Structural dynamics and protein interactions within the E. coli DNA polymerase III holoenzyme

Jacob Samuel Lewis

University of Wollongong

Follow this and additional works at: https://ro.uow.edu.au/theses1

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
Structural dynamics and protein interactions within the *E. coli* DNA polymerase III holoenzyme

by

Jacob Samuel Lewis

Bachelor of Medical Biotechnology (Advanced, Honours)

A thesis submitted in (partial) fulfilment of the requirements for the degree of Doctor of Philosophy

School of Chemistry
University of Wollongong

March 2017
To my parents for their love and support. I appreciate their sacrifices and would not have been able to reach this far without them.

To Mrs. Delaney for all your endless enthusiasm and boundless encouragement. Without you, I would not have discovered my love for science.
DECLARATION

I, Jacob Samuel Lewis, declare that this Thesis, submitted in accordance with the regulations of the University of Wollongong in (partial) fulfilment of the degree of Doctor of Philosophy. The work described herein is the author’s own work unless otherwise stated, and was carried out within the School of Chemistry, University of Wollongong, from March 2013 – June 2016. None of the material has been submitted in support of an application for any other degree.

Jacob Samuel Lewis

30th, March, 2017
ACKNOWLEDGEMENTS

While my name may be alone on the front cover of this Thesis, I am by no means its only contributor, rather there are many people who helped and deserve to be both acknowledged and thanked here.

I would like to express my deepest gratitude to my supervisor Professor Nicholas Dixon and co-supervisor Associate Professor Aaron Oakley, for their constant guidance, great patience and continuous support. Nick, you always had time to answer my questions not matter how small or confusing. Without your advice and seemingly endless knowledge, this Thesis would not have been possible. I thank you for giving me the opportunity to work in your laboratory and experience the fascinating and mystifying world of DNA replication.

I am forever indebted to my other mentor Dr Slobodan Jergic, you made me into the scientist I am today. You have taught me the fine art of experimental design, to approach problems with creative disciplined solutions, and a feeling of satisfaction that can only be experienced after you purify your first protein. I would also like to extend my gratitude to Drs Andrew Whitten and Anthony Duff for answering all my questions regarding SAXS and SANS, your patience, attention to detail and effort will not be forgotten.

The work in the latter part of this Thesis would not have been possible without the assistance and mentorship of Professor Antoine van Oijen. Antoine – I will be forever grateful for the opportunities you afforded me during my PhD and the confidence you had in my abilities. I would like to also thank Lisanne Spenkelink for her contributions and efforts while teaching a biochemist some microscopy.

To all the members of the Dixon, Oakley, Yu and van Oijen groups, past and present thank you for your help, friendship and for making this journey so enjoyable. To my friend and colleague Amy, thanks for being my confidant in the lab, you were always there to listen to my problems, concerns and anxieties.

Finally, thank you to my tolerant family and friends for your patience, love and support. I am especially indebted to my partner, Emma – thank you for your love, encouragement, and patience throughout the past 4 years, especially when I needed another “5 min” in the laboratory.
LIST OF CONFERENCE PRESENTATIONS AND PUBLICATIONS


ABBREVIATIONS

3D  three-dimensional
AAA+ ATPase associated with various cellular activities
ATPγS adenosine 5’-O-(3-thiol)triphosphate
Aλ Absorbance at wavelength λ
BSA bovine serum albumin
CBM clamp-binding motif
CLC clamp loader complex
cryo-EM cryo-electron microscopy
CTD C-terminal domain
CTS C-terminal segment
DEAE diethylaminoethyl
dmax maximum dimension computed from ρ(r)
ds double-stranded
EDTA ethylenediaminetetraacetic acid
ελ extinction coefficient at wavelength λ
ESI-MS electrospray ionisation-mass spectrometry
EtBr ethidum bromide
Exo IX Exonuclease IX
ε-YPet C-terminal fusion of fluorescent protein YPet to the ε subunit
FEN flap endonuclease
FPLC fast protein liquid chromatography
FRET fluorescence resonance energy transfer
FWHM full width at half maximum
GndHCl guanidine hydrochloride
HBD or DnaGC helicase-binding domain
HE holoenzyme
HOT homolog of theta
HSQC Heteronuclear Single Quantum Coherence
IPTG isopropyl-β-D-1-thiogalactopyranoside
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LB(T)</td>
<td>Lysogeny Broth broth with supplement (thymine)</td>
</tr>
<tr>
<td>MALLS</td>
<td>multi-angle laser light scattering</td>
</tr>
<tr>
<td>MCS</td>
<td>multiple cloning site</td>
</tr>
<tr>
<td>MD</td>
<td>Molecular dynamics</td>
</tr>
<tr>
<td>MWCO</td>
<td>molecular weight cut off</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>NOE</td>
<td>Nuclear Overhauser Effect</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>OB</td>
<td>oligosaccharide-binding</td>
</tr>
<tr>
<td>OFM</td>
<td>Okazaki fragment maturation</td>
</tr>
<tr>
<td>oriC</td>
<td>origin of replication</td>
</tr>
<tr>
<td>( p(r) )</td>
<td>pair-distance distribution function</td>
</tr>
<tr>
<td>P20</td>
<td>surfactant P20</td>
</tr>
<tr>
<td>PCNA</td>
<td>proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank</td>
</tr>
<tr>
<td>PHP</td>
<td>polymerase and histidinol-phosphatase</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethylsulfonyl fluoride</td>
</tr>
<tr>
<td>Pol I</td>
<td>polymerase I</td>
</tr>
<tr>
<td>Pol III</td>
<td>polymerase III (( \alpha\epsilon\theta ))</td>
</tr>
<tr>
<td>Pol III HE</td>
<td>polymerase III holoenzyme</td>
</tr>
<tr>
<td>Pol III*</td>
<td>HE lacking ( \beta_2 ) clamp</td>
</tr>
<tr>
<td>Q-Tof</td>
<td>Quadrupole–time-of-flight</td>
</tr>
<tr>
<td>RBS</td>
<td>ribosome binding site</td>
</tr>
<tr>
<td>RC</td>
<td>regenerated cellulose</td>
</tr>
<tr>
<td>( R_g )</td>
<td>radius of gyration</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>RPD</td>
<td>RNA polymerase domain</td>
</tr>
<tr>
<td>RU</td>
<td>resonance units</td>
</tr>
<tr>
<td>SANS</td>
<td>small angle neutron scattering</td>
</tr>
<tr>
<td>SAS</td>
<td>small angle scattering</td>
</tr>
<tr>
<td>SAXS</td>
<td>small angle X-ray scattering</td>
</tr>
<tr>
<td>SD</td>
<td>strand-displacement</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl-polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>SEC-SAXS</td>
<td>size exclusion chromatography coupled SAXS</td>
</tr>
<tr>
<td>Sm</td>
<td>single-molecule</td>
</tr>
<tr>
<td>SPR</td>
<td>surface plasmon resonance</td>
</tr>
<tr>
<td>SRC</td>
<td>Ser–Arg–Cys</td>
</tr>
<tr>
<td>ss</td>
<td>single-stranded</td>
</tr>
<tr>
<td>SSB</td>
<td>single-stranded binding protein</td>
</tr>
<tr>
<td>SSB-Ct</td>
<td><em>E. coli</em> highly conserved motif of C-terminus of SSB</td>
</tr>
<tr>
<td>TAE buffer</td>
<td>Tris/acetate/EDTA buffer</td>
</tr>
<tr>
<td>TBE buffer</td>
<td>Tris/borate/EDTA buffer</td>
</tr>
<tr>
<td>TIRF</td>
<td>total internal reflection</td>
</tr>
<tr>
<td>Tris</td>
<td>2-amino-2-hydroxymethyl-propane-1,3-diol</td>
</tr>
<tr>
<td>ts</td>
<td>temperature sensitive</td>
</tr>
<tr>
<td>ZBD</td>
<td>zinc-binding domain</td>
</tr>
</tbody>
</table>
ABSTRACT

DNA replication is an essential process that must proceed from start to end without failure. The *Escherichia coli* cellular replication machine or “replisome” is composed of a dozen or so different proteins. At the heart of the replisome is the DNA polymerase III holoenzyme (Pol III HE). We have established a detailed picture of the enzymatic abilities and functions of the Pol III HE, and as a result uncovered many of the protein–protein interactions essential for its function. How many of these protein–protein interactions allow the replisome to transition from one conformational state to another remains either speculative or unknown. The main aims of this project were to study the interactions that allow for a dynamic replisome, to better understand the hierarchy of strong and weak functional interactions within the Pol III HE.

In work reported in this Thesis, (1) the contribution of domain IVa of the τ subunit to the interaction with α was interrogated using surface plasmon resonance (SPR) and structural analysis by small angle X-ray (and Neutron) scattering. The additional residues in domain IVa of τ (τC24) were shown to increase the binding affinity 3–5-fold over a mutant containing the α binding domain V only (τC16). Examination of these residues by 1H15N-HSQC NMR showed few residues unique to domain IVa perturbed upon addition of α. This indicates these additional residues unique to domain IVa of τ make putative contacts with α. Structural analysis by SAXS and/or SANS of the α–τC16 and α–τC24 complexes revealed more compact structures than expected, and rigid-body models suggest a potential structural rearrangement of α upon τ binding in the absence of DNA.

Next, (2) SPR was used to measure the accurate $K_D$ values for complete, full-length α, αε and αετ complexes binding to β2 in combination with different clamp binding motifs (CBMs). Using a novel strategy, whereby binding of β2 to α and/or the αε complex immobilised via the α–τ interaction was measured, a contribution of τ to β2 binding was exposed. Various CBMs in ε and α were used to provide evidence for the ability of the C-terminal domains of τ to position/assemble α in a manner that excludes the ε CBM from accessing β2 in the absence of DNA. Although, it is currently unknown whether this effect is present once α binds primer–template DNA. Further, a direct physical interaction between immobilised τC24 and β2 was identified with an apparent $K_D$ of 110 ± 5 μM (β2 as a dimer). By using 1H15N-HSQC NMR and another truncation mutant of τ (τC22), the direct physical interaction between τ and β2 was confirmed. Moreover, a binding affinity of $K_D = 86 ± 54$ μM (β2 as a monomer) of the τ–β2 was derived from
\(^1\text{H}^{15}\text{N}-\text{HSQC} \) NMR data using \(^1\text{H} \) chemical shifts upon addition of decreasing concentrations of \(\beta_2\).

Next, (3) fluorescently labelled Pol III cores were produced by labelling either the \(\alpha\) subunit or \(\theta\) subunit using three typical labelling strategies including: labelling with fluorescent organic dyes, genetic fusion to a fluorescent protein and enzymatic labelling via a SNAP-tag. The potential uses of labelled Pol III cores were demonstrated using photoswitchable mKikGR-\(\alpha\) Pol III core in two different \textit{in vitro} and \textit{in vivo} experiments. These fluorescent reagents are a critical piece in the puzzle for the simultaneous observation of the composition and enzymatic activity of single replisomes in real-time.

Finally, (4) imaging of the reconstituted \textit{E. coli} DNA Pol III HE containing SNAP labelled Pol III cores at the single-molecule level were used to study the dynamics of DNA polymerases at the replication fork. Using a rolling-circle DNA amplification scheme to observe highly processive DNA synthesis in real-time, exchange of Pol III* at the replication fork was directly visualised by measuring the fluorescence intensity at the replisome spot as a function of time using 1:1 mixtures of red and green labelled Pol III*s. Additionally, Pol III* readily exchanges into the replisome when continually present at a rate that is dependent on its concentration. In juxtaposition, by pre-assembling replisomes on a rolling-circle template and initiating replication, we demonstrate the replisome is a stable entity, consistent with the current textbook view.

In addition, reassessment of the \(\tau\)--DnaB interaction by SPR showed binding is ~200-fold weaker than previously reported, with a measured \(K_D\) of 1.3 ± 0.2 \(\mu\)M (using \(\tau\delta\delta\delta\chi\chi\psi\) in 50 mM NaCl). This observation strongly supports Pol III* exchange as the \(\tau\)--DnaB interaction would not be an obstacle for efficient exchange.

The work in this Thesis deepens our knowledge of DNA replication, especially the structural dynamics of DNA polymerases at the replication fork in \textit{E. coli}. The \textit{E. coli} replisome is able to strike a balance between stability and plasticity that complies with established chemical principles. However, to fully understand mechanisms and functional interactions associated with such dynamic behaviour, more work involving real-time single-molecule \textit{in vitro} replication assays, \textit{in vivo} imaging and other new methods are required. By doing this we will be able to gain a complete picture of these protein–protein and protein–DNA interactions that orchestrate such short-lived intermediate states during coupled DNA replication.
TABLE OF CONTENTS

DEDICATION .......................................................................................................................... I
DECLARATION ........................................................................................................................... II
ACKNOWLEDGEMENTS ........................................................................................................ III
LIST OF CONFERENCE PRESENTATIONS AND PUBLICATIONS .................................. IV
ABBREVIATIONS .................................................................................................................. V
ABSTRACT ............................................................................................................................. VIII
TABLE OF CONTENTS ........................................................................................................... X
LIST OF FIGURES ................................................................................................................ XVIII
LIST OF TABLES .................................................................................................................. XXI

Chapter 1 General Introduction ........................................................................................ 1
1.1 Introduction ................................................................................................................... 3
1.2 Initiation at oriC, the origin of replication ................................................................. 6
1.3 The primosome, DnaB and DnaG .............................................................................. 6
1.4 The role of SSB at the replication fork .................................................................... 10
1.5 The elongation stage ............................................................................................... 13
1.5.1 Assembly and action of the DNA polymerase III holoenzyme ....................... 13
1.5.2 The β2 clamp ....................................................................................................... 14
1.5.3 The Pol III core, αεθ ...................................................................................... 18
1.5.4 The clamp loader complex ............................................................................ 24
1.6 Dynamics on the lagging strand ............................................................................. 30
1.6.1 Lagging strand priming .................................................................................. 32
1.6.2 The collision model ....................................................................................... 35
1.6.3 The signalling model .................................................................................... 35
1.6.4 Okazaki fragment maturation ....................................................................... 36
1.7 Concluding comments .......................................................................................... 38
1.8 Scope of this Thesis ............................................................................................. 39

Chapter 2 General Materials and Methods .................................................................... 42
2.1 Bacterial strains and plasmid vectors ........................................................... 43
  2.1.1 Bacterial strains .................................................................................. 43
  2.1.2 Plasmid vectors ................................................................................. 44
2.2 Chemicals and reagents .............................................................................. 45
2.3 Growth media ............................................................................................... 47
  2.3.1 Lysogeny Broth medium .................................................................... 47
  2.3.2 Lysogeny Broth solid medium (agar plates) ..................................... 47
  2.3.3 Lysogeny Broth autoinduction medium .......................................... 48
  2.3.4 SOC medium .................................................................................... 48
  2.3.5 Minimal medium for $^{15}$N labelling ............................................... 48
  2.3.6 ModC1 minimal medium for perdeuteration of proteins ................. 48
2.4 Custom oligonucleotides ............................................................................. 49
2.5 Molecular biology procedures ..................................................................... 49
  2.5.1 Preparation of plasmid DNA by alkaline lysis .................................. 49
  2.5.2 Quantification of DNA ..................................................................... 49
  2.5.3 Agarose gel electrophoresis .............................................................. 50
  2.5.4 Polymerase chain reaction (PCR) ...................................................... 50
    2.5.4.1 Cloning PCR ............................................................................. 50
    2.5.4.2 Colony PCR ............................................................................. 51
    2.5.4.3 Nucleotide sequence determination ......................................... 51
  2.5.5 Restriction endonuclease digestion .................................................... 51
  2.5.6 Purification of DNA .......................................................................... 52
    2.5.6.1 Extraction from agarose gels ...................................................... 52
    2.5.6.2 Extraction by spin column ......................................................... 52
  2.5.7 DNA ligation ...................................................................................... 52
  2.5.8 Transformation of competent E. coli cells ....................................... 52
2.6 Protein biochemistry methods .................................................................... 52
  2.6.1 Fast protein liquid chromatography (FPLC) purification of proteins ... 52
  2.6.2 Dialysis of proteins .......................................................................... 53
2.6.3 Sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE) ................................................................. 53

2.6.4 Staining of SDS-PAGE gels with Coomassie brilliant blue .......... 53

2.6.5 Concentration of proteins ........................................................................ 54

2.6.5.1 Ammonium sulphate precipitation.................................................... 54

2.6.5.2 Ultrafiltration............................................................................ 54

2.6.5.3 Acetone precipitation................................................................ 54

2.6.6 Quantification of proteins.................................................................. 54

2.6.6.1 Bradford assay................................................................. 54

2.6.6.2 UV absorbance at \( A_{280} \) for purified proteins......................... 55

2.7 Electrospray ionization (ESI)-mass spectrometry (MS) ................. 55

Chapter 3 Structural and Interaction Studies of Domain IVa of the \textit{Escherichia coli} \( \tau \) Subunit and its Contribution to the \( \alpha-\tau \) Interaction.................................................. 56

3.1 Introduction ............................................................................................. 58

3.1.1 The \( \tau \) subunit of the \textit{E. coli} clamp loader complex .......................... 58

3.1.2 \( \tau \) interacts with \( \alpha \) through its unstructured C-terminus........ 60

3.1.3 Using a surface plasmon resonance (SPR) biosensor to study protein–protein interactions .................................................. 62

3.1.4 Using small angle scattering to study biomolecular structures in solution 64

3.1.5 Aims of this chapter ........................................................................ 65

3.2 Materials and Methods ........................................................................ 67

3.2.1 Plasmid construction........................................................................ 67

3.2.2 Overproduction and purification of \textit{bio}^{\tau_{C16}} ........................................ 67

3.2.3 Overproduction and purification of \textit{bio}^{\tau_{C24}} ............................. 68

3.2.4 \textit{In vitro} biotinylation of \textit{bio}^{\tau_{C24}} ................................................ 69

3.2.5 Overproduction and purification of \( \tau_{C24} \) ..................................... 70

3.2.6 Overproduction and purification of \( ^{15} \text{N}-\text{labelled} \ \tau_{C24} \) ............. 70

3.2.7 Preparation of \( \tau_{C16} \)-agarose affinity column ................................. 71

3.2.8 Overproduction and purification of the \( \alpha \) subunit........................ 71
3.2.9 Overproduction and purification of perdeuterated α ........................................ 73
3.2.10 Reconstitution and purification of α−τC24 and α−τC16 complexes .............. 74
3.2.11 Analysis of protein–protein interactions by SPR methodology ............... 74
3.2.12 NMR spectroscopy and data processing ................................................... 75
3.2.13 Small angle X-ray scattering .................................................................... 76
3.2.14 Small angle Neutron scattering ............................................................... 76
3.3 Results ............................................................................................................ 77
  3.3.1 Biotinylation of bioτC24 and bioτC16 ........................................................... 77
  3.3.2 A direct assay for the α−τC24 interaction using SPR ............................... 78
  3.3.3 An examination of the α−τC24 interaction using NMR spectroscopy ....... 81
  3.3.4 Structural characterisation of α, α−τC16, and α−τC24 by SAXS and SANS 84
     3.3.4.1 The α subunit .................................................................................. 85
     3.3.4.2 The α−τC16 and α−τC24 complexes ............................................... 87
     3.3.4.3 SANS with contrast variation of the α−τC24 complex ....................... 89
     3.3.4.4 3D modelling of the α subunit from SEC-SAXS .............................. 93
     3.3.4.5 3D modelling of the α−τC16 complex from SEC-SAXS .................. 96
     3.3.4.6 3D modelling of the α−τC24 complex from SEC-SAXS and SANS ...... 98
3.4 Discussion .................................................................................................... 100
  3.4.1 Contribution of domain IV to the α−τ interaction .................................... 100
  3.4.2 Is domain IV of τ an actor in the processivity switch? ............................. 102
Chapter 4 The α Interaction Domain V of the Escherichia coli τ Subunit also Interacts
with the β2 Clamp .................................................................................................. 103
  4.1 Introduction ................................................................................................ 105
  4.2 Aims of this Chapter .................................................................................. 107
  4.3 Materials and Methods ............................................................................ 108
  4.3.1 Plasmid construction ............................................................................ 108
  4.3.2 Overproduction and purification of ε ................................................. 108
  4.3.3 Overproduction and purification of bio*ε ............................................ 110
  4.3.4 Overproduction and purification of bio*τQ ........................................ 110
4.3.5 Overproduction and purification of $^{15}$N-labelled $\tau_{C14}$ ........................................... 110
4.3.6 Analysis of protein–protein interactions by SPR methodology .................. 111
  4.3.6.1 The $\alpha$–$\varepsilon$ interaction ........................................................................ 112
  4.3.6.2 Interactions with the $\beta_2$ clamp .......................................................... 112
4.3.7 NMR spectroscopy and data processing .................................................. 113
4.4 Results ........................................................................................................ 114
  4.4.1 Biotinylation of bio*$\varepsilon_{WT}$ and bio*$\varepsilon_Q$ ........................................ 114
  4.4.2 A novel assay for measuring interaction between $\beta_2$ and $\alpha$ using $\varepsilon$ by SPR using immobilised $\varepsilon$ ............................................................... 114
  4.4.3 $\tau$ contributes to the interaction of $\beta_2$ with $\alpha$ ...................................... 117
  4.4.4 Addition of $\varepsilon$ to the $\alpha$–$\tau$ complex does not influence $\beta_2$ binding .......................................................... 120
  4.4.5 $\alpha_L \tau$ positions $\beta_2$ in a manner where the $\alpha_L$ CBM cannot access its binding site ........................................................................................................ 122
  4.4.6 A direct physical interaction between $\tau_{C24}$ and the $\beta_2$ clamp .......... 126
  4.4.7 An examination of the $\tau_{C}$–$\beta_2$ interaction using NMR spectroscopy ..... 126
4.5 Discussion .................................................................................................. 130
  4.5.1 The $\tau$–$\beta_2$ interaction ............................................................................ 130

Chapter 5 Development and Characterisation of Fluorescently Labelled Pol III Cores for Single-Molecule Imaging ............................................................................ 133

  5.1 Introduction ............................................................................................... 135
  5.1.1 The need for fluorescently labelled components in the Pol III HE ....... 135
    5.1.1.1 Organic fluorophores: maleimide-fluorophore conjugation ........... 136
    5.1.1.2 Gentic fusion of fluorescent proteins ............................................. 137
    5.1.1.3 Self-labelling through protein tags ............................................. 139
  5.1.2 In vitro DNA replication assays: A means for dissecting the functional properties of the DNA Pol III HE ................................................................. 140
5.2 Aims of this Chapter ............................................................................... 141
5.3 Materials and Methods ............................................................................ 142
  5.3.1 Construction of plasmids .................................................................. 142
  5.3.2 Overproduction and purification of $\theta_{cys}$ ........................................ 143
5.3.3 Thiopropyl Sepharose purification of \( \theta_{\text{cys}} \) ........................................ 144
5.3.4 Solid state labelling of \( \theta_{\text{cys}}^{*} \) with maleimide functionalised fluorescent dyes ............................................................... 145
5.3.5 Overproduction and purification of mKikGR\( -\alpha \) ........................................ 146
5.3.6 Overproduction and purification of SNAP\( -\alpha \) ........................................... 146
5.3.7 Enzymatic labelling of SNAP\( -\alpha \) ................................................................. 146
5.3.8 \textit{In vitro} reconstitution and purification of fluorescently labelled Pol III cores .................................................................................................................. 147
5.3.9 Determination of labelling efficiency by ESI-MS ........................................ 148
5.3.10 Degree of labelling determined by UV/Vis spectroscopy ................................ 148
5.3.11 NMR spectroscopy and data processing ...................................................... 148
5.3.12 Strand displacement DNA synthesis assays .................................................. 149
5.3.12.1 Production of primed M13 DNA template ............................................. 149
5.3.12.2 Replication reaction conditions .............................................................. 149
5.3.13 Fluorescent imaging of mKikGR\( -\alpha \) containing replisomes ................... 149
5.3.13.1 Single-molecule rolling-circle DNA replication assays ....................... 149
5.3.13.2 Single-molecule imaging of mKikGR\( -\alpha \) \textit{in vivo} ................................ 150

5.4 Results ........................................................................................................ 151
5.4.1 Non-thiol containing \( \theta_{\text{cys}} \) ................................................................. 151
5.4.2 Solid state labelling of \( \theta_{\text{cys}}^{*} \) is efficient and specific ................................ 153
5.4.3 Efficient labelling of SNAP\( -\alpha \) ................................................................. 154
5.4.4 S-galactosylation of \( \theta_{\text{cys}} \) \textit{in vivo} ................................................................. 155
5.4.5 Labelled Pol III cores are active in rolling-circle strand-displacement DNA synthesis ........................................................................................................ 157
5.4.6 Preliminary characterisation of mKikGR\( -\alpha \) \textit{in vitro} and \textit{in vivo} ......... 158

5.5 Discussion .................................................................................................. 161
5.5.1 Solid-state labelling protocol: An efficient methodology ...................... 161
5.5.2 S-galactosylation of \( \theta_{\text{cys}} \): An artificial case ........................................ 161
5.5.3 Potential uses for labelled Pol III cores .................................................. 163
Chapter 6 Single-molecule Visualisation of Fast Polymerase Turnover in the Bacterial Replisome

6.1 Introduction

6.1.1 The single-molecule rolling-circle assay

6.2 Aims of the Chapter

6.3 Methods

6.3.1 Protein expression and purification

6.3.2 Bulk helicase-independent Pol III strand-displacement DNA replication assays

6.3.3 Bulk leading- and lagging-strand DNA replication assays

6.3.4 Rolling-circle substrate preparation

6.3.5 Coverslip surface preparation

6.3.6 Flow-cell construction

6.3.7 TIRF microscope setup

6.3.8 Single-molecule rolling-circle replication reactions

6.3.9 Data analysis

6.3.10 SPR (BIAcore) measurements of the interactions between DnaB and immobilised clamp loader complexes

6.4 Results

6.4.1 Comparison of activities of wild-type and SNAP-labelled Pol III cores

6.4.2 Real-time fluorescence imaging of labelled Pol III during coupled DNA replication

6.4.3 Pre-assembled Pol III* complexes do not exchange Pol III cores

6.4.4 Pol III* readily exchanges into the replisome at a rate that is dependent on its concentration

6.4.5 Assessment of the DnaB–clamp loader complex interaction by SPR (BIAcore)

6.5 Discussion

6.5.1 Plasticity vs. stability

6.5.2 Perspectives and implications of a dynamic replisome
6.5.3 Future work ................................................................. 188

Chapter 7 Concluding Summary ...................................................... 190

7.1 Structural and interaction studies of domain IVa of the $\tau$ subunit and its contribution to the $\alpha-\tau$ interaction (Chapter 3) .................................................. 191

7.2 The $\alpha$ interaction domain V of the Escherichia coli $\tau$ subunit also interacts with the $\beta_2$ clamp (Chapter 4) ................................................................. 192

7.3 Development and characterisation of fluorescently labelled Pol III cores for single-molecule imaging (Chapter 5) .................................................. 194

7.4 Single-molecule visualisation of fast polymerase turnover in the bacterial replisome (Chapter 6) ......................................................................... 195

References .................................................................................. 197

Appendix ....................................................................................... 225
LIST OF FIGURES

Figure 1.1 | Architecture of the E. coli replisome at the chromosomal replication fork derived from in vitro studies and direct observation in vivo ................................................. 4
Figure 1.2 | Homology model of the E. coli DnaBζ–(DnaGC)3 primosomal complex based on the corresponding structure from Geobacillus stearothermophilis ......................... 8
Figure 1.3 | Structure of the E. coli SSB tetramer .......................................................... 12
Figure 1.4 | The β₂ sliding clamp in complex with primer–template DNA ....................... 16
Figure 1.5 | A model of the structural organisation of the E. coli αεθ–β₂–DNA complex ..................................................................................................................................... 21
Figure 1.6 | Crystal structures of the E. coli γδδζψ pep–DNA and χψ complexes .............. 26
Figure 2.1 | Maps of plasmids used in this Thesis .............................................................. 45
Figure 3.1 | The domain structures of α and τ subunits .............................................. 59
Figure 3.2 | The solution structure of τC₁₄ determined by NMR ..................................... 61
Figure 3.3 | The X-ray crystal structure of the αCTD–τC₁₆ complex ......................... 61
Figure 3.4 | Principle of surface plasmon resonance (SPR) and setup of a SPR biosensor ..................................................................................................................................... 63
Figure 3.5 | The τC₁₆-agarose affinity column removes the proteolytic C-terminal fragments of the α subunit ............................................................................................................ 73
Figure 3.6 | In vitro biotinylation of bioτC₂₄ was successful ............................................. 77
Figure 3.7 | SPR assessment of α–τC₁₆ and α–τC₂₄ interactions: τC₂₄ binds more strongly to α than τC₁₆ ........................................................................................................................ 79
Figure 3.8 | Sensorgrams showing experimentally determined Rmax values for bio*τC₁₆ and bio*τC₂₄ at saturating α concentrations ......................................................... 80
Figure 3.9 | SDS-PAGE analysis of purified [¹⁵N]τC₂₄ and the α subunit ......................... 82
Figure 3.10 | ¹H¹⁵N-HSQC NMR assessment of α–τC₂₄ interaction using uniformly ¹⁵N-labelled τC₂₄ ..................................................................................................................................... 83
Figure 3.11 | In vitro reconstitution of the α–τC₁₆ and α–τC₂₄ complexes from individual purified subunits .................................................................................................................. 84
Figure 3.12 | SEC-SAXS data for α ...................................................................................... 86
Figure 3.13 | SEC-SAXS data for α–τC₁₆ and α–τC₂₄ complexes ..................................... 88
Figure 3.14 | SANS with contrast variation can resolve α from τC₂₄ in the α–τC₂₄ complex ..................................................................................................................................... 91
Figure 3.15 | Ab initio envelopes and rigid-body model of the α subunit from SEC- SAXS ..................................................................................................................................... 95
Figure 3.16 | The α–τC16 complex in a DNA-free state is a closer fit to the experimental SEC-SAXS than the DNA-bound state................................................................. 97
Figure 3.17 | Structural modelling of α–τC24 derived from SEC-SAXS and SANS data.......................................................................................................................... 99
Figure 4.1 | The two domain arrangement of the ε proofreading subunit from E. coli. 106
Figure 4.2 | Purified bio*εWT and bio*εQ used in SPR experiments.........................114
Figure 4.3 | SPR assessment of the α–ε interaction..................................................115
Figure 4.4 | Immobilisation of the bio*εα complex to measure β2 binding by SPR ................................................................. 117
Figure 4.5 | Strengths of α–β2 interactions in the presence of the C-terminal domains of τ................................................................. 119
Figure 4.6 | The native CBM in ε does not significantly contribute to the α–β2 interaction in the presence of τ................................................................. 121
Figure 4.7 | A stronger CBM in ε strengthens the α–β2 interaction in the presence of τ................................................................. 122
Figure 4.8 | A strong CBM in α renders the CBM in ε unable to contribute to ταε–β2 interaction.................................................................................................................. 124
Figure 4.9 | SPR assessment of the τC24–β2 interaction ........................................ 126
Figure 4.10 | Solution structure of τC14 highlighting the residues for which 1HN chemical shift changes were observed for τC22 in the complex with β2 compared to τC22 alone............................................................................................................. 128
Figure 4.11 | The C-terminal domains of τ bind to the β2 clamp, causing local changes in the chemical environment of τC22................................................................. 129
Figure 4.12 | Cryo-EM structure of αε–β2–τC16 complex in DNA-free state......... 132
Figure 5.1 | Three labelling strategies used to fluorescently label proteins ........ 138
Figure 5.2 | Purification of θcys and detection of a modified version by ESI-MS after labelling with fluorescein-5-maleimide................................................................. 152
Figure 5.3 | Successful separation of θcys* from non-thiol-containing θcys ........ 153
Figure 5.4 | Efficient and specific maleimide-fluorophore conjugation to θcys* using solid-state labelling .................................................................................................. 154
Figure 5.5 | UV/Vis spectra of SNAP-α after treatment with O6-benzylguanine derivatives ............................................................................................................. 155
Figure 5.6 | NMR analysis confirming a galactose moiety is covalently attached to θcys in vivo .................................................................................................................. 156
Figure 5.7 | Agarose gel showing electrophoretic analysis of products of rolling-circle strand-displacement DNA replication assay comparing various labelled Pol III cores

Figure 5.8 | mKikGR-α has similar rates and processivity compared to wild-type α

Figure 5.9 | mKikGR-α co-localises with chromosomal fusion ε-YPet in replisome foci

Figure 6.1 | Established view of the organisation of the *E. coli* replication fork

Figure 6.2 | Single-molecule rolling-circle replication assay

Figure 6.3 | Purification and comparison of activities of wild-type and SNAP-labelled Pol III cores

Figure 6.4 | Real-time fluorescence imaging of coupled DNA replication

Figure 6.5 | Pre-assembled Pol III* complexes do not exchange Pol III core

Figure 6.6 | Rapid and frequent exchange of Pol III* is concentration dependent

Figure 6.7 | Pre-assembled replisomes participate in processive DNA replication

Figure 6.8 | SPR assessment of DnaB–CLC interactions: CLC containing τ3 binds stronger to DnaB than CLC containing τ1γ2 in buffer containing 50 mM NaCl

Figure 6.9 | The γ3δδ′ CLC does not bind DnaB

Appendix Figure 1 | Determination of deuteration level of α by ESI-MS under native conditions

Appendix Figure 2 | Representative field of view of SYTOX Orange-stained dsDNA from the single-molecule rolling-circle DNA replication assay

Appendix Figure 3 | Fluorescence intensity of replicating Pol III* complexes does not change at longer DNA lengths under near-TIRF imaging conditions

Appendix Figure 4 | Histograms of the stoichiometry of Pol III* at the replication fork
LIST OF TABLES

Table 1.1 | E. coli replisome components and associated functions. ......................... 5
Table 2.1 | General chemicals and reagents used. ....................................................... 45
Table 2.2 | ModC1 medium composition ................................................................. 49
Table 3.1 | Summary of the binding parameters for the $\alpha$–$\tau_{C16}$ and $\alpha$–$\tau_{C24}$ interactions determined by simultaneous fit of sensorgrams to Langmuir one-to-one or Langmuir one-to-one binding model with mass transfer models. ................................................ 80
Table 3.2 Summary of the binding parameters for the $\alpha$–$\tau_{C16}$ and $\alpha$–$\tau_{C24}$ interactions determined by simultaneous fit of sensorgrams with globally constrained $R_{max}$ values. ..................................................................................................................................... 81
Table 3.3 | Parameters extrapolated from the experimental SEC-SAXS data for the $\alpha$ subunit ......................................................................................................................................... 85
Table 3.4 | Parameters extrapolated from the experimental SEC-SAXS data for both $\alpha$–$\tau_{C16}$ and $\alpha$–$\tau_{C24}$ complexes .................................................................................................................................. 89
Table 3.5 | Parameters extrapolated from the experimental SANS data for the $\alpha$–$\tau_{C24}$ complex. .............................................................................................................................................. 92
Table 4.1 | Summary of the binding parameters for the $\alpha$–$\epsilon$ interaction determined by simultaneous fit of sensorgrams to a Langmuir one-to-one binding model ........................................ 116
Table 4.2 | Summary of the binding parameters for the $\tau_{C\alpha L}$–$\beta_2$ interaction determined by simultaneous fit of sensorgrams to a Langmuir one-to-one with mass transport binding model ............................................................................................................. 120
Table 4.3 | Summary of the binding parameters for the $\tau_{C\alpha L}$–$\beta_2$ or $\tau_{C\alpha L\epsilon}$–$\beta_2$ interactions determined by simultaneous fit of sensorgrams to a Langmuir one-to-one with mass transport binding model ................................................................................................................................... 125
Table 4.4 | Determined values of equilibrium constants ($K_D$) and interaction kinetics for interactions of $\alpha$, $\alpha\epsilon$ and $\epsilon$ complexes with $\beta_2$ ............................................................................................................................ 130
Table 6.1 | Equilibrium binding constants describing interactions between DnaB and immobilised CLCs in buffer containing 50 mM NaCl ......................................................... 185
Table 6.2 | Equilibrium binding constants describing interactions between DnaB and immobilised CLCs in buffer containing 25 mM NaCl ......................................................... 185
Table 7.1] Determined values of equilibrium constants \((K_D)\) and interaction kinetics for interactions of \(\alpha, \alpha\varepsilon\) and \(\tau\) complexes with \(\beta_2\) ........................................................... 193
Chapter 1
General Introduction
The E. coli DNA replication fork

Jacob S. Lewis, Slobodan Jergic and Nicholas E. Dixon

DNA replication in *Escherichia coli* initiates at oriC, the origin of replication and proceeds bidirectionally, resulting in two replication forks that travel in opposite directions from the origin. Here, we focus on events at the replication fork. The replication machinery (or replisome), first assembled on both forks at oriC, contains the DnaB helicase for strand separation, and the DNA polymerase III holoenzyme (Pol III HE) for DNA synthesis. DnaB interacts transiently with the DnaG primase for RNA priming on both strands. The Pol III HE is made up of three subassemblies: (i) The αεθ core polymerase complex that is present in two (or three) copies to simultaneously copy both DNA strands; (ii) The β₂ sliding clamp that interacts with the core polymerase to ensure its processivity; and (iii) the seven-subunit clamp loader complex that loads β₂ onto primer–template junctions and interacts with the α polymerase subunit of the core and the DnaB helicase to organise the two (or three) core polymerases. Here, we review the structures of the enzymatic components of replisomes, and the protein–protein and protein–DNA interactions that ensure they remain intact while undergoing substantial dynamic changes as they function to copy both the leading and lagging strands simultaneously during coordinated replication.

---

Material in this Chapter was originally published as:

Author contributions are as follows: J.S.L., S.J. and N.E.D. discussed and wrote the review. J.S.L designed and generated figures. I am also grateful for discussions with Antoine van Oijen, Aaron Oakley and other members of the extended DNA replication group at the University of Wollongong.
1.1 Introduction

All the genetic information essential for inheritance of traits that define the phenotypes of cells is packaged in the double helix of chromosomal DNA. A single DNA molecule from *Escherichia coli* contains just over 4.6 million base pairs (Mbp) [1] and is ~1.5 mm in length, roughly 1000 times the length of an individual cell. Replication of chromosomal DNA occurs remarkably fast and is a fundamentally complex, yet carefully orchestrated process that requires the assembly and action of a molecular machine at regions of the DNA called replication forks. In 1963, John Cairns reported the first autoradiograph of a replicating [3H]-thymidine-labelled *E. coli* chromosome [2]. This was the first time that complete circular chromosomes of *E. coli* were caught in the act of replicating their DNA. In these replicating chromosomes, two replication forks were identified, representing the sites of active DNA synthesis. At each of them, parental DNA must first be unwound into two template strands that are copied simultaneously with a high degree of accuracy and with great efficiency. Furthermore, the antiparallel structure of DNA and the requirement that DNA polymerases make DNA by extending a pre-existing primer from the 3'-OH end restrict the mechanisms that can be used at replication forks to copy DNA. At each fork, the two nascent DNA strands are necessarily synthesized by different processes, with the leading-strand synthesized continuously while the lagging-strand is synthesized discontinuously as short Okazaki fragments that are later joined [3,4]. Although the mechanisms on the two strands are different, their synthesis is believed to be coupled together, and to be carried out by the same replication apparatus, called the replisome.

Replisomes are multi-protein molecular machines, evolved to coordinate all the necessary enzymatic activities for coupled DNA replication. The *E. coli* replisome is undoubtedly the best understood across all species, and is composed of over a dozen individual subunits, some of which are present in multiple copies (Figure 1.1; Table 1.1). Once assembled, the *E. coli* replisome unwinds and duplicates DNA at remarkable speeds, approaching 1000 base pairs per second with an error rate of approximately one mutation for each $10^{-6}$ to $10^{-7}$ nucleotides incorporated (reviewed in [5]).

These individual subunits interact and exchange in a hierarchy of strong and weak functional interactions (pairwise $K_D$ values range from low pM to high µM), the kinetic properties of which are tightly controlled by external conditions such as cellular protein concentrations and nucleotide availability. This hierarchy of functional interactions
enables the replisome to transition through multiple conformational states to accomplish simultaneous, concerted copying of both DNA strands at each replication fork while its overall integrity is maintained [6].

Figure 1.1 | Architecture of the E. coli replisome at the chromosomal replication fork derived from in vitro studies and direct observation in vivo. The DnaB helicase is located at the apex of the replication fork on the lagging strand where it uses the energy of ATP hydrolysis to unwind dsDNA. The lagging-strand template produced by helicase action is protected by SSB. The DNA polymerase III holoenzyme (Pol III HE) uses the single strands of DNA produced by DnaB as templates to synthesize new DNA on both the leading and lagging strands. The $\beta_2$ sliding clamp confers high processivity on the Pol III HE by tethering the Pol III $\alpha\varepsilon\theta$ cores onto the DNA. The clamp loader complex (CLC) uses ATP hydrolysis to assemble the $\beta_2$ clamp onto RNA primer junctions on template DNA. Up to three Pol III cores are coupled through the $\tau$ subunit of the CLC through their extreme C-terminal domains, and the $\tau$ subunits also interact with DnaB, thus organizing and coupling the Pol III HE to DnaB. Due to the opposite polarity of the two DNA strands, the lagging strand is synthesized in a series of short Okazaki fragments in the opposite direction to the leading strand. DnaG primases interact with DnaB to synthesise RNA primers to initiate DNA synthesis on the lagging strand, signalling the start of synthesis of an Okazaki fragment. When an Okazaki fragment is complete, Pol I replaces RNA primers with DNA and DNA ligase A joins the fragments into a contiguous lagging-strand DNA chain. Figure adapted from [7] with permission.

These individual subunits interact and exchange in a hierarchy of strong and weak functional interactions (pairwise $K_D$ values range from low pM to high $\mu$M), the kinetic properties of which are tightly controlled by external conditions such as cellular protein concentrations and nucleotide availability. This hierarchy of functional interactions enables the replisome to transition through multiple conformational states to accomplish simultaneous, concerted copying of both DNA strands at each replication fork while its overall integrity is maintained [6].
Table 1.1 | *E. coli* replisome components and associated functions.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Subunit</th>
<th>Molecules/cell</th>
<th>Gene</th>
<th>Mass (kDa)</th>
<th>Interaction partners</th>
<th>Function</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pol III HE</td>
<td>(αβθ)2−τγδδ′ψχ−(β′2)2</td>
<td>10*</td>
<td></td>
<td>790.7</td>
<td></td>
<td>Replicates <em>E. coli</em> chromosome</td>
<td>[8]</td>
</tr>
<tr>
<td><strong>Pol III core</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>DNA synthesis and proofreading DNA polymerase 3′–5′ exonuclease</td>
<td>[9,10]</td>
</tr>
<tr>
<td>α</td>
<td>dnaE</td>
<td></td>
<td></td>
<td>129.9</td>
<td>τ, ε, β2, UmuD</td>
<td></td>
<td>[11-19]</td>
</tr>
<tr>
<td>ε</td>
<td>dnaQ</td>
<td></td>
<td></td>
<td>27.1</td>
<td>α, θ, β2</td>
<td></td>
<td>[14-16,20-25]</td>
</tr>
<tr>
<td>θ</td>
<td>holE</td>
<td></td>
<td></td>
<td>8.8</td>
<td>ε</td>
<td>Stimulates exonuclease activity</td>
<td>[25-30]</td>
</tr>
<tr>
<td><strong>Clamp loader</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Loads β2 clamp, organises Pol III HE ATPase, dimerizes Pol III cores</td>
<td>[31]</td>
</tr>
<tr>
<td>τ/γ</td>
<td>dnaX</td>
<td>140</td>
<td></td>
<td>71.1/47.5</td>
<td>δ, δ′, α, τ/γ, ψ, DnaB</td>
<td>ATPase, dimerizes Pol III cores</td>
<td>[17,32-39]</td>
</tr>
<tr>
<td>δ</td>
<td>holA</td>
<td>930</td>
<td></td>
<td>38.7</td>
<td>τ/γ, δ′, β2</td>
<td>ATPase, opens β2 clamp</td>
<td>[33,38,40-42]</td>
</tr>
<tr>
<td>δ′</td>
<td>holB</td>
<td>140</td>
<td></td>
<td>36.9</td>
<td>τ/γ, δ</td>
<td>ATPase</td>
<td>[33,40,41]</td>
</tr>
<tr>
<td>χ</td>
<td>holC</td>
<td>1200</td>
<td></td>
<td>16.6</td>
<td>ψ, SSB</td>
<td>Bridges clamp loader to SSB</td>
<td>[43,46]</td>
</tr>
<tr>
<td>ψ</td>
<td>holD</td>
<td>340</td>
<td></td>
<td>15.2</td>
<td>χ, τ/γ</td>
<td>Stabilises clamp loader</td>
<td>[41,43-45]</td>
</tr>
<tr>
<td><strong>Sliding clamp</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tethers polymerase on DNA</td>
<td></td>
</tr>
<tr>
<td>β</td>
<td>dnaN</td>
<td>300c</td>
<td></td>
<td>40.6</td>
<td>α, δ, ε, Hda, MutS, MutL, ligase A, Pol I, Pol II, Pol IV, Pol V</td>
<td>Processivity factor</td>
<td>[47-55]</td>
</tr>
<tr>
<td>DnaB</td>
<td>−</td>
<td>15–20</td>
<td>dnaB</td>
<td>52.4</td>
<td></td>
<td>5′–3′ helicase, activates DnaG</td>
<td>[37,56-64]</td>
</tr>
<tr>
<td>DnaG</td>
<td>−</td>
<td>50–100</td>
<td>dnaG</td>
<td>65.6</td>
<td></td>
<td>RNA primer synthesis</td>
<td>[46,65-70]</td>
</tr>
<tr>
<td>Pol I</td>
<td>−</td>
<td>400</td>
<td>polA</td>
<td>103.1</td>
<td>β2</td>
<td>Primer removal and gap filling</td>
<td>[9,71,72]</td>
</tr>
<tr>
<td>Ligase A</td>
<td>−</td>
<td>300</td>
<td>lig</td>
<td>73.6</td>
<td>β2</td>
<td>Joins nascent DNA fragments</td>
<td>[9,72-75]</td>
</tr>
<tr>
<td>SSB</td>
<td>−</td>
<td>~1000–2000d</td>
<td>ssb</td>
<td>18.8*</td>
<td></td>
<td>Protects ssDNA</td>
<td>[46,76-89]</td>
</tr>
</tbody>
</table>

* May be up to 20 molecules/cell. † Including 40 free Pol III cores/cell. ‡ As dimers of β. § As tetramers of SSB in wild-type *E. coli* cells during mid-log phase. ° As a monomer.
Our focus is on the *E. coli* replisome as a representative of those of other eubacteria. Recent work reviewed, for example in [90], indicates that all bacterial species use highly-conserved replisomal components to carry out conserved functions [91,92], but a picture is emerging where the hierarchy of protein interactions varies such that interactions that are strong in some species may be weaker in others, with the necessity for a multiplicity of interactions to enable dynamics while preserving replisomal integrity being satisfied in different ways. Access to three-dimensional (3D) structural information for most of the individual proteins and domains of the *E. coli* replisome over the last twenty-five years, combined with the ability to faithfully reconstitute replisomal complexes on various DNA templates *in vitro*, has led to the discovery of many dynamic protein–protein and protein–DNA interactions at the replication fork. This Chapter explores the assembly of replisomal subunits, their structures and their dynamic behavior.

### 1.2 Initiation at oriC, the origin of replication

Initiation of bacterial DNA replication occurs through highly coordinated assembly of nucleoprotein complexes at the unique origin of replication (*oriC*). DnaA is the specialist initiator protein that recognizes the origin and guides the formation of these nucleoprotein complexes [93-95]. To initiate DNA replication, this complex must be activated and undergo large conformational changes that result in the local unwinding of duplex DNA. DnaB, the ring-shaped replicative DNA helicase that unwinds double-stranded (ds) DNA, is opened and loaded onto the single-stranded (ss) DNA bubble ("open complex") from a complex with the helicase loader, DnaC. This loading requires DnaB-bound DnaC and *oriC*-bound DnaA to form a prepriming complex that eventually results in dissociation of DnaC and loading of DnaB to encircle the ssDNA. The helicase must then migrate along the ssDNA, opening the region to expose enough ssDNA for DnaG primase to interact with it to produce an RNA primer for the replicative polymerase, the DNA polymerase III (Pol III) holoenzyme (HE), to assemble on each replication fork to complete replisome assembly. The mechanics of initiation of DNA replication at *oriC* are described in detail in [96].

### 1.3 The primosome, DnaB and DnaG

DnaB is the replicative helicase and the first replisomal protein to assemble at the replication fork [97]. Following initial strand separation at *oriC* by DnaA, a DnaB molecule is loaded onto each of the single strands of the open complex. Once loaded, each DnaB initiates ATP-hydrolysis dependent unwinding of the duplex parental strands by translocating in the 5’–3’ direction on what will become the lagging-strand
template strands of the two replication forks that will invade the duplex DNA on either side of the origin. The result of translocation of DnaB (at a rate up to ~300 bp/s [98]) is progressive separation of the parental dsDNA into two ssDNA templates [99].

In *E. coli* and all other eubacteria, DnaB is a toroid of six identical 52-kDa subunits oriented in the same direction. The formation of a functional *E. coli* homohexamer is Mg²⁺-dependent, unlike bacteriophage T4 helicase that requires binding of ATP or GTP for hexamerisation [100]. DnaB has two domains. In *E. coli*, the N-terminal domain consists of a short, probably unstructured, region followed by a helical domain that although monomeric in isolation [60,61] forms a trimer of dimers in the hexamer to generate sites for interaction with up to three molecules of DnaG primase [62,68,101]. The C-terminal RecA-like motor domains of DnaB oligomerize as a ring-shaped hexamer to generate potentially six catalytic sites for ssDNA-dependent ATP hydrolysis [102,103]. Biochemical studies suggest that ~20 nucleotides (nt) of ssDNA pass through the central channel of DnaB when it is loaded on a forked DNA molecule with a 50-nt ssDNA tail [104-106]. In this configuration, the C-terminal motor domains are at the apex of each replication fork, and the trailing N-terminal domains are positioned to interact with DnaG to synthesize RNA primers complementary to the strands on which they are translocating. The first primer synthesized on each strand at oriC becomes the leading strand primer for the other replication fork, and subsequent primers initiate Okazaki fragment synthesis on the lagging strand.

Despite there being no available high-resolution 3D structure of the functional DnaB hexamer from *E. coli* alone or in any protein complex, several structures of full-length versions or domains of DnaB and complexes from other bacterial species are available [101,107-110]. The recent cryo-electron microscopy (cryo-EM) structure of the *E. coli* DnaB₆–(DnaC)₆ complex determined at 25 Å shows an open helical arrangement of DnaB–DnaC heterodimers, suggesting a mechanism for loading of the helicase at a nascent replication fork [111].

Early negative stain EM studies established the hexameric ring structure of *E. coli* DnaB₆ but raised questions about the exact conformation of the ring. Both C₃ and C₆ rotational symmetries were observed regardless of the presence of nucleotide cofactors [103,112,113]. Re-examination of DnaB alone with a bound ATP analog (ADP•BeF₃) by cryo-EM (at 25 Å) suggested C₃ symmetry (in the N-terminal domains) produced shapes most closely matching the reference-free 2D class averages and known crystal structures of other DnaB homologs [111]. All of these structures show a
wide open (dilated) central channel with the N-terminal domains appearing as a trimer of dimers (C₃ symmetry) while the C-terminal RecA-like motor domains have pseudo-six-fold symmetry. Subsequently, the Berger group reported the crystal structure of *Aquifex aelicus* DnaB, which showed a quite different arrangement of the N-terminal domains. Although their 3-fold symmetry is maintained, the central channel is severely constricted. This suggests that DnaB can adopt different arrangements of the N-terminal domains when it interacts with its different binding partners, and with different nucleotides, implying that transition between constricted and dilated structures may occur as DnaB translocates on ssDNA [114].

A much earlier structure of the N-terminal domain of DnaB (residues Lys24–Ser136) determined in solution by NMR [61] and the crystal structure of a similar domain construct [60] showed a dimer interface that assembles around Phe102, which was initially thought to contribute to dimerization of the N-terminal domain in the C₃ symmetrical state of the DnaB hexamer [115]. However, the crystal structures of full-length DnaB homologs and the recent cryo-EM structures disprove this, and the existence of the C₆ structure in the early micrographs appears now to be an artefact. In fact, Phe102 is now predicted to be part of the interface with the DnaG primase (Figure 1.2b, inset).

**Figure 1.2 | Homology model of the *E. coli* DnaB₆–(DnaGC)₃ primosomal complex based on the corresponding structure from *Geobacillus stearothermophilis* [101].** (a) Top view of a cartoon representation of the DnaB₆–(DnaGC)₃ complex showing the three DnaG molecules (red) bound at the periphery of DnaB’s N-terminal domains (blue). (b) Side view, rotated 90° around the y-axis of the view in a. The inset in b shows a zoomed view of the predicted interface between DnaGC (red) and DnaB (light blue), showing the position of Phe102 of DnaB (purple sticks). From an homology model prepared in collaboration with Thomas Huber, Australian National University. This and all later structural figures were prepared using PyMOL v1.8.0.5 (Schrödinger).
The translocating helicase interacts through its N-terminal domains with the primase, the specialist RNA polymerase that synthesizes short RNA primers on the ssDNA regions produced by helicase action [116-118]. In *E. coli*, the DnaB–DnaG interaction is weak and transient, though in some other species like the Firmicute *Geobacillus stearothermophilus*, it is stable and the complex can be crystallized [101] (Figure 1.2). Bacterial DnaG primases contain three domains that are linked together by regions predicted to be flexible. First is the N-terminal zinc-binding domain (ZBD), which is followed by the central catalytic RNA polymerase domain (RPD) and then the C-terminal helicase-binding domain (HBD or DnaGC). The ZBD contains a zinc-ribbon motif, which coordinates a Zn$^{2+}$ ion [119]. The ZBD is essential for primase activity and is thought to recognize trinucleotide priming sequences in the ssDNA template [120,121]. The bacterial RPD domains share high structural similarity with types IA and II topoisomerases and primase-like proteins (TOPRIM fold) [122,123], unlike those from evolutionarily and structurally distinct archaeo-eukaryotic primase RPD domains [124]. Within the RPD are two conserved catalytic motifs [67], one containing a conserved Glu presumed to act as a general base during ribonucleotide incorporation. The other contains two strictly conserved Asp (DxD) residues required for Mg$^{2+}$-mediated NTP-binding [125,126]. DnaG primase binds through the DnaGC domain to the N-terminal domain of the DnaB helicase to form the primosomal complex DnaB$_6$–(DnaG)$_3$ [62,68].

Although the ZBD and RPD are sufficient to catalyse template directed oligonucleotide synthesis in phage T7, *E. coli* and *A. aeolicus* systems [127-129], coordination of all three domains is required to allow effective initiation of *de novo* RNA primer synthesis. DnaGs preferentially utilize defined trinucleotide sequences in DNA. For *E. coli*, this is 5'-dCTG in the leading-strand and 5'-dC($^A$/$^T$)G in the lagging-strand templates [120,130,131]. Trinucleotide specificity varies somewhat among species, *e.g.*, primases from Firmicutes preferentially initiate synthesis from 5'-dCTA [132]. Acting alone, DnaG is slow to synthesize a 12–16 nt RNA primer once every 1000 s *in vitro* [132,133], yet much higher rates are required to support *in vivo* Okazaki fragment synthesis. The presence of DnaB stimulates DnaG activity to *in vivo* rates and limits the length of RNA primers to 10–14 nucleotides with 12 nucleotides being the predominant RNA primer length [131]. The ZBD recognizes and interacts with the trinucleotide recognition sequences, bringing the templates in close proximity to the RPD for RNA primer synthesis. This RNA priming activity is specifically stimulated by direct interaction with the N-terminal domain of DnaB through DnaGC [131]. Although *E. coli* DnaG has been shown to be monomeric [119], it is thought to bind to template...
DNA as a dimer [134] and to function in trans with another DnaG protomer bound to the DnaB hexamer, i.e., the ZBD of one protomer recognises the priming site in the DNA template while the RPD of the other protomer synthesises the primer [128].

There is no high-resolution structural information on a full-length primase from any eubacterial species, but structures of each of the individual domains of DnaG have been determined either by X-ray crystallography or NMR spectroscopy. These structures include the RPD and DnaGC domains from *E. coli* and other bacterial species, as well as the ZBD from *G. stearothermophilus* and RPD/ZBD from *A. aeolicus* [39,62,67,123,128,135,136]. Without a full-length DnaG primase structure, how each of the three domains in DnaG cooperate to enable efficient DnaB-dependent primer synthesis remains speculative.

The next step in replisome assembly at the fork requires the loading of the replicative polymerase, the Pol III holoenzyme (HE) at the primer termini. The polymerase never needs to dissociate from the template as it continuously and processively extends the first primer laid down on the leading strand at oriC. The situation on the other strand is quite different: replication of the lagging-strand needs to occur discontinuously, generating Okazaki fragments 1–2 kb in length that need to be processed by DNA polymerase I and joined by DNA ligase [9]. This means that lagging-strand, unlike leading-strand synthesis, generates considerable amounts of ssDNA, which is coated by single-stranded DNA-binding protein (SSB). The lagging-strand ssDNA is produced by helicase action, and consumed by DNA synthesis by the polymerase, and in a fully coupled situation where the rates of the two enzymes should be near identical, the amount of ssDNA at each replication fork would be nearly constant at about half of the average Okazaki fragment length, or 0.5–1 kb. Thus, the amount of SSB at each fork should also be relatively constant. It is appropriate at this stage to consider the properties of SSB and its still incompletely understood roles at the replication fork.

### 1.4 The role of SSB at the replication fork

The basic function of SSB is to bind preferentially to ssDNA with high affinity in a sequence-independent manner [77]. By doing this it is able to protect ssDNA from nucleolytic digestion and prevent intra-strand pairing such as hairpin formation, so as to preserve ssDNA in a conformation suitable for the action of DNA metabolic enzymes [87]. But SSB is also a pivotal interaction hub, interacting with a large number of proteins, directing and organizing them to sites of DNA replication, recombination and repair (reviewed in [87]). Historically SSB has been viewed as only providing inert
protection to regions of ssDNA. However, there is increasing evidence that SSB–ssDNA complexes are highly dynamic and have great functional significance. To illustrate, SSB is able to rapidly sample alternative binding topologies without dissociation [137,138], and to utilize a direct transfer mechanism between ssDNAs, hypothesized to enable recycling of SSB tetramers from old to new lagging-strand templates at the replication fork [139]. Earlier studies also demonstrated SSB’s role in RNA priming by DnaB–DnaG, which occurs randomly on naked ssDNA but is restricted, e.g., to defined origins of replication on SSB-coated single-stranded phage DNAs [140].

In E. coli, SSB is the product of the ssb gene, which is essential for cell viability [76,141,142]. SSB forms a stable homotetramer of 177 amino acid subunits [76], separated into two distinct domains. The N-terminal domain (112 residues) is structured, and forms a classic oligonucleotide (oligosaccharide)-binding (OB) fold responsible for ssDNA binding [81,143,144]. Several X-ray crystal structures of the OB-domain of E. coli SSB have been solved (Figure 1.3), both in isolation and bound to two 35-mer ssDNAs per tetramer [81,145,146]. The four ssDNA-binding domains in the tetramer enable it to bind tightly to ssDNA in a variety of modes with different binding properties depending on the monovalent or divalent cation concentrations, and input stoichiometry. The dominant binding modes observed in in vitro studies are referred to as SSB₆₅, SSB₅₆ and SSB₃₅ [147,148]. The subscript number reflects the average number of nucleotides bound by each tetramer [16,149-153]. The SSB₆₅ mode (Figure 1.3b), favoured by moderately high salt concentration (>2 mM Mg²⁺ or >200 mM Na⁺), utilizes all four ssDNA binding sites and exhibits little cooperativity between neighbouring ssDNA-bound tetramers [81,154-156]. The SSB₅₅ binding mode, on the other hand is favoured in low salt concentrations (<10 mM Na⁺) and is typified by high cooperativity of binding of neighbouring tetramers [152,154,155,157]. These two major binding modes and the inter-conversion between them have been visualised in real time using single-molecule fluorescence resonance energy transfer, smFRET [138], and SSB tetramers have been shown to translocate on ssDNA in the SSB₆₅ mode by a reptation mechanism [158].
Figure 1.3 | Structure of the E. coli SSB tetramer. a, Crystal structure of E. coli SSB (PDB: 1SRU). Each SSB monomer (residues 1–112; OB-fold) is depicted as a cartoon representation in a different colour, each with a disordered C-terminal tail (residues 113–177; not seen in the crystal structure). b, Model of the SSB tetramer in the 65-nt ssDNA binding mode, utilizing all four ssDNA binding sites on OB-folds of all four protomers for DNA binding and exhibiting little cooperativity between neighbouring SSB tetramers. For clarity, ssDNA is depicted as sticks (red) and disordered C-terminal segments including the conserved C-terminal tails (SSB-Ct) are omitted. Coordinates for the model [81] were generously provided by Gabriel Waksman (University College London and Birkbeck College, University of London).

In contrast, the C-terminal residues are not visible in the crystal structure of full-length SSB bound to ssDNA [145] and are largely disordered in solution [157]. The last eight amino acids at the C-terminus form a highly conserved acidic tail (SSB-Ct) that is the site of binding and recruitment to many of SSB’s protein partners to ssDNA [77,86,143,159-161]. Key binding partners in the replisome include DnaGC, the C-terminal helicase-binding domain of primase [70], and the χ subunit of the Pol III clamp loader complex [45,160,162,163]. In the absence of ssDNA, the SSB-Ct is thought to be sequestered from interactions with binding partners by interaction with the DNA-binding groove of the OB-domain, from which it is necessarily released on binding to ssDNA [143,157,160,164-167].

Given the ability of SSB to translocate on ssDNA, it seems likely that equilibration between the two major binding modes occurs rapidly under physiological conditions depending on the availability of ssDNA; conversion of SSB-coated DNA from the SSB35 to SSB65 mode could occur, for example, to grant access of proteins like the Pol III core or DnaB to ssDNA. Moreover, in the SSB35 mode, two high-affinity ssDNA sites are occupied and the other two-lower affinity sites are free. This provides a ready means of rapid transfer of SSB in this mode between two different ssDNA segments through a transient paired intermediate, simply by switching the affinity between pairs of DNA-binding sites [138,166]. It is thus tempting to speculate that cooperative transfer of SSBs in this mode could occur from in front of the Pol III core to behind the helicase on
13

the lagging-strand, playing a role in coupling of the two activities. It is not currently known how an advancing polymerase displaces SSB, or whether this is an active process.

1.5 The elongation stage

Watson and Crick first modelled the structure of dsDNA in 1953. Their proposal that DNA is arranged in an antiparallel helical structure composed of two stands led them directly to hypothesize that it is replicated in a semi-conservative fashion [168,169]. This meant that, upon separation of the two strands, each daughter cell contains dsDNA composed of parental (template) and complementary newly synthesized (nascent) polynucleotide chains. This hypothesis was proven by an elegant experiment conducted by Meselon and Stahl using an isotope of nitrogen to distinguish between parental and daughter stands of DNA [170]. During semi-conservative DNA replication, both strands are replicated simultaneously by polymerases that are believed to be physically linked; thus the replication machinery is faced with a directionality problem where the two DNA strands have opposite polarity. To explain how concerted replication of both strands can be physically coupled, a “trombone model” was proposed by Bruce Alberts [171,172]. In this model, the lagging strand at the replication fork forms a loop to reorient the lagging-strand polymerase so it advances in parallel with the one on the leading strand. This trombone loop grows and collapses during each cycle of Okazaki fragment synthesis. Hamdan and co-workers [173] have visualised the production and release of these replication loops in real time by the phage T7 replisome using single-molecule flow-stretching assays.

In *E. coli*, the replication forks move at close to 1000 bp/s, as revealed at the single-molecule level by both *in vitro* [14,174] and *in vivo* [175] studies. These rates are consistent with the need for each replisome to copy half of the 4.6 Mbp chromosome in ~40 min. Thus, each 1–2 kb Okazaki fragment is made in 1–2 s, which needs to include the time required for primer synthesis by DnaG (as discussed in [176]). We will return to consideration of the timing of lagging-strand priming later (Section 1.6.1).

1.5.1 Assembly and action of the DNA polymerase III holoenzyme

The last step in replisome assembly at *oriC* is loading of the chromosomal replicase, the DNA Pol III HE onto the first leading-strand primer. The Pol III HE is configured to be capable of simultaneous synthesis of both the leading and lagging strands. It is a remarkably efficient enzyme (its turnover number approaches 1000 per second per active site), it functions with almost perfect fidelity and achieves almost indefinite
processivity (never dissociating from primer–template DNA until synthesis is finished). The architecture of the *E. coli* replication fork is depicted in Figure 1.1 and the roles of the ten different Pol III HE subunits are listed in Table 1.1.

Unlike other macromolecular machines such as the ribosome, isolation of intact functional replisomes from cells has proven impossible; they need to be assembled for *in vitro* study from individual components. Although this reductionist approach makes for relatively slow progress, the meticulous determination of the exact composition of replisomes at a replication fork is now within our grasp. Nevertheless, with care [8,177,178], the Pol III HE can be isolated directly from cells and has an average composition: \((\alpha\varepsilon\theta)_2\alpha_2\gamma_2\delta'\psi_\chi_2\). For reasons described below, there has been some uncertainty whether this represents a single entity with a single defined composition [179]. For *in vitro* studies, the holoenzyme is conveniently assembled from separately overproduced and purified subunits, each of which can be made into stable and functionally distinct subassemblies – where \(\alpha\varepsilon\theta\) is the Pol III core, \(\beta_2\) is the homodimeric sliding clamp, and the clamp loader (or DnaX) complex (CLC, reviewed in [180]) has the composition \(\tau_{n}(3-n)\delta'\psi_\chi_2\) (where \(n = 0 \) or \(1\) in non-physiological forms, or \(2\) or \(3\) in the holoenzyme).

At least two Pol III cores are responsible for simultaneous replication of the leading and lagging strands. However, by themselves, the Pol III core is neither processive (10–20 nt [181]) nor rapid (\(~20\) nt/s, [13]) in DNA synthesis. It is the \(\beta_2\) clamp that, once assembled onto DNA by the ATPase activity of the clamp loader complex (CLC), confers efficient synthesis on all Pol III core subassemblies [182], e.g., the Pol III HE can replicate a 5 kb circle in a few seconds without dissociating [181], and more recent *in vitro* single-molecule measurements demonstrated processivities in excess of 100 kb [183,184]. Structures and functions of the three subassemblies of Pol III HE, and their variants, are discussed below.

### 1.5.2 The \(\beta_2\) clamp

Each round of chromosomal replication requires millions of nucleotides to be copied accurately to duplicate the parental DNA strand. Nevertheless, each *E. coli* cell is estimated to contain only 10–20 molecules of Pol III HE [8,9]. This low level implies that leading-strand synthesis must be a highly processive process where each holoenzyme at a replication fork can copy half of the chromosome without ever dissociating, or doing so only rarely. On the other hand, it also requires frequent exchange or recycling
of the polymerase from within the replisomal machinery on the lagging-strand during repeated Okazaki fragment synthesis.

The DNA polymerase duty cycle is composed of repetitive nucleotide incorporation and DNA translocation steps. It is associated strongly with the primer–template DNA and the complementary nucleotide to be incorporated in a ternary complex during the nucleotide incorporation phase, and its generally weaker and non-specific interaction with just the primer–template enables it to easily migrate along it during the translocation phase. Therefore, to reach the high processivity required for efficient chromosome duplication, the polymerase must be stably bound to the DNA template during both phases without compromising its ability to rapidly migrate along it between successive nucleotide incorporations. This contradiction is elegantly resolved by the polymerase binding to homodimeric ring shaped molecules known as the \( \beta_2 \) clamps that topologically bind (encircle) DNA in a sequence-independent manner, allowing them to slide freely along dsDNA [51,182].

The \( \beta \) clamp subunit is the 41 kDa product of the \( dnaN \) gene [8,185]. Early biochemical data demonstrated that the observed dependence of the longevity of the \( \beta_2 \)-DNA interaction on the topological state of DNA was not through the usual chemical forces but was mediated by its shape. It was proposed that sliding clamps encircle and freely diffuse along DNA in a sequence-independent manner, tethering their respective DNA polymerases to the DNA to enable high processivity [182]. This hypothesis was confirmed in 1992 when the first X-ray crystal structure of the \( E. coli \) \( \beta_2 \) clamp was reported [51] (Figure 1.4). The structure revealed two semicircular subunits arranged in a head-to-tail configuration to form a ring. The inner diameter of the ring is ~35 Å, large enough to encircle dsDNA. The head-to-tail arrangements of the subunits creates two structurally distinct faces, one of which contains the carboxyl termini and serves as an interaction hub for other proteins including the Pol III core [11,14,72,182,186,187]. Each \( \beta \) monomer is comprised of three domains with similar 3D folds and identical chain topology, although their amino acid sequences are quite distinct. Hence \( \beta_2 \) has pseudo six-fold symmetry (Figure 1.4a). The inner surface of the ring is lined with 12 \( \alpha \)-helices while the outer surface is composed of \( \beta \)-sheets.
**Figure 1.4 | The β₂ sliding clamp in complex with primer–template DNA.**

- **a.** Front view of the crystal structure of the *E. coli* β₂ sliding clamp in complex with primed DNA (PDB: 3BEP). Each β protomer in the dimer is depicted as a cartoon representation and coloured red (protomer A) or white (protomer B); bound primed DNA is shown as black sticks.  
- **b.** Side view, rotated 90° around the y-axis of the view in a. The dsDNA is tilted ~22° from the C₂ rotational axis of β₂. The N- and C-terminal faces are labelled.  
- **c.** Expanded view of a: residues Arg24 and Gln149 make direct interaction with dsDNA protruding from the loops on the C-terminal face, creating a positively charged channel.  
- **d.** The Pol III CBM peptide bound to β₂ (PDB: 3D1F). The solvent exposed surface of the β protomer is coloured white and the Pol III CBM is shown as pink sticks. The CBM binding pocket is organised into two subsites: subsite I (cyan) and subsite II (purple).

The ability of the clamp to slide freely on DNA in a sequence-independent manner is mediated though hydrogen-bonded water molecules and the electrostatic potential at the inner surface, which is strongly positive even though the clamp as a whole has a net negative charge [188]. The structure of the β₂ clamp bound to primed DNA has also been solved [52], revealing direct interaction with the DNA duplex is primarily through
Arg24 and Gln149 on protruding loops of the C-terminal face (Figure 1.4c). Interestingly, the DNA is highly tilted at 22° as it passes through the central channel (Figure 1.4b). Additionally, a crystal contact with an overhanging single-stranded portion of the DNA template is formed at the hydrophobic protein-binding pocket described in detail below [52]. It may be that in the absence of this connection during replication, $\beta_2$ may diffuse away from the 3’ terminus of the primer–template junction and that this ssDNA connection may act as a proxy to keep it at the primed site where it is needed for its interaction with polymerases. The symmetrical dimer interfaces that maintain the $\beta_2$ ring comprise six intermolecular ion pairs and four hydrogen bonds [42,51]. In solution the clamp exists as a closed ring with a measured $K_D$ for subunit dissociation < 60 pM [189,190]. This arrangement requires that one of the two dimer interfaces needs to be opened by the CLC to enable $\beta_2$ to be loaded at a primer terminus. Conditions are known, however, that enable subunit exchange to occur on reasonable time scales in the absence of the CLC, and these have been exploited to make hemi-mutant $\beta_2$ clamps for functional studies [14].

Enzymes that are central to the regulatory network intersecting DNA replication, recombination and repair are known to interact with the $\beta_2$ clamp on dsDNA; these include the CLC, the $\alpha$ and $\varepsilon$ subunits of the Pol III core and all other E. coli DNA polymerases (Pols I, II, IV and V), the mismatch repair proteins MutL and MutS, DNA ligase, and the Hda cell-cycle regulatory factor [11,14,50,72,182,186,187]. Thus the $\beta_2$ clamp can be seen as an interaction hub for proteins on dsDNA [92,191]. These protein interactions with the $\beta_2$ clamp all occur, at least partially, at the same two symmetry-related sites in the $\beta_2$ dimer (Figure 1.4d), through short penta- (optimally QLS/DFL) or hexa-peptide (broad consensus QxxΦxΦ, where x is any residue and Φ is hydrophobic) clamp-binding motifs (CBM) found in disordered segments or loops of $\beta_2$-binding proteins [11]. Many crystal structures of the E. coli $\beta_2$ clamp in complex with CBM peptides [53,192-194] and some larger protein domains including that from Pol IV [195] and the $\delta$ subunit of the clamp loader [42] have been determined, and an ordered binding mechanism of CBM peptide recognition has been probed crystallographically [193].

Structures of $\beta_2$ clamps from several other bacteria all show close structural similarity to E. coli $\beta_2$ [196-200]. Indeed, the archaeal and eukaryotic $\beta_2$ clamps, called PCNA (for proliferating cell nuclear antigen), have very similar structures, except that the pseudo-
six-fold symmetric rings are made up of three two-domain subunits, so they contain three symmetry-related “PIP-box” protein-binding sites [201].

In passing, we note that because the common protein binding sites in bacterial sliding clamps are conserved and participate in so many essential interactions, they represent excellent target sites for development of new antibacterial drugs [92]. Evolution of resistance to such drugs through target mutation would require simultaneous compensatory mutations in many $\beta_2$-binding partners, which is statistically highly improbable. Drug discovery efforts so far have produced additional structures of inhibitory compounds bound to the *E. coli* $\beta_2$ clamp [50,53,202-204]. Moreover, the target for griselimycin, a natural antibiotic with promise as a lead compound for treatment of tuberculosis, has recently been shown to be the *Mycobacterium tuberculosis* $\beta_2$ clamp [199]. Resistance, when forced to develop in the laboratory strain *Mycobacterium smegmatis*, occurred at very low frequency and by the unusual reversible amplification of the chromosomal region containing oriC as well as the dnaA and dnaN genes.

1.5.3 The Pol III core, $\alpha\varepsilon\theta$

The catalytic core of the Pol III HE consists of the $\alpha$, $\varepsilon$ and $\theta$ subunits in 1:1:1 stoichiometry [178]. The Pol III core ($\alpha\varepsilon\theta$) is a separately isolable complex that remains intact on the lifetime of *E. coli* chromosomal duplication. *In vivo* live cell imaging with fluorescent proteins show there are usually (approximately) three $\alpha\varepsilon\theta$ cores at each replication fork [205]. One is believed to be dedicated to leading-strand DNA synthesis, while the other (approximately) two may both participate in replication of the lagging-strand [206,207].

The large $\alpha$ subunit of the Pol III core is a family C polymerase from the DnaE family, and $\varepsilon$ is a separate 3’-5’ proofreading exonuclease subunit from the DnaQ family. The small $\theta$ subunit has a role in stabilizing $\varepsilon$ [27,29], but only exists in some bacterial species [191]. The $\alpha$ subunit catalyses the template-directed nucleotidyl transferase reaction that adds nucleotides to the 3’-OH of the growing DNA strand. The *E. coli* $\alpha$ subunit is the 130 kDa product of the *dnaE* gene and is organised in domains that confer its special properties. The N-terminal PHP (polymerase and histidinol-phosphatase) domain is a vestigial 3’-5’ exonuclease domain that is still functional as a proofreader in some species [19,208-211]. In *E. coli*, however, it has evolved to be the site of interaction of the separate $\varepsilon$ proofreading subunit [15,212]. Following the PHP
domain are the palm, thumb and fingers domains that characterize all DNA polymerases. The palm domain acts to identify the correct incoming nucleotide and contains two Mg²⁺ ions coordinated by three essential aspartic acids [213], forming the basis for the common two-metal catalytic mechanism of all DNA polymerases [214]. This mechanism has recently been captured directly by a detailed structural study of the DNA synthesis reaction by the error-prone human DNA polymerase η [215]. The thumb domain contacts the nascent dsDNA upstream of the active site. The extended fingers domain, adjacent to the palm, has direct contacts with the incoming nucleotide and, as in polymerases from other families [216,217], undergoes significant conformational changes once the nucleotide has been positioned near the Mg²⁺ ions, template and 3’ end of the nascent strand [216-218]. At the tip of the fingers domains is a conserved CBM that directly contacts the β2 clamp to ensure high processivity. Following this is an OB-fold domain that may interact with the ssDNA template [219-221] and the C-terminal β-binding domain [180,222] that interacts tightly with the C-terminal domain of a τ (but not γ) subunit of the CLC (Figure 1.5c) [17,Gao:2001gu 39].

There is no crystal structure available of the full-length E. coli α subunit; however, there is an X-ray structure of residues 1–917 that extends through the fingers domain, but lacks the internal CBM, OB-fold and τ-binding domains [19]. The full-length crystal structure of the (structurally) closely related Thermus aquaticus α (also a DnaE-type C family polymerase), both by itself [208] and in the ternary complex with primer–template DNA and incoming nucleotide has been solved [217]. The structures reveal a high degree of similarity in the basic architecture between the two Gram-negative (family C, DnaE-type) replicative polymerases. A large conformational change towards a more compact “closed” state is observed in T. aquaticus α upon DNA binding, primarily due to the large rotation in the fingers domains that repositions the (internal) β2-binding site located at its tip. A similar conformation is observed in the structure of a C family polymerase from Gram-positive Geobacillus kaustophilus PolC bound to both primer–template and incoming dNTP [223], suggesting a general tendency for transition toward a more “closed” state upon primer–template DNA and nucleotide binding [224,225]. Interestingly, although the OB-fold domains in DNA-bound structures from T. aquaticus and G. kaustophilus are located on the opposite sides of polymerase domains, both appear well positioned for delivery of the ssDNA template strand to the polymerase active site.
The θ subunit also binds to the N-terminal domain of ε [226]. A solution structure of the
ε186–θ complex (Figure 1.5b) has been determined by NMR and reveals a
hydrophobic interface responsible for the stable interaction of the two subunits
[28,227,228]. On the other hand, the unstructured εCTS contains the hexapeptide CBM
(residues Gln182–Phe187) that provides the second contact of the Pol III core with β2
[14,187], followed by the region through which ε interacts with the PHP domain of α
(Figure 1.5a) [15,16,212,226].

Jergic and co-workers uncovered a non-proofreading activity of ε [14]. Through
rigorous bulk biochemical and single-molecule studies, this activity was revealed to be
the result of interaction of the CBM in ε with one of the protein-binding sites in the β2
clamp. In conjunction with interaction of the internal CBM of α with the other symmetry-
related site in the clamp, this interaction maintains a closed state of the αεθ–β2
replicase complex in the polymerisation mode of DNA synthesis. The ε–β interaction,
which appears to also be maintained during proofreading [187], has apparently evolved
to be weak (K_D ~200 μM). This would enable its transient disruption, presumably to
facilitate access of alternative polymerases or other β2 clamp binding proteins, for
example during lesion bypass and DNA repair [229]. It is tempting to speculate that its
weak nature enables the transition between polymerisation and proofreading modes
whereby release of CBM in ε from β2 would coincide with transfer of the DNA template
between the active sites of α and ε.
Figure 1.5 | A model of the structural organisation of the E. coli αθ–β2–DNA complex [15]. Individual subunits α (orange), ε (blue), θ (purple), and β2 (red) are shown in cartoon representation and DNA is represented as black sticks. Inset a, Crystal structure of αPHP–εCTS fused via a flexible amino acid linker (PDB: 4GX8). The αPHP (orange) and εCTS (blue) domains are shown as cartoon representations. Residues 206–209 of εCTS that are not seen in the crystal structure are presented as blue dots. The εCTS is intrinsically unstructured in solution but adopts an α-helical structure on binding to the PHP domain of α. Inset b, NMR solution structure of the ε186–θ complex (PDB: 2XY8). A superposition of the 10 lowest-energy conformers is shown, and ε186 (blue) and θ (purple) are depicted as a cartoon representation. Active site residues in ε (Asp12, Glu14, Asp103, His162 and Asp167) are shown as red sticks. The θ subunit binds on the opposite side to the active site of ε at a hydrophobic interface. Inset c, Solution structure of the folded core of domain V of E. coli τ derived by deletion of the 18 C-terminal residues from domain V (PDB: 2AYA). A superposition of 20 lowest-energy NMR conformers (green cartoons) is shown. Domain V of τ interacts tightly (Kd ~260 pM) with the C-terminal region of α [17]; the last 18 residues of domain V of τ are intrinsically unstructured and are required for interaction with α. Secondary structure prediction and mutagenesis studies suggest formation of an α-helix upon binding to α [17].
The crystal structure of the $\alpha$PHP domain fused via a flexible linker to the proximal 44 C-terminal residues (Ser200–Ala243) of $\varepsilon$ reveals the $\varepsilon$CTS adopts an extended structure across one face of the PHP domain followed by an $\alpha$-helix (Figure 1.5a), while NMR experiments showed the intervening $\sim$14 residues between the CBM and $\alpha$-binding region of the $\varepsilon$CTS remain flexible even in the $\alpha\varepsilon\theta$–$\beta_2$ complex [15]. This region is predicted to contain a defined structural element termed a Q-linker, whose role is proposed to be to tether structurally distinct domains of proteins for multi-protein interactions [230]. This flexibility together with the high-affinity binding site of the $\varepsilon$CTS on the $\alpha$PHP domain being remote from the active site of the polymerase raises questions about how the exonuclease domain of $\varepsilon$ gains access to a mismatched primer terminus when proofreading is required. This $\alpha$PHP–$\varepsilon$CTS structure, when combined with the others described earlier and small-angle X-ray scattering data, allowed for rational modeling of the $\alpha\varepsilon\theta$–$\beta_2$ replicase complex with primer–template DNA in the polymerisation mode [15]. As in the structure of the $T. aquaticus$ $\alpha$ complex with primer–template DNA on which the relevant part of the model was based [217], the OB domain of $\alpha$ was positioned in the model to guide the ssDNA template into the polymerase active site. It therefore came as a surprise when a recent 8 Å cryo-EM structure of the complex of $E. coli$ $\alpha\varepsilon$–$\beta_2$ with primer–template DNA revealed a $>30$ Å movement of the OB domain far from the polymerase active site to establish a third non-canonical contact of the polymerase core with the sliding clamp (Figure 1.5) [222].

The gene that encodes the $\theta$ subunit is named $holE$ [29,30]. The protein is devoid of any enzymatic activity. It binds tightly to $\varepsilon$ but has no interaction with $\alpha$ [29,30,231], and its real function at the replication fork is not understood. Whereas $dnaE$ and $dnaQ$ are essential, the $holE$ gene is dispensable in $E. coli$ [26,232,233]. Analysis of $\delta$holE phenotypes in a $dnaQ^{49}$ background demonstrate indistinguishable growth rates compared to wild-type cells, normal nucleoid morphology and only a modest effect on mutation rates [26]. Only in a $dnaQ^{49\#}$ mutator background did $holE$ deletion result in a dramatic increase in mutation frequency, suggesting that the absence of $\theta$ increases the intrinsic thermal instability of the $dnaQ^{49}$ temperature-sensitive version of $\varepsilon$ [27]. In addition to genetic experiments, biochemical studies have shown that association of $\theta$ with full-length $\varepsilon$ promotes its exonuclease activity in vitro, increases its thermal stability, and also prevents its proteolysis both in vitro and in vivo [24,29,226].
The solution structures of θ [234] and of its complex with ε186 [28] have been determined (Figure 1.5b). Additionally, the structure of HOT, the homolog of theta from bacteriophage P1, has been solved in isolation by NMR spectroscopy [235] and in complex with ε186 by X-ray crystallography [236]. Taken together, these structures reveal a hydrophobic subunit interface and conformational invariability of the exonuclease domain of ε upon binding to either θ or HOT. The presence of the hydrophobic interface may suggest that θ (and HOT) may play roles in preventing misfolding and aggregation as a consequence of solvent exposure of hydrophobic residues of ε. Nevertheless, an interesting question still remains why phage P1 encodes HOT, when it relies on most of the E. coli replication machinery (except for DnaB and SSB) for its replication and does not encode any other subunit of Pol III HE [237].

A moonlighting function has recently been attributed to holE [238]. Genetic studies showed a correlation of transcriptional profiles of hha/yledT-like nucleoid-associated regulatory genes with holE. It was shown that holE and yedT down-regulate expression of the tna (tryptophanase) operon. It was hypothesised this may occur through ρ-dependent transcription termination in the tna operon leader region [238].

It is evident through extensive biochemical and structural analysis of the Pol III core that each subunit works cooperatively and stimulates the activity of others. For instance, the α subunit can stimulate the exonuclease activity of ε 10–80 fold through increasing the affinity of ε for a free 3'-OH terminus [239]. The θ subunit is also able to modestly stimulate exonuclease activity, probably by stabilizing ε [29], and association of ε with α induces a 2- to 3-fold increase in the polymerase activity [239]. This functional cooperatively is not a surprise, as most DNA polymerases contain separate intrinsic domains for the polymerase and exonuclease activities in a single polypeptide chain [240-242].

The partitioning mechanism between polymerisation and proofreading modes of the Pol III core is a large gap in our knowledge. In the two DNA polymerase families for which structures in both modes are known, family A (T. aquaticus Pol I in pol-mode [243] vs. E. coli Pol I in the exo-mode [71]) and family B (bacteriophage replicative polymerase RB69 in pol-mode [216] vs. RB69 in exo-mode [244]), partitioning involves a large movement of the primer–template (~30–40 Å) from one active site to the other and this is coupled to significant structural rearrangements. However, structures of the
family A and B polymerases are constrained by having an intrinsic proofreading domain, whereas the ε exonuclease domain in the Pol III core is flexibly tethered to α and may thus be able to sample a large volume of space close to the polymerase site, particularly when the interaction of the ε CBM with the β2 clamp is broken. This suggests that the transition between the two modes in *E. coli* and other organisms where exonuclease activity has evolved on a separate subunit may be different from the simpler replicative polymerases. This distinction may enable efficient proofreading to occur in the background of extremely high polymerisation rates, as is required for *E. coli*.

### 1.5.4 The clamp loader complex

The β2 clamp is a homodimeric toroid in solution and supports processive synthesis by topologically encircling DNA (Figure 1.4), so it must be deposited onto the DNA template by an activity that opens and closes the β2 ring around primer–template DNA. This is the primary role of the clamp loader complex (CLC). Five of the seven subunits of the *E. coli* CLC belong to the family of AAA+ ATPase family (Figure 1.6a) [245,246], and they form a horseshoe-shaped complex (Figure 1.6b) that binds and hydrolyzes ATP in a DNA dependent manner during clamp loading [247,248]. The complete CLC within DNA Pol III HE contains seven subunits and has the composition τₙγ(3–ₙ)δδ'–ψχ (n = 0–3, where physiologically relevant assemblies have n ≥ 2) [34,249].

The unique copies of δ and δ' in the CLC are encoded by the *holA* and *holB* genes [40], and are associated in the CLC with three copies of the *dnaX* gene product (τ and/or γ) that oligomerize via a common domain III [250]. In fact, the stable pentameric CLC core complex is held together via pairwise interactions of domains III of all of its AAA+ constituents [36]. While τ is a full-length product of *dnaX*, γ is a truncated version produced as a consequence of programmed ribosomal frameshifting during translation of *dnaX* mRNA [32,251,252]. The γ subunit therefore lacks the two unique C-terminal domains IV and V of τ responsible for binding the DnaB helicase [37] and the Pol III core [35], respectively. The τ subunit of the CLC thus dimerizes [253] the Pol III cores at the replication fork via the strong α–τ interaction [17,35,39,254]. Consequently the CLC plays a central functional and organizational role in the replisome, linking the leading- and lagging-strand polymerases with the replicative helicase. Interestingly, despite the CLC having only one copy of χ–ψ, approximately four copies were detected associated with the replisome in living *E. coli* cells [205].
Whether there are two or three τ subunits in the CLC and therefore two or three Pol III cores within the Pol III HE, is a subject of a debate. The use of in vivo single-molecule fluorescence imaging in live E. coli cells produced evidence to support that functioning replisomes contain three α subunits bound to the CLC associated with three τ subunits [205,255]. In further support, in vitro assays showed that use of Pol III HE containing three Pol III cores (“tripol”) eliminated the observation of residual Okazaki fragment gaps in rolling circle DNA replication reactions, and the “tripol” replisomes showed higher processivities [184,206] and produced shorter Okazaki fragments [256]. However, cells only able to assemble a CLC containing τ (with three Pol III cores associated) are less fit than their wild-type counterpart (with one γ subunit) in response to dsDNA breaks and damage induced by UV light, implying an important functional role of the γ subunit in vivo [179]. However, this specific functional role of γ upon irradiation by UV may be due to its involvement in DNA repair processes that could occur outside the context of the replication fork.

The minimal CLC cores proficient in clamp-loading are hetero-pentamers that lack both χ and ψ subunits [43,45,257-259]. They are comprised of homologous proteins evolved from a common ancestor and have a general composition of $\tau_\gamma(3-n)\delta\delta'$, (where $n = 0–3$), with $\gamma_3\delta\delta'$ being the simplest assembly. These various CLCs that vary in $\tau/\gamma$ stoichiometry can be resolved and isolated by ion-exchange chromatography [34], making them readily available for extensive in vitro biochemical studies.
Figure 1.6 | Crystal structures of the *E. coli* γδδ'ψpep–DNA and χ–ψ complexes. **a**, Cartoon representation of the crystal structure of δ' in the absence of nucleotide cofactors (PDB: 1A5T). Domains I, II and III are coloured blue, orange and purple respectively, and the P-loop segment is in red. This represents the general structural fold of the AAA+ proteins, δ, γ and δ' in the *E. coli* CLC. **b**, Structural arrangement of γδδ'ψpep bound to primer–template DNA (PDB: 3GLI). The structure shows an arrangement of ATPase subunits conforming to the helical structure of DNA. The location of ψpep on the clamp loader positions the χ–ψ assembly for interaction with SSB bound to the single-stranded template exiting the CLC. Below is a top view, obtained by 45° rotation around the x-axis of the view in **b**. The inset shows an expanded view of the interdomain coordination of the nucleotide analogue ADP•BeF₃ bound to Arg215 in the γ subunit. The adjacent SRC domain of a neighbouring γ subunit is highlighted in purple sticks. For clarity amino acids farther than 4 Å from ADP•BeF₃ have been omitted (PDB: 3GLI). **c**, Cartoon representation of the structure of *E. coli* γ–ψ, in pink and green, respectively, in complex with SSB-Ct shown as black sticks (PDB: 3SXU) [163]. The N-terminal 26 residues of ψ are flexible and do not appear in the crystal structure.
The χ (the holC gene product [43,44,260]) and ψ (holD [43,44,261]) subunits are accessory proteins that are not essential for the clamp loading reaction. Indeed, while CLCs and β2 clamps have universally conserved structure and function, χ and ψ homologs are more narrowly distributed in eubacteria [92]. The two subunits assemble in a strong 1:1 complex, the crystal structure of which has been reported (Figure 1.6c) [262]. The structure lacks the flexible [263] but conserved 26 N-terminal residues from ψ [262] (termed ψ_pep), otherwise responsible for the strong interaction of ψ–χ with the common domain III of γ and τ (Figure 1.6b) [36,41,263]. This interaction in turn stabilises the CLC by increasing the affinity of τ or γ for δ5’ [257]. The residues at the ψ–χ interface are highly conserved among several bacterial species [262]. Note that the χ subunit does not bind directly to the core of the CLC [264]. Rather, it interacts with the flexible C-terminus of SSB [165] via a hydrophobic pocket on the side opposite the interface with ψ [163]. Since ψ interacts with the CLC core on the side opposite its interface with χ (Figure 1.6c) [262], the ψ–χ complex serves as a molecular bridge that connects the CLC core with SSB, potentially positioning the CLC to scan the ssDNA template emerging from the helicase on the lagging strand for the presence of primers. In fact, the χ–SSB interaction promotes both processive polymerase elongation and β2 clamp loading [162]. Interestingly, binding of ψ to the CLC core also promotes its ATPase activity and affinity for both the β2 clamp and DNA [265]. Nevertheless, it appears that all of the determinants of functional contribution of ψ to the clamp loading reaction are captured by ψ_pep [41], arguing that no other major contacts of the structured region of ψ–χ with the CLC core is required.

CLC-associated χ (via ψ) has been proposed to control a primase-to-polymerase switch whereby it competitively displaces primase from synthesized RNA primers by switching from primase–SSB to χ–SSB interactions at the primer–template junction. Competitive displacement is proposed to destabilise the primase, which diffuses away to allow the CLC to deposit a β2 clamp onto the newly primed template [46]. Mutations designed to block χ–SSB complex formation result in defects in leading-strand synthesis and lead to generation of shorter Okazaki fragments on the lagging strand, alluding to a role of the χ–SSB complex in Pol III HE maintenance during lagging-strand cycling [163].
Nevertheless, the ψ subunit is dispensable in *E. coli* (but not in *Acinetobacter baylyi* [266] or *Pseudomonas aeruginosa* [267,268]). In these species, ψ proteins are significantly larger than, and show very little sequence homology to, *E. coli* ψ [191]. The additional domain in *P. aeruginosa* ψ relative to *E. coli* appears to be highly mobile and has been implicated in ssDNA binding. This may be a general property of ψ proteins from *Pseudomonadaceae* [269].

The first structures of the minimal pentameric CLC γ31–373δδ’ (~60 flexible residues from the C-terminus of each γ were omitted) were determined in nucleotide-free [38] and ATPγS-bound states [270]. Surprisingly, the two structures differ only slightly. Nucleotide (ATPγS) binding was expected to trigger a significant conformational change, suggesting that the determined nucleotide-bound structure may be in an inactive state, incompetent to bind the β2 clamp and DNA. It was only when the contribution of the ψ subunit [265] to the clamp-loading process was discovered and the determinants of its contribution were narrowed down to ψ pep [41] that the minimal CLC core in complex with ψ pep, the ATP analogue ADP•BeF3, and primer–template DNA was crystallized in an active state competent to bind the β2 clamp. The new active state structure revealed the intrinsically unstructured ψ pep folds to asymmetrically contact all three of the γ subunits [41]. Thus one copy of ψ (and therefore ψ–χ) binds across domains III of all three γ/τ subunits of the CLC, explaining the previous paradox in which ψ was known to interact with both γ and τ in isolation, but only one ψ–χ is present in the fully active, assembled seven-subunit CLC [253].

The CLC core structures in the active and inactive states reveal that the C-terminal domains III of the constituent AAA+ subunits interact to a form a collar, which supports the asymmetric arrangement of the N-terminal domains I and II that assemble the ATPase modules. Domains I and II contain motifs implicated in ATP binding, *i.e.*, the structurally conserved phosphate binding loop (P-loop), Walker A [271] and DExx motif (Walker B) [272,273] involved in metal binding, and sensor I [274] and sensor II [275] (located in domain II) motifs. In addition, these structures show that the ATP binding sites are located at subunit interfaces formed by conserved Ser–Arg–Cys (SRC) motifs from one subunit and P-loops from an adjacent one. This structural arrangement provides cooperativity between subunits. In pentameric CLCs, the three τ/γ subunits are capable of binding and hydrolysing ATP; δ’ lacks many of the motifs that comprise its AAA+ module (except the P-loop) so it cannot bind ATP but can donate its SRC
motif for hydrolysis of ATP at the interface with the neighbouring $\tau/\gamma$ subunit, as observed in the crystal structure of $\delta'$ alone [275]. The structure of the inactive ATP$\gamma$S-bound state of the CLC revealed that only two of the three nucleotide binding sites were occupied; the middle $\gamma$ subunit contained bound phosphate, thus failing to trigger the conformational change in the CLC that would render it capable of clamp binding [270]. However, the arrangement of AAA+ modules in the active state structure of the CLC on primer–template DNA revealed that the three copies of $\gamma$ and the $\delta'$ subunit formed a right-handed helical structure around dsDNA, a result of individual domains twisting with respect to the adjacent subunits. The absence of the clamp in the CLC structure could account for the lack of the anticipated engagement of the $\delta$ subunit in the ATPase spiral. The notion that the CLC locks onto the primed DNA in a screw-cap-like arrangement was proposed [41], and this organization seems to be critical for DNA recognition, DNA-dependent ATPase activity of the CLC, and clamp release [276]. A similar helical arrangement of ATP binding domains has been observed in the structure of the eukaryotic CLC–clamp (i.e., RFC–PCNA) complex [277], determined with bound ATP$\gamma$S in the absence of DNA.

In the ATP-bound state, $\delta$, the most diverged subunit of the *E. coli* CLC core, exposes its CBM for initial interaction with the $\beta_2$ clamp [42,258,278]. The RFC–PCNA structure suggests ATP binding induces further structural complementarity between the CLC and the $\beta_2$ clamp [276], which in turn further stimulates the other CLC subunits (i.e., $\tau/\gamma$) to contact $\beta_2$, albeit with weaker interactions compared to $\delta$ [279]. The CLC screw-cap-like arrangement and structural complementarity between the CLC and $\beta_2$ clamp induced by ATP binding were predicted to open one of the clamp’s two identical dimer interfaces, both in the in-plane and out-of-plane vectors, to mimic the shape of the CLC. Interestingly however, the ATP$\gamma$S-induced screw-cap-like arrangement in the RFC–PCNA crystal structure was insufficient alone to open the bound PCNA, which remained in the fully closed state [277].

Nevertheless, the subsequent crystal structure of the bacteriophage T4 CLC (gp44/62) in complex with its clamp gp45 in the presence of DNA and ATP analogue (ADP•BeF$_3$) revealed an open clamp structure that assumes the right-handed lock-washer shape. This helical shape matches the helical geometry of DNA such that $\alpha$-helices in the inner clamp surface track the minor groove of DNA [280]. Interestingly, although the trimeric gp45 clamp (by itself) is closed in its crystal structure [281], time-resolved FRET studies indicated that in solution, one of the interfaces is open wide enough to
accommodate dsDNA [282]. This propensity to remain open in solution may have facilitated entrapment of gp45 in the open state in the gp44/62 CLC structure.

Mechanistically, binding of three ATP molecules to the CLC is necessary to enable CLC–clamp binding [42], which occurs even in the absence of DNA. A FRET study of the RFC–PCNA–DNA complex reported that ATP hydrolysis drives the clamp-loading process to completion with transition of PCNA to a closed structure around DNA. On the other hand, non-hydrolysable ATPγS cannot drive closing of the PCNA interface, suggesting DNA binding alone is not responsible for clamp closing [283]. Single-molecule (sm)FRET experiments with the *E. coli* CLC have recently been used to visualise real-time conformational changes of individual β2 clamps as they are loaded onto DNA by minimal (γδδδ') CLCs [284]. Hydrolysis of one (or two) ATP molecules by the CLC in about 0.3 s results in ring closure around DNA, consistent with findings indicating a single hydrolysis step is sufficient to close the β2 clamp [285]. The remaining ATP molecule(s) are subsequently hydrolysed when the clamp is already closed and in 0.7 s allows dissociation of the CLC, leaving the loaded clamp on the DNA. These experiments were combined with single-molecule fluorescence polarization measurements at the interfacial domains of the β2 clamp, which reveal a rotation of ~8° as the β2 clamp opens [284]. The interaction between the CLC and primer–template junction is very similar in *E. coli* and bacteriophage T4, as evident from the crystal structures, and the rotation of the clamp as it closes is consistent with the T4 CLC–open clamp structure [280].

Thus 3D structures of CLCs and their respective clamps from all domains of life have now been reported, detailing basic architectures and compositions that have been conserved throughout evolution. Despite this, no full-length structure of the *E. coli* τ subunit has been solved either alone or in any CLC. Structural studies have proven a significant hurdle due to the dynamic behaviour and flexibility inherent in the τ subunit, and further effort is needed to answer many mechanistic and structural questions surrounding the spatial organisation of the Pol III HE on DNA as it cycles through various stages of DNA replication.

### 1.6 Dynamics on the lagging strand

At the replication fork, the leading strand is elongated continuously by the leading-stand Pol III core moving in the same direction as DnaB as it unwinds ds to ssDNA as it translocates on the other (lagging) strand. The chemistry and geometry of the DNA
molecule itself, combined with the requirement that DNA polymerases can only extend DNA in a 5’–3’ direction and the spatial restriction imposed by physical coupling of leading- and lagging-strand Pol III cores in one Pol III HE particle, all restrict the number of possible mechanisms used to copy the DNA at the replication fork. Lagging-strand replication loops can be formed to overcome these restrictions and support coupled leading- and lagging-strand synthesis by a single Pol III HE as the replisome progresses in one direction only. As a consequence, lagging-strand template replication occurs by discontinuous synthesis of short Okazaki fragments. A short RNA primer used to initiate each DNA fragment is generated by the specialist RNA polymerase, DnaG primase, as replicative DNA polymerases cannot initiate de novo synthesis of DNA. To assemble contiguous lagging-strand products, the RNA primers must be removed for genome maintenance, because RNA is inherently unstable to spontaneous hydrolysis cf. DNA, and then the nascent Okazaki fragments must be joined.

Processive and coupled DNA synthesis of both the leading- and lagging-strand templates to produce long DNA chains has been termed coordinated replication [286]. How these ordered molecular processes on the lagging strand occur in a timely fashion so that they keep pace with continuous leading-strand synthesis is discussed below.

Formation of replication loops on the lagging strand by partially reconstituted phage T4 replisomes was first hypothesized after simultaneous detection of long DNAs and Okazaki fragments during processive synthesis [171]. This discovery led Bruce Alberts to envision that as Okazaki fragment cycling proceeds, the lagging-strand loop grows and collapses analogous to a trombone slide extending and retracting, so replication loops have often been termed trombone loops [172]. Later, reconstitution of the E. coli replisome from purified components established that the Pol III HE was dimeric and asymmetric in architecture, providing a basis for formation of trombone loops. Moreover, the two (or three) constitutive Pol III cores utilize different processes to synthesize both strands simultaneously while progressing overall in one direction, depending on which strand they replicate. Based on similarity to the T4 replisome, trombone loop formation is thus expected also for the E. coli replisome. To date however, such loops have not been detected with any confidence in E. coli replication reactions. They have been directly visualised and their length distributions characterized in phages T7 [287] and T4 [288] replisomes using EM and later using single-molecule assays in the case of T7 [173,289].
Thus, despite great progress in our understanding of key processes and the players that contribute to Okazaki fragment synthesis, such as deposition of RNA primers by DnaG, primer handoff from DnaG to a lagging-strand polymerase for Okazaki fragment extension, and orchestrated release of the lagging-strand polymerase from one Okazaki fragment to be reused in synthesis of the next, the full mechanisms of these molecular processes are still matters of debate.

### 1.6.1 Lagging strand priming

During coordinated replication, synthesis of the leading strand and concurrent unwinding of dsDNA by DnaB helicase produces the lagging-strand ssDNA. Binding of SSB protects the ssDNA from degradation and unwanted secondary structure formation. Up to three DnaG primases bind concurrently to a single DnaB hexamer, where each DnaG monomer interacts across one of the three trimerization interfaces within the trimer of dimeric N-terminal domains in the hexamer (Figure 1.2). Two concurrently bound DnaGs are probably necessary and sufficient for priming. This indirect dimerization of otherwise monomeric primases via interaction with DnaB presumably allows DnaG to compete with SSB to bind ssDNA at specific trinucleotide priming sites behind the helicase to initiate primer synthesis. It is proposed that the ZBD of one DnaG molecule recognizes the priming site on nascent lagging-strand ssDNA to deliver it to the RPD domain of another DnaG molecule [128].

DnaG by itself is an extremely weak and error-prone DNA-dependent RNA polymerase that, if acting alone on DNA templates, takes around 20 min to synthesize RNA primers of physiological length [133]. DnaB stimulates RNA primer synthesis activity of primase ~5000-fold, 300-fold of which can be ascribed to stimulation of its template affinity and 15-fold to an increase in catalytic rate [131]. This implies that if a complete primer is synthesized while DnaG and DnaB are in contact, it would take at least 0.2 s to complete synthesis [290,291], corresponding to the time required for synthesis of ~200 bp of lagging-strand DNA. Maintenance of the primase–helicase contact during entire primer synthesis is mandated in phage T7 since the two activities are promoted by the same polypeptide molecule (gp4), but the timing of release of DnaG from DnaB is still unknown in *E. coli* DNA replication. In addition, binding to DnaB seems to increase the promiscuity of selection of initiation sequences by DnaG (otherwise found to be 5'-CTG-3' on the leading strand), perhaps due to increased stability of the lagging-strand ternary complex between DnaG, DnaB, and ssDNA, but with a distinct preference now for 5'-CAG-3', the complement of the preferred leading-strand sequence [292].
The precise mechanism by which the *E. coli* replisome coordinates lagging-strand priming with DNA synthesis on the two strands is not clearly understood [176]. There are three scenarios and variations of two of them. The first scenario is that helicase action and leading-strand synthesis may pause transiently during primer synthesis. In fact, transient pauses have been identified by single-molecule flow stretching experiments using reconstituted phage T7 replisomes [293], but in similar assays of leading-strand synthesis by the reconstituted *E. coli* replisome, no short pauses are observed at the time resolution used (data collection at 2 Hz) [294].

The second scenario is that DNA synthesis on both leading and lagging strands continues unimpeded during primer synthesis by the DnaG–DnaB complex. Experiments using reconstituted phage T7 and T4 replisomes have demonstrated that the primase–helicase remains in contact with the priming sequence during primer synthesis, with the constant supplement of new ssDNA as it is expelled from behind the helicase. The nascent lagging-strand template is thus organised into a new loop termed a “priming loop” [295,296]. An efficient hand-off of primer from primase to polymerase results in the collapse of the priming loop into the trombone loop. Furthermore, bacteriophage T7 lagging-strand polymerases were measured to outpace leading-strand polymerases [295], exactly as would be predicted to be needed for synthesis of the lagging strand to keep pace with the leading strand.

In *E. coli*, differences in rates between the two asymmetric Pol III cores during elongation is possible because of the difference in chemical environments characterizing the leading- and lagging-strand templates during the elongation phase. The lagging-strand Pol III core probably contacts and displaces SSB molecules in front, whereas it has been thought that there are no SSBs in front of the leading-strand polymerase. In addition, the eventual involvement of the structured core of the CLC with one of the two Pol III cores during DNA synthesis may also contribute to functional asymmetry, and therefore a difference in rates.

The third scenario is that priming loop formation is minimized by prompt dissociation of DnaG from DnaB as soon as a priming site is recognized, before or at an early stage of primer synthesis. However, this scenario does not obviate the need for the lagging strand polymerase to make up for the loss of lagging-strand replication during the time taken for primer synthesis.
The biochemical basis for differential rates of leading- and lagging-strand synthesis has not been adequately explained, so a less controversial variation of the last two scenarios is to have multiple Okazaki fragments being simultaneously synthesized by additional lagging-strand polymerases (and consequent loops) within the replisome [206,207,256]. Since neither priming loops nor trombone loops have been directly visualised as being abundantly associated with *E. coli* replisomes, it is far from certain how leading- and lagging-strand replication is actually coupled, the elegant views in textbooks notwithstanding.

There are many more Okazaki fragments than DnaG molecules in the cell, so DnaG must clearly separate in a timely way from its primer (and DnaB) to be recycled. The three-point-switch (primase-to-polymerase) model [46] explains how DnaG is recycled; the model posits that once a primer has been synthesized, DnaG dissociates from DnaB and then remains stably bound to the nascent primer through the interaction with an adjacent SSB tetramer on the lagging-strand template. DnaG-bound SSB eventually interacts with the $\chi$ subunit of the clamp loader complex (CLC) to displace DnaG from SSB and its primer. While this model appears elegant, the basis of the thermodynamic preference of $\chi$ for interaction with a particular SSB protomer bound to DnaG, as opposed to interaction with any other protomers of the same or neighbouring SSB tetramers, remains unclear. As described above, the three-point-switch model then infers that DNA synthesis rates by the two different Pol III cores must be different, *i.e.* the lagging-strand Pol III core would have to replicate faster to compensate for time that DnaG has spent associated with SSB and the additional time needed for loading of the $\beta_2$ clamp and initiation complex formation.

Following stable binding with DNA and processive elongation during coordinated synthesis, lagging-strand Pol III cores have to be able to rapidly dissociate from the ends of Okazaki fragments to associate with a $\beta_2$ clamp loaded onto the nascent primer–template of the next Okazaki fragment. A “processivity switch” must exist to allow a processive Pol III core to suddenly become distributive and dissociate from an Okazaki fragment so it can be re-used for synthesis of a new Okazaki fragment. There are two different models that have been promulgated as the source of the trigger for processivity switching: the collision model and the signalling model. The evidence for either model is still controversial.
1.6.2 The collision model

The collision model was first proposed whilst studying the *E. coli* replisome, after observations that the Pol III HE dissociates from the template after the polymerase replicates DNA to a nick and “collides” with the 5'-end of a previously extended primer [297]. This is a situation equivalent to collision with the 5' end of a previous Okazaki fragment. Subsequent studies indicated that the $\tau$ subunit of the CLC in particular is implicated in the switch [298]. A mechanism was proposed whereby $\tau$ and the $\beta_2$ clamp compete for interaction with the extreme C-terminus of $\alpha$ [18,299]. If $\tau$ could outcompete the $\beta_2$ clamp when $\alpha$ reaches the end of the Okazaki fragment (*i.e.*, at the nick) it could destabilise the interaction of $\alpha$ with the DNA, so it can promptly dissociate and be recycled [298]. Meanwhile, it was found that the potential interaction between the C-terminus of $\alpha$ and the $\beta_2$ clamp is dispensable for processive DNA replication. It is instead the interaction of the internal clamp-binding motif (CBM) in $\alpha$ that primarily determines polymerase processivity [300]. It was also demonstrated that the Pol III core cannot dissociate rapidly enough from a modelled end of an Okazaki fragment to support synthesis of a new Okazaki fragment every second or so [301]. Indeed, as the Pol III core synthesizes DNA to a nick, its affinity for the DNA template seems to gradually reduce, likely due to loss of interactions of the OB fold domain with ssDNA as it approaches the RNA primer of the previous Okazaki fragment [302]. Nevertheless, this weakening of interaction with the DNA was not enough to emulate a mechanism akin to the processivity switch [301]. Therefore, we currently have no direct evidence or detailed understanding of a collision mechanism for Pol III core recycling within the *E. coli* replisome.

1.6.3 The signalling model

The signalling model posits that the synthesis of a new primer by DnaG triggers the lagging-strand Pol III core to cycle, irrespective of whether the synthesis of the current Okazaki fragment has been completed [303]. The evidence for this model came from an observation that dilution of the Pol III to sub-nM concentrations resulted in gaps between Okazaki fragments [303]. Therefore, the signal is independent of Okazaki fragment completion and may precede a collision mechanism. Considering that DnaG concentration was previously found to determine the frequency of priming and lengths of Okazaki fragments [117], it was hypothesized that DnaG(s) at the replication fork provide the signal for processivity switching. Interestingly, both signalling and collision mechanisms appear to be operational during phage T7 replication [173].
The most recent studies with the *E. coli* replisome, however, suggest that the signal for switching does not come from primase *per se*, but from the presence of primer-template DNA [304]. The mechanism behind this is not yet determined, but it appears probable that it will involve the CLC. Future investigation may confirm whether the τ subunit in the CLC is indeed involved in processivity switching in a more substantive way, other than through its direct interactions with α and DnaB. Further studies are needed to unveil the mechanism behind signalling release of the Pol III core from the DNA template.

### 1.6.4 Okazaki fragment maturation

To assemble the continuous daughter DNA strand complementary to the lagging strand template from discontinuous Okazaki fragments, an RNA primer from each of them has to be replaced by DNA to leave nicks, and the nicks need to be ligated, in the process known as Okazaki fragment maturation (OFM). To achieve this, *E. coli* may potentially use any of three proteins to remove the RNA primers: Pol I, RNase H I and/or RNase H II. Pol I is equipped to replace the primers with DNA and DNA ligase A’s role is to seal the remaining nick by making a phosphodiester bond between free 3'-OH and 5'-phosphate groups.

Pol I was the first DNA polymerase discovered [305]. It bears the classical 5'–3' polymerase and 3'–5' proofreading exonuclease activities in a large C-terminal fragment that can be generated by proteolysis, and has been termed the Klenow fragment [71,306,307]. The smaller N-terminal fragment houses its unusual 5'–3' exonuclease activity [308,309], which is implicated in the removal of RNA primers. The two nuclease activities present in Pol I are fundamentally different; the 3'–5' nuclease is a classic exonucleolytic proofreader (like ε) that cleaves the last phosphodiester bond in a non-base-paired (mismatched or melted) 3'-end. The 5'–3' nuclease cleaves a bond only in a base-paired region.

Traditionally it has been thought that Pol I removes primers by nick translation, where the sequential (one-by-one) removal of 5'-rNMPs (or dNMPs) at a nick occurs concomitantly with replacement DNA synthesis in the 5'–3' direction. However, concurrent DNA synthesis was shown to stimulate the 5'–3' exonuclease domain to occasionally remove oligo- rather than mononucleotides [310], and subsequently, the small N-terminal fragment of Pol I was shown to be capable of specific hydrolysis of the phosphodiester bond at the junction of a 5'-single-stranded overhang (or flap), leaving a nick. This process is known as flap endonuclease (FEN) activity. FEN activity of Pol I
is consistent with identified homology between *E. coli* Pol I and both human FEN1 and yeast RAD27 FENs [311]. This is further reinforced by the homology between the *E. coli* Pol I FEN domain and the structurally characterized FEN domain from *T. aquaticus* Pol I that retains the structural fold of other known FENs [312]. The removal of oligonucleotides could operate by coupling Pol I strand-displacement activity [9] that generates ssDNA flaps at the 5'-end and subsequent FEN activity that removes them to leave a nick.

Finally, the nick is covalently sealed by the activity of DNA ligase A to form a contiguous nascent strand [74,75]. Interestingly, both Pol I and ligase A are suggested to exert their activities during OFM by binding to the β2 sliding clamp that is likely left in the wake of the replication fork following Pol III HE dissociation from the template-bound Okazaki fragment [72].

It appears that at least one functional FEN, that could be associated with OFM, is required for cell viability in bacteria [313]. Many eubacteria encode a second FEN paralog in addition to that present in Pol I [314]. FEN in bacteria is encoded by the *xni* (Exo nine) gene and the enzyme from Firmicutes contains clustered conserved acidic residues that comprise binding sites termed Cat1 and Cat2 that coordinate one divalent metal ion each. The Cat1 metal binding site is involved in catalytic activity whereas the Cat2-bound metal ion stabilises the enzyme–substrate complex and may not be involved directly in catalysis of the phosphoryl transfer reaction [315,316]. However, a subset of genera including *E. coli* encodes a different FEN (also termed Exo IX) that lacks the three aspartate residues that make up the Cat2 site [313,314]. Indeed, the structure of *E. coli* Exo IX bound to a flapped DNA template shows that the Cat2 site is absent, whereas the conserved Cat1 site contains a pair of oxo-bridged Mg$^{2+}$ ions [317], lending further support to the notion that the previously proposed two metal ion mechanism based on the structure of FEN from *T. aquaticus* Pol I [312] is conserved in FENs across phyla.

Whereas consensus was reached for Pol I being the essential enzyme required for processing RNA primers during OFM, *in vivo* studies later showed that the *polA* gene (encoding Pol I) is dispensable in *E. coli*, although deletion of *polA* leads to high mutation frequencies and a temperature-sensitive phenotype [313,318]. Further studies showed that it is the small (FEN) domain of Pol I that is more important for the viability of *E. coli* than the Klenow fragment [313,319]; the presence of only the FEN-coding region of the *polA* gene is sufficient to provide full viability [319]. Moreover, the FEN
domain of Pol I is absolutely required for cell viability in *Streptococcus pneumoniae* [320] and *Synechococcus elongates* [313], species that appear not to produce other FEN paralogs.

It has been further demonstrated that while the FEN-domain from either DNA Pol I or Exo IX is sufficient for cell viability in *E. coli*, double null-mutants in both FEN encoding domains are not [313]. Considering that it has been shown that in T5FEN, a FEN homolog of Exo IX in bacteriophage T5, Cat1 is essential and sufficient for endonucleolytic flap cleavage whereas both sites are required for 5'–3' exonuclease activity [321] on flapped templates, it can be concluded that the most indispensable activity of Pol I is not its polymerisation or 3'–5' exonuclease activities, but is in fact its flap endonuclease activity. It could thus be anticipated that the FEN endonuclease activity of Pol I is utilized during the normal OFM process and this way of processing Okazaki fragments must be critical for cell survival.

The other potential Okazaki fragment processing enzymes in *E. coli*, RNase Hl and RNase Hll, have been shown to be dispensable in the presence of Pol I in live *E. coli* cells [313]. In addition, the fact that unlike eukaryotic RNase H, *E. coli* RNase Hl can only digest RNA from the 3'–5' direction and leaves at least one ribonucleotide behind [322,323], combined with the lethal phenotype of the polA, xni double null mutant further strengthens the status of Pol I in the OFM process and raises doubt that RNase Hl and Hll have any role in it.

### 1.7 Concluding comments

In this Chapter, we have reviewed our own current understanding of how synthesis of DNA on the two separated DNA strands at replication forks may be coordinated to allow *E. coli* chromosomal DNA to be copied at rates of ~1 kbp/s with high processivity and without making mistakes. Nevertheless, our review is hardly comprehensive. There are a series of fork-related topics we have not covered, but which have nevertheless been at the forefront of recent studies. These include studies of how replication terminates when the two replication forks meet in the *Ter* region opposite *oriC* on the circular chromosome [175,324-327], how replisomes and forks are re-established when they collapse at replication impediments and roadblocks [328,329] and how replisomes displace protein roadblocks [330,331], bypass lesions in the template strands [332,333] and cope with collisions with other molecular machines like RNA polymerase [334,335].
Studies since the first \textit{in vitro} replication reactions were reconstituted with \textit{E. coli} replication proteins in the Kornberg and Hurwitz research groups \textasciitilde{}40 years ago have led to a very beautiful “textbook view” of a perfectly orchestrated series of events where proteins and enzymes come and go in an orderly manner as they are required. The textbook view is not incorrect, but it is a simplification, as it must be. It mostly derives from \textit{in vitro} experiments that have been simplified by staging the reactions in steps where isolable “initiation complexes” from a prior assembly step are used for the next stage. This strategy enables exchange in and out of the replisome of only the components that are deliberately added in excess during the stages studied (often just $\beta_2$ and/or DnaG primase).

The reason the textbook view \textit{must} be a simplification of reality is that the replisome must obey fundamental chemical and thermodynamic principles as it functions. We know that it is held together by a hierarchical series of weak and strong protein–protein and protein–nucleic acid interactions (that incidentally seem to vary in strength among bacterial genera). Multiple weak interactions enable dynamics – they enable binding partners to be exchanged within the replisome without the whole complex disintegrating, and they allow defective components to be replaced from solution. This means that in between the essentially irreversible steps involving nucleotide hydrolysis or incorporation (these include strand separation, clamp loading and primer synthesis, as well as dNMP incorporation into DNA), the replisome must equilibrate among all possible low energy conformational states where some interactions are transiently broken and vulnerable, for example, to exchange of components from solution. Thus the performance of the replisome may be rather messier than expected of an orchestra, but it is likely to be much more resilient to the odd mistake. Some of these concepts have recently been discussed elsewhere in similar terms [335].

1.8 Scope of this Thesis

Determination of the high-resolution structures of almost all the replisome components over the last twenty years has been accompanied by the discovery of many dynamic protein–protein interactions. In many cases intrinsically unstructured regions that become structured upon interactions with binding partners often mediate these. Many of the functional protein–protein interactions that allow the replisome to transition directly from one conformational state to another remain speculative or ill-defined. The work described in this Thesis is directed towards two broad aims: (i) The use of biophysical techniques to uncover and study the importance of functional protein–protein interactions, specifically those within the Pol III core ($\alpha\epsilon\theta$) and between Pol III core and
the τ subunit of the CLC; (ii) to use sensitive single-molecule based methods to understand how components of the Pol III core enter and leave the replication fork during coupled DNA replication.

Despite the extensive biochemical and structural analysis of the α–τ interaction, to date most of the research has been focussed on the α-binding domain of τ. Recent work in the Dixon lab has led to the crystallization of the τ-binding domain of α in complex with a truncation mutant of τ (τC16). This structure shows key residues in both α and τ involved in this interaction. Moreover, this structure reveals the close proximity of domain IV of τ to α. In Chapter 3, surface plasmon resonance (SPR) and nuclear magnetic resonance (NMR) spectroscopy are used to show putative interactions of additional residues in domain IVa of the τ subunit with the α subunit. Further, small angle X-ray scattering (SAXS) and small angle neutron scattering (SANS) were used to determine the overall size, shape, and build structural models of the α subunit alone, and in complex with the unique C-terminus of τ using either τC16 or τC24. Comparisons are made with the recently published cryo-EM structures of the αε–β2–τC16 complex.

In Chapter 4, SPR is used to determine accurate K_D values of complete, full-length complexes containing various CBMs. These data extend previously reported SPR and ESI-MS measurements where the stabilities of Pol III–β2 protein complexes were ranked, showing using various CBM containing peptides that the K_D for the αεθ–β2 interaction is dependent on the strength of CBM present in ε. The studies presented in Chapter 4 use full-length Pol III complexes immobilised via the τ subunit. Upon assessment of the ταε–β2 complexes containing strong CBMs, a novel weak interaction between the C-terminal domains IVa and V of τ and β2 was identified. Furthermore, NMR spectroscopy was used to confirm this weak interaction and allow identification of a putative binding surface in domain V of τ to β2. The discovery of this interaction may provide insight into short-lived intermediate states during the clamp loading process.

The enzymatic activities within the E. coli replisome and idea that the DNA polymerases required to synthesize new DNA on the leading and lagging strands are physically organised into one holoenzyme are well established. However, dynamic behaviours that govern transitions during replication remain speculative. Chapter 5 describes the development and potential uses of fluorescent reagents to simultaneously visualise DNA replication and Pol III core behaviour at the single-molecule level. Further, during the course of this work a serendipitous discovery was made, where a single cysteine mutant of the θ subunit becomes S-glycosylated with
attachment of a galactose molecule in vivo. Chapter 6 describes the use of one of these fluorescently labelled reagents. SNAP-labelled Pol III core was used to show that the replisome is able to strike a balance between stability and plasticity. Moreover, we demonstrate it is not the individual DNA polymerases that are exchanging, but that the entire holoenzyme rapidly exchanges with others in solution. However, if absent the replisome is able to retain its original holoenzyme forming a stable complex. We propose that the plurality of weak interactions that binds the holoenzyme to the replisome plays a key role in mediating rapid exchange while preventing the replisome from dissociating.
Chapter 2
General Materials and Methods
2.1 Bacterial strains and plasmid vectors

2.1.1 Bacterial strains

*Escherichia coli* strain AN1459 (F⁻ supE ilvC thr-1 leuB6 hsdR recA srlA::Tn10) [336] was used as a host strain for plasmid preparation, selection and maintenance. Although containing *supE* (amber suppressor), this strain also exhibits inefficient utilization of the opal stop codon (UGA), subsequently leading to overproduction of some longer versions of protein species by bypassing UGA stop codons. To circumvent this issue, the UGA stop codon is mutated to a more efficiently used ochre stop codon (UAA) or suppressor-free strains such as derivatives of BL21(λ.DE3) or used for overproduction of proteins.

The derivatives of the *E. coli* B strain BL21 [F⁻ ompT lon hsdS₈ (rB⁻ mB⁻)], BL21(λ.DE3)recA [recA srlA::Tn10] [337] or BL21(λ.DE3)/pLysS [338], were used as the expression host to overexpress proteins. Removal of cytoplasmic *lon* and *ompT* protease genes from these strains can stabilise target proteins during the cell lysis step of purification [339]. In addition, both strains contain the prophage λ.DE3 in their chromosome, which carries the T7 *gene 1* encoding bacteriophage T7 RNA polymerase (RNAP) under control of the host lacUV5 operator-promotor. The promotor may be induced by removal of the *lacI* repressor by the addition of a lactose derivative (*e.g.* isopropyl-β-D-thiogalactopyranoside, IPTG) in the log phase of growth, driving the production of T7 RNAP to promote the transcription of the gene of interested under control from the T7 promotor [338]. Expression host strain BL21(λ.DE3)recA also contains a chromosomal recA mutation, this acts to stabilise plasmids containing host genes by suppressing then integration into the *E. coli* chromosome. Additionally, it also prevents plasmid multimerization that can confound mutagenesis experiments.

The other host strain BL21(λ.DE3)/pLysS contains a pLysS plasmid (pACYC184 derivative) that contains the T7 *gene 10* encoding T7 lysozyme. The presence of T7 lysozyme enables toleration of relatively toxic genes by inhibiting their basal expression; it is a specific inhibitor of T7 RNAP [340]. As a bonus, expression of T7 lysozyme facilitates cell lysis during protein purification [338]. The pLysS plasmid also confers chloramphenicol resistance, and strains containing it are therefore supplemented with chloramphenicol to ensure its maintenance.
2.1.2 Plasmid vectors

Three vector scaffolds were used to carry target genes to construct plasmids in all experiments mentioned in this Thesis. The physical maps of these vectors (pND706 and pETMCSI are shown in Figure 2.1 respectively. The individual characteristics of each of these vectors are outlined below.

The vector pND706 contains tandem bacteriophage $\lambda$ $p_R$ and $p_L$ promoters and the $c^{\text{857}}$ gene. The $\phi 10$ translation initiation region from pET3 vectors [338] is linked to the $\lambda$ promoters and also contains a strong ribosome binding site (RBS) and unique NdeI site (5'-CATATG-3') containing an ATG initiation codon. The par region from pSC101 is also present, providing enhanced plasmid stability at high copy number in the absence of ampicillin selection [341]. When cells are transformed with derivatives of pND706 and are grown at 30 °C, transcription from both the $\lambda p_R$ and $p_L$ promoters is completely repressed by the $\lambda$ repressor. A rapid shift of temperature to 42 °C denatures the thermolabile repressor, which subsequently dissociates from the $\lambda p_R$ and $p_L$ promoters and high-level transcription of the inserted gene between the MCS is allowed [341,342]. With this system affords very stringent control of gene expression by temperature, thus, toxic proteins may be produced using this vector.

Vector pETMCSI [343] enables transcription of a gene as it appears in the Studier T7 expression systems [338]. In order to direct overproduction of a target protein, plasmids containing pETMCSI vector must be transformed into strains capable of induced expression of T7 RNAP (e.g. BL21($\lambda$DE3)recA).

Vector pKO1274 [17] is a pETMCSI derivative that allows fusion of a gene in-frame with a N-terminal biotin tag (sequence MAGLNDIFEAQKIEWHEH [344]) incorporated at the Ndel site.
2.2 Chemicals and reagents

MilliQ™ water [18.2 MΩ cm⁻¹ resistivity] (Millipore, USA) was used in all experiments and was sterilised by autoclaving at 121 °C for 20 min for molecular biology and protein biochemistry experiments. The purity of all chemicals and reagents were of at least molecular biology grade. All buffers were filtered through a 0.22 μm filter. Chemicals are listed below in Table 2.1.

Table 2.1 | General chemicals and reagents used.

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Chemical and reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ajax Finechem,</td>
<td>36% hydrochloric acid (HCl)</td>
</tr>
<tr>
<td>Australia</td>
<td>acetic acid</td>
</tr>
<tr>
<td></td>
<td>acetone</td>
</tr>
<tr>
<td></td>
<td>ammonium acetate (C₂H₃O₂NH₄)</td>
</tr>
<tr>
<td></td>
<td>ammonium sulfate (NH₄)₂SO₄</td>
</tr>
<tr>
<td></td>
<td>ethanol</td>
</tr>
<tr>
<td></td>
<td>methanol</td>
</tr>
<tr>
<td></td>
<td>glycerol</td>
</tr>
<tr>
<td></td>
<td>potassium dihydrogen orthophosphate (KH₂PO₄)</td>
</tr>
<tr>
<td></td>
<td>potassium monohydrogen orthophosphate (KHPO₄)</td>
</tr>
<tr>
<td></td>
<td>sodium hydrogen phosphate (Na₂HPO₄)</td>
</tr>
<tr>
<td>AMRESCO, USA</td>
<td>bromophenol blue</td>
</tr>
<tr>
<td></td>
<td>Coomassie brilliant blue R250</td>
</tr>
<tr>
<td></td>
<td>glycerol</td>
</tr>
<tr>
<td></td>
<td>sodium dodecyl sulphate (SDS)</td>
</tr>
<tr>
<td>Astral Scientific,</td>
<td>dithiotheritol</td>
</tr>
<tr>
<td>Australia</td>
<td></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</td>
</tr>
<tr>
<td>Bio-Rad,</td>
<td>ethidium bromide (EtBr)</td>
</tr>
<tr>
<td>Supplier</td>
<td>Chemical and reagents</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Australia</td>
<td>Precision Plus Protein Duel colour standards β-mercaptoethanol</td>
</tr>
<tr>
<td>Bioline, Australia</td>
<td>2'-deoxyadenosine-5'-triphosphate (dATP), lithium salt; 100 mM solution 2'-deoxycytidine-5'-triphosphate (dCTP), lithium salt; 100 mM solution 2'-deoxyguanosine-5'-triphosphate (dGTP), lithium salt; 100 mM solution 2'-deoxythymidine-5'-triphosphate (dTTP), lithium salt; 100 mM solution agarose (molecular grade)</td>
</tr>
<tr>
<td>Cambridge Isotope Laboratories, USA</td>
<td>ammonium chloride (15NH₄Cl)</td>
</tr>
<tr>
<td>Fermentus, USA</td>
<td>6× loading dye GeneRuler DNA ladder 100–10,000 bp</td>
</tr>
<tr>
<td>Fluka, USA</td>
<td>cesium iodide (CsI) formic acid phenylmethanesulfonyl fluoride (PMSF)</td>
</tr>
<tr>
<td>GE Healthcare, Sweden</td>
<td>10% (v/v) surfactant P20 (P20)</td>
</tr>
<tr>
<td>Jena Bioscience, Germany</td>
<td>adenosine 5'-triphosphate (ATP), sodium salt; 100 mM cytidine 5'-triphosphate (CTP), sodium salt; 100 mM guanosine 5'-triphosphate (GTP), sodium salt; 100 mM uridine 5'-triphosphate (UTP), sodium salt; 100 mM</td>
</tr>
<tr>
<td>Novagen, Australia</td>
<td>BugBuster master mix</td>
</tr>
<tr>
<td>Roche Diagnostics, Switzerland</td>
<td>ampicillin chicken egg lysozyme complete protease inhibitor cocktail tablets PCR buffer set</td>
</tr>
<tr>
<td>Sigma-Aldrich, USA</td>
<td>2-amino-2-hydroxymethyl-propane-1,3-diol (Trizma™/Tris) 2,2',2'',2'''-(ethane-1,2-diyldinitrilo)tetraacetic acid disodium salt (EDTA) α-lactose monohydrate adenosine 5'-diphosphate sodium salt adenosine-5'-[β,γ-imido]triphosphate (AMP-PNP) ammonium-d₄ deuteroxide boric acid bovine serum albumin (BSA) chloramphenicol copper sulfate (CuSO₄·5H₂O) D-biotin D-(+)-glucose iron sulfate (FeSO₄·7H₂O) glycerol-d₈ guanidine hydrochloride (GdnHCl) kanamycin L(+)-arabinose magnesium acetate [Mg(OAc)₂] magnesium sulfate (MnSO₄·H₂O)</td>
</tr>
<tr>
<td>Supplier</td>
<td>Chemical and reagents</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>Bioscientific, Australia</td>
<td>3-aminopropyltriethoxysilane</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Chemical and reagents</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>potassium chloride (KCl)</td>
</tr>
<tr>
<td></td>
<td>potassium L-glutamate</td>
</tr>
<tr>
<td></td>
<td>potassium hydroxide (KOH)</td>
</tr>
<tr>
<td></td>
<td>potassium sulfate (K&lt;sub&gt;2&lt;/sub&gt;SO&lt;sub&gt;4&lt;/sub&gt;)</td>
</tr>
<tr>
<td></td>
<td>sodium 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethane-1-sulfonate (HEPES Na&lt;sup&gt;+&lt;/sup&gt;)</td>
</tr>
<tr>
<td></td>
<td>sodium acetate (NaOAc)</td>
</tr>
<tr>
<td></td>
<td>sodium bicarbonate (NaHCO&lt;sub&gt;3&lt;/sub&gt;)</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 trihydrochloride</td>
</tr>
<tr>
<td></td>
<td>streptomycin sulfate</td>
</tr>
<tr>
<td></td>
<td>tetracycline</td>
</tr>
<tr>
<td></td>
<td>thiamine (vitamin B1)</td>
</tr>
<tr>
<td></td>
<td>thymine</td>
</tr>
<tr>
<td></td>
<td>tris(2-carboxyethyl)phosphine</td>
</tr>
<tr>
<td></td>
<td>hydrochloride (TCEP)</td>
</tr>
<tr>
<td></td>
<td>tri-sodium citrate</td>
</tr>
<tr>
<td></td>
<td>zinc chloride (ZnCl&lt;sub&gt;2&lt;/sub&gt;)</td>
</tr>
<tr>
<td></td>
<td>zinc sulfate (ZnSO&lt;sub&gt;4&lt;/sub&gt;·7H&lt;sub&gt;2&lt;/sub&gt;O)</td>
</tr>
</tbody>
</table>

2.3 Growth media

2.3.1 Lysogeny Broth medium

Lysogeny Broth (LB) culture medium [345], supplemented with 25 mg L<sup>-1</sup> thymine (LBT) and/or appropriate antibiotics as required (100 mg L<sup>-1</sup> ampicillin [LBTA; for strains AN1459 and BL21(λDE3)recA] and 34 mg L<sup>-1</sup> chloramphenicol [LBTAC for strain BL21(λDE3)/pLysS]), was used to grow E. coli strains. LBT media was autoclaved and then supplemented with antibiotic(s) (0.22 μm, filter sterilised) before culturing cells. Bacterial strains containing plasmids containing λ promoters were grown in broth cultures at 30 °C, whereas those containing T7 promoter plasmids were grown at 30 °C or 37 °C, with shaking at 210 rpm. An overnight culture (~18 h inoculation) was prepared prior to growth of 1 L cultures. The overnight culture (10 mL or one loop of cells taken directly from LBTA or LBTAC agar plates) was used to inoculate a fresh LBTA or LBTAC cultures (1 L) for the overproduction of proteins.

2.3.2 Lysogeny Broth solid medium (agar plates)

Cells containing vectors controlled by the λ promoters (pND706) or the T7 promoter (pETMCSI or pKO1274) were grown on LBT agar plates (15 g L<sup>-1</sup> agar in LBT medium with the appropriate antibiotics (Section 2.3.1)) in an incubator (Thermo Scientific, Australia) at 30 °C (pND706 or pETMCSI) or 37 °C (pETMCSI or pKO1274) overnight (~18 h).
2.3.3 Lysogeny Broth autoinduction medium

This medium was used to produce proteins when striving for high yields or when lower temperatures were required from their solubility. Typically, 1 L of this medium [346] consists of: ~930 mL LB medium, 20 ml of filtered 50 × 5052 mixture (25% (w/v) glycerol, 2.5% (w/v) glucose, 10% (w/v) α-lactose), 50 mL of filtered 20 × NPS mixture (0.5 M (NH₄)₂SO₄, 1 M KH₂PO₄, 1 M Na₂HPO₄ at pH 6.75), 25 mg L⁻¹ thymine, 1 mM MgSO₄, and trace metals [347], with 100 mg L⁻¹ ampicillin and 34 mg L⁻¹ chloramphenicol added when required.

2.3.4 SOC medium

SOC is a rich medium primarily used during recovery of E. coli cell transformations with DNA [348]. It was comprised of 2% (w/v) tryptone, 0.5% (w/v) yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgCl₂, 10 mM MgSO₄, and 20 mM glucose. This medium was used to recover transformed cells during expression of perdeuterated α.

2.3.5 Minimal medium for ¹⁵N labelling

This medium was used to prepare samples of uniformly ¹⁵N-labelled proteins. It was comprised of: 100 mM NaH₂PO₄/KH₂PO₄ at pH 7.1, trace salts, 1 mg L⁻¹ thiamine, 1 mM MgSO₄, 7.2 g L⁻¹ D(+)-glucose, 1 g L⁻¹ ¹⁵NH₄Cl as the sole source of nitrogen, and 100 mg L⁻¹ ampicillin [22]. Strain BL21(λDE3)recA containing λ-promoter plasmids were grown at 30 °C, while those with T7-promoter expression vectors were grown at 37 °C, as specified.

2.3.6 ModC1 minimal medium for perdeuteration of proteins

This medium was used to prepare perdeuterated α subunit. To prepare desired solutions of deuterated medium, 100% D₂O solutions were used. The composition of ModC1 is given in Table 2.2. Due to the propensity for ModC1 to precipitate, additive solutions A, B, C, and D are prepared separately then added to the bulk solution, with 100 mg L⁻¹ ampicillin. Strain BL21(λDE3)recA containing T7-promoter plasmids were grown at 37 °C, then reduced to 20 °C as specified.
Table 2.2 | ModC1 medium composition.

<table>
<thead>
<tr>
<th>Solution</th>
<th>Component</th>
<th>g L⁻¹</th>
<th>mM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bulk solution</td>
<td>NH₄Cl</td>
<td>2.58</td>
<td>48.23</td>
</tr>
<tr>
<td></td>
<td>KH₂PO₄</td>
<td>2.54</td>
<td>18.66</td>
</tr>
<tr>
<td></td>
<td>Na₂HPO₄</td>
<td>4.16</td>
<td>29.30</td>
</tr>
<tr>
<td></td>
<td>K₂SO₄</td>
<td>1.94</td>
<td>11.13</td>
</tr>
<tr>
<td></td>
<td>glycerol-d8</td>
<td>40</td>
<td>434</td>
</tr>
<tr>
<td>Additive A (1000 × stock)</td>
<td>FeSO₄·7H₂O</td>
<td>0.02</td>
<td>0.719</td>
</tr>
<tr>
<td></td>
<td>tri-sodium citrate</td>
<td>0.088</td>
<td>0.3410</td>
</tr>
<tr>
<td>Additive B (1000 × stock)</td>
<td>MnSO₄·H₂O</td>
<td>0.005</td>
<td>0.0348</td>
</tr>
<tr>
<td></td>
<td>ZnSO₄·7H₂O</td>
<td>0.0086</td>
<td>0.0299</td>
</tr>
<tr>
<td></td>
<td>CuSO₄·5H₂O</td>
<td>0.00076</td>
<td>0.00304</td>
</tr>
<tr>
<td>Additive C (1000 × stock)</td>
<td>Thiamine</td>
<td>0.048</td>
<td>0.1596</td>
</tr>
<tr>
<td>Additive D (1000 × stock)</td>
<td>MgSO₄·7H₂O</td>
<td>0.67</td>
<td>2.72</td>
</tr>
</tbody>
</table>

2.4 Custom oligonucleotides

Oligonucleotide primers were purchased from GeneWorks (Australia) or Integrated DNA Technologies (USA). Oligonucleotide primers were routinely dissolved in sterilised TE buffer (10 mM Tris.HCl pH 7.6, 1 mM EDTA) to a stock concentration of 100 μM and stored at −20 °C.

2.5 Molecular biology procedures

Commercially available kits mentioned below were used as described in the manufactures’ manuals, unless otherwise stated.

2.5.1 Preparation of plasmid DNA by alkaline lysis

Small-scale plasmid preparations (Mini-prep) were carried out using the commercially available QIAprep® spin mini-prep kit (Qiagen, USA). *E. coli* strain AN1459 containing the plasmid of interest was grown overnight on LBT plates containing the appropriate antibiotics and cells were harvested and used for plasmid extraction. Plasmid DNA was routinely eluted in warmed TE buffer and stored at −20 °C.

2.5.2 Quantification of DNA

The concentration of purified plasmids was determined spectrophotometrically at 260 nm, assuming 1.0 A₂₆₀ corresponds to 50 μg mL⁻¹ dsDNA, using a NanoDrop 2000c spectrophotometer (Thermo Scientific, Australia). The concentration of oligonucleotides was determined by using an the value of ε₂₆₀ based on the sequence rather than the general assumption that 1.0 A₂₆₀ corresponds to 33 μg mL⁻¹ ssDNA, using a NanoDrop 2000c spectrophotometer (Thermo Scientific, Australia).
2.5.3 Agarose gel electrophoresis

A Mini-Sub Cell GT agarose gel electrophoresis system (Bio-Rad, Australia) was used routinely to analyse the size and purity of DNA. Quantities of DNA were estimated by comparison with the DNA size marker GeneRuler DNA ladder. Different percentages 0.75–1% (w/v) of agarose gel in 1 x TBE buffer (44.5 mM Tris-boric acid, 1 mM EDTA) containing 0.5 μg mL⁻¹ EtBr or 1 x TAE buffer (40 mM Tris, 1 mM EDTA, 20 mM acetic acid) containing 0.5 μg mL⁻¹ EtBr [349] were set using the Bio-Rad apparatus based on the target size of DNA. The same 0.5 x TBE buffer containing 0.5 μg ml⁻¹ EtBr was used as running buffer for the electrophoresis. Samples were loaded with 6 x DNA loading dye (0.25% (w/v) bromophenol blue, 40% (w/v) sucrose; [349]). Electrophoresis was carried out at 50 V to allow samples to enter the gel matrix, and then at between 60–80 V (4–5.5 V cm⁻¹) controlled by a PowerPac Basic power supply (Bio-Rad) until appropriate resolution of bands had been obtained. The results were visualised under 302-nm UV light and recorded by a GelDoc™ XR UV transilluminator controlled by Quantity One® v4.6.9 software (Bio-Rad, Australia).

2.5.4 Polymerase chain reaction (PCR)

All PCR experiments were performed using a DNA Engine Peltier Thermal Cycler PTC-0200 equipped with 48/48 dual Alpha unit or S1000 Thermal Cycler with dual 48/48 fast reaction module (both from Bio-Rad, Australia). Thin-wall flat cap PCR tubes (200-μL) were used to carry out all reactions. Polymerases with their own supplied reaction buffers (PfuUltra High-Fidelity DNA polymerase from Agilent, USA; BIOTAQ Red DNA polymerase from Bioline, Australia) were used in different types of PCR described below. Deoxynucleotides (dNTPs) stock (final concentration 25 mM each) was made up by mixing equal volumes of each (100 mM dATP, dCTP, dTTP, and dGTP).

2.5.4.1 Cloning PCR

Cloning PCR was used to amplify a particular section of a gene from a template plasmid or PCR product. Reaction mixtures for PfuUltra High-Fidelity DNA polymerase contained 1 x reaction buffer, 0.2 μM each of the designed forward and reverse primers, 0.2 mM dNTPs, ~6 ng of plasmid template, and 2.5 U PfuUltra High-Fidelity DNA polymerase in a volume of 50 μL. The reaction cycles of the PCRs using PfuUltra High-Fidelity DNA polymerase were typically as follows: a single pre-denaturing step at 95 °C for 2 min, followed by 31 PCR cycles of denaturing step at 95 °C for 30 s, annealing step at 55 °C for 30 s, extension step at 72 °C for 1 min per kb of the target
fragment, and finishing with a single final extension at 72 °C for 10 min. Reaction mixtures were cooled to 4 °C after the completion of the PCR cycle.

2.5.4.2 Colony PCR

Colony PCR was used to screen potential positive clones obtained from cloning experiments. Reaction mixtures (19 μL) contained 1 × PCR buffer (Roche, Switzerland), 2 mM MgCl₂, 0.2 μM each forward and reverse primers, 0.2 mM dNTPs and 1 U of BIOTAQ Red DNA polymerase. Single colonies were picked, washed in 50 μL of sterilised MilliQ water then 1 μL of washed cells was then added to the 19 μL reaction mixture. The reaction cycles of PCRs using BIOTAQ Red DNA polymerase were the same as the cycling conditions for PfuUltra High-Fidelity DNA polymerase (Section 2.5.4). After PCR, the reaction mix (10 μL) was loaded directly on an agarose gel for electrophoresis (Section 2.5.3) as the BIOTAQ Red DNA polymerase contains loading dye. The products from cloning and colony PCR were analysed by agarose gel electrophoresis; resulting positive PCR products and colonies were identified by their size in comparison with the GeneRuler DNA ladder mix (100–10,000 bp).

2.5.4.3 Nucleotide sequence determination

Preparation of samples for determination of the nucleotide sequences of plasmid DNA was carried out using the BigDye™ Terminator v3.1 sequencing reaction kit (Applied Biosystems, USA). Reaction mixtures (20 μl) contained 1 × BigDye sequencing buffer, 0.16 μM of either a forward or a reverse primer, ~150 ng of plasmid template, and 1 × BigDye Terminator. The reaction mixture was purified by the commercially available DyeEx 2.0 Spin kit (Qiagen, USA). The eluate was dried using in a Savant SpeedVac SC110 concentrator equipped with a PH40-11 rotor (Thermo Scientific, Australia) for ~1 h and then submitted to the sequencing facility at the School of Biological Sciences, University of Wollongong for analysis by a 3130xl Genetic Analyzer (Applied Biosystems, USA). Resulting sequences were analysed using ChromasPRO (Technelysium, Australia).

2.5.5 Restriction endonuclease digestion

All restriction endonucleases and their reaction buffers were supplied by New England Biolabs (USA). Restriction digestion reactions (50 μL) using 10 U of enzyme were carried out at 37 °C for 3 h. Reactions were terminated by a 20 min heat inactivation at 65 °C or by using a spin column to isolate DNA products (Section 2.5.6.2). Products of the digestion were visualised by agarose gel electrophoresis (Section 2.5.6.1).
2.5.6 Purification of DNA

2.5.6.1 Extraction from agarose gels

DNA fragments from restriction endonuclease digestion that had been separated in an agarose gel containing EtBr were visualised under low-energy UV light (>320 nm) from a bench-top UV lamp and identified by comparison with the DNA standard GeneRuler DNA ladder. Target bands were excised using a sterilised scalpel blade. DNA was extracted from the gel using the QIAquick® gel extraction kit (Qiagen, USA).

2.5.6.2 Extraction by spin column

DNA fragments were also purified directly from reaction mixtures using the QIAquick® PCR purification kit (Qiagen, USA). It uses a silica based resin spin-column, which has selective binding properties to recover DNA and remove contaminants such as free nucleotides and enzymes. This kit was typically used after PCRs to remove primers, dNTPs and DNA polymerase or after restriction digestion to remove restriction-digested ends of inserts (<100 nucleotides) generated by PCRs and subsequent restriction endonuclease digestions.

2.5.7 DNA ligation

Vectors and inserts were ligated using T4 DNA ligase (2.5 U; New England Biolabs, USA). Unless otherwise specified a 1:3 molar ratio of vector to insert was used in ligation reactions. This reaction mixture (20 μL) was incubated at ~16 °C overnight (~18 h). After incubation, 7 μL of the ligation reaction was transformed into the appropriate E. coli strain.

2.5.8 Transformation of competent E. coli cells

Competent E. coli cells were prepared using CaCl₂ by the procedure described in [350]. Electrocompetent cells of E. coli were prepared as described by [351]. Competent E. coli cells were transformed with plasmid DNA as described in [350], except that heat-treatment was routinely at 30 °C for 2 min. Electrotransformation was as described by [351].

2.6 Protein biochemistry methods

2.6.1 Fast protein liquid chromatography (FPLC) purification of proteins

FPLC systems (GE Healthcare, Sweden) were used for protein purification. The ÄKTA™ systems (ÄKTApurifier™, ÄKTAFPLC™ or ÄKTApure™) were controlled by
UNICORN™ software, which allows real-time manual control or programmed methods control for protein purification. All chromatography was carried out in a cold room at 6 °C. Protein elution from the columns was monitored by absorbance at a particular wavelength (\(A_{280}\) for protein) and conductivity (mS cm\(^{-1}\)).

### 2.6.2 Dialysis of proteins

Dialysis was performed for buffer exchange of proteins during protein purification or in other experimental procedures. Spectra/Por® standard RC (regenerated cellulose) dialysis tubing (Spectrum Laboratories, USA) or a Slide-A-lyzer MINI dialysis units (Thermo Scientific, Australia) was used for dialysis of proteins. An appropriate molecular-weight-cut-off (MWCO) of the dialysis membrane was selected to be smaller than the molecular weight of the proteins of interest. Typically, three changes of buffer (2 L each, unless otherwise specified) were carried out at ~3 h intervals in a cold room.

### 2.6.3 Sodium dodecyl sulphate (SDS)-polyacrylamide gel electrophoresis (PAGE)

SDS-PAGE was used to identify the protein content in various samples either overproduced in cells or fractions collected off columns during FPLC (Section 2.6.1). Pre-cast SDS-PAGE gels (4–20% gradient) were used in all experiments and run using Mini-PROTEAN® Tetra cell gel apparatus (Bio-Rad, Australia). The protocol used was based on that described by [352]. The running buffer for SDS-PAGE was 192 mM glycine, 25 mM Tris base, 0.1% (w/v) SDS. Samples of cells taken before and after induction of protein synthesis were resuspended to \(A_{600} = 10\) in equal parts of 2 \(\times\) loading buffer (300 mM Tris base, 15% (v/v) glycerol, 0.6% (w/v) bromophenol blue, 50 mM dithiotheritol and 1–2% (w/v) SDS) and BugBuster master mix before loading 20 µL onto the gel. Protein samples were mixed 1:1 with 2 \(\times\) loading buffer (as described above). Gel electrophoresis was controlled by PowerPac Basic power supply (Bio-Rad, Australia) and were typically run at 180 V for 32 min or until the dye front had reached the bottom of the gel.

### 2.6.4 Staining of SDS-PAGE gels with Coomassie brilliant blue

Gels were stained in warm 40% (v/v) methanol, 10% (v/v) acetic acid, 0.3% (w/v) Coomassie brilliant blue for 10 min. Gels were subsequently destained in warm destaining solution (10% (v/v) propan-2-ol and 10% (v/v) acetic acid) until background stain was removed.
2.6.5 Concentration of proteins

2.6.5.1 Ammonium sulphate precipitation
Protein solutions were precipitated by addition of 0.45 g mL\(^{-1}\) of solid ammonium sulphate and stirring on ice in a cold room for at least 1 h. Precipitated proteins were pelleted using a Sorvall RC-6\(^{+}\) Superspeed centrifuge equipped with an SS-34 for <40 ml tubes or SE-12 rotor for <10 mL tubes (Thermo Scientific, Australia) at 40,000 \(\times\) \(g\) for 30 min at 6 °C. The pellet was then resuspended with a glass rod in the appropriate volume of buffer. The resuspended proteins were dialysed against storage or assay buffer for later use.

2.6.5.2 Ultrafiltration
Protein solutions were concentrated by ultrafiltration using Vivaspin\textregistered Ultra-4 or Ultra-15 centrifugal filter units (Satorius, Australia) with a suitable MWCO based on the molecular weights of the proteins. Prior to application of the protein solution, the membrane of the filter units was washed 3 \(\times\) with MilliQ water, then 3 \(\times\) with buffer by centrifugation at 3100 \(\times\) \(g\) for 10 min at 6 °C using a Scanspeed 1248R centrifuge (Labogene, Denmark) equipped with a swinging bucket rotor to remove the preserve – glycerol. Protein solutions were centrifuged under the same conditions until the target concentration was reached. If the volume of protein solution was larger than the device could take, multiple cycles of centrifugation were undertaken, and newly applied protein solution was mixed gently with the previously concentrated solution by inverting the device before centrifugation to avoid aggregation on the membrane.

2.6.5.3 Acetone precipitation
When required, protein solutions were concentrated by acetone precipitation for SDS-PAGE gel analysis. Proteins in one volume were precipitated by addition of four volumes of ice-cold acetone on ice. After 20 min, protein pellets were collected by centrifugation at 21,000 \(\times\) \(g\) for 20 min and dried for 10 min in air at room temperature.

2.6.6 Quantification of proteins

2.6.6.1 Bradford assay
The Quick Start Bradford Protein Assay (Bio-Rad, Australia) [353] was used to determine the concentrations of protein samples containing nucleotides in the buffer e.g. ATP or ADP. ATP absorbs UV light at 280 nm as protein does [354]. This can lead to over-estimates of the real protein concentration. Values of \(A_{595}\) were read using disposable transparent plastic cuvettes in a BioPhotometer (Eppendorf, Germany). The
concentration of the protein was determined by reference to a standard curve generated from known concentrations of BSA.

2.6.6.2 UV absorbance at $A_{280}$ for purified proteins

The UV absorbance of proteins at 280 nm ($A_{280}$) was directly measured using a NanoDrop 2000c (Thermo Scientific, Australia). Molar extinction coefficients of proteins at 280 nm ($\varepsilon_{280}$) were calculated according to the ExPASy Proteomics Server (ExPASy – ProtParam tool; URL: http://web.expasy.org/protparam/). Since reducing agent was always present in the buffer, no oxidised cysteine (disulfide linkage between two cysteine residues) was formed in the protein sample, so the number of cysteine residues was not considered to contribute to the molar extinction coefficients of a particular protein [355].

2.7 Electrospray ionization (ESI)-mass spectrometry (MS)

The accurate molecular mass of purified proteins was routinely analysed after purification using ESI-MS. This allows another means to confirm protein purity or site-directed mutation of the protein other than sequencing of the gene. Mass spectra were acquired using a SYNAPT® (in Q–Tof mode) (Waters, USA) in positive mode; the instrument was controlled by MassLynx v4.1 software (Micromass, UK). The instrument was calibrated using CsI (10 mg mL$^{-1}$ in 70% (v/v) propan-2-ol) over the same $m/z$ range. For accurate mass determination of proteins, 50 µL of protein sample (~0.5–10 µM) were dialysed extensively against 0.1% (v/v) formic acid (Section 2.6.2).
Chapter 3

Structural and Interaction Studies of Domain IVa of the *Escherichia coli* τ Subunit and its Contribution to the $\alpha$–$\tau$ Interaction
Structural and Interaction Studies of Domain IVa of the 
Escherichia coli \( \tau \) Subunit and its Contribution to the \( \alpha-\tau \) 
Interaction

The \( \alpha-\tau \) interaction is a critical interaction within the DNA Pol III HE, effectively coupling the leading and lagging strand polymerases together. The \( \tau \) subunit is composed of five distinct domains, each with a separate function. Detailed biochemical studies have shown that domain V of \( \tau \) is responsible for binding to \( \alpha \). Further, this interaction is mediated through an intrinsically unstructured region of \( \tau \) that becomes structured upon interaction with \( \alpha \). The preceding domain, domain IVa has been solely attributed to binding the DnaB helicase. In this chapter, kinetic and structural studies of domain IVa and its contribution to the \( \alpha-\tau \) interaction are presented. SPR assays were used to determine the effect of domain IVa on the \( \alpha-\tau \) interaction using a truncation mutant, \( \tau_{C24} \). These assays showed a 2–3-fold greater binding affinity and slower dissociation rate compared to \( \tau_{C16} \). Complementary NMR spectroscopy experiments show some residues in domain IVa are also perturbed upon binding to \( \alpha \). Finally, the overall size and shape of the \( \alpha-\tau_{C24} \) complex was determined by small angle scattering and compared to the \( \alpha-\tau_{C16} \) complex. Small angle scattering measurements reveal a more compact structure than expected, and suggest distinguishable conformational changes to \( \alpha \) upon \( \tau \) binding.

Contributions are as follows: SPR experiments were performed under the guidance of Dr Slobodan Jergic, University of Wollongong. NMR spectroscopy was performed by Dr Kioyshi Ozawa, Australian National University. Dr Anthony Duff expressed perdeuterated \( \alpha \) at the National Deuteration Facility, ANSTO. SANS data were reduced and analysed with Dr Andrew Whitten, Bragg Institute, ANSTO. A \( \alpha-\tau_{C16} \) model was prepared by Prof. Thomas Huber, Australian National University. DNA-free and DNA-bound models of \( \alpha-\tau_{C16} \) were prepared by A/Prof. Aaron Oakley, University of Wollongong.
3.1 Introduction

3.1.1 The $\tau$ subunit of the E. coli clamp loader complex

The DNA Pol III HE is separated into three distinct sub-assemblies, the three-subunit core polymerase ($\alpha\varepsilon\theta$), the processivity-determining homodimeric sliding clamp ($\beta_2$), and the seven-subunit clamp loader complex $\tau_{n\gamma(3-n)}\delta\delta^\prime\psi\chi$ (where $n = 2$ or $3$ in the HE). The overall composition of the minimal Pol III HE is believed to be ($\alpha\varepsilon\theta$)$_2$–$\tau_2$$\delta\delta^\prime\psi\chi$–($\beta_2$)$_2$. It contains two or three $\alpha\varepsilon\theta$ cores, one dedicated to synthesis of the leading strand and at least another to the lagging strand [205] (See Figure 1.1). For simultaneous coordinated DNA replication to occur, the two or three Pol III cores ($\alpha\varepsilon\theta$) are held in place by the interaction between the $\alpha$ and $\tau$ subunits, effectively dimerising the polymerases at the replication fork [Gao:2001gu 17,39,254]. Without these interactions, severe deficiencies within the replisome occur and it is unable to support full replicative functions. [254,356,357]. The interaction between $\alpha$ and $\tau$ is therefore a critical structural link in the replisome, bridging two $\alpha$ subunits to the CLC.

The $\tau$ (71 kDa) and $\gamma$ (47 kDa) subunits are encoded by the dnaX gene; the shorter $\gamma$ product results from programmed ribosomal frameshifting occurring 50% of the time in vivo [251,252,358]. Consequently, the C-terminal 212 amino acid residues of $\tau$ are not present in $\gamma$. The $\tau$ subunit is comprised of five domains, each providing a specialised function (Figure 3.1). The N-terminal domains I–III are common to $\gamma$; the additional 24 kDa C-terminal fragment unique to $\tau$ is connected to the first three domains by an apparently flexible proline rich linker region [359]. The unique 24 kDa region contains two domains: the 8 kDa domain IVa (in $\tau_{C24}$) has been associated with DnaB helicase binding [35] and the 16 kDa domain V ($\tau_{C16}$) binds to the C-terminal region of the $\alpha$ subunit [37].

The $\alpha$ subunit is the 130 kDa product of the dnaE gene and is ordered in domains that confer its special replicative properties (Figure 3.1). Together, these domains allow $\alpha$ to catalyse the nucleotidyl transferase reaction, adding nucleotides to the 3'-OH of the growing DNA strand.
Figure 3.1 | The domain structures of α and τ subunits. (Top panel) The domains present in DNA polymerase α. (Bottom panel) the domains of the τ and γ subunits; the position where γ terminates is indicated. The γ subunit shares the first 430 amino acids with τ but with the last amino acid Ser431 (present in τ) otherwise changed to a glutamic acid (Ser431Glu). Domains I–III shared by τ and γ are involved in loading of β₂ onto DNA. Domain IV is located between residues 413–496 and overlaps at the N-terminus by 17 C-terminal residues of γ. Only residues 430–496 unique to τ (except Lys430, shared by both τ and γ) are apparently responsible for binding to DnaB. Domain V (residues 497–643), only present in τ, are involved in binding the DNA polymerase α subunit. Truncation mutants of τ mentioned in this Chapter and following Chapters are labelled. The 24 kDa domain IVα and V truncation mutant τC24 is composed of a Met residue followed by the last 230 residues of τ (430–643). The mutant τC22 (residues 430–625) is identical to τC24 except that it is missing the 18 C-terminal residues. The 16.5 kDa domain V truncation mutant τC16 is composed of an initial Met residue and the extra 145 C-terminal residues of τ (residues 499–643). Finally, τC14 is identical to τC16 except that it is missing 18 C-terminal residues (residues 499–625).

There is no structure yet known of the full DNA Pol III HE complex in any of its replicative states. Progress has been made with various holoenzyme subassemblies trapped in some unique conformational states. These efforts have resulted in 3D structures of: α1–917 [19], δ“ [275], ε2–186 (ε186; [23], γ1–243 in nucleotide free and bound forms [360], θ [234], the homomeric binary complex β₂ in isolation [51] and with DNA [52], the heteromeric binary complexes δ–β₂ [42], χψ [262], and ε186–θ [25,28], and the pentameric (γ31–373)₃–δ–δ‘ complex in nucleotide free [38] and nucleotide bound forms [270]. Including these high-resolution structures, structural models of the αεθ–β₂ complex have also been described [15]. Unfortunately, the structure of the αεθ–
The solution structure of domain V of *E. coli* τ, derived from τC16 by deletion of 18 C-terminal residues (called τC14) has been determined by NMR spectroscopy. The fold is unique to τ subunits and assumes a KH type-II fold [39,361]. The fold of domain V of τ includes N- and C-terminal helices that extend from the globular core into the solvent (Figure 3.2a). These helices have a relative orientation that defines a fixed angle relationship between the domain termini (Figure 3.2b). The 26 C-terminal residues (residues 618–643 in the τ sequence) are intrinsically disordered and direct the interaction with α. Mutagenesis and deletion studies have proposed the induction of α-helices in this unstructured region that become ordered upon binding to α [17]. This hypothesis was proven correct by the X-ray crystal structure of α C-terminal domain (residues 963–1160 comprised of the OB and τ-binding domains; αCTD) bound to the τC16 fragment, where the αCTD was fused to the N-terminus of τC16 via a flexible 16 amino acid linker (Dr Zhi-Qiang Xu *et al.*, unpublished; Figure 3.3a,b).
Figure 3.2 | The solution structure of τC14 determined by NMR. a, Superposition of the 20 lowest energy NMR conformers of τC14 (PDB: 2AYA). b, The fold of τC14 has N- and C-terminal helices that extend from a globular core. Helices 1 (h1) and 6 (h6) form close contacts that define their relative orientation. Moreover, the C-terminal region of h6 is conformationally unstable and its mobility increases along the polypeptide chain towards the C-terminus.

The structure confirms this interaction occurs through an intrinsically unstructured region of the protein that folds into a defined structure upon interaction with its binding partner [362,363]. The previously unstructured C-terminal residues of τC16 adopt a helical structure followed by a highly twisted β-structure comprised of two β-strands linked by a helical turn. The interaction interface between αCTD and τC16 is extensive, containing 16 hydrogen bonds and 3 salt bridges (Dr Zhi-Qiang Xu, personal communication). Compared to the recent structure of T. aquaticus α–τC complex, E. coli τC16 and T. aquaticus τC have different structures and interact differently. This is not surprising considering the two τ proteins are not homologues, sharing less than 14% sequence identity in the C-terminal region [364].

Figure 3.3 | The X-ray crystal structure of the αCTD–τC16 complex. a, Cartoon representation of the structure of the αCTD (orange) and τC16 (blue) complex. The black dashed line represents the flexible linker. The inset in a shows a zoomed view of the unstructured region of τC16 that becomes structured upon binding to α. b, The interaction interface between αCTD and τC16 is extensive; residues participating in the interface are coloured red (Dr Zhi-Qiang Xu et al., unpublished).
3.1.3 Using a surface plasmon resonance (SPR) biosensor to study protein–protein interactions

SPR is a powerful technique commonly used to measure biomolecular interactions in real time in a label free environment. Biomolecular interactions are typically measured by immobilising one of the target proteins (ligand) on a carboxymethylated dextran coated thin gold layer on glass. The other protein partner (analyte) is passed over the ligand surface in a microfluidic flow chamber. In Chapters 3, 4 and 6, ProteOn XPR36 and BIAcore T200 instruments are used to investigate kinetics and thermodynamics of protein–protein interactions.

The output for a SPR instrument is based on a physical process that occurs when plane-polarised light is reflected from a gold film under total internal reflection conditions (Figure 3.4a). Briefly, a glass layer (has a high refractive index) coated with a thin layer of gold is directly in contact with buffer (low refractive index medium) in a flow channel. A laser light source is reflected under conditions of total internal reflection (TIR) from the back of the thin gold layer. Under these conditions an electric field (evanescent wave) propagates perpendicularly across the interface into the lower refractive index medium to a depth of ~150 nM [365]. At a well-defined resonance angle and wavelength of light, a surface plasmon is excited from the thin gold film and resonance is then induced. In turn, this causes an intensity decrease of the reflected light. The resonance angle is correlated with the evanescent wave and refractive index at the surface interface, and is simply related to changes of mass on the surface. The changes in mass on the sensor surface are recorded in real time and can be subsequently transformed into a sensorgram, where resonance units (RU) are proportional to mass (where 1 RU corresponds to ~1 pg mm$^{-2}$ of protein on the chip surface) and are plotted as a function of time. This allows kinetic rates (of association and dissociation) and thermodynamic (strength of binding) information to be derived.

A typical sensorgram generated from SPR is represented in Figure 3.4b. It depicts the time course of an interaction of analyte with an immobilised ligand. Firstly, the ligand is immobilised on the sensor chip surface. This is followed by injection of the analyte through the flow channels to measure the association phase of the interaction, followed by running buffer alone to measure the dissociation phase. The surface must then be regenerated using a buffer that washes off all the analyte to recover the original surface for subsequent interaction experiments. Described above is a typical data collection cycle. To acquire enough data for model fitting, an analyte concentration
series must be measured. The kinetic information of the interaction is obtained through model fitting. The thermodynamics of the interaction are calculated from the kinetics or for interactions that proceed to equilibrium, from a binding isotherm expressed as equilibrium values of $RU$ vs. analyte concentration. Additionally, as the $RU$ is proportional to the mass on the surface, other parameters such as stoichiometry of interactions can also be estimated from the maximum binding capacity of the ligand (in $RU$) and is denoted as $R_{\text{max}}$. 

![Diagram of surface plasmon resonance (SPR) and setup of a SPR biosensor.](image)

**Figure 3.4** | **Principle of surface plasmon resonance (SPR) and setup of a SPR biosensor.**  
**a**, The power of SPR technology comes from the ability to monitor changes in the refractive index within the immediate environment of the evanescent wave that causes a change of the angle where a drop in intensity can be observed in the reflected light at the detector. **b**, A graphical representation of a typical sensorgram of the signal produced during a binding experiment. I. A ligand immobilised on the sensor chip. Initially buffer is running over the sensor chip, and no change in signal is observed. II. The analyte is injected into the microfluidic chamber and associates with the ligand. This causes a change in the refractive index that is monitored as a change in angle where the decrease in intensity of the reflected light is observed. The change in angle is visualised in the sensorgram. III. Association and dissociation of the analyte and the ligand occur at equal rates during the equilibrium phase. IV. Analyte injection is replaced by buffer and dissociation of the analyte from the ligand is observed. As a result, a reduction in refractive index and the angle is observed. V. At any point of the dissociation, regeneration buffer can be injected, which will remove any analyte still associated with the chip surface. Figure adapted from [366].
3.1.4 Using small angle scattering to study biomolecular structures in solution

The use of small angle scattering of X-rays (SAXS) or neutrons (SANS) provides a robust method for analysing the overall structures of biological macromolecules in solution. Scattering is a phenomenon whereby a beam of radiation or particles is diverged from its initial trajectory by dissimilarities in a medium it crosses. A small angle scattering (SAS) experiment is conceptually simple and requires a highly collimated X-ray or neutron beam used to illuminate a sample. The scattered radiation is recorded on a detector and the direct beam is usually absorbed by a beam stop. Experiments involving biological macromolecules such as proteins in solution therefore require two separate measurements, one of the solvent alone (buffer) and the second of the macromolecules of interest in solution. It is assumed that the solvent is featureless with a constant scattering density. SAS is a low-resolution technique as it provides no information on atomic coordinates and is inherently constructed from spherically averaged scattering intensities or scattering length densities. It is important to recognise that the radial averaging of particles in solution during a SAS experiment is the limiting criterion for reconstruction of atomic details, more so than the resolution limits of the experiment. The power of SAS is in providing high-precision information regarding the size and shape of particles [367]. Furthermore, the scattering intensity can be easily computed for simple geometric shapes. SAS curves are used to calculate several overall model-independent parameters, reviewed in [368,369].

Faithful structure reconstruction of macromolecules using SAS is largely dependent on the ability to eliminate inter-particle repulsion/attraction, poly-dispersity, impurities, and buffer mismatches. Eliminating these factors are typically ensured by employing extensive and often elaborate purification procedures that come from thorough biochemical characterisation of the proteins and/or their protein–protein/nucleic acid complexes. However, many samples, especially intrinsically unstructured proteins and large macromolecular complexes, suffer from problems due to the presence of small sub-populations of either larger soluble aggregate forms or degraded complexes. In turn these sub-populations can adversely affect the quality of scattering data, often making interpretation difficult. Size exclusion chromatography coupled SAXS (SEC-SAXS) is a technique used to combat these issues and was first demonstrated by [370]. SEC-SAXS is a recent technical advancement that allows chromatography of samples immediately before irradiation with X-rays, allowing acquisition of reliable scattering data from otherwise non-homogenous proteins or large protein complexes.
susceptible to aggregation. Additionally, errors from incorrect buffer subtraction are also minimised. The experimental setup is analogous to SEC-MALLS (multi-angle laser light scattering) where the size exclusion column is linked upstream to the SAXS sample capillary so that continuous sample scattering can be recorded. The time resolution requirement combined with low scattering intensity caused by sample dilution caused by the chromatography step requires the high flux and sensitivity of synchrotron SAXS beamlines.

Fluctuations in electron density (in SAXS) and scattering length density (in SANS) in solution allow contrast in SAS experiments to be systematically manipulated by changing either the electron density and scattering length density of the environment (i.e. the solvent). This principle was exploited in early protein crystallography experiments to elucidate the molecular envelope of haemoglobin [371]. Later it was utilised in SAXS to determine the protein envelope of tomato bushy stunt virus [372]. In general, contrast variation in SAXS is not practical for biological macromolecules as they have very similar elemental compositions, meaning two different proteins will typically have very similar X-ray scattering density. However, the difference in scattering power (or contrast) between various constituents in a sample resulting from different scattering properties forms a fundamental aspect in SANS studies of biological samples [373].

In contrast to X-rays, different isotopes of the same element may have differing abilities to scatter neutrons (i.e. how “big” the nucleus appears to the neutron and therefore how strongly neutrons will be scattered from it). This is most notable in the large difference in the coherent scattering length of hydrogen ($^1$H) and its stable isotope deuterium ($^2$H) being $-0.56 \times 10^{10}$ and $6.33 \times 10^{10}$ cm$^{-2}$, respectively. When solvating H$_2$O is replaced by D$_2$O, exchangeable hydrogen atoms in the sample will be replaced by $^2$H, changing their scattering length density. Manipulation of these properties in combination with specific perdeuteration of individual components in a protein–protein or protein–nucleic acid complex provides an effective method to separate structural information such as shape and relative orientation of individual components within the same system in solution.

### 3.1.5 Aims of this Chapter

In this Chapter, I present SPR data that show a contribution of domain IVa of $\tau$ to the $\alpha–\tau$ interaction; this contribution causes a slower dissociation rate which results in an increase in binding affinity of 3–5-fold for $\tau_{C24}$ compared to $\tau_{C16}$ under the same
experimental conditions. I further examine the contribution of residues in domain IVa through changes in flexibility using $^1$H$^{15}$N-HSQC NMR spectroscopy. Finally, SAXS and SANS of the $\alpha-\tau_{C16}$ and $\alpha-\tau_{C24}$ complexes are used to measure the overall shape, and dimensions, and to derive structural models of both complexes. These models were subsequently compared with structures generated from both DNA-free and DNA-bound states of $\alpha-\tau_{C16}$ complex built from published cryo-EM structures.
3.2 Materials and Methods

3.2.1 Plasmid construction

For general methods related to plasmid construction, refer to Chapter 2. Specific vectors and promoters are described in Section 2.1.2. Unless otherwise specified, all plasmids were hosted in *E. coli* strain AN1459 (Section 2.1.1) during construction. Plasmid constructions were confirmed by nucleotide sequence determination (Section 2.5.4.3).

*Construction of plasmid pJSL1697 (encoding bio*τ*₂₄)*: This plasmid which directs overproduction of bio*τ*₂₄, was constructed by ligation of the 650 bp τ₂₄ NdeI–EcoRI fragment from pSJ1330 [17] between the corresponding sites in pKO1274 (Section 2.1.2) to give a T7 promoter N-terminal biotin-tagged τ₂₄ overexpression vector, pJSL1697.

3.2.2 Overproduction and purification of bio*τ*₁₆

Overproduction and purification of bio*τ*₁₆ was based on a previously described method [17]. Plasmid pKO1294 [17] is a λ-promoter vector that directs overproduction of N-terminally biotin-tagged bio*τ*₁₆. *E. coli* strain BL21(λDE3)recA/pKO1294 was grown at 30 °C in LBT medium containing ampicillin (100 mg L⁻¹) and 100 μM D-biotin to *A*₆₀₀ = 1.0. To induce overproduction of bio*τ*₁₆, the temperature was rapidly increased to 42 °C. Cultures were shaken for a further 2.5 h, then chilled in ice. Cells were harvested by centrifugation (16,900 × *g*; 8 min), frozen in liquid N₂ and stored at −80 °C.

Buffers used for purification of bio*τ*₁₆ were: lysis buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM dithiothreitol, 20 mM spermidine, 10% (w/v) sucrose); buffer A*τ*₁₆ (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM dithiothreitol, 5% (v/v) glycerol); buffer B*τ*₁₆ (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 3 mM dithiothreitol, 100 mM NaCl, 20% (v/v) glycerol).

After thawing, cells (6 g from 5 L of culture) were resuspended in 110 mL of lysis buffer and three protease inhibitor cocktail tablets and 0.7 mM PMSF were added to inhibit proteolysis. The cells were then lysed by being passed twice through a French press (12,000 psi). The lysate was clarified by centrifugation (35,000 × *g*; 30 min) to yield the soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.4 g mL⁻¹) and stirring for 60 min were collected by
centrifugation (35,000 \times g; 30 \text{ min}) and dissolved in buffer A_{TC16}+150 \text{ mM NaCl (35 mL)}. The solution was dialysed against three changes of 2 L of the same buffer, to yield Fraction II (70 mL).

Fraction II was applied at 1.5 mL min\(^{-1}\) onto a column (2.5 \times 10 \text{ cm}) of DEAE-650M resin that had been equilibrated in the same buffer. Fractions containing proteins that did not bind to the column were pooled and dialysed against three changes of 2 L of buffer A_{TC16}. The dialysate (Fraction III, 65 mL) was loaded at a flow rate of 1.5 mL min\(^{-1}\) onto a column (2.5 \times 10 \text{ cm}) of the same resin, now equilibrated in buffer A_{TC16}. After the column had been washed with 180 mL of the same buffer, proteins were eluted using a linear gradient (420 mL) of 0–160 mM NaCl in buffer A_{TC16}. Bio_{TC16} eluted in two discrete peaks of which the first one (dominant peak \(\alpha\)) eluted at about 50 mM NaCl while the second (peak \(\beta\)) eluted at 80 mM NaCl. Fractions from the two peaks containing purified proteins were separately pooled and dialysed against three changes of 2 L of buffer B_{TC16}, to give Fractions IVa (50 mL, containing 104 mg of protein) and IVb (30 mL, containing 34 mg of protein). Aliquots were frozen in liquid N\(_2\) and stored at –80 °C.

The molecular weights of purified proteins determined by ESI-MS from peak \(\alpha\) (18,542 ± 1 Da) and peak \(\beta\) (18,769 ± 2 Da) indicated the following:

(a) The N-terminal methionine had been removed from both samples;

(b) Only the protein in peak \(\beta\) was biotinylated (bio*_{TC16}) while the protein in peak \(\alpha\) did not have attached biotin (bio_{TC16});

(c) The complete separation of biotinylated and non-biotinylated protein had been achieved.

### 3.2.3 Overproduction and purification of bio_{TC24}

\textit{E. coli} strain BL21(\lambda,DE3) recA/pJSL1697 (Section 3.2.1) was grown at 37 °C in LBT medium containing ampicillin (100 mg L\(^{-1}\)) and 100 \muM D-biotin to \(A_{600} = 0.6\). To induce overproduction of bio_{TC24}, 1 mM IPTG was added to the shaking culture. Cultures were shaken for a further 3 h, and then chilled in ice. Cells were harvested by centrifugation (16,900 \times g; 8 min), frozen in liquid N\(_2\) and stored at –80 °C.

Buffers used for purification of bio_{TC24} were: lysis buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM dithiothreitol, 20 mM spermidine); buffer A_{TC24} (25 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM dithiothreitol, 5% (v/v) glycerol); buffer B_{TC24} (15 mM sodium...
phosphate, pH 6.5, 1 mM EDTA, 2 mM dithiothreitol, 5% (v/v) glycerol); buffer C\textsubscript{TC24} (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 3 mM dithiothreitol, 100 mM NaCl, 20% (v/v) glycerol).

After thawing, cells (4 g from 2 L of culture) were resuspended in 60 mL of lysis buffer and two protease inhibitor cocktail tablets and 0.7 mM PMSF were added to inhibit proteolysis. The cells were lysed by being passed twice through a French press (12,000 psi), cell debris was then removed from the lysate by centrifugation (35,000 \times g; 20 min) to yield the soluble Fraction I. Proteins in Fraction I were then fractionated by addition of solid ammonium sulphate (0.4 g mL\textsuperscript{−1}) and stirring for 60 min, harvested by centrifugation (35,000 \times g; 30 min) and dissolved in buffer A\textsubscript{TC24}+130 mM NaCl (30 mL). The solution was dialysed against two changes of 2 L of the same buffer, to yield Fraction II.

Fraction II was applied at 2.5 mL min\textsuperscript{−1} to a column (2.5 \times 10 cm) of DEAE-650M resin that had been equilibrated in buffer A\textsubscript{TC24}+150 mM NaCl. Fractions containing bio\textsubscript{TC24} did not bind to the resin and were pooled and dialysed against two changes of 2 L of buffer A\textsubscript{TC24}. The dialysate (Fraction III, 60 mL) was loaded at a flow rate of 2.5 mL min\textsuperscript{−1} onto a column (2.5 \times 10 cm) of the same resin, now equilibrated with buffer A\textsubscript{TC24}. The column was washed with buffer A\textsubscript{TC24} until bio\textsubscript{TC24} eluted between 1 and 1.5 column volumes in a sharp peak. Fractions containing bio\textsubscript{TC24} were pooled and dialysed against three changes of 2 L of buffer B\textsubscript{TC24}, to give Fraction IV (40 mL).

Fraction IV was applied at 0.6 mL min\textsuperscript{−1} onto a column (2.5 \times 14 cm) of phosphocellulose resin (Whatman P11) that had been equilibrated in buffer B\textsubscript{TC24}. After the column had been washed with 75 mL of buffer B\textsubscript{TC24}, bio\textsubscript{TC24} was eluted using a linear gradient (300 mL) of 0–550 mM NaCl in buffer B\textsubscript{TC24}. It eluted in a single peak at about 210 mM NaCl. Fractions containing purified bio\textsubscript{TC24} were pooled and dialysed against two changes of 2 L of buffer C\textsubscript{TC24}, to give Fraction V (12 mL containing 8 mg). Aliquots were frozen in liquid N\textsubscript{2} and stored at –80 °C.

### 3.2.4 In vitro biotinylation of bio\textsubscript{TC24}

Since I was unable to biotinylate bio\textsubscript{TC24} in vivo (no biotin has been incorporated, as determined by ESI-MS), in vitro biotinylation was performed. Highly purified His\textsubscript{6}-tagged E. coli biotin ligase was a generous gift from Dr Slobodan Jergic. Briefly, one part of biomix buffer (50 mM Tris.HCl, 250 mM bicine, pH 8.3, 50 mM ATP, 50 mM magnesium acetate, 250 mM D-biotin) was mixed with and part of 170 \mu M bio\textsubscript{TC24} in a
buffer (20 mM Tris.HCl, pH 7.6, 0.5 mM EDTA, 1 mM dithiothreitol, 150 mM NaCl) and three parts of MilliQ water, finally 1.8 μM of biotin ligase was added to the reaction mix. The mixture was treated at 25 °C for 3 h and then dialysed in 2 L of buffer E_{T_{C24}} (20 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM dithiothreitol, 5% (v/v) glycerol) at 6 °C to yield Fraction I.

Faction I was applied at 1.5 mL min⁻¹ onto a column (1.0 x 2 cm) of SuperQ-650M resin that had been equilibrated in the same buffer. After the column had been washed with 50 mL of buffer E_{T_{C24}}, biotinylated bio_{T_{C24}} (bio*_{T_{C24}}) was eluted as a single peak at ~60 mM NaCl using a linear gradient (85 mL) of 0–800 mM NaCl in buffer E_{T_{C24}}. Fractions containing purified bio*_{T_{C24}} were pooled and dialysed against two changes of 2 L of buffer C_{T_{C24}}, to give Fraction V (1.8 mL; containing 1.3 mg of protein).

The molecular weight of biotinylated bio*_{T_{C24}} determined by ESI-MS (26243 ± 0.3 Da) compares well to the theoretical mass of 26245.9 Da for bio*_{T_{C24}} and indicated that the protein was completely biotinylated.

## 3.2.5 Overproduction and purification of T_{C24}

Purification of T_{C24} was based on previously described method [17]. Plasmid pSJ1330 [17] is a λ-promoter vector that directs overproduction of T_{C24}. E. coli strain BL21(λDE3)recA/pSJ1330 (Section 2.3.1) was grown at 30 °C in LBT medium containing ampicillin (100 mg L⁻¹) to A₆₀₀ = 0.6. To induce overproduction of T_{C24}, the temperature was rapidly increased to 42 °C. Cultures were shaken for a further 2 h, and then chilled on ice. Cells were harvested by centrifugation (16,900 x g; 8 min), frozen in liquid N₂ and stored at –80 °C. After thawing, (6.5 g of cells from 4 L of culture), T_{C24} was purified according to the procedure used for bio{T_{C24}} (Section 3.2.3); a total of 65 mg of T_{C24} was obtained.

## 3.2.6 Overproduction and purification of \(^{15}\)N-labelled T_{C24}

A sample of uniformly \(^{15}\)N-labelled T_{C24} was prepared from cells of strain BL21(λDE3)recA/pSJ1330 grown in 2 L of minimal medium for \(^{15}\)N-labelling (Section 2.3.5) supplemented with 100 mg L⁻¹ ampicillin. Cultures were shaken overnight at room temperature to give A₆₀₀ = 1.2. To induce overproduction of \([^{15}\text{N}]-T_{C24}\) the temperature was rapidly increased to 42 °C. Cultures were shaken for a further 4 h,
then chilled on ice. Cells (4 g from 2 L culture) were harvested by centrifugation (16,900 x g; 8 min), frozen in liquid N₂ and stored at –80 °C.

[15N]-τC₂₄ was purified according to procedure used for bioτC₂₄ (Section 3.2.3); a total of 21 mg of [15N]-τC₂₄ was purified. The molecular weight of the product protein determined by ESI-MS (24,302 ± 1 Da) compares well to the theoretical mass of 24,306.2 Da for [15N]-τC₂₄ (assuming that the all N are 15N labelled). The purified sample was dialysed against three changes of 2 L of NMR buffer (Section 3.2.12), and then concentrated by ultrafiltration using an MWCO 4,000 Ultrafree-4 centrifugal filter unit. The 275 μM solution (~2 mL) of [15N]-τC₂₄ was used for NMR measurements (Section 3.2.12).

3.2.7 Preparation of τC₁₆-agarose affinity column

Following a procedure developed by Mr Nick Horan (University of Wollongong, unpublished). An affinity resin for purification of full-length (unproteolysed) α was prepared by conjugation of bio*τC₁₆ (α-binding domain V of τ) to high-capacity streptavidin-agarose resin (Pierce Biotechnology). bio*τC₁₆ (15 mL; 12 mg) was added drop-wise with gentle stirring into a suspension of 6 mL of resin in 11 mL of 50 mM Tris–HCl pH 7.6, 2 mM dithiothreitol, 1 mM EDTA, 50 mM NaCl at 6 °C over 20 min. Unconjugated streptavidin-agarose resin (2 mL) was added to a column and allowed to settle, then the suspension of τC₁₆-conjugated resin was poured over it. The column (1 x 10 cm) was then washed with 150 mL of 50 mM Tris.HCl pH 7.6, 2 mM dithiothreitol, 1 mM EDTA, 50 mM NaCl and stored at 6 °C in 50 mM Tris.HCl, 5 mM dithiothreitol, 1 mM EDTA, 50 mM NaCl, 0.03% (v/v) NaN₃.

3.2.8 Overproduction and purification of the α subunit

Plasmid pND517 [50] is a dnaE⁺ tac promoter vector that directs overproduction of the α subunit. The E. coli strain BL21(λDE3)recA/pND517 was grown at 30 °C in LBT medium containing ampicillin (100 mg L⁻¹; Section 2.3.1) to A₆₀₀ = 0.6. To induce overproduction of α, 1 mM IPTG was added to the shaking culture. Cultures were shaken for a further 3.5 h, and then chilled in ice. Cells were harvested by centrifugation (16,900 x g; 8 min), frozen in liquid N₂ and stored at –80 °C.

Buffers used for purification of α were: lysis buffer (50 mM Tris.HCl, pH 7.6, 2 mM EDTA, 2 mM dithiothreitol, 20 mM spermidine; 10% (w/v) sucrose); buffer Aα (50 mM
Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM dithiothreitol, 20% (v/v) glycerol); buffer Bα (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 10 mM dithiothreitol, 10% (v/v) glycerol); buffer Cα (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 3 mM dithiothreitol, 100 mM NaCl, 20% (v/v) glycerol).

After thawing, cells (45 g from 9 L of culture) were resuspended in lysis buffer (400 mL) and nine protease inhibitor cocktail tables and 0.7 mM PMSF were added to inhibit proteolysis. Cells were lysed by the addition of 0.2 mg mL⁻¹ chicken egg lysozyme to the suspension and left to stir at 6 °C for 20 min. The lysate was then heated to 37 °C for 10 min in a water bath. During heating, the cell suspension was divided into 50 mL portions and gently mixed each minute to promote even lysis. Samples were then pooled and left stirring at 6 °C for a further 20 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 (0.4 g mL⁻¹) and stirring for 60 min at 6 °C were collected by centrifugation (35,000 × g; 30 min) and dissolved in 60 mL of buffer Aα+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 at 1.5 mL min⁻¹ onto a column (2.5 × 10 cm) 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 chromatogram 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 (200 mL). This step removes contaminating nucleic acids that otherwise interfere with the next step.

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 60 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 α subunit eluted as a single peak at ~140 mM NaCl. Fractions under the peak were analysed using SDS-PAGE (Section 2.6.3) and those containing the α subunit were collected and pooled to yield fraction IV (75 mL). At this stage, α is only partially purified (Figure 3.5, lane 1) and still contains proteolytic fragments generated by removal of C-terminal residues; however, most contaminating proteases have been eliminated.
Fraction IV was directly applied at 1 mL min⁻¹ onto a column (1 × 10 cm) of τC₁₆-agarose that had been equilibrated in buffer Bα+20 mM MgCl₂. After the column had been washed with 15 mL of buffer Bα+0.6 M MgCl₂ and unbound proteins had been washed away, the highly purified α subunit was eluted using a linear gradient (20 mL) of 0.6–4.0 M MgCl₂ in buffer Bα. α eluted as a single peak at ~2.8 M MgCl₂. Fractions under the peak were immediately pooled and dialysed against two changes of 2 L of buffer Cα, to give Fraction V (50 mL, containing 207 mg of protein). Aliquots were frozen in liquid N₂ and stored at –80 °C.

Figure 3.5 | The τC₁₆-agarose affinity column removes the proteolytic C-terminal fragments of the α subunit. SDS-PAGE analysis of purified α (Coomassie blue stained). Lane 1: Fraction IV before loading onto the τC₁₆-agarose affinity column; Lanes 2 and 3: peak fractions of α after elution from the τC₁₆-agarose affinity column.

3.2.9 Overproduction and purification of perdeuterated α

A high density 1 L culture of E. coli expressing perdeuterated α subunit was produced using the bioreactor at the National Deuteration Facility at ANSTO (Lucas heights, Australia). The expression protocol is previously described in [374]. In brief, E. coli strain BL21(λDE3)recA was freshly transformed with the pND517 plasmid and rested in 0.3 mL of SOC medium for 2 h at 37 °C before being used to inoculate a 50% (v/v) D₂O ModC1 preculture containing glycerol-d₈ as the sole carbon source. The bacteria were adapted to D₂O at 37 °C, shaking at 220 rpm in sealed flasks filled to approximately 5% of capacity, by preparing successive precultures with increasing concentrations of D₂O (50, 90, and 100%), the first overnight and the latter for at least two generations (>2 h). The 100% D₂O preculture was obtained by gently pelleting the ice-chilled cells from the 90% D₂O preculture, followed by resuspension in 100% D₂O medium. The 100% D₂O adapted preculture was subsequently used to inoculate a bioreactor containing a 1 L final volume of perdeuterated ModC1 medium (Section 2.3.6) to an A₆₀₀ = 0.2. The bioreactor culture was maintained at 37 °C in aerobic conditions throughout, and 28%
(v/v) ammonium-d$_4$ deuter oxide was fed to maintain a minimum pH of 6.6. When $A_{600} = 16$ was reached, the temperature was reduced to 20 °C and expression was induced with IPTG to a final concentration of 1 mM. Samples were taken periodically to ascertain the expression levels of $\alpha$. The culture was grown until the carbon source was exhausted, then harvested by centrifugation ($8,000 \times g$; 30 min), frozen in liquid N$_2$ and stored at –80 °C. The total yield of cells from 1 L was 80 g.

After thawing, cells (23 g) were lysed and perdeuterated $\alpha$ was purified through Fraction V essentially as described for $\alpha$ (Section 3.2.8). The degree of deuteration was determined by ESI-MS comparison of hydrogenated and perdeuterated samples and was found to be 72% (Appendix Figure 1).

### 3.2.10 Reconstitution and purification of $\alpha$–C$_{24}$ and $\alpha$–C$_{16}$ complexes

To reconstitute the $\alpha$–C$_{16}$ and $\alpha$–C$_{24}$ complexes for both SAXS and SANS studies, separately purified $\alpha$, C$_{16}$ or C$_{24}$ were dialysed together in buffer A$\alpha$ (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM dithiothreitol, 10% (v/v) glycerol)+70 mM NaCl and left at 6 °C overnight. Proteins were mixed in a 1:3 molar ratio to form Fraction I; the excess of C$_{16}$ or C$_{24}$ ensures that all of $\alpha$ in the mixture has bound C$_{16}$ or C$_{24}$. To isolate intact protein complexes, Fraction I was directly applied at 1 mL min$^{-1}$ onto a column (1.5 x 5 cm) of SuperQ-650M resin equilibrated with the same buffer. After the column had been washed with 20 mL of Buffer A$\alpha$+70 mM NaCl to remove excess of C$_{16}$ or C$_{24}$, pure $\alpha$–C$_{16}$ or $\alpha$–C$_{24}$ complex was eluted in a single peak at 150 mM NaCl using a linear gradient (20 mL) of 70–1,600 mM NaCl in Buffer A$\alpha$. Fractions under the peak were pooled and dialysed against one change of 2 L of buffer C$\alpha$, to give Fraction II. Aliquots were frozen in liquid N$_2$ and stored at –80 °C.

### 3.2.11 Analysis of protein–protein interactions by SPR methodology

Methods were essentially as used previously [17], except that all experiments were carried out at 20 °C (instead of 25 °C) and a 6 x 6 multiplex ProteOn XPR-36 system (Bio-Rad, Australia) was used instead of a BIAcore 2000 instrument. The ProteOn XPR-36 instrument is advantageous as it utilises an array of 6 x 6 channels of flow cells on a single chip. This allows for interactions to be studied in a one-shot approach by simultaneous injection of different concentrations of analyte over the immobilised ligand surface, without intermittent regeneration.
All measurements used SPR buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM tris(2-carboxyethyl)phosphine, 300 mM NaCl, 0.005% (v/v) P20), with a ProteOn NLC (neutravidin-coated) sensor chip for immobilization of bio*τC16 and bio*τC24. All 36 interaction spots of the sensor chip were activated with three sequential injections of 1 M NaCl, 50 mM NaOH across six vertical (ligand) flow paths (40 s each at 40 μL min⁻¹) and six horizontal (analyte) flow paths (40 s each at 100 μL min⁻¹). The surface was further stabilised by two injections of 1 M MgCl₂ in each direction, with the same contact times and flow rates. Both bio*τC16 and bio*τC24 were diluted to 300 nM in SPR buffer and immobilised separately onto the six interaction spots of the vertical flow path (100 μL min⁻¹ for 15 s) in ligand channels 2 and 4, respectively. The sensorgrams verified that immobilised bio*τC16 and bio*τC24 remained stably attached to the surface. The chip was then rotated 90°, so flow then passes across all six horizontal (analyte) channels.

Interactions of α with either τC16 and τC24 were studied by injecting in the analyte direction using an appropriate concentration series in SPR buffer (zero and five concentrations of serially diluted samples; 4, 2, 1, 0.5, and 0.25 nM) at 45 μL min⁻¹ for 500 s, followed by dissociation in the same buffer for 6,000 s. The final sensorgrams were interspot and unmodified ligand flow path subtracted using ProteOn Manager™ Software v3.1.0.6. Binding affinity (dissociation constant, K_D) and binding kinetics (rate constants, k_a and k_d) parameters for the interaction of the α subunit with τC16 or τC24 were determined by global (simultaneous) fitting of five sensorgrams from the optimised concentration range using ProteOn Manager™ software and the appropriate interaction models; Langmuir one-to-one binding and Langmuir one-to-one binding with mass transfer models were used for fitting sensorgrams.

3.2.12 NMR spectroscopy and data processing

NMR experiments were carried out using 200 μM [¹⁵N]-τC24 (Section 3.2.6) in NMR buffer (10 mM sodium phosphate, pH 6.8, 1 mM dithiothreitol, 100 mM NaCl, 0.1 mM NaN₃) with 10% (v/v) D₂O. For the α–τC24, complex a 1.4 molar ratio of 40 μM [¹⁵N]-τC24 and 56 μM of unlabelled α was mixed in NMR buffer. NMR experiments were recorded at 298 K on a Bruker 800 MHz NMR spectrometer equipped with a ¹H/¹⁵N/¹³C-triple resonance cryoprobe using t₁,max = 32 ms, t₂,max = 158 ms and total recording time of 4 h. Data were processed using Topspin v3.5 (Bruker, Biospin) and Sparky3 v3.115 was used to overlay differently coloured contour plots of the NMR spectra recorded of the different samples.
3.2.13 Small angle X-ray scattering

Synchrotron SEC-SAXS data were collected at the SAXS/WAXS beamline of the Australian Synchrotron in 2013 and 2014. Proteins and their protein–protein complexes were measured in SAXS buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM tris(2-carboxyethyl)phosphine, 50 mM NaCl, 5% (v/v) glycerol) and 100 μL was injected at 0.2 mL min⁻¹ onto a Wyatt WTC-030S5 (7.8 × 300 mm) SEC column packed with 5 µm coated silica beads with a pore size of 300 Å and nominal separation range of 5–1250 kDa equilibrated at 12 °C with SAXS buffer. The A280 of the eluate was monitored immediately before its passage directly into the quartz measurement capillary (1.5 mm internal diameter) before passing into waste. The samples in the quartz capillary were illuminated by a collimated 11 keV (λ = 1.127 Å) X-ray beam with dimensions at the capillary surface of 0.25 × 0.15 mm (FWHM). X-ray scattering was recorded by a Pilatus 1 M area detector (DecTris, Switzerland) in 2 s frames from a position of 1600 mm behind the sample.

3.2.14 Small angle Neutron scattering

The reconstituted α−τC24 complex was loaded at 0.5 mL min⁻¹ onto a column (1.6 × 60 cm) of Superdex S200 resin (GE Healthcare, Sweden) pre-equilibrated with SANS buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM tris(2-carboxyethyl)phosphine, 50 mM NaCl, 5% (v/v) glycerol) to remove protein aggregates. Then the complex was concentrated by ultrafiltration using an MWCO 4,000 Ultrafree-4 centrifugal filter unit (Merk Millipore, USA) to 4.3 mg mL⁻¹ (1.2 mL) of α−τC24 complex prior to loading into either 1 or 2 mm quartz cuvettes for SANS measurements. SANS data were acquired using a complex of unlabelled τC24 and deuterium labelled α subunit (Section 3.2.9) in 0, 42, 60, 80 and 100% (v/v) D₂O solvent in SANS buffer containing either glycerol or glycerol-d₈ by using the SANS instrument QUOKKA [374] at ANSTO (Lucas Heights, Australia). SANS data were collected using a neutron wavelength λ = 5 Å. Neutron scattering was recorded on a 21000N Ordela ³He detector (Ordela, USA) at 6 °C. Two geometries were used to collect scattering data: 2.0 m sample-to-detector distance (2 h data collection time), and a 12.0 m sample-to-detector distance (4 h data collection time). The 2.0 m setting gives an accessible q range of 0.006–0.09 Å⁻¹, while the 12.0 m setting gives a range of 0.02–0.5 Å⁻¹.
3.3 Results

3.3.1 Biotinylation of bio\textsubscript{TC24} and bio\textsubscript{TC16}

Biotinylated proteins, bio\textsuperscript{*}\textsubscript{TC24} and bio\textsuperscript{*}\textsubscript{TC16}, were required for SPR measurements (Figure 3.6a). As previously described for bio\textsuperscript{*}\textsubscript{TC16} [17], complete resolution and purification of two species that originate from the same construct was achieved, with only one having biotin attached (Figure 3.6b). On the other hand, for the longer truncation mutant bio\textsubscript{TC24}, no \textit{in vivo} biotinylation was observed, even in the presence of a high concentration of D-biotin (100 μM; Figure 3.6c). To overcome this issue, \textit{in vitro} biotinylation was carried out with highly purified \textit{E. coli} biotin ligase and the reaction proceeded with ~100% efficiency as determined by ESI-MS (Figure 3.6d). Moreover, the presence of proteolysed products was not detected.

**Figure 3.6 | \textit{In vitro} biotinylation of bio\textsubscript{TC24} was successful.** a, SDS-PAGE analysis of purified bio\textsuperscript{*}\textsubscript{TC24} and bio\textsuperscript{*}\textsubscript{TC16} (Coomassie blue stained). b, The nano-ESI-mass spectrum of pooled fractions from peak β from bio\textsubscript{TC16} purification. The species has a measured molecular weight of 18,768.6 ± 0.1 Da; this agrees well with the theoretical mass of 18,769 Da. c, The nano-ESI spectrum of bio\textsubscript{TC24} purified from \textit{E. coli} cells did not contain any bio\textsuperscript{*}\textsubscript{TC24} with a measured molecular mass of 26,017.4 ± 1.7 Da; this is very close to the theoretical molecular mass of TC24 containing bio-tag only, 26,017.7 Da. d, The nano-ESI spectrum of bio\textsuperscript{*}\textsubscript{TC24} after \textit{in vitro} biotinylation and purification shows near complete biotinylation as determined by a measured molecular mass of 26,243.6 ± 0.2 Da, which is in excellent agreement with the theoretically calculated mass of 26,245.7 Da.
3.3.2 A direct assay for the $\alpha-\tau_{C24}$ interaction using SPR

The interaction between $\alpha$ and $\tau$ has been examined previously using a truncated domain mutant $\tau_{C16}$ [17]. In this study, bio*$\tau_{C16}$ was immobilised on a streptavidin coated SPR chip and serial dilutions of $\alpha$ were made to flow over the surface [17]. It was shown that the binding behaviour of $\alpha$ to $\tau_{C16}$ is a mass transfer limited interaction and very strong, with an apparent $K_D$ value of $264 \pm 1.60$ pM ($t_{1/2} = 29$ min at 300 mM NaCl) [17]. That work is confirmed and extended here by SPR using a ProteOn XPR36 instrument by examination of $\alpha$ binding to the $\tau_{C24}$. I hypothesised that domain IVa of $\tau$ may stabilise the interaction with the $\alpha$ subunit, as the structural fold of $\tau_{C16}$ places spatial restrictions on the N- and C-termini (Figure 3.3).

First bio*$\tau_{C16}$ (for direct comparison) and bio*$\tau_{C24}$ were immobilised on the neutravidin coated surface, to a density of 352 and 532 RU, respectively. Monomeric $\alpha$ binds monomeric $\tau_{C16}$ or $\tau_{C24}$ [17,359] and previously reported $k_a$ values were $>1 \times 10^6$ M$^{-1}$ s$^{-1}$, indicating the measured association rate may reflect the transfer of $\alpha$ to the dextran-matrix rather than the true association rate with $\tau$. Thus, sensorgrams were fit with both Langmuir one-to-one and Langmuir one-to-one with mass transfer models for comparison. Interaction kinetics were then calculated by simultaneously fitting of all five sensorgrams (Figure 3.7a–d); binding parameters determined are presented in Table 3.1.

To assess quality of fits to both Langmuir one-to-one binding and Langmuir one-to-one binding with mass transfer models, the $R_{\text{max}}$ was constrained to a globally constant value experimentally obtained by flowing $\alpha$ at saturating concentration (200 nM) over bio*$\tau_{C16}$ and bio*$\tau_{C24}$ (Figure 3.8a,b). The binding parameters derived from fitting with constrained $R_{\text{max}}$ were compared with experimentally derived values. The derived $K_D$ was within a factor of 1.5 of $K_D$ obtained from the kinetic parameters for both Langmuir one-to-one binding and Langmuir one-to-one binding with mass transfer models. In addition, the fit $R_{\text{max}}$ value of 1,298 RU compares well to the experimentally determined $R_{\text{max}}$ of 1,592 RU in the case of $\alpha-\tau_{C16}$, and 1,715 RU with experimentally determined $R_{\text{max}}$ of 2,105 RU for $\alpha-\tau_{C24}$ (Table 3.2). Moreover, the similarity in kinetic and thermodynamic values derived from both Langmuir one-to-one and Langmuir one-to-one binding with mass transfer models, including the fact that measured $k_a$ values were below $1 \times 10^6$ M$^{-1}$s$^{-1}$ implies higher flow rates used in these measurements significantly reduces the effects of mass transfer.
Figure 3.7 | SPR assessment of α–τC16 and α–τC24 interactions: τC24 binds more strongly to α than τC16. Sensorgrams show association and dissociation phases of α–τ interactions at ranges of α concentrations of serially-diluted samples of α. Curves, shown in gradient of colours were fit simultaneously (black curves) to various binding models. 

a α–τC16; the Langmuir one-to-one model was used to fit the data. 
b α–τC24, the Langmuir one-to-one model was used to fit the data. 
c α–τC16; the Langmuir one-to-one with mass transfer model was used to fit the data. 
d α–τC24; the Langmuir one-to-one with mass transfer model was used to fit the data. Derived parameters are shown in Table 3.1.

The calculated $K_D$ (α–τC16) value of 350.0 ± 0.8 pM is in good agreement with the previously reported value of 264 pM [17]. The slight discrepancy in $K_D$ values may be caused by differences in the flow rate and temperature as well as differences between instruments (BIAcore 2000 vs. ProteOn XPR36). Nevertheless, the data clearly demonstrate a strong interaction between τC24 and α. Specifically, the $k_d$ of α dissociating from τC24 is lower than that of τC16, whereas the $k_a$ values are similar, resulting in an overall 3–5-fold increase in binding affinity of τC24 compared to τC16. The change in binding affinity and slower dissociation rate may be a consequence of additional contributions of domain IVa of τ binding to α, beyond those previously detected with τC16.
Table 3.1 | Summary of the binding parameters for the $\alpha$–$\tau_{C16}$ and $\alpha$–$\tau_{C24}$ interactions determined by simultaneous fit of sensorgrams to Langmuir one-to-one or Langmuir one-to-one binding with mass transfer models. The Table displays determined variables: data collection parameters, equilibrium constant $K_D$, association rate constant $k_a$, dissociation rate constant $k_d$, mass transfer coefficient $k_t$, and saturation binding response $R_{\text{max}}$. All uncertainties are standard errors in parameters from fitting of complete data sets to appropriate binding models as indicated.

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>ProteOn XPR36</td>
<td></td>
</tr>
<tr>
<td>Acquisition rate (Hz)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Sensor chip</td>
<td>Neutravidin-coated NLC chip</td>
<td></td>
</tr>
<tr>
<td>Flow rate ($\mu$L min$^{-1}$)</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>Association time (s)</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td>Dissociation time (s)</td>
<td>6,000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Langmuir one-to-one</th>
<th>$\alpha$–$\tau_{C16}$</th>
<th>$\alpha$–$\tau_{C24}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D$ (pM)</td>
<td>$350.0 \pm 0.8$</td>
<td>$96.4 \pm 0.2$</td>
</tr>
<tr>
<td>$k_a$ (M$^{-1}$ s$^{-1}$)</td>
<td>$4.98 \pm 0.00 \times 10^5$</td>
<td>$5.76 \pm 0.00 \times 10^5$</td>
</tr>
<tr>
<td>$k_d$ (s$^{-1}$)</td>
<td>$17.5 \pm 0.0 \times 10^{-5}$</td>
<td>$5.55 \pm 0.00 \times 10^{-5}$</td>
</tr>
<tr>
<td>$k_t$ (RU M$^{-1}$ s$^{-1}$)</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (RU)</td>
<td>$1,298 \pm 1$</td>
<td>$1,715 \pm 1$</td>
</tr>
<tr>
<td>Stoichiometry</td>
<td>0.46</td>
<td>0.60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Langmuir one-to-one with mass transport</th>
<th>$\alpha$–$\tau_{C16}$</th>
<th>$\alpha$–$\tau_{C24}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D$ (pM)</td>
<td>$313 \pm 3$</td>
<td>$57.8 \pm 0.3$</td>
</tr>
<tr>
<td>$k_a$ (M$^{-1}$ s$^{-1}$)</td>
<td>$(7.38 \pm 0.05) \times 10^5$</td>
<td>$(15.5 \pm 0.5) \times 10^5$</td>
</tr>
<tr>
<td>$k_d$ (s$^{-1}$)</td>
<td>$(23.1 \pm 0.1) \times 10^{-5}$</td>
<td>$(8.99 \pm 0.07) \times 10^{-5}$</td>
</tr>
<tr>
<td>$k_t$ (RU M$^{-1}$ s$^{-1}$)</td>
<td>$(1.84 \pm 0.00) \times 10^9$</td>
<td>$(1.44 \pm 0.00) \times 10^9$</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (RU)</td>
<td>$1,232 \pm 2$</td>
<td>$1,426 \pm 1$</td>
</tr>
<tr>
<td>Stoichiometry</td>
<td>0.43</td>
<td>0.50</td>
</tr>
</tbody>
</table>

Figure 3.8 | Sensorgrams showing experimentally determined $R_{\text{max}}$ values for bio$^*\tau_{C16}$ and bio$^*\tau_{C24}$ at saturating $\alpha$ concentrations. Saturating concentrations of $\alpha$ (200 nM) were passed over multiple channels with immobilised bio$^*\tau_{C16}$ and bio$^*\tau_{C24}$. a The average response at saturation was measured to be 1,592 $RU$ for bio$^*\tau_{C16}$. b The average response at saturation was measured to be 2,105 $RU$ for bio$^*\tau_{C24}$. These $R_{\text{max}}$ values were used as a global constant value for fitting of data in Figure 3.7. Binding parameters calculated using then constant $R_{\text{max}}$ values are shown in Table 3.2.
Table 3.2 Summary of the binding parameters for the $\alpha$–TC$_{16}$ and $\alpha$–TC$_{24}$ interactions determined by simultaneous fit of sensorgrams with globally constrained $R_{\text{max}}$ values.

The table displays determined variables: equilibrium constant $K_D$, association rate constant $k_a$, dissociation rate constant $k_d$, mass transfer coefficient $k_t$, and saturation binding response $R_{\text{max}}$.

All uncertainties are standard errors in parameters from fitting of complete data sets to appropriate binding models as indicated.

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$–TC$_{16}$</th>
<th>$\alpha$–TC$_{24}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D$ (pM)</td>
<td>$477.0 \pm 0.2$</td>
<td>$141.0 \pm 0.4$</td>
</tr>
<tr>
<td>$k_a$ (M$^{-1}$ s$^{-1}$)</td>
<td>$(4.96 \pm 0.00) \times 10^5$</td>
<td>$(5.71 \pm 0.00) \times 10^5$</td>
</tr>
<tr>
<td>$k_d$ (s$^{-1}$)</td>
<td>$(17.6 \pm 0.0) \times 10^{-5}$</td>
<td>$(5.57 \pm 0.01) \times 10^{-5}$</td>
</tr>
<tr>
<td>$k_t$ (RU M$^{-1}$ s$^{-1}$)</td>
<td>$1,592$</td>
<td>$2,105$</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (RU)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stoichiometry</td>
<td>$0.55$</td>
<td>$0.73$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$–TC$_{16}$</th>
<th>$\alpha$–TC$_{24}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D$ (pM)</td>
<td>$320.0 \pm 0.7$</td>
<td>$57.6 \pm 0.4$</td>
</tr>
<tr>
<td>$k_a$ (M$^{-1}$ s$^{-1}$)</td>
<td>$(7.18 \pm 0.00) \times 10^5$</td>
<td>$(1.57 \pm 0.00) \times 10^6$</td>
</tr>
<tr>
<td>$k_d$ (s$^{-1}$)</td>
<td>$(23.0 \pm 0.0) \times 10^{-5}$</td>
<td>$(9.03 \pm 0.01) \times 10^{-5}$</td>
</tr>
<tr>
<td>$k_t$ (RU M$^{-1}$ s$^{-1}$)</td>
<td>$(2.94 \pm 0.00) \times 10^{22}$</td>
<td>$(4.74 \pm 0.00) \times 10^{22}$</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (RU)</td>
<td>$1,592$</td>
<td>$2,105$</td>
</tr>
<tr>
<td>Stoichiometry</td>
<td>$0.55$</td>
<td>$0.73$</td>
</tr>
</tbody>
</table>

3.3.3 An examination of the $\alpha$–TC$_{24}$ interaction using NMR spectroscopy

$^1$H$^{15}$N-HSQC NMR spectroscopy was used to probe the degree of perturbation within domain IVa of $\tau$ upon interaction with $\alpha$. The patterns of resonance frequencies of individual amides within a protein are unique, and thus can be used as a fingerprint of its structure. Chemical shifts in amide $^{15}$NH resonances are particularly sensitive to changes in the local environment. Perturbation of these shifts provides a highly sensitive tool for mapping interactions and binding surfaces of protein complexes. Therefore, monitoring the associated spectral changes in the $^1$H$^{15}$N-HSQC NMR spectra can identify residues involved in binding on a $^{15}$N-labelled protein on addition of an unlabelled partner protein. By monitoring the largest shifts/perturbations of cross-peaks upon complex formation (i.e., resonances that do not exhibit any changes arise from those regions that are not within this region), identification of the primary binding region within a complex is possible. To exploit this effect, a sample of uniformly $^{15}$N-labelled $\tau_{C24}$ was produced (Section 3.2.6; Figure 3.9a) and mixed with unlabelled $\alpha$ (Section 3.2.8; Figure 3.9b). Spectral changes were monitored in the $^1$H$^{15}$N-HSQC spectrum of [${}^{15}$N]$\tau_{C24}$ in the presence and absence of $\alpha$ to identify individual residues unique to domain IVa of $\tau$ that may contribute to the slower dissociation rate observed in SPR measurements.
The two-dimensional $^1$H$^{15}$N-HSQC NMR spectrum of $[^{15}$N]-τC24 (Figure 3.10a) was first compared to that of combinatorially labelled τC16 [361] by manual subtraction of cross-peaks common to both τC24 and τC16, leaving only cross-peaks unique to domain IVa. Comparison of τC16 with τC24 spectra shows that all amide peaks from combinatorially labelled τC16 were present in the $[^{15}$N]-τC24 spectrum with unchanged chemical shifts. Moreover, cross-peaks of many residues in domain IVa were located in the narrow region between 7.5 and 8.5 ppm in the proton dimension (Figure 3.10a,b). A lack of distribution in this region of the spectrum is characteristic of either α-helical or random coil conformations. This observation is consistent with previous NMR measurements of domain IVa in $^{15}$N-labelled τC22, a mutant of τC24 that lacks the 18 C-terminal residues from domain V (Figure 3.1; [39]). To identify any perturbations of flexible residues in domain IVa upon interaction with α, a $^{15}$N-HSQC spectrum was recorded of $[^{15}$N]-τC24 after addition of 1.5 equivalents of unlabelled purified α (Figure 3.10c). The global intensity of the signal decreased and most residues that belong exclusively to domain V within $[^{15}$N]-τC24 were broadened beyond detection. This observation is consistent with flexible C-terminal residues in domain V directly binding to α, especially the last 18 residues. Several cross-peaks unique to domain IVa became perturbed in intensity. Moreover, a few peptide and side chain amides became broadened beyond detection. As the cross-peaks in domain IVa have not yet been assigned to specific amino acids in the sequence, they could not readily be assigned to an amino acid in the sequence of τC24.
Figure 3.10 | $^1$H$^{15}$N-HSQC NMR assessment of $\alpha$-tc24 interaction using uniformly $^{15}$N-labelled $\tau$C24. a, The $^1$H$^{15}$N-HSQC NMR spectrum of $^{15}$N-labelled $\tau$C24. Clustering of cross-peaks in the narrow region between 7.5 and 8.5 ppm is characteristic of either $\alpha$-helical or random coil conformations, consistent with previous NMR measurements of domain IVa [39]. b, Superposition of the $^{15}$N-HSQC spectrum of combinatorially labelled $\tau$C16 (red) from [361] and the $^{15}$N-HSQC NMR spectrum of $\tau$C24 showing cross-peaks unique to domain IVa within $\tau$C24 (black). Unique cross-peaks are obtained by subtraction of cross-peaks common to $\tau$C16 and $\tau$C24. c, Superposition of $^{15}$N-labelled $\tau$C24 spectrum from b and $\alpha$+C24 (light blue). Addition of $\alpha$ results in global signal attenuation. Several cross-peaks show a high degree of attenuation, and in a few cases cross-peaks have been broadened beyond detection. Black arrows identify cross-peaks of interest.

Manual assignment based on typical chemical shifts for particular amides in $\alpha$-helices or random coils [375] suggests residues Trp478-Lys479-Ala480 in domain IVa are likely candidates. We cannot eliminate, however, the possibility that observed
perturbations of flexible residues in domain IVa are residues immediately preceding domain V. More in-depth NMR studies utilising selective combinatorially-labelled $\tau_{24}$ will be needed to clarify this. This would allow the assignment of unknown residues within domain IVa, a strategy previously utilised with great success for assignment of $\tau_{16}$ [361]. Therefore, we conclude that domain IVa remains mostly unperturbed and flexible in complex with $\alpha$.

3.3.4 Structural characterisation of $\alpha$, $\alpha-\tau_{16}$, and $\alpha-\tau_{24}$ by SAXS and SANS

To determine the overall size, shape, and build a structural model of the $\alpha$ subunit and reconstituted $\alpha-\tau_{16}$, $\alpha-\tau_{24}$ complexes (Figure 3.11), structural analysis by SEC-SAXS and SANS with contrast variation was undertaken. Accurate SEC-SAXS and SANS measurements were conducted on homogenous and monodisperse particles reconstituted from purified subunits and isolated by chromatography (Section 3.2.10). SEC-SAXS was utilised as it can effectively eliminate polydispersity arising from partial disassembly of non-covalent protein complexes between purification and measurement, enabling measurement of pure and intact macromolecular complexes in solution.

**Figure 3.11 | In vitro reconstitution of the $\alpha-\tau_{16}$ and $\alpha-\tau_{24}$ complexes from individual purified subunits.**

**a**, SDS-PAGE analysis of purified $\alpha-\tau_{16}$ and $\alpha-\tau_{24}$ complexes (Coomassie blue stained). **b**, SDS-PAGE analysis of purified deuterium–labelled $\alpha$ complexed with unlabelled $\tau_{24}$ (Coomassie blue stained).
3.3.4.1 The α subunit

The α subunit eluted as a single peak in SEC-SAXS (Figure 3.12a); selected frames were averaged across a range of high intensity scattering, and scattering of the buffer from a region prior to protein elution was subtracted to generate a 1D scattering curve (Figure 3.12b). The calculated radius of gyration ($R_g$) was constant across the $A_{280}$ peak (Figure 3.12a). The Guinier analysis shows little evidence of aggregation or inter-particle repulsion, as it does not deviate significantly from linearity (Figure 3.12c). The $R_g$ was calculated to be 38.3 ± 0.1 Å. The pair-distance distribution function, ($p(r)$), describes the distribution of interatomic distances in the scattered particle, i.e., the effective size of the scattering particle. The $p(r)$ curve shows α to be mostly a globular protein with a maximum dimension ($D_{max}$) of 125 ± 6 Å (Figure 3.12d). A summary of structural parameters determined from SEC-SAXS for the α subunit is given in Table 3.3.

Table 3.3 | Parameters extrapolated from the experimental SEC-SAXS data for the α subunit.

<table>
<thead>
<tr>
<th>Data collection parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
</tr>
<tr>
<td>Beam geometry</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
</tr>
<tr>
<td>q-range (Å⁻¹)</td>
</tr>
<tr>
<td>Exposure time (s)</td>
</tr>
<tr>
<td>Protein concentration (mg mL⁻¹)</td>
</tr>
<tr>
<td>Temperature (°C)</td>
</tr>
<tr>
<td>Standard</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structural parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>$I(0)$ (cm⁻¹) [from $p(r)$]</td>
</tr>
<tr>
<td>$R_g$ (Å) [from $p(r)$]</td>
</tr>
<tr>
<td>$I(0)$ (cm⁻¹) [from Guinier]</td>
</tr>
<tr>
<td>$R_g$ (Å) [from Guinier]</td>
</tr>
<tr>
<td>$D_{max}$ (Å)</td>
</tr>
<tr>
<td>Porod volume (Å³)</td>
</tr>
<tr>
<td>Dry volume (Å³) [from sequence]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecular mass determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial specific volume (cm³ g⁻¹)⁵</td>
</tr>
<tr>
<td>Contrast, $\Delta \rho$ (10¹⁰ cm⁻²)</td>
</tr>
<tr>
<td>Molecular mass $M_r$ [Fischer method]</td>
</tr>
<tr>
<td>Molecular mass $M_r$ [from sequence]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Software employed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary data reduction</td>
</tr>
<tr>
<td>Data processing</td>
</tr>
<tr>
<td>Ab initio modelling</td>
</tr>
<tr>
<td>Rigid body modelling</td>
</tr>
<tr>
<td>Three-dimensional graphics</td>
</tr>
</tbody>
</table>

⁵ Calculated from amino acid composition using MULCh [376].
Figure 3.12 | SEC-SAXS data for α. a, The elution profile of α. The protein eluted in a single peak, and exposures were averaged across the region of high scattering intensity and stable radius of gyration (green line). Buffer scattering was averaged over a region before protein elution. \( R_g \) and \( I(0) \) values shown here were calculated from individual, buffer-subtracted 2 s exposures. b, The experimental SAXS scattering profile of α. Each dot represents an intensity value for a given scattering angle. c, Guinier analysis of α showing a linear fit. d, Pair-distance distribution function \( p(r) \), derived in GNOM [377], from the scattering data (\( q \)-range 0.0110–0.299 Å\(^{-1}\)). The distribution is characteristic of a globular protein. All error bars represent the standard error of the mean intensity at that point.
3.3.4.2 The $\alpha-\tau_{C16}$ and $\alpha-\tau_{C24}$ complexes

Complexes of $\alpha$ bound to either $\tau_{C16}$ or $\tau_{C24}$ eluted as a single peak and selected frames were averaged across a range of high intensity scattering with stable $R_g$ (Figure 3.13a,b). Scattering of the buffer from a region prior to protein elution was subtracted to generate 1D scattering curves of each complex (Figure 3.13c). Guinier analysis was performed to verify that both $\alpha-\tau_{C16}$ and $\alpha-\tau_{C24}$ complexes were monodisperse during SEC-SAXS measurements.

The Guinier plots were linear at low $q$ and there is no evidence of aggregation or inter-particle effects for either complex (Figure 3.13d). The size of each complex was derived from the $p(r)$. Both $\alpha-\tau_{C16}$ and $\alpha-\tau_{C24}$ are slightly larger than $\alpha$ alone, reflected in the calculated $R_g$. The $p(r)$ analysis confirms the $\alpha-\tau_{C24}$ complex adopts a larger, more elongated average conformation than the $\alpha-\tau_{C16}$ complex in solution. This is consistent with a rather broad distribution of interatomic distances observed in the $p(r)$ (Figure 3.13e). The $\alpha-\tau_{C24}$ complex has a calculated $R_g$ of 50.0 $\pm$ 0.1 Å and $D_{max} = 160 \pm 8$ Å, larger than the calculated $R_g$ of $\alpha-\tau_{C16}$ being 43.3 $\pm$ 0.1 Å and $D_{max}$ of 150 $\pm$ 8 Å. Moreover, the $p(r)$ of $\alpha-\tau_{C24}$ has an asymmetric shoulder typical of a particle containing an elongated component, centred at $r \sim 50$ Å, followed by a progressive decrease down to zero at the $D_{max}$ value. These $D_{max}$ values infer a more compact structure than anticipated. A summary of structural parameters determined from SEC-SAXS for the $\alpha-\tau_{C16}$ and $\alpha-\tau_{C24}$ complexes is given in Table 3.4.
Figure 3.13 | SEC-SAXS data for α−TC16 and α−TC24 complexes. 

a and b. The α−TC16 and α−TC24 complexes eluted from SEC in a single peak, and exposures were averaged across the region of high scattering intensity and stable radius of gyration (green line). $R_g$ and $I(0)$ values shown are calculated from individual, buffer-subtracted 2 s exposures.

c. The experimental SAXS scattering profile for α−TC16 (red) and α−TC24 (black), where the α−TC24 data are offset by a factor of 50 for clarity.

d. Guinier analysis of α−TC16 (red dots) and α−TC24 (black dots) showing linear fits.

e. Pair-distance distribution function $p(r)$, derived for α−TC16 (red) and α−TC24 (black) in GNOM [377] from the scattering data ($q$-range 0.0110–0.299 Å$^{-1}$). The distribution of α−TC16 is characteristic of a globular particle whereas α−TC24 is characteristic of a more elongated particle. All error bars represent the standard error of the mean at that point.
Table 3.4 | Parameters extrapolated from the experimental SEC-SAXS data for both \( \alpha \rightarrow \tau_{C16} \) and \( \alpha \rightarrow \tau_{C24} \) complexes.

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th>( \alpha \rightarrow \tau_{C16} )</th>
<th>( \alpha \rightarrow \tau_{C24} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>SAXS-WAXS (Australian Synchrotron)</td>
<td></td>
</tr>
<tr>
<td>Beam geometry</td>
<td>Point</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.2713</td>
<td></td>
</tr>
<tr>
<td>q-range (Å(^{-1}))</td>
<td>0.01–0.50</td>
<td></td>
</tr>
<tr>
<td>Exposure time (s)</td>
<td>( \alpha \rightarrow \tau_{C16} ) (82 [41 × 2 s]) ( \alpha \rightarrow \tau_{C24} ) (26 [13 × 2 s])</td>
<td></td>
</tr>
<tr>
<td>Protein concentration (mg mL(^{-1}))</td>
<td>2.3 (loaded onto column)</td>
<td></td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Standard</td>
<td>Water</td>
<td></td>
</tr>
</tbody>
</table>

Structural parameters

| \( I(0) \) (cm\(^{-1}\)) [from \( p(r) \)] | \( 0.01277 \pm 0.00002 \) | \( 0.02782 \pm 0.00005 \) |
|\( R_g \) (Å) [from \( p(r) \)] | 43.3 ± 0.1 | 50.0 ± 0.1 |
|\( I(0) \) (cm\(^{-1}\)) [from Guinier] | \( 0.01270 \pm 0.00002 \) | \( 0.02797 \pm 0.00001 \) |
|\( R_g \) (Å) [from Guinier] | 42.8 ± 0.1 | 44.3 ± 0.3 |
|\( D_{\text{max}} \) (Å) | 150 ± 8 | 160 ± 8 |
| Porod volume (Å\(^3\)) | 252,000 ± 25,000 | 219,000 ± 22,000 |
|\( R_g \) (Å) [PDB: 5FKU] | 39.6 | — |
|\( D_{\text{max}} \) (Å) [PDB: 5FKU] | 135 | — |
| Dry volume (Å\(^3\)) [from sequence] | 179,000 | 189,000 |

Molecular mass determination

| Partial specific volume (cm\(^3\) g\(^{-1}\))\(^a\) | 0.739 | 0.739 |
| Contrast, \( \Delta \rho \) (10\(^{10}\) cm\(^{-2}\)) | 2.72 | 2.72 |
| Molecular mass \( M_r \) [Fischer method] | \( 210,000 \pm 21,000 \)\(^b\) | \( 179,000 \pm 18,000 \) |
| Molecular mass \( M_r \) [from sequence] | 146,000 | 154,000 |

Software employed

| Primary data reduction | ScatterBrain |
| Data processing       | PRIMUS and GNOM |
| Rigid body modelling  | SASREF |
| Three-dimensional graphics | PyMOL v1.8.0.5 |

\(^a\) Calculated from amino acid composition using MuLCh [376].

\(^b\) Analysis of \( M_r \) by Fischer methods tends to break down for flexible proteins, especially in cases where there are density changes, i.e., degrees of foldedness.

3.3.4.3 SANS with contrast variation of the \( \alpha \rightarrow \tau_{C24} \) complex

SEC-SAXS alone was not able to localise \( \tau_{C24} \) within the \( \alpha \rightarrow \tau_{C24} \) complex, a consequence of the large size of \( \alpha \) approximately five times larger than \( \tau_{C24} \). To overcome this, SANS with contrast variation was performed. The \( \alpha \) protein was labelled with deuterium (Section 3.2.9), to a level such that approximately 75% of the non-exchangeable hydrogen positions were occupied by deuterium. The deuterium labelled \( \alpha \) was complexed with unlabelled \( \tau_{C24} \) (Figure 3.11). Differences in scattering length densities between \(^1\)H and \(^2\)H mean that the labelled and unlabelled components of the complex scatter neutrons differently. Further, altering the \(^1\)H and \(^2\)H composition in the solvent modulates the extent each subunit scatters, to the point where it is possible to match the scattering length density of one subunit with the solvent (contrast matching), rendering it essentially invisible; thus, the resulting scattering is due to the
visible component alone. Here, the unlabelled \( \tau_{C24} \) was contrast matched at \(~42\%\) D\(_2\)O; hence, the scattering observed at this data point is due to the \( \alpha \) subunit, while the deuterium labelled \( \alpha \) subunit was matched at \(~100\%\) D\(_2\)O. Five neutron scattering contrast points were measured: 0, 42, 60, 80 and 100\% D\(_2\)O, where the additional contrast points yield information on the arrangement of each subunit in the complex. SANS data were merged from two different \( q \) ranges (0.006–0.09 and 0.02–0.5 Å\(^{-1}\)) and reduced with corrections made for solvent scattering, sample transmission, detector sensitivity and background radiation to yield 1D scattering curves (Figure 3.14a). The SANS data were of high quality, evident by linearity at low \( q \) values in the Guinier analysis at each neutron scattering contrast point (Figure 3.14b). SANS was able to localise the \( \alpha \) subunit from \( \tau_{C24} \) in the \( \alpha-\tau_{C24} \) complex from the measured contrast points, evident by the shifting of \( p(r) \) profiles from a distribution characteristic of a globular particle to of an elongated particle as the concentration of D\(_2\)O was increased. This is most intuitive when comparing 42\% and 100\% D\(_2\)O neutron scattering contrast points (Figure 3.14c). This change is also reflected in the calculated \( R_g \) at 42\% D\(_2\)O being 42.8 \( \pm \) 0.3 Å, whereas at 100\% D\(_2\)O the \( R_g = 36.4 \pm 1.4 \) Å. The maximum dimension of the particle (\( D_{\text{max}} \)) was consistent throughout all neutron scattering contrast points between 140 and 150 Å. The structural parameters derived from SANS measurements are summarised in Table 3.5.
Figure 3.14 | SANS with contrast variation can resolve $\alpha$ from $\tau_{24}$ in the $\alpha-\tau_{24}$ complex.

a, Neutron contrast variation scattering data at 0% D$_2$O (red), 42% D$_2$O (blue), 60% D$_2$O (green), 80% D$_2$O (purple), and 100% D$_2$O (orange) for $\alpha-\tau_{24}$. Data are shown on an absolute scale, where the 0–100% D$_2$O scattering data have been off-set by factors of $10^{-3}$, $10^{-5}$, $10^{-8}$, $10^{-11}$, and $10^{-14}$ for clarity.

b, Data over all neutron scattering contrast point are of high quality with linear Guinier regions.

c, Normalised pair-distance distribution function, $p(r)$, derived from the scattering data using GNOM [377] ($q$-range 0.0103–0.299 Å$^{-1}$) shows the $\alpha-\tau_{24}$ is characteristic of a more elongated particle – this characteristic is also observed by SEC-SAXS. $\tau_{24}$ is attributable to the extended conformation in the $\alpha-\tau_{24}$ complex shown by the $p(r)$ function shape at 100% D$_2$O (orange curve), because scattering at this neutron-scattering contrast point is dominated by unlabelled $\tau_{24}$. Other neutron scattering contrast points have been omitted for clarity. All error bars represent the standard error of the mean at that point.
Table 3.5 | Parameters extrapolated from the experimental SANS data for the \( \alpha-\text{TC24} \) complex.

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th>QUOKKA (ANSTO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>QUOKKA (ANSTO)</td>
</tr>
<tr>
<td>Beam geometry</td>
<td>Point</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>5.0</td>
</tr>
<tr>
<td>( q )-range (Å (^{-1}))</td>
<td>0.006–0.5</td>
</tr>
<tr>
<td>Protein concentration (mg mL(^{-1}))</td>
<td>3.8–4.3</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>6</td>
</tr>
<tr>
<td>Standard</td>
<td>Data placed on absolute scale by scaling against direct beam intensity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Structural parameters</th>
<th>0%</th>
<th>42%</th>
<th>60%</th>
<th>80%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( I(0) ) (cm(^{-1})) [from ( p(r) )]</td>
<td>1.604 ± 0.006</td>
<td>0.466 ± 0.002</td>
<td>0.2154 ± 0.0013</td>
<td>0.0296 ± 0.0007</td>
<td>0.0193 ± 0.0004</td>
</tr>
<tr>
<td>( R_g ) (Å) [from ( p(r) )]</td>
<td>44.3 ± 0.2</td>
<td>42.8 ± 0.3</td>
<td>41.4 ± 0.4</td>
<td>15 ± 5</td>
<td>36.4 ± 1.4</td>
</tr>
<tr>
<td>( I(0) ) (cm(^{-1})) [from Guinier]</td>
<td>1.594 ± 0.007</td>
<td>0.462 ± 0.003</td>
<td>0.2110 ± 0.0016</td>
<td>0.0311 ± 0.0006</td>
<td>0.0200 ± 0.0004</td>
</tr>
<tr>
<td>( R_g ) (Å) [from Guinier]</td>
<td>43.4 ± 0.3</td>
<td>42.1 ± 0.4</td>
<td>39.3 ± 0.5</td>
<td>26.1 ± 0.8</td>
<td>36.3 ± 1.3</td>
</tr>
<tr>
<td>( D_{\text{max}} ) (Å)</td>
<td>150 ± 8</td>
<td>150 ± 8</td>
<td>150 ± 8</td>
<td>150 ± 8</td>
<td>140 ± 8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecular mass determination</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Partial specific volume (cm(^{3})g(^{-1})) (^{a})</td>
<td>0.739</td>
<td>0.739</td>
<td>0.739</td>
<td>0.739</td>
<td>0.739</td>
</tr>
<tr>
<td>Contrast, ( \Delta \rho ) (10(^{10}) cm(^{-2}))</td>
<td>5.15</td>
<td>2.77</td>
<td>1.75</td>
<td>0.62</td>
<td>–0.52</td>
</tr>
<tr>
<td>Molecular mass ( M_r ) [Fischer method]</td>
<td>180,729</td>
<td>173,212</td>
<td>190,582</td>
<td>206,630</td>
<td>185,744</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Software employed</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary data reduction</td>
<td>ANSTO data reduction macros in ( \text{Igor} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data processing</td>
<td>( \text{PRIMUS} ) and ( \text{GNOM} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( ab \text{ initio} ) modelling</td>
<td>( \text{MONSA} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rigid body modelling</td>
<td>( \text{SASREF} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Three-dimensional graphics</td>
<td>( \text{PyMOL v1.8.0.5} )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{a}\) Calculated from amino acid composition using MuLCh [376].
3.3.4.4 3D modelling of the $\alpha$ subunit from SEC-SAXS

To date there is no full-length high-resolution structure of the $\alpha$ subunit of Pol III core to directly compare the experimental SEC-SAXS data with. However, a cryo-EM structure of the $\alpha\varepsilon-\beta_2-\tau_{C16}$ complex in the DNA-bound and DNA-free form has been recently published [222]. Therefore, rigid-body models of $\alpha$ were generated and used to compare $\alpha$ in DNA-bound and DNA-free states with experimental SEC-SAXS data. First, CRYSOL [378] was used to calculate the theoretical scattering profile of $\alpha$ alone in both the DNA-free and DNA-bound states from cryo-EM structures (PDB 5FKV and 5FKU). Subsequent fits of these models to the experimental scattering curve are shown in Figure 3.15a. The DNA-bound state of $\alpha$ is not a good fit to the experimental data because it begins to systematically deviate at $q \sim 0.1$ Å$^{-1}$. In fact, it is the most dissimilar to all models to the SEC-SAXS data with discrepancy $\chi^2 = 26.5$. On the other hand, the DNA-free state is a closer fit to the experimental SEC-SAXS data; however, it is still a poor fit with systematic deviations evident at $q \sim 0.1$ Å$^{-1}$ and discrepancy $\chi^2 = 9.2$.

The low-resolution shape of $\alpha$ was reconstructed by DAMMIN, a program used to compute ab initio dummy atom (bead) models from SAXS data [379]. DAMMIN was used to generate 15 independent models of $\alpha$, and the best DAMMIN model neatly fits the experimental data with discrepancy $\chi^2 = 1.4$. DAMAVER [380] was used to align models and build an average model. The averaged model (Figure 3.15b) converges on the shape reconstructed from DAMMIN. The shape is globular and agrees well with that of the crystal structure of $\alpha_{1–917}$ (PDB 2HQA). Further, rigid-body modelling of $\alpha$ was performed by implementing CORAL, a program that executes SAXS-based rigid-body modelling of macromolecular complexes, whose components lack some fragments (e.g. missing termini or unresolved inter-domain linkers). First, an N-terminal fragment was constructed using homology modelling of two existing X-ray crystal structures, $\alpha_{1–917}$ from E. coli [19] and residues 918–956 from the closely related T. aquaticus $\alpha$ structure [208]. Second, a C-terminal fragment comprised of residues 967–1160 from the unpublished X-ray crystal structure of $\alpha$CTD described in Section 3.1.2 was used. Finally, these two fragments was then joined by a flexible 10 residue ab initio bead linker (residues 957–966).
The best CORAL model of $\alpha$ is in good agreement with experimental scattering data having a discrepancy $\chi^2 = 1.9$ (Figure 3.15c). The superposition of the best CORAL model on the \textit{ab initio} envelopes demonstrates the good agreement between the \textit{ab initio} and rigid-body models. Inspection of the best CORAL $\alpha$ model suggests the C-terminal domain residues are positioned in an unexpected orientation. The OB fold occupies the region where dsDNA would typically be extruded during elongation (Figure 3.15d). Further, the TBD domain that interacts with $\tau$ is positioned near the internal CBM, such that if $\tau$ was bound it would block binding to the $\beta_2$ sliding clamp. This model may suggest there are at least two distinct conformational states of $\alpha$, one in the presence of DNA and the other in the absence of DNA and/or incoming nucleotide. Moreover, not only could DNA be responsible for changes to the conformational state of $\alpha$, but also the binding of the $\tau$ subunit (described in more detail in Section 3.3.4.4 and 3.3.4.5). Comparision by structural alignment of the best CORAL model of $\alpha$ and $\alpha$ alone from both DNA-free (5FKV) and DNA-bound (5FKU) cryoEM structures, show the global N-terminal structure is maintained between all structures (Figure 3.15d). Deviations in the alignment are apparent after the fingers domains. The OB fold and TBD are the most dissimilar among the models in their spatial arrangement. The large similarity in the large N-terminal region between the structures reflects the relatively small RMSD values of 2.8 Å and 3.2 Å of DNA-free and DNA-bound $\alpha$ structures, when compared against the CORAL model. Future studies will need to be done in the presence of a suitable DNA template, incoming nucleotide and in the absence/presence of $\tau$ to capture these different conformational states of $\alpha$ and understand their functional relevance.
Figure 3.15 | *Ab initio* envelopes and rigid-body model of the α subunit from SEC-SAXS. 

a, The experimental SAXS scattering profile for α is shown in black. Computed models and CRYSOL fits to the data are shown by coloured lines. b, Two orthogonal views of the models generated for α. Fifteen DAMMIN calculations [379] were performed; a typical fit is presented as the purple envelope (top and bottom panels), $\chi^2 = 1.4$, and averaged with DAMAVER [380] to produce the averaged and filtered shape shown as the grey envelope with mean normalised spatial discrepancy value = 0.690 ± 0.023 (top and bottom panels). The best rigid-body model produced by CORAL (as judged by the $\chi^2$) has been superposed (red cartoon) on the *ab initio* envelopes with SUPCOMB [381] and its fit to the experimental scattering data is in good agreement (red line in a; $\chi^2 = 1.9$). c, Larger view of best CORAL model from b with rigid bodies used in modelling labelled. d, Structural alignment of various structural models of α compared in SAXS scattering profiles in a. The CORAL model of α in b is shown in red, the cryoEM DNA-bound α is shown in green and cryoEM DNA-free α is shown in blue. The approximate path of DNA (orange cartoon) exiting the active site of α is shown. The N-terminal regions of α are very similar across all structures, deviating only after the fingers domain and being the most dissimilar in the C-terminal domains.
3.3.4.5 3D modelling of the α−τC16 complex from SEC-SAXS

A model of the α−τC16 complex in a “closed” conformation as a representative of the complex in the “polymerisation” mode from a previously published model was used as a starting template [15]. However, it was further refined by incorporating the αCTD−τC16 crystal structure described in Section 3.1.2 (Dr Zhi-Qiang Xu et al., unpublished). This new representative α−τC16 complex in a “closed” conformation was allowed to relax by performing an energy minimization step using classical molecular dynamics simulations (MD; denoted as refined model in figures). This minimization step ensures the system has no steric clashes or inappropriate geometries. Subsequently this model was compared with the experimental SEC-SAXS. CRYSOL was used to calculate the theoretical scattering profile of this α−τC16 model. The fit of the theoretical scattering of the model to the experimental scattering is shown in Figure 3.16a and is in good agreement with a discrepancy χ² = 3.3.

Next, the theoretical scattering profiles of models of α−τC16 complexes refined in real-space in the DNA-free (PDB 5FKU) and DNA-bound states (PDB 5FKV), where the αCTD and τC16 regions from published PDB coordinates were replaced with the X-ray crystal structure of αCTD−τC16 (Dr Zhi-Qiang Xu et al., unpublished) and the DNA removed were calculated using CRYSOL. Their fits to the experimental data are shown in Figure 3.16a. The DNA-bound complex was not a good fit to the experimental data with systematic deviations from experimental data in the region q = 0.1–0.2 Å⁻¹ and a discrepancy χ² = 25.1. On the other hand, this large systematic deviation at q = 0.1 Å⁻¹ is absent in the DNA-free complex. However, the fitting begins to deviate from the experimental scattering at q = 0.15 Å⁻¹. The DNA-free complex fits the experimental data better than the DNA-bound complex with a discrepancy χ² = 9.2. Based on these fits, the α−τC16 complex in a “closed” conformation is the most representative model of the SEC-SAXS data (Figure 3.16b). A structural alignment of the α−τC16 models in DNA-free and in “polymerisation” states described above shows a reasonable agreement in spatial positioning of the αCTD and bound τC16 molecule (Figure 3.16c,d). In the DNA-bound state, however, the OB fold has moved ~30 Å closer to the β2 sliding clamp presumably as a result of binding DNA. It is clear that the SEC-SAXS data are a closer match to α in a DNA-free state, and the experimental data most closely resembles the MD model of α in complex with τC16. This implies that a structural rearrangement upon DNA binding is likely. Whether the large movement of the OB-fold to a position closer the β2 clamp occurs during elongation with ssDNA template as α
forms a ternary complex with incoming nucleotide is yet to be demonstrated. Future structural studies will be needed to interrogate this.

Figure 3.16 | The \( \alpha-\tau_{16} \) complex in a DNA-free state is a closer fit to the experimental SEC-SAXS than the DNA-bound state. a, The CRYSOL fits of the refined model (red), DNA-bound (green), and DNA-free (blue) to experimental \( \alpha-\tau_{16} \) SEC-SAXS data. b, Cartoon representation of MD model built based on \( \alpha-\tau_{16} \) in polymerisation mode. Note the location of the \( \tau_{16} \) globular core in relation to \( \alpha \) fingers domain and positioning of the OB-fold to accept incoming ssDNA into the active site. The approximate path of incoming template ssDNA (green cartoon) and dsDNA exiting the active site of \( \alpha \) (black cartoon) is shown. c, Structural alignment of various \( \alpha-\tau_{16} \) complexes for which theoretical scattering was computed. The refined model (red) and DNA-free (green) models the OB-fold and and TBD occupy similar 3D space, while the DNA-bound state (purple) is quite different. d, Same as c, however, the OB fold and TBD from \( \alpha \) have been removed for clarity. It is clear that the \( \tau_{16} \) in the DNA-bound state is rotated outward.
3.3.4.6 3D modelling of the $\alpha$–$\tau_{C24}$ complex from SEC-SAXS and SANS

To reconstruct the 3D shape of the $\alpha$–$\tau_{C24}$ complex from SEC-SAXS and SANS measurements, a combination of ab initio bead and rigid-body modelling was implemented by MONSA [379] and SASREF [382], respectively. Models computed by MONSA, which allow simultaneous fitting of multiple curves (like a neutron contrast variation series), are in good agreement with experimental contrast variation series scattering curves. All 10 MONSA models fit the SANS and SEC-SAXS data well. The best MONSA model and its fits to the scattering data are shown in Figure 3.17a,c with all $\chi^2$ values $\leq 1.3$ for all neutron scattering contrast points. The model clearly resembles the globular shape of $\alpha$ and $\tau_{C24}$ is presented as an elongated shape spanning across $\alpha$. This basic shape is consistent with the position of $\tau_{C16}$ in the published cryo-EM models [222]. Rigid-body modelling was performed by using SASREF, a program that models quaternary structure of macromolecular complexes from multiple scattering curves [382]. The same modelling strategy used for the $\alpha$ subunit (Section 3.3.4.4) was implemented, instead using the unpublished X-ray crystal structure of the $\alpha$CTD–$\tau_{C16}$ complex and adding the preceding 66 amino acids of domain IVa of $\tau$ as 5 short rigid segments. The best SASREF model of the $\alpha$–$\tau_{C24}$ complex is a good match to the experimental scattering curves and its fits are shown in Figure 3.17b. For each contrast point, the $\chi^2$ values were $\leq 2.5$. Figure 3.17d shows that in the best SASREF model, the globular part of $\tau_{C16}$ is in a region close to the fingers domain of $\alpha$. This is similar to the location of the globular part of $\tau_{C16}$ in the DNA-free $\alpha\varepsilon$–$\beta\tau$–$\tau_{C16}$ cryo-EM structure, and the OB fold is positioned in a region where it can feed template ssDNA into the active site of $\alpha$. Moreover, the highly mobile domain IVa appears to protrude into space without contacting $\alpha$. 
Figure 3.17 | Structural modelling of $\alpha$–$\tau_{C24}$ derived from SEC-SAXS and SANS data. 

a, SEC-SAXS and SANS scattering profile of the $\alpha$–$\tau_{C24}$ complex with MONSA model fits. Ten MONSA calculations were performed, the best MONSA model neatly fits the X-ray and all neutron scattering contrast points; X-ray, 0, 42, 60, 80, 100% D$_2$O with discrepancy $\chi^2 = 1.2, 1.3, 0.9, 1.0, 1.0, 0.7$, respectively.

b, SEC-SAXS and SANS scattering profile of the $\alpha$–$\tau_{C24}$ complex with rigid-body modelling (SASREF model fits). Fifteen SASREF calculations were performed and the best SASREF model is in good agreement with X-ray and all neutron scattering contrast points; X-ray, 0, 42, 60, 80, 100% D$_2$O with discrepancy $\chi^2 = 2.5, 2.2, 1.4, 1.2, 1.2$, and 0.5, respectively.

c, Three orthogonal views of the best MONSA ab initio model showing the location of the $\alpha$ subunit (orange) and the $\tau_{C24}$ (blue). The panel on the right shows a $-45^\circ$ rotation about the $x$-axis in the middle panel.

d, Two orthogonal views of the best SASREF rigid-body model showing the $\alpha$ subunit (red) and the $\tau_{C24}$ (blue). Notice the position of the globular core of $\tau_{C24}$ is positioned in a similar position to $\tau_{C16}$ in the DNA-free cryo-EM structure. Also the known flexible domain IVa of $\tau$ is positioned out in space, away from $\alpha$. The panel on the right shows a $-90^\circ$ rotation about the $y$-axis in the left panel.
3.4 Discussion

This Chapter presents SPR and NMR chemical shift data that for the first time show a contribution of domain IV of \( \tau \) to the interaction with \( \alpha \). Moreover, structural studies by SEC-SAXS and SANS with \( \alpha \) alone and in complex with truncated versions of \( \tau \) have revealed a more compact structure than expected. Rigid body models suggest a structural rearrangement in \( \alpha \) upon \( \tau \) binding in the absence of DNA. This structural rearrangement is attributed to the \( \alpha \)CTD.

3.4.1 Contribution of domain IV to the \( \alpha-\tau \) interaction

The \( \alpha-\tau \) interaction is critical for simultaneous leading and lagging strand DNA replication by the *E. coli* DNA Pol III HE. Detailed biochemical studies have clearly demonstrated that the extreme C-terminus of \( \tau \) is responsible for binding to the extreme C-terminus of \( \alpha \). More recent high-resolution structural studies have confirmed that the \( \alpha-\tau \) interaction is an example of an intrinsically unstructured region becoming ordered upon binding to its partner. The X-ray crystal structure of the \( \alpha \)CTD–\( \tau_{C16} \) complex shows that the N- and C-termini of \( \tau_{C16} \) are proximal, imposing a structural limitation for the maximum spatial distance between two \( \alpha \) subunits bound to DnaB via \( \tau_{C24} \) in the replisome during coordinated DNA replication.

The replisome contains at least two \( \tau \) molecules but only one DnaB hexamer. Therefore, at least two \( \tau \) molecules will be available to bind to the same DnaB hexamer. In the context of the Pol III* complex (HE lacking the \( \beta_2 \) clamp), the N-terminal domains of the \( \tau \) subunits are arranged close together in the clamp loader complex (CLC). Consideration of previous results concluding that domain IV of \( \tau \) is the DnaB-binding site and domain V is mostly rigid when interacting with \( \alpha \), and that both \( \alpha \) (~130 kDa) and DnaB (310 kDa) are large proteins, the complex including DnaB, \( \tau \) and both leading and lagging strand Pol III cores may assume a well-defined structure.

Both \( \alpha \) subunits and the DnaB helicase could be positioned near each other, then interactions would be mediated through domain IV of \( \tau \). This, together with the SPR and NMR data presented in this Chapter and previous studies observing that the monomeric truncation mutant \( \tau_{C24} \) can also interact with DNA [17] support such a model. Another key observation in support of this hypothesis is that domain IVa of \( \tau \) can be positioned in such a manner as to contribute to the \( \alpha-\tau \) interaction, resulting in a slower dissociation rate from \( \alpha \). Moreover, the NMR experiments show perturbations.
in unassigned residues in domain IVa of $\tau$ upon addition of full length $\alpha$. Further support for this model comes from the $\tau$–DnaB interaction in *B. subtilis*, where atomic force microscopy shows two neighbouring DnaB domains interacting with two different $\tau$ subunits [383].

The spatial constraint of domain IV to domain V upon binding to $\alpha$ may also play roles in stabilisation of the $\alpha$–$\tau$ interaction during the translocation phase of the polymerase duty cycle, locking the OB fold into a conformation that readily accepts the parental ssDNA template. In addition, the DNA itself may also contribute to stability of the $\alpha$–$\tau$ interaction, which may heavily depend on the positively charged residues in domain IV of $\tau$. In this scenario, one could imagine that these unstructured residues in domain IV act as an electrostatic probe to guide the ssDNA template into the active site of $\alpha$, maintained through its direct interaction with DnaB.

A recent cryo-EM structure of the $\alpha\beta\gamma\delta\epsilon\beta_2\tau_{C16}$ complex in the presence and absence of primer–template DNA has been published. These structures reveal a large conformational change where the $\alpha$CTD moves 35 Å closer to the $\beta_2$ clamp upon binding DNA. A model is put forth to explain this large movement in the context of coupled DNA replication, whereby the $\alpha$CTD acts as a switch to cycle the polymerase off the DNA during Okazaki fragment synthesis [222]. Models computed using SEC-SAXS and SANS data suggest that in the absence of DNA, $\tau_{C16}$ is positioned in a manner where the globular part is able to interact with the fingers domain of $\alpha$, an observation also apparent in the DNA-free cryo-EM structure. The global structure of the DNA-bound state does not fit the experimental SEC-SAXS or SANS data, echoing a large conformational change in the complex in response to binding the primer–template junction. Together, structural studies by SAS imply a separate conformational change in $\alpha$ upon binding $\tau$, presumably occurring through the $\alpha$CTD. Whether this change prepares $\alpha$ for subsequent interactions with proteins or nucleic acid, or promotes a series of intermediate states of the $\alpha$–DNA–incoming nucleotide tertiary complex are yet to be directly observed. It is important to point out that all of the work presented in this Chapter is done in the absence of DNA, and is meant to provide a foundation for future studies. If we are to disentangle and understand the role of domain IV and its contributions to the $\alpha$–$\tau$ interaction in the context of the replisome during replication, studies presented here must be repeated in the presence of appropriate DNA templates.
3.4.2 Is domain IV of $\tau$ an actor in the processivity switch?

It is tempting to speculate based on data presented in this Chapter that in the absence of DNA domain IV of $\tau$ may have additional roles apart from binding to DnaB. The fact that the $\tau$ subunit plays roles in both the CLC that loads the $\beta_2$ clamp on the primer junction and simultaneously binds the polymerase makes it suitable to play a role in signalling. I propose such a role, implicating domain IV in provision of a signalling mechanism that causes a rearrangement of the $\alpha$CTD during switching from polymerisation and/or proofreading modes to strand-displacement synthesis mode, which is likely a discrete conformation that $\alpha$ transitions to as it reaches the end of an Okazaki fragment (Prof Nick Dixon, personal communication). In this process, the regions of domain IV not bound to DnaB may sense the DNA structure of the primer template upstream of the OB-fold. Once this switch is engaged, an $\alpha$CTD rearrangement may occur to better position $\alpha$ for strand-displacement synthesis. If this hypothesis is to be validated, primer–templates must be incorporated into the, $\alpha$–$\tau$, $\alpha$–$\beta_2$ and $\alpha$–$\beta_2$–$\tau$ models to better understand the molecular details of such a signal in the $\alpha$–$\beta_2$–DNA and $\alpha$–$\beta_2$–$\tau$–DNA complexes. In doing so, we will also extend and refine our model into a physiologically relevant context. The outcomes of such experiments will also facilitate understanding of any changes in conformations of the $\alpha$–$\tau$ complex, particularly to the C-terminal domains to a DNA-bound state that are not evident in current models of $\alpha$ in the absence of DNA. By doing this we will be able to ultimately produce a molecular movie of $\alpha$–DNA complex formation. A key challenge is attaining a tight $\alpha$–$\tau$–DNA and $\alpha$–$\beta_2$–DNA complex for equilibrium-based measurements, this is due to the relatively weak affinity of $\alpha$ for DNA.

To overcome this challenge a single point mutant of $\alpha$ ($\alpha$E612K) may be exploited. The single point mutation Glu612 to Lys612 has been characterised [384], where apart from being hyper-processive and resistant to replication pausing, the $\alpha$E612K mutant binds primer–template DNA more tightly than wild-type $\alpha$. Thus, the $\alpha$E612K mutant could be exploited to enable reconstruction of stable Pol III core–DNA complexes for structural studies. Preliminary SANS experiments incorporating $\alpha$E612K have demonstrated the ability to form stable $\alpha$–DNA complexes for timescales amenable to equilibrium-based measurements (unpublished data). Hence, future work will focus on this mutation to study structural changes and characterise the effects of DNA on the $\alpha$–$\tau$ interaction. This may help confirm a role of domain IV as a signaller during coupled DNA replication.
Chapter 4

The $\alpha$ Interaction Domain V of the *Escherichia coli* $\tau$ Subunit also Interacts with the $\beta_2$ Clamp
The α Interaction Domain V of the *Escherichia coli* τ Subunit also Interacts with the β2 Clamp

A physical interaction of the ε proofreading subunit and β2 has been reported. This weak interaction provides stability for the Pol III core on template DNA and is maintained in both the polymerisation and proofreading modes during DNA synthesis. The ε–β2 interaction has likely evolved to be weak enough to be disrupted during transitions to other conformational states, to allow for other enzymes to access β2 during the replication and repair processes. SPR and ESI-MS studies have ranked the stabilities of Pol III–β2 protein complexes. These studies showed using various CBM containing peptides that the dissociation constant $K_D$ of the αεθ–β2 interaction is dependent on the strength of the CBM present in ε. Accurate $K_D$ values of complete, full-length complexes have not yet been reported. Data described in this Chapter rectify this, by using full-length protein complexes containing various CBMs and extending previous measurements. Unexpected, a novel interaction between the C-terminal domains IVa and V of τ and β2 was also identified. Moreover, SPR and NMR spectroscopy show the τ–β2 interaction is weak, with a dissociation constant of ~110 μM. In addition, NMR spectroscopy also allowed identification of residues within domain V of τ that contact β2 in this weak pairwise interaction. The newly identified interaction between τ and β2 may have functional significance in clamp loading pathways.

Contributions are as follows: SPR experiments were performed under the guidance of Dr Slobodan Jergic, University of Wollongong. NMR spectroscopy was performed and analysed under the guidance of Prof Gottfried Otting, Australian National University. Proteins: β2, αL, εL and [15N]-τC22 were generous gifts from Dr Jergic.
4.1 Introduction

It has become quite clear over the past 20 years of research that the interplay between multiple pairwise protein–protein interactions stabilise the replisome. Many of these interactions occur at sites that interact with multiple binding partners. This suggests that the exchange of binding partners in a particular order during replication is critical for proper functioning. It is believed that weak interactions between the Pol III core and $\beta_2$ allow transitions between discrete conformational states, by breaking and reforming interactions without complex collapse, or dissociation from the DNA template – thereby providing fundamental design rules for replisomal function. Consequently, a network of pairwise protein–protein interactions can have reduced apparent dissociation constants ($K_D$) within replisomal subassemblies, relative to $K_D$ values for their individual pairwise interactions. Therefore, it is important to combine biophysical studies with functional assays to identify weak protein interactions, to build a more detailed picture of how these networks of interactions affect the ordering of replication events and transitions between different conformational states.

The Pol III core–$\beta_2$ complex on primer–template DNA undergoes transitions to at least two major conformational states during replication. The first ensures efficient and processive DNA synthesis (polymerisation mode) and the other ensures fidelity by exonucleolytic editing of mismatched nucleotides (proofreading mode). It is currently unknown how transitions occur to each major state and what protein interactions facilitate such changes. During replication, the clamp loader complex (CLC) binds the DNA polymerase (Pol III core) and assembles $\beta_2$ clamps onto the primer template. Once loaded onto DNA, the $\beta_2$ clamp enables the holoenzyme (HE) to gain impressive replicative properties. Ultimately, the $\beta_2$ clamp is a processivity factor that tethers the Pol III core to the template DNA, allowing the polymerase to remain stably bound on the DNA template for thousands of nucleotide incorporations. This protein–protein interaction is critical during nucleotide incorporation and the translocation phases of the DNA polymerase duty cycle. It enables stability without compromising the ability of $\alpha$ to rapidly migrate along DNA between successive nucleotide incorporations. In the polymerisation mode, one of the two symmetry-related sites in the $\beta$ dimer is occupied by $\alpha$ [300] via a short peptide motif (sequence QADMF). It is important to note that related clamp binding motifs (CBMs) occur in disordered segments or loops in the many $\beta$-binding proteins [11]. The two equivalent sites in the homodimeric ring of $\beta_2$ enable it to bind two different proteins at the same time. Thus, it has been suggested to
be important for reversible handover to other repair polymerases (e.g., Pol II, IV, or V) during bypass of a lesion on the template DNA [299,385].

The location of the ε proofreading subunit within the Pol III core–β2 complex in some of its conformational states is currently ill-defined. The ε subunit is divided into two discrete domains (Figure 4.1). The N-terminal region contains the exonuclease domain (residues 2–180) and is also the site of interaction with the θ subunit [21]. The C-terminal domain is flexible, and the residues following Ala209 in the unstructured C-terminal domain of ε are responsible for the interaction with the PHP domain of α [16]. By challenging the replicase to make DNA under difficult conditions, the contribution of weak interactions becomes more apparent and new protein–protein interactions can be discovered. Using this strategy, Jergic et al. [14] have discovered a novel helicase-independent Pol III strand-displacement rolling circle assay that is exclusively dependent on the presence of the ε subunit, however, does not require the N-terminal exonuclease domain. By careful examination using bulk phase biophysical techniques and sensitive single-molecule tethered-bead assays, this activity was traced to a narrow region in the εCTS just after the exonuclease domain that interacts with β2 (termed CBM; Figure 4.1). Thus, the ε–β interaction provides stability for the Pol III core on template DNA and is maintained in the polymerisation mode of DNA synthesis. Complementary studies performed by the Lamers’ laboratory indicate that the ε–β2 interaction is also maintained in the proofreading mode [187]. We speculate that its weak nature enables ε to be easily disrupted from β2 in transitions between various conformational states.

Figure 4.1 | The two domain arrangement of the ε proofreading subunit from E. coli. A CBM is present in the C-terminal domain of the ε subunit. The flexible C-terminal segment (εCTS) interacts strongly with α. Figure adapted from [14].
4.2 Aims of this Chapter

The use of mutant versions of ε that contain weakened and strengthened CBMs have proven invaluable in probing the interactions involved in maintaining polymerisation and proofreading modes [14,15]. In addition, these mutants of ε may also provide insights into possible mechanisms that facilitate transitions between the two modes (Park et al., personal communication), that allow a Pol III core that stalls upon encountering a lesion in the DNA template, to transfer the DNA template to repair polymerases Pol II or Pol IV [229,386]. This Chapter makes use of native, strong and weak CBM mutants of ε and α to determine accurate $K_D$ values of complete, full-length $αε–β_2$ complexes. In the process, it was discovered that the τ subunit (unique C-terminal domains IVa and V) contributes to β₂ binding if present within the complex. Furthermore, the unique C-terminal domains IVa and V of τ are able to out compete the interaction between ε and β₂ within $ταεθ–β_2$ complex. Next, reliable evidence is presented of a novel physical interaction between τ and β₂ using both SPR and NMR assays. Both techniques confirm the same weak binding affinity of $τ_{C16}/τ_{C24}$ for β₂. This is an intriguing result that awaits further investigation to help narrow the spatial and temporal parameters of the τ–β₂ interaction, in the mosaic of Pol III HE conformational states and dynamics that exist at the replication fork.
4.3 Materials and Methods

4.3.1 Plasmid construction

For general methods related to plasmid construction, refer to Chapter 2. Specific vectors and promoters are described in Section 2.1.2. Unless otherwise specified, all plasmids were hosted in *E. coli* strain AN1459 (Section 2.1.1) during construction. Plasmid constructions were confirmed by nucleotide sequence determination (Section 2.5.4.3).

Construction of plasmid pJSL2172 (encoding bio*ε): A N-terminal bio-tag fusion (encoded sequence MAGLNDIFEAQKIEWHEH) of *dnaQ* (encoding ε) was constructed by PCR using pSH1017 [22] as template. An *Nde*I site was incorporated at the start codon of *dnaQ* using primers 650 (5’-AAAAAAAAAAAGCATATGAGCAGCTAATACACGCCAG-3’; *Nde*I site italicised) and outside vector primer pET4 (5’-CCTTTTCGCGCTTTGTTAGCAG-3’). The PCR-generated fragment was then digested with *Nde*I and *Eco*RI and inserted between *Nde*I and *Eco*RI sites of pKO1274 (Section 2.1.2). This yields plasmid pJSL2172, which directs overproduction of bio*ε* under transcriptional control of the phage T7 φ10 promoter.

Construction of plasmid pJSL2164 (encoding bio*ε*Q): The ε mutant εQ contains a single Q to A mutation in the CBM of ε, which greatly weakens the e–b2 interaction [14]. Plasmid pJSL2164 was constructed identically to pJSL2172, except that pSJ1445 was used as template for PCR.

4.3.2 Overproduction and purification of ε

Overproduction and purification of ε was based on a modified purification protocol previously published [14,20]. Plasmid pSH1017 is a T7-promoter vector that directs overproduction of wild-type ε. *E. coli* strain BL21(λ,DE3)/pLysS/pSH1017 was grown at room temperature in LB autoinduction medium (Section 2.3.3) containing ampicillin (100 mg L⁻¹) and chloramphenicol (34 mg L⁻¹). Protein production was autoinduced over a period of 48 h. Cells were then chilled in ice, harvested by centrifugation (16,900 x g; 8 min), frozen in liquid N₂ and stored at –80 °C.

Buffers used for purification of the ε were: lysis buffer (150 mM Tris.HCl pH 7.6, 2 mM dithiothreitol, 1 mM EDTA); refolding buffer (25 mM Tris.HCl pH 7.6, 1 mM EDTA, 5 mM dithiothreitol, 20% (v/v) glycerol); buffer Aε (25 mM Tris.HCl pH 7.6, 2 mM
After thawing, cells (6.3 g from 400 mL of culture) were resuspended in lysis buffer (80 mL) and lysed by being passed twice through a French press at 12,000 psi. The lysate was clarified by centrifugation (35,000 × g; 30 min) to yield the insoluble Fraction I. The ε subunit is completely insoluble even when cells are lysed in the presence of lysis buffer + 1 M NaCl. The next step was to wash the pellet.

In wash one, Fraction I was resuspended in 50 mL of lysis buffer + 1 M NaCl, stirred at 6 °C for 30 min and homogenized by passing twice through a French press at 12,000 psi. The suspension was centrifuged (35,000 × g; 30 min) and the pellet was collected. In wash two, the collected pellet was again resuspended in 50 mL of lysis buffer + 1 M NaCl, stirred in the cold room at 6 °C for 30 min, then passed twice through a French press operated at 12,000 psi and pelleted as in the previous step. Next, in wash three, the pellet was resuspended in 50 mL of lysis buffer with no NaCl, stirred again 6 °C for 30 min, homogenised by passing twice through a French press operated at 12,000 psi and pelleted as in previous steps to yield the insoluble Fraction II. Finally, the pellet was resuspended in 5 mL of lysis buffer by stirring at 6 °C for 30 min to yield Fraction III. This suspension was frozen in liquid N2 and stored in 1 mL aliquots –80 °C.

After thawing, 1 mL of Fraction III was resuspended with 9 mL of refolding buffer and solubilised (unfolded) by addition of 10 mL of refolding buffer + 6 M GndHCl to yield Fraction IV (final GndHCl concentration of 3 M). The transparent Fraction IV was stirred at 6 °C for 30 min. Fraction IV (20 mL) was added dropwise to 40 mL of vigorously stirring refolding buffer to reach a final concentration of 1 M GndHCl, then two complete protease inhibitor cocktail tablets (Roche) were added to yield Fraction V. Fraction V was dialysed for 5 h against 2 L of refolding buffer + 500 mM NaCl, then transferred to dialyse against 2 L of refolding buffer containing no NaCl overnight. The dialysate was clarified by centrifugation (35,000 × g; 40 min) to yield Fraction VI; ~20–30% of ε was found to remain soluble. Fraction VI was applied at 1 mL min⁻¹ to a column (1.5 × 5 cm) of SuperQ-650M resin that had been equilibrated in buffer Aε. After the column was washed with 10 mL of buffer Aε, a linear gradient from 50–400 mM NaCl over 40 mL in buffer Aε was applied, ε eluted as a single peak at ~85 mM NaCl. Fractions containing ε were pooled and dialysed against buffer Bε to give Fraction VII (5.2 mL, containing 12 mg of protein). Aliquots were frozen in liquid N2 and stored at –80 °C.
4.3.3 Overproduction and purification of bio*ε

*E. coli* strain BL21(λ.DE3)/pLysS/pJSL2172 was grown at room temperature in LB autoinduction medium (Section 2.3.3) containing ampicillin (100 mg L⁻¹) and chloramphenicol (34 mg L⁻¹). Protein production was autoinduced over a period of 48 h. After 48 h, cells were chilled in ice, harvested by centrifugation (16,900 × g; 8 min), frozen in liquid N₂ and stored at −80 °C. After thawing, cells (5 g from 400 mL) were lysed and bio*ε was purified as described for ε (Section 4.3.2). A total of 5 mg of bio*ε was obtained. Aliquots were frozen in liquid N₂ and stored at −80 °C.

4.3.4 Overproduction and purification of bio*εQ

*E. coli* strain BL21(λ.DE3)/pLysS/pJSL2164 was grown at room temperature in LB autoinduction medium (Section 2.3.3) containing ampicillin (100 mg L⁻¹) and chloramphenicol (34 mg L⁻¹). Protein production was autoinduced over a period of 48 h. After 48 h, cells were chilled in ice, harvested by centrifugation (16,900 × g; 8 min), frozen in liquid N₂ and stored at −80 °C. After thawing, cells (5.5 g from 400 mL) were lysed and bio*εQ was purified as described for ε (Section 4.3.2). A total of 6 mg of bio*εQ was obtained. Aliquots were frozen in liquid N₂ and stored at −80 °C.

4.3.5 Overproduction and purification of 15N-labelled τC₁₄

Overproduction and purification of uniformly 15N-labelled τC₁₄ was based on a purification protocol previously published [39]. Plasmid pSJ1308 is a λ-promoter vector that directs overproduction of τC₁₄. *E. coli* strain BL21(λ.DE3) recA / pSJ1308 was grown in 3 L of minimal medium for 15N-labelling (Section 2.3.5) supplemented with 100 mg L⁻¹ ampicillin. Cultures were shaken overnight at room temperature to give A₆₀₀ = 1.2. Protein production was induced for 4 h at 42 °C, and then cells were chilled in ice. Cells were harvested by centrifugation (16,900 × g; 8 min), frozen in liquid N₂ and stored at −80 °C.

Buffers used for purification of [15N]-τC₁₄ were: lysis buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM dithiothreitol, 20 mM spermidine); buffer A_{τC₁₄} (35 mM Tris.HCl, pH 7.6, 1 mM EDTA, 2 mM dithiothreitol); buffer B_{τC₁₄} (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 3 mM dithiothreitol, 100 mM NaCl, 10% (ν/ν) glycerol).

After thawing, cells (3 g from 3 L of culture) were resuspended in 50 mL of lysis buffer, one complete protease inhibitor cocktail tablet (Roche) and PMSF added to 0.7 mM to
inhibit proteolysis. The cells were then lysed by being passed twice through a French press (12,000 psi). The lysate was clarified by centrifugation (35,000 × g; 30 min) to yield the soluble Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulfate (0.4 g mL⁻¹) and stirring for 60 min were collected by centrifugation (35000 × g; 30 min) and dissolved in buffer Aᵦ₋₁₃₀ mM NaCl (35 mL). The solution was dialysed against three changes of 2 L of the same buffer, to yield Fraction II (40 mL).

Fraction II was applied at 1.5 mL min⁻¹ onto a column (2.5 × 10 cm) of DEAE-650M resin that had been equilibrated in the same buffer. Fractions containing proteins that did not bind to the column were pooled and dialysed against three changes of 2 L of buffer Aᵦ₋₁₃₀. The dialysate (Fraction III, 65 mL) was loaded at a flow rate of 1.5 mL min⁻¹ onto a column (2.5 × 10 cm) of the same resin, now equilibrated in buffer Aᵦ₋₁₃₀. The column was washed with buffer Aᵦ₋₁₃₀ until [¹⁵N]-TC₁₄ eluted between 1.5–2 column volumes in a sharp peak considering that the purification is based on late elution (with an observed avidity effect). Fractions containing purified [¹⁵N]-TC₁₄ were pooled and dialysed against three changes of 2 L of buffer Bᵦ₋₁₃₀, to give Fraction IV (6 mL, containing 17 mg of protein). Aliquots were frozen in liquid N₂ and stored at –80 °C.

The purified sample was dialysed against three changes of 2 L of NMR buffer (20 mM Tris pH 6.8, 1 mM dithiothreitol, 0.1 mM NaN₃), and then concentrated by ultrafiltration using an MWCO 4000 Ultrafree-4 centrifugal filter unit to a concentration of 1.3 mM solution (1.3 mL) suitable for NMR measurements with β₂. Highly purified β₂ used in NMR experiments (Section 4.3.7) was a generous gift from Dr Slobodan Jergic, University of Wollongong.

4.3.6 Analysis of protein–protein interactions by SPR methodology

All experiments were carried out at 25 °C on a 6 × 6 multiplex ProteOn XPR-36 system (Bio-Rad, Australia). The ProteOn XPR-36 instrument is advantageous as it utilizes an array of 6 × 6 channels of flow cells on a single chip. This allows for interactions to be studied in a one-shot approach without intermittent regeneration, by simultaneous injection of different concentrations of analyte over the immobilised ligand surface (see Section 3.1.3 for a detailed description).

All measurements of the α–ε interaction were performed at 25 °C in SPRa buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM dithiothreitol, 300 mM NaCl, 0.005% (v/v)
P20), with a ProteOn NLC (Neutravidin-coated) sensor chip for immobilization of bio*ε.

All 36 interaction spots of the sensor chip were activated with three sequential injections of 1 M NaCl, 50 mM NaOH across six vertical (ligand) flow paths (40 s each at 40 μL min⁻¹) and six horizontal (analyte) flow paths (40 s each at 100 μL min⁻¹). The surface was further stabilised by two injections of 1 M MgCl₂ in each direction, with the same contact times and flow rates.

4.3.6.1 The α–ε interaction

Bio*ε was diluted to 300 nM in SPRa buffer and immobilised onto all of the six interaction spots of the vertical flow path (60 μL min⁻¹ for 50 s) in ligand channel 5. The sensorgrams verified that immobilised bio*ε remained stably attached to the surface. The chip was then rotated 90°, so flow will pass across all six horizontal (analyte) channels. The interaction with ε was assessed by injection with α in the analyte direction, using an concentration series in SPRa buffer (zero and five concentrations of serially diluted samples; 4, 2, 1, 0.5, and 0.25 nM) at 50 μL min⁻¹ for 300 s, followed by dissociation in the same buffer for 10,000 s. The final sensorgrams were interspot and unmodified ligand flow path subtracted using ProteOn Manager™ Software v3.1.0.6. Binding affinity (dissociation constant; K_d) and parameters of binding kinetic rate constants (k_a and k_d) for the interaction of the α subunit with the ε subunit were determined by global (simultaneous) fitting of five sensorgrams from the optimized concentration range using a Langmuir one-to-one binding model (ProteOn Manager™ software).

4.3.6.2 Interactions with the β2 clamp

The α–τ interaction was used to reliably immobilise α or αε complexes to the chip surface. All αε complexes used in this study were made in situ by mixing 100 nM α with 200 nM ε, followed by a 5 min incubation at room temperature. Both bio*τC16 and bio*τC24 (Section 3.3.1) were diluted to 150 nM in SPRb buffer (50 mM Tris.HCl, pH 7.6, 1 mM EDTA, 1 mM dithiothreitol, 100 mM NaCl, 0.005% (v/v) P20) and immobilised (70 μL min⁻¹ for 20 s) onto all six interaction spots of the vertical flow path in ligand channels 2 and 3, respectively. The sensorogram verified that immobilised α or αε complexes remained stably bound to the surface. The chip was then rotated 90°, so flow will pass across all six horizontal (analyte) channels. The final average densities were 272 RU for bio*τC16 and 770 RU for bio*τC24 across all six interaction spots. Interactions with different αε complexes and β2 were carried out by firstly saturating both τC16 and τC24 containing channels with α (100 nM) and ε (200 nM) when
appropriate, at 30 μL min⁻¹ for 500 s in SPRa buffer. This was followed by sequential injections of one or two appropriate concentration series of β₂ in SPRa buffer (zero and five concentrations of serially diluted samples) in the analyte direction at 50 μL min⁻¹ for 60 s, followed by dissociation in the same buffer until the signal had returned to baseline (typically 200 s). Thus, there was no need for regeneration of surfaces between successive injections. The final sensorgrams were first interspot and unmodified ligand flow path subtracted. Then the signals were normalised to the ratio of α bound at saturation on the six ligand interaction spots of either τ₁₆ or τ₂₄, and bound α on the interaction spot at which the sensorgram was recorded. Then they were finally zeroed using BIAevaluation Software v3.1. Binding affinity (dissociation constant, K_D and kinetic (rate constants, k_a and k_d) parameters for the binding of all complexes were either determined by global (simultaneous) fitting of at least five sensorgrams from the optimised concentration range using BIAevaluation Software v3.1 and the appropriate interaction models: Langmuir one-to-one binding model, Langmuir one-to-one binding model with mass transfer correction, or steady state one-to-one Langmuir binding model.

4.3.7 NMR spectroscopy and data processing

NMR experiments were carried out using 150 μM [¹⁵N]-%₁₆ in NMR buffer (20 mM Tris.HCl, pH 6.8, 1 mM dithiothreitol, 0.1 mM NaN₃) with 10% (v/v) D₂O, in the absence or presence of unlabelled β₂ at 150, 300, and 600 μM as a monomer which was also dialysed into NMR buffer. The NMR experiments were recorded at 298 K on a Bruker 800 MHz NMR spectrometer equipped with a ¹H/¹⁵N/¹³C-triple resonance cryoprobe. Data were processed using Topspin v3.5 (Bruker, Biospin). The dissociation constant was derived using the following equation:

\[ \delta(b) = \delta_{(\text{max})} \times \frac{(K_D + t + b - \sqrt{(K_D + t + b)^2 - 4tb})}{2t} \]

Where \( \delta(b) \) is the ¹H chemical shift change in Hz, \( K_D \) is the dissociation constant, t is the concentration of τ₁₆ and b is the concentration of β [387].
4.4 Results

4.4.1 Biotinylation of bio\(^{*}\)\(_{\text{WT}}\) and bio\(^{*}\)\(_{\text{EQ}}\)

Biotinylated proteins, bio\(^{*}\)\(_{\text{WT}}\) and bio\(^{*}\)\(_{\text{EQ}}\), were required for SPR measurements. Bio\(^{*}\)\(_{\text{WT}}\) and bio\(^{*}\)\(_{\text{EQ}}\) were produced and purified as described in Section 4.3.2 (Figure 4.2). Interestingly, we could resolve by anion exchange chromatography and purify two species that originated from the same constructs (for both bio\(^{*}\)\(_{\text{WT}}\) and bio\(^{*}\)\(_{\text{EQ}}\)). One contained biotin attached to the bio-tag and the other did not. The presence of biotin attached to the bio-tag on the proteins was confirmed by injection onto a streptavidin coated sensor chip as biotinylated bio\(^{*}\)\(_{\text{WT}}\) and bio\(^{*}\)\(_{\text{EQ}}\) did not ionize well during attempts to measure accurate masses by denaturing ESI-MS.

![Figure 4.2](image)

**Figure 4.2** | Purified bio\(^{*}\)\(_{\text{WT}}\) and bio\(^{*}\)\(_{\text{EQ}}\) used in SPR experiments. SDS-PAGE analysis of purified pooled fractions of biointylated bio\(^{*}\)\(_{\text{WT}}\) and bio\(^{*}\)\(_{\text{EQ}}\) subunits used in subsequent SPR measurements (Coomassie blue stained).

4.4.2 A novel assay for measuring interaction between \(\beta_2\) and \(\alpha\) using \(\varepsilon\) by SPR using immobilised \(\varepsilon\)

A physical interaction of \(\varepsilon\) with \(\beta_2\) has been previously reported. In those studies, SPR and ESI-MS ranked the stabilities of protein complexes and showed, using various CBM containing peptides, that the dissociation constant \(K_D\) is dependent on the strength of the CBM present in \(\varepsilon\) [14]. Accurate \(K_D\) values of complete, full-length complexes, however, were not reported. The internal CBM in \(\alpha\) maintains a closed state of the Pol III core−\(\beta_2\) complex during the polymerisation mode of DNA replication. Therefore, understanding the stability and strength of the \(\varepsilon−\beta_2\) interaction in the Pol III core−\(\beta_2\) complex will be important to unravel functional states of the replisome. Data described in this Chapter aim to rectify this lack of knowledge by extending previous measurements, instead using full-length protein complexes containing various CBMs.
To establish accurate $K_D$ values for the interaction of $\varepsilon$ containing either wild-type or strong CBM and/or $\alpha$ and $\beta_2$, and to better understand the rate of dissociation of $\beta_2$ within this complex for assembly of stable complexes for X-ray crystallography, interactions were measured by SPR. Complexes of $\alpha\varepsilon$ were immobilised on the surface to measure $\beta_2$ binding. Using this strategy, comparisons with previous kinetic measurements of the $\alpha-\varepsilon$ interaction could then be made. In a first step, bio*$\varepsilon$ was immobilised on a neutravidin coated surface to a surface density of 368 RU. Then, a range of serially diluted samples of $\alpha$, including a zero concentration, was injected over immobilised $\varepsilon$ as described in Section 4.2.6.1. Since monomeric $\alpha$ binds monomeric $\varepsilon$ (1:1 binding), the preconditions for fitting sensorgrams using Langmuir one-to-one binding model were met. Thus, this model was used to simultaneously fit all five sensorgrams. The resulting sensorgrams of the direct interaction of $\alpha$ with $\varepsilon$ are presented in Figure 4.3; binding parameters subsequently determined are presented in Table 4.1.

![Figure 4.3](image)

**Figure 4.3 | SPR assessment of the $\alpha-\varepsilon$ interaction.** Sensorgrams show association and dissociation phases of the $\alpha-\varepsilon$ interactions at ranges of $\alpha$ concentrations of serially-diluted samples at 4, 2, 1, 0.5, and 0.25 nM of $\alpha$ shown by purple, orange, red, green and blue curves respectively. Curves, shown in different colours, were fit simultaneously (black lines) to a Langmuir one-to-one binding model. Derived binding parameters are shown in Table 4.1.

Our results differ significantly from those reported previously [212]. For instance, the $K_D$ value of $36 \pm 2$ pM differs by 200-fold from the reported value of 7 nM. Discrepancies in the results and low binding stoichiometry are likely due to effects of steric hindrance imposed by the surface matrix. In the previous report $\alpha$ was immobilised via its N-terminus, which is the site of interaction with $\varepsilon$. Since sensorgrams were not reported, it is hard to compare and access the source of the large difference in $K_D$ values. Nevertheless, the data here clearly demonstrate that the $\alpha-\varepsilon$ interaction is extremely strong and stable with a derived $t_{1/2}$ of 11.7 h. It is important to note that we could not
find regeneration conditions that would dissociate the bound $\alpha$ from $\varepsilon$ completely without denaturing $\varepsilon$.

Table 4.1 | Summary of the binding parameters for the $\alpha$–$\varepsilon$ interaction determined by simultaneous fit of sensorgrams to a Langmuir one-to-one binding model. The Table displays determined variables: data collection parameters, equilibrium constant $K_D$, association rate constant $k_a$, dissociation rate constant $k_d$ and binding level if all immobilised $\varepsilon$ molecules that interact with $\alpha$ ($R_{\text{max}}$). All uncertainties are standard errors in parameters from fitting of the complete data set in Figure 4.3.

### Data collection parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>ProteOn XPR36</td>
</tr>
<tr>
<td>Acquisition rate (Hz)</td>
<td>3</td>
</tr>
<tr>
<td>Sensor chip</td>
<td>Neutravidin-coated NLC chip</td>
</tr>
<tr>
<td>Flow rate ($\mu$L min$^{-1}$)</td>
<td>50</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>25</td>
</tr>
<tr>
<td>Association time (s)</td>
<td>300</td>
</tr>
<tr>
<td>Dissociation time (s)</td>
<td>10,000</td>
</tr>
<tr>
<td>$K_D$ (pM)</td>
<td>$35.6 \pm 0.2$</td>
</tr>
<tr>
<td>$k_a$ (M$^{-1}$ s$^{-1}$)</td>
<td>$(4.64 \pm 0.00) \times 10^5$</td>
</tr>
<tr>
<td>$k_d$ (s$^{-1}$)</td>
<td>$(1.65 \pm 0.00) \times 10^{-5}$</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (RU)</td>
<td>257.2 $\pm$ 0.4</td>
</tr>
<tr>
<td>Stoichiometry</td>
<td>0.17</td>
</tr>
</tbody>
</table>

After establishment that the $\alpha$–$\varepsilon$ complex is stable on timescales amenable to measurements of $\beta_2$ binding to $\alpha$ and $\varepsilon$ under our experimental conditions. The $\alpha\varepsilon$–$\beta_2$ interaction was examined using $\alpha$ immobilised via bio$^*$-$\varepsilon$ by injecting $\beta_2$ at nine concentrations of serially diluted samples; 32, 16, 8, 4, 2, 1, 0.5, 0.125 and 0.0625 $\mu$M, as dimers (Figure 4.4a). Detailed kinetic studies of the $\alpha$–$\beta_2$ interactions have been previously described, exhibiting fast on/fast off kinetics and an apparent $K_D$ value of 0.8 $\pm$ 0.1 $\mu$M (in 100 mM potassium glutamate at 20 °C). In our measurements $\beta_2$ interacted with and dissociated from $\alpha$ too quickly to permit determination of rate constants, so the $K_D$ was derived from the equilibrium levels of $\beta_2$ bound to $\varepsilon$–$\alpha$ as a function of $\beta_2$ concentration. In accord with previous measurements [300] we also observed fast on/fast off kinetics even though our measurements are performed in the presence of the $\varepsilon$ subunit (Figure 4.4b). The presence of $\varepsilon$ and the physical interaction between $\varepsilon$ and $\beta_2$ did not visibly slow down the dissociation as expected. The $K_D$ value measured was 3.1 $\pm$ 0.5 $\mu$M (Figure 4.4b inset). The measured affinity of $\alpha$ for $\beta_2$ in the presence of $\varepsilon$ is inconsistent with a 4-fold stronger binding in the absence of $\varepsilon$ [182,300,388]. A likely reason for this discrepancy the difference in salt concentration used for measurements. Previous measurements were performed in 100 mM potassium glutamate while interactions were measured here at 100 mM NaCl. Potassium glutamate is the salt of a weak acid (glutamic acid) and does not contribute much to the ionic strength of the buffer. In addition to this, it is the anions in solution...
that will affect protein–protein interactions, not the cations [389]. For these reasons the apparent \( K_D \) values reported previously in the literature can not be rationally compared with those presented here.

**Figure 4.4 | Immobilisation of the bio*εα complex to measure β2 binding by SPR.**

**a.** Experimental design: First bio*ε was immobilised on the SPR chip surface, then α is introduced at saturating concentrations to form the bio*εα complex. Next β2 is injected over the surface at varying concentrations. **b.** Sensorgrams show association and dissociation phases of bio*εα–β2 interaction at a range of β2 concentrations. Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset).

### 4.4.3 τ contributes to the interaction of β2 with α

The CBM in ε is close to the N-terminal globular domain (Figure 4.1), and although the \( K_D \) derived for the α–β2 interaction after assembly on bio*ε is within the expected \( K_D \) range, an alternate means to immobilise α on the surface was used to alleviate the possibility of dextran-matrix induced steric hindrance on the chip surface. To increase the distance of the CBM (both in α and ε) from the surface, α was immobilised via bio*τC16 and/or bio*τC24. Data presented in Chapter 3 demonstrate the strength and stability of the α–τ interaction; e.g. α–τC24 has a \( t_{1/2} \) of 3.5 h in 300 mM NaCl (Section 3.3.2). Thus, the α–τ interaction is stable on timescales amenable to immobilise αε complexes on the chip surface to measure β2 binding under these experimental conditions.

To first identify any kinetic differences after increasing the distance of the CBM from the surface, β2 binding to α immobilised via \( τ_{C16} \) or \( τ_{C24} \) was performed. First, bio*τC16 or bio*τC24 were immobilised to surface densities of 276 and 770 RU, respectively, followed by injection of α (100 nM) to saturation. Next, binding of β2 to ατC16 was measured by injecting β2 at ten concentrations of serially diluted samples; 32, 16, 8, 4,
2, 1, 0.5, 0.25, 0.125, and 0.0625 μM as a dimer plus a blank. Binding at equilibrium was used to derive the $K_D$ value considering the fast on/fast off kinetics of the interaction (Figure 4.5a). Interestingly, a slight decrease in $K_D$ for $\alpha\tau_{C16}-\beta_2$ binding was observed compared to that of $\alpha\varepsilon-\beta_2$ (2.1 vs. 3.1 μM). This outcome was unexpected considering that, unlike the $\varepsilon-\beta_2$ interaction, binding of $\tau_{C16}$ to $\beta_2$ has not been previously reported in the literature. A significantly weaker $\alpha\tau_{C16}-\beta_2$ interaction would have been expected when compared to $\alpha\varepsilon-\beta_2$ because of the additional CBM in $\varepsilon$. Moreover, a further decrease in $K_D$ was observed when $\tau_{C24}$ was used to immobilise $\alpha$ on the surface ($K_D = 0.89 \pm 0.10$ μM; Figure 4.5b). This further 3-fold increase in binding affinity confirmed the specific contribution of $\tau_{C24}$ to the $\alpha-\beta_2$ interaction and indicated that residues in domain IVa of $\tau$ either: (i) further contribute to binding to $\beta_2$ in a specific manner, or (ii) sufficiently distance domain V of $\tau$ from the chip surface to exacerbate its effect in the absence of potential surface (dextran) induced steric hindrances. Taken together with previous data these measurements suggest for the first time a potential weak $\tau-\beta_2$ interaction in the *E. coli* replisome.

Next, the $\alpha_L-\beta_2$ interaction was investigated; this $\alpha$ mutant ($\alpha_L$) contains a strong CBM (sequence QLDLF). Previously, the introduction of this strong CBM in $\alpha$ at the internal $\beta$-binding site was shown to increase the binding affinity of $\alpha$ to $\beta_2$ by 120-fold compared to its wild-type counterpart [300]. Further, changing the internal CBM to QLDLF from QADMF does not affect the $K_D$ of the $\alpha-\tau$ interaction when directly measured [300]. Surprisingly, under our assay conditions and in the presence of $\tau_{C16}$ or $\tau_{C24}$, the $\tau_{C\alpha_L}-\beta_2$ binding affinities were ~1000-fold greater compared to $\tau_{C\alpha}-\beta_2$ (on either $\tau_{C16}$ or $\tau_{C24}$) when the strong CBM was present in $\alpha$. The stronger affinity is the result of a much slower dissociation rate, and very fast association rate. Therefore, the $K_D$ was determined by simultaneously fitting sensorgrams with a Langmuir one-to-one binding model with mass transport (Figure 4.5c,d; determined binding parameters are presented in Table 4.2). The $\alpha_L-\beta_2$ binding in the presence of $\tau_{C16}$ or $\tau_{C24}$ is ~8-fold stronger compared to the strength of interaction reported in [300]. Although our measurements are performed at significantly higher ionic strength (100 mM NaCl vs. 100 mM potassium glutamate), this would be expected only to weaken the $\alpha_L-\beta_2$ interaction. This further underscores the existence of $\tau-\beta_2$ contacts within the $\tau\alpha(\varepsilon)-\beta_2$ complex. Will the $\alpha\varepsilon-\tau$ complex bind $\beta_2$ with greater affinity?
Figure 4.5 | Strengths of α–β2 interactions in the presence of the C-terminal domains of τ. a, Sensorgrams shows association and dissociation phases of bio-τC16α–β2 interaction at a range of β2 concentrations. Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset). b, Sensorgrams shows association and dissociation phases of bio-τC24α–β2 interaction at a range of β2 concentrations. Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset). c, Sensorgrams show association and dissociation phases of the bio-τC16α–β2 interaction at a range of β2 concentrations. Sensorgrams, shown in blue gradient of colours, were fit simultaneously (black curves) to a Langmuir one-to-one binding model with mass transfer correction. d, Sensorgrams show association and dissociation phases of the bio-τC24α–β2 interaction at a range of β2 concentrations. Curves, shown in green gradient of colours were fit simultaneously (black curves) to a Langmuir one-to-one binding model with mass transfer correction. Derived kinetic parameters are shown in Table 4.2.
Table 4.2 | Summary of the binding parameters for the $\tau\alpha-L-\beta_2$ interaction determined by simultaneous fit of sensorgrams to a Langmuir one-to-one with mass transport binding model. The Table displays determined variables: data collection parameters, equilibrium constant $K_D$, association rate constant $k_a$, dissociation rate constant $k_d$ and saturation binding response $R_{\text{max}}$. All uncertainties are standard errors in parameters from fitting of complete data set as indicated.

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th>Instrument</th>
<th>ProteOn XPR36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition rate (Hz)</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Sensor chip</td>
<td>Neutravidin-coated NLC chip</td>
<td></td>
</tr>
<tr>
<td>Flow rate ($\mu$L min$^{-1}$)</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>Association time (s)</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Dissociation time (s)</td>
<td>300</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Langmuir one-to-one with mass transport</th>
<th>$\tau_\text{C16}\alpha-L-\beta_2$</th>
<th>$\tau_\text{C24}\alpha-L-\beta_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D$ (nM)</td>
<td>$1.77 \pm 0.76$</td>
<td>$0.86 \pm 0.17$</td>
</tr>
<tr>
<td>$k_a$ (M$^{-1}$ s$^{-1}$)</td>
<td>$(6.06 \pm 1.28) \times 10^7$</td>
<td>$(7.79 \pm 0.77) \times 10^7$</td>
</tr>
<tr>
<td>$k_d$ (s$^{-1}$)</td>
<td>$(0.11 \pm 0.02)$</td>
<td>$(0.07 \pm 0.01)$</td>
</tr>
<tr>
<td>$k_t$ (RU M$^{-1}$ s$^{-1}$)</td>
<td>$(2.70 \pm 0.36) \times 10^9$</td>
<td>$(2.30 \pm 0.02) \times 10^9$</td>
</tr>
<tr>
<td>$R_{\text{max}}$ (RU)</td>
<td>$105 \pm 3$</td>
<td>$402 \pm 1$</td>
</tr>
</tbody>
</table>

4.4.4 Addition of $\varepsilon$ to the $\alpha-\tau$ complex does not influence $\beta_2$ binding

To examine the impact of $\varepsilon_{\text{WT}}$ to the strength of $\alpha-\beta_2$ interaction, $\alpha-\varepsilon_{\text{WT}}$ complexes were made in situ and flowed at saturating concentrations (100 nM $\alpha$ and 200 nM $\varepsilon$) over bio*$\tau_{\text{C16}}$ or bio*$\tau_{\text{C24}}$. Then $\beta_2$ was injected at eight concentrations of serially diluted samples: 2, 1, 0.5, 0.25, 0.125, 0.0625, 0.0313, and 0.0156 $\mu$M as a dimer. Binding at equilibrium was used to derive the $K_D$ value (Figure 4.6b,c).

Past studies suggest the presence of $\varepsilon$ in the $\alpha\varepsilon$ complex results in a greater binding affinity of $\alpha$ to $\beta_2$ [14,182,300,388] (lower $K_D$ value). However, the $K_D$ values obtained for the $\alpha-\beta_2$ interaction in the absence of $\varepsilon_{\text{WT}}$ was similar to those measured in its presence (Figure 4.5 and 4.6). The relatively close $K_D$ values of $\alpha-\beta_2$ interaction in the presence of $\varepsilon$ or $\tau$ (Figure 4.4 and 4.5) does not lend itself to synthesis of a reliable conclusion. Nevertheless, the apparent $K_D$ values in particular of $\tau_{\text{C24}}\alpha-\beta_2$ in the presence and absence of $\varepsilon$ are so close ($1.1 \pm 0.0$ and $0.89 \pm 0.10$ $\mu$M respectively). These values are ~3-fold lower compared to the $\varepsilon\alpha-\beta_2$ interaction ($K_D = 3.1 \pm 0.5$ $\mu$M) and indicate that the CBM in $\varepsilon$ may not have access to the binding pocket in $\beta_2$ in the presence of $\tau$ under our experimental conditions. On the other hand, a 45-fold greater $K_D$ value at equilibrium is seen when $\varepsilon_{\text{WT}}$ is replaced with $\varepsilon_{\text{L}}$ containing the strong CBM (QTSMAF *cf.* QLSLPL; [14] under the same experimental conditions (Figure 4.7a,b).
This implies that the strong CBM in ε_L is now able to access the binding site in β_2, even in the presence of the C-terminal domains IVa and/or V of τ.

Figure 4.6 | The native CBM in ε does not significantly contribute to the α−β_2 interaction in the presence of τ. a. SDS-PAGE analysis of purified ε used in make αε complexes in situ in subsequent SPR assays presented in this Chapter (Coomassie blue stained). b. Sensorgrams show association and dissociation phases of bio−τ_C16αε−β_2 interaction at a range of β_2 concentrations (blue gradient). Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset). c. Sensorgrams show association and dissociation phases of bio−τ_C24αε−β_2 interaction at a range of β_2 concentrations (green gradient). Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset).
Figure 4.7 | A stronger CBM in ε strengthens the α–β2 interaction in the presence of τ. a, Sensorgrams show association and dissociation phases of bio−τC16εL−β2 interaction at a range of β2 concentrations (blue gradient). Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset). b, Sensorgrams show association and dissociation phases of bio−τC24εL−β2 interaction at a range of β2 concentrations (green gradient). Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset).

4.4.5 αLε positions β2 in a manner where the εL CBM cannot access its binding site

To access the effect of a strong CBM in α to the strength of ταL−β2 interaction, and the ability of ε to access the β2 binding pocket, αLεWT or αLεL complexes were mixed in situ and introduced at saturating concentrations (100 nM αL and 200 nM εWT or εL) over bio*τC16 or bio*τC24. Then β2 was injected at five or six concentrations of serially diluted samples: 3.125, 1.563, 0.781, 0.391, 0.195, and 0.098 nM for interactions containing αLεWT and 1.563, 0.781, 0.391, 0.195, and 0.098 nM as a dimer for interactions containing αLεL. The resulting sensorgrams are presented in Figure 4.8a–d; binding parameters are shown in Table 4.3. Both sets of sensorgrams were simultaneously fit with a Langmuir one-to-one model with mass transport as $k_a$ values were $>1 \times 10^6$ M$^{-1}$ s$^{-1}$. These $k_a$ values suggest that the measured association rate may reflect the diffusion-limited transfer of β2 from the bulk solution to the matrix, rather than
the true association rate with $\tau \alpha \varepsilon$. The presence of a strong CBM in $\varepsilon$ had no effect on the apparent $K_D$ values when compared to those derived from $\varepsilon_{WT}$ containing complexes. Further, the $K_D$ of $\sim 1$ nM for both $\alpha L$ and $\alpha L\varepsilon L$ complexes (immobilised on bio$^\ast \tau_{C16}$ or bio$^\ast \tau_{C24}$) is very similar to the apparent $K_D$ values measured in $\beta_2$ binding to $\tau-\alpha L$ complexes, where the $\varepsilon$ subunit was absent (Section 4.3.3). Similar $K_D$ values measured for $\tau \alpha L-\beta_2$ and $\tau \alpha L\varepsilon L-\beta_2$ indicated that the presence of the $\tau$ subunit in contact with $\alpha$ renders the other hydrophobic binding pocket in $\beta_2$ (one is occupied by $\alpha$) inaccessible.

The apparent incongruence between the results showing a 45-fold stimulation in $\tau \alpha-\beta_2$ interaction by the presence of $\varepsilon L$ and practically no stimulation at all in the case of $\tau \alpha L-\beta_2$ could be rationalised by the presence of different pathways for assembly of complexes with $\beta_2$. For example, in the case of $\tau \alpha \varepsilon L$, the primary contact with $\beta_2$ is likely established initially through the CBM in $\varepsilon L$ that has a significantly stronger binding affinity to $\beta_2$ compared to the native CBM in $\alpha$. This should result in a faster rate of association and thus lower $K_D$. One possibility is that this multi-site binding role of $\varepsilon$ prevents dissociation (slows dissociation) of $\beta_2$ if the $\alpha-\beta_2$ interaction is transiently broken, allowing for its rapid reformation (re-binding). It remains speculative whether once the first contact with $\varepsilon L$ is made and secondary contact with the CBM in $\alpha$ is established, that the interaction between $\tau$ and $\beta_2$ would render its binding site inaccessible to $\varepsilon L$. Alternatively, the first contact of $\varepsilon L$ with $\beta_2$ may prevent this assembly pathway, i.e., by locking the complex in a conformation where $\tau$ would not have clear access to $\beta_2$. However, in the case of the $\tau \alpha L-\beta_2$ interaction in the presence of $\varepsilon L$, first the contact with $\beta_2$ is likely established via $\tau \alpha L$, that locks the complex in a conformation rendering the $\varepsilon L-\beta_2$ contact less probable.

After observation that in the presence of $\tau \alpha L$, the $\varepsilon L$ CBM cannot access $\beta_2$, we decided to directly determine the impact of $\tau$ on the $\tau \alpha-\beta_2$ interaction by measuring the $\alpha L-\beta_2$ interaction in the absence of $\tau$. To do this we exploited a biotinylated $\varepsilon$ mutant ($\varepsilon_O$) containing a weak CBM (sequence ATSMAF); the $\varepsilon_O$ mutant should not significantly contribute to $\beta_2$ binding due to its very weak affinity for the $\beta_2$ binding pocket ($K_D > 2$ mM) [14].
Figure 4.8 | A strong CBM in α renders the CBM in ε unable to contribute to ταε−β2 interaction. a, Sensorgrams show association and dissociation phases of the bio-τC16L2ε−β2 interaction at a range of β2 concentrations. Black curves show simultaneous fitting to a Langmuir one-to-one with mass transport binding model. b, Sensorgrams show association and dissociation phases of the bio-τC24L5ε−β2 interaction at range of β2 concentrations. Black curves show simultaneous fitting to a Langmuir one-to-one with mass transport binding model. c, Sensorgrams show association and dissociation phases of the bio-τC16L2ε−β2 interaction at a range of β2 concentrations. Black curves show simultaneous fitting to a Langmuir one-to-one with mass transport binding model. d, Sensorgrams show association and dissociation phases of the bio-τC24L5ε−β2 interaction at range of β2 concentrations. Black curves show simultaneous fitting to a Langmuir one-to-one with mass transport binding model. e, Sensorgrams show association and dissociation phases of the bio-τC24L5ε−β2 interaction at a range of β2 concentrations (purple gradient). Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset). Derived kinetic parameters are given in Table 4.3.
Table 4.3 | Summary of the binding parameters for the τCαLε–β2 and τCαLεLε–β2 interactions determined by simultaneous fit of sensorgrams to a Langmuir one-to-one with mass transport binding model. The Table displays determined variables: data collection parameters, equilibrium constant $K_D$, association rate constant $k_a$, dissociation rate constant $k_d$, and saturation binding response $R_{max}$. All uncertainties are standard errors in parameters from fitting of complete data sets as indicated.

**Data collection parameters**

<table>
<thead>
<tr>
<th>Instrument</th>
<th>ProteOn XPR36</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acquisition rate (Hz)</td>
<td>3</td>
</tr>
<tr>
<td>Sensor chip</td>
<td>Neutravidin-coated NLC chip</td>
</tr>
<tr>
<td>Flow rate (μl min⁻¹)</td>
<td>50</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>25</td>
</tr>
<tr>
<td>Association time (s)</td>
<td>60</td>
</tr>
<tr>
<td>Dissociation time (s)</td>
<td>250</td>
</tr>
</tbody>
</table>

**Langmuir one-to-one with mass transport**

<table>
<thead>
<tr>
<th></th>
<th>τCαLε–β2</th>
<th>τCαLεLε–β2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D$ (nM)</td>
<td>1.43 ± 0.13</td>
<td>1.72 ± 0.28</td>
</tr>
<tr>
<td>$k_a$ (M⁻¹ s⁻¹)</td>
<td>(2.76 ± 1.43) × 10⁶</td>
<td>(7.97 ± 0.70) × 10⁶</td>
</tr>
<tr>
<td>$k_d$ (s⁻¹)</td>
<td>(0.0396 ± 0.00)</td>
<td>(0.0137 ± 0.00)</td>
</tr>
<tr>
<td>$k_l$ (RU M⁻¹ s⁻¹)</td>
<td>(10.5 ± 1.4) × 10⁹</td>
<td>(3.96 ± 0.52) × 10⁹</td>
</tr>
<tr>
<td>$R_{max}$ (RU)</td>
<td>106 ± 3</td>
<td>328 ± 10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>τCαLεLε–β2</th>
<th>τCαLεLε–β2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_D$ (nM)</td>
<td>1.35 ± 0.32</td>
<td>1.73 ± 0.11</td>
</tr>
<tr>
<td>$k_a$ (M⁻¹ s⁻¹)</td>
<td>(3.27 ± 0.38) × 10⁷</td>
<td>(3.77 ± 0.12) × 10⁷</td>
</tr>
<tr>
<td>$k_d$ (s⁻¹)</td>
<td>(0.0437 ± 0.00)</td>
<td>(0.0653 ± 0.00)</td>
</tr>
<tr>
<td>$k_l$ (RU M⁻¹ s⁻¹)</td>
<td>(1.59 ± 0.00) × 10⁹</td>
<td>(1.13 ± 0.00) × 10⁹</td>
</tr>
<tr>
<td>$R_{max}$ (RU)</td>
<td>402 ± 14</td>
<td>900 (constrained)</td>
</tr>
</tbody>
</table>

Under these conditions, most of the detected binding could be attributed to the αL–β2 interaction. First, bio*εQ was immobilised on the surface to density of 290 RU, followed by sequential injection of αL at a saturating concentration (100 nM). Then β2 was injected at nine concentrations of serially diluted samples; 160, 80, 40, 20, 10, 5, 2.5, 1.25 and 0.625 nM as a dimer. The sensorgrams are presented in Figure 4.8e. In the absence of τ, the interaction between αL and β2 has different kinetic characteristics, rather resembling a fast on/fast off interaction. Secondly, the $K_D$ value of 14 ± 3 nM calculated from binding at equilibrium is 10-fold weaker than the $K_D$ value obtained for α–β2 immobilised via τ. This clearly demonstrates that when the C-terminal domains of τ are omitted from the αL–β2 interaction and replaced by εQ, the strength of the α–β2 interaction is diminished. This result confirms a contribution of τ to the α–β2 interaction. Collectively, our results clearly indicate that the presence of the τ C-terminal domains both stabilises the α–β2 complex and somehow reduces the incidence of the CBM in ε (when present) finding its binding pocket in β2. The next logical question was whether this new role of τ in the α–β2 interaction is exerted via a direct contact with β2?
4.4.6 A direct physical interaction between τC24 and the β2 clamp

Data in Section 4.3.5 showed that in the absence of τ, the αL–β2 interaction is 10-fold weaker, thus it is reasonable to think there might be a direct interaction between τ and β2. To confirm a direct physical interaction between the C-terminal domains of τ and the β2 clamp, the interaction was examined by SPR using a BIACore T200 instrument (GE Healthcare, Sweden). First, bio*τC24 was immobilised on a fresh streptavidin coated sensor chip to a surface density of 680 RU. We hypothesised that if τC and β2 do physically bind, the interaction would be weak. Hence β2 was injected at five concentrations of serially diluted samples: 200, 100, 50, 25, and 12.5 μM, calculated as a dimer. Due to the instability of the SPR response at these high β2 concentrations, the dissociation constant $K_D$ was then calculated by binding at equilibrium using the first part of the sensorgrams highlighted in grey in Figure 4.9. To our surprise, τC24 did interact with β2 with an apparent $K_D$ of 110 ± 5 μM (as a dimer). To date, this is the first direct evidence for a physical τ–β2 interaction.

**Figure 4.9 | SPR assessment of the τC24–β2 interaction.** Sensorgrams show association and dissociation phases of the bio–τC24–β2 interaction at a range of β2 concentrations (orange gradient). Responses at equilibrium were fit using a steady state one-to-one Langmuir binding model (inset). Fitting resulting in a derived $K_D$ of 110 ± 5 μM with β2 (as a dimer). The grey box indicates the region of the SPR response at equilibrium used for $K_D$ determination.

4.4.7 An examination of the τC–β2 interaction using NMR spectroscopy

It is important to realise that using high protein concentrations above 20 μM pushes the SPR technology to the limit. Artefacts like non-specific binding, re-binding and signal instability occur frequently at these high concentrations. Such artefacts may propagate into derived kinetic and thermodynamic calculations (this is especially true for derived $k_d$ values) so we looked for an alternative method to confirm the τ–β2 interaction.
NMR spectroscopy is widely used to measure protein–protein interactions and map their interfaces. A popular technique is analysis of chemical shift perturbations due to relative ease of execution and sensitivity to subtle changes in chemical environments of proteins. Moreover, an advantage of using this method to measure protein–protein interactions is that the range of measurable interactions is extended into the very high \( \mu \text{M} \) range, a region not accessible by traditional SPR and difficult to measure by other biophysical binding assays. To firstly verify the \( \tau-\beta_2 \) interaction, and secondly access the reliability of the \( K_D \) value for the \( \tau-\beta_2 \) interaction determined by SPR (Section 4.3.6), chemical shift perturbations in \( ^1\text{H}^{15}\text{N}\)-HSQC NMR spectra of the C-terminal domains of \( \tau \) were monitored in a serial dilution series of \( \beta_2 \) concentrations. Proton and \( ^{15}\text{N} \) resonance assignments of \( \tau_{C14} \) (a structured portion of domain V of \( \tau \) that misses the 18 C-terminal residues responsible for \( \alpha \) binding; Section 3.1.2) were previously determined in collaboration with the Otting laboratory [39], allowing potentially easy identification of individual residues in this region of \( \tau \) responsible for binding to \( \beta_2 \).

A sample of uniformly labelled \( ^{15}\text{N} \)-labelled \( \tau_{C14} \) was prepared as described in Section 4.3.5. In addition, considering that consistently stronger \( \alpha-\beta_2 \) interactions were detected by SPR when \( \tau_{C24} \) was used instead of \( \tau_{C16} \), we utilised another truncation mutant of \( \tau \), \( \tau_{C22} \). This protein also misses the 18 C-terminal residues of \( \tau \) responsible for the interaction with \( \alpha \), but contains the preceding domain IVa originally found to be flexible by NMR [39]. In contrast to residues in domain V, those in domain IVa have not been assigned. \( ^{15}\text{N} \)-\( \tau_{C22} \) was generously supplied by Dr Slobodan Jergic (University of Wollongong). Since smaller proteins yield better signals (shaper peaks) in NMR, \( ^1\text{H}^{15}\text{N}\)-HSQC spectrum of a sample of 150 \( \mu \text{M} \) \( ^{15}\text{N} \)-\( \tau_{C14} \) and 600 \( \mu \text{M} \) \( \beta \) (calculated as a monomer) was recorded. Comparison of the measured spectrum to that of \( ^{15}\text{N} \)-\( \tau_{C14} \) alone revealed very little change in peak intensities or chemical shifts, suggesting only a very weak or no interaction between \( \tau_{C14} \) and \( \beta_2 \).

Due to the physical limitation to further increase the \( \beta_2 \) concentration in the sample (\( \beta_2 \) can fortuitously be concentrated to \( \sim \)1 mM and produced in abundance), we turned to \( ^{15}\text{N} \)-\( \tau_{C22} \) which we predicted to interact with \( \beta_2 \) more strongly. Consistent with this prediction (based on previous SPR measurements), an \( ^1\text{H}^{15}\text{N}\)-HSQC spectrum of a sample containing 150 \( \mu \text{M} \) \( ^{15}\text{N} \)-\( \tau_{C22} \) and 600 \( \mu \text{M} \) \( \beta_2 \) (as monomer) revealed pronounced cross-peak attenuations and changes in chemical shifts to some of the residues in \( \tau_{C22} \) compared to the recorded \( ^1\text{H}^{15}\text{N}\)-HSQC spectrum of \( \tau_{C22} \) in the absence
of $\beta_2$. This suggested the $\tau_{C22}-\beta_2$ interaction occurred on a timescale suitable to measure by NMR. To further identify the residues in domain V previously assigned in $\tau_{C22}$, important for binding to $\beta_2$, the $^1$H$^{15}$N chemical shifts associated with $\beta_2$ binding were monitored in a dilution series of decreasing $\beta_2$ concentrations. Superposition of the two-dimensional $^1$H$^{15}$N-HSQC NMR spectra recorded of 150 $\mu$M $[^{15}$N]$\tau_{C22}$ with decreasing concentrations (600, 300, 150, and 0 $\mu$M) of $\beta_2$ (as a monomer) are presented in Figure 4.11a. Residues in $\tau_{C22}$ that show distinguishable changes in $^1$H-$^{15}$N peak positions (chemical shifts) and/or peak intensities are highlighted. Residue Ile619 demonstrated the largest peak shifts coupled with intensity attenuation as a function of $\beta$ concentration and was used to derive the dissociation constant, $K_D$ for binding of $\tau_{C22}$ to $\beta_2$. The $^1$H chemical shift changes are 12.9, 22.9 and 24.9 Hz for 150, 300 and 600 $\mu$M $\beta$ giving a $K_D$ value of 85 ± 54 $\mu$M using the equation in Section 4.3.7. This result is in reasonable agreement with the SPR data (Section 4.4.6). Figure 4.11b shows peak attenuations of backbone amide $^{15}$N-HSQC cross peaks while Figure 4.11c shows the changes in $^1$H chemical shifts, in the presence of 600 $\mu$M $\beta$. The $^1$HN chemical shifts $>$0.007 ppm and attenuation of side chain amides define a $\tau$–$\beta$ binding interface when mapped onto the solution structure of $\tau_{C14}$ (Figure 4.10). In summary, both SPR and NMR confirmed the same weak binding affinity, providing reliable evidence for the physical interaction between the $\beta_2$ clamp and domain V of $\tau$.

Figure 4.10 | Solution structure of $\tau_{C14}$ highlighting the residues for which $^1$HN chemical shift changes were observed for $\tau_{C22}$ in the complex with $\beta_2$ compared to $\tau_{C22}$ alone. Residues for which $^1$HN displayed chemical shift changes $>$0.007 ppm in the presence vs. the absence of $\beta_2$ are shown in a space filling representation. A clear interaction surface is evident centred in the globular region (indicated by the red line; right panel). Chemical shifts associated with residues located in helix 1 and 6 are assumed to be a result of conformational changes to the helices with respect to one another. It is likely they do not participate directly in the interaction with $\beta_2$. PDB: 2AYA [39].
Figure 4.11 | The C-terminal domains of τ bind to the β2 clamp, causing local changes in the chemical environment of τC22. a, Superposition of $^1$H$^{15}$N-HSQC spectra of uniformly $^{15}$N-labelled samples of τC22 at 150 μM (black) with a titration series of 150, (orange) 300, (blue) and 600 (red) μM β (as a monomer) shows both chemical shifts and signal attenuation in many cross-peaks. This is clearly seen in bottom inset. (Bottom inset) Residue Ile619 demonstrated the largest peak shifts and intensity attenuations as a function of β concentration and was used to derive the dissociation constant, $K_D$ for the τC22–β interaction. Labelled residues represented those in domain V that have backbone $^1$HN chemical shift changes >0.007 ppm in $^{15}$N-labelled τC22 in the presence of 600 μM β vs. the absence of β. (Top inset) The $^1$H chemical shift changes are 12.9, 22.9, and 24.9 Hz for 150, 300, and 600 μM β2 to give an apparent $K_D$ value of 85 ± 54 μM for the τC22–β interaction. b, Peak attenuation of the backbone amide $^1$H$^{15}$N-HSQC cross-peaks in the presence of 600 μM. c, Difference in $^1$HN chemical shifts of cross peaks of τC22 in the presence of 600 μM β. The difference in chemical shift for Ile619 is denoted by a red star.
4.5 Discussion

This Chapter presents the equilibrium constants ($K_D$) and interaction kinetics derived by SPR for interactions of $\alpha$, $\alpha\varepsilon$ and $\tau$ (either $\tau_{C16}$ or $\tau_{C24}$) complexes with either their native sequences or variants containing mutated CBMs. Further, using a combination of SPR and NMR spectroscopy, reliable evidence for a novel interaction between domain V of $\tau$ and $\beta_2$ is presented. A summary of the results presented in this Chapter is shown in Table 4.4.

Table 4.4 | Determined values of equilibrium constants ($K_D$) and interaction kinetics for interactions of $\alpha$, $\alpha\varepsilon$ and $\tau$ complexes with $\beta_2$. The Table displays determined variables: equilibrium constant $K_D$, association rate constant $k_a$, and dissociation rate constant $k_d$. All uncertainties are standard errors in parameters from fitting of complete data sets.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>$k_a$ (M$^{-1}$ s$^{-1}$)</th>
<th>$k_d$ (s$^{-1}$)</th>
<th>$K_D$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bio*\varepsilon--\alpha</td>
<td>(4.64 ± 0.00) $\times 10^5$</td>
<td>(1.65 ± 0.00) $\times 10^{-5}$</td>
<td>0.0356 ± 0.002</td>
</tr>
<tr>
<td>bio*\varepsilon--\beta_2</td>
<td>$-$</td>
<td>$-$</td>
<td>3130 ± 490</td>
</tr>
<tr>
<td>bio*\varepsilon--\beta_2</td>
<td>$-$</td>
<td>$-$</td>
<td>2080 ± 200</td>
</tr>
<tr>
<td>bio*$\tau_{C24}\varepsilon$--$\beta_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>890 ± 10</td>
</tr>
<tr>
<td>bio*$\tau_{C16}\varepsilon$--$\beta_2$</td>
<td>(6.06 ± 1.28) $\times 10^7$</td>
<td>(0.11 ± 0.02)</td>
<td>1.77 ± 0.76</td>
</tr>
<tr>
<td>bio*$\tau_{C24}\varepsilon$--$\beta_2$</td>
<td>(7.79 ± 0.77) $\times 10^7$</td>
<td>(0.07 ± 0.01)</td>
<td>0.86 ± 0.17</td>
</tr>
<tr>
<td>bio*$\tau_{C16}\varepsilon$--$\beta_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>1300 ± 110</td>
</tr>
<tr>
<td>bio*$\tau_{C24}\varepsilon$--$\beta_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>1140 ± 50</td>
</tr>
<tr>
<td>bio*$\tau_{C16}\varepsilon$--$\beta_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>29 ± 0.83</td>
</tr>
<tr>
<td>bio*$\tau_{C24}\varepsilon$--$\beta_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>29 ± 2.23</td>
</tr>
<tr>
<td>bio*$\tau_{C16}\varepsilon$--$\beta_2$</td>
<td>(2.76 ± 1.43) $\times 10^6$</td>
<td>(0.0396 ± 0.00)</td>
<td>1.43 ± 0.13</td>
</tr>
<tr>
<td>bio*$\tau_{C24}\varepsilon$--$\beta_2$</td>
<td>(7.97 ± 0.70) $\times 10^6$</td>
<td>(0.0137 ± 0.00)</td>
<td>1.72 ± 0.28</td>
</tr>
<tr>
<td>bio*$\tau_{C16}\varepsilon$--$\beta_2$</td>
<td>(3.27 ± 0.38) $\times 10^7$</td>
<td>(0.0437 ± 0.00)</td>
<td>1.35 ± 0.32</td>
</tr>
<tr>
<td>bio*$\tau_{C24}\varepsilon$--$\beta_2$</td>
<td>(3.77 ± 0.12) $\times 10^7$</td>
<td>(0.0653 ± 0.00)</td>
<td>1.73 ± 0.11</td>
</tr>
<tr>
<td>bio*$\tau_{C16}\varepsilon$--$\beta_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>14 ± 3</td>
</tr>
<tr>
<td>bio*$\tau_{C24}\varepsilon$--$\beta_2$</td>
<td>$-$</td>
<td>$-$</td>
<td>110000 ± 5000</td>
</tr>
</tbody>
</table>

4.5.1 The $\tau$–$\beta_2$ interaction

When the $\alpha$ subunit is mutated to contain a strong CBM positioned at the internal $\beta_2$ binding site (i.e., in $\alpha_L$), a 120-fold decrease in the apparent $K_D$ value of the $\alpha$–$\beta_2$ complex is observed over the native CBM sequence [300]. Interestingly, when $\tau_{C16}$ or $\tau_{C24}$ are present in the $\alpha_L$–$\beta_2$ complex, a ~1000-fold greater binding affinity of ~ 1 nM was observed relative to the $\alpha$–$\beta_2$ complex. However, the derived $K_D$ value of $\beta_2$ binding to the $\tau$–$\alpha$ (on either $\tau_{C16}$ or $\tau_{C24}$) complex is significantly weaker, with a measured $K_D$ value in the low $\mu$M range. Surprisingly, in the absence of $\tau$ the $\alpha_L$–$\beta_2$ interaction is 10-fold weaker with an apparent $K_D$ value of 14 nM. This difference in binding affinities align well with other SPR measurements using $\varepsilon_L$ and $\alpha$ that show a
45-fold decrease in the determined $K_D$ value for $\beta_2$ binding (29 nM) compared to the $\alpha\varepsilon$ complex containing the native CBM. These apparent differences in $K_D$ values suggest that a strong CBM in $\alpha$ renders the CBM in $\varepsilon$ unable to contribute to $\tau\alpha-\beta_2$ interactions, at least in the absence of DNA. The CBM in $\varepsilon$ is flexible in the $\alpha\varepsilon\theta$ complex [15] and is free to sample a large conformational space. Alternatively, these results may simply be explained by the ability of $\tau$ and/or $\alpha$ to position $\beta_2$ in a manner so that the CBM in $\varepsilon$ is inaccessible to the $\beta_2$ binding pocket. Further evidence for this line of thinking comes from comparing the measured $K_D$ values of complexes, such as those containing $\alpha_L-\beta_2$ with either $\varepsilon$ or $\varepsilon_L$, which do not differ in $K_D$ values.

Collectively, these SPR measurements lead us to believe the C-terminal domains of $\tau$ contribute to $\beta_2$ binding to the Pol III complex. In agreement with this, a weak interaction between $\tau_{C24}$ and $\beta_2$ was observed by SPR. These measurements showed equilibrium responses from which a binding affinity constant of $K_D = 110 \pm 5 \mu M$ (as a dimer) was derived. To further confirm this, $^{15}$N-HSQC NMR spectroscopy was performed using a sample of uniformly $^{15}$N-labelled $\tau_{C22}$. A weak $\tau-\beta_2$ binding affinity of $K_D = 86 \pm 54 \mu M$ (as a monomer) was derived from the $^1$H chemical shifts measured during these experiments. The derived binding affinity from these NMR measurements are in rough agreement with the SPR measurements, given that different buffer conditions were used. Derived $K_D$ values by these two different techniques provides reliable evidence that the interaction between the $\beta_2$ clamp and the unique C-terminus of $\tau$ is genuine. It is important to note that the difference in binding strengths between $[^{15}$N]$\tau_{C22}$ and $[^{15}$N]$\tau_{C14}$ in $^{15}$N-HSQC NMR suggests residues in domain IVa also contact and possibly contribute to binding to $\beta_2$. However, it is not clear whether these are simple non-specific or real interactions that have functional implications. This is an intriguing conclusion, because it suggests a new interaction between the $\tau$ and $\beta_2$ subunits may occur at some stage of replisomal dynamics during DNA replication.

A recent cryoEM structure of the $\alpha\tau_{C16}-\beta_2$ complex [222] showed a putative contact between the globular region of $\tau_{C16}$ (residues 562–566) and the fingers domain of $\alpha$ (residues 657–667) in the absence of DNA. Interestingly, residues His652 and Asn564 in $\tau$ identified in this interaction interface (Figure 4.12) share similarities with the interaction interface identified by $^{15}$N-HSQC NMR (Figure 4.11).
Figure 4.12 | Cryo-EM structure of $\alpha_2\beta_2-\tau_{C16}$ complex in DNA-free state. Residues in $\tau_{C16}$ identified by $^1$H$^{15}$N-HSQC NMR to interact with the $\beta_2$ clamp (purple sticks) likely interact with the fingers domain in $\alpha$. Residues identified in this Chapter may have both structural and functional implications in $\alpha$ binding in a DNA-free state. PDB 5FKU has been refined in real space with unpublished $\alpha$CTD–$\tau_{C16}$ X-ray crystal structure by A/Prof Aaron Oakley (University of Wollongong).

Thus, it is tempting to consider this new interaction between $\tau$ and $\beta_2$, and the identified binding surface, plays a role in the $\beta_2$ clamp loading or polymerase loading pathways onto template DNA. This would provide support of a switch-like model proposed by Fernandez-Leiro and co-workers [222], where during clamp loading the $\tau$ binding surface interacts with $\beta_2$, a conformation that may be further stabilized by the presence of domain IV of $\tau$. Upon formation of the $\alpha\beta_2$–DNA complex, a conformational switch is flipped. This allows the polymerase fingers domain in $\alpha$ to rebind the globular region in domain V of $\tau$. Future mutagenesis studies will need to be performed to ascertain the functional relevance of residues Arg561 and His562 in $\tau$, in processes like $\beta_2$ clamp and polymerase loading, using in vitro replication assays.
Chapter 5

Development and Characterisation of Fluorescently Labelled Pol III Cores for Single-Molecule Imaging
Development and Characterisation of Fluorescently Labelled Pol III Cores for Single-Molecule Imaging

The synchronisation and molecular ordering of enzymatic events at the replication fork are incompletely understood. Due to the challenges associated with directly observing replication complexes in action, rare events and short-lived intermediates are often difficult to detect using traditional biochemical methods. In many instances these methods conceal transitions through different conformational states and dynamic behaviour. Simultaneous observation of composition and enzymatic activity of single replisomes is needed to gain a complete understanding of the events of coupled DNA replication by the Pol III HE. Thus, development of new single-molecule fluorescence tools to study replisome dynamics and capture these short-lived intermediate states is required. The production and characterisation of different labelled complexes within the DNA Pol III HE is the first step in achieving this goal. Work presented in this Chapter describes the cloning, purification and labelling of two components of the Pol III core, (α and θ) with high specificity and efficiency. Individually labelled subunits were then reconstituted into Pol III cores and further purified by ion-exchange chromatography. The activity of the labelled Pol III cores was accessed by in vitro DNA replication assays, which show that all labelled cores are functional. Potential in vitro and in vivo uses of fluorescent regents produced in this Chapter are demonstrated by subjecting photo-switchable mKikGR labelled Pol III cores to single-molecule rolling-circle replication assays, as well as live cell imaging. Additionally, an unanticipated consequence of adding a C-terminal cysteine to the θ subunit was uncovered, whereby the additional cysteine becomes modified by attachment of a galactose moiety in vivo.

Contributions: NMR spectroscopy and analysis was performed by Prof Gottfried Otting, Australian National University. Preliminary single-molecule rolling-circle replication assays with mKikGR-α were performed by Dr Karl Duderstadt, Max Planck Institute of Biochemistry. *E. coli* strain MG1655 carrying chromosomal fusion ε-YPet was a generous gift from Dr Andrew Robinson, University of Wollongong. Proteins: β2, SSB4, τδδδγ CLC, and primed ssDNA template were generous gifts from Dr Slobodan Jergic, University of Wollongong.
5.1 Introduction

Great efforts have been made to understand the function and composition of the *E. coli* replisomal subassemblies during DNA replication. Still, the synchronisation and molecular ordering of enzymatic events at the replication fork are not fully understood. This is due to the challenges associated with directly observing replication complexes in action. The textbook view of these processes is constantly being challenged by the development of new techniques. Mounting evidence indicates that the replisome composition is highly dynamic, with many components rapidly exchanging during synthesis [184,390-393]. These studies imply that the replisome does not operate as a stable, static machine; rather it can respond to the changing cellular environment. If we are to construct a comprehensive understanding of the processes and coordination at the replication fork, we must develop and deploy new methods to investigate short-lived intermediate states and dynamic events that occur at the fork. The development of new experimental toolkits in single-molecule biophysics has seen great progress toward achieving these goals, allowing researchers to visualise and characterise single replisomes using both *in vitro* and *in vivo* systems. For instance, *in vitro* studies using flow-stretching tethered bead assays have revealed loop formation and release during Okazaki fragment synthesis [173,392] and fluorescence-based assays have shown that, the normally exchanged $\beta_2$ clamp could be reused when no free protein was available in solution, suggesting plasticity of the replisome depending on its local environment [184]. Likewise, new fluorescence imaging techniques have allowed for single replisomes to be followed in real-time *in vivo*, permitting more accurate quantitative measurements on replisome composition inside the cell [205].

5.1.1 The need for fluorescently labelled components in the Pol III HE

Traditional biochemical DNA replication assays require agarose gel electrophoresis and/or incorporation of radioactive nucleotides to monitor rates of DNA synthesis and processivity, by observation of DNA products. These assays average over a large ensemble of replication events, making it difficult to characterise distributions of enzymatic rates and processivities. Compounding this issue, rare events and short-lived intermediates are usually impossible to detect using these traditional methods.

Simultaneous observation of composition and enzymatic activity of single replisomes will be needed to gain a complete understanding of the events that unfold during coupled DNA replication. The imaging of individual fluorescently-labelled proteins and their interactions with substrates or binding partners reveal aspects of biochemical
reactions that are inaccessible in ensemble experiments such as dynamics or short-lived conformations (reviewed in [394]). Single-molecule techniques already exist to visualise DNA products and short-lived DNA structures during synthesis, such as priming loops [295,296]. Furthermore, these methods have been used to assess the activities of individual replisomes using traditional rolling-circle DNA templates, by flow-stretching the DNA products and visualising them by staining with an DNA-intercalating dye [183,395]. By combining labelling of individual proteins within the replication machinery (e.g. DNA polymerases) and staining DNA products, information about the both composition and structure can be accessed and studied. Such methods have been applied to visualise real-time conformational changes to individual \(E. coli\) \(\beta_2\) clamps as they are loaded by the \(\gamma^3\delta\delta'\) CLC [284]. These types of fluoresecent single-molecule assays are not limited to the \(E. coli\) replication machinery; other replisomes like that of phage T7 have also been studied using these techniques. For instance, labelled T7 gp5 was used to reveal consistent functional exchange of DNA polymerases at the replication fork [396]. Such experiments need to be conducted using the \(E. coli\) replisome, but have largely been limited by the inability to produce active fluorescently labelled replisomal proteins.

Therefore, a prerequisite for single-molecule fluorescence-based experiments is conjugation of a fluorescent probe to the target protein/DNA, usually at one location with high efficiency. Ideally, fluorescent probes need to be long-lived (i.e. slow photobleaching kinetics) and bright (reviewed in [397]). Presented in this Chapter are three typical methods used for specific fluorescent labelling of proteins. These methods include conjugation of maleimide containing organic dyes to thiols on the amino acids (cysteine), genetic fusion with a photo-switchable fluorescent protein, and self-labelling using a SNAP-tag. Below is a brief outline of each labelling strategy employed in this Chapter.

5.1.1.1 Organic fluorophores: maleimide-fluorophore conjugation

Synthetic organic fluorophores are small and can have high photo-stability and brightness. However, chemical coupling of synthetic dyes can be limited by the efficiency of reactions, in specificity as well as the labelling stoichiometry. Cysteines afford the greatest flexibility for location-specific labelling due to the their relatively infrequent occurrence in protein amino acid sequences. Additionally, site-directed mutagenesis can be used to introduce or remove cysteines at specific locations. Maleimide-fluorophore coupled dyes are highly specific for the thiol group of cysteine, and allow for rapid labelling under mild conditions (typical buffers at near physiological
pH; Figure 5.1a). The protein to be labelled must be maintained in a reduced state by chemicals like dithiothreitol or tris[2-carboxyethyl]phosphine prior to the maleimide coupling reaction to avoid inactivation of thiol groups by formation of disulphide bridges or further oxidation. On the other hand, reducing agents used to maintain thiol groups must be removed before conjugation so that they do not compete with target thiols in the proteins, inhibiting the labelling reaction.

In general, chemical conjugation by maleimide-based chemistry (and others) requires precise control of reaction conditions. Labelling efficiency and stoichiometry are crucial factors for single-molecule measurements and may require many optimisation steps to experimental parameters like ionic strength, pH, temperature, reaction time, and protein-to-dye concentration to achieve sufficient labelling. Given that reactions are unlikely to proceed to 100% and a large molar excess of dye to protein is needed, the labelled protein must be purified from excess free dye. Removal of excess free dye and unlabelled populations are typically necessary to minimise fluorescence background and/or the effects of dark populations of proteins on subsequent studies. Since chemical conjugation to organic fluorophores is likely to change the electrostatic properties of the protein and fluorescent dyes are small (typically ≤1 kDa), excess dye molecules can be removed by either ion-exchange or gel filtration chromatography.

5.1.1.2 Genetic fusion of fluorescent proteins

The most popular method to fluorescently label proteins is genetic fusion to a fluorescent protein. Genetic “in frame” labelling affords unique advantages over other options such as chemical coupling. For example, it is absolute in its labelling specificity (genetically encoded) and enables non-invasive tracking of cellular proteins in time and space in specific cells or sub-cellular compartments in living cells. Intense development and refinement over the past two decades has led to a wide variety of fluorescent proteins, offering choices in brightness, pH sensitivity, photophysical properties, and excitation and emission spectra [398]. Thus, the emergence of so-called photoactivatable (and switchable) fluorescent proteins has occurred. Photoactivatable fluorescent proteins are capable of significant changes in their spectral properties in response to irradiation with specific wavelengths of light. Photo-switchable fluorescent proteins have inspired developments in imaging methods like PhADE [399], where a monomeric mutant of KikGR (mKikGR; Figure 5.1b) has been used to increase the concentration range of wide-field single-molecule fluorescence imaging by two orders of magnitude. This has allowed researchers to study eukaryotic DNA replication
dynamics at the single-molecule level using Xenopus egg extracts [399]. Compared to organic dyes fluorescent proteins are limited by their relatively low brightness, photostability (blinking) and fast photobleaching rates. In the case of photostability, improvements may be made in in vitro experiments by addition of oxygen scavenging systems; however, this is not yet possible for in vivo applications.

Figure 5.1 | Three labelling strategies used to fluorescently label proteins. a, (Top panel) Example of an organic fluorescent dye Alexa Fluor 488 (sticks representation). (Bottom panel) Maleimide-fluorophore conjugation is a popular method for specific labelling as it is highly specific for the thiol group of the cysteine, where thiolate anion acts as a nucleophile and attacks the double bond on the maleimide group forming a thioether. This reaction allows for rapid labelling under mild conditions (pH 7–7.5). b, (Top panel) Shows an example of fluorescent protein mKikGR represented as green cartoon (PDB: 2DDD) Fluorescent proteins are larger than organic dyes and allow for genetic fusion to the protein of interest. The surge in development of photoactivatable fluorescent proteins for both in vitro and in vivo use has resulted in studies of photoconversion mechanisms combined with directed evolution to enhance favourable photo-physical characteristics such as brightness, lifetime and maturation (folding) time. (Lower panel) Detailed studies have elucidated mKikGR’s molecular mechanism which allows photo-conversion from a green-state to a red-state after activation by a 405 nm pulse of light. A tripeptide within mKikGR (His62–Tyr63–Gly64), forms a green chromophore that can be photo-converted to a red one via a β-elimination and trans–cis isomerisation which subsequently extents a conjugated π-system (PDB 2DDD, 2DDC). c, Example of a self-labelling tag, the SNAP-tag, modelled with bound SNAP surface 647 substrate (top panel; PDB 3LOO). The SNAP-tag has gained popularity as reactions proceed by a well-defined mechanism, and have predictable stoichiometry and rapid kinetics irrespective of the protein attached to it. (Bottom panel) Proposed mechanism of SNAP-tagging. Chemical or optical probes consist of benzylguanine substrates bearing an appropriate probe attached at the periphery of the benzylic ring; the guanine nucleotide acts as a leaving group.
5.1.1.3 Self-labelling through protein tags

Given the limitations of both chemical conjugation and use of fluorescent proteins, development of another class of protein-based fusion tags that catalyse the covalent self-attachment of an organic fluorophore in either *in vitro* or *in vivo* settings has occurred. Such tags are not innately fluorescent and only become fluorescent when exposed to a specific ligand. This lends itself to several advantages over other labelling strategies. For example, labelling may be restricted in both time and space, and sequential labelling schemes can be used. Moreover the labelling reaction rates (enzymatic speeds) are fast, lower protein:dye ratios can be employed, and a high degree of specificity and efficiency can be achieved.

Many self-labelling protein tags are commercially available including Halo [400], SNAP and CLIP [401,402] tags. The SNAP-tag has been extensively used in work described in this Thesis. The SNAP-tag is an engineered variant of the human DNA repair protein O⁶-alkylguanine-DNA alkyltransferase. The modified SNAP enzyme now catalyses the irreversible attachment of O⁶-benzylguanine derivatives bearing a chemical or optical probe to a cysteine in the active site, while guanine acts as the leaving group (Figure 5.1c). So far, several fluorescent imaging applications have utilised the self-labelling SNAP-tag, including FRET [403,404], chromophore-assisted inactivation [405], biosensors for ion movement [406], and super resolution microscopy [407,408].
5.1.2 *In vitro* DNA replication assays: A means for dissecting the functional properties of the DNA Pol III HE

Nearly every process inside cells from DNA replication and recombination to functions of motor proteins involved in intracellular transport are being studied using *in vitro* systems reconstructed from purified components. Importantly, *in vitro* systems allow precise control over the concentration of each element and each component included to the reaction. Further, the general "noise" arising from natural side reactions or competing processes in the cell is mitigated by elimination of proteins or other biomolecules that catalyse other reactions. The first *in vitro* DNA replication reactions with the Pol III HE were conducted in the Kornberg and Hurwitz research groups ~40 years ago, evolving in complexity throughout the years. These studies began with partial replication reactions, where only a subset of proteins required for the full reaction was present, and the DNA template was in a single-stranded form [9]. These reactions progressed in complexity as the reconstitution of the entire complement of proteins needed for Pol III HE assembly and subsequent replication on primed dsDNA templates was achieved in the early 1990s [254,409].

The use of *in vitro* DNA replication assays has revealed most of the functions of various protein subunits. Important insights into the network of protein–protein interactions which exist (often transiently) within the replisome have also been made using these assays. Many of the discoveries have been made using a rolling-circle DNA template e.g. the discovery that the DNA Pol III HE is an asymmetric dimer, or more recently the weak interaction between \( \varepsilon \) and \( \beta_2 \). The rolling-circle template in its design allows almost indefinite DNA synthesis, being only limited by the processivity of the replisome rather than the length of the template itself. It has proven invaluable in tracking DNA synthesis in real time in ensemble assays, by separation of DNA products on an agarose gel or by incorporation of radioactive nucleotides. More recently single-molecule assays using the rolling-circle DNA template have afforded unique opportunities to follow replication of a single DNA molecule in real time over hundreds of thousands of nucleotide incorporations [183,184,395].

Processive DNA replication by Pol III HE on ssDNA templates is rapid (~1000 bp s\(^{-1}\)) and usually stops when it encounters a duplex region [410]. The Pol III HE also possesses a weak intrinsic strand displacement (SD) activity on 5'-flap-containing templates [14,411-414]. SD synthesis occurs when the Pol III core meets a forked
structure after it completes replication of full circular ssDNA template [14] (See Figure 5.7). Normally, without a helicase to separate the two DNA strands at the fork, the whole replication complex stalls and is blocked. Conditions are now known that allow the Pol III core to proceed through the fork and perform helicase-independent SD synthesis. Moreover, SD synthesis is highly dependent on the presence of several proteins including full-length SSB, ε, τ, and the χ and ψ subunits of the CLC [14,414]. Compared with synthesis on ssDNA templates, higher dNTP concentrations (~250 μM) are also needed. For efficient SD synthesis to occur a Pol III–τ–ψ–χ–SSB bridge that mediates an interaction of the leading-strand polymerase with the lagging-strand template must be formed [414]. Thus, SD synthesis is highly dependent on strong and weak pair-wise interaction networks, that help stabilise the replisome at the fork during its conformational transitions. Consequently, it is a very challenging mode of DNA synthesis for the polymerase to maintain, resulting in distributive DNA replication. When used in combination with traditional rolling-circle replication assays on primed ssDNA templates (where rates, processivity, and length of Okazaki fragments can be monitored), SD rolling-circle DNA replication assays provide an additional tool to interrogate functional consequences arising from introduction of mutation that may affect functions and/or structures of protein–protein interactions within the Pol III HE. Before proceeding to single-molecule DNA replication assays, this SD assay was used as a benchmark to study the activity of the labelled Pol III cores (mKikGR-αεθ, SNAP-αεθ, αεθcys) produced for work described in this Thesis.

5.2 Aims of this Chapter

Traditional in vitro rolling-circle DNA replication assays are only able to monitor the production of DNA, so the dynamics of individual proteins and their subassemblies within in the Pol III HE in E. coli have not been directly visualised. This Chapter explains in detail the expression, purification, and fluorescent labelling of two subunits of the Pol III core, α and θ. Once assembled, these labelled complexes will be used in development of new single-molecule fluorescence tools to study replisome dynamics using established rolling-circle DNA replication assays (see Chapter 6). Procedures were based on work that has been previously done within the Dixon laboratory, but modifications were sometimes necessary, often due to challenges associated with genetic engineering or attachment of fluorescent probes. Potential applications of labelled Pol III cores and individually labelled subunits are given and further discussed.
5.3 Materials and Methods

5.3.1 Construction of plasmids

For general methods related to plasmid construction, refer to Chapter 2. Specific vectors and promoters are described in Section 2.1.2. Unless otherwise specified, all plasmids were hosted in *E. coli* strain AN1459 (Section 2.1.1) during construction. Plasmid constructions were confirmed by nucleotide sequence determination (Section 2.5.4.3).

Construction of plasmid pJSL2131 (encoding $\theta_{cys}$): The *holE* gene encoding single C-terminal cysteine mutant of $\theta$ was constructed by PCR using pCM869 [415] as a template. An *NdeI* site was incorporated at the start codon, TGA stop codon replaced with a TGC cysteine codon, and both a TAA stop codon and *EcoRI* site incorporated at the 5' end of *holE* using primers 651 (5'-AAAAAAAAAAACATATGCTGAAGAATCTGG
CTAAACTG-3'; *NdeI* site italicised) and 652 (5'-AAAGAATTCTTACATTTAAAGTTTG
GGCTCGTAAGG-3'; *EcoRI* site italicised, TAA stop codon underlined). The PCR-generated fragment was then digested with *NdeI* and *EcoRI* at introduced sites and inserted between the *NdeI* and *EcoRI* sites of pETMCSI. This yields plasmid pJSL2131, which directs overproduction of $\theta_{cys}$.

Construction of plasmid pBOB1 (encoding wild-type $\alpha$): The *tac* promoter plasmid pND517 contains the *dnaE* gene between a pair of *BamHI* restriction sites [50]. In addition, the *BamHI* site following the gene overlaps with an *NcoI* site such that previous digestion with *NcoI* eliminates it. To incorporate an *NdeI* site at the start codon of *dnaE*, pND517 was used as template for PCR amplification of the 5'-portion of *dnaE* gene using primers 671 (5'-AAAAAGATCCTAAGGAGGTTGCTATG
TCTGAACCACGTTC-3'; *BamHI* and *NdeI* sites are italicised, ribosome-binding site underlined) and 673 (5'-CGTTTGGCGATCTCAACGGTGT-3'). The PCR product (Fragment I; 522 bp) was isolated from an agarose gel following digestion with *BamHI* and *XhoI*. Next, pND517 was digested with *NcoI*, and the purified linearised product digested independently with *XhoI* to generate Fragment II (3063 bp) and with *BamHI* to yield Fragment III (5129 bp). Fragments I–III were ligated to yield pBOB1.

Construction of pJSL2156 (encoding mKikGR-$\alpha$): Plasmid pCM1444 (Dr Claire Mason, unpublished) is a derivative of pND706 (Section 2.1.2) containing a gene construct encoding the N-terminal mKikGR fusion protein mKikGR-SSB. The 729 bp
*NdeI–BamHI* fragment of pCM1444 encoding mKikGR with a C-terminal flexible peptide linker (sequence: TRESGSIGS [337], the 502 bp *BamHI–Xhol* fragment of pKO1479wt [15] encoding the N-terminal segment of α and the large *NdeI–XhoI* fragment of pBOB1, encoding the remainder of α, were ligated together in a 3:2:1 molar ratio to generate pJSL2156, which directs overproduction of mKikGR-α.

**Construction of plasmid pJSL2197 (encoding SNAP-α):** A modified snap26b gene was amplified from pSNAP-tag(T7)-2 (New England Biolabs) by strand overlap PCR. In the first PCR, an *NdeI* site was incorporated at the start codon and an internal *MluI* site was removed by silent mutation using primers 728 (5’-AAAAAAAAACATATGG ACAAAAGATTGCGAA-3’) and 729 (5’-TGGAAATAGGCAGTTACCCGCGTTCCGC-3’), yielding Fragment I. A second PCR used primers 730 (5’-TGGCTGAACGCCTATTTTC ATCAGCCGGAAGC-3’) and 732 (5’-AAAAGGGATCCGATAGAGCCAGCTCAGCGT TCCCAGACCCGG-3’) to generate fragment II, removing the TGA stop codon and incorporating a sequence encoding a flexible peptide linker (sequence: TRESGSIGS [337]) flanked by *MluI* and *BamHI* sites at the 3’ end of snap26b.*Δ*MluI*. Equimolar amounts of isolated fragments I and II were then used as templates for PCR with the outside primers 728 and 732 to generate a product that was digested with *NdeI* and *BamHI* and isolated from a gel. This fragment was ligated in 3:2:1 molar ratio with the 502 bp *BamHI–Xhol* fragment of pKO1479wt [15] encoding the N-terminal segment of α and the large *NdeI–XhoI* fragment of pBOB1, encoding the remainder of α, to generate pJSL2197, which directs overproduction of SNAP-α.

5.3.2 Overproduction and purification of θ_{cys}

Purification of θ_{cys} was based on previously described method for purification of wild-type θ [23, 24].

Strain BL21(λDE3)/pLysS/pJSL2131 was grown at room temperature in LB autoinduction medium (Section 2.3.3) containing ampicillin (100 mg L⁻¹) and chloramphenicol (34 mg L⁻¹). Protein production was autoinduced over a period of 48 h. Cells were then chilled in ice, harvested by centrifugation (16,900 × g; 8 min), frozen in liquid N₂ and stored at −80 °C.

Buffers used for purification of θ_{cys} were: lysis buffer: 50 mM Tris.HCl pH 7.6, 2 mM dithiothreitol, 1 mM EDTA, 20 mM spermidine; buffer Aθ (40 mM Tris.HCl pH 7.6, 2 mM dithiothreitol, 1 mM EDTA, 5% (v/v) glycerol; buffer Bθ (20 mM sodium phosphate pH
After thawing, cells (9 g from 1 L of culture) were resuspended in lysis buffer (100 mL) and lysed by two passages through a French press operated at 12,000 psi. The lysate was clarified by centrifugation (35,000 × g; 30 min), to yield Fraction I. Proteins that were precipitated from Fraction I by addition of solid ammonium sulphate (0.40 g mL⁻¹) and stirring for 60 min, were collected by centrifugation (35,000 × g; 30 min) and dissolved in buffer Aθ+150 mM NaCl (35 mL). The solution was dialysed against three changes of 2 L of the same buffer, to yield Fraction II.

The dialysate was applied at 1 mL min⁻¹ onto a column (2.5 × 10 cm) of DEAE-650M that had been equilibrated in buffer Aθ+150 mM NaCl. The column was washed with 100 mL of the same buffer. Fractions containing proteins that did not bind to the column were pooled and dialysed against three changes of 2 L of buffer Aθ to yield Fraction III. The dialysate (Fraction III, 70 mL) was again applied at 1 mL min⁻¹ onto a column (2.5 × 10 cm) of the same resin, now equilibrated in buffer Aθ. The column was washed with 100 mL of the same buffer. Fractions containing already relatively pure θcys were pooled and dialysed against three changes of 2 L of buffer Bθ to yield Fraction IV. The dialysate (Fraction IV, 80 mL) was applied at 0.5 mL min⁻¹ onto a column (5 × 14 cm) of Whatman P11 phosphocellulose that had been equilibrated in buffer Bθ. After the column had each been washed with 450 mL of the same buffer, θcys was eluted in a linear gradient of 0–900 mM NaCl over 1,000 mL in buffer Bθ+1 M NaCl. Fractions containing θcys, which eluted in a single peak at ~220 mM NaCl, were pooled and dialysed against three changes of 2 L of buffer Cθ to yield Fraction V (24 mL containing 243 mg of protein). Aliquots were frozen in liquid N₂ and stored at –80 °C.

5.3.3 Thiopropyl Sepharose purification of θcys

Thiopropyl Sepharose resin contains reactive 2-thiopyridyl disulphide groups attached through a chemically stable ether linkage to the Sepharose matrix. This resin allows solutes containing thiol groups to react under mild conditions to form mixed disulphides, allowing reversible immobilisation of thiol containing molecules. After thawing, proteins (90 mg; 10 μmol) in θcys Fraction V were precipitated by addition of solid ammonium sulphate (0.40 g mL⁻¹) and further stirred for 40 min at 6 °C, the solution was centrifuged (35000 × g; 30 min). The precipitate was resuspended in 5
mL of buffer Dθ (100 mM sodium phosphate pH 7.3, 200 mM NaCl, 1 mM EDTA) that had been simultaneously sonicated and deoxygenated while purging with Ar gas to give Fraction II. Fraction II (5 mL) was added drop-wise with gentle stirring into a suspension of 4 mL of Thiopropyl Sepharose 6B resin (capable of binding 20 μmol of thiols) suspended in buffer Dθ at 6 °C for 1 h. The resin was centrifuged (500 × g; 5 min) and supernatant containing modified θcys was removed to yield Fraction III. The remaining resin containing conjugated θcys was resuspended in degassed and deoxygenated buffer Dθ (20 mL) and poured into a column. The column (1.5 × 3 cm) was then washed with 50 mL of the same buffer, θcys was eluted using a step gradient of buffer Dθ (21 mL) containing 25 mM dithiothreitol to yield Fraction IV. Fraction III containing modified θcys and Fraction IV containing θcys were dialysed against three changes of 2 L of buffer (50 mM Tris.HCl pH 7.6, 3 mM dithiothreitol, 1 mM EDTA, 100 mM NaCl, 20% (v/v) glycerol) to yield Fraction V (4 mL containing 16.8 mg of modified θcys) and Fraction VI (12 mL containing 42.6 mg of unmodified θcys (θcys*)). Aliquots were frozen in liquid N₂ and stored at –80 °C.

5.3.4 Solid state labelling of θcys* with maleimide functionalised fluorescent dyes

Methods described below were adapted from [416]. Three different fluorescent probes were used to label θcys*: Fluorescein-5-maleimide (Pierce Biotechnology), CF488a (Biotium) and CF647 (Biotium). First, a total of 2.2 mg of θcys* was reduced with 3 mM tris(2-carboxyethyl)phosphine (pH 7.6) in buffer Eθ (100 mM sodium phosphate pH 7.3, 200 mM NaCl, 1 mM EDTA, 70% (w/v) ammonium sulphate) at 6 °C for 1 h with gentle rotation to yield Fraction I. Fraction I was centrifuged (21,000 × g; 15 min) at 6 °C and the supernatant carefully removed. The precipitate was washed with ice cold buffer Eθ that had been extensively degassed by sonication and deoxygenated using Ar gas, then pelleted by centrifugation (21,000 × g; 15 min) at 6 °C and supernatant removed to yield Fraction III. The labelling reaction was carried out on Fraction III, now devoid of reducing agent, using 5-fold molar excess of maleimide conjugated dyes with 490 μM θcys* in 500 μL of deoxygenated and degassed buffer Fθ (100 mM sodium phosphate pH 7.3, 200 mM NaCl, 1 mM EDTA). The reaction was allowed to proceed for 3 h at 23 °C, followed by further incubation at 6 °C overnight with gentle rotation (in the dark). The reaction was subsequently quenched using 30 mM dithiothreitol for 1 h at 6 °C yielding Fraction IV. Fraction IV was applied at 1 mL min⁻¹ to a column (1.5 × 10 cm) of Superdex G-25 resin equilibrated with buffer Gθ (50 mM Tris.HCl pH 7.6, 3 mM
dithiothreitol, 1 mM EDTA, 100 mM NaCl). Fractions containing the labelled \( \theta_{\text{cys}} \) were pooled and dialysed into buffer H\( \theta \) (50 mM Tris.HCl pH 7.6, 3 mM dithiothreitol, 1 mM EDTA, 100 mM NaCl, 20% \( v/v \) glycerol). The degree of labelling was determined by ESI-MS (Section 2.7).

### 5.3.5 Overproduction and purification of mKikGR-\( \alpha \)

*E. coli* strain BL21(\( \lambda \).DE3)recA/pJSL2156 was grown at 30 °C in LB medium supplemented with thymine (25 mg L\(^{-1}\)) and ampicillin (100 mg L\(^{-1}\)). Upon growth to \( A_{600} = 0.6 \), 1 mM IPTG was added and cultures were shaken for a further 3.5 h, then chilled in ice. Cells (26 g from 6 L of culture) were harvested by centrifugation (16,900 \( \times \) g; 8 min), frozen in liquid N\(_2\) and stored at –80 °C.

After thawing, cells were lysed and mKikGR-\( \alpha \) was purified through Fraction IV essentially as described for wild-type \( \alpha \) [50] (Section 3.2.8). Fraction IV (50 mL) was dialysed against two changes of 2 L of buffer A\( \alpha \) (25 mM Tris.HCl pH 7.6, 2 mM dithiothreitol, 1 mM EDTA, 10% \( v/v \) glycerol) and applied at 1 mL min\(^{-1}\) onto a column (2.5 \( \times \) 12 cm) of heparin-Sepharose [50] that had been equilibrated with buffer A\( \alpha \). The column was washed with 30 mL of buffer A\( \alpha \) and proteins were eluted using a linear gradient (150 mL) of 0–400 mM NaCl in buffer A\( \alpha \). mKikGR-\( \alpha \) eluted as a single peak at \( \sim 40 \) mM NaCl. Fractions were collected and pooled to yield Fraction V, which was applied directly at 1 mL min\(^{-1}\) onto the column of \( \tau_{\text{C16}} \)-agarose affinity resin as previously described in Section 3.2.7. A total of 32 mg (32 mL) of mKikGR-\( \alpha \) was purified. Aliquots were frozen in liquid N\(_2\) and stored at –80 °C.

### 5.3.6 Overproduction and purification of SNAP-\( \alpha \)

*E. coli* strain BL21(\( \lambda \).DE3)recA/pJSL2197 was grown at 30 °C in LB medium supplemented with thymine (25 mg L\(^{-1}\)) and ampicillin (100 mg L\(^{-1}\)). Upon growth to \( A_{600} = 0.6 \), 1 mM IPTG was added and cultures were shaken for a further 3.5 h, then chilled in ice. Cells (30 g from 6 L of culture) were harvested by centrifugation (16,900 \( \times \) g; 8 min), frozen in liquid N\(_2\) and stored at –80 °C. After thawing, cells were lysed and SNAP-\( \alpha \) was purified as described for mKikGR-\( \alpha \) (Section 5.3.5). A total of 68 mg (40 mL) of SNAP-\( \alpha \) was purified. Aliquots were frozen in liquid N\(_2\) and stored at –80 °C.

### 5.3.7 Enzymatic labelling of SNAP-\( \alpha \)

Two different fluorescent probes, SNAP-Surface 649 (red) and SNAP-Surface Alexa Fluor 488 (green; New England Biolabs), were used to label SNAP-\( \alpha \). All labelling
reactions were carried out using a 2-fold molar excess of dye with 27 μM SNAP-α in 1 mL of 50 mM Tris.HCl pH 7.6, 2 mM dithiothreitol, 100 mM NaCl, 5% (v/v) glycerol (buffer Aα) for 2 h at 23 °C, followed by 6 °C overnight with gentle rotation (in dark). Following the coupling, the reaction mixture was supplemented with 1 mM EDTA and excess dye was removed by gel filtration at 1 mL min⁻¹ through a column (1.5 × 10 cm) of Sephadex G-25 (GE Healthcare) in buffer Aα+1 mM EDTA. Fractions containing the labelled SNAP-α were pooled and dialysed against 2 L of buffer Bα (50 mM Tris.HCl pH 7.6, 3 mM dithiothreitol, 1 mM EDTA, 100 mM NaCl, 20% (v/v) glycerol), frozen in liquid N₂ and stored in aliquots at −80 °C. The degree of labelling was measured to be 90% for SNAP-α649 and 83% for SNAP-α488 by UV/Vis spectrophotometry.

5.3.8 In vitro reconstitution and purification of fluorescently labelled Pol III cores

To reconstitute the various labelled Pol III cores, individually purified subunits (either mKikGR-α, labelled SNAP-α or α, ε and θ, or labelled θcys) were mixed in a molar ratio of 1:2.5:5 (α:ε:θ) and dialysed (in the dark) into Buffer APol III (50 mM Tris.HCl pH 7.6, 2 mM dithiothreitol, 1 mM EDTA, 70 mM NaCl, 5% (v/v) glycerol) overnight to yield Fraction I. Excess of ε relative to α and also that of θ relative to both ε and α is critical to ensure that all of α in the mixture indeed forms the Pol III core complex. To remove excess ε and θ from the reconstituted Pol III core, Fraction I was loaded at 1.5 mL min⁻¹ on to a column (1 × 2 cm) of SuperQ-650M equilibrated with buffer APol III. The column was washed (in the dark) until the A₂₈₀ had returned to base line indicating excess ε and θ had passed through the column. The Pol III core complex was then eluted from the column using a sharp linear gradient of 0–1.6 M NaCl over 4 mL in buffer APol III to avoid complex collapse and elute the various labelled Pol III cores as a single peak (at ~160 mM NaCl). Fractions containing labelled Pol III cores were pooled and dialysed (in the dark) into 2 L of buffer BPol III (50 mM Tris.HCl pH 7.6, 3 mM dithiothreitol, 1 mM EDTA, 100 mM NaCl, 25% (v/v) glycerol) to yield Fraction II. Aliquots were frozen in liquid N₂ and stored at −80 °C.

5.3.9 Determination of labelling efficiency by ESI-MS

Labelling efficiency was determined by electrospray ionization mass spectrometry (ESI-MS) by analysing the molecular mass. ESI-MS is accurate to < 0.01% of the total molecular mass of the protein (i.e. 1 Da error for a protein of 10,000 Daltons). Firstly, labelled θcys samples were extensively dialysed into 0.1% (v/v) formic acid at 6 °C. Mass spectra were acquired on a SYNAPT (Waters, USA) mass analyser (in Q-Tof
mode); the instrument was under the control of MassLynx v4.1 software (Microsmass UK). The SYNAPT instrument was calibrated using CsI (10 mg mL\(^{-1}\)) in 70% propan-2-ol over the same \(m/z\) range. Calculated masses were computed using the ProtParam tool from the ExPASy Proteomics Server (ExPASy–ProtParam tool; URL: http://web.expasy.org/protparam/) by inputting the amino acid sequence of proteins and adding the molecular weight of maleimide-fluorophore coupled dyes obtained from the manufacturer’s certificate of analysis.

5.3.10 Degree of labelling determined by UV/Vis spectroscopy

The concentration of the and degree of labelling was measured by UV/Vis spectroscopy for SNAP-\(\alpha\) and \(\theta\)cys* conjugated to fluorescent dyes using the following equations.

The concentration of conjugated protein is given by:

\[
[p\text{rotein (mg L}^{-1}\text{)}] = [A_{280} - (CF_{280} \times \lambda_{\text{max}})] \times \left(\frac{\text{MW}}{\varepsilon_{\text{prot}}}\right)
\]

Where \(A_{280}\) is the measured absorbance at 280 nm, \(CF_{280}\) is the correction factor of the fluorophore’s contribution at 280 nm according to the manufacturers certificate of analysis, \(\lambda_{\text{max}}\) is the measured absorbance of the fluorophore at its maximum absorbance, \(\text{MW}\) is the molecular weight of the protein in Da, and \(\varepsilon_{\text{prot}}\) is the extinction coefficient of the protein in M\(^{-1}\) cm\(^{-1}\).

The degree of labelling (DOL) is calculated using the following equation:

\[
\text{DOL} = \frac{\lambda_{\text{max}} \times \varepsilon_{\text{prot}}}{\text{corrected}A_{280} \times \varepsilon_{\text{dye}}}
\]

Where \(\lambda_{\text{max}}\) is measured absorbance of the fluorophore at its maximum absorbance, \(\varepsilon_{\text{prot}}\) is the extinction coefficient of the protein in M\(^{-1}\) cm\(^{-1}\), \(\text{corrected}A_{280}\) is the absorbance at 280 nm corrected for the fluorophore’s contribution, and \(\varepsilon_{\text{dye}}\) is the extinction coefficient of the dye in M\(^{-1}\) cm\(^{-1}\) according the manufacturers certificate of analysis.

5.3.11 NMR spectroscopy and data processing

NMR experiments were carried out using 1.8 mM modified \(\theta\)cys in NMR buffer (10 mM sodium phosphate pH 6.5 and 0.1 mM NaN\(_3\)) with 5% D\(_2\)O. Proton chemical shifts were referenced to water. NMR experiments were recorded at 298 K on a Bruker 800 MHz NMR spectrometer equipped with a TCI cryoprobe. The spectra recorded included...
TOCSY, COSY, NOESY, $^{13}$C-HSQC, and $^{13}$C-HSQC with TOCSY relay. Data were processed using Topspin v3.5 (Bruker, Biospin).

5.3.12 Strand displacement DNA synthesis assays

5.3.12.1 Production of primed M13 DNA template
The 5$'$-tailed primer-template DNA for the Pol III SD replication assays was prepared using an oligonucleotide primer with the sequence (5$'$-T$^{36}$-TATGTACCCCGGTTGATA ATCAGAAAAGCCCCA-3$'$) that was annealed to a circular, single-stranded wild-type M13 DNA template (6407 nucleotides) such that the 36 nucleotides highlighted remained unhybridised and formed a 5$'$ tail. A 30-fold molar excess of the primer was mixed with M13 ssDNA (35 nM) in a buffer of final composition 30 mM Tris.HCl, pH 7.6, 15 mM MgCl$_2$, 130 mM NaCl and 0.1 mM EDTA. The mixture was treated at 55 °C for 10 min, then slowly cooled over 8 h to allow hybridisation before being stored at –20 °C until use.

5.3.12.2 Replication reaction conditions
Conditions for the standard coupled strand extension and Pol III SD replication reaction were adapted from described methods [14]. Briefly, reactions contained 2.5 nM primed DNA template, 1 mM ATP, 0.5 mM of each dNTP, 30 nM $^{\tau_3\delta\delta'\chi\psi}$, 150 nM Pol III (wild-type or labelled), 200 nM $\beta_2$ and 800 nM SSB$_4$ in 25 mM Tris.HCl pH 7.6, 10 mM MgCl$_2$, 10 mM dithiothreitol and 130 mM NaCl, in a final volume of 13 µL. Components (except DNA) were mixed and treated at room temperature, then cooled in ice for 5 min before addition of DNA. Reactions were initiated at 30 °C, and quenched at time points by addition of EDTA to ~100 mM and SDS to ~1%. Products were separated by 0.66% agarose gel electrophoresis in a running buffer of 180 mM Tris.HCl pH 7.6, 40 mM acetic acid and 4 mM EDTA. The gel was run at 60 V for 150 min, then stained with SYBR-Gold (Invitrogen) per the manufacturer’s directions. After staining, gels were illuminated with 302 nm UV light and photographed using a GelDoc XR+ imaging system (Bio-Rad). The extent of DNA synthesis for individual reactions was assessed qualitatively, based on visual inspection of the stained DNA products.

5.3.13 Fluorescent imaging of mKikGR-α containing replisomes

5.3.13.1 Single-molecule rolling-circle DNA replication assays
Rolling-circle DNA replications assays were performed are described in detail in Chapter 6, Section 6.3.8.
5.3.13.2 Single-molecule imaging of mKikGR-α in vivo

*E. coli* K12 strain MG1655 expressing a chromosomal C-terminal fusion of Pol III ε with YPet (ε-YPet) was cultured overnight at 30 °C in EZ rich medium (Teknova) with glycerol as the carbon source and 100 mg mL⁻¹ ampicillin. Overnight cultures were reset in fresh EZ medium with glycerol with a 1:100 dilution and growth at 30 °C for at least 2 h before imaging. Cells were then introduced onto a silanised microscope slide followed by imaging. Samples were imaged on an inverted microscope (Nikon Eclipse Ti-E) with a CFI Apo TIRF x 100 oil-immersion TIRF objective (NA 1.49, Nikon) equipped with an EM-CCD camera (Photometrics, USA) under control of NIS-elements AR v4.3 (Nikon, Japan). The mKikGR-α was imaged first by being continuously activated by 405 nm laser light (Coherent, Obis 647-100 CW) at 3.16 W cm⁻², then excited with a 568 nm laser (Coherent, Sapphire 568-200 CW) at 441.5 mW cm⁻². Both 405 nm and 568 nm lasers were shuttered and ε-YPet was visualised by excitation with a 514 nm laser (Coherent, Sapphire 568-200 CW) at 16.7 W cm⁻². The analysis was done with ImageJ v1.5c4 using in-house built plugins.
5.4 Results

5.4.1 Non-thiol containing $\theta_{\text{cys}}$

The Pol III $\theta$ subunit is a small 8.8 kDa protein without any native cysteines. The $\theta$ subunit binds to the proofreading subunit $\epsilon$ in the Pol III core forming a very strong 1:1 complex. The lack of cysteines in $\theta$ allows for introduction of a single cysteine for site specific labelling using maleimide-fluorophore conjugation. Therefore, labelling of a single cysteine mutant of $\theta$ could be used as a fluorescent reporter for the Pol III core during DNA synthesis in single-molecule replication assays.

A single cysteine mutant of $\theta$ ($\theta_{\text{cys}}$) was generated by genetic incorporation of an additional cysteine residue to the C-terminus and was purified (Figure 5.2a). Conventional labelling protocols using a desalting step to remove reducing agent prior to labelling yielded low labelling efficiencies of $\leq 10\%$. Since the introduced cysteine is in a highly flexible region of $\theta$, it is likely that rapid oxidation of the sulphur on the cysteine by reactive oxygen species to unreactive states (e.g. disulphide or sulphonic acid) was the cause of the poor labelling efficiencies. To prevent reoxidation of the target thiols, a solid-state labelling protocol was investigated. This procedure has been reported to give high specificity (>90%) for target thiols and high coupling efficiencies ranging from 70–90% [416]. Following this protocol, purified $\theta_{\text{cys}}$ is firstly reduced by tris(2-carboxyethyl)phosphine and precipitated with ammonium sulphate through the salting out effect [417,418] under reducing conditions. Next, centrifugation removes the tris(2-carboxyethyl)phosphine (the reducing agent) in the supernatant and the $\theta_{\text{cys}}$ pellet is recovered. Finally, the $\theta_{\text{cys}}$ is resolubilised in buffer containing maleimide-fluorophore dyes and allowed to react. The excess dye is then removed by gel filtration (Figure 5.2b). Fluorescein-5-maleimide was used to evaluate the effectiveness of this solid-state labelling procedure for labelling $\theta_{\text{cys}}$. Accurate mass determination by ESI-MS was subsequently used to determine labelling efficiencies.

The specificity of solid-state labeling was determined by a comparison of the ESI-MS spectra of the labelled $\theta_{\text{cys}}$ with a sample of wild-type $\theta$. The labelling of the wild-type (cysteine free) $\theta$ was not detectable. Surprisingly, when $\theta_{\text{cys}}$ was reacted with fluorescein-5-maleimide two species were identified by ESI-MS analysis, one being $9,376.2 \pm 0.2$ Da, equivalent to the theoretical molecular weight of $\theta_{\text{cys}}$ covalently bound to a single fluorescein-5-maleimide molecule. The other had a measured mass of $9,111.3 \pm 0.2$ Da, indicating an increase in molecular weight of 162 Da compared to the
theoretical molecular weight of θcys (Figure 5.2c). The latter molecular weight species was also detected by ESI-MS in the purified θcys preparation prior to labelling (Figure 5.2d).

Figure 5.2 | Purification of θcys and detection of a modified version by ESI-MS after labelling with fluorescein-5-maleimide. a, SDS-PAGE analysis of highly purified θcys (Coomassie blue stained). b, Strategy of solid state cysteine labelling. (I) The purified protein is reduced by tris(2-carboxyethyl)phosphine and precipitated with ammonium sulphate under reducing conditions. Centrifugation removes tris(2-carboxyethyl)phosphine in the supernatant and recovers the protein as a pellet. (II) Proteins are dissolved in a buffer containing thiol reactive dyes and allowed to react. (III) Excess dye is removed by gel filtration. c, The nanoESI-mass spectrum of θcys labelled with fluorescein-5-maleimide using solid-state labelling. Two different molecular weight species are present, one corresponding to successful labelling of a single fluorescein-5-maleimide to θcys (green lines) the other 162 Da larger (black lines). d, The nanoESI-mass spectrum of θcys sample containing 3 mM tris(2-carboxyethyl)phosphine prior to labelling also shows two species, one corresponding to the correct size of θcys (red lines) and the other larger molecular weight species seen in c.

To test whether high labelling efficiencies were prohibited by the presence of a population of θcys containing a covalent modification blocking the target thiol, θcys was subjected to thiopropyl chromatography. This type of chromatography can be used to separate thiol-containing proteins from non-thiol-containing proteins. Proteins containing non-reactive thiol groups will pass directly though the column and proteins containing reactive thiol groups will bind to the resin, by formation of a mixed disulphide bond with immobilised 2-pyridyl disulphide. Washing with moderate concentrations of reducing agent such as dithiothreitol can elute bound proteins by disrupting the disulfide bonds. Indeed, the modified θcys species was present in the flow through,
confirming this larger molecular weight species contained a non-reactive thiol (Figure 5.3a). On the contrary, the thiol-containing \( \theta\text{cys} \) (\( \theta\text{cys}^* \)) was retained on the column and eluted upon addition of buffer containing 25 mM dithiothreitol (Figure 5.3b). The addition of a thiopropyl column proved an extremely effective step in removal of non-thiol-containing species from \( \theta\text{cys}^* \). All subsequent solid-state labelling reactions were performed using \( \theta\text{cys}^* \).

**Figure 5.3 | Successful separation of \( \theta\text{cys}^* \) from non-thiol-containing \( \theta\text{cys} \).**

(a) The nano-ESI-mass spectrum of pooled fractions of modified \( \theta\text{cys} \) that did not bind to the thiopropyl column. The species has a measured molecular weight of 9,110.5 ± 0.3 Da. (b) The nano-ESI-mass spectrum of pooled fractions of \( \theta\text{cys}^* \) that were bound and eluted from the thiopropyl column had a measured molecular weight of 8,948.9 ± 0.2 Da.

### 5.4.2 Solid state labelling of \( \theta\text{cys}^* \) is efficient and specific

Removal of non-thiol-containing \( \theta\text{cys}^* \) significantly improved labelling efficiencies. Unlabelled \( \theta\text{cys}^* \) was not detectable by ESI-MS after reaction with maleimide functionalised rhodamine dyes, fluorescein-5-maleimide (\( \theta\text{cysF}5\text{M} \)) and CF488a (\( \theta\text{cys}488 \)) using the solid-state labelling procedure (Figure 5.4a,b). On the other hand, \( \theta\text{cys}^* \) coupled to cyanine dye CF647 did not ionise well, resulting in an insufficient amount of ions to determine an accurate molecular weight. Although unreliable, ESI-MS analysis did show two species present in the sample, one corresponding to the approximate weight of CF647 conjugated to \( \theta\text{cys}^* \) and the other smaller, possibly a partial fragment of the CF647 dye molecule (e.g. the maleimide functional group and spacer without fluorescent dye; Figure 5.4c). In addition, \( \theta\text{cys} \) has a small extinction coefficient at 280 nm (\( \epsilon = 8,480 \text{ M}^{-1} \text{ cm}^{-1} \)) whereas CF647 absorbs at 280 nm at ~3% of its peak absorbance (650 nm; \( \epsilon_{647} \) (CF647) = 240,000 M\(^{-1}\) cm\(^{-1}\)). The low \( \epsilon_{280} \) of \( \theta\text{cys}^* \) makes measuring labelling efficiency at low concentrations difficult, especially with the contaminating signal from the CF647 dye. Estimation of labelling efficiency by UV/Vis spectroscopy was found to be ~30% (Figure 5.4d). However, SDS-PAGE analysis does not show a smaller sized protein band when compared directly to unlabeled \( \theta\text{cys} \), suggesting a much higher labelling efficiency. Hence, solid-state labelling proved an effective procedure to label \( \theta\text{cys}^* \) with maleimide-fluorophore.
coupled dyes from two distinct chemical classes (rhodamine and cyanine) with high specificity and efficiency, making both \( \theta_{\text{cys488}} \) and \( \theta_{\text{cys647}} \) suitable for future single-molecule fluorescence studies.

![Figure 5.4](image)

**Figure 5.4** | Efficient and specific maleimide-fluorophore conjugation to \( \theta_{\text{cys}} \) using solid-state labelling. **a**, The nano-ESI spectrum of \( \theta_{\text{cys}} \) conjugated to a single fluorescein-5-maleimide molecule with a measured mass of 9,376.2 ± 0.2 Da. **b**, The nano-ESI spectrum of \( \theta_{\text{cys}} \) conjugated to a single CF488a molecule with a measured mass of 9,984.5 ± 0.1 Da. **c**, The nano-ESI spectrum of \( \theta_{\text{cys}} \) conjugated to a single CF647 molecule (blue lines) and another species possibly a partial fragment of the CF647 dye molecule (black lines), with measured masses of 9,812.8 ± 0.1 Da and 9,107.8 ± 0.2 Da respectively. **d**, The UV/Vis spectrum of \( \theta_{\text{cys}} \) conjugated to CF647 (blue line). Inset SDS-PAGE gel analysis of unlabeled \( \theta_{\text{cys}} \) (2 \( \mu \)g loaded; left lane) and \( \theta_{\text{cys647}} \) (2 \( \mu \)g loaded; right lane). Due to the small size of \( \theta_{\text{cys}} \) addition of the ~800 Da dye results in an obviously reduced mobility.

### 5.4.3 Efficient labelling of SNAP-\( \alpha \)

Due to the inability of the \( \alpha \) subunit to be ionised in 0.1% (v/v) formic acid using ESI-MS, UV/Vis spectroscopy was used to determine the degree of labelling of SNAP-\( \alpha \). SNAP-\( \alpha \) has a much higher extinction coefficient than \( \theta_{\text{cys}} \) (\( \varepsilon_{280} \) (SNAP-\( \alpha \)) = 123,080 M\(^{-1}\) cm\(^{-1}\)), thus the \( A_{280} \) peak, and labelling efficiency can be confidently measured. The degree of labelling was measured to be 90% for SNAP-\( \alpha \) bound to SNAP-surface 649 (SNAP-\( \alpha_{649} \)) and 83% for SNAP-\( \alpha \) bound to Alexa Fluor 488 (SNAP-\( \alpha_{488} \); Figure 5.5a,b). These high labelling efficiencies make SNAP-\( \alpha_{649} \) and SNAP-\( \alpha_{488} \) suitable for future single-molecule fluorescence studies (see Chapter 6).
Figure 5.5 | UV/Vis spectra of SNAP-α after treatment with O6-benzylguanine derivatives. 

a, UV/Vis spectrum of SNAP-α488 in storage buffer after removal of excess dye by gel filtration. 
b, UV/Vis spectrum of SNAP-α649 in storage buffer after removal of excess dye by gel filtration.

5.4.4 S-galactosylation of θcys in vivo

Two lines of evidence suggested that the higher mass species in θcys might be post-translationally S-glycosylated. Firstly, the additive mass of 162 Da is consistent with addition of a monosaccharide hexose (or hexosamine) unit. Secondly, the high mass species does not contain a reactive thiol group. To identify the chemical composition of the covalent modification to θcys NMR spectroscopy was used. A sample of purified θcys at natural isotopic abundance containing only the non-thiol-containing protein (1.8 mM) was prepared in NMR buffer (Section 5.3.11). To our surprise the 1D proton NMR spectrum shows the expected multiplet fine structures characteristic of a galactose moiety, with the following 1H NMR assignments: (H1) 5.49 ppm, 4.10 ppm (H2), 3.68 ppm (H3), 3.97 ppm (H4), 4.21 ppm (H5) and 3.76 ppm (H6) (Figure 5.6a). A NOESY spectrum shows NOEs with the anomeric proton that point to the two β-protons of Cys at 2.99 and 3.05 ppm (Figure 5.6b). Further, NOESY cross-peaks confirm the conformation of the hexose. The COSY cross-peaks show weak connectivities between H3 and H4, and H4 and H5 protons, typical for the small $^3J_{HH}$ couplings expected for galactose (Figure 5.6c). The TOCSY-relayed HSQC spectrum shows cross-peaks from H1 to H2 and H3, identifying their $^{13}$C chemical shifts at natural abundance (Figure 5.6d). Astonishingly, θcys can be glycosylated in vivo by addition of a galactose moiety to the engineered cysteine at the C-terminus. Prokaryotes typically glycosylate proteins by O- or N-linked glycans, whereas S-glycosylation has only been reported recently for prokaryotes in Gram-positive bacteria, *Bacillus subtilis* and the KW30 strain of *Lactobacillus plantarum* [419,420]. Here, we present the first example of the occurrence of protein S-glycosylation in a Gram-negative bacterium, *E. coli.*
Figure 5.6 | NMR analysis confirming a galactose moiety is covalently attached to $\theta_{\text{cys}}$ in vivo. 

a, 1D $^1$H NMR spectrum showing the chemical shifts expected for galactose with the characteristic anomic proton of the sugar at 5.49 ppm. 

b, NOESY spectrum of S-glycosylated $\theta_{\text{cys}}$ shows NOEs between the anomic proton of galactose and the two $\beta$-protons of a single cysteine residue at $\sim$3 ppm, which also make NOEs to a backbone amide at 8.05 ppm attributed to the amide proton of the cysteine residue. The NOEs with the anomic proton agree with that of galactose. 

c, COSY spectrum depicting the spin system of galactose. 

d, TOCSY-relayed HSQC spectrum of S-glycosylated $\theta_{\text{cys}}$, showing the correlations with the anomic proton.
5.4.5 Labelled Pol III cores are active in rolling-circle strand-displacement DNA synthesis

Fluorescently labelled cores were reconstituted and purified as described in Section 5.3.8. To assess the activity of reconstituted fluorescently labelled Pol III cores (αεθ) compared to an unmodified wild-type Pol III core, Pol III SD rolling-circle assays were performed. In this assay, DNA synthesis by Pol III HE on oligonucleotide-primed circular ss M13 DNA is monitored by separating DNA products on an agarose gel and staining them with SYBR-Gold, a dye that detects both ss and dsDNA. Beyond the expected strand extension reaction to the fully ds circular product (TFII), helicase-independent synthesis of products greater than the length of ds M13 are observed, following a 20 min reaction (Figure 5.7).

The products in Figure 5.7 that are longer than TFII form arise from Pol III SD DNA synthesis, a mode of replication described in [14,414]. Pol III SD products produced from labelled Pol III cores were not as robust as those from wild-type Pol III core under the standard reaction conditions, regardless of the subunit labelled (α or θ). The SD synthesis reaction is challenging in terms of its requirement for all but one of the DNA
Pol III subunits (θ), so it is not surprising that slight differences in activity were detected using this assay. SD synthesis reactions are sensitive to small deficiencies in weak-pair-wise interactions within the Pol III HE upon loading and extension on short DNA templates [14].

The mKikGR-α containing Pol III core had the closest activity to wild-type. Interestingly, it is the only construct without a charged organic molecule attached. Although the θ subunit is dispensable under the standard assay conditions [14], θcys* labelled with CF647 showed reduced SD synthesis compared to wild-type and θcys* labelled with CF488a. This may be a consequence of more negative surface charges on θ coupled to θ’s close vicinity to the active centre of the polymerase α. These additional charges may perturb electrostatic contacts with the DNA template as it enters the active site. Further, these differences in SD activity may arise from steric hindrances, especially for bulky mKikGR and SNAP-tag labelled α. Although the mKikGR and SNAP-tag genetic fusions are spatially constrained by short 9 amino acid linkers, they are still free to explore conformational space and may impact important conformational transitions of the Pol III core during SD synthesis (i.e. loading and delivery of DNA into the active site of α).

5.4.6 Preliminary characterisation of mKikGR-α in vitro and in vivo

The N-terminal genetic fusion mKikGR-α was tested both in vitro and in vivo as a proof of concept for future studies. First purified, reconstituted Pol III core containing mKikGR-α was examined in the rolling-circle DNA replication assay. In these assays replication is visualised by real-time near-total internal reflection fluorescence (TIRF) imaging of stained dsDNA. This method allows quantification of the rates and processivities of individual replisomes. Replisomes reconstituted using mKikGR-α were active in coupled replication, having similar rate (474 ± 32 vs. 492 ± 25 bp/s) and processivity (35 ± 9 vs. 73 ± 25 kb) distributions to wild-type replisomes (Figure 5.8 and Figure 6.4).
Figure 5.8 | mKikGR-α has similar rates and processivity compared to wild-type α. a, Kymograph of an individual DNA molecule (43.2 kb) undergoing coupled leading- and lagging-strand replication. Refer to Chapter 6 for experimental details. The arrow indicates the direction of flow. b, Histogram showing the distribution of rates of replication for mKikGR Pol III (474 ± 32 bp s⁻¹) fit to a Gaussian distribution. c, Histogram of the processivity of replication of mKikGR Pol III (35 ± 9 kb), fit with a single exponential decay function. All uncertainties represent the standard deviation.

Next mKikGR-α was imaged in vivo to demonstrate its usefulness as a photo-switchable fluorescent marker for replisome foci in living cells. First plasmid pJSL2156 (encoding mKikGR-α) was transformed into strain MG1655 expressing ε-YPet chromosomal fusion and imaged. The basal expression level of mKikGR-α in pJSL2156 made it possible to observe fluorescent foci in living E. coli cells without high fluorescent background. To reduce fluorescence bleed-through from unswitched mKikGR-α into the ε-YPet channel, mKikGR-α was first photoswitched to its red state by continuous excitation by a 405 nm laser and imaged. Then ε-YPet was subsequently imaged. On average cells contained four replisome foci per cell, which is typical of E. coli cells undergoing active DNA replication and division in log phase growing in a EZ-glycerol growth medium. Chromosomal fusion ε-YPet has been used as a fluorescent marker for active replication forks [205,255]. To demonstrate that mKikGR-α is able to occupy the same cellular position as ε-YPet during a given time, co-localisation analysis of mKikGR-α foci with ε-YPet foci was performed. Fitting to a 2D Guassian peak of fluorescent foci was used to determine the degree of co-
localisation between ε-YPet and mKikGR-α. An mKikGR-α focus was determined to be co-incident with ε-YPet if the centre of the fitted peak (within 2 pixels of both $x$ and $y$ coordinates) was coincident with those determined for ε-YPet foci. Positions of mKikGR-α were coincident with ε-YPet with 57% co-localisation ($N = 14$ cells; Figure 5.9). This degree of co-localisation is outside the range expected by chance, based on the area of the cell occupied by replisome foci, ~5–10% (Dr Andrew Robinson, personal communication). The simplest interpretation of this observation is that mKikGR-α can be incorporated into active replication forks in vivo from a plasmid based system.

**Figure 5.9** | mKikGR-α co-localises with chromosomal fusion ε-YPet in replisome foci. Representative cell showing colocalisation (white spots) of mKikGR-α and ε-YPet at replisome foci. Individual channels are labelled below each image. White arrows indicate co-localised replication foci.
5.5 Discussion
This Chapter presents the cloning, purification and labelling of two components of the Pol III core (α and θ) with high specificity and efficiency. Subsequently, labelled individual subunits were reconstituted into Pol III cores and further purified by ion-exchange chromatography. Pol III SD DNA replication assays show that all labelled cores are able to perform strand extension and participate in SD DNA synthesis. During the production of a single cysteine mutant, θcys, we unexpectedly uncovered that it can be S-glycosylated in vivo by addition of a galactose moiety onto the artificial cysteine on the C-terminus. Finally, the potential uses of mKikGR-labelled Pol III cores, both in vitro and in vivo are demonstrated.

5.5.1 Solid-state labelling protocol: An efficient methodology
Fluorescent labelling by modification of solvent accessible cysteine residues using thiol-reactive dyes is an attractive method for site-specific conjugation of fluorophores. Bottlenecks, however, in this reaction include maintenance of reactive thiol groups without oxidation before labelling and the effective removal of excess dye molecules prior to fluorescence studies. Implementation and modification of a solid-state labelling protocol enabled high labelling efficiency (>90%) and specificity of the θ subunit of the Pol III core with two distinct chemical classes of organic dyes, rhodamine and cyanine. I hypothesise that the transition of reduced target thiols from a local environment that is highly solvent exposed to the solid-state by salting out with ammonium sulphate efficiently prevents oxidation to unreactive forms. This results in high labelling efficiencies. This is likely a consequence of rapid removal of reducing agent and replacement with reactive dye, combined with the restricted movement of the thiol without access to oxygen rich solvent. Moreover, solid-state labelling is advantageous in that multiple samples can be labelled simultaneously. Simply, the crucial steps can be easily synchronized, the reducing agent is rapidly removed, and the concentration of protein and maleimide-fluorophore can be precisely controlled and maintained throughout the protocol. In conventional methods, however, this is more difficult and much slower. I believe this method is versatile and can be easily expanded to other thiol labelling chemistries like haloacetyl and alkyl halide with limited amendments to the procedure.

5.5.2 S-galactosylation of θcys: An artificial case
The S-glycosylation of θcys with a single galactose moiety shows for the first time the existence of a S-glycosylation pathway in E. coli. Glycosylation is typically observed on
asparagine or serine/threonine residues. However, modification of a cysteine with a glycan has been reported on human glycopeptides but is extremely rare. Two independent studies have reported the discovery of S-glycosylation of bacterial peptides. The first, was an antimicrobial peptide produced by *Bacillus subtilis* 168 called sublancin, which becomes modified with a glucose moiety [421,422]. Further, enzyme SunS, formally an uncharacterised gene product of *yolJ*, was found to add glucose to the sulphur on sublancin cysteine 22. The second, was glycocin F, a bacteriocin produced in *Lactobacillus plantarum* KW30 bearing an acetylhexosamine attached to its C-terminal cysteine residue [419,420]. It is noteworthy that SunS does not share sequence homology with any *E. coli* proteins. It is curious that in both reported cases of S-glycosolation in bacteria the modified cysteine residue is located at either the extreme C-terminus or in a flexible region, suggesting penetration into a binding pocket during attachment of the sugar.

The expression procedure may provide some insight into the mechanism by which θcys may become modified in vivo. Expression by autoinduction relies on the preferential use of carbon sources by *E. coli* and the regulation of glucose/lactose levels in the growth medium [346]. Autoinduction of protein production using the T7 promoter system under control of the lac operon (expressing T7 RNA polymerase) requires a functional LacY transporter and active β-galactosidase, which performs transglycosylation of lactose to allolactose (the actual inducer) as well as hydrolysis of lactose to glucose and galactose. Consequently, β-galactosidase is constantly being expressed during target protein expression. Usually, β-galactosidase catalyses reactions with β-D-galacto-pyranosides with an oxygen glycosidic bond [423,424]. The enzyme is promiscuous, also reacting with substrates having other glycosidic linkages including nitrogen [425] and sulphur, but with much reduced catalytic efficiency [426]. It is thus tempting to speculate that β-galactosidase may possess a moonlighting function in *E. coli*. Mechanistically it is easy to speculate how a such a reaction could occur. First, the reactive thiol on θcys would behave directly as the nucleophile during the cleavage of lactose into the products galactose and glucose, followed by transfer of the galactose molecule to the artificial cysteine in θcys. Preliminary experiments support this hypothesis. Expression of θcys in BL21(λDE3)AI, an *E. coli* strain which carries the T7 RNA polymerase gene in the araB locus allowing T7 RNA polymerase to be regulated by the araBAD promoter inducible by L-arabinose, [427] eliminates S-galactosylation of θcys (data not shown). Although outside the scope of this Thesis, it would be interesting
to pursue further, using o-nitrophenyl β-D-galactopyranoside as a substrate in colourimetric assays with β-galactosidase.

5.5.3 Potential uses for labelled Pol III cores

The advent of single-molecule fluorescence techniques has allowed dissection of roles of many protein–protein and protein–DNA interactions in great kinetic detail. Though classical biochemical studies have elucidated most of the numerous protein functions and interactions within the replisome, several properties can be more readily examined through direct visualisation of the components and activities of individual complexes. Examples are the dynamic behaviours that govern transitions through multiple conformational states at the replication fork, in particular how and when the Pol III core enters and leaves the fork. The current view is of a stable complex, where the same Pol III core is re-used for the recurrent synthesis of lagging strand Okazaki fragments as shown by in vitro experiments. Surprisingly, rapid and frequent exchange of DNA polymerases in and out of phage replisomes has been observed using sensitive single-molecule techniques [396,428,429]. Access to labelled polymerases through mKikGR-α and SNAP-α fusions, as well as labelled Pol III cores (through labelled θ), will allow us to probe similar polymerase exchange mechanisms that may also apply to the E. coli replisome. This subject will be explored in detail in Chapter 6.

Secondly, θ’s role at the replication fork is yet to be fully understood. Labelled θ can be used to follow θ through different modes of DNA synthesis (polymerisation and proofreading) either using traditional rolling-circle based assays, or other techniques such as fluorescence correlation spectroscopy. Such experiments may illuminate the exact role of θ during replication. Questions concerning proofreading dynamics may also be investigated using single-molecule FRET, considering the close proximity of labelled θ to the ε active site. A possible result from these studies would be determination of distances and the timing of the transfer of a mismatched primer template DNA from α to ε.

This Chapter has demonstrated that genetic fusion of a fluorescent protein to the N-terminus of α has little effect on its native function or its ability to be purified and reconstituted into Pol III cores for in vitro studies. The genetic fusion of mKikGR-α was used to probe the effect of N-terminal fusions to α both in vivo and in vitro. In vivo experiments where mKikGR-α was expressed using a plasmid-based system showed it to be actively incorporated into replisomes and could be used as a replisome marker.
(Section 5.4.6). Preliminary in vitro single-molecule characterisation of mKikGR-α Pol III cores show similar rates and processivities compared to wild-type Pol III core (Section 5.4.6). These observations suggest α may tolerate other fluorescent proteins and tags at its N-terminus with limited effects on function. Moreover, PhADE imaging [399] is now tractable using mKikGR-α, which may be used to track polymerase dynamics at high Pol III core concentrations (>10 nM) under both standard and challenging conditions. It should be noted that previously the C-terminal fusion α-YPet, has been used for in vivo single-molecule experiments. While growth and chromosome processing parameters were shown to be similar to wild-type E. coli K12 strain AB1157 [205,255], no effort was made to assess the implications of such a fusion on the α–τ interaction. Considering there has been no in vitro characterisation of α-YPet yet, it remains unclear whether fusion of a fluorescent protein to the C-terminus of α impairs the complex interaction networks within the Pol III HE.

In summary, the use of labelled Pol III cores described in this Chapter in future single-molecule studies will allow us to disentangle the heterogeneity and follow individual molecules throughout a DNA replication experiment. Using single-molecule based techniques will allow the examination of other questions like: “What are the polymerase dynamics at the replication fork?” and will propel us towards the goal of watching rare and short-lived states of individual replisomes in action.
Chapter 6
Single-molecule Visualisation of Fast Polymerase Turnover in the Bacterial Replisome
Single-Molecule Visualisation of Fast Polymerase Turnover in the Bacterial Replisome

Jacob S. Lewis, Lisanne M. Spenkelink, Slobodan Jergic, Nicholas P. Horan, Karl E. Duderstadt, Andrew Robinson, Nicholas E. Dixon & Antoine M. van Oijen

#These authors contributed equally.

The replisome is a multi-protein machine that has evolved to coordinate all the necessary enzymatic activities for DNA replication. Its components responsible for unwinding, priming and synthesising nucleic acids are functionally conserved across all domains of life. The *Escherichia coli* replisome has been used as a road map to uncover the design rules that enable DNA duplication with high efficiency and fidelity [430]. Although the enzymatic activities of the Pol III holoenzyme, the chromosomal replicase in *E. coli* are well understood, its dynamics within the replisome are not [7,431]. Here we test the widely accepted view that the holoenzyme remains stably associated within the replisome during replication and is efficiently recycled without dissociation during successive cycles of Okazaki fragment synthesis on the lagging strand. We use single-molecule replication assays with fluorescently labelled polymerases to demonstrate that Pol III* (holoenzyme lacking only the $\beta_2$ sliding clamp) is rapidly exchanged at replication forks during coupled leading and lagging strand DNA synthesis. Nevertheless, the replisome is highly resistant to dilution in the absence of Pol III* in solution. These observations suggest a mechanism that facilitates replacement of replisomal components dependent on their availability in the environment. This concentration-dependent dissociative mechanism provides balance between stability and plasticity.

The contents of this Chapter are published in collaboration with Prof Antoine van Oijen's laboratory at the University of Wollongong and are presented here with modifications.


I also thank Dr Harshad Ghodke for helpful discussions and Mr Enrico Monachino, Dr Yao Wang and Dr Zhi-Qiang Xu for contributing additional reagents.
6.1 Introduction

The *E. coli* replisome requires participation of thirteen different proteins, ten of which form the DNA polymerase III (Pol III) holoenzyme (HE). The HE is arranged into three functionally distinct and stably-bound subassemblies: $\alpha\beta\theta$ forms the Pol III core that has DNA polymerase activity, $\beta_2$ is the sliding clamp needed for stable association with the primer–template DNA, and $\tau_n\gamma(3-n)\delta\psi\chi$ (where $n = 2$ or $3$ in the HE) is the clamp loader complex (CLC) that loads $\beta_2$ onto DNA and is the central organiser of the replisome (Figure 6.1) [430]. The CLC interacts with two or three Pol III cores via the $\alpha-\tau$ interaction, forming stable complexes termed Pol III* (i.e., HE lacking only the $\beta_2$ sliding clamp). Pol III* ensures the organisation of the cores needed for coordinated DNA synthesis on the two template strands [286] and is essential for cell survival [356]. Although physical coupling of leading and lagging strand cores in one HE particle requires the lagging strand polymerase to undergo cycles of release and rebinding from one Okazaki fragment to the next, the molecular mechanisms underlying its cycling are still debated [301]. There is, however, consensus that Pol III is reused rather than replaced for successive Okazaki fragment synthesis [183,184,294,395]. Thus, the replisome is believed to be a highly stable entity.

**Figure 6.1 | Established view of the organisation of the *E. coli* replication fork.** The DnaB helicase encircles the lagging strand, facilitates unwinding of dsDNA through ATP hydrolysis, and recruits DnaG primase for synthesis of RNA primers that initiate synthesis of 1–2 kb Okazaki fragments on the lagging strand. The extruded single-stranded (ss) DNA is protected by ssDNA-binding protein, SSB. The Pol III HE uses the ssDNA as a template for coupled, simultaneous synthesis of new DNA on both strands. The $\beta_2$ sliding clamp confers high processivity on the Pol III HE by tethering the $\alpha\beta\theta$ Pol III core onto the DNA. The clamp loader complex (CLC) assembles the $\beta_2$ clamp onto RNA primer junctions. Up to three Pol III cores interact with the CLC through its $\tau$ subunits to form the Pol III* complex, and the $\tau$ subunits also interact with DnaB, thus coupling the Pol III HE to the helicase.

The key observations that support efficient Pol III recycling derive from *in vitro* replication assays in the absence of free polymerase [183,184], and are consistent with
the high stability of the $\alpha-\tau$ interaction that binds cores to the CLC ($K_D = 0.3 \text{ nM}; t_{1/2} = 29 \text{ min in } 300 \text{ mM NaCl}$) [17]. Nevertheless, the introduction of high concentrations of catalytically dead Pol III* (still able to bind primed DNA) inhibits ongoing replication [393] so the physiological relevance of Pol III* exchange remains uncertain. Reconciling these different observations, we here demonstrate the presence of a novel exchange mechanism that allows Pol III* to remain stably associated with the replisome under conditions of high dilution, yet facilitates rapid exchange at nanomolar concentrations.

6.1.1 The single-molecule rolling-circle assay

Using the minimal set of 12 proteins required to support coupled leading and lagging strand synthesis, we allow active replisomes to self-assemble onto pre-formed replication forks [294,395]. A 5'-flap at the end of a 7.2 kb double-stranded (ds) circular DNA substrate is anchored to the surface of a microfluidic flow cell and replication is initiated by introducing a laminar flow of buffer with the components required for coupled leading and lagging strand synthesis. Once the circular template has been replicated, the polymerase acting at the 3' end will encounter the 5' end of the original primer and synthesis will continue by displacing the previously synthesised DNA as ssDNA, this ssDNA tail will become the lagging strand in coordinated DNA synthesis (Figure 6.2a). With the lagging strand attached to the surface and the continuously growing DNA product stretched in the buffer flow, the dsDNA circle moves away from the anchor point. Flowing low concentrations (pM) of M13 DNA into the chamber leaves hundreds of DNA molecules tethered in a single field of view ($1024 \times 1024 \mu m$), each of which can serve as a substrate for the replisomal proteins introduced. Replication is visualised by real-time near-TIRF fluorescence imaging of stained dsDNA (Figure 6.2b, Appendix Figure 2). This strategy allows quantification of the rates of individual replisomes and their processivities (Figure 6.2c).
Figure 6.2 | Single-molecule rolling-circle replication assay. a, Schematic representation of the experimental design. 5’-Biotinylated M13 DNA is coupled to the passivated surface of a microfluidic flow cell through a streptavidin linkage. Addition of the E. coli replication proteins and nucleotides initiates DNA synthesis. The elongated hydrodynamically flow-stretched DNA products are stained with DNA stain (SYTOX Orange) and visualised using fluorescence microscopy. b, Kymograph of an individual DNA molecule undergoing coupled leading- and lagging-strand replication. The grey area indicates the fluorescence intensity of stained DNA. c, Single-molecule trajectory obtained from the kymograph in b is used to quantify the rates and processivities of replication events. The magenta box represents an example line segment used to determine replication rates.

6.2 Aims of the Chapter

Examples of dynamic polymerase exchange have been observed and characterised in other model systems, bacteriophages T4 and T7. Inspired by these observations, we investigated whether the replicative E. coli DNA polymerase, Pol III is reused or frequently exchanged during coupled DNA replication. This Chapter explains in detail the use of a single-molecule approach to directly visualise the dynamics of Pol III complexes at the replication fork, whereas SPR experiments demonstrate the DnaB–τ interaction is ~200-fold weaker than previously reported. This observation suggests the DnaB–τ interaction will not present an obstacle for frequent polymerase exchange. A rolling-circle DNA amplification scheme is used to observe highly processive DNA synthesis in real time, while imaging Pol III* complexes reveals they are rapidly exchanged at replication forks. On the contrary, upon dilution the replisome is stable and able to replicate hundreds of kilobases without dissociation in the absence of free Pol III*. Together, these results are used to comment on the balance between plasticity and stability of the bacterial replisome and other macromolecular complexes.
6.3 Methods

6.3.1 Protein expression and purification

*E. coli* DNA replication proteins were produced as described previously: the $\beta_2$ sliding clamp [188], SSB [166], the DnaB6(DnaC)6 helicase–loader complex [14], DnaG primase [119], the Pol III $\tau_3\delta\psi\chi$ CLC [294], and Pol III $\alpha\epsilon\theta$ core [294]. Preparation and purification of SNAP-labelled Pol III cores are described in Chapter 5. Highly purified *E. coli* Pol I and DNA ligase A were generous gifts of Dr Yao Wang (University of Wollongong). Two different biotinylated CLCs with stoichiometries of bio-$\chi\psi\tau_3\delta\psi'$, bio-$\chi\psi\tau_1\gamma_2\delta\delta'$ and components for assembly of the third: bio-$\chi$, $\psi$ and $\gamma_3\delta\delta'$ were generous gifts from Drs Slobodan Jergic and Allen Lo (University of Wollongong).

6.3.2 Bulk helicase-independent Pol III strand-displacement DNA replication assays

Conditions for the standard Pol III SD reaction were adapted from described methods [14] and are described in detail in Section 5.3.12.

6.3.3 Bulk leading- and lagging-strand DNA replication assays

Helicase-dependent coupled leading and lagging strand DNA synthesis reactions were set up in replication buffer (25 mM Tris.HCl, pH 7.9, 50 mM potassium glutamate, 10 mM Mg(OAc)$_2$, 40 $\mu$g/mL BSA, 0.1 mM EDTA and 5 mM dithiothreitol) and contained 1.5 nM of a 2-kb circular dsDNA template, 1 mM ATP, 250 $\mu$M CTP, GTP, and UTP, and 50 $\mu$M dCTP, dGTP, dATP, and dTTP, 6.7 nM wild-type or SNAP-labelled Pol III*, 30 nM $\beta_2$, 300 nM DnaG, 100 nM SSB$_4$, and 30 nM DnaB6(DnaC)$_6$ in a final volume of 13 $\mu$L. Components (except DNA) were mixed and treated at room temperature, then cooled in ice for 5 min before addition of DNA. Reactions were initiated at 30 °C and quenched after 30 min by addition of 7 $\mu$L 10% SDS and 6 $\mu$L DNA loading dye (6 mM EDTA, 300 mM NaOH, 0.25% (w/v) bromocresol green, 0.25% (w/v) xylene cyanol FF, 30% (v/v) glycerol). The quenched mixtures were loaded into a 0.6% (w/v) agarose gel in alkaline running buffer (50 mM NaOH, 1 mM EDTA). Products were separated by agarose gel electrophoresis at 14 V for 18 h. The gel was then neutralised in 1 M Tris.HCl, pH 7.6, 1.5 M NaCl and stained with SYBR Gold. Okazaki fragment length distribution was calculated by normalising the intensity as a function of DNA length.
6.3.4 Rolling-circle substrate preparation

To construct the rolling circle template [395], a 66-mer 5'-biotin-
T₃₆AATTCGTAATCATGGTCATAGCTGTTTCCT-3' oligonucleotide (Integrated DNA
Technologies) was annealed to M13mp18 ssDNA (New England Biolabs) in TBS buffer
(40 mM Tris.HCl, pH 7.5, 10 mM MgCl₂, 50 mM NaCl) at 65 °C. The primed M13 was
then extended by adding 64 nM T7 polymerase gp5 (New England Biolabs) in 40 mM
Tris.HCl pH 7.6, 50 mM potassium glutamate, 10 mM MgCl₂, 100 μg mL⁻¹ BSA, 5 mM
dithiothreitol and 600 μM dCTP, dGTP, dATP and dTTP at 37 °C for 60 min. The
reaction was quenched with 100 mM EDTA and the DNA was purified using a PCR
Purification Kit (Qiagen).

6.3.5 Coverslip surface preparation

Microscope glass coverslips are prepared as previously described [432]. Briefly, 24 ×
24 mm microscope glass coverslips (Marienfeld, Germany) are first cleaned by
sonication in a staining jar for 20 min in ethanol followed by 1 M potassium hydroxide
(repeated twice); coverslips are rinsed with MilliQ water between steps. Next,
coverslips are washed in acetone (repeated twice) to remove any traces of water and
then treated for 2–3 min with 2% (v/v) 3-aminopropyl-triethoxysilane in acetone with
gentle agitation. The reaction is quenched by addition of 2 L of MilliQ water poured
directly into the container. Coverslips are then dried with N₂ gas and baked at 110 °C
for 30 min. Now the glass is coupled to the alkoxy group of an aminosilane, creating a
surface with reactive amine groups that can subsequently be coated with a mixture of
biotinylated and non-biotinylated succinimidyl propionate-PEG. Next, methylated (M-
SPA-5000) and biotinylated (Biotin-PEF-CO₂NHS-5000) PEG are mixed in a ratio of
50–100:1, respectively in 100 mM NaHCO₃ pH 8.2 to a final concentration of 0.2%
(w/v) biotinylated PEG. This solution is used to treat a silanised coverslip for 3 h at
23 °C. Next, biotinylated functionalised coverslips are rinsed thoroughly with MilliQ
water and dried under a flow of N₂ gas.

6.3.6 Flow-cell construction

Microfluidic flow cells were prepared as described [432]. Briefly, a PDMS flow chamber
was placed on top of a PEG-biotin-functionalised microscope coverslip that has been
incubated with 1 mg mL⁻¹ streptavidin in buffer containing 10 mM phosphate, 2.7 mM
potassium chloride and 137 mM sodium chloride at pH 7.4. To help prevent non-
specific interactions of proteins and DNA with the surface, the chamber was blocked
with buffer containing 20 mM Tris.HCl, pH 7.5, 2 mM EDTA, 50 mM NaCl, 0.2 mg mL\(^{-1}\) BSA, and 0.005% (v/v) Tween-20.

6.3.7 TIRF microscope setup

The chamber was placed on an inverted microscope (Nikon Eclipse Ti-E) with a CFI Apo TIRF 100× oil-immersion TIRF objective (NA 1.49, Nikon) under control of NIS-elements AR v4.3 (Nikon) and connected to a syringe pump (Adelab Scientific) for flow of buffer. Double-stranded DNA was visualised in real-time by staining it with 150 nM SYTOX Orange (Invitrogen) excited by a 568-nm laser (Coherent, Sapphire 568-200 CW) at 150 \(\mu\)W cm\(^{-2}\). The red and green Pol III* were excited at 700 mW cm\(^{-2}\) with 647 nm (Coherent, Obis 647-100 CW) and 488 nm (Coherent, Sapphire 488-200 CW) lasers, respectively. The signals were separated via dichroic mirrors and appropriate filter sets (Chroma). Imaging was done with an either an EMCCD (Photometics, Evolve 512 Delta) or sCMOS camera (Andor, Zyla 4.2).

6.3.8 Single-molecule rolling-circle replication reactions

Conditions for coupled DNA replication under continuous presence of all proteins were adapted from previously described methods [294,395]. Briefly, 30 nM DnaB\(_6\)(DnaC)\(_6\) was incubated with 1.5 nM biotinylated ds M13 substrate in replication buffer (25 mM Tris.HCl, pH 7.9, 50 mM potassium glutamate, 10 mM Mg(OAc)\(_2\), 40 \(\mu\)g mL\(^{-1}\) BSA, 0.1 mM EDTA and 5 mM dithiothreitol) with 1 mM ATP at 37 °C for 30 s. This mixture was loaded into the flow cell at 100 \(\mu\)L min\(^{-1}\) for 40 s and then at 10 \(\mu\)L min\(^{-1}\). An imaging buffer was made with 1 mM UV-aged Trolox, 0.8% (w/v) glucose, 0.12 mg mL\(^{-1}\) glucose oxidase, and 0.012 mg mL\(^{-1}\) catalase (to increase the lifetime of the fluorophores and reduce blinking), 1 mM ATP, 250 \(\mu\)M CTP, GTP, and UTP, and 50 \(\mu\)M dCTP, dGTP, dATP, and dTTP in replication buffer. Pol III* was assembled in situ by incubating \(\tau_3\delta\delta'\chi\psi\) (410 nM) and SNAP-labelled Pol III cores (1.2 \(\mu\)M) in imaging buffer at 37 °C for 90 s. Replication was initiated by flowing in the imaging buffer containing 6.7 nM Pol III* (unless specified otherwise), 30 nM \(\beta_2\), 300 nM DnaG, 250 nM SSB\(_4\), and 30 nM DnaB\(_6\)(DnaC)\(_6\) at 10 \(\mu\)L min\(^{-1}\). Reactions were carried out at 31 °C, maintained by an electrically heated chamber (Okolab, Italy).

Conditions for the pre-assembly replication reactions were adapted from published methods [52,184]. Solution 1 was prepared as 30 nM DnaB\(_6\)(DnaC)\(_6\), 1.5 nM biotinylated ds M13 substrate and 1 mM ATP in replication buffer. This was incubated at 37 °C for 3 min. Solution 2 contained 60 \(\mu\)M dCTP and dGTP, 6.7 nM red Pol III*,
and 74 nM $\beta_2$ in replication buffer (without dATP and dTTP). Solution 2 was added to an equal volume of solution 1 and incubated for 6 min at 37 °C. This was then loaded into the flow cell at 100 $\mu$L min$^{-1}$ for 1 min and then 10 $\mu$L min$^{-1}$ for 10 min. The flow cell was washed with replication buffer containing 60 $\mu$M dCTP and dGTP. Replication was finally initiated by flowing in the imaging buffer containing 60 $\mu$M of each dNTP, 50 nM $\beta_2$, 300 nM DnaG and 250 nM SSB$_4$ at 10 $\mu$L min$^{-1}$.

6.3.9 Data analysis

The analysis was done with ImageJ using in-house built plugins. The rate of replication of a single molecule was obtained from its trajectory and calculated for each segment that had constant slope. Processivity was determined by de-convolving the length of the M13 DNA template from the final length of the extended DNA and converted to bp using the known length of tethered λ-phage DNA (48.5 kb) with our cameras and using the calibrated pixel size in bp, where 1 pixel = 214.5 bp for sCMOS and 470 bp for EMCCD cameras.

To determine the stoichiometry of Pol III* at the replisome, the average intensity of a single labeled Pol III core (6 pM) was calculated by immobilization on the surface of a cleaned microscope coverslip in imaging buffer. The imaging was under the same conditions as used during the single-molecule rolling-circle experiments. Using ImageJ with in-house built plugins, we calculated the integrated intensity for every Pol III core in a field of view after applying a local background subtraction. The histograms obtained were fit with a Gaussian distribution function using MATLAB 2014b, to give a mean intensity of 5100 ± 2000 units for the red Pol III core and 1600 ± 700 units for the green Pol III core (Figure 6.7e). To measure the intensity of the fluorescent spot at the replication fork, we tracked its position and integrated the intensity for both colors simultaneously over time. Given there is no decay in fluorescence intensity of labeled Pol III cores as a function of DNA length under near-TIRF imaging conditions during DNA replication (Appendix Figure 3), we calculated the total number of Pol III*’s at every time point during coupled DNA replication by dividing these intensities by the intensity of a single Pol III*. Subsequent histograms were fit to four (6.7 nM) or three (0.3 nM) Gaussians centred at integral numbers of Pol III* (Appendix Figure 4) using MATLAB 2014b (Mathworks).
6.3.10 SPR (BIAcore) measurements of the interactions between DnaB and immobilised clamp loader complexes

A BIAcore T200 instrument (GE Healthcare) was used to study the interactions of DnaB with various CLCs. All studies were carried out at 20 °C and data collected at 10 Hz. A streptavidin-coated sensor chip (GE Healthcare) was activated with three sequential injections of 1 M NaCl, 50 mM NaOH (1 min each at flow rate of 5 μL min⁻¹).

A solution of 6.8 nM bio-χψτδδ' clamp loader was immobilised using biotin–streptavidin interaction in SPR buffer (25 mM Tris.HCl, pH 7.6, 0.25 mM DTT, 50 mM NaCl, 0.2 mM ATP 5 mM MgCl₂ and 0.005% (v/v) P20) at 5 μL min⁻¹ to a density of 1450 RU on one flow cell. On another channel a solution of 6 nM bio-χψτγδδ' clamp loader was immobilised at 5 μL min⁻¹ to a density of 1400 RU. The other two flow cells were left unmodified to serve as controls for buffer subtraction. DnaB interactions with bio-χψτδδ' and bio-χψτγδδ' were carried out by sequential injections of solutions containing DnaB in SPR injection buffer (25 mM Tris.HCl, pH 7.6, 0.25 mM DTT, 50 mM NaCl, 1 mM ADP 5 mM MgCl₂ and 0.005% (v/v) P20), with a zero and nine concentrations of serially diluted samples; 1600, 800, 400, 200, 100, 50, 25, 12.5, and 6.25 nM (as hexamer) at a flow rate of 30 μL min⁻¹ for 30 s. Dissociation in the same buffer was measured over 500 s.

In a separate experiment a lower [NaCl] buffer was used to reassess interactions by stimulating binding of CLC including γ₃ CLC to DnaB. It has been previously reported that domain IV of τ is responsible for binding to DnaB [37], therefore the γ₃ CLC should not interact. SPR measurements were carried out as above but using a low [NaCl] SPR buffer (25 mM Tris.HCl pH 7.6, 0.25 mM DTT, 25 mM NaCl, 0.2 mM ATP, 5 mM MgCl₂ and 0.005% (v/v) P20). First the Bio-χψτδδ' was immobilised to a density of 1200 RU on one flow cell. DnaB was injected in a low [NaCl] SPR injection buffer (25 mM Tris.HCl, pH 7.6, 0.25 mM DTT, 25 mM NaCl, 1 mM ADP, 5 mM MgCl₂ and 0.005% (v/v) P20) over all flow cells using the same concentration series and association and dissociation times as described above. Conversely, the γ₃ CLC was assembled in situ on the SPR chip surface on another flow cell (Figure 6.9a). First, a solution of 10 nM Bio-χ was injected in low NaCl SPR buffer at 5 μL min⁻¹ to a density of 240 RU; next two sequential injections of 50 nM ψ in the same buffer at 5 μL min⁻¹ saturated all immobilised χ's (177 RU). Finally, a solution of 66 nM γδδ' complex was injected over immobilised bio-χψ at 5 μL min⁻¹ to a density of 1600 RU. DnaB was injected at 1 μM
in low [NaCl] SPR injection buffer at a flow rate of 30 μL min\(^{-1}\) for 40 s, followed by 500 s dissociation time in the same buffer. As before, another unmodified flow cell served as a control.

Interactions between DnaB and the various clamp loader complexes exhibited fast kinetics, so only equilibrium data could be extracted from the sensorgrams. The equilibrium dissociation constants (\(K_D\)) describing interactions were obtained using BIAevaluation software v3.1 (GE Healthcare) by fitting the responses at equilibrium against concentration using the steady-state one-to-one Langmuir binding model.
6.4 Results

6.4.1 Comparison of activities of wild-type and SNAP-labelled Pol III cores

In Chapter 5, I described the production of fluorescently labelled Pol III α subunit following its fusion to a SNAP-tag, separately labelled in >80% yields to red and green fluorophores (Section 5.4.3) that were then reconstituted from individual SNAP-α, ε and θ subunits and isolated chromatographically (Figure 6.3a–d) [294]. To compare the activities of wild-type with fluorescently labelled Pol III cores, labelled Pol III cores were assembled into single-colour Pol III*s in situ with separately-isolated τ3-CLC and subjected to helicase independent Pol III SD DNA replication assays [14]. Labelled Pol III*s were active in SD assays with all starting templates being fully replicated, with SD products evident after 1.5 min, consistent with wild-type Pol III* (Figure 6.3e). While the data clearly show labelled Pol III* participate in SD synthesis, this does not yet establish their activity in coupled DNA replication. To demonstrate this, we used an additional coupled DNA replication assay. The labelled Pol III*s were also active in coupled DNA replication, producing Okazaki fragments of similar sizes to wild-type polymerases (Figure 6.3f).

6.4.2 Real-time fluorescence imaging of labelled Pol III during coupled DNA replication

Real-time single-molecule observation of rolling-circle DNA replication was used to characterise the dynamics by which individual polymerases associate and dissociate during replication. Simultaneous imaging of the stained DNA and red Pol III* shows that the polymerase spot is located at the tip of the growing DNA, confirming that the labelled Pol III is a functional component of reconstituted replisomes (Figure 6.4a). A kymograph (Figure 6.4b) shows the fluorescence of the red Pol III* during rolling-circle replication; it supports replication at rates similar to the untagged wild-type enzyme (Figure 6.4c).
Figure 6.3 | Purification and comparison of activities of wild-type and SNAP-labelled Pol III cores. a, SDS-PAGE of final fraction from the $\tau_{C_{16}}$ affinity chromatography. b and c, SDS-PAGE of pooled fractions of SNAP labelled Pol III cores; a representative chromatography profile is shown in d. e, Agarose gel of products of Pol III SD DNA synthesis, a demanding assay for Pol III* activity [14]. The time course of flap-primer extension on M13 ssDNA depicts products larger than unit length of dsDNA (TFII) generated by Pol III SD DNA synthesis. f, Alkaline agarose gel of coupled DNA replication. Reactions were performed on a 2-kb circular dsDNA template with wild-type (w.t.) Pol III*, w.t. Pol III* + SYTOX Orange, red SNAP-labeled Pol III*, and green SNAP-labelled Pol III*. (Left panel) The gel was stained with SYBR-Gold. (Right panel) Intensity profiles of lanes 2–5 of the left panel. The Okazaki fragment size distribution is centred at $1.3 \pm 0.4$ kb. Intensity profiles have been corrected for the difference in intensity of staining of different size fragments using the ladder as a standard.
Figure 6.4 | Real-time fluorescence imaging of coupled DNA replication. a, Representative kymograph of simultaneous imaging of double-stranded DNA stained with SYTOX Orange (grey scale) and of Pol III labelled with a red fluorophore (magenta) in real-time. The kymograph demonstrates the fluorescent spot corresponding to Pol III co-localises with the tip of the growing DNA product (evident as a white spot) where the replication fork is located. The arrow indicates the direction of flow. b, Kymograph of the distribution of red-labelled Pol IIIs on an individual DNA molecule. The Pol III moves with the replisome in the direction of flow as it elongates the DNA, visible as a bright magenta spot moving away from the surface anchor point. Additional Pol IIIs are left behind the moving replisome, seen as horizontal lines on the kymograph. c, Histograms of the rate of replication for wild-type Pol III (492 ± 23 bp s⁻¹) and red Pol III (561 ± 27 bp s⁻¹) fit to Gaussian distributions. d, Kymograph of the distribution of red Pol III on an individual DNA molecule in the presence of 150 nM Pol I and 100 nM DNA ligase. Prolonged Pol III spots behind the replisome are no longer observed due to the action of Pol I in Okazaki fragment processing. e, Fluorescence intensity as a function of time of individual red Pol IIIs immobilised on the surface of a coverslip (lower trace; black line is an exponential fit with lifetime = 14.1 ± 0.4 s), and of the replisomal spot in b (upper trace). The fluorescence lifetime of red Pol III at the replisome is much longer than the photobleaching lifetime of the dye at 700 mW cm⁻². The errors represent the standard errors of the mean.

We also observe Pol III that remains bound to the DNA behind the replisome, evident as horizontal lines in Figure 6.4b. We reasoned that these correspond to its retention at the 3'-termini of some Okazaki fragments and repeated the experiment in the presence of Pol I and/or DNA ligase; Pol I replaces RNA primers with DNA and ligase seals the remaining nick. In the presence of Pol I (with or without ligase), Pol III binding behind the replisome is no longer observed (Figure 6.4d), consistent with Pol I efficiently displacing Pol III during Okazaki fragment maturation. Surprisingly, the fluorescent Pol
III at the growing tip of the rolling circle is highly resistant to photobleaching. Its fluorescence in the replisome has a much longer lifetime compared to that of labelled Pol III cores immobilised on a surface and subjected to the same excitation intensity (Figure 6.4e). Since the experiments in Figure 6.4 are performed with 6.7 nM Pol III* in solution, this observation suggests that the polymerase exchanges into the replisome from solution to replace photobleached Pol III.

6.4.3 Pre-assembled Pol III* complexes do not exchange Pol III cores

To characterise the dynamic behaviour of Pol III at the fork and directly visualise its exchange in real time, we used mixtures of red and green Pol III*s. To demonstrate that green Pol III cores in a Pol III* complex do not exchange with the red ones from another Pol III*, we combined them in a 1:1 ratio for 30 min at 37 °C (Figure 6.5 bottom right panel), and imaged the mixture on the surface of a coverslip at the single-molecule level (Figure 6.5). Consistent with the stable interaction between α in the core and τ in the CLC, exchange of Pol III cores was not observed.

![Figure 6.5](image_url) | Pre-assembled Pol III* complexes do not exchange Pol III core. (Bottom right) Red and green Pol III* are separately pre-assembled by treatment at 37 °C for 15 min (30 nM Pol III core and 10 nM τ3-CLC). These are then mixed in equal ratios and kept at 37 °C for 1 hour prior to dilution to 6 pM Pol III* for imaging. (Top left) Red Pol III* complexes; (middle right) green Pol III* complexes; (bottom left) merged red and green channels. No co-localization is observed, demonstrating the α–τ interaction within the Pol III* complex remains intact for the duration of the DNA replication assays. (Middle right) Red and green Pol III cores are mixed before adding the CLC (30 nM Pol III core and 10 nM τ3-CLC). Pol III* is formed by treatment at 37 °C for 15 min. Complexes are then allowed to equilibrate for 1 hour at 37°C prior to dilution to 6 pM for imaging. (Middle bottom) Red and green Pol III cores co-localize (white spots). No co-localisations are observed, demonstrating the α–τ interaction within the Pol III* complex remains intact for the duration of the DNA replication assays. White scale bars represent 5 μm.
6.4.4 Pol III* readily exchanges into the replisome at a rate that is dependent on its concentration

We visualised exchange of Pol III* at the replication fork by measuring the fluorescence intensity at the replisome spot as a function of time using 1:1 mixtures of red and green Pol III*s. At a total Pol III* concentration of 6.7 nM (Figure 6.6a), the replisomal spot exhibits fast dynamics displaying both colours, while at a lower concentration of 0.3 nM (Figure 6.6b), the dynamics appear less prominent and distinct exchange events are visible. The longer persistence of a single colour at the lower concentration demonstrates that Pol III* exchange is concentration dependent. Given that Pol III* remains intact on time scales much longer than the duration of our experiment, these observations can only be explained by wholesale exchange of Pol III* at the replication fork.

![Figure 6.6](image)

**Figure 6.6 | Rapid and frequent exchange of Pol III* is concentration dependent.** a and b, Kymographs of the distributions of red Pol III* (magenta) and green Pol III* (green) on an individual DNA molecule at a total Pol III* concentration of 6.7 (a) or 0.3 nM (b). Co-localisation of the two signals is shown as a bright white fluorescent spot.

Our demonstration of rapid exchange of entire Pol III*s at the replication fork, however, does contradict observations that both leading and lagging strand Pol III cores remain stably associated during coupled DNA replication [183,184,294,395]. Those studies used assays in which replisomes were assembled, replication initiated, and the reactions rapidly diluted to measure the stability of synthesising replisomes on DNA. To reconcile our observations of dynamic exchange of Pol III* with the previous work, we carried out single-molecule pre-assembly replication assays [183,184] using the red Pol III*. In this experiment, the replisome is pre-assembled onto the rolling-circle template in solution. Subsequently, the template is attached to the surface of a flow cell, which is then washed to remove all unbound proteins. Replication is initiated by introduction of a replication solution that omits Pol III* and helicase. Since absence of free Pol III* in solution makes polymerase exchange impossible, we hypothesised that Pol III would be recycled within the replisome, enabling its sustained participation in processive DNA replication.
Figure 6.7 | Pre-assembled replisomes participate in processive DNA replication. 

a, Kymograph of a pre-assembled replisome containing red Pol III*. The intensity of the signal from the replisomal spot decreases after a Pol III* is left behind. It subsequently bleaches and the signal does not recover. 

b, Histograms of the rates of replication with Pol III* present in solution ($561 \pm 27$ bp s$^{-1}$) and under pre-assembly conditions ($445 \pm 195$ bp s$^{-1}$), each fit to a Gaussian distribution. 

c, Histograms of the processivity of replication with Pol III* present in solution ($73 \pm 25$ kb) and under pre-assembly conditions ($76 \pm 26$ kb), each fit with a single exponential decay function. 

d, Histograms for conditions with pre-assembled replisomes (no polymerases in solution; processivity = $76 \pm 26$ kb) and under conditions where pre-assembled replisomes are challenged with 10 nM Pol III core ($3.5 \pm 0.6$ kb), each fit with a single exponential decay function. The data show that actively replicating Pol III* can be easily displaced when challenged with entities that bind to the clamp, but cannot support coordinated leading and lagging strand synthesis. The errors represent the errors of the fit. 

e, Histograms of the intensity distribution of single Pol III cores; histograms are fit with Gaussian distribution functions to give a mean intensity of $5100 \pm 2000$ for the red Pol III core and $1600 \pm 700$ for the green Pol III core. 

f, Histograms of the stoichiometry of Pol III* at the replication fork. The histograms are fit to four (6.7 nM) or three (0.3 nM) Gaussians centred at integral numbers of Pol III* calculated from single Pol III core intensities (Appendix Figure 4). The black lines represent the sums of these distributions. All uncertainties represent the standard deviation.

These conditions indeed support highly processive DNA replication (Figure 6.7a), with synthesis rates and processivities identical to a situation with Pol III* in solution (Figure 6.7b,c). Further, we observed photobleaching without recovery, consistent with the original, pre-assembled Pol III* remaining stably associated within the replisome. As further confirmation of the robustness of the pre-assembled replisome in the absence of competing polymerases and the easy displacement of Pol III* upon challenge, we initiate replication by pre-assembly of replisomes, normally supporting highly processive synthesis, and challenge them with Pol III core. The observation of a sharp
reduction in processivity is consistent with the displacement of the Pol III* from the replication fork by the Pol III cores, which are unable to support coordinated leading- and lagging-strand synthesis [286] (Figure 6.7d). Next, we quantified the number of Pol III*s at the replication fork by normalising the fluorescence intensity of the replisomal spot to the intensity of a single Pol III* calculated from single Pol III core intensities (Figure 6.7e). The peaks of the distributions are at one Pol III* per replisome (Figure 6.7f), consistent with *in vivo* observations [205].

6.4.5 **Assessment of the DnaB–clamp loader complex interaction by SPR (BIAcore)**

It is reported that in the presence of the τ subunit, DnaB at the replication fork moves at least ten-fold faster [433]. Proteolysis and kinetic binding studies (using SPR) have determined that the small 8 kDa domain IV of τ is responsible for mediating the interaction between τ and DnaB. It was shown that binding of DnaB from solution to surface immobilised τ_{C24} was dependent on density of immobilised ligand, suggesting that this binding mechanism is multivalent [37] with an apparent $K_D$ value of 5 ± 0.4 nM. This implies that true one-to-one binding affinities between DnaB and τ may be too weak to be observed under the experimental conditions. Due to the incompatibility of Pol III* exchange with the strong τ–DnaB interaction and slow dissociation, we have reassessed the τ–DnaB interaction by measuring physiologically relevant binding between CLCs containing full-length τ and DnaB. Here, work from the McHenry laboratory is confirmed and extended by SPR using a BIAcore T200 instrument by examination of various CLCs binding to the DnaB helicase.

First bio-$\chi_{\psi}\tau_3\delta^\prime$ and bio-$\chi_{\psi}\tau_1\gamma_2\delta^\prime$ were immobilised on the streptavidin coated chip surface to 1450 and 1400 RU, respectively. Addition of the nucleotide (ATP) to the SPR running buffer aids to stabilise the CLC on the surface; more specifically it stabilises the $\tau_{1-n}\delta^\prime$ ($n = 1–3$) on the surface as the $\chi_{\psi}$ interaction is extremely strong, taking days to dissociate (Dr Slobodan Jergic, unpublished). Next, binding of DnaB to bio-$\chi_{\psi}\tau_3\delta^\prime$ CLC was measured by injecting DnaB at eight concentrations of serially diluted samples; 8, 4, 2, 1, 0.5, 0.25, 0.125, and 0.063 μM (as hexamer) plus a blank. Binding at equilibrium was used to derive the $K_D$ considering the fast on/fast off kinetics of the interaction. The sensorgrams were fit with a steady-state one-to-one Langmuir binding model (Figure 6.8b inset, Table 6.1). Following this, binding of DnaB to the bio-$\chi_{\psi}\tau_1\gamma_2\delta^\prime$ CLC was measured by injecting DnaB at six concentrations of serially diluted samples; 8, 4, 2, 1, 0.5, and 0.25 μM (as hexamer) plus a blank. Binding at equilibrium
was used to derive the $K_D$ by fitting sensorgrams as before (Figure 6.8c inset; Table 6.1). Thus, in the context of the CLC, DnaB binding exhibits fast on/ fast off kinetics. This is inconsistent with previous measurements of the $\tau$–DnaB interaction that reported a stable interaction, whereby the complex can be isolated by gel filtration or where the $K_D$ of the $\tau_{C24}$–DnaB interaction is in the low nM range [434].

Figure 6.8 | SPR assessment of DnaB–CLC interactions: CLC containing $\tau_3$ binds stronger to DnaB than CLC containing $\tau_1\gamma_2$ in buffer containing 50 mM NaCl. a, Experimental design. Biotinylated $\tau_3$ CLC was immobilised on the SPR chip surface, then DnaB is injected over the surface at varying concentrations. b, Sensorgrams show association and dissociation phases of DnaB–bio-$\chi_{\psi\tau\ddot{\delta}\delta'}$ interactions at a range of serially-diluted DnaB concentrations. Responses at equilibrium were fit using a steady-state one-to-one Langmuir binding model (inset). The derived $K_D$ was $1.29 \pm 0.18 \mu M$. c, Sensorgrams show association and dissociation phases of DnaB–bio-$\chi_{\psi\tau\gamma_2\ddot{\delta}\delta'}$ interactions at a range of serially-diluted DnaB concentrations. Responses at equilibrium were fit as in a (inset); the derived $K_D$ was $4.06 \pm 0.29 \mu M$. Grey boxes shows the average response values used for calculation of $K_D$. A summary of derived thermodynamic parameters is shown in Table 6.1.

Next we assembled the bio-$\chi_{\psi\gamma_3\ddot{\delta}\delta'}$ CLC in situ on the SPR chip surface through sequential interactions of components (Figure 6.9a). We observed that even at 1 $\mu M$ DnaB, no binding to the $\gamma_3\ddot{\delta}'$ CLC was detected (Figure 6.9a). This result confirms that the unique C-terminus of the $\tau$ subunit is important for binding to DnaB, although it is
still unclear if oligmerisation of τ is specifically required for this interaction to occur. Moreover, it is probable the interaction is dominated by electrostatic contacts, as reduction of the NaCl concentration from 50 to 25 mM increases the binding affinity of τ₃ containing CLC 3.6-fold from 1.29 ± 0.18 to 0.35 ± 0.07 μM (Figure 6.9b, Table 6.2). Whereas, at 200 mM NaCl the interaction cannot be seen.

![Figure 6.9](image)

**Figure 6.9 | The γδδ³' CLC does not bind DnaB.** **a.** In situ assembly of immobilised bio-χΨτδδ³' CLC on the SPR chip surface. First biotinylated χ (10 nM) was injected to a density of 240 RU, followed by two sequential injections of ψ (50 nM) in the same buffer to saturate χ to ~177 RU. Finally, the γδδ³' complex (66 nM) was introduced to a saturating density of 1600 RU. Injection of DnaB at 1 μM did not illicit a response on immobilised Bio-χΨτδδ³' CLC (inset; blue region). **b.** Sensorgrams show association and dissociation phases of DnaB–bio-χΨτδδ³' interactions of serially-diluted DnaB concentrations in buffer containing 25 mM NaCl. Responses at equilibrium for the faster process in biphasic interaction were fit using a steady state one-to-one Langmuir binding model (inset). The derived $K_D$ was 351 ± 73 nM. It is clear that the DnaB–CLC interaction is extremely sensitive to ionic strength; previously measured data at 50 mM NaCl showed a $K_D$ of 1.29 ± 0.18 μM whereas at 200 mM NaCl the interaction cannot be seen. The grey box shows average response values used for calculation of $K_D$. A summary of derived thermodynamic parameters is shown in Table 6.2.
Table 6.1 | Equilibrium binding constants describing interactions between DnaB and immobilised CLCs in buffer containing 50 mM NaCl. All uncertainties are standard errors in parameters from fitting of the complete data set as indicated.

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>BIACore T200</td>
</tr>
<tr>
<td>Acquisition rate (Hz)</td>
<td>10</td>
</tr>
<tr>
<td>Sensor chip</td>
<td>Sensor chip SA</td>
</tr>
<tr>
<td>Flow rate (μL min⁻¹)</td>
<td>30</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>20</td>
</tr>
<tr>
<td>Association time (s)</td>
<td>30</td>
</tr>
<tr>
<td>Dissociation time (s)</td>
<td>500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DnaB–bio-χψττδδ′</th>
<th>DnaB–bio-χψττδδ′</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K_D (μM)</strong></td>
<td>1.29 ± 0.18</td>
</tr>
<tr>
<td><strong>R_max (RU)</strong></td>
<td>437 ± 20</td>
</tr>
</tbody>
</table>

Table 6.2 | Equilibrium binding constants describing interactions between DnaB and immobilised CLCs in buffer containing 25 mM NaCl. All uncertainties are standard errors in parameters from fitting of complete data set as indicated.

<table>
<thead>
<tr>
<th>Data collection parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Instrument</td>
<td>BIACore T200</td>
</tr>
<tr>
<td>Acquisition rate (Hz)</td>
<td>10</td>
</tr>
<tr>
<td>Sensor chip</td>
<td>Sensor chip SA</td>
</tr>
<tr>
<td>Flow rate (μL min⁻¹)</td>
<td>30</td>
</tr>
<tr>
<td>Temperature (°C)</td>
<td>20</td>
</tr>
<tr>
<td>Association time (s)</td>
<td>40</td>
</tr>
<tr>
<td>Dissociation time (s)</td>
<td>500</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DnaB–bio-χψττδδ′</th>
<th>DnaB–bio-χψττδδ′</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>K_D (nM)</strong></td>
<td>351 ± 73</td>
</tr>
<tr>
<td><strong>R_max (RU)</strong></td>
<td>580 ± 46</td>
</tr>
</tbody>
</table>
6.5 Discussion

This Chapter presents imaging of the reconstituted \textit{E. coli} DNA Pol III HE at the single-molecule level to study the dynamics of labelled DNA polymerases at the replication fork and SPR assessment of the $\tau$–DnaB interaction in the context of various CLCs. We show rapid cycling of Pol III*s in and out of the replication fork during coupled DNA replication, while SPR data strongly support Pol III* exchange as the $\tau$–DnaB interaction would not be an obstacle to efficient exchange. Both findings therefore, challenge and add to the classical textbook picture of the single \textit{E. coli} DNA Pol III HE being stably bound to the helicase at the replication fork and performing virtually all the DNA synthesis. Rather, we propose the \textit{E. coli} replisome strikes a balance between stability and plasticity, which complies with established chemical principles rather than our natural perception of the beauty of simplicity.

6.5.1 Plasticity vs. stability

We propose that the plurality of weak interactions that stabilise the HE within the replisome (DnaB) plays a key role in mediating rapid exchange while preventing the replisome from dissociating. For instance, in the absence of Pol III* in solution, the replisome retains its original polymerases and forms a highly stable complex resistant to dilution. Pol III* readily exchanges into the replisome when continually present at a rate that is dependent on its concentration. Such a concentration dependent dissociative mechanism seems counterintuitive, but can be rationalised through a complex protein–protein and protein–DNA interaction network controlled and maintained by multiple dynamic interactions. Under dilute conditions, transient disruption of any one of these interactions would be followed by its rapid re-formation, preventing dissociation. If, however, there are competing Pol III*s in close proximity to the fork, one of these can bind at a transiently vacated binding site (\textit{e.g.}, on the $\beta_2$ sliding clamp or DnaB helicase) and consequently be at a sufficiently high local concentration to compete out the original Pol III* for binding to the other sites. As its concentration increases, the binding equilibria are pushed towards occupancy of all binding sites and more than one Pol III* is associated with the replisome. At lower concentrations, Pol III* still exchanges, but the average number of Pol III*s is reduced. Further evidence of this mechanism comes from comparison of the number of Pol III*s in or near the replisome at different concentrations. Quantification of the number of Pol III*s at the replication fork by normalising the fluorescence intensity of the replisomal spot to the intensity of a single Pol III* shows the peak of the distributions at one Pol III* per replisome. Nevertheless, we find that often more than one Pol III* is present in the
Additionally, SPR experiments presented here, examining the DnaB–CLC interaction suggest that it is ~200-fold weaker than previously reported [37], and requires the unique C-terminus of τ, presumably domain IV. The transient nature of this interaction (fast on/fast off kinetics) likely plays a critical role within the replisome. It was shown that the τ–DnaB complex is amenable to gel filtration [434]; however, the present results suggest that this would not be possible at the concentrations used. Data from Kim et al., [434] may have been misinterpreted as a consequence of various oligomeric states of τ (including τ₄) that could be co-eluting with DnaB. Since the τ subunit interacts with both Pol III core and the DnaB helicase, multiple copies of the CLC containing the τ and attached Pol III core might bind DnaB helicase and to a Pol III* repository awaiting incorporation into the replication fork (as proposed in the T7 system [429]), allowing for efficient polymerase exchange at the replication fork. Such concentration dependent exchange has recently been reported for other systems [392,396,435-437] and mathematically described by multisite competitive exchange mechanisms [438,439], suggesting this mechanism may be important for the functioning and maintenance of other macromolecular machineries.

6.5.2 Perspectives and implications of a dynamic replisome

This study paints a messier picture than the textbook models indicate, but is consistent with the pragmatism of evolution to select for a balance between plasticity and efficiency. Such properties predict that future in vivo single-molecule experiments investigating polymerase dynamics will detect various levels of exchange at progressing replication forks that will depend merely on the availability (or not) of polymerases in the environment. Hence, it will be important to design experimental approaches that are able to distinguish between Pol III and Pol III*. This will be challenging as other exogenous Pol IIIs required for other processes associated with the nascent replication product, such as mismatch repair, may also be present. Previous studies by the McHenry laboratory showed that free Pol III core cannot exchange with Pol III at the replication fork [393,440]. Thus, involvement of the CLC in Pol III* exchange indicates that interactions via the τ-containing CLC are important for polymerase exchange. A simple explanation for this would be that domain IV and/or V unique to τ interacts with a replicating Pol III HE and triggers exchange, possibly by chaperoning Pol III off nascent Okazaki fragments when a new primer becomes available [304]. Alternatively, direct contacts between the invading Pol III*, whether through direct protein–protein or protein–DNA–protein interactions with the “to be
displaced" Pol III HE might be required. Although Pol IV displaces Pol III from the β2 clamp at concentrations corresponding to those relevant during the SOS DNA damage response aided by a secondary rim contact [229], a mutant Pol IV lacking its CBM and thus a secondary binding site on the rim of β2 did not reduce the Pol III processivity (unlike its wildtype counterpart). Efficient exchange occurs at higher concentrations (as in SOS induced conditions), and shared interactions between Pol IV and a component of Pol III*, such as a single cleft of β during competition between Pol III and Pol IV bound at additional sites facilitate dissociation and subunit exchange. It is uncertain whether exchange occurs through distinctly ordered or simultaneous interaction pathway. It is reasonable to think such mechanisms could induce conformational changes in Pol III* to make one of the two mutual binding pockets in β2 available. Alternatively, the mechanism of Pol III* exchange may differ from the pathway used by Pol IV, consistent with different protein requirements for exchange [393].

The concept of a dynamic replisome should be considered in the context of robust replication in a cellular environment where success not only depends not only on the efficiency and accuracy of duplication, but also on the capacity of the replication machinery to navigate obstacles on the chromosome. Evolution of a highly dynamic replisome that is able to exist as a stable entity with rapidly exchangeable components allows for defective components to be easily replaced from solution and provides access to other potential binding partners, e.g., translesion synthesis polymerases like Pol IV [441], without violating fundamental chemical and thermodynamic principles. The discovery of new mechanisms that rely on plasticity to allow replisomes to overcome obstacles during the cell cycle [442,443], such as a leading strand lesion [444] have solidified the importance of future work to elucidate the balance between polymerase stability and plasticity during DNA synthesis. Regardless of the mechanism, further advances in understanding the structural basis of the τ–DnaB interaction will help illuminate possible exchange mechanisms and their roles during elongation.

### 6.5.3 Future work

Finally, future work with the goal to understand the molecular mechanism of Pol III* exchange at the replication fork in *E. coli* should progress in three main directions:

(i) Biochemical and structural studies to identify the exact amino acids/regions involved in the DnaB–τ interaction.

(ii) Quantification of timescales of Pol III* exchange, which may be extended by comparing exchange timescales with fluorescently labelled CLCs and,
(iii) Assessment of labelled Pol III cores and CLCs using single-molecule *in vivo* approaches, with a broader goal to bridge the gap in knowledge between *in vitro* and *in vivo* replisome dynamics.

Experiment (i) will be a significant investment in resources. However, the novel SPR assay described in Section 6.4.5 may be used to test mutants of $\tau$ generated to identify residues that participate in binding to DnaB. Experiment (ii) will require an efficient methodology for quantification of exchange timescales that is yet to be developed. Previously this has been done by de-convoluting random noise from repeated patterns in fluorescence intensity signals using an autocorrelation function [396]. This approach could provide a great starting point. Also, fluorescent labelling of a subunit of the CLC, possibly via $\delta'$, would be useful for future assessments. The $\delta'$ subunit may be an efficient fluorescent marker for the CLC compared to already existing *dnaX*-fusions as only one subunit is incorporated into CLCs regardless of *dnaX* stoichiometry and the polypeptide chain is not subject to extensive proteolysis during isolation, as is the $\tau$ subunit [37].

The incorporation of *in vivo* studies will present a unique set of challenges. While some variants of *E. coli* carrying fluorescent protein gene fusions to replication proteins on their chromosome [205,255] already exist, others will need to generated and characterised (e.g. genetic fusion of the *holB* gene encoding the $\delta'$ subunit). Considering most genetic fusions have not been rigorously tested using *in vitro* replication assays, it is hard to make accurate comparisons and de-convolve heterogeneity between *in vitro* and *in vivo* experiments. Further, other influences such as unwanted side reactions from other DNA metabolism processes like DNA repair and recombination within the cell may also affect replisome dynamics. Moreover, measurement of polymerase exchange rates at forks *in vivo* could provide a measure of the concentration of free Pol III$^*$ in their vicinity. Nonetheless, establishing the general rules and mechanisms that govern polymerase exchange promises to be a challenging, yet fruitful area of investigation.
Chapter 7

Concluding Summary
The *E. coli* DNA Pol III holoenzyme (HE) is assembled from three different subassemblies. The individual subunits that form these subassemblies interact and exchange in a hierarchy of strong and weak functional interactions, the functional properties of which are strongly controlled by external conditions such as cellular protein concentrations and nucleotide availability. These properties enable the replisome to transition through multiple conformational states to accomplish simultaneous concerted copying of both DNA strands at each replication fork, while its overall integrity is maintained [6]. Many of the enzymatic activities of the Pol III HE have been studied in great detail. Although these studies have provided many insights into the complexity of the Pol III HE, for instance the various enzymatic activities of individual subunits, it remains largely unknown which functional interactions orchestrate the conformational transitions and how these interactions are maintained during replication. This Thesis aimed to study some of the protein–protein interactions within the *E. coli* DNA Pol III HE critical for its proper functioning, with a focus on those interactions within the Pol III core (αεθ) and between the Pol III core and the τ subunit of the clamp loader complex (CLC).

### 7.1 Structural and interaction studies of domain IVa of the τ subunit and its contribution to the α–τ interaction (Chapter 3)

Determination of the high-resolution structures of almost all of the replisome components over the last twenty-five years has been accompanied by the discovery of many dynamic protein–protein interactions. These dynamic interactions are typically mediated by intrinsically unstructured regions that become structured upon interactions with their respective binding partners. The α–τ interaction is an example of such an interaction. Extensive biochemical and structural studies of the α–τ interaction have been conducted. However, most of the research has been focussed on the α-binding domain of τ. The recently solved X-ray crystal structure of the τ-binding domain of α in complex with a truncation mutant of τ (τ_{C16}) provides much needed insights into the structural arrangement of the αCTD upon binding τ. This structure shows key residues in both αCTD and domain V of τ involved in the interaction (Dr Zhi-Qiang Xu et al., unpublished). The interaction interface results in spatial constraints on the N- and C-termini of τ_{C16} with the αCTD. This spatial constraint implies the preceding domain IV is in close proximity to α.
SPR (ProteOn XPR36) measurements show a direct contribution of domain IVa of \(\tau\) to the \(\alpha-\tau\) interaction. The inclusion of the additional 66 amino acids of domain IVa results in a slower dissociation rate and a 3–5-fold increase in binding affinity of \(\tau_{C24}\) over \(\tau_{C16}\) binding to \(\alpha\), with an apparent \(K_D\) of 96 pM in 300 mM NaCl. The increase in \(K_D\) suggests these additional residues present in domain IVa of \(\tau\) make contacts with \(\alpha\).

Further examination of \(\tau_{C24}\) upon interaction with \(\alpha\) by \(^{15}\)N-HSQC NMR showed that signals for a few residues in domain IVa become broadened beyond detection. The cross-peaks in domain IVa have not yet been assigned to specific amino acids in the sequence. Consequently, the perturbed cross-peaks in \([^{15}\text{N}]\tau_{C24}\) upon binding to \(\alpha\) could not be precisely identified.

In addition to SPR and NMR measurements, structural analyses of the \(\alpha-\tau_{C16}\) and \(\alpha-\tau_{C24}\) complexes by SAXS and SANS were performed. SAXS and SANS experiments allowed precise measurements of the overall shapes and dimensions of each complex as well as \(\alpha\) alone. Moreover, these data were also used to derive structural models and comment on the similarities of other previously proposed models and recent cryo-EM structures. Both SAXS and SANS data revealed more compact structures than expected, and rigid-body models suggest a potential structural rearrangement of \(\alpha\) upon \(\tau\) binding in the absence of DNA. We speculate this is a consequence of the rearrangement of the \(\alpha\)CTD. To confirm this, high-resolution approaches such as cryo-EM and single-particle reconstruction will need to be employed.

### 7.2 The \(\alpha\) interaction domain V of the *Escherichia coli* \(\tau\) subunit also interacts with the \(\beta_2\) clamp (Chapter 4)

A physical interaction has recently been reported between the \(\varepsilon\) and \(\beta_2\) subunits. This weak interaction helps provide stability for the Pol III core on template DNA while in the polymerisation mode during DNA synthesis. It has evolved to be weak enough to be disrupted during transitions to other conformational states. This property is likely critical to allow other enzymes access to \(\beta_2\) during the replication cycle and during DNA repair processes. Jergic and co-workers [14] used peptides containing various CBMs (strong, native and weak) to measure the dissociation constant, \(K_D\) of the \(\varepsilon-\beta_2\) interaction. It is clear from both SPR and ESI-MS data that the strength of the \(\varepsilon-\beta_2\) interaction is dependent on the strength of the CBM present in \(\varepsilon\). The peptides that were used for \(K_D\) measurements provide a good estimate, but are not a true representative measure of the strength of binding in the context of the native protein complexes.
To measure accurate $K_D$ values for complete, full-length $\alpha$, $\alpha\epsilon$ and $\alpha\epsilon\tau$ complexes binding to $\beta_2$, various protein complexes containing the different CBMs were assembled on SPR chip surfaces and kinetic and/or thermodynamic parameters were derived. A summary of the results is given in Table 7.1.

<table>
<thead>
<tr>
<th>Interaction</th>
<th>$k_a$ ($M^{-1} s^{-1}$)</th>
<th>$k_d$ ($s^{-1}$)</th>
<th>$K_D$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bioC–$\alpha$</td>
<td>$(4.64 \pm 0.00) \times 10^5$</td>
<td>$(1.65 \pm 0.00) \times 10^{-5}$</td>
<td>$0.036 \pm 0.002$</td>
</tr>
<tr>
<td>BioCG–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$3130 \pm 490$</td>
</tr>
<tr>
<td>bioTC16α–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$2080 \pm 200$</td>
</tr>
<tr>
<td>bioTC24α–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$890 \pm 10$</td>
</tr>
<tr>
<td>bioTC16αL–$\beta_2$</td>
<td>$(6.06 \pm 1.28) \times 10^7$</td>
<td>$(0.107 \pm 0.02)$</td>
<td>$1.77 \pm 0.76$</td>
</tr>
<tr>
<td>bioTC24αL–$\beta_2$</td>
<td>$(7.79 \pm 0.77) \times 10^7$</td>
<td>$(0.07 \pm 0.01)$</td>
<td>$0.86 \pm 0.17$</td>
</tr>
<tr>
<td>bioTC16αC–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$1300 \pm 110$</td>
</tr>
<tr>
<td>bioTC24αC–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$1140 \pm 50$</td>
</tr>
<tr>
<td>bioTC16αL–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$29 \pm 0.83$</td>
</tr>
<tr>
<td>bioTC24αL–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$29 \pm 2.23$</td>
</tr>
<tr>
<td>bioTC16αLC–$\beta_2$</td>
<td>$(2.76 \pm 1.43) \times 10^6$</td>
<td>$(0.0396 \pm 0.00)$</td>
<td>$1.43 \pm 0.13$</td>
</tr>
<tr>
<td>bioTC24αLC–$\beta_2$</td>
<td>$(7.97 \pm 0.70) \times 10^6$</td>
<td>$(0.0137 \pm 0.00)$</td>
<td>$1.72 \pm 0.28$</td>
</tr>
<tr>
<td>bioTC16αL–$\beta_2$</td>
<td>$(3.27 \pm 0.38) \times 10^7$</td>
<td>$(0.0437 \pm 0.00)$</td>
<td>$1.35 \pm 0.32$</td>
</tr>
<tr>
<td>bioTC24αL–$\beta_2$</td>
<td>$(3.77 \pm 0.12) \times 10^7$</td>
<td>$(0.0653 \pm 0.00)$</td>
<td>$1.73 \pm 0.11$</td>
</tr>
<tr>
<td>bioCGC–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$14 \pm 3$</td>
</tr>
<tr>
<td>bioTC24–$\beta_2$</td>
<td>–</td>
<td>–</td>
<td>$110000 \pm 5000$</td>
</tr>
</tbody>
</table>

First, the $\alpha$–$\epsilon$ interaction was reassessed using biotinylated $\epsilon$ instead of biotinylated $\alpha$ [212]. Our results differed significantly from those previously reported using a N-terminally biotinylated $\alpha$. The measured $K_D$ value of $36 \pm 2$ pM is 200-fold lower than previously reported when the interaction is measured using immobilised $\epsilon$ instead of immobilised $\alpha$.

Using a novel strategy, whereby we measured binding of $\beta_2$ to $\alpha$ and/or $\alpha\epsilon$ complex immobilised via the $\alpha$–$\tau$ interaction, a contribution of $\tau$ binding to $\beta_2$ was exposed. A strong CBM in $\epsilon$ ($\alpha_L$) was able to negate this contribution. However, if replaced by a strong CBM in $\alpha$ ($\alpha_L$) the nature of the CBM in $\epsilon$ did not affect the strength of the interaction with $\beta_2$. This strongly suggests that the C-terminal domains of $\tau$ position/assemble $\alpha$ in a manner that excludes the $\epsilon$ CBM from accessing $\beta_2$ in the absence of DNA. It is currently unknown whether this effect is present once $\alpha$ binds primer–template DNA. In later SPR (BIAcore) measurements, a direct physical
interaction between immobilised \( \tau_{C24} \) and \( \beta_2 \) was identified, with an apparent \( K_D \) of 110 ± 5 μM (\( \beta_2 \) as a dimer). This is a fascinating observation, and suggests a novel \( \tau-\beta_2 \) interaction may exist during some part of the DNA replication cycle.

Using another truncation mutant of \( \tau \) (\( \tau_{C22} \)), the very weak \( \tau-\beta_2 \) binding affinity of \( K_D = 86 \pm 54 \) μM (\( \beta_2 \) as a monomer) was derived from \(^{15}\)N-HSQC NMR using \(^1\)H chemical shift perturbations upon addition of various concentrations of \( \beta_2 \). Importantly the derived \( K_D \) value was in good agreement with SPR measurements. Furthermore, considering proton and \(^{15}\)N resonance assignments of domain V of \( \tau \) (through solution structure of \( \tau_{C14} \)) were previously determined, the putative binding surface of the \( \tau-\beta_2 \) interaction could be identified on domain V of \( \tau \). Using this technique, we could identify individual residues in this region of \( \tau \) responsible for binding to \( \beta_2 \). Interestingly, residues identified by \(^{15}\)N-HSQC NMR experiments coincide with residues in \( \tau \) that have been reported to contact the fingers domain of \( \alpha \) in a recent cryo-EM structure of the \( \alpha\varepsilon-\beta_2-\tau_{C16} \) complex (Figure 4.12). It is thus tempting to consider this interaction between \( \tau \) and \( \beta_2 \) plays a role in \( \beta_2 \) clamp loading and polymerase loading onto template DNA. In fact, these observations may support a switch like model proposed by Fernandez-Leiro and co-workers [222]. It will be especially revealing in the future to examine the effects on these processes by mutagenesis studies.

7.3 Development and characterisation of fluorescently labelled Pol III cores for single-molecule imaging (Chapter 5)

Traditional biochemical DNA replication assays require agarose gel electrophoresis and/or incorporation of radioactive nucleotides to monitor rates of DNA synthesis and processivity by observation of DNA products. Rare events and short-lived intermediates are often difficult to detect using these traditional methods, which conceal transitions through discrete conformational states and dynamic behaviour that may occur. Simultaneous observation of composition and enzymatic activity of single replisomes will be needed to gain a complete understanding of the events of simultaneous leading and lagging strand DNA replication by the Pol III HE. Thus, production of different labelled complexes within the DNA Pol III HE are needed to develop new single-molecule fluorescence tools to study replisome dynamics and capture these short-lived intermediate states.

Chapter 5 describes the cloning, purification and labelling of two components of the Pol III core (\( \alpha \) and \( \theta \)) with high specificity and efficiency. Subsequently, labelled individual
subunits were reconstituted into Pol III cores and further purified by ion-exchange chromatography. Analyses by bulk-phase replication assays show that all labelled Pol III cores can perform strand extension and participate in SD DNA synthesis. During labelling of $\theta_{\text{cys}}$, it was discovered that it could be S-glycosylated \textit{in vivo} by addition of a galactose moiety onto the artificial cysteine at the extreme C-terminus. This astounding observation is the first evidence of a S-glycosylation pathway in \textit{E. coli}. Although outside the scope of this Thesis, detailed biochemical and genetic studies are needed to further uncover the glycosyltransferase responsible. Finally, the potential uses of labelled Pol III cores using a genetic fusion of mKikGR to the $\alpha$ subunit is demonstrated by \textit{in vitro} and \textit{in vivo} experiments. Replisomes reconstituted \textit{in vitro} using mKikGR-\$\alpha$ show it is active in single-molecule rolling-circle replication reactions. As expected, it has similar rate and processivity distributions compared to replisomes reconstituted with wild-type Pol III cores. Further, mKikGR-\$\alpha$ expressed at basal levels from pJSL2156 (Section 5.3.1) co-localised with an $\varepsilon$-YPET chromosomal fusion protein, showing in living, replicating \textit{E. coli} cells that mKikGR-\$\alpha$ expressed from a plasmid can be incorporated into active replication forks \textit{in vivo}.

### 7.4 Single-molecule visualisation of fast polymerase turnover in the bacterial replisome (Chapter 6)

Dynamic polymerase exchange has been observed and characterised in other model systems, such as bacteriophages T4 and T7. Inspired by these observations, I investigated whether the \textit{E. coli} DNA polymerase is reused or frequently exchanged during coupled DNA replication.

Reconstituted \textit{E. coli} DNA Pol III HE containing SNAP labelled Pol III cores (produced in Chapter 5) were used to study the dynamics of DNA polymerases at the replication fork at the single-molecule level. By using a rolling-circle DNA amplification scheme to observe highly processive DNA synthesis in real-time, exchange of Pol III$^*$ at the replication fork was visualised by measuring the fluorescence intensity at the replisome spot as a function of time using 1:1 mixtures of red and green labelled Pol III$^*$'s. Additionally, Pol III$^*$ readily exchanges into the replisome when continually present, at a rate that is dependent on its concentration. Contrary to this observation, by pre-assembling replisomes on a rolling circle template, we demonstrate the replisome is a stable entity. This is clearly shown by its ability to replicate thousands of nucleotides without dissociation in the absence of free Pol III$^*$, an observation in agreement with many other studies [183,184,294,395]. Assessment of $\tau$–DnaB interaction by SPR shows binding is $\sim$200-fold weaker than previously reported [37], with an apparent $K_D$. 


value of $1.3 \pm 0.2 \mu M$ (using $\tau_3 \delta \delta' \chi' \psi$ in 50 mM NaCl). Interestingly, in the context of binding the CLC, DnaB exhibits fast on/fast off kinetics. This observation strongly supports Pol III* exchange on DnaB as the $\tau$–DnaB interaction would not be an obstacle for efficient exchange by a concentration dependent mechanism. Further, Pol III cores within a Pol III* complex do not exchange with the cores from another Pol III*. Both findings therefore challenge and add to the classical textbook model of the single *E. coli* DNA Pol III HE. Where previously a single DNA Pol III HE was believed to be stably bound at the replication fork to the helicase, performing virtually all of the DNA synthesis in the cell. In contrast to this classical view, we propose the *E. coli* replisome strikes a balance between stability and plasticity that complies with established chemical principles. It will be important in the future to establish the general rules and mechanisms that govern polymerase exchange within the Pol III HE by concentration dependent mechanisms.
Chapter 8
References


Kitani, T., Yoda, K., Ogawa, T., Okazaki, T., Evidence that discontinuous DNA replication in *Escherichia coli* is primed by approximately 10 to 12 residues of RNA starting with a purine, *J. Mol. Biol.* 184 (1985) 45–52.


Lohman, T.M., Bujalowski, W., Negative cooperativity within individual tetramers of *Escherichia coli* single strand binding protein is responsible for the transition between the (SSB)$_{35}$ and (SSB)$_{56}$ DNA binding modes, *Biochemistry*. 27 (1988) 2260–2265.


Dohrmann, P.R., McHenry, C.S., A bipartite polymerase-processivity factor interaction: only the internal β binding site of the α subunit is required for processive replication by the DNA polymerase III holoenzyme, *J. Mol. Biol.* 350 (2005) 228–239.

Dohrmann, P.R., Manhart, C.M., Downey, C.D., McHenry, C.S., The rate of polymerase release upon filling the gap between Okazaki fragments is inadequate to support cycling during lagging strand synthesis, *J. Mol. Biol.* 414 (2011) 15–27.


Appendix
Appendix Figure 1 | Determination of the deuteration level of $\alpha$ by ESI-MS under native conditions. 

**a**, The nano-ESI spectrum of $\alpha$ in buffer containing 150 mM ammonium acetate pH 7.7, 1 mM $\beta$-ME. The measured mass of 129,770.3 ±1.8 Da is in excellent agreement with the theoretical mass of 129,773 Da. This measurement was used to calibrate the measurement the perdeuterated a sample. 

**b**, The nano-ESI spectrum, of perdeuterated $\alpha$ in hydrogenated buffer containing 150 mM ammonium acetate pH 7.7, 1 mM $\beta$-ME. A molecular mass of 134810.69 ±16.7 was measured for perdeuterated $\alpha$, based on this measurement it was calculated that perdeuterated $\alpha$ was deuterated 72% non-exchangeable $^2$H in H$_2$O containing buffer, which would exchange to 100% in a 100% D$_2$O containing buffer.

Appendix Figure 2 | Representative field of view of SYTOX Orange-stained dsDNA from the single-molecule rolling-circle DNA replication assay. Efficient DNA replication proceeds in the presence of the full complement of replication reaction mix, including the reconstituted *E. coli* replisome, NTPs and dNTPs. Note both the length and number of products.
Appendix Figure 3 | Fluorescence intensity of replicating Pol III* complexes does not change at longer DNA lengths under near-TIRF imaging conditions. The fluorescence intensity of labeled Pol III* complexes does not change as a function of DNA length during single-molecule rolling-circle DNA replication under constant flow. The errors represent the standard deviation.

Appendix Figure 4 | Histograms of the stoichiometry of Pol III* at the replication fork. a, Intensity distribution at 6.7 nM Pol III* and b, intensity distribution at 0.3 nM Pol III*. The histograms are fit with either four (6.7 nM) or three (0.3 nM) Gaussian distribution functions centred at integral numbers of Pol III*.