Investigations into the Architecture and Function of the Bacterial Replisome

Nicholas Horan
University of Wollongong

UNIVERSITY OF WOLLONGONG
COPYRIGHT WARNING

You may print or download ONE copy of this document for the purpose of your own research or study. The University does not authorise you to copy, communicate or otherwise make available electronically to any other person any copyright material contained on this site. You are reminded of the following:

This work is copyright. Apart from any use permitted under the Copyright Act 1968, no part of this work may be reproduced by any process, nor may any other exclusive right be exercised, without the permission of the author.

Copyright owners are entitled to take legal action against persons who infringe their copyright. A reproduction of material that is protected by copyright may be a copyright infringement. A court may impose penalties and award damages in relation to offences and infringements relating to copyright material. Higher penalties may apply, and higher damages may be awarded, for offences and infringements involving the conversion of material into digital or electronic form.

Recommended Citation

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au
Investigations into the Architecture and Function of the Bacterial Replisome

A thesis submitted in fulfilment of the requirements for the award of the degree

Doctor of Philosophy

from

the University of Wollongong

by

Nicholas Horan

Bachelor of Biotechnology (Honours)

School of Chemistry

2016
# TABLE OF CONTENTS

SUMMARY.................................................................ix

ACKNOWLEDGMENTS..................................................xi

DECLARATION..........................................................xiii

PUBLICATIONS............................................................xiii

ABBREVIATIONS............................................................xiv

Chapter 1  GENERAL INTRODUCTION..............................1

1.1 DNA replication in *E. coli*........................................2

1.2 Distribution and organisation of the bacterial chromosome ....3

1.3 Replication of DNA....................................................4

  1.3.1 Initiation.............................................................4

  1.3.2 Elongation..........................................................7

    1.3.2.1 Okazaki fragment synthesis..................................9

    1.3.2.2 Cycling and signalling on the lagging strand.................10

    1.3.2.3 Efficiency and adaptability of the *E. coli* replisome ....12

  1.3.3 Termination.......................................................13

1.4 *E. coli* DNA polymerases.........................................15

1.5 *E. coli* DNA polymerase III......................................16

  1.5.1 Structure and organisation of components......................17

    1.5.1.1 The α subunit.................................................17

    1.5.1.2 The ε subunit..................................................22

    1.5.1.3 The θ subunit..................................................24

    1.5.1.4 The βε subunit.................................................26

    1.5.1.5 The clamp-loader complex......................................29

    1.5.1.6 Single-stranded DNA binding protein (SSB)....................33

1.6 Studying complex multi-protein machines........................34

1.7 Scope of Work......................................................36

Chapter 2  GENERAL MATERIALS AND METHODS.....................37

2.1 General Materials................................................38

  2.1.1 Bacterial strains.................................................38

  2.1.2 Plasmid vectors and expression systems........................39

  2.1.3 Chemicals and reagents.........................................41

2.1.4 Growth media....................................................43

    2.1.4.1 LB liquid medium.............................................43

    2.1.4.2 LB solid medium (agar plates)..............................43

    2.1.4.3 LB autoinduction medium....................................43

    2.1.4.4 Minimal medium for $^{13}$C,$^{15}$N labelling................44

2.2 General Methods..................................................44
Chapter 3  STRUCTURE OF THE E. COLI ε PROOFREADING SUBUNIT C-TERMINAL SEGMENT IN COMPLEX WITH THE α POLYMERASE PHP DOMAIN .......................... 55

3.1 Introduction .................................................................................................................. 56
3.2 Materials and Methods, .............................................................................................. 61
  3.2.1 Oligonucleotide primers, ...................................................................................... 61
  3.2.2 Plasmids for protein expression, .......................................................................... 61
  3.2.3 Protein purification buffers, .................................................................................. 62
  3.2.4 Purification of ε:α270 variants, ............................................................................. 63
3.2.5 Growth of protein crystals ...................................................................................... 67
3.3 Results .......................................................................................................................... 68
  3.3.1 Data collection and refinement ............................................................................. 68
  3.3.2 An atomic structure of the α–ε interaction ............................................................ 71
  3.3.3 The Extended Construct Reveals Additional Sites of Interactions ................. 72
  3.3.4 Positioning the εCTS within the Context of the α Subunit .................................. 75
  3.3.5 Inversing the order of linked domains induces the formation of domain-swapped multimers ............................................................ 76
3.4 Discussion .................................................................................................................... 78
3.4.1 Protein purification and crystallography trials ........................................... 78
3.4.2 The extent of the α-ε interaction can be confidently defined ....................... 80
3.4.3 The αε270ε209 ‘reverse’ construct ................................................................ 81
3.4.4 A crystal structure allows for informed model building ......................... 82

Chapter 4    MODELLING THE STRUCTURE OF THE E. COLI
β2αεθ REPLICATION COMPLEX ................................................................. 83
4.1 Introduction ................................................................................................. 84
  4.1.1 Core:clamp conformational changes during replication .......................... 86
  4.1.2 Generating a structural model of β2αεθ .................................................. 87
  4.1.2 Refining the position of the ε N-terminal domain .................................. 88
  4.1.3 Verification of structural model by SAXS ............................................. 89
  4.1.4 Stabilisation of the complex for crystallography ................................... 90
4.2 Materials and Methods ................................................................................ 91
  4.2.2 Plasmids and strains required for protein expression ............................... 91
    4.2.2.1 Nuclear magnetic resonance (NMR) experiments ......................... 91
    4.2.2.1 Pol III core:β2 production ............................................................. 91
  4.2.3 Production and purification of proteins .................................................. 92
    4.2.3.1 Purification of the θ subunit ......................................................... 92
    4.2.3.2 Production and purification of 13C,15N-θ ....................................... 93
    4.2.3.3 Production and purification of ε .................................................. 94
    4.2.3.4 Production and purification of 13C15N ε ....................................... 95
    4.2.3.5 Purification of alternately 13C,15N-double-labelled εθ complexes ..... 96
    4.2.3.6 Purification of the εL subunit ...................................................... 97
    4.2.3.7 Purification of the β2 subunit ..................................................... 99
    4.2.3.8 Purification of the αE612K subunit .......................................... 101
  4.2.4 Protein-protein affinity purification of the αL subunit ......................... 104
    4.2.4.1 Purification of biotinylated τs ..................................................... 104
    4.2.4.2 Assembly of a biotinylated τs-streptavidin column ...................... 105
    4.2.4.3 Purification of the αL subunit using protein-protein affinity chromatography ................................................................. 106
  4.2.5 Purification of αLεLθ core complexes ................................................. 108
  4.2.6 Preparation of the β2αLθ complex for SAXS experiments ................. 109
  4.2.7 Nuclear magnetic resonance (NMR) experimental conditions .......... 110
  4.2.8 Assaying polymerase activity by strand displacement DNA synthesis ................................................................. 110
  4.2.9 Conditions for SAXS experiments ....................................................... 111
    4.2.10 Preparation of the β2αLθ complex for crystallography .................. 113
    4.2.11 Trial growth of crystals ............................................................... 114
4.3 Results .......................................................................................................... 115
  4.3.1 Purification of the αL subunit .............................................................. 115
4.3.1.1 Affinity chromatography ................................................................. 115
4.3.1.2 Assessing the polymerase activity of α by strand displacement DNA synthesis .................................................. 117
4.3.2 Only the CBM residues of ε interact with the β-clamp ....................... 118
4.3.3 The θ subunit interacts with βC ......................................................... 120
4.3.4 Building a model of the E. coli βααθ:DNA complex ......................... 120
4.3.5 SAXS experiments and model validation ........................................... 123
4.3.6 Crystallography trials ....................................................................... 125
4.4 Discussion ............................................................................................. 127
4.4.1 Protein production and purification procedures ................................ 127
4.4.1.1 Protein subunit and complex purification ...................................... 127
4.4.1.3 Protein–protein affinity purification of the αL subunit .................. 127
4.4.2 The εNTD:θ complex appears highly mobile .................................. 129
4.4.3 SAXS validation of the complex model ............................................ 130
4.4.4 Crystallography trials ....................................................................... 131
4.4.5 Structural modelling advancements and future research directions .. 132

Chapter 5  A STRATEGY FOR FLUORESCENT LABELLING OF PROTEINS .................................................................................. 135

5.1 Introduction .......................................................................................... 136
5.1.1 Use of the formylglycine generating enzyme for fluorescent labelling ................................................................. 137
5.1.2 Single-molecule Förster resonance energy transfer (FRET) .............. 139
5.2 Materials and Methods ....................................................................... 140
5.2.1 Cloning overview ........................................................................... 140
5.2.2 Oligonucleotide Primers ................................................................. 140
5.2.3 Construction of plasmid for FGE8-tagged α .................................. 141
5.2.4 Preparation of a FGE8/pETMCSI cloning vector ............................ 141
5.2.5 Construction of plasmids for FGE8-tagged θ, εWT, εL and εQ ............ 143
5.2.6 Production and Purification of αFGE and θFGE ................................. 144
5.2.7 Production and purification of formylglycine generating enzyme (FGE) .......................................................... 145
5.2.8 Production and purification of Δ5εWTFGE, Δ5εLFGE, and Δ5εQFGE 147
5.2.9 Enzymatic conversion and fluorescent labelling protocol ............... 148
5.2.10 Establishing the efficiency of conversion and labelling .................. 149
5.3 Results ................................................................................................. 150
5.3.1 Assessment of polymerase activity .................................................... 150
5.3.2 Assessing the efficiency of conversion and troubleshooting ............ 151
5.4 Discussion and current status of project ............................................. 152

Chapter 6  CONCLUDING REMARKS ......................................................... 154
REFERENCES ......................................................................................... 159
APPENDIX ................................................................................................. 183
# LIST OF FIGURES

## Chapter 1

1.1 Replication initiator protein DnaA ........................................... 5  
1.2 Domain organisation of the DnaB hexamer surrounding a ssDNA oligo ............................................................. 7  
1.3 Assembly of the _E. coli_ replisome and elongation of DNA ...... 8  
1.4 Polarity of replication termination in _E. coli_ ........................... 14  
1.5 Structural basis for the formation of the Tus-Ter lock .............. 15  
1.6 The subunits of the multimeric DNA polymerase III holoenzyme ............................................................................. 17  
1.7 The α polymerase subunit of DNA polymerase III .................. 19  
1.8 The ε subunit of DNA polymerase III ..................................... 23  
1.9 The θ subunit of DNA polymerase III ..................................... 25  
1.10 The β subunit of DNA polymerase III ..................................... 27  
1.11 Crystal structure of the γ3δδ'–DNA complex with bound nucleotide analogue ADP-BeF₃ ...................................................... 30  
1.12 The SSB tetramer bound to ssDNA in the 65 mode ................. 33  

## Chapter 2

2.1 Plasmid maps of pETMCSI, pCE30, pND517 and pBOB ....... 41  

## Chapter 3

3.1 Domain organisation schematic of the _E. coli_ ε proofreading subunit ............................................................................. 56  
3.2 Variations of linked ε:α constructs presented as a simplified domain schematic .......................................................... 60  
3.3 SDS-PAGE analysis of purified linked ε–α₁₁₂₀ variants ........... 66  
3.4 Selected protein crystals of linked εCTSα₁₁₂₀ constructs .......... 69
3.5 Electron density map of the ε200α270P protein..........................72
3.6 2.15 Å crystal structure of the ε200α270P fusion protein.................73
3.7 Map of a.a. interactions between εCTS and α270 from the crystal structure of ε200α270P..........................................................74
3.8 Location of ε interaction with the α subunit..............................75
3.9 Positive ion nanoESI mass spectrum under native conditions of (A) ε209α270 and (B) α270ε209 .............................................77

Chapter 4
4.1 A structural model of E. coli β2 binding to α.............................84
4.2 SDS-PAGE analysis of purified protein products........................94
4.3 Purification of the αL subunit by protein-protein affinity chromatography..........................................................116
4.4 Strand displacement (SD) DNA synthesis shows the relative activity of α polymerase subunits and mutant variants following purification..................................................117
4.5 13C,15N-labelled ε193 in complex with purified θ interacts with β2 only through the CBM..................................................119
4.6 Models of the β2αLεLθ complex with primer-template DNA in the polymerisation mode..............................................122
4.7 SEC-SAXS elution profile of β2αLεLθ. SDS-PAGE analysis of a reproduction of the SEC-SAXS elution of β2αLεLθ......................124
4.8 Comparison of experimental SEC-SAXS pattern of β2αLεLθ in the presence and absence of DNA/dATP/ddCTP.................125
4.9 Crystal growth from samples containing the β2αLεLθ complex in selected conditions.....................................................126
4.10 Cryo-electron microscopy (cryo-EM) models of the β2αεε complex.............................................................................133
Chapter 5

5.1 Formylglycine generating enzyme (FGE) recognises the amino-acid consensus sequence LCTPSR and catalyses the oxidation of the cysteine thiol to an aldehyde to generate formylglycine .................................................................137

5.2 The Pictet-Spengler reaction...............................................................138

5.3 Partial nucleotide sequence of pNH2158........................................143

5.4 SDS-PAGE analysis of purified proteins ............................................145

5.5 Strand displacement (SD) DNA synthesis shows the relative activity $\alpha_{FGE}$ and $\alpha_{WT}$ polymerase subunits following purification ...........................................................................................................150

5.6 Positive ion nanoESI mass spectrum in 0.1% formic acid of $\theta_{FGE}$ following conversion with the FGE enzyme ..............152
LIST OF TABLES

Chapter 2

2.1 List of chemicals and reagents used .........................................................41

Chapter 3

3.1 Final purified protein yields and mass spectrometric (MS) measured molecular weights (MW) in formic acid [0.1 % (v/v)] of linked ε/α270 constructs .................................................................66

3.2 X-ray data collection and refinement statistics for ε200α270P .........70

3.3 Hydrogen-bonded and electrostatic contacts between εCTS and α270 in the crystal structures of constructs ε200α270P and ε209α270P .................................................................74
SUMMARY

The accurate and efficient replication of *Escherichia coli* DNA is catalysed by a 17-subunit assembly of proteins termed the DNA polymerase III holoenzyme (Pol III HE). The Pol III HE is further organised into three sub-assemblies: the clamp-loader complex, the β2 sliding clamp and the Pol III core complex (Pol III core). The Pol III core is comprised of the α polymerase, the ε exonuclease and the small θ subunit of an as yet undefined function. The β2 subunit encircles dsDNA and locks the Pol III core onto a template DNA by interaction with clamp-binding motifs (CBMs) located in the α and ε subunits.

A detailed understanding of how the β2αεθ complex interacts with DNA and transforms to accommodate both DNA synthesis and proofreading functions is of great interest to researchers in the field. In this project, I used a combination of experimental techniques including protein crystallography, small-angle X-ray scattering (SAXS), nuclear magnetic resonance (NMR) and a variety of additional biochemical methods to build and validate an atomic resolution model of the structure of the β2αεθ complex. I further investigated a strategy for the fluorescent labelling of Pol III core proteins for use in future single-molecule experiments.

Chapter 1 describes the background to this work and gives an overview of bacterial DNA replication, focusing specifically on the organism *E. coli*. The steps involved in replication initiation, elongation and termination are outlined and a brief overview of bacterial DNA polymerases is given. The component proteins of the Pol III HE are described in detail, with particular focus on the α, ε, θ and β2 subunits. A brief discussion of
relevant experimental techniques that are used to study multi-protein machines is also included.

Chapter 2 details the general materials and methods used in this work. The methods involved in molecular cloning, protein purification and analysis are described.

Chapter 3 presents a 2.15 Å crystal structure of a linked-protein construct displaying the site of interaction between the N-terminal PHP domain of the α subunit and the C-terminal segment of the ε subunit. The crystal structure is modelled onto the full-length α subunit and how this structure contributes to the building of an informed model of the β2αεθ complex is discussed.

Chapter 4 details the building of an atomic resolution structural model of the β2αεθ complex utilising data from protein crystallography, NMR and additional biochemical studies. This model was then validated by SAXS after co-purification of a stabilised version of the β2αεθ complex, where both the α and ε CBMs had been mutated to bind more tightly to the β2 subunit. Crystallography trials of this complex were also pursued, so far without success.

Chapter 5 outlines work I have performed to investigate a protein fluorescent labelling strategy where a specific protein tag was incorporated into Pol III core protein subunits for site specific enzymatic sulfhydryl to aldehyde modification by the formylglycine generating enzyme (FGE). Tagged versions of the α and ε proteins were purified; however, the enzymatic conversion and subsequent conjugation of the aldehyde bearing protein with a fluorescecent dye were problematic and this project remains ongoing.

Chapter 6 provides a brief concluding discussion to this Thesis and outlines possible future experimental directions for investigation of the β2αεθ complex and Pol III HE structure and function.
ACKNOWLEDGEMENTS

I would like to extend my sincere thanks and appreciation to my supervisors at the University of Wollongong, Professor Nicholas Dixon and Dr Slobodan Jergic, whose support and guidance, both academic and personal, has been greatly valued.

I am privileged to have been given the opportunity to work within a large group of talented and motivated scientists. My research has involved many collaborations and exposure to a wide-range of experimental techniques. Thank you to Associate Professor Aaron Oakley for enthusiastically sharing your deep knowledge of crystallography. Thank you to Dr Thitima Urathamakul and Dr Celine Kelso for mass spectrometry help and direction. Many thanks to Professor Samir Hamdan and Professor Antoine van Oijen for friendly encouragement with single-molecule experiments. Thank you to Dr Kiyoshi Ozawa and Dr Thomas Huber for performing NMR and structural modelling experiments. Thanks to Dr Jong-Bong Lee for assistance with fluorescent tagging of proteins. Special cheers to Dr Zhi-Qiang Xu and Dr Flynn Hill for your collegiality and good natures throughout my PhD.

I would also like recognise all members of the Dixon, Oakley and van Oijen groups, past and present, and many colleagues in the UoW School of Chemistry for your friendship and support over the years.

Most importantly, I acknowledge the love and encouragement of my family during my lengthy student ‘career’.

I have been supported during this project by an Australian Postgraduate Award.
DECLARATION

I, Nicholas Horan, declare that this thesis, submitted in fulfilment of the requirements for the award of Doctor of Philosophy, in the School of Chemistry, University of Wollongong, is wholly my own work unless otherwise referenced or acknowledged. This document has not been submitted for qualifications at any other academic institution.

Nicholas Horan

31st August 2016
PUBLICATIONS


### ABBREVIATIONS

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Å</td>
<td>Ångström</td>
<td></td>
</tr>
<tr>
<td>AAA+</td>
<td>ATPase</td>
<td>associate with a variety of cellular activities</td>
</tr>
<tr>
<td>a.a.</td>
<td>Amino acid</td>
<td></td>
</tr>
<tr>
<td>A&lt;sub&gt;λ&lt;/sub&gt;</td>
<td>Absorbance at wavelength λ</td>
<td></td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine triphosphate</td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>Base pair(s)</td>
<td></td>
</tr>
<tr>
<td>Da</td>
<td>Daltons</td>
<td></td>
</tr>
<tr>
<td>DEAE</td>
<td>Diethylaminoethyl</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>dNTP</td>
<td>Deoxyribonucleoside triphosphate</td>
<td></td>
</tr>
<tr>
<td>dsDNA</td>
<td>Double-stranded DNA</td>
<td></td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
<td></td>
</tr>
<tr>
<td><em>E. coli</em></td>
<td><em>Escherichia coli</em></td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
<td></td>
</tr>
<tr>
<td>ESI-MS</td>
<td>Electrospray ionisation mass spectrometry</td>
<td></td>
</tr>
<tr>
<td>FPLC</td>
<td>Fast protein liquid chromatography</td>
<td></td>
</tr>
<tr>
<td>FRET</td>
<td>Förster resonance energy transfer</td>
<td></td>
</tr>
<tr>
<td>IPTG</td>
<td>Isopropyl β-D-1-thiogalactopyranoside</td>
<td></td>
</tr>
</tbody>
</table>
kb  Kilobase(s)
kDa  Kilodalton(s)
$K_D$  Dissociation constant
LMW  Protein molecular weight markers
MW  Molecular weight
MWCO  Molecular weight cut off
$m/z$  Mass-to-charge ratio
NH$_4$OAc  Ammonium acetate
NMR  Nuclear magnetic resonance
nt  Nucleotide(s)
OD$_{600}$  Optical density at 600 nm
PAGE  Polyacrylamide gel electrophoresis
PCR  Polymerase chain reaction
PDB  Protein Data Bank
PEG 3350  Polyethylene glycol with an average molecular weight of 3350 Da
PEG 4000  Polyethylene glycol with an average molecular weight of 4000 Da
PHP  Polymerase and histidinol-phosphatase domain (of the Pol III α subunit)
PMSF  Phenylmethylsulphonyl fluoride
Pol III core  The αεθ complex of *E. coli* DNA polymerase III
Pol III HE  The DNA polymerase III holoenzyme
Q-TOF  Quadrupole time-of-flight
SAXS  Small angle X-ray scattering
SDS-PAGE  Sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SSB  Single-stranded DNA-binding protein
ssDNA  Single-stranded DNA
TCEP-HCl  Tris(2-carboxyethyl)phosphine hydrochloride
Tris-HCl  Tris(hydroxymethyl)aminomethane hydrochloride
α  The α subunit of Pol III HE
β  The β subunit of Pol III HE
γ  The γ subunit of Pol III HE
δ  The δ subunit of Pol III HE
ε  The ε subunit of Pol III HE
θ  The θ subunit of Pol III HE
τ  The τ subunit of Pol III HE
ψ  The ψ subunit of Pol III HE
χ  The χ subunit of Pol III HE
Chapter 1

GENERAL INTRODUCTION
1.1 DNA replication in *E. coli*

The ability of an organism to replicate its genetic material is essential to the propagation and success of all life. As a cell divides, it must organise the efficient and faithful copying of its deoxyribonucleic acid (DNA) genome to ensure each daughter progeny has the complete set of genetic instructions necessary to maintain and propagate life. Failure to accomplish this task leads to cell death.

DNA is copied by enzymes called polymerases (first isolated by Kornberg *et al.*, 1956 a and b), which generate new double-stranded DNA (dsDNA) based on the sequence of a single-stranded DNA (ssDNA) template by following hydrogen-bonding rules as described by James Watson and Francis Crick (Watson and Crick, 1953). Replication of DNA is a complex process that requires the coordinated action of over 30 separate protein subunits (Marians, 1992) participating in a highly dynamic network of protein-protein and protein-DNA interactions. The fundamental nature of this field of research has stimulated over 50 years of intense scientific interest and made DNA replication one of the best understood biological processes. Technological advances in a variety of biochemical and biophysical techniques are driving ongoing research that is continually delving deeper into the workings of the DNA replication machinery, permitting an ever-clearer picture of a complex and essential biological process.

The Gram-negative gamma proteobacterium *Escherichia coli* (*E. coli*) is widely used as a model organism for the study of cellular processes as it is easily cultured and its genetics are easily manipulated and well understood (Cronan, 2014). This review will give an overview of prokaryotic DNA replication and, unless noted, all molecular processes described herein specifically relate to those in *E. coli*. 
1.2 Distribution and organisation of the bacterial chromosome

Bacterial chromosomes are generally single circular chains 1–5 mm in length and are condensed roughly 1000-fold to fit within cellular dimensions (Trun and Marko, 1998). In contrast to the highly ordered morphological appearance of the eukaryotic chromosomes, the prokaryotic chromosome (or nucleoid) has a more diffuse and visually random arrangement within the cell. A cross-sectional view of the genome has an appearance variously compared to a bottlebrush or a transection of coral, with a moderately compact central core from which numerous filamentous DNA segments extend (Jackson et al., 2012). Although the bacterial genetic material lacks the extremely tight compaction and defined segmentation of eukaryotic chromosomes, many of the same organisational principles are shared across all forms of life; facilitating the regulation and distribution of DNA within the cell and the controlled separation of new sister chromosomes during replication of genomic DNA and cellular division (Viollier et al., 2004; Nielsen et al., 2006).

The two principal elements of hierarchal control known to organise the bacterial chromosome and regulate the arrangement of DNA within the cell are DNA supercoiling and the interaction of the chromosome with a variety of small DNA-binding proteins. Negative supercoiling (due to under-winding) of naturally helical DNA molecules induces bending and distortion of higher ordered DNA structure (Vologodskii and Cozzarelli, 1994), creating segmented topological domains – of an average size of approximately 10 kb in E. coli – and compacting chromosomal DNA (Postow et al., 2004; Stein et al., 2005). Additionally, a variety of nucleoid-associated proteins (such as IHF, FIS, HU, and SMC; Hardy and Cozzarelli, 2005) are also known to assist in the condensation process by inducing DNA bending
or by holding the compacted DNA together and providing physical links between adjacent DNA segments (Dillon and Dorman, 2010).

Chromosomal replication requires the unwinding of DNA, which causes a build-up of positive supercoiling (due to over-winding) in front of the advancing replication machinery (Peter et al., 1998). In bacteria, the enzyme gyrase relaxes coiling tension that builds up ahead of the progressing replication fork by breaking, untwisting and then resealing a double strand of DNA (Khodursky et al., 2000). The enzyme topoisomerase IV also utilises this same mechanism to separate two newly generated circular sister chromatids, which remain linked together (termed catenanes) as a result of semi-conservative DNA replication (see Section 1.3.2) and require separation before cellular fission can occur (Yan et al., 1999). Positive supercoiling generated during replication can also be relieved by the opposing action of the DNA gyrase enzyme, which introduces negative supercoils and helps maintain the chromosome in an overall slightly underwound state (Drlica, 1992; Nollmann et al., 2007).

These (and other) mechanisms provide a high level of control of the compaction, distribution and separation of genomic DNA in the bacterial cell, assisting in ensuring highly regulated replication of new DNA and orderly cellular division (see Jackson et al., 2012 for recent review).

1.3 Replication of DNA

1.3.1 Initiation

Assembly of the *E. coli* replication machinery on chromosomal DNA first requires the association of molecules of the initiator protein DnaA (Figure 1.1A) with a unique 254 bp sequence termed the oriC (Hwang and Kornberg, 1990). Multiple distributed 9 bp sequence motifs within oriC (termed dnaA boxes) each bind a DnaA monomer in a contiguous
arrangement that promotes formation of a multimeric nucleoprotein complex (Echols, 1990; Fujikawa et al., 2003; Duderstadt et al., 2011; Figure 1.1). In the presence of ATP and additional proteins (such as integration host factor, IHF) the protein-DNA complex melts dsDNA at an adjacent A-T rich region (DNA-unwinding element, DUE) and engages ssDNA in a stable complex (Figure 1.1C), creating the beginnings of a ssDNA ‘bubble’ at the replication origin and permitting access to additional replication factors such as the DnaB helicase (reviewed in Messer et al., 2001; Messer, 2002; Duderstadt and Berger 2008; Duderstadt and Berger 2013).

**Figure 1.1: Replication initiator protein DnaA.** (A) Cartoon representation of the DnaA monomer structure from *Aquifex aeolicus* displaying domains I–III in pale green and DNA-binding domain IV in a darker green (Erzberger et al., 2002; PDB ID: 1L8Q). (B) Cartoon representation of the domain IV structure of DnaA from *E. coli* in complex with DnaA-box sequence dsDNA (Fujikawa et al., 2003; PDB ID: 1J1V). Helix α5 binds to the major groove of dsDNA and is indicated in (A) and (B). (C) Space-filling representation of a DnaA tetramer form structure from *Aquifex aeolicus* with ssDNA (Duderstadt et al., 2011; PDB ID: 3R8F). Each DnaA monomer is differently coloured and ssDNA is coloured orange.
The DnaB helicase is a hexameric ring-shaped ATPase that is recruited to the origin as a complex with the helicase loader DnaC, through interaction between DnaC and ssDNA-bound DnaA (Carr and Kaguni, 2001; Fang et al., 1999; Learn et al., 1997). Loading of a DnaB molecule onto each of the ssDNA strands is an ATP dependent process that leads to subsequent dissociation of hydrolysed DnaC-ADP and permits the loaded helicase molecule to then bind the DnaG primase enzyme (Wahle et al., 1989 a and b). Crystallographic modelling shows the DnaB hexamer, once loaded, adopts an ascending/descending right-handed spiral conformation surrounding ssDNA, such that the first and last monomers in the circular sequence are slightly offset (Bailey et al., 2007). This arrangement facilitates 5’ to 3’ end ATP-driven translocation along the ssDNA template in the 5’ to 3’ direction as the helicase rotates, in much the same way a helical screw thread enables lateral movement of a nut on a bolt as it spins. Structural data with bound molecules of the semi-hydrolysed NTP analog GDP-AlF₄ suggest translocation occurs in 2–nucleotide steps per ATP molecule hydrolysed as the hexamer is sequentially rotated through each monomer (Itsathitphaisarn et al., 2012).

The two unidirectional helicases move outwards from oriC, further enlarging the replication bubble as both complexes travel away from each other by encircling one strand and occluding the other. As strand separation occurs, the helicase-coupled DnaG primase enzyme synthesises short RNA oligonucleotide primers (10–12 bases long) on each newly available ssDNA template, giving the genomic replication complex DNA polymerase III holoenzyme (DNA Pol III HE) a point to begin the elongation phase of replication (Zechner et al., 1992 a and b).
1.3.2 Elongation

One copy each of the large multimeric Pol III HE remains bound to both respective helicase enzymes as they travel outwards and away from the origin (Prescott and Kuempel, 1972), forming two replication ‘forks’ at which additional replication proteins are assembled. Precursor deoxyribonucleoside triphosphates (dNTPs) are added sequentially to the free 3’ end of the growing primer terminus to synthesise a new complementary DNA strand. DNA replication is therefore bidirectional and semiconservative, as each dsDNA product contains one parental and one daughter strand (Meselson and Stahl, 1958).

At each replication fork, a single Pol III HE complex copies both primed template strands simultaneously. Two copies of the polymerase
subunit (α) are held together by a multipart subunit architecture (discussed in detail in Section 1.3), which coordinates the loading and dissociation of other replication associated protein subunits and enables the highly-organised parallel replication of both template strands (reviewed in McHenry, 2011; Duderstadt et al., 2014; Lewis et al., 2016).

Figure 1.3: Assembly of the E. coli replisome and elongation of DNA. (A) The DnaA protein binds then multimerises at the origin of replication (oriC) to melt dsDNA. Two copies of the DnaB/DnaG primosome are then recruited to the origin and begin translocating away from one another, separating the two strands of DNA and forming the beginnings of a replication ‘bubble’. (B) One copy each the DNA polymerase III holoenzyme (Pol III HE) travels in the same direction as each helicase as it replicates both template strands simultaneously, extending the replication bubble. (C) Okazaki synthesis of DNA on the lagging strand requires the repeated formation of ssDNA loops and the continual reassociation of β-clamp and Pol III core subunits (αβθ) onto new RNA primers made by the DnaG primase. Segments of ssDNA are protected during loop formation by single-stranded DNA-binding protein (SSB; Figure reformatted from Robinson and van Oijen, 2013).
1.3.2.1 Okazaki fragment synthesis

All known polymerases are able only to extend a polynucleotide chain in the 5’ to 3’ direction (Wu et al., 1992; Zechner et al., 1992 a and b), providing a topological challenge for the holoenzyme due to the anti-parallel backbone-chemistry of the two DNA strands. One strand, termed the leading strand, is replicated continuously in the same direction as fork progression while the other, termed the lagging strand, must be looped to provide the correct orientation for coordinated DNA synthesis of both strands as the polymerase progresses in one direction only (Figure 1.3C). Lagging strand replication is achieved through synthesis of discontinuous sections of complementary DNA known as Okazaki fragments (Okazaki and Okazaki, 1969; Kornberg and Baker, 1991). In E. coli, these are approximately 1–2 kb in length, a little shorter than T4 and T7 bacteriophage Okazaki fragments of approximately 2.3 kb in length but significantly longer than those found in eukaryotic and archaean replication systems, which are 100–200 bp in length (Duderstadt et al., 2014). The E. coli replisome, progressing at a rate of 600–1000 bp per second (Kornberg and Baker, 1991; Breier et al., 2005), must therefore be able to synthesise each Okazaki fragment on a timescale of 1–2 seconds.

It has long been conceptually proposed that Okazaki fragment synthesis is enabled by the formation of lagging-strand loops, where the single-stranded template is folded back into the appropriate orientation for polymerase synthesis, then drawn out and replicated in repeated cycles of loop growth and release for each fragment generated (Sinha et al., 1980; Alberts et al., 1983). This model (Figure 1.3 B and C) has since been supported by electron microscopic visualisation of replication intermediates in the T4 and T7 bacteriophage systems and in vitro single-molecule observations of dynamic cycling of lagging-strand extension and release in
the T7 bacteriophage replication system (Park et al., 1998; Chastain et al., 2000 and Chastain et al., 2003; Hamdan et al., 2009). This elegant solution, however, creates complications, as each cycle of loop formation and release requires the dissociation and reloading of several Pol III HE subunits, as well as the synthesis of a new RNA primer by the primase enzyme. The Pol III HE remains linked during coupled synthesis (Kornberg and Baker, 1991), obliging lagging-strand progression to maintain pace with leading-strand synthesis and helicase translocation, despite the additional molecular acrobatics required to synthesise lagging strand Okazaki fragments. Precisely how the replication machinery accomplishes this is matter of current debate, potentially involving a number of different, non-exclusive, molecular mechanisms that may vary to some extent between different organisms (reviewed in Lewis et al., 2016)

1.3.2.2 Cycling and signalling on the lagging strand

Due to their relative simplicity, the T4 and T7 bacteriophage replication systems are comparatively well characterized. In contrast to the ten protein types that constitute the E. coli replisome, coordinated DNA synthesis can be performed using only seven different T4 replication proteins; with the T7 system requiring just four (Richardson, 1983; Cha and Alberts, 1989). The T7 replication apparatus is therefore frequently utilised as a model system for interrogating the complex kinetics of lagging strand Okazaki fragment cycling.

The streamlined T7 replisome has both its helicase and primase activity consolidated within the single polypeptide chain of the hexameric gp4 protein (Dunn et al., 1983; Mendelman et al., 1992). The synthesis of a primer by the T7 gp4 primase domain has been shown to be particularly slow in isolation, requiring about one second to complete (Frick et al., 1999),
with the process of enzymatic handover from the primase to the polymerase at the primer terminus of each lagging stand Okazaki fragment also requisite before extension of the primer by the polymerase can occur (Kato et al., 2004). These two events occur on a timescale far too slow to support continuous coordinated synthesis without progression of the helicase/primase being either temporarily halted until the completion of each, counter-directionally synthesised, tetraribonucleotide primer (Tabor and Richardson, 1981) or requiring the formation of a ‘priming loop’ to resolve the directionality problem as the helicase continues to advance; in much the same manner as proposed for the formation of replication loops during Okazaki fragment synthesis. Intriguingly, experimental evidence suggests both of these two, non-exclusive, models for the management of lagging strand Okazaki fragments are in operation in the T7 replication system (Lee et al., 2006; Pandey et al., 2009).

The resolution of Okazaki fragments also requires a mechanism to prompt ordered release of replication and/or priming loops and regulate the dissociation of polymerase subunits from template DNA during lagging strand cycling. Competing explanations for how the replisome achieves this include a molecular signal initiated by the primase enzyme at some point following binding to template DNA, or by collision of the lagging strand polymerase with the 5’-end of the preceding Okazaki fragment (Tougu and Marians, 1996; Salinas and Benkovic, 2000; Lee et al., 2002). Experimental evidence, once more, suggests that multiple signalling mechanisms or kinetic pathways may be utilised by prokaryotic replisomes, potentially providing a level of redundancy during Okazaki fragment processing as a failsafe system (Li and Marians, 2000; Tanner et al., 2008; Hamdan et al., 2009).

The more complex E. coli replisome comprises separate primase and helicase protein subunits (termed the primosome when associated), which permits the dissociation of the primase enzyme once bound to the ssDNA
template and may allow synthesis of the primer polyribonucleotide whilst uncoupled from the Pol III HE/helicase complex and direction of fork progression. The precise role of primase binding, synthesis and release in regulating the Okazaki fragment timing in the T4, T7 and E. coli (or any other) replication systems is unknown and current research remains inconclusive as to the sequence of molecular events regulating these processes. To enable efficient Okazaki fragment turnover, it is likely that different organisms employ multiple mechanistic solutions, which are dependent on individual replisome architectures and/or local environment conditions (T7 lagging stand cycling is reviewed in Hamdan and van Oijen, 2010; Lee and Richardson, 2011).

1.3.2.3 Efficiency and adaptability of the E. coli replisome

The 17-subunit E. coli replisome contains dedicated component proteins that facilitate loading and unloading of replication associated enzymes from a DNA template (the DnaX complex: τ, γ, δ, δ', χ and ψ subunits), edit/remove incorrect DNA sequence to improve fidelity (the ϵ subunit) and increase the stability and processivity of the polymerase (α) subunit on a template DNA (the β2 clamp). The separation of these functions from the main polymerase provides the replication machinery with versatility and adaptability, enabling it to efficiently operate at a rapid rate (up to 1000 nt/sec for over 150 kb without dissociation; Chandler et al., 1975; Mok and Marians, 1987), with high fidelity (an error range of under \(10^{-9}\) per base pair incorporated; Drake et al., 1969) and the capability to overcome a variety of topological and chemical challenges in a crowded cellular environment (stalls, blocks, DNA nicks and breaks) by switching or adjusting, where required, to accommodate variable tasks. These (and other) replisome components will be discussed in detail in Section 1.3.
1.3.3 Termination

Synthesis of new DNA continues as both replication forks travel outward around the circular chromosome at approximately the same speed. Replication is terminated when each fork arrives at a DNA region comprised of Ter sites, located approximately 180 degrees from the oriC (Hill et al., 1988). The E. coli chromosome contains ten 23-base pair Ter sites, arranged in two oppositely orientated groups of five, which each bind a molecule of the replication termination protein Tus (Hidaka et al., 1989; Coskun-Ari and Hill, 1997). Binding of the Tus protein creates a unidirectional impediment for the replication machinery, which blocks progression of forks approaching from one direction (non-permissive face), yet allows for displacement of the Tus protein when encountered from the other (permissive face; Figure 1.4).

Atomic structures of the Tus protein bound to a wild-type Ter sequence (Figure 1.5A) and to a forked oligonucleotide with a (C) oligonucleotide at the (6) position of the non-permissive face (Figure 1.6B) illustrate the proposed molecular basis for unidirectional fork arrest. The cytosine residue at this position is able to ‘flip’ into a binding pocket present in the Tus protein as dsDNA is melted by the replication machinery approaching the complex from one direction. It is this interaction that generates the extremely stable interaction ($K_D$ in the pM range) between the Tus protein and Ter sequence and enables the ‘trapping’ of replication forks approaching from one direction within a certain region of the chromosome but not the other, preventing over replication while awaiting the alternate fork to reach the termination region (Mulcair et al., 2006).
Recent single-molecule experiments have deepened the kinetic understanding of this process (Elshenawy et al., 2015; Pandey et al., 2015) and demonstrated that successful dissolution of lock formation and removal of the Tus protein from a Ter site is a process also dependent on the speed of an approaching replisome from the ‘non-permissive’ direction and competition between rates of Tus displacement by strand-separation and the efficiency of ‘C6 base-flipping’ that generates the tight Tus-Ter lock. It is surprising that blockage of replication forks approaching from the non-permissive direction occurs only 50% of the time, both in vivo and in vitro (Kaplan and Bastia, 2009; Elshenawy et al., 2015), indicating that the presence of ten Ter sites (five of each polarity) in the E. coli chromosome provides a high-level of redundancy and strongly promotes termination of replication within the proper region of the genome. Replication is terminated when the second fork encounters the first blocked fork (Hill and Marians, 1990).
Figure 1.5: Structural basis for the formation of the Tus-Ter lock. The Tus protein is shown in purple using a ribbon representation. Oligonucleotide DNA is shown in orange and the ‘C(6)’ base is highlighted using yellow spheres. (A) The ‘unlocked’ form of the Tus-Ter complex (Kamada et al. 1996; PDB 1ECR). (B) The ‘locked’ form of the Tus-Ter complex (Mulcair et al. 2006; PDB: 2EWJ) shows the C(6) base ‘flipped’ and now in contact with a binding pocket in the Tus protein. Green and red arrows indicate permissive and non-permissive orientations, respectively.

1.4 *E. coli* DNA polymerases

Named in the order of their discovery, *E. coli* produces five known DNA polymerases. DNA polymerase I (Kornberg et al., 1956) is the 103 kDa product of the *polA* gene and its main role in replication is to remove RNA primers laid down by DnaG primase at the beginning of each Okazaki fragment, replacing them with a new strand of DNA in the process known as nick translation. DNA polymerase II, the 90 kDa product of the *polB* gene (Kornberg and Gefter, 1970; Moses and Richardson, 1970) and DNA polymerase V, the product of the *umuD* and *umuC* genes (Tang et al. 1999; Reuven et al., 1999), are induced during the SOS response to DNA damage and are involved in DNA repair. DNA polymerase IV (also SOS-induced) is the product of the *dinB* gene and is a relatively low fidelity polymerase, able to skip an area of DNA damage and temporarily substitute for other high fidelity polymerases (Kim et al., 1997; Napolitano et al., 2000). The multi-subunit DNA polymerase III holoenzyme is the enzyme principally
responsible for chromosomal DNA replication (Gefter et al., 1971; Kornberg and Gefter, 1972)

1.5 *E. coli* DNA polymerase III

DNA polymerase III holoenzyme (Pol III HE) is a large, dynamic 17 subunit complex consisting of 10 different types of protein subunit (Figure 1.6). It is commonly divided into three subassemblies: (i) the Pol III core, (ii) the $\beta_2$ sliding clamp, and (iii) the clamp loader complex.

The Pol III core consists of the three subunits $\alpha$, $\varepsilon$, and $\theta$ (McHenry and Crow, 1979), with each Pol III HE containing two Pol III core complexes, one each for synthesis of the leading and lagging strands (Studwell-Vaughan and O'Donnell, 1991). The $\alpha$ subunit contains the catalytic site which polymerises dNTPs and extends the nucleotide chain of a primed template DNA in the 5'→3' direction (Maki and Kornberg, 1985). It is bound to the $\varepsilon$ subunit, which is the 3'→5' proofreading exonuclease responsible for editing nucleotide insertion errors made by $\alpha$ (Scheuermann and Echols, 1984). The small $\theta$ subunit binds to $\varepsilon$, but not to $\alpha$, and is thought to contribute to the stability of $\varepsilon$ (Studwell-Vaughan and O'Donnell, 1993; Taft-Benz and Schaaper, 2004).

When it is actively synthesising DNA, each Pol III core is bound to a $\beta_2$ sliding clamp, which locks the core polymerase onto the growing DNA strand, improving stability and dramatically increasing polymerase processivity (Johnson and O'Donnell, 2005). The loading of sliding clamp and core polymerase subunits onto DNA is performed by the clamp loader or DnaX complex $[\tau_2(\gamma/\tau)\delta\delta'\epsilon\psi]$, which maintains transient contact with the advancing helicase enzyme and both Pol III cores to hold the holoenzyme together (O'Donnell, 2006) and coordinate the simultaneous synthesis of leading and lagging DNA strands by cycling clamp and polymerase subunits.
on the lagging strand for each new Okazaki fragment that is produced (López de Saro et al., 2003).

**Figure 1.6: The subunits of the multimeric DNA polymerase III holoenzyme.** The 17-subunit enzyme consists of ten different subunits with the overall formula: \((αεθ)2β2τ4δδ'ψ\). During the replication cycle, the holoenzyme maintains transient contact with a selection of replication-associated proteins, including DnaB and SSB (shown here). For clarity, lagging-strand ssDNA has been presented in the same orientation as leading-strand ssDNA and the Okazaki fragment ssDNA loop has been omitted (Figure illustration based on Schaeffer et al., 2005).

### 1.5.1 Structure and organisation of components

#### 1.5.1.1 The α subunit

The α subunit is the 1160-residue (Wieczorek and McHenry, 2006), 130 kDa product of the dnaE gene (Welch and McHenry, 1982) and contains the catalytic site responsible for the 5’→3’ DNA polymerase activity of the Pol III core (Maki and Kornberg, 1985). The crystal structure of a large fragment (residues 1–917) of the α subunit (Lamers et al., 2006) reveals a distinctive overall fold that resembles the appearance of a cupped right hand (Figure 1.7A). Accordingly, the three catalytic domains within the subunit are
commonly denoted as the Palm (residues 271–432 and 511-560), Thumb (433–510) and Fingers (561–911) domains. These three domains form a deep cleft where the active site for DNA synthesis resides; primarily comprised of residues from the palm domain (Figure 1.7C). The fold of the protein creates a central passage through which a template DNA substrate can neatly pass, encompassed by the Palm, Thumb and Fingers domains (Figure 1.7B).

The Fingers domain comprises four subdomains, duly termed the Index (residues 641–756), Middle (561–640 and 757–778), Ring (779–838) and Little (839–911) fingers. The Fingers bind incoming ssDNA and direct the template ssDNA to the active site, while the thumb domain guides the newly formed dsDNA as it leaves (Lamers et al., 2006). A fourth PHP (polymerase and histidinol-phosphatase) domain is located at the N-terminal end of the α subunit (residues 2–270). Deletion studies have shown that this region binds to the ε subunit but has no direct role in the synthesis of DNA (Weiczorek and McHenry, 2006).

All bacterial replicative polymerases belong to the C family of DNA polymerases and can be classified as one of two major forms: in Gram-negative bacteria, such as E. coli, the replicative polymerase is the polymerase III α-subunit (DnaE family), while in Gram-positive bacteria the replicative polymerase is termed Pol C (Ito and Braithwaite, 1991). The C family polymerases are the most efficient of all the polymerase families, and together with the contribution of the β2-clamp and dnaB helicase, they are able to synthesise new DNA exceptionally rapidly (>600 nt/s at 30°C; Breier et al., 2005), with high fidelity and they display remarkable processivity (>150 kb; Mok and Marians 1987 a and b). Although completely distinct from eukaryotic replicative polymerases (B family polymerases) in sequence and structural characteristics, the C family replicative polymerases, Pol C and Pol III α, share many similarities and almost certainly share a common
Figure 1.7: The α polymerase subunit of DNA polymerase III. (A) Crystal structure of residues 2–917 displayed using a space-filling representation (Lamers et al., 2006; PDB ID: 2HQA). The PHP domain is displayed in green, the Thumb domain in pink and the Palm domain in purple. The four subdomains that comprise the Fingers domain (Index, Middle, Ring and Little) are displayed in differing shades of yellow/orange. (B) Space-filling representation of (A), now wholly coloured pale orange and presented with the α C-terminal segment (CTS) from the aligned crystal structure of the full-length *Thermus aquaticus* α subunit (coloured red; Bailey et al., 2006; PDB ID: 2HPI) and the aligned crystal structure of the *Thermus aquaticus* α subunit with a short DNA oligo (coloured orange and blue; Wing et al., 2008; PDB ID: 3E0D). Structural alignments were performed using PyMOL software. (C) Domain schematic of the full-length *E. coli* α subunit. The PHP domain is known to bind the ε subunit of Pol III HE and precedes the catalytic Palm (P), Thumb (Th) and Fingers domains. The locations of internal and external β-binding motifs found within the CTS are highlighted in blue. All domains are coloured in account with their designation in (A) and (B).

evolutionary origin (Leipe et al., 1999; Bailey et al., 2006; Wing et al., 2008). Both families contain correspondingly ordered PHP, Palm, Thumb and Fingers domains, followed by a C-terminally located β-binding motif (or motifs, depending on classification). Notably, however, the OB-fold (oligonucleotide/oligosaccharide binding) domain, which probably binds
incoming ssDNA, is found within the C-terminal segment (CTS) of Pol III α, while the same domain is found N-terminal to the PHP domain in Pol C – a probable remnant of an ancient gene duplication event (Timinskas et al., 2013). Furthermore, it is of interest to note the absence of a separate proofreading exonuclease subunit in Pol C and the majority of Pol III α polymerases. It is instead integrated into a catalytically active PHP domain (Aravind and Koonin, 1998; Rock et al., 2015). The separation of the proofreading and polymerisation functions into different subunits in E. coli and some other Gram negative species appears to represent a phylogenetic outlier and an integrated proofreading function within the polymerase PHP domain is more widely conserved in eubacteria (Rock et al., 2015).

The universal presence of the PHP domain within all C family polymerases suggests an essential function supplementary to exonuclease activity. Indeed, a high degree of structural similarity remains between Pol C and Pol III α PHP domains, regardless of the absence or presence of exonuclease activity or the full or partial loss of metal binding ability. The E. coli PHP domain displays no measureable exonuclease activity and has lost metal-binding residues at its quiescent active site (Lamers et al., 2006). Remarkably, however, the metal-binding ability of the domain can be restored experimentally by just three point mutations (Barros et al., 2013), demonstrating a high level of structural similarity between C-family PHP domains, which exhibit metalloenzymatic activity, and the E. coli Pol III α PHP domain, which has ceded its enzymatic activity to the separate Pol III ε subunit. Furthermore, the PHP domain has also been shown to contribute significantly to the global stability and polymerisation activity of Pol III α (Barros et al., 2013), indicating an important role within the context of the replisome that is yet to be fully understood.

The polymerase active site is highly conserved within C family polymerases, with active site residues that coordinate two magnesium ions,
D401, D403 and D555 in E. coli α, absolutely conserved (Pritchard and McHenry, 1999; Timinskas et al., 2013). Intriguingly, the coordination of active site residues, magnesium ions and incoming nucleotide is near identical to that of the X family polymerases, such as the eukaryotic repair polymerase Pol β, which display a significantly reduced catalytic rate of synthesis (Pelletier et al., 1994; Lamers et al., 2006). Consequently, the comparatively rapid rate of DNA synthesis by bacterial C family replicative polymerases cannot be explained by the composition of the active site alone. It has been suggested that the extended Fingers domain of Pol C family polymerases (a distinctive feature much larger than in any other known polymerase) may have a significant role in enhancing catalytic rate (Lamers and O’Donnell, 2008).

The E. coli α polymerase subunit makes several interactions with a variety of different replisomal protein subunits, facilitating its dynamic loading onto and dissociation from DNA during cycles of replication and repair, as well as maintaining polymerase stability during processive DNA synthesis. Loading of Pol III α onto a template DNA is facilitated by the τ subunit of the Pol III clamp loader complex, which interacts with the αCTS (Studwell-Vaughan and O’Donnell, 1991). The internal β-binding motif of the α subunit, located immediately following the α Fingers domain (Figure 1.7C), makes contact with the β2 clamp during processive DNA replication through one of two hydrophobic binding pockets present in the clamp dimer (discussed further in Section 1.3.1.4). A second ‘external’ β-binding motif, located at the extreme C-terminal end of the α subunit (Figure 1.7C), has previously been proposed to bind the remaining β subunit hydrophobic pocket during processive replication (Leu et al., 2003), however, recent in vitro experiments (Jericic et al., 2013) have shown this position to be occupied by a similar β-binding sequence motif present within the ε subunit during
processive replication. The physiological relevance, if any, of the external β-binding motif of the α polymerase remains unknown.

The ε proofreading exonuclease subunit binds tightly ($K_D = 5$ nM; Weiczorek and McHenry, 2006) to the α polymerase within the N-terminal PHP domain, with the pair remaining in contact throughout processive DNA synthesis and proofreading events. Binding of ε to α has been shown to increase polymerase activity two-fold and proofreading activity 10- to 80-fold (Maki and Kornberg, 1987).

1.5.1.2 The ε subunit

The ε subunit of *E. coli* DNA Pol III is the 242 residue (27.5 kDa) product of the dnaQ gene (Scheuermann *et al.*, 1983) and is solely responsible for the 3′→5′ exonuclease activity of the *E. coli* core (Scheuermann and Echols, 1984; Maki and Kornberg, 1985). The structured N-terminal domain (residues 2–180) is connected via a flexible Q-linker region identified by NMR (Ozawa *et al.*, 2008) to an intrinsically unstructured C-terminal domain (residues 201–243, at least), which contains the α binding site (Taft-Benz and Schaaper, 1999; Figure 1.8). The flexible interdomain region includes the hexapeptide sequence [QTSMAF] located from residues 182–187 that is consistent with an established clamp binding motif found in other proteins that bind the β2 subunit (Dalrymple *et al.*, 2001; discussed further in Section 1.3.1.4). It has been demonstrated that the ε subunit interacts *in vitro* with the β2 subunit during processive DNA synthesis and contributes to the overall stability of the Pol III αεθβ2 complex on template DNA (Jergic *et al.*, 2013; Ozawa *et al.*, 2013; Toste Rêgo *et al.*, 2013).

The crystal structure of the N-terminal domain (ε180) of the ε subunit has been solved, showing a five-stranded β sheet core that is a common feature of members of the DnaQ superfamily, to which the Pol III ε subunit
belongs (Hamdan et al., 2002a). The reaction mechanism at the catalytic site relies upon a binuclear divalent metal (Mn$^{2+}$ or Mg$^{2+}$) site, similar to the active site of DNA polymerases (Cisneros et al., 2009). The proofreading capability of the ε subunit contributes to an overall nucleotide insertion error rate of $10^{-10}$ per base pair replicated for the αεθ polymerase core (Drake, 1991). Deletion of the dnaQ gene has been shown to result in poor cell growth and highly increased levels of spontaneous mutation (Slater et al., 1994).

Figure 1.8: The ε subunit of DNA polymerase III. The ε subunit is comprised of two domains. The N-terminal domain (Hamdan et al., 2002a; PDB ID: 1J54) contains the 3'-5' exonuclease active site and the θ subunit binding site. The intrinsically unstructured C-terminal domain contains the site of interaction with the α subunit and is connected to the N-terminal domain via a flexible Q-linker. A β-binding motif (QTSMAF; highlighted in pink) is located within the linker region from amino acid residues 182–187.

The distinctive linked-domain arrangement of the ε subunit proposes the possibility of an interesting functional significance. The tight interaction ($K_D = 5$ nM; Wieczorek and McHenry, 2006) between the ε C-terminal segment (εCTS) and the α PHP domain (Taft-Benz and Schaaper, 1999) provides a stable anchor point for the ε subunit, while the comparatively weak interaction ($K_D = 210$ μM; Jergic et al., 2013) between the ε interdomain β-binding motif and the hydrophobic binding pocket of the βε clamp
ostensibly governs a more transient interaction and may allow the catalytic domain of the ε subunit to completely dissociate from the β2 clamp, yet still remain localized at the template DNA as the holoenzyme shifts between different modes of operation. Structural models of the αεθβ2 complex on a template DNA oligo show a distance >70 Å between the active sites of the α polymerase subunit and the ε N-terminal domain, indicating a significant conformational change would be expected to occur when transferring a mismatched template DNA between the two sites (Ozawa et al., 2013). The flexibility provided to the ε catalytic domain by a long and flexible linking region can be speculated to play a role in primer-template handover; however, a detailed understanding of the dynamics of the ε N-terminal catalytic domain during cycles of replication, editing and repair remains unclear.

The dynamics of the ε subunit are likely highly complex and may involve additional undiscovered transient interactions with other replication associated proteins, including intermediate modes of interaction that are of functional interest. The potential short-lived and weak nature of many replisomal protein interactions (and motor proteins in general) makes the identification and full characterization of their physiological roles a significant experimental challenge. Novel methodologies and multifaceted experimental approaches are required to better understand the functional role of individual subunits within large motor proteins. Several experimental approaches are applied in this Thesis in the investigation of the structural and functional characteristics of the ε subunit.

1.5.1.3 The 0 subunit

The small 0 subunit (8.8 kDa) is the product of the holE gene and is not highly conserved in bacterial Pol III cores (Carter et al., 1993; Robinson et al.
The θ subunit binds tightly to the ε subunit but not to α, as demonstrated originally by gel filtration studies (Studwell-Vaughan and O'Donnell, 1993). A structure of the θ subunit in complex with ε has been solved (Figure 1.7A) by NMR (nuclear magnetic resonance) and shows that the N-terminal two thirds of the protein is comprised of three helical segments. Helix 1 (Q10 to Y31) is mostly hydrophobic, while helices 2 (A37 to R42) and 3 (R49 to R68) are amphipathic. The extreme N-terminal (M1 to L8) and extreme C-terminal ends (P70 to K76) of the protein appeared to be unstructured. The site of θ interaction with ε is in the N-terminal region of helix 1 and the C-terminus of helix 3 (Keniry et al., 2006).

![Figure 1.9: The θ subunit of DNA polymerase III.](image)

Initial experiments involving a strain carrying a null mutation of the *holE* gene showed no effects on cell viability, indicating that the θ subunit is largely dispensable (Slater *et al.*, 1994). However, sequence analysis of the *holE* gene showed homologs present in the genomes of other enterobacteria, suggesting a conserved function. Indeed, the bacteriophage P1 homolog *hot*
gene has been shown to be capable of substituting for \textit{holE} (Chikova and Schaaper, 2005) and displays a very similar NMR structure (Keniry \textit{et al.}, 2006; Mueller \textit{et al.}, 2005). A crystal structure of the HOT protein in complex with the \(\varepsilon\) N-terminal catalytic domain has also been solved, providing an insight into how the \(\theta\) subunit may interact with and stabilise the \(\varepsilon\) proofreading subunit (Kirby \textit{et al.}, 2006).

Further investigations into the role of the \(\theta\) subunit have shown that it plays a role in stabilising the \(\varepsilon\) subunit and improving fidelity (Studwell-Vaughan and O’Donnell, 1993; Taft-Benz and Schaaper, 2004) as well as enhancing \(\varepsilon\) solubility during cell free protein synthesis (Ozawa \textit{et al.}, 2008). These results go some way to elucidating the full role of \(\theta\), the function of which is still largely unknown.

\textbf{1.5.1.4 The \(\beta_2\) subunit}

The \(\beta_2\) sliding clamp is a ring-shaped dimer of two identical 40.6 kDa \(\beta\) subunits coded by the \textit{dnaN} gene (Burgers \textit{et al.}, 1981). The clamp’s two 366 amino acid residue subunits combine to form a ring-shaped dimer, which encircles and diffuses freely (random translocation by thermal fluctuation) along double-stranded DNA (Figure 1.10; Kong \textit{et al.}, 1992; Cho \textit{et al.}, 2014). A variety of replication associated proteins, including all five \textit{E. coli} polymerases, are known to interchangeably bind to the \(\beta_2\) clamp at slightly different contact points within a common hydrophobic binding pocket present on each \(\beta\) subunit monomer (Johnson and O’Donnell, 2005; Lopez de Saro \textit{et al.}, 2006; Georgescu \textit{et al.}, 2008; Wolff \textit{et al.}, 2011). Proteins that interact with the \(\beta_2\) clamp do so through a universal pentapeptide or hexapeptide linear clamp binding motif (CBM) with the formula QL\(x_1\)L\(x_2\)F/L (where \(x\) is any amino acid; S/D preferred at \(x_1\); \(x_2\) may be absent), that are found within disordered segments or loops and are highly conserved
throughout eubacteria (Dalrymple et al., 2001; Yin et al., 2013). Binding of the Pol III HE α polymerase to the β₂ clamp locks it onto the growing nucleotide chain, dramatically increasing its processivity and speed (Wagner et al., 2000; Johnson and O’Donnell, 2005).

Figure 1.10: The β subunit of DNA polymerase III. (A) Crystal structure of the β₂ dimer displayed using a cartoon representation. Each monomer is differently coloured in red or purple. Lines of symmetry (S; coloured green) and six-fold pseudo symmetry (PS; coloured black) are also shown (Oakley et al., 2003; PDB ID: 1MMI). (B) Space-filling representation of (A), where the hydrophobic binding pocket present between domains 2 and 3 of each monomer is highlighted in yellow. (C) Cartoon representation of the crystal structure of the β₂ dimer (now coloured grey) together with a short dsDNA oligo (coloured orange; Georgescu et al., 2008; PDB ID: 3BEP). The dsDNA is observed here to bind at an angle of 22°, relative to the horizontal plane of the clamp. Side-chains of residues commonly involved in binding within the hydrophobic pocket are displayed as yellow spheres for clarity. The C- and N-terminal faces of the molecule are indicated.
The crystal structure of the β2 sliding clamp of *E. coli* shows it contains six similar subdomains (three within each monomer), with 12 α helices lining the internal face of the ring and β sheets predominating on the outer surface (Figure 1.10A, B and C; Kong *et al.*, 1992; Oakley *et al.*, 2003). The internal face carries a net positive charge that interacts with both single- (*K_D* = 450 nM) and double-stranded (*K_D* = 120 nM) DNA. Once loaded around the nascent DNA template, the C-terminal face of the dimer makes contact with the α polymerase through one of the symmetrically related hydrophobic binding pockets located on each monomer (Figure 1.10C; La Duca *et al.*, 1986; Dohrmann and McHenry, 2005). Newly synthesised dsDNA is extruded from the N-terminal face of the clamp as polymerisation proceeds.

It has been recently demonstrated that the alternate binding pocket of the clamp is occupied by the ε exonuclease subunit during processive DNA synthesis (Jergic *et al.*, 2013). The ε-β interaction is quite weak, however (approximately 200 fold weaker than the α-β interaction), and it is probable that this site becomes transiently available for other proteins that are known to bind the clamp, providing access to the DNA template as the replisome switches from modes of polymerization, editing and repair. O’Donnell and colleagues have previously proposed a ‘toolbelt’ model for the clamp, offering an analogy for the ability of the two binding pockets of the β2 dimer to interchangeably bind and localize a number of replication associated proteins (such as the lesion bypass polymerase Pol IV) to the primed template junction (Indiani *et al.*, 2005). Interestingly, it has also been shown that Pol III and Pol IV can swap positions at the same clamp interaction pocket in experiments using a β2 heterodimer that has only one functioning protein binding site (Heltzel *et al.*, 2009), suggesting that the clamp is capable of switching protein binding partners at either or both sites, depending on what task is to be performed. Surprisingly, DNA was shown to bind at quite a steep angle (22°) relative to the rotational axis of the clamp (Georgescu *et
al., 2008; Figure 1.10C), which may assist in switching between polymerases III and IV during lesion bypass. It remains unclear precisely what role the clamp dimer plays in organising and recruiting additional proteins to the primed template DNA as the replisome alternates between modes of replication and repair.

The β2 clamp is loaded onto primed DNA by interaction with the δ subunit of the clamp loader complex (see Section 1.3.1.5) in an ATP dependent reaction. Binding of the δ subunit to β2 occurs at one of the two hydrophobic binding pockets of the β dimer, as well as interacting with a helical segment that is thought to destabilise one of the two clamp dimer interfaces, opening the ring (Jeruzalmi et al., 2001) and positioning it around dsDNA. It is proposed that interactions between DNA and residues in the binding pocket may displace δ as the clamp is being loaded, breaking connections with the clamp loader complex and closing the ring around DNA in the process (Jeruzalmi et al., 2001). Thus, the process of clamp loading appears to be modulated by DNA and is based on the existence of common binding sites within the clamp (Naktinis et al., 1996).

The α subunit on its own interacts with the clamp quite weakly ($K_D = 800 \text{ nM}$) while the core complex including ε and θ interacts more strongly ($K_D = 250 \text{ nM}$) and is further stabilised by the addition of DNA ($K_D = 5 \text{ nM}$; Naktinis et al., 1995, 1996; Kim and McHenry, 1996).

1.5.1.5 The clamp-loader complex

The clamp loader complex (CLC) is a seven subunit protein complex within the Pol III HE, with the composition $\tau_\gamma \tau_3 \cdot \delta \delta' - \chi\psi$ (where $n = 0–3$, where physiologically relevant assemblies have $n \geq 2$) encoded by five genes (reviewed in McHenry, 2011; Lewis et al., 2016). The complex utilises ATP to load β2 sliding clamps around dsDNA and chaperone polymerase cores ($\alpha\theta$)
into position to associate with the loaded clamp and template DNA (Ason et al., 2003). The clamp-loader complex also serves as a central organiser of coupled leading and lagging strand replication through interactions with both the α polymerase subunits and the DnaB helicase enzyme during processive DNA synthesis (see Figure 1.6 for a simplified schematic illustration of its domains). Five of the seven subunits (γ/δ, δ and δ’) of the CLC belong to the AAA+ ATPase family and form a horse-shoe shaped complex that binds and hydrolyses ATP in a DNA dependent manner during clamp-loading (Figure 1.11; Simonetta et al., 2009).

![Crystal structure of the γδδ'-DNA complex with bound nucleotide analogue ADP-BeF₃.](image)

*Figure 1.11: Crystal structure of the γδδ’-DNA complex with bound nucleotide analogue ADP-BeF₃. Each γ subunit is coloured in a different shade of blue, the δ subunit is coloured yellow, the δ’ subunit is coloured green and the centrally positioned primer-template DNA is in orange. Molecules of bound ADP-BeF₃ are shown in pink as spheres.*

The δ subunit of the clamp loader is the 38.7 kDa product of the holA gene (Dong et al., 1993) and, during clamp loading, makes contact with the C-terminal face of the β₂ clamp between β domains I and II and also within the common hydrophobic β-binding pocket at which a raft of proteins interact (Jeruzalmi et al., 2001). Opening the clamp ring occurs via an allosteric interaction modulated by ATP binding and hydrolysis by the clamp loader (Naktinis et al., 1995). Closing of the clamp around DNA and
detachment from the clamp loader complex involves the subsequent hydrolysis of up to two further ATP molecules (Hingorani et al., 1999; Cho et al., 2014).

The δ' subunit of the clamp loader is the 36.9 kDa product of the holB gene (Dong et al., 1993) that binds the δ subunit and either one τ or one γ subunit (both products of the dnaX gene; see following paragraph) of an assembled clamp loader complex (Song and McHenry, 2001; Bullard et al., 2002). The δ' subunit forms a tight 1:1 complex with the δ subunit that is isolable as a stable complex by gel filtration (Onrust and O'Donnell, 1993). Binding of the δ' subunit to δ is thought to modulate the interaction between δ and the β2 clamp (Turner et al., 1999). The interaction between δ' and τ/γ is not as tight in the absence of the δ subunit; however, a complex of δ', γ and the clamp loader accessory protein ψ was stable enough for copurification by gel filtration (Onrust et al., 1995a).

The τ (71 kDa) and γ (47 kDa) subunits are both products of the dnaX gene (McHenry, 1982; Kodaira et al., 1983; Mullin et al., 1983). The γ subunit is produced by a −1 ribosomal translational frameshift caused by a hairpin following a “slippery” oligo(A) sequence in the mRNA. This frameshift has an efficiency of 40% (80% in an overproducing cell) and can be abolished by a mutation in the oligo(A) region (Flower and McHenry, 1990; Tsuchihashi and Kornberg, 1990). The τ subunit comprises five domains, with domains four and five being absent in the truncated γ subunit (Jeruzalmi et al., 2001b; Gao and McHenry, 2001). Domains I and II are involved in ATPase activity, while domain III participates in oligomerisation (Kelman and O’Donnell, 1995; Gao and McHenry, 2000). The 24 kDa C-terminal fragment unique to the τ subunit contains two additional domains, with domain IV responsible for interaction with the DnaB6 helicase (Gao and McHenry, 2001b) and domain V containing the site of interaction with the α subunit (Su et al., 2007; Jergic et al., 2007) to chaperone the polymerase onto template DNA. Each τ
subunit of the clamp loader interacts tightly with the core polymerase at the C-terminal end of the α subunit \( (K_D = 260 \text{ pM}; \text{Jergic et al., 2007}) \), forming a bridge between the two \( \alpha\theta\beta_2 \) assemblies on the leading and lagging strands and centrally positioning the clamp loader complex at the replication fork.

Functional clamp loader complexes are often reconstituted \textit{in vitro} using either three copies of the τ subunit or, more traditionally, two copies of the τ subunit and one of the γ subunit (McHenry, 1982; Cull and McHenry, 1995). The presence of the truncated γ subunit within clamp loaders \textit{in vivo}, however, has been a source of debate. Experimental evidence exists that suggests clamp loaders under physiological conditions can assemble with three τ subunits (McInerney \textit{et al.}, 2007; Reyes-Lamothe \textit{et al.}, 2010), which is in opposition to the long held view that the complex must contain one copy of the γ subunit (reviewed by McHenry, 2011). The presence of three copies of τ would offer the interesting possibility that the holoenzyme can bind three \( \alpha\theta\theta \) cores at one time, thereby keeping one as a ‘backup’ for efficient exchange during Okazaki fragment turnover. Recent \textit{in vivo} experiments by McHenry and colleagues, however, where the only source of γ in the cell was biotin-tagged, quantified one molecule of γ per clamp assembly (Dohrmann \textit{et al.}, 2016). It remains possible that clamp loader composition \textit{in vivo} may be largely stochastic, with some holoenzymes containing three molecules of τ and others two.

Both the χ (the 17 kDa product of the \textit{holC} gene) and ψ (the 15 kDa product of the \textit{holD} gene) subunits (Carter \textit{et al.}, 1993b and c; Xiao \textit{et al.}, 1993) are accessory proteins that are not essential for clamp loading. The χ subunit however, does play a role in enhancing replisome processivity by binding to single-stranded DNA binding (SSB) proteins (see Section 1.3.1.6). Omission of χ from the complex prevents contacts with SSB and lowers the efficiency of the clamp loader under elevated salt conditions. A peptide of 15 C-terminal residues of SSB has been shown to be sufficient to bind to χ.
demonstrating the extreme C-terminal end of SSB to be the site of interaction with χ. This interaction presumably directs the Pol III HE to SSB coated sites of ssDNA (Kelman et al., 1998).

### 1.5.1.6 Single-stranded DNA binding protein (SSB)

The *E. coli* single-stranded binding protein (SSB) is encoded by the *ssb* gene and has a subunit weight of 18.9 kDa (Sancar et al., 1981). Four monomers assemble into a functional tetramer in solution as a dimer of dimers, with crystallographic modelling displaying two different intersubunit binding interfaces (Raghunathan et al., 1997). The SSB tetramer binds non-specifically to ssDNA in three known binding modes, sequestering either 35 ± 2 nucleotides [(SSB)$_{35}$; see Figure 1.12], 56 ± 3 nucleotides [(SSB)$_{36}$] or 65 ± 3 nucleotides [(SSB)$_{65}$; Lohman and Overman, 1985; Bujalowski and Lohman, 1986]. It plays a protective role by shielding ssDNA from damage or self-annealing during replication (Meyer and Laine, 2000). The *in vitro* binding mode preferred has been shown to be dependent on a number of variables, including temperature, salt concentration, pH and protein concentration. Alternate binding modes have been theorized to play a role in different cellular processes (Bujalowski and Lohman, 1986).

![Figure 1.12: The SSB tetramer bound to ssDNA in the 65 mode.](image)

In this mode, the four SSB monomers together bind and sequester 65 ± 3 nucleotides. Each protein monomer is coloured differently in purple, green, blue and salmon and ssDNA is shown in orange. The image is displayed using a cartoon representation (PDB ID: 1EYG; Raghunathan et al., 2000).
In addition to the well characterized interaction with ssDNA, the SSB tetramer also interacts with over a dozen genomic maintenance proteins, including DnaG primase, the χ accessory subunit of the clamp loader complex, Pol II, Pol V and exonuclease I (reviewed by Shereda et al., 2008). The common site for all characterised interactions is a highly conserved C-terminal motif (in E. coli this motif is located from residues 169–177 of each of the 177 amino acid monomer chains: -Met-Asp-Phe-Asp-Asp-Ile-Pro-Phe; Lu and Keck, 2008) that remains disordered in full-length crystal structures of SSB tetramers, suggesting this C-terminal ‘tail’ remains flexible in the context of the tetramer (Matsumoto et al., 2000; Savvides et al., 2004). The SSB tetramer therefore plays an important role in replication as an organizer protein that interacts with a network of protein subunits, localizing replication enzymes to ssDNA or the replication fork in general during DNA synthesis and repair.

1.6 Studying complex multi-protein machines

The sophisticated architecture of Pol III HE represents a level of molecular organisation and complexity in the cell that was impossible to envision as recently as a few decades ago. Many important processes in the cell are now known to involve the cooperative action of many protein subunits that function in concert as molecular machines. Molecular-scale machines function on many of the same principles as human-scale machines to perform work; comprising multiple interlinking moving parts, each with their own discrete function, utilising energy in an organised way to perform work.

Many cellular operations require highly specialised and intricate molecular machineries that are a challenge to investigate experimentally. Complexes of proteins are often very large (into the MDa molecular weight
range) and contain multiple unstructured and highly flexible peptide regions, making static structure determination of complete complexes by crystallography difficult and often unfeasible. Structural information provided by nuclear magnetic resonance (NMR) is generally limited to proteins of a molecular weight under approximately 50 kDa (notable exceptions to this limit exist; Fiaux et al., 2002) and the techniques of small-angle X-ray scattering (SAXS) is resolution limited and more suitable for studying the overall shape and dimensions of a protein of protein complex. The resolution limit of electron microscopy (EM) is rapidly improving, generating structural models of under 10 Å resolution and allowing for the accurate building of atomic resolution multi-protein complex models by docking crystallographic or NMR structures of individual component subunits or domains (Fernandez-Leiro et al., 2015).

Individual protein subunits within a large dynamic complex are held together by a network of interactions that are organised in a hierarchy of strength. The tightest and most stable of these interactions (\( K_D \) in the nM range or lower) are more readily identified and characterised using traditional biochemical techniques such as affinity purification or gel-filtration, while more advanced techniques such as surface plasmon resonance (SPR) and native electrospray ionisation (ESI) mass spectrometry can detect significantly weaker interactions (\( K_D \) in the \( \mu \)M range or lower). The development of single-molecule methodologies over the last decade has provided researchers with a powerful new approach to investigate complex enzyme kinetics, without the averaging inherent in ensemble experiments. A suite of single-molecule based techniques, including fluorescence and force manipulation experiments, are now routinely exploited in the laboratory (Deniz et al., 2008; Tanner et al., 2008, van Oijen and Dixon 2015).

Modelling the atomic structures of large multi-subunit protein complexes and identifying and understanding the physiological significance
of weak interactions within the context of a protein complex such as Pol III HE involves a manifold experimental approach. Structural and biochemical information acquired from a range of techniques can be pieced together gradually in an effort to refine structural models and generate new hypotheses regarding the function of large protein assemblies. A range of structural and functional experimental techniques are utilised within the work described in this Thesis, including protein crystallography, SAXS, NMR and investigations into a fluorescence-based single-molecule technique.

1.7 Scope of this Thesis

The scope of my PhD programme was to study the structure and function of the subunits of the β2αεθ complex, specifically focusing on the structure and functional role of the ε proofreading subunit within the Pol III HE. In Chapter 3 I present a crystal structure displaying the site of interaction between the α and ε subunits. In Chapter 4 I utilise this structural information, along with additional structural and biochemical data, to produce an informed structural model of the β2αεθ complex. I then co-purify a stabilised version of this complex and use small-angle X-ray scattering to validate this model. In Chapter 5 I investigate a protein labelling strategy by incorporating a protein tag into Pol III core protein subunits that can be site-specifically modified by the formylglycine generating enzyme and then conjugated to a chemically modified fluorescent dye. This work is ongoing in the Dixon laboratory and I present my current contribution to the project.
Chapter 2

General materials and methods
2.1 General Materials

2.1.1 Bacterial strains

The *E. coli* strain AN1459 (F- *ilv leu thr supE recA srlA::Tn10; Vasudevan et al., 1991) was used as a host for plasmid preparation and selection during molecular cloning as well as a long-term storage strain for maintenance. The commercially available strain DH5α [fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80Δ (lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1] (ThermoFisher Scientific) was also used at times for the same purposes. Both strains lack the necessary T7 gene 1 required to transcribe genes downstream of a phage T7 promoter and are utilised where required to eliminate ‘leaky’ expression of genes contained within the frequently used pET expression plasmid system. The dnaN gene (coding for the *E. coli* β2 sliding clamp) was expressed under the control of heat-inducible phage λ promoters in the AN1459 strain (Vasuvedan et al., 1991).

The pET expression system places genes under control of the T7 promoter and enables overexpression of proteins by induction with the allolactose mimic isopropyl β-D-1-thiogalactopyranoside (IPTG). The *E. coli* strains BL21(λDE3)recA [recA srlA::Tn10] (Williams et al., 2002) or BL21(λDE3)/pLysS (Studier et al., 1990) were used as hosts for protein overproduction of genes contained within a pET expression vector. These strains lack the *lon* and *ompT* protease genes and limit proteolysis of proteins during purification. The *recA* mutation stabilizes plasmids containing host genes by suppressing their integration in the *E. coli* chromosome. The pLysS plasmid codes for the production of T7 lysozyme, which inhibits T7 RNA polymerase and prevents the ‘leaky’ expression of potentially toxic genes prior to the induction of synthesis of a target protein. The pLysS plasmid confers the host strain with chloramphenicol resistance.
2.1.2 Plasmid vectors and expression systems

The plasmid vector pETMCSI (Neylon et al., 2000; Figure 2.1A) was used as the primary vector scaffold for most cloning and gene expression experiments reported in this Thesis. This vector contains a multiple cloning site (MCS) flanked upstream by the T7 ϕ10 promoter and downstream by the Tϕ transcription terminator (Studier et al., 1990). Genes under control of the T7 expression system were induced by the addition of IPTG in derivatives of the BL21(λDE3) host strain described above (Section 2.1.1). Alternatively, genes under control of the T7 expression system were at times overexpressed using an autoinduction growth medium (Section 2.1.4.3). Using this approach, cell cultures were grown at lower temperatures (<25°C) to high densities using a phosphate buffering system with glucose and glycerol present in the medium as the primary metabolic carbon sources. Once these preferential carbon sources are exhausted, cells begin to metabolise lactose present in the autoinduction medium, generating allolactose that induces production of T7 RNA polymerase under the control of the lac promoter on the λDE3 prophage and stimulates high-level production of a protein of interest. The plasmid pETMCSI contains a ColE1 origin of replication (ori) and confers the host cell with resistance to ampicillin by transcription and translation of the bla gene.

Plasmid vector pCE30 (Love et al., 1996; Figure 2.1B) was utilised only once in this work, as a host plasmid for the dnaN gene (coding for the β subunit) that had been inserted within the MCS. This plasmid construct is designated as pND261 (Oakley et al., 2003). The parental vector pCE30 contains tandem pr and pl promoters from the bacteriophage λ, which enable the high-level transcription of a downstream gene. Transcription from these promoters is controlled by the product of the cl857 gene, which is able to effectively eliminate expression of genes under the control of the pr and pl.
promoters by binding to operator sequences at 30°C. Heating of cultures to 42°C stimulates dissociation of the temperature-sensitive repressor protein from the operators and permits high-level transcription of a downstream gene(s) of interest.

Plasmid vector pND517 (Figure 2.1C) contains the wild-type dnaE gene (coding for the E. coli α polymerase subunit) under control of the IPTG-inducible tac promoter and had been previously constructed to optimise the translation initiation region upstream of the dnaE gene and boost overproduction of α (as described in Wijffels et al., 2004). The plasmid pKO1538 (coding for the α mutant αE612K; overproduction and purification described in Section 4.2.3.8) had also previously been constructed by generating a single-point mutation in the dnaE gene of pND517. This plasmid was provided as a kind gift from Dr Kiyoshi Ozawa. The α variant of the dnaE gene (overproduction and purification described in Section 4.2.4.3) was contained within the parental vector pETMCSI (Jergic et al., 2013).

Plasmid vector pBOB (Figure 2.1D) was constructed by Dr Slobodan Jergic by removing the dnaE gene from pND517 and adding an NdeI restriction site at its ATG start codon. This vector construct enables simple transfer of gene inserts between pET based vectors (i.e. pETMCSI) and the pBOB vector, which contains the optimised translation initiation region found in pND517. The vector pSJ2132 (Dr Slobodan Jergic) contains the dnaE gene within the pBOB vector and was utilised in this Thesis for the one-step transfer of the dnaE gene into a pETMCSI based vector containing a nucleotide region coding for a specific protein tag (described in detail in Section 5.2.4).
Figure 2.1: Plasmid maps of (A) pETMCSI, (B) pCE30, (C) pND517 and (D) pBOB. Selected genes and regulatory regions on each plasmid are highlighted using coloured arrows or bars (each is discussed in further detail in the accompanying text). Commonly utilised restriction sites within each plasmid are annotated using a pink dash (Figure 2.1C and D) or listed in sequential order as part of a multiple cloning site (MCS; Figure 2.1A, B and D).

2.1.3 Chemicals and reagents

Table 2.1: List of chemicals and reagents used.

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Chemicals and reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ajax Finechem, Australia</td>
<td>acetone, acetic acid, ammonium sulfate (NH₄)₂SO₄, ethanol, potassium dihydrogen orthophosphate (KH₂PO₄), dipotassium hydrogen orthophosphate (K₂HPO₄), monosodium phosphate (NaH₂PO₄·2H₂O), disodium phosphate (Na₂HPO₄·2H₂O), propan-2-ol (isopropanol), 36% hydrochloric acid (HCl), magnesium chloride (MgCl₂)</td>
</tr>
<tr>
<td>Company</td>
<td>Product Details</td>
</tr>
<tr>
<td>------------------------</td>
<td>---------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>AMRESCO, USA</td>
<td>bromophenol blue, sodium salt</td>
</tr>
<tr>
<td></td>
<td>sodium dodecyl sulphate (SDS)</td>
</tr>
<tr>
<td>Astral Scientific, Australia</td>
<td>isopropyl-β-D-1-thiogalactopyranoside (IPTG)</td>
</tr>
<tr>
<td></td>
<td>dithiothreitol (DTT)</td>
</tr>
<tr>
<td></td>
<td>ethidium bromide (EtBr; 10 mg/mL in water)</td>
</tr>
<tr>
<td>Bacto, Australia</td>
<td>Bacto™ tryptone</td>
</tr>
<tr>
<td></td>
<td>Bacto™ yeast extract</td>
</tr>
<tr>
<td></td>
<td>Difco™ agar technical</td>
</tr>
<tr>
<td></td>
<td>Difco™ Luria-Bertani (LB) broth, Miller</td>
</tr>
<tr>
<td>Bio-Rad, Australia</td>
<td>Precision Plus Protein™ Dual Color standards (protein molecular weight markers)</td>
</tr>
<tr>
<td></td>
<td>native PAGE sample buffer + β-mercaptoethanol</td>
</tr>
<tr>
<td>Bioline, Australia</td>
<td>ammonium acetate</td>
</tr>
<tr>
<td></td>
<td>glycerol</td>
</tr>
<tr>
<td></td>
<td>deoxynucleotide triphosphate mix (dNTP's)</td>
</tr>
<tr>
<td>Chem-supply, Australia</td>
<td>agarose</td>
</tr>
<tr>
<td>Fluka, Australia</td>
<td>cesium iodide (CsI)</td>
</tr>
<tr>
<td></td>
<td>dimethyl sulfoxide (DMSO)</td>
</tr>
<tr>
<td></td>
<td>formic acid</td>
</tr>
<tr>
<td></td>
<td>phenylmethanesulfonyl fluoride (PMSF)</td>
</tr>
<tr>
<td>Geneworks, Australia</td>
<td>oligonucleotide primers</td>
</tr>
<tr>
<td>New England Biolabs, USA</td>
<td>restriction endonucleases (BamHI, EcoRI, Ncol, Ndel, Xbal, Xhol)</td>
</tr>
<tr>
<td></td>
<td>antarctic phosphatase</td>
</tr>
<tr>
<td></td>
<td>T4 DNA ligase</td>
</tr>
<tr>
<td>Roche Diagnostics, Switzerland</td>
<td>ampicillin</td>
</tr>
<tr>
<td></td>
<td>Completer Protease Inhibitor Cocktails</td>
</tr>
<tr>
<td>Sigma-Aldrich, USA</td>
<td>α-lactose monohydrate</td>
</tr>
<tr>
<td></td>
<td>2-amino-2-hydroxymethylpropan-1,3-diol (Trizma™/Tris)</td>
</tr>
<tr>
<td></td>
<td>β-mercaptoethanol (BME)</td>
</tr>
<tr>
<td></td>
<td>bovine serum albumin (BSA)</td>
</tr>
<tr>
<td></td>
<td>boric acid</td>
</tr>
<tr>
<td></td>
<td>calcium chloride (CaCl₂)</td>
</tr>
<tr>
<td></td>
<td>Coomassie brilliant blue R</td>
</tr>
<tr>
<td></td>
<td>chloramphenicol</td>
</tr>
<tr>
<td></td>
<td>Ethylenediaminetetraacetic acid (EDTA)</td>
</tr>
<tr>
<td></td>
<td>D-(+)-glucose</td>
</tr>
<tr>
<td></td>
<td>glycine</td>
</tr>
<tr>
<td></td>
<td>imidazole</td>
</tr>
<tr>
<td></td>
<td>magnesium chloride (MgCl₂)</td>
</tr>
<tr>
<td></td>
<td>magnesium sulphate (MgSO₄)</td>
</tr>
<tr>
<td></td>
<td>2-(N-morpholino)ethanesulfonic acid (MES)</td>
</tr>
<tr>
<td></td>
<td>polyethyleneglycol-3350 (PEG-3350)</td>
</tr>
<tr>
<td></td>
<td>potassium chloride (KCl)</td>
</tr>
<tr>
<td></td>
<td>sodium 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethane-1-sulfonate (HEPES Na⁺)</td>
</tr>
<tr>
<td></td>
<td>sodium acetate (NaOAc)</td>
</tr>
<tr>
<td></td>
<td>sodium chloride (NaCl)</td>
</tr>
<tr>
<td></td>
<td>sodium hydroxide (NaOH)</td>
</tr>
<tr>
<td></td>
<td>spermidine·3HCl</td>
</tr>
<tr>
<td></td>
<td>thymine</td>
</tr>
<tr>
<td></td>
<td>zinc chloride (ZnCl₂)</td>
</tr>
<tr>
<td>Thermo Scientific, Australia</td>
<td>GeneRuler™ DNA Ladder Mix (DNA molecular weight markers)</td>
</tr>
</tbody>
</table>
2.1.4 Growth media

2.1.4.1 LB liquid medium

E. coli strains were grown in autoclaved LB medium (Luria and Burrous, 1957) supplemented with 25 mg/L thymine (LBT) and, as required, ampicillin (100–200 mg/L; LBTA) and/or chloramphenicol (35 mg/L; LBTA/C). Strains containing plasmids with gene expression under the strict control of λ promoters (e.g. pND261) were grown at 30°C and rapidly heated to 42°C to stimulate protein over-expression. Cells containing tac promoter or T7 expression vectors (i.e. pETMCSI derivatives) were grown at 37°C or, where required, 30 or 39°C to limit unwanted protein expression and/or increase solubility. Genes under the control of tac and T7 promoters were induced with the addition of 0.5 mM IPTG to growth medium.

2.1.4.2 LB solid medium (agar plates)

E. coli strains were grown on LB agar plates [15 g/L agar in LB medium, supplemented with appropriate antibiotics (Section 2.1.4.1)] in a temperature-controlled incubator at 30°C (for strains containing pCE30 derivatives) or 37°C overnight.

2.1.4.3 LB autoinduction medium

This medium was used to enhance culture density and increase protein yield and/or when protein expression was required at lower temperatures (< 30°C) for improved protein solubility. One litre of medium (Studier, 2005) comprises 930 mL of sterile LB medium (Section 2.1.4.1), 20 mL of sterile filtered 50× 5052 mixture (25% w/v glycerol, 2.5% w/v glucose, 10% w/v α-lactose), 50 mL of sterile filtered 20× NPS mixture (0.5 M (NH₄)₂SO₄, 1 M KH₂PO₄, 1 M Na₂HPO₄ at pH 6.75), 25 mg/L thymine, 1 mM
MgSO₄, supplemented with 100–200 mg/L ampicillin and 35 mg/L chloramphenicol, where required.

2.1.4.4 Minimal medium for ¹³C,¹⁵N labelling

Double-labelled ¹³C,¹⁵N proteins were expressed in a M9 minimal medium (Miller, 1972), containing 100 mM NaH₂PO₄ at pH 7.1, trace salts, 1 mM MgSO₄, 100 mg/L ampicillin, 35 mg/L chloramphenicol, 1.5 g/L ¹³C D(+) glucose (CortecNet, France) as the only carbon source and 1.5 g/L ¹⁵NH₄Cl (Cambridge Isotope Laboratories, USA) as the sole source of nitrogen.

2.2 General Methods

2.2.1 Molecular cloning procedures

2.2.1.1 Polymerase chain reaction (PCR)

Amplification of target DNA fragments was achieved using standard in vitro DNA replication and thermal cycling techniques, refined from those first described by Kleppe et al. (1971) and Saiki et al. (1988). Reaction mixtures contained, in general, 4–8 ng of purified plasmid template DNA, 0.3 μM each of forward and reverse primers (15–25mer), 1 mM dNTP mix, 2.5 units of PfuTurbo DNA polymerase (Stratagene/Agilent, USA) in 1× supplied buffer (20 mM Tris-HCl pH 8.8, 2 mM MgSO₄, 10 mM KCl, 10 mM (NH₄)₂SO₄, 1 % Triton X-100, 1 mg/mL BSA), made up to 50 μL volume with sterile Milli-Q H₂O. PCR cycle conditions were: a single 2 min denaturation step at 95°C, followed by 34 cycles of 30 s denaturation at 95°C, 30 s annealing at 55–65°C (dependant on calculated primer melting temperatures) and 1 min/kb extension at 72°C, before concluding with a
single 10 min extension period at 72°C. Reaction products were observed using agarose gel electrophoresis (Section 2.2.1.9).

2.2.1.2 Colony PCR

Following ligation and transformation, plasmids in transformants were screened for the presence of insert DNA directly from colonies using a modified PCR protocol to that described in Section 2.2.1.1. Cells from a selected colony were suspended in 50 μL of sterile Milli-Q H₂O and 1 μL of cell mix was added to 1 unit of REDTaq® (Sigma-Aldrich, St. Louis, MO, USA) together with 1× supplied buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.1 mM MgCl₂, 0.01% gelatin), 0.3 μM each of forward and reverse primers flanking the target insert region (15–25mers), 1 mM dNTP mix, with Milli-Q H₂O added to give a total reaction volume of 10 μL. PCR cycle conditions were as described in Section 2.2.1.1, with an extended 10 min initial denaturation step to ensure cell lysis and release of plasmid DNA.

2.2.1.3 Site directed mutagenesis PCR

Site directed mutagenesis was performed using the QuikChange® mutagenesis kit (Stratagene). The PCR cycling protocol used was similar to that outlined in Section 2.2.1.1 for a standard polymerase chain reaction, with the exceptions of an increased extension time (1 min per kb) required to replicate the full circle of plasmid and only 18 thermal cycles were performed. The plasmid DNA products were subsequently transformed into E. coli strains AN1459 or DH5α for selection (Section 2.1.1).
2.2.1.3 Transformation of competent E. coli cells

Chemically competent E. coli cells were prepared using CaCl$_2$ following the procedure outlined by Morrison (1979). Competent cells and plasmid DNA were mixed on ice for 30 min then subjected to heat shock treatment at 42°C for 1 min. Following heat shock, cells were recovered in LB medium (Section 2.1.4.1) at 37°C for 1 h then plated on LB agar containing appropriate antibiotic(s) (Section 2.1.4.2). Single colonies of cells were selected and struck-purified to ensure genetically homogenous cell lines. Less frequently, electrocompetent cells prepared and utilized as described by Miller and Nickoloff (1995) were used for the transformation of plasmid DNA using electroporation. When required, electrocompetent BL21pLysS cells (Section 2.1.1) were obtained courtesy of Dr Slobodan Jergic.

2.2.1.4 Plasmid extraction

Small-scale plasmid preparations (up to 10 μg) were prepared from cells grown overnight on LB plates (Section 2.1.4.2) containing relevant antibiotics. Cells were collected, lysed and purified using the commercially available QIAprep Spin Miniprep Kit (Qiagen, Germantown, MD, USA). Plasmid DNA obtained using this method was subsequently used for restriction endonuclease digestion, sequencing PCR or transformation of competent cells.

2.2.1.5 Restriction endonuclease digestion

All restriction endonucleases (e.g. NdeI, EcoRI and BamHI) and their recommended buffers were supplied by New England Biolabs (NEB, USA). Digestions of DNAs were carried out in the appropriate buffer as recommended and supplied by NEB. Reactions were generally performed at
37°C for 2–4 h using 10–20 units of the required enzyme(s) and 2–5 μg of plasmid DNA, dependant on molecular weight. Following digestions, reactions were terminated using heat inactivation at 65–75°C for at least 20 min. Digested vector DNA samples were frequently treated with 1–5 units of Antarctic phosphatase (NEB) at 37°C for 20 min per reaction mixture to remove 5ʹ-phosphate monoesters to prevent self-ligation. Digestions were assessed for completion using agarose gel electrophoresis (Section 2.2.1.9) and the desired fragment(s) isolated directly from the gel or by using a DNA binding spin-column (Section 2.2.1.11).

2.2.1.7 Isolation of DNA fragments

DNA fragments generated by restriction digestion were recovered and purified from agarose gels using the commercially available QIAquick Gel Extraction Kit (Qiagen). Alternatively, fragments between 100 bp and 10 kb in length were separated from smaller nucleotide contaminants (i.e. after digestion of a PCR product) using the QIAquick PCR Purification Kit (Qiagen).

2.2.1.6 Ligation

Ligation of DNA fragments was performed using T4 DNA ligase (NEB) together with supplied ligation buffer, as specified by NEB and described by Sambrook et al. (1989). A standard reaction contained 60–80 ng of vector DNA and 3–6 units of T4 ligase enzyme. To maximise efficiency, a 3:1 molar ratio of insert to vector DNA was targeted. Reaction mixtures (20 μL) were initiated at room temperature (22°C) then placed in a fridge within an insulated container to cool slowly to 6°C overnight.
2.2.1.8 Nucleic acid quantification

Concentration of double-stranded (ds) DNA was determined spectrophotometrically using a NanoDrop 2000c UV-Vis spectrophotometer with packaged control and reporting software (Thermo Scientific), assuming a solution with $A_{260} = 1.0$ contains 50 μg/mL of dsDNA. Plasmid or oligonucleotide DNA was routinely stored in TE buffer (10 mM Tris.HCl, pH 7.6, 1 mM EDTA) at –20°C or –80°C.

2.2.1.9 Agarose gel electrophoresis

Plasmid and oligomeric DNA were separated for preparative and analytic purposes using a Mini-Sub Cell GT submarine horizontal electrophoresis unit (Bio-Rad, Australia). Gels were prepared using 0.7–1.5% agarose (depending on the DNA separation characteristics required; Bioline, Australia) in TAE buffer (40 mM Tris-acetate, pH 8.2, 1 mM EDTA; Aaij and Borst, 1972; Hayward and Smith, 1972) containing 0.5 μg/mL ethidium bromide (EtBr; Bio-Rad). Prior to loading, samples were mixed with a DNA loading dye (10 mM Tris-HCl, pH 7.6, 0.03 % bromophenol blue, 0.03 % xylene cyanol FF, 60% glycerol, 60 mM EDTA; ThermoFisher Scientific). Electrophoresis was performed at 40–80 V until the required resolution of DNA fragments was achieved. DNA was visualised using a UV transilluminator (Universal Hood II Gel Doc, Bio-Rad) and band sizes compared to a commercial DNA marker (GeneRuler 100–10,000 bp; ThermoFisher Scientific) that was electrophoresed concurrently with experimental samples.
2.2.1.10 Extraction from agarose gels

DNA fragments were visualised in agarose gels containing EtBr with long-wavelength UV light. DNA bands of interest were excised from the gel using a sterile scalpel blade and purified from the gel using the QIAquick® gel extraction kit (Qiagen).

2.2.1.11 Purification by spin-column

DNA fragments were purified from a reaction mixture using the QIAquick® PCR purification kit (Qiagen). The kit was most often used to remove contaminants following a PCR reaction or enzymatic digestion.

2.2.1.12 Nucleotide sequencing PCR

Plasmid or oligonucleotide DNA was prepared for sequencing by PCR using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA). Vector specific primers designed to flank gene inserts were commonly used to prepare samples for sequencing of whole genes. Primers designed to anneal within a gene of interest were also employed, where required, to extend sequencing coverage over a long gene or target a region of interest within a gene. Reaction mixtures and thermal cycling conditions were as recommended in the protocols accompanying the sequencing kit (Applied Biosystems). Following PCR cycling, DNA was purified from the reaction mixture by precipitation with ethanol, centrifugation and the removal of supernatant by pipetting and evaporation. Samples were submitted to the sequencing facility at the School of Biological Sciences, University of Wollongong for analysis using a 3130xl Genetic Analyzer (Applied Biosystems). Sequencing data were analysed using

2.2.1.13 Custom oligonucleotides

Oligonucleotide primers for cloning and sequencing procedures were obtained from GeneWorks (Hindmarsh, SA, Australia). Dried oligo samples were prepared for use by the addition of TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA) to a stock concentration of 100 μM and stored at −20°C.

2.2.2 Protein biochemistry methods

2.2.2.1 Fast protein liquid chromatography (FPLC) purification of proteins

Chromatographic purification of overproduced proteins was performed utilising columns containing a range of stationary-phase resins and other materials (e.g. hydroxylapatite) that were connected to an automated FPLC instrument (ÄKTApurifier or ÄKTAFPLC, GE). The delivery of mobile-phase solutions and the monitoring and collection of eluent from connected columns was monitored, controlled and evaluated in real-time or by executing programmed methods using UNICORN software (GE Healthcare). All chromatography was performed in a cold room at 4–6°C.

2.2.2.2 Concentration of proteins

Purified protein samples were most often concentrated by precipitation with ammonium sulphate or by centrifugation using a size exclusion filter unit. Proteins were precipitated by the addition of solid ammonium sulphate to samples (0.45 g/mL) then centrifugation and resuspension of protein pellets in a reduced volume of buffer.
Contaminating ammonium sulphate was removed from the protein pellet by thorough dialysis into an appropriate buffer. Alternatively, centrifugal filter units (Amicon Ultra-4 or Ultra-15; Millipore, Germany) of a desired molecular weight cut-off (MWCO) were used to reduce the volume of buffer in a sample while retaining the desired protein.

On occasion, samples of very dilute protein were required to be concentrated for analysis by SDS-PAGE using precipitation with acetone. Four volumes of ice-cold acetone were added to samples, which were then left on ice for at least 15 min before centrifugation (20 000 x g; 20 min) to collect protein pellets. Following centrifugation, the supernatant was removed and protein pellets were air-dried. An appropriate volume of protein loading buffer (Section 2.2.2.4) was then added to each pellet to achieve the desired protein concentration.

The selection of method used for the concentration of proteins was based either on prior experience handling a particular protein or the inferred stability of a protein.

2.2.2.3 Dialysis of proteins

Dialysis of samples for buffer exchange was performed routinely during protein purification and on occasion for experimental sample preparation. Regenerated cellulose dialysis tubing (Spectra/Por, Spectrum Laboratories, USA) was most often utilised for sample volumes >500 μL, while smaller volumes required for mass spectrometric measurements were buffer exchanged using Slide-A-Lyzer mini dialysis units (ThermoFisher Scientific). Depending on the extent of buffer exchange required, samples were most often dialysed against two or three changes of buffer (volume normally at least 1 L each) for a minimum of 4 h intervals. An appropriate molecular weight cut-off (MWCO) dialysis membrane was selected to ensure
protein molecules of interest were retained within the tubing or mini dialysis units.

2.2.2.4 Sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE)

SDS-PAGE was used extensively throughout this Thesis, most often for monitoring the purity of a protein following each purification step in a multi-step procedure. The electrophoresis protocol used was based on that first outlined by Laemmli (1970). Pre-cast SDS-PAGE gradient resolving gels (4–12%) were used on all occasions and were supplied by either Bio-Rad or by Invitrogen/ThermoFisher. Electrophoresis was performed using either the Mini-PROTEAN Tetra cell for Bio-Rad gels or the XCell4 SureLock midi-Cell Runner box system for Invitrogen gels. The running buffer used for Bio-Rad gels was 25 mM Tris base, 190 mM glycine and 0.1% SDS and that used for Invitrogen gels was 50 mM MES, 50 mM Tris base, 0.1% SDS, 1 mM EDTA. Precision Plus Protein Dual Color standards (Bio-Rad) were used as a molecular weight marker on all occasions. Prior to loading within gel wells, samples were mixed with 2 x protein loading dye (300 mM Tris base, 15% (v/v) glycerol, 0.6% (w/v) bromophenol blue, 1% SDS; Sambrook et al., 1989) and fresh β-mercaptoethanol reducing agent to ≈200 mM immediately before loading. Samples containing significant levels of unwanted protein or DNA contamination were heated at 95°C for at least 3–5 min to facilitate easier loading of viscous or partially precipitated samples. This step was not necessary for samples containing highly purified protein. Gels were electrophoresed at 160–180 V for 30–50 min using a connected PowerPac Basic power supply unit (Bio-Rad) until the desired resolution was obtained.

Following electrophoresis, acrylamide gels were removed from their plastic scaffolding and washed well with Milli-Q water. Gels were then
stained for 10–15 min with gentle shaking in a staining solution [0.2% (w/v) Coomassie brilliant blue R, 40% (v/v) methanol, 10% (v/v) acetic acid] that was warmed in a microwave to just under boiling. Gels were then removed from the staining solution and washed well with Milli-Q water before being placed in a destaining solution [40% (v/v) propan-2-ol, 10% (v/v) acetic acid] with gentle shaking that was then similarly heated. To achieve optimal results, the destaining solution was often replaced 4–5 times and left overnight before capturing a final image by scanning.

2.2.2.5 Electrospray ionization (ESI) mass spectrometry (MS)

The accurate molecular mass of purified proteins was assessed using ESI-MS to confirm the identity of the protein produced (required for newly constructed proteins) and assess the success of targeted sequence changes made during molecular cloning procedures. This procedure was most often performed after the sample had been dialysed extensively into a solution of 0.1% formic acid to completely unfold proteins and provide molecules with a positive charge necessary for detection. Using this approach, the molecular mass of purified protein samples could usually be determined to <0.01% of the total molecular mass of the protein. Alternatively, protein samples were also analysed by mass spectrometry in their native conformation following extensive dialysis (up to 5 changes of 0.5–1 L) into a buffer of ammonium acetate (100–500 μM) and the reducing agent β-mercaptoethanol (1 mM). This approach was required where the native conformation of proteins was required to be maintained to assess intra- or intermolecular protein-protein interactions or where a protein was unamenable to accurate mass determination in a formic acid buffer (e.g. the full-length α subunit and its derivatives). Molecular weights determined using this method are generally
less accurate than those determined using formic acid (<0.1% of the total molecular mass of the protein).

Mass spectra were acquired using a quadrupole-time-of-flight (Q-Tof) SYNAPT (Waters, USA) controlled by MassLynx v4.1 software (Micromass, UK). The instrument was calibrated using a sample of 10 mg/mL cesium iodide over the same m/z range required for the detection of sample proteins. Protein samples were analysed at concentrations of ≈1–10 μM.
Chapter 3

STRUCTURE OF THE *E. coli* ε PROOFREADING SUBUNIT C-TERMINAL SEGMENT IN COMPLEX WITH THE α POLYMERASE PHP DOMAIN
3.1 Introduction

The ε exonuclease subunit of DNA Pol III HE is a dual-domain protein comprised of a structured catalytic N-terminal globular domain (residues 1–182) and an intrinsically unstructured C-terminal segment (CTS) of ~61 residues (residues 183–243). C-terminal ε residues (in whole or in part) following the structured N-terminal domain are broadly referred to here as εCTS. The ε subunit binds tightly to the α polymerase subunit ($K_D = 5$ nM) by interaction with residues of the εCTS (Perrino et al., 1999; Wieczorek and McHenry, 2006) and to the θ subunit through its N-terminal catalytic domain (Studwell-Vaughan and O'Donnell, 1993). Solution NMR has demonstrated that some of the residues in the εCTS (183–201, at least) remain flexible even when full-length ε is in complex with the α and θ subunits (Ozawa et al., 2008). The location of these residues proposes a permanently flexible interdomain region (termed a Q-linker because it contains multiple Gln residues) between the α-binding and catalytic domains of ε (Figure 3.1).

![Figure 3.1: Domain organisation schematic of the *E. coli* ε proofreading subunit.](image)

An atomic resolution structure of the αεθ core complex has, so far, remained elusive. Crystal structures are available for the catalytic domain of ε (residues 1–186; Hamdan et al., 2002) and for a C-terminally truncated version (residues 2–917) of the *E. coli* α subunit (Lamers et al., 2006).
Crystallisation of the full-length ε subunit has likely been impeded by the flexibility of the εCTS and, in a similar manner, the flexible interdomain region of the ε subunit would be expected to prevent co-crystallisation of full-length ε with full-length or a C-terminally truncated version of the α subunit. Additionally, the inherent lack of structure in the εCTS and its propensity to aggregate precludes its isolation as a recombinant peptide expressed in vivo (Perrino et al., 1999).

The ε subunit is known to bind within the first 320 residues of the α subunit (Wieczorek and McHenry, 2006), which includes the N-terminal PHP domain (residues 1–270; Lamers et al., 2006). Extensive solution NMR and cross-linking studies performed by Dr Kiyoshi Ozawa (outlined in Ozawa et al., 2008 and Ozawa et al., 2013) identified candidate residues of the εCTS involved in binding to the α N-terminal domain. Residues of the εCTS from Thr183–Thr201 remain highly flexible when in the presence of the α PHP domain, while those between Thr201 and Ala209 appear to be less mobile and potentially comprise a region of transient interaction with α. Those residues C-terminal to Gln208 seem to remain in permanent contact with the α PHP domain, facilitating the tight and stable interaction between the pair. In response to these data, an unconventional approach was taken in an attempt to obtain a crystal structure displaying this interaction.

A chimeric protein construct comprising both the εCTS (ε residues 209–243) and the α N-terminal PHP domain (α residues 1–270) was generated by placing a 9 a.a linker (−TRESGSIGS−) between the two domains, positioned immediately following the C-terminus of εCTS and preceding the N-terminus of the α PHP domain. This linker was chosen for this project as it had been shown previously to successfully cyclise interacting protein domains while remaining highly flexible, as demonstrated by NMR (Williams et al., 2002 and Williams et al., 2005). This approach artificially localised the two interacting regions of the ε and α subunits, facilitating
binding and generating a soluble protein. A crystal structure (1.7 Å) of this molecule was solved in conjunction with Dr Andrew Robinson and Associate Professor Aaron Oakley during an Honours research project (N. Horan, Honours Thesis, 2010; PDB ID: 4GX8), demonstrating the nature of the interaction between these two domains. Subsequent analysis of the structure revealed an unintentional leucine to proline amino acid substitution at a.a position 21 of the α chain, originating from a previously undiscovered PCR generated point mutation. This protein construct is referred to here as ε209α270P.

The crystal structure of ε209α270P revealed an extended area of binding between the εCTS and the αPHP domains (buried surface area 513.84 Å²; see Figure 3.5 for presentation of the lengthened structure of ε200α270P). The contact was characterised by a region of random coil at the C-terminal end of the εCTS, followed by a single α-helix and a second short segment of random coil immediately preceding the interdomain linking peptide (Ozawa et al, 2013 and N. Horan, Honours Thesis, 2010). A region of comparatively poor electron density around the N-terminus of the construct (ε residues Ala209 and Ser210) presented some uncertainty about the path taken by the remaining linker residues towards the N-terminal catalytic domain of ε and its proximal attachment point at the β-binding pocket of the β; sliding clamp (involving ε residues 182–187; Jergic et al. 2013). The twenty-one intermediate linker residues unavailable in the structure of ε209α270P (ε residues 188–208) have the potential to span a contour length >60 Å, providing a high degree of flexibility for conformation changes involving the ε N-terminal catalytic domain that likely occur during switching between polymerisation and proofreading modes. The characterisation of additional sites of interaction involving the ε interdomain linker region and the αPHP domain may further assist in informed model building of the ε subunit within the context of the
β_2αεθ complex and generate new hypotheses concerning the dynamic movement of the proofreading subunit during the replication cycle.

This experimental chapter describes the construction and purification of a number of ε:α_270 variants, comprising successively longer interdomain linker regions beginning at ε residues 200, 189, 180 and 1 (full-length). It was anticipated that a crystal structure of one, or more, extended constructs might provide additional structural information useful for further defining the path of the linker region. In addition to these proteins, a version of ε_209α_270P was produced with the unintentional leucine to proline mutation in the α_270 chain reverted to wild-type (designated here as ε_209α_270). Although this mutation was far from the site of interaction between the two domains and is unlikely to have any effect on the nature of interaction between them, a crystal structure of the unmutated construct would nevertheless be desirable. The corrected protein was also predicted to be reasonably easy to produce and purify in tandem with the additional linked variants.

Localising the εCTS and α_270 domains with a flexible 9 amino acid linker successfully enabled their co-purification as the soluble and crystallisable construct ε_209α_270P. Constraining the two domains with an artificial linker might perturb the site of interaction or prevent the two domains from interacting at their native sites. To test for this possibility a ‘reverse’ linked construct, where the ε_209 and α_270 domains were positioned in the opposite order about the interdomain linking peptide, was produced, purified and analysed here alongside ε_209α_270 using mass spectrometry under native conditions. It was hypothesised that the alternate linker placement within this construct (termed α_270ε_209) would be unable to facilitate stable binding between the two domains and this construct would therefore form domain-swap dimers more readily in comparison to ε_209α_270.
Figure 3.2: Variations of linked ε:α constructs presented as a simplified domain schematic. The α270 domain with and without the Leu21→Pro mutation (designated by the letter “P”) is shown in green. The ε domain (or part thereof) is coloured blue with the subscript number denoting the first a.a in the sequence. The flexible 9-residue inter-domain linker is coloured red. The construct ε209α270P (of which a crystal structure was solved during Honours research) was not investigated within the current work and has been omitted.
3.2 Materials and Methods

3.2.1 Oligonucleotide primers

564: 5’-TTAGGTTACGTTACAGCG-3’ (24-mer)

564A: 5’-AAAGAATTCTTAGGTTACGTTACAGCG-3’ (32-mer; EcoRI site indicated)

565: 5’-TGCTCGCCAGCGGCAACT-3’ (24-mer)

566: 5’-AGTTGCCTCTGGCGAGCA-3’ (24-mer)

568: 5’-AAAAAAATGCATATGGCAACAATTCAGCCATTG (33-mer; NdeI site indicated)

570: 5’-AAAAAAATGCATATGGAAGGAGAGACACAAC (30-mer; NdeI site indicated)

572: 5’-AAAAAAATGCATATGGGTGCTAAACGTCGATG (32-mer; NdeI site indicated)

574: 5’-AAAAAAATGCATATGGCAGCAGCAATTACAC (30-mer; NdeI site indicated)

0002: 5’-AAGGATCCATGTCTGAACCACGTTTCG (27-mer, BamHI site indicated)

3.2.2 Plasmids for protein expression

The plasmid pNH1543 (coding for the linked-construct ε_{200a270P}) was made by first PCR amplifying the linker-α_{270} fragment from the plasmid pKO1479 (coding for ε_{209a270P}; Ozawa e al., 2013) with primers 564 and 566 and the PCR amplification of the ε_{210-243} C-terminal fragment from the plasmid pSH1017 (containing the wild-type dnaQ gene, coding for the full-length ε subunit; Scheuermann et al., 1983) with primers 565 and 568. The resulting PCR products were mixed together and re-amplified using primers 564A and 568 to generate a single product flanked by NdeI and EcoRI restriction sites. Following cleavage with appropriate enzymes, the fragment was
inserted between corresponding sites in pETMCSI (Neylon et al., 2000; Section 2.1.2). The plasmids pNH1544, pNH1545 and pNH1546 (coding for the linked constructs \( \varepsilon^{189}\alpha^{270P}, \varepsilon^{180}\alpha^{270P} \) and \( \varepsilon\alpha^{270P} \), respectively) were similarly generated by amplification of the plasmid pSH1017 from the appropriate site in the \( dnaQ \) gene. On each occasion the primer 568 was replaced with either primer 570, 572 or 574 to generate the three successively longer constructs required.

DNA sequencing of the plasmid pKO1479 (used to generate the linker–\( \alpha^{270} \) fragment) had revealed a single-point mutation corresponding to a leucine to proline change at a.a. position 21 in the \( \alpha^{270} \) segment. Consequently, the four extended constructs described above also contained this mutation. This error was corrected in the plasmid pKO1479 by amplification of the wild-type \( \alpha^{270} \) segment from plasmid pKO1538 with primers 0002 and 564A to generate a PCR product flanked by \( BamHI \) and \( EcoRI \) restriction sites. Following cleavage with these enzymes, the fragment was reinserted into the plasmid pKO1479 in place of the mutated \( \alpha^{270P} \) segment to give the plasmid pNH1560.

The plasmid pKO1539 (coding for the construct \( \alpha^{270}\varepsilon^{209} \)) had been prepared previously (Ozawa et al. 2013) and was kindly made available for use by Dr Kiyoshi Ozawa.

### 3.2.3 Protein purification buffers

Buffers used were as follows: lysis buffer (50 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM DTT, 20 mM spermidine, 0.5 mM PMSF, 10 % (w/v) sucrose); resuspension buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 20 % glycerol, 160 mM NaCl; filtered); chromatography buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 15 % glycerol; filtered); crystallography buffer (10 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT,
100 mM NaCl; filtered); storage buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 20 % glycerol, 100 mM NaCl; filtered).

Chromatography buffers were supplemented with NaCl as required in concentrations noted within the purification protocols to follow.

3.2.4 Purification of $\varepsilon:\alpha_{270}$ variants

Linked $\varepsilon:\alpha_{270}$ protein variants were overproduced and purified using a comparable protocol to that initially devised for the production and purification of $\varepsilon_{209}\alpha_{270P}$ (N. Horan, Honours Thesis, 2010). A common methodology will be described here only once and any relevant modifications noted within. A summary of each purification procedure including yield, purity, molecular weight characterisation etc. can be found in Table 3.1 and Figure 3.3.

Phage T7 promotor plasmids pKO1539, pNH1560, pNH1543, pNH1544, pNH1545 and pNH1546 (coding for protein constructs $\alpha_{270}\varepsilon_{209}$, $\varepsilon_{209}\alpha_{270}$, $\varepsilon_{200}\alpha_{270P}$, $\varepsilon_{189}\alpha_{270P}$ and $\varepsilon\alpha_{270P}$, respectively) were individually transformed into chemically competent E. coli BL21/pLysS cells and single colonies of each were isolated and propagated. Small-volume cultures of each strain were grown overnight in LBTA/C medium at 37°C and used to inoculate large volume overproduction cultures (10 mL added per 1000 mL of medium).

The E. coli strain BL21/pLysS/pKO1539 (coding for the protein $\alpha_{270}\varepsilon_{209}$) was grown in 6 × 1 L of LBTA/C medium at 37°C to an OD$_{600}$ of 0.5. To induce overproduction, IPTG was added to 0.5 mM and the cultures were shaken for a further 4 h at 37°C. Cells were harvested by centrifugation (11000 × g; 7 min), frozen in liquid nitrogen and stored at -80°C. To improve yield, the E. coli strains BL21/pLysS/pNH1560, /pNH1543, /pNH1544, /pNH1545 and /pNH1546 (coding for proteins $\varepsilon_{209}\alpha_{270}$, $\varepsilon_{200}\alpha_{270P}$, $\varepsilon_{189}\alpha_{270P}$ and
The εα270P, respectively) were grown in 1L each of LB autoinduction medium (Section 2.1.4.3) containing ampicillin (100 mg/L) and chloramphenicol (35 mg/L). Cultures were grown at 27°C for 48 h to final OD₆₀₀ readings of 8–9, with the exception of the culture producing the εα270 variant, which obtained a final OD₆₀₀ reading of 17.5. Cells were harvested by centrifugation (17,000 × g; 7 min), frozen in liquid nitrogen and stored at −80°C. In small volume culture trials, the protein construct εα270P was found to be wholly insoluble when expressed in BL21/pLysS at 37°C and 70–80 % insoluble when expressed at 22°C. It was decided not to pursue further expression or purification trials with this construct (discussed further in Section 3.4.1).

Each of the 5 soluble variants was purified separately, using the following protocol performed at 6–8°C throughout: Frozen cells (10–30 g) were thawed and resuspended in lysis buffer (10 mL per gram of cell pellet) before being passed twice through a French press (12,000 psi) to lyse cells. The lysate was clarified by centrifugation (17,000 × g; 40 min) to yield the soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.23 g/mL) and stirring for 60 min were collected by centrifugation (38,000 × g; 60 min) and dissolved in resuspension buffer (30 mL). The solution was dialysed against 1 L of the same buffer overnight, to yield Fraction II.

Fraction II was applied by gravity (at ~1 mL/min) onto a column (2.5 × 16 cm) of Toyopearl DEAE-650M resin that had been equilibrated with chromatography buffer + 160 mM NaCl. Fractions containing proteins that did not bind to the column were pooled and dialysed against 2 L of chromatography buffer containing no NaCl to yield Fraction III (80–100 mL). Fraction III was applied by gravity to the same column, now equilibrated in chromatography buffer. After the column had been washed in 80 mL of chromatography buffer + 20 mM NaCl, the bound proteins were eluted using a linear gradient (320 mL) of 20–200 mM NaCl in chromatography buffer. On
each occasion, the target protein eluted as a broad peak and fractions of interest were assessed for purity by SDS-PAGE. In the case of Fraction III containing the protein ε\textsubscript{209}α\textsubscript{270}, the total was divided into 3 x 30 mL samples that were loaded separately in three purification runs to avoid exceeding the binding capacity of the available DEAE resin. Selected fractions of purified protein were pooled to yield Fraction IV (final volume of 90–200 mL).

Fraction IV was diluted 2–3 fold with chromatography buffer to give a final NaCl concentration of ~60 mM NaCl and divided into smaller aliquots, each containing approximately 40–50 mg of protein (3–7 aliquots, depending on estimated protein amount). Significantly less of the α\textsubscript{270}ε\textsubscript{209} protein was present (see Table 3.1) and, as such it was divided into only two volumes. Each volume was loaded using a 50 mL superloop connected to an ÄKTA system onto an 8 mL 10/100 GL monoQ column (GE Lifesciences) that had been pre-equilibrated in chromatography buffer. After the column had been washed with 40 mL of the same buffer, bound proteins were eluted using a linear gradient (90 mL) of 0–400 mM NaCl in chromatography buffer. Target proteins generally eluted in two separate peaks, with a main peak at 190 mM and a smaller, yet significant, additional peak at 100 mM NaCl. The α\textsubscript{270}ε\textsubscript{209} ‘reverse’ protein construct eluted less cleanly as a number of conflated smaller peaks over the range 130–170 mM NaCl. Protein purification characteristics are discussed further in Section 3.4.1.

Fractions containing highly purified target protein were pooled (Figure 3.3), divided where appropriate and dialysed into 2 changes of 1 L crystallography and/or storage buffers. Prior to dialysis, samples for crystallography were concentrated by precipitation with ammonium sulphate (0.35 g/mL), resuspended in crystallography buffer and later adjusted where needed to achieve the desired concentrations. Aliquots were frozen in liquid nitrogen and stored at ~80°C. Final yields and mass spectrometric analysis of each protein can be found in Table 3.1.
Table 3.1: Final purified protein yields and mass spectrometric (MS) measured molecular weights (MW) in formic acid [0.1 % (v/v)] of linked ε/α270 constructs. Storage and crystallography buffer conditions are as detailed in Section 3.2.3.

<table>
<thead>
<tr>
<th>Protein Construct</th>
<th>α270ε209</th>
<th>ε209α270</th>
<th>ε200α270P</th>
<th>ε185ε270P</th>
<th>ε189ε270P</th>
<th>ε189α270P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell weight harvested</td>
<td>10 g</td>
<td>30 g</td>
<td>14 g</td>
<td>16 g</td>
<td>13 g</td>
<td>Nil</td>
</tr>
<tr>
<td>Purified protein yield</td>
<td>4 mg</td>
<td>240 mg</td>
<td>150 mg</td>
<td>75 mg</td>
<td>145 mg</td>
<td>Insoluble</td>
</tr>
<tr>
<td>Storage buffer</td>
<td>0.8 mg/mL (2.5 mL)</td>
<td>5.8 mg/mL (45 mL)</td>
<td>15.2 mg/mL (8 mL)</td>
<td>12.2 mg/mL (4 mL)</td>
<td>14.5 mg/mL (8 mL)</td>
<td>Nil</td>
</tr>
<tr>
<td>Crystallography buffer</td>
<td>8.5 mg/mL (150 μL)</td>
<td>9.1 mg/mL (5 mL)</td>
<td>10 mg/mL (3 mL)</td>
<td>10 mg/mL (3 mL)</td>
<td>10 mg/mL (3 mL)</td>
<td>Nil</td>
</tr>
<tr>
<td>MS measured MW (Da)</td>
<td>34 338.3</td>
<td>34 476.0</td>
<td>35 521.5</td>
<td>36 769.5</td>
<td>37 621.7</td>
<td>n/a</td>
</tr>
<tr>
<td>Theoretical MW (Da)</td>
<td>34 340.9</td>
<td>34 472.1 (-Ile)</td>
<td>35 522.3 (-Met)</td>
<td>36 768.8 (-Met)</td>
<td>37 619.3 (-Met)</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Figure 3.3: SDS-PAGE analysis of purified linked ε/α270 variants. Protein samples were standardised to 10 μg per lane and are arranged in increasing molecular weight from left to right as follows: (I) ε270ε209 (II) ε209ε270 (III) ε200ε270P (IV) ε185ε270P (V) ε189ε270P. A reference protein standard is presented in the far left lane (M). This gel was stained with Coomassie blue.
3.2.5 Growth of protein crystals

Diffraction quality crystals utilised to solve the previous structure of ε\textsubscript{209α270F} had been grown in the condition CX: [0.2 M MgCl\textsubscript{2}, 0.1 M Tris-HCl pH 8.0, 18% (w/v) PEG 3350, 3 mM TCEP-HCl at 4°C]. During the screening process it was noted that high-quality crystals also appeared in many similar conditions to the above (CX) in which the PEG3350 concentration and/or pH of Tris-HCl had been varied slightly. In pursuit of crystals of the additional five purified ε:α\textsubscript{270} protein variants, it was decided to screen each around a similar set of conditions.

Aliquots of purified protein samples stored in crystallography buffer were recovered and thawed on ice. Each protein was screened across 24 separate conditions using a grid formation in which the concentration of PEG3350 was varied along one axis [12, 14, 16 and 18 % (w/v)] and the pH of 0.1 M Tris-HCl varied along the other (pH 8.0, 8.2, 8.4, 8.6, 8.8 and 9.0). The concentrations of MgCl\textsubscript{2} and TCEP in condition CX remained the same in all screening conditions. Screening was performed using a sitting-drop configuration with 4.5 μL of protein sample mixed with an equal volume of reservoir solution. Crystal trays were placed at 4°C and monitored daily for crystal growth.
3.3 Results

3.3.1 Data collection and refinement

Crystals of $\varepsilon_{200}\alpha_{270}$P commonly appeared after 3–4 days and resembled rectangular prisms 300–400 $\mu$m in length (Figure 3.4 A). The observed morphology and arrangement of crystals was similar to crystals obtained of $\varepsilon_{209}\alpha_{270}$P (Figure 3.4 B) and both often grew as multiple ‘stacked’ layers of individual crystals. Typical crystals of $\varepsilon_{200}\alpha_{270}$P showed X-ray diffraction patterns due to complex arrangements of individual lattices (Figures 3.4 C and D) and only a small fraction were suitable for X-ray diffraction. Within the conditions of the screen, superior crystal growth generally occurred in higher PEG 3350 concentrations [16 and 18 % (w/v)] and lower pH conditions (Tris-HCl pH 8.0–8.4).

Protein constructs $\varepsilon_{189}\alpha_{270}$P, $\varepsilon_{180}\alpha_{270}$P were occasionally seen to form a microcrystalline ‘shower’ under some conditions within the screen but remained soluble under the majority of conditions. The protein constructs $\varepsilon_{209}\alpha_{270}$ and $\alpha_{270}\varepsilon_{209}$ remained soluble in all conditions tested. As a result of structural information gathered for the $\varepsilon_{200}\alpha_{270}$P protein, no further attempts to optimise crystal growth conditions for these proteins were made (discussed in Section 3.4.1).
Suitable crystals of ε2200Ω270P were collected by Dr Nan Li by two transfers into a cryoprotectant solution of reservoir solution supplemented with 15 % (w/v) PEG400 before being flash cooled at 100 K for data collection. X-ray data were kindly collected by Dr David Jacques at the Australian Synchrotron Facility on Beamline MX2 at wavelength 0.95369 Å using an exposure time of 1 sec per frame over a total of 155° at 0.25° per frame.

Data refinement and model building were performed by Dr Nan Li and Associate Professor Aaron Oakley. The structure of ε200Ω270P was solved at 2.15 Å resolution by molecular replacement, using the refined structure of ε200Ω270P at 1.7 Å resolution as a starting model. Final models were obtained following cycles of refinement using REFMAC (Murshudov et al., 1996) and manual building using COOT (Emsley et al., 2010). Data collection and refinement statistics are given in Table 3.2.
Table 3.2: X-ray data collection and refinement statistics for $E_{20037}$IP (molecular replacement). Table reproduced from Ozawa et al. 2013.

<table>
<thead>
<tr>
<th>PDB ID:</th>
<th>4GX9</th>
</tr>
</thead>
</table>

**Data collection**

<table>
<thead>
<tr>
<th>Space group</th>
<th>$P2_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell dimensions</td>
<td></td>
</tr>
</tbody>
</table>
  $a, b, c$ (Å) | 83.02, 56.98, 35.07 |
  $\alpha, \beta, \gamma$ (') | 90.0, 93.78, 90.0 |
| Resolution (Å) | 3.5–2.15 (2.19–2.15)¹ |
| $R_{\text{sym}}$ or $R_{\text{merge}}$ | 9.5 (49.4) |
| CC1/2 | 0.837 (2.19–2.15) |
| $R_{\text{free}}$ | 0.337 (2.19–2.15) |
| $I / \sigma(I)$ | 10.4 (1.6) |
| Completeness (%) | 97.2 (96.5) |
| Redundancy | 3.1 (2.9) |

**Refinement**

| Resolution (Å) | 35–2.15 |
| No. reflections | 67,126 |
| No. reflections in highest res. bin | 3361 |
| $R_{\text{work}} / R_{\text{free}}$ | 22.6 / 29.1 |
| $R_{\text{work}} / R_{\text{free}}$ highest res. bin | 27.8 / 34.3 |
| No. atoms |  
  Protein | 9801 |
  Ligand/ion | 0 |
  Water | 277 |
| $B$-factor |  
  Protein | 21.9 |
  Ligand/ion |  
  Water | 21.8 |
| R.M.S. deviations |  
  Bond lengths (Å) | 0.019 |
  Bond angles (°) | 1.75 |

**Ramachandran plot**

| residues in: core region | 89.2% |
| additional allowed region | 10 / 4% |
| generously allowed region | 0 / 3% |
| disallowed region | 0% |

¹ Values in parentheses are for highest-resolution shell.
3.3.2 An atomic structure of the α–ε interaction

The ε209α270P protein crystallised with space-group symmetry P2₁, with four molecules per unit cell (Figure 3.6 A). Consistent with the structure of ε209α270P, all four monomers within the unit cell displayed continuous electron density from residue Lys211 of the ε chain onwards through the εCTS (Figure 3.5), linker region and α270 domain. Of these, two monomers (designated Monomers 2 and 3 in the PDB file) exhibited additional weak electron density interpreted as the tetrapeptide segment Ile202–Ile205 in Monomer 3 and the tripeptide segment Ile202–Arg204 in Monomer 2. In both Monomers 2 and 3 the residue Ser210 was also fully ordered.

The refined structure of Monomer 3 (displayed in Figures 3.6 B, C and D) shows the εCTS forming an extended or random coil conformation across one face of the α270 PHP domain, followed by a perpendicularly orientated α-helix extending from ε residues Thr218 to Gly237 and an additional region of random coil from ε residues Ser238 to Ala243 (Figure 3.6 C). The 9 a.a. interdomain linker is fully ordered between εAla243–αMet1 and solvent exposed, appearing unconstrained and able to span the distance from the ε C-terminus to the α N-terminus. It is possible, however, that some constraints may be placed on C-terminal ε residues following the α-helix, particularly Arg242 and Ala243, which are seen not to interact with the α270 domain in either structure and may be disrupted from their native binding site/s. The Leu to Pro mutation at position 21 of the α chain is positioned far from the sites of εCTS interaction with α270 (Figure 3.6 B) and is unlikely to have any significant effect on protein interaction. The αPro21 residue in all four monomers makes contact with an alanine residue in a neighbouring unit-cell, which most likely promotes crystal formation through stabilisation during crystal packing. The linked protein construct with this mutation
corrected (ε200α270) was not readily crystallisable under the conditions of the screen.

![Electron density map](image)

**Figure 3.5: Electron density map of the εCTS (green) and linker (yellow) domains of the ε200α270P protein, calculated using molecular replacement and contoured at 1 σ. The protein is observed using a stick representation.**

### 3.3.3 The Extended Construct Reveals Additional Sites of Interaction

In supplement to the structure of ε200α270P, the structure of ε200α270P allowed ε residue Ser210 to be positioned with high confidence, while ε residues Ile202–Ile205 were built into additional weak density visible in monomer 3. While it is difficult to conclusively state that the short, isolated tetrapeptide segment is part of the same peptide chain as the remaining εCTS residues and not part of a neighbouring molecule forming an intermolecular contact, it is almost certain that this area of contact between the two domain displays the path taken by the εCTS as it navigates towards the permanently flexible linking region. The distance between the α-carbon atoms of residues Ile205 and Ser210 (12.1 Å) is consistent with the corresponding distance of the fully structured region between Lys211 and Phe216 (14.8 Å), demonstrating the region of poor electron density from residues Val206–Ala209 is comparable to the length of an adjacent random-coil tetrapeptide segment and that the missing segment in the structure of ε200α270P is consistent with the dimensions of the expected gap.
Hydrogen-bonded and electrostatic interactions between the εCTS and α270 domains for the linked constructs ε200α270P and ε209α270P are summarised in Table 3.3 and illustrated in Figure 3.7.

Figure 3.6: 2.15 Å crystal structure of the ε200α270P fusion protein. (A) Cartoon depiction showing a crystallographic unit cell of four monomers (coloured orange, purple, red and turquoise). (B) An individual monomer presented using a cartoon representation. The α270 domain (green) is connected to the εCTS domain (blue) by a 9 a.a linker (red). The location of the PCR generated L→P mutation at position 21 of the α chain is highlighted using magenta spheres. (C) A monomer viewed from the same orientation as (B), now displaying the α270 domain using a space-filling representation and coloured grey. The first ordered residue of the structure (εlle202), as well as the first (εThr218) and last (εGly237) α-helix forming residues are labelled. (D) Rotated view of (C) with side-chains of interacting residues from εCTS displayed in orange. Interacting residues of the random-coil region prior to the α-helix are annotated. A blue dashed line represents an approximate prospective pathway for the peptide backbone of ε residues Val206–Ala209.
**Table 3.3**: Hydrogen-bonded and electrostatic contacts between εCTS and α270 in the crystal structures of constructs ε200α270P and ε209α270P (right column only; PDB 4GX8). Table reproduced from Ozawa et al. 2013.

<table>
<thead>
<tr>
<th>ε200α270P only</th>
<th>ε209α270P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ε residue</td>
<td>α residue</td>
</tr>
<tr>
<td>ε Gln203</td>
<td>α Glu179</td>
</tr>
<tr>
<td></td>
<td>α His183</td>
</tr>
<tr>
<td></td>
<td>α Arg172</td>
</tr>
<tr>
<td>ε Arg204</td>
<td>α Asp252</td>
</tr>
<tr>
<td></td>
<td>α Leu249</td>
</tr>
<tr>
<td></td>
<td>α Glu248</td>
</tr>
<tr>
<td>ε Ile205</td>
<td>α Asp252</td>
</tr>
<tr>
<td>ε Ser210</td>
<td>α Glu255</td>
</tr>
<tr>
<td>ε Arg213</td>
<td>α Asp164</td>
</tr>
<tr>
<td>ε Val215</td>
<td>α Asn259</td>
</tr>
<tr>
<td>ε Ala217</td>
<td>α Glu262</td>
</tr>
<tr>
<td>ε Glu221</td>
<td>α Arg266</td>
</tr>
<tr>
<td>ε His225</td>
<td>α Lys63</td>
</tr>
<tr>
<td>ε Arg228</td>
<td>α Val269</td>
</tr>
<tr>
<td></td>
<td>α His58</td>
</tr>
<tr>
<td>ε Leu240</td>
<td>α Gly61</td>
</tr>
<tr>
<td>ε Trp241</td>
<td>α Glu3</td>
</tr>
</tbody>
</table>

**Figure 3.7**: Map of a.a. interactions between εCTS and α270 from the crystal structure of ε200α270P. The protein is displayed using a cartoon representation with the α270 domain coloured green, the εCTS coloured blue and the linker presented in red. Interacting a.a. residues are highlighted in magenta in the α270 domain and orange in the εCT.
3.3.4 Positioning the εCTS within the Context of the α Subunit

The structure of ε200a270P was superimposed onto the crystal structure of the α subunit (residues 1–917; Lamers et al, 2006; PDB 2HQA) by aligning the near-homologous structures of the α PHP domain from both molecules. A composite model of the two proteins shows a site of ε interaction that is far removed from where the catalytic N-terminal domain of the ε subunit might be expected to reside, between the β2 sliding clamp and the active centre of the polymerase. The circuitous route taken by the εCTS wraps almost entirely around one side of the α PHP domain towards the site of α-helical attachment.

Figure 3.8: Location of ε interaction with the α subunit as viewed from different perspectives (A) and (B): The ε C-terminal domain (a.a. residues 202–243) is shown in blue using a cartoon representation. The α subunit (a.a. residues 1–917) is displayed using a space-filling depiction with the PHP domain (a.a. residues 1–270) coloured green and the remaining residues coloured light orange. The a.a. residue Ile202 of the ε chain has been indicated in both (A) and (B) for reference and the inter-domain linker has been removed for clarity. A prospective position for the ε N-terminal catalytic domain is shown in blue and the approximate locations of β-interacting regions of α and ε are shown in purple. Illustrations were created by structural alignment of the ε200a270P protein with a previously published structure of the E. coli α subunit (a.a. residues 1–917; Lamers et al, 2006; PDB 2HQA) using PyMOL molecular graphics visualisation software (Schrödinger).
3.3.5 Inversing the order of linked domains induces the formation of domain-swapped multimers

The protein constructs $\varepsilon_{209}\alpha_{270}$ and $\alpha_{270}\varepsilon_{209}$ were assessed for the formation of potential domain-swapped multimers using mass spectrometry under native conditions (Figure 3.9). Both samples showed the presence of a predominant monomer species, with smaller proportions of dimer also detectable in each. Under the conditions tested, the $\alpha_{270}\varepsilon_{209}$ protein dimer was present at approximately twice the abundance relative to monomer to that of the $\varepsilon_{209}\alpha_{270}$ protein. The sample of $\alpha_{270}\varepsilon_{209}$ also showed the presence of detectable levels of a trimeric species that were not observed in any significant level in the sample of $\varepsilon_{209}\alpha_{270}$. 
Figure 3.9: Positive ion nanoESI mass spectrum under native conditions of (A) ε_{209}α_{270} and (B) α_{270}ε_{209}. Proteins (20 μM) were characterised following dialysis against 500 mM NH₄OAc, 1 mM β-mercaptoethanol, pH 7.6. Peak envelopes corresponding to the molecular masses of monomers (34.5 or 34.3 kDa, respectively), dimers (69.0 or 68.7 kDa, respectively) and trimer (103.0 kDa, for (B) only) conformations are bracketed in purple and respective charge states are indicated above each peak. Simplified cartoon schematics of monomer and probable domain-swap multimer formations are presented above each envelope.
3.4 Discussion

3.4.1 Protein purification and crystallography trials

The linking of the otherwise insoluble εCTS to the α270 domain with a short flexible peptide was a pleasingly successful strategy to obtain well-behaved protein samples. The shorter (and crystallisable) versions of this construct, ε209α270P and ε200α270P, were relatively easily produced and purified in high yields. Two further extended versions of the linked construct, ε189α270P and ε180α270P, were also purified in high yields. However, some unusual behaviour was observed when performing a final ion-exchange column purification step of these proteins using Mono-Q resin. Both proteins consistently eluted from the resin in two distinct peaks at different salt concentrations (100 mM and 190 mM NaCl; data not shown), which may be explained by the presence of a highly mobile and unbound region of linker peptide causing unstable or variable interaction of the protein with the ion-exchange resin. Although this reasoning is speculative, such an explanation is consistent with NMR and crystallographic data that show these extended peptide regions very likely remain permanently flexible in the presence of the α270 domain. A highly mobile linker region interfering with crystal packing contacts is also the probable explanation for the failure of these two extended constructs to readily crystallise under the conditions explored here.

The shorter of the six constructs, ε209α270, was easily produced and purified with a high yield; however, with the absence of the leucine to proline mutation in the α270 chain, the construct was not readily crystallisable. Closer analysis of the two crystal structures containing the unintentional mutation, ε209α270P and ε200α270P, show this proline residue in each of the four monomers in a unit cell making a crystal contact with an alanine residue from a symmetry mate. While the corrected (unmutated) protein may have
crystallised under different conditions if optimised from a new set of sparse screens, it appears probable that the additional crystal contacts formed in the mutated protein were crucial in the formation of stable crystals suitable for collection of high-quality diffraction data. With the exception of the site of the leucine to proline mutation, the structures of both linked constructs that were crystallised showed the α270 domain structure was completely conserved when compared to the same domain from the structure of α (residues 1–917; Lamers et al., 2006). These data, together with a distant location of the sites of domain interaction from the site of mutation, indicated that further pursuit of an unmutated crystal structure would be unnecessary.

The full-length wild-type ε protein is largely insoluble when overexpressed in vivo and is purified using a refolding protocol (Scheuermann and Echols, 1984; described in Section 4.2.3.5). It was therefore unsurprising to find that the construct εα270P was also mostly insoluble, even when produced by autoinduction at a low temperature (22°C). With the understanding that this linked protein would comprise two relatively large globular domains (the ε N-terminal catalytic domain and the α N-terminal PHP domain) separated by a highly flexible interdomain peptide, it was reasoned that investigating such a refolding strategy to produce sufficient protein for crystallographic experiments would be challenging. A lack of detectable interaction between these two domains (discussed further in Section 4.3.2) likely limits the additional information that could be gained from a crystal structure of this construct, were it to become available.

In contrast to the behaviour of linked constructs, in which the εCTS was placed N-terminal in sequence in relation to the α270 domain, the ‘reversed’ construct with the domains located in an opposite order (α270ε209) was much more difficult to produce and purify and significantly reduced levels of protein production and unstable behaviour during ion exchange.
chromatography were observed. These characteristics would be consistent with a flexible construct where the εCTS remains unbound from the α270 domain and unable to reach the correct binding site as a consequence of linker placement.

3.4.2 The extent of the α-ε interaction can be confidently defined

The interaction between the E. coli α polymerase and the ε proofreading exonuclease is particularly tight and stable, with a published $K_D = 5\,\text{nM}$ (Wieczorek and McHenry, 2006) and possibly as low as 40 pM (S. Jergic, unpublished work). It is likely that, once formed within the cell, the two proteins never dissociate during the replication cycle. The key ε residues involved in this interaction are those C-terminal to, and likely including εSer210, with crystal structure and NMR data (Ozawa et al., 2008; Ozawa et al., 2013) indicating this region of interaction with the α PHP domain is permanent once formed. In their 2006 paper, Wieczorek and McHenry identified that a construct of the first 255 residues in the α chain was insufficient to display any measurable interaction with the ε subunit. This correlates well with the crystal structures presented here, which show that α residues Asn259, Glu262, Arg266 and Val269 form critical contacts with five of the eight ε residues from εArg213 onwards seen to interact with the α subunit, including two helix forming residues (summarised previously in Table 3.3 and displayed in Figure 3.7).

The ε C-terminal region immediately preceding these residues, however, appears to retain some mobility when in complex with α. The weak electron density observed for residues Ile202–Arg204 in the crystal structure of ε200(α270P is what may be expected for a peptide region with transient interactions with α. A lack of appreciable electron density from where ε residues Ile205–Ala209 is presumed to reside similarly indicates a region of contact with α that is likely transient. However, it is improbable that this
short pentapeptide region remains permanently flexible when in complex with α as it is immediately bounded by two regions that participate in significant interaction with the α subunit.

The linker residues of ε from Gln182–Phe187 have been shown to interact with a hydrophobic pocket on the β2 sliding clamp (Jergic et. al., 2013). When part of the β2αεθ core-clamp assembly, the permanently flexible interdomain region of the ε subunit has been shown here to span the ε residues Ala188–Ala200, at least (Ozawa et al., 2013). It is likely that ε residues Thr201–Ala209 comprise a region of transient interaction with α, providing a comparatively small contribution to the binding energy of the overall ε-α interaction. It is conceivable that this region provides the proofreading subunit with some reserve flexible length to employ when the β2αεθ shifts between polymerisation and proofreading modes, allowing for significant conformational adjustments to occur.

3.4.3 The α270ε209 ‘reverse’ construct

The crystal structures of ε209α270P and ε200α270P show extensive contacts between the εCTS and the α PHP domains, all of which make sense in the context of the larger β2αεθ complex and correlate well with previously predicted regions of interaction. The introduction of an artificial linker between the two domains, however, had the potential to disrupt their native site of interaction by constraining the two domains and limiting the availability of residues. It was hypothesised that a construct with the domains reversed in sequential order (α270ε209) may ideally provide a crystal structure of a domain-swapped dimer (or multimer), displaying the same site of interaction as observed in the structures of ε209α270P and ε200α270P and giving a strong indication that the initial structural models were correct. Diffraction quality crystals of α270ε209 were unable to be grown under similar screening conditions to that required for crystal growth of the ε209α270P and
proteins. This may be explained by difficulties in forming a stable domain-swapped crystal or due to the absence of the leucine to proline mutation critical for the crystal growth in the structures determined, or both. In place of constructing and purifying a ‘reverse’ construct with the leucine to proline incorporated for crystallisation trials, a simple comparative analysis between the ε209α270 and α270ε209 proteins was performed using mass spectrometry under native conditions to assess the relative levels of monomers and multimers in each sample. These results showed the α270ε209 ‘reverse’ protein construct has a propensity to form multimers more readily than the ε209α270 protein under the native conditions tested here. While the difference in multimer ratios observed between the two proteins is not especially significant (approximately double), this observation is consistent with what may be expected for linked constructs of the εCTS and α PHP domains that are more able to form an interprotein contact in one domain arrangement than another.

3.4.4 A crystal structure allows for informed model building

The structures of ε209α270 and ε200α270 provide a convincing structural model for the pathway taken by the εCTS across the α270 subunit, with the extended construct displaying this interaction with added clarity. Accurately defining the extent of this structural interaction is of fundamental value to model building efforts of the β2αεθ complex and for generating new hypotheses concerning the functional dynamics of the ε proofreading domain during error correction and processive DNA synthesis. The following experimental Chapter within this Thesis will extend on this notion and utilise the data presented in the current chapter to inform the building and validation of a β2αεθ structural model.
Chapter 4

Modelling the structure of the

\textit{E. coli} $\beta_2\alpha\epsilon\theta$ replication complex
4.1 Introduction

The *E. coli* Pol III HE β<sub>2</sub>εθ complex is comprised of the α polymerase, the ε exonuclease, the small θ accessory subunit and the ring-shaped β<sub>2</sub> dimer that encircles primer-template DNA. The α and θ subunits both bind tightly to the ε subunit (low nM range dissociation constants) to form what is termed the Pol III core. Both the α and ε subunits of the Pol III core contain clamp-binding motifs (CBMs) that interact comparatively weakly with the β<sub>2</sub> subunit (high μM range dissociation constants) through two equivalent hydrophobic binding pockets located within each β monomer of the dimeric clamp (Figure 4.1).

![Image](image.png)

**Figure 4.1:** A structural model of *E. coli* β binding to α in the (A) ‘closed’ DNA-bound form modelled on the structure of the Taq α with DNA and dNTP (Wing et al., 2008) and (B) ‘open’ conformation of *E. coli* β binding to α without DNA (based on Bailey et al., 2006). In both (A) and (B) the εNTD:θ complex can be accommodated between the α PHP domain (orange) and the hydrophobic binding pocket of β (yellow). The location of Leu888 in Taq α (Val832 in *E. coli*) within the β-binding domain of α (light orange) in also annotated. Strengthening of the CBMs in α and ε (to form the α<sub>L</sub> and ε<sub>L</sub> mutant subunits) stabilises the interaction between the Pol III core and the β-clamp. Structural models were reproduced from Jergic et al., 2013.

High resolution (<4 Å) structures of all component subunits of β<sub>2</sub>εθ are known in whole or in part from different bacterial species. However, a complete structure of β<sub>2</sub>εθ or αεθ from any organism remains elusive.
Diffraction quality crystal growth of the αεθ core is likely impeded by the long and flexible inter-domain linking peptide of the ε subunit (discussed in Chapter 3). In addition, the weak native interactions between the β₂ subunit and the CBMs of the α and ε subunits would also be expected to preclude crystal growth of the larger β₂αεθ complex.

Stronger CBMs have been identified in nature in some of the variety of replication-associated proteins that interact with the β-clamp (Dalrymple et al., 2001; Wijffels et al., 2004 and 2011). Previous work performed by Dixon and colleagues and others has substituted wild-type CBMs found in both the α and ε subunits with known tighter-binding peptide motifs. The wild-type pentameric α CBM (QADMF) was substituted for the sequence QLDLF to create the construct designated here as αL (Dohrmann and McHenry, 2005). The wild-type hexameric ε CBM (QTSMAF) was substituted for the sequence QLSLPL, the strong CBM from the replication initiation regulatory protein Hda (Kurz et al., 2004), to create the construct termed εL (Jergic et al., 2013). The αεθ core is routinely co-purified in the Dixon Lab and is sufficiently stable to be isolated as an intact complex by ion-exchange chromatography. The weak native interaction of the β-clamp with the Pol III core has, so far, presumably prevented co-purification of the β₂αεθ complex. It was anticipated that the incorporation of stronger-binding CBMs into the α and ε subunits would stabilise the Pol III core:clamp complex sufficiently for purification by ion-exchange chromatography. A stable and pure sample of β₂α₁ε₁θ would provide the opportunity to pursue further structural investigation of this complex, both in isolation and also with the potential addition of a template DNA molecule.
4.1.1 Core:clamp conformational changes during replication

The Pol III $\beta_2\alpha\epsilon\theta$ complex is thought to engage in two main conformational arrangements on a template DNA, supporting both a stable and efficient polymerisation mode, where the $\alpha$ subunit is operative, and also a proofreading mode, where the $\epsilon$ subunit is positioned to excise incorrectly inserted nucleotides. The precise spatial arrangement of subunits during each mode is uncertain, as is the sequence of molecular events that stimulates the structural transition between these two (and probably other) conformational states.

A number of proteins, including all known repair polymerases, interact with a hydrophobic binding pocket in the $\beta$ subunit. The dimeric $\beta$-clamp contains two such sites, which is presumed to allow for exchange of protein subunits at one site while the other maintains contact with either the $\alpha$ or $\epsilon$ subunit, thus keeping the $\beta_2\alpha\epsilon\theta$ complex intact during bypass of a lesion in the template DNA (López de Saro et al., 2003; Indiani et al., 2005). It is less apparent what role these sites play during transition between polymerisation and proofreading modes. The substantial length of the flexible inter-domain region in the $\epsilon$ subunit may be sufficient to absorb a structural rearrangement that disengages the $\alpha$ subunit from DNA and moves the $\epsilon$ subunit into place to perform its proofreading function without the need to break the $\epsilon$-$\beta$ interaction. Alternatively, the $\epsilon$NTD may need to detach from $\beta_2$ to move toward the DNA template, rearranging the $\beta_2\alpha\epsilon\theta$ complex from, what is termed here, a ‘closed’ to an ‘open’ configuration (Figure 4.1).

A comparison of the crystal structures of Taq $\alpha$, with and without a primer-template DNA (Bailey et al., 2006; Wing et al., 2008), shows that the polymerase has an open structure that closes in the presence of DNA. It has also been shown that a suppressor mutation in the $dnaE$ gene corresponding
to a Val to Gly change at a.a. position 832 in the peptide sequence of E. coli α (αV832G) is able to partially compensate for the absence of ε during replication and improve the stability of the complex, presumably by promoting polymerase closing in the absence of the ε-β interaction (Lancy et al., 1989; Lifscis et al., 1992; Slater et al., 1994; Jergic et al., 2013). Closure of the β2αεθ complex into a more compact arrangement is likely stimulated by both the presence of a template DNA and also by the contribution of the ε-β interaction and is mediated, at least in part, at a location in the domain of α that contains V832.

4.1.2 Generating a structural model of β2αεθ

Extensive biochemical research has identified the principal sites of interaction between subunits of the β2αεθ complex and a number of crystallographic and NMR structures have been reported of individual components. A crystal structure of the full-length α subunit from Thermus aquaticus (Bailey et al., 2006) and a C-terminally truncated version (residues 2–917 of 1160) of E. coli α (Lamers et al., 2006) have been solved. Of the two potential CBMs that have been identified within the α subunit, only the internal site (Chapter 1; Figure 1.7) has been shown to bind one of two hydrophobic binding pockets of the β-clamp during processive replication (Dohrmann and McHenry 2005). The E. coli β2 subunit is particularly amenable to crystallisation and high-resolution crystal structures exist of the β-clamp in isolation (Kong et al., 1992; Oakley et al., 2003), bound to many different CBM peptides (Wolff et al., 2011; Yin et al., 2013) and also with a template dsDNA molecule (Georgescu et al., 2008). A crystal structure of the N-terminal catalytic domain of the ε subunit (εNTD; residues 2–186) has been solved alone (Hamdan et al., 2002) and also in complex with HOT, the
phage P1 homolog of the *E. coli* θ subunit (DeRose *et al.*, 2004; Kirby *et al.*, 2006). In addition, a complex of the ε NTD and θ subunit has been determined by NMR (Keniry *et al.*, 2006; Pintacuda *et al.*, 2006), displaying the site of interaction between ε and θ. A crystal structure revealing the site of interaction between the ε C-terminal segment (εCTS) and the α subunit is presented here in Chapter 3 and has also been published (Ozawa *et al.*, 2013). The long and flexible inter-domain Q-linker peptide of the ε subunit (also described in Chapter 3) contains a CBM that interacts with the alternate hydrophobic binding pocket of a single β2 dimer during processive DNA synthesis (Jergic *et al.*, 2013).

Steitz and colleagues have previously presented an informed model of the α and β2 subunits in complex with a template DNA molecule (Wing *et al.*, 2008), proposing that the secondary hydrophobic pocket on the β2 dimer may be occupied by the external CBM of the α subunit. The newly discovered interaction between the ε subunit and the β-clamp has instead positioned the ε CBM at this location when the Pol III HE is in a stable polymerisation mode. Together with a crystal structure showing the precise location of binding between the ε and α subunits (Chapter 3), this new information has allowed members of the Dixon group and collaborators to extend on the model presented by the Steitz group to attempt to accurately position the ε and θ subunits within the complex.

### 4.1.2 Refining the position of the ε N-terminal domain

Extensive site-specific Bpa (*p*-benzoyl-*L*-phenylalanine) photocrosslinking (Chin *et al.*, 2002) and NMR experiments have been previously performed by Dr. Kiyoshi Ozawa with the aim of defining the extent of the inter-domain flexible region of the ε subunit and probing for any potential additional weak interactions between subunits of β2αεθ (Ozawa *et al.*, 2008.
and 2013). These data indicated that the ε inter-domain Q-linker region (ε residues 188–208, at least) remains highly mobile even within the context of the β2αεθ complex. It was also shown that the ε\textsubscript{186} domain displays no detectable interaction with the α subunit. The location of the ε CBM anchors the globular εNTD:θ complex in close proximity to the β-clamp, raising the possibility of additional weak interactions between β\textsubscript{2} and the ε or θ subunits that may constrain the parameters of a β2αεθ model. This will be investigated here by NMR using isotopically labelled versions of the εNTD:θ complex to map chemical shift changes in the presence of unlabelled β\textsubscript{2}.

4.1.3 Verification of structural model by SAXS

The inclusion of two artificially strengthened CBMs (creating the α\textsubscript{L} and ε\textsubscript{L} subunit variants) was anticipated to stabilise the β2αεθ complex sufficiently for isolation by chromatography. Small-angle X-ray scattering (SAXS) measurement of a highly purified sample of the β2αLεLθ complex is described within this Chapter, providing low-resolution structural data to compare with structural models of β2αεθ. The addition of a dsDNA construct (based on an oligo design reported in the structure of Taq α and DNA; Wing et al., 2008), containing a 7 nt overhang and designed to trap the β2αLεLθ complex in the polymerisation mode when in the presence of dATP, ddCTP and MgCl\textsubscript{2} is also investigated here. Measurements by SAXS of β2αLεLθ:DNA could provide structural insight into the overall shape and dimensions of the core:clamp, both with and without DNA, assisting in model validation and refinement. In addition, a stable complex containing a DNA molecule may improve the chance of crystallisation of the whole β2αLεLθ:DNA complex.
4.1.4 Stabilisation of the complex for crystallography

A crystal structure of the core:clamp complex, with or without DNA, would be a very significant outcome. The most immediate challenge in pursuing this possibility is the production of a sufficient quantity of highly pure $\beta_2\alpha_L\epsilon_L\theta$ complex. Reconstituting and concentrating large protein complexes from individual components is a multi-step process that often results in low yields and thus requires relatively large initial amounts of purified subunits. Highly pure samples of the $\beta_2$, $\epsilon_L$ and $\theta$ subunits can all be produced in large quantities following well established protocols (see Sections 4.2.3.1, 4.2.3.6 and 4.2.3.7). The $\alpha_L$ subunit (or any other polymerase variant) is generally much more difficult to purify as it is susceptible to proteolysis from the C-terminus and the products of proteolysis are problematic to remove using existing ion-exchange protocols available. We investigate here the use of a novel protein-protein affinity purification technique that exploits the tight interaction ($K_D = 260$ pM; Jergic et al., 2007) between the 16 kDa C-terminal domain V of the E. coli Pol III HE $\tau$ subunit (here termed $\tau_{16}$) and the C-terminal $\tau$-binding domain of the $\alpha_L$ polymerase (or any mutant variant of $\alpha$ with preserved $\tau$-binding capacity). This approach was attractive as the interaction was anticipated to be highly specific and would also allow for any C-terminal proteolytic products from $\alpha_L$ to be removed, an important consideration when pursuing protein crystallisation.
4.2 Materials and Methods

4.2.2 Plasmids and strains required for protein expression

4.2.2.1 Nuclear magnetic resonance (NMR) experiments

The plasmids pCM869 and pND261 (coding for the wild-type *E. coli* θ and β subunits, respectively) were as described (Oakley *et al.*, 2003; Keniry *et al.*, 2006). The production of purification of the unlabelled θ subunit and the β	extsubscript{2} subunit are described in Section 4.2.3.6 and 4.2.3.7. The production of \textsuperscript{13}C,\textsuperscript{15}N double-labelled θ is described below in Section 4.2.3.2. The plasmid pSJ1447 (coding for the ε\textsubscript{193} protein) was constructed by Dr Kiyoshi Ozawa and Dr Slobodan Jergic. The production and purification of unlabelled and \textsuperscript{13}C,\textsuperscript{15}N-double-labelled ε\textsubscript{193} is outlined below in Sections 4.2.3.3 and 4.2.3.4. The co-purification of \textsuperscript{13}C,\textsuperscript{15}N-ε\textsubscript{193}:θ and \textsuperscript{13}C,\textsuperscript{15}N-0:ε\textsubscript{193} alternately labelled complexes required for NMR experiments is described here in Section 4.2.3.5.

4.2.2.1 Pol III core: β	extsubscript{2} production

The plasmids and strains required for the production of the θ, ε\textsubscript{L}, β	extsubscript{2}, τ\textsubscript{16} and α\textsubscript{L} subunits were as described (Oakley *et al.*, 2003; Keniry *et al.*, 2006; Jergic *et al.*, 2007; Jergic *et al.*, 2013). The production and purification of each is outlined below from Sections 4.2.3.1 and 4.2.3.6–4.2.3.8 and 4.2.4. The α\textsubscript{E612K} polymerase active site mutant was not required for any experimental work described in this Thesis, nevertheless, the production and purification of this subunit is described here in Section 4.2.3.8 as an example of a ‘traditional’ methodology for the purification of the polymerase. In contrast, a novel protein-protein affinity methodology was utilised for purification of the α\textsubscript{L} subunit (Section 4.2.4).
4.2.3 Production and purification of proteins

4.2.3.1 Purification of the θ subunit

The method for purification of the θ subunit was derived from Hamdan et al. (2002) and Keniry et al., (2006) and is a combination of two protocols. *E. coli* strain BL21(λDE3)/pLysS/pCM869 (Keniry et al., 2000) was grown at 25°C in 1 L LB autoinduction medium (Section 2.1.4.3) containing ampicillin (100 mg/mL) and chloramphenicol (35 mg/mL) for a period of 48 h. Cells were harvested by centrifugation (11 000 × g, 7 min), frozen in liquid nitrogen and stored at –80°C.

Frozen cell pellets (7.5 g in total) were resuspended in 100 mL of lysis buffer (50 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM DTT, 20 mM spermidine, 0.5 mM PMSF, 10% sucrose) before being passed twice through a French press (12 000 psi) to lyse cells. The lysate was clarified by centrifugation (38 000 × g; 30 min) to yield soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.40 g/mL) and stirring for 60 min were collected by centrifugation (38 000 × g; 45 min) and dissolved in 35 mL of buffer Aθ (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 20% glycerol) + 160 mM NaCl. The solution was dialysed against 2 L of the same buffer overnight, to yield Fraction II.

Two flow-through chromatography steps at high and low NaCl concentration were performed to remove contaminating DNA and proteins as completely as possible, with the aim of avoiding overloading the binding capacity of the resin. Fraction II was applied by gravity (at ~1 mL/min) onto a column (2.5 × 16 cm) of Toyopearl DEAE-650M resin that had been equilibrated with buffer Aθ + 160 mM NaCl. Fractions containing proteins that did not bind to the column were pooled and dialysed against two changes of 2 L of buffer Aθ containing no NaCl to yield Fraction III (80 mL). Fraction III was reapplied to the same column, as above, now equilibrated in
buffer Aθ with no NaCl. The θ protein (theoretical pI = 9.15) does not bind to
DEAE resin at pH 7.6 and was collected in flow-through fractions. Fractions
containing proteins which did not bind to the DEAE resin in buffer Aθ were
pooled and dialysed against three changes of 2 L of buffer P0 (20 mM
sodium phosphate, pH 6.5, 1 mM EDTA, 1 mM DTT, 10% glycerol) to yield
Fraction IV (120 mL).

Fraction IV was applied at 0.5 mL/min onto a column of
phosphocellulose (Whatman P-11) resin (2.5 × 13 cm) that had been
equilibrated in buffer P0. After the column had been washed with 150 ml of
the same buffer, bound protein was eluted using a linear gradient (400 mL)
of 0–1 M NaCl in buffer P0. The θ protein eluted as a single peak at ~250 mM
NaCl and fractions containing highly purified protein were identified using
SDS-PAGE and pooled, before being dialysed against two changes of 2 L of
buffer Sθ (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 20 % glycerol)
to give Fraction V (20 mL).

Fraction V containing ~150 mg of θ (7.8 mg/mL) was aliquoted then
frozen in liquid nitrogen and stored at −80°C. A final A$_{260}$/A$_{280}$ of 0.62 was
calculated. SDS-PAGE analysis of the final purified protein product can be
viewed in Figure 4.2.

4.2.3.2 Production and purification of $^{13}$C,$^{15}$N-θ

The *E. coli* strain BL21(λDE3)/pLysS/pCM869 (coding for the wild-
type *E. coli* θ subunit) was grown in 1 L of minimal medium (Section 2.1.4.4)
at 30°C for 24 h to an OD$_{600}$ of 1.1. To induce overproduction, IPTG was
added to 1 mM and the cultures were shaken for a further 5 h at 30°C to a
final OD$_{600}$ of 2.2. Cells were harvested by centrifugation (11000 × g; 7 min),
frozen in liquid nitrogen and stored at −80°C.
Frozen cells (3.25 g) were thawed and protein purification was carried out as outlined for the purification of the θ subunit in Section 4.2.3.1. A total of 22 mg of highly purified \(^{13}\text{C,}^{15}\text{N-0}\) protein was collected at a concentration of 1.1 mg/mL. Aliquots were frozen in liquid nitrogen and stored at \(\sim\)80°C.

![Figure 4.2: SDS-PAGE analysis of purified protein products.](image)

Protein samples were standardised to \(\sim\)10 μg per lane and are arranged as follows: (I) \(\beta_2\), (II) \(\alpha_6\), (III) \(\epsilon_6\), (IV) \(\theta\), (V) \(\alpha\epsilon612\), (VI) \(\beta\epsilon\epsilon\), (VII) \(\epsilon193\), (VIII) \(\epsilon193\), (IX) \(\epsilon193\), (X) \(\epsilon193\). For clarity, lane (IX) has been digitally altered by laterally translating the reference protein standard lane (M) adjacent to (IX) so to remove intervening lanes of unrelated protein samples. A reference protein standard is presented in the far left lane (M) of each panel and is annotated only once for simplicity.

4.2.3.3 Production and purification of \(\epsilon193\)

The \(E.\ coli\) strain BL21pLysS/(ADE3)/pSJ1447 (coding for the truncated \(E.\ coli\) \(\epsilon193\) protein) was grown in LB medium and induced to overexpress using IPTG at an earlier date by Dr Slobodan Jergic. A frozen cell pellet was obtained from \(\sim\)80°C storage courtesy of Dr Jergic for purification of the \(\epsilon193\) protein.

Frozen cells (4.2 g) were thawed and resuspended in 60 mL of lysis buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 10 mM spermidine, 10% (w/v) sucrose, 200 mM NaCl) before being passed twice through a French press (12 000 psi) to lyse cells. The lysate was clarified by
centrifugation (38 000 g; 20 min) to yield the soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.35 g/mL) and stirring for 60 min were collected by centrifugation (38 000 × g; 45 min) and dissolved in 25 mL of buffer A_ε193 (25 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM DTT, 10% glycerol, 5 mM MgCl₂, 75 mM NaCl; filtered). The solution was dialysed against 2 L of the same buffer overnight, to yield Fraction II.

Fraction II was applied by gravity (at ~1 mL/min) onto a column of DEAE-650M resin that had been equilibrated with buffer A_ε193. Fractions containing proteins that did not bind to the column were pooled and dialysed against 2 L of buffer B_ε193 (25 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM DTT, 10% glycerol) to give Fraction III (50 mL).

Fraction III was loaded using a 50 mL Superloop connected to an ÄKTA system onto an 8 mL 10/100 GL Mono Q column that had been pre-equilibrated in buffer B_ε193. After the column had been washed with 40 mL of the same buffer, bound proteins were eluted using a linear gradient (70 mL) of 0–200 mM NaCl in buffer B_ε193. The ε_193 protein eluted as a broad set of peaks and SDS-PAGE analysis of fractions of interest showed significant protein contamination, however, as the protein was to be subsequently mixed with the ^13C,^15N-0 protein and purified as a complex (Section 4.2.3.5), a high degree of purity was not essential at this stage.

A total of 50 mL of sample was pooled as Fraction IV, with a concentration of ε_193 protein (estimated following SDS-PAGE analysis) of 0.1–0.2 mg/mL (5–10 mg yield). The molecular weight of ε_193 in 0.1 % formic acid was determined by ESI-MS as 21 349.8 Da and may be compared to the theoretical calculated value of 21 351.3 Da (N-terminal methionine removed).

4.2.3.4 Production and purification of ^13C^15N ε_193
The *E. coli* strain BL21/(λDE3)pLysS/pSJ1447 (coding for the truncated *E. coli* ε193 protein) was grown in 1 L of minimal medium (Section 2.1.4.4) at 30°C for 24 h to an OD600 of 0.8. To induce overproduction, IPTG was added to 1 mM and the cultures were shaken for a further 5 h at 30°C to a final OD600 of 1.3. Cells were harvested by centrifugation (11000 × g; 7 min), frozen in liquid nitrogen and stored at −80°C.

Frozen cells (2.2 g) were thawed and protein purification was carried out as outlined for the purification of the ε193 subunit in Section 4.2.3.3. A total of 12 mL of sample was pooled as Fraction IV, with an estimated 13C,15N-ε193 concentration of 0.5 mg/mL (~6 mg yield).

### 4.2.3.5 Purification of alternately 13C,15N-double-labelled ε193:θ complexes

Complexes of ε193:13C,15N-θ and 13C,15N-ε193:θ were each prepared using a comparable method that was derived, with modifications, from the co-purification procedure of the αεθ Pol III core complex (Section 4.2.7).

Immediately following the purifications of ε193 and 13C,15N-ε193, pooled samples (Fraction IV in each case) were mixed with the corresponding 13C,15N-θ or unlabelled θ proteins required (Sections 4.2.3.1 and 4.2.3.2), in a targeted molar ratio of 1:3 as follows: (i) 5–10 mg of ε193 was mixed with 5 mg of 13C,15N-θ and (ii) ~6 mg of 13C,15N-ε193 was mixed with 7.8 mg of θ. Complexes were then dialysed overnight against 2 L of complex buffer (25 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM DTT, 10% glycerol) before being loaded onto a 1 mL 5/50 GL Mono Q column (GE Lifesciences) that had been pre-equilibrated in the same buffer. Proteins were loaded onto the column using a 10 mL Superloop connected to an ÄKTA system and the column was washed well with complex buffer (10–15 mL) to remove excess θ or 13C,15N-θ, which were not expected to be retained by anionic interactions using a
MonoQ column at pH 7.6. Bound complex was eluted using a gradient (20 mL) of 0–200 mM NaCl in complex buffer. The $^{13}\text{C},^{15}\text{N}$-$\epsilon_{193}$:0 complex eluted as a single peak at approximately 30 mM NaCl. The $\epsilon_{193}$:$^{13}\text{C},^{15}\text{N}$-0 complex, however, did not bind to the column under these conditions and was collected in flow-through fractions. To facilitate binding, the previous purification step was repeated after dialysis against complex buffer adjusted to pH 8.2. Following this modification, the complex was retained by the Mono Q column and eluted as a single peak at approximately 45 mM NaCl during the gradient elution.

Fractions containing highly purified protein complexes were identified by SDS-PAGE and pooled. A total of 7.5 mL of $\epsilon_{193}$:$^{13}\text{C},^{15}\text{N}$-0 was collected at a concentration of 0.6 mg/mL (4.5 mg yield). A total of 3.5 mL of $^{13}\text{C},^{15}\text{N}$-$\epsilon_{193}$:0 was collected at a concentration of 1.1 mg/mL (3.9 mg yield). Aliquots were frozen in liquid nitrogen and stored at $-80^\circ\text{C}$. SDS-PAGE analysis of the final purified protein products can be viewed in Figure 4.2.

### 4.2.3.6 Purification of the $\epsilon_{L}$ subunit

The $\epsilon_{L}$ subunit is completely insoluble when overproduced and is purified using a refolding method, first described by Scheuermann and Echols (1984) for the purification of wild-type $\epsilon$ and later optimised for increased yield as described in Hamdan et al. (2002).

The *E. coli* strain BL21/(λDE3)/pLysS/SJ1446 (containing the modified *dnaQ* gene coding for the $\epsilon_{L}$ subunit; Jergic *et al.*, 2013) was grown at room temperature (22°C) in LB autoinduction medium (Section 2.1.4.3) containing ampicillin (100 mg/L) and chloramphenicol (35 mg/L). Protein production was induced over a period of 72 h as described by Studier (2005). Cells were harvested by centrifugation (11 000 × g, 7 min), frozen in liquid nitrogen and stored at $-80^\circ\text{C}$. 
Frozen cell pellets (15.7 g in total) were resuspended in 100 mL of lysis buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM DTT) before being passed twice through a French press (12 000 psi) to lyse cells. The lysate was clarified by centrifugation (38 000 × g; 15 min) to yield the insoluble Fraction I (ε is completely insoluble and remains in the pellet). The pellet was then washed to remove soluble material as completely as possible.

In wash step 1, Fraction I was resuspended as completely as possible in 50 mL of lysis buffer + 1 M NaCl and passed through a French press twice (12 000 psi). The suspension was centrifuged (38 000 × g, 30 min) and the pellet collected.

In wash step 2, the collected pellet was resuspended in lysis buffer (this time with no NaCl) and the suspension was homogenised, centrifuged and the pellet collected as described in wash step 1.

Wash steps 1 and 2 were repeated to yield a thoroughly washed, insoluble pellet. The pellet was then resuspended in 7.5 mL of lysis buffer to give a milky white suspension that was frozen in liquid nitrogen and stored at −80°C as Fraction II.

After thawing, Fraction II was diluted by the addition of 67.5 mL of buffer Aε (25 mM Tris.HCl, pH 7.6, 1 mM EDTA, 5 mM DTT, 20% glycerol) at 6°C with stirring. Proteins in the suspension were solubilised by adding 75 mL of denaturation buffer Aε + 6 M guanidine hydrochloride (GuHCl) to yield Fraction III (final GuHCl concentration of 3 M). The solution turned transparent during continual stirring for 45 min.

To assist refolding, the sample was transferred to room temperature while stirring and allowed to warm slightly over 30 min before slowly adding 300 mL of buffer Aε (at 6°C) over a further 30 min to give Fraction IV (final GuHCl concentration of 1 M). Fraction IV (450 mL) had a slightly turbid appearance as insoluble material began to precipitate out of solution. The sample was divided equally into two dialysis bags and each dialysed
against 2 L of buffer Aε + 500 mM NaCl for 5 h. Samples were then further
dialysed together against 3 changes of 2 L buffer Bε (25 mM Tris.HCl, 1 mM
EDTA, 1 mM DTT, 10% glycerol) for 4 h minimum each. Following dialysis,
samples were pooled and centrifuged (11 000 × g, 90 min) to pellet a
significant amount of insoluble material that had precipitated during the
refolding process. A total of ~500 mL of clarified supernatant was collected
as Fraction V.

Fraction V was applied by gravity (at ~2.5 mL/min) onto a column (2.5
× 13 cm) of Toyopearl Super-Q resin that had been equilibrated in buffer Bε.
The column was then washed in 80 mL of buffer Bε and bound proteins were
eluted using a linear gradient (200 mL) of 0–300 mM NaCl in buffer Bε. A
large amount of ε protein was still apparent in flow-through fractions
analysed by SDS-PAGE and these fractions were pooled, diluted by a factor
of 1.5 to ensure binding, reapplied to the Super-Q resin and eluted as
described above.

On each occasion, fractions containing highly purified εL protein
eluted as a single peak at ~100 mM NaCl. Following each purification run,
relevant fractions were identified using SDS-PAGE, then pooled, frozen in
liquid nitrogen and stored at −80°C. A total volume of 35 mL containing 90
mg of εL (2.6 mg/mL) was obtained from the first elution and 40 mL
containing 72 mg of εL (1.8 mg/mL) was obtained from the second. A final
A_{260}/A_{280} of 0.60 was calculated. SDS-PAGE analysis of the final purified
protein product can be viewed in Figure 4.2.

4.2.3.7 Purification of the β2 subunit

The β2 subunit was purified following the protocol outlined in Oakley
et al. (2003). The plasmid pND261 places the dnaN gene that encodes β under
the control of the temperature inducible bacteriophage λ promoter.
Optimisation of the nucleotide distance between the ribosome-binding site and gene start codon enables high-level expression of the β subunit.

The *E. coli* strain AN1459/pND261 was grown in $6 \times 1$ L of LBTA medium (Section 2.1.4.1) at 30°C to an OD$_{600}$ of 0.5. To induce overproduction, cultures were rapidly heated to 42°C and shaken for a further 3 h. Cells were harvested by centrifugation (11000 $\times$ g; 7 min), frozen in liquid nitrogen and stored at −80°C.

Frozen cell pellets (10.5 g in total) were thawed and resuspended in 110 mL of lysis buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 10 mM spermidine) before being passed twice through a French press (12 000 psi) to lyse cells. The lysate was clarified by centrifugation (38 000 $\times$ g; 20 min) to yield soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.30 g/mL) and stirring for 60 min were collected by centrifugation (38 000 $\times$ g; 60 min) and dissolved in 30 mL of buffer Aβ (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 10% glycerol) + 150 mM NaCl. The solution was dialysed against 2 L of the same buffer overnight, to yield Fraction II.

Fraction II was applied by gravity (at ~1 mL/min) onto a column of DEAE-650M resin that had been equilibrated with buffer Aβ + 150 mM NaCl. Fractions containing proteins that did not bind to the column were pooled and dialysed against 2 L of buffer Aβ containing no NaCl to yield Fraction III (60 mL).

Fraction III was reapplied onto the same column, now equilibrated in buffer Aβ, under the same conditions described above. The column was then washed with 100 mL of buffer Aβ and bound proteins were eluted using a linear gradient (250 mL) of 0–400 mM NaCl in buffer Aβ. Eluted fractions were analysed using SDS-PAGE and those containing the β protein were pooled and dialysed against three changes of 1 L of buffer Pβ (20 mM sodium phosphate pH 7.6, 1 mM DTT) to yield Fraction IV (60 mL).
Fraction IV was passed at 1 mL/min through a column of hydroxyapatite (2.5 × 14 cm; BioRad) that had been equilibrated in buffer Pβ. Under these conditions, the β2 subunit passes unretarded through the column and was collected in the flow-through fractions. Purification using hydroxyapatite is a critical purification step that removes contaminating proteins from the β2 subunit, including suspected proteases. Samples purified without this final step have been noted to have reduced or eliminated crystal growth. Flow-through fractions were analysed by SDS-PAGE and those containing the β2 subunit were pooled and dialysed against two changes of 1 L of buffer Sβ (30 mM Tris.HCl, pH 7.6, 0.5 mM EDTA, 1 mM DTT, 20% glycerol, 50 mM NaCl) to yield Fraction V (45 mL).

Fraction V containing ~480 mg of β2 (10.6 mg/mL) was aliquoted then frozen in liquid nitrogen and stored at −80°C. A final A260/A280 of 0.67 was calculated. SDS-PAGE analysis of the final purified protein product can be viewed in Figure 4.2.

4.2.3.8 Purification of the αE612K subunit

The *E. coli* strain BL21(λDE3)/pLysS/pKO1538 contains a modified version of the *dnaE* gene encoding a single amino Glu→Lys change at a.a. position 612 of the α protein. The plasmid pND517 (Wijfells *et al.*, 2004) contains the *dnaE* gene and was used as a template for construction of the plasmid pKO1538. Following the protocol described in Wijfells *et al* (2004), the strain was grown in 6 × 1 L of LBTA/C medium (Section 2.1.4.1) at 37°C to an OD600 of 0.5. After cultures were rapidly cooled to 30°C on ice, overproduction was induced by addition of IPTG to 0.5 mM and continual shaking for a further 4 h. Cells were harvested by centrifugation (16 900 × g; 7 min), frozen in liquid nitrogen and stored at −80°C.
Frozen cell pellets (30 g in total) were thawed and resuspended completely in 300 mL of lysis buffer (50 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM DTT, 20 mM spermidine, 10% (w/v) sucrose). Cells were lysed by the addition of chicken egg lysozyme (Sigma) to the suspension to 0.2 mg/mL, then divided into 50 mL volumes and heated to 30°C for 5 min in a water bath. Each sample was gently mixed each minute to promote even lysis. Samples were then pooled and left stirring at 6°C for a further 30 min before the lysate was clarified by centrifugation (16 900 × g; 30 min) to yield soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.36 g/mL) and stirring for 60 min were collected by centrifugation (16 900 × g; 60 min) and dissolved in 60 mL of buffer Aα (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 20% glycerol) + 190 mM NaCl. The solution was dialysed against two changes of 2 L of the same buffer overnight, to yield Fraction II.

Fraction II was divided into two equal portions and one applied by gravity (at ~1 mL/min) onto a column of DEAE-650M resin that had been equilibrated with buffer Aα + 190 mM NaCl. This procedure was repeated for the second portion of Fraction II and fractions from each purification run containing proteins that did not bind to the column were pooled and dialysed against 2 L of buffer Aα containing no NaCl to yield Fraction III (100 mL).

Fraction III was reapplied onto the same column, now equilibrated in buffer Aα + 20 mM NaCl, under the same conditions described above. The column was then washed with 50 mL of buffer Aα + 20 mM NaCl and bound proteins were eluted using a linear gradient (300 mL) of 20–500 mM NaCl in buffer Aα. The αE612K protein eluted under multiple A280 peaks, centred at ~140 mM NaCl. Eluted fractions were analysed using SDS-PAGE and those containing the αE612K protein were pooled and dialysed against three
changes of 1 L buffer of Pa (20 mM sodium phosphate pH 6.5, 1 mM EDTA, 1 mM DTT, 10% glycerol) to yield Fraction IV (30 mL).

Fraction IV was centrifuged (38 000 × g; 10 min) to remove a small amount of precipitate generated during dialysis and then applied by gravity at 0.5 mL/min onto a column of phosphocellulose resin that had been equilibrated in buffer Pa + 10 mM NaCl. After the column had been washed with 50 mL of the same buffer, bound proteins were eluted using a linear gradient (400 mL) of 0–700 mM NaCl in buffer Pa. The αE612K protein eluted as a broad peak with some persistent contamination from suspected proteolytic fragments. Eluted fractions were analysed using SDS-PAGE and selected samples were pooled and dialysed against three changes of 1 L of buffer Bα (25 mM Tris.HCl, pH 7.6, 0.5 mM EDTA, 1 mM DTT, 10% glycerol) to yield Fraction V (40 mL).

Fraction V was applied by gravity at ~1 mL/min to a column of heparin-Sepharose (GE Life Sciences) that had been equilibrated in buffer Bα. The column was washed with 50 mL of buffer Bα and bound proteins were eluted using a linear gradient (150 mL) of 0–500 mM NaCl. The full-length αE612K protein binds tightly to this resin and can be mostly separated from proteolytic contaminants. The αE612K protein eluted at ~190 mM NaCl and fractions under the main peak were assessed for quality using SDS-PAGE. Selected fractions were pooled and dialysed against 1 L of buffer Sα (30 mM Tris-HCl, pH 7.6, 0.5 mM EDTA, 1 mM DTT, 15% glycerol, 150 mM NaCl) before the sample was aliquoted, frozen in liquid nitrogen and stored at −80°C. A total of ~4.5 mg of αE612K in a volume of 20 mL was concentrated by centrifugation to 2.4 mL at a concentration of 1.9 mg/mL. SDS-PAGE analysis of the final purified αE612K protein product can be viewed in Figure 4.2.
4.2.4 Protein-protein affinity purification of the α1 subunit

4.2.4.1 Purification of biotinylated τ16

The 16 kDa C-terminal domain V of the *E. coli* Pol III HE τ subunit (termed τ16) interacts tightly with the C-terminal domain of the α polymerase subunit. A version of the τ16 domain that had been modified to include an N-terminal amino acid biotinylation sequence (MAGLNDIFEAQKIEWHEH; Beckett *et al.*, 1999; construct is here termed tag-τ16) was over-expressed in the presence of biotin and then purified to investigate its use as a chromatography affinity tag for enhanced purification of the α subunit.

The λ promoter plasmid pKO1294 was as described in Jergic *et al.*, 2007. The *E. coli* strain BL21(λDE3)recA/pKO1294 was grown in 6 × 1 L of LBTA medium at 30°C to an OD₆₀₀ of 0.1. At this point, cultures were supplemented with 80 μM D-biotin (Sigma) and shaking continued until an OD₆₀₀ of 0.5 was reached. To induce overproduction, cultures were rapidly heated to 42°C and shaken for a further 4 h. Cells were harvested by centrifugation (11000 × g; 7 min), frozen in liquid nitrogen and stored at –80°C.

Frozen cell pellets (7.3 g in total) were thawed and resuspended in 100 mL of lysis buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 20 mM spermidine, 10% (w/v) sucrose, 0.7 mM PMSF) before being passed twice through a French press (12 000 psi) to lyse cells. The lysate was clarified by centrifugation (18 000 × g; 30 min) to yield soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.36 g/mL) and stirring for 45 min were collected by centrifugation (38 000 × g; 40 min) and dissolved in 30 mL of buffer Aτ (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 20% glycerol) + 160 mM NaCl. The solution was dialysed against 1 L of the same buffer overnight, to yield Fraction II.
Fraction II was applied by gravity (at ~1 mL/min) onto a column of DEAE-650M resin that had been equilibrated with buffer Aτ + 160 mM NaCl. Fractions containing proteins that did not bind to the column were pooled and dialysed against three changes of 1 L of buffer Aβ containing no NaCl to yield Fraction III (80 mL).

Fraction III was reapplied onto the same column now equilibrated in buffer Aτ, under the same conditions described above. The column was then washed with 120 mL of buffer Aτ and bound proteins were eluted using a linear gradient (420 mL) of 0–160 mM NaCl in buffer Aτ. SDS-PAGE analysis indicated the tag-τ16 protein eluted from the column over two partially separated peaks: a primary peak at approximately ~90 mM NaCl and a smaller secondary peak at ~110 mM NaCl. It was reasoned that one of these peaks was comprised largely of biotinylated protein and the other unbiotinylated. Selected fractions were analysed for accurate weight determination by denaturing nanoESI-MS in 0.1% (w/v) formic acid. Biotinylated tag-τ16 (here termed bio-τ16) was the predominant species in the smaller, second elution peak while unbiotinylated tag-τ16 was mostly present in fractions from the first, larger peak. The appropriate fractions were pooled, frozen in liquid nitrogen and stored at −80°C. A total of ~38 mg of tag-τ16 in 50 mL of buffer Aτ + 90 mM NaCl was retained for potential future in vitro biotinylation. A total of ~14 mg of bio-τ16 in 40 mL of buffer Aτ + 110 mM NaCl was recovered. SDS-PAGE analysis of the final purified bio-τ16 protein product can be viewed in Figure 4.2.

4.2.4.2 Assembly of a biotinylated τ16 - streptavidin column

All procedures to follow were performed at 6°C: 10 mL of high-capacity streptavidin-agarose resin (Pierce Biotechnology) was mixed with 10 mL of buffer Pt (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM DTT, 10%
glycerol, 200 mM NaCl, 0.02% (w/v) sodium azide). Approximately 4 mL of the resin-buffer suspension was poured into a column (1 cm diameter) and allowed to settle with continued washing with additional buffer Pt. The remaining 14 mL of solution was mixed with half the thawed sample of purified bio-τ₁₆ (~7 mg) and the additional resin was allowed to settle on top of the unconjugated resin as it was continually washed with buffer Pt. During washing of the resin, the flow-through was collected and passed through the column once more to ensure immobilisation of as much bio-τ₁₆ as possible. Unconjugated streptavidin-agarose resin was packed below the conjugated resin to help retain within the column any dissociated or unbound bio-τ₁₆. After the resin had been thoroughly washed, the column was assembled and stored in buffer Pt for future use.

Following successful purifications of the αL subunit using this column (described below in Section 4.2.4.3), the remaining sample of purified bio-τ₁₆ (~7 mg) was passed through the column at a later date to boost the capacity of the resin (here termed SA-τ₁₆ resin). The column was stored in buffer Pt for long-term storage following each use.

4.2.4.3 Purification of the αL subunit using protein-protein affinity chromatography

Two productions and purifications of the αL subunit were performed; the first as a trial of the recently assembled SA-τ₁₆ affinity column and the second as a supplementary purification after the second portion of bio-τ₁₆ had been added to the affinity column to boost capacity. As both methodologies were near identical, the procedure will be described here only once and relevant differences will be noted.

Plasmid pSJ1392 (Jergic et al., 2013) contained the modified dnaE gene encoding αL under control of a T7 promoter. The E. coli strain
BL21(λDE3)pLysS/pSJ1392 was grown in 6 × 1 L of LBTA/C medium (Section 2.1.4.1) at 37°C to an OD_{600} of 0.4. After cultures were rapidly cooled to 30°C on ice, overproduction was induced at an OD_{600} of 0.5 by addition of IPTG to 0.5 mM with continued shaking for a further 4 h. Cells were harvested by centrifugation (16 900 × g; 7 min), frozen in liquid nitrogen and stored at –80°C.

Frozen cell pellets (13 g in total from first production and 16 g from the second) were thawed and resuspended completely in 10 mL of lysis buffer (50 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM DTT, 20 mM spermidine, 10% (w/v) sucrose) per gram of cells. Cells from each production were lysed and proteins were partially purified by precipitation with ammonium sulphate and two-step DEAE-650M resin anion-exchange chromatography as described previously for the purification of αE612K (Section 4.2.3.8).

Directly following gradient NaCl elution of proteins from the DEAE resin, fractions containing the impure αL protein (now in buffer Aα + 150 mM NaCl) were applied onto the SA-τ16 resin by gravity at ~1 mL/min. The column was then washed with 15 mL of buffer Aα + 20 mM MgCl₂ and bound protein eluted using a linear gradient (30 mL) of 20–4000 mM MgCl₂. The αL protein bound tightly to the SA-τ16 resin and eluted as a single peak at ~2.8 M MgCl₂. Fractions under the peak were immediately pooled and dialysed against 1 L of buffer Sατ (30 mM Tris-HCl, pH 7.6, 0.5 mM EDTA, 1 mM DTT, 15% glycerol, 150 mM NaCl, 10 mM DTT), followed by two changes of 1 L of buffer Sα. A minimum DTT concentration of 5–10 mM is critical when the protein is in high MgCl₂ concentrations as batches of αL and α proteins produced using lower DTT concentrations (below 2 mM) displayed significantly reduced activity (discussed further in Sections 4.3.1.2 and 4.4.1.3). Following dialysis, the sample was aliquoted, frozen in liquid nitrogen and stored at –80°C. A total of 12 mg of αL at 0.4 mg/mL (30 mL)
was obtained from the first purification and 24 mg at 1.6 mg/mL (15 mL) was obtained from the second purification. A final $A_{280}/A_{260}$ of 0.59 was calculated for the second purification. A step-by-step SDS-PAGE analysis of the αL purification process can be viewed in Figure 4.2.

4.2.5 Purification of αLεLθ core complexes

Purified core complexes were prepared by ion-exchange chromatography using a method optimised by Dr Slobodan Jergic for purification of the wild-type αεθ complex (Jergic et al., 2013). Two separate samples of the αLεLθ complex were prepared, one following each αL purification (Section 4.2.4.3). On each occasion, individual subunits were mixed in an approximate molar ratio of 1:3:9 (αL:εL:θ, respectively) and dialysed overnight against 2 L buffer Aαεθ (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 10% glycerol, 65 mM NaCl). In the first purification, 10 mg of αL was mixed with 7 mg of εL and 7 mg of θ. In the second purification, 24 mg of αL was mixed with 15 mg of εL and 15 mg of θ.

Following dialysis, samples were loaded by gravity at ~0.5 mL/min onto a column Super-Q resin that had been washed and equilibrated thoroughly in buffer Aαεθ. The column was then washed well with 20 mL of the same buffer and bound complex was eluted from the column using a linear gradient (20 mL) of 65–1600 mM NaCl. On each occasion, the protein complex eluted as a single sharp peak, very soon after the beginning of the linear increase in salt concentration (approximately 100 mM NaCl). Fractions under the peak were pooled and dialysed overnight against 1 L of buffer Sαεθ (50 mM Tris.HCl, pH 7.6, 0.5 mM EDTA, 1 mM DTT, 20% glycerol, 100 mM NaCl). Following dialysis, the sample was aliquoted, frozen in liquid nitrogen and stored at –80°C. A total of 8.1 mg of αLεLθ at 0.6 mg/mL (13.5
mL) was obtained from the first purification and 17.6 mg at 1.6 mg/mL (11 mL) was obtained from the second. SDS-PAGE analysis of the final purified \( \alpha_L \varepsilon_L \theta \) core complex can be viewed in Figure 4.2.

**4.2.6 Preparation of the \( \beta_2 \alpha_L \varepsilon_L \theta \) complex for SAXS experiments**

A small-scale trial co-purification of \( \beta_2 \alpha_L \varepsilon_L \theta \) was performed as follows: core \( \alpha_L \varepsilon_L \theta \) complex (1 mg at 0.6 mg/mL) was mixed with \( \beta_2 \) (1.5 mg at 10.6 mg/mL) to target an approximate 1:3 molar ratio of core complex to \( \beta_2 \) subunit, respectively. The sample was dialysed overnight in 1 L of buffer CB (50 mM Tris.HCl, pH 7.6, 0.5 mM EDTA, 2 mM DTT, 15% glycerol) + 30 mM NaCl to generate Fraction I.

Fraction I was applied to a 1 mL column of Mono Q resin, which was then washed well with 10 mL of buffer CB + 30 mM NaCl and bound protein eluted using a gradient (30 mL) of 30–800 mM NaCl in buffer CB. Unresolved problems with the integrity of the particular column used resulted in high-pressure, slow loading and poor resolution. However, the chromatogram did give an indication that the complex bound more tightly than expected to the Mono Q resin. Consequently, all fractions under areas of the chromatogram displaying a significant \( A_{280} \) reading were pooled and dialysed against buffer CB + 100 mM NaCl overnight to generate Fraction II. Fraction II was applied to a different 1 mL column of Mono Q resin at 0.5 mL/min, which was then washed with 25 mL of buffer CB + 100 mM NaCl and bound protein was eluted using a gradient (50 mL) of 100–1000 mM NaCl in buffer CB. Excess \( \beta_2 \) protein eluted from the column as a single sharp peak at ~180 mM NaCl. The \( \beta_2 \alpha_L \varepsilon_L \theta \) complex eluted very soon after at ~230 mM NaCl as a sharp primary peak, followed by a secondary shoulder peak that considerably broadened the volume over which the \( \beta_2 \alpha_L \varepsilon_L \theta \) complex
eluted. SDS-PAGE analysis of fractions under each peak confirmed the stoichiometric integrity of the β2εθ complex.

A total of 3 mL of sample was pooled and concentrated using 3 kDa MWCO centrifugal concentrators (Amicon) to a final volume of 100 μL at a concentration of 2.15 mg/mL. The sample was centrifuged at 15 000 × g for 30 min at 6°C to pellet a small amount of precipitate, then frozen in liquid nitrogen and stored in dry ice in provision for SAXS experiments (Section 4.3.4)

4.2.7 Nuclear magnetic resonance (NMR) experimental conditions

All NMR experiments and data analysis were performed by Dr Kiyoshi Ozawa at the Research School of Chemistry, Australian National University. All NMR spectra were recorded at 25°C using Bruker 600 and 800 MHz NMR spectrometers equipped with cryoprobes, using 200 μl solutions in 3 mm sample tubes. 15N-HSQC spectra used $t_{1\text{max}} = 32 \text{ ms}$, $t_{2\text{max}} = 102 \text{ ms}$ and total recording times of 1–13 h. 2D HN(CO) spectra and 3D HN(CO)CA and HNCA spectra were recorded in 20–24 h per spectrum. D2O was added to all samples to a final concentration of 10% (v/v) prior to NMR measurements.

The $^{13}$C,$^{15}$N-ε193θ and $^{13}$C,$^{15}$N-θ:ε193 complexes (34 μM) were titrated with 34 and 68 μM β2 and $^{15}$N-HSQC spectra recorded. Spectra were also recorded of a sample of $^{15}$N-Gln, Thr, Ser, Met, Ala, Phe labelled ε193:unlabeled θ (27 μM) sample produced by Dr Kiyoshi Ozawa (Ozawa et al., 2013) with and without added β2 at 30 μM.

4.2.8 Assaying polymerase activity by strand displacement DNA synthesis

The novel protein-protein affinity purification protocol outlined in Section 4.2.4.3 requires high concentrations of MgCl2 (up 3 M) to elute bound
polymerase from the SA-τ16 resin. Each batch of α polymerase subunit (and its mutant variants) purified using this method was assayed for strand displacement (SD) DNA synthesis using the double-stranded plasmid pSCW01 (2.03 kb; Geng et al., 2011) that had been modified to produce a template construct with a 25 nt ssDNA gap and a 59 nt ssDNA 5′ overhang (obtained courtesy of Mr. Enrico Monachino, University of Wollongong). Alternatively, oligonucleotide-primed circular single-stranded phage M13 DNA (6.4 kDa) prepared by Dr Slobodan Jergic (Jergic et al., 2013) was also employed as a template for assessment of the activity of the tagged polymerase mutant αFGE (see Chapter 5, Section 5.3.1).

A standard Pol III SD reaction contained 2.5 nM primed DNA template, 1 mM ATP, 0.5 mM of each dNTP, 30 nM τ3δδ′ψχ clamp loader, 100 nM α or mutant polymerase variant, 150 nM ε, 200 nM β2, and 850 nM SSB4 in 25 mM Tris.HCl pH 7.6, 10 mM MgCl2, 10 mM DTT and 100 mM NaCl, in a final volume of 15 μL (Jergic et al., 2013). Components (except DNA) were mixed and treated for 5 min at room temperature, cooled in ice and DNA added. Reactions were initiated at 30°C, and quenched after 20 min (unless indicated otherwise) by addition of a loading buffer containing EDTA to 100 mM and SDS to 1%. Products were separated by agarose gel electrophoresis and stained with SYBR Gold (Invitrogen, Carlsbad, CA, USA).

4.2.9 Conditions for SAXS experiments

Scattering data were recorded on the SAXS/WAXS beamline at the Australian Synchrotron by Dr Flynn Hill by size-exclusion chromatography-coupled small-angle X-ray scattering (SEC-SAXS). Two samples of the β2αεεεθ complex were prepared using condition A and condition B. Condition A contained 50 mM Tris.HCl pH 8.0, 0.1 M NaCl, 1 mM EDTA, 1 mM TCEP, 5% (v/v) glycerol. In condition B, a double-stranded 5′-fluorescein-labelled
primer-template DNA with 28 nucleotides on the template strand (5’ Fluorescein-TTTTTGTGGCACTGGCCGTCGTTGTCG) was paired with a complementary 21-mer (5’ CGACAACGACGCCAGTCCA) to form a 7 nt 5’ overhang. The oligo construct was added in slight excess to the protein sample 5 minutes before loading to reach 27 μM, and a mixture was then added to give a final concentration of 50 μM dATP, 50 μM ddCTP, 5 mM MgCl₂. Incorporation of the ddCTP was expected to extend the primer strand to 22 nt and dATP is the next incoming dNTP. In condition B, the protein was eluted in the a buffer containing the same components as the buffer in condition A, plus 5 mM MgCl₂ and 100 µM dATP. A 70 µl sample of complex at 3.7 mg/mL in each condition were injected at 0.5 ml/min onto a Wyatt WTC-030S5 size exclusion chromatography column (7.8 × 300 mm) equilibrated at 12°C in the same buffer. A₂₅₀ of the eluate was monitored immediately prior to its passage through a quartz capillary that was illuminated by a collimated 11 keV X-ray beam, λ = 1.127 Å. Scattering from the sample was measured by a Pilatus 1 M detector (Dectris, Switzerland) that recorded 2D scattering images in 2 s exposures from a position 3349 mm behind the sample. For all the frames used for the data analysis, no protein damage induced by X-rays was observed. Scattering from the buffer eluate was stable as averaged from 10 exposures prior to and after elution of protein, to give the buffer scattering. Following radial averaging and buffer subtraction, the radii of gyration, Rₛ, of five exposure bins were determined by Guinier analysis using AUTORG (Petoukhov et al., 2007) and plotted against elution volume. Sample scattering was averaged across the region of Rₛ stability, which corresponded to the main UV absorption peak in the elution profile and encompassed 20 exposures. The scattering pattern was truncated within the range 0.012 ≤ Q ≤ 0.16 Å⁻¹. The theoretical SAXS patterns, radii of gyration and envelope volumes of various atomic models
were calculated using CRYSOL (Svergun et al., 1995), for comparison with experimental data.

4.2.10 Preparation of the $\beta_2\alpha_1\varepsilon_1\theta$ complex for crystallography

For crystallography experiments, purification of larger amounts of $\beta_2\alpha_1\varepsilon_1\theta$ complex was required. A total of 21 mg of $\alpha_1\varepsilon_1\theta$ core complex was mixed with 31.5 mg of $\beta_2$, targeting an approximate molar ratio of 1:3. The sample was dialysed overnight in 1 L buffer CB (Section 4.2.8) + 100 mM NaCl to generate Fraction I (~28 mL).

Fraction I was divided into four volumes of 7 mL and each was separately loaded onto a 1 mL column of Mono Q resin, then washed with 12–15 mL of buffer CB + 100 mM NaCl. On the first occasion the $\beta_2$ protein was eluted from the column using a gradient of 100–200 mM NaCl in 15 mL buffer CB. Following $\beta_2$ elution, the gradient was immediately increased to 200–500 mM NaCl in 10 mL buffer CB in an effort to reduce the broadened elution profile observed in Section 4.2.8; however, a partially conflated ‘double-peak’ was still observed and it was decided to investigate the use of a step-gradient to elute the $\beta_2\alpha_1\varepsilon_1\theta$ complex in the second purification run.

The second sample was loaded onto the Mono-Q column and washed with buffer as described above and the $\beta_2$ protein then eluted using a gradient of 100–200 mM NaCl in 10 mL of buffer CB. Immediately after $\beta_2$ had eluted from the column, the concentration of NaCl of buffer CB was increased to 500 mM and the $\beta_2\alpha_1\varepsilon_1\theta$ complex eluted as a single, sharp peak over approximately 2 mL (2 column volumes). Subsequently, the third and fourth samples of Fraction I were subjected to the same purification procedure and accordingly gave similar elution profiles.
Samples under the $\beta_2\alpha_1\epsilon\theta$ peaks from each of the four purification runs were assessed for composition by SDS-PAGE and relevant fractions were pooled and dialysed against buffer CBS (50 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM DTT, 15% glycerol, 100 mM NaCl) to generate Fraction II (12 mL). Approximately 5 mL of Fraction II at a concentration of 1.1 mg/mL was aliquoted, frozen in liquid nitrogen and stored at −80°C. The remaining sample was dialysed overnight into 1 L of crystallography buffer (10 mM Tris.HCl pH 7.6, 1 mM EDTA, 2 mM DTT, 100 mM NaCl) and then concentrated using 3 kDa MWCO centrifugal concentrators (Amicon) to a final volume of 700 μL at a concentration of 7.8 mg/mL. A final SDS-PAGE analysis of the purified complex is shown in Figure 4.2.

4.2.11 Trial growth of crystals

A 200 μL sample of $\beta_2\alpha_1\epsilon\theta$ at 7.8 mg/mL in crystallography buffer was sent to the Collaborative Crystallisation Centre (C3) in Melbourne, Australia. Sparse screening was performed using a sitting drop vapour diffusion method, using 200 nL drops in 50 μL of reservoir solution. The following screens available at the facility were utilised at both 4°C and 20°C: JCSG*, PACT, ProPlex and PEG/ion-PEG/ion2. Images of protein wells were acquired automatically at the facility and monitored remotely for crystal growth.
4.3 Results

To investigate the structure of $\beta_2\alpha\varepsilon\theta$ complex, I first developed a novel protein-protein chromatography methodology for purifying the $\alpha$ protein on a large scale and with improved purity. I then utilised a strand-displacement DNA synthesis assay to confirm polymerase activity. In conjunction with Dr Kiyoshi Ozawa, I produced alternately $^{13}$C,$^{15}$N double-labelled complexes of $\varepsilon_{193}\theta$ to probe for any undiscovered interaction between the $\beta_2$ subunit and the catalytic N-terminal domain of the $\varepsilon$ or $\theta$ subunits using NMR. These data, along with multiple other sources of structural and biochemical information, were used to build a model of the $\beta_2\alpha\varepsilon\theta$ complex which was validated by analysing a co-purified sample of the stabilised $\beta_2\alpha_L\varepsilon_L\theta$ complex, alone and in the presence of a DNA oligo, using small-angle X-ray scattering coupled to size-exclusion chromatography (SEC-SAXS). A larger amount of the $\beta_2\alpha_L\varepsilon_L\theta$ complex was then co-purified and entered into crystallography trials to investigate if stabilisation of the interaction between the Pol III core and the $\beta$-clamp would allow for the formation of diffraction quality protein crystals.

4.3.1 Purification of the $\alpha_L$ subunit

4.3.1.1 Affinity chromatography

A novel protein-protein affinity chromatography purification method was investigated to purify the $\alpha_L$ polymerase subunit by exploiting the tight interaction between the C-terminal 16 kDa segment of the $\tau$ protein ($\tau_{16}$) and the C-terminal $\tau$-binding domain of $\alpha$ (or any mutant variant of $\alpha$ with preserved $\tau$-binding capacity). A number of standard purification steps (i.e. ammonium sulphate precipitation and DEAE resin chromatography) were retained to remove as much contamination and potential proteolytic
enzymes as possible prior to passage of the sample through the SA-\(\tau_{16}\) resin (Figure 4.3 B; lanes I–V). Full-length \(\alpha_L\) bound tightly to the column and was eluted using a steep MgCl\(_2\) gradient to 4 M concentration, eluting as a sharp peak centred at approximately 2.8–3.0 M MgCl\(_2\) (Figure 4.3 A). SDS-PAGE analysis of fractions under the eluted peak showed highly pure full-length \(\alpha_L\) with an absence of any visible proteolytic fragments (Figure 4.3 B; lane VI).

![Chromatogram showing the elution profile of the \(\alpha_L\) protein from SA-\(\tau_{16}\) resin using a high-concentration of MgCl\(_2\)].

**Figure 4.3: Purification of the \(\alpha_L\) subunit by protein-protein affinity chromatography.** (A) Chromatogram showing the elution profile of the \(\alpha_L\) protein from SA-\(\tau_{16}\) resin using a high-concentration of MgCl\(_2\). The region of \(\alpha_L\) protein collected and pooled is shaded in polka-dots. (B) Step-by-step SDS-PAGE analysis of samples following: (I) Whole cells after induction with IPTG, (II) clarification of soluble fraction after cell lysis, (III) precipitation with (NH\(_4\))\(_2\)SO\(_4\), (IV) DEAE resin high-salt flow-through, (V) DEAE resin binding with gradient salt elution and (VI) binding and elution from SA-\(\tau_{16}\) resin. The protein standard presented in the far left lane (M) has been digitally added to the figure as an approximate reference and was not electrophoresed concurrently with samples I–VI.
4.3.1.2 Assessing the polymerase activity of α by strand displacement DNA synthesis

Protein-protein affinity purification of the α polymerase (and related mutant variants, including αL and the triple-mutant αKGB) required the use of harsh ionic conditions (up to 4 M MgCl₂) to disrupt the tight α-τ₁₆ interaction. Strand displacement (SD) DNA synthesis was performed to assess the comparative activity of each batch of polymerase produced using this method against a sample of wild-type α polymerase produced using a well-established ion-exchange chromatography protocol (described in Jergic et al., 2013; Figure 4.4 far left lane).

**Figure 4.4:** Strand displacement (SD) DNA synthesis shows the relative activity of α polymerase subunits and mutant variants following purification. Flap-primer extension on a dsDNA template (pSCW01) shows SD as larger than unit-length DNA products. Assessing the quantity of high molecular weight products in each lane establishes the relative efficiency of synthesis between batches of purified polymerase as compared to a sample of αWT (far left lane) purified using ‘traditional’ methods. The two lanes marked with a red asterisk both display a reduced level of activity for the αWT and αKGB polymerase variants that were purified using the protein-protein affinity methodology described in Section 4.2.4 with a low concentration (1 mM) of DTT. Activity was significantly improved by raising the concentration of DTT to 10 mM in the presence of high MgCl₂ concentrations (far right lane). Gels were stained with SYBR gold nucleic acid stain (Invitrogen).
The SD synthesis (Jergic et al., 2013) is one of the most demanding in vitro DNA replication assays and is routinely used to test the activities of the full Pol III HE. Nucleotide extension products were separated by agarose gel electrophoresis and assessed for products longer than unit length. Samples of polymerase purified through the SA-τ₁₆ step with a low buffer concentration of DTT (1 mM; Figure 4.4, central two lanes highlighted with a red asterisk) displayed a noticeably reduced activity under the conditions of this assay. Increasing the buffer concentration of DTT to 10 mM in the presence of high MgCl₂ concentrations during the purification of αₗ (Figure 4.4, far right lane) increased polymerase activity to a level comparable to a sample of αWT purified using the traditional method.

4.3.2 Only the CBM residues of ε interact with the β-clamp

The protein ε₁₉₃ contains the structured N-terminal exonuclease domain of ε (residues 2–180; Hamdan et al., 2002 a and b) and the clamp-binding motif (CBM, residues 182–187; Jergic et al., 2013). A sample of ε₁₉₃ selectively labelled with ¹⁵N-glutamine, threonine, serine, methionine, alanine and phenylalanine was prepared by Dr Kioyshi Ozawa using cell-free synthesis in the presence of excess unlabelled θ subunit (Ozawa et al., 2013) The ¹⁵N-HSQC spectra of this complex in the absence and presence of the β₂ subunit (30 μM) were compared to the ¹⁵N-HSQC spectrum of uniformly in vivo labelled ¹³C,¹⁵N-labelled ε₁₉₃:unlabelled θ complex prepared as described here in Sections 4.2.3.5 (Figure 4.5).
Figure 4.5: $^{13}$C,$^{15}$N-labelled ε₁₉₃ in complex with purified θ interacts with $\beta_2$ only through the CBM. Superimposition of $^{15}$N-HSQC spectra of uniformly in vivo $^{15}$N,$^{13}$C-labelled ε₁₉₃:unlabelled θ (black spectrum, selected resonance assignments in black) and of ε₁₉₃:θ (27 μM) labelled specifically in ε₁₉₃ with $^{15}$N-glutamine, threonine, serine, methionine, alanine and phenylalanine in the absence (blue spectrum) and presence (red spectrum) of $\beta_2$ (30 μM). Cross-peaks were observed for 52 of 59 Gln, Thr, Ser, Met, Ala and Phe residues in the structured ε₁₈₆ domain, and all were unaffected by addition of $\beta_2$ (selected signals labeled in purple); those of Ser2 and Thr3 in the disordered N-terminus could not be assigned, while Ala100, Thr128, Ser144, Ala164 and Thr179 had low intensity even in the absence of $\beta_2$. Signals in the CBM that broaden beyond recognition in the presence of $\beta_2$ (red spectrum; i.e., Gln182, Thr183, Ser184, Met185, Ala186, Phe187) are labeled in green, while assignments for flexible residues at the N- and C-terminals (Ala4, Ala188 and Thr193) that are unaffected by $\beta_2$ are labeled in orange. Figure and legend from Ozawa et al., 2013.

Under these conditions, addition of $\beta_2$ led to the disappearance of signals corresponding to residues of the CBM (ε residues 182–187) only. No significant changes were observed in the spectrum of the structured ε N-terminal proofreading domain or in ε residues beyond the CBM (ε residues 188–193). Addition of $\beta_2$ at 34 μM and 68 μM to $^{13}$C,$^{15}$N-labelled ε₁₉₃:unlabelled θ at 34 μM (data not shown) also only showed significant changes in the $^{15}$N-HSQC spectra in residues corresponding to the region of the ε CBM. These data indicate that the exonuclease domain of ε remains mobile in complex with $\beta_2$ and is only tethered to the clamp through interaction with the residues of the ε CBM.
4.3.3 The θ subunit interacts with β₂

The $^{15}$N-HSQC spectra of the $^{13}$C,$^{15}$N-$\theta$ε193 complex (34 μM; prepared as described in Section 4.2.3.5) in the absence and presence of β₂ were recorded by Dr Kiyoshi Ozawa (data not shown). Peaks corresponding to θ residues Arg60, Leu61, Ser67, Leu69, Tyr31 and Asn32 disappeared at molar ratios of complex to β₂ of 1:2, consistent with an interaction between these residues and the β₂ subunit on a timescale measurable by NMR. These measurements were recorded at a later date and, as such, were not included in the publication by Ozawa et al. 2013. They indicate that the θ subunit interacts with the β-clamp through a number of residues, although it is not possible to know from these data where on β₂ this interaction is located.

4.3.4 Building a model of the E. coli β₂αεθ:DNA complex

A 3-dimensional structural model of the E. coli β₂αεθ:DNA complex in a polymerisation mode was assembled by Dr Thomas Huber (Australian National University) by overlaying known structural models of (i) the E. coli β₂-DNA complex (Georgescu et al., 2008; PDB ID: 3BEP), (ii) full-length *Thermus aquaticus* (Taq) α with primer-template DNA (Wing et al., 2008; PDB ID: 3E0D), (iii) full-length Taq without DNA (Bailey et al., 2006; PDB ID: 2HPI), (iv) E. coli α (residues 2–917; Lamers et al., 2006; PDB ID 2HNH), the crystal structures of (v) ε²₀⁹α²⁷₀P (PDB ID: 4GX8) and ε²₀₀₀α²⁷₀P (PDB ID: 4GX9) described here in Chapter 3 and published in Ozawa et al., 2013, (vi) the NMR structure of θ in the ε₁₈₆:θ complex (Kirby et al., 2006; PDB ID: 2AXD) and (vii) the crystal structures of ε₁₈₆ (Hamdan et al., 2002; PDB IDs: 1J53 and 1J54). Changes around the L21P mutation in the α chain of the structures of ε²₀⁹α²⁷₀P and ε²₀₀₀α²⁷₀P were ignored and the two clamp-binding motifs (CBMs) located in α (residues 920–924) and ε (residues 182–187) were modelled based
on related structures of β2-peptide complexes (Jeruzalmi et al., 2001; PDB ID: 1JQL and Wolff et al., 2011; PDB ID: 3Q4L). A large number of modelled random coil peptides corresponding to the flexible linker of ε (residues 188–201) were able to span the gap from ε residues 202–205 at the site of interaction with the PHP domain of α and ε CBM residues 182–187, modelled into one hydrophobic binding pocket of the β-clamp.

The ε N-terminus:θ (εNTD:θ) sub-complex is anchored close to the β-clamp by ε CBM residues (Gln182–Phe187), which immediately follow the terminal structured α-helix forming residue (Gly180) of the globular ε N-terminal domain. In the absence of any detectable interaction between the εNTD and either the α or β2 subunits, this allows for significant free rotation about this point in the space between the α PHP domain and the β-clamp. A number of residues of the θ subunit were shown by NMR to potentially interact with the β2 subunit (Section 4.3.3). However, without knowing where on β2 these residues interact, no attempt was made at this stage to further constrain the position of the εNTD:θ sub-complex in the model.

A thousand conformers of this model were generated, allowing for free rotation around the backbone dihedral angles of ε residues Gly180–Gln182 in the space provided between the α PHP domain and the β-clamp. A final model of the E. coli β2αεθ:DNA is displayed with one randomly selected linker and εNTD:θ complex conformation selected (Figure 4.6 A and B) and also with multiple sterically allowed linker and εNTD:θ complex conformations superimposed (Figure 4.6 C).
Figure 4.6: Models of the $\beta_{\text{LE}}\theta$ complex with primer-template DNA in the polymerisation mode. (A) Compact form of the ‘closed’ structure with the globular $\varepsilon\text{NTD}:\theta$ complex positioned in the space between the $\beta$ clamp and the $\alpha$ subunit. Each subunit is coloured differently with the $\alpha$ PHP domain coloured orange to highlight the attachment point for the $\varepsilon\text{CTD}$. The molecule is displayed using a cartoon representation. (B) Alternate view of (A). The two clamp-binding motifs (CBMs) present within the $\alpha$ and $\varepsilon$ subunits are highlighted using pink spheres. Residues of the $\theta$ subunit that interact with the $\beta$ clamp are highlighted using purple spheres. (C) The molecule displayed using a space-filling representation for the $\alpha$ subunit and $\beta$ clamp as viewed in (A). The $\alpha$ PHP domain is contrasted with the remainder of the $\alpha$ here using a darker shade of blue. Multiple conformations are shown for the 22-residue linker segment that connects the two domains of the $\varepsilon$ subunit. All conformations of the linker are sterically allowed. Multiple conformations of the $\varepsilon\text{NTD}:\theta$ complex are also shown using a semi-transparent overlay to display a range of sterically permitted conformations.
4.3.5 SAXS experiments and model validation

I showed in Section 4.2.6 that the stabilised β2αεθ complex could be isolated chromatographically as an intact complex. Size-exclusion chromatography coupled to SAXS (SEC-SAXS) measurements were performed of the β2αεθ complex alone and also in the presence of a fluorescently labelled primer-template DNA oligo. The data were compared to theoretical models of the β2αεθ complex, generated as described in Section 4.3.4. The collection and processing of SEC-SAXS data was performed by Dr Flynn Hill and will be outlined here in brief. A more detailed description of experimental procedures and data processing can be found in Hill (2013) and Ozawa et al (2013).

SEC-SAXS measurements were averaged from exposures taken from a distance of 3349 mm and within the regions of protein elution where the radius of gyration was stable (Figures 4.7 A). Final scattering patterns were acquired after subtracting regions of buffer only scattering from protein scattering. A replication of protein chromatography elutions performed at a later date showed that in both condition A and B (without and with DNA) the β2αεθ protein complex elutes from the column in the expected subunit stoichiometry (2:1:1:1) over the range of the main peak (Figure 4.7 A; top and bottom panels) as determined by SDS-PAGE analysis of eluted fractions (Figure 4.7 B). Fractions corresponding to the peak for condition B with DNA present (Figure 4.7 A; bottom panel) displayed no visible fluorescence under UV illumination, indicating no fluorescein-tagged DNA was present in the protein peak. A standard sample of fluorescein-tagged DNA at the expected amount gave bright fluorescence.

Guinier analysis of β2αεθ indicated radii of gyration of $49.7 \pm 0.2 \text{ Å}$ and $50.7 \pm 0.2 \text{ Å}$ in conditions A and B, respectively, while the pair distance distributions indicated radii of gyration ($R_g$) of $48.9 \pm 0.2 \text{ Å}$ and $49.9 \pm 0.4 \text{ Å}$
and maximum dimensions of 159.5 and 177.6 Å in conditions A and B, respectively. By comparison, the averaged theoretical scattering of 1000 β2αεθ models (generated as described in Section 4.3.4) gave a mean $R_g$ of 47.6 Å and mean envelope diameter of 154.2 Å.

Figure 4.7: (A) Top: SEC-SAXS elution profile of β2αεθ recorded at a camera length of 3349 mm. The radius of gyration was stable across the protein elution peak. Twenty exposures were averaged across the range of high forward scatter intensity, 7.045—7.370 mL. Bottom: SEC-SAXS elution profile of β2αεθ in the presence of DNA, MgCl₂, dATP and ddCTP. Twenty exposures were averaged across the range of high forward scatter intensity, 7.100—7.438 mL. (B) SDS-PAGE analysis of a reproduction of the SEC-SAXS elution of β2αεθ. Fractions corresponding to the fractions collected show the complex was present in the expected 2:1:1:1 monomer stoichiometry across the range included in data processing (included volume). Void volume fractions containing aggregated complexes (not shown in A) were not included in data processing. A reference protein standard is presented in the far left lane (M).
The ensemble of structural models described here in Section 4.3.3 were compared with experimental SAXS data by calculating the theoretical scattering of each structure within the ensemble. One thousand theoretical $I(Q)$ patterns were generated and averaged to generate the mean theoretical scattering pattern, which was then compared with experimentally acquired SAXS patterns. Representative “open/loose” and “closed/compact” models were also generated for comparison with experimental data (Figure 4.8).

![A graph showing the comparison of experimental and theoretical SAXS patterns](image)

**Figure 4.8**: Comparison of experimental SEC-SAXS pattern of $\beta_2\theta_{L\theta}$ in the presence and absence of DNA/dATP/ddCTP with the theoretical $I(Q)$ curves of both ‘loose’ and ‘compact’ models of $\beta_2\theta_{L\theta}$.

### 4.3.6 Crystallography trials

Small needle-like crystals (most likely of salt) of a maximum length of 150–200 μM were grown in a number of conditions, generally having a wide pH range of 4–8, containing PEG3350 or PEG4000 at a concentration of 12–20% (w/v), and a salt (commonly either sodium malonate, ammonium sulphate or diammonium tartrate). All crystal growth seen was consistent with the needle-like morphologies viewed in Figure 4.9. No crystals were
observed in any condition in a time period shorter than approximately 20 days.

A number of attempts were made over the following 12 months to replicate crystal growth using a larger volume scale in the hope of obtaining diffraction quality crystals of increased size. Sitting-drop vapour diffusion was performed using protein sample volumes of 1–5 μL and numerous conditions were screened based on positive ‘hits’ from samples crystallised at the Facility. Unfortunately, no further crystal growth of was observed in any condition tested on a larger scale and no crystal suitable for diffraction have yet been obtained.

![Crystal growth images](image)

**Figure 4.9: Crystal growth from samples containing the βαLεLθ complex in selected conditions:** (left panel) 0.2 M diammonium tartrate pH 7.5, 20% (w/v) PEG 3350 at 20°C after 37 days and (right panel) 0.2 M sodium malonate pH 5, 20% (w/v) PEG 3350 at 20°C after 37 days.
4.4 Discussion

4.4.1 Protein production and purification procedures

4.4.1.1 Protein subunit and complex purification

Both the double-labelled and unlabelled samples of the ε_{193} protein showed significant contamination following purification by ion-exchange chromatography. However, contamination was efficiently removed during isolation of the double-labelled \textsuperscript{13}C,\textsuperscript{15}N-ε_{193};θ and ε_{193};\textsuperscript{13}C\textsuperscript{15}N-θ complexes. The β_{2}, ε_{L} and θ subunits were all purified in large amounts following well-established protocols (Oakley et al., 2003; Keniry et al., 2006; Jergic et al., 2013) and each was free from any evident contamination. The α_{L}ε_{L}θ Pol III core complex was also uncomplicated to isolate by utilising a commonly used ion-exchange chromatography protocol, refined by Dr Slobodan Jergic (Jergic et al., 2013) The β_{2}α_{L}ε_{L}θ complex was successfully co-purified in an apparently stoichiometrically correct ratio of subunits using ion-exchange chromatography, although the elution profiles observed during chromatography (not shown) were often convoluted, contributing to low yields (particularly during the large-scale purification of the complex) that could likely be improved by further optimisation of chromatography conditions.

4.4.1.3 Protein-protein affinity purification of the α_{L} subunit

A novel protein-protein affinity chromatographic method was developed for purification of the α_{L} subunit by exploiting the stable interaction between the polymerase and the C-terminal segment of the τ subunit. This strategy was remarkably successful at improving both the yield and purity of the final purified α_{L} protein product. A major source of
contamination by proteolytic fragments was removed during this chromatography step as only those polymerase subunits that were unproteolysed from their C-termini were selectively immobilised by the SA-\(\tau_{16}\) resin. The stable interaction between the polymerase and the conjugated resin also ensured other contaminants could be thoroughly removed from the sample prior to elution of the \(\alpha\)L protein using a high concentration of MgCl\(_2\). A typical production of a polymerase or mutant variant subunit using a ‘traditional’ methodology (as described in Section 4.2.3.8 for the purification of the \(\alpha_{E612K}\) protein) usually resulted in a final yield <10 mg and often with significant visible proteolytic contamination. By comparison, yields approaching 50 mg and with greatly improved purity have been reported by members of the Dixon group when utilising the SA-\(\tau_{16}\) resin.

Initial attempts to purify polymerase subunits (the wild-type \(\alpha\) subunit and a mutant polymerase termed \(\alpha_{KGB}\) each not required for any experimental work in this Thesis) using the protein-protein affinity methodology resulted in a decrease in polymerase activity compared to a reference sample of wild-type \(\alpha\) when samples were assessed using a strand-displacement (SD) DNA synthesis assay. It was suspected that heavy metal contaminants (up to 200 \(\mu\)M combined Pb, Fe, Cu, Zn and Mn) originating from the high concentration of MgCl\(_2\) (up to 4 M) in the elution buffer was sequestering molecules of DTT and lowering the effective reducing power of the buffer (Krężel \textit{et al}., 2000). Polymerase activity was recovered under the conditions of the SD assay following an increase in the concentration of DTT from 1 mM to 10 mM in elution and dialysis buffers containing high concentrations of MgCl\(_2\).
4.4.2 The εNTD:θ complex appears highly mobile

We investigated whether the catalytic N-terminal domain of the ε subunit or the θ subunit participated in any additional weak interactions with the β-clamp, further to the recognised interaction between the ε CBM and the β₂ subunit. NMR detected no interaction between the first 181 residues of the ε catalytic domain and the β-clamp when the complex was in the presence of equimolar or excess β₂ and it was not possible to further constrain the position of the εNTD within the β₂αεθ complex from these data. A similar set of NMR measurements were recorded at a later date, this time using a sample of the ε₁₉₃:θ complex where the θ subunit was uniformly ¹³C,¹⁵N-labelled. Intriguingly, a number of residues of the θ subunit were observed to interact with the β-clamp in the presence of a 2:1 excess molar ratio of the β subunit. These residues are clustered towards one end of the roughly oblong-shaped θ subunit (Figure 4.6), suggesting this portion of θ – and by implication, the εNTD – is oriented towards the β-clamp at a stage of the replication cycle. Without currently knowing where on the β-clamp these residues interact, no attempt was made to include this interaction in a theoretical structural model to be compared with SAXS measurements of the β₂αεθ complex (Section 4.4.3, below). It is, however, interesting to speculate how the β₂αεθ complex may accommodate both this interaction and the interaction between the ε CBM and the β₂ subunit as the two sites are structurally located far from one another in the εNTD:θ complex. It may be that the εNTD:θ complex is brought into the proximity of DNA by an interaction between θ and the β₂ subunit, laying the exonuclease domain flat against the clamp as the ε CBM maintains contact with β₂. It is also possible that both interactions occur during separate time points in the replication cycle and the θ-β₂ interaction is relatively transient. Locating the site of interaction on β₂ would require further investigation using chemical cross-
linking experiments or solid-state NMR measurements with isotopically labelled $\beta_2$.

### 4.4.3 SAXS validation of the complex model

Analysis of eluted fractions showed that the $\beta_2\alpha\varepsilon\iota\theta$ complex was not DNA-bound in condition B and SEC-SAXS measurements recorded under this condition were of the $\beta_2\alpha\varepsilon\iota\theta$ complex alone. A redesign of the DNA construct may be required before proceeding with further SAXS or crystallography experiments. The small differences between the elution profiles of the $\beta_2\alpha\varepsilon\iota\theta$ complex both with and without DNA (Figure 4.7 A) may be explained by the presence of dATP, ddCTP and MgCl$_2$ in the condition with DNA (condition B), causing a putative conformational change in this complex.

The experimental SEC-SAXS patterns offer validation of the general features of the model presented in Section 4.3.4, showing agreement between overall dimensions measured in experimental data and that of a ‘closed’ or ‘compact’ model of $\beta_2\alpha\varepsilon\iota\theta$. At $Q$-values in the range of 0.07–0.12 Å$^{-1}$, the experimental SEC-SAXS patterns showed better agreement with the averaged calculated theoretical SAXS pattern of 1000 $\beta_2\alpha\varepsilon\iota\theta$ models than that of the ‘compact’ $\beta_2\alpha\varepsilon\iota\theta$ model alone, suggesting that the $\varepsilon$NTD:0 domain is not restrained in a single conformation within the complex but rather is largely free to rotate at the $\varepsilon$ CBM attachment point on the $\beta$-clamp.
4.4.4 Crystallography trials

The growth of protein crystals was pursued following the development of a method to co-purify a stable complex of $\beta_2\alpha_L\varepsilon_L\theta$ on a large scale. A sample of the complex screened at the C3 Facility in Melbourne, Australia, surprisingly gave growth of crystals under a number of conditions following a time period of at least 20 days. We were sceptical of the make-up of the crystals observed due to the length of time taken to grow, the high number of conditions crystals formed in and the known low probability of growing crystals of a large multi-protein complex with a highly flexible interdomain protein region and relatively weak inter-protein contacts (between $\alpha_L/\varepsilon_L$ and $\beta_2$). It is highly likely that the crystals grown were of salt, as many of the conditions in which crystals were observed in were simple, containing only PEG, a buffering chemical and a salt. The $\beta_2$ subunit is readily crystallisable and it is also possible that crystals were of the $\beta_2$ subunit alone. Although a number of attempts were made over the following year to scale-up crystal growth to obtain a suitable crystal for diffraction, no further crystals were observed under any condition tested.

We also investigated producing a protein construct comprising the full-length $\alpha_L$ subunit linked to the C-terminal region of the $\varepsilon_L$ subunit by the same 9 amino acid flexible linker used for the construction of linked proteins described in Chapter 3. The construct was designed to begin at residue 180 of the $\varepsilon_L$ chain and thus incorporate both of the stronger CBMs into the $\varepsilon$ and $\alpha$ subunits, while removing the globular $\varepsilon$NTD thought to be highly flexible within the context of the core:clamp complex. It was anticipated that a sample of this construct co-purified with the $\beta_2$ subunit may be more amenable to crystallography than the $\beta_2\alpha_L\varepsilon_L\theta$ complex. Unfortunately, this construct appears to be highly toxic to cell growth and sequencing of clones showed the insertion of stop condons at various positions in the portion of
the modified gene encoding the αL protein. No further attempts were made to pursue the production and purification of this protein construct.

4.4.5 Structural modelling advancements and future research directions

The Lamers group has recently made significant progress in modelling the structure of the *E. coli* β2αεθ complex, both with and without a primer-template DNA and also inclusive of the C-terminal segment of the τ subunit that binds α (τC). Chemical cross-linking experiments provided distance restraints that enabled the informed building of a structural model of the polymerase, clamp and exonuclease and revealed the novel interaction between the ε and β2 subunits, contemporaneously to work published by the Dixon group (Jergic *et al.*, 2013; Toste Rêgo *et al.*, 2013).

The Lamers group has also published experimental results using the technique of cryo-electron microscopy (cryoEM) to build a model of the β2αετC complex in both a DNA-bound and DNA-free state (Fernandez-Leiro *et al.*, 2015; Figure 4.10). Interestingly, these models reveal a novel interaction between the C-terminal OB-fold domain of the α subunit and the β2 subunit that, together with the CBMs located in the α and ε subunits, forms a third point of contact between the clamp and the Pol III core. The C-terminal OB-fold and τ-binding regions of the α subunit are also shown to undergo a large structural change on binding to DNA, moving more than 30 Å towards the clamp. The OB-fold of the α subunit has previously been shown to bind isolated ssDNA and was proposed to act as a sensor that regulates the affinity of the Pol III core to the clamp (Georgescu *et al.*, 2009a). The EM model of the DNA-bound complex, however, positions the OB-fold of the α subunit ~40 Å away from the ssDNA portion of the DNA template,
suggesting that regulation of clamp interaction with the Pol III core does not occur through direct binding of the OB-fold to ssDNA.

![Figure 4.10: Cryo-electron microscopy (cryo-EM) models of the β_{2αεC} complex. (A) Surface representation of the complex in (left to right) a DNA-free, DNA-bound and DNA-bound without tail states. (B) Individual crystal structures of the subunit components were fitted into the cryo-EM map. The β_{2} subunit is coloured green, the ε subunit in yellow and the α subunit in orange (residues 2–959), red (the OB-fold) and blue (ε-binding domain). DNA is coloured black in the centre and right panels. Figure from Fernandez-Leiro et al., 2015.](image)

When the complex is in both a DNA-bound and DNA-free state the εNTD is wedged between the clamp and N-terminal PHP domain of α, maintaining contact with the β_{2} subunit through the εCBM, although rotating slightly as the complexes closes. The position of the εNTD during proofreading is still unclear. Modelling suggests a distance of over 70 Å between the polymerase and exonuclease active sites (Jergic et al., 2013), indicating a large conformational change must occur to transfer the DNA from one site to the other when shifting between polymerisation and proofreading modes. This may be accomplished by a repositioning of the DNA within the complex, a significant protein conformational change, or
both. The long and flexible Q-linker region between the two domains of the ε subunit suggests an important functionality and it is reasonable to propose that this distinctive domain arrangement may be vital to the transfer of the catalytic N-terminal domain of the ε subunit to the DNA template when Pol III HE switches to a proofreading mode. How this occurs, however, is still an open question.

The Dixon group and collaborators are pursuing the use of single-molecule techniques to help uncover novel protein-protein interactions within the replisome and understand the dynamic rearrangement of subunits within the Pol III core-β2 complex during the transition from synthesis to editing. Work I have performed towards the progress of a FRET (Förster resonance energy transfer) assay using a novel peptide-tag for the fluorescent labelling of proteins is described in Chapter 5.
Chapter 5

A strategy for fluorescent labelling of proteins
5.1 Introduction

The *E. coli* β2αεθ complex is a dynamic protein assembly that seems to undergo multiple conformational rearrangements when engaged with a template DNA molecule to facilitate the transition between alternate modes of enzymatic activity, including DNA synthesis and proofreading. I have shown in Chapters 3 and 4 that the techniques of protein crystallography and small-angle X-ray scattering, combined with data from several other biochemical experimental sources, can be used to build and validate an informed static model of this complex (as published in Ozawa et al., 2013). Current research is expanding on our understanding of the β2αεθ complex and providing new structural insights into the molecular mechanisms that enable the efficient and accurate synthesis of new DNA (Fernandez-Leiro et al., 2015).

Static molecular structures provide researchers with valuable information from which new hypotheses can be generated regarding the functional dynamics of a protein or protein assembly. Testing these hypotheses, however, often requires a different set of experimental approaches, including single-molecule techniques. In this experimental chapter, I outline work that I have performed towards refining a protein fluorescent labelling procedure, the principles of which are outlined in Section 5.1.1. We aimed to fluorescently label the α and θ proteins, as well as three variations of the ε exonuclease subunit: the wild-type ε subunit (εWT), the εL subunit containing a stronger CBM (described in Section 4.1) and the εQ subunit, where the wild-type CBM had been weakened by a Q→A mutation at the first a.a. position in the CBM motif (*i.e.* QTSMAF→ATSMAF; described in Jergic et al., 2013). These fluorescently labelled proteins were intended for use in the laboratory of Professor Jong-Bong Lee (POSTECH, Korea) using the technique of Förster resonance energy transfer (FRET). A
brief outline of the experimental technique of FRET and how this approach can utilise fluorescently tagged proteins to study the β2αεθ assembly is outlined in Section 5.1.2.

5.1.1 Use of the formylglycine generating for fluorescent labelling

The formylglycine generating enzyme (FGE) recognises the polypeptide consensus sequence CxPxR (where x is any amino acid) and specifically oxidises the cysteine within the sequence to the aldehyde formylglycine (Figure 5.1; Carlson et al., 2008; Wu et al., 2009). The FGE recognition sequence can be placed at the N- or C-terminus of a protein, or at an exposed site within a polypeptide sequence. Conversion can occur either in vivo by co-expression of a target protein with the FGE protein (Rabuka et al. 2012) or in vitro by reacting a purified protein bearing the recognition sequence with a purified sample of FGE (Holder et al., 2015). An aldehyde group is an example of a ‘chemical handle’ that can be selectively reacted with an α-nucleophile to form a site-specific conjugate with a chemical cargo, such as a fluorescent dye.

![Diagram](image_url)

**Figure 5.1:** The formylglycine generating enzyme (FGE) recognises the amino-acid consensus sequence LCTPSR and catalyses the oxidation of the cysteine thiol to an aldehyde to generate formylglycine. In vitro treatment of a protein containing this consensus sequence with FGE generates formylglycine, which can be conjugated with a reactive fluorescent dye.
In the current work, I describe the construction of plasmids coding for the proteins \( \alpha_{\text{FGE}}, \theta_{\text{FGE}}, \Delta5_{\text{WT FGE}}, \Delta5_{\text{EL FGE}} \) and \( \Delta5_{\text{EQ FGE}} \), where the FGE amino acid recognition sequence ‘LCTPSR’ (Rabuka et al. 2012) was incorporated at the N-terminus of each and \( \Delta5 \) indicates the first five amino acids in the \( \varepsilon \) sequence were removed. The overproduction and purification of each protein is investigated and the overproduction and purification of FGE protein from \textit{Mycobacterium tuberculosis} is also described. An enzymatic conversion and labelling protocol (described in Section 5.2.9) was performed on the \( \alpha_{\text{FGE}} \) and \( \theta_{\text{FGE}} \) proteins to label both proteins with a fluorescent dye.

The fluorescent dyes used in this project were obtained from Redwood Bioscience and are modified forms of the Alexa Fluor 555 (AF555) and Alexa Fluor 647 (AF647) dyes (ThermoFisher) that contain a reactive \( \beta \)-arylethylamine. The Pictet-Spengler reaction (Figure 5.2; reviewed by Stöckigt et al., 2011) is a condensation reaction between a \( \beta \)-arylethylamine and an aldehyde or ketone. The aldehyde functional group of formylglycine can selectively react with a functionalised dye molecule to covalently bind to a target protein which has been enzymatically modified by FGE. Only the AF555 dye was utilised in labelling experiments in the work described here.

---

**Figure 5.2: The Pictet-Spengler reaction.** A \( \beta \)-arylethylamine undergoes ring closure in a condensation reaction with an aldehyde or a ketone.
5.1.2 Single-molecule Förster resonance energy transfer (FRET)

Single-molecule techniques are used to study the properties of individual molecules, avoiding the ensemble averaging inherent in large populations. Researchers make use of a number of single-molecule experimental approaches to study protein-protein and protein-DNA complexes, including manipulation of molecules by force and tracking by fluorescence.

Förster resonance energy transfer (FRET) describes a physical mechanism by which energy is transferred nonradiatively from an excited donor fluorophore to an acceptor fluorophore. This process is distance dependant and changes in the fluorescence of both the donor and acceptor fluorophore report on changes in relative distance between the pair over a range of 10–100 Å; a range particularly suitable for studying pairwise distances between different protein subunits of a multi-protein complex (reviewed by Sekar and Periasamy et al., 2003).

The laboratory group of Professor Jong-Bong Lee (POSTECH, Korea) uses the technique of single-molecule FRET to study the dynamics of protein-protein and protein-DNA interactions (e.g., see Cho et al., 2014; Lee et al., 2014). By fluorescently labelling individual subunits in the β2αεθ complex, the Dixon and Lee groups aim to study the dynamic rearrangement of subunits during transition between polymerisation and proofreading activities on a template DNA using FRET. Strengthening of the ε subunit CBM has been shown to stabilise the β2αεθ complex on a template DNA during processive synthesis (Jergic et al., 2013) and it would be of particular interest to incorporate the three ε variants, εWT, εL and εQ, into a single-molecule FRET proofreading assay to assess how the interaction between the ε and β2 subunits contributes to exonuclease activity and to study the transition between polymerisation and exonuclease modes.
5.2 Materials and Methods

5.2.1 Cloning overview

The a.a. sequence, ‘MSLCTPSR’ (here termed FGE8) contains the FGE recognition sequence LCTPSR and was placed at the N-terminus of the wild-type α protein by introduction of the nucleotide sequence 5’-ATGAGCCTGTGCACCCCCGTCTAGA (here termed FGE24; XbaI site indicated) upstream of the dnaE gene, using overlap extension PCR with partially complementary oligonucleotide primers and standard PCR conditions (Section 2.2.1.1).

The FGE24 sequence was also placed upstream of the genes coding for the θ, ε, εl and εQ proteins by first modifying the T7-promoter plasmid pETMCSI (Section 2.1.2) by introducing the FGE24 sequence between Ndel and EcoRI restriction sites. Then, by making use of an XbaI restriction site present in FGE24, relevant genes already within a pETMCSI-type vector were PCR amplified to include the XbaI restriction site at the 5’-end, then easily inserted between XbaI and EcoRI restriction sites in the new plasmid vector. Cloning details for all constructs are detailed below in Sections 5.2.3–5.2.5.

5.2.2 Oligonucleotide primers

pBOBftag: 5’-AAAAAAAAAAÇATATGAGCCTGTGCACCCCCGTCTAGATCTGAACCACGTTTCGTACAC-3’ (58-mer, NdeI site indicated by a dashed-line, XbaI site indicate by a solid underline)

pND517rev: 5’-CGGCACGATAGTCATGC-3’ (18-mer)

XbaIinsertF: 5’-TATGAGCCTGTGCACCCCCGTCTAGATCG-3’ (28-mer, XbaI site indicated by a solid underline and ‘sticky ends’ by a dashed underline)

XbaIinsertR: 5’-AATTCGATCTAGACGGGGTGCACAGGCTCA-3’ (30-mer, XbaI site indicated by a solid underline, ‘sticky ends’ indicated by a dashed underline)
XbaIEpsilon: 5’-AAAAATCTAGAACACGCCAGATCGTTCTCGATAC-3’
(34-mer, XbaI site indicated).

XbaITheta: 5’-AAAAATCTAGACTGAAGAATCTGGCTAAACTGG-3’
(33-mer, XbaI site indicated).

pET6: 5’-GGAGCCACTATCGACTACGC-3’ (20-mer)

5.2.3 Construction of plasmid for FGE8-tagged α

The plasmid pSJ2132 contains the full-length wild-type dnaE gene (coding for the α subunit) between NdeI and NcoI restriction sites in the plasmid vector pBOB1 (Section 2.1.2; S. Jergic, unpublished work). The dnaE gene from plasmid pSJ2132 was PCR amplified with primer oligonucleotides pBOBftag and pET6 to generate an approximate 3.7 kb fragment, which was then restriction digested with Ndel and XhoI to produce a DNA fragment comprising FGE24 directly upstream of the first 499 nucleotides of the dnaE gene. The plasmid pSJ2132 was separately restriction digested with XhoI and NcoI and the DNA fragment beginning from nucleotide position 500 of the dnaE gene and ending immediately downstream of the dnaE gene was gel purified. A separate preparation of pSJ2132 was restriction digested with Ndel and NcoI and the digested pBOB1 vector (with the dnaE gene excised) was gel purified. The three fragments were ligated concurrently and plasmids containing the FGE24 sequence upstream of the full-length dnaE gene were identified by colony PCR and stored as plasmid pNH2153.

5.2.4 Preparation of a FGE24/pETMCSI cloning vector

A purified sample of plasmid vector pETMCSI (Section 2.1.2) was generated by restriction digestion of the plasmid pNH2145 (containing an unrelated gene) with Ndel and EcoRI, followed by gel purification (Section
2.2.1.7) of the 4.7 kb vector fragment following agarose gel electrophoresis. The oligonucleotides XbalinsertF and XbalinsertR were mixed together at a concentration of 10 μM each, then heated at 95°C for 10 min in a heating block. Samples were allowed to cool slowly overnight to promote selective annealing of complementary oligonucleotide fragments. Each oligonucleotide was designed, once annealed, to create short nucleotide overhangs or ‘sticky ends’ at each end of the dsDNA fragment to facilitate ligation with the Ndel and EcoRI restriction digested pETMCSI plasmid vector.

Subsequently, the short dsDNA insert was mixed with the restriction digested pETMCSI vector in a molar ratio of 3:1 and ligated. Following transformation and colony PCR, plasmids of interest were retained and relevant regions of DNA were sequenced.

While pursuing further cloning experiments requiring this construct, two previously unconsidered sequence complications were discovered: firstly, an Xbal site was already present in the pETMCSI vector 55 bp upstream of the site newly introduced and secondly, the inserted nucleotide sequence contained within it the recognition sequence (GATC) for deoxyadenosine methylase, which wholly blocks the action of Xbal by methylation of adenine (Geier and Modrich, 1979). Two successive rounds of quick-change PCR using mismatched primers to introduce a single-point mutation at each site rectified these problems. DNA sequencing confirmed the integrity of the final construct and the plasmid was propagated and stored as pNH2158 (see Figure 5.3 below for inserted nucleotide sequence and corrections made).
5.2.5 Construction of plasmids for FGE8-tagged $\theta$, $\text{ewt}$, $\epsilon_{L}$ and $\epsilon_{Q}$

The $\text{holE}$ gene (coding for the wild-type $\theta$ protein) was amplified by PCR from plasmid pCM869 (Keniry et al., 2000) using the primers XbaITheta and pET6 to generate a DNA fragment flanked upstream by a XbaI restriction site and containing an EcoRI restriction site downstream of the gene. The PCR product was digested with XbaI and EcoRI and ligated between corresponding restriction sites in the plasmid pNH2158 (Section 5.2.4) to create the plasmid pNH2162.

The plasmids pSH1017 (Hamdan et al., 2002), pSJ1445 and pSJ1446 (Jergic et al., 2013) contain variations of the $\text{dnaQ}$ gene that code for the proteins $\text{ewt}$, $\epsilon_{Q}$ and $\epsilon_{L}$, respectively (discussed in Section 5.1). Each gene was similarly PCR amplified using primers XbaIEpsilon and pET6 to generate DNA fragments that were flanked by an XbaI restriction site upstream of the gene and an EcoRI restriction site downstream. The sequence of the primer XbaIEpsilon was designed to amplify each gene beginning at nucleotide position 16 of $\text{dnaQ}$ (or variations) to remove the nucleotide sequence coding for the first 5 amino acids of each $\epsilon$ protein that are known to be flexible in solution (Pintacuda et al., 2006). It was considered desirable to minimise flexibility of the N-terminus of $\epsilon$ to enable easier interpretation of subsequent FRET data. Following restriction digestion with XbaI and EcoRI, each PCR
product was ligated between corresponding restriction sites in the plasmid pNH2158 to create the plasmids pNH2159, pNH2160 and pNH2161, coding for the proteins termed here as Δ5εWTFGE, Δ5εLFGE, and Δ5εQFGE, respectively.

5.2.6 Production and purification of αFGE and θFGE

The proteins αFGE and θFGE were overproduced and purified using comparable method to those used in Sections 4.2.4.3 and 4.2.3.1 for the production and purification of αL and θ proteins, respectively. Any relevant differences in protocols and yields obtained for each are outlined below. SDS-PAGE analysis of the final purified protein products can be viewed in Figure 5.4.

The *E. coli* strain BL21(λDE3)/pLysS/pNH2153 (coding for the αFGE protein) was grown in 6 × 1 L of LBTA/C medium and protein overproduction induced by IPTG. Following expression, 18 g of cells were collected. Cells were lysed and purified as described Section 4.2.3.4, yielding a total of ~23 mg of purified αFGE that was aliquoted and stored at −80°C. The molecular weight of αFGE in 150 mM ammonium acetate was determined by ESI-MS as 130 518.8 Da and may be compared to the theoretical calculated value of 130 518.2 Da (N-terminal methionine removed).

The *E. coli* strain BL21(λDE3)/pLysS/pNH2162 (coding for the θFGE protein) was grown in 9 L of LBTA/C medium and protein overproduction was induced by IPTG. Following expression, 21 g of cells were collected. Cells were lysed and the θFGE protein was purified as described in Section 4.2.3.1, yielding a total of ~10 mg of purified θFGE that was aliquoted and stored at −80°C. The molecular weight of θFGE in 0.1% formic acid was determined by ESI-MS as 9460.1 Da and may be compared to the theoretical calculated value of 9459.9 Da (N-terminal methionine removed).
5.2.7 Production and purification of formylglycine generating enzyme (FGE)

The plasmid that directs production of the *Mycobacterium tuberculosis* FGE sulfatase was obtained from Professor Jong-Bong Lee (POSTECH, Korea). The gene is contained within a plasmid vector derived from pBR322 (Bolivar *et al.*, 1977) and enables transcription of a gene downstream of the phage T7 φ10 promotor and generates translated proteins that are N-terminally tagged with the a.a sequence ‘MGSSHHHHHHHHSSGLVPRGSH’; incorporating a His<sub>6</sub> purification tag and thrombin cleavage site. For clarity, the tagged version of the *M. tuberculosis* FGE protein described here is simply referred to as ‘FGE’. Upon receipt, the plasmid was stored as pNH2176. A protocol for the production and purification the FGE protein was obtained from the Lee group and is outlined here, with some modifications.
The *E. coli* strain BL21(λDE3)/pLysS/pNH2176 was grown in 1 L of LB autoinduction medium (Section 2.1.4.3) containing ampicillin (100 mg/L) and chloramphenicol (35 mg/L) at 25°C until an OD$_{600}$ of 0.2 was reached. The culture was then rapidly cooled to 16°C and protein production was induced over a further period of 90 h as described by Studier (2005). Cells were harvested by centrifugation (11 000 × g, 7 min), frozen in liquid nitrogen and stored at −80°C.

Frozen cell pellets (24 g in total) were resuspended in 250 mL of lysis buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM DTT, 10% sucrose, 100 mM NaCl) before being passed twice through a French press (12 000 psi) to lyse cells. The lysate was clarified by centrifugation (11 000 × g; 30 min) to yield soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.40 g/mL) and stirring for 60 min were collected by centrifugation (11 000 × g; 60 min) and dissolved in 70 mL of buffer $A_{\text{FGE}}$ (25 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 10% glycerol) + 160 mM NaCl. The solution was dialysed against 2 L of the same buffer overnight, to yield Fraction II.

Fraction II was divided into two aliquots of equal volume and each separately applied by gravity (at ~1 mL/min) onto a column (2.5 × 16 cm) of Toyopearl DEAE-650M resin that had been equilibrated with buffer $A_{\text{FGE}}$ + 160 mM NaCl. Fractions containing proteins that did not bind to the column were pooled and dialysed against 2 L buffer $B_{\text{FGE}}$ (25 mM Tris.HCl, pH 7.6, 0.5 mM DTT, 10% glycerol, 300 mM NaCl) + 20 mM imidazole to yield Fraction III (150 mL).

Fraction III was analysed by SDS-PAGE and estimated to contain a significant amount of FGE protein (>200 mg). Subsequently, approximately 100 mL of Fraction III was immediately frozen in liquid nitrogen and stored at −80°C as a partially purified sample for future use. The remaining 50 mL of sample was further divided into three aliquots and each then passed
separately at 1 mL/min through a 1 mL His-trap column that had been pre-equilibrated with buffer BFGE + 20 mM imidazole. On each occasion, after loading of the sample the column was washed with a further 10 mL of the same buffer and bound proteins were eluted using a linear gradient (10 mL) of 20–500 mM imidazole in buffer BFGE. Proteins eluted as a single peak after application of the imidazole gradient and were collected, pooled and dialysed against 2 L of buffer SFGE (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM DTT, 20% glycerol, 150 mM NaCl) to give Fraction IV (13 mL).

Fraction IV containing ~65 mg of FGE (5 mg/mL) was aliquoted then frozen in liquid nitrogen and stored at –80°C. The molecular weight of FGE in 0.1% formic acid was determined by ESI-MS as 34 774.5 Da, which may be compared to the theoretical calculated value of 34 775.8 Da (N-terminal methionine removed). SDS-PAGE analysis of the final purified protein product can be viewed in Figure 5.4.

5.2.8 Production and purification of Δ5εWTFGE, Δ5εLFGE, and Δ5εQFGE

The three variations of FGE8-tagged ε protein (construction of plasmids for each as described in Section 5.2.5) were overproduced using similar procedures to that outlined in Section 4.2.3.6 for the production of the εL subunit. In comparison to εL, the expression of each was significantly reduced (not shown) and each appeared highly toxic to cell growth, strains reaching an OD₆₀₀ < 2 after a 72 h autoinduction. By comparison, a OD₆₀₀ > 10 is typically achieved when producing the full length εWT, εL or εQ subunits. Purification of each protein was pursued using the refolding procedures described in Section 4.2.3.6. However, on each occasion, the procedure was unsuccessful at refolding any insoluble Δ5εWTFGE, Δ5εLFGE or Δ5εQFGE protein using the prescribed conditions. No further attempts to overexpress or purify
these proteins were made. Presumably, deletion of five amino acids from the N-terminus of ε is sufficient to disrupt normal refolding of the protein.

5.2.9 Enzymatic conversion and fluorescent labelling protocol

Fluorescent labelling of the proteins αFGE and θFGE (Section 5.2.6) was investigated following a protocol obtained from the Lee group (POSTECH, Korea) and based on that described by Holder et al. (2015). The procedure can be divided into four steps: In step 1, a target protein (αFGE or θFGE) was mixed in a 1:1 molar ratio with the FGE enzyme and dialysed in 1 L of conversion buffer (50 mM Tris.HCl pH 8.0, 100 mM NaCl, 20 mM arginine, 0.25 mM DTT) at 4°C for 48 h. A volume of 6 mL of the αFGE protein at a concentration of 1 mg/mL was mixed 350 μL of the FGE protein at 5 mg/mL. A volume of 5 mL of the θFGE protein at a concentration of 0.4 mg/mL was mixed with 1.5 mL of the FGE protein at 5 mg/mL. In step 2, the sample was removed from the dialysis tubing and imidazole was added to a concentration of 20 mM before it was passed through a 1 mL His-Trap column (GE Healthcare) with washing with conversion buffer. A280 values of the eluate were monitored and flow-through fractions containing the target protein were pooled. In step 3, the pooled sample was dialysed into 1 L of labelling buffer (100 mM potassium phosphate pH 6.4, 300 mM NaCl and 0.25 mM DTT) and dialysed overnight then concentrated to a small volume (< 1 mL) using a centrifugal concentrator. The sample was then mixed with an appropriate AF555 fluorescent dye (stock sample dissolved in H2O) to a final dye concentration of 0.5 mM and incubated at 4°C for 48 h. In step 4, to remove excess dye, the sample was then passed through a column of Sephadex G-25 resin (1.5 × 10 cm; GE Healthcare) that had been equilibrated in buffer Dy (50 mM Tris.HCl pH 7.6, 0.5 mM EDTA, 1 mM DTT, 5% glycerol, 150 mM NaCl). The sample was eluted from the column in buffer Dy and the eluate was monitored by eye for the passage of excess dye and by
A<sub>280</sub> reading for the passage of protein through the column. Fractions corresponding to a significant A<sub>280</sub> reading were collected and stored at –80°C.

5.2.10 Establishing the efficiency of conversion and labelling

Following the protocol described in Section 5.2.9, protein samples were assessed visually for fluorescence following SDS-PAGE (Section 2.2.2.4; prior to staining with Coomassie blue) against a sample of AF555 dye alone at an equivalent concentration to that expected at 100% and 10% labelling efficiency. Each gel was subsequently stained with Coomassie blue (as described in Section 2.2.2.4) to evaluate the purity of the αFGE or θFGE proteins.

Additional trouble-shooting experiments were performed by following step 1 (Section 5.2.9), with alterations. A small volume of the θFGE protein (500 µL) was mixed with the FGE protein at molar ratios of 10:1, 1:1 and 1:2 θFGE to FGE, respectively, and aliquots of each were incubated by dialysis in conversion buffer as described in 5.2.9 for time periods of 2 h, 12 h and 24 h. To estimate the efficiency of enzymatic conversion of cysteine to formlyglycine by the FGE protein, the θFGE protein was assessed for accurate mass determination following step 1 (Section 5.2.9) by ESI-MS in 0.1% formic acid. The αFGE protein was not examined for conversion by the FGE protein as it is not readily detected by ESI-MS in formic acid and the large molecular weight of the protein (~130 kDa) would also make it unfeasible to accurately detect a molecular weight change of 18 Da, corresponding to the cysteine to formylglycine conversion.
5.3 Results

After purification the $\alpha_{FGE}$ and $\theta_{FGE}$ proteins, it was first necessary to establish the enzymatic activity of $\alpha_{FGE}$. Efforts to fluorescently label both proteins were made following the protocol outlined in 5.2.9; however, both attempts were unsuccessful. The efficiency of enzymatic conversion by the FGE protein was subsequently investigated by ESI-MS. Results are outlined below in Sections 5.3.1 and 5.3.2.

5.3.1 Assessment of polymerase activity

The $\alpha_{FGE}$ protein was purified using protein-protein affinity chromatography, as described in Section 4.2.4.3 for the purification of the $\alpha_t$ protein. Next, a strand-displacement (SD) synthesis assay was performed to assess the comparative activity of the $\alpha_{FGE}$ protein produced against a sample of wild-type $\alpha$ polymerase produced using a well-established ion-exchange chromatography protocol (described in Jergic et al., 2013; Figure 5.5).

![Figure 5.5: Strand displacement (SD) DNA synthesis shows the activity of $\alpha_{FGE}$ and $\alpha_{WT}$ polymerase subunits following purification. Extension on a primed M13 ssDNA template shows SD as larger than unit-length DNA products. Assessing the quantity of high molecular weight products in each lane establishes the relative efficiency of synthesis between $\alpha_{FGE}$ purified using a novel protein-protein affinity methodology as compared to a sample of $\alpha_{WT}$ purified using ‘traditional’ methods. Both samples displayed comparable activity. Gels were stained with SYBR gold nucleic acid stain (Invitrogen)](image-url)

Figure 5.5: Strand displacement (SD) DNA synthesis shows the activity of $\alpha_{FGE}$ and $\alpha_{WT}$ polymerase subunits following purification. Extension on a primed M13 ssDNA template shows SD as larger than unit-length DNA products. Assessing the quantity of high molecular weight products in each lane establishes the relative efficiency of synthesis between $\alpha_{FGE}$ purified using a novel protein-protein affinity methodology as compared to a sample of $\alpha_{WT}$ purified using ‘traditional’ methods. Both samples displayed comparable activity. Gels were stained with SYBR gold nucleic acid stain (Invitrogen)
The assay was performed as described in Section 4.2.8, using a primed single-stranded phage M13 DNA as a template for replication. Under the conditions of this assay, the αFGE protein displayed enzymatic activity comparable to the wild-type α polymerase.

5.3.2 Assessing the efficiency of conversion and troubleshooting

Samples of both the αFGE and θFGE protein displayed no visible fluorescence when assessed by SDS-PAGE, prior to staining with Coomassie blue (Section 5.2.10). In contrast, samples of AF555 dye alone at a 100% and 10% concentration relative to that which would be expected in a wholly labelled sample of protein showed bright fluorescence (not shown). This demonstrated that neither the αFGE nor θFGE proteins were labelled to any appreciable level (well below 10%) following the protocol described in Section 5.2.9.

We reasoned that poor labelling efficiency may be a consequence of low enzymatic conversion by the FGE protein. To test this, the θFGE protein was incubated with different molar ratios of FGE enzyme and for varying lengths of time before being assessed for accurate mass determination by ESI-MS in 0.1% formic acid (Section 5.2.10). The conversion of a sulfhydryl group to an aldehyde results in a decrease in molecular weight of 18 Da and in all spectra recorded under these conditions, 10–30% of species observed corresponded to the MW of converted θFGE (a sample spectrum is shown in Figure 5.6). No trend was identified between the ratio of FGE used in the reaction or length of time samples were incubated. We further speculated that a fraction of converted species may contain a hydrated aldehyde, increasing the molecular weight by 18 Da and possibly disguising the level of enzymatic conversion. It remains problematic to accurately determine the efficiency of conversion by the FGE enzyme. At this point, my contribution to this project concluded.
5.4 Discussion and current status of project

Work on this project is currently ongoing and significant progress has been made by members of the Dixon Laboratory. Dr Zhi-Qiang Xu has since generated plasmid constructs coding for full-length εWTGE, εLFGE, and εQFGE proteins, where the first five amino acids in the ε were included. Dr Slobodan Jergic has successfully expressed and purified the εWTGE protein using the refolding methodology described in Section 4.2.3.6, indicating that the N-terminal five residues are critical for correct folding of the ε protein. Dr Jergic pursued labelling of the εWTGE protein by following the protocol outlined in Section 5.2.9 with an extended incubation time of the protein with fluorescent dye (step 3) from 48 h (2 days) to 6 days. Increasing the incubation period with the dye resulted in a labelling efficiency of the εWTGE protein of approximately 15–20%, indicating that the current labelling conditions are inefficient and could be improved by optimisation of the
labelling protocol and/or FGE purification procedure. Recent studies by Holder et al. (2009) have shown that the FGE enzymes from *Streptomyces coelicolor* and *Homo sapiens* both contain a copper co-factor that is required for high substrate turnover rates. Purification of the FGE enzyme using a Cu-NTA resin (as opposed to the Ni-NTA resin contained in the His-trap column used in Section 5.2.7) may avoid stripping the enzyme of copper during purification and improve activity. The Dixon group also aims to pursue further improvements to the labelling procedure and investigate methodologies to improve the labelling efficiency of the εLFGE, εQFGE, θFGE and αFGE proteins for use in future fluorescence experiments.
Chapter 6

CONCLUDING REMARKS
Work described in this Thesis has contributed to:

(a) The crystal structure of a linked-protein construct displaying the site of interaction between the α polymerase subunit PHP domain and the ε exonuclease subunit C-terminal domain.

(b) The building of an atomic resolution structural model of the β2αεθ complex of the Pol III HE utilising the crystal structure in (a), additional NMR experiments and a range of previously published biochemical and structural data.

(c) The development of a novel methodology to purify the α polymerase by using a novel protein-protein affinity purification method that exploits the tight interaction between the C-terminal domain of the τ clamp-loader protein the C-terminal region of the α subunit.

(d) The successful co-purification of the β2αLεLθ complex by enhancing the stability of the interaction between the β-clamp and Pol III core by mutation of the clamp-binding motifs found in α and ε.

(e) The validation of the structural model built in (b) by measurement of a co-purified sample of the β2αLεLθ complex by small-angle X-ray scattering.

(f) The development of a strategy to label Pol III core proteins with a fluorescent dye by incorporation of a peptide tag that can be selectively modified by the formylglycine generating enzyme (FGE) to generate a reactive aldehyde group for chemical conjugation with a dye.
In Chapter 3 we investigated the use of a 9 amino acid flexible linker to join the α PHP and ε exonuclease C-terminal (CT) domains into a single polypeptide chain. It was anticipated that by artificially localising the interaction between the two subunits, we could overcome experimental limitations conferred by the insolubility of the ε C-terminal domain alone. This strategy was successful in producing a range of soluble protein constructs comprised of the α PHP domain and differing length segments of the εCTD. A structure was solved of a protein construct beginning from amino acid 200 of the full-length ε subunit and then linked to the α PHP domain through the 9 amino acid flexible linker. Analysis of the structure showed continuous density from residue Ser210 through the remainder of the ε C-terminal domain, with additional density through residues Ile202–Ile205, suggesting that residues Ile202–Ala209 may be transiently bound. These data were consistent with prior experimental NMR experimental data (Ozawa et al., 2013). We are confident that we have assigned the full extent of interaction between the α PHP and ε CT domains and, consequently, the full-length of the flexible inter-domain Q-linking region between the two domains of the ε subunit.

In supplement to these results, I also purified a linked protein comprising the α PHP and ε CT domains in the opposite sequential arrangement to the protein crystallised. This was investigated as a negative control to confirm that the inter-peptide linker had not artificially constrained the interaction between the two protein domains to a non-native site. Native mass spectrometry of this ‘reversed’ protein construct showed the presence of significant domain-swapped dimeric species, suggesting that the interaction displayed in the crystal structure is preferred.
In Chapter 4 we made use of the structural information gathered in Chapter 3 to position the ε subunit within the β2αεθ and generate a structural model of the complex. We used NMR to probe for any undiscovered interaction between the β-clamp and either the ε N-terminal domain or θ subunit in an attempt to further constrain the parameters of the structural model. However, at this stage we were unable to position the globular εNTD:θ domain within the model.

To further structurally investigate the β2αεθ complex, I developed a novel methodology for purification of the α polymerase subunit using protein-protein affinity purification. This approach worked remarkably well and has since enabled researchers in the Dixon group to obtain the polymerase subunit (and mutant variants) in high yield and purity. I co-purified a stabilised complex of β2αεθ where the two clamp-binding interactions between the Pol III core and β-clamp had been strengthened (β2αLεLθ) and utilised the technique of small-angle X-ray scattering to confirm that the overall dimensions of this complex are consistent with the theoretical model. The β2αLεLθ complex was also entered into crystallisation trials. However, so far we have obtained no diffraction quality protein crystals.

In Chapter 5 I investigated a strategy to fluorescently label the proteins of the Pol III core, with the aim of using these for future Förster resonance energy transfer (FRET) experiments in collaboration with the laboratory group of Professor Jong-Bong Lee (POSTECH, Korea). This approach made use of a protein tag that is recognised by the FGE enzyme to site-specifically convert a cysteine to a formylglycine. The reactive aldehyde group can then be chemically conjugated to a modified fluorescent dye. The production and purification of the α and ε proteins incorporating this tag was successful. However, we were unable to successfully fluorescently label
these proteins. It remains unclear if low enzymatic conversion by the FGE enzyme or inefficiencies in the dye conjugation reaction is the source of poor labelling efficiencies. This project is currently ongoing in the Dixon laboratory and Dr Slobodan Jergic has since been successful in labelling a tagged version of the ε protein with an efficiency of ~15–20%.

The experimental challenges involved in identifying the many weak and transient interactions and fundamental characteristics that are essential to the operation of multi-part protein assemblies are driving the development of new experimental techniques and technologies that will provide great benefit to scientific researchers across a variety of fields studying protein-machinery systems. The work outlined in this Thesis has highlighted the utility of using a multi-faceted experimental approach to determining the structure of large protein complexes. The use of SAXS, in particular, was an efficient and effective tool to validate the overall shape and dimensions of an atomic resolution structural model.

Structural investigation of proteins can provide critical insights into the workings of protein assemblies and assist researchers in generating new functional hypotheses. It is anticipated that a combined approach of techniques including SAXS, electron microscopy as well as single-molecule studies will have a growing influence on the direction of future experimental studies of multi-protein complexes, including the Pol III HE.


Cronan, J.E. (2014). *Escherichia coli* as an experimental organism. eLS. DOI: 10.1002/9780470015902.a0002026.pub2


Appendix

Amino acid sequence of the ε subunit of *E. coli* DNA polymerase III (243 a.a.). The N-terminal methionine undergoes cleavage subsequent to production.

MSTAITRQIVLDTETTGMNQIGAHYEGHKIIIEIGAVEVVNRRLTGNNFHV YLKPDRILVDPEAFGVHGIADEFLDKPTFAEVADEFMDYIRGAELVHNA AFDIGFMDYEFSSLKRDIPKTNTFCKVTDSLAVARKMFPGRNSLDALCA RYEIDNSKRTLHGALLDAQILAEVYLAMTGGQTSMAFAMEGETQQQG EATIQRIVRQASKLRVVVFATDEEIAAHEARLDLVQKKGGSCLWRA

Colour Key:

ε N-terminal domain (Ser2–Gly181): β-binding motif (QLSLPL in εL; ALSLPL in εQ); Flexible interdomain region: ε C-terminal domain (Gln196–Ala243).

Final amino acid in the sequence of ε193 coloured pink
Amino acid sequence of the fusion protein $\epsilon_\alpha^{270P}$ (522 a.a).

MSTAITRQIVLDTETTMNQIGAHYEGHKIEIGAVEEVNVRLTNFGHVYLKPDRLVDPEAFGVHGIADEFLLDKPFTAEVADEFMDYIRGAELVIHNAAFDIFMDYEFSSLKKRDIPTNTFCKVTDSLVARKMFPGRKRNSLDAALCARYEIDNSKRTLGALLDAQILAEYVLAMTGGQTSMAFAMEGETQQQGEATIQRIVRQASKLRVFATDEEIAAHEARLDLVQQKGGSCLLWRATRESGSIGMSEPRFVHLRVDYSDSIDMPAKTAPLVKKAAALGMMPALAITDFTNLCGLVKFYGAGHGAGIKPIVGAFNVQCDLLGDELTHLTVLAANNTGYQLTLLISKAYQRYGAAGPIIDRDLWLIENGLISGGRMDVGRSLLRGNSALVDECVAFYEEHPDVFLELIRTGRPDEESYHLAHEAEARGLPVTVATNDVRFIDSSDFDAHEIRVAIHDGFTLDSDKPRNYSQPQQYMRSHEEMCELFADIPEALANTVEIAKRCNVTTRESGSIGSASKLRVFATDEEIAAHEARLDLVQQKGGSCLLWRA

Colour Key:

$\epsilon$CT + Linker + $\alpha^{270P}$ (Met1–Thr270).

Leu to Pro mutation indicated in pink.

First amino acid in sequence for linked construct variants beginning at residue numbers 180, 189, 200 and 209 of the $\epsilon$CT are highlighted in red.

Amino acid sequence of the fusion protein $\alpha^{270}\epsilon^{209}$ (314 a.a)

MSEPRFVHLRVDYSDSIDMPAKTAPLVKKAAALGMMPALAITDFTNLCGLVKFYGAGHGAGIKPIVGAFNVQCDLLGDELTHLTVLAANNTGYQLTLLISKAYQRYGAAGPIIDRDLWLIENGLISGGRMDVGRSLLRGNSALVDECVAFYEEHPDVFLELIRTGRPDEESYHLAHEAEARGLPVTVATNDVRFIDSSDFDAHEIRVAIHDGFTLDSDKPRNYSQPQQYMRSHEEMCELFADIPEALANTVEIAKRCNVTTRESGSIGSASKLRVFATDEEIAAHEARLDLVQQKGGSCLLWRA
Amino acid sequence of the $\theta$ subunit of *E. coli* DNA polymerase III (76 a.a.). The N-terminal methionine undergoes cleavage subsequent to production.

MLKNLAKLDQTEMKDKNVLAAAGVAFKERYNMPVIAEAEVEREQPEHLRSWFRERLIAHRLASVNLSSLPYEPKLK

Amino acid sequence of the $\beta$ subunit of *E. coli* DNA polymerase III (366 a.a).

MKFTVEREHLLKPLQQVGPGLGGRPTPLPLGNLLLQVADGTLSTGTDLEREMVARVALVQHPGATTVPARKFDIRGCGTPEGAEAIAVQLEGERMLVRSGRSRSFLSLPLADFPNLDDWQSEVEFTPLQATMKRLIEATQFSMAHQUVARRYLNMLFETEGEELRTVATDGHRALAVCSMPIGGQLPSHSVIVPRKGVIELMRMLDGDNPRLVRQIGSNNIRAHVGDIFTSKLVDGRFPDHYRRVLPKNPDKHLEAGDLKQAFARAALSNEKFRGVRLYVSENQLKITANNPEQEAEIEILDVYSGAEMEGFNPVSYVLDVLNALKCENVMLMTDSVSSVQJEDASSQAAYVVMPML

Amino acid sequence of the $\tau_{16}$ fragment of the $\tau$ subunit of *E. coli* DNA polymerase III (146 a.a).

MKALEHEKTPELAAKLAEEAIERDPWAAQVQLSPKLVALQVAWNKESSDAVCLHLRSSQRHLNNRGAQQKLAEEALSMLKSTVELTIVEDDNPAVRTPLEWRQAIYEKLAQARESIIADNNIQTTLRFFDAELDEISIRPI

Amino acid sequence of the His-tagged formylglycine-generating enzyme (FGE) from *Mycobacterium tuberculosis* (390 a.a).

MGSSHHHHHHSSGLVPRGSERMTELVDGGSGFRMGSTRFYPEEAPIHTVTVRAFAVERHPVTNAQFASFVTASEQPLDPGGLYPDAAIDLCPGAMVFCTAGPVDLRDWRQWWWDVVPAGRCRWHPGRDSDIADRGHPPVQVAYPDAYVARWAGRRPLEAEWEYARDGTTAYAWGDQEKGPGMLMANTWQGRFPYRNDALGTVGTSTPSGFPANGFGLLDMIGNVWEWTTEFPYHHRIDPSTACCAPVKLATAADPMISQTLKGSHLCAPAYCHRYRPAARSPQSDQDTATTHIGFRVCADPVSG
Amino acid sequence of the α subunit of *E. coli* DNA polymerase III (1160 a.a). The N-terminal methionine undergoes cleavage subsequent to production.

```
MSEPRFVHLRVHSDYSMIDGLAKTAPLVKKAAALGMPALAITDFTNLCG
LVKFYGAGHGAGIKPVGADFNVQCDLLLGDELTHLTVLAANNTGYQNL
TLISSKAYQRGYGAAGPIIDRWLIELNEGLLILLSGGRMGDVGRSLLRGN
ALVDECVAFYEHHDFPRLRFTGDPDFDEESYLHAAVELAEARGLPVAT
NDVRFIDSSDFDAHEIRAIHDGFTLDDPKRPYNPSQPQQYMSEEMCELF
ADIEALANTVEIAKCNVTVRLGEYFLPQPFTGDMSTEDYLVKRAKEGL
EERLALFDPDEEERLKRREYDERLETLEQVINQMGPFGYFLVMWFQWSK
DNGVPVGPGRGSGAGSLVAYALKTDLDPLEFDLLFERFLNPERVSMPDF
DVDFCMKEKRDQVIEHVADMYGRDAVQITFGTMAAKAVIRDVGRVLG
HPYGFDVDRISKLIPPDPGMLTAKAFEAEPQLPEIYEADDEEVKALIDMARK
EGVTRNAGKHAGGVVIAPTKITDFAPLYCDEEGKHPQTQFDSDKVEYAG
LVKFDFLGLRTLTIINWALEMINKRRAKNGEPPDLLIAIPLDDKKSFDMLQ
RSETTAVFQLESRGMKDLIKLRQPDCFEDMIALVALFRPGPLLQGMVDNF1
DRKHGREIESYPDVQVQHESLKPVLEPTYGIILYEQVQMIAVQLSGYTLG
GADMRLRAGMKKKPEEMAKQRSVFAEGAEKNGINELAMKIFDLVEKF
AGYGNKFSHSAAYALVSYQTLWLKAHYPAEMFAAAMTDNDYKVTV
GLVDECRWMGLKILPPPDSNGLYHFFHVNDDEIYVIGIAKGVGEPIA1I
IEARNKGGYFRELFDLCARTDTKRLNRRVLEKLMAGFDRILGPMRAALM
NSLGDALKAADDQHAKAEAIQGADMGVLANEPEQIESQYASCQWPWEQ
VVLDERETLGLYLTHPHINQYKLIEERYVGVRKDMHTERPGRKVITA
AGLVVAAARVMVTKRGNIGICTLLDRSGLERVMLFTDALKYQQLEKDKRI
LIVSGVSIFDFSGGLKMTARVMDIJDEAREKYARGLALISLTDQIDDQLL
NRLRQSLEPHRSGTIPVHLVYYQRADARARLRFGATWRVSPSDRLLLNRG
LIGSESEQVELEFD
```

Position of the αE612K mutation is highlighted in pink

The internal clamp binding motif (CBM) is highlighted in green. In αL this sequence is QLDLF.