discussed in Chapter 5.
4.6 hPCNA-WRN co-crystals

Only one form of hPCNA-WRN co-crystals were screened by diffraction quality: diamond shaped crystals formed in conditions around 2 M sodium malonate and 0.1 M sodium acetate pH4.6 (Figure 3.5). In total, 25-50 separate crystals were screened. At our home source the best crystals found only diffracted to 8-10 Å (Figure 4.4). The crystal's diffraction had well shaped reflections but had a severe resolution fall off, with absolutely no reflections higher in resolution than 8 Å. 10 of the best diffracting crystals
were taken to be screened at the synchrotron. The synchrotron improved the resolution limit to about 6-7 Å, but there was still a severe resolution fall-off with no reflections at higher resolution. These crystals were, therefore, inherently limited in resolution and no complete data set of value could be collected.

![Diffraction pattern of hPCNA-WRN peptide co-crystals. This 20 minute exposure shows that the resolution limit only extends to ~10 Å. No data set was collected because the resolution was not adequate.]

4.7 hPCNA-RecQ5 co-crystals
Only one form of hPCNA-RecQ5 co-crystals were screened by diffraction quality: the rectangular plates grown in conditions around 2 M ammonium sulfate and 0.1 M sodium acetate pH 4.6 (Figure 3.6). Only 5 co-crystals were screened as the crystals had not been completely optimized. Diffraction extended to ~3.5 Å, and reflections were well shaped (Figure 4.5). Several crystals exhibited overlapping lattice lines, as crystals were difficult to separate from each other; the mounted crystal may have been two or several epitaxial crystals together. No data set was collected on these crystals. Further optimization of this crystal form may lead to better diffraction suitable for structure determination.
Figure 4.5 Diffraction pattern of hPCNA-RecQ5 peptide co-crystals. This 20 minute exposure shows the resolution extends below 3.5 Å but the overlapping lattice lines show epitaxial twinning.
Chapter 5: Diffraction data Processing

5.1 Indexing, scaling, and merging of hPCNA-p66(452:466) diffraction data

All 90° of data collected at the synchrotron was used for data processing. Because the reflections at low resolution were sizably larger than those at high resolution, some trouble in reflection selection occurred during indexing. This problem was largely overcome and data was suitably indexed. All data was indexed, scaled, and merged using Crystal Clear and d*trek (Pflugrath, 1999). The space group was determined to be P3_21 which is the same space group as that published for hPCNA-p21(139:160). The unit cell parameters were determined to be 82.5 x 82.5 x 203.8 Å and 90° x 90° x 120°, which is also very similar to the unit cell parameters reported for the published hPCNA-p21(139:160) structure. The Rmerge was significantly improved by the synchrotron data as compared to home source data: using our home source the Rmerge overall was 0.117 whereas using the synchrotron source the Rmerge overall was 0.081. From this, we can see that not only did the synchrotron improve the resolution limits of this crystal but also the overall quality of the data. The chi^2 was 0.8 and the mosaicity was 0.8. The remaining data processing statistics can be found in Table 5.1. The completeness, average I/σ, and redundancy are all within acceptable values for a quality data set.

5.2 Indexing, Scaling, and merging of hPCNA-FEN1(331:350) diffraction data
All 180° of synchrotron data was used for indexing, scaling, and merging. All indexing, scaling, and merging was carried out with Denzo and Scalepack (Otwinowski and Minor, 1997). The space group was difficult to determine immediately due to the twinning. The data could be merged with an $R_{\text{merge}}$ overall of 0.064 in P6$_3$ and also merged in P6$_3$22 with an $R_{\text{merge}}$ of 0.11. Systematic absences were consistent with both space groups. We determined the true space group to be P6$_3$ and the apparent space group to be P6$_3$22. The data was therefore processed in P6$_3$. The unit cell parameters were found to be 81.5 x 81.5 x 67.5 Å and 90° x 90° x 120°. The data set was of much higher quality than that seen for the hPCNA-p21(139:160) and hPCNA-p66(452:466) structures as seen by the statistics: the $R_{\text{merge}}$ overall of 0.064, a very high average intensity over background, and a very high overall completeness even in the highest resolution shell. This structure is also by far the highest resolution structure seen of hPCNA to date. The summary of the data processing statistics can be found in Table 5.1.
<table>
<thead>
<tr>
<th>Table 5.1 Data processing statistics.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Collection</strong></td>
</tr>
<tr>
<td>Beam line</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>Space Group</td>
</tr>
<tr>
<td>Unit Cell (Å)</td>
</tr>
<tr>
<td>Resolution (Å)</td>
</tr>
<tr>
<td>Unique reflections</td>
</tr>
<tr>
<td>Average Redundancy</td>
</tr>
<tr>
<td>Completeness (%)</td>
</tr>
<tr>
<td>Rmerge</td>
</tr>
<tr>
<td>I/σ</td>
</tr>
</tbody>
</table>

1Statistics apply to data prior to detwinning. Numbers in parenthesis represent statistics for the highest resolution shell.
Chapter 6: Crystal Twining

6.1 Background on crystal twinning

Crystals generally contain homogeneous orientations of unit cell domains related by crystallographic symmetry (Figure 6.1A). Twinning occurs when a crystal is composed of two or more specific domains related by specific twinning operations; this often leads to higher apparent symmetry (for a review see Yeates and Fam, 1999). The proportion of each twin domain is referred to as the twinning fraction (α) and is expressed in decimal form from 0 to 1, with 0 being 0% and 1 being 100% of the particular twin domain orientation. The twin domain can also be thought of as the physical three-dimensional volume of the crystal occupied by each twin domain. Several forms of twinning exist including epitaxial twinning, merohedral twinning, and pseudomerohedral twinning. Epitaxial twinning generally occurs on the macroscopic level when two or more crystals are physically juxtaposed to each other. Epitaxial twinning is generally easy to see both physically as well as on the diffraction level. Crystals with epitaxial twinning will appear as a hybrid of two or more crystals under a light microscope, and upon exposure to x-rays will give diffraction patterns with split reflections, overlapping or nonsensical lattice lines, or appear as two separate diffraction patterns overlapped. Merohedral twinning occurs on a macroscopic level with twin domains overlapping exactly three-dimensionally to produce higher symmetry; a single crystal is composed of two or more lattice domains which are systematically
interpenetrated with each other. Merohedral twinning can happen when the lattice

![A B C](image)

**Figure 6.1** Schematic representation of hemihedral twinning within theoretical crystal lattices. For simplicity, the figures represent a 2-dimensional cross-section of 3-dimensional crystal domains. Crystal domains are designated by small squares, with arrows indicating their orientation. Two orientations of the lattice domain exist, shown by up and down arrows A Example of a normal lattice with no twinning B Example of partial twinning with only a small proportion of lattice domains with different orientations ($\alpha=1/6$) C Example of perfect twinning with equal proportions of each twin domains ($\alpha=0.5$).

symmetry exceeds the underlying symmetry of the crystal. With merohedral twinning the lattices of the apparent and true space groups belong to the same Laue class.

Hemihedral twinning is merohedral twinning with exactly 2 twin domains. Most
macromolecular twinning is of this type and the two domain orientations are related by a 180° rotation. Perfect hemihedral twinning is defined as having a twin fraction of 0.5; lattices consist of equal proportions of twin domains. Partial hemihedral twinning is defined as having any twin fraction less than 0.5. The crystals generally appear single and diffraction patterns look normal, although they are not. The reflections contain intensity contributions from both domains, with each twin domain contributing proportionally to its volume. This is directly related to the twin fraction as well as the twinning symmetry relations. Furthermore, these factors are the key to detwinnng diffraction data for structure solution (Grainger, 1969) discussed in section 6.2.

Pseudomohedral twinning is very similar to merohedral twinning, but occurs less often; it is defined by having twin domains which come together to produce apparent symmetry in different lattice classes (i.e., monoclinic and hexagonal) and often encompass lattices related by other rotations other than 180°.

Twinning may be difficult to overcome but twinned structures can be solved. The best way to overcome twinning is to find crystal forms that are not twinned or have a lower twin fraction. Several additives have been found that have been shown to, in many separate cases, remove or alleviate twinning (reviewed by Sixma, 2001). These additives are included in the crystal drop in low amounts, usually 0.1-5%. The additives include small organic molecules such as dioxane, DMSO, ethanol, isopropanol, and methanol. Detergents and small molecular weight PEGs (molecular weights less than 1000) have also been shown to be helpful. When non-twinned crystals cannot be found, there are ways to detwin the data. In fact, sometimes twinned structures are solved as normal
structures, unbeknownst to the crystallographer, most likely due to a low twinning fraction

6.2 Detwinning

The true intensities can be recovered from data measured from crystals exhibiting hemihedral twinning if the twin fraction is less than 0.5; this process is known as detwinning. The mathematical basis for detwinning was pioneered by Grainger (Grainger, 1969). The principal element behind detwinning lies in real space. Crystal domains within a twinned crystal are related by a symmetry operation. Likewise, the reflections resulting from each twin domain are related by a reciprocal space symmetry operation. This symmetry relationship between twin domains is known as the twinning operation. Because the diffraction of the twin domains do not interfere with each other, observed intensities are simply a linear combination of the two twin domain's intensities. Their relationship is examined in equations 1 and 2.

\[ p = (1-\alpha)I_1 + \alpha I_2 \quad (1) \]
\[ q = \alpha I_1 + (1-\alpha)I_2 \quad (2) \]

p and q are reflections related by the twinning operation. \( I_1 \) and \( I_2 \) are the untwinned intensities that compose the observed reflections, p and q, and overlap exactly. Thus, once the twinning operation and \( \alpha \) are known accurately, the true reflections can be recovered. The process is not perfect and as \( \alpha \) approaches 0.5 the detwinning becomes more difficult because the measurement errors increase by \( 1/(1-2\alpha) \), which increase significantly as \( \alpha \) approaches 0.5 (Yeates, 1999). If the twin fraction is far from 0.5, the
smaller twin domain does not make a major contribution to the observed intensities. Due to
the sensitivity of the detwinning process to error, the twin fraction must be known with
little error as well. In addition, detwinning requires the identification of the true space
group as well as the apparent space group to identify the twin domain symmetry relations
for detwinning. Several statistical methods exist for finding α, and detwinning of data is
possible in most commonly used refinement suites: CNS (Brunger et al., 1998), CCP4
(Collaborative Computational Project, 1994), and SHELXL (Herbst-Irmer and Sheldrick,
1998) included. These methods use data on the level of the reflection file which has
already been indexed, scaled, and merged.

6.3 Warning signs of twinning and twinning detection

Perhaps the most subjective detection of twinning is on the level of crystal
morphology. Because detecting twinning by morphology is subjective and twinned
crystals show only subtle differences, this method for detecting twinning is the least
useful and is only reserved for expert crystallographers. Crystal packing and defects on
the lattice level have effects on the overall morphology of a crystal. As such, crystals
inflicted with merohedral twinning often have a concave morphology (Yeates, 1997).

Twining will also lead to difficulty when multiple data sets are needed. This
makes twinning harder to overcome when using MIR. Different crystals will inevitably
have different twinning fractions, even if only slightly different. These differences will
multiply difficulties when the scaled data sets are merged because the relationships of the
observed intensities are not the same.
Twinning can also be seen on the level of data processing in the form of $R_{\text{merge}}$. Data will merge well in a related space group of higher symmetry. If the twinning fraction is less than 0.5, the $R_{\text{merge}}$ will be slightly higher in the higher symmetry space group. The true space group is represented by contributions of the dominant twin domain in the lower symmetry space group with a lower $R_{\text{merge}}$ and the apparent space group is represented in a space group of higher but related symmetry with a higher, but acceptable, $R_{\text{merge}}$ due to the contributions of the other twin domain.

Hemihedral twinning can only occur in certain space groups with specific twin operations. Trigonal, hexagonal, and tetragonal space groups are the most common space groups which support hemihedral twinning.

Because partially twinned data often goes unnoticed, it is not until the level of refinement that symptoms of twinning are seen. If a structure is generally difficult to solve, cannot be solved, or is highly disordered, it may be twinned. Distorted reflections from twinning will have a direct result on electron density quality and respective refinement.

Another clue that a crystal may be twinned is an abnormal or impossible Matthew's coefficient. If a given protein cannot physically fit within the given unit cell parameters, the space group may have been misidentified due to false apparent symmetry caused by twinning.

Finally, the most specific criteria to determine the presence of twinning is that of intensity statistics. The intensity statistics of twinned data sets are abnormal and deviate from expected values. Reflections in the presence of twinning do not have appropriate intensities due to the fact that they are receiving contributions from two lattice domains.
Two common tests exist for examining intensity statistics with twinning. The first is $<I^2>/<I>^2$ where $I$ is intensity. For a perfectly twinned structure this ratio will be 1.5 and that for a non-twinned structure will be 2. The second mathematical test is the cumulative $H$ (equation 3) distribution over twin related reflections, which differs based on the twin fraction; this method requires that the true space group is known so that the twin related operations can be used in the calculations. Just as there are symmetry related reflections, there are twin related reflections, and their intensities are directly related to the twin fraction. $H$ is the fractional difference between a pair of twin related intensities,

$$H = |p-q| / (p+q)$$  \hspace{1cm} (3)

where $p$ and $q$ are the observed intensities of twin related reflections. Thus, $H$ has different distributions for different twin fractions and can be compared to expected values of these distributions for different twin fractions. Complications can occur, and these calculations are not perfect. For instance, non-crystallographic symmetry can potentially emulate a twinning operation. In addition, anisotropic diffraction can distort intensity statistics and make twinning harder to detect by intensity statistics alone.

6.4 Detection of hPCNA-FEN1(331:350) peptide crystal twinning

Several factors led us to the understanding that the hPCNA-FEN1(331:350) crystals were twinned. As is the case with many crystallographers dealing with twinned crystals, nothing appeared abnormal until the structure would not refine properly. Our molecular replacement search model did not contain the FEN1 or p21 peptide, and the presence of peptide density in our density maps was henceforth used a measure of proper
space group determination and detwinning. Prior to detwinning, molecular replacement solutions were easily found, but R-factors hung well above 30%. In addition, FEN1 peptide density was largely uninterpretable, and hPCNA density was of low quality.

Sliding clamps have a history of twinning. The Arcael PCNA was shown to have hemihedral twinning with a true space group of P6₃ (Chapados et al., 2004). Even more relevant, was that human PCNA was shown to have trigonal twinning tendencies (Gulbis et al., 1996). Given that human PCNA has a tendency to twin, and our space group was trigonal or hexagonal, we were further convinced that our crystals may be twinned.

No visible signs of twinning were present when crystals were observed under the light microscope. Crystals did not appear concave; they appeared completely normal.

Another clue came when reexamining our data processing. When data was scaled and merged in P6₃ the R_{merge} was 0.06. When data was scaled and merged in P6₂22 the R_{merge} was 0.11, which is higher than the value for P6₃ but still a viable R_{merge} value. This is consistent with hemihedral twinning; space group P6₃ is in the same lattice class as P6₂22 with the higher symmetry space group having the slightly higher R_{merge}.

The intensity statistics described above also confirmed that our crystals were twinned. The statistics were calculated using the Yeates twinning server. \(<I^2>/<I>^2\), often used to detect perfect twinning, was found to be 1.8 for all reflections. Again, for non-twinned structures this should be 2 while for perfectly twinned structures the value should be 1.5. This suggested that the structure exhibited twinning but it was not perfect twinning. This was confirmed by the \(<H>\) distribution (shown in Figure 6.2), that indicated partial twinning with a twinning fraction of 0.36. This value of \(\alpha\) was again confirmed by the CCP4 Detwin program (Collaborative Computational Project, 1994).
Figure 6.2 Intensity statistics ($<H>$) distribution for various twin fractions. Our data, shown in red, shows a statistical distribution that would give it a twinning fraction between 0.3 and 0.4. Data produced using CCP4 Detwin program.

True space group determination was challenging for several reasons. The Matthew's coefficient is often a good determinant of proper space group determination, especially in the face of twinning. The Matthew's coefficient ($V_m$) is given in Å³/Dalton and is a measure of the proportion of protein mass volume to solvent volume within a given unit cell. The $V_m$ should ideally be around 2.2 (values of 1.7-3.5 are acceptable) for proteins with a solvent content of 40-60%. The hPCNA-p21(139:160) structure was solved with hPCNA as a crystallographic trimer. Based on our $R_{merge}$ values, if our crystal had hemihedral twinning, the true space group was either P6$_3$ or P6$_3$22. The
Matthew's calculations showed that the hPCNA trimer could not adequately fit in these unit cells with a $V_m = 0.695$ for P6$_3$ and $V_m = 0.347$ for P6$_3$22. This indicated that the twinning was not merohedral but rather pseudomeroedral with a true space group in a lower lattice group such as monoclinic, or that hPCNA was crystallographically a monomer in which the biological trimer was formed by the symmetry operations of the space group. The $V_m$ for the hPCNA monomer in P6$_3$ was 2.1, suggesting that P6$_3$ could very well be the true space group. We also tried to determine the true space group by means of close observation of systematic absences. Unfortunately, the observed systematic absences are the same for P6$_3$, P6$_3$22, and the monoclinic case of P2$_1$ (0, 0, 2n). Systematic absences could not be used for space group determination, but rather only as a conformation of knowledge already gained by out $R_{merge}$ values.

From this point, twinning was largely overcome by educated experimentation using quality FEN1 peptide density as a measure of proper detwinning. In the pseudomeroedral case, the reflection file was scaled and merge in P2$_1$, without direct detwinning of the reflection file, and refined using SHELXL, inputting the proper twinning operators with the TWIN command (Herbst-Irmer and Sheldrick, 1998). No suitable structure solution was achieved with this method as seen by the quality of the density maps and the high $R_{factors}$ in refinement. In the second case, the reflection file was detwinned directly in the CCP4 suite using the program Detwin (Collaborative Computational Project, 1994). An $\alpha$ value of 0.36 and a twinning operator of 2 along a, b, a*, b* was used with the true space group as P6$_3$ and the apparent space group as P6$_3$22. The detwinned reflection file was subsequently used for molecular replacement and refinement in CNS as would be done normally without twinning. FEN1 peptide density
was subsequently clearly visible and refinement continued smoothly to an appropriate end.
Chapter 7: Structure solution and refinement

7.1 hPCNA-p66(452:466) structure solution

The scaled and merged reflection file from d*trek had 5% of its data flagged for Free R calculations: this reflection file was used for molecular replacement as well as all refinement. The hPCNA-p21(139:160) structure (1AXC.pdb) was used as the search model with the p21 peptide and all waters removed. All molecular replacement and refinement was carried out in CNS (Brünger et al., 1998).

<p>| Table 7.1 Cross-rotation solutions using the hPCNA-p66(452:466) data.¹ |
|----------------|----------------|----------------|----------------|</p>
<table>
<thead>
<tr>
<th>Index</th>
<th>theta 1 (°)</th>
<th>theta 2 (°)</th>
<th>theta 3 (°)</th>
<th>RF-function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>354.821</td>
<td>0.677</td>
<td>4.143</td>
<td>0.2035</td>
</tr>
<tr>
<td>14</td>
<td>0.000</td>
<td>0.677</td>
<td>9.323</td>
<td>0.0634</td>
</tr>
<tr>
<td>19</td>
<td>308.067</td>
<td>0.000</td>
<td>317.390</td>
<td>0.0589</td>
</tr>
<tr>
<td>23</td>
<td>168.326</td>
<td>0.000</td>
<td>177.648</td>
<td>0.0550</td>
</tr>
<tr>
<td>24</td>
<td>32.741</td>
<td>0.000</td>
<td>42.064</td>
<td>0.0549</td>
</tr>
<tr>
<td>28</td>
<td>98.196</td>
<td>0.000</td>
<td>107.518</td>
<td>0.0543</td>
</tr>
<tr>
<td>32</td>
<td>84.688</td>
<td>-1.865</td>
<td>94.010</td>
<td>0.0512</td>
</tr>
<tr>
<td>33</td>
<td>239.065</td>
<td>11.865</td>
<td>248.388</td>
<td>0.0511</td>
</tr>
<tr>
<td>36</td>
<td>196.378</td>
<td>0.000</td>
<td>205.700</td>
<td>0.0495</td>
</tr>
<tr>
<td>38</td>
<td>359.065</td>
<td>11.865</td>
<td>8.388</td>
<td>0.0485</td>
</tr>
</tbody>
</table>

¹ The hPCNA trimer (without p21-derived peptides and waters) was used as the search model (PDB:1AXC). The correct solution is highlighted in red. Data was produced in CNS (Brünger et al., 1998).
Cross-rotation was carried out first. A fast direct rotation search was used extending from 15 to 4 Å. The top ten cross-rotation solutions are listed in the Table 7.1.

The cross-rotation list was then used to carry out translation molecular replacement. A general method translation search was performed extending from 15 to 4 Å. One solution stood out well above the others as measured by the monitor above the mean. The top ten translation solutions are listed in Table 7.2. Solution #1 was chosen for further refinement. All refinement was initially carried out with a resolution range of 15 to 2.4 Å. Next, 20 cycles of rigid body refinement were carried out ending with $R_{\text{free}}=40.88$, $R=41.24$. This was followed by group B-factor refinement which significantly improved R-factors, with $R_{\text{free}}=37.59$ and $R=38.05$. 200 cycles of minimization refinement
followed next with a small improvement in R-factors, with $R_{\text{free}}=37.09$ and $R=32.82$. After minimization refinement, slow-cool simulated annealing was carried out with a starting temperature of 3000 K, and R-factors improved very little ending at $R_{\text{free}}=36.99$ and $R=32.54$. At this point, Fo-Fc, 2Fo-Fc, 3Fo-2Fc, and composite omit density maps were calculated. p66 peptide density was clearly visible in the Fo-Fc density map, as well as the other density maps, and residues 451-455 were built into the model. Following peptide rebuilding, simulated annealing, individual B-factor refinement, and minimization refinement were carried out. Addition of the peptide significantly improved the R-factors with, $R_{\text{free}}=34.75$ and $R=29.40$. Next, several rounds of manual rebuilding based on composite omit maps, followed by individual B-factor refinement as well as minimization refinement were carried out. Several small areas of the model were disordered and made model building in these areas difficult; these areas included the N-terminus of hPCNA (residues 1-2), the C-terminus of hPCNA (residues 256-261), parts of the interdomain connector loop of hPCNA (residues 123-133), and hPCNA residues 63-65, 91-97, 106-109, 162-165, and 186-192. In addition, waters were gradually added to the structure. Initially, only 103 waters could be added for optimal R-factor improvement; even so, the addition of waters at this point only improved the free R by only approximately a half percent. After waters were added, more rounds of manual rebuilding followed by individual B-factor refinement and minimization refinement were carried out. At this point, the free R-factor converged at around 32% and special considerations had to be taken into account to allow refinement to continue further.

First, we considered the resolution range of our refinement. We monitored R and $R_{\text{free}}$ by resolution bin. Although our data scaled and merged well to 2.4 Å, refinement
R-factors in resolution bins higher than 2.6 Å did not meet acceptable values. The unacceptable R-factor values for these high resolution bins were contributing to the overall high R-factor values, as these reflections were not as well measured as they appeared in data processing. To this end, data was re-processed, indexed, scaled, and merged to 2.6 Å. From this point forward, all refinement was carried out using a resolution range of 10 to 2.6 Å. In addition, we increased the sigma cutoff criteria for structure factors used in refinement, so that only reflections that were well measured were used for map and R-factor calculation. This criteria was carried out by accepting only $|F| \geq 2\sigma|F|$

In addition, our structure was characterized by high B-factors. The average protein B-factor was 65.25 Å$^2$. During our B-factor refinement some of the parameters were optimized to better refine the B-factors and improve R-factors. The target of B-factor refinement in CNS is not the R-factor, atom position, or B-factor value. B-factor refinement has a target which is defined as R.m.s. deviations from ideality of bonded main-chain atoms, bonded side-chain atoms, angle main-chain atoms, and angle side-chain atoms. These targets are typically set at 1.5 Å$^2$ for bonded main-chain, 2.0 Å$^2$ for bonded side-chain, 2.0 Å$^2$ for angle-main-chain, and 2.5 Å$^2$ for angle-side chain. These parameters have been used in CNS for historical reasons, but since their inception have been shown to need revision in certain circumstances. Yu & Karplus showed by means of molecular dynamics calculations that these restraints are too tight by about a factor of 2 for main-chain parameters and by a factor of 3 for side-chain parameters (Yu and Karplus, 1985). In addition, the constraints do not take into consideration absolute B-factor values; structures with large B-factors would require loosened R.m.s deviations.
An R.m.s deviation of 1.5 Å² for a structure with an average B-factor of 10 Å² makes sense, while for a structure with a 65 Å² average B-factor does not. To this end, we loosened our target restraints in individual B-factor refinement as follows: 3 Å² for bonded main-chain atoms, 6 Å² for bonded side-chain atoms, 4 Å² for angle main-chain atoms, and 7.5 Å² for angle side-chain atoms.

In addition, our previous water picking was unusual in two respects: very few waters were picked (although we could see clear density for more water in the density) and the R-factors went up if more than 103 waters were picked. To solve this problem two strategies were employed. Waters were picked by manual inspection and not relying on the automated water pick program of CNS (Brunger et al., 1998). In addition, once waters were picked their position was fixed and not refined. In the end, a total of 602 waters were picked.

A non-crystallographic symmetry restraint was applied towards the end of refinement, as hPCNA has near perfect three-fold symmetry. The NCS restraint was not very tight and was weighted as 100 in CNS (In CNS, NCS restraints range from 0 to 300 with 0 as not very tight and 300 as a very tight restraint, restraining the monomers as almost identical). The NCS restraint helped the free R-factor only minimally, less than 0.5%.

Using all of these strategies, several rounds of manual rebuilding followed by individual B-factor refinement and minimization refinement were carried out. The R-factors came down and the structure was able to now be refined to acceptable completion. The final R-factors were $R_{\text{free}}=27.83$ and $R=24.05$. The structure also had excellent geometry and no Ramachandran outliers. Refinement statistics can be found in Table
7.5. Some residues could not be modeled due to disorder and therefore not included in the final PDB coordinates. They are: A187-190, A258-261, B452, B465-466, C108, C186-190, C258-261, D452, D466, E187-192, E260-261, F452, and F466. The atomic coordinates and structure factors have been deposited as (1U76) for the hPCNA-pol-δ p-66 subunit residues (452-466).

7.2 hPCNA-FEN1(331:350) structure solution

Once the reflection file was detwinned as described in Chapter 6, molecular replacement and refinement proceeded smoothly in CNS (Brunger et al., 1998). 5% of the reflection file was set aside for \( R_{free} \) calculations. The reflection file with the removed Free R flags was subsequently used for molecular replacement and refinement. Monomer A of the published hPCNA-p21(139:160) (PDB:1AXC) structure was used as the search model for molecular replacement with the p21 peptide and water molecules removed. All refinement was carried using the hPCNA monomer; the biological trimer is formed by the symmetry operators of space group P6₃.
**Table 7.3 Cross-rotation solutions using the hPCNA-FEN1(331:350) data.**

<table>
<thead>
<tr>
<th>Index</th>
<th>theta 1</th>
<th>theta 2</th>
<th>theta 3</th>
<th>RF-function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>81.707</td>
<td>176.611</td>
<td>83.365</td>
<td>0.0772</td>
</tr>
<tr>
<td>6</td>
<td>7.500</td>
<td>1.659</td>
<td>22.428</td>
<td>0.0556</td>
</tr>
<tr>
<td>9</td>
<td>308.293</td>
<td>166.659</td>
<td>306.635</td>
<td>0.0541</td>
</tr>
<tr>
<td>12</td>
<td>94.928</td>
<td>153.389</td>
<td>40.000</td>
<td>0.0518</td>
</tr>
<tr>
<td>15</td>
<td>288.389</td>
<td>173.293</td>
<td>296.683</td>
<td>0.0509</td>
</tr>
<tr>
<td>16</td>
<td>89.622</td>
<td>6.706</td>
<td>104.551</td>
<td>0.0505</td>
</tr>
<tr>
<td>17</td>
<td>226.079</td>
<td>26.611</td>
<td>211.151</td>
<td>0.0493</td>
</tr>
<tr>
<td>18</td>
<td>209.492</td>
<td>6.706</td>
<td>224.421</td>
<td>0.0492</td>
</tr>
<tr>
<td>20</td>
<td>80.000</td>
<td>166.659</td>
<td>25.072</td>
<td>0.0490</td>
</tr>
<tr>
<td>26</td>
<td>194.807</td>
<td>16.659</td>
<td>179.878</td>
<td>0.0481</td>
</tr>
<tr>
<td>33</td>
<td>181.537</td>
<td>26.611</td>
<td>196.465</td>
<td>0.0466</td>
</tr>
</tbody>
</table>

*The hPCNA trimer (without p21-derived peptides and waters) was used as the search model (PDB:1AXC). The correct solution is shown in red. Data was produced in CNS (Brunger et al., 1998).*

---

**Table 7.4 Translation solutions using the hPCNA-FEN1(331:350) data.**

<table>
<thead>
<tr>
<th>Solution #</th>
<th>theta 1</th>
<th>theta 2</th>
<th>theta 3</th>
<th>trans X</th>
<th>trans Y</th>
<th>trans Z</th>
<th>Monitor</th>
<th>Packing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>71.73</td>
<td>176.91</td>
<td>72.03</td>
<td>11.69</td>
<td>-24.12</td>
<td>0.54</td>
<td>0.485</td>
<td>0.5607</td>
</tr>
<tr>
<td>2</td>
<td>60.61</td>
<td>1.02</td>
<td>324.76</td>
<td>5.70</td>
<td>11.61</td>
<td>-1.07</td>
<td>0.118</td>
<td>0.4883</td>
</tr>
<tr>
<td>3</td>
<td>286.02</td>
<td>167.69</td>
<td>282.35</td>
<td>36.46</td>
<td>-1.01</td>
<td>-1.13</td>
<td>0.122</td>
<td>0.4716</td>
</tr>
<tr>
<td>4</td>
<td>87.72</td>
<td>155.73</td>
<td>36.22</td>
<td>-50.46</td>
<td>46.06</td>
<td>3.58</td>
<td>0.094</td>
<td>0.5325</td>
</tr>
<tr>
<td>5</td>
<td>27.84</td>
<td>174.30</td>
<td>34.66</td>
<td>-17.49</td>
<td>0.26</td>
<td>-5.52</td>
<td>0.122</td>
<td>0.5008</td>
</tr>
<tr>
<td>6</td>
<td>181.21</td>
<td>4.47</td>
<td>12.81</td>
<td>-20.60</td>
<td>9.10</td>
<td>3.03</td>
<td>0.130</td>
<td>0.4231</td>
</tr>
<tr>
<td>7</td>
<td>224.26</td>
<td>22.42</td>
<td>216.26</td>
<td>-46.63</td>
<td>40.03</td>
<td>-2.09</td>
<td>0.092</td>
<td>0.4642</td>
</tr>
<tr>
<td>8</td>
<td>181.38</td>
<td>4.39</td>
<td>253.61</td>
<td>-10.15</td>
<td>-3.11</td>
<td>-1.89</td>
<td>0.132</td>
<td>0.5046</td>
</tr>
<tr>
<td>9</td>
<td>72.11</td>
<td>173.61</td>
<td>13.35</td>
<td>27.21</td>
<td>39.27</td>
<td>5.52</td>
<td>0.141</td>
<td>0.5146</td>
</tr>
<tr>
<td>10</td>
<td>219.78</td>
<td>15.82</td>
<td>148.65</td>
<td>36.03</td>
<td>-26.65</td>
<td>6.24</td>
<td>0.101</td>
<td>0.4185</td>
</tr>
</tbody>
</table>

*The hPCNA trimer (without p21-derived peptides and waters) was used as the search model (PDB:1AXC) with the cross-rotation list given in Table 7.3. The translation solution used for refinement is highlighted in red. Data produced in CNS (Brunger et al., 1998).*
Cross-rotation was carried out first. A fast direct rotation search was used extending from 15 to 4 Å. The top ten cross-rotation solutions are listed in the Table 7.3. The cross-rotation list was then used to carry out translation molecular replacement. A general method translation search was performed extending from 15 to 3 Å. One solution stood well above the others as seen by the monitor above the mean. The top ten translation solutions are listed in the Table 7.4. Solution #1 was chosen for further refinement. All refinement was carried out with a resolution range of 10 to 1.85 Å. 20 cycles of rigid body refinement were carried out first with the R-factors ending as \( R_{\text{free}} = 45.81 \) and \( R = 44.03 \). Next, 100 minimization cycles of refinement were carried out with R factors becoming \( R_{\text{free}} = 40.65 \) and \( R = 35.31 \). After minimization, group B-factor refinement was carried out and R-factors became \( R_{\text{free}} = 35.69 \) and \( R = 30.77 \). Slow cool simulated annealing starting at 2500 K was carried out and R-factors became \( R_{\text{free}} = 33.65 \) and \( R = 29.22 \). Individual B-factors were next refined and R-factors again dropped with \( R_{\text{free}} = 32.76 \) and \( R = 27.75 \). After individual B-factor refinement, 100 cycles of minimization refinement were carried out and R-factors became \( R_{\text{free}} = 32.53 \) and \( R = 27.61 \). At this point, density maps were made and included: Fo-Fc, 2Fo-Fc, 3Fo-2Fc, and composite omit density maps. At least 10 residues of the FEN1 peptide were visible in the Fo-Fc map, as well as the other density maps. Residues 337-343 of the FEN1 peptide were built into the model. More individual B-factor refinement was carried out as well as minimization refinement and R-factors came down to \( R_{\text{free}} = 30.02 \) and \( R = 25.38 \). From this point, several more rounds of manual rebuilding based on composite omit maps followed by individual B-factor refinement, simulated annealing, minimization refinement, and water picking was carried out. Refinement and R-factors
converged with $R_{\text{free}}=26.24$ and $R=21.31$ with a 131 water molecules. The first five residues of the peptide were disordered and could not be modeled. In addition, the last 2 residues of the peptide could not be modeled as well due to disorder; based upon electron density they may have been occupying more than one conformation. In addition, hPCNA residues 187-190 and 256-261 could not be modeled due to disorder. Refinement statistics are located in Table 7.5. The atomic coordinates and structure factors have been deposited as (1U7B) for hPCNA-FEN1(331:350).

7.3 Comparison of hPCNA structure quality

The only hPCNA-p21(139:160) co-structure and was published in 1996. The structure was solved to 2.6 Å with R-factors of $R_{\text{free}}=28.9$ and $R=19.2$. Because the $R_{\text{free}}$ and R-factors were ~10% apart, the structure was most likely hyper-refined and may have been subject to model bias. Crystallographic refinement methods have improved significantly since 1996, and subsequently our structures now present a higher quality hPCNA model. Our R-factors are lower than the hPCNA-p21(139:160) structure and not as hyper-refined. In addition, our 1.85 Å structure represents a much higher resolution view of hPCNA than was available previously.
Table 7.5 Refinement statistics for hPCNA-p66(452-466) and hPCNA-FEN1(331:350).

<table>
<thead>
<tr>
<th></th>
<th>hPCNA-p66</th>
<th>hPCNA-FEN1¹</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>10-2.6</td>
<td>10-1.85</td>
</tr>
<tr>
<td>$R_{\text{work}}$²</td>
<td>24.05</td>
<td>21.31</td>
</tr>
<tr>
<td>$R_{\text{free}}$³</td>
<td>27.83</td>
<td>26.24</td>
</tr>
<tr>
<td>Number of atoms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>6106</td>
<td>2037</td>
</tr>
<tr>
<td>Water</td>
<td>602</td>
<td>134</td>
</tr>
<tr>
<td>Average B-factor</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>65.25</td>
<td>18.59</td>
</tr>
<tr>
<td>Solvent</td>
<td>72.79</td>
<td>72.47</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bonds(A)</td>
<td>0.007364</td>
<td>0.005611</td>
</tr>
<tr>
<td>Angles(A)</td>
<td>1.28878</td>
<td>1.26151</td>
</tr>
<tr>
<td>Ramachandran analysis (%)⁴</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Most favored</td>
<td>89.5</td>
<td>89.1</td>
</tr>
<tr>
<td>Additionally allowed</td>
<td>9.7</td>
<td>10.5</td>
</tr>
<tr>
<td>Generously allowed</td>
<td>0.8</td>
<td>0.4</td>
</tr>
<tr>
<td>Disallowed</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

¹Values for hPCNA/FEN refinement are for data that was detwinned
²$R_{\text{work}} = \Sigma |F_o-F_c| / \Sigma |F_o|$ for all data with $F_o > 2 \sigma (F_o)$ excluding data to calculate $R_{\text{free}}$
³$R_{\text{free}} = \Sigma |F_o-F_c| / \Sigma |F_o|$, for all data with $F_o > 2 \sigma (F_o)$ excluded from refinement
⁴Calculated using PROCHECK
Chapter 8: Isothermal titration calorimetry (ITC)

8.1 Background of p21 PIP-box peptide ITC

Our studies were designed to compare and contrast several PIP-box peptides to correlate our structural studies with binding constants and thermodynamic properties. Because the p21-derived PIP-box peptide's binding had been measured by ITC previously, we used this experiment as a proof of concept that our material and methodology could be used to recreate the published binding constants and thermodynamic properties of the p21 peptide binding. Our first goal was to use the published conditions and methodology to reproduce the previous numbers with our own protein materials and ITC equipment. Only after we had accomplished this proof of concept, did we begin to experiment with the other PIP-box peptides.

Prior to our ITC studies, only one ITC study had been published studying hPCNA with PIP-box derived peptides (Zheleva et al., 2000). The study solely characterized the p21 PIP-box peptide interaction with hPCNA. The p21-derived peptide $^{141}$KRRQTSMTDFYHSKRRLIFS$^{160}$ (conserved PIP-box residues highlighted in red) was used at a concentration of 200 μM in the syringe and injected 15-30 times at 5 μL per injection into 10 μM hPCNA (hPCNA in PBS buffer) and heat changes were measured. Data was analyzed and fit using Origin software (MicroCal, LLC) and association constants as well as thermodynamic parameters were calculated. The association constant was found to be $1.14 \times 10^7$ M$^{-1}$ (Kd=87.7 nM), $\Delta H=-9.758$ kcal mol$^{-1}$, and $\Delta S=0.085$ cal mol$^{-1}$ °C. This agreed with studies that showed full length p21 bound
to hPCNA with a lower affinity (Kd=3-15 nM) (Knibiehler et al., 1996; Zheleva et al., 2000). Besides the wild type p21 peptide, two mutant peptides were also used in the work by Zheleva and colleagues: Met147Ala and Asp149Ala. The conserved PIP-box residue Met147, when mutated, abrogated binding to such a degree that binding constants could not be measured. The Asp149Ala mutant severely decreased binding affinity, and the association constant was measured to be 7.82 x 10^5 M^{-1} (Kd=1.28 μM) with ΔH=−11.270 kcal mol^{-1} and ΔS=−10.24 cal mol^{-1} °C.

8.2 ITC methodology

We used several PIP-box derived peptides in our ITC experiments, which include those derived from p21, FEN1, p66, RecQ5, and WRN. The RecQ5 and WRN derived peptides exhibited such weak binding that adequate heat changes were not observed and quality association constants and thermodynamic properties could not be derived. Two p66 derived peptides were used, p66a and p66b. p66a was derived from the native protein which terminates at residue 466. p66b was designed by adding the C-terminal residues of p21, 155RRLIFS160, to the C-terminus of p66a; these are the residues that form the anti-parallel beta-sheet with the interdomain connector loop of hPCNA and are thought to be important for binding (Gulbis et al., 1996). The p66b peptide would allow us to determine how much strength of association would be gained from the potential
Table 8.1 Peptides used in ITC experiments.

<table>
<thead>
<tr>
<th>Peptide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>p21a</td>
<td>GRKRRQTSMTDFYHSKRRLIFS</td>
</tr>
<tr>
<td>p21b</td>
<td>KRRQTSMTDFYHSKRRLIFS</td>
</tr>
<tr>
<td>p66a</td>
<td>GKNARQVSITGFQFRK</td>
</tr>
<tr>
<td>p66b</td>
<td>GKNARQVSITGFQFRKRRRLIFS</td>
</tr>
<tr>
<td>FEN1</td>
<td>RQGSTQGRVLDVFVTGSLSSA</td>
</tr>
<tr>
<td>WRN</td>
<td>EGDQWKLRLDFDILKLFN</td>
</tr>
<tr>
<td>RecQ5</td>
<td>VKEEAQNLIRHHFHGR</td>
</tr>
</tbody>
</table>

Peptides used in our ITC experiments. Conserved residues of the PIP-box are highlighted grey. Residues of p21 that form the anti-parallel β-sheet with the hPCNA are shown in black. The numbers to the left of the peptide sequence indicate the starting residue number in their native protein. p21a is the p21-derived peptide used in our experiments; p21b is the peptide used by Zheleva et al., 2000.

anti-parallel beta-sheet formation with the p66 derived peptide. All peptides used in our ITC experiments are shown in Table 8.1.

All ITC experiments were thermostated at 30°C and heat changes were measured using a MicroCal, LLC VP-ITC instrument (MicroCal, LLC). All peptide and hPCNA buffers consisted solely of PBS (phosphate-buffered saline). Data was analyzed and fit using Origin 7 software (MicroCal, LLC). In order to produce proper baselines, the heat of injection must be subtracted from experimental heat changes. To this end, peptide to PBS buffer injections using time scales and injection volumes equal to the experimental values were subtracted from experimental data prior to analysis and fitting with Origin 7 software (MicroCal, LLC).

For the p21 derived peptide, a1 μL aliquot followed by 29 2 μL aliquots of 400 μM p21 peptide were injected into 1.4 mL of 10 μM hPCNA. For the FEN1 derived peptide, 15 titrations of 400 μM FEN1 peptide were injected into 1.4 mL of 17 μM hPCNA. For p66a, 2 μL followed by 59 5 μL titrations of 200 μM peptide were injected.
into 1.4 mL of 10 μM hPCNA. For p66b, a 2 μL aliquot followed by 49 5 μL aliquots were injected into 1.4 mL of 10 μM hPCNA. The ITC curves and heat changes can be seen in Figure 8.1. Thermodynamic parameters for peptide association can be seen in Table 8.2.

8.3 Comparison of PIP-box peptide association constants and thermodynamic properties

Once the heat of peptide binding to hPCNA data was analyzed (Figure 8.1), we were able to compare and contrast binding affinities and thermodynamic properties of the various PIP-boxes to hPCNA. Fitting of the p21 binding data gave an association constant much higher than the other peptides (1.2 x 10^7 M^{-1}) and a stoichiometry of 1:1; this is in close agreement with the previously published data which used a very similar p21 derived peptide (1.14 x 10^7 M^{-1}) (Zheleva et al., 2000). p66a gave an association constant of 6.41 x 10^4 M^{-1} while p66b gave an association constant of 6.51 x 10^5 M^{-1}. This allowed us to conclude that although the C-terminal region of p21 that engages in an anti-parallel beta-sheet with hPCNA does contribute to binding affinity, it is not significant enough to explain the marked difference in affinity between p21 and p66 derived peptides. The FEN1 peptide showed an even lower association constant of 1.67 x 10^4 M^{-1}. Because FEN1 also contains a C-terminal region that could form an anti-parallel beta-sheet with hPCNA, it is
Figure 8.1 ITC measurements of binding for peptides derived from the PIP-box of p66, FEN1, and p21 to hPCNA. All data were fit to a one binding site per hPCNA monomer model. A Binding of peptide p21 to hPCNA. B Binding of peptide FEN1 to hPCNA. C ITC measurement of the binding of peptide p66a to hPCNA. D Binding of peptide p66b to hPCNA.
Table 8.2  hPCNA isothermal titration calorimetry with p21, p66a, p66b, and FEN1.

<table>
<thead>
<tr>
<th></th>
<th>p21</th>
<th>p66a</th>
<th>p66b</th>
<th>FEN1</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (stoichiometry)</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
</tr>
<tr>
<td>K (affinity constant)(M^{-1})</td>
<td>$1.2 \times 10^7$</td>
<td>$6.4 \times 10^4$</td>
<td>$6.5 \times 10^5$</td>
<td>$1.7 \times 10^4$</td>
</tr>
<tr>
<td>$\Delta G$ (free energy)(kcal mol^{-1})</td>
<td>-9.8</td>
<td>-6.7</td>
<td>-8.1</td>
<td>-5.9</td>
</tr>
<tr>
<td>$\Delta H$ (enthalpy)(kcal mol^{-1})</td>
<td>-29</td>
<td>-19</td>
<td>-10</td>
<td>-20</td>
</tr>
<tr>
<td>$\Delta S$ (entropy)(cal mol^{-1} K^{-1})</td>
<td>-64</td>
<td>-41</td>
<td>-7.1</td>
<td>-47</td>
</tr>
<tr>
<td>$-T\Delta S$ (kcal mol^{-1})</td>
<td>19</td>
<td>12</td>
<td>2.2</td>
<td>14</td>
</tr>
</tbody>
</table>

unlikely this region contributes much strength to association and, instead, the strength of association is found in the structural differences within the more conserved region of the PIP-box. The FEN1 peptide also shows a several fold difference in association from the full length FEN1 protein from which it is derived (the full length FEN1/hPCNA interaction Kd=0.02μM) (Gomes et al., 2000). Although we can interpret how well each PIP-box contributes to binding, we can not rule out the possibility that there are other portions of the full length proteins which make contacts with hPCNA and contribute to the overall strength of binding.

All peptides fit a 1:1 stoichiometry (1 peptide per hPCNA monomer), which follows previous research showing that the trimeric hPCNA can accommodate one PIP-box peptide per monomer (Zheleva et al., 2000). The p21 interaction showed a largely favorable $\Delta G$ value of -9.8 kcal mol^{-1}. The weaker binding peptides p66a, p66b, and FEN1 also showed favorable $\Delta G$ values of -6.7, -8.1, and -5.9 kcal mol^{-1}. p66a, p66b, and FEN1 all showed sizably less favorable $\Delta G$ values than p21, corresponding to their much lower affinity constants. The enthalpic ($\Delta H$) and entropic ($-T\Delta S$) terms contributing to Gibbs free energy for the peptide interactions are also presented in Table
8.3. The values of $\Delta H$ are largely negative for all the peptides. The largely negative enthalpic terms are the driving force for the favorable Gibbs free energy term. The $\Delta S$ terms are all disfavorably negative but compensated for by the largely negative enthalpy term. p66b has the least negative $\Delta H$ value (-10.2 cal mol$^{-1}$) but is compensated for by having the least negative $\Delta S$ value (-7.1 kcal mol$^{-1}$). Peptides with weaker binding affinities correlated with less favorable enthalpic terms; such changes in enthalpy have been shown to be associated with solvent reorganization and alterations in packing volume among protein-protein and protein-ligand interactions (Baldwin et al., 1998). Although definitive and accurate comparisons of the different peptide enthalpy values are complicated by differences in their primary sequences, the differences in packing volumes as seen in the corresponding crystallographic structures is discussed in Chapter 10.1.
Chapter 9: hPCNA-p66(452:466) and hPCNA-FEN1(331:160)

co-crystal structures

9.1 Overall structure comparison

Comparison of the overall hPCNA monomer structure of hPCNA-p66(452:466) with that of hPCNA bound to FEN1(331-350) or with p21(139-160) co-structures shows that hPCNA bound to all three peptides remains largely unchanged (Figure 9.1). Superimposition of 249 Cαs from the hPCNA monomer of the hPCNA-p21(139:160) structure to those of hPCNA-p66(452:466) and hPCNA-FEN1(331:350) resulted in root mean square distances (r.m.s.d.) of 0.6 and 0.9 Å, respectively. The most notable difference in the structures is the variability of the IDCL, which makes numerous contacts with the PIP-box peptides. The IDCL residues 119-134 of the hPCNA-p66(452:466) and hPCNA-FEN1(331:350) have an r.m.s.d. of 1.2 and 1.4 Å respectively, which contrasts to the low deviations for the hPCNA monomer as a whole. In both structures, regions of the IDCL were characterized by high temperature factors as a whole and electron density consistent with increased mobility. The structures of the PIP-box peptides are similar in topology (Figures 9.1, 9.2, 9.3) and consist of three segments: an extended N-terminal region, a central conserved region containing the hydrophobic residues and 310 helix, and a C-terminal region of variable length. The peptide extends across the interdomain connector loop with the peptide N-terminus forming contacts with the hPCNA C-terminus. The central conserved hydrophobic residues insert into a
hydrophobic plug between the interdomain connector loop and other internal loops (βC₁-βD₁ and βG₂-βH₂, these loops

**Figure 9.1** Overall structure comparison of hPCNA bound to peptides derived from human DNA polymerase-δ p66-subunit, FEN1, and p21 (Gulbis et al., 1996). A The trimeric hPCNA bound to the p66-subunit peptide (residues 452-466). hPCNA (green), p66-subunit (452-466) (red). B Stereo Cα trace of superimposed hPCNA co-structures hPCNA-p66 (452-466) with hPCNA (slate), p66 (yellow); hPCNA-FEN1 (331-350) with hPCNA grey, FEN1 (red); and hPCNA-p21 (139-160) with hPCNA (pink) and p21 (cyan). All hPCNA monomers showed similarity with r.m.s.d. values of less than 1.0 Å, although the interdomain connector loop (IDCL) of hPCNA (residues 119-134) showed appreciably more variability.
Figure 9.2 Stereo view of electron density from 2Fo-Fc maps (contoured at 1σ) of peptides derived from p66 and FEN1 overlayed on a ribbons diagram of the hPCNA monomer (gray). N represents the N-terminus of the peptide and C represents the C-terminus of the peptide. 

A  p66 peptide (453-465) bound to hPCNA. Only residues 453-465 could be modeled due to disorder at the termini. B  FEN1 peptide (335-349) bound to hPCNA. Residues 335-349 could be readily modeled whereas residues 346-349 could only be modeled at lower contour and suggest greater mobility.
are defined in Figure 1.4). Also, The C-terminal residues form contacts with the interdomain connector loop. Both peptides have a PIP-box motif (Q\textit{xx}(M/L/I)\textit{xx}F(Y/F) (Table 1.1) that makes good hydrophobic contacts from a $3_{10}$ helix to a binding pocket on PCNA and maintains a hydrogen bonding network. The buried surface area of the p66 and FEN1 peptides was 1,593 and 1,360 Å$^2$ corresponding to 88% and 80% of their respective total surface area. This is consistent with their role as docking peptides for sliding clamp-protein interactions as compared to other sliding clamp co-structures with similar binding surface interfaces discussed more thoroughly in Chapter 11.

### 9.2 p66 and FEN1 PIP-box Interactions

Although the overall fold of the p66 and FEN1 peptides are similar to the p21 peptide, interesting structural differences can be found in the details (Figure 9.3). The interactions are most easily compared in three segments: the N-terminal region, the $3_{10}$ helical region and the C-terminal region.

#### 9.2a Comparison of PIP-box N-terminal regions

Several residues of the N-terminal portion of p66 that were disordered in the homologous region of the FEN1 and p21 peptide (all residues but one N-terminal to the conserved PIP-box glutamine) are clearly seen in our structure as well as several residues of the C-terminal region of hPCNA (256-259) as seen in figure 9.4. Perhaps the most interesting aspect of the p66 peptide structure is the N-terminal region which turns in a
Figure 9.3 Molecular recognition surfaces of hPCNA in complex with peptides derived from the C-terminus of PCNA-interacting proteins p66, FEN1, and p21. The hPCNA monomer is shown in slate with the other contacting monomers of the hPCNA trimer shown in the background (wheat and salmon). The surface contacts of hPCNA within 5 Å of the interacting peptides are shown (lime). The PIP-Box containing peptides are displayed as yellow sticks with the conserved residues of the PIP-box highlighted in red. A hPCNA surface contacted with the p66 peptide (residues 453-465). B hPCNA surface contacted with the FEN1 peptide (residues 335-349). C hPCNA surface contacted with the p21 peptide (residues 139-160).
different direction than seen in the FEN1 and p21 peptides and snakes around the C-terminal region of hPCNA (Figure 9.4). This looping around the C-terminus of hPCNA is supported by Arg 455 and Asn 454 of p66 which form main-chain hydrogen bonds with main-chain hPCNA C-terminal residues while p66 residue Ala 453 makes no contacts with hPCNA. The conserved glutamine of all three PIP-box peptides form an almost identical conformation and hydrogen bonding scheme (Figure 9.5). This conserved glutamine residue of the PIP-box makes two hydrogen bonds from it's side-chain to a water molecule which is in turn hydrogen bonded to the main-chain nitrogen of hPCNA Ala 208. Both the FEN1 and p66 peptides differ from the p21 residue Thr 145 at their homologous residues (seen in Figure 10.2A). A Gly is found at this position in FEN1, while p66 is occupied by a Val; both peptides lack the hydrogen bond here that forms from the OH group of Thr 145 to the backbone carbonyl of Pro 253 in the p21 structure. Finally, the side chain of Arg 339 of the FEN1 peptide extends much deeper into the solvent region than that of p21 Ser 146 and p66 Ser 458, making a contact with a water molecule.
Figure 9.4 The p66 peptide N-terminus wraps around C-terminus of hPCNA. hPCNA shown as a wheat-colored surface representation and the p66 peptide as a stick representation. This portion of the hPCNA C-terminus
Figure 9.5 Superimposition of conserved PIP-box Gln from p66(blue), FEN1(green), and p21(yellow). Hydrogen bonding is shown by red dots and waters shown in cyan. All three PIP-box Gln residues are nearly identical in position and form similar hydrogen bonds to the water and hPCNA Ala 252.

9.2b Comparison of PIP-box $3_{10}$ helical regions

Although the p66, FEN1, and p21 all form a $3_{10}$ helix, the side-chains show notable differences. The two conserved PIP-box residues of FEN1 and p66 that insert into the hydrophobic pocket of hPCNA have primary sequence differences that have significant structural consequences. Conserved residue Tyr151 in p21 forms two
hydrogen bonds from its OH group: one to a water molecule and one to the amino group of the hPCNA Gln131 side-chain, strengthening its placement in the hydrophobic core. In p66 and FEN1, the equivalent residues are both phenylalanine and therefore lack the ability to make these two contacts. This is shown in Figure 10.2B. Furthermore, the non-aromatic conserved PIP-box residue of p21 is Met 147, while in p66 it is Ile 459 and in FEN1 it is Leu 441. Because p66 and FEN1 lack a sulfur at this site it packs less deeply and fills less space due to its smaller size. This is explored more in depth in chapter 10.

9.2c Comparison of PIP-box C-terminal regions

Finally, both the p66 and FEN1 peptides do not form a C-terminal beta sheet anti-parallel to the interdomain connector loop of hPCNA. Although p66 residue Gln 456 forms a hydrogen bond from its main-chain carbonyl to hPCNA Gly127 main-chain nitrogen as does p21, p66 does not and cannot form an extended anit-parallel β-sheet with hPCNA because of its biological termination at residue Lys 458. In addition, the two residues C-terminal to the conserved aromatic PIP-Box residue of p66 form different interactions than FEN1 and p21: the rotamer of p66 Gln 464 flips in an opposite direction of the homologous FEN1 and p21 structure and Arg 465 is not making any visible contacts with hPCNA as is the homologous residue in p21. The C-terminus of FEN1 begins to form an anti-parallel beta sheet with the interdomain connector loop of hPCNA but then becomes disordered, much as does the p66 peptide prior to its termination. Lys 345 of FEN1 finds itself in nearly the same position as the homologous p21 His 152 which differs from the p66 Gln 464. Lys 345 of FEN1 forms a carbonyl mediated
hydrogen bond with hPCNA main-chain nitrogen Gly 127. Although the following C-terminal residues Val 346, Thr 347, and Gly 348 of FEN1 could be modeled in a position parallel to the interdomain connector loop of hPCNA, no hydrogen bonding contacts consistent with beta-sheet formation are seen. Furthermore, FEN1 residues 346-348 appear to be making no contacts at all with the hPCNA interdomain connector loop, which may account for their disorder.

9.2d Conclusions

In conclusion, the p66, FEN1, and p21 PIP-box peptides share several differences in primary sequence, hydrogen bonding schemes, ionic interactions, hydrophobic interactions and packing volumes, the significance of which will be discussed in the following chapters. Due to primary sequence differences, FEN1 and p66 lack hydrogen bonding capacity from residues homologous to p21's residues Thr 145 and Tyr 151. In addition the 3$_{10}$ helix packs much less tightly against the hPCNA surface in the FEN1 and p66 structures. Finally, p66 and FEN1 lack the anti-parallel beta-sheet and ionic interactions found in their C-terminal region.
Chapter 10: p21 is a model for a high affinity PIP-box

10.1 $3_{10}$ helix packing volumes of p66, FEN1, and p21 PIP-box peptides

The $3_{10}$ helix appears to be the only universally conserved structural motif of the PIP-Box peptide family. The $3_{10}$ helix interacts with hPCNA through a largely hydrophobic molecular recognition surface and has been referred to as a "hydrophobic plug" that fits into hPCNA using the conserved hydrophobic residues of the PIP-Box consensus sequence (Figure 10.1). As shown in Figure 10.1, the $3_{10}$ helices of the pol-δ p66-subunit and FEN1 PIP-boxes nestle into the hPCNA surface with substantially less packing efficiency than that of p21. The conserved hydrophobic side chains Ile 459 of p66 and Leu 340 of FEN1 do not fill the hydrophobic pocket of hPCNA as effectively as the methionine side chain of p21. Gomes et. al (2000), showed that mutating hPCNA residues Leu 126 and Ile 128 of the hydrophobic pocket to hydrophobic residues with smaller packing volumes significantly decreased hPCNA's ability to interact with FEN1 in vitro. Additionally, the second conserved aromatic residue of the PIP-box of p66 and FEN1 is a phenylalanine. The phenylalanine side chains of p66 and FEN1 insert into hPCNA but cannot make the specific hydrogen bonds made by the Tyr 151 of p21 to Gln 131 and via a water to Tyr 133 of hPCNA (Figure 10.2). The edge of the hydrophobic
pocket of hPCNA is partially occupied by the polar residues Tyr 133 and Gln 131, and in our structure, the non-polar phenylalanine side chains of the PIP-Box peptides pack less efficiently than the analogous Tyr 151 of p21. Packing space and solvent exposed surface area are good indicators of stability within a hydrophobic pocket (Baldwin et al., 1998).
Figure 10.1 Packing of the conserved $3_{10}$ helix of p66 (452-466) and FEN1 (331-350) with hPCNA. Conserved hydrophobic residues of p66 and FEN1 (red) of the PIP-Box pack more loosely with the hydrophobic core of hPCNA than does the comparable PIP-Box peptide of p21. Gaps between the conserved hydrophobic residues which insert into the hydrophobic core and hPCNA were calculated for p66, FEN1, and p21 using the program SURFNET (Laskowski, 1995). To illustrate the difference in packing between p21 (139-160) and FEN1 (331-350) and p66 (452-466) their gap volumes were subtracted and are shown in green. PIP-interacting peptides are shown as yellow sticks with the conserved hydrophobic residues show in red. hPCNA surface is shown in wheat. A The p66 PIP-box gap volume (with p21 gap volume subtracted) B The FEN1 PIP-box and gap volume (with p21 gap volume subtracted)
The gaps between the PIP-box residues of the $3_{10}$ helix and hPCNA were calculated for the p66-hPCNA, FEN1-hPCNA, and p21-hPCNA co-structures and were found to be 378.8, 456.6, and 307.5 Å$^3$, respectively; these calculations show a trend that correlates the larger gap size in the hydrophobic pocket with lower binding affinity. The buried surface area and packing density of the respective $3_{10}$ helices shows that p21 is a better match to the molecular recognition surface of hPCNA (Figure 10.1).

10.2 Hydrogen bonding and ionic interactions

Although the overall fold of the p66, FEN1 and p21 peptides bound to hPCNA is similar, subtle differences can be found in the details that can explain the respective affinities of these PIP-Boxes for hPCNA. Important determinants for the affinity of the PIP-Box sequences to PCNA arise from differences in primary sequence within the PIP-Box and flanking sequences that govern their in vivo interaction.

Our structures show that the C-terminal regions of p66 and FEN1 do not support an extensive β-sheet structure, as seen in the PCNA-p21 co-structure. The C-terminus of FEN1 begins to form β-sheet-like interactions with the IDCL of hPCNA but then becomes disordered. In all three hPCNA co-structures, the residue following the last conserved PIP-Box aromatic residue makes hydrogen bonds to Gly-127 of hPCNA, and in p21, this residue is at the start of the extended β-sheet structure. This residue’s position is anchored next to the conserved aromatic residue, but it is the last well-ordered contact point in the hPCNA co-structures with p66 (452-466) and FEN1 (331-350). Our
crystallographic and ITC results with pol-δ p66 and FEN-1 C-terminal peptides suggest that the high affinity to hPCNA and extended β-sheet of the p21 C-terminus might be a special case among proteins containing the PIP-Box motif.

The role that the extended β-sheet, C-terminal, 155RRLIFS160 sequence of p21 plays in increasing its affinity for hPCNA was investigated by extending the p66 peptide (p66a) to include these additional sequences (p66b) (Table 8.2) and measuring the effect on its affinity for hPCNA. As seen in Table 8.2, extending the p66-subunit sequence to include the residues of p21 does increase affinity for hPCNA 10-fold and suggests that this region is at least partly responsible for the higher affinity of the p21 PIP-Box.

A combination of binding data and structural comparisons suggests that the interaction of hPCNA with p21 is exceptionally tight and that p21 makes several contacts to hPCNA not typical of most PIP-Boxes (see discussion of 3₁₀ helix interactions). Binding and structural studies of the p21 (139-160) peptide clearly demonstrate that it binds with high affinity (82.6 nM) and makes more extensive contacts with hPCNA than the analogous peptide from the p66-subunit and FEN1.

There is also better overall electrostatic complementation between the p21 C-terminus and the hPCNA IDCL than that of the p66 or FEN1 PIP-Box motifs (not shown). The FEN-1 C-terminus has two uncharged residues (Gly-348 and Ser-349) in positions that are occupied by two arginines in p21 (Arg-155,156). Arg-155 and -156 of p21 make ionic interactions to hPCNA resulting in additional affinity. In principle, FEN-1 should be able to form a good, anti-parallel, β-sheet which suggests that ionic and hydrophobic interactions, and not the formation of a β-sheet may be responsible for the increased affinity of p21 through this region. Collectively, the additional p21 sequences
increased the affinity of p66 for hPCNA 10-fold and suggests that the subtle differences in packing to the PCNA surface, and formation of a hydrogen bonding network between p21 residues Tyr-151 and Thr-145 and hPCNA shown in Figure 10.2 make up the remaining 20-fold increase in affinity.

Figure 10.2 Alignment of PIP-box structures in hydrogen bonding networks. p66 is shown in blue, FEN1 in green, p21 in yellow, and waters in cyan. A The N-terminal Thr 145 of p21 makes a hydrogen bond that p66 Val 457 and FEN1 Gly 338 cannot make to the hPCNA proline main-chain carbonyl. B The conserved aromatic residue of the PIP-box, Tyr 151 in p21, is a phenylalanine in the homologous residues of the p66 and FEN1 structures and cannot make hydrogen bonds. This also causes the boundaries of the hydrophobic pocket to be extended outward slightly.
We propose that the variable PIP-Box sequences found in many proteins will likely result in a range of protein to PCNA affinities (at least 200-fold) that may directly mediate \textit{in vivo} interactions necessary for the proper regulation of DNA replication and repair.
Chapter 11: PIP-boxes have a small contact surface with PCNA

Comparing all sliding clamp and PIP-box protein co-crystal structures raises an important question as to the number of amino acids flanking the PIP-box motif required for binding. Based upon co-structures of eukaryotic sliding clamps with PIP-Box containing proteins (yeast PCNA-RFC) (Bowman et al., 2004) or peptides (p21, p66, FEN-1) (Gulbis et al., 1996), the typical molecular recognition surface on PCNA is small and is contacted by 11-13 residues. Superimposition of the $3_{10}$ helices of the known PIP-Box containing peptides clearly shows that the N- and C-terminal regions are typically disordered and/or break away from the sliding clamp surface (Figure 11.1). Both our human polymerase as well as the prokaryotic polymerase structures show that the C-terminal region, which in the hPCNA-p21(139:160) form the anti-parallel beta-sheet with the interdomain connector loop, are inherently lacking the extended beta-sheet. This is also seen in our human FEN1 structure as well as the Archae FEN1 structure. The Archae FEN1 structure lacks most of the C-terminal region due to its experimental design and most of our human FEN1 C-terminal region is disordered and not making any clear beta-sheet structure. FEN1 may or may not make this C-terminal beta-sheet with PCNA, but our data suggests that it does not. The recent yeast RFC-PCNA structure is the only eukaryotic sliding clamp co-structure with a full length protein (Bowman et al., 2004). The PIP-Box motif of RFC is internal, not at the C-terminus. RFC makes contacts to PCNA through 10 residues and is in excellent agreement with our co-structures. Of course p21 does make use of an extended region flanking the PIP-Box to contact PCNA,
but as we have discussed earlier, this appears to be fairly unusual. Furthermore, the extent
to which the N-terminal region of PIP-box peptides makes important contacts with PCNA
is also unclear. As shown in Figure 11.1, only two-three residues N-terminal of the
conserved glutamine make contacts with PCNA in all structures. In some cases this is
due to disorder as with our FEN1 structure and the p21 structure, while in other cases it is
due to the experimental design of the peptide as in the case of the Archae FEN1 peptide
and our p66 peptide. Perhaps the area just N-terminal of the conserved PIP-Box Gln is a
breaking point for the PIP-box proteins wherein it halts contact with the sliding clamp
and positions itself back into the solution. The PIP-Box fulfils its function of providing a
specific protein-protein contact to PCNA through a small stretch of residues whose
affinity for PCNA can be dramatically affected by even modest changes in sequence.
Figure 11.1 Superimposition of the PIP-box regions from six PIP-Box to sliding clamp co-crystal structures. The superimposition shows that the region of sliding clamp interacting proteins that contacts the sliding clamp is only eleven to thirteen residues, outside of which the proteins diverge widely. Pol-δ p66-subunit (453-465) (cyan), human FEN1 (336-348) (magenta), p21 (139-160) (PDB:1AXC) (lime), yRFC (388-407) (PDB:1SXJ) (yellow), RB69 DNA polymerase (893-903) (PDB:1B77) (orange), and Archaeoglobulus fulgidus FEN1 (PDB:1RXM) (red).
Chapter 12: Comparison of DNA polymerase-sliding clamp interactions

12.1 The Pol IV/DinB co-structure

To date, only one crystal structure with a full-length polymerase bound to a sliding clamp exists: that of the *E. coli* translesion polymerase DinB with the *E. coli* sliding clamp (β) seen in Figure 12.2 (Bunting et al., 2003). The *E. coli* β clamp is a dimer and the crystal structure shows one polymerase bound to each monomer (Figure 12.1). The conserved motif of the *E. coli* polymerase for interaction with the clamp is located at the extreme C-terminus of the polymerase: \(3_{46}^{QLVLGL_{351}}\), with residue Leu 351 being the last residue of the polymerase. The conserved motif is completely extended and lies within a cleft located between the second and third subdomains of the β clamp subunits. The major contacts of the interaction motif reside only within 10 residues and consist mainly of hydrophobic interactions (Figure 12.2). Conserved residues Leu 349 and Leu 351 of the polymerase are buried within a hydrophobic cavity of the β clamp (Leu 155, Thr 172, Leu 177, Pro 242, Val 247, Val 360, and Met 362). Gly 350 makes no interactions but acts as a tight bridge that positions Leu 349 and Leu 351 in close proximity allowing a hydrophobic contact between their side-chains. Leu 351 of the polymerase makes an electrostatic interaction with Arg 152 of the β clamp. Val 348 of the polymerase makes no interactions. Leu 347 of the polymerase also buries its side-chain into a hydrophobic cleft of the β clamp (Val 344, Met 362, Pro 363, and
Arg 355) as does polymerase residue Met 343 (buried by Gly 280, Val 281, Arg 282, Gly 318, Met 364, and Leu 365 of the β clamp). Finally, the side-chain of the conserved glutamine of the polymerase is buried and makes only one hydrogen bond to the carbonyl of Met 363 of the β clamp.

12.2 Comparison of the prokaryotic and human clamp interactions

Interesting differences are seen between the interactions of hPCNA and the DNA polymerase-δ p66-subunit, as compared to the *E. coli* pol IV/Dinb and *E. coli* sliding clamp (β-subunit) structure (Bunting et al., 2003). Although the two sliding clamps share a toroidal shape, their primary sequences are very divergent. While the *E. coli* clamp can accommodate only two polymerases, the human PCNA has the potential to accommodate three polymerases due to its trimer nature. In addition, the primary sequence and structure of their PIP-boxes share considerable divergence. The interaction of our DNA polymerase δ p66-subunit (452-466) with PCNA more closely resembles that seen for hPCNA with p21(Gulbis et al., 1996), FEN1, and RFC (Bowman et al., 2004) and for the bacteriophage RB69 sliding clamp (Shamoo and Steitz, 1999) to its cognate DNA polymerase.

The conserved C-terminal region of *E. coli* polIV that interacts with the β-subunit (346QLVLGL351) has very little sequence homology with the C-terminal PIP-Box motifs of other proteins (Warbrick, 2000). The overall topology of the interacting region of the pol IV/Dinb polymerase is almost completely extended and lacks the 3₁₀ helix in its central conserved region that is found in all other sliding clamp co-structures. The pol
IV/DinB C-terminus binds within an extended cleft of the β-subunit as opposed to the p66 peptide which follows the pattern seen in eukaryotic and bacteriophage RB69 PIP-Boxes, where the $3_{10}$ helix plugs into a small hydrophobic patch on PCNA. Although the interactions of *E. coli* pol IV/Dinb and the β-subunit are also hydrophobic, they differ from those made between p66 and the PIP-Box consensus sequence. The hydrophobic clefts of the β-clamp are also larger and comprised of a larger number of residues. The pol IV/Dinb tail has no conserved phenyl group as does p66 (Phe463), and forms three important buried hydrophobic interactions (Leu349, Leu 351, and Met 343) whereas p66 only has two (Phe463 and Iso459). p66 also has a conserved Phe 462 which blocks the hydrophobic cleft from the solvent which is lacking in the *E. coli* clamp structure. The pol IV/DinB C-terminus resides within an extended depression of the β-clamp whereas the p66 c-terminus has several more solvent exposed regions along the extremities of the peptide. Also, the N-terminus of the prokaryotic PIP-box does not twist around the C-terminus of the sliding clamp as does our human structure. Finally, although the prokaryotic polymerase has a conserved glutamine making interactions, the environment and number of hydrogen bonds made differs from that of the human PIP-box conserved glutamine. The differences between the human and prokaryotic systems underscore the importance of increasing our structural understanding of the human system. The structural differences between sliding clamp interactions of Eukaryotes and Prokaryotes may be useful for the development of novel antibiotics.
Figure 12.1 Crystal structure of the *E. coli* beta-clamp (monomers shown in blue and green) with the pol IV/DinB(LF domain) shown in red. The prokaryotic clamp interaction domain is located at the C-terminus of polIV/DinB and is seen contacting the surface of the clamp (PDB:1UNN) (Bunting et al., 2003).
Figure 12.2 Contact surface of polIV/DinB with the beta-clamp. Conserved residues of the clamp interaction motif are colored red, the clamp is illustrated as a surface representation (brown) with the area located within 5 Å of polIV/DinB highlighted in green. PolIV/DinB is shown as a stick representation. The contacts are quite different from the PIP-box/hPCNA interaction (PDB:1UNN) (Bunting et al., 2003).
Chapter 13: Biological significance of hPCNA-p66(452:466), hPCNA-FEN1(331:350), and hPCNA-p21(139:160)

With the addition of our two structures, there now exist only three structures of PIP-box peptides with the human sliding clamp. No full-length PIP-box containing protein in complex with the human sliding clamp exists. The question therefore arises, although the PIP-box of FEN1 and p66 are necessary for interaction, are they sufficient? Are contacts outside of the PIP-box, in the case of FEN1 and p66, required for in vivo interactions?

Although the PIP-box has been shown to be sufficient for in vivo interactions in the case of p21, human RFC, and DNA Ligase I, whether the dozen or so PIP-Box residues alone are sufficient for the interaction of PCNA and its other interacting partners, such as p66 and human FEN1, remains the subject of investigation (Maga, 2003). Proteins can certainly make additional and extensive contacts to a sliding clamp. For example, the E. coli polIV/Dnb clearly interacts with the β-subunit at two distinct contact surfaces (Bunting et al., 2003). However, the recent yPCNA to RFC co-structure (Bowman et al., 2004) shows only interactions through the PIP-Box motif, which suggests that further contacts are not essential.

Previous studies have shown that PCNA loaded onto DNA stimulates the enzymatic activities of DNA ligase (Tom et al., 2001), FEN1 (Tom et al., 2000), and the processivity of DNA polymerase δ (Ducoux et al., 2001), whereas PCNA free in solution does not (Frank et al., 2001; Gomes and Burgers, 2000). Paradoxically, deletion of the PIP-Box motif typically abolishes these stimulatory effects, although mutation of the
region can produce little or no effect (Frank et al., 2001; Gomes and Burgers, 2000). For example, Frank et al. showed that mutating the conserved phenylalanine residues of the human PIP-box abolished the FEN1-hPCNA interaction \textit{in vitro}, while the same mutations \textit{in vivo} had little effect on hPCNA-mediated stimulation of the FEN1 nuclease activities. These findings suggest that the \textit{in vivo} binding of PCNA to other proteins results not only from direct PCNA-protein interactions through the PIP-Box motif, but also another set of contacts that occur when the sliding clamp is loaded onto DNA. One possibility is that additional contacts exist on PCNA that are outside the PIP-Box binding surface, as suggested by site-directed mutagenesis studies on yPCNA (Gomes, and Burgers, 2000). Based on our work, earlier binding studies (Tom et al., 2000), and the structure of yPCNA-RFC, we favor an alternate model, one in which the second contact is between the protein interacting with PCNA and DNA. DNA polymerases, ligases and repair enzymes possess both specific binding affinities for their substrates, as well as non-sequence specific affinity for DNA in general. DNA polymerases are a well-characterized and familiar example. \textit{In vitro} studies have shown that most replicative DNA polymerases have nanomolar affinity for the primer-template junction of DNA but much weaker binding to duplex DNA, often less than micromolar. The interaction of PCNA with DNA polymerase should be regarded from the \textit{in vivo} situation, where PCNA is already loaded on DNA. Under these circumstances, DNA polymerase’s affinity for DNA and its processivity is dramatically enhanced by the micromolar PIP-Box to PCNA interaction since the free energies of the affinity of DNA polymerase with DNA and with PCNA will be effectively additive. Once loaded, the sliding clamp is topologically trapped, and the local concentrations of PCNA and DNA polymerase are much higher
than when they are free in solution (i.e. not bound to DNA). The overall effect of the p66-
subunit’s PIP-Box interaction to DNA polymerase has the potential to increase pol-δ’s
affinity for DNA 64,000-fold. This hypothesis is also supported by kinetic studies that
show that PCNA is able to lower the $K_M$ of FEN1 for substrate by stabilizing FEN1
binding at its site of action (Tom et al., 2000). This model suggests that additional
contacts to PCNA are unnecessary to explain in vivo function. Even small changes in
sequence among PIP-Box containing regions can result in broader binding partner
preferences in vivo. In this context, the PIP-Box motif may function as a readily
adaptable docking peptide to mediate the wealth of PCNA interactions with DNA
replication and repair proteins without requiring more elaborate recognition surfaces.

These questions will not be adequately answered until crystal structures of full-
length PIP-box proteins bound to the human sliding clamp are brought to light. The
difficulty of hPCNA purification and crystallization has most certainly impeded progress
of crystal structure solution in the human sliding clamp field. This problem is
compounded by the addition of DNA in clamp containing crystallization experiments.
Clamps must be loaded unto DNA by clamp loaders. In addition, clamps are by nature
highly mobile once loaded onto DNA. High mobility often impedes crystallization. In
short, human sliding clamp structural biology faces extreme challenges.
Chapter 14: PCNA: a putative anti-cancer drug target

Nearly 200,000 people will die of lung cancer in the next year, with breast cancer, prostrate cancer, and colorectal cancer close behind in combined mortality rates (Greenlee et al., 2001). The large numbers of deaths resulting from these types of cancer on a yearly basis imply that current treatments are not entirely effective. Mortality rates show that new targets for chemopreventive drugs and tumor suppressing drugs are in dire need.

PCNA has been shown to have a pivotal role in carcinogenesis. Although PCNA has not been found to have mutations in cancer cells, its expression patterns change in cancer cells (Khurti et al., 2001; Kawahira, 1999; Iatropoulos and Williams, 1996; Minardi et al., 2004). PCNA expression is largely up-regulated in cancer cells to accommodate the increased need for DNA synthesis. In fact, PCNA up-regulation is so widespread among different cancer types that it has become a marker of dysregulated proliferation as well as a key identifier of carcinoma; cancer tissues can be identified by being exposed to immunohistochemical staining of PCNA using anti-PCNA monoclonal antibodies. Perhaps the most interesting aspect of PCNA up-regulation in cancer cells is that it is so widespread among various cancer types, identifying it as a very fundamental aspect of cancer biology. PCNA up-regulation has been identified in nearly all cancers it has been studied including leukemia, lung, melanoma, prostrate, lymph node, breast, and colorectal cancers.

Our research has opened the door for new cancer drug development. We propose the PIP-box binding surface of PCNA represents an excellent cancer drug target. Cancer
cells require increased levels of DNA synthesis to propagate themselves. p21 has been shown to block DNA synthesis by its interaction with PCNA. Our research has shown what aspects of p21's interaction on a structural level allow for it to block the p66 interaction required for DNA synthesis. It was only after structure solution of our PCNA-p66 structure that the p21 and p66 PIP-box interactions could be compared in order to understand this competition on a structural level. The structural aspects of p21's interaction which allow for p66 functional inhibition on the PCNA surface could be used in drug development. Inhibitor drugs of this nature would bind PCNA, blocking p66 interaction and function, subsequently preventing DNA synthesis and cancer cell proliferation.

Several other aspects of PCNA cell biology make it a good cancer drug target. First, PCNA has no downstream regulation: PCNA interacts directly in DNA synthesis. In contrast, proteins such as p53, shown to be one of the most highly mutated proteins in cancer cells and the target of many cancer studies, have many downstream targets. Cancer drug development which targets highly regulated proteins is complicated because cancer cells can compensate for it's activity in pathways downstream. In contrast, p66 and PCNA are directly involved in DNA regulation and cannot be compensated for by other mechanisms. Finally, PCNA represents a good cancer drug target because it is found upregulated in so many different cancer cells, allowing development of a broad use cancer drug. It is often the case that different cancers have characteristically different protein mutations, making cancer drugs that target these proteins useful only to specific types of cancer.
References


