University of St Andrews

Structural studies of ArnA, a bifunctional enzyme from *Escherichia coli,* and an archaeal heterotrimeric sliding clamp from *Sulfolobus solfataricus*

Gareth Williams

A thesis submitted for the degree of Doctor of Philosophy

February 2006

Th F310

Declarations

I Gareth Williams, hereby certify that this thesis, which is approximately 40000 words in length, has been written by me, that it is a record of work carried out by me and that it has not been submitted in any previous application for a higher degree

Date...21/6/06........... Signature of Candidate.

I was admitted as a research student in September 2002 and as a candidate for the degree of Ph.D. in October 2003; the higher study for which this is a record was carried out at the University of St.Andrews between 2002 and 2005

Date...21/6/06........... Signature of Candidate.

I hereby cerify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of Ph.D. in the University of St.Andrews and that the candidate is qualified to submit this thesis in application for that degree

Date 21/6/06........... Signature of supervisoi

In submitting this thesis to the University of St. Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any *bona fide* library or research worker.

Date...21/6/06......... Signature of Candidate.

Abstract

Several pathogenic Gram-negative bacteria can modify the structure of Lipid A in their outer membrane to avoid the host immune response. One of these modifications includes addition of 4-amino-4-deoxy-L-arabinose, resulting in resistance to cationic antimicrobial peptides. Essential to this modification are both decarboxylase and formyltransferase functions of the bifunctional enzyme ArnA. The decarboxylase and formyltransferase domains of ArnA have been expressed and purified as discrete proteins and the crystal structures solved to 2.3 and 1.2 Å respectively. The formyltransferase structure has been determined in complex with the substrate and cofactor analogues, UDP and N-5-formyl-tetrahydrofolate. The decarboxylase structure represents the apo form of the enzyme and key catalytic residues for decarboxylation were identified from the superimposition of this structure with the related enzyme RmlB. These residues were mutated and assayed for activity compared to the wild type protein, providing insights into ArnA decarboxylation and hydride transfer in short chain dehydrogenases / reductases.

That ArnA catalyses two, non-sequential reactions raises the possibility that the intermediate is channelled from ArnA to another protein, ArnB, which catalyses the intervening step, and then back to ArnA. This possibility was explored using gel filtration assays to determine whether a stable complex formed between ArnA and ArnB.

Sliding clamps are essential to all domains of life and their toroidal structure is conserved, providing the basis for their non-specific encircling of duplex DNA. In archaea and eukaryotes PCNA acts as the clamp for the replicative polymerase and is also involved in other pathways such as DNA repair. The archaean *Sulfolobus solfataricus* is

unique in that it has evolved heterotrimeric PCNA, although a heterotrimeric clamp also exists in eukaryotes with a specialised function in the DNA damage response. The structure of PCNA from *S. solfataricus*, has been solved and progress made towards the structural solution of this clamp in complex with an interacting domain of XPF. This is the first structure of a heterotrimeric sliding clamp and these preliminary results are the basis for future studies investigating the subunit specific interactions of *S. solfataricus* PCNA. This study will provide insights into the function of eukaryotic clamps and substrate specificity of XPF homologues.

Abbreviations

Amino acids

A – Ala – alanine

C – Cys – cysteine

D – Asp – aspartic acid

E – Glu - glutamic acid

F – Phe – phenylalanine

G – Gly – glycine

H – His – histidine

I – Ile – isoleucine

K – Lys – lysine

L – Leu – leucine

M – Met – methionine

N –Asn – asparagines

P – Pro – praline

Q – Gln – glutamine

R – Arg – arginine

S – Ser – serine

T – Thr – threonine

V – Val - valine

W – Trp – tryptophan

Y – Tyr – tyrosine

General

6-4PP(s) - 6-4 photoproduct(s)

9-1-1 complex - Rad9-Rad1-Hus1 heterotrimer

A - adenine

A. aeolicus – Aquifex aeolicus

A. pernix – Aeropyrum pernix

ADH – alcohol dehydrogenase

AICAR - 5-aminoimidazole-4-carboxamide

ASADH - aspartate semialdehyde dehydrogenase

ATP - adenine triphosphate

BER - base excision repair

Bis-tris - Bis(2-hydroxyethyl)amino-tris(hydroxymethyl)methane

C - cytidine

C. elegans - Caenorhabditis elegans

CAF1 - chromatin assembly factor 1

CAMP(s) - cationic antimicrobial peptide(s)

CDK(s) – cyclin dependent kinase(s)

CDC – centres for disease control and prevention

CMP - Cytodine monophosphate

CPD(s) - cyclobutane pyrimidine dimer(s)

COFS - cerebro-oculofacio-skeletal syndrome

CS - cokayne syndrome

D. melanogaster – Drosophila melanogaster

DADH - *D. melanogaster* ADH

DHFR – dihydrofolate reductase

dsDNA - double stranded DNA

E. coli - Escherichia coli

ESI MS - electrospray ionisation mass spectrometry

FDH – formyltetrahydrofolate dehydrogenase

FEN1 - flap endonuclase 1

FMT – 10-formyltetrahydrofolate-L-methionyl-tRNA N-formyltransferase

G - guanine

Gal - galactose

GAR(T) – glycinamide ribonucleotide (transformylase)

GGR - global genome repair

Glc- glucose

GlcNAc - N-acetyl-D-glucosamine

GlcUA – glucuronic acid

H. pylori - Helicobacter pylori

HEPES – N-(2-hydroxyethyl)piperazine-N'-(2- ethanesulfonic acid)

HhH - Helix-hairpin-helix

HIV – Human immune-deficiency virus (p2)

HNPCC - hereditary non-polyposis colorectal cancer

HPLC - high pressure liquid chromatography

HR - homologous recombination

IMPCH - inosine monophosphate cyclohydrolase

IPTG - isopropyl-ß-D-thiogalactopyranoside

Kdo – 3-deoxy-D-manno-octulosonic acid

LADH - horse liver ADH

L-Ara(f)4N – (formyl) 4-amino-4-deoxy-L-arabinose

L-Ara4O – β-L-threo-pentapyranosyl-4'-ulose

LB - luria broth

LBHB - low barrier hydrogen bond

LPS - Lipopolysaccharide

M. thermoautotrophicus - Methanothermobacter thermoautotrophicus

MAD -multiwavelength anomalous diffraction/dispersion

MCMT - cytosine 5-methyltransferase

MDR – medium chain dehydrogenases/reductases

MES - 2-(N-Morpholino)ethanesulfonic acid

MMR - mismatch repair

MRSA - Methacillin resistant *S. aureus*

N. meningitides - Neisseria meningitides

N-5-/N-10-fTHF – N-5-/N-10-formyl tetrahydrofolate

$NAD(^+)$ - nicotinamide adenine nucleotide oxidised (reduced) form

$NADP(^+)$ - nicotinamide adenine nucleotide phosphate oxidised (reduced) form

NER - nucleotide excision repair

P. aeruginosa – Pseudomonas aeruginosa

PBP(s) – penicillin binding protein(s)

PCNA - proliferating cell nuclear antigen

PEG – polyethylene glycol

PIP – PCNA interacting protein

Pol - DNA polymerase

R. leguminosarum – Rhizobium leguminosarum

Rad - Radiation sensitive mutant

RFC - Replication factor C

RMSD – root mean square deviation

RNAPII - RNA polymerase II

RPA - replication protein A

rRNA – ribosomal ribonucleotide

S. aureus – Staphylococcus aureus

S. cerevisiae – Sacchromyces cerevisiae

S. pombe - Schizosacchromyces pombe

S. typhimurium – Salmonella typhimurium

Se-Met – selenomethionine modified

SDR(s) - short chain dehydrogenases/reductases

ssDNA - single stranded DNA

*Sso*PCNA – *S. solfataricus* PCNA

*Sso*XPF – *S. solfataricus* XPF

T-T - Thymine-thymine

TCR - transcription coupled repair

TEAA - triethylammonium acetate

tRNA – transfer RNA

TTD - trichothiodystrophy

Udg – UDP-Glc dehydrogenase

UDP - Uracil diphosphate

UMP - uracil monophosphate

VRE - Vancomycin resistant *enerococcus*

VRSA - Vancomycin resistant *S. aureus*

WHO – World Health Organisation

XP - Xeroderma pigmentosum

XPF(HhH)$_2$ - HhH domain of XPF

Contents

Chapter 1

Introduction

INTRODUCTION

1.1   ANTIBIOTIC RESISTANCE

Bacteria have a typical genome size of around 4 megabases. The majority of their genomes encode functional genes (> 90 %) with very little non-coding regions. Bacteria have a relatively high mutation rate, its DNA polymerase makes around one mistake in a million[1] and due to the nature of their genomes any mutation occurring has a high probability of altering a gene product. The small genome and mutation rate, coupled to their replication rate, makes bacteria particularly adaptable[2]. As a result bacteria have colonised almost all corners of the world; from the arctic and deep sea vents, to the guts of animals. This adaptability has been harnessed by pathogenic bacteria, which utilise an arsenal of extracellular factors to colonise their host and promote the release of nutrients on which they survive. These pathogens are the cause of infectious disease that has plagued higher organisms since their own evolution.

Prior to the mid-1940's deaths from infectious bacteria were common, with infections such as *Staphylococcus aureus* having a mortality rate of around 80 %[3]. The discovery of the antibiotic properties of *Penicillium* in 1897 in Duchesne's Doctoral thesis, its rediscovery by Alexander Fleming in 1928[4] and the synthesis of Penicillin by Florey and Chain[5,6], ushered in a new era for human kind, the defeat of pathogenic bacteria. This era lasted less than 40 years; predictably bacteria have adapted to the use of antibiotics. Unfortunately pharmaceutical interest in developing novel antibiotics declined during the golden era of antibiotic therapy and has yet to recover. Although the widespread use of antibiotics has contributed to an increase of

the average human lifespan by 8 years[7], over the last 15 years hospital admissions and

mortality rates due to bacterial infections have risen steadily (figure 1.1.1). This rise

is due to the spread of antibiotic resistance in bacteria and an increase in the number

of immune suppressed individuals, largely as a result of old age, HIV and

chemotherapy (World Health Organisation (WHO), Centres for Disease Control and

Prevention (CDC)).
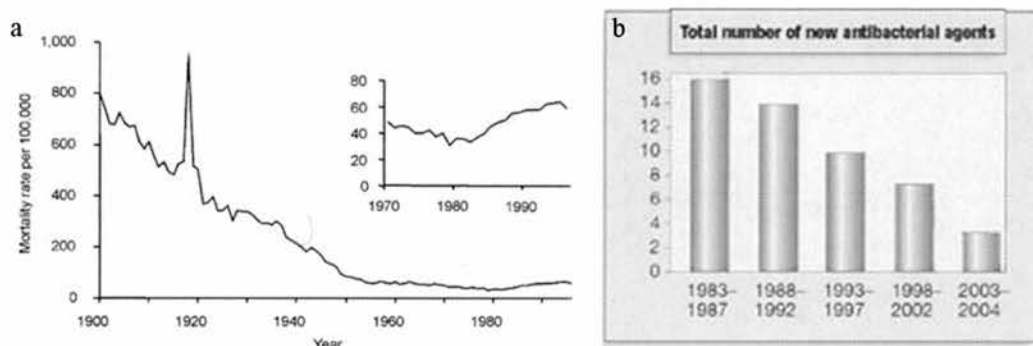
Figure 1.1.1: Graphs showing: (a) the decrease in deaths from infectious disease

over the last 100 years in America. The large peak around 1920 is due to the avian flu

epidemic. Inset charts the recent rise in deaths, largely due to the spread of antibiotic

resistance (WHO, CDC). (b) The decline of antibacterial drugs produced in the last 20

years[8].

1.1.1   Antibiotics

There are four main cellular processes that current antibiotics target protein synthesis,

DNA replication, DNA repair and cell-wall synthesis.

Several antibiotics target protein synthesis to kill bacteria, by inhibition of one

of several steps catalysed by the ribosome during translation. These include

macrolides of the erythromycin class[9], aminoglycosides[10], tetracyclines [11] and a more

recent class of antibiotics, the oxazolidinones[3]. Fundamental to the use of antimicrobials that target the ribosome are the large differences that exist between this machinery in prokaryotes and eukaryotes[12]. Tetracycline utilises the differences between bacterial and eukaryotic ribosomal RNA (rRNA), an important component of the ribosome[13]. The specific binding of tetracycline to bacterial rRNA prevents the association of transfer RNA (tRNA) with the ribosome[14]. As a result translation elongation and therefore protein synthesis is inhibited in prokaryotes but not eukaryotes[14].

Similarly to protein synthesis, the differences between prokaryotic and eukaryotic DNA replication and repair can be exploited in the use of antibiotics to specifically inhibit bacterial machinery[12]. The quinolone, coumarin and cyclothialidine classes of antibiotics target DNA replication, specifically through inhibition of the topoisomerase DNA gyrase[15]. DNA gyrase has the unique ability to introduce supercoils into DNA, a process essential to prokaryotic DNA replication due to the circular nature of their genome[15]. Although evolutionarily related to other topoisomerases, close homologues of DNA gyrase do not exist in eukaryotes and therefore these antibiotics only inhibit bacterial replication[15,16].

The cell wall and outer surface of bacterial cells is different to that found in eukaryotes and is therefore an excellent antibiotic target. In Gram-positive bacteria the peptidoglycan layer protects the cell and provides the mechanical strength for the cell wall to resist osmotic lysis[17]. It is a mesh of peptide and glycan strands. Penicillin binding proteins (PBPs), containing both transpeptidase and glycosylase activities[17,18], link peptide chains and extend sugar chains. The substrate for the transpeptidase is a pentapeptide, with a dipeptide D-Ala-D-Ala terminus, the crosslinking of which gives mechanical strength to the cell[17]. Antibiotics such as

penicillins, cephalosphorins and vancomycin[19,20] interfere with the transglycosylation or transpeptidation processes, weakening the bacterial cell wall, leading to lysis due to changes in osmotic pressure.

Gram-negative bacteria have a smaller peptidoglycan layer than Gram-positive bacteria. However, they have an outer cell membrane, the outer surface of which contains the lipopolysaccharide (LPS) (chapter 2.1.1). This layer provides a more effective barrier to their environment and as a result some antibiotics are less effective against these organisms. Enzymes involved in the biosynthesis of the lipids and sugars that make up the LPS are potential drug targets against Gram-negative pathogens, discussed in chapter 2.

1.1.2 Mechanisms of Antibiotic Resistance

Most antibiotics are natural products that have evolved in bacteria or fungi as a method of chemical warfare to kill competing organisms. The evolution of these antibiotics has led to the co-evolution of resistance so that the antibiotic producer is not killed by its own product[21-23]. Further, since antibiotics target fundamental processes in bacteria that involve proteins, mutations in genes which confer resistance are quickly selected and replicated in rapidly dividing bacterial cells. As a result antibiotic resistance is inevitable and several common mechanisms have evolved (figure 1.1.2).
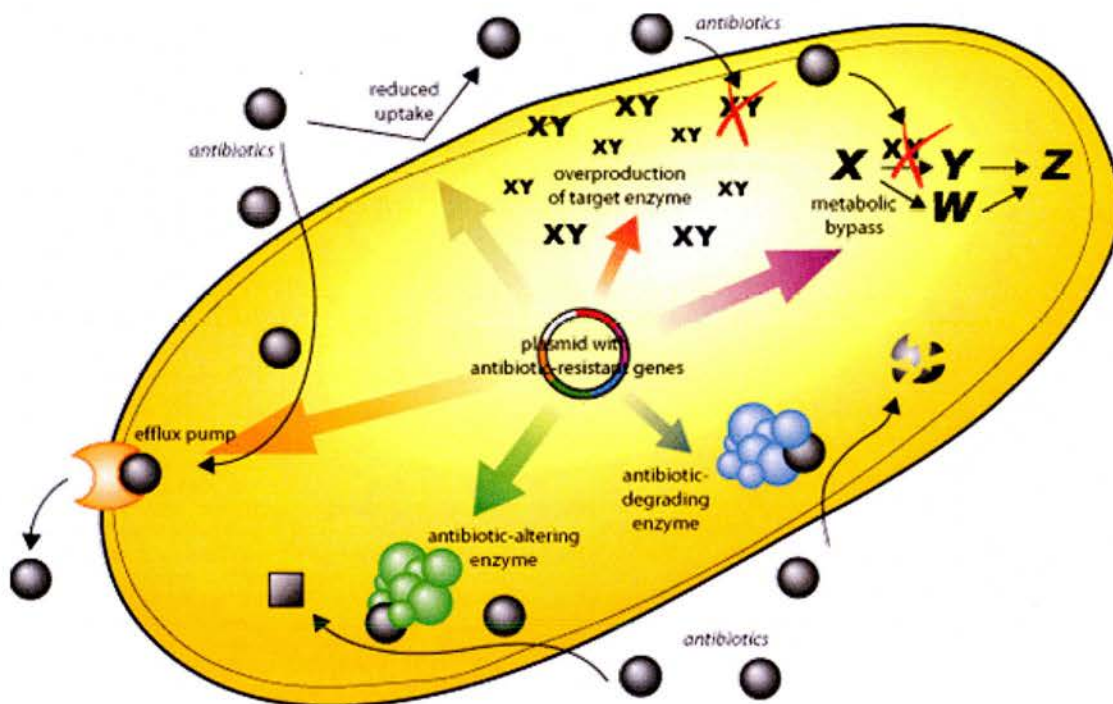
Figure 1.1.2: Mechanisms of resistance to antibiotics. Taken from http://www.bioteach.ubc.ca/Biomedicine/CationicPeptides.

The effect of the antibiotic can be neutralised by altering the antibiotic molecule so that it is no longer effective. The β-lactam ring of Penicillins and Cephalosphorins is the active part of these antibiotics. These antibiotics act as pseudosubstrates for the transpeptidase reaction catalaysed in the formation of the peptidoglycan layer, resulting in the acylation of the transpeptidase active site[18,24]. The penicilloylated transpeptidase deacylates very slowly and so the enzyme active site is occupied, inhibiting the normal crosslinking of peptide chains in the peptidoglycan layer[18,24]. This action is prevented in some resistant bacteria by expression of β-lactamase enzymes which hydrolyse the β-lactam ring to produce penicillinoic acid, which does not inhibit transpeptidases[24]. Alternatively, instead of breaking down the active group, the antibiotics can be covalently modified with

6

moieties that prevent its function[25]. This is how aminoglycosides are neutralised, with resistant bacteria expressing enzymes that can add AMP, PO_3^{2-} or acyl groups to the antibiotic, reducing its affinity for RNA and thus its effectiveness[25].

The concept of pumping is similar to neutralisation; if an active form of the antibiotic is not present at its target, it is unable to have its effect. Thus efflux pumps, which pump out the antibiotic, have evolved as a mechanism of resistance[21-23]. By pumping the antibiotic out of the cell at a faster rate than it can diffuse in, the concentration of antibiotic is never high enough to be effective[12]. These pumps are membrane proteins that are related to those normally produced[21,23] and are a common mechanism for the export of antibiotics by antibiotic producers, so that they themselves are not affected[21-23]. The protein synthesis targeting antibiotics tetracycline[21,26] and erythromycin[21,27] are removed by this mechanism in some resistant bacteria.

The antibiotic target itself can be mutated or modified to prevent antibiotic binding[12]. For example mono- or dimethylation of the 23S ribosome subunit at a specific adenosine by the methyltransferase Erm does not affect protein synthesis, but lowers the affinity for erythromycin and pristinamycin antibiotics [28]. Alternatively mutations to the target protein that lowers the affinity for an antibiotic can result in resistance. For example mutation of PBPs can render bacteria resistant to penicillin[29]. The basis for methacillin resistant *S. aureus* (MRSA) is the acquisition of the MecA protein, which encodes a different version of a PBP, with low affinity for almost all β-lactam antibiotics[30,31].

Vancomycin resistance works in a different, more complicated way, in which the building blocks used for constructing the peptidoglycan layer are changed[19,20]. Vancomycin acts by binding with high affinity to the terminal dipeptide D-Ala-D-Ala

of the substrate for transpeptidases, inhibiting the crosslinking of the peptide chains in the peptidoglycan layer[19,20]. In vancomycin resistant bacteria vanHAX encodes enzymes that reduce pyruvate to D-lactate (VanH), link this to D-Ala (VanA) and hydrolyse the normal dipeptide D-Ala-D-Ala (VanA)[19,20]. In these resistant bacteria only D-Ala-D-lactate is elongated to produce a pentapeptide that can act as a substrate for transpeptidases, with no effect on the cross-linking efficiency but with an 1000-fold reduced affinity for vancomycin[19,20].

1.1.3   Multi-drug resistance

The spread of antibiotic resistance is driven by the selective pressure placed on the bacteria by the continued use of antibiotics and the promiscuity of gene transfer between bacteria. As discussed above; there are many ways that bacteria can evolve resistance to antibiotics and other antimicrobials and as a result there is an almost inevitability that given enough time, resistance will evolve[32,33]. Vancomycin was introduced in 1958 and clinical isolates of resistant bacteria were not observed until 1987, 29 years later. This is very unusual as clinically significant resistance usually appears within months to a few years after introduction of an antibiotic[34], particularly when it only targets one gene product. The unusual time delay for vancomycin resistance is almost certainly due to the complexity of its resistance mechanism, requiring the action of five gene products[19,20].

It is common for the genes that confer resistance to any particular antibiotic to be collected on plasmids[33] or transposons[19], commonly termed mobile DNA elements. These plasmids are independently replicated within the cell and are passed between bacterial cells, species and sometimes genera. This is the principle mechanism

leading to the rapid spread of resistance genes through bacterial populations[33]. Transposons are able to excise themselves from one genetic locus and integrate into another, allowing resistance genes encoded within the transposon to move promiscuously between plasmids and the genome[19]. Such mobile DNA elements allow the accumulation of more than one set of resistance genes in one organism, resulting in the evolution of multi-drug resistant bacteria[12]. It is these multi-drug resistant bacteria that are currently posing the greatest threat to human health.

There are almost weekly reports in the British press about the pandemic of MRSA in hospitals. *S. auereus* causes serious infections during invasive surgery, with a higher mortality rate associated with MRSA than susceptible infections[35]. The last line of defence against MRSA is the glycopeptide vancomycin, however the emergence of strains of first *enterococci*[36,37] and subsequently *S. aureus*[38] that are resistant to vancomycin (VRE and VRSA respectively) is extremely worrying as there is no effective treatment for these infections. The spread of multi-drug resistant tuberculosis[39] and the Gram–negative *Salmonella typhimurium* DT104[39] are also particularly worrying as they are two of the leading causes of mortality and morbidity worldwide (WHO).

Over the last 40 years more could have been done to avert this crisis we are currently facing as, from an early time after the introduction of penicillin, the problem of bacterial resistance became clear. It was in the 1940's that products secreted from bacteria were identified that break down penicillin[40] and Alexander Fleming, in a New York Times article, warned of the misuse of antibiotics that could select for and lead to the outgrowth of resistant bacteria. These warning signs were not heeded and a combination of antibiotic misuse in both medical and agricultural sciences[12] and a general lack of investment in developing new antibiotics by pharmaceutical

companies[41,42], requires us to develop new drugs to combat infectious disease. Pharmaceutical companies are reluctant to invest in such programs as the almost inevitable evolution of resistance means that any new antibiotic may have limited use in a relatively short time and any successful broad ranging drug would be limited in its use.

1.1.4 Future perspectives

So what can be done to tackle the increasing problem of drug resistant bacteria? With recent technological advances, coupled with the success of sequencing the genomes of pathogenic bacteria (The Wellcome Trust Sanger Institute) and structural genomics[43], there may be many ways that this problem can be addressed. Each protein that is essential for growth, survival, pathogenicity or antibiotic resistance of a pathogen, is a potential drug target. As the success of genomic and proteomic projects continues, where more genomes of pathogenic bacteria are sequenced and the function of the encoded genes are established, more targets will be discovered. Structural genomic projects, which are still in their infancy, will result in the structural solution of more and more of these potential targets. In recent years the availability of high performance computing platforms has transformed the way lead compounds are identified and prioritised[44]. By molecularly docking tens or hundreds of thousands of potential inhibitors from a virtual chemical library onto an enzyme binding site, the mode and affinity of binding can be predicted. This allows compounds to be prioritised for biological screening and has the potential to cut both the cost and time needed to develop new drugs. The combination of structural genomics and virtual screening will provide a greater base to start from for structure based drug design. In

the future it may therefore be anticipated that academia will play a bigger role in the initial drug discovery and improvement stages of finding new antibiotics and then, in collaboration with industry, go on to test these drugs.

In addition, the potential of cationic antimicrobial peptides (CAMPs) may help in the fight against resistant bacteria. CAMPs are an important part of innate immunity in animals and in sufficient concentrations can kill a wide range of pathogens through permeabilisation of their cell walls (chapter 2.1.4) [45,46]. They therefore have the potential of either being used as an antibiotic/anti-infective or in combination, to increase the permeability of the cell wall of pathogens, helping in the delivery of current and future drugs.

1.2   MACROMOLECULAR ASSEMBLIES

Over the last 10 years the success of genomic projects in sequencing the genomes of the important model organisms *Escherichia coli*[47], *Saccharomyces cerevisiae*[48], *Caenorhabditis elegans*[49], *Drosophila melanogaster*[50], mouse[51] and the human genome project[52,53], has given us a wealth of information. These data have been invaluable to the geneticist, biochemist, molecular biologist and structural biologist; however it only scratches the surface in understanding how this genomic information encodes a functional organism. The human genome project[52,53] revealed that, despite a genome size over 30 times larger, we have a similar number of genes as the invertebrate *C. elegans*[49], around 30000[52,53]. It is also apparent that the number of human disorders actually exceeds the number of genes we possess[54]. These observations suggest that the 'complexity' of humans compared to lower eukaryotes must be in how our ensemble of expressed gene products (proteome) is utilised, as

11

opposed to gene number. It is therefore not only the elucidation of function of every gene product, but how they are combined to create an organism, that represents a major challenge for post-genomic biology.

Macromolecular interactions and assemblies, including those between protein-protein, protein-DNA and protein-RNA, play major roles in almost all of cellular biology. These include mediation of cell-cell interactions[55], signal transduction[56], cell-cycle control, apoptosis[57], organelle structure[58], lipid[59] and amino acid metabolism[60], protein degradation[61], DNA replication[59], DNA repair[62], transcription[63], translation[64] and splicing[65]. Many of these pathways are present in all three domains of life; however there exists a general trend that the more complicated an organism, the more interactions necessary to mediate and carry out these processes. The DNA repair pathway nucleotide excision repair (NER) is an example of one of these processes that involves a range of protein-protein interactions and is discussed in more detail below (1.3).

It is interesting to think of how this complexity has evolved considering the underlying structural similarity that exists between related proteins in all domains of life. As of January 2006 the protein data bank (PDB) contains nearly 35000 structures from the three domains of life. It is estimated that the number of distinct structural folds contained in the proteome is approximately 1000[66] and those deposited in the PDB represent around 80 % of these[67]. In contrast the estimation of the number of distinct structural types of interactions that can occur is in the region of 10000[68], with only 20-30 % represented in the PDB[68,69]. Thus it seems organisms have evolved proteins that are made up by a relatively small number of structural folds that can interact with other proteins in a much more complicated manner. A feature of eukaryotic proteins over prokaryotic homologues is that, although the

overall fold is largely conserved, extra features of the structure are present, for example loops and N- or C-terminal extensions, which are important for protein-protein interactions, allowing greater control and functionality of these proteins.

Protein-protein interactions are driven by thermodynamics with an equilibrium existing between the free and complexed proteins. The extremes of this equilibrium result in transient or very weak and very stable complex formation. An important factor leading to the evolution of these protein-protein interactions is the conditions in which they have evolved within the cell. In *E. coli* it is estimated that 20-40 % of the cytoplasm is taken up by macromolecules, with an overall protein concentration between 200 and 400 mg ml^{-1} [70] and a similar concentration exists in the compartments of eukaryotic cells[71]. The high levels of proteins and other macromolecules within the cell results in an environment in which there is very little free space available and this macromolecular crowding[72] provides a non-specific force for macromolecular compaction and association[73-75]. An example of the importance of macromolecular crowding is demonstrated by the *in vitro* reconstitution of DNA replication and transcription, with the presence of an *in vitro* crowding agent required to obtain similar properties as observed *in vivo*[76]. There are also many other examples of *in vitro* crowding agents significantly effecting both kinetic rates[77,78] and macromolecular interactions[75,79,80]. In general *in vitro* conditions are a poor reflection of those *in vivo* as the solutions are very dilute, with protein concentrations rarely above 10 mg ml^{-1}. It is therefore important to consider the effect of macromolecular crowding when drawing important *in vivo* conclusions from such assays[81].

The formation of protein-protein complexes can give several advantages. The protein complex can alter the kinetics of a process which can affect; substrate binding, catalysis or allosteric properties[82]. One example is the binding of PCNA to DNA

polymerase, which alters the processivity of this enzyme[83]. Protein-protein interactions can also facilitate substrate channelling, which can result in several advantages including an increased rate of consecutive reactions. Protein-protein interactions also allow regulation of pathways through inactivation of proteins, such as NFκB bound to its repressor[84]. They can also change the substrate specificity of another protein as is the case for transcriptional regulation, where the interaction of transcription factors with RNA polymerase directs the transcription of different genes[85]. Another major advantage of multi-subunit complexes is the use of different subunits to regulate or alter responses, such as differences observed in the RNA polymerase complex during initiation of transcription, elongation[63,85] and upon stalling when a block is encountered[86].

The daunting task of attempting to work out the entire 'interactome' of an organism, i.e. how all of the protein-protein interactions occurring within (and between) cells makes up an organism, has already begun, with *S. cerevisiae*[87,88] and *D. melanogaster*[89] being used as model organisms. These studies have focused on the use of yeast-two-hybrid screens to look for the interaction between bait proteins, made from as many open reading frames as possible from an organism, and prey. Figure 1.2.1 shows the kind of picture that is emerging from these studies and although their accuracy is debateable, due to the number of false positives observed and not every ORF being suitable as bait, it is clear the 'interactome' of an organism is extremely complex and will take many years to elucidate and verify. The structural biologists' role investigating these interactions and how they relate to cellular biology will be one of the major features of structural biology in the 21st century.
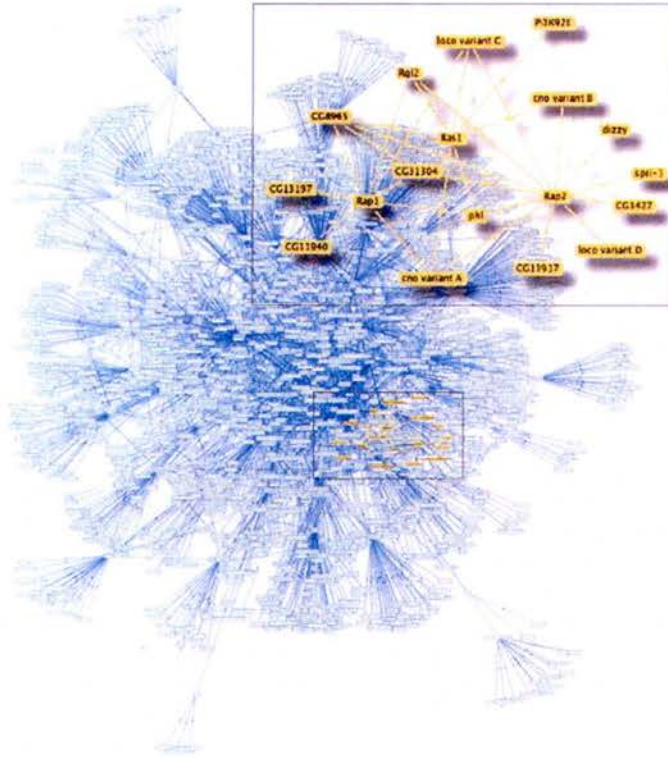
Figure 1.2.1: Protein interaction map (PIM) linking protein-protein interactions in *D. melanogaster*, generated by the PIM Walker graphical interface, a tool developed to display networks. The proteins in the map that bear an RA (Ras Association) or RBD (Raf-like Ras-binding) domain (coloured in orange) define a discrete subnetwork around Ras-like GTPases (colored in yellow). This diagram demonstrates the complexity of protein-protein interactions within the cell Taken from Formstecher et al.[90].

1.3 REPAIR OF UV-DAMAGE

1.3.1 DNA damage

DNA damage arises from many different factors including ultra-violet (UV), X-ray and ionising radiation, mutagenic compounds and endogenous factors, for example errors in replication by DNA polymerase and damage from reactive oxygen species

generated during metabolism[91] [92]. This damage can occur anywhere in the genome and if it is not efficiently repaired it results in permanent mutation following replication. Although it is the selection of beneficial traits arising from mutations that drives evolution, deleterious mutations have serious consequences for the cell and are one of the major causes of cancer[91]. Figure 1.3.1 shows some of the types of DNA damage that can occur including inter-strand crosslinks by chemotherapy agents such as cis-platin, mismatches due to errors in replication, production of cyclobutane pyrimidine dimers (CPDs) or 6-4 photoproducts (6-4PPs) by UV exposure and single and double strand breaks by X-rays[93]. The number of such mutations occurring in the human genome (~3 x 10^9 bases) per day, per cell is astonishingly high, with estimations around 25000[94]. Since the direct result of this damage is mutation or blocking of replication and transcription, if this damage is not corrected the cell quickly dies. Several mechanisms of DNA repair have evolved to deal with the variety of DNA damage that can occur (reviewed in[62,93]). These include: single repair enzymes catalysing direct reversal of damage[62], recombination repair[95], both homologous recombination[95] (HR) and non-homologous end joining [96,97], mismatch repair (MMR)[98,99], base excision repair (BER)[100] and nucleotide excision repair (NER)[101], the mechanism of which is discussed in detail here. The importance of DNA repair in humans is demonstrated by a number of inherited disorders that are defective in DNA repair. These include hereditary non-polyposis colorectal cancer (HNPCC)[102] and Xeroderma pigmentosum (XP)[101], defective in MMR and NER respectively, both of which result in the early onset of cancer.

DNA damaging agent

X-rays	UV light	X-rays	Replication errors
Oxygen radicals	Polycyclic	Anti-tumour agents	
Alkylating agents	aromatic	(cis-platinum, MMC)	
Spontaneous reactions	hydrocarbons		

DNA damage

Uracil	6-4PP	Interstrand cross-link	*A-G* mismatch
Abasic site	CPD	Double-strand break	*T-C* Mismatch
8-Oxoguanine	Bulky adduct		Insertion
Single-strand break			Deletion

Repair mechanism

Base excision repair	Nucleotide excision repair	Recombinational repair	Mismatch repair

Figure 1.3.1: Summary of DNA damaging agents and the main type of lesion they cause along with the most common forms of repair employed to deal with this damage. Adapted from Hoeijmakers[93].

1.3.2   Nucleotide excision repair

NER is unique in that it has evolved almost exclusively to remove exogenous DNA damage such as the UV induced damage CPDs and 6-4PPs, as well as other bulky lesions[94,103]. The common feature of these types of damage is the distortion of the DNA helix[104] and it is this distortion that is thought to be recognised by the DNA damage recognition proteins that initiate this process[104,105]. The importance of NER in humans is highlighted by the severe genetic disorders associated with defects in this process including XP, Cokayne syndrome (CS) trichothiodystrophy (TTD) and cerebro-oculofacio-skeletal syndrome (COFS) [94,106,107]. The phenotypes of these disorders are detailed in table 1.3.2, with patients suffering from XP, CS, TTD and their combination exhibiting extreme sensitivity to UV light, however only XP patients have the early onset and high occurrence of skin lesions[101]. Defective DNA repair in fibroblast cells from XP patients was first described by Cleaver in 1968[108]. Since then cell lines from all of these disorders have shaped the research of NER over the last 40 years, with identification of the mutated genes and the variation in phenotypes of the different disorders allowing great insights into this complex process[109].

Observations over 20 years ago suggested that regions of DNA containing transcribed genes were repaired at a faster rate than non-transcribed regions[110,111], with further research showing that this faster repair is specifically associated with the transcribed strand[112,113]. This phenomenon has been termed transcription coupled repair (TCR) (figure 1.3.5) and the CS, TTD, XP/CS and XP/TTD human disorders link this process to NER[94,106]. As a result, TCR has been studied as a sub-pathway of NER, specifically repairing UV-damage and bulky lesions in the transcribed gene,

with the NER pathway required for the repair of damage in non-transcribed regions called global genome repair (GGR) (figure 1.3.3).

Clinical symptoms	XP	CS	TTD	XP/ CS	XP/ TTD	COFS
UV sensitivity	++	++[a]	++	++	++	?
Increased Freckling	++	-	-	++	++	?
Skin cancer	++	-	-	++	+	?
Cachectic dwarfism	-	++	++	++	+	+++
Microcephaly	-	++	++	++	?	+++
Progressive cognitive impairment	-	++	++	++	++	+++
Sensorineural deafness	-	++	++	++	-	+++
Eye abnormalities	-	++	++	++	?	+++
Skeletal abnormalities	-	+	+	+	?	++
Spasticity	-	++	++	++	?	+++
Ataxia	-	++	++	++	-	+++
Axonal neuropathy	-	+/-	?	+	?	?
Demyelinating neuropathy	-	++	++	++	?	?
Myopathy	-	-	-	+/-	-	?
Brain calcification	-	++	++	++	?	+++
Hypogonadism	-	++	++	++	?	?
Brittle hair and nails	-	-	++	-	+	?
Hyperkeratosis	-	-	++	-	+	?
Progeria	-	++	++	++	?	?

Table 1.3.2: Summary of the clinical symptoms of the NER disorders XP, CS, TTD and COFS and the combination disorders XP-CS and XP-TTD, adapted from Andressoo and Hoeijmakers[107].
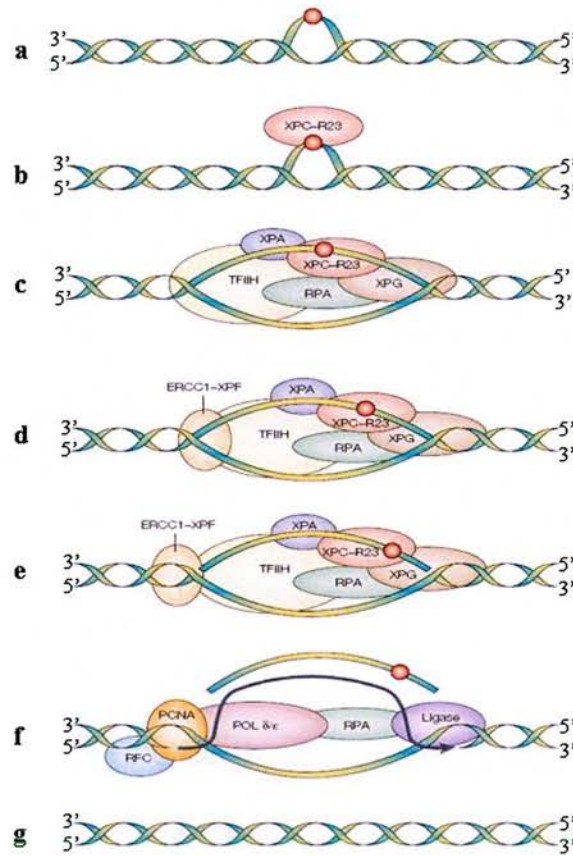
Figure 1.3.3: Global genomic repair by the NER pathway. The helix distortion resulting from DNA damage (a) is recognised by the XPC heterodimer (b), resulting in the recruitment of other NER proteins that open up the helix around the distortion (c). The XPG and XPF DNA-structure specific endonucleases make incisions 3' and 5' to the damage respectively, resulting in the excision of the damaged DNA (d and e). DNA polymerase δ or ε are then loaded by the actions of the clamp loader (RFC) and sliding clamp (PCNA), allowing gap synthesis, before DNA ligase completes the process. Taken from Friedberg[94].

1.3.3   Damage recognition

During the GGR process of NER the protein that initially recognizes the damage appears to be the XPC-hHR23B complex (figure 1.3.2b). This has been shown biochemically by the decreased repair time of DNA which had been pre-incubated with XPC-hHR23B[114] and observed during the sequential assembly of NER factors *in vivo*[115]. The helix distortion generated by 6-4PPs is greater than that of CPDs[104] and it has been shown that XPC-hHR23B binds to 6-4PPs with a greater affinity than CPDs[105,116]. As a result the repair efficiency of 6-4PPs is 5 times greater than CPDs[104], despite UV irradiation generating more CPDs than 6-4PPs[117-119]. These observations are all consistent with the notion that it is the helix distortion, not the specific damage, which is the basis for the GGR damage recognition step. Mutations in the genes encoding the damaged DNA binding protein heterodimer, consisting of the 128 (p128/DDB1) and 48 kDa (p48/DDB2) subunits, results in the XP group E disorder[120,121 122]. Although this heterodimer is not essential for NER *in vitro*[123], the XP-E phenotype demonstrates its importance *in vivo*[94]. It is thought that XPE may play a role in assisting XPC in the recognition of damage, particularly CPDs[116,124,125].

After damage recognition other factors that are crucial for NER are recruited; to verify the damage, XPA and replication protein A (RPA)[123,126] and to locally unwind a region of the DNA around the lesion, the TFIIH transcription factor[127,128]. XPA has the ability to bind UV-damaged DNA[129]. However that XPC-hHR23B initiates NER[114,115] suggests XPA acts downstream of XPC to synergistically recognise the presence of damage[130]. XPA's affinity for UV lesions increases with the extent of helical distortion[131] and the observation that XPC-hH23B binding to a lesion further increases helical distortion[114] is probably how XPA is recruited to the

site of damage. The interaction of XPA with other NER proteins including TFIIH[132] and RPA[133] leads to their recruitment to the complex. RPA is a heterotrimeric single stranded DNA (ssDNA) binding protein that is involved in DNA repair, recombination and replication[134]. Its function during NER is to bind the undamaged strand, essential for the full opening of the double helix[135]. Along with XPA, RPA is thought to also play a role in distinguishing the damaged and undamaged strands[130].

TFIIH is a multi-subunit protein complex involved in both NER and transcription initiation[136], containing 9 polypeptides[137,138] including the XPB[139] and XPD[140] helicases, which are essential for both processes. These helicases have opposite polarities with XPB unwinding DNA in a 3'-5' direction[141] and XPD 5'-3'[142] around the DNA damage. The action of the XPB and XPD helicases results in the formation of the NER open complex (figure 1.3.1c), however the complete opening of this region requires the presence of XPA, RPA and also XPG, which is the endonuclease that makes an incision 3' to the DNA damage[143,144].

1.3.4   Damage excision

The function of two DNA-structure specific endonuclease enzymes, XPG[145] and the XPF-ERCC1 heterodimer[146-148], results in the excision of a patch of DNA between 24 and 32 nucleotides in length[149]. XPG interacts with TFIIH[150] and makes an incision on the damaged strand 3' to the damage, recognising the splayed duplex structure formed during the initiation steps of NER[151]. XPG is found associated with the proliferating cell nuclear antigen (PCNA)[151], although PCNA appears to have little role in stimulating its activity in NER[151]. XPG's recruitment to the NER machinery is independent of XPA[115], however it is only in the presence of XPA, probably mediated

by RPA[148], that XPG makes the incision between 2-8 nucleotides from the damage[144]. This incision is followed by the XPF-ERCC1 endonuclease activity, approximately 15-25 nucleotides 5' to the damage[144,152]. XPF-ERCC1 is recruited to the NER open complex through its interactions with the XPA protein[115] and the concerted action of XPA and RPA is thought to facilitate the XPF-ERCC1 incision on the damaged strand[153].

1.3.5   Re-synthesis and ligation
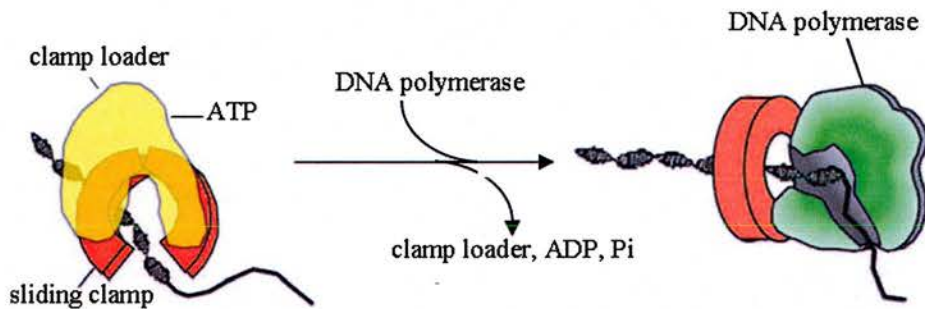
Figure 1.3.4: PCNA loading by the ATP-dependent clamp loader RFC. Once loaded onto the ds/ssDNA junction PCNA can recruit the replicative polymerase Polδ for initiation of replication. Adapted from Bowman *et al.*[154].

After the 3' and 5' incisions are made the damaged strand is thought to dissociate along with the NER proteins, with the exception of RPA. The excised ssDNA would be unable to sustain the NER complex of proteins so it is likely that at this point the complex falls apart, releasing the damaged DNA fragment and individual NER components. The damaged stretch of DNA would then be digested by nucleases and the NER proteins available for recruitment to other sites of UV-damage. RPA is likely to remain bound to the undamaged ssDNA to protect it from nucleases. Two

double stranded (ds)/ssDNA junctions are formed after excision, with the junction 5'
to the damage having ssDNA leaving in a 3'-5' direction. This junction is recognised
by replication factor C (RFC), a component of the DNA polymerase (Pol) δ or ε
holoenzyme[126]. RFC loads the PCNA sliding clamp[155], which recruits DNA
polymerase δ or ε (figure 1.3.4). The damaged segment can then be re-synthesised
and the resulting nick ligated by DNA ligase I (figure 1.3.3f and g)[156,157 158].

1.3.6 Transcription coupled repair
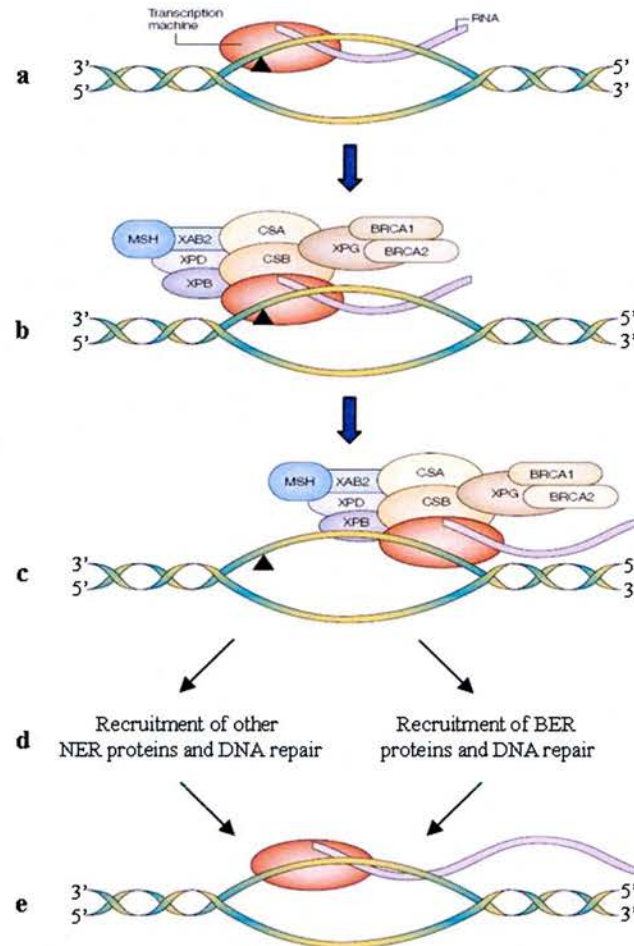
Figure 1.3.5: Transcription coupled repair of UV-damage. Stalling of RNAPII during transcription (a) results in the recruitment of a host of proteins (b) that is thought to back-track the polymerase, uncovering the lesion (c). NER proteins are then recruited to repair the lesion (d), as shown in figure 1.3.3c-f, before transcription is re-initiated (e). Taken from Friedberg[94].

Transcription coupled repair (TCR) (figure 1.3.5) has largely been studied as a sub-branch of NER, specifically involved in the removal of UV lesions on the transcribed strand of genes (see review[159]). The human disorders CS and TTD demonstrate the importance of TCR[94,106]. However the lack of a suitable *in vitro* system for studying

this process makes the actual mechanism of action less well understood than GGR[107]. Extensive studies using cells derived from CS and XP/CS patients have given the greatest insights into this process with certain mutations in 5 genes encoding; CSA, CSB, XPB, XPD and XPG, resulting in TCR defects[160]. The current model of TCR of UV-damage is that the proteins involved in GGR, discussed above, are conserved with the exception of the XPC-hHR23B and XPE proteins. It has been shown that the helix distortion introduced by CPDs and 4-6PPs, as well as other bulky lesions, act as a hindrance to transcription, resulting in the stalling of RNA polymerase II (RNAPII)[104,161-163]. It is thought that the stalled RNAPII acts as the initial DNA damage recognition step of TCR[104,112,164], resulting in the recruitment of NER proteins[94,165]. CPDs are more efficiently repaired by TCR than GGR[112,113] because their recognition is more efficient by RNAPII, which will stall at every CPD, than XPC-hHR23B, which has a relatively weak affinity for CPDs[105,116].

It is also possible that the packaging of DNA into chromatin plays a major role in determining the rate of NER *in vivo* (reviewed in[166]) as cell free experiments using human cell extracts show that DNA packaging onto nucleosomes inhibits the rate of NER[167]. *In vitro* experiments also demonstrate that the core GGR factors; XPC-hHR23b, XPA, RPA, TFIIH, XPG and XPF-ERCC1, which are sufficient for NER on naked DNA templates, do not repair 6-4PPs on nucleosomal DNA[168]. It is therefore evident that other factors are important for NER in a chromatin context *in vivo* and this is likely to involve chromatin remodelling machinery[166,169], with the XPE heterodimer implicated in this process[166,170,171]. Although DNA packaging into chromatin structure would affect GGR, no such problems would be associated with TCR as the DNA needs to be relaxed from chromatin for normal transcription. This could be a major factor influencing the faster rate of TCR of CPDs *in vivo*.

26

Key to the process of TCR is the action of the CSA and CSB proteins[118,172,173], which are essential for the restart of transcription after UV-damage[174], although the exact role of these proteins is unclear. CSA is thought to be important in facilitating protein-protein interactions due to the presence of five WD-40 repeats, interacting directly with RNAPII, CSB and a subunit of TFIIH[175]. CSB is a large, 168 kDa protein that belongs to the DNA-dependent ATPase SWI/SNF family of proteins that are involved in chromatin remodelling[176,177]. Similar to this family of proteins, CSB exhibits DNA dependent ATPase activity[178] and is able to remodel chromatin *in vitro*, providing a possible link to TCR and chromatin remodelling[179]. DNA damage results in the translocation of CSA to the nuclear matrix in a CSB dependent manner, co-localising with RNAPII[180], thought to be important for the initiation of TCR *in vivo*. CSB may be a component of the RNAPII elongation complex[178,181,182], with its indirect interaction with RNAPII *in vitro* important for the elongation of transcription on a template containing pause sites[183] and may function in TCR by recruiting TFIIH[184]. TFIIH is involved in transcription initiation[136] and is thought to be lost from the transcription elongation complex after approximately 50 bases[185,186]. The observation that CSB deficient cells can efficiently remove CPDs from areas close to RNA polymerase initiation start sites, but are unable to do so during transcriptional elongation[187] are consistent with its role in recruiting TFIIH to the site of stalled RNAPII[184,188].

Biochemical assays show that RNAPII stalled at a lesion actually impairs NER[104,162,189], presumably because it blocks the access of the repair proteins to the site of DNA damage. However the requirement of both XPA and RPA for TCR[101,190] suggests that upon recruitment of TFIIH to the site of stalled RNAPII, the lesion must be uncovered. This is hypothesised to be achieved by a backtracking of RNAPII

rather than its dissociation[104,191,192]. Backtracking of the RNAPII, possibly mediated by CSA, CSB and TFIIH proteins[159], would allow XPA and RPA access to the site of damage and NER could then proceed as discussed above for GGR[104,191,192] (figure1.3.3c-f). The backtracking of RNAPII may also involve one of two recently discovered, novel XPA binding proteins (XAB1 and 2), with XAB2 shown to interact with CSA, CSB and RNAPII[159,193]

As well as being important for TCR it has been shown that mutations in genes resulting in CS or TTD also lowers the overall levels of transcription[194-196], suggesting their importance in normal transcription. It is thought that the reason patients with CS and TTD do not exhibit the increased susceptibility to developing cancer after UV exposure (as seen for XP) is because any cells with significant amounts of damage will result in RNAPII stalling which, if TCR cannot occur, will activate pathways resulting in apoptosis[196]. Consistent with this idea, high levels of apoptosis are observed in CS patients and this may contribute to the rapid aging and other symptoms observed[195,196].

Until recently TCR has only been studied as a sub-branch of NER, to repair UV-damage and bulky lesions on transcribed strands of DNA. However cell lines from HNPCC patients, a disorder associated with MMR defects, are also deficient in TCR[197]. It has also been shown that TCR can repair oxidative lesions[198], usually repaired by BER. These results suggest that, rather than being a branch of NER, TCR may be a more general repair pathway. However this idea is more contentious as the involvement of BER proteins in TCR of oxidative lesions has yet to be shown and although MMR proteins from *E. coli* have been shown to be involved in TCR[197], such a role in eukaryotes is unclear[199-201]. Further, cell free systems also show that RNAPII does not stall upon encountering the 8-oxyguanine[202] and thymine glycol[203,204] lesions

usually repaired by BER. Thus TCR of such lesions would presumably require a different recognition step to classical TCR.

The description of NER above has focused on that in human cells, however homologues of most of the NER proteins are conserved in eukaryotes[86,103], with a few notable exceptions (4.1.4)[86,205]. In prokaryotes a NER pathway exists, although the proteins carrying out this process are not conserved. In *E. coli* the UvrAB complex is important for UV induced damage recognition[206] and the UvrC endonuclease makes the incisions both 3' and then 5' to the DNA damage[206,207]. The damaged fragment, of between 8-12 nucleotides[208], is then removed and the resulting nick ligated by the concerted action of the UvrD helicase, PolI holoenzyme and DNA ligase [209].

The process of NER in the third kingdom of life, archaea, is less well characterised and is discussed in Chapter 4.1.4.

Observations in a number of patients diagnosed with XP but did not have a defective NER pathway[210], XP-variants, hampered the elucidation of the NER pathway[86] It was not until 1999 that the gene responsible for the XP-V complementation group was elucidated, revealing an alternative pathway for coping with UV-damage within the cell.

1.3.7   Replication bypass of DNA lesions
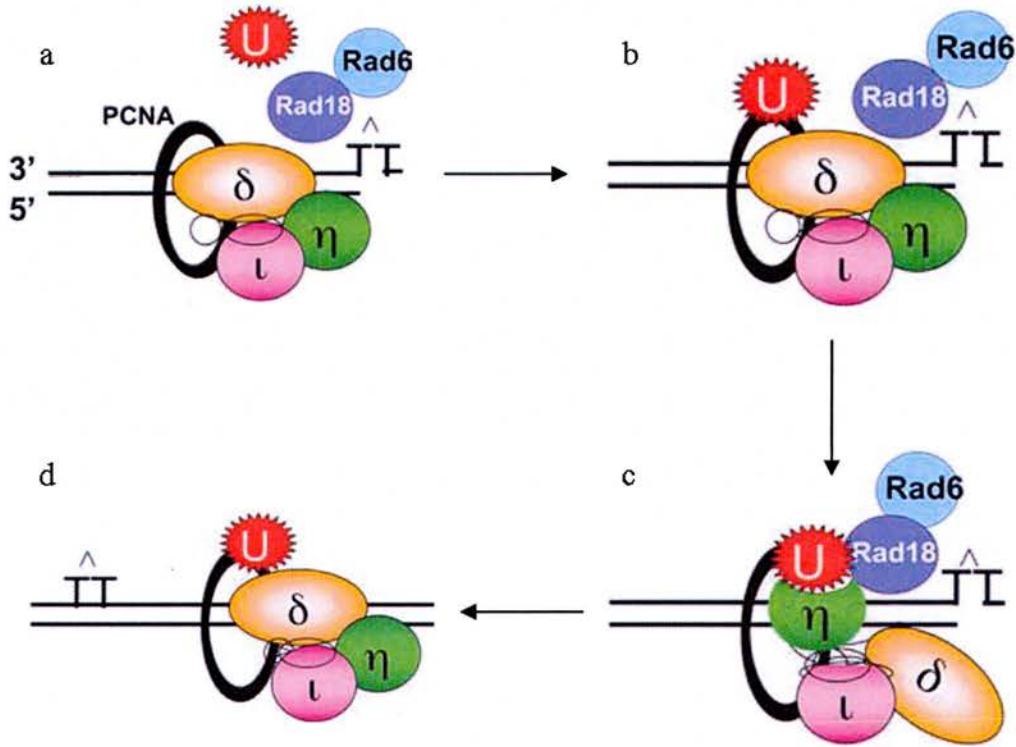
Figure 1.3.6:   Replication bypass of a CPD by translesion synthesis.   On encountering a CPD the replication fork stalls.   The cellular response to this results in the ubiquitination of PCNA, mediated by the Rad6 family of proteins[211] (a and b). The ubiquitination of PCNA increases its affinity for Polη, leading to a polymerase switch from Polδ to Polη (c).   Polη then replicates the CPD and a few extra bases, until the replication fork is free of the damage, before there is another polymerase switch back to Polδ, resuming normal replication (d).   It is thought that the damage polymerases may be part of the normal replication machinery and the activities of different damage polymerases (for example Polι) may be employed depending on the damage encountered.   Taken from Lehmann[212].

Post-replication repair is a term used for the observation that intact daughter strands can be synthesised in the presence of persisting DNA damage *in vivo*[213,214]. The predominant mechanism for replicating this damage in *E. coli* involves homologous recombination[215,216], however this is not the case for mammalian cells[213]. The key to understanding this process in eukaryotes was the identification of the gene that resulted in the XP-V phenotype, as cells from these patients were shown to be defective in post-replication repair[217-220]. It was not until 1999 that the gene product missing in XP-V patients was isolated and shown to be a DNA polymerase (Polη) capable of replicating thymine-thymine (*T-T*) CPDs[221], immediately suggesting a mechanism of post-replication repair of this damage in mammalian cells. The Polη polymerase was cloned and found to be homologous to radiation sensitive mutant (Rad) 30 in *S. cerevisiae*[222-224] and has since been extensively characterised *in vitro* (reviewed in[214,225,226]).

The discovery of Polη[222,223] led to the subsequent identification of several other similar polymerases including Polι from humans[227], the *E. coil dinB* and *umuC* gene products, PolIV and PolV respectively[223,228] and *dinB* homologues in humans (Polκ)[228] and archaea[229,230]. These polymerases have been grouped into the Y-family of polymerases and while they share sequence identity and similarity with one another they are unrelated in sequence to classical polymerases[214] Two other important DNA polymerases involved in replicating damage have also been identified, Polζ and Rev1[86,214]. The typical error rate of the DNA-damage polymerases is approximately 10^{-2} on both damaged and undamaged templates, compared to $10^{-5} - 10^{-7}$ for the replicative polymerase which is only active on undamaged DNA[223,224]. Due to their low fidelity they have been termed error-prone polymerases.

31

Although the error-prone polymerases have been well characterised *in vitro*, their *in vivo* function is less clear, with only Polη's role well understood[212,214]. An important question is why would the cell want to use these mutagenic polymerases? The XP-V phenotype, along with biochemical studies[231,232], suggests that the *in vivo* role of Polη is for the error-free replication of UV-induced CPDs. This is achieved as Polη appears to have evolved a preference for adding the correct nucleotides, two adenines, opposite the most common CPD, *T-T*[214,233]. Indeed it is thought that the use of this polymerase actually represents the predominant and most efficient way of repairing this damage *in vivo*[233]. The mutagenesis observed in XP-V individuals is likely to be due to the use of other damage-polymerases that replicate CPDs less efficiently and less accurately. It has been speculated that the role of the different damage-polymerases in the cell is to accurately replicate other DNA lesions; however there is currently little *in vivo* evidence to support this[212]. Importantly, the use of these polymerases must be under extremely tight control as their misuse during normal replication would have drastic consequences for genomic stability and cell viability.

The structural solution of the catalytic regions of *S. cerevisiae* Polη homologue Rad 30[234] and two *dinB* homologues from *S. solfataricus*[229,230], one in complex with DNA[229], have provided structural insights into the mechanism of action of the Y-family of DNA polymerases[235]. Despite no sequence similarity, the structures of these polymerases reveal that they have the finger, thumb and palm domains conserved in other DNA polymerases, as well an additional domain[229,230,234,235]. The key difference between the error-prone and high fidelity polymerases is that the active site of the Y-family of polymerases is slightly larger,

forming a more open conformation that implies a mechanism for the accommodation of altered/damaged bases within it[229,230,234,235].

In vivo studies looking at the localisation of Polη before and after UV exposure show that upon irradiation it moves from a more disperse nuclear location to discrete foci in the nucleus, with the foci co-localising with PCNA and replicating DNA [231,232]. This is consistent with the proposed mechanism of post-replication repair by these error-prone polymerases (figure 1.3.6). It is thought that the error prone polymerases are associated with the replication machinery[212]. Upon stalling of the replication fork, due to its encounter with DNA damage, the error-prone polymerase is utilised to synthesise the damaged DNA and just a handful of other bases, before reverting back to the high fidelity replication polymerase[101,236]. Key to this process is the interaction of these polymerases with PCNA, which increases their polymerase activity without increasing their processivity[237]. It is thought that PCNA plays a crucial role in mediating the switch between different polymerases required during lesion bypass and that this may be controlled by ubiquitination[211,214].

The elucidation of the pathways that have evolved to deal with UV-induced DNA damage within the cell reveal a fascinating story in which the processes of replication, transcription and NER play important roles. The emerging evidence suggesting TCR is not unique to damage repaired by NER and the identification of other error-prone polymerases that can replicate several types of lesions[214], suggest that this overlapping role of replication, transcription and repair in maintaining the genome is not unique to UV-induced damage.

Chapter 2

ArnA, a bifunctional enzyme involved in the biosynthesis of 4-amino-4-deoxy-L-arabinose:

Crystallisation, data collection and investigation of a protein-protein interaction

2.1. INTRODUCTION

2.1.1 Lipopolysaccharide
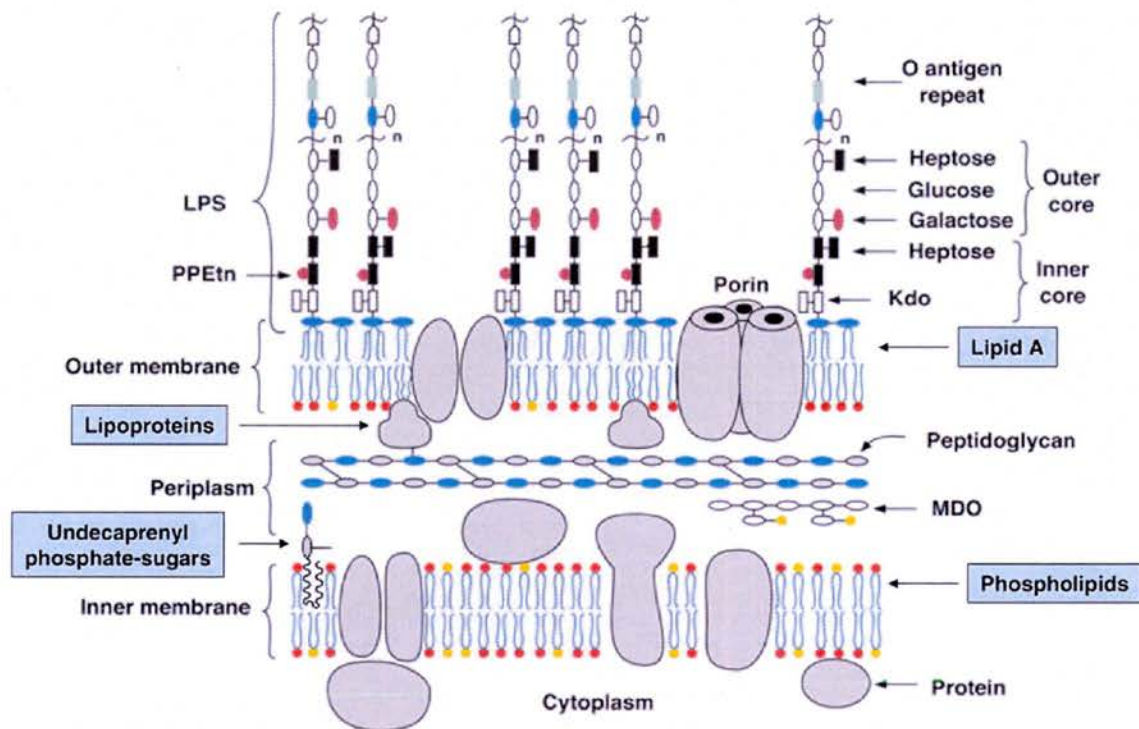
Figure 2.1.1: Schematic model of the inner and outer membranes of *E. coli* K12 taken from Raetz and Whitfield[238]. The LPS makes up the outer leaflet of the outer membrane and something similar exists in almost all Gram-negative bacteria[239].

Figure 2.1.1 shows a diagram of the membrane of *E. coli* K12, which is fairly typical for Gram-negative bacteria. It consists of inner and outer membranes separated by a thin peptidoglycan layer, with the Lipopolysaccharide (LPS) being attached to the outer monolayer of the asymmetric outer membrane[238,239]. The vast majority of Gram–negative bacteria have some recognisable form of the LPS[240-243] and the function of this is important to the survival of the organism[242,244]. In pathogens this role is particularly important; it provides a barrier which results in impermeability of the outer membrane to large molecules and hydrophobic compounds, including

antimicrobial agents[245-248] and also protection against the action of phagocytic cells[249]. In addition, the LPS facilitates interactions between the bacterium and its host environment, which are often important in establishing infection[250,251] and avoiding the host's immune response[251,252]. The LPS consists of 3 regions: the outer segment is called the O-antigen and is made up of repeating polysaccharide units[253-255]; it is this region that first interacts with the host[256,257]. The composition of the repeating polysaccharide units varies greatly between different bacterial species[258,259] and even strains[260-262]. It is this variability that results in the wide range of modes of pathogenicity and host targets of Gram-negative bacteria[256-258].

The middle section of the LPS consists of the core oligosaccharide unit[263], a short chain of sugars which, with minor variations[264], are common between species but vary from genus to genus[258,264,265]. The core and O-antigen are attached to the outer membrane through the third region of LPS, the hydrophobic anchor Lipid A[239,242] which is an important structural component of the cell membrane[266]. The structure of Lipid A (figure 2.1.2 and 2.1.6) is relatively well conserved between almost all Gram–negative bacteria[238,242,267]. The components of the LPS are biosynthesised by enzymes within the cytosol or the inner surface of the inner membrane[242,268,269] and are subsequently transported out of the cell, initially to the outer surface of the inner membrane[239]. This process is not well characterised but it is hypothesised that the essential inner membrane protein MsbA[270], structures of which have recently been published[271,272], may transport Lipid A[269,270,273] and the flippase RfbX is a possible candidate for export of the O-antigen[274]. After assembly of the LPS at the inner membrane[275-277] it is transported through the periplasm and integrated into the outer leaflet of the outer membrane by an unknown process[239]. In contrast, the biosynthesis of the individual components has been relatively well characterised

(reviewed in[238,255,264,278,279]), particularly for Lipid A which was first described in 1984[280-282].

2.1.2   Lipid A and its Biosynthesis

The structure of Lipid A from *E. coli* K12 is shown in figures 2.1.2 and 2.1.6. It consists of a β-1',6-linked disaccharide of glucosamine, phosphorylated at the 1'- and 4'-positions and acylated at the 2-, 3-, 2'- and 3'- positions with R-3-hydroxymyristate[242,283,284]. Glycosylation of Lipid A at the 6'- position with two 3-deoxy-D-manno-octulosonic acid (kdo) moieties represents a minimum form of the LPS that is required for growth of *E. coli* and most Gram-negative bacteria[239,285]. There are some exceptions including: the survival of *Thermotoga maritima*[286] and *Neisseria meningitides* Type B[287] mutants that are unable to biosynthesise Lipid A and the complete absence of Lipid A biosynthetic genes in some spirochetes[288,289].

The biosynthesis of *E. coli* K12 Lipid A, shown in figure 2.1.2, has been well characterised and is used as a model system as recognisable *lpx* genes are present in almost all Gram-negative bacterial genomes sequenced[239].
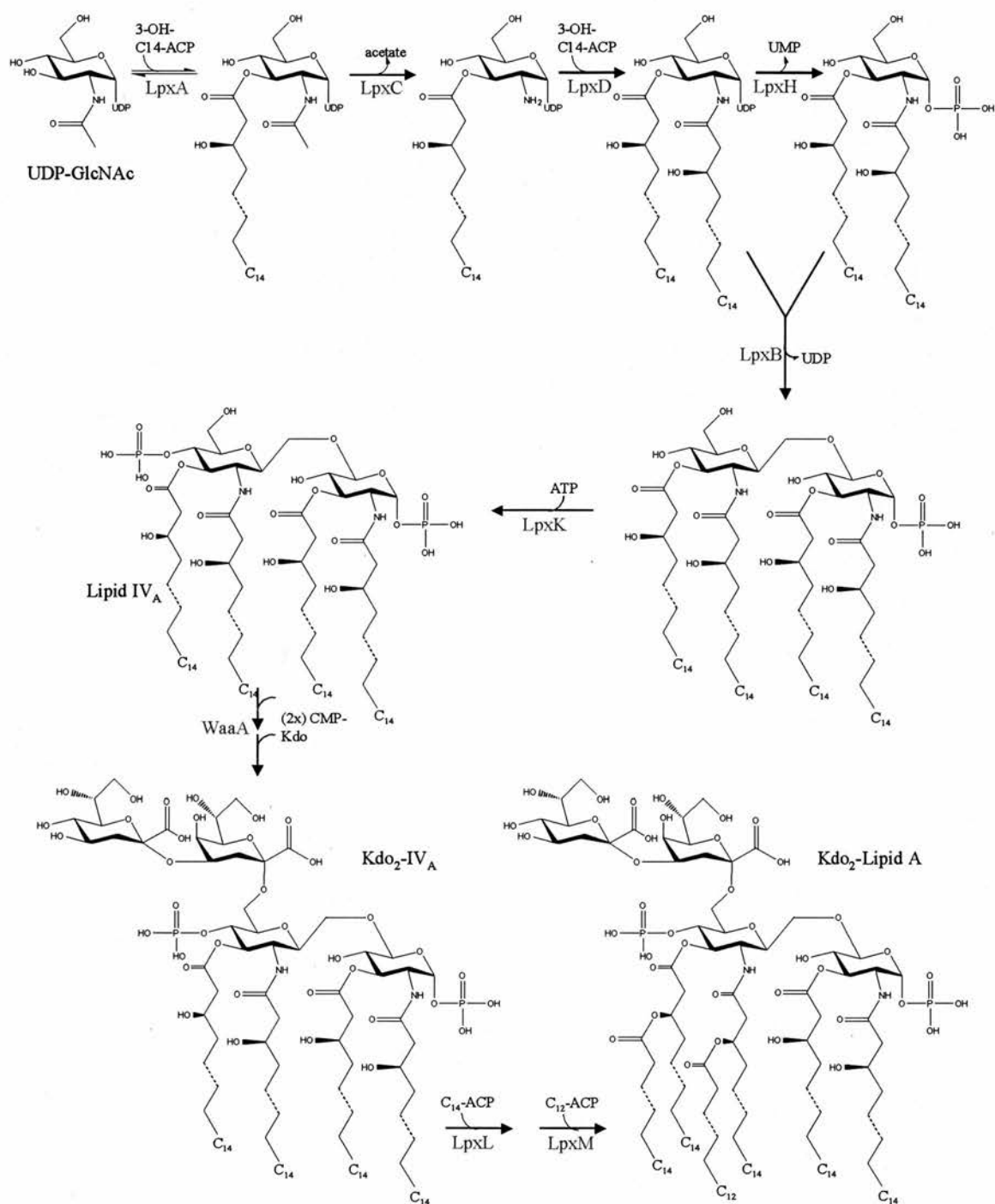
Figure 2.1.2: Biosynthetic pathway for the biosynthesis of Lipid A, enzymes involved are labelled blue. Adapted from Raetz and Whitfield[238].

The first step of Lipid A biosynthesis involves the acylation of UDP- (uracil diphosphate) N-acetyl-D-glucosamine (UDP GlcNAc), catalysed by the enzyme

LpxA, which requires acyl carrier protein thioester as a donor substrate and is selective for β-hydroxymyristate[290,291]. LpxA has an interesting structure (figure 2.1.3), which represents a novel fold, termed a left-handed parallel β helix[292,293]. This fold is formed from multiple contiguous hexad repeats in the sequence and violates the protein folding constraints for right handed crossovers between strands of parallel β-sheets[293]. The active site of this enzyme lies between coils in adjacent subunits and the length of these coils acts as a hydrocarbon ruler[294]. In *E.-coli* there are 10 coils stacked upon one another, which specifically incorporates 14 carbon acyl chain[295], with sequence differences in other bacteria leading to the preferential incorporation of carbon acyl chains of different length[294,296].
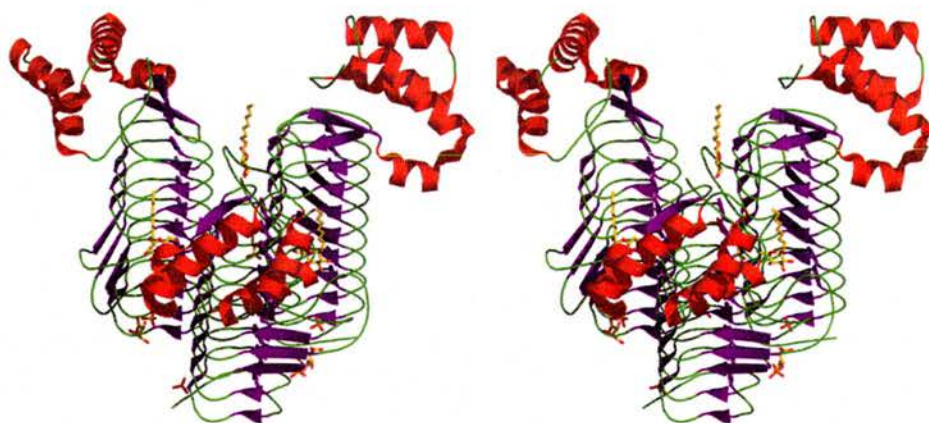
Figure 2.1.3: Stereo view of the LpxA structure from *Helicobacter pylori* in complex with 1-s-octyl-β-D-thioglucoside, tartaric acid and a sulphate ion (PDB code 1J2Z)[292].

The LpxA product, UDP-3-O-(acyl)-GlcNA, is then deacetylated by the deacetylase LpxC, which is a zinc metalloenzyme[297]. The equilibrium of the LpxA reaction strongly favours substrate[290] and it is the LpxC reaction that is the committed step of Lipid A biosynthesis[290,297,298]. LpxC bears no sequence resemblance to other deacetylases and has been shown to be an excellent drug target[285,299].

LpxD is the next enzyme acting in this pathway and it incorporates a second β-hydroxymyristate moiety, forming UDP-2,3-diacylglucosamine[300]. The highly selective pyrophosphatase, LpxH, then cleaves off the nucleotide and adds pyrophosphate yielding 2,3-diacylglucosamine-1-phosphate, commonly named Lipid X[301]. LpxH is one of the least well conserved *lpx* genes in Gram-negative bacteria, present in approximately 70 % of sequenced genomes and it is thought additional isoenzymes must exist in bacteria lacking LpxH to catalyse UDP-2,3-diacylglucosamine hydrolysis[238].

The disaccharide synthase enzyme LpxB then makes the β,1'-6'- linked disaccharide by condensation of UDP-2,3-diacylglucosamine with lipid X[280,281]. The LpxK kinase then phosphorylates the disaccharide at the 4' position, to produce Lipid IVA[302].

Lipid IVA is then transported to the inner membrane where the bifunctional enzyme WaaA transfers the two kdo moieties[303] from CMP-kdo, synthesised by KdsA and KdsB[304,305], producing Kdo_2-lipid IVA.

The final stages in the biosynthesis are specific to kdo modified Lipid IVA and involves the transfer of lauroyl and myristoyl residues, by the respective transferases LpxL and LpxM, to the distal glucosamine unit[306-308]. The Kdo modified Lipid A is also known as the endotoxin as it is this molecule that activates the host immune response[239,242,309].

2.1.3 The Immune Response to Lipid A

During infection by Gram–negative bacteria the LPS-binding protein associates with Lipid A[310] and this complex is detected at picomolar levels by CD14 receptors on

macrophages and endothelial cells[311-314]. This results in a signal transduction cascade that activates the host's innate immune response (figure 2.1.4)[242,312,313]. This triggers the production of diverse mediators of inflammation including cytokines[315,316], chemokines[317] and cellular adhesion molecules[318]. The overproduction of these mediators and clotting factors can damage small blood vessels and may result in Gram–negative sepsis[319]. This is accompanied by disseminated intravascular coagulation and in severe cases can lead to organ failure and even death[319]. Another important class of molecules produced by the innate immune response are cationic antimicrobial peptides (CAMPs)[320,321], which can directly target bacteria through interactions with Lipid A[46,322,323].
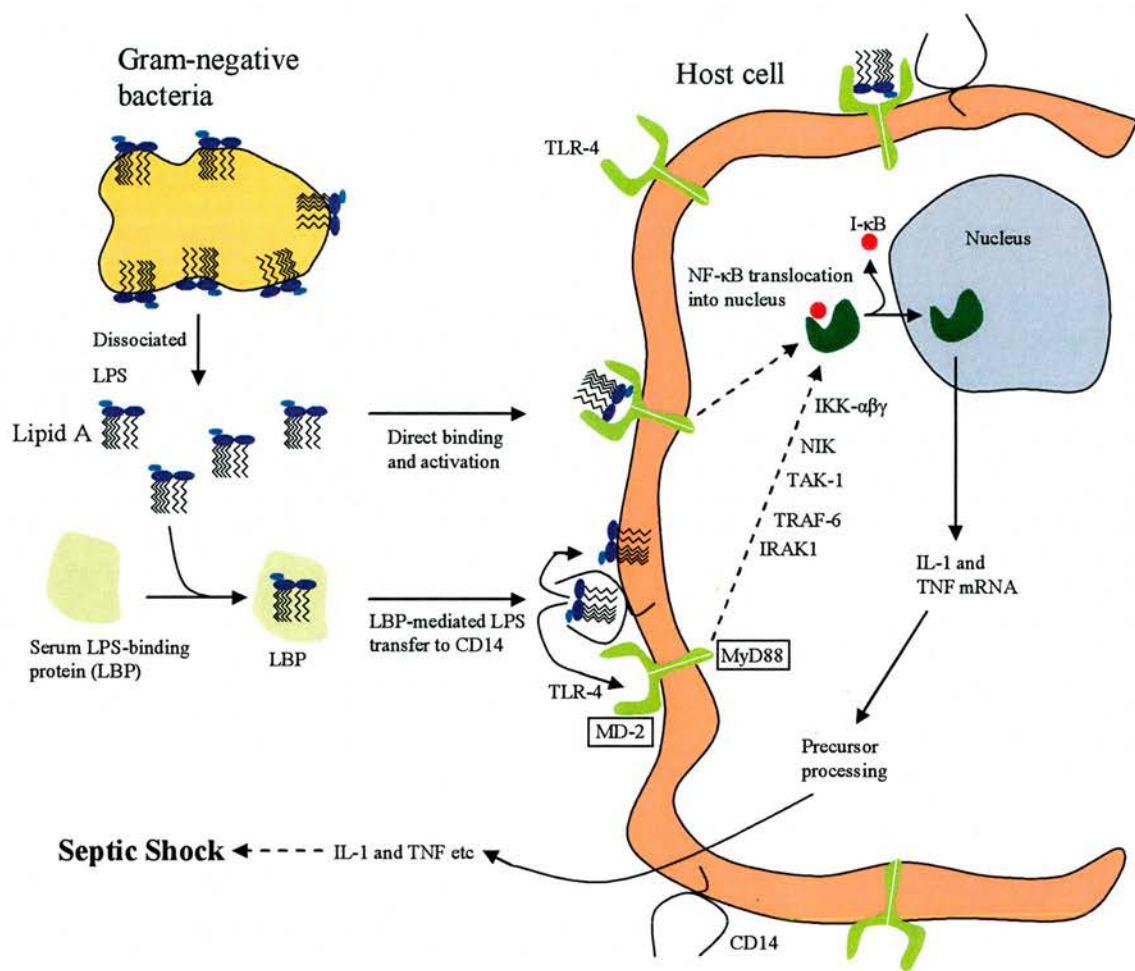
Figure 2.1.4: Immune response to Lipid A adapted from Raetz and Whitfield[238] and based on research using mouse[324] and human cells[325,326].

2.1.4   Cationic Antimicrobial peptides

As well as being important in innate immunity[46,322,323] CAMPs are widespread in nature and are produced in several human cell types including those of the digestive tract[327], skin[328] and lungs[329,330]. They also represent a class of antibiotics that include Polymyxin B, used in some creams and lotions[331]. CAMPs are small, amphiphatic, positively charged peptides [320], which have a whole range of antimicrobial activities that can kill a diverse range of organisms including bacteria[332], fungi[333], and eukaryotic parasites[45,334]. CAMPs target cells through electrostatic interactions with

acidic phospholipids in the outer membrane (figure 2.1.5)[335] and in Gram–negative bacteria this is specifically through affinity with phosphate groups of the Lipid A molecule[46,335,336]. Upon interacting with Lipid A, the amphiphatic nature of CAMPs allows them to traverse the outer membrane, with several CAMPs coming together to form a pore in the membrane, resulting in solubilisation and ultimately death of the bacterium[46,323,335,336].
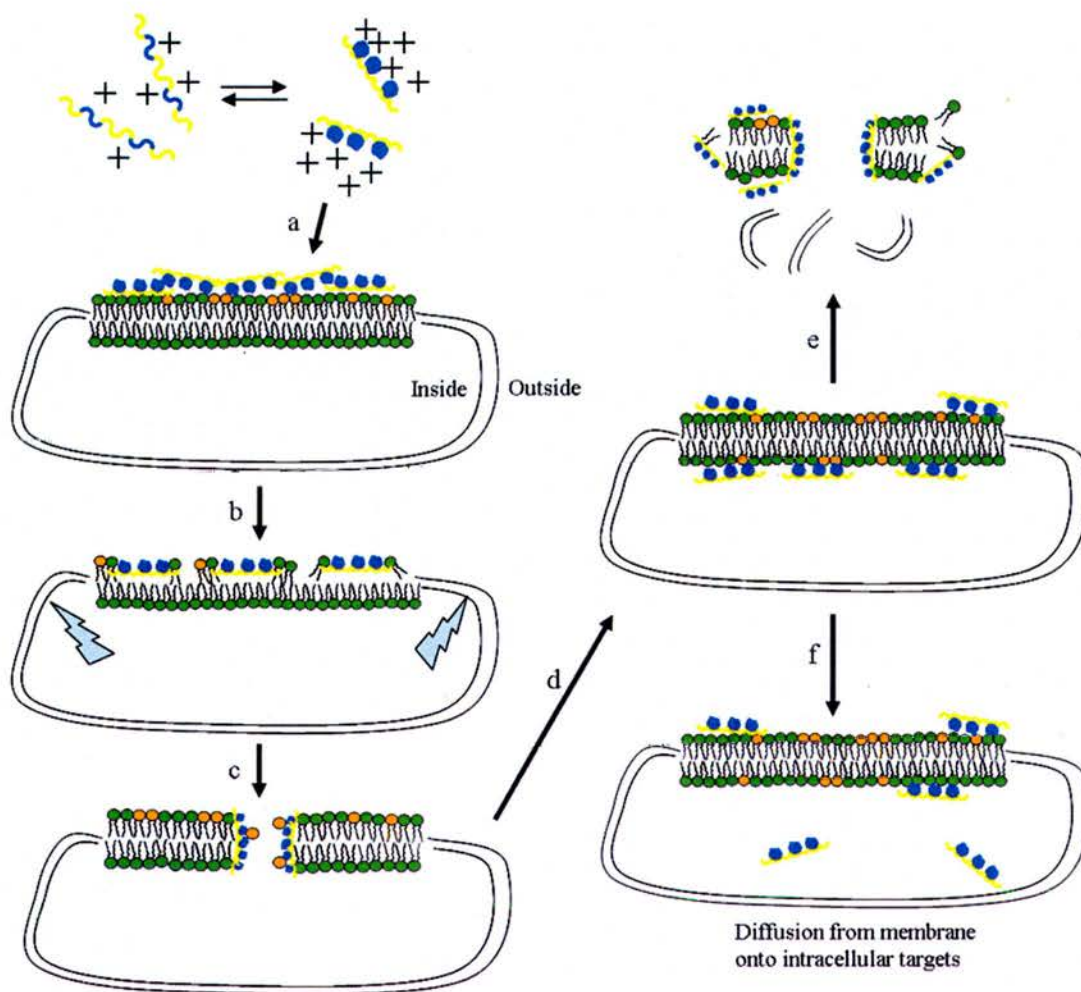
Figure 2.1.5: Mechanism of bacterial cell permealisation by the action of CAMPs, adapted from Zasloff[46]. Most cationic antimicrobial peptides are amphiphatic molecules with charged amino acids on one side (blue spheres) and hydrophobic side chains on the other (yellow rods), allowing them to interact with the microbial membrane (a and b). Electrostatic interactions and the transmembrane bioelectric field result in a 'carpet' of peptides on the membrane surface, which strains the membrane (b, shown by lightning bolts). The strain is lowered by the transition of the peptides into another arrangement (c), forming membrane pores. If the concentration of CAMPs is high enough this results in lysis (e). It is thought that sub-lethal levels of CAMPs may allow them to enter the cell without killing it, resulting in downstream effects such as alteration of transcription.

2.1.5 Regulated modifications of Lipid A
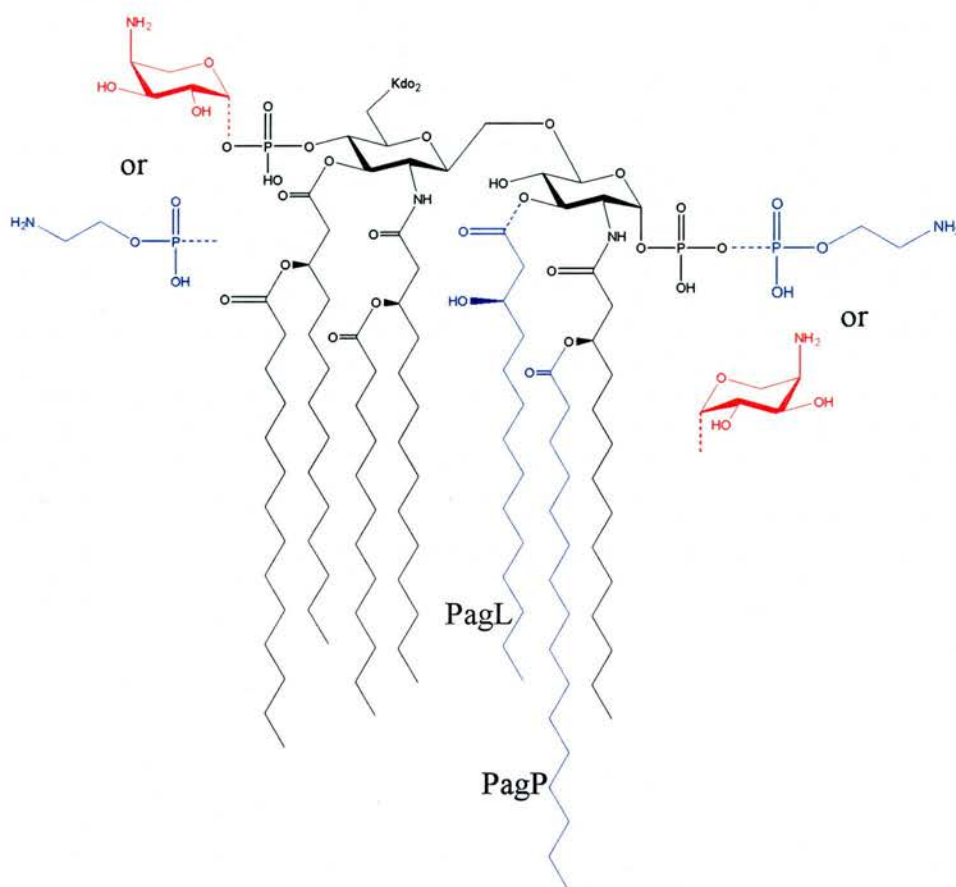
Figure 2.1.6: Structure of the Lipid A endotoxin with regulated modifications shown by dashed lines, adapted from Raetz and Whitfield[238]. The modifications of the 4' or 6' phosphate groups with L-Ara4N or phosphatidylethanolamine[337-339] are shown as is the addition of palmitoyl by PagP and removal of hydroxymyristate by PagL. Modifications coloured in blue and red are under the primary control of PhoP/PhoQ and PmrA/PmrB respectively.

The molecular structure of Lipid A can be covalently modified in response to environmental stimuli, altering the outer surface of the bacterium. This allows adaptation to the local environment, which can vary considerably during infection[238,279]. These modifications of Lipid A are under the general control of the PhoP/PhoQ two component regulatory system[338], which is required for *Salmonella*

spp. virulence[338,340]. Environmental changes, such as the low magnesium ion concentrations found in phagolysosomes[341,342], are detected by PhoQ, a sensory kinase, resulting in its activation and the phosphorylation of PhoP[340]. PhoP is a transcription factor and when phosphorylated it regulates the expression of over 40 genes[340]. As part of this adaptation to the environment, the PhoP/PhoQ system controls the resistance of *S. typhimurium* to CAMPs including defensin NP1[343], magainin-2[344], melittin[343] and those contained in crude neutrophil granule extracts[345].

These modifications usually include the removal or decoration of phosphate groups or fatty acyl chains and are carried out by additional enzymes to the nine required for Lipid A-Kdo biosynthesis in *E. coli*. Alterations in the acylation pattern of Lipid A has an impact on the fluidity of the outer membrane and is important in adaptation to different temperatures[346-348]. For example, growth of *E. coli* at 12 °C results in the replacement of LpxL activity with LpxP, a different acetyltransferase that adds an acyloxyacyl linked palmitoleate to the 2' position[347]. An outer membrane protein PagP, a serine hydrolase, has also been identified [342] that can transfer a palmitoyl group to the glucosamine 2' position (fgure 2.1.5)[349], resulting in hepta-acylated Lipid A[350]. PagP mutants of *S. typhimurium* show sensitivity to certain CAMPs of the innate immune response[350] and it is thought that this modification leads to tighter packing of the LPS and an increase in the integrity of the outer membrane[338,350].

It has been shown that the two most important features of Lipid A that result in activation of the hosts immune response, both *in vivo* and *in vitro*, are the phosphate groups and fatty acyl chains[306,314,351,352]. These experiments demonstrate that, despite the relatively well conserved structure of Lipid A, relatively small changes in either fatty acid composition or to the phosphate groups can impact on the

detection of LPS by the hosts defence system and it is unsurprising that evolution has taken advantage of this. For example, *Salmonella* strains contain a lipase, PagL that removes the R-3-hydroxymyristate from Lipid A (figure 2.1.5)[353]. It is thought that this enzyme may also catalyse deacylation of Lipid A, resulting in Lipid A species that illicit a reduced immune response by the host, thereby prolonging infection[279,353]. Further, 3-*O*-deacylase activities have been detected in the pathogen *Pseudomonas aeruginosa*[354] and may also be present in *Helicobacter pylori*[355,356] and *Porphyromonas gingivalis*[241,357]. There are also several examples of Lipid A that lacks either one or both phosphate groups in their Lipid A including *Rhizobium leguminosarum* [306,358,359] and *H. pylori*[355,356]. Additionally phosphatases that remove the 1' phosphate of Lipid A have been discovered in *R. leguminosarum*[360] and *H. pylori*[279] and orthologues of these enzymes may exist in the pathogens *Francisella tularensis*, *Brucella melitenisis* and *Legionella pneumophila*[360]. Removal of the phosphate groups from Lipid A not only alters the immune response during infection, but also results in a lower affinity for CAMPs [279,338]. Mechanisms of resistance to CAMPs have also evolved where, instead of removal, the 1' and/or 4' Lipid A phosphate groups are masked with either phosphatidylethanolamine and/or 4-amino-4-deoxy-L-arabinose (L-Ara4N)[337,361], neutralising their negative charges. This modification is also under the general control of the PhoP/PhoQ system[362], but is more tightly regulated by the iron sensing two component regulatory system PmrA/PmrB[337,362,363].

PmrA is a transcription factor and its activation in *S. typhimurium* by either constitutive mutation[337], PhoP[362], extra-cytoplasmic iron levels through the iron sensing protein PmrB[364] or mildly acidic conditions[365], results in elevated levels of phosphatidylethanolamine/L-Ara4N Lipid A modifications and resistance to the

action of certain CAMPs[343,361,366]. This modification is important *in vivo* and it has been shown that not only do *pmrA* constitutive mutants of *S. typhimurium* survive longer inside neutrophils[367], but *pmrA* deficient mutants have reduced virulence[368]. How PmrA regulates the addition of phosphatidylethanolamine is still unclear although an enzyme transferring this moiety to Lipid A phosphates has recently been detected in *E. coli* membranes[279] and the *lptA* gene from *N. meningitides* is necessary for this addition[369,370]. In contrast the PmrA regulated addition of L-Ara4N to Lipid A is well understood; in 1998 Gunn *et al.* first described 2 gene loci, *pmrE* and *pmrF*, that are activated by PmrA, resulting in this modification[363].

2.1.6   L-Ara4N modification of Lipid A

The *pmrE* locus encodes the well characterised UDP-glucose dehydrogenase [363] (Ugd) that catalyses the production of UDP-glucuronic acid (UDP-GlcUA) from UDP-Glucose (UDP-Glc)[371,372]. The *pmrF* locus is a multicistronic operon encoding 7 genes *pmrHFIJKLM*[363], all of which, with the exception of *pmrM*, along with Ugd are essential for L-Ara4N modification of Lipid A [363,368]. These latent genes have also been identified in *E. coli* K12[371] and the pathogens *Yersinia pestis*[373], *Bordetella pertussis*[374] and *P. aeruginosa*[375]. The discovery of these genes[363] allowed the hypothesis of a biosynthetic pathway for L-Ara4N and its addition to Lipid A, based on the similarity of the encoded proteins to other enzymes[349,371] and this pathway has since been largely validated (figure 2.1.7)[349,371,376-380].
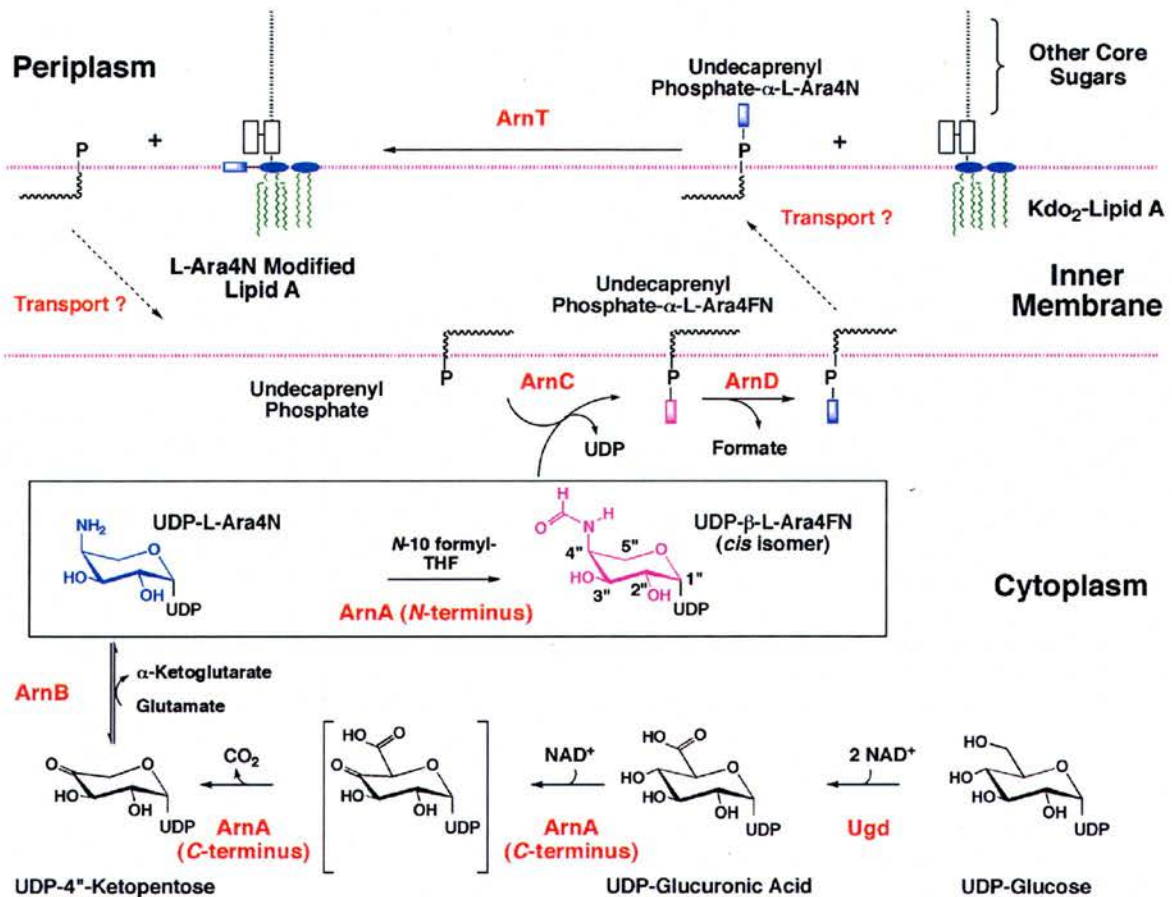
Figure 2.1.7: Currently proposed pathway for the biosynthesis of L-Ara4N from UDP-Glc, its transport to the outer membrane and its subsequent addition to Lipid A. Taken from Williams *et al.*[381].

The first unique step in the biosynthesis of L-Ara4N is carried out by the C-terminal decarboxylase domain of ArnA[377]. The sequence of ArnA predicts that it belongs to the SDR superfamily of proteins that are characterised by an initial hydride transfer step carried out by a conserved triad of S/T, Y and K (3.1.8)[382]. Based on this homology it was originally predicted that ArnA oxidises UDP-GlcUA and that there would be a subsequent decarboxylation reaction to produce UDP-β-L-threo-pentapyranosyl-4'-ulose (L-Ara4O)[377]. This decarboxylation step was thought to be catalysed by the enzyme ArnJ[371,383]. However, Breazeale *et al.* showed biochemically that ArnA alone was sufficient for the oxidative decarboxylation of UDP-GlcUA,

suggesting that the decarboxylation step was either spontaneous or catalysed by ArnA[377].

UDP-L-Ara4O acts as the substrate for ArnB, a pyridoxyl dependent aminotransferase, which transfers an amino group from glutamate to the C4' ketone, producing UDP-LAra4N[380,384]. Noland *et al.* have solved the structure of ArnB in complex with PLP and PMP cofactors (figure 2.1.8)[384] and its overall fold is similar to type 1 aminotransferases such as aspartate aminotransferase[385].
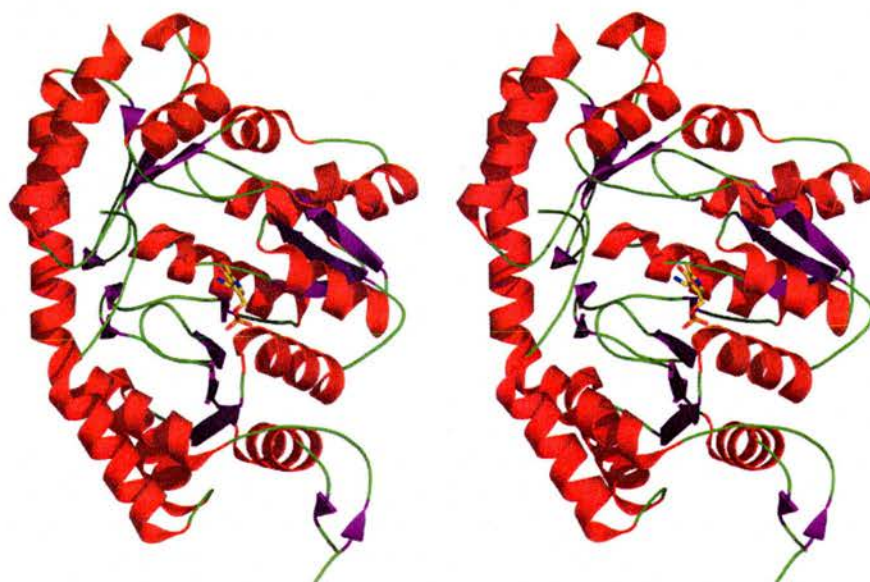
Figure 2.1.8: Stereo view of the ArnB structure in complex with PLP (PDB code: 1MDX). This enzyme belongs to the type 1 aminotransferase superfamily[384,385].

The activities of the C-terminal domain of ArnA and ArnB produce the L-Ara4N carbohydrate attached to the nucleotide UDP it therefore remains for this carbohydrate to be transferred from the cytosol to the outer surface of the inner membrane where it is added to Lipid A. The next step in the pathway was originally unclear[377]. ArnA has an N-terminal domain, which shares homology with formyltransferase enzymes and is conserved between all bacteria known to modify

their Lipid A with L-Ara4N[377,381]. It was shown that ArnA could transfer a formyl group from N-10-fTHF to UDP-L-Ara4N *in vitro*. However until recently it was unknown whether this reaction was essential *in vivo*[376,377]. Breazeale *et al.* cloned the ArnA C-terminal decarboxylase and N-terminal formyltansferase as two separate proteins and showed that in ArnA defective *E. coli* mutants, L-Ara4N modification of Lipid A could only be restored if both ArnA domains were reintroduced[376]. Further, they have developed an enzymatic assay for the next enzyme in the pathway, ArnC and shown it only uses UDP-L-Ara4fN and not UDP-L-Ara4N as a substrate[376]. ArnC is a membrane protein and it transfers the L-Ara4fN moiety from UDP to undecaprenyl phosphate[376]. This is then transferred to the outer surface of the inner membrane by an as yet undetermined pathway and at some point it is deformylated by the ArnD deformylase[376]. The final step of the pathway requires the transfer of L-Ara4N, from the novel carrier-lipid undecaprenyl phosphate-α-L-Ara4N, to one of the Lipid A phosphate groups[386]. This is catalysed by ArnT, an inner membrane arabinotransferase[378].

2.1.7   Lipid A biosynthesis/modification as an antibiotic target

Enzymes involved in Lipid A biosynthesis are good antibiotic targets as, if you stop its biosynthesis, you will stop the growth and virulence of the bacteria[245,387,388] and inhibition of this pathway results in increased sensitivity to other antibiotics[285]. The case of LpxC really highlights the potential of these enzymes as antibiotic targets, with inhibitors of LpxC from *E. coli* and related bacteria such as *S. typhimurium* having potent inhibitory effects of around 50 nM[285,389]. Broader ranging inhibitors that are effective against LpxC from *E. coli* through to the most diverse LpxC in *A.*

aeolicus have also been developed, but are not yet fully optimised[299]. The recent solution of the LpxC structure from *A. aeolicus*[390,391] should help in the development of even more potent, broad ranging inhibitors of LpxC that may be useful in future therapies.

Although enzymes that carry out regulated modifications of Lipid A are only conserved between a relatively small subset of Gram-negative bacteria and are not essential for their growth, they are also good drug targets due to their importance in avoiding the immune response. This is highlighted by the clinical isolation of *P. aeruginosa* with L-Ara4N modified Lipid A, from the lungs of cystic fibrosis patients, with associated increase in the inflammatory response and CAMP resistance[392,393]. As the overall effect of these modifications is to reduce the Lipid A affinity for CAMPs, inhibition of the enzymes responsible would lead to greater susceptibility to these antimicrobial agents, thus helping the host immune response and would also be useful as a combination therapy. All of the enzymes involved in the pathway of L-Ara4N biosynthesis and its addition to Lipid A are therefore potential drug targets, with the accessibility of soluble proteins to X-ray crystallography singling out ArnA and ArnB for structural studies with an eventual aim for rational drug design.

2.2 MATERIALS AND METHODS/RESULTS

2.2.1 Over-expression of ArnA constructs and ArnB
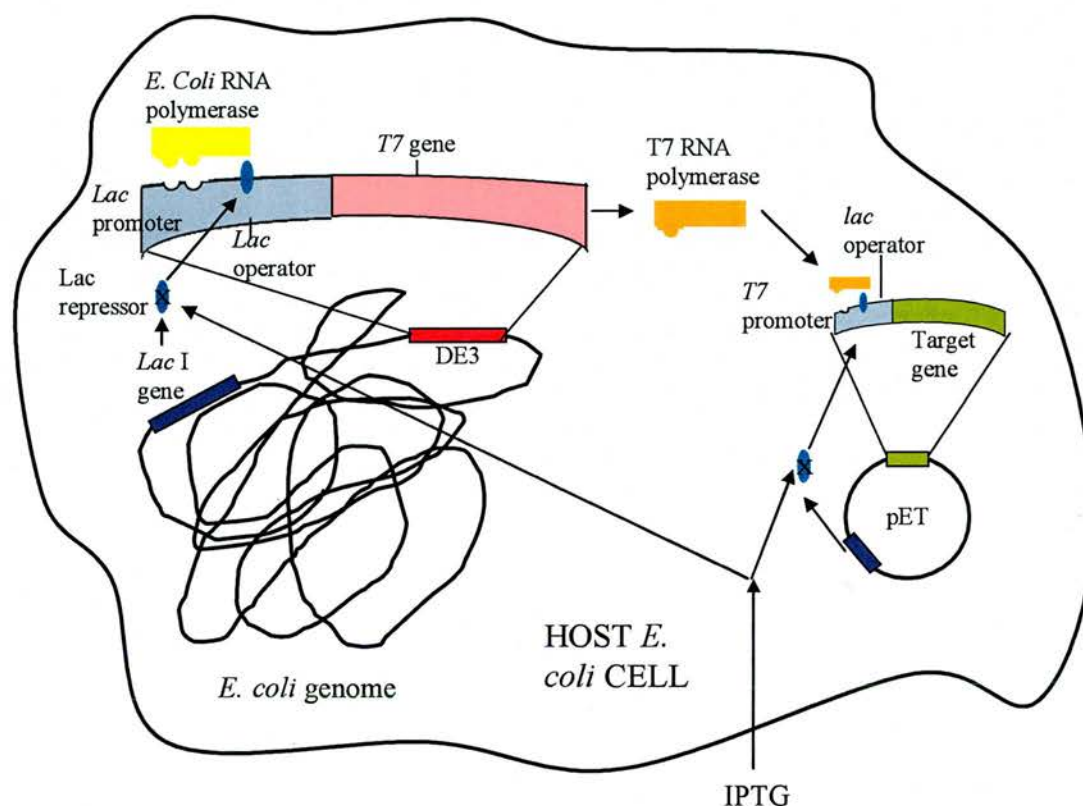
Figure 2.2.1: Protein over-expression using the pET expression system. *E. coli* used for protein expression, for example BL21 (DE3) (the designation DE3 indicates that the host is a lysogen of λDE3, and therefore carries a chromosomal copy of the T7 RNA polymerase gene under control of the *lacUV5* promoter. The pET plasmid includes the *lacI* and target genes, with the target gene been cloned immediately downstream of the *T7* promoter and lac operator. Under normal conditions expression from the *Lac* and *T7* promoters is repressed by the Lac repressor. Induction by the lactose analogue IPTG inhibits the Lac repressor, allowing expression of T7 RNA polymerase and subsequently the target gene. Adapted from the pET handbook from Novagen.

Table 2.2.2 details the full length ArnA, decarboxylase and formyltransferase domains of ArnA and ArnB constructs we received from our collaborators in North Carolina. The following procedures successfully produced soluble protein from all constructs. *E. coli* BL21(DE3) cells (genotype F-, *ompT*, *hsdS*$_\beta$(r_β-m$_\beta$-), *dcm*, *gal*, (DE3) tonA) were transformed with plasmid DNA encoding the relevant protein to be over-expressed. Single colonies were grown in 10 ml Luria broth (LB) supplemented with 50 μgml^{-1} kanamycin overnight. Overnight cultures were used to inoculate 500 ml of LB in 2 l flasks and grown to an OD$_{600}$ of 0.6-0.8 at 37 °C. The temperature was then reduced to 18 °C and protein over-expression was achieved by overnight induction with 0.5 mM isopropyl-ß-D-thiogalactopyranoside (IPTG) (figure 2.2.1). Cells were harvested at 10500 x g and those used to over-express histidine tagged and native protein were re-suspended in 50 mM HEPES pH 8.0, 10 mM imidazole and 300 mM NaCl or 50 mM HEPES ph 8.0 and 50 mM NaCl respectively. Soluble proteins were extracted by incubation at room temperature for 1 hour with 100 μgml^{-1} lysozyme and 20 μgml^{-1} DNase (Sigma) before 4 x 30 sec sonication passes on ice and subsequent centrifugation for 30 min at 75500 x g.

Protein	Vector	Tag	Restriction Sites and modifications	Antibiotic Resistance	Amino Acids Cloned
ArnA	pET 28b	N6H	NdeI, XhoI	Kanamycin	M_1- N_{660}
ArnA	pET 24b	C6H	NdeI, XhoI Stop codon mutated to L	Kanamycin	M_1- N_{660}
ArnA	pET 24b	-	NdeI, XhoI	Kanamycin	M_1- N_{660}
ArnB	pET 28b	N6H	NdeI, XhoI	Kanamycin	M_1- N_{660}
ArnB	pET 24b	-	NdeI, XhoI	Kanamycin	M_1-H_{393}
ArnA decarboxylase domain	pET 24b	-	NdeI, XhoI	Kanamycin	M_1-N_{305}
ArnA formyltransferase domain	pET 24b	-	NdeI, XhoI L_{316} mutated to M	Kanamycin	M_{316}- N_{660}

Table 2.2.2: Constructs obtained from our collaborators in North Carolina. All genes are from *E. coli* K12. N6H and C6H represents N- and C-terminal 6 x histidine tagged constructs, respectively. pET vectors were obtained from Novagen.

2.2.2   Purification

Tagged Constructs

The same procedure was used to purify all tagged proteins. Supernatant containing target protein was applied to a charged Histrap Nickel Sepharose, high performance column (Amersham Biosciences), which had been equilibrated with 50 mM HEPES pH 8.0, 300 mM NaCl and 10 mM imidazole. Unbound and weakly bound proteins were removed by extensive washing with buffer including 10 mM and then 80 mM imidazole. Essentially pure target protein was eluted with an imidazole gradient from 80 - 300 mM. Eluted protein was dialysed against 50 mM HEPES pH 8.0, 300 mM

NaCl and 2 mM dithiothreitol (DTT), before concentration and further purification by S200 gel filtration.

Untagged Constructs

The purification protocols of all four native proteins were very similar. Supernatant containing target protein was first applied to an HQ anion exchange affinity column (Applied Biosystems) and bound proteins were eluted with an increasing concentration of NaCl (0.05-1 M). ArnB, full length ArnA and the decarboxylase domain of ArnA all bound to the anion exchange column, whereas the formyltransferase domain of ArnA did not bind. NaCl was added to a final concentration of 4.5 M to fractions containing the target protein and then applied to an HP hydrophobic column (Applied Biosystems). The target protein was eluted with a decreasing NaCl gradient, dialysed into 50 mM HEPES pH 8.0 containing 300 mM NaCl and further purified by S200 gel filtration chromatography (Amersham Biosciences). Each step of purification was monitored by sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE). Figure 2.2.3 shows pure ArnA proteins after S200 gel filtration, as judged by Coomassie blue staining.
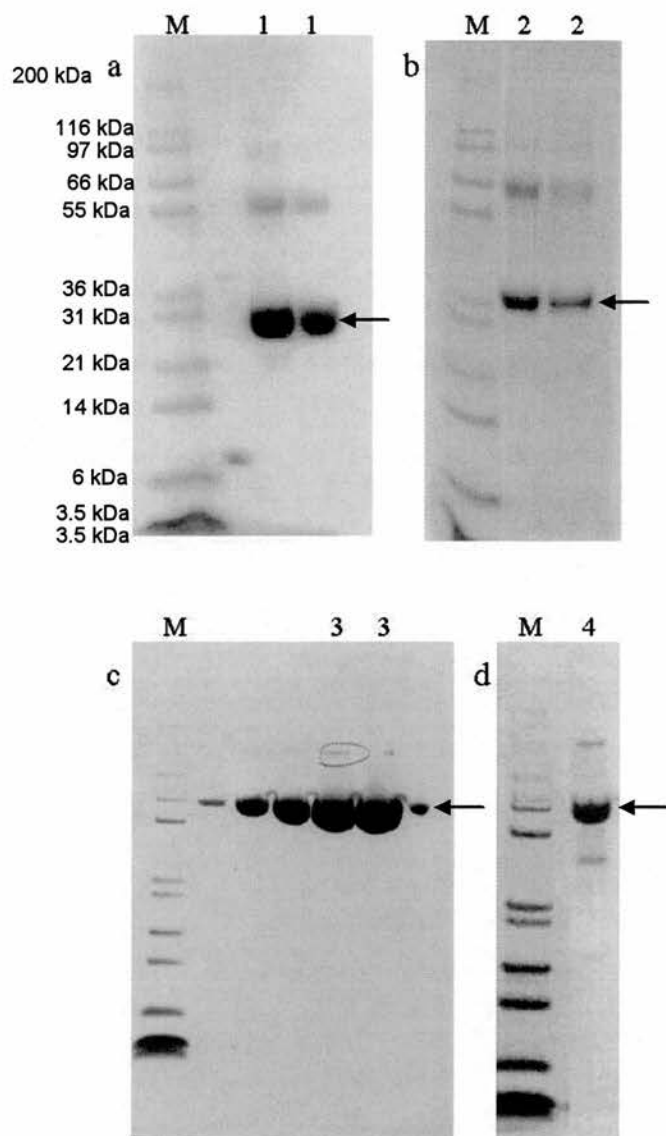
Figure 2.2.3: Coomassie Blue stained SDS PAGE gels. (a) ArnA formyltransferase, (b) ArnA decarboxylase, (c) His-tagged full length ArnA and (d) untagged full length ArnA. The protein standard markers are labelled in (a) and pure protein is highlighted by an arrow.

2.2.3 Crystallisation

Full length ArnA

Attempts were made to crystallise the three different constructs of full length ArnA. It was possible to concentrate the tagged and untagged ArnA protein to 10 mg ml^{-1} and 6 mg ml^{-1} respectively. After concentration ArnA protein was dialysed into 20 mM HEPES pH 7.5, 50 mM NaCl and 2 mM DTT. This protein, at a range of concentrations between 3 and 10 mg ml^{-1}, or protein that had been incubated at 4 °C overnight with 5-10 mM NAD$^+$, UDP-GlcUA and N-5-fTHF (Sigma), or some combination of these, was used for crystallisation screening. The sitting drop vapour diffusion technique was used with protein drops of between 2 and 6 μl and a 1:1 ratio of protein:precipitant, set up against 100 μl of reservoir solution. Commercially available screens from Hampton, Emerald Biosystems and Molecular Dimensions were screened at 20 and 4 °C.

After extensive screening of both tagged and untagged full length ArnA the most promising crystallisation conditions were at 20 °C and typically contained between 10 and 20 % PEG 6000 – 8000, buffer at pH 6 - 8 and 200 mM salt. These conditions resulted in the formation of spherulites (figure 2.2.5) which, upon addition of izit dye (Hampton) stained blue suggesting they were protein. Resolving several spherulites by SDS PAGE and staining with Coomassie blue confirmed them to be full length ArnA. Extensive attempts were made to optimise these initial hits by varying pH, precipitant and salt as well as their concentrations, stock and streak seeding and additive screens (Hampton). No crystals suitable for X-ray diffraction were obtained. This could be due to flexibility in the linker connecting the two

separate domains of ArnA. Two approaches attempted to try and lock the full length ArnA protein into a more ordered conformation, with the hope of aiding its crystallisation. Firstly a variety of ArnA substrates, cofactors and their analogues were used in co-crystallisation screens. It was hoped that binding of these ligands might reduce any flexibility in the active sites of the two domains and put the ArnA protein in a more homogeneous, ordered state that would be more likely to crystallise. Also, as discussed below, we thought that ArnA and ArnB might interact with one another. Therefore co-purification of these enzymes was attempted in the hope that binding of ArnB would order the ArnA protein, allowing the crystallisation of the complex (4.2.6). Neither attempts improved the crystallisation of ArnA and therefore structural studies were focused on the individual domains, expressed and purified separately.

Individual domains of ArnA

The pure decarboxylase or formyltransferase protein were concentrated to 5 mg ml^{-1} and 8 mg ml^{-1} respectively and dialysed into 25 mM HEPES pH 7.5, 50 mM NaCl and 2 mM DTT. 5 – 10 mM NAD$^+$ and either UDP-GlcUA, UDP-Glc, UDP galactose (Gal) or UDP (Sigma) were added to the decarboxylase protein and 5 – 10 mM N-5-fTHF and UDP (Sigma) were added to the formyltransferase protein and mixtures were incubated overnight at 4°C. Crystallisation screens were set up as for full length ArnA at 20 °C.

Screening of decarboxylase protein was much more successful than for full length ArnA. Initial crystals were obtained from the Hampton Index screen condition 9 (3 M NaCl and 0.1 M Bis-tris pH 5.5), using drops containing 1 μl protein and 1 μl

mother liquor. An optimisation screen around this condition gave good quality crystals suitable for X-ray diffraction obtained in a condition containing 3.2 M NaCl, 0.1 M Bis-tris pH 5.2, using a drop with 4 µl protein and 4 µl precipitant equilibrated against a reservoir of 100 µl mother liquor. The improvement of the size and quality of the crystals can be seen in figure 2.2.5, with many small crystals present before (ci) and a large, single, well defined crystal present after the optimisation screen (cii).

After extensive screening of ArnA formyltransferase protein, good quality crystals were obtained using a Hamilton robot in the Hampton Index screen condition 65 (17 % PEG 10000, 0.1 M Bis-tris pH 5.5 and 0.1 M ammonium acetate). Extensive screens around this condition failed to reproduce these crystals and therefore a drop containing 4.5 µl protein and 4.5 µl precipitant was set up against 100 µl of solution from the same well of the block that was used in the initial robot screen. This experiment gave rise to excellent quality crystals (figure 2.2.5).
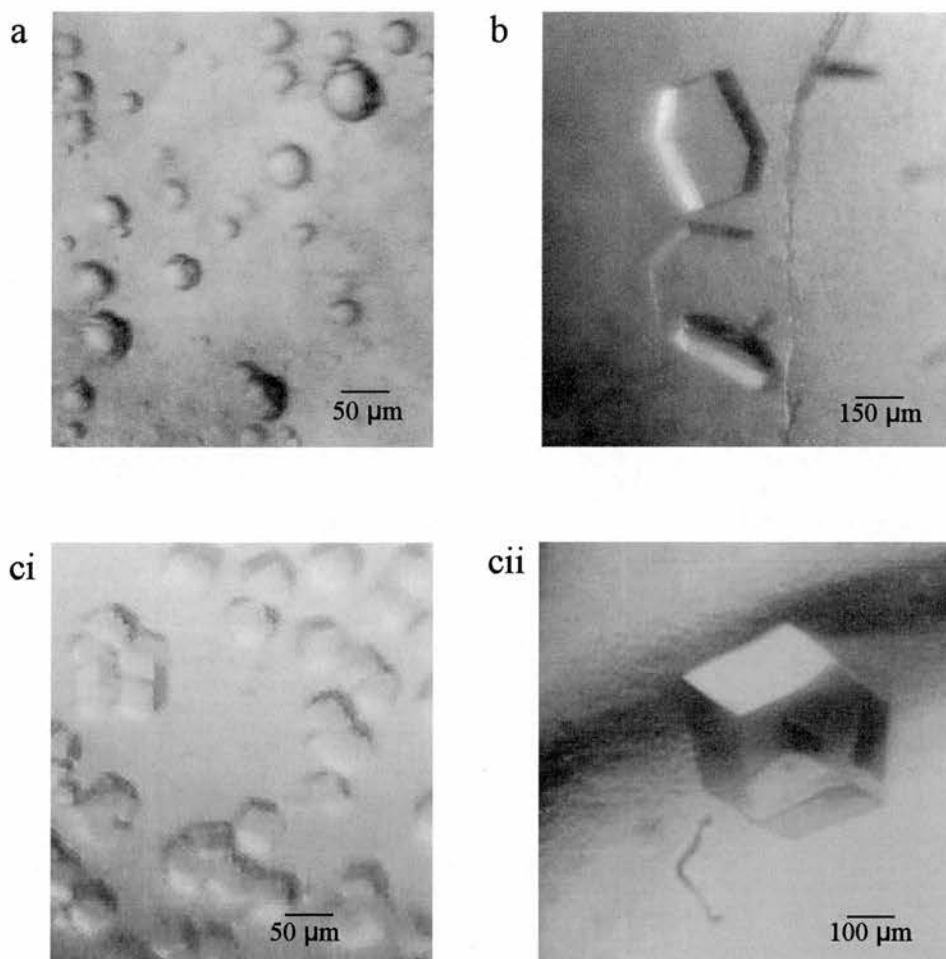
Figure 2.2.4: Figure showing protein spherulites formed by full length ArnA (a), crystals of formyltransferase protein (b) and crystals of dehydrogenase protein before (ci) and after (cii) optimisation.

2.2.4 Initial X-ray diffraction

An in house X-ray generator (Rigaku 007 rotating anode) was used to screen initial decarboxylase and formyltransferase crystals to test their diffraction quality. A small formyltransferase crystal was cryoprotected with 20 % glycerol added to the mother liquor and 20 minute exposures with 0.5 ° oscillations produced the diffraction image in figure 2.2.6a, with spots to 2.5 Å at the edge. Because it was a small crystal that

61

was tested, this result suggested that larger crystals would diffract to a high resolution at the European Synchrotron Radiation Facility (ESRF).

Cryoprotectants were screened against the conditions giving rise to decarboxylase crystals; 20 % glycerol, MPD and ethylene glycol mixed with mother liquor were unsuccessful but a short soak in 4 M Na formate gave good cryoprotection with no ice formation. A large decarboxylase crystal diffracted to 3.3 Å on our in house source using a 20 min exposure with 0.5 ° oscillation (figure 2.2.6b). The relatively poor diffraction of these crystals, despite their high quality appearance (figure 2.2.5c) suggested that a synchrotron source would be required to obtain data above 3 Å.
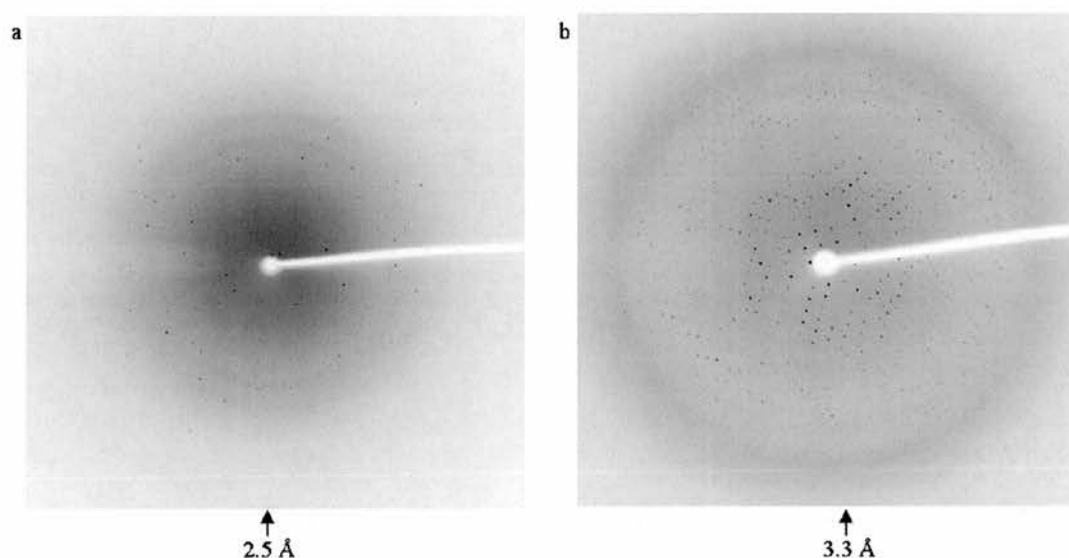
Figure 2.2.6: Initial in house x-ray diffraction obtained to 2.5 Å and 3.3 Å from formyltransferase (a) and decarboxylase (b) crystals respectively.

2.2.5   Data collection and processing

Once a cryocooled crystal diffracts satisfactorily a data collection strategy needs to be calculated to ensure that the dataset collected can be processed. The first step in this

process is to calculate the unit cell dimensions from the diffracted spots on the detector. The programs used to calculate this do so using autoindexing, which also calculates the orientation of the mounted crystal[394-396]. Whether the unit cell parameters are accurate can be tested by predicting spots based on the theoretical diffraction and examining whether these approximately fit the observed spots. Estimation of the mosaicity of the crystal allows an improvement of this prediction, which should subsequently predict almost all of the observed spots[394,397,398]. If the prediction looks good then a data collection strategy can be designed to give a range of oscillation angles that gives the most complete amount of unique data with non-overlapping lunes. Most data processing programs have functions that can calculate this, which test the percentage of overlaps observed for different oscillation angles and give the completeness of data over a range of angles[395,396,399]. Ideally, as large an oscillation range as possible should be selected that gives zero overlaps. The larger oscillation range not only increases the number of fully recorded reflections, but also reduces the exposure of the crystal to the X-ray beam. The data collection strategy will give the range of oscillation angles giving the most complete amount of unique data. Once the data collection strategy has been calculated a complete dataset can then be collected.

Once the data has been collected it is important to refine the cell parameters before integrating the data, to ensure that very accurate cell parameters are defined, in a process known as post-refinement[394,395,397]. This is important as the initial autoindexing only calculates the cell from a small number of images (often one or two) and as a result the cell parameters estimated are not as accurate as required. Post-refinement takes segments of data, usually from different areas of the crystal separated by around 90°, and integrates these images. Typically the differences

between the observed and calculated spot positions (or the non-intergerness of the indices) are minimised. The result of post-refinement is accurate cell constants to within 0.01 Å and an accurate estimate of the mosaicity[395,397]. In most cases of data processing these parameters can then be fixed during integration of the data[395,396,399].

The overall aim of data processing is to obtain estimates, which have been appropriately corrected, of the intensities of the Bragg reflections contained in a set of two-dimensional detector images[395]. The measurement of intensity is made by subtracting the background level of intensity from the spot intensity. Ideally this should be a measurement of the background behind the spot but this is impossible. Instead the pixels within the prediction box are used for spot-profiling, with the average intensity of outer pixels used as the background[395,397]. It is therefore important to pick a box size (a range of pixels) that adequately covers the spot, since otherwise the background calculation will include some of the diffraction intensity. It is also important not to have too large a box as this will result in a reduction of the signal-to-noise ratio and also increases the chance that pixels belonging to other reflections will be included within the box for the current reflection[394,395,397]. During integration of the data several other experimental parameters are also refined including the crystal to detector distance, Y-scale and the tilt and twist of the detector. This is important as these parameters can change slightly throughout the course of the experiment[395-397].

At the end of integration a data file is output with a measure of the intensity of each miller index, with an estimate of the error. Despite all of the corrections made during post-refinement and integration, further processing is necessary before the intensity measurements can be used for structure determination or refinement[400,401]. Systematic errors that have occurred during the experiment, including changes in the

incident beam intensity or size and absorption of the diffracted beam by the sample, need to be corrected so that all of the measured intensities are on the same scale[398,400,401]. The process of scaling compares the intensity estimates of symmetry related reflections and calculates scale factors to get the best internal consistency of these intensities[400,401].

The data file from integration is a measurement of the intensity of all reflections however, what is needed is a single estimate of the intensity of all symmetry-related observations of a given reflection. This data-reduction is carried out during merging and requires the correct space group to be entered into the program so that symmetry related observations can be properly grouped and the mean intensity and error can be calculated. If the space group is unknown the axial reflections, listed in the output from the scaling/merging program, gives an indication of systematic absences, which can be used to determine the correct space group (by comparison of expected systematic absences for each space group as listed in the international tables for crystallography[402]). Many crystallography applications, such as refinement, require not the intensity but the amplitude of each unique reflection[403]. The amplitudes, which are proportional to the square root of the intensities[398], are therefore calculated and output in the final data file from some merging programs.

The scaling and merging procedures are important in assessing data quality and the standard error in the final result and result in data that can be used for crystallagraphic applications such as phasing and refinement.

2.2.6 Data Collection and analysis

Formyltransferase domain of ArnA

Crystals were cryocooled to 100 K using a cryoprotectant of mother liquor containing 20 % glycerol. High resolution data to 1.2 Å was collected on native crystals at ESRF beamline ID14-2 240, 0.5 ° oscillation images were collected, using an exposure time of 3 s and a crystal to detector distance of 91 mm. An example of the diffraction pattern obtained is shown in figures 2.2.7 and 2.2.8. Images were indexed in MOSFLM[396], which suggested the crystal lattice belonged to the P222 Laue group with cell dimensions of a = 67.5 Å, b = 90.7 Å, c = 97.6 Å, $\alpha = \beta = \gamma = 90$ °. Data were processed and integrated using these parameters and then scaled in SCALA and merged in TRUNCATE (table 2.2.9)[401]. Observation of the systematic absences after merging in P222 revealed that the space group is $P2_12_12_1$. Data were reindexed in this space group and subsequent merging gave the data statistics presented in table 2.2.9
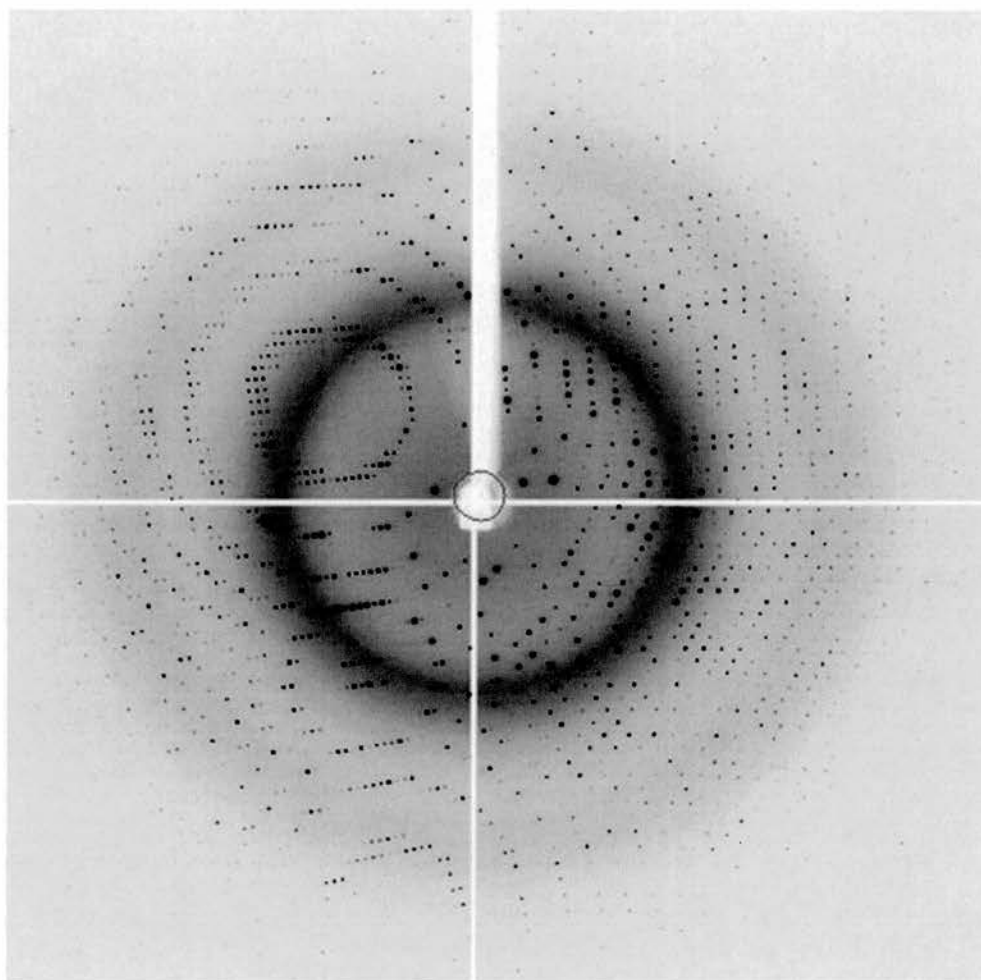
Figure 2.2.7: X-ray diffraction of formyltransferase crystals obtained at the ESRF beamline ID14-EH1, diffraction to 1.6 Å is shown for clarity.

1.2 Å
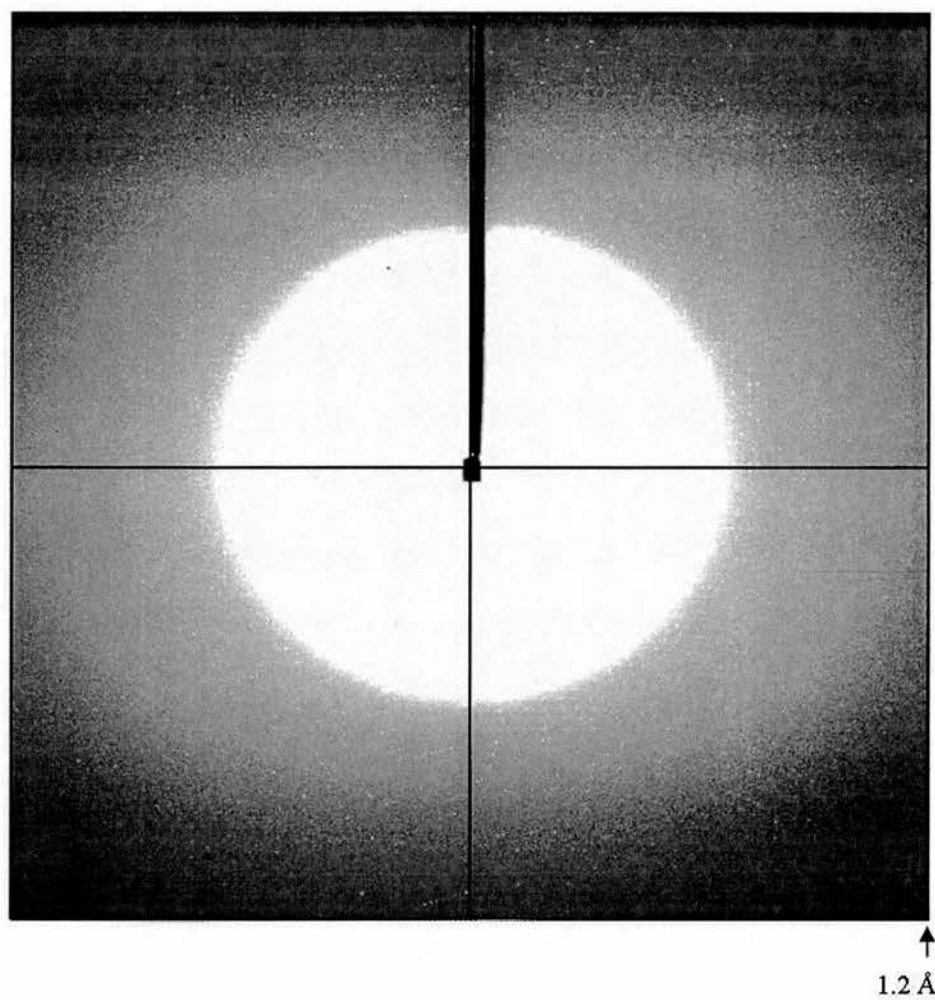
Figure 2.2.8: Formyltransferase diffraction with higher contrast than figure 2.2.7 and white spots on a black background, showing that the diffraction goes to 1.2 Å at the edge of the detector.

Data collection	ArnA formyltransferase domain
Resolution (Å)	28-1.2
Highest shell (Å)	1.26-1.2
Wavelength (Å)	0.934
Unit cell	a=67.5, b=90.7, c=97.6 $\alpha=\beta=\gamma=90°$
Space group	$P2_12_12_1$
Unique reflections	183672
Multiplicity	4.9 (4.9)
Completeness (%)	98.3 (98.3)
Rmerge* (%)	6.2 (29.7)
I/σ	7 (2.4)

Table 2.2.9: Native formyltransferase data statistics, numbers in () correspond to the highest resolution cell. $^{*}R_{merge} = \sum \sum I (h)i - (I (h)) / \sum \sum I (h)i$, where $I(h)i$ is the measured diffraction intensity and the summation includes all observations.

Decarboxylase Data Collection and Analysis

A large decarboxylase crystal was cryocooled to 100 K using a cryoprotectant of 4 M sodium formate. Data to 2.3 Å were collected on native crystals on beamline ID14-2 at the ESRF using an exposure time of 2 s and crystal to detector distance of 217 mm. An example of the diffraction pattern obtained is shown in figure 2.2.10. Data were treated similarly to the formyltransferase data, with indexing and integration in the P23 laue group in MOSFLM[396]. Reindexing of the data into several possible spacegroups with higher symmetry in SCALA[401] and subsequent merging of the data in TRUNCATE determined the space group to be either $P4_132$ or $P4_332$, with cell dimensions a = b = c = 149.4 Å, $\alpha = \beta = \gamma = 90°$, data statistics are shown in table 2.3.11.

2.3 Å
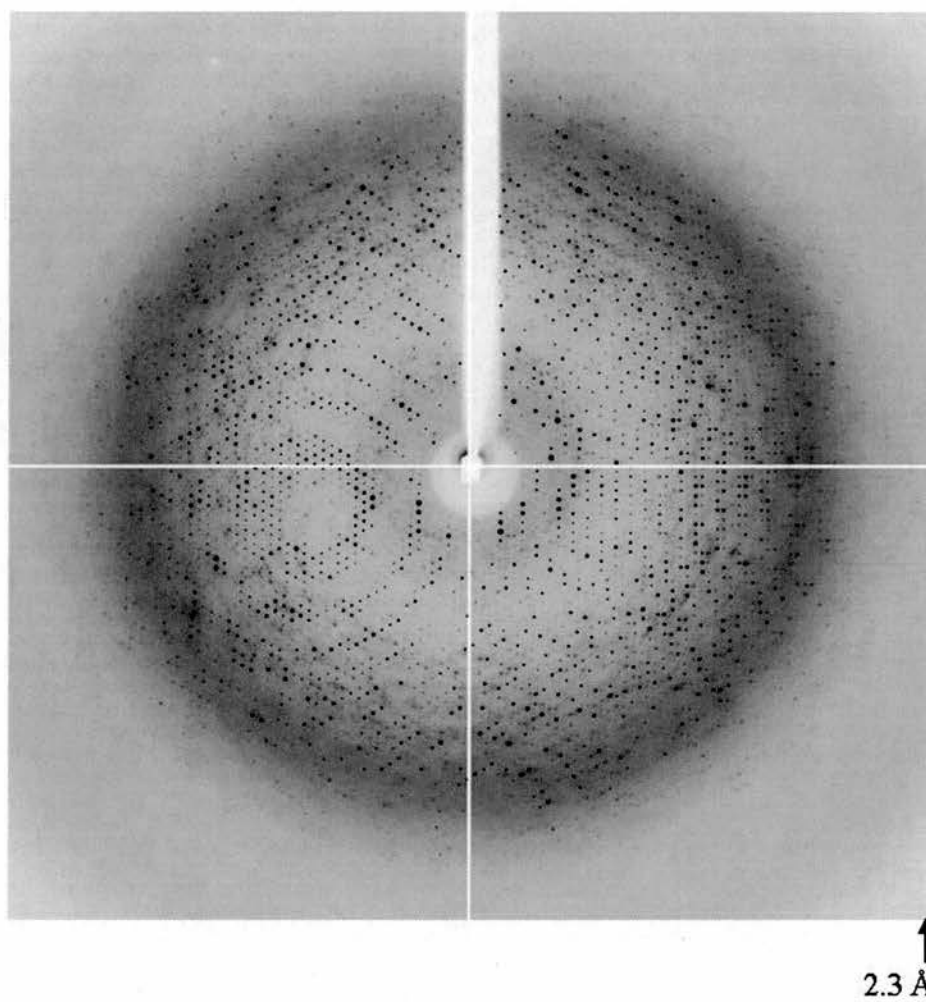
Figure 2.2.10: X-ray diffraction of decarboxylase crystals to 2.3 Å at the ESRF beamline ID14-EH2.

Data collection	ArnA decarboxylase domain
Resolution (Å)	88-2.30
Highest shell (Å)	2.42-2.30
Wavelength (Å)	0.933
Unit Cell	a=b=c=149.5 Å $\alpha=\beta=\gamma=90°$
Space group	$P4_132$
Unique reflections	25245
Multiplicity	16.3 (11.4)
Completeness (%)	97.8 (100)
Rmerge (%)	7.0 (36.9)
I/σ	7.8 (2.1)

Table 2.2.11 (above): Native decarboxylase data statistics, numbers in () correspond to the highest resolution cell.

2.2.7   Investigation of Protein – Protein Interactions

A stable complex between ArnA and ArnB

The activities of both domains of ArnA are required for the L-Ara4N modification of Lipid A *in vivo*. As shown in figure 2.1.7 the UDP-L-Ara4O intermediate, the product of ArnA decarboxylase activity, is converted to UDP-L-Ara4N by a separate enzyme ArnB, before going back to ArnA where the formyltransferase activity produces UDP-L-Ara4fN. Although fusion proteins are common in biology, it is unusual for a bifunctional enzyme to require an intervening enzymatic step. We therefore investigated if purified full length ArnA formed a complex with ArnB, to facilitate the process of substrate channelling between these enzymes. The purified ArnA and ArnB proteins, at approximately 3 mg ml^{-1} (0.075 mM), were mixed and incubated for 1-3 hours at room temperature in buffer containing 50 mM HEPES pH 7.5 and 250 mM NaCl. The mixture was applied to an S200 gel filtration column and

the elution of proteins was followed by A_{280}. This experiment was repeated in the presence of 5 mM UDP-GlcUA, 5 mM NAD^+, 5 mM glutamic acid, 5 mM pyridoxal 5'-phosphate and 5 mM N-5-fTHF (Sigma) (figure 2.2.12). ArnA and ArnB proteins that have been mixed elute at the same point as unmixed controls (figure 2.2.12), demonstrating no stable complex is formed in these conditions. We conservatively estimate that the lower level of detection from an S200 gel filtration column is about 0.06 mg ml^{-1} (1 µM), correlating to a lower limit of the K_d as 5 mM.
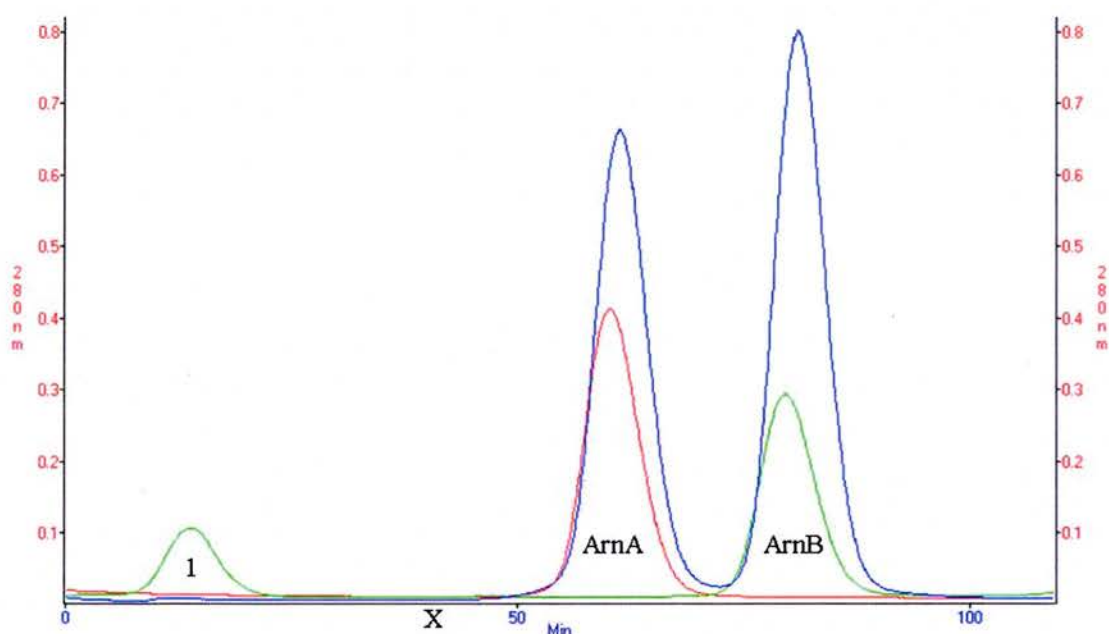
Figure 2.2.12: Traces from the gel filtration column investigating complex formation between full length ArnA and ArnB. The blue trace represents the elution after incubation of ArnA with ArnB, compared with individually eluted ArnA (red) and ArnB (green). The slight shift in peaks is caused by use of the S200 gel filtration column on different days and a slight difference in the volume of material loaded. X marks the void volume of the column and it is clear no material elutes around this area, as would be expected for any complex formation. Peak 1 is the elution of material left on the column from a previous run.

Our collaborators in North Carolina also performed Nickel pull down experiments using the Penta•His antibody (Qiagen) to investigate whether N-terminally His tagged ArnB pulled down over-expressed, untagged ArnA from cell lysate. They also observed no complex formation between ArnA and ArnB as judged by silver staining[381].

Individual domains of ArnA

To investigate whether the individually isolated ArnA domains interact with each other similar gel filtration experiments were performed as for full length ArnA and ArnB, both with and without substrates, cofactors and cofactor analogues. Similarly to ArnA and ArnB, no other peak apart from those for the individual decarboxylase and formyltransferase proteins was observed (data not shown).

2.3 Discussion

Genes encoded by the *PmrF* operon are essential for the L-Ara4N modification of Lipid A and, in all bacteria known to carry out this modification, ArnA is encoded by a single gene. However, although this modification has yet to be detected in *Ralstonia solanacearum* and *metallidurans*[238,376], the recent sequencing of their *pmrF* operons annotates the decarboxylase and formyltransferase domains as discrete genes[404]. This raises the possibility that full length ArnA may not be present in some bacteria, which is consistent with biochemical results. Breazeale *et al.* showed that the individual domains of ArnA, expressed individually, still catalyse their respective reactions and further, they can compensate for the action of the full length enzyme *in vivo*[376,377]. It was therefore attempted to structurally characterise ArnA by crysallising full length ArnA and sub-cloned decarboxylase and formyltransferase enzymes, with no crystallographic artefacts expected by pursuing the separated domains. Good quality data were collected on both formyltransferase and decarboxylase proteins and used for structural solution, detailed in Chapter 3. However no diffraction quality crystals were obtained for any of the full length ArnA enzyme.

Although biochemical data suggest that there is no reason for the individual domains to be present in a bifunctional enzyme[238,376], that full length ArnA is well conserved suggests there is an evolutionary advantage for this. ArnA catalyses two, non-consecutive reactions that are essential for the cytosolic biosynthesis and subsequent transportation of L-Ara4N to the inner membrane[376,377]. The decarboxylation and formyl transfer reactions are separated by an intermediary amino transfer reaction, catalysed by ArnB[380,384]. This is unusual; enzymes that have two distinct activities within a biosynthetic pathway usually catalyse consecutive

reactions[59], with substrate channelling providing the evolutionary advantage for this[405]. Although unusual, bifunctional enzymes catalyzing non-consecutive reactions are not without precedent. In *E. coli* a bifunctional enzyme AK-HDH I, with aspartokinase and homoserine dehydrogenase activities, catalyses the first and third reactions in the biosynthesis of amino acids belonging to the aspartate family[406,407]. The intervening reaction is catalysed by aspartate semialdehyde dehydrogenase (ASADH)[408]. Kinetic analysis of the AK-HDH I and ASADH reactions with increasing amounts of inactive ASADH and also by the creation of inactive bifunctional enzymes suggest that, although interactions between AK-HDH I and ASADH have not been detected, these enzymes interact to facilitate substrate channelling[406,407]. Another example of a bifunctional enzyme catalysing non-consecutive reactions is IspDF, which is a fusion of IspD[409] and IspF[410] proteins that exists in some organisms[409,411] and is involved in the non-mevalonate pathway of isoprenoid biosynthesis[412]. Within this pathway the IspE kinase catalyses the intervening reaction [413] and complex formation between IspE and both IspDF and the mono-enzymes has been detected *in vitro* [411]. This complex formation has been proposed to be the basis of substrate channelling within this pathway that could bring together up to 18 catalytic centers[411].

We therefore thought that there was a possibility that the ArnA and ArnB enzymes may be part of a larger complex to facilitate substrate channelling. In such a complex the L-Ara4O product of the decarboxylase reaction would be passed to the ArnB enzyme, where an amino group would be transferred. L-Ara4N would then be channelled back to ArnA, its formyltransferase domain, where the formyl group would be added. If such a complex did exist it may help explain the apparent evolutionary pressure for full length ArnA in organisms known to modify their Lipid

A with L-Ara4N. Breazeale *et al.* developed an enzymatic assay for ArnB and show that the forward reaction to produce UDP-Lara4N is thermodynamically unfavourable with an equilibrium constant of approximately 0.1[380]. Thus, if a complex does form to facilitate substrate channelling, it would help to compensate for the unfavourable ArnB reaction and act to drive all of the UDP-L-Ara4O intermediate into product. There are other potential advantages substrate channelling could have in this pathway. Breazeale *et al.* suggested investigating the use of UDP-LAra4O as an inhibitor or alternative substrate for sugar-nucleotide utilising enzymes to ArnA[377]. If either UDP-Lara4O or UDP-Lara4N intermediates could bind other enzymes *in vivo*, substrate channelling in the ArnA enzyme would prevent this. Initial, but by no means exhaustive, experiments were carried out to investigate this intriguing possibility of a protein:protein complex forming between ArnA and ArnB. Gel filtration experiments were performed to see whether a stable complex formed between these enzymes, with the hope that if such a complex formed structural studies could be undertaken. Our results suggest that ArnA and ArnB do not form a stable complex and that any interaction that may occur between them must be transient with the equilibrium, *in vitro*, strongly favouring the individual proteins. Due to advances in other aspects of my research I did not look beyond gel filtration experiments to investigate complex formation. However there are several experiments that could be performed in the future to carry out a more extensive investigation into whether these enzymes interact *in vivo* to facilitate substrate channelling.

2.4 Future work

A concept overlooked in the gel filtration experiments performed above is the *in vivo* phenomenon of crowding, discussed in chapter 1.2. Within the cell the high concentrations of proteins and other macromolecules can significantly alter the kinetics of reactions[73,75,414] and protein associations[74], due to the limited volume available. In particular the formation of specific protein-protein interactions is enhanced due to the non-specific effect of macromolecular crowding[74,414,415], as a protein complex excludes less volume than the two individual components[414]. Therefore conditions utilised *in vitro*, with relatively dilute buffer and protein concentrations, are often inadequate to mimic *in vivo* conditions. In the future if the possible interaction of ArnA with ArnB was to be further investigated it would be well worth carrying out experiments in the presence of *in vitro* crowding agents, reviewed in Minton, 2001[75] and Hall and Minton, 2003[81]. One example of such an experiment could be a frontal analysis gel filtration experiment similar to that employed by Morgunov and Srere, 1998[80]. They were able to observe complex formation between citrate synthase and malate dehydrogenase when they applied both proteins to a gel filtration column that had been pre-equilibrated with 0.2 mg ml^{-1} of citrate synthase and 10 % polyethylene glycol (PEG)[80]. Since the column is pre-equilibrated with one of the proteins there will always be the potential for an interaction to occur, whereas in the ArnA-ArnB gel filtration experiments the different sizes of ArnA and ArnB may result in their separation early in the column and any transient interaction may be missed.

Several other experiments could also be performed to investigate whether substrate channelling occurs in this pathway[416], although since substrate channelling

is dependent on at least a transient interaction between enzymes it is still important to demonstrate such an interaction. One recent method for investigating substrate channelling between separate enzymes is assaying the reaction in the presence of a large excess of an inactive version of the second (or later) enzyme, which is still able to bind substrate[406,417]. For ArnA such an experiment would require an inactive ArnB mutant that can still bind substrate and have the same (presumed) ability to bind ArnA as wild type ArnB. The excess inactive ArnB would out-compete wild-type ArnB for the hypothetical binding sites to ArnA and would only decrease the rate of reaction if substrate channelling was occurring between the two enzymes.

The further investigation of substrate channelling within this pathway is intriguing and would be an interesting direction to take in the future.

Chapter 3

Structural solution of the individual formyltransferase and decarboxylase domains of ArnA:

Insights into the mechanism of decarboxylation by site-directed mutagenesis and kinetic assay

3.1   INTRODUCTION

Summary

Attempts to crystallise the full length ArnA protein were unsuccessful, whereas crystals suitable for data collection were obtained for the individual decarboxylase and formyltransferase domains. This chapter describes the structural solution of the decarboxylase domain to 2.3 Å and the formyltransferase domain, in complex with UMP and a cofactor analogue N-5-fTHF, to 1.2 Å. Based on the superimposition of the native decarboxylase structure with RmlB, E_{434} and S_{434} active site mutants were designed to probe the oxidative decarboxylation mechanism. Our results suggest that oxidative decarboxylation is enzyme catalysed and that E_{434} appears to be essential for the reaction, with the second decarboxylation step driving oxidation.

3.1.1   Formyltransferases

Within the cell there are only a handful of reactions that utilise N-10-formyltetrahydrofolate (N-10-fTHF) for formyl transfer. These include glycinamide ribonucleotide (GART)[418] and 5-aminoimidazole-4-carboxamide (AICAR)[419] formyltransferases, both involved in purine biosynthesis and 10-formyltetrahydrofolate:L-methionyl-tRNA N-formyltransferase (FMT)[420], involved in formylating the amino acid in methionyl-tRNA (figure 3.1.2). N-10-fTHF is also hydrolysed to produce THF by the nicotinamide adening phosphate ($NADP^+$) dependent, bifunctional enzyme formyltetrahydrofolate dehydrogenase (FDH), with N-terminal hydrolase and C-terminal dehydrogenase domains[421,422]. The N-terminal

domain of FDH is homologous to FMT and is capable of hydrolysing N-10-fTHF *in vitro* [422]. The C-terminal domain is homologous to aldolase dehydrogenase and it is only the activity of both domains, present as a bifunctional enzyme, that is thought to be functional *in vivo*[422].

The N-terminal sub-domain of ArnA transfers a formyl group from N-10-fTHF to UDP-L-Ara4N, a reaction that is essential for the L-Ara4N modification of Lipid A (Chapter 2.1.5). ArnA shows 29 % identity to FMT, 26% identity to FDH, 25 % identity to GART (figure 3.4.1), but no significant identity to AICAR. Figure 3.1.3 shows the structures of GART[418], AICAR[419] and another formyltransferase PurT[423], that formylates β-GAR to produce formyl-GAR in an N-10-fTHF independent manner. The structures of these enzymes reveal that they share no overall structural similarity and represent three different folds that nature has evolved to catalyse these very similar reactions. Each enzyme also employs a different mechanism of formyltransfer[423-425], demonstrating the importance of this chemistry in biology. It is because of this importance that there have been extensive studies into these interesting enzymes, in particular GART. As a result these enzymes have been relatively well characterised both structurally and mechanistically, providing insights into how the formyltransferase domain of ArnA may function.
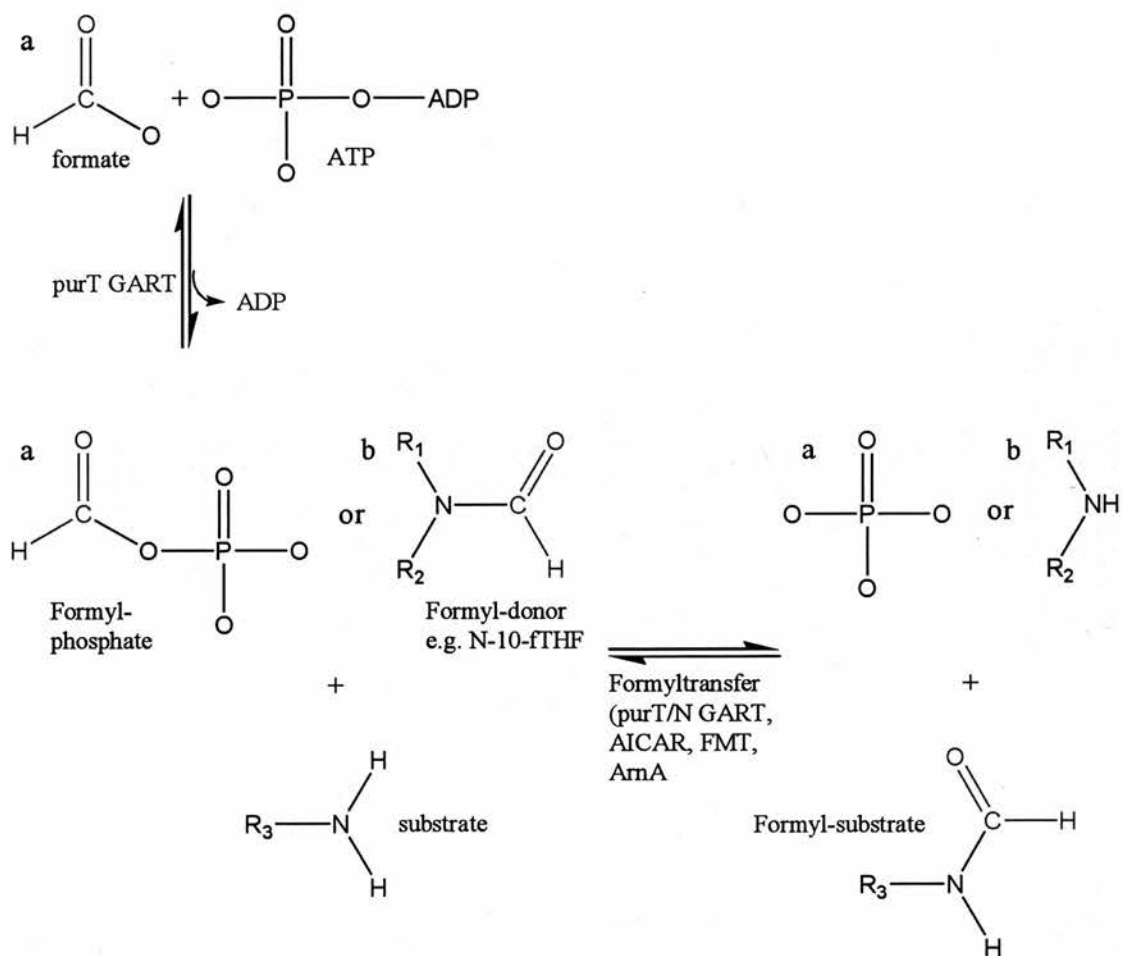
Figure 3.1.1: Chemical diagram of the one carbon transfer catalysed by (a) purT

GART and (b) purN GART, AICAR, FMT and ArnA.

3.1.2 Folate metabolism

(6S)-5,6,7,8-tetrahydrofolate

Level of C_1 oxidation	C_1–complex formation	C_1–complex interconversion	C_1–complex utilization

methanol

homocysteine → methionine

methionine synthase → H_4folate

CH_3

5-CH_3-H_4folate

methylenetetrahydrofolate reductase

$+ NADP^+$
$+ NADPH$
$+ H^+$

formaldehyde

H_4folate

serine → glycine

glycine hydroxy-methyltransferase

CH_2

5,10-CH_2-H_4folate

dUMP → dTMP

thymidylate synthase → H_2folate

methylenetetrahydrofolate dehydrogenase

$+ NADP^+$
$+ NADPH$
$+ H^+$

formate

CHO

5-HCO-H_4folate

ATP → ADP $+ P_i$

5-formyl-tetrahydrofolate cyclo-ligase

CH

5,10-CH^+-H_4folate

glutamate Formimidoyl-transferase

L-glutamate

NH_3

methenyltetrahydrofolate cyclohydrolase

$\pm H_2O$

See figure 3.1.1b

N-formyl-L-glutamate

formimidoyl-tetrahydrofolate cyclodeaminase

CHO

Met-tRNA^{fMet}
fMet-tRNA^{fMet}

H_4folate

H_4folate

HC=NH

FMT

Histidine catabolism

5-NHCH-H_4folate

10-HCO-H_4folate

$NADP^+ + H_2O$

$CO_2 + NADPH + H^+$

FDH

formyltetrahydrofolate synthetase

$ADP + P_i$

H_4folate   formate + ATP

H_4folate

b

ATP +glycine   ADP + P$_i$

phosphoribosylamine-
glycine ligase

5-phospho-D-ribosylamine

ribose
activation

10-HCO-H$_4$folate

H$_4$folate

GART

N^1-(5-phospho-D-ribosyl)glycinamide

ATP +Gln

ADP + Glu + P$_i$

N^2-formyl-N^1-(5-phospho-D-ribosyl)glycinamide

FGAR
amidotransferase

ATP

ADP + P$_i$

2-formamido-N^1-(5-phospho-D-ribosyl)acetamidine

Inosine 5'-phosphate

(AICAR)

5-amino-1-(5-phospho-D-ribosyl)imidazole

Figure 3.1.2: (a) Folate coenzymes with important functions within the cell. Only the chemical formula of folate within the dotted box (top) is shown in the diagram. (b) Early stages of purine biosynthesis. Enzymes are coloured in blue, with those discussed in the text (GART, FMT and AICAR) coloured red. Adapted from the International Union of Biochemistry and Molecular Biology website: http://www.chem.qmul.ac.uk/iubmb.

GART (purN), FDH, FMT and AICAR enzymes are all involved in folate metabolism (figure 1.1.1). This is essential to cell viability as folate derivatives are required for

84

the biosynthesis of the DNA and RNA nucleotide precursors, purines and pyrimidines[426], as well as some amino acids[420]. In many cancer cells enzymes involved in folate metabolism are often found up- or, in the case of FDH, down-regulated[421]. In turn this has created a lot of interest in these enzymes and their potential as good drug targets[427]. Dihydrofolate reductase (DHFR) is involved in maintaining the folate cofactor in its biologically relevant, fully reduced state[427,428] and the first inhibitors of folate metabolism, aminopterin and methotrexate[429], targeted this enzyme. These compounds, along with related drugs that inhibit species-specific DHFR, have widespread clinical uses as anti-cancer, ant-bacterial and anti-protozoal agents[428]. As a result of the success of anti-DHFR drugs other enzymes involved in folate metabolism have also been targeted including; GART[430,431], FDH[422] and AICAR[432,433].

GART is a particularly well studied formyltransferase, transferring a formyl group to its substrate β-GAR in the first step of puridine synthesis[434]. In 1992 a folate inhibitor against human GART was in clinical trials when the homologous *E. coli* GART enzyme was solved[418] (figure 3.1.3). Since then the human GART subdomain has also been solved[435]. These structures have been important in understanding the mechainism of formyl transfer from N-10-fTHF in this class of enzymes and have also led to the structure based design of GART specific inhibitors that are potential anti-cancer drugs[436].

Figure 3.1.3 (below): Three different structural folds that have evolved to transfer a formyl group.

(a) The bifunctional AICAR-IMPCH enzyme (PDB code: 1G8M)[419]. The individual AICAR and inosine monophosphate cyclohydrolase (IMPCH) enzymes are active when separated, however although IMPCH is functional as a monomer (coloured green and red), AICAR is only functional as a dimer (coloured orange). It is this dimer that is required for formyl transfer from N-10-fTHF in this enzyme[419]. The three shades of orange represent three distinct sub-domains in AICAR, with substrate and N-10-fTHF binding at the dimer interface.

(b) purT GART (PDB code: 1EZ1) uses ATP and formate for formyl transfer. ATP binds between domains A and B, coloured in lime and yellow respectively, with substrate (β-GAR) binding at the C-terminal end of the β-sheet in domain A, coloured blue. The ATP and formate co-complex is shown[423].

(c) purN GART is a monomer with one functional domain (coloured violet). It is shown with an inhibitor[418] in the folate binding site (PDB code: 1CDE). Substrate binds below the N-10-fTHF site for formyl transfer.

Labels in figure:
a — AICAR, IMPCH
b — PurT GART, A, B, C
c — PurN GART
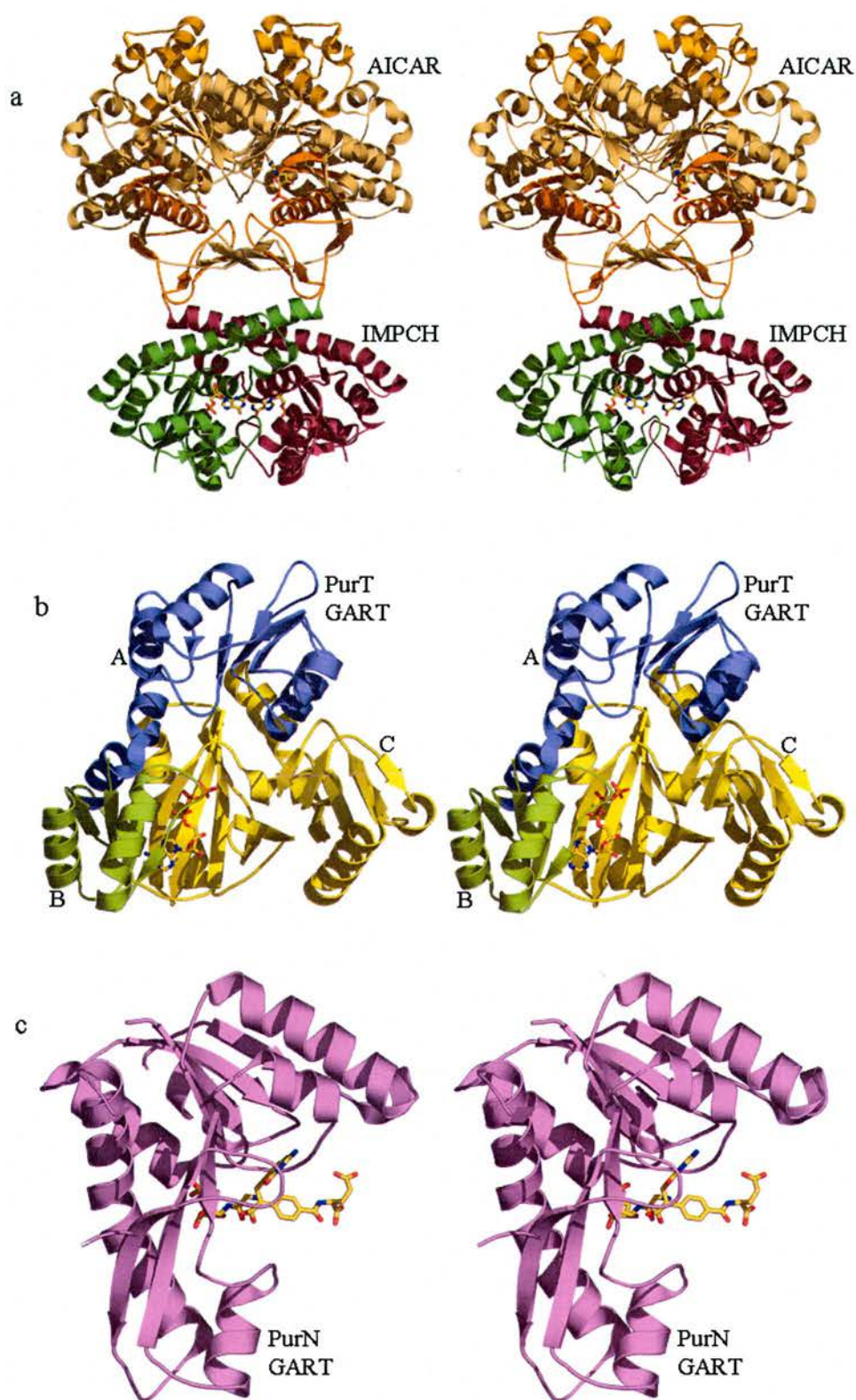
Figure 3.1.3: Figure legend above.

3.1.3   The GART family of formyltransferases

Of the N-10-fTHF utilising formyltransferases there are two sub-families that do not share sequence or structural similarity. One sub-family includes AICART, which transfers a formyl group to 5-amino-imidazole 4-caboxamide in the penultimate step of purine biosynthesis[431]. The other sub-family represents the better studied class of N-10-fTHF utilising formyltransferases that includes GART, FMT and FDH.

The GART sub-family shares relatively low sequence identity, around 15-30 %, however structurally they are quite similar (figure 3.3.2)[422]. All of these enzymes share a folate binding domain. In GART this constitutes the whole enzyme[418] (figures 3.1.3 and 3.3.2). In FMT the N-terminal sub-domain binds folate, with the C-terminal sub-domain involved in substrate binding[420]. Similarly to FMT, the N-terminal sub-domain of the FDH hydrolase domain binds folate; however the function of its C-terminal sub-domain, although important for activity, is unclear (unpublished results[422]).

The folate binding domain has a 7-stranded β-sheet flanked by α-helices with 6 of the 7 strands making up the characteristic Rossmann fold. There is also an extra α-helix outside of this core, which is only ordered in GART upon N-10-fTHF binding, but makes interactions with the C-terminal sub-domain in FMT and FDH [422]. This domain is structurally very similar between these enzymes and in particular the core 7-stranded β-sheet is largely superimposible. Chumanevich et al. found that superimposition of just the 7-stranded β-sheet of FDH (~45 Cα's) with the corresponding region in GART and FMT gave a rool mean square deviation (RMSD) of just 0.8 and 1 Å respectively[422]. In contrast the RMSD of all common Cα's (~155) of the folate binding domain of FDH superimposed with GART and FMT, is 3.5 and

3.7 Å respectively[422]. This demonstrates that, in common with many enzymes, the chemical 'core' of the enzyme is structurally well conserved.

Bicyclic ring   Benzoyl   Glutamate
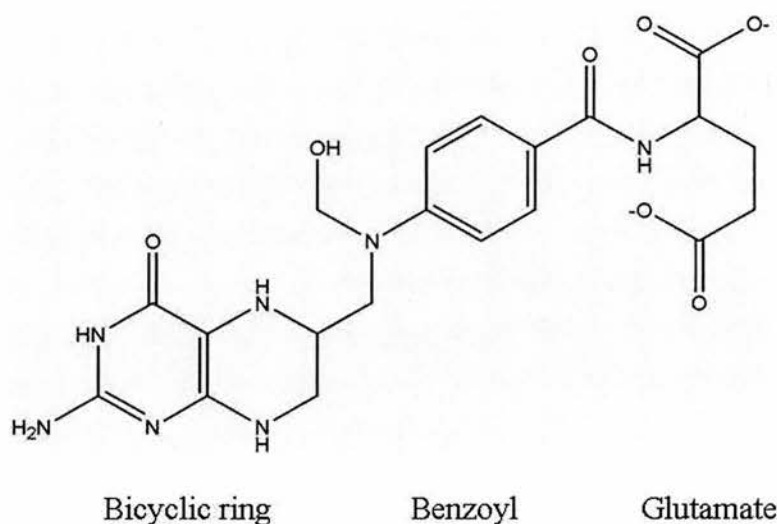
Figure 3.1.4: The N-10-fTHF cofactor contains three regions; a bicyclic ring, a benzoyl ring and glutamate.

There are two major structural differences between the FMT and GART structures. In FMT there is an extra loop protruding from the active site, composed of residues 34-49 and a C-terminal sub-domain, residues 209-314, attached to the N-terminal sub-domain by a linker, residues 290-208[437]. In the apo structure of FMT the protruding loop is disordered[420] but becomes ordered upon tRNA substrate binding[437]. In the tRNA-FMT complex structure, several interactions occur between residues on the ordered loop with the acceptor-stem of the tRNA (figure 3.4.3), correctly orientating the adenosine base, with esterified formyl-methionyl group attached, into the active site[437]. Interestingly the C-terminal sub-domain is also involved in substrate binding; this domain is made up of 2 large and two smaller β-sheets, which fold together to form a small β-barrel, similar to the OB-fold[437]. There are extensive interactions between the tRNA substrate and the C-terminal subdomain. In particular

89

the protein contacts the minor grooves of the acceptor-arm and the D-arm as well as to the D-stem of the tRNA (figure 3.4.3). Important to the D-arm interaction is an electropositive channel that leads towards the N-terminus of FMT, including the residues K_{209}, K_{246}, K_{291}, R_{303} and R_{304}[437]. The interactions between the tRNA and protein clamp the tRNA molecule in place.

3.1.4 Mechanism of formyltransfer

Early biochemical studies of the GART mechanism showed that it did not form either cofactor amidine (5-10-methyenyltetrahydrofolate) or enzyme-amidinium intermediate species[438]. Instead the mechanism of formyl transfer was suggested to involve direct nucleophilic attack of the β-GAR amine on the formyl carbon of N-10-fTHF[438]. Subsequent studies are consistent with this mechanism of formyl transfer and structural solution of GART and FMT give insights into how this reaction is catalysed by the enzyme[425,434]. The structures of GART[418], FMT[437] and FDH[422] in complex with their substrates, reveal a conserved location of the active site, with two strictly conserved residues, H_{108} and D_{144} (numbered as GART), ideally located to assist in formyl transfer from N-10-f-THF. A further residue, N_{106}, which is conserved in GART and FMT but absent in FDH, is also positioned to assist in the reaction[425,439,440]. That N_{106} is not conserved in FDH probably reflects the differences in the reaction of this enzyme with other formyltransferases of this family, as the formyl group is not transferred to a substrate.

Extensive mutational analysis has been carried out on these catalytic residues in GART[425,439,440]. These studies show that although all three residues are important for full activity of the enzyme, none are absolutely required[439,440], consistent with the

idea that these residues are assisting the direct attack of the β-GAR amine on N-10-fTHF. D_{144} is the most important residue in catalysis[440] and is located near to the N-10 position of the folate, however it is not close enough and sterically blocked from the α-amine of GAR and thus cannot extract a proton[425,439]. Instead it is thought H_{108} donates a proton to the formyl oxygen, initiating the tetrahedral intermediate formation resulting from attack of the amino group of GAR which, due to the low-dielectric state of the active site, is in its free base form[425] (figure 3.4.2). H_{108} would need to be protonated, a situation that is facilitated by the salt bridge between D_{144} and H_{108}, raising the pKa of H_{108} to approximately 9.8[425]. A proton is then transferred to the N-10 of the folate from the β-GAR substrate through a catalytic water molecule, the position of which is mediated by D_{144}[425]. Tetrahedral intermediate breakdown and product release follows this proton transfer. N_{106} is located close to the formyl group and it is thought that this residue plays a role in stabilising the negatively charged oxygen of the intermediate in GART and FMT mechanisms (figure 3.4.2)[425].

3.1.5   SDR enzymes

```
ArnADH:---------- MRVLILGVNG FIGNHLTERL LREDHYEVYG LDIG------ SDAISRFLNH
RmlB   :---------- MKILVTGGAG FIGSAVVRHI INNTQDSVIN VDKLT-YAGN LESLTEIENN
GalE   :---------- MRVLVTGGSG YIGSHTCVQL LQNGHDVIIL DNLCN-SKRS VLPVIERLGG
UDP-XS:(X83)SEKDR KRILITGGAG FVGSHLTDKL MMDGHEVTVV DNFFT-GRK- -RNVEHWIGH
MLCR   :---MKLNFSG LRALVTGAGK GIGRDTVKAL HASGAKVVAV TRTNS----D LVSLAKECP-
DADH   :---MSFTLTN KNVIFVAGLG GIGLDTSKEL LKRDLKNLVI LDRIEN-PAA IAELKAINPK

ArnADH:PHFHFVEGDI SIHSEWIEY- HVKKCDVVLP LVAIATPIEY TRN-PLRVFE LDFEENLRII
RmlB   :ERYKFEHADI CDSVAIANIF AHHQPDAIMH LAAESHVDRS ITG-PADFIE TNIVGTYILL
GalE   :KHPTFVEGDI RNEALMTEIL HDHAIDTVIH FAGLKAVGES VQK-PLEYYD NNVNGTLRLI
UDP-XS:ENFELINHDV VEPLYI---- ---EVDQIYH LASPASPPNY MYN-PIKTLK TNTIGTLNML
MLCR   :-GIEPVCVDL GDWDATEKAL GGIGPVDLLV NNAALVIMQP FLEVTKEAFD RSFSVNLRSV
DADH   :VTVTFYPYDV TVPIAETTKL LKTIFAQLKT VDVLINGAGI LDD---HQIE RTIAVNYTGL

ArnADH:RYCV------ ----KYRKRI IFPSTSEVYG MCSDKYFDED HSN------L IVGPVNKPRW
RmlB   :EEARKYWLAL SEDRKGAFRF HHISTDEVYG D---LPHPDE VSSDTILPLF TEQTSYSPSS
GalE   :SAMR------ ---AANVKNF IFSSSATVYG DQPKIPYVES FPTGTPQSPY GKSKLMVEQI
UDP-XS:GLAK------ ----RVGARL LLASTSEVYG DPEVHPQSED Y--------W GHVNPIGPRA
MLCR   :FQVS------ ---QMVARDM INRGVPGSIV NVSSMVAHVT FPN------- --------LI
DADH   :VNTT------ ----TAILDF WDKRKGGPGG IICNIGSVTG FN-------- ----AIYQVP

ArnADH:IYSVSKQLLD RVIWAYGEKE GLQFTLFRPF NWMGPRLDNL NAARIGSSRA ITQLILNLVE
RmlB   :PYSASKASSD HLVRAWRRTY GLPTIVTNCS NNYGPYHFPE K--------L IPLIILNAIA
GalE   :LYDLQKAQPD WSIALLR-YF NPVGAHPSGD MGEDPQGIPN N--------L MPYIAQVAVG
UDP-XS:CYDEGKRVAE TMCYAYMKQE GVEVRVARIF NTFGPRMHMN DGR------V VSNFILQALQ
MLCR   :TYSSTKGAMT MLTKAMAMEL GPHKIRVNSV NPT------- -------VV LTDMGKKVSA
DADH   :VYSGTKAAVV NFTSSLAKLA PITGVTAYTV NPGITRTT-- --------L VHKFNSWLDV

ArnADH: G-SPIKLID- -----GGKQK RCFTDIRDGI EALYRIIENA GNRCDGEIIN IGNPENEASI
RmlB   : G-KLLPVYG- -----NGEQI RDWLYVEDHA R------PGH DLRYAIDASK IKQDLG---W
GalE   : RRDSLAIFGN DYPTEDGTGV RDYIHVMDLA DGHVVAMEKL ANKPGVHIYN LGAGVGNSVL
UDP-XS: G-EPLTVYG- -----SGSQT RAFQYVSDLV NG----LVAL MNSNVSSPVN LGNPEEHTIL
MLCR   : DPEFARKLK- -----ERHPL RKFAEVEDVV NSILFLLSDR SASTSGGGIL VDAGYLAS--
DADH   : EPQVAEKLL- -----AHPTQ PSLACAENFV KAIELNQNGA IWKLDLGTLE AIQWTKHWDS

ArnADH:EELGEMLLAS FEKHPLRHHF PPFAGFRVVE SSSYYGKGYQ DVEHIKPSIR NAHRCLDWEP
RmlB   :VPQETFETGI TKTI------ ---------- HWYLNNK--- EWWQRVMDGS YAGERLGLVE
GalE   :DVVNAFSKAC GKPV------ ---------- NYHFAPRREG DLPAYWADAS KADRELNWRV
UDP-XS:EFAQLIKNLV GSGS------ ---------- EIQFLSEAQD DPQKIKPDIK KAKLMLGWEP
  MLCR   :(end)
  DADH   :GI(end)

  ArnADH:KIDMQETIDE TLDFFLRTVD LTDKPS(end)
  RmlB   :(end)
  GalE   :TRTLDEMAQD TWHWQSRHPQ GYPD(end)
  UDP-XS:VVPLEEGLNK AIHYFRKELE YQANNQ(X18)
```

Figure 3.1.5: Sequence alignment of ArnA decarboxylase (ArnADH) with the similar *E. coli* enzymes RmlB, GalE and human UDP glucuronic decarboxylase (UDP-XS). The classical SDR's mouse lung carbonyl reductase (MLCR) and DADH are shown for comparison.

Figure 3.1.5 continued: The well conserved GxxxGxG and YxxxK motifs are highlighted in red. Residues conserved in the ArnADH, RmlB, GalE and UDP-XS sequences are highlighted in light blue and those similar are coloured green. The residues thought to be important in ArnA decarboxylase are highlighted dark blue, with corresponding residues in RmlB and GalE highlighted in grey. The residues important in NADP$^+$ binding (MLCR) and NAD$^+$ in the other enzymes are coloured yellow.

Sequence analysis of the C-terminal decarboxylase domain of ArnA predicted that this domain belongs to the SDR superfamily of enzymes (figure 3.1.5). This is one of the largest superfamilies in nature and contains enzymes that catalyse a diverse range of reactions, using NAD(P)$^+$ as a cofactor. Despite this diversity, which is related in sequence, the structures of SDR enzymes are homologous, reflecting the conserved chemistry of hydride transfer catalysed by SDR enzymes.

NAD(P)$^+$ dependent alcohol dehydrogenase (ADH) activity is very widespread in nature and important in all domains of life. The independent evolution of three main, mechanistically and structurally different, enzyme superfamilies reflects the importance of this reaction[441,442]. The three superfamilies are the iron dependent ADH's representing the smallest and least studied group[443,444], the medium chain dehydrogenases/reductases (MDR), requiring a metal ion (usually zinc) as a cofactor[445] and the short chain dehydrogenases/reductases (SDR), representing the largest group[445]. The birth of the SDR superfamily as a class of enzymes was in 1981 and began with the report of an ADH from *D. melanogaster* (DADH)[445]. Although DADH carries out the same reaction as horse liver ADH (LADH), it shares very little sequence identity and does not require a zinc ion[445]. DADH is significantly shorter in

length than horse LADH but has a conserved Rossmann fold. This fold is present in the N-terminal region of DADH as opposed to the C-terminal end in LADH (figure 3.1.6), suggesting DADH belonged to an unrelated family of alcohol dehydrogenases[445]. Another member of this family was also identified in 1981, a prokaryotic ribitol dehydrogenase[446]. Over the last 25 years the SDR superfamily has grown to over 3000 enzymes, making it one of the largest families in nature[382,447]. This is reflected by the wide substrate spectrum of these enzymes, ranging from alcohols, through sugars and steroids to xenobiotics[447,448]. The main structural difference between SDR and MDR enzymes is that SDRs have one structural domain, sufficient for both binding cofactor and catalysis, whereas MDRs have distinct cofactor binding and catalytic domains (figure 3.1.6). Despite catalysing homologous reactions and having a conserved Rossmann fold, the enzyme classes have distinct mechanisms[449,450].

The SDR enzymes typically share low sequence identity of around 15 – 30 % and although there are few strictly conserved residues there are several recognizable sequence motifs[382,448] (table 3.1.7). The two most important motifs are the coenzyme binding motif GxxxGxG[451] and the active site motif YxxxK (highlighted in figures 3.1.5 and 3.1.6) which, along with less well conserved S / T, makes up the catalytic triad. The SDR superfamily can be further classified based on sequence and is split into two main families, the classical and extended SDRs (table 3.1.7)[382]. The classical SDRs include oxidoreductases such as steroid dehydrogenases and are typically around 250 amino acids in length. Extended SDRs are usually slightly larger, around 350 amino acids, and include isomerases and lyases (dehydratases), such as galactose epimerases and glucose dehydratases respectively, as well as some oxidoreductases[382,452]. A more recent study of SDR enzymes by Kallberg *et al.*

identified three new families, divergent, intermediate and complex SDRs and further classifies the classical and extended SDR families into 7 and 3 sub-families respectively[448]. This classification is based on more subtle differences in the conserved motifs, designed to functionally assign the SDRs[448].
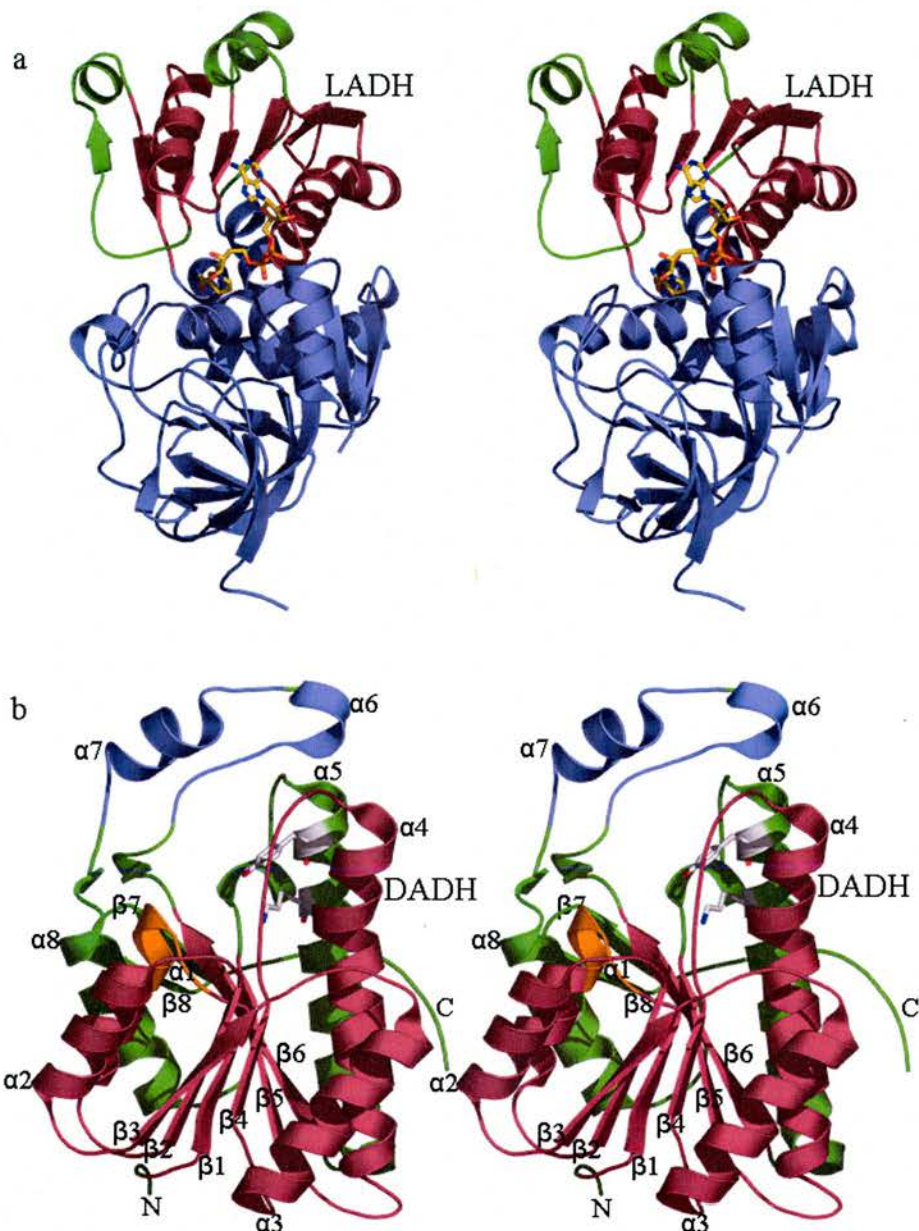
Figure 3.1.6: Stereo representation of (a) Horse liver ADH (PDB code: 1MGO)[453] and (b) *Drosophila* ADH (PDB code: 1SBY)[454] structures (b), members of the MDR and SDR superfamilies respectively. The Rossmann fold in each enzyme is coloured in red. The substrate binding and catalytic domain of the MDR and the substrate binding, functional sub-domain of the SDR enzymes are coloured blue. In DADH the conserved GxxxGxG motif is coloured orange and the strictly conserved active site tyrosine and lysine are shown in stick.

Secondary structure	SDR motifs			Possible function
	Classical	Extended		
β1 + α1	TGxxxGhG	TGxxGhaG		Structural role in coenzyme binding
β3 + α3	Dhx[cp]	DhxD		Adenine ring binding of coenzyme
β4	GxhDhhhNNAGh	[DE]xhhHxAA		Structural role stabilising central β-sheet
α4	hNhxG	hNhhGTxxhhc		Part of active site
β5	GxhhxhSSh	hhhxSSxxhaG		Part of active site
α5	Yx[AS][ST]K	PYxx[AS]Kxxh[DE]		Part of active site
β6	h[KR]h[NS]xhxPGxxxT	H[KR]xxNGP		Structural role

Figure 3.1.7: Table of major sequence motifs in classical and extended SDR families[448]. Labelling of secondary structure is consistent with figures 3.1.6b and 3.3.5. Adapted from Kalberg et al.[448].

3.1.6   Structure of SDR enzymes

The structure of a typical SDR enzyme is shown in figure 3.1.6b. It has a 7-stranded β-sheet in the centre of the protein, flanked on either side by three α-helices. Two α-helices are not part of the main body of the structure and these, along with the C-terminus, are involved in substrate recognition and binding[452,455,456]. The 7-stranded β-sheet makes a hydrophobic core and binds $NAD(P)^+$. These sheets contain the conserved GxxxGxG motif[451] as well as several conserved hydrophobic residues important in cofactor binding[382,457]. Six strands of the β-sheet, along with 4 of the flanking α-helices, make up the characteristic Rossmann fold, which has two βαβαβ units with the βαβ motifs having the usual right handed crossover[458] (figure 3.1.6). The seventh strand, which is separated by the two C-terminal α-helices, is linked by a left-handed crossover[452]. The core Rossmann fold displays higher sequence similarity

between SDRs (figure 3.1.5) than the C-terminal region and in the structures solved is superimposible. The major structural difference between members of the extended family of SDRs (and SDRs in general) is in the C-terminus, with the two smaller α-helices not superimposing well, indicative of the wide substrate diversity[455,456]. There are a small number of examples of monomeric SDR enzymes[459], but the vast majority exist as either a dimer or tetramer (reviewed by Tanaka *et al.*[452]).

3.1.7 Coenzyme specificity

The βαβ fold making up part of the Rossmann fold is found in enzymes that utilise nucleotide cofactors such as NAD(P)(H) and FAD[451], however side chain variations of amino acids surrounding this motif have evolved, determining which cofactor binds. In NAD(H)$^+$ utilising enzymes there is usually an acidic residue, D or E, at the C-terminal end of the second β-strand (β2), which hydrogen bonds to the 2'- and 3'-hydroxyl groups of the adenine ribose moiety (figure 3.1.5)[448]. NADP$^+$ utilising SDRs lack this acidic residue as it would electrostatically repel the negative 2'-phosphate group and instead have two positive residues, usually R or K, which stabilise NADP$^+$ binding[448,452]. These two residues are found at **X** on the Gxx**X**GxG motif and immediately following the position where the acidic residue is conserved in NAD(H)$^+$ utilising enzymes (figure 3.1.5). It is not surprising that the roles of these negative and positive residues have evolved to select the use of NAD$^+$ or NADP$^+$ as the only structural difference between these cofactors is a large, negative phosphate group at the O2 position of the adenine ring. Importantly in all SDR enzymes, binding of the cofactor orientates the nicotinamide ring close to the active site catalytic triad, S/T and YxxxK.

3.1.8   Hydride Transfer
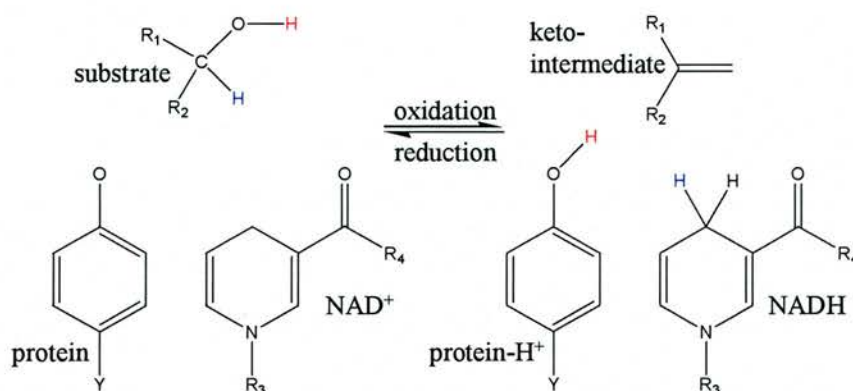
Figure 3.1.8: Proton and hydride transfer catalysed by SDR enzymes.

SDRs usually catalyse oxidoreductase reactions but also include dehydratases, epimerases and decarboxylases. Common to all of these reactions is the transfer of a hydride and a proton between substrate and protein/NAD(P)$^+$ (figure 3.1.8)[382]. Extensive mutational and structural studies have confirmed the importance of the catalytic triad, with mutations in Y, K or S/T leading to complete loss or reduction of enzyme activity[460-464]. This triad is located close to the nicotinamide ring of the NAD(P)$^+$ and the carbon of the substrate. This permits Y, which acts as a base, to extract a proton. The conserved K plays a role in the binding and positioning of the nicotinamide moiety of the NAD(P)$^+$ cofactor and, along with the positive group of NAD(P)$^+$, lowers the acidity of the Y to a pKa ~7. This results in the tyrosine side chain being deprotonated, allowing it to transfer a proton to and from substrate[463]. Concomitant with proton transfer, hydride is transferred between NAD(P)$^+$ and substrate (figure 3.1.8)[461,465]. The precise role of the less stringently conserved S/T in the reaction is disputed.

Early reports based on the structures of *E. coli* GalE, showed that the active site Y was too far from the hydroxyl group of the galactose substrate to directly

extract a proton[460,466-468]. It was proposed that the S/T residue acted to relay the proton from the galactose hydroxyl group to Y , the so-called shuttling mechanism[463,469]. However crystal structures of human GalE[463], RmlB[470] and SQD1[471] (an SDR that converts UDP-glucose to UDP-sulfoquinovose, important for sulfolipid biosynthesis) have shown that Y acts directly as a base. In the dimeric human GalE structure the catalytic Y is found 3.1 Å and 4.1 Å from the glucosyl C'4-hydroxyl group in different monomers[463]. This suggests that the position of the catalytic Y in the *E. coli* GalE structures is a crystallographic artifact and that the S/T residue has a different role in SDR oxidation[469]. Further crystal structures of similar SDR enzymes, RmlB[470] and RmlD[472], reveal that this residue makes close hydrogen bonds with the substrate, leading to the proposition that it forms a low barrier hydrogen bond (LBHB)[470,471]. The LBHB is thought to help stabilise any reaction intermediates or transition states[473,474]. The structure of RmlD proposed that a LBHB is crucial but exists between the transition state of the substrate and catalytic Y[472]. The role of S/T is suggested to fine tune the pKa of the transition state to facilitate proton transfer[469,472].

3.1.9   GalE catalysed epimerisation

GalE catalyses the interconversion of UDP-galactose and UDP-glucose (figure 3.1.9). It does this by oxidising C4', allowing the movement of the C4' hydroxyl group from down to up positions[461]. The catalytic Y_{149} (labelling consistent with *E. coli* GalE) has its pKa lowered sufficiently by the influence of K_{153} to allow it to either abstract or donate a proton[461,463,475,476]. Upon extraction of the C4' hydroxyl proton a hydride is transferred from C4' to NAD$^+$, producing a 4'-ketopyranose intermediate and NADH. Key to the whole reaction is the ability of the carbohydrate to rotate within the active site and be re-reduced by NADH at the opposite face.
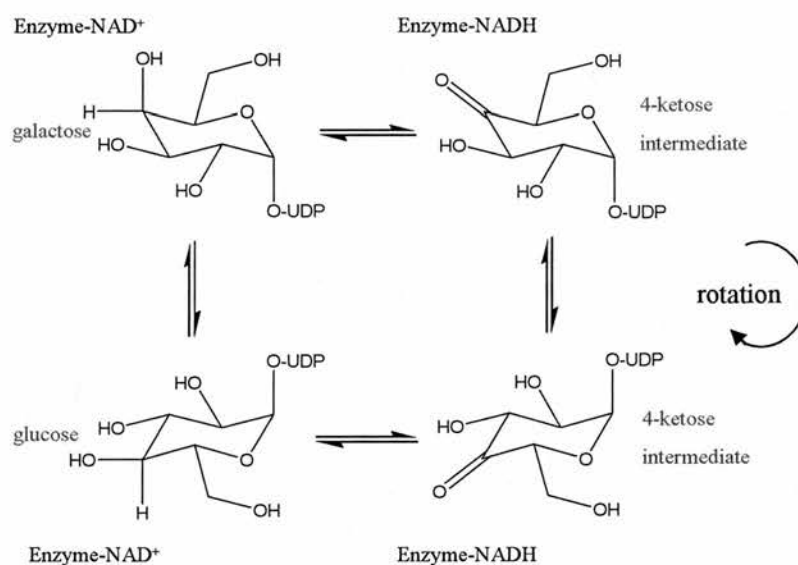
Figure 3.1.9: Mechanism of the *E. coli* GalE enzyme, which catalyses the interconversion of UDP-galactose and UDP-glucose[476]. Adapted from Berger *et al.*[476].

3.1.10 RmlB dehydratase

The committed catalytic step for the biosynthesis of most deoxy sugars is the C6'
deoxygenation catalysed by NDP-hexose 4,6-dehydratases such as RmlB, which takes
dTDP-glucose and produces dTDP-4-keto-6-deoxy-D-glucose[470,477,478]. The
chemistry of this reaction involves three separate steps; oxidation by NAD^+ at C4',
followed by concerted elimination of water from C6' and C5' and finally reduction at
C6' by NADH to produce product and regenerated NAD^{+}[479]. The reaction
mechanism is shown in figure 3.1.10. High resolution structural analysis along with
ab initio electronic structure calculations show that in RmlB the NADH formed is
buckled, thereby increasing this enzymes reducing potential, activating hydride
transfer[480]. Mutational and structural studies have shown that E_{135} and N_{134} are
essential for the second step of the reaction mechanism, elimination of water. The
formation of the keto intermediate leads to a lowering of the pKa of C3' and C5'
protons which are α to the C4' keto group. The C5' proton points towards the side
chain of E_{135}, which is proposed to extract this proton, with the resulting negative
charge stabilised as an enolate[470,480]. Elimination of water requires the C6' hydroxyl
group to be protonated, with D_{134} supplying this proton. RmlB-like enzymes have a
conserved N that is located close to this hydroxyl group and since a negative charge
would partly remove the proton from the OH group, preventing C6'-O6' cleavage, this
residue must itself be protonated[470]. Elimination of water leads to the formation of a
glucosene intermediate which is reduced by NADH. This mechanism requires
rotation of the carbohydrate within the active site to move the C6' position of the
glucosene close enough to NADH to allow hydride transfer to occur[470].
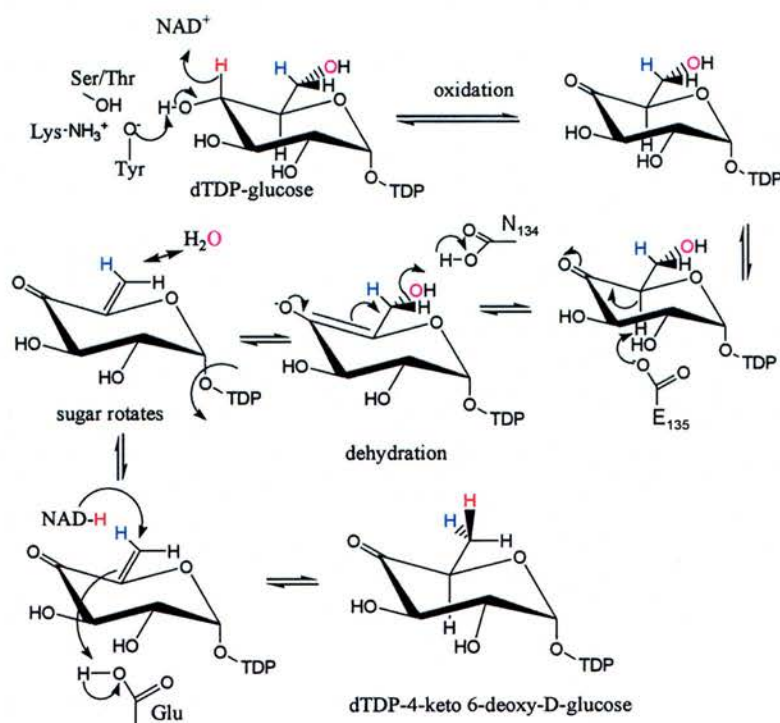
Figure 3.1.10: Reaction mechanism of the RmlB enzyme consisting of three distinct steps; oxidation, elimination of water and reduction. Amino acid labelling is consistent with RmlB from *S. typhimurium*. Taken from Allard *et al.*[470].

3.1.11   ArnA

The ArnA residues T_{432} and $Y_{463}SVSK_{467}$ represent the conserved catalytic triad S/T and YxxxK. This would be consistent with an initial oxidation step carried out by the ArnA decarboxylase domain to form a keto-sugar intermediate, which then decomposes by decarboxylation. Breazeale *et al.* first demonstrated that the oxidative decarboxylation of UDP-GlcUA to UDP-LAra4O was catalysed by ArnA[377]. It was originally postulated that ArnA catalysed the oxidation of UDP-GlcUA and a second enzyme, PmrJ, catalysed the decarboxylation step[363,371]. By showing that ArnA was sufficient for both the oxidation and subsequent decarboxylation reactions they raised the question as to whether decarboxylation occurred spontaneously or was enzyme

103

catalysed? UDP-glucuronic acid decarboxylases, which are also called UDP-xylose synthases, reutilise NADH to reduce the 4C'-keto group of the keto-intermediate, producing UDP-xylose[481]. In contrast ArnA releases UDP-LAra4O and NADH as products[377]. Determination of the ArnA decarboxylation structure is important for several reasons. Firstly, as mentioned in Chapter 2, it is a potential drug target due to its role in the L-Ara4N modification of Lipid A, a mechanism of resistance to CAMPs. Secondly, the structure would help determine whether the decarboxylation step is enzyme catalysed and if so, how so? Thirdly, the structure may give insights into why NADH is not reutilised in ArnA.

3.2 MATERIALS AND METHODS

3.2.1 Selenomethionine expression and purification

Because of the relatively low sequence identity of the decarboxylase and formyltransferase domains to other structures in the PDB database and the failure of initial molecular replacement attempts, selenomethionine protein was grown. 250 ml M9 medium, supplemented with 50 µg ml^{-1} kanamycin in 2 l flasks was inoculated with an overnight culture of *E. coli* BL21(DE3), transformed with either ArnA_DH or ArnA_FT plasmids (table 2.2.2), for growth of selenomethionine modified (Se-Met) protein. Se-Met protein was produced using a standard protocol as described by Doublie[482]. Decarboxylase and formyltransferase Se-Met protein was purified to homogeneity and prepared for crystal trials using the same procedures as for native protein (chapter 2.2), but 7 mM DTT was added to all buffers to prevent oxidation of the selenomethionine. A small aliquot of pure protein was sent for whole protein mass spectrometry analysis and confirmed full selenomethionine incorporation, with the formyltransferase results shown in figure 3.2.1.

a

b
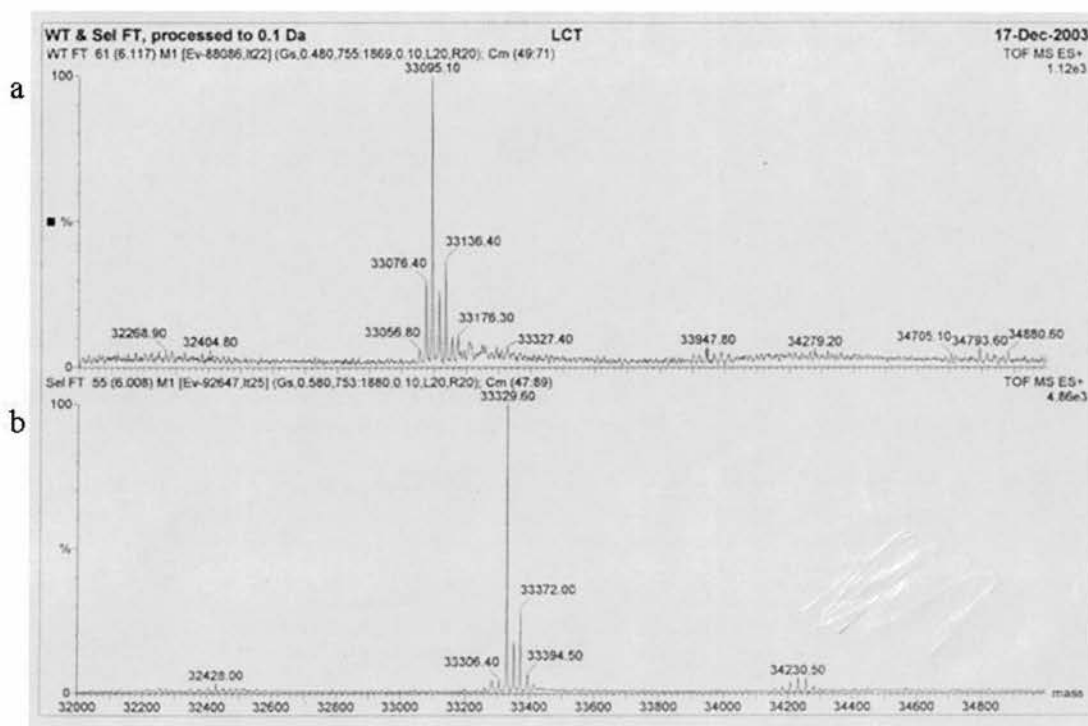
Figure 3.2.1: Whole protein mass spectrometry analysis of Se-Met (b) and native (a) formyltransferase protein, with the mass difference of 234.5 corresponding to full incorporation of 5 Se.

3.2.2   Crystallisation of Se-Met protein

Se-Met formyltransferase protein was used to screen for new crystallisation conditions, at both 20 and 4 °C, as conditions giving rise to native crystals (2.2.4) did not produce Se-Met crystals. Crystals suitable for X-ray diffraction formed at 4 °C in 0.22 M ammonium fluoride and 22 % PEG 3350, with drops containing 3 μl protein and 3 μl precipitant against a well of 100 μl mother liquor. Decarboxylase Se-met protein crystallised in 2.9 M NaCl and 0.1 M Bis-tris pH 5.6, a similar condition to native protein.

3.2.3   **MAD data collection of formyltransferase and decarboxylase SeMet crystals**

In house screening determined that the croprotectant, space group and unit cell of the Se-Met formyltransferase and decarboxylase crystals were essentially the same as native crystals (2.2.3). A single Se-Met formyltransferase crystal was used to collect multiwavelength anomalous diffraction/dispersion (MAD) data to 1.6 Å at three wavelengths. The wavelengths were chosen as peak 0.9793 Å, inflection 0.9796 Å and remote 0.9393 Å (table 3.2.2), as measured by fluoresence scan on the ESRF beamline ID14-4. Similarly, a single Se-Met decarboxylase crystal was used to collect MAD data to 3.35 Å at three wavelengths (table 3.2.7) on beamline BM14 at the ESRF. Data were indexed and integrated in MOSFLM[396], scaled in SCALA[401] and merged in TRUNCATE. Scaled but unmerged data were used for structure determination. Formyltransferase and decarboxylase data collection statistics are shown in tables 3.2.2 and 3.2.7 respectively.

3.2.4   **Structural solution and refinement**

Formyltransferase domain

Running SOLVE[483] with data from all three wavelengths cut to 2.5 Å located 8 of the 10 selenomethionines in the asymmetric unit, with sites listed in table 3.2.3 corresponding to 2 formyltransferase molecules as predicted by Matthew's coefficient. The electron density map output from SOLVE is shown in figure 3.2.4a and there are two separate regions within the map that are slightly darker than the background noise

(circled) which, along with the good statistics from SOLVE (table 3.2.2 and 3.2.3), suggested a solution. The initial phases from SOLVE were improved in RESOLVE, which also carries out density modification including solvent flattening[484]. Figure 3.2.4b shows the electron density map output from RESOLVE and with the solvent flattened the density corresponding to the two formyltransferase monomers is much clearer. Closer inspection of electron density maps in O[485] showed continuous electron density consistent with protein secondary structure, with clear density for side chains confirming a structural solution.

Data Collection	Peak	Inflection	Remote
Wavelength (Å)	0.9793	0.9796	0.9393
Resolution (Å)		24-1.6 (1.63-1.6)	
Unique reflections	77800	77914	77900
Space Group		$P2_12_12_1$	
Cell Dimensions		a=67.5, b=90.7, c=97.6 $\alpha=\beta=\gamma=90°$	
Beamline		ESRF ID14-4	
Detector		ADSC-CCD	
I/σ	8.8 (6.3)	8.9 (6.2)	8.8 (6.2)
Multiplicity	4.4 (4.5)	4.4 (4.5)	4.3 (4.4)
Completeness (%)	96.8 (95.8)	96.9 (95.8)	96.8 (95.8)
R_{merge}* (%)	5.1 (11.5)	5.1 (11.3)	5.1 (11.3)
SOLVE			
Se sites		8	
Refined f'/f''	-7.8 / 4.9	-10.2 / 2.6	-3.5 / 3.1
Z score		21.3	
Phasing power		0.64	

*$R_{merge} = \sum \sum I (h)i - (I (h)) / \sum \sum I (h)i$, where $I(h)i$ is the measured diffraction intensity and the summation includes all observations

Table 3.2.2: Data statistics for the formyltransferase 3 wavelength MAD experiment.

X	Y	Z	Occupancy	B-	Peak Height (σ)
0.112	0.221	0.180	1.000	10.1	28.2
0.102	0.682	0.021	0.991	25.5	21.6
0.393	0.642	0.031	0.819	15.4	23.9
0.751	0.113	0.234	0.943	9.7	25.7
0.229	0.547	0.032	0.969	10.8	25.0
0.012	0.802	0.217	0.809	13.5	20.1
0.868	0.541	0.143	0.997	15.1	24.6
0.281	0.270	0.250	0.898	20.5	19.9

Table 3.2.3: 8 Se sites found by SOLVE in the formyltransferase asymmetric unit.
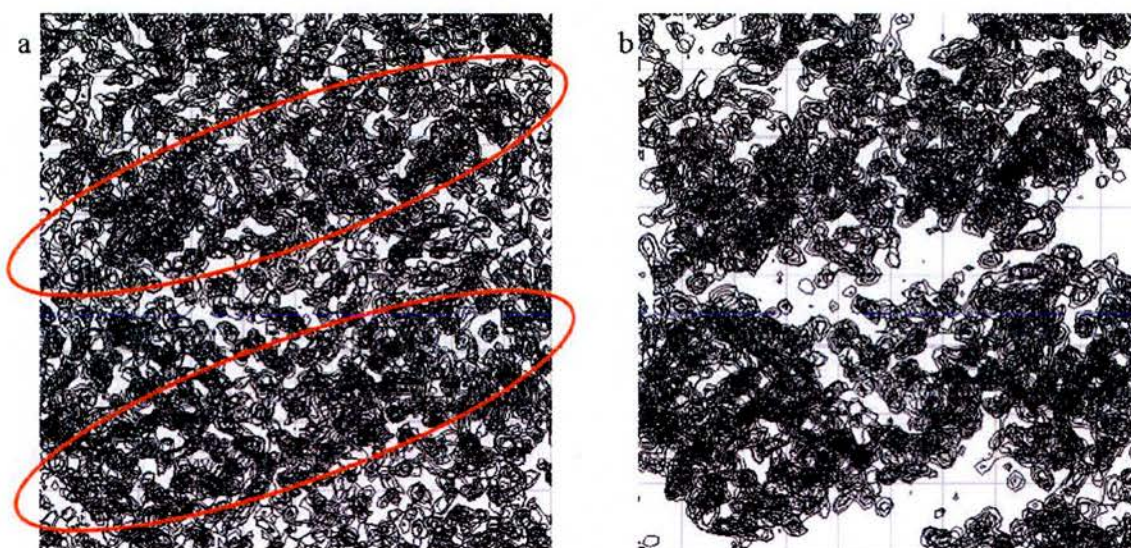
Figure 3.2.4: Initial electron density maps of the formyltransferase structure from SOLVE[483] (a) and from RESOLVE[484] after solvent flattening (b). The darker regions correspond to two formyltransferase monomers in the asymmetric unit (circled in a).

The 1.6 Å MAD data were merged with the high resolution 1.2 Å dataset, that was previously collected on a native formyltransferase crystal (2.2.5), using the CAD program as implemented in the CCP4 program suite[486]. Phases were extended to 1.2 Å and further density modification carried out in the DM program[487], before automatic chain tracing was carried out in ARP/wARP[488]. Over 90 % of the structure

was traced and side-chains were docked using guiSIDE[489]. Missing residues that had electron density were fitted manually by hand in O[485] or COOT[490]. During the building process it was evident that extra density was present in the structure that could not be accounted for by protein or water molecules and it was possible to model N-5-fTHF and uracil monophosphate UMP into these regions (figure 3.2.6). Libraries for these ligands were made in PRODRG[491] for use in refinement using the REFMAC5 program[403]. Initially rigid body refinement was carried out before TLS groups were defined and cycles of restrained anisotropic, TLS refinement and rebuilding were performed[403]. The final refined R factor and R_{FREE} values of 13.5 and 15.8 were reached before validation by PROCHECK[492] and WHATIF[493] and structure deposition into the PDB. The refinement statistics are presented in table 3.2.5. Figure 3.3.1 shows the structure of the two formyltransferase monomers in complex with N-5-fTHF and UMP that are found in the asymmetric unit discussed in 3.4.

Refinement	Formyltransferase
Resolution	1.2 Å (1.26-1.2)
RMSD bond length (Å)/angles (°)	0.019/1.93
Rfactor/R_{FREE} (%)	13.5/15.7 (14.4/18.2)
Residues in most favored regions[1] (%)	92
Residues in allowed regions[1] (%)	8
PDB code	2bln

Table 3.2.5: Refinement statistics for the formyltransferase structure. [1] These refer to the Ramachandran plot and are defined by PROCHECK[492]. Values in parenthesis correspond to the highest resolution shell.
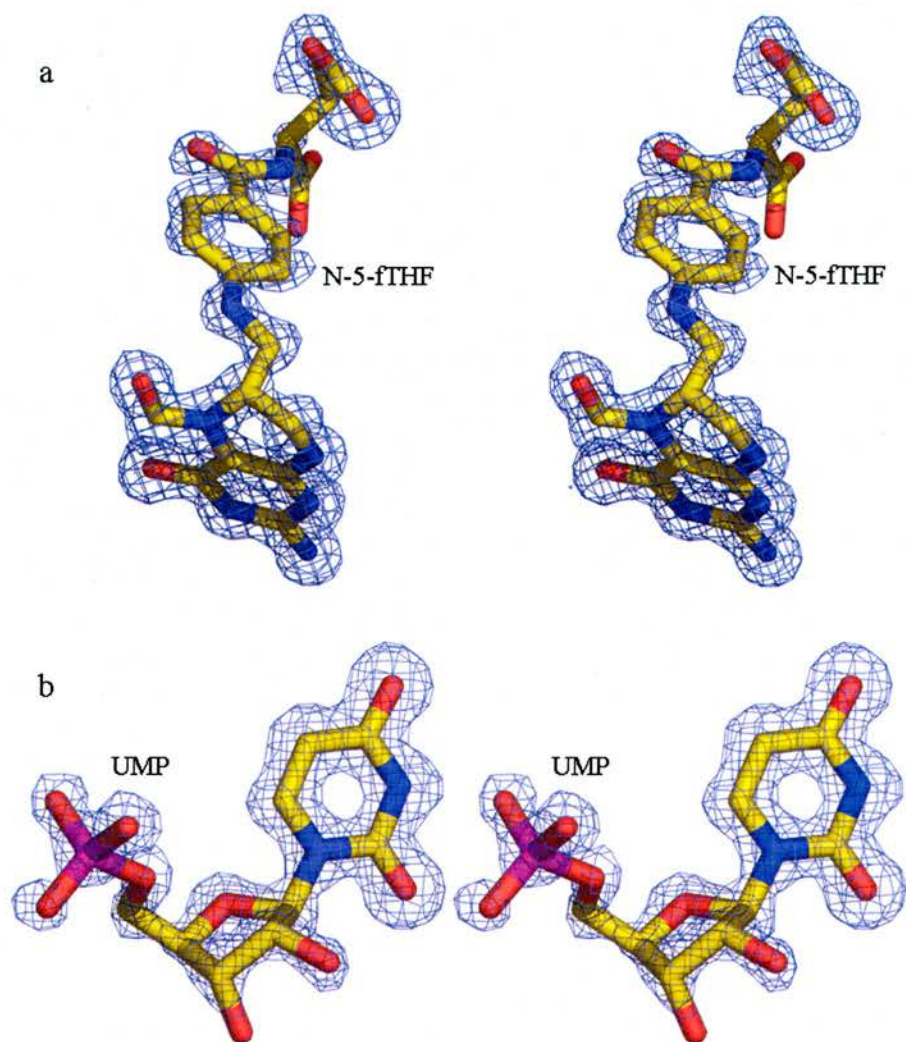
Figure 3.2.6: Stereo representation of N-5-fTHF (a) and UMP (b) modeled into their respective unbiased Fo-Fc electron density maps, found in the structure of the formyltransferase domain. The molecules are shown in stick with atoms coloured; carbon-yellow, nitrogen-blue, oxygen-red and phosphorous-purple. The map is contoured at 3σ (0.22eÅ^{-3}).

Decarboxylase domain

Data Collection	Peak	Inflection	Remote
Wavelength (Å)	0.9798	0.9710	0.9392
Resolution (Å)		61-3.35 (3.43-3.35)	
Unique reflections	15066	15065	15047
Space Group		P4₁32	
Cell Dimensions		a=b=c=149.5 α=β=γ=90°	
Beamline		ESRF BM14	
Detector		ADSC-CCD	
I/σ	4.4 (2.2)	4.7 (1.5)	5.3 (2.5)
Multiplicity	65 (67)	21 (21)	21 (21)
Completeness (%)	100 (100)	100 (100)	100 (100)
R$_{merge}$* (%)	14.3 (31.1)	12.4 (27.1)	13.4 (29.7)
SOLVE			
Se sites		4	
Refined f'/f''	-9.5 / 5.4	-10.1 / 2.5	-2.6 / 3.7
Z score		13.8	
Phasing power		0.52	

*R$_{merge}$ = $\sum \sum I$ (h)i – (I (h)) / $\sum \sum I$ (h)i, where I(h)i is the measured diffraction intensity and the summation includes all observations

Table 3.2.7: Data statistics for the decarboxylase 3 wavelength MAD experiment.

Values in parenthesis correspond to the highest resolution shell.

Because the space group was either P4₁32 or P4₃32 (2.2.5) SOLVE[483] was run in both hands. Good statistics were only obtained using the space group P4₁32, with SOLVE locating all 4 selenomethionines (tables 3.2.7 and 3.2.8), corresponding to one monomer in the asymmetric unit. Observation of the raw map from SOLVE identifies a darker region indicating that this area is above noise, suggesting that a correct solution had been found (figure 3.2.9). RESOLVE was run to improve the phases and carry out solvent flattening and the map from RESOLVE clearly shows the protein region (figure 3.2.9)[484].

X	Y	Z	Occupancy	B-	Peak Height (σ)
0.459	0.423	0.078	0.459	35.9	38.6
0.797	0.097	0.039	0.797	60.0	38.4
0.708	0.223	0.063	0.708	60.0	29.2
0.595	0.157	0.053	0.595	60.0	28.0

Table 3.2.8: 4 Se sites found by SOLVE in the decarboxylase asymmetric unit.
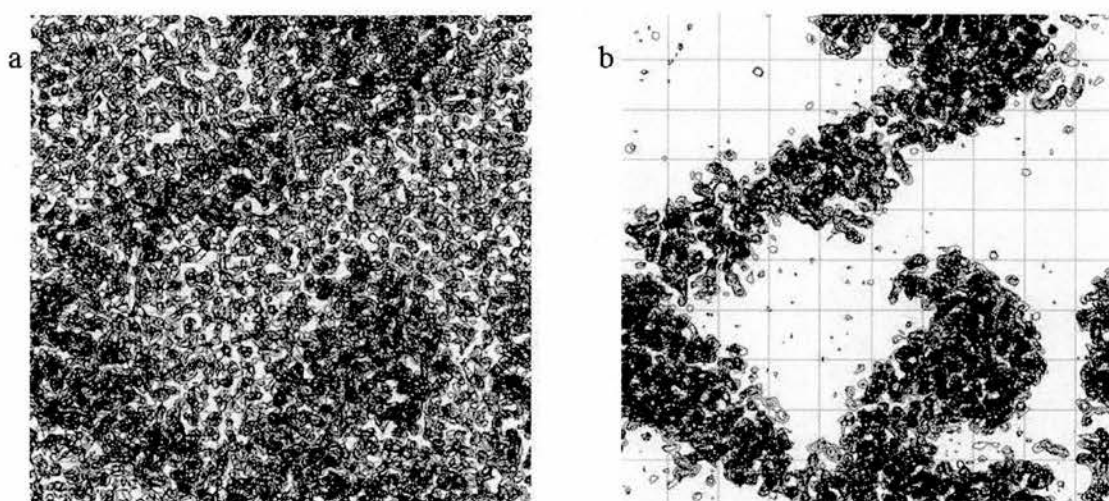
Figure 3.2.9: Initial electron density maps of the decarboxylase structure from SOLVE[483] (a) and RESOLVE[484] after solvent flattening (b). The darker regions correspond to protein.

The 3.35 Å data was merged with the higher 2.3 Å data using CAD from the CCP4 suite [486], with phase extension and further density modification performed in DM[487]. The 2.3 Å electron density maps were viewed in O[485] and interpretation of this density confirmed that the decarboxylase domain had been solved. This data, as well as the decarboxylase domain sequence file, was then input into ARP/wARP[488], which built over 90 % of the structure and automatically fitted side chains[488]. Missing residues that had electron density were modelled by hand in O or COOT[490] and water

atoms were added using ARP/wARP[488]. Refinement cycles were carried out in REFMAC5 using first rigid body and then restrained TLS refinement[403] to give final Rfactor and R_{FREE} values of 18.6 and 22.9 % respectively, with some refinement statistics shown in table 3.2.10. Manual inspection of electron density maps shows no density that is not accounted for by protein or water molecules and the apo-structure is presented in figure 3.3.5.

Refinement	Decarboxylase
Resolution	2.3 Å (2.42-2.30)
RMSD bond length (Å)/angles (°)	0.019/1.63
Rfactor/R_{FREE} (%)	18.6/23.0 (21.1/26.5)
Residues in most favored regions[1] (%)	90
Residues in allowed regions[1] (%)	10
PDB code	2b11

Table 3.2.10: Refinement statistics for the formyltransferase structure. [1] These refer to the Ramachandran plot and are defined by PROCHECK[492]. Values in parenthesis correspond to the highest resolution shell.

114

3.2.5   Site directed mutagenesis
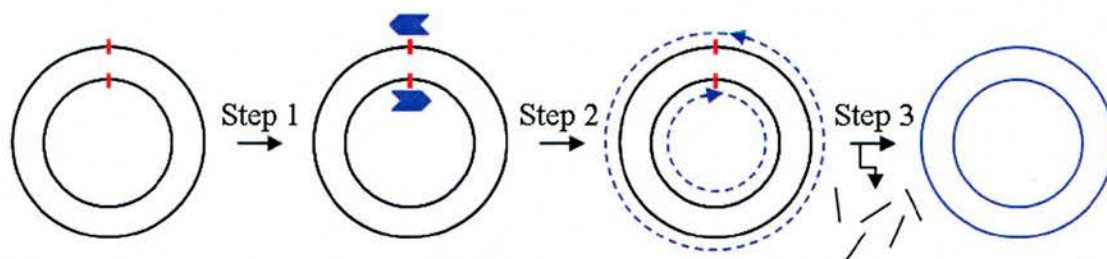
Figure 3.2.11: Major steps of site directed mutagenesis. Forward and reverse primers (blue arrows) are designed around the mutation site (red line) and used to amplify both strands of the methylated DNA template (black line) (steps 1 and 2). The newly synthesised, un-methylated DNA base pairs to form a nicked plasmid and the template DNA is digested by the restriction endonuclease DpnI (step3). This nicked vector containing the mutant sequence is then transformed into super-competent *E. coli* which ligates the nick and the new plasmid can be harvested.

Based on the superimposition of the RmlB active site with ArnA (figure 3.3.6), and the conservation of these residues in other UDP-GlcUA decarboxylases, it was decided to mutate the S_{433} and E_{434} residues to assess their role in the ArnA reaction. S_{433} and E_{434} were both mutated to A and E_{434} was additionally mutated to the isosteric Q, using the QuikChange® protocol, outlined in figure 3.2.11. The reaction mixture included the ArnA-DH plasmid (table 2.2.2),P*pfu* DNA polymerase, dNTP's and the relevant oligonucleotides (table 3.2.12). The QuikChange® protocol uses a PCR based technique to amplify the whole vector using forward and reverse primers. These primers contain the relevant base pair changes to incorporate the mutation and are designed around the mutation site. The PCR reaction uses the high fidelity DNA polymerase *Pfu*, however it does not correct the mismatched sequence because this is

located close to the 3' end of the primer, where the proof reading ability of the enzyme is poor. The result of this PCR is the amplification of both strands of the vector, with the new sequence containing the relevant mutation. Importantly, the newly synthesised DNA is un-methylated and complimentary base pairs with one another and the template DNA. As the template DNA is methylated (by the *E. coli* cells it is extracted from) it can be specifically digested by the restriction enzyme DpnI, which only digests methylated and hemi-methylated duplex DNA. This leaves a mixture of unmethylated, nicked plasmids containing the desired mutation and digested DNA fragments, which is used to transform the super-competent *E. coli* TamI cells (Novagen). The *E. coli* cells further digest the fragments of template DNA and the nicked plasmid is ligated together by DNA ligase.

DNA sequencing of the vectors extracted from TamI cells after site-directed mutagenesis confirmed the presence of the mutations, with no other mutations having been inserted elsewhere in the gene. These new vectors were used to transform *E. coli* BL21(DE3) cells for over-expression and subsequent purification of the mutant proteins, following the same procedures as for wild-type protein (2.2.2 and 2.2.3). Pure protein of the three mutants, along with native decarboxylase protein, was sent to the Protein Characterisation Facility, Institute of Biomedical & Life Sciences, University of Glasgow for Circular dichroism and fluorescence spectra analysis. The spectra for the mutant and wild-type proteins are very similar (figure 3.2.13), strongly suggesting that all of the proteins have the same secondary structure and are properly folded.

S433A   3'-CGTCAACT**(C)-GCA**GAAGTTTATGGG-5'

3'-CCCATAAACTTC**TG(G)-C**AGTTGACG-5'

E434A   3'-CGTCAACTTCA**G(A)-CA**GTTTATGGG-5'

3'-CCCATAAAC**TG(A)-C**TGAAGTTGACG-5'

E434Q   3'-CCCGTCAACTTCA**(G)-CA(A)-G**GTTTATGGG-5'

3'-CCCATAAAC**(A)-CT(C)-G**TGAAGTTGACGGG-5'

Table 3.2.12: Oligonucleotides used for site-directed mutagenesis. Underlined and in bold is the mutated codon, with the original base in ().
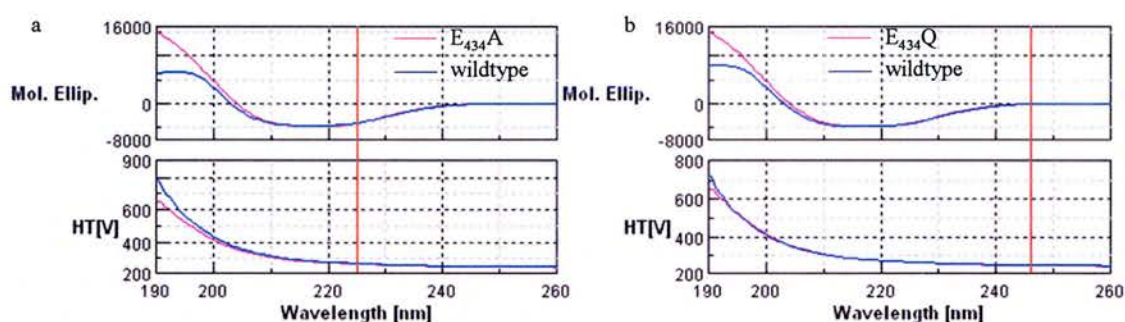
Figure 3.2.13: Circular dichroism-spectra of ArnA decarboxylase wild-type, $E_{434}A$ (a) and $E_{434}Q$ (b) mutant proteins. The far-UV spectra show that the secondary structure of the mutant proteins is the same as wild-type. The $S_{433}A$ spectra is identical to (b) and its structure (PDB code 1Z73) shows correct folding. The slight differences at low wavelength (used to determine tertiary structure) are because the decarboxylase protein didn't exhibit a CD signal in the near UV region. This situation usually occurs because the aromatic side chains are not particularly rigidly held in the tertiary structure of the protein.

3.2.6 Kinetic assay

The activity of wild-type and mutant ArnA decarboxylase enzymes to catalyse the C4' oxidation of UDP-GlcUA was assayed by following the rate of NADH production at A_{340}. The assays contained 50 mM HEPES pH 7.5, 300 mM NaCl, 2 mM DTT, 0.2 mg ml^{-1} bovine serum albumin, 8 mM NAD$^+$, 4.5 mM UDP-GlcUA and either 0.5 μM wild-type or 2.5 μM mutant protein. Reactions were carried out in 1 ml cuvettes at 30°C and the initial rate of reaction was measured over 10 minutes. These assays allowed us to assess the V_{max} for the decarboxylase enzymes as both NAD$^+$ and UDP-GlcUA are in a large excess. Apparent K_m values were determined by varying the amount of initial UDP-GlcUA or NAD$^+$ from 1/5 x to 5 x K_m, whilst keeping the unvaried substrate in excess. The results of these assays are shown in table 3.2.14. The experiments to calculate the reaction rate for native and mutant proteins were first carried out by our collaborators in North Carolina. We then repeated these results and calculated the V_{max} and K_m for the $E_{434}A$ mutant (table 3.2.14), while our collaborators calculated the V_{max} and K_m for native and $S_{433}A$ proteins. It is apparent that the $S_{433}A$ mutant and $E_{434}A$ mutants do retain some activity, approximately 40x and 100x reduction in activity compared to wild-type, whereas surprisingly the $E_{434}Q$ mutant has no detectable activity. Given that the assay measures the rate of NADH production it is possible that the $S_{433}A$ and $E_{434}A$ mutants are only defective in the decarboxylation step and therefore accumulate keto-intermediate. In order to probe this possibility the reaction was monitored by high pressure liquid chromatography (HPLC).

Enzyme	Specific activity (μMmin^{-1}mg^{-1})	UDP-GlcUA		NAD$^+$		Protein (μM)
		V_{max} (μMmin^{-1})	Apparent K_m (mM)	V_{max} (μMmin^{-1})	Apparent K_m (mM)	
ArnA Decarboxylase domain	350	6.5	0.7	6.8	1.3	0.5
S$_{433}$A Decarboxylase mutant	8.8	0.7	0.2	0.8	1.4	2.5
E$_{434}$A Decarboxylase mutant	3.3	0.3	0.4	0.3	1.6	2.5
E$_{434}$Q Decarboxylase mutant	No Activity	N/A	N/A	N/A	N/a	2.5

Table 3.2.14: Results of the kinetic assay, showing that although the E$_{434}$Q mutant is inactive, the activity of the S$_{433}$A and E$_{434}$A mutatnts are reduced around 40 and 100 times the wild-type protein respectively, without significantly perturbing UDP-GlcUA or NAD$^+$ binding.

3.2.7 HPLC assay

Wild-type and mutant reactions were carried out as described above and left overnight to allow accumulation of products, a time scale in which the wild-type reaction is approximately 50 % complete. Reaction substrates and products were separated from the protein by spinning the reaction in a concentrator with a 30 kDa molecular weight cut off, which retained the decarboxylase enzyme. The flow through from the concentrator was diluted 50 times and subjected to HPLC analysis, performed using a Varian Prostar machine. Ion-pair reversed-phase HPLC on a Varian OmniSpher C-18 column (250 x 4.6 mm) at a flow rate of 1 ml min^{-1}, using a method based on that of

Mäki et al.[494], was used to analyse the nucleotide sugars in the reaction mix. The column was equilibrated with 20 mM triethylammonium acetate (TEAA), pH 6.0. After isocratic elution with 20 mM TEAA (pH 6.0) for 7 min, nucleotide sugars were eluted with a linear gradient of 0 – 1 % acetonitrile in 20 mM TEAA (pH 6.0) over 28 min, NAD^+ and NADH were then eluted with a linear gradient of 1 – 25 % acetonitrile in 20 mM TEAA (pH 6.0) over 10 min. Effluent was monitored with UV detection at 254nm with NAD^+, NADH and UDP-GlcUA being identified in the reaction by co-elution with commercial standards (Sigma). This allowed the determination of the elution point of UDP-LAra4O from the wild-type reaction (fig. 3.2.15a) and this peak is clearly present in the $E_{434}A$ mutant reaction (figure 3.2.15b), the trace of which is identical to that of the $S_{433}A$ reaction. This analysis demonstrates that the reactions carried out by the mutant enzymes are identical to the wild-type reaction and that there is no additional build up of ketone intermediate.

The results of the HPLC assays were complemented with electrospray ionisation mass spectrometry (ESI MS) carried out by our collaborators in North Carolina. In this analysis wild-type, $S_{433}A$ and $E_{434}A$ reactions were left to run for 24 hours before dilution and ESI MS analysis. ESI MS identified masses corresponding to the UDP-GlcUA and NAD^+ substrates and hydrated UDP-L-Ara4O and NADH products, but not for ketone or hydrated ketone intermediates.
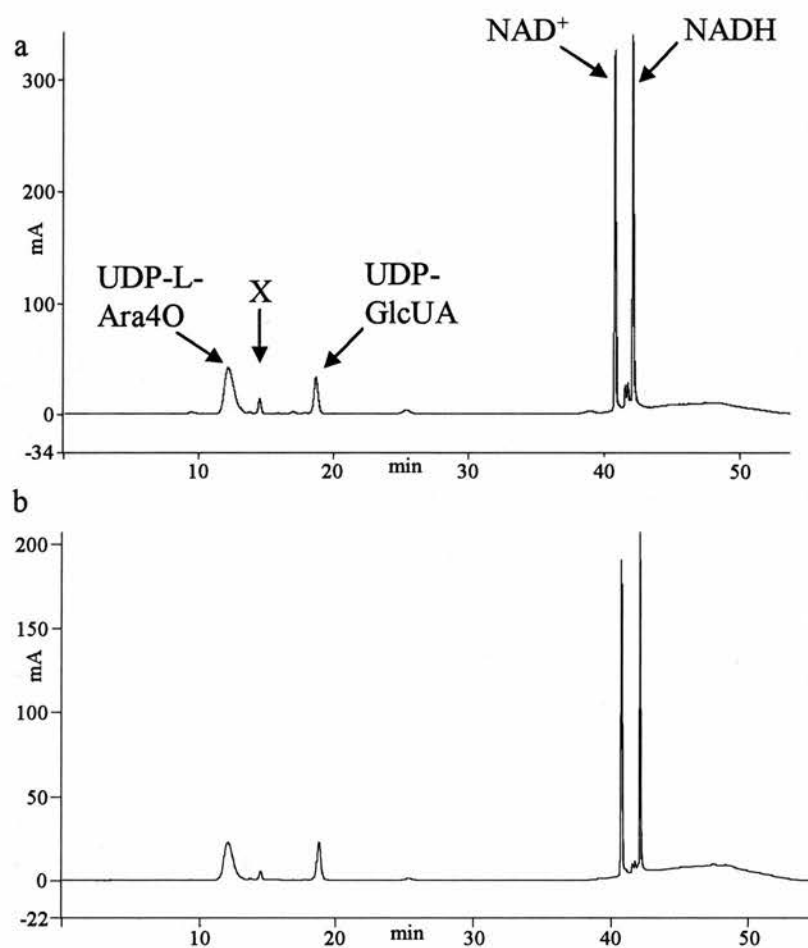
Figure 3.2.15: HPLC traces of wild-type (a) and E$_{434}$A (b) mutant reactions, showing that the reactions are the same and that there is no additional build up of ketone intermediate. The small peak marked X is unidentified, however its presence is consistent in both wild-type and mutant reactions.

3.3   RESULTS

3.3.1   Structure of the ArnA formyltransferase domain
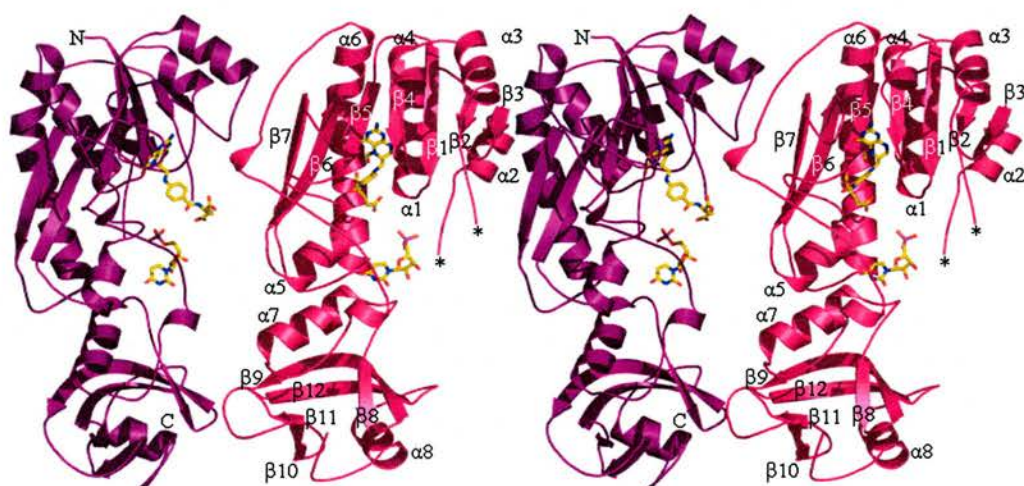
Figure 3.3.1:   The structure of the formyltransferase domain.   The protein is represented as ribbons and the secondary structure elements are numbered as the text. A loop (residues N_{35}-A_{40}) that is missing in the structure is labeled *.   The two subdomains are visible, the N-terminal subdomain is at the top of the figure and extends to the uridine ring, the C-terminal subdomain sits below the ring.   The two monomers found in the asymmetric unit are shown.   The decarboxylase domain is attached to the C-terminus (labeled C) in full length ArnA.

The formyltransferase domain of ArnA crystallises with two monomers in the asymmetric unit and buried surface calculations suggest that a functional dimer is not formed.   This is consistent with the structure of the apo-formyltransferase solved by Gatzeva-Topolova *et al.*, which crystallises with one monomer in the asymmetric unit[495]. The overall ArnA formyltransferase structure consists of two sub-domains; an

N-terminal domain (residues $M_1 - K_{181}$) and a smaller C-terminal domain (residues $F_{207} - N_{305}$), which are linked by a loop (residues $H_{182} - S_{206}$). This is a very similar structure to FMT[437], which also crystallises with two monomers in the asymmetric unit, and the N-terminal hydrolase domain of FDH[422]. This places ArnA formyltransferase into the family of formyltransferases similar to GART, characterised by a structurally conserved N-10-fTHF binding domain, rather than AICAR (figure 3.1.3). In ArnA formyltransferase the characteristic domain is made up of the N-terminal sub-domain and is folded into a 7-stranded β-sheet with 5 parallel β-strands (β1-5) and 2 anti-parallel strands (β6-7), flanked by a total of 6 α-helices, forming a version of the classic Rossmann fold[418,458]. The C-terminal sub-domain of ArnA formyltransferase is absent in GART but structurally conserved in FMT[420,437] and the hydrolase domain of FDH[422]. In ArnA it consists of 2 large β-sheets (β8 and β9), 2 small β-sheets (β10 and β11) and 2 α-helices (α7 and α8). The β-sheets fold to form a small β-barrel, which is flanked on either side by one of the helices, with α7 located parallel to α5 of the N-terminal domain (Figure 3.3.1). A structural alignment of the ArnA formyltransferase domain with FMT, the hydrolase domain of FDH and GART was carried out using DALI (figure 3.3.2) with RMSD's of 1.7, 2.2 and 2.2 Å respectively.
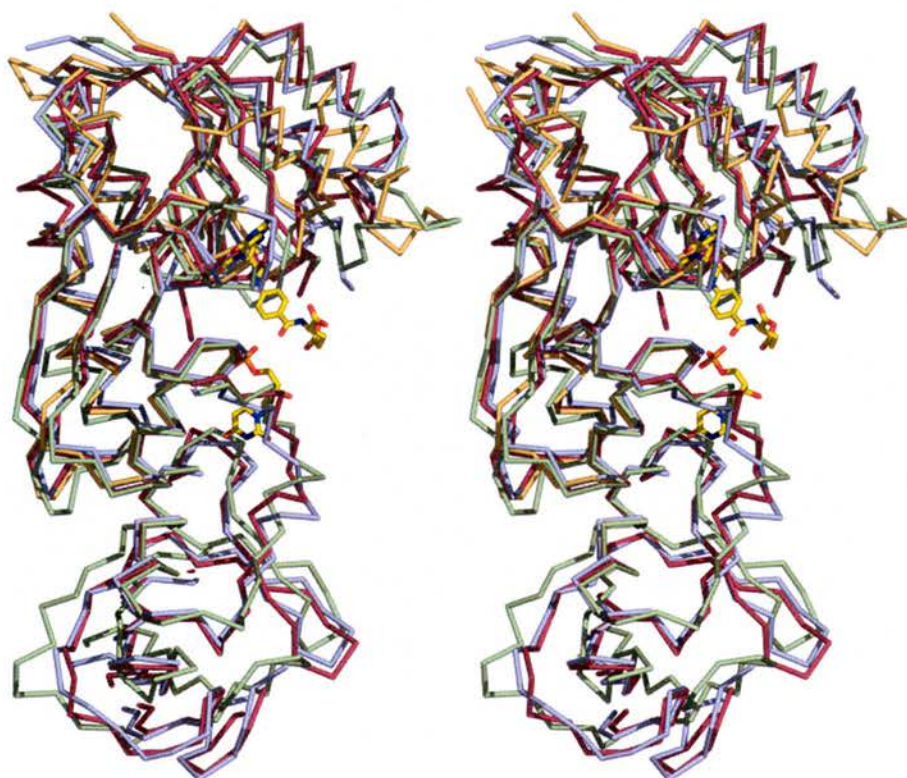
Figure 3.3.2: Superimposition of ArnA formyltransferase (red) with the structurally related enzymes FMT (green), FDH (blue) and GART (salmon). N-5-fTHF and UMP found in the ArnA active site are shown in stick and coloured as in figure 3.2.6. The larger N-terminal folate binding domain superimposes particularly well. Overall RMSD values are 2.2, 1.7 and 2.2 Å for GART, FMT and FDH respectively.

3.3.2 Ligand Binding

During refinement of the structure clear electron density that did not correspond to protein or water molecules was found and N-5-fTHF and UMP were modelled into this density, shown in figure 3.2.6. Although we have been unable to get a true substrate co-crystal, the location of these ligands in the structure and superimposition with GART and FMT, allows us to locate the ArnA formyltransferase active site. Relative to the protein fold, the active site of ArnA is in the same position as in FMT,

GART and the hydrolase domain of FDH, consistent with ArnA belonging to this class of N-10-fTHF utilising enzymes. In ArnA, UMP is bound mainly by the N-terminal domain and is adjacent to N-5-fTHF (Figure 3.3.1 and 3.3.3). Y_{42} is hydrogen bonded to O^{2P} and R_{85} to O^{1P} of the UMP phosphate, with a further hydrogen bond between O4 of uracil and R_{201} (Figure 3.3.3). Hydrophobic interactions occur between uracil and N_{118}, T_{202} and W_{228}. R_{201} and T_{202} are part of the loop connecting the N- and C-terminal sub-domains and W_{228} is part of $\alpha 7$ of the C-terminal subdomain (Figure 3.3.1).

N-5-fTHF binds adjacent to $\beta 4$ and $\beta 5$ with hydrogen bonds between the bicyclic ring and two loops, one between $\beta 4$ and $\alpha 4$ (residues I_{88} and H_{86}) and the other between $\beta 6$ and $\beta 7$ (residues V_{136} and D_{140}). Hydrophobic interactions between N-5-fTHF and residues L_{87} (on loop between $\beta 4$ and $\alpha 4$), N_{102} (on $\beta 5$), M_{135} and A_{139} (on loop between $\beta 6$ and $\beta 7$), also assist in the binding of the cofactor analogue. Figure 3.3.3 shows these interactions in more detail.
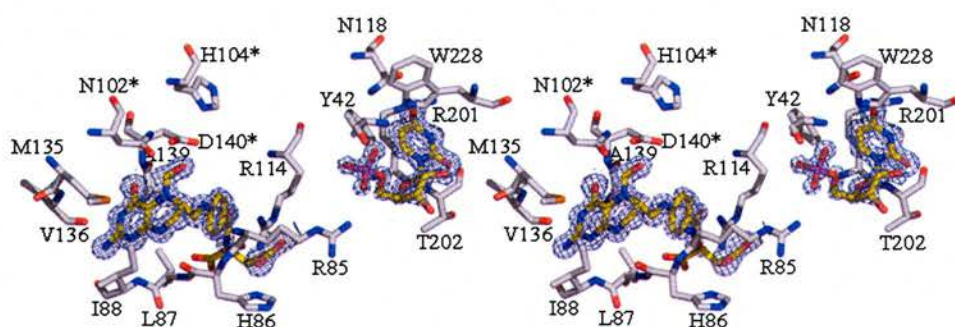
Figure 3.3.3: N-5-fTHF and UMP binding by ArnA formyltransferase. The map coloured in blue is the unbiased Fo-Fc map contoured at 3σ ($0.22e\text{Å}^{-3}$)., as shown in figure 3.2.6.

Analysis of the thermal factors of the N-5-fTHF atoms indicates the bicyclic ring is very well ordered, while the benzoyl ring is less ordered with the glutamate displaying the least order. This is in agreement with the proposal of Almassy et al.. that the bicyclic ring is more tightly bound than the benzoyl or glutamate moieties in GART [418]. Figure 3.3.4 shows a space-filled model of ArnA formyltransferase and this clearly shows why this is the case. There is a pocket into which the bicyclic ring is bound, effectively clamping this region in place. The benzoyl ring exits this pocket and the glutamate is more exposed to the surface. Interestingly there is a short tunnel leading from the UMP moiety towards the folate. This gives an insight into how the L-Ara4N moiety would be positioned towards the formyl group of N-10-fTHF, which has been modelled (figure 3.3.4).
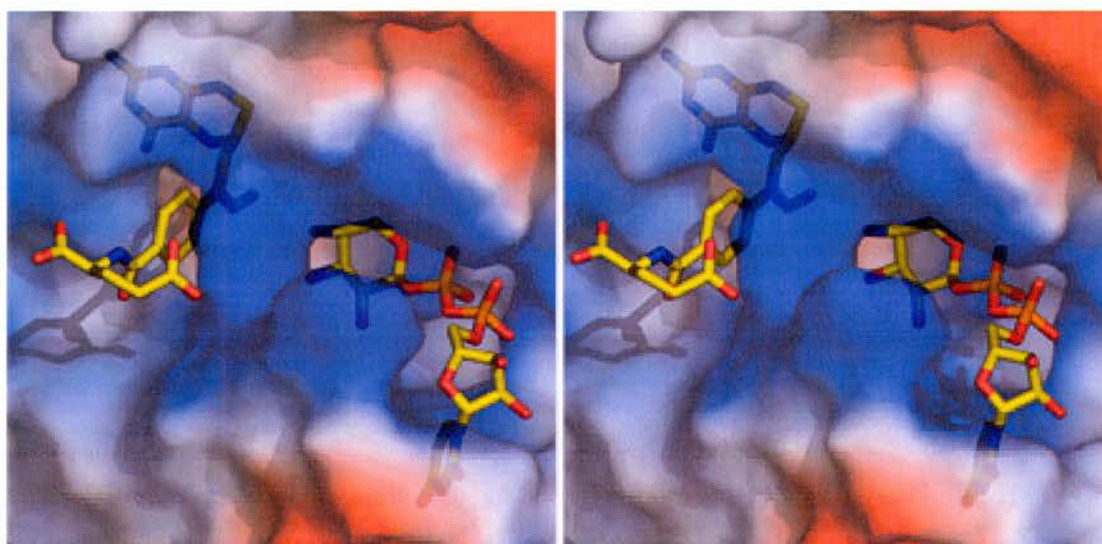
Figure 3.3.4: Electrostatic space fill model of ArnA formyltransferase active site with N-10-fTHF and UDP-Lara4N modelled in. The bicyclic ring of N-10-fTHF is buried in a cavity and there is a clear tunnel into which the carbohydrate of UDP-LAra4N binds, positioning it close to the N-10 formyl group. N-10-ffTHF and L-

Ara4N are shown in stick and coloured as in 3.2.6. Negative, positive and neutral amino acid side chains are coloured red, blue and white respectively.

3.3.3 Structure of the ArnA decarboxylase domain
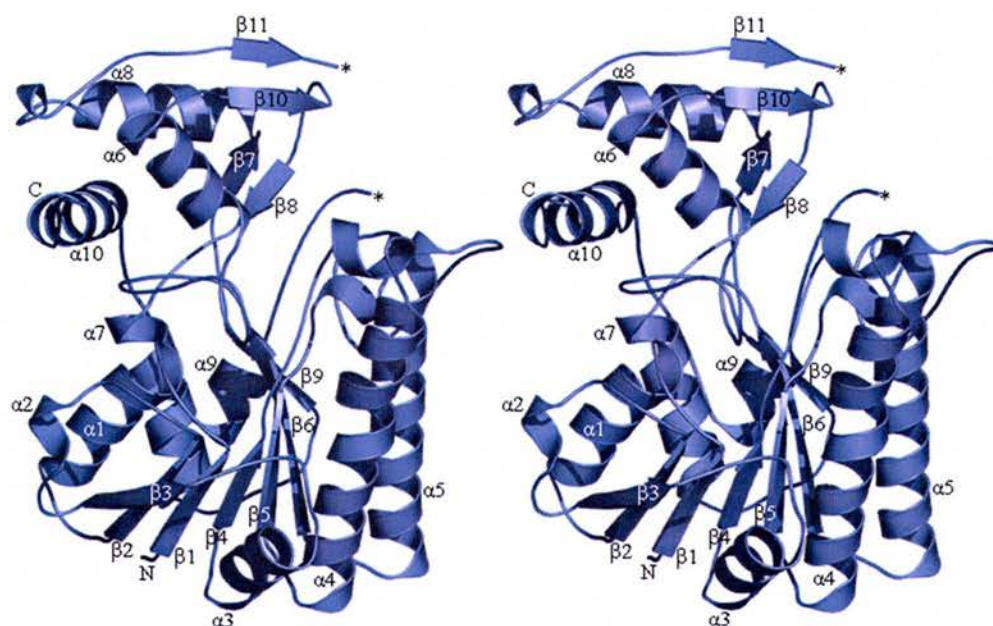
Figure 3.3.5.: The structure of the ArnA decarboxylase domain. The protein is represented as ribbons and the secondary structure elements are labeled as in the text. A loop (residues V_{604}-D_{615}) that is missing in the structure is labeled *. The monomeric unit is shown. The N-terminus of this domain (labeled N) is attached to the formyltransferase domain in full length ArnA.

The overall structure of the ArnA decarboxylase domain is shown in figure 3.3.5 and reveals that, as expected, it is a member of the SDR enzyme superfamily. It has a modified version of the classic Rossmann fold typical for SDR enzymes, to which an α-helix (α9) and β-strand (β9) are contributed from the C-terminal functional carbohydrate binding domain. Gatzeva-Topolova *et al.* have also solved the apo-decarboxylase domain of ArnA (PDB code 2BLL) and their structure is essentially the same as that presented here with an RMSD of 0.2 Å for all Cα's. Superimposition of ArnA with RmlB from *Streptococcus suis* in complex with TDP-Glc and NAD$^+$

(RMSD 2.0 Å) allows a conceptual model of UDP-GlcUA and NAD^+ in the ArnA decarboxyase active site to be made (figure 3.3.6). The superimposition reveals that the ArnA residues S_{433} and E_{434} superimpose with the RmlB residues N_{134} and E_{135}, which are important in the well characterised RmlB reaction (3.1.10)[462,470]. In the RmlB reaction the catalytic tyrosine directly extracts a proton, forming the keto intermediate, before the glutamic acid deprotonates the C'5 position[462,470]. The initial oxidation step in the ArnA reaction is expected to be the same as in RmlB as the catalytic triad is conserved. Mutations remote from the catalytic triad and NAD^+ binding would not be expected to perturb hydride transfer. We therefore expected that mutating S_{433} and E_{434} would not significantly impact hydride transfer but, if these residues were important in the decarboxylation reaction, would effectively cause the reaction to stall here and release NADH and the keto-intermediate. However we observed that these mutants significantly reduced the ability of this enzyme to transfer hydride, without significantly affecting NAD^+ or UDP-GlcUA binding (table 3.2.14). Furthermore, the $S_{433}A$ and $E_{434}A$ mutants still produce the decarboxylated L-Ara4O product and do not accumulate the keto-UDP-GlcUA intermediate (figure 3.2.15).
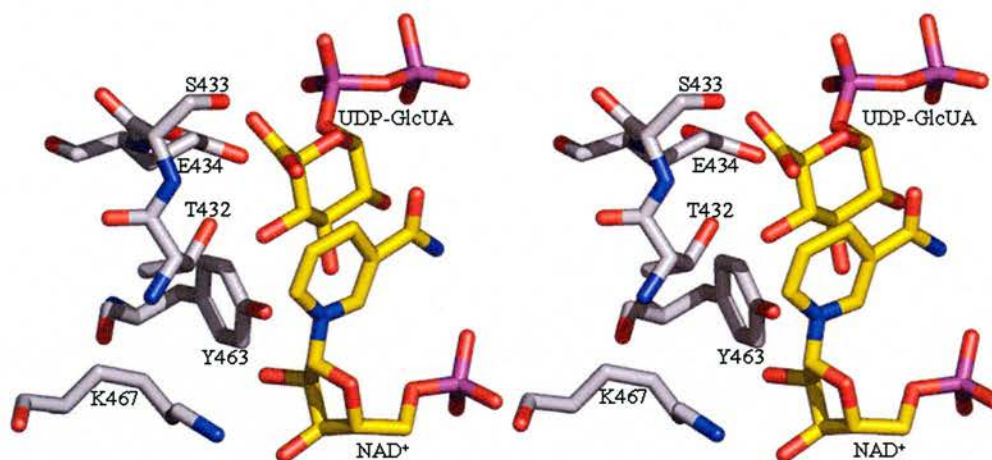
Figure 3.3.6: The active site of the decarboxylase enzyme. We have modeled in the UDP-GlcUA substrate to generate a conceptual model. The model is based on a superposition of the RmlB substrate complex[470]. The catalytic triad residues are shown as are the residues we predict to be important in the decarboxylation step (S_{433} and E_{434}). These residues superimpose with N_{134} and E_{135} of RmlB, which are responsible for catalysis of the dehydration step in RmlB transformation[462,470]. The molecules are shown in stick atoms for NAD^+ and UDP-GlcUA are coloured as figure 3.3.6 and residues are coloured; carbon grey, nitrogen blue and oxygen red.

3.4   DISCUSSION

3.4.1   Formyltransferase domain of ArnA

ArnA is a bifuntional enzyme with two distinct activities, both of which are essential for L-Ara4N modification of Lipid A and polymyxin resistance in Gram-negative bacteria (Chapter 2.1)[376,377]. The interest in this field is high and in September of 2004 a group from Colorado published the structure of the apo-decarboxylase enzyme[496]. Further, whilst our results were in press[381], the same group published two papers; one on the apo-formyltransferase domain with active site mutational analysis[495] and the other with the full length ArnA enzyme, with the decarboxylase domain in complex with UDP-GlcUA and ATP. This study also included mutational analysis of the decarboxylase active site[497], suggesting a different mechanism of decarboxylation. The discussion compares these results with those obtained here.

Figure 3.4.1 shows the sequence alignment between the formyltransferase domain of ArnA, GART and FMT. Highlighted are N_{102}, H_{104} and D_{140} (in ArnA), the three conserved residues thought to be involved in catalysis. Their sequence and structural conservation suggests a similar mechanism of formyl transfer as described for GART by Shim et $al.$[425,434], with the major differences between the proteins being how they bind their respective substrates UDP-L-Ara4N (ArnA), β-GAR (GART) and M-tRNA (FMT).

```
             1                                                    41#
ArnA  ----MKTVVF  AYHDMGCLGI  EALLAAGYEI  SAIFTHTDNP  ---GEKAFYG
FMT   MSESLRIIFA  GTPDFAARHL  DALLSSGHNV  VGVFTQPDRP  AGRGKKLMPS
GART  --MNIVVLIS  GNGSNLQAII  DACKTNKIKG  TVRAVFSNKA  DAFGLERA-R

      44                                           ‡‡‡‡    93
ArnA  SVARLAAERG  IPVYAPDNVN  HPLWVERIAQ  LSPDVIFSFY  YRHLIYDEIL
FMT   PVKVLAEEKG  LPVFQPVSLR  PQENQQLVAE  LQADVMVVVA  YGLILPKAVL
GART  QAGIATHTLI  ASAFDSREAY  DRELIHEIDM  YAPDVVVLAG  FMRILSPAFV

      94        *      *              ‡    #              ‡‡    *143
ArnA  QLAPAGAFNL  HGSLLPKYRG  RAPLNWVLVN  GETETGVTLH  RMVKRADAGA
FMT   EMPRLGCINV  HGSLLPRWRG  AAPIQRSLWA  GDAETGVTIM  QMDVGLDTGD
GART  SHYAGRLLNI  HPSLLPKYPG  LHTHRQALEN  GDEEHGTSVH  FVTDELDGGP

      144                                                    193
ArnA  IVAQLRIALA  PDDIAITLHH  KLCHAARQLL  EQTLPAIKHG  NILEIAQREN
FMT   MLYKLSCPIT  AEDTSGTLYD  KLAELGPQGL  ITTLKQLADG  TAKPEVQDET
GART  VILQAKVPVF  AGDSEDDITA  RVQTQEHAIY  PLVISWFADG  RLK---MHEN

      194      ##           #                             243
ArnA  EATCFGRRTP  DDSFLEWHKP  ASVLHNMVRA  VADPWPGAFS  YVGNQKFTVW
FMT   LVTYAEKLSK  EEARIDWSLS  AAQLERCIRA  F-NPWPMSWL  EIEGQPVKVW
GART  AAWLDGQRLP  PQGYAADE--  ----------  ----------  ----------

      244                                                    291
ArnA  SSRVHPHASK  AQPGSVISVA  P--LLIACGD  GALEIVTGQA  GDGITMQGSQ
FMT   KASVIDTATN  AAPGTILEAN  KQGIQVATGD  GILNLLSLQP  AGKKAMSAQD
GART  ----------  ----------  ----------  ----------  ----------

      292           305
ArnA  LAQTLG--LV  QGSRLN
FMT   LLNSRREWFV  PGNRLV
GART  ----------  ------
```

Figure 3.4.2: Sequence alignment of ArnA formyltransferase domain, FMT and GART, numbering corresponds to full length ArnA. The key conserved catalytic residues are highlighted in red and marked * as in Figure 3.3.3. Conserved residues between all three enzymes are highlighted blue and further residues conserved between ArnA and FMT are highlighted in green. ArnA residues that recognize N-5-fTHF and UMP are marked with ‡ and # respectively.

In ArnA N_{102}, H_{104} and D_{140} are thought to assist in formyl transfer. In this mechanism an important salt bridge formed between H_{104} and D_{140} raises the pKa of H_{104} to around 9.8 so that the imidazole group is protonated. This allows hydrogen bonding of the formyl group of N-10-fTHF to H_{104} as well as to N_{102}, resulting in activation of the carbonyl carbon, allowing direct nucleophilic attack by the primary amine of UDP-L-Ara4N. A tetrahedral intermediate is formed, with hydrogen bonding to H_{104} and N_{102} stabilising the oxyanion. D_{140} then plays a further important role by mediating, along with a water molecule, the transfer of a proton from UDP-L-Ara4N to the N-10 position of the folate. This is followed by the breakdown of the intermediate and release of products. Mutational analysis carried out by Gatzeva-Topolova *et al.* demonstrate that N_{102}, H_{104} and D_{140} play assisting roles in the reaction[495].
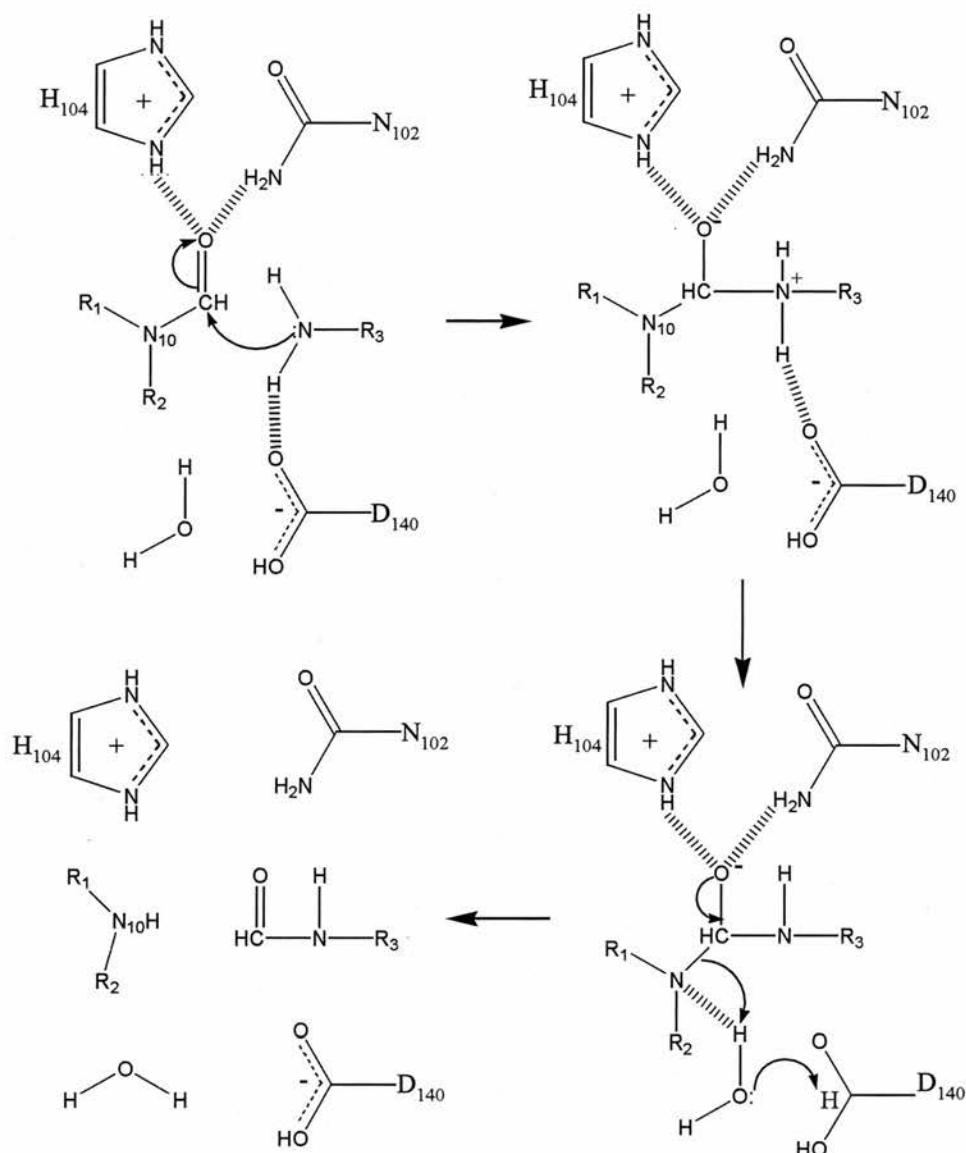
Figure 3.4.1: Plausible mechanism for formyltransfer in ArnA as proposed by Gatzeva-Topolova *et al.*[495]. R_1 is the benzoyl ring and glutamate regions and R_2 the bicyclic ring of N-10-fTHF (figure 3.1.4). R_3 is the carbohydrate and UDP region of UDP-L-Ara4N.

It is interesting that the C-terminal sub-domain is conserved in ArnA, FMT and the hydrolase domain of FDH yet absent in GART (figure 3.3.2). In FMT the role of this small β-barrel domain has been well characterised. It extensively binds to the tRNA molecule, stabilising the interaction between the enzyme and this large substrate and allows the correct positioning of adenosine (3.1.3)[437]. This C-terminal

sub-domain seems to play a similar role in ArnA formyltransferase, interacting with a nucleotide, UMP, presumably to correctly orientate the L-Ara4N moiety in the active site, allowing formyl transfer. These interactions are made by two residues on a loop linking the N- and C-terminus and one residue on the C-terminus proper.

Further comparison of ligand binding in FMT and ArnA also suggests a lesser role for the ArnA C-terminal sub-domain in substrate binding. Figure 3.4.3 shows the superimposition of apo-FMT and tRNA-complexed FMT structures[420,437] and apo- and complexed ArnA formyltransferase structures[381,495]. There is an obvious shift of the C-terminal sub-domain in FMT upon substrate binding, caused by a hinge like motion of the loop connecting the two sub-domains, with an RMSD between these structures of 1.8 Å. This, along with the ordering of a loop in the active site, effectively clamps the tRNA substrate between C- and N-terminal sub-domains[437]. Upon binding of UMP and N-5-fTHF to ArnA there are no such conformational changes, with the apo[495] and complexed[381] structures being largely superimposible, with a RMSD of 0.7 Å (figure 3.4.3). It is possible that binding of the true substrate, UDP-L-Ara4N, may cause the ordering of the active site loop and this might trigger the C-terminal shift, but such a conformational change would seem unnecessary as the binding of UMP and the modeling of UDP-L-Ara4N in the active site shows no reason why such a shift would be necessary (figure 3.3.4). It is interesting that GART only requires the folate binding domain to both bind its substrate, β-GAR, which is much closer in size to UDP-L-Ara4N than tRNA, and catalyse formyl transfer. It might be envisaged that removal of the C-terminal sub-domain in ArnA may not greatly hinder formyl transfer to UDP-L-Ara4N. However unpublished results by Gatzeva Topolova et al. show that such a mutant does lack activity[495]. Therefore it is

clear that the small number of interactions between the C-terminal sub-domain of

ArnA and substrate are important for either substrate recognition or formyl transfer.

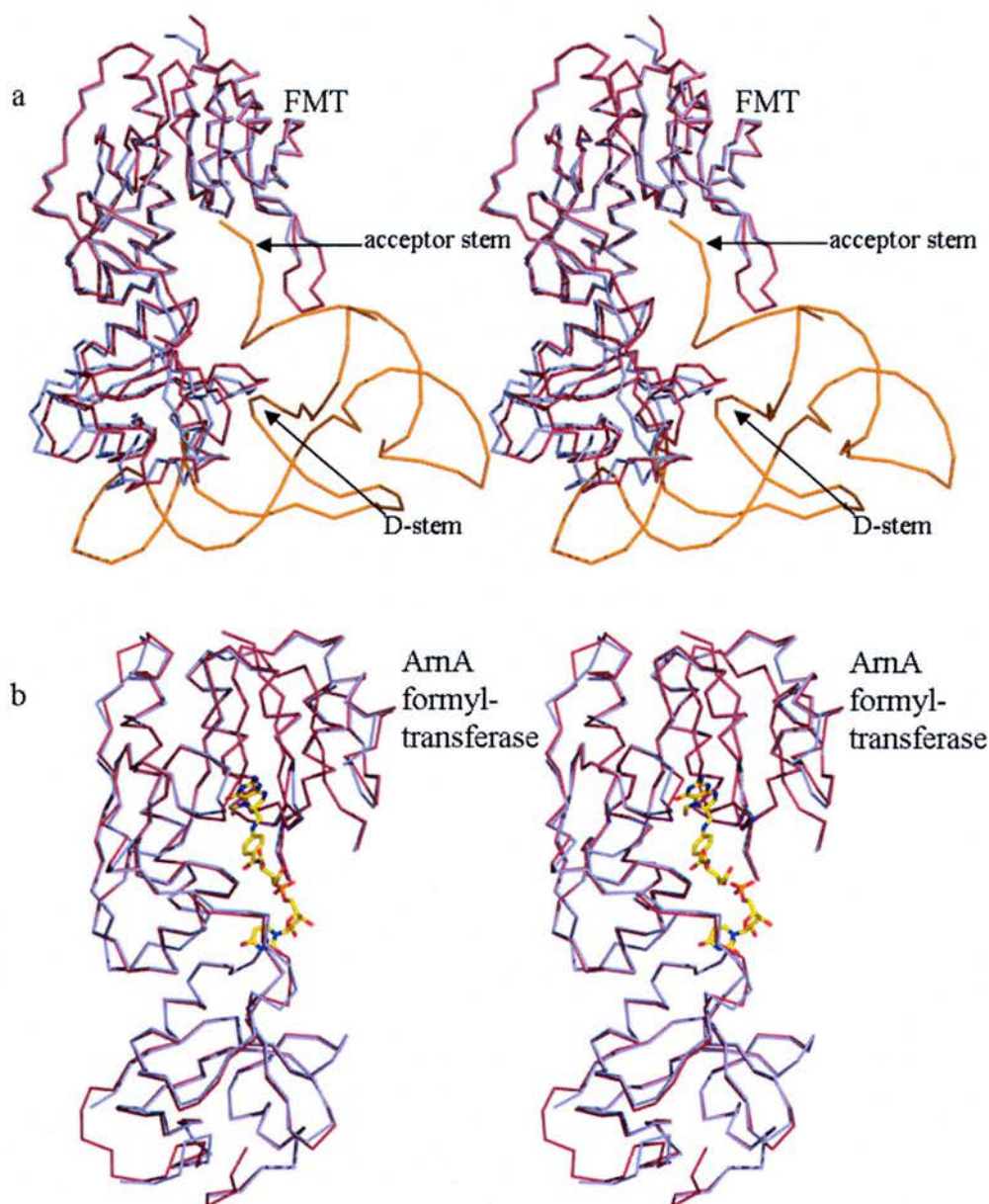***Figure 3.4.3:*** C-terminal sub-domain shift and ordering of active site loop occurring upon binding of tRNA substrate (orange) to FMT (a) but not upon UMP and N-5-fTHF binding to ArnA formyltransferase (b). Apo- and complexed structures are shown in blue and red respectively. The tRNA D-stem and acceptor stem that are part of the D- and acceptor-arms respectively, referred to in the text, are labelled.

FDH is another bifunctional enzyme with N-terminal hydrolase and C-terminal dehydrogenase domains homologous to the formyltransferase domain of ArnA and aldolase dehydrogenase respectively[422]. Importantly the C-terminal sub-domain of the hydrolase domain matches that in ArnA formyltransferase and FMT. This is an interesting case as it catalyses the $NADP^+$ dependent removal of formyl from N-10-fTHF to produce THF[422]. In this reaction it is thought that hydrolysis of N-10-fTHF is carried out by the N-terminal hydrolase domain (figure 3.3.2) and that $NADP^+$ is utilized to oxidise the formyl to CO_2, allowing product release[422]. The FDH hydrolase domain therefore only utilises N-10-fTHF as a substrate but, despite this, it appears the C-terminal sub-domain is essential for activity (unpublished data)[422]. The crystal structure of the hydrolase domain gives no insight into why this domain is important, but it may have some role in the communication of the hydrolase and dehydrogenase domains as both of these are necessary for function *in vivo*[422]. These results suggest that even when not directly involved in substrate binding this C-terminal sub-domain may have important functions. This may therefore be the case in ArnA too. The C-terminal sub-domain is present in the middle of the full length ArnA enzyme and could play a critical role in spatially separating the formyltransferase and decarboxylase domains, allowing them to be active as a bifunctional enzyme. The conservation of the full length protein suggests that despite the formyltransferase and decarboxylase domains being active as discrete enzymes, evolution seems to preferentially select the full length enzyme. There must be a reason for this and we hypothesised that some extent of substrate channelling may occur from the decarboxylase domain of ArnA to ArnB and back to the ArnA formyltransferase domain. Although gel filtration experiments (2.2.4) failed to detect the necessary complex formation between ArnA and ArnB, the limitations of this

technique do not rule out the possibility of substrate channelling (2.3). The full length structure of ArnA (figure 3.4.7) is very interesting as it forms a large hexameric structure. This structure provides few insights into why full length ArnA is conserved as no substantial interactions are provided by the formyltransferase domain in the oligomerisation of the enzyme[497]. It is tempting to speculate that, if substrate channelling provided the selection pressure for full length ArnA (2.4), the unusual structure of ArnA may form a platform for a larger multi-enzyme complex, which would include ArnB.

3.4.2 Decarboxylase domain of ArnA

The biochemical data presented in 3.2 and 3.3 allows us to draw important conclusions about the ArnA decarboxylation reaction. First, decarboxylation of 4-keto-UDP-GlcUA would appear not to occur spontaneously, but is indeed enzyme catalysed. If the reaction was spontaneous then one would not expect mutations that do not perturb either NAD^+ or UDP-GlcUA binding, are not conserved in SDR enzymes, or implicated in hydride transfer, to alter the reaction rate. Secondly, if these mutants are perturbing decarboxylation one would perhaps expect accumulation of the keto-intermediate. That this is not observed (figure 4.2.15) provides new insights into the thermodynamics of hydride transfer.

The inactivity of $E_{434}Q$ and reduced activity of $S_{433}A$ and $E_{434}A$ mutants (40x and 100x reduction from wild-type activity respectively) (table 4.2.14), suggests that the NAD^+ dependent oxidation of UDP-GlcUA has not been decoupled from decarboxylation leading to UDP-L-Ara4O product. It was postulated by Gaetzeva-Topolova *et al.* that the E_{434} residue acted as a steric hindrance to prevent rotation of

139

the substrate and keto-intermediate in the active site, in a mechanism similar to GalE[496]. Implicit in this mechanism is that the rates of oxidation and decarboxylation are significantly slower than rotation of the carbohydrate in the active site. The reduced activity of the $E_{434}A$ mutant would, at least in part, be explicable by such a mechanism. However the isosteric $E_{434}Q$ mutant would be able to block rotation of the carbohydrate in the active site just as effectively as the wild-type enzyme. That this mutant is inactive suggests that E_{434} plays a different role in the deacrboxylation step than merely blocking rotation of the glucuronic moiety. It is possible that E_{434} could act as a base, ensuring deprotonation of the carboxylic acid, thus promoting decarboxylation on formation of the 4'-keto intermediate. The residual activity of the $E_{434}A$ mutant may be accounted for by water molecules filling the created void to provide a bridge between the carboxylate of UDP-GlcUA and a second unidentified base in ArnA. In $E_{434}Q$ there would be no room for a water molecule, therefore there would be no bridge to a second base, thus decarboxylation could be inhibited. We suggest that S_{433}, by hydrogen bonding to the carboxylate group of the GlcUA, may aid decarboxylation (figure 3.4.4).
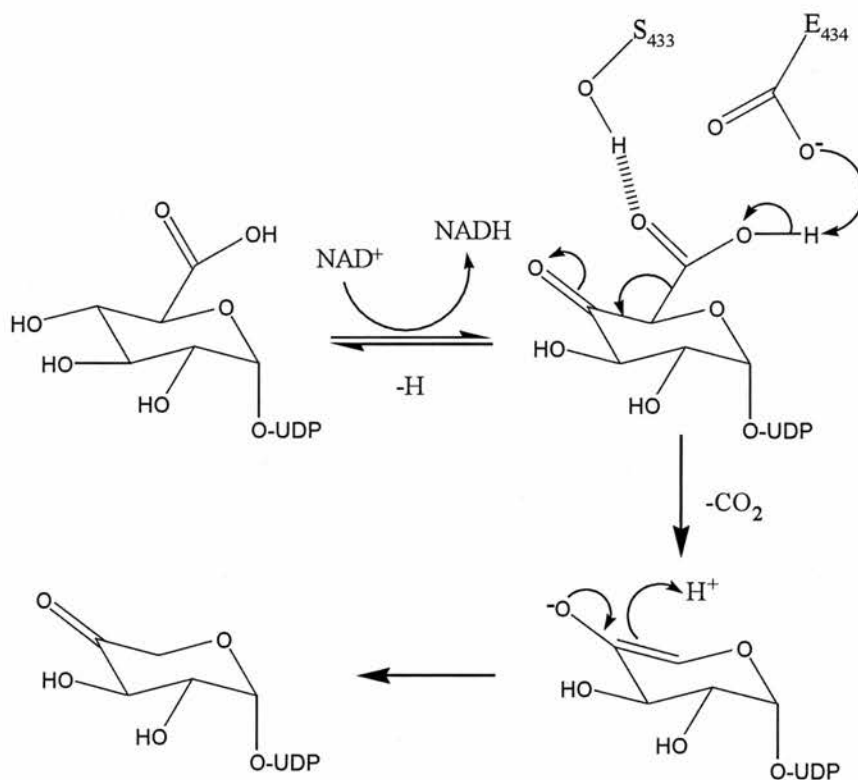
Figure 3.4.4: Potential mechanisms of the ArnA decarboxylation reaction using E_{345} as a base, proposed from our kinetic study[381].

Gatzeva-Topolova *et al.* originally implicated the ArnA residues S_{433} and R_{619} in the decarboxylation reaction based on their superimposition of the ArnA decarboxylase domain with *E. coli* GalE and because R_{619} is conserved in other UDP-GlcUA decarboxylases (UDP-xylose synthases)[496]. In their more recent paper detailing the full length ArnA structure they test their hypothesis by making several mutations of these residues[497]. Encouragingly the results they obtain for the $S_{433}A$ mutant are very similar to ours in that there is a 40-fold reduction in activity. Interestingly they find that mutation of the R_{619} residue also has a significant impact on catalysis, reducing the overall activity around 400 fold compared to wild-type, strongly implicating this residue in the ArnA decarboxylase reaction[497]. However, although they use a large excess of substrate (6 mM), they do not calculate K_m values

for NAD^+ or UDP-GlcUA binding[497]. These results led to their proposed decarboxylase mechanism presented in figure 3.4.5, involving the deprotonation of R_{619}. Although the pKa of the enolate would be higher than that of R_{619} (pKa ~12), deprotonation of arginine in enzyme mechanisms is rare.
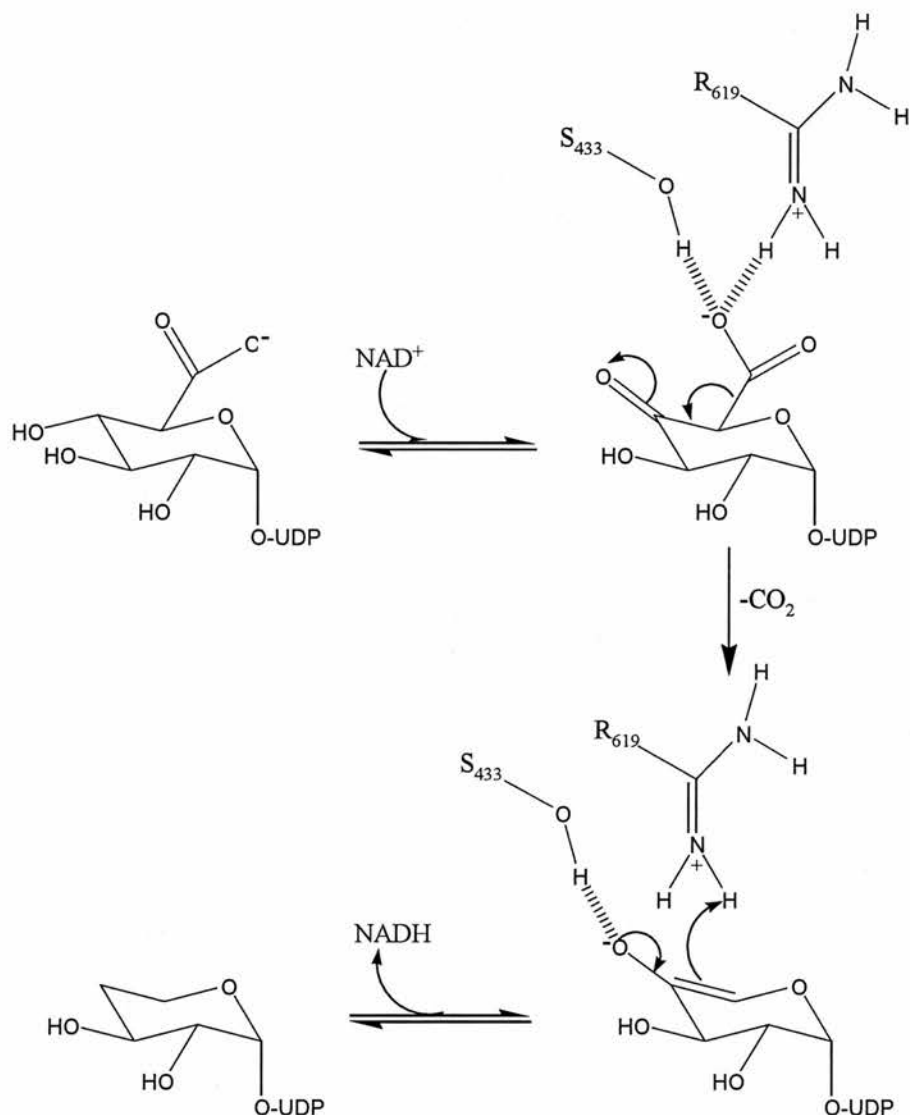

Figure 3.4.5: Alternative mechanism of ArnA decarboxylation using R_{619} as an acid. Proposed by Gatzeva-Topolova *et al.*[497].

Along with their mutational analysis Gatzeva-Topolova *et al.* presented the full length structure of ArnA, with the decarboxylase domain in complex with adenine triphosphate (ATP) and the UDP-GlcUA substrate. This provides insights into the

decarboxylase active site, which further confuses the mechanistic picture. In the structure R_{619} is around 3.5 Å from C5' (figure 3.4.6), putting it in a position where it may play a direct role in the decarboxylation reaction. However E_{434} is 6.5 Å from the carboxylic acid (figure 3.4.6), suggesting a more indirect role for this residue, which is not consistent with our $E_{434}Q$ mutant. In their full length structure of ArnA the active site Y is 4.8 Å from the C4' hydroxyl group. Therefore Gaetzeva-Topolova *et al.* hypothesise the proton shuttle mechanism of hydride extraction[468] through T_{432}[497]. However, structural studies of Human GalE[463], SQD1[471] and RmlB[462,470] have shown direct hydride extraction by the catalytic Y, consistent with mutagenesis studies (3.1.8)[460,464]. It is now generally accepted that earlier crystallographic results suggesting C4' deprotonation may occur through a shuttle mechanism[466,467] were an artifact and are inconsistent with mutagenesis[460]. The positioning of Y in the ArnA active site is therefore likely to be an artifact and thus the picture of the decarboxylase active site presented by Gatzeva-Topolova *et al.*[497] may not accurately represent the true positioning of catalytic residues during the reaction. It is therefore possible that hat although E_{434} is over 6 Å from the carboxylic acid in the full length ArnA crystal structure, this may not reflect its position during the decarboxylase reaction. One reason why these distances are observed may be because ATP, a relatively poor analogue of NAD^+ (figure 3.4.7) is bound in the crystal structure. It is possible that the binding of the nicotinamide moiety of NAD^+ in the active site may rearrange the positioning of these residues and such a co-complex would be desirable to get a more accurate representation of the ArnA active site.
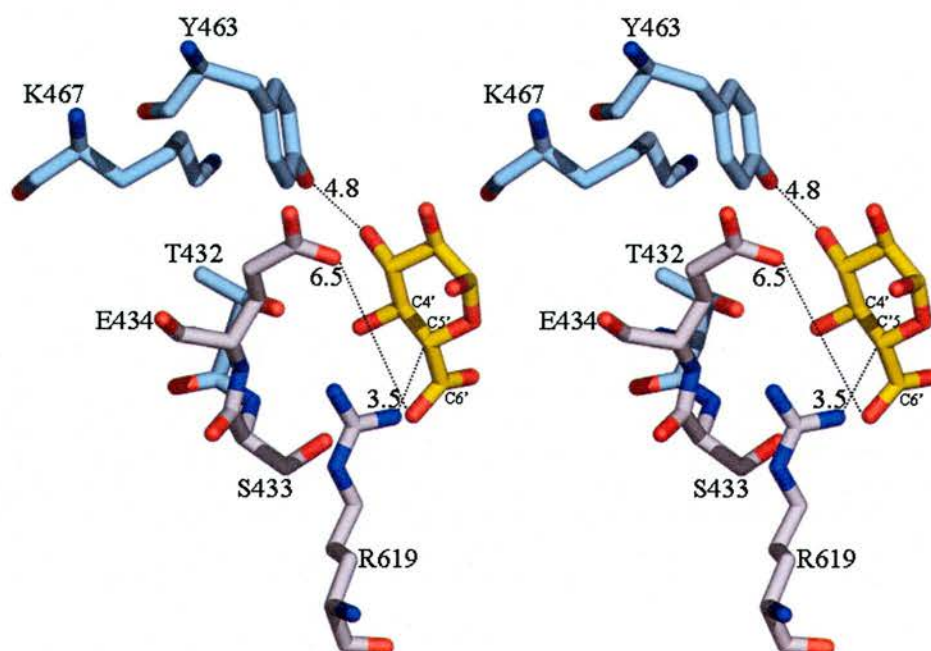
Figure 3.4.6: Important active site residues of the decarboxylase domain of ArnA, adapted from the full length structure of ArnA with the decarboxylase domain in complex with UDP-GlcUA and ATP[497]. Residues are labelled with nitrogen blue, oxygen red and carbon either yellow, light blue or grey for the UDP-GlcUA moiety, the conserved catalytic triad or residues proposed to be important for the decarboxylation respectively. Distances shown by dotted line are in Å.
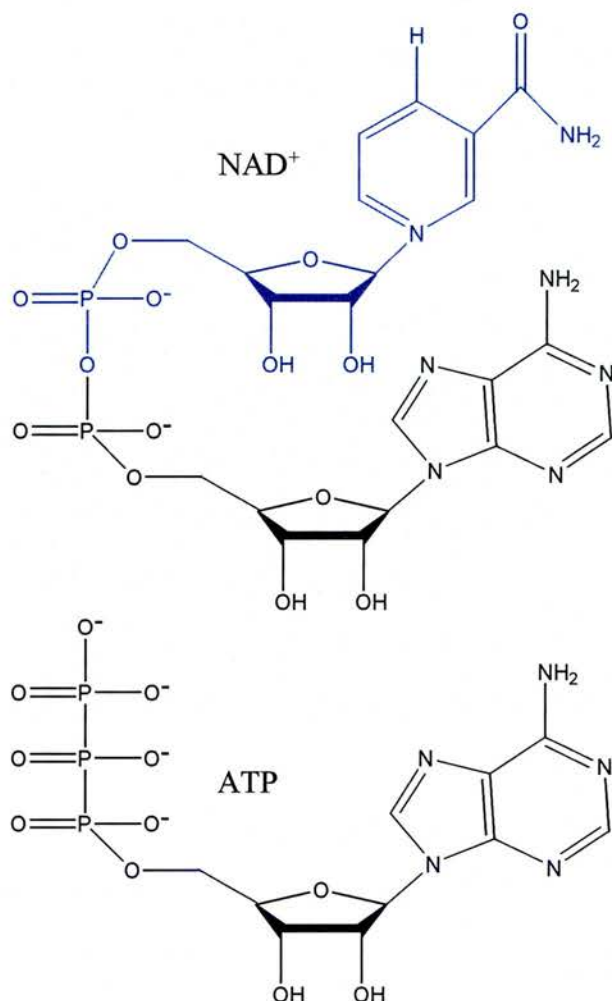
Figure 3.4.7: Chemical structures of ATP (a) and NAD$^+$ (b).

The second conclusion from our results, which is consistent with the data published by Gatzeva-Topolova *et al.*, is more subtle. If S_{433} and E_{434} are involved in decarboxylation it would be expected that the reactions catalysed by these mutants should transfer a hydride in the initial oxidation step. This would result in the build up of NADH as the 4-keto-UDP-GlcUA intermediate was produced. The same would be true for the R_{619} mutant of Gatzeva-Topolova[497]. Like us the assay utilised by Gatzeva-Topolova *et al.* monitors NADH production, not product formation. Thus all of the R_{619}, E_{434} and S_{433} mutants, which are presumed only to affect the second step of the reaction, all affect the initial oxidation. We show that there is no accumulation

145

of keto-intermediate for any of the mutants we made, suggesting a very tight coupling of the oxidation and decarboxylation reactions. The question is why, in the stepwise reaction of oxidative decarboxylation, does decarboxylation affect the preceding step of hydride transfer? The answer must be that, since NADH production does not occur, the initial hydride transfer has an unfavourable free energy. It seems that the equilibrium for the oxidative step lies far to the substrate UDP-GlcUA than the intermediate UDP-4-keto-GlcUA. It must therefore be the decarboxylation step, which is effectively irreversible, that drives the reaction and serves to pull substrate to product. When this irreversible step is abolished or significantly slowed, the hydride remains principally on the substrate, consistent with a slower rate of NADH production. Examining mechanisms of other carbohydrate utilising SDR's suggests that this may be a common theme. In RmlB an irreversible step, the elimination of water, occurs after hydride transfer from substrate[470]. Prior to this elimination the reaction is reversible, with both substrate and keto sugar intermediate shown to exist[498,499]. GalE oxidses and reduces its substrates without ever releasing the keto-sugar intermediate[466,467]. Thus in all three cases the equilibrium between oxidised substrate and NAD^+ is finely balanced. High resolution structure analysis and electronic structure calculations have shown that SDR enzymes fine tune the redox potential of NAD(P)H by distortion of the nicotinamide ring[480]. ArnA, RmlB and GalE all operate on quite similar carbohydrates, GlcUA, glucose and galactose respectively, and all three have similar active site volumes. Thus, by poising the equilibrium of hydride transfer towards substrate such that only a second irreversible step will create flux through the pathway, a subtle substrate specificity can be created by the second irreversible step. The structure of ArnA does not reveal any reason why UDP-Glc should not bind; it is slightly smaller than UDP-GlcUA. The presence

of E_{434} and S_{433} in ArnA would prevent catalysis of the dehydration of UDP-Glc, since the proton source required for water elimination is absent. By ensuring in ArnA, that the equilibrium for the first oxidation reaction favours substrate, accumulation of reactive and undesired 4′-keto sugars closely related in structure to the true substrate are prevented. This may well be a general feature of these very similar SDR enzymes.

3.4.3   Apo vs complexed ArnA decarboxylase
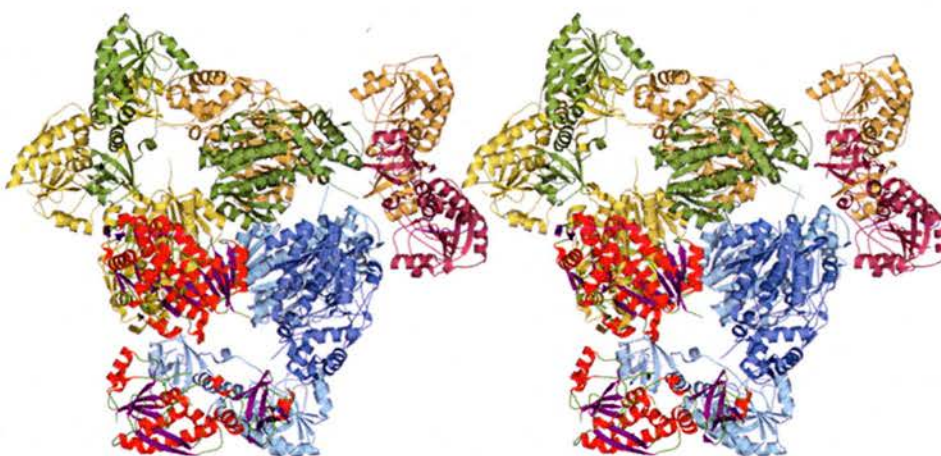
Figure 3.4.7: Overall structure of the full length ArnA hexamer. One monomer is labeled with red α-helices and purple β-sheets, another has the decarboxylase domain coloured in slate and formyltransferase domain coloured in raspberry. Cyan, orange, yellow and green are used to individually colour the both the decarboxylase and formyltransferase domains of the other 4 monomers. Adapted from Gaetzeva-Topolova *et al.*[497].

As the full length ArnA structure published by Gatzeva *et al.* is a co-complex with ATP and UDP-GlcUA they compare the complexed and apo decarboxylase structures[497].

Full length ArnA crystallises as a hexamer and this multimeric form of ArnA is confirmed by gel filtration and multiangle light scattering analysis[497]. The structure is effectively a dimer of trimers and the hexamer adopts the shape of a three-blade propeller with dimensions of 162 x 87 Å. The core of the structure is made up of the C-terminal decarboxylase domains, with the N-terminal formyltransferase domains sticking out on the periphery. The majority of the interactions between the individual components are contributed by the decarboxylase domain which, from observation of B-factors, is more rigid than the formyltransferase domains[497]. The major contacts between the subunits of each trimer are mediated by hydrogen bonding and electrostatic interactions; these include salt bridging between K_{381} and K_{382} with D_{368} and E_{366} in adjacent protomers and hydrogen bonding between E_{374} with S_{373} and E_{374}. The dimerisation of each trimer occurs across a hydrophobic surface made up by residues 448-456 of adjacent trimers and results in the formation of a two stranded antiparallel β-sheet. A 4-helix bundle also forms between neighbouring protomers, formed by residues 401-423 and 460-480. A few interactions are also contributed by the formyltransferase domain, but these are relatively minimal[497].

Comparison of the apo and complexed ArnA decarboxylase structures reveals several key differences. Most strikingly a loop, residues 500-509, that appears to block the NAD^+ binding pocket in the apo structure, has undergone a large conformational change, with I_{505}, in the middle of the loop, being displaced by 17.5 Å. Interestingly in the complexed form of the enzyme this loop takes on the form of a short α-helix positioned over UDP-GlcUA, trapping the substrate in its binding pocket. This suggests an ordered sequential mechanism of substrate binding, with UDP-GlcUA binding first and NAD^+ subsequently binding. That the ArnA reaction does not reutilise the NADH formed in a subsequent reduction step is unusual for

such SDR enzymes. Gatzeva-Topolova *et al.* suggest that after the initial hydride transfer NADH may dissociate, with the helix trapping UDP-GlcUA keeping it bound to the enzyme, explaining why NADH is not utilised to reduce the ketone product. However such a mechanism would perhaps suggest the decoupling of the oxidation and decarboxylation steps, inconsistent with the biochemical data. In RmlB[462,470] and GalE[467] the further reduction by NADH relies on the rotation of the substrate carbohydrate moiety in the active site. It may be that in ArnA such a rotation cannot occur and this may be the reason why NADH is not further utilised.

Additional conformational changes also contribute to substrate binding. A loop, residues 605-616, becomes ordered upon UDP-GlcUA and ATP binding, making several contacts to the uracil group of UDP-GlcUA[497]. There is also a small hinge-like motion of the overall enzyme between the substrate and NAD^+ binding functional sub-domains of the decarboxylase enzyme. This causes the enzyme to close in slightly around both UDP-GlcUA and ATP[497].

3.5   Future work

This study has provided insights into the decarboxylase reaction catalysed by ArnA, however the results of Gatzeva-Topolova *et al.* leave unanswered questions. Further structural and mutational studies are required to determine whether the mechanism of decarboxylation catalysed by ArnA involves E_{434} or R_{619}. The catalytic tyrosine is ~5 Å from the C4' position and therefore unable to directly extract a proton. This suggests that the active site presented by Gatzeva-Topolova *et al.* is different to that formed during the oxidative decarboxylation reaction of ArnA. It is therefore important to get a co-complex of ArnA decarboxylase with both UDP-L-Ara4N and

NAD$^+$ bound, to give a more accurate representation of the active site. Mutational analysis of the catalytic triad, particularly the tyrosine and threonine, will also be important to confirm that the ArnA reaction proceeds via direct proton extraction by tyrosine.

As the functions of both the decarboxylase and formyltransferase domains are essential for the resistance of several pathogens to cationic antimicrobial peptides, their structural solution provides two templates for structure-based drug design. The robust assays that exist for both the decarboxylase (presented here) and formyltransfer reactions[376,495], will be important for validation of any potential inhibitors.

Chapter 4

Structural studies of heterotrimeric PCNA from *Sulfolobus solfataricus*:

Structural solution of a heterotrimeric sliding clamp and
initial crystallisation of this clamp in complex with the
Helix-hairpin-helix domain of XPF

Chapter 4

4.1   Introduction

4.1.1   Sliding Clamps
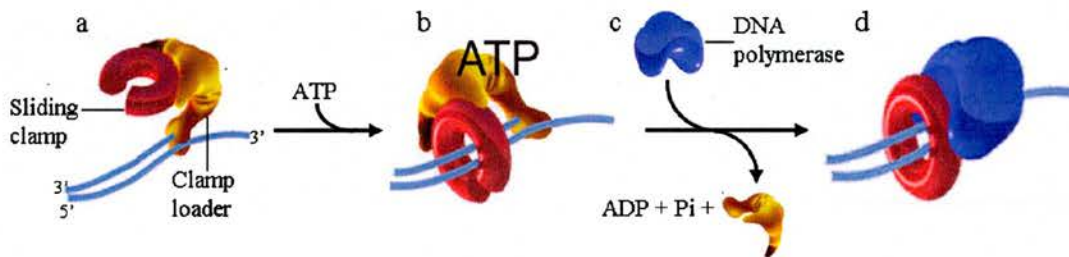
Figure 4.1.1: Initiation of DNA replication. (a) The clamp loader recognises the ds/ssDNA junction and (b) uses ATP to load the sliding clamp. (c) The sliding clamp recruits the replicative DNA polymerase and processive replication can occur. Taken from Bruck and O'Donnell[500]

The replicative DNA polymerase III (PolIII) holoenzyme from *E. coli* is capable of non-disociative replication of long stretches of DNA at an astonishing rate of over 750 bases per second[501-504]. A similar processivity is found for the archaeal and eukaryotic replicative polymerase holoenzymes, PolB[504] and Polδ or ε (Polδ or ε)[505] respectively. This processivity is achieved due to the activity of two associated proteins that are part of the holoenzyme, the sliding clamp and clamp loader[506-508]. The clamp loader utilises ATP to load the sliding clamp onto a primed section of DNA[508] and the clamp interacts with the core DNA polymerase tethering it to DNA, resulting in processive replication[504] that continues after the removal of the clamp loader (figure 4.1.1)[509]. In prokaryotes the β-subunit of the PolIII holoenzyme acts as the sliding clamp[506,507,509] and it is the proliferating cell nuclear antigen (PCNA)[510] in

archaea[505] and eukarya[83,511-513]. PCNA was discovered over 25 years ago[510,514] and its name derives from its essential role in cellular proliferation[510].

In *E. coli* the β-subunit clamp is also essential for lagging strand synthesis, where PolIII is passed from one β-subunit to the next at newly primed Okazaki fragments[506,515]. Biochemical studies show that both the β-subunit[509] and PCNA[516] can be cross-linked to circular duplex DNA *in vitro* but, upon linearisation, this association is not seen[509,516]. The reconstitution of DNA replication *in vitro* also shows that PCNA can initiate processive replication without the aid of a clamp loader on a linearised template[516]. These studies suggested a mechanism for the sliding clamp in which it non-specifically encircles DNA and can freely slide in either direction along the duplex[509]. The structure of the β-subunit was solved in 1992[517] and, consistent with the biochemical data[509,516], it forms a 'donut' or toroidal structure, with a central hole big enough to accommodate the B and A forms of DNA[517] that are found during replication[518]. The β-subunit forms a head to tail dimer with each monomer consisting of three globular domains. These domains are structurally almost identical, despite low sequence identity of less than 15 %[517]. Each domain consists of two anti-parallel α-helices flanked by two, four stranded β-sheets, with the 12 α-helices and 12 β-sheets in the dimer lining the inner and outer surface of the toroidal structure respectively[517]. The β-subunit has an overall negative charge. However, more locally, the surface that lines the central cavity is electrostatically positive suggesting, consistent with biochemical studies[509,516,517], that DNA passes through the centre of the sliding clamp. Further structural features also suggest this method of DNA binding[517] and it is thought that the β-subunit only interacts with DNA through non-specific interactions, mediated mainly by water molecules[517]. Several electrostatic, as well as hydrophobic and hydrogen bond, interactions

contribute to the dimer interface and results in the formation of a very stable β-subunit[517]. It is this stability of the dimer that results in the requirement of an ATP dependent accessory protein, the five subunit γ-complex to load the β-subunit onto DNA[519]. Structural studies suggest that the γ-complex acts as a molecular wrench, forcing the dimer apart at one of the interfaces, allowing β-clamp loading to DNA[519].

Eukarya and archaea have evolved the use of a trimeric sliding clamp in replication[506,520], with each PCNA monomer approximately two thirds of the molecular weight of the β-subunit monomers (~28 as opposed to ~40 kDa)[509]. These observations led to the prediction that the PCNA trimer also consists of 6 structural domains that, despite the lack of sequence conservation, are structurally similar to those of the β-subunit[517]. The structural solution of the PCNA sliding clamp in *S. cerevisiaea*[521], humans[522] and archaea[523] as well as the bacteriophage T4 sliding clamp gene45 protein[524] showed this to be the case, with a remarkable conservation of topology between sliding clamps conserved right the way through evolution. This structural conservation is depicted in the superimposition of the β-subunit and *S. cerivisiae* PCNA (figure 4.1.2). The structural features suggesting non-specific encircling of DNA for the β-subunit are conserved in PCNA. The major differences between the structures of these two sliding clamps are the subunit interfaces, with PCNA utilising more hydrophobic, rather than electrostatic, interactions and extra loop regions that are present in PCNA. It is these loop regions that have been shown both biochemically (see table 4.1.2) and structurally[522,525] to be important in mediating protein-protein interactions. Indeed eukaryotic PCNA is found to have many more interactions with other cellular proteins than the β-subunit sliding clamp, which is exclusively involved in replication. This suggests PCNA has diverse cellular roles, with some of the known partners shown in table 4.1.2..
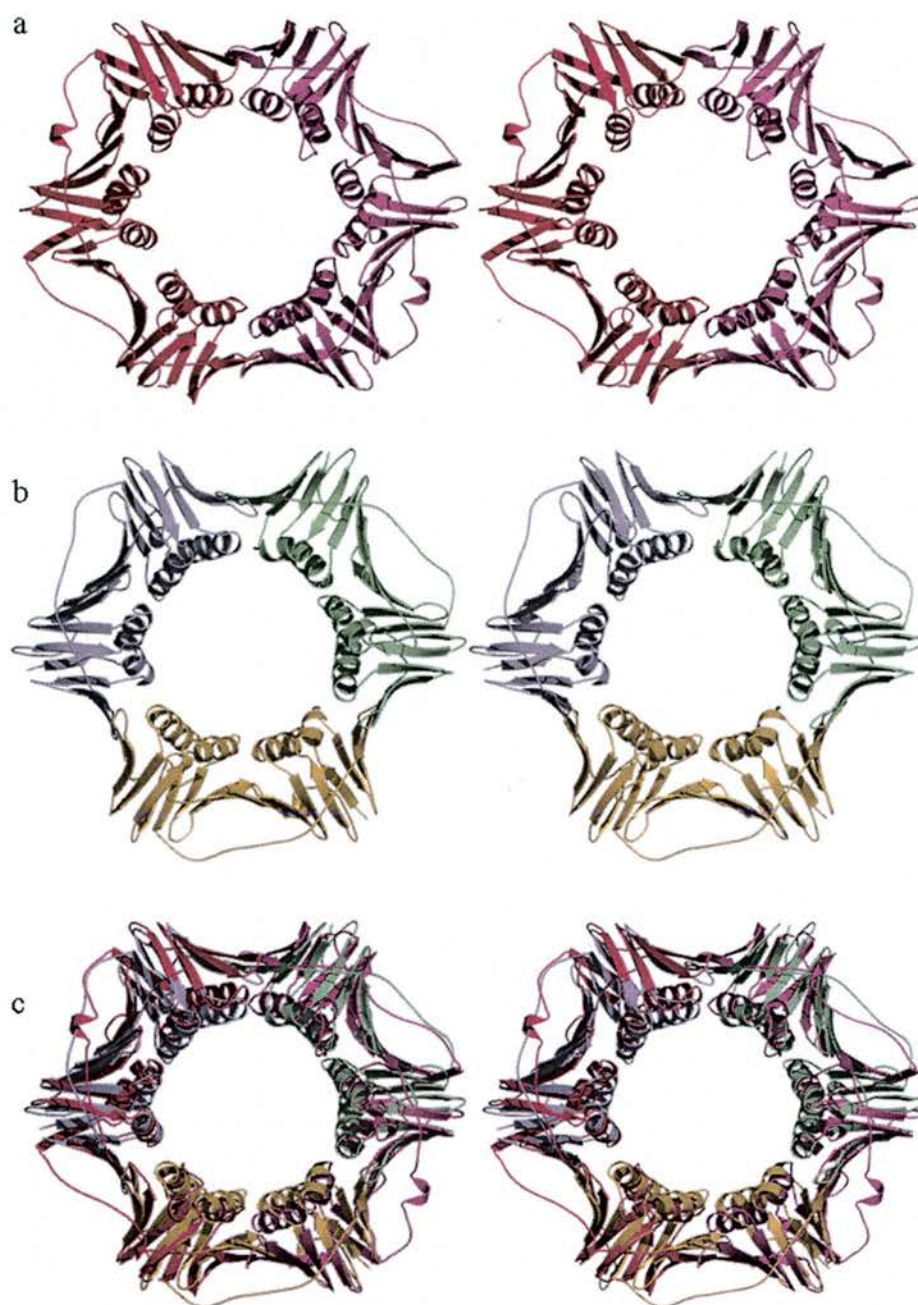
Figure 4.1.2: Structural superimposition (c) of the dimeric β-subunit[517] from *E. coli* (PDB code: 2POL)[517] (a) and trimeric PCNA from *S. cerivisiaea* (PDB code: 1PLQ)[521] (b). Monomers are coloured differently. Despite the lack of sequence conservation and different oligomeric states the two structures superimpose remarkably well. Superimpositions were carried out in COOT[490].

4.1.2 PCNA interactions:

Protein	Effect of PCNA interaction	Interacting region		Reference
		PCNA	Partner	
DNA replication				
Elongation				
Polδ	Stimulation of polymerase	C-ter and IDL*	Multiple subunits	526,527
Polε	Stimulation of polymerase	C-ter and IDL	Multiple subunits	527,528
RFC	Loading of PCNA to DNA	C- and N-term	Multiple subunits	529
RPA		IDL	N-ter	530
Lagging strand synthesis				
DNA ligase I	Stimulation of ligase	IDL	N-ter	531,532
FEN1	Stimulation of FEN1	IDL	C-ter	525,533
Post-replication events				
Chromatin assembly factor-1 (CAF1)	Promotes chromatin assembly	IDL	N-ter	534,535
DNA cytosine-methyltransferase (MeCTr)	Stimulates DNMT activity	C-ter and IDL	N-ter	536,537
Chromosomal transmission fidelity 18	PCNA loading to DNA		Not known (NK)	538,539
DNA repair and recombination				
Apurinic/apyrimidinic endonuclease 1 and 2 (APN1, APN2)	Stimulates APE1 and 2 activities involved in BER	C-ter and IDL	C-ter	540,541
Repair DNA polymerases Polβ, λ, η, κ, ι	Stimulation of polymerase in translesion synthesis	N-terminal and IDL	C-ter	237,542-545
Growth arrest and DNA damage (Gadd45)	Hinders growth arrest and apoptosis	IDL	NK	546
MutL homologue 1 (MLH1)	Role in mismatch repair	IDL	NK	547
MutS homologue 2, 3 and 6 (MSH2, 3 and 6)	Role in MMR	IDL	N-ter	548,549
Radiation sensitive mutant 5 and 18 (RAD5 and 18)	Ubiquitination of PCNA:DNA repair role		NK	211
Ubiquitin conjugating enzyme 9 (UBC9)	Ubiquitination/SUMOylaton of PCNA: DNA repair role	IDL	NK	211
Uracil DNA glycosylase (UDG)	Role in BER		N-ter	550
Xeroderma pigmentosum G (XPG)	Essential for NER	IDL	C-ter	151
Regulatory proteins				
P21	Cell cycle control: inhibition of PCNA dependent DNA replication	IDL	N-ter	522,551
P300	Facilitation of PCNA function in DNA repair	C-ter	NK	552
Cyclin D	Cell cycle control: inhibits cell growth	N- and C-ter	NK	553,554
Cyclin dependent kinases (CDK))	Cell cycle control: stimulates phosphorylation of replication proteins	C-ter	NK	554,555

IDL* represents the interdomain linker region of PCNA

Table 4.1.3 (above): Some of the diverse range of proteins that have been shown to interact with PCNA from eukaryotes, adapted from Vivona and Kelman[556] and Maga and Hübscher[557].

The range of proteins shown to interact with PCNA in eukaryotes link PCNA into almost all forms of DNA maintenance including; replication[83,512,513,526], repair[151,211,542], the DNA damage response[522,551], chromatin remodelling[534-537], as well as apoptosis[546,558] and cell cycle progression[551,554,559].

PCNA is essential for leading and lagging strand synthesis[560] and the addition of antisense PCNA oligonucleotides stops both DNA synthesis and cellular progression[561,562]. In addition to its role in tethering the replicative polymerase to DNA, PCNA is a central coordinator for the processing of Okazaki fragments on the lagging strand[563,564]. In this process PCNA coordinates the sequential actions of Polδ, flap endonuclease 1 (FEN1) and DNA ligase I (figure 4.1.4)[563-565].
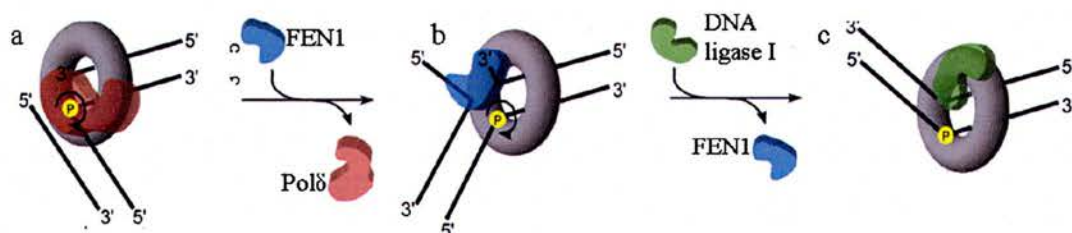
Figure 4.1.4: PCNA's role in coordinating the sequential actions of Polδ, FEN1 and DNA ligase I during Okazaki fragment maturation in eukaryotes. In the rotary handoff mechanism PCNA coordinates the switch from (a) Polδ to (b) FEN1 and finally (c) DNA ligase I. Adapted from Chapados *et al.*[566].

During replication DNA needs to be relaxed and then re-compacted following passage of the replication fork, a process carried out by the chromatin remodelling machinery (reviewed in[567]). PCNA is implicated in the re-compaction

step due to its interactions with several proteins involved in chromatin remodelling and inheritance including DNA cytosine 5-methyltransferase (MCMT)[536] and chromatin assembly factor 1 (CAF1)[534,535]. MCMT specifically recognises hemimethylated DNA and is essential for the inheritance of the methylation pattern of DNA[568,569]. CAF1 is a molecular chaperone that loads H3 and H4 histones onto newly synthesised DNA, marking the new strand for chromatin assembly[535]. Some PCNA molecules are left behind on newly synthesised DNA and it is thought that these act as a topological marker, allowing methylation inheritance and chromatin assembly (figure 4.1.5)[537,570]. In addition to its role in mediating chromatin remodelling following DNA replication, PCNA is also thought to remodel chromatin during DNA repair, allowing access of the repair machinery to the site of damage. In this role the interaction of PCNA with p300, part of chromatin remodelling machinery, is thought to be important[552,557].

PCNA is an important accessory protein for many DNA repair enzymes and is involved in almost all forms of repair (reviewed in Maga and Hubscher[557]). Its major role is in mediating the DNA resynthesis that is required for BER[542], NER[123], MMR[548,549,570,571], double strand break repair[572] and replication repair involving lesion bypass[237,543,544] (1.3). In addition to mediating resynthesis, PCNA is also thought to be involved in early events of DNA repair. For example in NER PCNA has been shown to co-localise on DNA with XPA *in vivo*, implicating PCNA in early NER events[573,574]. PCNA also forms a complex with the XPG endonuclease enzyme[151], an interaction that is unlikely to result in resynthesis but may important for NER efficiency[151,575]. Also during MMR PCNA has been shown to interact with and increase the mismatch binding specificity of the MSH2, MSH3 and MSH6 proteins[548,571]. Further, a PCNA-MSH2-MSH6 complex is found to be stable on

newly synthesised DNA, with the MSH2-MSH6 complex actively transferred to mismatched bases[570]. This further suggests PCNA can act as a topological marker for DNA, as thought to be important for MCMT and CAF1 interactions, allowing newly synthesised DNA to be scanned for mismatches (figure 4.1.5).

PCNA is also involved in cell-cycle control through interactions with p21[522,551], cyclinD[553,554] and cyclin dependent kinases (CDKs)[576]. The PCNA CDK2-cyclinA complex bridges to other proteins which can then be phosphorylated[576]. This may be an essential mechanism for the control of the replication fork.
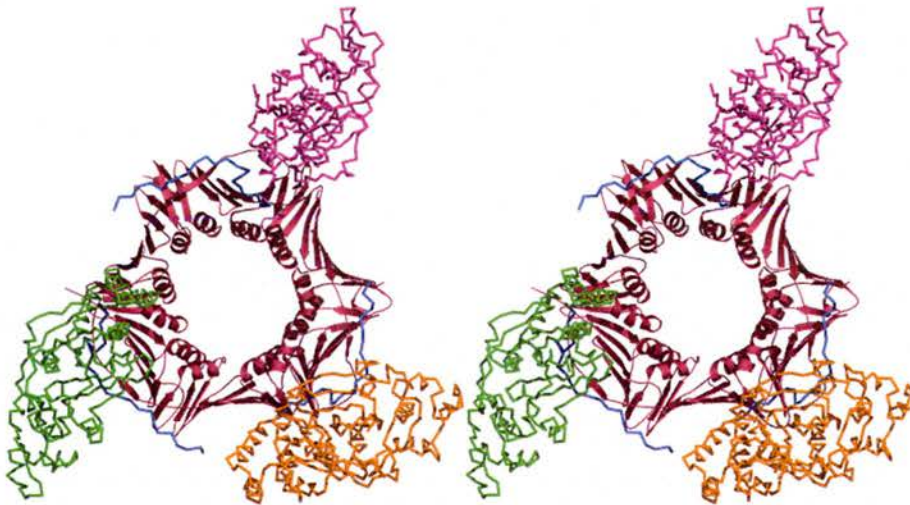
Figure 4.1.5: Stereo view of the human PCNA-FEN1 complex, adapted from Sakuri *et al.* (PDB code: 1UL1)[525]. PCNA is coloured in red with the FEN1 subunits coloured magenta, green and orange. The C-terminal of each FEN1 subunit is coloured in blue and this is the PIP motif that superimposes well with the p21 PIP peptide solved in complex with human PCNA[522], which previously provided the basis for modelling PCNA interactions.
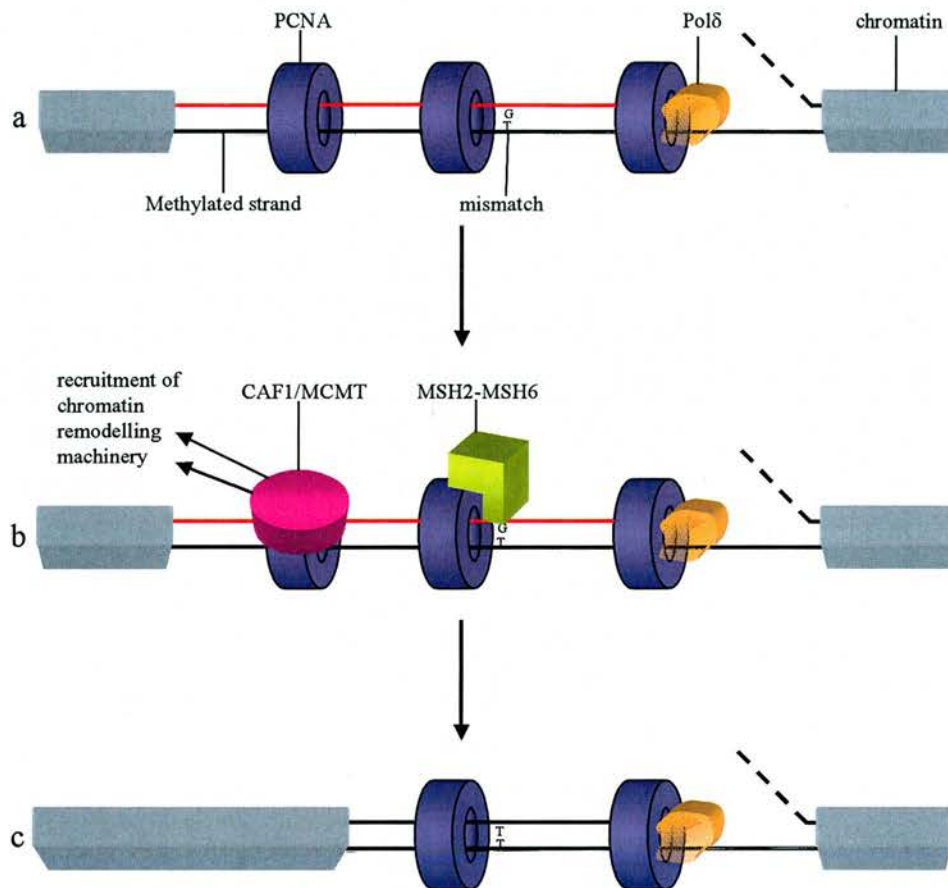
Figure 4.1.5: Schematic of PCNA's role in post-replication events, mediating chromatin remodelling and mismatch repair. For simplicity only lagging strand synthesis is shown (dotted line represents the lagging strand). (a) PCNA is left on DNA following replication (new, unmethylated strand coloured red). (b) This PCNA recruits mismatch repair proteins and proteins involved in chromatin inheritance. (c) mis-matches are repaired and DNA packaged back into its chromatin state.

4.1.3   PCNA control

An important question is how are the vast numbers of interactions that can occur with PCNA controlled within the cell? p21 is a key cell cycle control protein and plays a major role. Under normal cellular conditions p21 is expressed at low levels during S-phase of the cell cycle, a time where PCNA levels are high and localise with

replication foci[522]. Upon initiation of the DNA damage response the cellular levels of p21 increases in both p53 dependent[577,578] and independent[557,579] pathways. This upregulation results in the co-localisation of p21 with PCNA, inhibiting DNA replication[580]. Importantly, although an inhibition to replication, p21 levels do not inhibit PCNA's function in DNA repair[580,581]. Thus initiation of DNA damage checkpoints resulting in p21 upregulation directs PCNA away from its role in replication, prolonging S-phase, giving the cell more time to complete DNA repair before replication proceeds[522].

p21 contains the same PCNA interaction motif as the replicative polymerase, FEN1 and DNA ligase I and is thought to inhibit replication by out-competing these enzymes for the same PCNA binding site[582]. The structural solution of the C-terminal peptide of p21 in complex with PCNA[522], shows this site to be the interdomain linker region of PCNA (figure 4.1.5). This region is responsible for the majority of interactions made by PCNA (table 4.1.3), however the N- and C-terminal regions are also known to bind other proteins such as cyclin D[553] and GADD45[546] respectively. That p21 does not inhibit PCNA's role in DNA repair suggests that its recruitment to repair pathways relies on interaction away from the IDL of PCNA. However the situation within the cell cannot be as simple as that since a major part of PCNA's function in DNA repair is the resynthesis step, involving the replicative polymerase. Presumably at some point during the repair process p21 must dissociate or be removed to allow this resynthesis step to occur.

4.1.4   DNA repair in the archaea

Bacteria   **Archaea**   **Eukarya**

Crenarchaea   Euryarchaea

LUCA
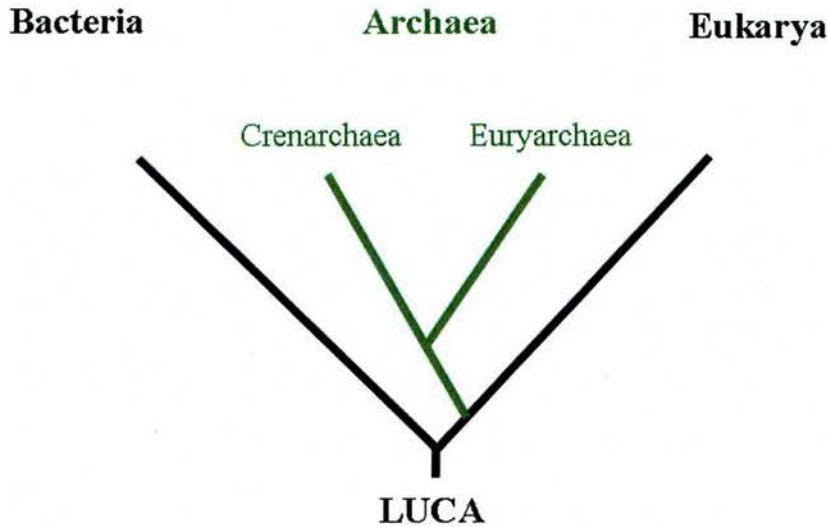
* Last universal common ancestor

Figure 4.1.6: The tree of life, showing that the last common ancestor between archaea and eukaryotes was more recent than either to bacteria, adapted from White[583].

The separation of cellular life into three domains; bacteria; archaea and eukarya, was first proposed by Woese in 1977, based on rRNA sequence divergences[584]. tRNA sequence differences[585] and biochemical studies[585,586] further validated this three-domains of life theory. Since then a large amount of phylogenetic research[587] has provided evidence that archaea and eukarya share a more recent common ancestor than either does to bacteria[588,589], as depicted in figure 4.1.6. Archaea can further be divided into euryarchaea and crenarchaea[587]. Euryarchaea includes methanogens, sulphate-reducing species and extreme halophiles. Crenarchaea includes acidophiles, extreme thermophiles and sulphate dependent archaea, originally classified due to their entirely thermophilic niches[587]. As many archaea can survive at high

temperatures it may suggest that the common ancestor between eury- and crenarchaea was thermophilic in nature.

The close phylogenetic relationship between archaea and eukarya is particularly striking for the nucleic acid metabolism pathways of translation[588-590], transcription[591-593] and replication[594,595]. Similarities between DNA repair pathways were thought to be less. However evidence is emerging that these processes may be more closely related[583].

HR double strand break repair is conserved throughout the three domains of life, driven by the homologous recombinases RecA[596], RadA[597] and Rad51[598] in bacteria, archaea and eukarya respectively. Closer similarity of RadA to Rad51[599], along with archaeal homologues of the eukaryotic Rad50/Mre11 proteins[600], suggests that HR-double strand break repair is more similar to the eukaryotic equivalent[601]. The enzymes involved in BER are conserved throughout all three domains of life, centred around DNA glycosylases that have specialised to remove specific DNA lesions[602]. The MMR process is more puzzling as, although homologues of the proteins involved are conserved between bacteria and eukaryotes, they are not present in archaea[583], suggesting that novel mechanisms may have evolved in archaea to deal with this type of damage. NER in archaea has been subject to the greatest debate in this field, with both prokaryotic and eukaryotic NER protein homologues found in some archaea[583,603].

The ability of archaea to carry out NER was first demonstrated in *Methanothermobacter thermoautotrophicus*[604] and shown to have a similar excision fragment to prokaryotes. Genomic analysis of *M. thermoautotrophicum*[603], reveals that they have a set of bacterial Uvr enzymes (1.3.6), as do a select group of other archaea including mesophilic methagens and halophiles[583]. However the majority of

163

archaea, including all crenarchaea, lack Uvr enzymes but instead have orthologues of the eukaryotic NER enzymes XPF, XPG, XPD and XPB (1.3), which are also present in some archaea that have Uvr enzymes[583]. It is now widely thought that the Uvr enzymes have been obtained from bacteria by horizontal gene transfer and that the original NER pathway in archaea was eukaryal in nature[583]. It is interesting to note the absence of any obvious homologues of the eukaryal NER DNA damage recognition proteins XPA, XPC and XPE[602]. The published genome of the primitive eukaryote *Plasmodium falciparum*[605] also lacks such homologues and *S. cerevisiae* lacks XPE, with the heterodimer Rad7-Rad16, along with the XPC homologue Rad4, providing the DNA damage recognition step[103]. This suggests that these DNA damage recognition proteins may have evolved in eukaryotes[583] and that other DNA damage sensing proteins may exist in archaea.

No patch repair, characteristic of prokaryotic and eukaryotic NER, has been detected in archaea lacking Uvr homologues. However the ability of the model archaean *S. solfataricus* (lacking Uvr enzymes) to repair the UV induced lesions CPDs and 6-4 PPs in the dark has been shown *in vivo*[606], suggesting a form of NER exists. Analysis of the *S. solfolobus* transcriptional response to UV damage, reveals that XPF, XPB and XPG are upregulated[606]. XPF and XPG are also upregulated after exposure to actinomycin D, an inter-collating agent that also inhibits transcription, although the XPB level is unaffected[606]. This implicates these enzymes in the damage response. However the varying degrees of upregulation suggests diverse roles of these proteins, for example XPB in transcription, XPG in Okazaki fragment maturation and XPF in recombination[606].

4.1.5   Archaeal PCNA

PCNA but not β-subunit homologues exist in archaea, sharing approximately 20-30 % sequence identity with eukaryotic PCNA[607]. The sequenced genomes of euryarchaea all contain one copy of PCNA, shown to form a homotrimer[523]. In the crenarchaea the situation is more complicated with two homologues existing in *Pyrobaculum aerophilium*[608] and three homologues in *Aeropyrum pernix*[609] *S. tokodaii*[583], and *S. solfataricus*[607]. While it has been shown that in *A. pernix* PCNA can exist as both homo- and heterotrimers[609], *S. solfataricus* only forms a PCNA1, PCNA2 and PCNA3 heterotrimer [607].

Protein	Effects of interaction with PCNA	Interaction detection method	References
DNA replication			
Elongation			
DNA polymerase B	Stimulation of polymerase activity	Co-IP, *in vitro* functional assays, yeast-2-hybrid, GST pulldown	607,610,611
DNA polymerase D	Stimulation of polymerase activity	Co-IP, *in vitro* functional assays	610,611
RFC	Loading of PCNA onto DNA	Co-IP, *in vitro* functional assays, yeast-2-hybrid, structural solution	607,610,612
Okazaki fragment maturation			
DNA ligase	Ligase stimulation	GST-pulldown	607
FEN1/XPG	FEN-1 stimulation	GST-pulldown, yeast-2-hybrid, *in vitro* functional assay	607,613,614
RNase HII	RNase HII stimulation	Yeast-2-hybrid	610
DNA repair and recombination			
DNA polymerase Y1	Polymerase stimulation	*in vitro* functional assays	615
RAD2	-	Yeast-2-hybrid	610
UDG1	Stimulation of glycosylase activity	GST-pulldown	613
FEN1 / XPG	Stimulation of FEN1 endouclease	*in vitro* functional assays	607,613,614
XPF	XPF stimulation	*in vitro* functional assays, GST-pulldown	601

Table 4.1.7: Known PCNA interaction proteins in archaea, adapted from Vivona and Kelman 2003[556].

The formation of the *S. solfataricus* PCNA heterotrimer (*SsoPCNA*) first requires the heterodimerisation of PCNA1 and PCNA2, which form a stable complex with a K_D of 1.2×10^{-11} M[607]. This heterodimer, but not the PCNA1 or PCNA2 monomers, can then interact with PCNA3 forming a complex with a K_D of 2.7×10^{-7} M[607]. No homotrimer formation is observed for any individual PCNA subunit[607]. This heterotrimer is functional in DNA replication[607] and repair[607,616] and has been shown to make important interactions with some of the proteins shown in table 4.1.7[601].

The heterotrimeric nature of *SsoPCNA* suggests that the 3 individual subunits will be able to make subunit specific interactions and this hypothesis was tested by Bell and co-workers in the original paper describing the heterotrimer[607]. They found strong interactions between PCNA1 and FEN1, PCNA2 and PolB and PCNA3 with DNA ligase, they also showed that the PCNA heterotrimer could simultaneously bind these enzymes[607]. These results suggest that in some archaea the machinery involved in Okazaki fragment maturation[563,564] may assemble with heterotrimeric PCNA forming a functional machine. The observation that *SsoPCNA* can bind uracil DNA glycosylase (UDG1) through PCNA3 suggests that a similar machine may also form during processive replication to scan newly synthesised DNA for uracil incorporation[616]. It is unlikely that such an interaction is important for a global form of BER by UDG1 *in vivo* as, in the presence of archaeal chromatin-like binding proteins, the PCNA-UDG1 activity is strongly repressed[616].

The archaeal clamp loader (RFC) consists of one large subunit (RFC-l) and a homotetrameric small subunit (RFC-s)[612,617]. The interactions formed between *SsoPCNA* and RFC are also subunit specific, with PCNA3 interacting with RFC-l and only the PCNA1-PCNA2 heterodimer interacting with RFC-s[607]. This suggests a

mechanism of loading where either RFC-l pulls PCNA3 away from the PCNA1-PCNA2 heterodimer or the RFC-l and RFC-s subunits bind their partners and keep the heterotrimer apart until loaded onto DNA[607] (figure 4.1.8).
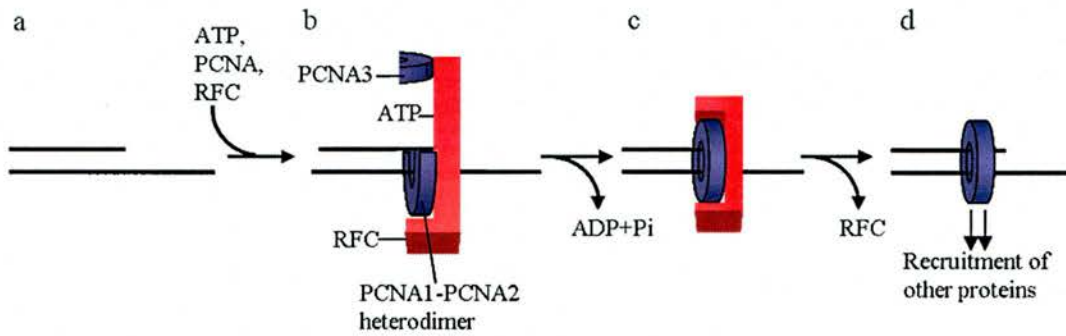
Figure 4.1.8: Cartoon representation of the possible mechanism for the loading of *Sso*PCNA onto DNA. (a and b) RFC pulls or keeps PCNA3 apart from the PCNA1-PCNA2 heterodimer and recognises the ds/ssDNA junction. (c) The heterotrimeric clamp is loaded onto DNA in an ATP dependent reaction. (d) PCNA can then recruit interacting proteins to DNA.

*Sso*PCNA also interacts with *S. solfataricus* XPF (*Sso*XPF)[601], which was unexpected as this interaction is not conserved in euryarchaea or eukaryotes, unlike the XPG-PCNA interaction[151,601]. This interaction is dependent on the extreme C-terminus of *Sso*XPF and is essential for the endonuclease activities observed for *Sso*XPF[601,618,619]. GST-fusion pull-down experiments show that *Sso*XPF specifically interacts with *Sso*PCNA through the PCNA1 and PCNA3 subunits[601].

A heterotrimeric sliding clamp also exists in eukarya, although this clamp does not function in replication but is specialised to the DNA damage response (4.1.6)[620,621].

4.1.6   The 9-1-1 complex

The 9-1-1 heterotrimeric complex consists of the eukaryotic proteins that are homologous to the *Schizosaccharomyces pombe* proteins Rad9, Rad1 and Hus1[620,621]. Structural modelling techniques predicted that, despite limited sequence conservation, all three proteins had a PCNA-like fold, suggesting that this heterotrimer formed a sliding clamp called the 9-1-1 complex[620,621]. This is supported by biochemical evidence[622] and transmission electron microscopy show that the 9-1-1 complex forms a ring-like structure which is indistinguishable from that of PCNA[623].

In response to DNA damage the damage checkpoint is activated, resulting in the arrest of the cell cycle and activation of DNA repair pathways[624,625] as well as a variety of other DNA metabolic functions[626] and apoptosis (figure 4.1.9)[627,628]. This checkpoint coordinates the repair process and gives the cell time to repair its genome. The proteins that sense the DNA damage have yet to be identified although the 9-1-1 complex, along with Rad17, are all implicated in these early events[626]. Rad17 shares homology with the 5 subunits of RFC and can interact with the 4 RFC-s subunits[629]. Rad17 can bind to chromatin, an association that is greatly up-regulated in response to DNA damage[626,630]. It can also interact with the 9-1-1 complex, which is essential for the recruitment of this complex to chromatin[626], suggesting that it functions as a clamp loader for the 9-1-1 complex[631,632]

Sensing of DNA damage results in the activation of the ATM and ATR kinases, which phosphorylate a variety of substrates involved in the damage response[626,633] including, in a 9-1-1 complex dependent manner, Rad17[632] and the Chk1 and Chk2 kinases[629]. This has led to the hypothesis that the 9-1-1 complex acts as a central mediator of these early events[632], bridging interactions between the ATR

and ATM kinases and their substrates, reminiscent of PCNA's role in cell cycle control (4.1.2).
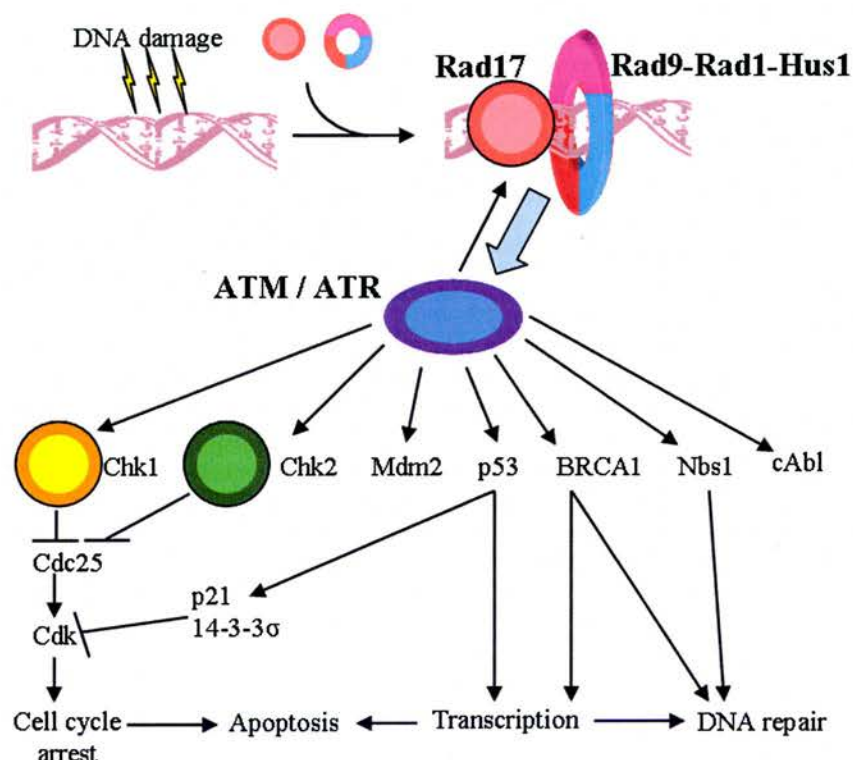
Figure 4.1.9: Initial DNA damage response events mediated by the Rad9-Rad1-Hus1 sliding clamp. Adapted from Zhou[134].

4.1.7 XPF homologues

Figure 4.1.10 shows the domain architecture and oligomeric nature of XPF orthologues from crenarchaea, euryarchaea and eukarya. Crenarchaeal XPF has an N-terminal nuclease domain and two consecutive C-terminal helix-hairpin-helix (HhH) domains (XPF(HhH)$_2$), separated by an 11 amino acid linker[601,634]. Each of the HhH domains contains two consecutive HhH motifs[601,634]. The crenarchaeal XPF has a PCNA interaction motif at its C-terminus and its interaction with *S. solfataricus*

heterotrimeric PCNA is essential for its nuclease activity, but not its DNA binding[601,634].

Euryarchaea have an extra N-terminal domain that shows sequence similarity to DNA helicases and indeed possesses DNA dependent ATPase activity[635]. The C-terminal nuclease and XPF(HhH)$_2$ domain is conserved between cren- and euryarchaea[601]. However euryarchaea lack the C-terminal PCNA interaction motif, suggesting that the fusion of a helicase with XPF nuclease during evolution dispensed with the PCNA dependence of the XPF nuclease activity[601].

Eukaryotic PCNA is similar in domain architecture to euryarchaea, however the sequence of the N-terminal DNA helicase domain has diverged such that it is no longer active[636]. Unlike archaeal XPF, which forms a functional homodimer, eukaryotic XPF forms a heterodimer with the ERCC1 protein[148], which acts as a bridge in the NER pathway by interacting with RPA[135,637]. ERCC1 is distantly related to XPF, containing the nuclease and HhH$_2$ domains, although the nuclease activity has been lost[638]. The conservation of these domains suggests that during the evolution of eukaryotes the nuclease and HhH$_2$ domains of XPF have duplicated and the product diverged to evolve the ERCC1 protein[583]. A similar process has also occurred to evolve another eukaryotic XPF homologue, Mus81, which forms a functional heterodimer with either Eme1/Mms4[639] or cds1[640,641]. A major difference between the domain architecture of Mus81 and other XPF nuclease domains is that instead of having the two HhH motifs at its C-terminus, Mus81 is flanked by single HhH motifs (figure 4.1.10d)[634].
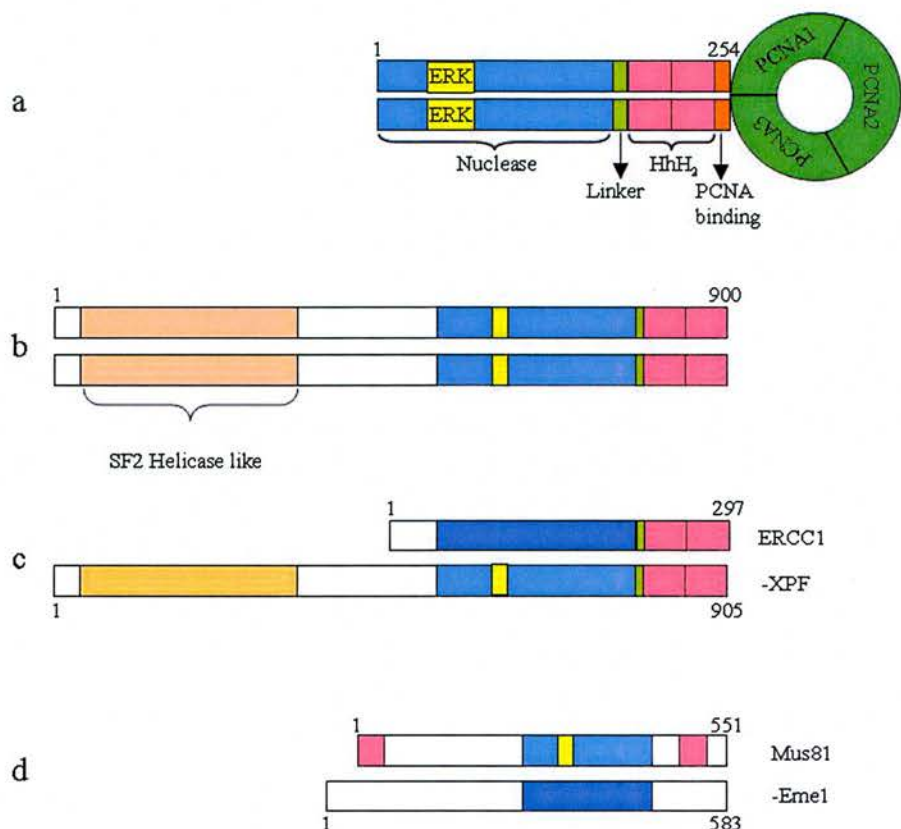
Figure 4.1.10: Functional XPF from crenarchaea (a), euryarchaea (b) eukaryotes (c) and the XPF homologue Mus81-Eme1 (d). Domains are coloured as labelled for cren- or euryarchaea and non-functional domains are coloured a shade darker. PCNA is shown bound to XPF in crenarchaea as it is essential for it activity[601]. Adapted from Roberts *et al.*[601] and Newman *et al.*[634].

XPF homologues have been the subject of extensive biochemical studies including the XPF-ERCC1[148], Mus81[639,640] and *S. solfataricus* XPF enzymes[601,618,619]. XPF-ERCC1 endonuclease activity relies on both the structure and sequence of its DNA substrate. It cleaves 3' to a pyrimidine nucleotide[148], upstream of a ds/ssDNA junction where the ssDNA leaves in a 5'-3' direction, such as the DNA structure formed 5' to DNA damage during NER[148,642]. Mus81 shows no sequence specificity for its DNA substrate and is most active on 3' flaps with both upstream and

downstream duplex present and DNA junctions, which are formed from the processing of stalled replication forks[639-641]. There appears to be no overlap in the substrate specificity of XPF-ERCC1 and Mus81, with each enzyme evolved to specifically process different DNA junctions that occur within the cell.

S. solfataricus XPF (*Sso*XPF) shows similar sequence specificity for its DNA substrate as XPF-ERCC1 and it can also cleave a splayed duplex (figure 4.1.11), the XPF-ERCC1 DNA substrate[601]. Further kinetic analysis of the substrate specificity of *Sso*XPF demonstrates that it is more active on Mus81-like substrates, particularly 3' flaps, nicked duplexes, 3- and 4-way junctions[619]. These results suggest that crenarchaeal XPF is an ancient form of both XPF-ERCC1 and Mus81, sharing features of both enzymes. The recent structural solution of XPF from the crenarchaea *A. pernix*, which is closely related to *S. solfataricus*[634], both in complex and in the absence of DNA provides insights into how crenarchaeal XPF recognises its DNA substrates and the evolution of the XPF family of endonucleases.
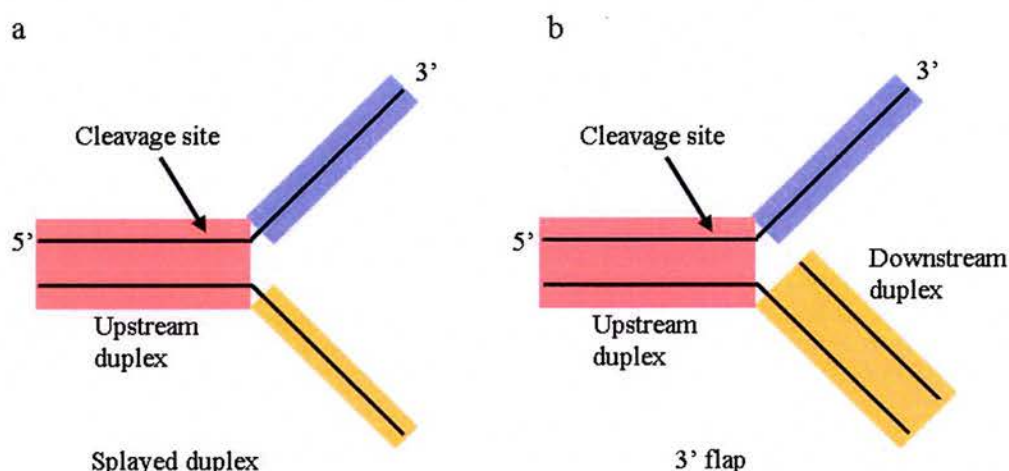
Figure 4.1.11: Substrates cleaved most efficiently by the eukaryotic XPF homologues XPF-ERCC1 (a) and Mus81 (b). *Sso*XPF can cleave both junctions; however it is more active on the 3' flap (b). Adapted from Newman *et al.*[634].
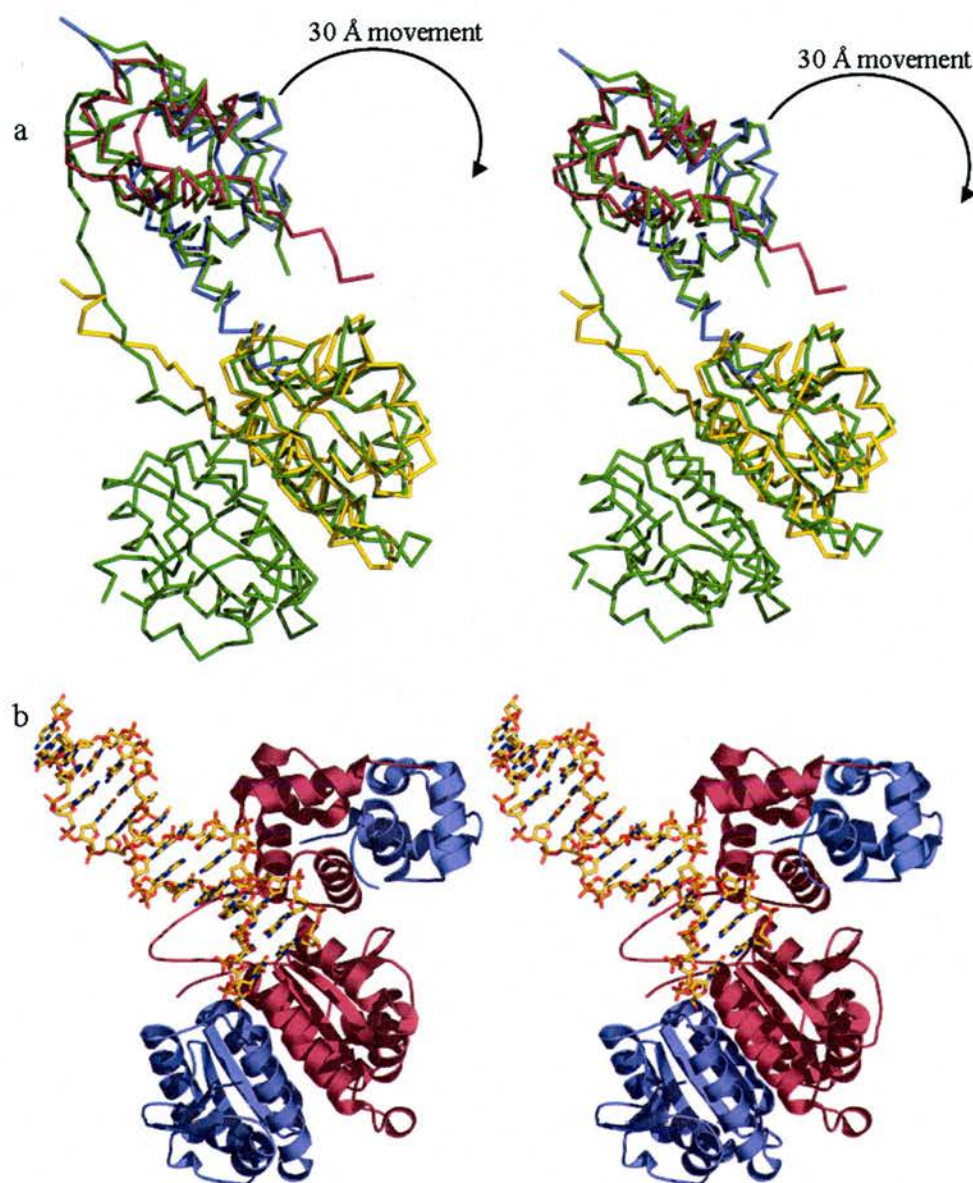
Figure 4.1.12: Stereo views of (a) the superimposition of apo-XPF from *A. pernix* (green) with the ERCC1 nuclease domain (yellow) and XPF (red) - ERCC1 (blue) HhH$_2$ domains. (b) DNA-complexed XPF from *A. pernix*. One monomer is coloured in red and the other in blue, the picture is orientated with the XPF(HhH)$_2$ and nuclease domains at the top and bottom of the picture respectively. The compaction of the monomer coloured in red is clearly visible upon binding DNA. Adapted from Newman *et al.*[634].

Attempts to structurally characterise eukaryotic XPF homologues have been hampered due to instability of the enzyme, although structures of both the ERCC1 nuclease domain and HhH$_2$ domains of the XPF-ERCC1 heterodimer[643] have recently been reported. Archaeal XPF has therefore been a useful model for structural studies (figure 4.1.12, 4.1.13). The structures of the nuclease domain of XPF from *Pyrococcus furiosus*[644] and both apo- and DNA complexed, essentially full length XPF from the crenarchaean *Aeropyru. pernix*[634] (figure 4.1.12) have been published. These structures, particularly of *A. pernix* XPF, provide insights into the DNA-structure specificity of these enzymes[634].

The *A. pernix* XPF-DNA complex is with a duplex of DNA which is not a substrate. Upon DNA binding one of the protomers folds up, with the XPF(HhH)$_2$ domains moving 30 Å (figure 4.1.12), interacting with DNA and the nuclease domain[634]. The authors commented on a two-fold axis between the XPF(HhH)$_2$ domains of the dimer, suggesting that there is the potential for a DNA duplex to bind to the XPF(HhH)$_2$ domain of the extended protomer[634]. Modelling of this second DNA duplex positions the 5' end into the nuclease active site of the extended protomer, with an angle of approximately 90° between the modelled and complexed duplexes (figure 4.1.13). The authors hypothesise that the duplex DNA bound in the structure represents the downstream duplex of an XPF substrate. There is a hydrophobic cleft adjacent to this duplex DNA which is thought to bind the short region of ssDNA produced by XPF cleavage[634].

The separate structures of the nuclease and HhH$_2$ domains of human ERCC1 show that the fold is the same as archaeal XPF (figure 4.1.12)[643]. Based on these structures and biochemical observations the authors propose a model for XPF-ERCC1 binding to a splayed duplex in which the ssDNA arm leaving in a 3'-5' direction is

bound by both the nuclease and HhH_2 domain of ERCC1, with the other arm bound by the $XPF(HhH)_2$ domain (figure 4.1.13)[643]. In this model the nuclease domain of XPF binds the ds/ssDNA junction where the ssDNA leaves in a 5'-3' direction, where it makes its incision[643]. This model is in contrast with that proposed by Newman *et al.* who drew analogies from their XPF structures from *A. pernix*. In this model they suggest that the extended protomer represents ERCC1, with the HhH_2 domain proposed to bind the upstream dsDNA, with $XPF(HhH)_2$ binding the downstream ssDNA[634].

Although these models of archaeal XPF and XPF-ERCC1 substrate recognition are attractive they are not conclusive. Further characterisation of XPF from crenarchaea, both in complexes with DNA and with PCNA, is therefore important to fully understand the structural basis of substrate recognition in this ancient XPF enzyme.
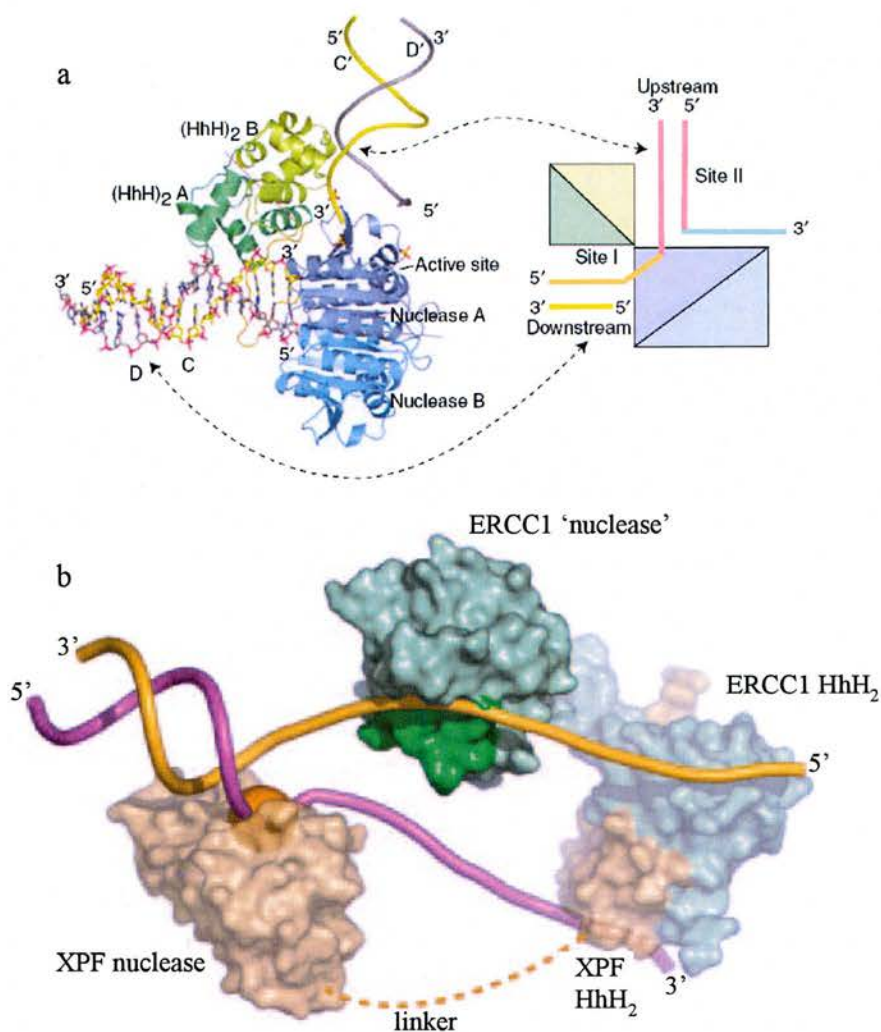
Figure 4.1.13: Models of substrate binding by XPF homologues. (a) XPF from the crenarchaean *A. pernix*, taken from Newman *et al.*[634]. (b) Human XPF-ERCC1, taken from Tsodikov *et al.*[643].

4.2 MATERIALS AND METHODS

4.2.1 Over-expression of PCNA1, PCNA2, PCNA3 and XPF(HhH)$_2$

IPTG induction (for tagged and untagged constructs)

Table 4.2.1 details the PCNA and XPF constructs we received from our collaborators in St Andrews. Soluble protein was over-expressed from all constructs using the same conditions. *E. coli* Rosetta (DE3) cells, a BL21 (DE3) derivative which contains a plasmid (pRARE) containing the 6 rare codons tRNAs for the codons AGG, AGA, AUA, CUA, CCC, and GGA, were transformed with plasmid DNA encoding the relevant protein to be over-expressed. The Rosetta (DE3) cell line enhances the the expression of genes in *E. coli* that contain codons not commonly found in *E. coli*. Single colonies were grown in 10 ml LB overnight, supplemented with 50 μgml^{-1} kanamycin or 100 μgml^{-1} ampicillin for tagged and untagged constructs respectively. Overnight cultures were used to inoculate 500 ml of LB in 2 l flasks and grown to an OD$_{600}$ of 0.8-1.0 at 37 °C. The temperature was then reduced to 25 °C and protein over-expression was achieved by overnight induction with 0.2 mM IPTG.

Auto-induction (for tagged constructs)

1 ml of a 10 ml overnight, grown as above, was used to inoculate 50 ml of LB and cells were harvested, after overnight incubation at 37 °C, at 2000 g and re-suspended in 20 ml PBS. This was then used to inoculate auto-induction medium for over-

expression of protein, as detailed by Studier[645]. Briefly, 400 ml of auto-induction growth medium in a 2 l baffled flask, supplemented with 100 μgml^{-1} kanamycin, was inoculated with 5 ml re-suspended cells and grown at 37 °C, with shaking at 300 r.p.m, for 6 hours to start bacterial growth. The temperature was then lowered to 20 °C and left for 72 hours, allowing bacterial growth to reach saturation and the subsequent self-induction of protein expression[645].

Protein	Vector	Tag	Restriction Sites and modifications	Antibiotic Resistance	Amino Acids Cloned
PCNA1 (SSO0397)	pETDuet	No tag	NcoI / BamHI	Ampicillin	M_1 - L_{249}
PCNA1 (SSO0397)	pET19b	N-terminal 6x histidine	NcoI / BamHI	Kanamycin	M_1 - L_{249}
PCNA2 (SSO1047)	pETDuet	No tag	NcoI / BamHI	Ampicillin	M_1 - D_{246}
PCNA2 (SSO1047)	pET19b	N-terminal 6x histidine	NcoI / BamHI	Kanamycin	M_1 - D_{246}
PCNA3 (SSO0405)	pETDuet	No tag	NcoI / BamHI	Ampicillin	M_1 - V_{259}
PCNA3 (SSO0405)	pET19b	N-terminal 6x histidine	NcoI / BamHI	Kanamycin	M_1 - V_{259}
XPF(HhH)$_2$ (SSO0729)	pET28b	N-terminal 6x histine with thrombin cleavage site	L_{150} mutated to M_{150} NcoI / BamHI	Kanamycin	M_{150} - L_{233}

Table 4.2.1: Constructs of the PCNA and XPF(HhH)$_2$ genes obtained from out collaborators Jana Rudolph and Professor Malcolm White in St Andrews. In brackets are the gene annotations from the *S. solfataricus* genome.

Cells grown by either technique were harvested at 10500 g and were re-suspended in 50 mM HEPES pH 7.5, 10 mM imidazole and 300 mM NaCl (His-tagged) or 50 mM MES pH 5.7 and 50 mM NaCl (untagged). Soluble proteins were

extracted by incubation at room temperature for 1 hour with 100 μgml^{-1} lysozyme and 20 μgml^{-1} DNase (Sigma) before lysis, either by 4 x 30 sec sonication passes on ice or two passes through a Constant cell disrupter (Constant systems). The soluble and insoluble fractions were separated by centrifugation for 30 min at 75500 g and target protein subsequently purified.

4.2.2   Purification

Individual components

The same procedure was used to purify all tagged proteins as detailed for tagged ArnA in chapter 2.2.2 (figure 4.2.2).

Lysate containing target untagged protein was applied to a 5 ml Heparin HP column (Amersham Biosciences), pre-equilibrated with 50 mM MES pH 5.7 and 50 mM NaCl. Extensive washing with 50 mM MES pH 5.7 and 1 M NaCl removed almost all contaminating proteins and essentially pure PCNA1, 2 or 3 was eluted from the column with an increasing salt concentration up to 2 M. After concentration of relevant fractions and dialysis into 50 mM MES pH 5.7, 300 mM NaCl and 2 mM DTT, the individual proteins were purified to homogeneity by S200 gel filtration (Amersham Biosciences).

The XPF(HhH)$_2$ construct contains a thrombin cleavage site between the histidine tag and the start of the XPF(HhH)$_2$ gene. Untagged XPF(HhH)$_2$ was obtained by overnight incubation of histidine tagged protein, eluted from the HisTrap column and dialysed into 50 mM Tris pH 8.2 and 250 mM NaCl, with 1 unit of thrombin. Untagged XPF(HhH)$_2$ was separated from any uncleaved protein, the

cleaved tag and thrombin by passing it back through the HisTrap column, before further purification by S200 gel filtration.
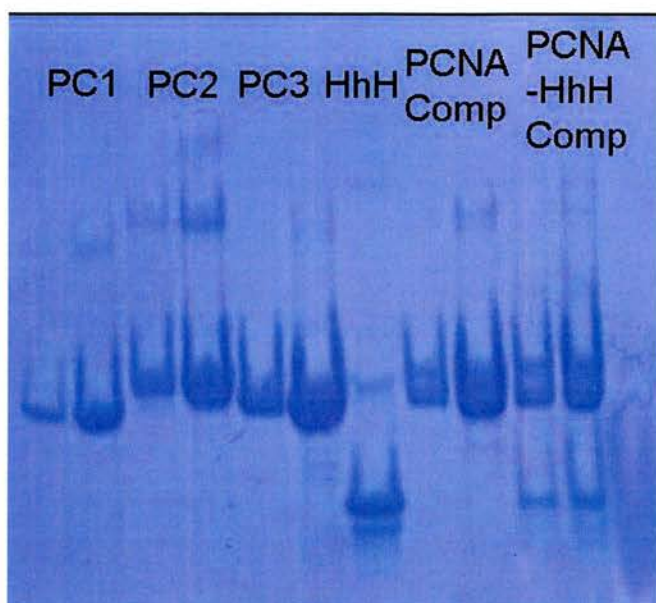
Figure 4.2.2: Coomssie blue stained SDS PAGE gel showing the purified tagged individual PCNA proteins, XPF(HhH)$_2$ (HhH), the tagged PCNA heterotrimer and tagged PCNA-XPF(HhH)$_2$ heterotetramer.

As it took so long to purify large amounts of the three PCNA components and their complex (in excess of 15 gel filtration columns), this purification protocol was optimised and the throughput significantly improved without impacting on final purity, as judged by Coomassie blue stained SDS-PAGE. In the optimised protocol the PCNA1, 2 and 3 soluble lysates, containing a slight excess of PCNA3, were mixed for one hour before heat treatment for 30 min at 65 °C. Precipitated protein was removed by centrifugation at 75500 g for 20 min and the PCNA heterotrimer was purified as for an individual PCNA subunit.

The pure heterotrimer was mixed with an excess of purified XPF(HhH)$_2$, either tagged or untagged, for 1 hour at room temperature, before the PCNA heterotetramer was purified by S200 gel filtration. In solution XPF(HhH)$_2$ exists as a homodimer of ~20 kDa and the PCNA heterotetramer is ~105 kDa, therefore

XPF(HhH)$_2$ was mixed in excess as it is easier to resolve species of ~105 and ~20 kDa than ~105 and ~75 kDa by gel filtration (figure 4.2.3).
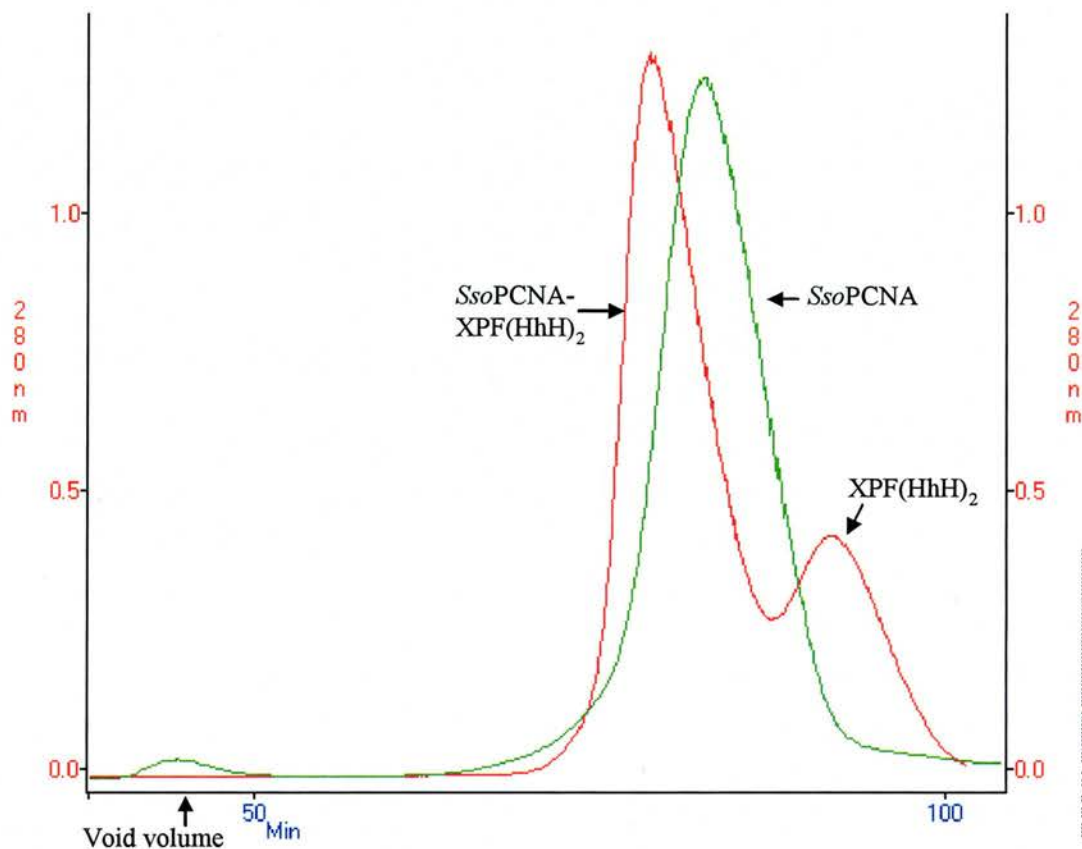
Figure 4.2.3: S200 gel filtration traces of the *Sso*PCNA heterotrimer (green) and *Sso*PCNA-XPF(HhH)$_2$ (red). The overlapping traces of the heterotetramer and heterotrimer show the importance of using an excess of XPF-(HhH)$_2$.

4.2.3 Crystallisation

Purified heterotrimer and heterotetramer were concentrated to 20 mg ml^{-1} as judged by Bradford assay (Sigma) and dialysed into 20 mM HEPES pH 7.5, 20 mM NaCl and 2 mM DTT prior to screening for crystallisation conditions. The sitting drop method was employed for crystal screening, using a drop size of 2 μl, containing 1 μl protein and 1 μl precipitant, against 100 μl of mother liquor. Crystal screens 1 and 2,

Index, Peg/Ion, Natrix (Hampton), Wizard 1 and 2 (Emarald Biosciences) and protein-protein complex (Sigma) sparse matrix screens were employed for initial crystal screening.

PCNA Hetereotrimer

Tagged PCNA heterotrimer formed small clusters of crystals in several initial conditions, which all contained between 10 and 25 % PEG (between PEG 1000 and PEG 4000), 100-300 mM salt and a pH between 6 and 8. The clusters were cryoprotected in 10 – 20 % (2R,3R)-(-)2,3-butanediol (Sigma) and subjected to X-ray diffraction on our in-house source (Rigaku 007 rotating anode) using an RAxis IV detector. A 20 min exposure revealed that these crystals were protein, diffracting to ~8 Å, figure 4.2.3a. Extensive optimisation of these conditions was employed including variation of precipitant, salt and buffer, along with changing the concentrations of these and the pH of the buffer. Different batches of purified PCNA heterotrimer was also used during optimisation trials at concentrations varying from 3 to 20 mg ml^{-1}. Eventually this optimisation, using a protein concentration between 10 and 20 mg ml^{-1} from a batch of protein over-expressed using the auto-inducing medium, gave rise to large, single crystals that had well defined edges (figure 4.2.3b). This condition was 11 % PEG 3350, 0.18 M ammonium acetate, 0.01 M calcium chloride and 100 mM Bis-tris pH 6.5. A single crystal was looped out and cryoproteced as above to test for diffraction quality in house and at the ESRF. Despite their excellent appearance, these crystals only diffracted to between 4.5 and 5 Å (figure 4.2.3b). Further optimisation using streak and stock seeding as well the

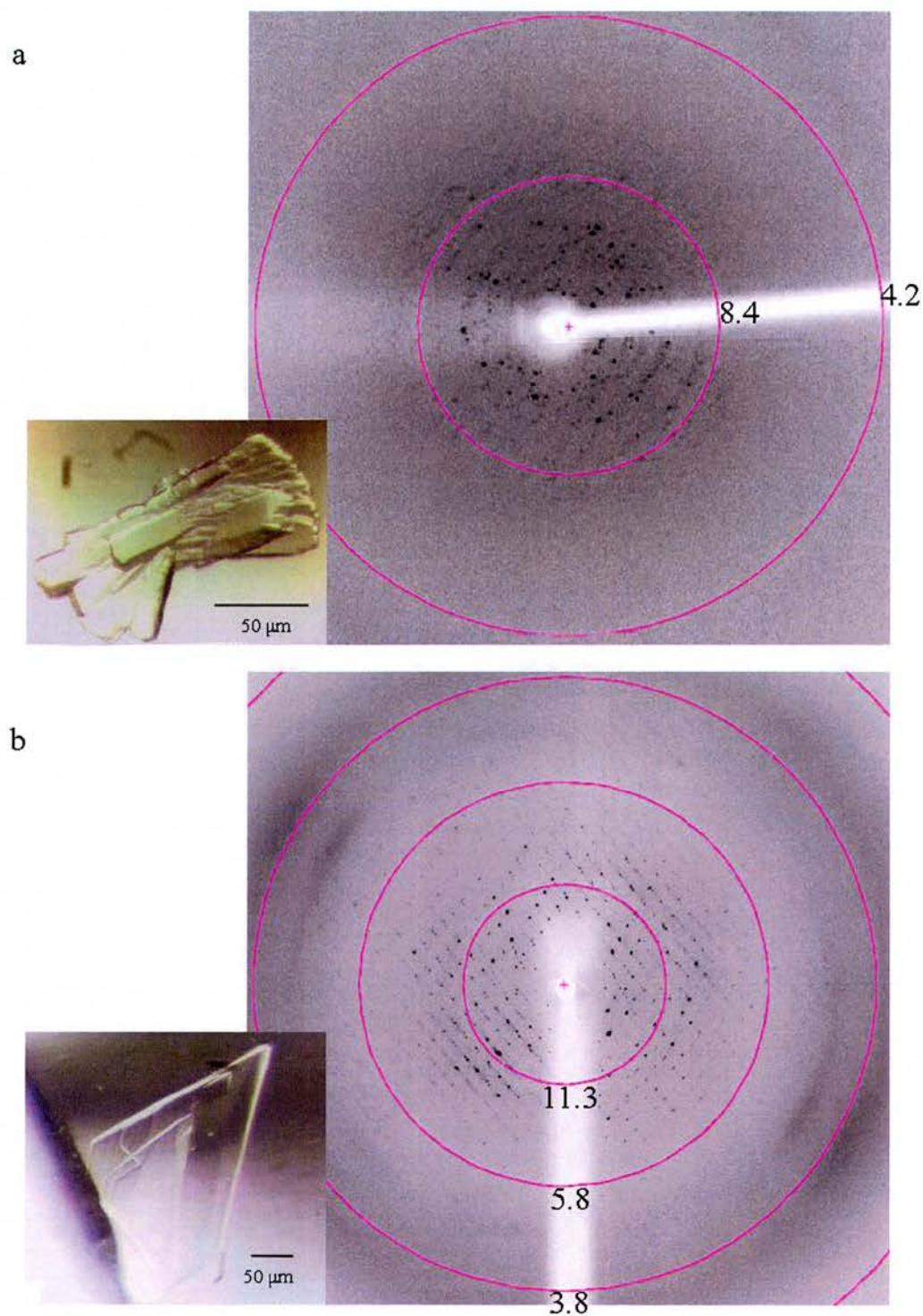Hampton additive and detergent screens failed to improve the diffraction of these crystals.

Figure 4.2.3: Initial crystals of the PCNA heterotrimer and the resulting diffraction on our in house source (a) and after optimisation at the ESRF (b). Labels on the diagrams are in Å.

The Sigma protein-protein complex screen was designed based on the structures of macromolecular complexes in the PDB as of 2001[646]. Their finding was that, in general, protein-protein complexes are more likely to crystallise with a lower precipitant concentration than normal soluble proteins, with a preference for PEG as a precipitant and a smaller pH range between 4 and 9[646]. Observation of the conditions that gave rise to PCNA heterotrimer crystals in the initial screening process was consistent with this, revealing a preference for PEG at a lower concentration than present in most commercial screens. The Nextal PEG screen contains the 48 20 % PEG 3350 conditions that are present in the Hampton Peg/Ion screen as well as a further 48 conditions, screening different PEGs against buffers. Further crystal trials were focussed around the Nextal PEG screen with the 96 conditions undiluted, diluted by 30 % or diluted by 50 % with water, screened against PCNA heterotrimer at 7, 10 and 15 mg ml^{-1}. The screens using 50 % diluted Nextal PEG solutions gave rise to crystals that looked good quality within 1 week, however these crystals diffracted poorly (figure 4.2.3b). Two other conditions in the 50 % diluted screen also gave rise to crystals, 0.05 M sodium acetate pH 4.6 and either 10 % PEG 10000 or 7.5 % PEG 20000. Their quality appears poor, forming very thin sheets that clustered together. However careful manipulation allowed removal of a single crystal, which was cryoprotected with 20 % (2R,3R)-(-)2,3-butanediol prior to X-ray diffraction. A 30 min exposure on an in-house source resulted in diffraction to 3.3 Å (figure 4.2.9a). Optimisation of these crystals resulted in larger, slightly thicker plates and although they were always found as a cluster, good single crystals could be removed and used for data collection (figure 4.2.9) (4.2.4).

PCNA-XPF(HhH)₂ complex

Initial crystals were obtained for the tagged PCNA-XPF(HhH)$_2$ complex in 25 % PEG 3350, 0.2 M ammonium acetate and Bis-tris pH 5.5. These small crystals were cryoprotected in 20 % glycerol and a 20 min exposure to X-rays from our in-house source showed them to be protein, with some spots diffracting to ~11 Å (figure 4.2.6a). Optimisation of this initial condition gave rise to single crystals of an approximate size of 0.2 by 0.2 by 0.1 mm, that took between 3 and 5 weeks to grow. These crystals were cryoprotected with 20 % glycerol and frozen at 100 K. X-ray diffraction was obtained on our in-house source to ~6 Å, with a 20 min exposure. Increasing the exposure time to 40 min resulted in the diffraction at 0° to 3.7 Å (figure 4.2.6b). However, because the crystal was damaged, the 90° diffraction was only to ~5 Å and was therefore not suitable for data collection.

It was hoped that taking the crystals to a synchrotron X-ray source would improve their diffraction. These crystals grow on the bottom or side of the sitting drop plate and are very fragile and attempts to tease the crystals off the plate usually resulted in cracked crystals, which diffracted poorly (around 5 Å). One small crystal was successfully removed from a drop without cracking, and in some directions this crystal diffracted to 3 Å. This diffraction was very anisotropic and the data collected was of poor quality and unsuitable for structural solution. The fragility of the crystals at the ESRF was not observed when mounting the crystals in house and it is likely that the transportation of these crystals to the ESRF resulted in their deterioration. Hanging drop plates were also utilised in the optimisation of these crystals to prevent their growth on the surface of the crystal plates. Although relatively large single

crystals could be obtained and looped from the drop without damage, they degraded

upon cryoprotection at the ESRF, diffracting to only 5 Å.
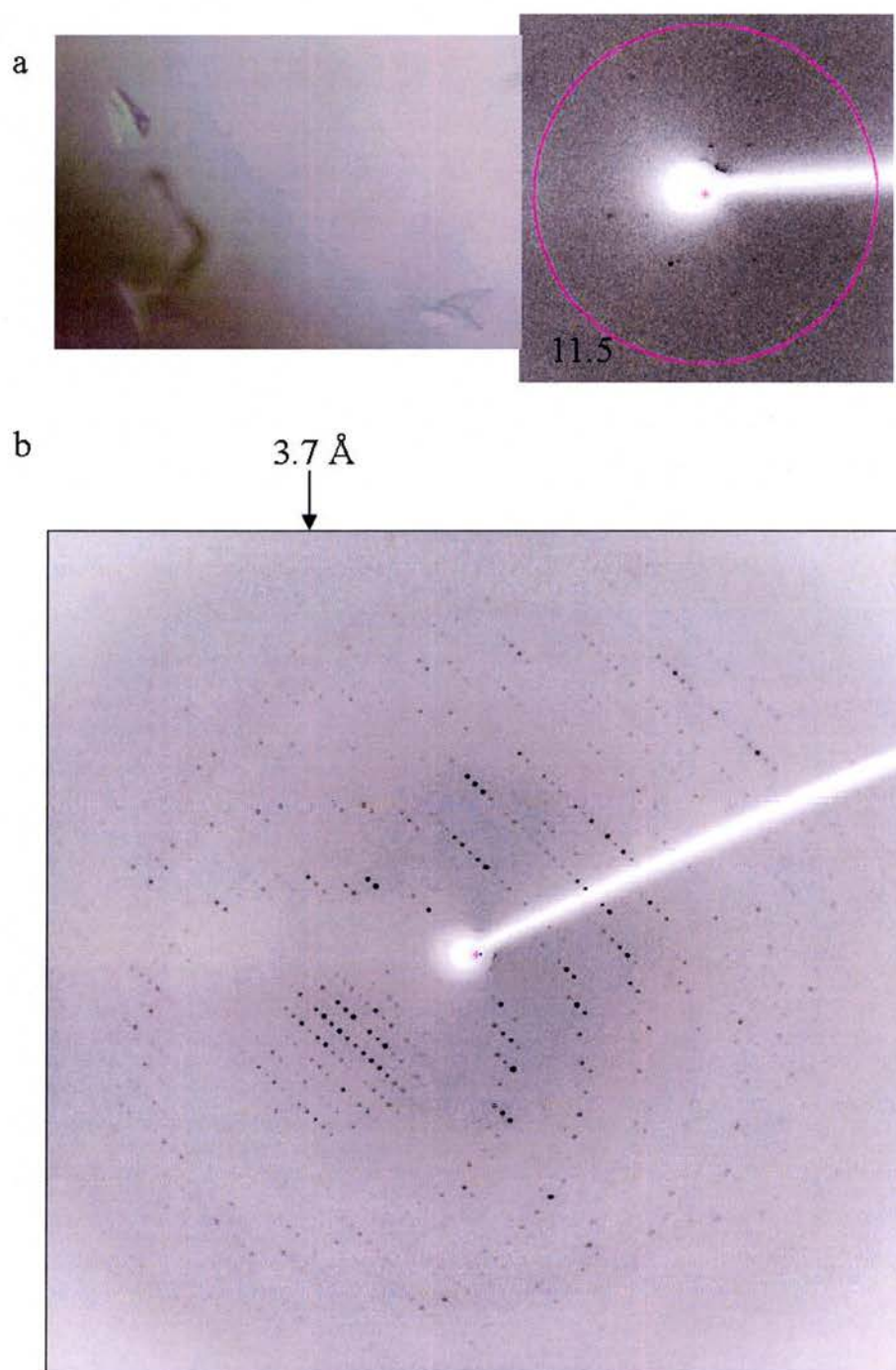
Figure 4.2.6: Initial diffraction from PCNA-XPF(HhH)₂ crystals (a), then after
optimisation to 3.7 Å (b).

Attempts to repeat this crystal condition have failed. Many other conditions have yielded crystals that look of excellent quality however, similarly to the heterotrimeric PCNA these only diffract, at best, to 4.5- 5 Å at the synchrotron, shown in figure 4.2.7.

Very recently untagged PCNA-XPF(HhH)$_2$, at a concentration of 8 mg ml^{-1}, formed a small cluster of crystals (figure 4.2.8) in 12.5 % PEG 3350, 0.1 M ammonium sulphate and 0.05 M HEPES pH 7.5. This cluster of crystals was cryoprotected in 20 % (2R,3R)-(-)2,3-butanediol and initial diffraction to 3.0 Å was observed (figure 4.2.8), although due to the presence of multiple lattices this crystal is not suitable for data collection. Optimisation trials for improving these promising crystals are currently underway.
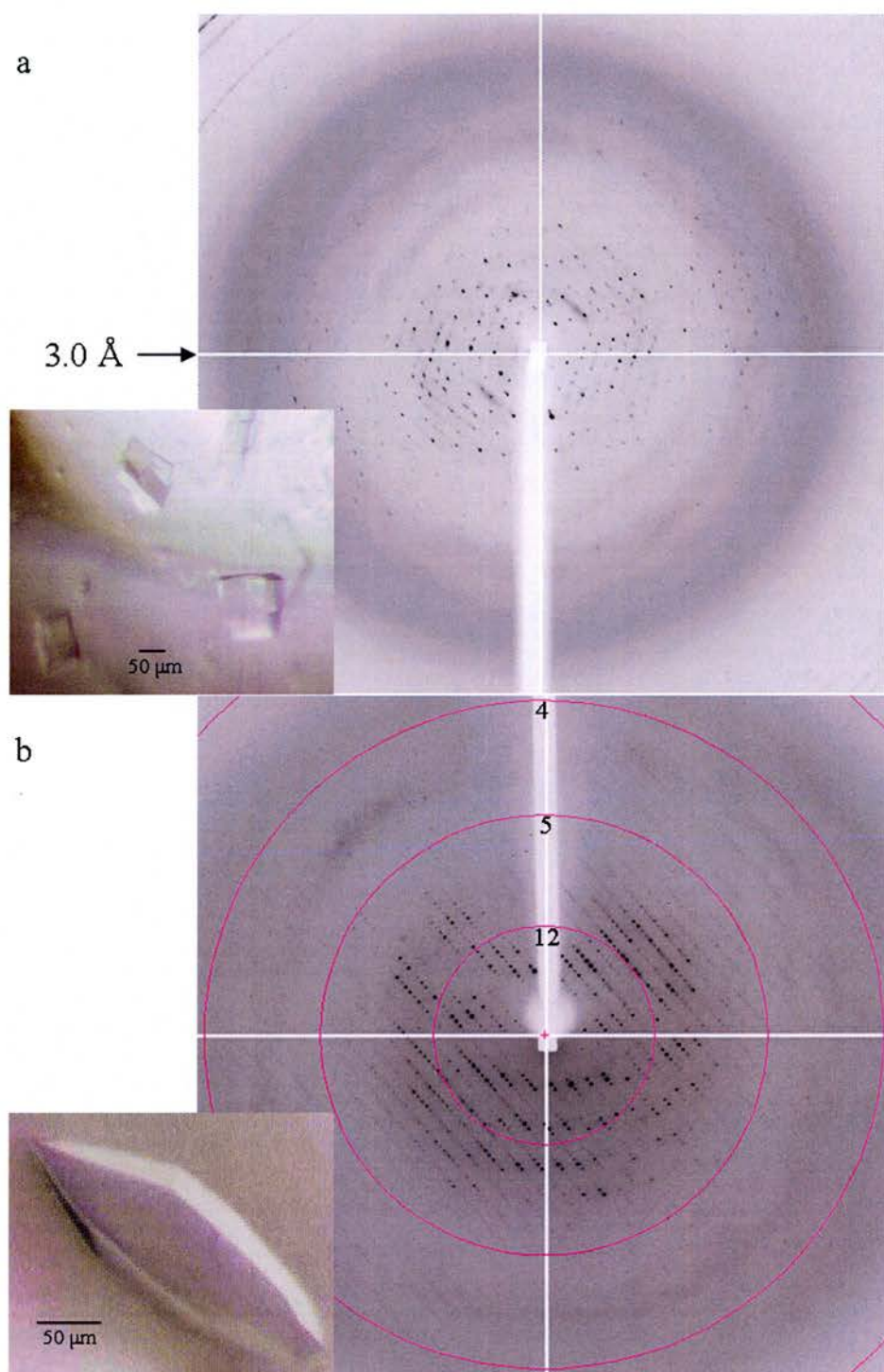
Figure 4.2.7: Anisotropic diffraction of PCNA-XPF(HhH)$_2$ crystals at the ESRF (a) and new crystals that, despite their excellent appearance, only diffract to ~5 Å (b).
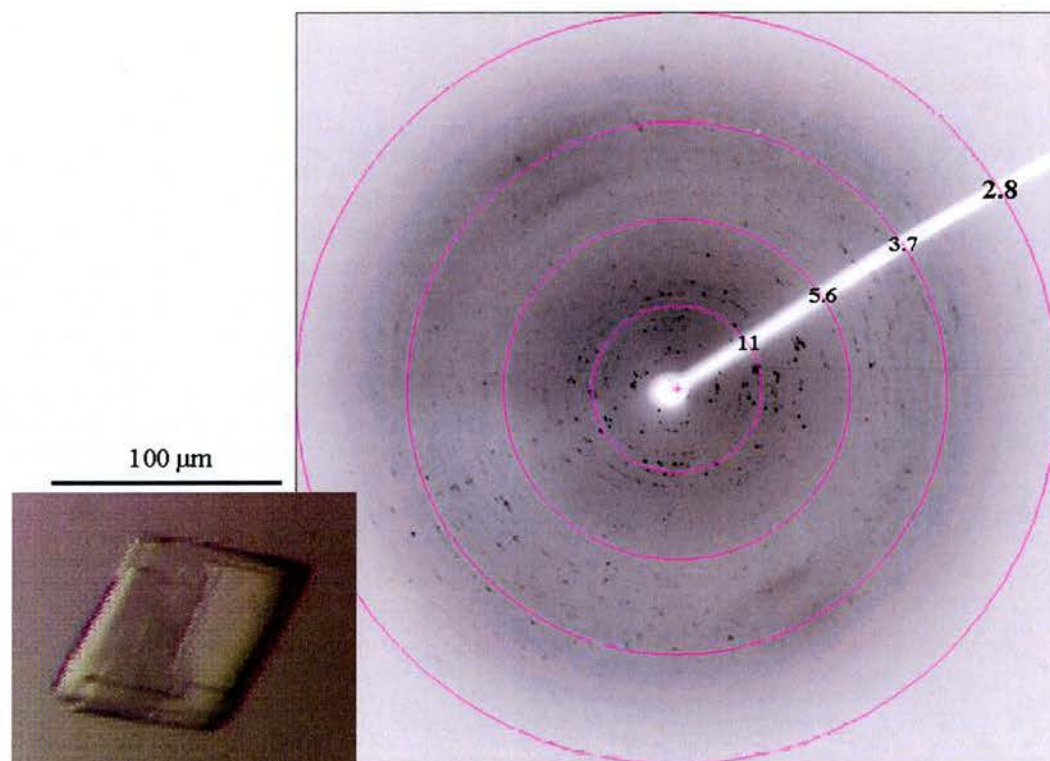
Figure 4.2.8: Recent diffraction of a cluster of crystals formed by the untagged PCNA-XPF(HhH)$_2$ heterotetramer. These crystals diffract in-house to approximately 3 Å at 0° and 90° apart.

4.2.4 Data collection

A large *Sso*PCNA crystal, from an optimised condition containing 0.07 M sodium acetate pH 4.9, was used to collect data to 2.8 Å (fig 4.2.9b) on beamline BM14 at Daresbury light source. MOSFLM[396] identified the space group as belonging to the P222 Laue group, with the unit cell a = 45.1 b = 79.3 c = 235.0, α = β = γ = 90.0°. 360 images were collected with an oscillation angle of 0.5 ° on a CCD image plate using an exposure time of 50 sec. Data were integrated with MOSFLM[396], scaled with SCALA[401] and merged in TRUNCATE, giving the data statistics as shown in table 4.2.10.
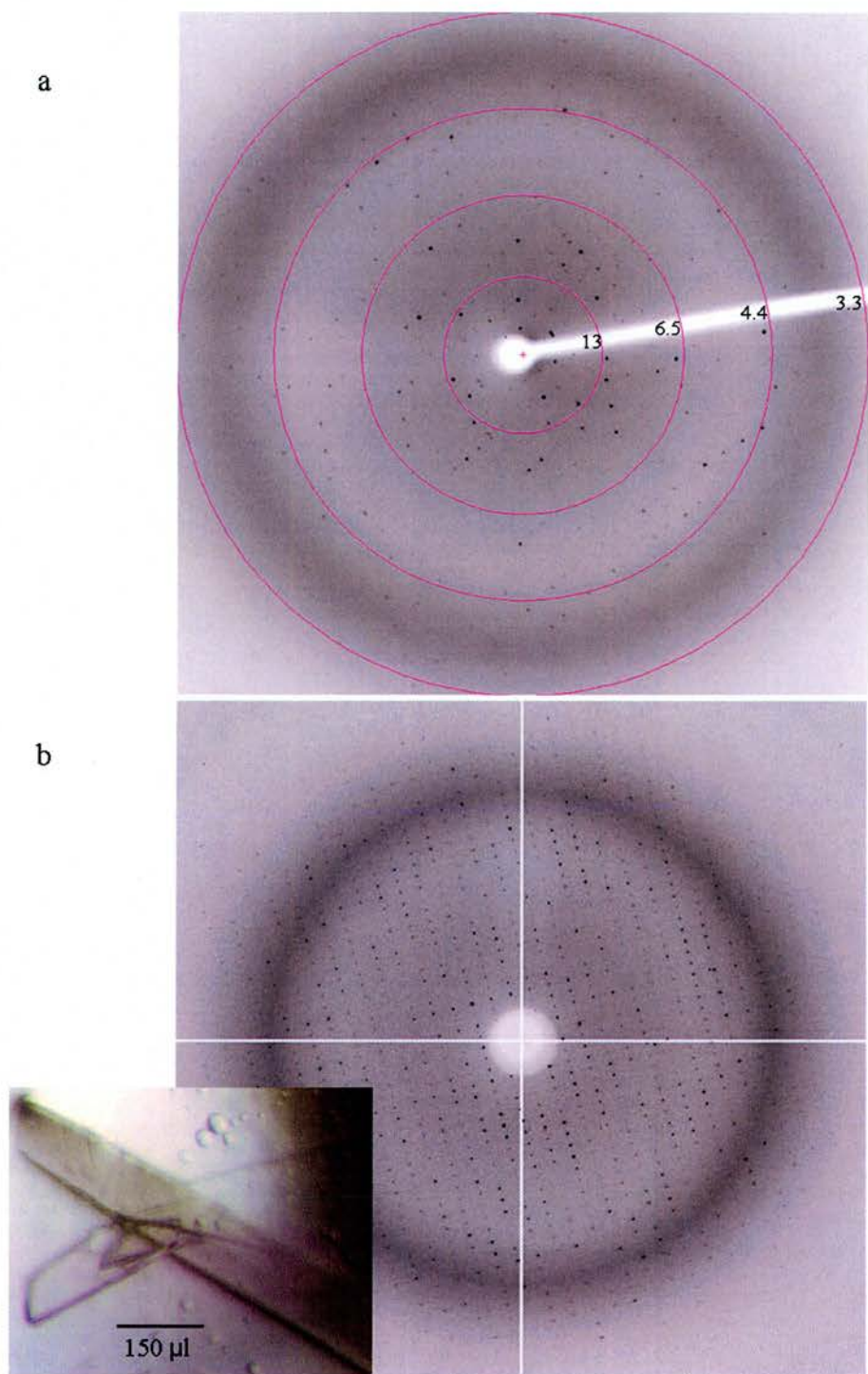
Figure 4.2.9: Good quality diffraction was obtained to 3.3 Å for initial PCNA heterotrimer crystals obtained in 0.05 M sodium acetate and 5 % PEG 2000 (a). Optimised crystals were used to collect data at Daresbury to 2.8 Å at the edge of the plate (b).

Data collection	PCNA Heterotrimer
Resolution (Å)	79.4 - 2.8
Highest shell (Å)	2.85 - 2.8
Wavelength (Å)	0.979
Unit cell	a = 45.1 b = 79.3 c = 235.0 $\alpha = \beta = \gamma = 90.0°$
Space group	$P2_12_12_1$
Unique reflections	22483
Multiplicity	4.9 (5.3)
Completeness (%)	77.8 (77.8)
Rmerge (%)	9.5 (33.5)
I/σ	6.5 (2.3)

Table 4.2.10: Data statistics for the heterotrimeric PCNA crystal, collected at Daresbury light source, beam station BM14.

4.3   RESULTS

4.3.1   Structural solution of PCNA heterotrimer

Figure 4.3.1 shows the sequence alignment of PCNA from *S. cerivisiae* with human

PCNA and *Sso*PCNA. PCNA1, 2 and 3 do no not share high sequence identity with

one another. PCNA1 is 24 % and 22 % identical to PCNA2 and 3 respectively with

PCNA2 sharing 23 % identity to PCNA3. BLAST searches of all three sequences

against structures in the PDB reveal several potential search models. The closest

homologue found was from *S. tokodaii* (PDB code 1UD9), sharing 63 % identity with

PCNA3, 27 % with PCNA2 and 22 % with PCNA1. Although the identity is not very

high for PCNA1 or 2 it was hoped that during molecular replacement, if the PCNA3

orientation was found then this would help the determination of the PCNA1 and 2

positions. Default search parameters in PHASER[647] were employed with the

following modifications. The resolution for the molecular replacement search was cut

to 3 Å and the number of clashes allowed within the packing of the structural

solutions was increased from 0 to 25 due to the low similarity between PCNA1 and

PCNA2 with the search model. The poly-alanine trace of 1UD9 was used as a search

model. The statistics output from PHASER looked very promising (table 4.3.2), with

a Z-score greater than 8 and a log likelihood gain of over 250, strongly suggesting a

solution was found.

```
        1                                                                50
YEAST: MLEAKFEEAS  LFKRIIDGFK  DCVQLVNFQC  KEDGIIAQAV  DDSRVLLVSL
HUMAN: MFEARLVQGS  ILKKVLEALK  DLINEACWDI  SSSGVNLQSM  DSSHVSLVQL
PCNA1: MFKIVYPNAK  DFFSFINSIT  NVTDSIILNF  TEDGIFSRHL  TEDKVLMAIM
PCNA2: MMKAKVIDAV  SFSYILRTVG  DFLSEANFIV  TKEGIRVSGI  DPSRVVFLDI
PCNA3: -MKVVYDDVR  VLKDIIQALA  RLVDEAVLKF  KQDSVELVAL  DRAHISLISV

        51                                                              100
YEAST: EIGVEAFQEY  RC-DHPVTLG  MDLTSLSKIL  RCGNNTDTLT  LIADNTPDSI
HUMAN: TLRSEGFDTY  RC-DRNLAMG  VNLTSMSKIL  KCAGNEDIIT  LRAEDNADTL
PCNA1: RIPKDVLSEY  SI-DSPTSVK  LDVSSVKKIL  SKASSKKA-T  IELTETDSGL
PCNA2: FLPSSYFEGF  EVSQEKEIIG  FKLEDVNDIL  KRVLKDDTLI  L-SS-NESKL
PCNA3: NLPREMFKEY  DVN-DEFKFG  FNTQYLMKIL  KVAKRKEAIE  I-ASESPDSV

        101                                                             150
YEAST: ILLFEDTKKD  RIAEYSLKLM  DIDADFLKIE  ELQYDSTLSL  PSSEFSKIVR
HUMAN: ALVFEAPNQE  KVSDYEMKLM  DLDVEQLGIP  EQEYSCVVKM  PSGEFARICR
PCNA1: KIIIRDEKSG  AKSTIYIKAE  KGQVEQLTEP  KVNLAVNFTT  DESVLNVIAA
PCNA2: TLTFDGEFT-  --RSFELPLI  QVESTQPPSV  NLKFPFKAQL  LTITFADIID
PCNA3: IINITGSTN-  --REFNVRNL  EVSEQEIPEI  NLQFDISATI  SSDGFKSAIS

        151                                                             200
YEAST: DLSQLSDSIN  IMITKETIKF  VADGDIGSGS  VIIKPFVDME  HPETSIKLEM
HUMAN: DLSHIGDAVV  ISCAKDGVKF  SASGELGNGN  IKLSQTSNVD  KEEEAVTIEM
PCNA1: DVTLVGEEMR  ISTEEDKIKI  EA-GEEGKRY  VAF-----LM  KDKPLKELSI
PCNA2: ELSDLGEVLN  IHSKENKLYF  EVIGDLSTAK  VELST-----  DNGTLLEAS-
PCNA3: EVSTVTDNVV  VEGHEDRILI  KAEGES-EVE  VEFSK-----  DTGGLQDLEF

        201                                                             250
YEAST: DQPVDLTFGA  KYLLDIIKG-  SSLIDRVGIR  LSSEAPALFQ  FDLKSGGFLQ
HUMAN: NEPVQLTFAL  RYLNFFTKA-  TPLSSTVTLS  MSADVPLVVE  YKIADMGHLK
PCNA1: DTSASSSYSA  EMFKDAVKGL  RGFSAPTMVS  FGENLPMKID  VEAVSGGHMI
PCNA2: GADVSSSYGM  EYVANTTKM-  RRAFDSMELY  FGSQIPLKLR  FKLPQEGYGD
PCNA3: SKESKNSYSA  EYLDDVLSL-  TKLIDYVKIS  FGNQKPLQLF  FNMEGGGKVT

        251                  211
YEAST: FFLIKFNDE  E
HUMAN: YYLIKIEDE  EGS
PCNA1: FWIIRL
PCNA2: FYIIRAD
PCNA3: YLLIKV
```

Figure 4.3.1: Sequence alignment of *S. cerevisiae* (yeast) PCNA with human PCNA,

PCNA1, PCNA2 and PCNA3. Residues conserved in three out of five and all

sequences are highlighted in yellow and blue respectively. Numbers correspond to

the *S. cerevisiae* PCNA residues. Adapted from Dionne et al.[607].

	Z-score	Log-likelihood gain (LLG)	LLG refined
PCNA3 search			
Fast rotation function	**4.69**	31.01	-
Fast translation function	**9.05**	76.63	*78.2*
PCNA2 search			
Fast rotation function	**3.67**	89.67	-
Fast translation function	**10.42**	147.26	*163.09*
PCNA1 search			
Fast rotation function	**3.53**	172.76	-
Fast translation function	**8.10**	208.40	*265.31*

Table 4.3.2: Output from PHASER. The different sequence identities between the PCNA monomers and search model allowed each subunit to be searched for separately. As judged by manual inspection of the solution PHASER found the positions of PCNA3, followed by PCNA2 and finally PCNA1. The Z-scores (highlighted in bold) for the rotation function demonstrate that PCNA3 is more easily located than PCNA 2 and both are more easily located than PCNA1. The refined log likelihood gain (highlighted in bold and italic) is accumulative.

PHASER output one solution that contains one dimer and one monomer in the asymmetric unit (figure 4.3.3). Generation of the symmetry related molecules in COOT[490] creates a heterotrimer that immediately suggested a correct solution was found (figure 4.3.5). Observation of the electron density maps confirmed this solution as side chains could be identified corresponding to each PCNA monomer.
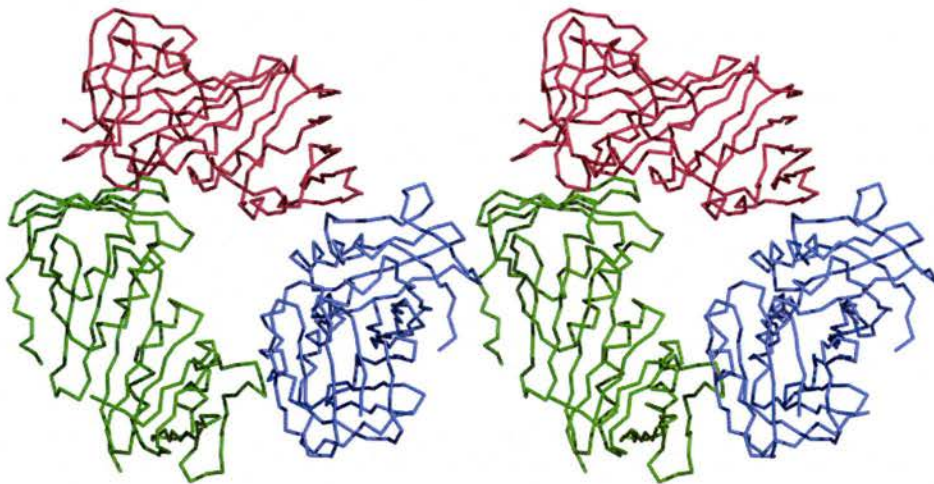
Figure 4.3.4: PCNA subunits found by PHASER in the asymmetric unit with *PCNA*1, 2 and 3 coloured blue, red and green respectively. Generation of the symmetry related objects generates the heterotrimer, shown in figure 4.3.4.

4.3.2   Overall structure

Figure 4.3.3: Superimposition of PCNA1 (yellow), PCNA2 (blue) and PCNA3 (red) shows that their overall structure is very similar.

The electron density map has been initially traced by hand in COOT[490] and O[485] and side chains for PCNA1-3 fitted. Initial rigid body and restrained refinement has been

carried out in REFMAC5[403] to an R_{factor} of 35 % and R_{FREE} of 40 %, however further refinement of the model is needed before detailed analysis of the structure can be performed. The overall fold of the PCNA1-3 monomers are essentially the same as each other (figure 4.3.4). The two domains within each monomer fold to form the typical βαββ fold found in all sliding clamps. PCNA3 superimposes with a RMSD of 1.7 and 1.3 Å between PCNA1 and PCNA2 respectively, with PCNA1 and PCNA2 superimposing with a RMSD of 1.7 Å. The structure formed by heterotrimeric PCNA is very similar to that of other sliding clamps, superimposing well with *S. cerevisiae* PCNA and *E. coli* β-subunit (figure 4.3.5) with RMSD of 2.5 and 3.2 Å respectively.
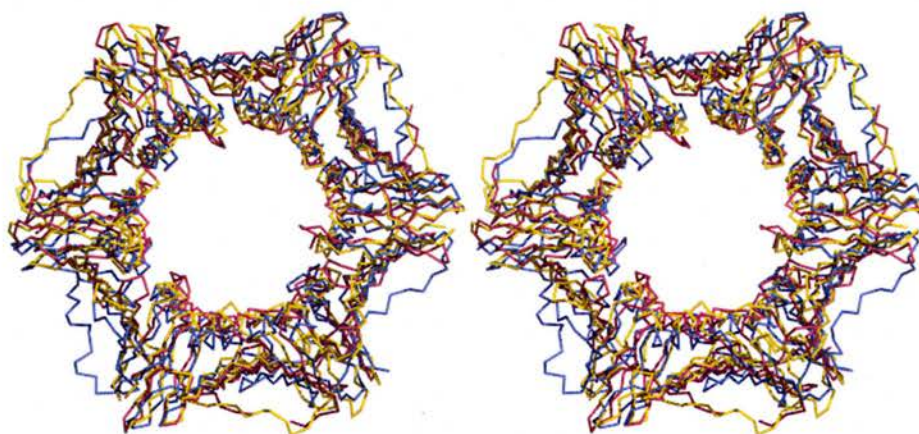
Figure 4.3.4: Superimposition of *Sso*PCNA (red) with *S. cerevisiaea* PCNA (blue) and the β-subunit (yellow). The superimposition shows that, despite the heterotrimeric nature of *Sso*PCNA, the overall toroidal structure formed is essentially the same as other clamps.

Tracing the electron density reveals some areas that have poorly defined electron density. These areas map to the interdomain linker of each PCNA monomer, suggesting, along with several other PCNA structures that also lack density for these

loops (PDB code 1UD9), the flexibility of this region. It is possible that as the refinement progresses these areas may become traceable.

4.3.3 Sequence alignments with the 9-1-1 complex

Venclovas and Thelen carried out structure-based predictions of the Rad1, Rad9 and Hus1 proteins and found they were likely to contain a PCNA-like fold[621]. They performed a complex sequence alignment of these proteins with PCNA from a number of eukaryotes[621]. Based on this alignment and that for PCNA1-3 with *S. cerevisiae* PCNA (figure 4.3.1) allows PCNA1-3 to be tentatively aligned with Rad1 (figure 4.3.3a), Hus1 (figure 4.3.3b) and Rad9 (figure 4.3.3c) to see if there is any significant sequence similarity between these proteins.

These sequence alignments reveal that Rad1, Hus1 and Rad9 protein sequences only share around 8–12/10-15 % identity/similarity to *Sso*PCNA. PCNA1 is 8.8/13.3, 9.2/13.3 and 11.2/14.9 % identical to Rad1, Hus1 and Rad9 respectively. PCNA2 is 11.4/13.8, 12.2/12.2 and 7.7/13.4 % identical to Rad1, Hus1 and Rad9 respectively. PCNA3 is 10.4/13.1, 10.0/11.6 and 11.6/10.0 % identical to Rad1, Hus1 and Rad9 respectively. This tentatively suggests that PCNA2 is most closely related to Hus1 and Rad1, PCNA1 to Rad9 and PCNA3 shares similar identity / similarity to all three proteins.

```
RAD1 : -KFSASTVHLE HITTALSCLT PFGSKDDVLI FIDADGLSFV REN-NHVIKI
PCNA1: MFKIVYPNAK DFFSFINST NVTD--SIIL NFETEDGIFSR HLTEDKVLMA
PCNA2: MMKAKVIDAV SFSYILRTVG DELS--EANF IVTKEGIRVS GIDPSRVVFL
PCNA3: -MKVVYDDVR VLKDIIQA-A RLVD--EAVL KFKQISVEL ALDRAHISL

RAD1 : QLLSRELFM SYSYRNETE- DHMKLCVKIN HILDSVSVMN ------RNSD
PCNA1: IMRIPKDVLS EYSI-----D SPTSVKLDVS SVKKILSK-- --------AS
PCNA2: DIFLPSSYFE GFEVS----Q EKEIIGFKLE DVNDILKR-- --------VL
PCNA3: SVNLPREM-K EYDVN----- DEFKFGENTQ YLMKILKV- --------AK

RAD1 : DIVECTLSYD GHGSPFVLIF ED--SFISER VEYSTYLIKD FDTNGLELDR
PCNA1: SKKA-TIELT ETDSGLKIII RDEKSGAKST IYIKAEKGQV EQ-LTEPKVN
PCNA2: KDDTLIL-SS -NESKLTLTF DGEFT---RS FELPLIQVES TQ-PPSVNLK
PCNA3: RKEAIEI-AS ESPDSVIINI TGSTN---RE ENVRNLEVSE QE-IPEINLQ

RAD1 : ERISFEAIIK GEALHSALKD LKEIGCKECY VYAKTEANDE NVFALISKSQ
PCNA1: --LAVNFTTD ESVLNVIAAD VTLVG-EEMR ISTEE----- DKIKIEA-GE
Pcna2: --FPFKAQLL TITFADIIDE LSDLG-EVLN IHSKE----- NKLYFEVIGD
Pcna3: --FDIS-T S SDGFK-AISE VSTVT-DNV VEGHE----- DRILIKAEGE

RAD1 : LGFSKIKLPS NRSY--LEKL QVFDGDSTTV IDGFAVIGFF DFTSFDKIRK
PCNA1: EGKRYVAF-- ---LMKDPL KELSID----- --TSASSSY SAEMFKDAVK
PCNA2: LSTAKVELST -----DNGTL LEAS-G----- ---ADVSSSY GMEYVANTTK
PCNA3: S-EVEVEFSK -----DTGGL QDLEFS---- ---KESKNSY SAEYLDDVLS

RAD1 : S-TKIASKVL FRMDVHGVLS VNILSQTDDV I(20)PKDYP G-IVIEVCML
PCNA1: GLRGFSAPTM VSFGENLPMK IDVEAV----- --------- SGGHMIFWIA
PCNA2: M-RRASDSME LYFGSQIPLK LRFKLP----- --------- QEGYGDFYIA
PCNA3: L-IPLSDY K ISFGNQKP Q LFFNME----- --------- -GGKVTYLLA

RAD1 : EKESID(83)-
PCNA1: PRL
PCNA2: PRAD
PCNA3: PKV
```

Figure 4.3.3a

```
HUS1 : MKLKLIVNGC EAPDYKLLR TTINTVASLR KTAILRRNSE RLTIISTPK(13
PCNA1: -MFKIVYP-- -NAKDFSFEI NSITNVTD-- -SIILNFETED GIFSRHLTE
PCNA2: -MMKAKVI-- -DAVSFYIL RTVGDELS-- -EANFIVTKE GIRVSGIDP
PCNA3: --M-VVYD-- -DVRVL-DII QALARLVD-- -EAVIK-KQD SVELVALDR

HUS1 : GDTGQLWCTI PHD-VFRLYT VISARELNTI TMECNCDSLL SVFKRY(7)S
PCNA1: DKVLMAIMRI PKDV-LSEYS ID----SPT SVKLDVSSVK KILSKA--SS
PCNA2: SRVVFLDIFL PSSY-FEGFE VSQ----EKE IIGFKLEDVN DILKRV--LK
PCNA3: AHISLISVNL PREM-TKEYD VN-----DEF KFGF-TQYLM KILAVA--KR

HUS1 : SSNMIKLQS M(25)PICAL GITFEEIVH(32)NKVIMHSF KVPVKLLFRA
PCNA1: KKA-TIELTE -----TDSGL KIIIRDE---- --KSGAKS TIYIKAEKGQ
PCNA2: DDTLIL-SS- ----NESKL TLTFDGEFT---- ----R SFELPLIQVE
PCNA3: KEAIE-ASE S-----PDSV I NITGSTN---- ----R EFN-RN-EVS

HUS1 : QDTRIQEPMI NYIQLMMY(12)FHGFIRRVE RYSNVNHIHL MGVKK(8)DV
PCNA1: VEQLTEPKVN -LAVNFTTD--ESVLNVIAA DVTLVG-EEM RISTE-EDKI
Pcna2: STQPPSVNLK --FPFKAQLL--TITFADIID ELSDLG-EVL NIHSK-ENKL
Pcna3: EQEIPEINLQ -FDISATIS---SDGFKSAIS EV-TVT-DNV VEGH-EDRI

HUS1 : ELKIIVNKLD WHGEICWNGP LDSVIQRQEG LTDNPSQN(86)STHEVIIRC
PCNA1: KIEA-GEEGK RYVAF---- ---LMKDPL KELSID---- --TSASSSYSA
PCNA2: YFEVIGDLST AKVELST--- -----DNGTL LEAS----- --GADVSSSYGM
PCNA3: LIVAEGES-E VEV-FSK--- -----DTGGL QDLEFS--- ---KESKNSYSA

HUS1 : KDWKVCSKL YAAFEVVLA ISHDESCVFH CSLDRGSL(8)PRER-GQIIY
PCNA1: EMFKDAVKGL RGFSAPTMVS FGENLPMKID VEAV----- ---SGGHMIF
PCNA2: EYVANTTKM- RRASDSMELY FGSQIPLKLR FKLP----- ----QEGYGDF
PCNA3: EYLDDVLSI- TKLSDYKIS FGNQKPLQLF NME------- ----GG-KVT

HUS1 : YIARSKGL
PCNA1: W
PCNA2: Y
PCNA3: L
```

Figure 4.3.3b

```
RAD9 : MSFKATITES GKQNIWFRAI YVLSTIQDDI KITVTTNELI AWSMNETDTT
PCNA1: ---MFKIVYP N-AKDFFSFI NSITNVTDSI ILNFTEDGIF SRHLTEDKVL
PCNA2: ---MMKAKVI D-AVSFSYIL RTVGDFLSEA NFIVTKEGIR VSGIDPSRVV
PCNA3: ----MKVVYD D-VRVLKDII QALARLVDEA VLKFKQDSVE LVALDRAHIS

RAD9 : LCQVRFQKSF FEEYEFKPHE (15)GNSHKL YSFRVNGRHL TTISRKPDGD
PCNA1: MAIMRIPKDV LSEYSID--- --------SP TSVKLDVSSV KKILSKASS-
PCNA2: FLDIFLPSSY FEGFEVSQ-- --------EK EIIGFKLEDV NDILKRVLK-
PCNA3: LISVNLPREM FKEYDVN--- --------DE FKFGFNTQYL MKILKVAKR-

RAD9 : GIKSFTIAVN NTSTCPESLA NRLIVVIEMD SLIV(12)DP IIINLKY-KR
PCNA1: ---------- ---------- -----KKA-T IELTET-DSG LKIIIRDEKS
PCNA2: ---------- ---------- -----DDTLI L-SS--NESK LTLTFDGEFT
PCNA3: ---------- ---------- -----KEAIE I-ASES-PDS VIINITGSTN

RAD9 : RFLDVFGTAA S(23)TSALF NEE(16)INY ICCNSTLLKN FLDNCVNVT
PCNA1: GAKSTIYIKA EKGQVEQLTE PKVN--LAVN FTTDESVLNV IAADVTL-VG
Pcna2: ---RSFELPL IQVESTQPPS VNLK--FPFK AQLLTITFAD IIDELSD-LG
Pcna3: ---REFNVRN LEVSEQEIPE INLQ--FDIS ATISSDGFKS AISEVST-VT

RAD9 : DEVKLEINVH RLSITAFTKA VYGKNNDLLR NALSMSNTIS T-----LDLE
PCNA1: EEMRISTEED KIKIEA-GE- ---------E GKRYVAF--- --LMKDKPLK
PCNA2: EVLNIHSKEN KLYFEVIGD- ---------L STAKVELST- ----DNGTLL
PCNA3: DNVVVEGHED RILIKAEGES ---------- -EVEVEFSK- ----DTGGLQ

RAD9 : HYCL(22)SI IFKLKDFKNF ITI-GP(9)D NISLWFCHPG DPILMQMQKP
PCNA1: ELSIDTSASS SYSAEMFKDA VKGLRGF-SA PTMVSFGE-N LPMKIDVEAV
PCNA2: EAS-GADVSS SYGMEYVANT TKM-RRA-SD SMELYFGS-Q IPLKLRFKLP
PCNA3: DLEFSKESKN SYSAEYLDDV LSL-TKL-SD YVKISFGN-Q KFLQLFFNME

              RAD9 : GVK(11)NDD I-LEGKFIKT A(218)
              PCNA1: ---------- SGGHMIFW
              PCNA2: ---------- QEGYGDFY
              PCNA3: ---------- GGGKVTYL
```

Figure 4.3.5: Sequence alignment of Rad1 (a), Hus1 (b) and Rad9 (c) with PCNA1-3 Residues that are identical to Rad1, Hus1 or Rad9 in PCNA1, 2 and 3 are coloured red, green and yellow respectively. Those that are coloured blue are similar.

4.4   Discussion

4.4.1   *Sso*PCNA

Sliding clamps have been the focus of intense structural interest due to their involvement in processive DNA replication and unique ability to non-specifically encircle duplex DNA and freely slide in either direction along it. The structures of clamps involved in replication, including the β-subunit from Gram–negative[517] and Gram–positive bacteria[648], gene45 from bacteriophage[524] and PCNA from archaea[523] and eukaryotes[522], all form a ring-shaped structure with a central cavity that is thought to encircle DNA. The majority of these clamps are homodimeric (bacteria) or homotrimeric (bacteriophage, archaea and eukaryotes), however nature has also evolved heterotrimeric sliding clamps. More than one homologue of PCNA exists in some crenarchaea, which have been shown to form a heterotrimer that acts as the sliding clamp for their replicative polymerase[607]. A heterotrimeric sliding clamp with a similar overall structure to PCNA is also thought to exist in eukaryotes[621], made up of the Rad1, Hus1 and Rad9 proteins and called the 9-1-1 complex (4.1.6). This sliding clamp is unique in that it is not an accessory protein for DNA replication, instead it is involved in initial DNA damage response events[626]. Presented here is the partially refined 2.8 Å structure of the first heterotrimeric sliding clamp, the PCNA heterotrimer from *S. solfataricus*. This structure superimposes well with other sliding clamps (figure 4.3.5), suggesting that the topological nature of encircling DNA is the essential feature of all clamps.

The *Sso*PCNA structure provides the initial basis of an interesting structural biology study. It demonstrates that this heterotrimer is amenable to crystallographic

studies and the fully refined structure will give further insights into this sliding clamp family of proteins. It is also highly desirable to optimise the SsoPCNA crystals further to collect data at a higher resolution, as this structure will act as a model for future structural studies. As the 2.8 Å structure presented here is not fully refined, a detailed structural analysis of SsoPCNA is not carried out. Instead this discussion focuses on the directions this project may take.

The two most interesting features of the heterotrimeric clamp are the interactions between individual PCNA protomers at the three dimer interfaces and the subunit specific interactions that are important for its cellular function. The binding affinities of the individual PCNA monomers as well as their specific interaction with RFC suggest that PCNA3 is pulled apart from the PCNA1-PCNA2 heterodimer during clamp loading (figure 4.1.5). This is likely to be reflected in the structure, with the PCNA3-PCNA2 and PCNA3-PCNA1 interfaces presumably weaker than the PCNA1-PCNA2 interface. The fully refined SsoPCNA structure will allow this to be assessed in detail and will provide insight into the clamp loading of this sliding clamp. This may have further implications for the heterotrimeric sliding clamp in eukaryotes. It would be interesting to calculate the affinities of the binding of each component within the 9-1-1 complex as, although some groups observe dimerisation of each component with the other two[622], this is not always seen. It has been reported that Hus1 and Rad1 from S. pombe do not interact with one another in the absence of Rad9[649,650]. This suggests that the components of the 9-1-1 complex interact with one another with different affinities, which is likely to have an effect on the clamp loading of this complex, reminiscent of SsoPCNA.

4.1.2 ***Sso*PCNA-XPF**

The progress made towards the structural solution of the *Sso*PCNA-XPF(HhH)$_2$ complex suggests that the interactions of this sliding clamp with cellular partners are also accessible to X-ray crystallography. Such subunit-specific interactions are also essential to the 9-1-1 complex[626] and PCNA's function in at least some of its cellular roles (4.1.2), for example the bridging of the CDK2-cyclinA complex with its substrate[577]. The *Sso*PCNA-XPF(HhH)$_2$ structure would provide the first insights into subunit-specific interactions made by a sliding clamp and would provide a useful model for understanding these interactions in eukaryotes, particularly for the 9-1-1 complex.

There are also several other co-complexes that would be desirable to help characterise the subunit specific interactions made by *Sso*PCNA. Firstly, peptides could be designed based on the PCNA interacting motifs of PolB, FEN-1 and DNA ligase from *S. solfataricus* that will specifically bind PCNA2, PCNA1 and PCNA3 respectively. The structural solution of these peptides in complex with *Sso*PCNA would provide insights into the Okazaki fragment maturation machine thought to form during replication of *S. solfataricus*[616]. A more ambitious aim would be the elucidation of this okazakaki fragment maturation machine in its entirety. That *Sso*PCNA forms a hexameric complex with PolB, FEN1 and DNA ligase[616] suggests that this may be a possibility.

Although structures of archaeal XPF in complex with DNA[634] and the ERCC1(HhH)$_2$-XPF(HhH)$_2$[643] (4.1.9) have provided insights into the substrate recognition of the XPF family of endonucleases, they are not conclusive. The solution of the SsoPCNA-XPF(HhH)$_2$ structure will allow accurate modelling of

*Sso*PCNA's interaction with both apo- and DNA complexed XPF. This will either add confidence to the model of substrate recognition hypothesised by Newmann *et al.*[634] or suggest an alternative mechanism. As *Sso*XPF is active against both XPF-ERCC1[601] and Mus81[619] like DNA substrates the SsoPCNA-XPF(HhH)$_2$ structure will act as an excellent model, providing insights into substrate recognition by this family of enzymes. Further structural work of the *Sso*PCNA-XPF interaction would also be interesting. Most importantly would be the solution of *Sso*PCNA with full length XPF and this complex with a variety of DNA substrates. These structures would further characterise the substrate specificities of this class of enzymes.

References

1   Alberts, B.M. (1987) Prokaryotic DNA replication mechanisms. *Philos Trans R Soc Lond B Biol Sci* 317 (1187), 395-420

2   Morris, A. et al. (1998) The superbugs: evolution, dissemination and fitness. *Curr Opin Microbiol* 1 (5), 524-529

3   G., M. et al. (2000) Staphylococcus aureus (including toxic shock syndrome). In: Principles and practice of infectious diseases. Ed. Waldvogel. *Philadelphia: Churchill Livingstone* 5th Ed, 2065-2092

4   Fleming, A. (1929) On the antibacterial actions of a Penicillium, with special reference to their use in the isolation of *B. influenza*. *Br. J. Exp. Path.* 10, 226-236

5   Chain, E. et al. (1993) Penicillin as a chemotherapeutic agent. 1940. *Clin Orthop Relat Res* (295), 3-7

6   Chain, E. et al. (1940) *Lancet* II

7   Hancock, R.E. and Knowles, D. (1998) Are we approaching the end of the antibiotic era? *Curr Opin. in Microbiol* 1, 493-494

8   Nathan, C. (2004) Antibiotics at the crossroads. *Nature* 431 (7011), 899-902

9   Brisson-Noel, A. et al. (1988) Mechanism of action of spiramycin and other macrolides. *Antimicrob. Chemother.* 22, 13-23

10 Fourmy, D. et al. (1996) Structure of the A site of Escherichia coli 16S ribosomal RNA complexed with an aminoglycoside antibiotic. *Science* 274 (5291), 1367-1371

11 Chopra, I. (1985) The tetracyclines, handbook of experimental pharmacology (Eds Hlavaka, J. J. and Boothe, J. H. *Springer Berlin, 1985* 78

12 Walsh, C. (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature* 406 (6797), 775-781

13 Bogdanov, A.A. et al. (1995) Structure and function of 5S rRNA in the ribosome. *Biochem Cell Biol* 73 (11-12), 869-876

14 Auerbach, T. et al. (2002) Antibiotics targeting ribosomes: crystallographic studies. *Curr Drug Targets Infect Disord* 2 (2), 169-186

15 Shen, L.L. (1992) In: Quinolone antibacterial agents (Eds Hooper, D. C. and Wolfson, J. S. *American society for Microbiology, Washington DB, 1992* 77 (2nd Edn)

16 Wigley, D.B. (1995) Structure and mechanism of DNA topoisomerases. *Annu Rev Biophys Biomol Struct* 24, 185-208

17 Nakagawa, J. et al. (1984) Functional biosynthesis of cell wall peptidoglycan by polymorphic bifunctional polypeptides. Penicillin-binding protein 1Bs of Escherichia coli with activities of transglycosylase and transpeptidase. *J Biol Chem* 259 (22), 13937-13946

18 Spratt, B.G. and Cromie, K.D. (1988) Penicillin-binding proteins of gram-negative bacteria. *Rev Infect Dis* 10 (4), 699-711

19 Arthur, M. and Courvalin, P. (1993) Genetics and mechanisms of glycopeptide resistance in enterococci. *Antimicrob Agents Chemother* 37 (8), 1563-1571

20 Walsh, C.T. et al. (1996) Bacterial resistance to vancomycin: five genes and one missing hydrogen bond tell the story. *Chem Biol* 3 (1), 21-28

21 Paulsen, I.T. et al. (1996) Proton-dependent multidrug efflux systems. *Microbiol Rev* 60 (4), 575-608

22 Borges-Walmsley, M.I. and Walmsley, A.R. (2001) The structure and function of drug pumps. *Trends Microbiol* 9 (2), 71-79

23 Levy, S.B. (1992) Active efflux mechanisms for antimicrobial resistance. *Antimicrob Agents Chemother* 36 (4), 695-703

24 Phillippon, A. et al. (1985) Extended spectrum beta-lactamases. *Atimicrob. Agents Chemother* 28 (302-307)

25 Shaw, K.J. et al. (1993) Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes. *Microbiol Rev* 57 (1), 138-163

26 Krulwich, T.A. et al. (2001) Functions of tetracycline efflux proteins that do not involve tetracycline. *J Mol Microbiol Biotechnol* 3 (2), 237-246

27 Ross, J.I. et al. (1990) Inducible erythromycin resistance in staphylococci is encoded by a member of the ATP-binding transport super-gene family. *Mol Microbiol* 4 (7), 1207-1214

28 Bussiere, D.E. et al. (1998) Crystal structure of ErmC', an rRNA methyltransferase which mediates antibiotic resistance in bacteria. *Biochemistry* 37 (20), 7103-7112

29 Spratt, B.G. (1994) Resistance to antibiotics mediated by target alterations. *Science* 264 (5157), 388-393

30 Song, M.D. et al. (1987) Evolution of an inducible penicillin-target protein in methicillin-resistant Staphylococcus aureus by gene fusion. *FEBS Lett* 221 (1), 167-171

31 Chu, D.T. et al. (1996) New directions in antibacterial research. *J Med Chem* 39 (20), 3853-3874

32 Levy, S.B. (1998) The challenge of antibiotic resistance. *Sci Am* 278 (3), 46-53

33 Davies, J. (1994) Inactivation of antibiotics and the dissemination of resistance genes. *Science* 264 (5157), 375-382

34 Davies, J. (1996) Bacteria on the rampage. *Nature* 383 (6597), 219-220

35 Menichetti, F. (2005) Current and emerging serious Gram-positive infections. *Clin Microbiol Infect* 11 Suppl 3, 22-28

36 Leclercq, R. et al. (1988) Plasmid-mediated resistance to vancomycin and teicoplanin in Enterococcus faecium. *N Engl J Med* 319 (3), 157-161

37 Uttley, A.H. et al. (1989) High-level vancomycin-resistant enterococci causing hospital infections. *Epidemiol Infect* 103 (1), 173-181

38 Paterson, D.L. (1999) Reduced susceptibility of Staphylococcus aureus to vancomycin--a review of current knowledge. *Commun Dis Intell* 23 (3), 69-73

39 Dye, C. et al. (2002) Erasing the world's slow stain: strategies to beat multidrug-resistant tuberculosis. *Science* 295 (5562), 2042-2046

40 Abraham, E.P. and Chain, E. (1940) An enzyme from bacteria able to destroy Penicillin. *Nature* 146 (837)

41 Projan, S.J. (2003) Why is big Pharma getting out of antibacterial drug discovery? *Curr Opin Microbiol* 6 (5), 427-430

42 Coates, A. et al. (2002) The future challenges facing the development of new antimicrobial drugs. *Nat Rev Drug Discov* 1 (11), 895-910

43 Chandonia, J.M. and Brenner, S.E. (2006) The impact of structural genomics: expectations and outcomes. *Science* 311 (5759), 347-351

44 Waszkowycz, B. (2002) Structure-based approaches to drug design and virtual screening. *Curr Opin Drug Discov Devel* 5 (3), 407-413

45   Hancock, R.E. (2001) Cationic peptides: effectors in innate immunity and novel antimicrobials. *Lancet Infect Dis* 1 (3), 156-164

46   Zasloff, M. (2002) Antimicrobial peptides of multicellular organisms. *Nature* 415 (6870), 389-395

47   Blattner, F.R. et al. (1997) The complete genome sequence of Escherichia coli K-12. *Science* 277 (5331), 1453-1474

48   Goffeau, A. et al. (1996) Life with 6000 genes. *Science* 274 (5287), 546, 563-547

49   (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. *Science* 282 (5396), 2012-2018

50   Adams, M.D. et al. (2000) The genome sequence of Drosophila melanogaster. *Science* 287 (5461), 2185-2195

51   Waterston, R.H. et al. (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420 (6915), 520-562

52   Lander, E.S. et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409 (6822), 860-921

53   Venter, J.C. et al. (2001) The sequence of the human genome. *Science* 291 (5507), 1304-1351

54   Miles, E.W. et al. (1999) The molecular basis of substrate channeling. *J Biol Chem* 274 (18), 12193-12196

55   Husi, H. et al. (2000) Proteomic analysis of NMDA receptor-adhesion protein signaling complexes. *Nat Neurosci* 3 (7), 661-669

56   Wijnhoven, B.P. et al. (2000) E-cadherin-catenin cell-cell adhesion complex and human cancer. *Br J Surg* 87 (8), 992-1005

57   Houtman, J.C. et al. (2005) Examining multiprotein signaling complexes from all angles. *Febs J* 272 (21), 5426-5435

58   Rout, M.P. and Aitchison, J.D. (2000) Pore relations: nuclear pore complexes and nucleocytoplasmic exchange. *Essays Biochem* 36, 75-88

59   Srere, P.A. (1987) Complexes of sequential metabolic enzymes. *Annu Rev Biochem* 56, 89-124

60   Welch, G.R. and Gaertner, F.H. (1980) Enzyme organization in the polyaromatic-biosynthetic pathway: the arom conjugate and other multienzyme systems. *Curr Top Cell Regul* 16, 113-162

61   Verma, R. et al. (2000) Proteasomal proteomics: identification of nucleotide-sensitive proteasome-interacting proteins by mass spectrometric analysis of affinity-purified proteasomes. *Mol Biol Cell* 11 (10), 3425-3439

62   Christmann, M. et al. (2003) Mechanisms of human DNA repair: an update. *Toxicology* 193 (1-2), 3-34

63   Cramer, P. et al. (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292 (5523), 1863-1876

64   Yusupov, M.M. et al. (2001) Crystal structure of the ribosome at 5.5 A resolution. *Science* 292 (5518), 883-896

65   Neubauer, G. et al. (1998) Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet* 20 (1), 46-50

66   Chothia, C. (1992) Proteins. One thousand families for the molecular biologist. *Nature* 357 (6379), 543-544

67   Andreeva, A. et al. (2004) SCOP database in 2004: refinements integrate structure and sequence family data. *Nucleic Acids Res* 32 (Database issue), D226-229

68   Aloy, P. and Russell, R.B. (2004) Ten thousand interactions for the molecular biologist. *Nat Biotechnol* 22 (10), 1317-1321

69   Sali, A. et al. (2003) From words to literature in structural proteomics. *Nature* 422 (6928), 216-225

70   Zimmerman, S.B. and Trach, S.O. (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of Escherichia coli. *J Mol Biol* 222 (3), 599-620

71   Sear, S. (2004) The cytoplasm of living cells: a functional mixture of thousands of components. *J. Phys: Condens. Matter* 17 (S3587-S3595)

72   Minton, A.P. et al. (1992) Model for the role of macromolecular crowding in regulation of cellular volume. *Proc Natl Acad Sci U S A* 89 (21), 10504-10506

73   Ellis, R.J. (2001) Macromolecular crowding: obvious but underappreciated. *Trends Biochem Sci* 26 (10), 597-604

74   Minton, A.P. (2000) Implications of macromolecular crowding for protein assembly. *Curr Opin Struct Biol* 10 (1), 34-39

75   Minton, A.P. (2001) The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J Biol Chem* 276 (14), 10577-10580

76   Zimmerman, S.B. (1993) Macromolecular crowding effects on macromolecular interactions: some implications for genome structure and function. *Biochim Biophys Acta* 1216 (2), 175-185

77   Ovadi, J. and Saks, V. (2004) On the origin of intracellular compartmentation and organized metabolic systems. *Mol Cell Biochem* 256-257 (1-2), 5-12

78   Zimmerman, S.B. and Minton, A.P. (1993) Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu Rev Biophys Biomol Struct* 22, 27-65

79   Murphy, L.D. and Zimmerman, S.B. (1995) Condensation and cohesion of lambda DNA in cell extracts and other media: implications for the structure and function of DNA in prokaryotes. *Biophys Chem* 57 (1), 71-92

80   Morgunov, I. and Srere, P.A. (1998) Interaction between citrate synthase and malate dehydrogenase. Substrate channeling of oxaloacetate. *J Biol Chem* 273 (45), 29540-29544

81   Hall, D. and Minton, A.P. (2003) Macromolecular crowding: qualitative and semiquantitative successes, quantitative challenges. *Biochim Biophys Acta* 1649 (2), 127-139

82   Phizicky, E.M. and Fields, S. (1995) Protein-protein interactions: methods for detection and analysis. *Microbiol Rev* 59 (1), 94-123

83   Prelich, G. et al. (1987) Functional identity of proliferating cell nuclear antigen and a DNA polymerase-delta auxiliary protein. *Nature* 326 (6112), 517-520

84   Brown, K. et al. (1993) Mutual regulation of the transcriptional activator NF-kappa B and its inhibitor, I kappa B-alpha. *Proc Natl Acad Sci U S A* 90 (6), 2532-2536

85   Tansey, W.P. (2001) Transcriptional activation: risky business. *Genes Dev* 15 (9), 1045-1050

86   Cleaver, J.E. et al. (2001) Nucleotide excision repair "a legacy of creativity". *Mutat Res* 485 (1), 23-36

87   Ito, T. et al. (2002) Roles for the two-hybrid system in exploration of the yeast protein interactome. *Mol Cell Proteomics* 1 (8), 561-566

88   Gavin, A.C. et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415 (6868), 141-147

89   Giot, L. et al. (2003) A protein interaction map of Drosophila melanogaster. *Science* 302 (5651), 1727-1736

90   Formstecher, E. et al. (2005) Protein interaction mapping: a Drosophila case study. *Genome Res* 15 (3), 376-384

91   Ames, B.N. et al. (1995) The causes and prevention of cancer. *Proc Natl Acad Sci U S A* 92 (12), 5258-5265

92   Bertram, J.S. (2000) The molecular biology of cancer. *Mol Aspects Med* 21 (6), 167-223

93   Hoeijmakers, J.H. (2001) Genome maintenance mechanisms for preventing cancer. *Nature* 411 (6835), 366-374

94   Friedberg, E.C. (2001) How nucleotide excision repair protects against cancer. *Nat Rev Cancer* 1 (1), 22-33

95   Thompson, L.H. and Schild, D. (2002) Recombinational DNA repair and human disease. *Mutat Res* 509 (1-2), 49-78

96   Critchlow, S.E. and Jackson, S.P. (1998) DNA end-joining: from yeast to man. *Trends Biochem Sci* 23 (10), 394-398

97   Barnes, D.E. (2001) Non-homologous end joining as a mechanism of DNA repair. *Curr Biol* 11 (12), R455-457

98   Modrich, P. and Lahue, R. (1996) Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu Rev Biochem* 65, 101-133

99   Umar, A. and Kunkel, T.A. (1996) DNA-replication fidelity, mismatch repair and genome instability in cancer cells. *Eur J Biochem* 238 (2), 297-307

100   Lindahl, T. (2001) Keynote: past, present, and future aspects of base excision repair. *Prog Nucleic Acid Res Mol Biol* 68, xvii-xxx

101   Cleaver, J.E. (2005) Cancer in xeroderma pigmentosum and related disorders of DNA repair. *Nat Rev Cancer* 5 (7), 564-573

102   Jiricny, J. (1998) Eukaryotic mismatch repair: an update. *Mutat Res* 409 (3), 107-121

103   Prakash, S. and Prakash, L. (2000) Nucleotide excision repair in yeast. *Mutat Res* 451 (1-2), 13-24

104   Tornaletti, S. and Hanawalt, P.C. (1999) Effect of DNA lesions on transcription elongation. *Biochimie* 81 (1-2), 139-146

105   Sugasawa, K. et al. (1997) Two human homologs of Rad23 are functionally interchangeable in complex formation and stimulation of XPC repair activity. *Mol Cell Biol* 17 (12), 6924-6931

106   Cleaver, J.E. et al. (1999) A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Hum Mutat* 14 (1), 9-22

107   Andressoo, J.O. and Hoeijmakers, J.H. (2005) Transcription-coupled repair and premature ageing. *Mutat Res* 577 (1-2), 179-194

108   Cleaver, J.E. (1968) Defective repair replication of DNA in xeroderma pigmentosum. *Nature* 218 (142), 652-656

109   Hoeijmakers, J.H. (1993) Nucleotide excision repair I: from E. coli to yeast. *Trends Genet* 9 (5), 173-177

110   Bohr, V.A. et al. (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell* 40 (2), 359-369

111 Mellon, I. et al. (1986) Preferential DNA repair of an active gene in human cells. *Proc Natl Acad Sci U S A* 83 (23), 8878-8882

112 Mellon, I. et al. (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. *Cell* 51 (2), 241-249

113 Sweder, K.S. and Hanawalt, P.C. (1992) Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription. *Proc Natl Acad Sci U S A* 89 (22), 10696-10700

114 Sugasawa, K. et al. (1998) Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell* 2 (2), 223-232

115 Volker, M. et al. (2001) Sequential assembly of the nucleotide excision repair factors in vivo. *Mol Cell* 8 (1), 213-224

116 Kusumoto, R. et al. (2001) Diversity of the damage recognition step in the global genomic nucleotide excision repair in vitro. *Mutat Res* 485 (3), 219-227

117 Patrick, M.H. (1977) Studies on thymine-derived UV photoproducts in DNA--I. Formation and biological role of pyrimidine adducts in DNA. *Photochem Photobiol* 25 (4), 357-372

118 van Hoffen, A. et al. (1995) Transcription-coupled repair removes both cyclobutane pyrimidine dimers and 6-4 photoproducts with equal efficiency and in a sequential way from transcribed DNA in xeroderma pigmentosum group C fibroblasts. *Embo J* 14 (2), 360-367

119 Balajee, A.S. et al. (1999) DNA repair of pyrimidine dimers and 6-4 photoproducts in the ribosomal DNA. *Nucleic Acids Res* 27 (12), 2511-2520

120 Hwang, B.J. and Chu, G. (1993) Purification and characterization of a human protein that binds to damaged DNA. *Biochemistry* 32 (6), 1657-1666

121 Nichols, A.F. et al. (1996) Mutations specific to the xeroderma pigmentosum group E Ddb- phenotype. *J Biol Chem* 271 (40), 24317-24320

122 Rapic-Otrin, V. et al. (2003) True XP group E patients have a defective UV-damaged DNA binding protein complex and mutations in DDB2 which reveal the functional domains of its p48 product. *Hum Mol Genet* 12 (13), 1507-1522

123 Aboussekhra, A. et al. (1995) Mammalian DNA nucleotide excision repair reconstituted with purified protein components. *Cell* 80 (6), 859-868

124 Tang, J.Y. et al. (2000) Xeroderma pigmentosum p48 gene enhances global genomic repair and suppresses UV-induced mutagenesis. *Mol Cell* 5 (4), 737-744

125 Fitch, M.E. et al. (2003) In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product. *J Biol Chem* 278 (47), 46906-46910

126 Shivji, M.K. et al. (1995) Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA. *Biochemistry* 34 (15), 5011-5017

127 Winkler, G.S. et al. (1998) Affinity purification of human DNA repair/transcription factor TFIIH using epitope-tagged xeroderma pigmentosum B protein. *J Biol Chem* 273 (2), 1092-1098

128 Coin, F. et al. (1999) Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. *Embo J* 18 (5), 1357-1366

129   Tanaka, K. et al. (1990) Analysis of a human DNA excision repair gene involved in group A xeroderma pigmentosum and containing a zinc-finger domain. *Nature* 348 (6296), 73-76

130   Missura, M. et al. (2001) Double-check probing of DNA bending and unwinding by XPA-RPA: an architectural function in DNA repair. *Embo J* 20 (13), 3554-3564

131   Asahina, H. et al. (1994) The XPA protein is a zinc metalloprotein with an ability to recognize various kinds of DNA damage. *Mutat Res* 315 (3), 229-237

132   Nocentini, S. et al. (1997) DNA damage recognition by XPA protein promotes efficient recruitment of transcription factor II H. *J Biol Chem* 272 (37), 22991-22994

133   Stigger, E. et al. (1998) Functional analysis of human replication protein A in nucleotide excision repair. *J Biol Chem* 273 (15), 9337-9343

134   Wold, M.S. (1997) Replication protein A: a heterotrimeric, single-stranded DNA-binding protein required for eukaryotic DNA metabolism. *Annu Rev Biochem* 66, 61-92

135   de Laat, W.L. et al. (1998) DNA-binding polarity of human replication protein A positions nucleases in nucleotide excision repair. *Genes Dev* 12 (16), 2598-2609

136   Drapkin, R. et al. (1994) Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* 368 (6473), 769-772

137   Egly, J.M. (2001) The 14th Datta Lecture. TFIIH: from transcription to clinic. *FEBS Lett* 498 (2-3), 124-128

138   Gerard, M. et al. (1991) Purification and interaction properties of the human RNA polymerase B(II) general transcription factor BTF2. *J Biol Chem* 266 (31), 20940-20945

139   Weeda, G. et al. (1990) A presumed DNA helicase encoded by ERCC-3 is involved in the human repair disorders xeroderma pigmentosum and Cockayne's syndrome. *Cell* 62 (4), 777-791

140   Weber, C.A. et al. (1990) ERCC2: cDNA cloning and molecular characterization of a human nucleotide excision repair gene with high homology to yeast RAD3. *Embo J* 9 (5), 1437-1447

141   Roy, R. et al. (1994) The MO15 cell cycle kinase is associated with the TFIIH transcription-DNA repair factor. *Cell* 79 (6), 1093-1101

142   Schaeffer, L. et al. (1994) The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor. *Embo J* 13 (10), 2388-2392

143   Evans, E. et al. (1997) Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *Embo J* 16 (21), 6559-6573

144   Evans, E. et al. (1997) Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. *Embo J* 16 (3), 625-638

145   Mudgett, J.S. and MacInnes, M.A. (1990) Isolation of the functional human excision repair gene ERCC5 by intercosmid recombination. *Genomics* 8 (4), 623-633

146   Westerveld, A. et al. (1984) Molecular cloning of a human DNA repair gene. *Nature* 310 (5976), 425-429

147   Sijbers, A.M. et al. (1996) Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* 86 (5), 811-822

148   de Laat, W.L. et al. (1998) Mapping of interaction domains between human repair proteins ERCC1 and XPF. *Nucleic Acids Res* 26 (18), 4146-4152

149   Wood, R.D. (1997) Nucleotide excision repair in mammalian cells. *J Biol Chem* 272 (38), 23465-23468

150   Araujo, S.J. et al. (2001) Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome. *Mol Cell Biol* 21 (7), 2281-2291

151   Gary, R. et al. (1997) The DNA repair endonuclease XPG binds to proliferating cell nuclear antigen (PCNA) and shares sequence elements with the PCNA-binding regions of FEN-1 and cyclin-dependent kinase inhibitor p21. *J Biol Chem* 272 (39), 24522-24529

152   Mu, D. et al. (1996) Reaction mechanism of human DNA repair excision nuclease. *J Biol Chem* 271 (14), 8285-8294

153   Saijo, M. et al. (1996) Sequential binding of DNA repair proteins RPA and ERCC1 to XPA in vitro. *Nucleic Acids Res* 24 (23), 4719-4724

154   Bowman, G.D. et al. (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature* 429 (6993), 724-730

155   Wood, R.D. and Shivji, M.K. (1997) Which DNA polymerases are used for DNA-repair in eukaryotes? *Carcinogenesis* 18 (4), 605-610

156   Hunting, D.J. et al. (1991) DNA polymerase delta mediates excision repair in growing cells damaged with ultraviolet radiation. *Biochem Cell Biol* 69 (4), 303-308

157   Coverley, D. et al. (1992) A role for the human single-stranded DNA binding protein HSSB/RPA in an early stage of nucleotide excision repair. *Nucleic Acids Res* 20 (15), 3873-3880

158   Tomkinson, A.E. and Levin, D.S. (1997) Mammalian DNA ligases. *Bioessays* 19 (10), 893-901

159   Mellon, I. (2005) Transcription-coupled repair: a complex affair. *Mutat Res* 577 (1-2), 155-161

160   Friedberg, E.C. (1996) Relationships between DNA repair and transcription. *Annu Rev Biochem* 65, 15-42

161   Corda, Y. et al. (1993) Spectrum of DNA--platinum adduct recognition by prokaryotic and eukaryotic DNA-dependent RNA polymerases. *Biochemistry* 32 (33), 8582-8588

162   Donahue, B.A. et al. (1994) Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. *Proc Natl Acad Sci U S A* 91 (18), 8502-8506

163   Brooks, P.J. et al. (2000) The oxidative DNA lesion 8,5'-(S)-cyclo-2'-deoxyadenosine is repaired by the nucleotide excision repair pathway and blocks gene expression in mammalian cells. *J Biol Chem* 275 (29), 22355-22362

164   Mellon, I. and Hanawalt, P.C. (1989) Induction of the Escherichia coli lactose operon selectively increases repair of its transcribed DNA strand. *Nature* 342 (6245), 95-98

165   Tsutakawa, S.E. and Cooper, P.K. (2000) Transcription-coupled repair of oxidative DNA damage in human cells: mechanisms and consequences. *Cold Spring Harb Symp Quant Biol* 65, 201-215

166   Reed, S.H. (2005) Nucleotide excision repair in chromatin: the shape of things to come. *DNA Repair (Amst)* 4 (8), 909-918

167 Sugasawa, K. et al. (1993) Cell-free repair of UV-damaged simian virus 40 chromosomes in human cell extracts. I. Development of a cell-free system detecting excision repair of UV-irradiated SV40 chromosomes. *J Biol Chem* 268 (12), 9098-9104

168 Hara, R. et al. (2000) DNA damage in the nucleosome core is refractory to repair by human excision nuclease. *Mol Cell Biol* 20 (24), 9173-9181

169 Vignali, M. et al. (2000) ATP-dependent chromatin-remodeling complexes. *Mol Cell Biol* 20 (6), 1899-1910

170 Rapic-Otrin, V. et al. (2002) Sequential binding of UV DNA damage binding factor and degradation of the p48 subunit as early events after UV irradiation. *Nucleic Acids Res* 30 (11), 2588-2598

171 Liu, W. et al. (2000) Nuclear transport of human DDB protein induced by ultraviolet light. *J Biol Chem* 275 (28), 21429-21434

172 Venema, J. et al. (1990) The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc Natl Acad Sci U S A* 87 (12), 4707-4711

173 Troelstra, C. et al. (1992) ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* 71 (6), 939-953

174 Mayne, L.V. and Lehmann, A.R. (1982) Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Res* 42 (4), 1473-1478

175 Henning, K.A. et al. (1995) The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH. *Cell* 82 (4), 555-564

176 Troelstra, C. et al. (1990) Molecular cloning of the human DNA excision repair gene ERCC-6. *Mol Cell Biol* 10 (11), 5806-5813

177 Matson, S.W. et al. (1994) DNA helicases: enzymes with essential roles in all aspects of DNA metabolism. *Bioessays* 16 (1), 13-22

178 Selby, C.P. and Sancar, A. (1997) Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II. *J Biol Chem* 272 (3), 1885-1890

179 Citterio, E. et al. (2000) ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Mol Cell Biol* 20 (20), 7643-7653

180 Kamiuchi, S. et al. (2002) Translocation of Cockayne syndrome group A protein to the nuclear matrix: possible relevance to transcription-coupled DNA repair. *Proc Natl Acad Sci U S A* 99 (1), 201-206

181 van Gool, A.J. et al. (1997) The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. *Embo J* 16 (19), 5955-5965

182 Balajee, A.S. et al. (1997) Reduced RNA polymerase II transcription in intact and permeabilized Cockayne syndrome group B cells. *Proc Natl Acad Sci U S A* 94 (9), 4306-4311

183 Tantin, D. et al. (1997) Recruitment of the putative transcription-repair coupling factor CSB/ERCC6 to RNA polymerase II elongation complexes. *Mol Cell Biol* 17 (12), 6803-6814

184 Tantin, D. (1998) RNA polymerase II elongation complexes containing the Cockayne syndrome group B protein interact with a molecular complex

containing the transcription factor IIH components xeroderma pigmentosum B and p62. *J Biol Chem* 273 (43), 27794-27799

185 Goodrich, J.A. and Tjian, R. (1994) Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* 77 (1), 145-156

186 Zawel, L. et al. (1995) Recycling of the general transcription factors during RNA polymerase II transcription. *Genes Dev* 9 (12), 1479-1490

187 Tu, Y. et al. (1997) Sequence-specific and domain-specific DNA repair in xeroderma pigmentosum and Cockayne syndrome cells. *J Biol Chem* 272 (33), 20747-20755

188 Leadon, S.A. (1999) Transcription-coupled repair of DNA damage: unanticipated players, unexpected complexities. *Am J Hum Genet* 64 (5), 1259-1263

189 Selby, C.P. and Sancar, A. (1990) Transcription preferentially inhibits nucleotide excision repair of the template DNA strand in vitro. *J Biol Chem* 265 (34), 21330-21336

190 Batty, D.P. and Wood, R.D. (2000) Damage recognition in nucleotide excision repair of DNA. *Gene* 241 (2), 193-204

191 Hanawalt, P.C. et al. (1994) Repair and transcription. Collision or collusion? *Curr Biol* 4 (6), 518-521

192 Tornaletti, S. et al. (1997) Nucleotide sequence context effect of a cyclobutane pyrimidine dimer upon RNA polymerase II transcription. *J Biol Chem* 272 (50), 31719-31724

193 Nakatsu, Y. et al. (2000) XAB2, a novel tetratricopeptide repeat protein involved in transcription-coupled DNA repair and transcription. *J Biol Chem* 275 (45), 34931-34937

194 Dianov, G.L. et al. (1997) Reduced RNA polymerase II transcription in extracts of cockayne syndrome and xeroderma pigmentosum/Cockayne syndrome cells. *Nucleic Acids Res* 25 (18), 3636-3642

195 de Boer, J. and Hoeijmakers, J.H. (1999) Cancer from the outside, aging from the inside: mouse models to study the consequences of defective nucleotide excision repair. *Biochimie* 81 (1-2), 127-137

196 Hanawalt, P.C. (2000) DNA repair. The bases for Cockayne syndrome. *Nature* 405 (6785), 415-416

197 Mellon, I. et al. (1996) Transcription-coupled repair deficiency and mutations in human mismatch repair genes. *Science* 272 (5261), 557-560

198 Le Page, F. et al. (2000) Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome. *Cell* 101 (2), 159-171

199 Leadon, S.A. and Avrutskaya, A.V. (1998) Requirement for DNA mismatch repair proteins in the transcription-coupled repair of thymine glycols in Saccharomyces cerevisiae. *Mutat Res* 407 (2), 177-187

200 Leadon, S.A. and Avrutskaya, A.V. (1997) Differential involvement of the human mismatch repair proteins, hMLH1 and hMSH2, in transcription-coupled repair. *Cancer Res* 57 (17), 3784-3791

201 Sweder, K.S. et al. (1996) Mismatch repair mutants in yeast are not defective in transcription-coupled DNA repair of UV-induced DNA damage. *Genetics* 143 (3), 1127-1135

202 Viswanathan, A. and Doetsch, P.W. (1998) Effects of nonbulky DNA base damages on Escherichia coli RNA polymerase-mediated elongation and promoter clearance. *J Biol Chem* 273 (33), 21276-21281

203 Cooper, P.K. et al. (1997) Defective transcription-coupled repair of oxidative base damage in Cockayne syndrome patients from XP group G. *Science* 275 (5302), 990-993

204 Tornaletti, S. et al. (2001) Effect of thymine glycol on transcription elongation by T7 RNA polymerase and mammalian RNA polymerase II. *J Biol Chem* 276 (48), 45367-45371

205 Costa, R.M. et al. (2003) The eukaryotic nucleotide excision repair pathway. *Biochimie* 85 (11), 1083-1099

206 Lin, J.J. and Sancar, A. (1992) (A)BC excinuclease: the Escherichia coli nucleotide excision repair enzyme. *Mol Microbiol* 6 (16), 2219-2224

207 Verhoeven, E.E. et al. (2000) Catalytic sites for 3' and 5' incision of Escherichia coli nucleotide excision repair are both located in UvrC. *J Biol Chem* 275 (7), 5120-5123

208 Sancar, A. and Rupp, W.D. (1983) A novel repair enzyme: UVRABC excision nuclease of Escherichia coli cuts a DNA strand on both sides of the damaged region. *Cell* 33 (1), 249-260

209 Friedberg, E.C. et al. (1995) DNA repair and mutagenesis. *ASM Press, Washington DC*

210 Cleaver, J.E. (1972) Xeroderma pigmentosum: variants with normal DNA repair and normal sensitivity to ultraviolet light. *J Invest Dermatol* 58 (3), 124-128

211 Hoege, C. et al. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 419 (6903), 135-141

212 Lehmann, A.R. (2005) Replication of damaged DNA by translesion synthesis in human cells. *FEBS Lett* 579 (4), 873-876

213 Lehmann, A.R. (1972) Postreplication repair of DNA in ultraviolet-irradiated mammalian cells. *J Mol Biol* 66 (3), 319-337

214 Lehmann, A.R. (2002) Replication of damaged DNA in mammalian cells: new solutions to an old problem. *Mutat Res* 509 (1-2), 23-34

215 Goodman, M.F. (2000) Coping with replication 'train wrecks' in Escherichia coli using Pol V, Pol II and RecA proteins. *Trends Biochem Sci* 25 (4), 189-195

216 Cox, M.M. et al. (2000) The importance of repairing stalled replication forks. *Nature* 404 (6773), 37-41

217 Cordeiro-Stone, M. et al. (1997) Replication fork bypass of a pyrimidine dimer blocking leading strand DNA synthesis. *J Biol Chem* 272 (21), 13945-13954

218 Ensch-Simon, I. et al. (1998) Bypass of a site-specific cis-Syn thymine dimer in an SV40 vector during in vitro replication by HeLa and XPV cell-free extracts. *Biochemistry* 37 (22), 8218-8226

219 Svoboda, D.L. et al. (1998) Defective bypass replication of a leading strand cyclobutane thymine dimer in xeroderma pigmentosum variant cell extracts. *Cancer Res* 58 (11), 2445-2448

220 Cordonnier, A.M. et al. (1999) Impaired translesion synthesis in xeroderma pigmentosum variant extracts. *Mol Cell Biol* 19 (3), 2206-2211

221 Masutani, C. et al. (1999) Xeroderma pigmentosum variant (XP-V) correcting protein from HeLa cells has a thymine dimer bypass DNA polymerase activity. *Embo J* 18 (12), 3491-3501

222 Masutani, C. et al. (1999) The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. *Nature* 399 (6737), 700-704

223 Johnson, R.E. et al. (1999) hRAD30 mutations in the variant form of xeroderma pigmentosum. *Science* 285 (5425), 263-265

224 Matsuda, T. et al. (2000) Low fidelity DNA synthesis by human DNA polymerase-eta. *Nature* 404 (6781), 1011-1013

225 Livneh, Z. (2001) DNA damage control by novel DNA polymerases: translesion replication and mutagenesis. *J Biol Chem* 276 (28), 25639-25642

226 Friedberg, E.C. et al. (2000) The many faces of DNA polymerases: strategies for mutagenesis and for mutational avoidance. *Proc Natl Acad Sci U S A* 97 (11), 5681-5683

227 McDonald, J.P. et al. (1999) Novel human and mouse homologs of Saccharomyces cerevisiae DNA polymerase eta. *Genomics* 60 (1), 20-30

228 Gerlach, V.L. et al. (1999) Human and mouse homologs of Escherichia coli DinB (DNA polymerase IV), members of the UmuC/DinB superfamily. *Proc Natl Acad Sci U S A* 96 (21), 11922-11927

229 Ling, H. et al. (2001) Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell* 107 (1), 91-102

230 Zhou, B.L. et al. (2001) Crystal structure of a DinB lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Mol Cell* 8 (2), 427-437

231 Thakur, M. et al. (2001) DNA polymerase eta undergoes alternative splicing, protects against UV sensitivity and apoptosis, and suppresses Mre11-dependent recombination. *Genes Chromosomes Cancer* 32 (3), 222-235

232 Kannouche, P. et al. (2001) Domain structure, localization, and function of DNA polymerase eta, defective in xeroderma pigmentosum variant cells. *Genes Dev* 15 (2), 158-172

233 Masutani, C. et al. (2000) Mechanisms of accurate translesion synthesis by human DNA polymerase eta. *Embo J* 19 (12), 3100-3109

234 Trincao, J. et al. (2001) Structure of the catalytic core of S. cerevisiae DNA polymerase eta: implications for translesion DNA synthesis. *Mol Cell* 8 (2), 417-426

235 Friedberg, E.C. et al. (2001) Error-prone DNA polymerases: novel structures and the benefits of infidelity. *Cell* 107 (1), 9-12

236 Kannouche, P.L. et al. (2004) Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. *Mol Cell* 14 (4), 491-500

237 Haracska, L. et al. (2001) Physical and functional interactions of human DNA polymerase eta with PCNA. *Mol Cell Biol* 21 (21), 7199-7206

238 Raetz, C.R. and Whitfield, C. (2002) Lipopolysaccharide endotoxins. *Annu Rev Biochem* 71, 635-700

239 Raetz, C.R.H. (1996) Bacterial lipopolysaccharides; a remarkable family of bioactive macroamphiphiles. *In Eschericia coli and Salmonella. Cellular and Molecular Biology, vol 1, edn 2. Edited by Neidhardt FCEA. Washington DC: American Society for Microbiology*, 1035-1063

240   Mattsby-Baltzer, I. et al. (1992) Lipid A in Helicobacter pylori. *Infect Immun* 60 (10), 4383-4387

241   Ogawa, T. (1993) Chemical structure of lipid A from Porphyromonas (Bacteroides) gingivalis lipopolysaccharide. *FEBS Lett* 332 (1-2), 197-201

242   Raetz, C.R. (1990) Biochemistry of endotoxins. *Annu Rev Biochem* 59, 129-170

243   Wilkinson, S.G. (1977) Surface carbohydrates of the procaryotic cell. ed. Sutherland, I. W. *Academic Press Inc. New York*, 97-105

244   Mayer, H. (1989) Bacterial lipopolysaccharides. *Pure & Appl. Chem.* 61 (7), 1271-1282

245   Nikaido, H. and Vaara, M. (1985) Molecular basis of bacterial outer membrane permeability. *Microbiol Rev* 49 (1), 1-32

246   Nikaido, H. (1996) Outer membrane; *Escherichia coli* and *Salmonella*: *cellular and molecular biology*, vol 1, edn 2. Edited by Neidhardt FCEA. *Washington DC: American Society for Microbiology*, 29-47

247   Leive, L. (1974) The barrier function of the gram-negative envelope. *Ann N Y Acad Sci* 235 (0), 109-129

248   Nikaido, H. (1979) Permeability of the outer membrane of bacteria. *Angew Chem Int Ed Engl* 18 (5), 337-350

249   Moriyon, I. and Lopez-Goni, I. (1998) Structure and properties of the outer membranes of Brucella abortus and Brucella melitensis. *Int Microbiol* 1 (1), 19-26

250   Chaby, R. et al. (2005) Interactions between LPS and lung surfactant proteins. *J Endotoxin Res* 11 (3), 181-185

251   Frirdich, E. and Whitfield, C. (2005) Lipopolysaccharide inner core oligosaccharide structure and outer membrane stability in human pathogens belonging to the Enterobacteriaceae. *J Endotoxin Res* 11 (3), 133-144

252   Alexander, C. and Rietschel, E.T. (2001) Bacterial lipopolysaccharides and innate immunity. *J Endotoxin Res* 7 (3), 167-202

253   Orskov, I. et al. (1977) Serology, chemistry, and genetics of O and K antigens of Escherichia coli. *Bacteriol Rev* 41 (3), 667-710

254   Schnaitman, C.A. and Klena, J.D. (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. *Microbiol Rev* 57 (3), 655-682

255   Whitfield, C. (1995) Biosynthesis of lipopolysaccharide O antigens. *Trends Microbiol* 3 (5), 178-185

256   Reeves, P. (1995) Role of O-antigen variation in the immune response. *Trends Microbiol* 3 (10), 381-386

257   Reeves, P.R. (1992) Variation in O-antigens, niche-specific selection and bacterial populations. *FEMS Microbiol Lett* 79 (1-3), 509-516

258   Lerouge, I. and Vanderleyden, J. (2002) O-antigen structural variation: mechanisms and possible roles in animal/plant-microbe interactions. *FEMS Microbiol Rev* 26 (1), 17-47

259   Verma, N.K. et al. (1988) O-antigen variation in Salmonella spp.: rfb gene clusters of three strains. *J Bacteriol* 170 (1), 103-107

260   Raymond, C.K. et al. (2002) Genetic variation at the O-antigen biosynthetic locus in Pseudomonas aeruginosa. *J Bacteriol* 184 (13), 3614-3622

261   Liu, D. and Reeves, P.R. (1994) Presence of different O antigen forms in three isolates of one clone of Escherichia coli. *Genetics* 138 (1), 6-10

262   Rahman, M.M. et al. (1997) A virulent isolate of Salmonella enteritidis produces a Salmonella typhi-like lipopolysaccharide. *J Bacteriol* 179 (7), 2126-2131

263   Fuller, N.A. et al. (1973) The biosynthesis of cell wall lipopolysaccharide in Escherichia coli. VII. Characterization of heterogeneous "core" oligosaccharide structures. *J Biol Chem* 248 (22), 7938-7950

264   Heinrichs, D.E. et al. (1998) Molecular basis for structural diversity in the core regions of the lipopolysaccharides of Escherichia coli and Salmonella enterica. *Mol Microbiol* 30 (2), 221-232

265   Wilkinson, S.G. (1983) Composition and structure of lipopolysaccharides from Pseudomonas aeruginosa. *Rev Infect Dis* 5 Suppl 5, S941-949

266   Labischinski, H. et al. (1985) High state of order of isolated bacterial lipopolysaccharide and its possible contribution to the permeation barrier property of the outer membrane. *J Bacteriol* 162 (1), 9-20

267   Rietschel, E.T. et al. (1994) Bacterial endotoxin: molecular relationships of structure to activity and function. *Faseb J* 8 (2), 217-225

268   Raetz, C.R. and Dowhan, W. (1990) Biosynthesis and function of phospholipids in Escherichia coli. *J Biol Chem* 265 (3), 1235-1238

269   Jann, K. and Jann, B. (1987) Polysaccharide antigens of Escherichia coli. *Rev Infect Dis* 9 Suppl 5, S517-526

270   Polissi, A. and Georgopoulos, C. (1996) Mutational analysis and properties of the msbA gene of Escherichia coli, coding for an essential ABC family transporter. *Mol Microbiol* 20 (6), 1221-1233

271   Chang, G. and Roth, C.B. (2001) Structure of MsbA from E. coli: a homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* 293 (5536), 1793-1800

272   Reyes, C.L. and Chang, G. (2005) Structure of the ABC transporter MsbA in complex with ADP.vanadate and lipopolysaccharide. *Science* 308 (5724), 1028-1031

273   Zhou, Z. et al. (1998) Function of Escherichia coli MsbA, an essential ABC family transporter, in lipid A and phospholipid biosynthesis. *J Biol Chem* 273 (20), 12466-12475

274   Liu, D. et al. (1996) An O-antigen processing function for Wzx (RfbX): a promising candidate for O-unit flippase. *J Bacteriol* 178 (7), 2102-2107

275   MacLachlan, P.R. et al. (1991) Cloning, characterization, and DNA sequence of the rfaLK region for lipopolysaccharide synthesis in Salmonella typhimurium LT2. *J Bacteriol* 173 (22), 7151-7163

276   Collins, L.V. and Hackett, J. (1991) Molecular cloning, characterization, and nucleotide sequence of the rfc gene, which encodes an O-antigen polymerase of Salmonella typhimurium. *J Bacteriol* 173 (8), 2521-2529

277   Klena, J.D. et al. (1992) Role of Escherichia coli K-12 rfa genes and the rfp gene of Shigella dysenteriae 1 in generation of lipopolysaccharide core heterogeneity and attachment of O antigen. *J Bacteriol* 174 (22), 7297-7307

278   Samuel, G. and Reeves, P. (2003) Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydr Res* 338 (23), 2503-2519

279   Trent, M.S. (2004) Biosynthesis, transport, and modification of lipid A. *Biochem Cell Biol* 82 (1), 71-86

280   Ray, B.L. et al. (1984) The biosynthesis of gram-negative endotoxin. Formation of lipid A disaccharides from monosaccharide precursors in extracts of Escherichia coli. *J Biol Chem* 259 (8), 4852-4859

281   Bulawa, C.E. and Raetz, C.R. (1984) The biosynthesis of gram-negative endotoxin. Identification and function of UDP-2,3-diacylglucosamine in Escherichia coli. *J Biol Chem* 259 (8), 4846-4851

282   Anderson, M.S. et al. (1985) The biosynthesis of gram-negative endotoxin. Formation of lipid A precursors from UDP-GlcNAc in extracts of Escherichia coli. *J Biol Chem* 260 (29), 15536-15541

283   Takayama, K. et al. (1983) Complete structure of lipid A obtained from the lipopolysaccharides of the heptoseless mutant of Salmonella typhimurium. *J Biol Chem* 258 (21), 12801-12803

284   Imoto, M., Kusumoto, S., Shiba, T., Naoki, T., Iwashita, T., Rietschel, E. T., Wollenweber, H.- W., Galanos, C. and Lüderitz, O. (1983) Chemical structure of *E. coli* lipid A: linkage site of acyl groups in the disaccharide backbone. *Tetrahedron Lett.* 24, 4017-4020

285   Onishi, H.R. et al. (1996) Antibacterial agents that inhibit lipid A biosynthesis. *Science* 274 (5289), 980-982

286   Nelson, K.E. et al. (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of Thermotoga maritima. *Nature* 399 (6734), 323-329

287   Steeghs, L. et al. (1998) Meningitis bacterium is viable without endotoxin. *Nature* 392 (6675), 449-450

288   Fraser, C.M. et al. (1998) Complete genome sequence of Treponema pallidum, the syphilis spirochete. *Science* 281 (5375), 375-388

289   Takayama, K. et al. (1987) Absence of lipopolysaccharide in the Lyme disease spirochete, Borrelia burgdorferi. *Infect Immun* 55 (9), 2311-2313

290   Anderson, M.S. et al. (1993) UDP-N-acetylglucosamine acyltransferase of Escherichia coli. The first step of endotoxin biosynthesis is thermodynamically unfavorable. *J Biol Chem* 268 (26), 19858-19865

291   Anderson, M.S. and Raetz, C.R. (1987) Biosynthesis of lipid A precursors in Escherichia coli. A cytoplasmic acyltransferase that converts UDP-N-acetylglucosamine to UDP-3-O-(R-3-hydroxymyristoyl)-N-acetylglucosamine. *J Biol Chem* 262 (11), 5159-5169

292   Lee, B.I. and Suh, S.W. (2003) Crystal structure of UDP-N-acetylglucosamine acyltransferase from Helicobacter pylori. *Proteins* 53 (3), 772-774

293   Raetz, C.R. and Roderick, S.L. (1995) A left-handed parallel beta helix in the structure of UDP-N-acetylglucosamine acyltransferase. *Science* 270 (5238), 997-1000

294   Wyckoff, T.J. et al. (1998) Hydrocarbon rulers in UDP-N-acetylglucosamine acyltransferases. *J Biol Chem* 273 (49), 32369-32372

295   Wyckoff, T.J. and Raetz, C.R. (1999) The active site of Escherichia coli UDP-N-acetylglucosamine acyltransferase. Chemical modification and site-directed mutagenesis. *J Biol Chem* 274 (38), 27047-27055

296   Dotson, G.D. et al. (1998) Expression cloning of a Pseudomonas gene encoding a hydroxydecanoyl-acyl carrier protein-dependent UDP-GlcNAc acyltransferase. *J Bacteriol* 180 (2), 330-337

297   Young, K. et al. (1995) The envA permeability/cell division gene of Escherichia coli encodes the second enzyme of lipid A biosynthesis. UDP-3-

O-(R-3-hydroxymyristoyl)-N-acetylglucosamine deacetylase. *J Biol Chem* 270 (51), 30384-30391

298 Anderson, M.S. et al. (1988) Biosynthesis of lipid A in Escherichia coli: identification of UDP-3-O-[(R)-3-hydroxymyristoyl]-alpha-D-glucosamine as a precursor of UDP-N2,O3-bis[(R)-3-hydroxymyristoyl]-alpha-D-glucosamine. *Biochemistry* 27 (6), 1908-1917

299 Jackman, J.E. et al. (2000) Antibacterial agents that target lipid A biosynthesis in gram-negative bacteria. Inhibition of diverse UDP-3-O-(r-3-hydroxymyristoyl)-n-acetylglucosamine deacetylases by substrate analogs containing zinc binding motifs. *J Biol Chem* 275 (15), 11002-11009

300 Kelly, T.M. et al. (1993) The firA gene of Escherichia coli encodes UDP-3-O-(R-3-hydroxymyristoyl)-glucosamine N-acyltransferase. The third step of endotoxin biosynthesis. *J Biol Chem* 268 (26), 19866-19874

301 Babinski K. J. and Raetz, C.R.H. (1998) Identification of a gene encoding a novel *Eschericia coli* UDP-2,3-diacylglucosamine hydrolase. *FASEB J* 12, A1288

302 Ray, B.L. and Raetz, C.R. (1987) The biosynthesis of gram-negative endotoxin. A novel kinase in Escherichia coli membranes that incorporates the 4'-phosphate of lipid A. *J Biol Chem* 262 (3), 1122-1128

303 Clementz, T. and Raetz, C.R. (1991) A gene coding for 3-deoxy-D-manno-octulosonic-acid transferase in Escherichia coli. Identification, mapping, cloning, and sequencing. *J Biol Chem* 266 (15), 9687-9696

304 Ray, P.H. and Benedict, C.D. (1982) CTP:CMP-3-deoxy-D-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase). *Methods Enzymol* 83, 535-540

305 Strohmaier, H. et al. (1995) Expression of genes kdsA and kdsB involved in 3-deoxy-D-manno-octulosonic acid metabolism and biosynthesis of enterobacterial lipopolysaccharide is growth phase regulated primarily at the transcriptional level in Escherichia coli K-12. *J Bacteriol* 177 (15), 4488-4500

306 Brozek, K.A. and Raetz, C.R. (1990) Biosynthesis of lipid A in Escherichia coli. Acyl carrier protein-dependent incorporation of laurate and myristate. *J Biol Chem* 265 (26), 15410-15417

307 Clementz, T. et al. (1996) Function of the htrB high temperature requirement gene of Escherchia coli in the acylation of lipid A: HtrB catalyzed incorporation of laurate. *J Biol Chem* 271 (20), 12095-12102

308 Clementz, T. et al. (1997) Function of the Escherichia coli msbB gene, a multicopy suppressor of htrB knockouts, in the acylation of lipid A. Acylation by MsbB follows laurate incorporation by HtrB. *J Biol Chem* 272 (16), 10353-10360

309 Raetz, C.R. (1993) Bacterial endotoxins: extraordinary lipids that activate eucaryotic signal transduction. *J Bacteriol* 175 (18), 5745-5753

310 Tobias, P.S. et al. (1986) Isolation of a lipopolysaccharide-binding acute phase reactant from rabbit serum. *J Exp Med* 164 (3), 777-793

311 Ulevitch, R.J. and Tobias, P.S. (1999) Recognition of gram-negative bacteria and endotoxin by the innate immune system. *Curr Opin Immunol* 11 (1), 19-22

312 Medzhitov, R. and Janeway, C., Jr. (2000) Innate immunity. *N Engl J Med* 343 (5), 338-344

313 Aderem, A. and Ulevitch, R.J. (2000) Toll-like receptors in the induction of the innate immune response. *Nature* 406 (6797), 782-787

314   Loppnow, H. et al. (1989) IL-1 induction-capacity of defined lipopolysaccharide partial structures. *J Immunol* 142 (9), 3229-3238

315   Arend, W.P. and Massoni, R.J. (1986) Characteristics of bacterial lipopolysaccharide induction of interleukin 1 synthesis and secretion by human monocytes. *Clin Exp Immunol* 64 (3), 656-664

316   Dinarello, C.A. (1991) Interleukin-1 and interleukin-1 antagonism. *Blood* 77 (8), 1627-1652

317   Tamura, M. et al. (1992) Lipopolysaccharides of Bacteroides intermedius (Prevotella intermedia) and Bacteroides (Porphyromonas) gingivalis induce interleukin-8 gene expression in human gingival fibroblast cultures. *Infect Immun* 60 (11), 4932-4937

318   Delvos, U. et al. (1987) Effect of lipopolysaccharides on cultured human endothelial cells. Relationship between tissue factor activity and prostacyclin release. *Blut* 55 (2), 101-108

319   Parrillo, J.E. (1993) Pathogenetic mechanisms of septic shock. *N Engl J Med* 328 (20), 1471-1477

320   Zasloff, M. (1992) Antibiotic peptides as mediators of innate immunity. *Curr Opin Immunol* 4 (1), 3-7

321   Bevins, C.L. (1994) Antimicrobial peptides as agents of mucosal immunity. *Ciba Found Symp* 186, 250-260; discussion 261-259

322   Ganz, T. (2003) Defensins: antimicrobial peptides of innate immunity. *Nat Rev Immunol* 3 (9), 710-720

323   Hoffmann, J.A. et al. (1999) Phylogenetic perspectives in innate immunity. *Science* 284 (5418), 1313-1318

324   Poltorak, A. et al. (1998) Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* 282 (5396), 2085-2088

325   Medzhitov, R. et al. (1997) A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. *Nature* 388 (6640), 394-397

326   Chow, J.C. et al. (1999) Toll-like receptor-4 mediates lipopolysaccharide-induced signal transduction. *J Biol Chem* 274 (16), 10689-10692

327   Jones, D.E. and Bevins, C.L. (1992) Paneth cells of the human small intestine express an antimicrobial peptide gene. *J Biol Chem* 267 (32), 23216-23225

328   Stolzenberg, E.D. et al. (1997) Epithelial antibiotic induced in states of disease. *Proc Natl Acad Sci U S A* 94 (16), 8686-8690

329   Diamond, G. et al. (1993) Airway epithelial cells are the site of expression of a mammalian antimicrobial peptide gene. *Proc Natl Acad Sci U S A* 90 (10), 4596-4600

330   Goldman, M.J. et al. (1997) Human beta-defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. *Cell* 88 (4), 553-560

331   Hancock, R.E. and Chapple, D.S. (1999) Peptide antibiotics. *Antimicrob Agents Chemother* 43 (6), 1317-1323

332   Cole, A.M. et al. (1999) Innate antimicrobial activity of nasal secretions. *Infect Immun* 67 (7), 3267-3275

333   Zanetti, M. et al. (2002) Cathelicidin peptides as candidates for a novel class of antimicrobials. *Curr Pharm Des* 8 (9), 779-793

334   Chalk, R. et al. (1995) Brugia pahangi: the effects of cecropins on microfilariae in vitro and in Aedes aegypti. *Exp Parasitol* 80 (3), 401-406

335   Matsuzaki, K. (1999) Why and how are peptide-lipid interactions utilized for self-defense? Magainins and tachyplesins as archetypes. *Biochim Biophys Acta* 1462 (1-2), 1-10

336   Shai, Y. (1999) Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides. *Biochim Biophys Acta* 1462 (1-2), 55-70

337   Vaara, M. et al. (1981) Characterization of the lipopolysaccharide from the polymyxin-resistant pmrA mutants of Salmonella typhimurium. *FEBS Lett* 129 (1), 145-149

338   Guo, L. et al. (1997) Regulation of lipid A modifications by Salmonella typhimurium virulence genes phoP-phoQ. *Science* 276 (5310), 250-253

339   Helander, I.M. et al. (1996) Characterization of lipopolysaccharides of polymyxin-resistant and polymyxin-sensitive Klebsiella pneumoniae O3. *Eur J Biochem* 237 (1), 272-278

340   Miller, S.I. et al. (1989) A two-component regulatory system (phoP phoQ) controls Salmonella typhimurium virulence. *Proc Natl Acad Sci U S A* 86 (13), 5054-5058

341   Garcia Vescovi, E. et al. (1996) Mg2+ as an extracellular signal: environmental regulation of Salmonella virulence. *Cell* 84 (1), 165-174

342   Bishop, R.E. et al. (2000) Transfer of palmitate from phospholipids to lipid A in outer membranes of gram-negative bacteria. *Embo J* 19 (19), 5071-5080

343   Miller, S.I. et al. (1990) Characterization of defensin resistance phenotypes associated with mutations in the phoP virulence regulon of Salmonella typhimurium. *Infect Immun* 58 (11), 3706-3710

344   Groisman, E.A. et al. (1992) Resistance to host antimicrobial peptides is necessary for Salmonella virulence. *Proc Natl Acad Sci U S A* 89 (24), 11939-11943

345   Fields, P.I. et al. (1989) A Salmonella locus that controls resistance to microbicidal proteins from phagocytic cells. *Science* 243 (4894 Pt 1), 1059-1062

346   Aussel, L. et al. (2000) Novel variation of lipid A structures in strains of different Yersinia species. *FEBS Lett* 465 (1), 87-92

347   Carty, S.M. et al. (1999) Effect of cold shock on lipid A biosynthesis in Escherichia coli. Induction At 12 degrees C of an acyltransferase specific for palmitoleoyl-acyl carrier protein. *J Biol Chem* 274 (14), 9677-9685

348   Krasikova, I.N. et al. (1999) Effects of growth temperature and pVM82 plasmid on fatty acids of lipid A from Yersinia pseudotuberculosis. *Biochemistry (Mosc)* 64 (3), 338-344

349   Zhou, Z. et al. (1999) Lipid A modifications characteristic of Salmonella typhimurium are induced by NH4VO3 in Escherichia coli K12. Detection of 4-amino-4-deoxy-L-arabinose, phosphoethanolamine and palmitate. *J Biol Chem* 274 (26), 18503-18514

350   Guo, L. et al. (1998) Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. *Cell* 95 (2), 189-198

351   Golenbock, D.T. et al. (1991) Lipid A-like molecules that antagonize the effects of endotoxins on human monocytes. *J Biol Chem* 266 (29), 19490-19498

352   Pohlman, T.H. et al. (1987) Deacylated lipopolysaccharide inhibits neutrophil adherence to endothelium induced by lipopolysaccharide in vitro. *J Exp Med* 165 (5), 1393-1402

353   Trent, M.S. et al. (2001) A PhoP/PhoQ-induced Lipase (PagL) that catalyzes 3-O-deacylation of lipid A precursors in membranes of Salmonella typhimurium. *J Biol Chem* 276 (12), 9083-9092

354 Basu, S.S. et al. (1999) A deacylase in Rhizobium leguminosarum membranes that cleaves the 3-O-linked beta-hydroxymyristoyl moiety of lipid A precursors. *J Biol Chem* 274 (16), 11150-11158

355 Moran, A.P. et al. (1997) Structural characterization of the lipid A component of Helicobacter pylori rough- and smooth-form lipopolysaccharides. *J Bacteriol* 179 (20), 6453-6463

356 Suda, Y. et al. (1997) Chemical structure of lipid A from Helicobacter pylori strain 206-1 lipopolysaccharide. *J Biochem (Tokyo)* 121 (6), 1129-1133

357 Kumada, H. et al. (1995) Structural study on the free lipid A isolated from lipopolysaccharide of Porphyromonas gingivalis. *J Bacteriol* 177 (8), 2098-2106

358 Que, N.L. et al. (2000) Two-dimensional NMR spectroscopy and structures of six lipid A species from Rhizobium etli CE3. Detection of an acyloxyacyl residue in each component and origin of the aminogluconate moiety. *J Biol Chem* 275 (36), 28017-28027

359 Que, N.L. et al. (2000) Purification and mass spectrometry of six lipid A species from the bacterial endosymbiont Rhizobium etli. Demonstration of a conserved distal unit and a variable proximal portion. *J Biol Chem* 275 (36), 28006-28016

360 Karbarz, M.J. et al. (2003) Expression cloning and biochemical characterization of a Rhizobium leguminosarum lipid A 1-phosphatase. *J Biol Chem* 278 (41), 39269-39279

361 Vaara, M. (1981) Increased outer membrane resistance to ethylenediaminetetraacetate and cations in novel lipid A mutants. *J Bacteriol* 148 (2), 426-434

362 Gunn, J.S. and Miller, S.I. (1996) PhoP-PhoQ activates transcription of pmrAB, encoding a two-component regulatory system involved in Salmonella typhimurium antimicrobial peptide resistance. *J Bacteriol* 178 (23), 6857-6864

363 Gunn, J.S. et al. (1998) PmrA-PmrB-regulated genes necessary for 4-aminoarabinose lipid A modification and polymyxin resistance. *Mol Microbiol* 27 (6), 1171-1182

364 Wosten, M.M. et al. (2000) A signal transduction system that responds to extracellular iron. *Cell* 103 (1), 113-125

365 Soncini, F.C. and Groisman, E.A. (1996) Two-component regulatory systems can interact to process multiple environmental signals. *J Bacteriol* 178 (23), 6796-6801

366 Vaara, M. (1981) Effect of ionic strength on polymyxin resistance of pmrA mutant of Salmonella. *FEMS Microbiol Lett.* 11, 321-326

367 Stinavage, P. et al. (1989) O antigen and lipid A phosphoryl groups in resistance of Salmonella typhimurium LT-2 to nonoxidative killing in human polymorphonuclear neutrophils. *Infect Immun* 57 (12), 3894-3900

368 Gunn, J.S. et al. (2000) Constitutive mutations of the Salmonella enterica serovar Typhimurium transcriptional virulence regulator phoP. *Infect Immun* 68 (6), 3758-3762

369 Mackinnon, F.G. et al. (2002) Identification of a gene (lpt-3) required for the addition of phosphoethanolamine to the lipopolysaccharide inner core of Neisseria meningitidis and its role in mediating susceptibility to bactericidal killing and opsonophagocytosis. *Mol Microbiol* 43 (4), 931-943

370 Cox, A.D. et al. (2003) Phosphorylation of the lipid A region of meningococcal lipopolysaccharide: identification of a family of transferases

that add phosphoethanolamine to lipopolysaccharide. *J Bacteriol* 185 (11), 3270-3277

371   Baker, S.J. et al. (1999) The Salmonella typhi melittin resistance gene pqaB affects intracellular growth in PMA-differentiated U937 cells, polymyxin B resistance and lipopolysaccharide. *Microbiology* 145 (Pt 2), 367-378

372   Oppenheimer, N.J., and Handlon, A. L. (1992) In The Enzymes (Sigman, D. S., ed), 3rd Ed, Vol 20. *Academic Press Inc, San Diego CA*, 453-504

373   Parkhill, J. et al. (2001) Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* 413 (6855), 523-527

374   Parkhill, J. et al. (2003) Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. *Nat Genet* 35 (1), 32-40

375   Stover, C.K. et al. (2000) Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen. *Nature* 406 (6799), 959-964

376   Breazeale, S.D. et al. (2005) A formyltransferase required for polymyxin resistance in Escherichia coli and the modification of lipid A with 4-Amino-4-deoxy-L-arabinose. Identification and function oF UDP-4-deoxy-4-formamido-L-arabinose. *J Biol Chem* 280 (14), 14154-14167

377   Breazeale, S.D. et al. (2002) Oxidative decarboxylation of UDP-glucuronic acid in extracts of polymyxin-resistant Escherichia coli. Origin of lipid a species modified with 4-amino-4-deoxy-L-arabinose. *J Biol Chem* 277 (4), 2886-2896

378   Trent, M.S. et al. (2001) An inner membrane enzyme in Salmonella and Escherichia coli that transfers 4-amino-4-deoxy-L-arabinose to lipid A: induction on polymyxin-resistant mutants and role of a novel lipid-linked donor. *J Biol Chem* 276 (46), 43122-43131

379   Zhou, Z. et al. (2001) Lipid A modifications in polymyxin-resistant Salmonella typhimurium: PMRA-dependent 4-amino-4-deoxy-L-arabinose, and phosphoethanolamine incorporation. *J Biol Chem* 276 (46), 43111-43121

380   Breazeale, S.D. et al. (2003) Origin of lipid A species modified with 4-amino-4-deoxy-L-arabinose in polymyxin-resistant mutants of Escherichia coli. An aminotransferase (ArnB) that generates UDP-4-deoxyl-L-arabinose. *J Biol Chem* 278 (27), 24731-24739

381   Williams, G.J. et al. (2005) Structure and function of both domains of ArnA, a dual function decarboxylase and a formyltransferase, involved in 4-amino-4-deoxy-L-arabinose biosynthesis. *J Biol Chem* 280 (24), 23000-23008

382   Jornvall, H. et al. (1995) Short-chain dehydrogenases/reductases (SDR). *Biochemistry* 34 (18), 6003-6013

383   Gunn, J.S. (2001) Bacterial modification of LPS and resistance to antimicrobial peptides. *J Endotoxin Res* 7 (1), 57-62

384   Noland, B.W. et al. (2002) Structural studies of Salmonella typhimurium ArnB (PmrH) aminotransferase: a 4-amino-4-deoxy-L-arabinose lipopolysaccharide-modifying enzyme. *Structure (Camb)* 10 (11), 1569-1580

385   Jansonius, J.N. (1998) Structure, evolution and action of vitamin B6-dependent enzymes. *Curr Opin Struct Biol* 8 (6), 759-769

386   Trent, M.S. et al. (2001) Accumulation of a polyisoprene-linked amino sugar in polymyxin-resistant Salmonella typhimurium and Escherichia coli: structural characterization and transfer to lipid A in the periplasm. *J Biol Chem* 276 (46), 43132-43144

387   Raetz, C.R. (1986) Molecular genetics of membrane phospholipid synthesis. *Annu Rev Genet* 20, 253-295

388   Vaara, M. (1996) Lipid A: target for antibacterial drugs. *Science* 274 (5289), 939-940

389   Kline, T. et al. (2002) Potent, novel in vitro inhibitors of the Pseudomonas aeruginosa deacetylase LpxC. *J Med Chem* 45 (14), 3112-3129

390   Coggins, B.E. et al. (2003) Structure of the LpxC deacetylase with a bound substrate-analog inhibitor. *Nat Struct Biol* 10 (8), 645-651

391   Whittington, D.A. et al. (2003) Crystal structure of LpxC, a zinc-dependent deacetylase essential for endotoxin biosynthesis. *Proc Natl Acad Sci U S A* 100 (14), 8146-8150

392   Ernst, R.K. et al. (1999) Specific lipopolysaccharide found in cystic fibrosis airway Pseudomonas aeruginosa. *Science* 286 (5444), 1561-1565

393   Govan, J.R. and Deretic, V. (1996) Microbial pathogenesis in cystic fibrosis: mucoid Pseudomonas aeruginosa and Burkholderia cepacia. *Microbiol Rev* 60 (3), 539-574

394   Rossmann, M.G. and van Beek, C.G. (1999) Data processing. *Acta Crystallogr D Biol Crystallogr* 55 (Pt 10), 1631-1640

395   Leslie, A.G. (2006) The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* 62 (Pt 1), 48-57

396   Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. Joint CCP4 and ESF-EAMCB. *Newsletter on Protein Crystallography* 26

397   Leslie, A.G. (1999) Integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr* 55 (Pt 10), 1696-1702

398   Rhodes, G. (2000) Crystallography Made Crystal Clear. *Book - Academic Press* 3

399   Leslie, A.G. et al. (2002) Automation of the collection and processing of X-ray diffraction data -- a generic approach. *Acta Crystallogr D Biol Crystallogr* 58 (Pt 11), 1924-1928

400   Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr D Biol Crystallogr* 62 (Pt 1), 72-82

401   Evans, P.R. (1997) "SCALA" Joint CCP4 and ESF-EAMCB. *Newsletter on Protein Crystallography* 33, 22-24

402   International tables for crystallography. Volume A: Space group symmetry. *Book - Springer* A (5)

403   Murshudov, G.N. et al. (1999) Efficient anisotropic refinement of macromolecular structures using FFT. *Acta Crystallogr D Biol Crystallogr* 55 (Pt 1), 247-255

404   Salanoubat, M. et al. (2002) Genome sequence of the plant pathogen Ralstonia solanacearum. *Nature* 415 (6871), 497-502

405   Ovadi, J. (1991) Physiological significance of metabolic channelling. *J Theor Biol* 152 (1), 1-22

406   James, C.L. and Viola, R.E. (2002) Production and characterization of bifunctional enzymes. Substrate channeling in the aspartate pathway. *Biochemistry* 41 (11), 3726-3731

407   James, C.L. and Viola, R.E. (2002) Production and characterization of bifunctional enzymes. Domain swapping to produce new bifunctional enzymes in the aspartate pathway. *Biochemistry* 41 (11), 3720-3725

226

408 Karsten, W.E. and Viola, R.E. (1991) Chemical and kinetic mechanisms of aspartate-beta-semialdehyde dehydrogenase from Escherichia coli. *Biochim Biophys Acta* 1077 (2), 209-219

409 Rohdich, F. et al. (1999) Cytidine 5'-triphosphate-dependent biosynthesis of isoprenoids: YgbP protein of Escherichia coli catalyzes the formation of 4-diphosphocytidyl-2-C-methylerythritol. *Proc Natl Acad Sci U S A* 96 (21), 11758-11763

410 Herz, S. et al. (2000) Biosynthesis of terpenoids: YgbB protein converts 4-diphosphocytidyl-2C-methyl-D-erythritol 2-phosphate to 2C-methyl-D-erythritol 2,4-cyclodiphosphate. *Proc Natl Acad Sci U S A* 97 (6), 2486-2490

411 Gabrielsen, M. et al. (2004) Hexameric assembly of the bifunctional methylerythritol 2,4-cyclodiphosphate synthase and protein-protein associations in the deoxy-xylulose-dependent pathway of isoprenoid precursor biosynthesis. *J Biol Chem* 279 (50), 52753-52761

412 Rohmer, M. (1999) The discovery of a mevalonate independent pathway for isoprenoid biosynthesis in bacteria, algae and higher plants. *Nat. Prod. Rep.* 16, 565-574

413 Luttgen, H. et al. (2000) Biosynthesis of terpenoids: YchB protein of Escherichia coli phosphorylates the 2-hydroxy group of 4-diphosphocytidyl-2C-methyl-D-erythritol. *Proc Natl Acad Sci U S A* 97 (3), 1062-1067

414 Minton, A.P. and Wilf, J. (1981) Effect of macromolecular crowding upon the structure and function of an enzyme: glyceraldehyde-3-phosphate dehydrogenase. *Biochemistry* 20 (17), 4821-4826

415 Minton, A.P. (1981) Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers* 20, 2093-2120

416 Spivey, H.O. and Ovadi, J. (1999) Substrate channeling. *Methods* 19 (2), 306-321

417 Geck, M.K. and Kirsch, J.F. (1999) A novel, definitive test for substrate channeling illustrated with the aspartate aminotransferase/malate dehydrogenase system. *Biochemistry* 38 (25), 8032-8037

418 Almassy, R.J. et al. (1992) Structures of apo and complexed Escherichia coli glycinamide ribonucleotide transformylase. *Proc Natl Acad Sci U S A* 89 (13), 6114-6118

419 Greasley, S.E. et al. (2001) Crystal structure of a bifunctional transformylase and cyclohydrolase enzyme in purine biosynthesis. *Nat Struct Biol* 8 (5), 402-406

420 Schmitt, E. et al. (1996) Structure of crystalline Escherichia coli methionyl-tRNA(f)Met formyltransferase: comparison with glycinamide ribonucleotide formyltransferase. *Embo J* 15 (17), 4749-4758

421 Krupenko, S.A. and Oleinik, N.V. (2002) 10-formyltetrahydrofolate dehydrogenase, one of the major folate enzymes, is down-regulated in tumor tissues and possesses suppressor effects on cancer cells. *Cell Growth Differ* 13 (5), 227-236

422 Chumanevich, A.A. et al. (2004) The crystal structure of the hydrolase domain of 10-formyltetrahydrofolate dehydrogenase: mechanism of hydrolysis and its interplay with the dehydrogenase domain. *J Biol Chem* 279 (14), 14355-14364

423 Thoden, J.B. et al. (2002) PurT-encoded glycinamide ribonucleotide transformylase. Accommodation of adenosine nucleotide analogs within the active site. *J Biol Chem* 277 (26), 23898-23908

424   Wolan, D.W. et al. (2002) Structural insights into the avian AICAR transformylase mechanism. *Biochemistry* 41 (52), 15505-15513

425   Shim, J.H. and Benkovic, S.J. (1999) Catalytic mechanism of Escherichia coli glycinamide ribonucleotide transformylase probed by site-directed mutagenesis and pH-dependent studies. *Biochemistry* 38 (31), 10024-10031

426   Warren, M.S. et al. (1996) The transformylase enzymes of *de novo* purine biosynthesis. *Pure & Appl. Chem.* 68 (11), 2029-2036

427   Banerjee, D. et al. (2002) Novel aspects of resistance to drugs targeted to dihydrofolate reductase and thymidylate synthase. *Biochim Biophys Acta* 1587 (2-3), 164-173

428   Elion, G.B. (1989) The purine path to chemotherapy. *Science* 244 (4900), 41-47

429   Kisliuk, R.L. (2003) Deaza analogs of folic acid as antitumor agents. *Curr Pharm Des* 9 (31), 2615-2625

430   Zhang, Y. et al. (2003) Rational design, synthesis, evaluation, and crystal structure of a potent inhibitor of human GAR Tfase: 10-(trifluoroacetyl)-5,10-dideazaacyclic-5,6,7,8-tetrahydrofolic acid. *Biochemistry* 42 (20), 6043-6056

431   Beardsley, G.P. et al. (1989) A new folate antimetabolite, 5,10-dideaza-5,6,7,8-tetrahydrofolate is a potent inhibitor of de novo purine synthesis. *J Biol Chem* 264 (1), 328-333

432   Cheong, C.G. et al. (2004) Crystal structures of human bifunctional enzyme aminoimidazole-4-carboxamide ribonucleotide transformylase/IMP cyclohydrolase in complex with potent sulfonyl-containing antifolates. *J Biol Chem* 279 (17), 18034-18045

433   Xu, L. et al. (2004) Crystal structure of avian aminoimidazole-4-carboxamide ribonucleotide transformylase in complex with a novel non-folate inhibitor identified by virtual ligand screening. *J Biol Chem* 279 (48), 50555-50565

434   Shim, J.H. and Benkovic, S.J. (1998) Evaluation of the kinetic mechanism of Escherichia coli glycinamide ribonucleotide transformylase. *Biochemistry* 37 (24), 8776-8782

435   Zhang, Y. et al. (2002) Crystal structures of human GAR Tfase at low and high pH and with substrate beta-GAR. *Biochemistry* 41 (48), 14206-14215

436   Beardsley, G.P. et al. (1998) Structure and functional relationships in human pur H. *Adv Exp Med Biol* 431, 221-226

437   Schmitt, E. et al. (1998) Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *Embo J* 17 (23), 6819-6826

438   Smith, G.K. et al. (1982) Direct transfer of one-carbon units in the transformylations of de novo purine biosynthesis. *Biochemistry* 21 (12), 2870-2874

439   Warren, M.S. et al. (1996) A rapid screen of active site mutants in glycinamide ribonucleotide transformylase. *Biochemistry* 35 (27), 8855-8862

440   Warren, M.S. and Benkovic, S.J. (1997) Combinatorial manipulation of three key active site residues in glycinamide ribonucleotide transformylase. *Protein Eng* 10 (1), 63-68

441   Persson, B. et al. (1994) A super-family of medium-chain dehydrogenases/reductases (MDR). Sub-lines including zeta-crystallin, alcohol and polyol dehydrogenases, quinone oxidoreductase enoyl reductases, VAT-1 and other proteins. *Eur J Biochem* 226 (1), 15-22

442   Riveros-Rosas, H. et al. (2003) Diversity, taxonomy and evolution of medium-chain dehydrogenase/reductase superfamily. *Eur J Biochem* 270 (16), 3309-3334

443   Persson, B. et al. (1991) Characteristics of short-chain alcohol dehydrogenases and related enzymes. *Eur J Biochem* 200 (2), 537-543

444   Krozowski, Z. (1994) The short-chain alcohol dehydrogenase superfamily: variations on a common theme. *J Steroid Biochem Mol Biol* 51 (3-4), 125-130

445   Jornvall, H. et al. (1981) Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. *Proc Natl Acad Sci U S A* 78 (7), 4226-4230

446   Neuberger, M.S. and Hartley, B.S. (1981) Structure of an experimentally evolved gene duplication encoding ribitol dehydrogenase in a mutant of Klebsiella aerogenes. *J Gen Microbiol* 122 (2), 181-191

447   Oppermann, U. et al. (2003) Short-chain dehydrogenases/reductases (SDR): the 2002 update. *Chem Biol Interact* 143-144, 247-253

448   Kallberg, Y. et al. (2002) Short-chain dehydrogenases/reductases (SDRs). *Eur J Biochem* 269 (18), 4409-4417

449   Jornvall, H. (1999) Multiplicity and complexity of SDR and MDR enzymes. *Adv Exp Med Biol* 463, 359-364

450   Jornvall, H. et al. (1999) SDR and MDR: completed genome sequences show these protein families to be large, of old origin, and of complex nature. *FEBS Lett* 445 (2-3), 261-264

451   Wierenga, R.K. et al. (1985) Interaction of pyrophosphate moieties with alpha-helices in dinucleotide binding proteins. *Biochemistry* 24, 1346-1357

452   Tanaka, N. et al. (2001) SDR: Structure, mechanism of action, and substrate recognition. *Curr. Org. Chem.* 5, 89-111

453   Rubach, J.K. and Plapp, B.V. (2003) Amino acid residues in the nicotinamide binding site contribute to catalysis by horse liver alcohol dehydrogenase. *Biochemistry* 42 (10), 2907-2915

454   Benach, J. et al. (2005) Drosophila alcohol dehydrogenase: acetate-enzyme interactions and novel insights into the effects of electrostatics on catalysis. *J Mol Biol* 345 (3), 579-598

455   Duax, W.L. et al. (2000) Steroid dehydrogenase structures, mechanism of action, and disease. *Vitam Horm* 58, 121-148

456   Duax, W.L. et al. (1996) The fascinating complexities of steroid-binding enzymes. *Curr Opin Struct Biol* 6 (6), 813-823

457   Oppermann, U.C. et al. (1997) Structure-function relationships of SDR hydroxysteroid dehydrogenases. *Adv Exp Med Biol* 414, 403-415

458   Rossmann, M.G. et al. (1975) Evolutionary and structural relationships among dehydrogenases. *Enzymes* 1, 61-102

459   Ghosh, D. et al. (2001) Porcine carbonyl reductase. structural basis for a functional monomer in short chain dehydrogenases/reductases. *J Biol Chem* 276 (21), 18457-18463

460   Liu, Y. et al. (1997) Mechanistic roles of tyrosine 149 and serine 124 in UDP-galactose 4-epimerase from Escherichia coli. *Biochemistry* 36 (35), 10675-10684

461   Arabshahi, A. et al. (1988) Uridine diphosphate galactose 4-epimerase. pH dependence of the reduction of NAD+ by a substrate analog. *J Biol Chem* 263 (6), 2638-2643

462 Allard, S.T. et al. (2001) The crystal structure of dTDP-D-Glucose 4,6-dehydratase (RmlB) from Salmonella enterica serovar Typhimurium, the second enzyme in the dTDP-l-rhamnose pathway. *J Mol Biol* 307 (1), 283-295

463 Thoden, J.B. et al. (2000) Crystallographic evidence for Tyr 157 functioning as the active site base in human UDP-galactose 4-epimerase. *Biochemistry* 39 (19), 5691-5701

464 Somoza, J.R. et al. (2000) Structural and kinetic analysis of Escherichia coli GDP-mannose 4,6 dehydratase provides insights into the enzyme's catalytic mechanism and regulation by GDP-fucose. *Structure Fold Des* 8 (2), 123-135

465 Naismith, J.H. (2004) Chemical insights from structural studies of enzymes. *Biochem Soc Trans* 32 (Pt 5), 647-654

466 Thoden, J.B. et al. (1996) Crystal structures of the oxidized and reduced forms of UDP-galactose 4-epimerase isolated from Escherichia coli. *Biochemistry* 35 (8), 2557-2566

467 Thoden, J.B. et al. (1996) Molecular structure of the NADH/UDP-glucose abortive complex of UDP-galactose 4-epimerase from Escherichia coli: implications for the catalytic mechanism. *Biochemistry* 35 (16), 5137-5144

468 Thoden, J.B. et al. (1996) High-resolution X-ray structure of UDP-galactose 4-epimerase complexed with UDP-phenol. *Protein Sci* 5 (11), 2149-2161

469 Field, R.A. and Naismith, J.H. (2003) Structural and mechanistic basis of bacterial sugar nucleotide-modifying enzymes. *Biochemistry* 42 (25), 7637-7647

470 Allard, S.T. et al. (2002) Toward a structural understanding of the dehydratase mechanism. *Structure (Camb)* 10 (1), 81-92

471 Mulichak, A.M. et al. (1999) Crystal structure of SQD1, an enzyme involved in the biosynthesis of the plant sulfolipid headgroup donor UDP-sulfoquinovose. *Proc Natl Acad Sci U S A* 96 (23), 13097-13102

472 Blankenfeldt, W. et al. (2002) Variation on a theme of SDR. dTDP-6-deoxy-L- lyxo-4-hexulose reductase (RmlD) shows a new Mg2+-dependent dimerization mode. *Structure* 10 (6), 773-786

473 Cleland, W.W. et al. (1998) The low barrier hydrogen bond in enzymatic catalysis. *J Biol Chem* 273 (40), 25529-25532

474 Cleland, W.W. and Kreevoy, M.M. (1994) Low-barrier hydrogen bonds and enzymic catalysis. *Science* 264 (5167), 1887-1890

475 Thoden, J.B. et al. (1997) Structural analysis of UDP-sugar binding to UDP-galactose 4-epimerase from Escherichia coli. *Biochemistry* 36 (21), 6294-6304

476 Berger, E. et al. (2001) Acid-base catalysis by UDP-galactose 4-epimerase: correlations of kinetically measured acid dissociation constants with thermodynamic values for tyrosine 149. *Biochemistry* 40 (22), 6699-6705

477 Johnson, D.A. and Liu, H. (1998) Mechanisms and pathways from recent deoxysugar biosynthesis research. *Curr Opin Chem Biol* 2 (5), 642-649

478 Liu, H.W. and Thorson, J.S. (1994) Pathways and mechanisms in the biogenesis of novel deoxysugars by bacteria. *Annu Rev Microbiol* 48, 223-256

479 Gabriel, O. and Lindquist, L.C. (1968) Biological mechanisms involved in the formation of deoxy sugars. IV. Enzymatic conversion of thymidine diphosphoglucose-4T to thymidine diphospho-4-keto-6-deoxyglucose-6T. *J Biol Chem* 243 (7), 1479-1484

480 Beis, K. et al. (2003) The structure of NADH in the enzyme dTDP-d-glucose dehydratase (RmlB). *J Am Chem Soc* 125 (39), 11872-11878

481 Moriarity, J.L. et al. (2002) UDP-glucuronate decarboxylase, a key enzyme in proteoglycan synthesis: cloning, characterization, and localization. *J Biol Chem* 277 (19), 16968-16975

482 Doublie, S. (1997) Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol* 276, 523-530

483 Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Crystallogr D Biol Crystallogr* 55 (Pt 4), 849-861

484 Terwilliger, T.C. (2000) Maximum-likelihood density modification. *Acta Crystallogr D Biol Crystallogr* 56 (Pt 8), 965-972

485 Jones, T.A. et al. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A* 47 (Pt 2), 110-119

486 Collaborative Computational Project, N. (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50 (Pt 5), 760-763

487 Cowtan, K. (1994) An automated procedure for phase improvement by density improvement. *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* 31, 34-38

488 Perrakis, A. et al. (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Biol* 6 (5), 458-463

489 Morris, R.J. et al. (2004) Breaking good resolutions with ARP/wARP. *J Synchrotron Radiat* 11 (Pt 1), 56-59

490 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60 (Pt 12 Pt 1), 2126-2132

491 Schuttelkopf, A.W. and van Aalten, D.M. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr D Biol Crystallogr* 60 (Pt 8), 1355-1363

492 Laskowski, R.A. et al. (1993) Main-chain bond lengths and bond angles in protein structures. *J Mol Biol* 231 (4), 1049-1067

493 Vriend, G. (1990) WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8 (1), 52-56, 29

494 Maki, M. et al. (2003) Cloning and functional expression of a novel GDP-6-deoxy-D-talose synthetase from Actinobacillus actinomycetemcomitans. *Glycobiology* 13 (4), 295-303

495 Gatzeva-Topalova, P.Z. et al. (2005) Crystal structure and mechanism of the Escherichia coli ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry* 44 (14), 5328-5338

496 Gatzeva-Topalova, P.Z. et al. (2004) Crystal structure of Escherichia coli ArnA (PmrI) decarboxylase domain. A key enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry* 43 (42), 13370-13379

497 Gatzeva-Topalova, P.Z. et al. (2005) Structure and mechanism of ArnA: conformational change implies ordered dehydrogenase mechanism in key enzyme for polymyxin resistance. *Structure (Camb)* 13 (6), 929-942

498 Hegeman, A.D. et al. (2002) Concerted and stepwise dehydration mechanisms observed in wild-type and mutated Escherichia coli dTDP-glucose 4,6-dehydratase. *Biochemistry* 41 (8), 2797-2804

499 Hegeman, A.D. et al. (2001) Probing catalysis by Escherichia coli dTDP-glucose-4,6-dehydratase: identification and preliminary characterization of

functional amino acid residues at the active site. *Biochemistry* 40 (22), 6598-6610

500 Bruck, I. and O'Donnell, M. (2001) The ring-type polymerase sliding clamp family. *Genome Biol* 2 (1), REVIEWS3001

501 Fay, P.J. et al. (1981) Size classes of products synthesized processively by DNA polymerase III and DNA polymerase III holoenzyme of Escherichia coli. *J Biol Chem* 256 (2), 976-983

502 Burgers, P.M. and Kornberg, A. (1982) ATP activation of DNA polymerase III holoenzyme from Escherichia coli. II. Initiation complex: stoichiometry and reactivity. *J Biol Chem* 257 (19), 11474-11478

503 O'Donnell, M.E. and Kornberg, A. (1985) Complete replication of templates by Escherichia coli DNA polymerase III holoenzyme. *J Biol Chem* 260 (23), 12884-12889

504 Kornberg, A. and Baker, T.A. (1991) DNA replication. *New York: W. H. Freeman*

505 Ellison, V. and Stillman, B. (2001) Opening of the clamp: an intimate view of an ATP-driven biological machine. *Cell* 106 (6), 655-660

506 O'Donnell, M.E. (1987) Accessory proteins bind a primed template and mediate rapid cycling of DNA polymerase III holoenzyme from Escherichia coli. *J Biol Chem* 262 (34), 16558-16565

507 Maki, S. and Kornberg, A. (1988) DNA polymerase III holoenzyme of Escherichia coli. III. Distinctive processive polymerases reconstituted from purified subunits. *J Biol Chem* 263 (14), 6561-6569

508 O'Donnell, M. and Studwell, P.S. (1990) Total reconstitution of DNA polymerase III holoenzyme reveals dual accessory protein clamps. *J Biol Chem* 265 (2), 1179-1187

509 Stukenberg, P.T. et al. (1991) Mechanism of the sliding beta-clamp of DNA polymerase III holoenzyme. *J Biol Chem* 266 (17), 11328-11334

510 Miyachi, K. et al. (1978) Autoantibody to a nuclear antigen in proliferating cells. *J Immunol* 121 (6), 2228-2234

511 Prelich, G. et al. (1987) The cell-cycle regulated proliferating cell nuclear antigen is required for SV40 DNA replication in vitro. *Nature* 326 (6112), 471-475

512 Tan, C.K. et al. (1986) An auxiliary protein for DNA polymerase-delta from fetal calf thymus. *J Biol Chem* 261 (26), 12310-12316

513 Bravo, R. et al. (1987) Cyclin/PCNA is the auxiliary protein of DNA polymerase-delta. *Nature* 326 (6112), 515-517

514 Bravo, R. and Celis, J.E. (1980) A search for differential polypeptide synthesis throughout the cell cycle of HeLa cells. *J Cell Biol* 84 (3), 795-802

515 Studwell, P.S. and O'Donnell, M. (1990) Processive replication is contingent on the exonuclease subunit of DNA polymerase III holoenzyme. *J Biol Chem* 265 (2), 1171-1178

516 Tinker, R.L. et al. (1994) Detecting the ability of viral, bacterial and eukaryotic replication proteins to track along DNA. *Embo J* 13 (22), 5330-5337

517 Kong, X.P. et al. (1992) Three-dimensional structure of the beta subunit of E. coli DNA polymerase III holoenzyme: a sliding DNA clamp. *Cell* 69 (3), 425-437

518 Saenger, W. (1984) Principles of nucleic acid structure. *New York: Springer:Verlag*

519 Turner, J. et al. (1999) The internal workings of a DNA polymerase clamp-loading machine. *Embo J* 18 (3), 771-783

520 Jarvis, T.C. et al. (1989) Structural and enzymatic studies of the T4 DNA replication system. I. Physical characterization of the polymerase accessory protein complex. *J Biol Chem* 264 (21), 12709-12716

521 Krishna, T.S. et al. (1994) Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA. *Cell* 79 (7), 1233-1243

522 Gulbis, J.M. et al. (1996) Structure of the C-terminal region of p21(WAF1/CIP1) complexed with human PCNA. *Cell* 87 (2), 297-306

523 Matsumiya, S. et al. (2002) Physical interaction between proliferating cell nuclear antigen and replication factor C from Pyrococcus furiosus. *Genes Cells* 7 (9), 911-922

524 Shamoo, Y. and Steitz, T.A. (1999) Building a replisome from interacting pieces: sliding clamp complexed to a peptide from DNA polymerase and a polymerase editing complex. *Cell* 99 (2), 155-166

525 Sakurai, S. et al. (2005) Structural basis for recruitment of human flap endonuclease 1 to PCNA. *Embo J* 24 (4), 683-693

526 Zhang, S.J. et al. (1995) A conserved region in the amino terminus of DNA polymerase delta is involved in proliferating cell nuclear antigen binding. *J Biol Chem* 270 (14), 7988-7992

527 Eissenberg, J.C. et al. (1997) Mutations in yeast proliferating cell nuclear antigen define distinct sites for interaction with DNA polymerase delta and DNA polymerase epsilon. *Mol Cell Biol* 17 (11), 6367-6378

528 Lee, S.H. et al. (1991) Synthesis of DNA by DNA polymerase epsilon in vitro. *J Biol Chem* 266 (33), 22707-22717

529 Zhang, G. et al. (1999) Studies on the interactions between human replication factor C and human proliferating cell nuclear antigen. *Proc Natl Acad Sci U S A* 96 (5), 1869-1874

530 Loor, G. et al. (1997) Identification of DNA replication and cell cycle proteins that interact with PCNA. *Nucleic Acids Res* 25 (24), 5041-5046

531 Tom, S. et al. (2001) DNA ligase I and proliferating cell nuclear antigen form a functional complex. *J Biol Chem* 276 (27), 24817-24825

532 Levin, D.S. et al. (1997) An interaction between DNA ligase I and proliferating cell nuclear antigen: implications for Okazaki fragment synthesis and joining. *Proc Natl Acad Sci U S A* 94 (24), 12863-12868

533 Hegde, P. et al. (1999) Mutational analysis of the Shab-encoded delayed rectifier K(+) channels in Drosophila. *J Biol Chem* 274 (31), 22109-22113

534 Moggs, J.G. et al. (2000) A CAF-1-PCNA-mediated chromatin assembly pathway triggered by sensing DNA damage. *Mol Cell Biol* 20 (4), 1206-1218

535 Shibahara, K. and Stillman, B. (1999) Replication-dependent marking of DNA by PCNA facilitates CAF-1-coupled inheritance of chromatin. *Cell* 96 (4), 575-585

536 Chuang, L.S. et al. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* 277 (5334), 1996-2000

537 Iida, T. et al. (2002) PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. *Genes Cells* 7 (10), 997-1007

538 Ohta, S. et al. (2002) A proteomics approach to identify proliferating cell nuclear antigen (PCNA)-binding proteins in human cell lysates. Identification of the human CHL12/RFCs2-5 complex as a novel PCNA-binding protein. *J Biol Chem* 277 (43), 40362-40367

539 Venclovas, C. et al. (2002) Molecular modeling-based analysis of interactions in the RFC-dependent clamp-loading process. *Protein Sci* 11 (10), 2403-2416

540 Dianova, II et al. (2001) Interaction of human AP endonuclease 1 with flap endonuclease 1 and proliferating cell nuclear antigen involved in long-patch base excision repair. *Biochemistry* 40 (42), 12639-12644

541 Unk, I. et al. (2002) Stimulation of 3'-->5' exonuclease and 3'-phosphodiesterase activities of yeast apn2 by proliferating cell nuclear antigen. *Mol Cell Biol* 22 (18), 6480-6486

542 Kedar, P.S. et al. (2002) Direct interaction between mammalian DNA polymerase beta and proliferating cell nuclear antigen. *J Biol Chem* 277 (34), 31115-31123

543 Haracska, L. et al. (2001) Targeting of human DNA polymerase iota to the replication machinery via interaction with PCNA. *Proc Natl Acad Sci U S A* 98 (25), 14256-14261

544 Haracska, L. et al. (2002) Stimulation of DNA synthesis activity of human DNA polymerase kappa by PCNA. *Mol Cell Biol* 22 (3), 784-791

545 Maga, G. et al. (2002) Human DNA polymerase lambda functionally and physically interacts with proliferating cell nuclear antigen in normal and translesion DNA synthesis. *J Biol Chem* 277 (50), 48434-48440

546 Vairapandi, M. et al. (2000) Characterization of MyD118, Gadd45, and proliferating cell nuclear antigen (PCNA) interacting domains. PCNA impedes MyD118 AND Gadd45-mediated negative growth control. *J Biol Chem* 275 (22), 16810-16819

547 Umar, A. et al. (1996) Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* 87 (1), 65-73

548 Clark, A.B. et al. (2000) Functional interaction of proliferating cell nuclear antigen with MSH2-MSH6 and MSH2-MSH3 complexes. *J Biol Chem* 275 (47), 36498-36501

549 Kleczkowska, H.E. et al. (2001) hMSH3 and hMSH6 interact with PCNA and colocalize with it to replication foci. *Genes Dev* 15 (6), 724-736

550 Muller-Weeks, S.J. and Caradonna, S. (1996) Specific association of cyclin-like uracil-DNA glycosylase with the proliferating cell nuclear antigen. *Exp Cell Res* 226 (2), 346-355

551 Waga, S. et al. (1994) The p21 inhibitor of cyclin-dependent kinases controls DNA replication by interaction with PCNA. *Nature* 369 (6481), 574-578

552 Hasan, S. et al. (2001) Transcription coactivator p300 binds PCNA and may have a role in DNA repair synthesis. *Nature* 410 (6826), 387-391

553 Fukami-Kobayashi, J. and Mitsui, Y. (1999) Cyclin D1 inhibits cell proliferation through binding to PCNA and cdk2. *Exp Cell Res* 246 (2), 338-347

554 Matsuoka, S. et al. (1994) D-type cyclin-binding regions of proliferating cell nuclear antigen. *J Biol Chem* 269 (15), 11030-11036

555 Zhang, H. et al. (1993) Proliferating cell nuclear antigen and p21 are components of multiple cell cycle kinase complexes. *Mol Biol Cell* 4 (9), 897-906

556 Vivona, J.B. and Kelman, Z. (2003) The diverse spectrum of sliding clamp interacting proteins. *FEBS Lett* 546 (2-3), 167-172

557 Maga, G. and Hubscher, U. (2003) Proliferating cell nuclear antigen (PCNA): a dancer with many partners. *J Cell Sci* 116 (Pt 15), 3051-3060

558 Jonsson, Z.O. and Hubscher, U. (1997) Proliferating cell nuclear antigen: more than a clamp for DNA polymerases. *Bioessays* 19 (11), 967-975

559 Flores-Rozas, H. et al. (1994) Cdk-interacting protein 1 directly binds with proliferating cell nuclear antigen and inhibits DNA replication catalyzed by the DNA polymerase delta holoenzyme. *Proc Natl Acad Sci U S A* 91 (18), 8655-8659

560 Prelich, G. and Stillman, B. (1988) Coordinated leading and lagging strand synthesis during SV40 DNA replication in vitro requires PCNA. *Cell* 53 (1), 117-126

561 Liu, Y.C. et al. (1989) Requirement for proliferating cell nuclear antigen expression during stages of the Chinese hamster ovary cell cycle. *Biochemistry* 28 (7), 2967-2974

562 Jaskulski, D. et al. (1988) Inhibition of cellular proliferation by antisense oligodeoxynucleotides to PCNA cyclin. *Science* 240 (4858), 1544-1546

563 Ayyagari, R. et al. (2003) Okazaki fragment maturation in yeast. I. Distribution of functions between FEN1 AND DNA2. *J Biol Chem* 278 (3), 1618-1625

564 Jin, Y.H. et al. (2003) Okazaki fragment maturation in yeast. II. Cooperation between the polymerase and 3'-5'-exonuclease activities of Pol delta in the creation of a ligatable nick. *J Biol Chem* 278 (3), 1626-1633

565 Bae, S.H. and Seo, Y.S. (2000) Characterization of the enzymatic properties of the yeast dna2 Helicase/endonuclease suggests a new model for Okazaki fragment processing. *J Biol Chem* 275 (48), 38022-38031

566 Chapados, B.R. et al. (2004) Structural basis for FEN-1 substrate specificity and PCNA-mediated activation in DNA replication and repair. *Cell* 116 (1), 39-50

567 Varga-Weisz, P. (2005) Chromatin remodeling factors and DNA replication. *Prog Mol Subcell Biol* 38, 1-30

568 Bestor, T.H. and Verdine, G.L. (1994) DNA methyltransferases. *Curr Opin Cell Biol* 6 (3), 380-389

569 Adams, R.L. (1995) Eukaryotic DNA methyltransferases--structure and function. *Bioessays* 17 (2), 139-145

570 Lau, P.J. and Kolodner, R.D. (2003) Transfer of the MSH2.MSH6 complex from proliferating cell nuclear antigen to mispaired bases in DNA. *J Biol Chem* 278 (1), 14-17

571 Flores-Rozas, H. et al. (2000) Proliferating cell nuclear antigen and Msh2p-Msh6p interact to form an active mispair recognition complex. *Nat Genet* 26 (3), 375-378

572 Goetz, J.D. et al. (2005) Reduced repair of DNA double-strand breaks by homologous recombination in a DNA ligase I-deficient human cell line. *DNA Repair (Amst)* 4 (6), 649-654

573 Aboussekhra, A. and Wood, R.D. (1995) Detection of nucleotide excision repair incisions in human fibroblasts by immunostaining for PCNA. *Exp Cell Res* 221 (2), 326-332

574 Miura, M. and Sasaki, T. (1996) Effect of XPA gene mutations on UV-induced immunostaining of PCNA in fibroblasts from xeroderma pigmentosum group A patients. *Mutat Res* 364 (1), 51-56

575 Nichols, A.F. and Sancar, A. (1992) Purification of PCNA as a nucleotide excision repair protein. *Nucleic Acids Res* 20 (13), 2441-2446

235

576 Koundrioukoff, S. et al. (2000) A direct interaction between proliferating cell nuclear antigen (PCNA) and Cdk2 targets PCNA-interacting proteins for phosphorylation. *J Biol Chem* 275 (30), 22882-22887

577 Liu, M. and Pelling, J.C. (1995) UV-B/A irradiation of mouse keratinocytes results in p53-mediated WAF1/CIP1 expression. *Oncogene* 10 (10), 1955-1960

578 el-Deiry, W.S. et al. (1993) WAF1, a potential mediator of p53 tumor suppression. *Cell* 75 (4), 817-825

579 Chen, J. et al. (1995) Separate domains of p21 involved in the inhibition of Cdk kinase and PCNA. *Nature* 374 (6520), 386-388

580 Podust, V.N. et al. (1995) Mechanism of inhibition of proliferating cell nuclear antigen-dependent DNA synthesis by the cyclin-dependent kinase inhibitor p21. *Biochemistry* 34 (27), 8869-8875

581 Shivji, M.K. et al. (1998) Resistance of human nucleotide excision repair synthesis in vitro to p21Cdn1. *Oncogene* 17 (22), 2827-2838

582 Warbrick, E. (1998) PCNA binding through a conserved motif. *Bioessays* 20 (3), 195-199

583 White, M.F. (2003) Archaeal DNA repair: paradigms and puzzles. *Biochem Soc Trans* 31 (Pt 3), 690-693

584 Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74 (11), 5088-5090

585 Fox, G.E. et al. (1977) Classification of Methanogenic Bacteria by 16S Ribosomal RNA Characterization. *Proc Natl Acad Sci U S A* 77 (10), 4537-4541

586 Kandler, O. and Hippe, H. (1977) Lack of peptidoglycan in the cell walls of Methanosarcina barkeri. *Arch Microbiol* 113 (1-2), 57-60

587 Woese, C.R. et al. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A* 87 (12), 4576-4579

588 Hori, H. and Osawa, S. (1987) Origin and evolution of organisms as deduced from 5S ribosomal RNA sequences. *Mol Biol Evol* 4 (5), 445-472

589 Hori, H. and Osawa, S. (1979) Evolutionary change in 5S RNA secondary structure and a phylogenic tree of 54 5S RNA species. *Proc Natl Acad Sci U S A* 76 (1), 381-385

590 Brown, J.R. and Doolittle, W.F. (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc Natl Acad Sci U S A* 92 (7), 2441-2445

591 Rowlands, T. et al. (1994) The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science* 264 (5163), 1326-1329

592 Huet, J. et al. (1982) Spot-immunodetection of conserved determinants in eukaryotic RNA polymerases. Study with antibodies to yeast RNA polymerases subunits. *J Biol Chem* 257 (5), 2613-2618

593 Iwabe, N. et al. (1989) Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci U S A* 86 (23), 9355-9359

594 Edgell, D.R. and Doolittle, W.F. (1997) Archaea and the origin(s) of DNA replication proteins. *Cell* 89 (7), 995-998

595 Keeling, P.J. and Doolittle, W.F. (1995) Archaea: narrowing the gap between prokaryotes and eukaryotes. *Proc Natl Acad Sci U S A* 92 (13), 5761-5764

596   Bell, C.E. (2005) Structure and mechanism of Escherichia coli RecA ATPase. *Mol Microbiol* 58 (2), 358-366

597   Seitz, E.M. et al. (1998) RadA protein is an archaeal RecA protein homolog that catalyzes DNA strand exchange. *Genes Dev* 12 (9), 1248-1253

598   Baumann, P. and West, S.C. (1998) Role of the human RAD51 protein in homologous recombination and double-stranded-break repair. *Trends Biochem Sci* 23 (7), 247-251

599   Seitz, E.M. and Kowalczykowski, S.C. (2000) The DNA binding and pairing preferences of the archaeal RadA protein demonstrate a universal characteristic of DNA strand exchange proteins. *Mol Microbiol* 37 (3), 555-560

600   Constantinesco, F. et al. (2002) NurA, a novel 5'-3' nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea. *EMBO Rep* 3 (6), 537-542

601   Roberts, J.A. et al. (2003) An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA. *Mol Microbiol* 48 (2), 361-371

602   Aravind, L. et al. (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucleic Acids Res* 27 (5), 1223-1242

603   Grogan, D.W. (2000) The question of DNA repair in hyperthermophilic archaea. *Trends Microbiol* 8 (4), 180-185

604   Ogrunc, M. et al. (1998) Nucleotide excision repair in the third kingdom. *J Bacteriol* 180 (21), 5796-5798

605   Gardner, M.J. et al. (2002) Genome sequence of the human malaria parasite Plasmodium falciparum. *Nature* 419 (6906), 498-511

606   Salerno, V. et al. (2003) Transcriptional response to DNA damage in the archaeon Sulfolobus solfataricus. *Nucleic Acids Res* 31 (21), 6127-6138

607   Dionne, I. et al. (2003) A heterotrimeric PCNA in the hyperthermophilic archaeon Sulfolobus solfataricus. *Mol Cell* 11 (1), 275-282

608   Fitz-Gibbon, S.T. et al. (2002) Genome sequence of the hyperthermophilic crenarchaeon Pyrobaculum aerophilum. *Proc Natl Acad Sci U S A* 99 (2), 984-989

609   Daimon, K. et al. (2002) Three proliferating cell nuclear antigen-like proteins found in the hyperthermophilic archaeon Aeropyrum pernix: interactions with the two DNA polymerases. *J Bacteriol* 184 (3), 687-694

610   Motz, M. et al. (2002) Elucidation of an archaeal replication protein network to generate enhanced PCR enzymes. *J Biol Chem* 277 (18), 16179-16188

611   Cann, I.K. et al. (1999) Functional interactions of a homolog of proliferating cell nuclear antigen with DNA polymerases in Archaea. *J Bacteriol* 181 (21), 6591-6599

612   Kelman, Z. and Hurwitz, J. (2000) A unique organization of the protein subunits of the DNA polymerase clamp loader in the archaeon Methanobacterium thermoautotrophicum deltaH. *J Biol Chem* 275 (10), 7327-7336

613   Yang, H. et al. (2002) Direct interaction between uracil-DNA glycosylase and a proliferating cell nuclear antigen homolog in the crenarchaeon Pyrobaculum aerophilum. *J Biol Chem* 277 (25), 22271-22278

614   Hosfield, D.J. et al. (1998) Structure of the DNA repair and replication endonuclease and exonuclease FEN-1: coupling DNA and PCNA binding to FEN-1 activity. *Cell* 95 (1), 135-146

615   Gruz, P. et al. (2001) Synthetic activity of Sso DNA polymerase Y1, an archaeal DinB-like DNA polymerase, is stimulated by processivity factors proliferating cell nuclear antigen and replication factor C. *J Biol Chem* 276 (50), 47394-47401

616   Dionne, I. and Bell, S.D. (2005) Characterization of an archaeal family 4 uracil DNA glycosylase and its interaction with PCNA and chromatin proteins. *Biochem J* 387 (Pt 3), 859-863

617   Cann, I.K. and Ishino, Y. (1999) Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* 152 (4), 1249-1267

618   Roberts, J.A. and White, M.F. (2005) DNA end-directed and processive nuclease activities of the archaeal XPF enzyme. *Nucleic Acids Res* 33 (20), 6662-6670

619   Roberts, J.A. and White, M.F. (2005) An archaeal endonuclease displays key properties of both eukaryal XPF-ERCC1 and Mus81. *J Biol Chem* 280 (7), 5924-5928

620   Thelen, M.P. et al. (1999) A sliding clamp model for the Rad1 family of cell cycle checkpoint proteins. *Cell* 96 (6), 769-770

621   Venclovas, C. and Thelen, M.P. (2000) Structure-based predictions of Rad1, Rad9, Hus1 and Rad17 participation in sliding clamp and clamp-loading complexes. *Nucleic Acids Res* 28 (13), 2481-2493

622   Burtelow, M.A. et al. (2001) Reconstitution and molecular analysis of the hRad9-hHus1-hRad1 (9-1-1) DNA damage responsive checkpoint complex. *J Biol Chem* 276 (28), 25903-25909

623   Shiomi, Y. et al. (2002) Clamp and clamp loader structures of the human checkpoint protein complexes, Rad9-1-1 and Rad17-RFC. *Genes Cells* 7 (8), 861-868

624   Cortez, D. et al. (1999) Requirement of ATM-dependent phosphorylation of brca1 in the DNA damage response to double-strand breaks. *Science* 286 (5442), 1162-1166

625   Lim, D.S. et al. (2000) ATM phosphorylates p95/nbs1 in an S-phase checkpoint pathway. *Nature* 404 (6778), 613-617

626   Zhou, B.B. and Elledge, S.J. (2000) The DNA damage response: putting checkpoints in perspective. *Nature* 408 (6811), 433-439

627   Lowe, S.W. et al. (1993) p53 is required for radiation-induced apoptosis in mouse thymocytes. *Nature* 362 (6423), 847-849

628   Hirao, A. et al. (2000) DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science* 287 (5459), 1824-1827

629   O'Connell, M.J. et al. (2000) The G2-phase DNA-damage checkpoint. *Trends Cell Biol* 10 (7), 296-303

630   Kai, M. et al. (2001) Fission yeast Rad17 associates with chromatin in response to aberrant genomic structures. *Mol Cell Biol* 21 (10), 3289-3301

631   Green, C.M. et al. (2000) A novel Rad24 checkpoint protein complex closely related to replication factor C. *Curr Biol* 10 (1), 39-42

632   Zou, L. et al. (2002) Regulation of ATR substrate selection by Rad17-dependent loading of Rad9 complexes onto chromatin. *Genes Dev* 16 (2), 198-208

633   Elledge, S.J. (1996) Cell cycle checkpoints: preventing an identity crisis. *Science* 274 (5293), 1664-1672

634   Newman, M. et al. (2005) Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *Embo J* 24 (5), 895-905

635   Komori, K. et al. (2002) Novel endonuclease in Archaea cleaving DNA with various branched structure. *Genes Genet Syst* 77 (4), 227-241

636   Sgouros, J. et al. (1999) A relationship betweena DNA-repair/recombination nuclease family and archaeal helicases. *Trends Biochem Sci* 24 (3), 95-97

637   Li, L. et al. (1995) Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. *Mol Cell Biol* 15 (4), 1993-1998

638   Gaillard, P.H. and Wood, R.D. (2001) Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. *Nucleic Acids Res* 29 (4), 872-879

639   Boddy, M.N. et al. (2001) Mus81-Eme1 are essential components of a Holliday junction resolvase. *Cell* 107 (4), 537-548

640   Boddy, M.N. et al. (2000) Damage tolerance protein Mus81 associates with the FHA1 domain of checkpoint kinase Cds1. *Mol Cell Biol* 20 (23), 8758-8766

641   Chen, X.B. et al. (2001) Human Mus81-associated endonuclease cleaves Holliday junctions in vitro. *Mol Cell* 8 (5), 1117-1127

642   Davies, A.A. et al. (1995) Role of the Rad1 and Rad10 proteins in nucleotide excision repair and recombination. *J Biol Chem* 270 (42), 24638-24641

643   Tsodikov, O.V. et al. (2005) Crystal structure and DNA binding functions of ERCC1, a subunit of the DNA structure-specific endonuclease XPF-ERCC1. *Proc Natl Acad Sci U S A* 102 (32), 11236-11241

644   Nishino, T. et al. (2003) X-ray and biochemical anatomy of an archaeal XPF/Rad1/Mus81 family nuclease: similarity between its endonuclease domain and restriction enzymes. *Structure* 11 (4), 445-457

645   Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif* 41 (1), 207-234

646   Radaev, S. and Sun, P.D. (2002) Crystallization of protein-protein complexes. *Applied Crystallography* 35

647   McCoy, A.J. et al. (2005) Likelihood-enhanced fast translation functions. *Acta Crystallogr D Biol Crystallogr* 61 (Pt 4), 458-464

648   Argiriadi, M.A. et al. (2006) Crystal structure of a DNA polymerase sliding clamp from a Gram-positive bacterium. *BMC Struct Biol* 6 (1), 2

649   Kostrub, C.F. et al. (1998) Hus1p, a conserved fission yeast checkpoint protein, interacts with Rad1p and is phosphorylated in response to DNA damage. *Embo J* 17 (7), 2055-2066

650   Caspari, T. et al. (2000) Characterization of Schizosaccharomyces pombe Hus1: a PCNA-related protein that associates with Rad1 and Rad9. *Mol Cell Biol* 20 (4), 1254-1262