Structural studies of three enzymes: telomerase, the methyltransferase CobJ and pectate lyase

To, Teng Teng

Structural studies of three enzymes: Telomerase, the methyltransferase CobJ and Pectate lyase

Teng Teng To

Thesis submitted to the University of London for the Degree of Doctor of Philosophy

Abstract

This thesis investigates the structure and function of three enzymes of biotechnological and biomedical interest: telomerase from *Caenorhabtidis elegans*, pectate lyase from *Bacillus subtilis* and the methyltransferase CobJ from *Rhodobacter capsulatus*.

Telomerase is a ribonucleoprotein found in all eukaryotes and its function is to maintain telomere length, sustain chromosome integrity and circumvent the end-replication problem. The protein requires two subunits to function, telomerase reverse transcriptase (TERT), the catalytic component, and an intrinsic RNA template (TR). The TR makes telomerase a unique reverse transcriptase using the template in the synthesis of short iterative sequences which cap the ends of telomeres. This work reports the successful cloning of a small and therefore potentially crystallisable TERT from *C. elegans* and expression trials of this catalytic component.

Cobalamin (vitamin B_{12}) is an intricate small molecule belonging to a group of compounds called cyclic tetrapyrroles. Its biosynthesis is achieved through a complex pathway encompassing over thirty different enzyme-mediated reactions. Within this pathway there are seven methyltransferases which add eight S-adenosyl-methionine (SAM) derived methyl groups to the macrocycle. CobJ catalyses the methylation of C17 and ring contraction at C20, this reaction which exudes C20 from the tetrapyrrole ring is unprecedented in nature. In this thesis I report the crystallisation of native CobJ and refinement and validation of a high resolution structure along side co-crystallisation and soaking experiments aimed at capturing an enzyme-tetrapyrrole complex.

Pectate lyase (BsPel) is an enzyme secreted from the bacterium *B. subtilis,* it is one of many enzymes secreted by plant pathogens that is responsible for soft rot disease in plants and vegetables. The lyase utilises anti β-elimination chemistry to cleave an α-1,4-glycosidic link present in polygalacturonate the major component of the plant cell wall.

The structure of BsPel in complex with hexagalacturonate and a cobalt metal has been solved confirming the position and role of the putative catalytic base Arg 279 in the abstraction of a proton from C5 in galacturonate.

Acknowledgements

It has been a real experience jumping on board this PhD train. I remember my first day walking into the lab as a freshly graduated student with my bright white lab coat hopeful and optimistic, but at the same time I knew it wasn't going to be a walk in a park. It has been a rollercoaster of emotions, one minute you're up in the clouds next minute you're down in the dumps.

A big thank you goes to my family especially mum and dad, I hope I've done you proud. Thank you Nathan you have been a great help to me and I believe will continue to be so for a long time still ^_^. Thanks to my childhood friends who lended me their ears listening to me moan and grumble, sorry, you know who you are! Thank you all for your support much much appreciated.

I was lucky enough to land myself into an incredibly friendly and helpful lab full of generous welcoming people. The first of these people is my supervisor Prof Richard Pickersgill, thank you for allowing me to dabble my feet in research and for guiding me throughout. Also a massive thank you must go to Arefeh and Yumi the two most helpful, supportive, strong and understanding people I have met, amazing women! Thank you to everyone in the lab Tom, Kerem, Salyha, Shuang, Allan, Saima, Nadine and Claire a bunch of happy crazy scientists! I'll miss the tea breaks, girly chats and the not so surprise birthday celebrations. Good luck to the ones who have yet to finish their PhDs! Salyha stay off the chocolate and Nadine stay off the coffee!

Finally thank you to Queen Mary University of London and BBSRC for providing the funds and facilities for me to carry out my PhD.

List of Figures

List of Tables

Abbreviations

ALA	Aminoleavulinic acid
DEAE	Diethylaminoethyl
DTT	Dithiothreitol
GalA	Galacturonate
GTP	Guanosine triphosphate
GDP	Guanosine diphosphate
GST	Glutathione-S-transferase
HBA	hydrogenobyrinic acid
HGA	Homogalacturonan
IFD	Insertion of fingers domain
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MES	4-Morpholinoethanesulfonic acid
PBG	Porphobilinogen
PBS	Phosphate buffered saline
PEG	Polyethyleneglycol
SDS	Sodium dodecyl sulphate
SAH	S-adenosyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SUMT	S-adenosyl-L-methionine uroporphyrinogen III methyltransferase
RG-I	Rhamnogalacturonan-I
RG-II	Rhamnogalacturonan-II
RMS	Root mean square
RT	Reverse transcriptase
TBS	Tris buffered saline
TEMED	Tetramethylethylenediamine
TERT	Telomerase reverse transcriptase
TR	Telomerase RNA
Uro'gen III	Uroporphyrinogen III

Chapter One:

Introduction

In this Thesis the structural studies of three enzymes are described. The first results chapter focuses on the cloning and expression of the protein subunit of telomerase reverse transcriptase (TERT). A small telomerase was chosen for study since it may be easier to produce and crystallise and yet still would illuminate the general principles of telomerase action. The next chapter details the structural studies on the methyltransferase CobJ specifically the refinement of a high resolution structure of CobJ and the synthesis of tetrapyrroles for the production of CobJ-tetrapyrrole complexes in the crystal to understand the process of ring contraction, which is unique to cobalamin and catalysed by the enzyme CobJ. In the last results chapter I detail the production and structure of a Michaelis complex of pectate lyase (BsPel) that illuminates the mechanism of this enzyme. This introduction mirrors the three results chapters by providing first background on telomerase before describing the cobalamin methyltransferase CobJ and finally pectate lyase.

1.1 Introduction: Telomerase

1.1.2 Telomerase and the end replication problem

DNA replication is a fundamental biological process that requires extreme accuracy in order to preserve the integrity of the genome from generation to generation. Replication occurs at a specific point called the origin where the DNA double strands unwind forming a replication fork, a leading strand and a lagging strand. In eukaryotes the DNA is linear and due to the uni-directional action of polymerase and the requirement for RNA primer initiation via RNA primase, the resultant daughter strands from the parental lagging strand will have DNA information compromised at its most 5' end (see *Figure* 1.1.1). Continuation of such a trend would ultimately lead to gradual shortening of telomeres and eventual loss of chromosomal DNA and cell senescence (Watson 1972; Blackburn 1984).

Nature has developed a way to combat the end-replication problem by capping the ends of chromosomes with short repetitive DNA sequences thereby maintaining telomere length. Human telomeres are made up of tandem repeats of TTAGGG nucleotides and can range from 15 to 20 kbp at birth and can be reduced to 5 kbp during chronic disease (Collins 2000). In order for telomeres to remain protected and prevent them from being recognised as broken or damaged DNA, the ends form a T loop structure as a defence mechanism (Van steensel 1998). The enzyme responsible for catalysing the repeat synthesis of short tandem DNA repeats is telomerase and this enzyme action was first discovered in *Tetrahymena* (Greider 1985). Telomerase adds short G-rich iterative DNA sequences onto the 3' ends of the parent strands allowing RNA priming and DNA replication of the extreme 5' ends of daughter DNA strands (see *Figure* 1.1.2). Telomerase maintains the chromosome integrity by maintaining telomere length and preventing chromosome end-to-end fusion, without telomerase telomeres will become successively shorter after every cell division which leads to loss of DNA information and eventually replicative senescence.

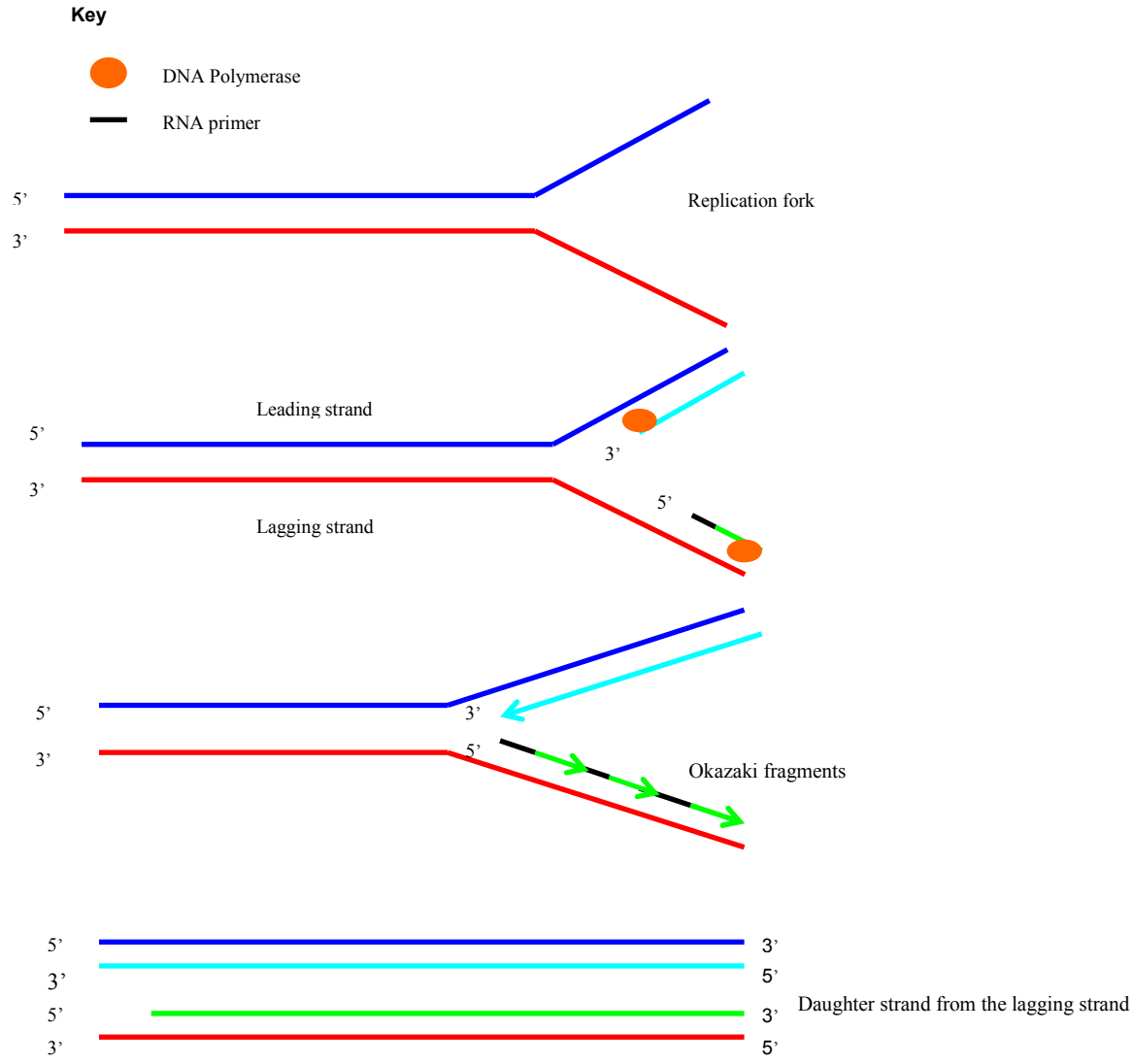


Figure 1.1.1 Diagram depicting the end-replication problem. Double stranded DNA unwinds forming a replication fork separating the strands into a leading strand and a lagging strand. Due to the uni-directional 5’-to-3’ action of DNA polymerase formation of Okazaki fragments occur requiring the removal of primers and joining of fragments via ligase. Lagging strand synthesis requires RNA priming and at the most 5’ end of the daughter strand chromosomal DNA is not copied, figure adapted from (Fanti 1999).

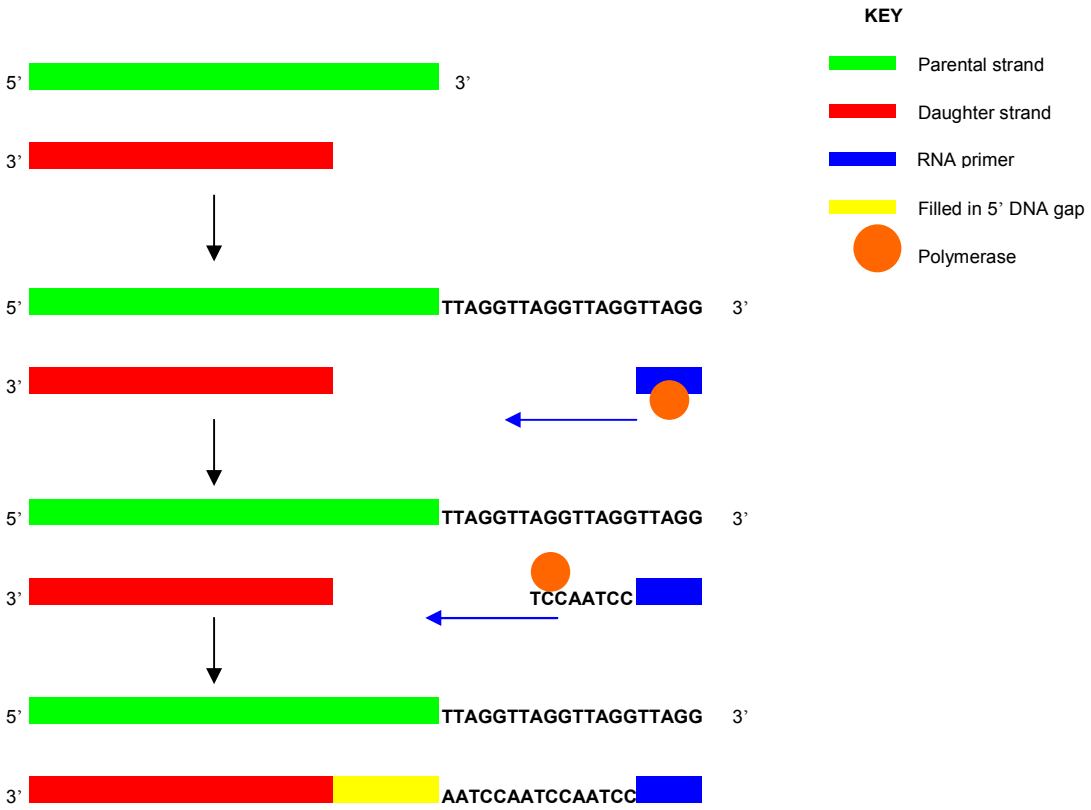


Figure 1.1.2 Telomerase action provides the ability to cap and maintain telomere length by synthesising tandem repeats of a short sequence specific for different organisms. To help circumvent the end replication problem telomerase adds these short repeats to the 3' end of parental strands providing RNA primer initiation and allowing DNA polymerase to read through to the portion of DNA not copied and filling in the gap at the 5' end of the daughter strand, figure adapted from (Autexier 2006).

1.1.3 Telomerase independent survival

Telomerase dependent chromosomal DNA maintenance is by far the most widely studied mechanism for telomere replication and in most cases upregulation of telomerase is found in cancer cells. However, it has been shown in 10-15 % of cancer cells continual telomere lengthening was achieved by an alternative mechanism other than telomerase action, called alternative lengthening of telomeres (ALT) (Cesare 2010). ALT uses recombinational telomere elongation and due to the repetitive nature of telomeres these loci are prone to intra- and interchromosomal recombination. Studies have shown *Sacchromyces cerevisiae* and *Kluyveromyces lactis* mutants lacking telomerase activity use ALT mechanisms (Lundblad 1993).

1.1.4 Telomerase reverse transcriptase subunit

Telomerase is an unusual reverse transcriptase since it possesses its own RNA template, telomerase RNA (TR), which it reverse transcribes from. This subunit along with telomerase reverse transcriptase (TERT) are the two subunits which are minimally required for enzymatic function. Together they work in tandem to synthesise short DNA telomere repeats (Shay 2002).

TERT is the catalytic protein subunit of telomerase identified in a number of organisms including humans, mice, plants, yeast and ciliated protozoa. The gene for TERT was initially cloned from the yeast *S. cerevisiae* and the ciliated protozoa *Euplotes aediculatus* (Lingner 1996; Lingner 1997). Sequence comparison studies have shown telomerase carries the same seven conserved motifs in the central region of TERT (see Figure 1.1.3) that are universally common in all reverse transcriptases (Lingner 1997). Within the central region of TERT are sequence motifs A and C which appear to harbour three aspartate residues important for catalysis also commonly found in other reverse

transcriptases (RTs) such as HIV-1 RT. These motifs are needed for metal binding and polymerase chemistry (Joyce 1994).

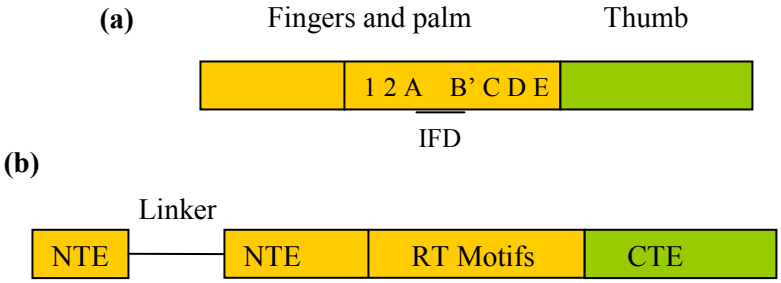


Figure 1.1.3 N-terminal (NTE) and C-terminal domain (CTE) highlighted in yellow and green respectively. **(a)** General reverse transcriptases possess a thumb and fingers and palm domains with seven conserved motifs in the central domain. Insertion in fingers domain (IFD) is found between A and B'. **(b)** TERTs have a similar structure; the only difference is the addition of a large NTE domain separated into a distal and proximal NTE domain joined by a linker and a smaller CTE domain.

The difference in structural features which sets telomerase and other RTs apart is a large insertion between two of the conserved motifs termed A and B'. In retroviral RTs the distance between the two motifs is 20 residues, in TERTs the distance between A and B' ranges from 70 to 120 residues. Since the two motifs are situated in a region called the palm and fingers this insertion is also known as the insertion in fingers domain (IFD). It is thought that IFD is involved in RNA recognition and telomere lengthening (Lue 2003).

Other than the central region of TERT harbouring the seven motifs, there are also N- and C-terminal domains present which serve to provide the different functions of telomerase including maintaining telomere length, integrity and structure. The N-terminal region is approximately 400 amino acids in size and can be divided into two portions; distal and proximal which are separated by a large linker. The C-terminal domain is smaller in size typically 150 amino acids, this holds true for the majority of TERTs examined in a number of studies (Nakamura 1997; Bryan 2000). Two motifs found in the proximal N-terminus are T and CP motifs which are highly conserved among phylogenetic groups

and they appear to be essential for RNA recognition and binding, together these motifs make up the RNA-binding domain (Lai 2001). Both T and CP motifs are involved in telomerase RNA binding, however they appear to bind non-template regions of TR and allow the securing of TR to TERT whilst allowing RNA to move through the active site as confirmed by mutational studies (Bryan 2000). In humans and *Tetrahymena* an additional RID2 (RNA interaction domain 2) is found in the proximal N-terminus involved in RNA template interactions regulating telomerase assembly. It contains a vertebrae specific RNA binding motif and RID2 mutants failed to reconstitute activity (Moriarty 2002).

RNA interaction domain 1 (RID1) is found in the distal portion of the N-terminus separated from RID2 via a linker. RID1 appears to be involved in repeat addition processivity and provides an anchoring point for TR (Moriarty 2004). Within RID1 region is a sequence named N-DAT, which has been reported to dissociate the biological and catalytic activities of telomerase (Counter 1998) and may be important for the recruitment of telomerase to telomeres. Studies have suggested that RID1 may work in tandem with the C-terminal of TERT which is implicated in repeat addition processivity and physical interactions with RID1.

The yeast catalytic subunit is encoded by the EST2 (ever shorter telomeres) gene. Mutational analysis revealed shortening of telomeres and senescence in yeast and so it was confirmed as the TERT component of telomerase, termed Est2p (Lingner 1997). As with other reverse transcriptases, conserved aspartates in motifs A and C are essential for catalytic activity. Within the proximal N-terminal region of yeast, motif GQ has been identified as essential for telomerase catalysis and sequence motifs CP and QFP are required for efficient binding of TR (Bosoy 2003).

In summary TERT is generally made up of four domains including the central portion of TERT harbouring the common RT motifs, distal and proximal N-terminal domains and

the C-terminal domain (Autexier 2006). Within these domains are characteristic sequences making up motifs which carry specific telomerase activities as summarised in *Table* 1.1.1, motifs A and C are commonly found in all RTs.

Domain	Sequence motif	Suggested Function
RT motif	A and C (across all TERTs)	Catalysis
	IFD (across all TERTs)	TR recognition and telomere lengthening
Proximal N-terminal	T and CP (across all TERTs)	Binding of non-template regions TR
	RID2 (humans)	Interacts with TR and regulates telomerase assembly
	CP (yeast and *tetrahymena)*	Catalysis and efficient binding TR
	QFP (yeast)	
Distal N-terminal	RID1 (humans)	Repeat addition processivity
	N-DAT (humans)	Recruitment of telomerase to telomeres
	GQ (yeast)	Catalysis and efficient binding TR
C-terminal	None (humans and yeast)	Repeat addition processivity

Table 1.1.1 Summary of the domains and sequence motifs found in TERT and their suggested functions and roles in telomerase activity.

1.1.5 Telomerase RNA domain

As mentioned earlier telomerase uses an intrinsic telomerase RNA template to reverse transcribe G-rich portions of tandem DNA repeats which are added onto the 3' ends of

telomeres. TR was first discovered in *T.thermophila* (Yu 1990) and subsequently identified in several ciliates, yeast and vertebrates. The size of TR varies widely from 150 nucleotides to over 1300 nucleotides in fungi and the primary sequence of TR differs greatly with little sequence homology between groups, but within groups that are closely related there are some sequence similarities. Phylogenetic analysis has revealed a number of conserved secondary structures making up the core of TR including a template from which telomeric repeats are synthesised which are approximately eight to eleven nucleotides long, a 5' template boundary that regulates the sequence to be transcribed and a putative pseudoknot structure, which together make up a large loop structure by base pair interaction (Romero 1991).

To accomplish template recognition, template boundary sequences are used to enable proper template usage essential for the synthesis of the correct telomere repeats. In most vertebrates long-range base pair interaction forms a P1 helix. In humans, this is divided into two regions; helix P1a and P1b connected via an internal loop. Helix P1 was suggested to have a role in defining the template boundary. To test the function of the P1 helix in humans, truncations of bases were made at various points at the 5' and 3' ends and were used in telomerase reconstitution experiments. The results showed helix P1b and P1a were required for correct template utilisation and are critical for template boundary definition (Chen 2003a). In human TR the pseudoknot and a domain called CR4/CR5 are needed for catalytic activity and TERT binding. The CR4/CR5 domain consists of two helices and a hairpin loop (p6.1). The hairpin is thought to be important for protein interaction (Tesmer 1999; Moriarty 2004). The pseudoknot structure in human TR is highly conserved and is required for telomerase activity, it includes the template region, template boundary region and a folded extended helical arrangement making the pseudoknot (Theimer 2005), see *Figure* 1.1.4.

Various studies ranging from phylogenetic and mutational analyses to NMR have all contributed towards the uncovering of a number of TRs and structures of TR domains.

In *Tetrahymena* telomerase a sequence termed TRE was suggested as the regulating element. Mutation of the TRE sequence prevents effective rounds of telomeric repeat synthesis, but it is still able to bring the template to the active site of telomerase. The mutated TRE sequence prevents accurate priming and in return disallows the promotion of repeat addition processivity (Miller 2002). Another region is *T. thermophila* identified as being important for defining proper template boundary include the tight helix stem-loop II (Richards 2006a). Another TR structure domain from *T.thermophila* identified is the stem-loop IV, this region has been linked to TERT binding and proper pseudoknot folding and processivity. This domain forms a helical structure with a highly kinked region and deletion of this region renders the telomerase inactive, and therefore stem-loop IV must be kinked in order to function (Sperger 2001; Richards 2006b), see *Figure* 1.1.4.

Yeast TR otherwise known as TLC1 telomerase RNA from *S. cerevisiae* is one of the largest at 1157 nucleotides long. The structure contains three large arms that originate from the central core enclosing the template and Est2p the binding site of yeast TERT, see *Figure* 1.1.4. Although TLC1 is large in size and possesses long arms, studies have shown that a smaller approximately 500 nucleotide long mini TR consisting of the template, Est1p, Est2p and Ku arm is able to still function and maintain activity in *vivo* and in *vitro,* however telomeres were shortened and overall cell fitness was reduced (Zapulla 2005). The Est1p and Ku arm appears to be a DNA binding motif which helps recruit telomerase to the telomere ends (Peterson 2001; Lundblad 2003).

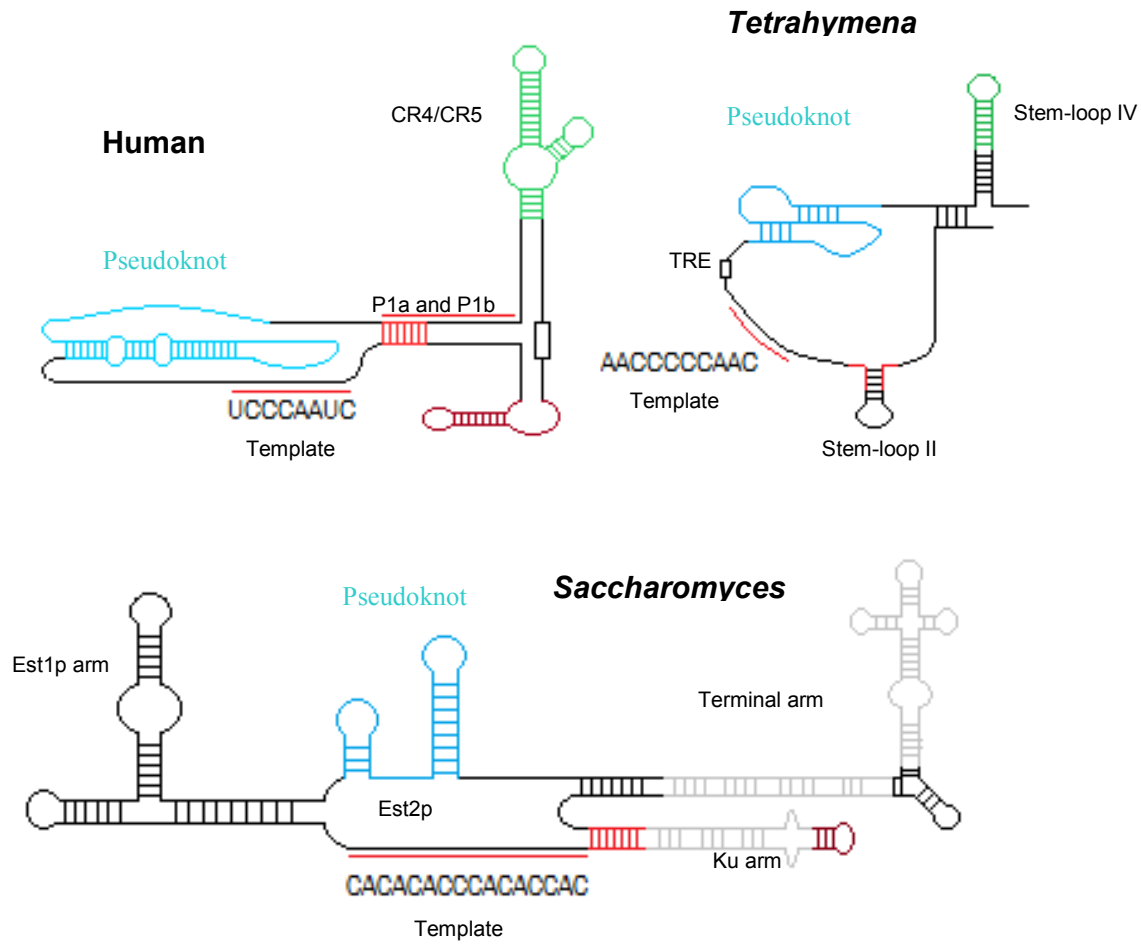


Figure 1.1.4 TRs do not have highly conserved sequences and the sizes vary widely however their secondary structures hold some similarity and a common core consists of a template region, pseudoknot (blue) and template boundary (red). The trans activation domain is shown in green. Here the TRs for human, *Saccharomyces* and *Tetrahymena* are shown. In humans the P1a and P1b helix defines the template boundary and the psuedoknot and CR4/CR5 structure are required for catalytic activity. In *T. thermophila* the template boundary, TRE and stem-loop II helix has been identified and without these elements effective rounds of telomere synthesis can not take place. *S. cerevisiae* TR is one of the largest possessing long arms Est1p, Est2p and Ku. Est2p is the binding site for TERT and the Ku and Est1p arms help recruit telomerase to the ends of telomeres (Autexier 2006).

1.1.6 Repeat addition processivity

Repeat addition processivity is a unique biochemical attribute of telomerase and the mechanism can be broken down into two types of movement; after each nucleotide addition the RNA-DNA complex is simultaneously translocated away from the active site and translocation of the RNA template occurs to realign the 3' end of DNA to 3' end of RNA after each cycle of template copying. The first type of movement has been referred to as type I and the second movement as type II translocation. Type I nucleotide addition is common in all RTs and type II nucleotide repeat addition processivity is only found in telomerase (Collins 1999; Chen 2003b).

The large insertion between motifs A and B' (IFD) located in the RT domain (see Figure 1.1.3) is required for telomerase function in *vivo* and in *vitro* in *S. cerevisiae.* Mutagenesis analysis revealed residues important for primer utilisation and type II translocation (Lue 2003). Another domain identified as important for processivity is RID1 located in the distal portion of N-terminal domain (RNA interaction domain) in human TERT, it interacts with the conserved pseudoknot template in human TR (Moriarty 2004). The C-terminal domain of TERT in humans and *S. cerevisiae* has also been recognized as essential for telomerase catalytic activity and repeat addition processivity (Huard 2003). In total there are three groups of TERT domains or motifs that have been demonstrated to promote repeat addition processivity; IFD, RID1 and C-terminal domain.

Elongation requires the alignment of the 3' end of telomere-primer and TR-template facilitated by an anchor site in TERT. Once the template has been copied the extended telomere dissociates and realigns to the 3' end of the template, see *Figure* 1.1.5 (Chen 2003b). The anchor site is predicted to prevent dissociation of telomerase from its DNA substrate during the translocation step of repeat addition processivity; the anchor site may also facilitate alignment or positioning of primers in the active site and promote the

elongation of partially telomeric and non-telomeric substrates (Harrington 1991; Morin 1991; Wang 1998).

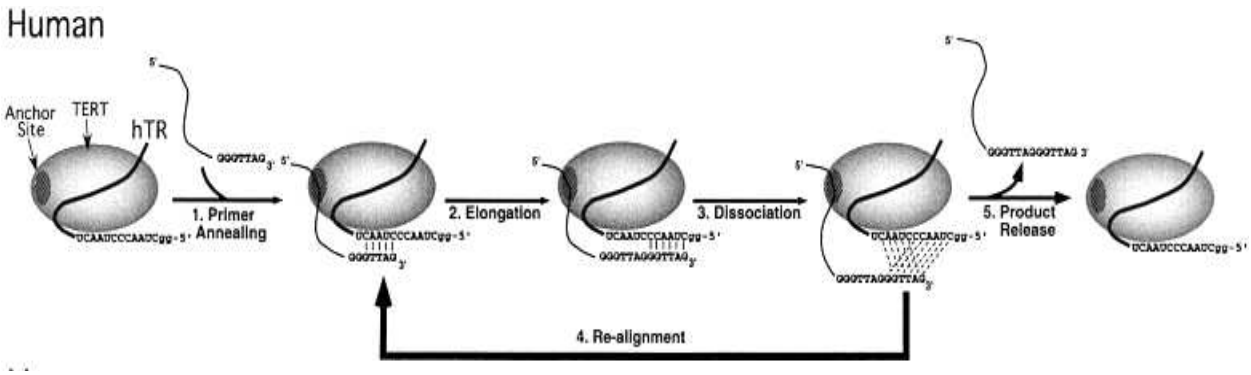


Figure 1.1.5 A model for the repeat addition processivity mechanism for human telomerase. An anchor site tethers the incoming telomere-primer allowing binding to the TR template region. The annealed primer is then elongated followed by dissociation and re-alignment of telomere to the 3' end of TR facilitated by base pair interaction (as represented by the dashed lines). The final step is the release of lengthened telomere ends (Chen 2003b).

1.1.7 Biomedical significance

Telomeres are thought to be the marker that controls the number of cell divisions that a given cell can complete. The Hayflick limit describes the phenomena of the number of times a cell is able to divide before the cell stops dividing. The number of times a cell can divide is linked to the length of the telomeres (Hayflick 1961). Telomerase can provide cells with immortality and in doing so, allow cancerous cells to proliferate.

Telomerase activity is low or undetectable in somatic cells (Wright 1996) but is found active in highly proliferative stem cells, germline, epithilial and haemopoietic cells (Forsyth 2002). Interestingly telomerase activity is upregulated in the majority of cancer cells by approximately 90 %. Screening carried out shows most types of human cancer are associated with the presence of telomerase activity and malignancy (Shay 1997).

Since many cancers have active telomerase function it provides a motivation for cancer therapy development as a drug target. If successful, telomerase can be down-regulated in cancerous cells.

1.2 The cobalamin methyltransferase CobJ

Vitamin B_{12} biosynthesis is achieved via an intricate and complicated pathway harbouring approximately 30 enzyme-mediated reactions. CobJ is one of many enzymes in the vitamin B_{12} pathway and acts as a methyltransferase as well as performing an auxiliary reaction catalysing a unique ring contraction step unprecedented in nature. The following sections discuss the cobalamin biosynthetic pathway and CobJ in more detail.

1.2.1 Vitamin B_{12} pathway

The ability to synthesise vitamin B_{12} (cobalamin), a cobalt-containing modified tetrapyrrole, is only seen in some archaea and eubacteria and no evidence suggests that eukaryotes can synthesise vitamin $B_{12,}$ although it is a known cofactor for a number of enzymes mediating methylation, reduction and intramolecular rearrangements (Kadner 1977). Cobalamin is vital for some protists and humans but plants and fungi neither make nor use vitamin B_{12} and it appears to have no role in some bacterial metabolism (Roth 1996). Cobalamin was first isolated and crystallised in 1948 by two different groups (Rickes 1948; Smith 1948). The steps in the pathway include methylation, ring contraction, amidation and decarboxylation reactions (Scott 2003).

Over time the synthesis of cobalamin has diverged into two routes; an aerobic pathway with late-on cobalt insertion and an anaerobic route with early-on cobalt chelation (Muller 1991; Frank 2005; Heldt 2005). In total there are eight methylations catalysed by six methyltransferases. The first methylations occur at two positions C2 and C7 catalysed by S-adenosyl-L-methionine uroporphyrinogen III methyltransferase (SUMT) also known as CobA which forms precorrin-2 from uroporphyrinogen III (uro'gen III). Only after this point does the pathway take its separate paths into aerobic or anaerobic synthesis of cobalamin, see *Figure* 1.2.1 (Vevodova 2004). The anaerobic equivalent to

CobA is the enzyme CysG. For a summary of the other methyltransferases present in the aerobic and anaerobic pathway see *Table* 1.2.1.

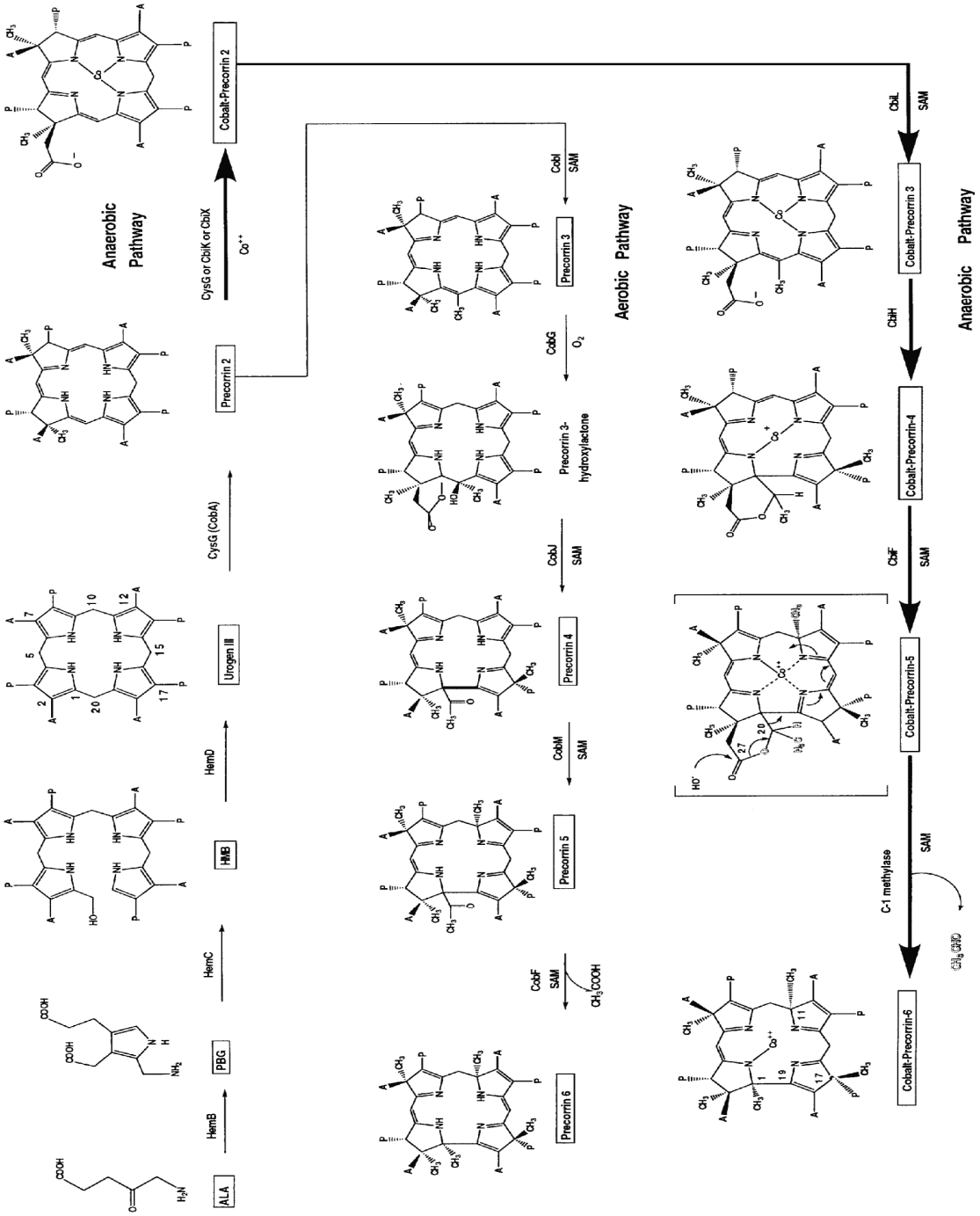


Figure 1.2.1 Incomplete vitamin B_{12} pathway showing the point of diversion from aerobic to an anaerobic pathway after the synthesis of precorrin-2 (Scott 2002).

Aerobic	Anaerobic	Methylation position	Product
CobA	CysG	C2 and C7	Precorrin-2
CobI	CbiL	C20	Precorrin-3a
CobJ	CbiH	C17	Precorrin-4
CobM	CbiF	C11	Precorrin-5
CobF	CbiD	C1	Precorrin-6a
CobL	CbiE and CbiT	C5 and C15	Precorrin-8

Table 1.2.1 Methyltransferases present in the aerobic and anaerobic vitamin B_{12} pathway indicating the site of methylation and the corresponding product.

1.2.3 Modified Tetrapyrroles

All modified tetrapyrroles are derived from uroporphyrinogen III (uro'gen III) (Warren 1990), see *Figure* 1.2.3, which is synthesised from two molecules of 5-aminoleavulinic acid (ALA) via a condensation reaction catalysed by HemB or ALA dehydratase forming porphobilinogen (PBG). The next step in the formation of uro'gen III is the polymerisation of four molecules of PBG via PBG deaminase or HemC forming preuroporphyrinogen, a linear tetrapyrrole. The final step is the generation of the unsymmetrical isomer uro'gen III via uro'gen III synthase (Warren 1990), see *Figure* 1.2.2.

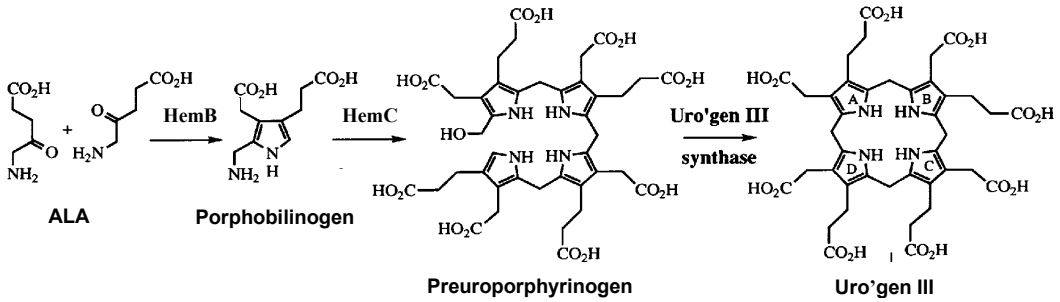


Figure 1.2.2 Schematic diagram depicting the initial steps of uro'gen III production from ALA (Patrick 1996).

Vitamin B_{12} is a large, complex, modified tetrapyrrole belonging to a class of compounds called cyclic tetrapyrroles which includes haem, sirohaem, chlorophyll and cofactor F_{430}, see *Figure* 1.2.3. The core of vitamin B_{12} consists of a corrin ring made up of four pyrrole subunits and in the middle of this ring is a cobalt metal making four bonds with the pyrrole group nitrogens. Cobalamin is generally found in three forms adenosylcobalamin involved in reductase and rearrangement reactions, methylcobalamin involved in transfer of methyl groups between compounds (former and latter being biologically active forms) and cyanocobalamin (commercially prepared vitamin B_{12}) (Ludwig 1997).

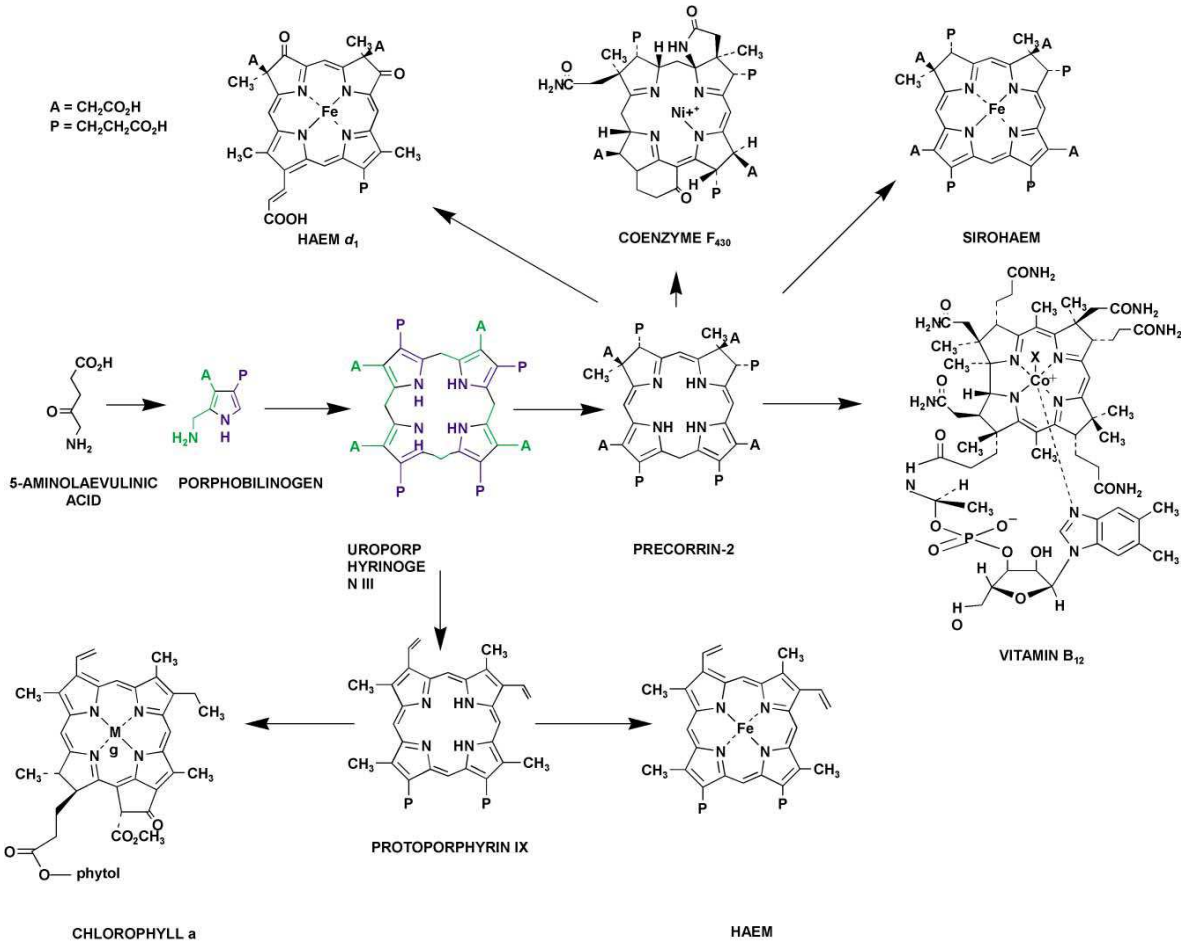


Figure 1.2.3 All modified tetrapyrroles are derived from uro'gen III where it branches into different pathways undergoing transformation into either haem, coenzyme F_{430}, sirohaem, chlorophyll or vitamin B_{12} (Warren 2002).

Adenosylcobalamin has a 5' deoxyadenosyl moiety attached covalently to the upper (Coβ) axial ligand joining it to cobalt and the lower Coα axial ligand is 5,6-dimethylbenzimidazole attached covalently as a loop. Methylcobalamin and cyanocobalamin on the other hand have a methyl group and a cyanide in place of the 5' deoxyadenosyl moiety respectively (Schneider 1987) see *Figure* 1.2.4.

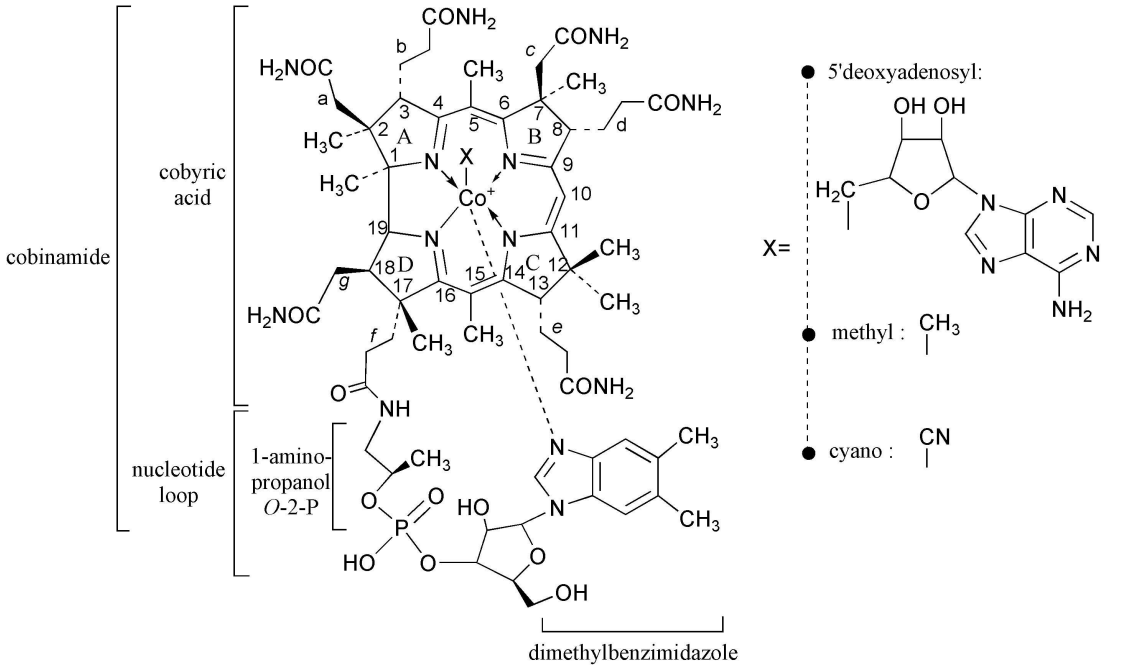


Figure 1.2.4 The structure cobalamin highlighting the corrin ring system, cobalt metal, lower axial ligand 5,6-methylbenzimidazole ribonucleotide and upper axial ligand shown with 'X' can be replaced by either 5' deoxyadenosyl moiety, methyl group or cyanide in the case of adenosylcobalamin, methylcobalamin and cyanocobalamin respectively. The names referring to the derivatives of incomplete cobalamin are also shown (Warren 2002).

1.2.4 Aerobic pathway

As mentioned earlier, cobalamin biosynthesis is divided into two pathways, the difference in the two being the requirement for molecular oxygen and the timing of cobalt insertion. Since a component from the aerobic pathway was studied in this Thesis the next sections look at the steps of the aerobic pathway in more detail.

1.2.4.1 Conversion of uro'gen III into cobyric acid

The branch point between the aerobic and anaerobic pathways occur after the production of precorrin-2; therefore the first step in the biosynthesis of cobalamin is the production of precorrin-2 in both routes. S-adenosyl-L-methionine uro'gen III methyltransferase (SUMT) encoded by CobA gene (Crouzet 1990), catalyses the conversion of uro'gen III into precorrin-2, also known as dihydrosirohydrochlorin, methylated at two positions C2 and C7. The product is also thought to be the last common intermediate for the synthesis of all other modified tetrapyrroles apart from chlorophyll, see *Figure* 1.2.3.

The committed step into aerobic synthesis is via the enzyme CobI which catalyses the SAM-dependant methylation at position C20 see *Figure* 1.2.5. This produces the tetrapyrrole precorrin-3a followed by the insertion of a hydroxyl group via CobG, which is characteristic of the aerobic pathway since there is a requirement for molecular oxygen during this step (Debussche 1993; Scott 1993). The next step is a ring contraction step with the exuding of C20 and methylation at C17 by CobJ forming precorrin-4 (Warren 2002). Precorrin-4 is converted to precorrin-5 by CobM, methylating at C11 followed by deacetylation of C1 methyl ketone and methylation at C1 via CobF provides precorrin-6a (Thibaut 1990; Debussche 1993). A NADPH dependant reduction of the C18-C19 double bond converts precorrin-6a to precorrin-6b catalysed by CobK followed by methylation at two positions, C5 and C15, by CobL (Blanche 1992). The corrin ring synthesis is complete after hydrogenobyrinic acid (HBA) is made via CobH isomerase with the rearrangement of C11 and C12 methyl groups (Thibaut 1990). With these steps in place the corrin ring is now complete forming cobyric acid; the next steps are the ring decoration steps which will form cobinamide.

1.2.4.2 Formation of cobinamide from cobyric acid

The first reaction in the decoration of the corrin ring is catalysed by cobyrinic acid a,c-diamide synthase (CobB), which converts cobyric acid to cobyrinic acid a,c-diamide with amide groups donated by glutamine to positions C2 and C7 producing hydrogenobyrinic acid a,c-diamide (Debussche 1990). The next stage is cobalt insertion by cobalt chelatase which consists of a complex CobN, S and T, as their separate entities the enzymes are inactive but together they catalyse the cobalt chelation step forming co(II)cobyric acid a,c-diamide intermediate late on in the pathway characteristic of the aerobic route (Debussche. 1992). CobR reduces co(II)cobyric acid a,c-diamide to Co(I) which is more nucleophilic and suitable for adenosylation by CobO. The adenosyl group is donated from ATP to the upper axial position of cobalt producing adenosylcob(III)yrinic acid a,c-diamide or ado-cobyrinic acid a,c-diamide (Debussche 1991; Lawrence 2008). The last step of corrin ring decoration is the amidation of side chains via CobQ, see *Figure* 1.2.5 (Blanche 1991). After this point the aerobic and anaerobic pathways rejoin in their route to vitamin B_{12} biosynthesis and four more enzymes are required in the full conversion of uro'gen III to adenosylcobalamin.

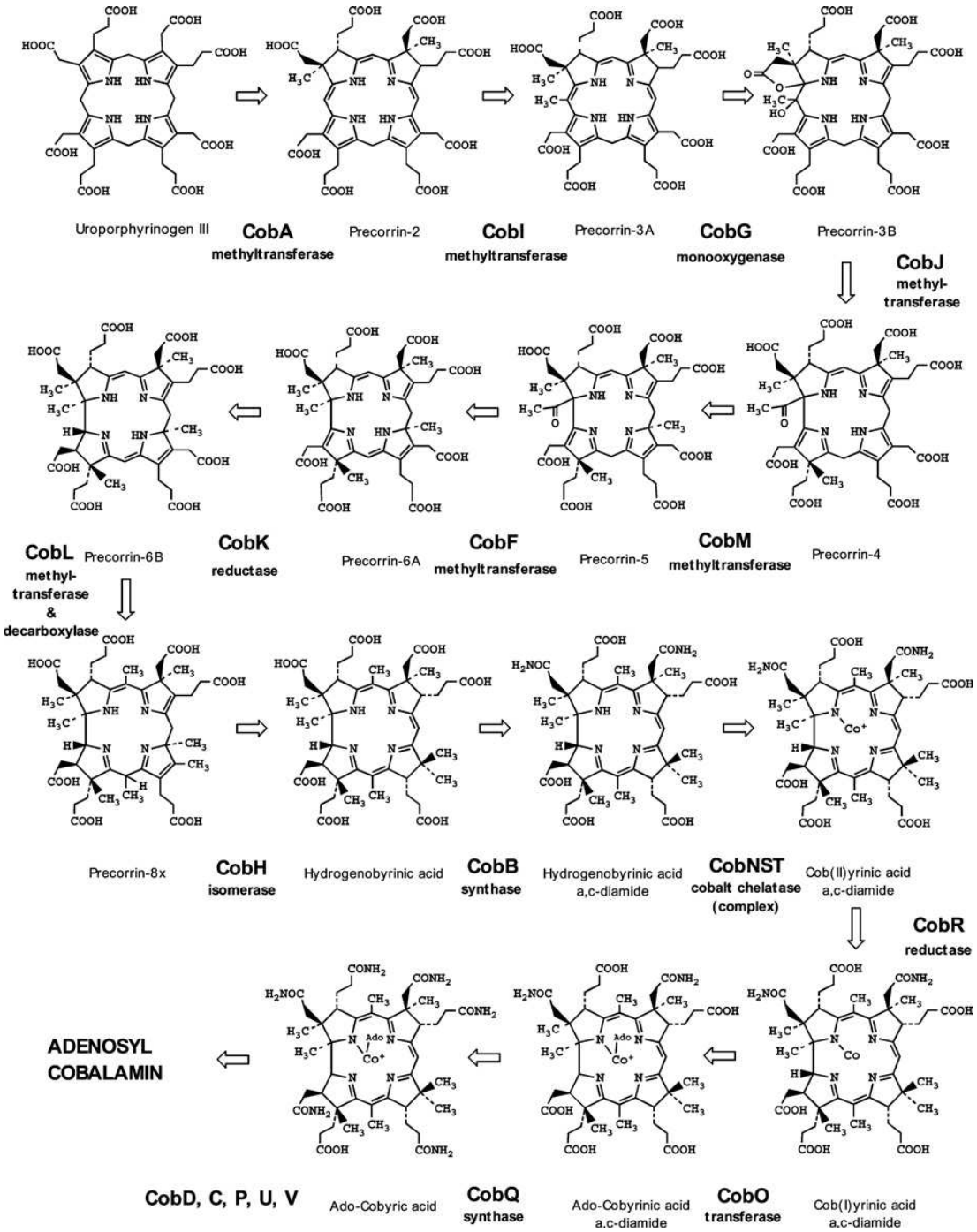


Figure 1.2.5 Summary of the aerobic pathway and the enzymes involved in converting uro'gen III into adenosylcobalamin. The corrin ring synthesis requires the action of eight enzymes CobA through to CobH producing HBA (Heldt 2005).

1.2.4.3 Conversion of cobinamide into adenosylcobalamin

The final steps of converting cobinamide into adenosylcobalamin are depicted in *Figure* 1.2.6 with CobC/D, CobP and CobV. The first step is catalysed by CobC/D where aminopropanol from threonine is transferred to the propionic acid side chain of ring D, (see Figure 1.2.2) and this reaction requires R-1-aminopropan-2-ol and Mg^{2+}/ATP for activity, producing adenosylcobinamide. Following this step is the phosphorylation of the intermediate by CobP and transfer of a guanidine monophosphate group from GTP giving rise to adenosylcobinamide-GDP (Blanche 1991; Crouzet 1991). The final step is the exchange of the guanidine monophosphate group with α-ribazole 5'-phosphate forming cobalamin 5'-phosphate and dephosphorylation of cobalamin 5'-phosphate generating adenosylcobalamin (Cameron 1991).

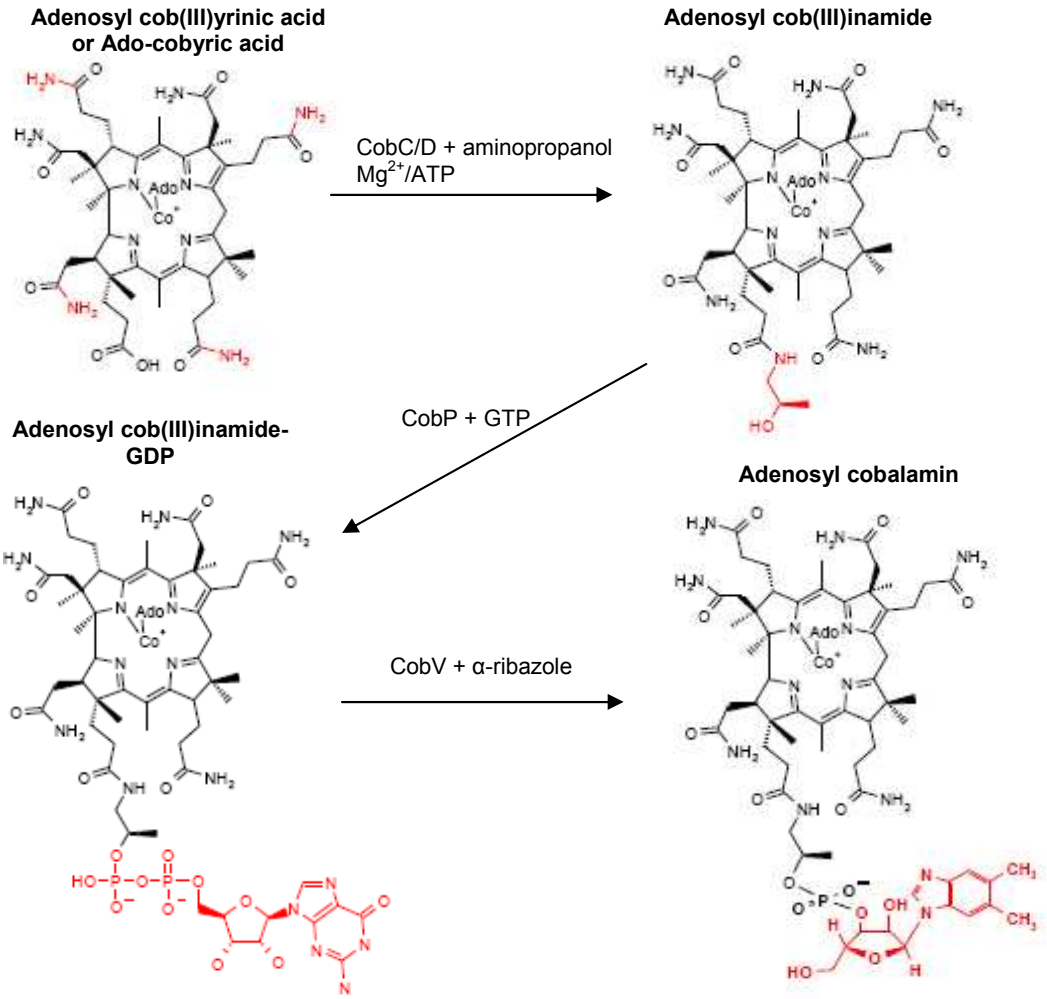


Figure 1.2.6 Diagram depicting the final common enzymatic steps for aerobic and anaerobic biosynthesis of adenosylcobalamin from Adenosyl cob(III)yrinic acid.

1.2.5 Biological significance

Cobalamin was first discovered in 1948 by its ability to cure pernicious anaemia, a condition where vitamin B_{12} is not absorbed into the body and a lack of red blood cells ensues (Smith 1948). In the present day, three classes of vitamin B_{12} enzymes have been recognised and these are adenosylcobalamin dependent isomerases, methylcobalamin dependent methyltransferases and the reductive dehalogenases. Humans require vitamin

B_{12} but are unable to synthesise it, in order to obtain the vitamin it is ingested through foodstuffs rich in the vitamin. The recommended daily intake of vitamin B_{12} is approximately 5 μg a day. In humans cobalamin is needed for methionine synthase (Allen 1993) and methylmalonyl CoA (Ledley 1990). Cobalamin deficiency can manifest as neurological dysfunction, pernicious anaemia and is associated with asthma, depression, multiple sclerosis and tinnitus.

1.3 Pectate lyase

Pectate lyase is secreted from bacteria causing soft rot disease and crop spoilage. The last results chapter in this Thesis focuses on the enzyme pectate lyase from *Bacillus subtilis* and the reaction mechanism used for cleavage of glycosidic bonds.

1.3.1 Plant cell wall

The cell wall is important for maintaining the form and structural elements of the plant, it is made up of two types of polysaccharide networks, pectin and cellulose (Carpita 1993; Vincken 2003). The pectin network is the major component of the plant cell wall in dicotyledonous plants forming the middle lamella and is composed of up to 17 different monosaccharides. The monosaccharides are organised into different smooth and hairy regions forming homogalacturonan (HGA), rhamnogalacturonan-I (RG-I) and rhamnogalacturonan-II polysaccharide backbones (RG-II) (O'Neill 1990; Mohnen 1999).

HGA is a polysaccharide made up of GalA (galacturonate) residues α-1,4 linked together and is synthesised by the Golgi apparatus followed by the deposition in the cell wall (Zhan 1998). RG-I is composed of a repeating backbone disaccharide 4)-α-D-galacturonic acid-(1,2)-α-L-rhamnose(1 (Vincken 2003). RG-II has a backbone of at least eight GalA residues α-1,4 linked with four structurally distinct oligoglycosyl side chains of known consistent lengths attached to the backbone (Vidal 2000). Enzymes which degrade the cellulose and pectate network belong to two different classifications, the hydrolases and polysaccharide lyases, respectively.

1.3.2 CAZy database

A CAZy (Carbohydrate-Active Enzymes Database) classification system identifies each polysaccharide lyase as a member of a PL family based on the amino acid sequence similarities reflecting the structural features (Cantarel 2009). There are 22 families of polysaccharide lyases based on the CAZy classification database and only three types of topologies have been observed for pectate lyases, the $(\alpha/\alpha)_7$ barrel in PL-2, parallel β-helix observed in PL-1, 3 and 9 and $(\alpha/\alpha)_3$ barrel in PL-10. BsPel belongs to the polysaccharide lyase family 1 (PL-1). All enzymes belonging to PL-1 carry out a β-elimination reaction on the glycosidic bond with subsequent formation of a carbon-carbon double bond at the non-reducing end.

1.3.3 Pectic enzymes

Enzymes secreted from plant pathogens cause devastating diseases in plants and vegetation such as soft rot which destroys the integrity of plant tissues. There are a number of enzymes capable of crop spoilage.

Pectins are naturally degraded by pectic enzymes and there are three types of de-esterifying enzymes known as; pectinesterases, depolymerising enzymes (hydrolases and lyases) and protopectinases. BsPel is a depolymerising enzyme, therefore the next sections cover hydrolase and lyase enzymes in more detail.

1.3.3.1 Hydrolysis of the glycosidic bond

In nature oligosaccharides and polysaccharides can be assembled in many different ways, adopting a multitude of stereochemical variations. Living organisms take advantage of this ability and use oligosaccharide and polysaccharides for specific roles in storage, structure and signalling (Laine 1994).

Glycosidic bond hydrolysis commonly requires a proton donor and a nucleophile (base). There are two possible mechanisms, retaining and inverting which refer to the retention or inversion of the configuration about the anomeric carbon centre. In the former mechanism the base is in close vicinity of the sugar anomeric carbon and in the inverting reaction the base is further away to allow for the accommodation of a water molecule between the base and the sugar, see *Figure* 1.3.1. The positioning of the proton donor is the same for both retaining and reverting mechanisms where both proton donors are within hydrogen-bonding distance of the glycosidic oxygen (Davies 1995).

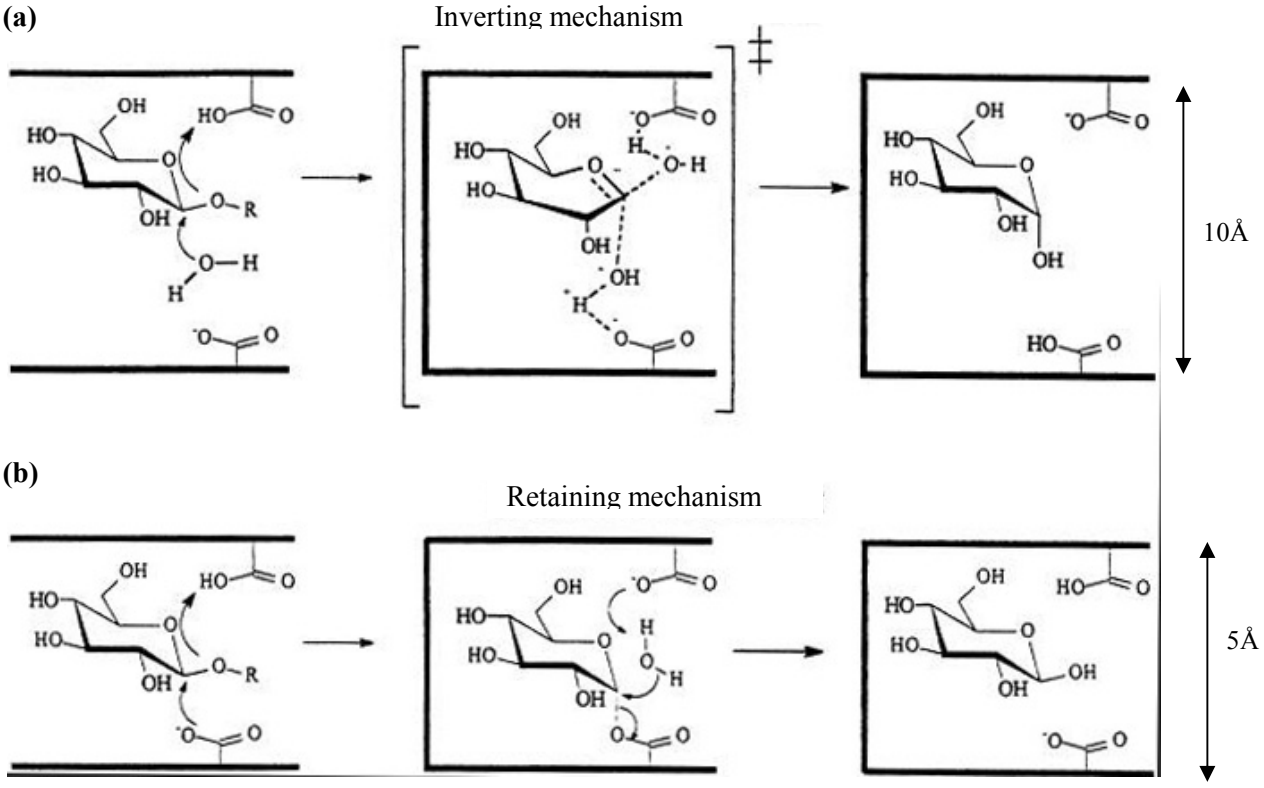


Figure 1.3.1 The distances between the catalytic residues are 5 Å and 10 Å for retaining and inverting mechanisms respectively. **(a)** The inverting mechanism requires the attack of water molecule protonating the glycosidic oxygen. **(b)** In the retaining mechanism the glycosidic oxygen is protonated by an acid catalyst and a base assists in the departure of galacturosyl group (Williams 2009).

1.3.3.2 Elimination of the glycosidic bond

Pectate lyases, otherwise known as pectate transeliminases, were first discovered in *Erwinia carotovora* and *Bacillus polymyxa* (Starr 1962). The enzyme is known to catalyse the eliminative cleavage of pectin (Carpita 1993). Unlike hydrolases, lyases have a basic pH optimum. The enzyme harnesses an *anti* β-elimination reaction which involves the abstraction of the α-proton from C5 of polygalacturonic acid resulting in the elimination of the β-leaving group and cleavage of the C4 glycosidic oxygen bond. This generates a 4,5-unsaturated galacturonosyl product residue at the non-reducing end of the polysaccharide see *Figure* 1.3.2. The reaction requires the presence of calcium ions (Collmer 1986). This mechanism uses a base in close vicinity to the C5 proton adjacent to the C4 glycosidic oxygen bond to be cleaved, previous studies have shown the base to be a conserved arginine, and mutation of this catalytic amino acid residue renders the enzyme inactive (Charnock 2002). Protonation of the leaving group is also expected to be a feature of a catalyst accelerating the rate of reaction.

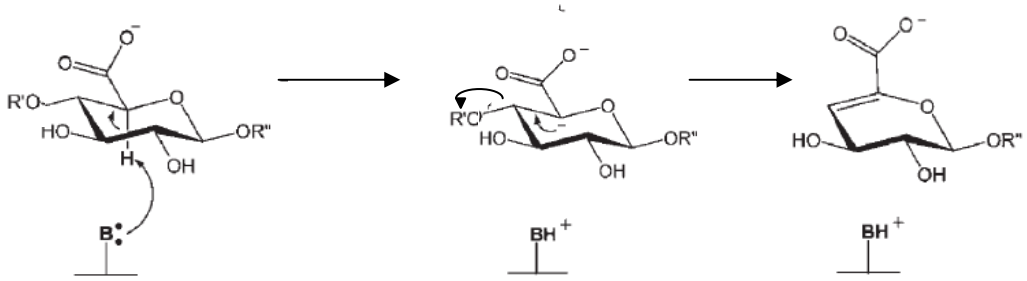


Figure 1.3.2 General *anti* β-elimination mechanism adopted by lyases making use of a catalytic base to abstract a proton from C5 cleaving the glycosidic bond, figure adapted from (Yip 2004).

1.3.4 The structures of polysaccharide lyases

Pectate lyase PelC (Yoder 1993) and PelE (Lietzke 1994) both from *Erwinia chrysanthemi* and BsPel from *Bacillus subtilis* (Pickersgill 1994) all have the then-unusual structure consisting of β strands folded into a large right-handed helix termed the parallel β-helix. Later another two topologies were identified in pecate lyase; $(\alpha/\alpha)_3$ barrel and $(\alpha/\alpha)_7$. Other enzymes which break the glycosidic bond with the same parallel β-helix structure are summarised in *Table* 1.3.1.

Enzyme	**Short name**	**Origin**	**Family**	**Reference**
Pectate lyase	PelC	*E. chrysanthemi*	PL-1	(Yoder 1993)
	BsPel	*B. subtilis*		(Pickersgill 1994)
	PelE	*E. chrysanthemi*		(Lietzke 1994)
	PelA	*E. chrysanthemi*		(Thomas 2002)
	Pel-15	*Bacillus* sp. KSM-P15	PL-3	(Akita 2001)
	PelI	*E. chrysanthemi*		(Creze 2008)
	Pn1B	*Aspergillus niger*		(Vitali 1998)
	Pel9A	*E. chrysanthemi*	PL-9	(Jenkins 2004)
Rhamnogalacturonase	RGase A	*Aspergillus aculeatus*	PL-4	(Petersen 1997)
Polygalacturonase	PehA	*Erwinia carotivora*	GH-28	(Pickersgill 1998)
Polygalacturonase II	PG II	*A. niger*		(Van Santen 1999)
Pectin methylesterase	PemA	*E. chrysanthemi*	CE-8	(Jenkins 2001)

Table 1.3.1 Pectic enzymes with the parallel β-helix structure as identified from the CAZy database. The families which possess the parallel β-helix fold include pectate lyases (PL), carbohydrate esterases (CE) and glycoside hydrolases (GH).

The crystal structure of PelC mutant in complex with a pentagalacturonate sugar has been solved and the fragment binds in a pocket interacting with positively charged amino acids and four calcium ions. From this structure Arg218 was highlighted as being responsible for initiating proton abstraction during β-elimination of the glycosidic bond, since the R218K mutation rendered the enzyme inactive (Scavetta 1999).

Cevibrio japonicus lyase Pel10A from PL-10 was solved revealing a topology that was predominantly α-helical with a distorted $(\alpha/\alpha)_3$ barrel. A Michaelis complex of an inactive mutant with trigalacturonate and calcium ions bound confirmed the importance and catalytic activity of Arg254 (Charnock 2002). Comparison with an inactive mutant of Pel1C (Herron 2000) from the PL-1 family in complex with tetragalacturonate and calcium ions, revealed an identical location for the putative catalytic bases Arg254 and Arg218 for Pel10A and Pel1C respectively. The mechanism involving the use of an arginine acting as a base is supported by the observation of converging geometry in the active site as seen in the comparison between PL-10 and PL-1 family lyase structures (Charnock 2002).

1.3.5 Industrial significance

Pectate lyases play a pivotal role in recycling and remodelling of plant material which is important for maintaining the biosphere (Scavetta 1999). Apart from acting as cementing agents in the plant cell wall, pectins also play a role in the texture of fruits and vegetables. Industrially, pectic substances account for one-quarter of the world's food enzyme production and are used frequently in the food industry for increasing the yield and clarity of fruit juices and as gelling components (Panchev 1988; Alkorta 1998). Other areas of commercial use include the textile industry, fermentation of coffee and tea, oil extractions and treatment of pectic waste water (Kashyap 2001). There is potential to exploit the enzymes, rather than using crude preparations, for the more selective modification of pectins and the production of more specific pectin additives and ingredients. In the developed world, phytopathogenic bacteria can devastate crops, Erwinia infections are now well established in Europe as the summers become milder and wetter, so inhibitors based on transition state analogues may be useful in protecting the security of our crop supply.

Chapter Two:

Materials and methods

2.1 Materials

2.1.1 Water

The water used for all crystallisation experiments was purified to 18.2 MΩ.cm (ddH_2O) and all buffers were made up using water purified to 15.0 MΩ.cm (dH_2O) both from Purelab ELGA water purifier system. For molecular biology molecular grade water was used where possible or autoclaved ddH_2O.

2.1.2 Plasmids

All plasmids used for cloning were pET vectors from Novagen (Table 2.1) see Appendix 1 for vector maps.

Vector	Promoter	Antibiotic resistance	Tags
pET3d	T7	ampicillin	N terminal T7 tag
pET14b	T7	ampicillin	N terminal His tag
pET41	T7 *lac*	kanamycin	N terminal GST tag
pETcoco2	T7	ampicillin	N terminal His tag

Table 2.1 List of pET vectors used for molecular cloning.

2.1.3 Polymerases

A number of polymerases were used for cloning and site-directed mutagenesis reactions including KOD Hot start DNA polymerase (Novagen), PfuTurbo DNA polymerase (Stratagene) and HotStar Taq (Qiagen).

2.1.4 Restriction Enzymes

Restriction enzymes were used to cut specific sites by recognising characteristic sequences of nucleotides. *Table* 2.2 lists all the restriction enzymes and sites used.

Restriction enzyme	Restriction site
NdeI	5' CA/TATG 3' 3' GTAT/AC 5'
BamHI	5' G/GATCC 3' 3' CCTAG/G 5'
XhoI	5' C/TCGAG 3' 3' GAGCT/C 5'
BlpI	5' GC/TNAGC 3' 3' CGANT/CG 5'

Table 2.2 The restriction enzymes and restriction sites used where '/' indicate the site of cleavage (New England BioLab).

2.1.5 Luria Bertani medium

Medium used for bacterial cell growth was Luria Bertani medium (LB medium). Tryptone and yeast extract were obtained from Oxoid (Thermo Fisher Scientific, UK), salt was from Fisher scientific. LB was made up to one litre using deionised water 15.0 MΩ.cm purity and autoclaved to sterilise, see below for details.

Per Litre
5 g NaCl
10 g Tryptone
10 g Yeast extract

Agarose plates were made using the same LB medium but with added 1.5 % agar (Sigma Aldrich) sterilised by autoclaving.

2.1.6 Protein Purification

2.1.6.1 Buffers

All buffers were made using deionised water of 15.0 MΩ.cm purity and filtered using 0.2 µm filter and buffers were pH corrected using 5 M NaOH or neat HCl using a MP230 pH meter from Mettler Toledo. For detailed composition of buffers used for protein purification see *Tables* 2.3 to 2.5.

Buffer name	Buffer composition	pH
Binding buffer	5 mM Imidazole 20 mM Tris 400 mM NaCl	8.0
Wash buffer I	50 mM Imidazole 20 mM Tris 400 mM NaCl	8.0
Wash II	100 mM Imidazole 20 mM Tris 400 mM NaCl	8.0
Elution buffer	400 mM Imidazole 20 mM Tris 400 mM NaCl	8.0

Table 2.3 The buffers required for nickel immobilised purification. Imidazole was from Sigma Aldrich, Tris and NaCl were from Fisher Scientific.

Buffer name	Buffer composition	pH	Usage
Buffer A	20 mM MES	6.0	Cation exchange
Buffer B	20 mM MES 1 M NaCl	6.0	
Buffer A	20 mM Tris 100 mM NaCl	8.0	Size exclusion
Wash	PBS	7.5	Glutathione sepharose
Elution	10 mM reduced glutathione 50 mM Tris	8.0	
Low salt	20 mM Tris 100 mM NaCl	8.0	pd-10
High salt	20 mM Tris 500 mM NaCl	8.0	
Activation	Ethanol 1 % Acetic acid	—	CP-18
Wash	1 % Acetic acid	—	
Elution	90 % Ethanol	—	

Table 2.4 The buffers used for various purification procedures. MES was from Fisher Scientific, PBS (phosphate buffered saline) tablets were purchased form Sigma-Aldrich and reduced glutathione was from Calbiochem.

Buffer name	Composition	pH
Wash I	20 mM Tris	8.0
Wash II	20 mM Tris 100 mM NaCl	8.0
Wash III	20 mM Tris 200 mM NaCl	8.0
Wash IV	20 mM Tris 300 mM NaCl	8.0
Elution	20 mM Tris 400 mM NaCl	8.0

Table 2.5 The buffers required for purification on a DEAE (Diethylaminoethyl Sephacol) column.

2.1.6.2 Columns

The different types of columns used during the entirety of this thesis are listed in *Table* 2.6. All columns and resins were from GE Healthcare with exception to DEAE column which is from Sigma Aldrich and CP-18 column from Waters.

Type of column	Column name
Size exclusion	Superdex 200 10/300 GL
Affinity	HiTrap chelating HP
Affinity batch method	Glutathione sepharose Immobilised nickel
Ion exchange	HiTrap SP HP DEAE
Reverse phase	CP-18
De-salting	pd-10

Table 2.6 Columns used for various purification procedures.

2.1.7 Competent cells

Transformations were carried out using *E. coli* cells where up-taking of recombinant plasmid DNA allows for cloning and expression by using the host's machinery. Various competent cells were used for replication and protein production, see *Table* 2.7.

Purpose	Competent cells
DNA replication	XL-1 Blue Nova Blue α- Select Nova F-
Protein production	BL21 (DE3) BL21 (DE3) pLysS BL21 star (DE3) pLysS BL21 codonplus (DE3)- ril Origami (DE3) Rosetta Blue (DE3) pLysS C43 (DE3) pLysS

Table 2.7 A summary of the competent cells used for cloning and expression. All cells were from Novagen except for α-Select (Bioline), BL21 codonplus (DE3)-ril (Stratagene) and BL21 star (DE3) pLysS (Invitrogen).

2.1.8 Antibodies

Primary antibody anti-His mouse from Novagen was used at a working dilution of 1:1000. The secondary antibody was goat anti-mouse IgG alkaline phosphatase conjugate from Novagen used at 1:5000 dilution. Colourimetric alkaline phosphatase detection tablets SigmaFast BCIP/ NBT were purchased from Sigma-Aldrich.

2.1.9 Antibiotics

Antibiotics were used during DNA cloning and expression of cells, a list of antibiotics with the stock and working concentrations can be found at *Table* 2.8.

Antibiotics	Stock concentration (mg/ ml)	Final concentration (µg/ml)
Ampicillin	100	100
Chloramphenicol	35	35
Tetracycline	12.5	12.5
Kanamycin	30	30

Table 2.8 Antibiotics used and the corresponding stock and final concentrations.

2.1.10 Crystallisation

2.1.10.1 Buffers

For the purpose of crystallisation all buffers were made from deionised water of purity 18.2 MΩ.cm (see *Table* 2.9).

Buffer	Stock concentration (M)	pH	Stock (%) w/v	Purchased from
Sodium acetate	1	4.6	__	BDH
Tris	1	8.5	__	Fisher Scientific
Ammonium acetate	1	4.6	__	BDH
Sodium cacodylate	0.1	6.5	__	Fluka
Calcium acetate	1	__	__	BDH
PEG 4000	__	__	50	Hampton Research
PEG 8000	__	__	40	Hampton Research

Table 2.9 Buffers and precipitants used during optimisation crystallisation trials.

2.2 Methods

2.2.1 Molecular cloning and site-directed mutagenesis

2.2.1.1 Primer design

Parameters were followed for the design of primers for molecular cloning such that primer length was between 25 to 30 base pairs long, the ends of the primers ended with a G (guanine) or a C (cytosine), the overall GC content was 40 to 60 %, melting temperature was between 52 to 58° C, complimentarity between the forward and reverse primers were minimised and ensuring no frameshift took place. By adhering to such rules formation of primer-dimers and hair-pin loops were minimised and allowed for optimum primer annealing parameters and primer stability to be obtained. A free program called NetPrimer (www.premierbiosoft.com) was used to check these parameters.

Primers used for site-directed mutagenesis were designed in the same way as described above with the exception to a change of a single base situated near the middle of the sense and antisense primers to allow for a point mutation.

The primers used for molecular cloning and site-directed mutagenesis can be found in section 3.2.1 and 5.2.1. All primers were ordered from MWG Biotech. Primers were stored at -20 °C at 100 mM stock and diluted to 10 mM using molecular grade water when required for PCR reactions.

2.2.1.2 Polymerase chain reaction

The general cycling parameters for the different DNA polymerases used are listed in *Table* 2.10. All PCR reactions were carried out using a PTC-150 MiniCycler by MJ Research.

A typical PCR reaction set up consists of the following:

2 µl DNA template
5 µl 10× reaction buffer
1.5 µl forward primer
1.5 µl reverse primer
1 µl dNTP
38 µl ddH_2O
1 µl polymerase

DNA polymerase	Step	Temperature (ºC)	Time
PfuTurbo polymerase (Stratagene)	1. Activation	95	2 min
	2. Denaturation	95	30 sec
	3. Annealing	Primer T_m-5 ºC	30 sec
	4. Elongation	72	1 min for targets ≤1 kb
	5. Final elongation	72	10 min
	Repeat steps 2 to 4 for 30 cycles.		
KOD Hot Start polymerase (Novagen)	1. Activation	95	2 min
	2. Denaturation	95	20 sec
	3. Annealing	Lowest primer T_m	10 sec
	4. Elongation	70	20 sec/kb
	Repeat steps 2 to 4 for 20 to 40 cycles.		
HotStar Taq polymerase (Qiagen)	1. Activation	95	15 min
	2. Denaturation	94	0.5 to 1 min
	3. Annealing	50 to 68	0.5 to 1 min
	4. Elongation	72	1 min
	5. Final elongation	72	10 min
	Repeat step 2 to 4 for 25 to 35 cycles.		

Table 2.10 Summarises the general cycling conditions for the different DNA polymerases used for cloning, for specific optimised cycling parameters refer to later results chapters.

2.2.1.3 Direct cloning

Direct cloning required the introduction of two unique restriction sites to either side of the gene by PCR amplification. Choosing the appropriate restriction sites were important and care was taken to ensure these sites were not present in the gene of interest, but found on the destination vector to allow for ligation. To achieve ligation the PCR reaction mixture was gel purified first using a Qiagen gel purification kit followed by double digestion of the gene of interest and vector using suitable restriction enzymes. These were ligated using Takara ligase (Takara Bio Inc).

Double digestion and ligation set-ups are described below, restriction enzymes and buffers were from NEB. Subsequent ligated insert and vector were transformed into DNA cloning competent *E. coli* cells.

Double digestion of plasmid vector or gene insert:

10 µl vector or PCR product of gene
1 µl 10× buffer
1 µl restriction enzyme one
1 µl restriction enzyme two
7 µl dH_20

Ligation

7 µl gene insert
1 µl vector
8 µl Takara

The ligation mix was left to incubate at 16 ºC for 30 minutes using the PCR machine.

2.2.1.4 Digestion and ligation independent cloning

This particular technique uses an Ek/LIC cloning system (Novagen) which carries the advantage of eliminating the need for digestion and ligation. This is valuable since the number of steps are reduced thereby making the cloning process quicker and reducing the likelihood of errors.

The primers used for Ek/LIC system are specially designed to incorporate extensions compatible with the vector pET41 (see section 3.2.2). KOD Hot Start polymerase was used for insert amplification, the general cycling parameters can be found in *Table* 2.10. The PCR mixture and subsequent steps for Ek/LIC strategy cloning are listed as follows.

PCR reaction mixture

5 µl Buffer
3 µl $MgSO_4$
5 µl dNTP
1.5 µl forward primer
1.5 µl reverse primer
2 µl DNA template
1 µl KOD Hot Start polymerase
31 µl dH_2O

T4 Treatment

2 µl insert
2 µl 10× T4 DNA polymerase buffer
2 µl dATP
1 µl DTT
12.6 µl dH_2O
0.4 µl T4 DNA polymerase

The mixture was incubated for 30 minutes at 22 ºC to activate the enzyme and then incubated at 75 ºC for 20 minutes to inactivate the enzyme.

Annealing

1 µl Ek/LIC vector pET41

2 µl insert

This was incubated for five minutes at 22 ºC then 1 µl of EDTA was added and further incubated for another five minutes at 22 ºC. The ligated product was then ready for transformation (see 2.2.3).

2.2.2 Site-directed mutagenesis

For site-directed mutagenesis PCR reactions PfuTurbo polymerase or KOD Hot Start polymerase was used as listed in *Table* 2.10 with the general cycling parameters. For different target genes specific cycling parameters were used see section 3.2.2 and 4.2.7.

Once the target gene was amplified with the appropriate primers for single base mutation 1µl DpnI was added, gently mixed and incubated at 37 ºC for one hour to digest the parental non-mutated template. The DpnI-treated DNA was then used in a transformation.

2.2.3 Transformation

Cloned genes were transformed into *E. coli* cells for DNA replication followed by plasmid isolation using QIAprep Spin Miniprep Kit from Qiagen. For protein production the plasmid was transformed into the appropriate *E. coli* competent cells. The general transformation protocol used during the entirety of the project for different cell lines is as follows:

1. 50 µl of cells were thawed on ice and 1 µl plasmid DNA was added to the cells and left on ice for five minute.
2. The cells were then heat-shocked (see *Table* 2.11).
3. Immediately after heat treatment the cells were placed on ice for two minutes.
4. 250 µl of LB media was added and incubated for one hour at 37 ºC with shaking at 200 rpm.
5. 50 µl of the cells were plated onto LB agarose plates with the appropriate antibiotics, the remaining cells were centrifuged and most of the supernatant was decanted and the pellet was resupsended. The plates were left in a 37 ºC incubator overnight.

Competent cells	Heat-shock temperature (ºC)	Duration (sec)	Antibiotic resistance
XL-1 Blue	42	45	__
Nova Blue	42	30	__
Nova F	42	30	__
α-Select	42	45	__
BL21 (DE3)	42	30	__
BL21 (DE3) pLysS	42	30	Cam
BL21 star (DE3) pLysS	42	30	Cam
BL21 codonplus (DE3)-ril	42	20	Cam
Origami (DE3)	42	30	Tet, Str, Kan
Rosetta Blue (DE3) pLysS	42	30	Tet, Cam
C43 (DE3) pLysS	42	30	Cam

Table 2.11 Summarises the heat-shock temperatures and duration, also listed are the antibiotic resistance of each cell line.

2.2.4 Protein production

2.2.4.1 Protein growth conditions

Starter cultures of 5 ml of LB media supplemented with the appropriate antibiotics were inoculated and grown overnight at 37 ºC. In baffled two litre flasks the starter cultures were used to start protein growth in a litre of LB media at variable temperatures depending on the cell strain used and protein growth requirements. The cells were then induced with IPTG once an $O.D_{600}$ of 0.6 was reached. Induction took place at a suitable temperature and duration. For specific growth and induction conditions see sections 3.3.2, 3.3.3, 4.2.1 and 5.3.1.

Cells were then harvested by centrifugation at 5,000 rpm for 20 minutes using a Sorvall RC-5B refrigerated centrifuge from Du Pont Instruments. The pellet was stored in a falcon tube at -80 ºC.

2.2.4.2 Cell lysis

To lyse the cells from crude lysate a French press from, SLM-Aminco, Rochester, N.Y was used. The system used high pressures of 20,000 psi to disrupt the cells, for sufficient lysis the crude lysate was passed through the press twice. To pellet cell debris the solution was centrifuged at 18,000 rpm for 20 minutes (Sorvall RC-5B).

An alternative method for releasing protein was sonication. A sonicator from sonics VibraCell was used to pulse the resupsended pellet in binding buffer at 15 second increments and then centrifuged at 18,000 rpm for 20 minutes (Sorvall RC-5B) to pellet the cell debris. The lysate was ready for column application.

2.2.5 Protein purification

2.2.5.1 Nickel immobilised column

For His-tagged protein, cell pellets were resuspended in binding buffer (see Table 2.3) and lysed, the lysate was loaded onto a nickel column. Two types of nickel columns used; batch method with nickel immobilized on a column of Fast Flow Chelating Sepharose from GE Healthcare and AKTA assisted HiTrap chelating HP column. Both methods required charging of the column with $NiSO_4$.

The method for $NiSO_4$ charging for batch method is listed below.

1. Load 4 ml of Fast Flow Chelating Sepharose into a clean plastic column.
2. Wash the beads with five column volumes of dH_2O.
3. Load 2.5 ml of 500 mM $NiSO_4$.
4. Wash the beads with five column volumes of dH_2O followed by five column volumes of binding buffer to equilibrate.

To charge HiTrap chelating HP column the same procedure was followed from step number two, but two column volumes of $NiSO_4$ were used for charging the column.

Once charged the lysate was loaded onto the column with extensive washing using binding buffer and then a gradient elution was used for HiTrap chelating HP column with increasing concentrations of imidazole ranging from 5 mM to 400 mM. The batch method required washing and elution sequentially using the buffers listed in *Table* 2.3.

2.2.5.2 Glutathione sepharose column

For GST-tagged proteins a batch method glutathione sepharose column was used. In a clean plastic column 3 ml of glutathione sepharose beads from GE Healthcare were loaded and washed with 30 ml of PBS. Lysate was then added to the column and washed with 60 ml of PBS. Reduced glutathione was added at 1.5 ml and incubated for ten minutes at 4 ºC (see *Table* 2.4 for list of buffers used). The fraction was then collected, if necessary the incubation step was repeated until no more protein was eluted. This was checked using a BioRad assay (Bradford 1976).

2.2.5.3 Ion exchange columns

Proteins exposed to a pH lower than its pI will carry a positive charge and therefore will bind to a cation exchanger column HiTrap SP HP. For protein purification the column must first be equilibrated with buffer A (see *Table* 2.4 for list of buffers). Lysate was then loaded onto the column and gradient eluted using buffer B.

The DEAE column was a batch method purification technique requiring 4 ml of DEAE beads loaded into a clean plastic column. The beads were equilibrated with five column volumes of wash buffer I (see *Table* 2.5 for list of buffers). Lysate was loaded onto the column carefully without disturbing the resin. Washes followed using buffers of increasing amounts of NaCl and finally eluted in 400mM NaCl.

2.2.5.4 pd-10 column

The pd-10 is a pre-packed column with Sephadex G-25 Medium used for de-salting, buffer exchanging and removal of low molecular weight compounds. The column was

equilibrated with 50 ml of either high salt or low salt buffer (see *Table* 2.4) depending on the sample to be loaded onto the column. After equilibration 2.5 ml of sample was loaded onto the column followed by 3.5 ml of high or low salt buffer to elute depending on the desired buffer to be transferred into, the initial ten drops were allowed to flow through, the rest was collected.

2.2.5.5 Size exclusion column

The size exclusion column Superdex 200 10/300 was used as a final purification step serving as a buffer exchange system, removal of any aggregated proteins and separation of protein by size between 10 kDa to 600kDa. Protein was concentrated down to 1 ml using concentrators from VivaSpin with various molecular weight cut offs including 5 kDa and 10 kDa. Concentrated samples were loaded onto a 1 ml loop and injected onto a pre-equilibrated size exclusion column (see *Table* 2.4 for buffer). The protein was eluted over two column volumes of buffer collecting 1 ml fractions.

2.2.5.6 CP-18 column

The CP-18 column is a reverse phase column used for de-salting and obtaining samples in an organic solvent for drying. The column was firstly activated using 15 ml of hexane followed by 15 ml ethanol with 1 % (v/v) acetic acid. Acetic acid was added to the sample to a final of 1 % this was loaded onto the column and washed with 20 ml 1 % acetic acid. To elute 90 % ethanol was used.

2.2.6 SDS-PAGE gel

In order to keep track of the expression and purification of a protein it was essential that SDS-PAGE gels were used after each purification step and for visualising the initial pre- and post-induction of a protein.

Throughout the project 12 and 15 % polyacrylamide gels were used, the specific volumes of the various chemicals can be found in *Table* 2.12. The resolving gel solution was made adding TEMED (tetramethylethylenediamine) last, using a BioRad SDS-PAGE gel casting system the resolving gel solution was poured between two glass plates. A layer of ethanol was added to the top of the resolving gel to provide a level surface for the stacking gel to set. Once the resolving gel had set the ethanol was poured off and washed with dH_2O and blotted dry. Stacking gel was added and a comb with 10 to 15 lanes was inserted and left to set.

Solution components	12 % resolving gel component volumes (ml)	15 % resolving gel component volumes (ml)	5 % stacking gel component volumes (ml)
H_2O	1.6	1.1	2.7
30 % acrylamide	2.0	2.5	0.67
1.5 M Tris (pH 8.8)	1.3	1.3	_
1.0 M Tris (pH 6.8)	_	_	0.5
10 % SDS	0.05	0.05	0.04
10 % ammonium persulfate	0.05	0.05	0.04
TEMED	0.002	0.002	0.004

Table 2.12 Volumes for one SDS-PAGE gel (5 ml total) are given for 12 % and 15 % resolving gels and 5 % stacking gel (total volume 4 ml).

To run the gel 1× SDS-running buffer was used, see below:

Per litre
30.2 g Tris pH 8.3
188 g Glycine
10 g SDS

Protein samples were prepared by adding 1× SDS-loading buffer in a 1:1 ratio, the components of 1× SDS-loading buffer is listed below:

Per 10 ml
100 mM Tris
200 mM DTT
4 % (w/v) SDS
0.2 % (w/v) bromophenol blue
20 % (v/v) glycerol

The samples were then heated on a heat block at 100 °C for five minutes; the samples were then loaded onto the gels and ran at 100 V until the dye front ran out of the gel. For pre- and post-induction samples a pellet from 1 ml sample was resuspended in 400 µl of SDS loading buffer denatured at 100 °C for five minutes and then 3 µl samples were loaded. SeeBlue plus (Invitrogen) protein ladder was used. The gels were then stained in Comassie blue dye for 10 minutes, the stain solution was made up as follows:

Per litre
1 g Comassie blue
30 % (v/v) Methanol
10 % (v/v) acetic acid

After staining, the gels were destained in a small amount of destain so that protein bands can be observed, which consisted of:

Per litre (v/v)
Acetic acid 30 %
Methanol 40 %

For both stain and destain the remaining volume was made up to one litre using dH_2O.

2.2.7 Western blot

Western blotting is a technique used to identify a target protein, in this thesis the His-tag in the construct was exploited. Primary antibodies specific to His-tagged proteins were used, once the antibody was bound specific secondary conjugate antibodies were then applied to allow for colour detection and amplification of the chemiluminescense signal (see section 2.1.8).

Proteins to be probed were run on a polyacrylamide gel. The gel, together with two pads and a piece of nitrocellulose membrane (Invitrogen) were soaked in running buffer (see below) for ten minutes before blotting.

Per 500 ml
15.1g Tris pH 8.3
94g Glycine
20 % (v/v) Methanol

Next the gel was placed in a Trans Blot SD semi-dry transfer cell from BioRad in a stack consisting of a pad, nitrocellulose membrane, polyacrylamide gel and a pad (see *Figure* 2.2.1).

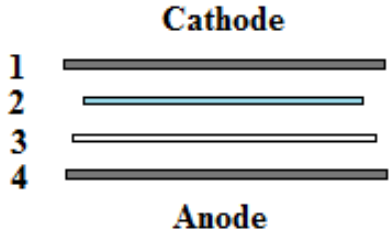


Figure 2.2.1 The set-up for a western blot using a semi-dry blotter consists of 1) pad 2) polyacrylamide gel 3) nitrocellulose membrane 4) pad.

The gel was left to blot for 30 minutes at 200 mA. After 30 minutes the membrane was blocked with 5 % (w/v) skimmed milk powder (oxoid, Thermo Fisher) and 0.1 % (v/v) tween (Fisher Scientific) made up in PBS for an hour to block off non-specific proteins. The membrane was then incubated in 15 µl anti-His mouse antibody in 10 ml 5 % skimmed milk powder and 0.1 % tween for one hour at 4 °C with rocking, after one hour the membrane was washed three times each time for ten minutes with PBS. Next 3 µl of goat anti-mouse alkaline phosphatase conjugate secondary antibody was added to 10 ml of TBS buffer (tris buffered saline) consisting of 50 mM Tris pH 7.5, 150 mM NaCl and 0.1 % tween and used to incubate the membrane for one hour at 4 °C with rocking. Again the membrane was washed three times successively for ten minutes each with TBS buffer. For the detection the membrane was incubated with one SigmaFast BCIP/ NBT tablet dissolved in 20 ml ddH_2O. Once bands start to appear on the membrane care was taken to prevent over detection as non-specific bands may start to appear.

2.2.8 Substrate production

2.2.8.1 Precorrin-3a

Precorrin-3a (referred to as PC-3a hereafter) is a tetrapyrrole formed by the methylation of carbon 20 of precorrin-2 via a CobI methyltransferase. PC-3a is the substrate synthesised for use in soaking experiments into CobJ since the substrate is readily made under anaerobic conditions although the immediate tetrapyrrole utilised by CobJ in the vitamin B_{12} pathway is precorrin-3b (PC-3b). The difference between PC-3a and PC-3b is the addition of a lactone ring.

To synthesise PC-3a a multi-enzyme assay using enzymes on the vitamin B_{12} pathway including Hem B, C, D and CobA with the addition of CobI are required. Details of the process can be found in section 2.2.8.5.

2.2.8.2 Plasmid preparation of pET14b-BmCobI (*Brucella melitensis*)

pET14b-BmCobI was transformed into DH5α competent cells (see section 2.2.4 and Table 2.15) and a single colony was picked and grown in 10 ml LB media consisting of ampicillin and chloramphenicol and grown at 37 ºC overnight with 250 rpm shaking. The cells were harvested for plasmid preparation (see section 2.2.4 for plasmid preparation protocol).

2.2.8.3 Protein production of pET14b-BmCobI

pET14b-BmCobI was transformed into BL21-Star (DE3) pLysS competent cells (see section 2.2.3 and Table 2.11) a single colony was picked and grown in a 10 ml LB

media starter culture supplemented with ampicillin and chloramphenicol at 37 ºC overnight with shaking at 250 rpm. The starter culture was used to grow a one litre baffled flask of LB media with ampicillin and chloramphenicol at 37 ºC until an $O.D_{600}$ of 0.6 was reached. Induction was carried out at 19 ºC overnight with 0.4 mM IPTG final concentration. The cells were harvested and resuspended in 15 ml binding buffer (Table 2.3) and stored at -20 ºC ready for cell lysis and purification.

2.2.8.4 Plasmid preparation of pETcoco2-Hem B, C, D and CobA

pETcoco2 plasmid with Hem B, C, D and CobA (pETcoco2-abcd) was transformed into Nova F- competent cells (see section 2.2.3 and Table 2.11), colonies were picked and grown in 10 ml LB cultures with 0.2 % glucose supplemented with ampicillin. These were grown at 37 ºC overnight. An aliquot was removed and diluted in a 1:50 dilution into pre-warmed fresh LB media with ampicillin and no glucose. This was grown at 37 ºC with shaking at 250 rpm until an $O.D_{600}$ of 0.2-0.4 was reached, to induce plasmid replication L-arabinose was added to a final concentration of 0.01 %. This was grown for 4-5 hours and then harvested for plasmid purification.

2.2.8.5 Protein production of pETcoco2-abcd

pETcoco2-abcd plasmid was transformed into BL21-Star (DE3) pLysS competent cells (see section 2.2.3 and *Table* 2.11). A single colony was transferred into 10 ml LB containing ampicillin, chloramphenicol and 0.01 % L-arabinose and grown at 37 ºC overnight with 250 rpm shaking. This was used to inoculate one litre of LB medium with ampicillin, chloramphenicol and 0.01 % L-arabinose and grown to an $O.D_{600}$ of 0.6 and then induced with 0.4 mM IPTG overnight at 19 ºC. Cells were harvested by

spinning down at 5,000 rpm for 20 minutes and the pellet was resuspended in 15 ml binding buffer (see *Table* 2.3) and used directly in purification or kept frozen at -20 ºC.

2.2.8.6 Multi-enzyme assay

A multi-enzyme assay was carried out in an anaerobic glove box (Belle Technology). All buffers brought into the glove box were degassed for 30 minutes to one hour using argon prior to use. DEAE and pd-10 columns were purged with nitrogen for 30 minutes in the purge box and placed into the chamber a day in advance to allow for removal of oxygen and equilibration of column.

Over expressed pETcoco2-abcd and pET14b CobI pellets were resuspended in binding buffer, sonicated, centrifuged and the lysate was loaded onto a batch method nickel immobilised column individually for purification. CobI was eluted with wash buffer II abcd was eluted with elution buffer (see *Table* 2.3 and section 2.2.5.1). This step was carried out aerobically. The samples were then transferred into the anaerobic chamber along with 500 µl of 18 mg/ml SAM (S-adenosyl-L-methionine) and 500 µl of 6 mg/ml ALA (aminolevulinic acid) both dissolved in 20 mM Tris pH 8.0 and 100 mM NaCl. SAM required pH adjusting to pH 8.0 using 2.0 M NaOH.

Two separate pd-10 columns were pre-equilibrated with 50 ml of either high salt buffer (for CobI) or low salt buffer (for abcd) for buffer exchange and purification (see *Table* 2.4 for buffers and section 2.2.5.4 for the use of pd-10 column). Once fractions were collected the concentration of CobI was determined by measuring the absorbance at 280 nm via a UV-spectrophotometer (U-3010 by Jencons PLS).

The multi-enzyme assay was composed of abcd, CobI, SAM, ALA and low salt buffer, three reactions were set up; A, B and C in glass tubes. Reaction A was a negative control where no CobI was added, reaction B and C contained 40 µM and 80 µM CobI

respectively, see *Table* 2.13 for the final volumes of the different components of the reaction.

Vial	**A**	**B**	**C**
ABCD (µl)	1000	1000	1000
CobI (µl)	-	311	622
SAM (µl)	50	50	100
ALA (µl)	50	50	100
Buffer (µl)	3900	3589	2178

Table 2.13 A summary of the final volumes of each component for the three reactions set up for PC-3a production in the glove box. Three vials of the multi-enzyme assay were set up as listed above. Vial A was the negative control with no CobI present, 1.8 mg/ml SAM and 0.6 mg/ml ALA, vial C contained 40 µM CobI, 1.8 mg/ml SAM and 0.6 mg/ml ALA, vial B contained 80 µM CobI, 3.6 mg/ml SAM and 1.2 mg/ml ALA. The buffer was a low salt buffer.

The tubes were left overnight in the glove box, since PC-3a is light sensitive the tubes were wrapped in aluminium foil. Once the reaction was complete each solution was scanned on the UV-spectrophotometer from 300 to 700 nm for characteristic trace of PC-3a (see section 4.3.1.1). To minimise oxygenation and unreliable scans each reaction was measured one at a time in cuvettes (UV-Cuvette micro from Brand) with plastic stoppers parafilmed at the edges. Once PC-3a had been identified the rest of the sample was purified on a DEAE column (see section 2.2.5.3). Fractions were collected and buffer exchanged into a low salt buffer using a pd-10 column to make the substrate suitable for co-crystallisation and crystal soaking experiments.

For concentrated PC-3a it was necessary to use a CP-18 column to remove the high salt concentrations and allow for evaporation of ethanol until a more concentrated sample could be obtained in the glove box (see section 2.2.5.6 and *Table* 2.4).

2.2.8.7 Precorrin-3b

Precorrin-3b (referred to as PC-3b hereafter) is a tetrapyrrole formed by the reaction of PC-3a with monooxygenase CobG forming PC-3b. The synthesis of CobG follows the same protocol for the production of PC-3a (see section 2.2.8.6) with the addition of purified and concentrated CobG to PC-3a in the glove box. To allow for oxygenation the reaction was taken out of the anaerobic chamber.

2.2.9 BsPel spectroscopy activity assay

To check the activity of BsPel a spectroscopic enzyme assay was used to monitor the cleavage of the α-1,4 glycosidic bond of polygalacturonic acid. The absorbance at 235 nm was measured which is characteristic of the formation of a carbon-carbon double bond.

Table 2.14 lists the assay solution used for the assay which was pH adjusted to 8.5 of which 1 ml of the assay solution was added to a cuvette followed by 20 µl of purified BsPel at 0.2 mg/ml. The mixture was then measured at 235 nm for duration of three minutes; the gradient tells us the activity of the enzyme.

Chemical	Stock (M)	Volume (ml)	Amount (mg)
Tris pH8.5	1	1.0	_
$CaCl_2$	1	0.02	_
Polygalacturonic acid	_	_	100

Table 2.14 Summarises the assay solution used for BsPel enzyme assay. A final volume of 20 ml was achieved by adjusting with dH_2O. Polygalacturonic acid was obtained from Sigma-Aldrich.

2.2.10 Crystallography

X-ray crystallography is a powerful technique used to solve the three-dimensional structures of proteins at atomic resolution. In order to produce images of individual atoms of bond distances of approximately 1.5 Å, X-rays are used since the wavelength typically falls between 0.5 and 1.6 Å.

Crystallographers are unable to focus X-rays using lenses like a microscope can focus light, therefore diffraction spot position and intensity must be recorded in order to obtain an image of a protein structure. The relative phase of each diffracted reflection must also be measured or calculated before an electron density map of the structure can be synthesised. The first step for a crystallographer is to form protein crystals and crystallisation is often the major bottleneck in crystal structure solution as without a suitably diffracting crystal there is no prospect of a crystal structure. Crystals consist of regular arrangement of molecules in space. The asymmetric unit when operated on by the crystal symmetry generates the unit cell which when translated by the crystallographic vectors *a, b*, and *c* generates the crystal.

2.2.10.1 Crystallisation

Protein used for crystallisation trials should be pure and homogeneous, both chemically and conformationally. All buffers and precipitants used were made using deionised water of 18.2 MΩ.cm purity (see *Table* 2.9). Different reservoir conditions were made up in each well of a pre-greased 24-well plate (plates from Hampton Research and high vacuum grease from Dow Corning). A hanging drop vapour diffusion technique was used throughout; typically 2 µl mother liquor was mixed with 2 µl of protein on a clean siliconated glass slide (Hampton Research) and placed over the well to form an air-tight seal. The trays were left in a temperature controlled room at 19 ºC.

During this time wells were left to equilibrate with the protein drops, over time gradual diffusion of water occurs from the drops and causes a slow increase in the precipitant and protein concentration encouraging crystallisation see *Figure* 2.2.2.

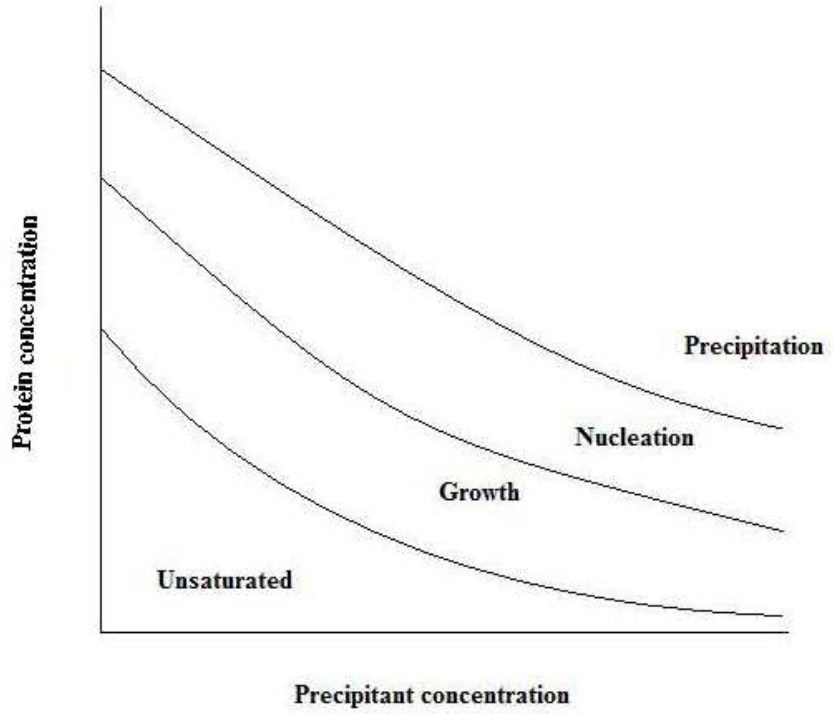


Figure 2.2.2 Phases of crystallisation where unsaturated regions indicate concentrations of protein and precipitant yielding no crystals. Increase in concentrations in protein and precipitant results in growth and nucleation however too much precipitant and protein results in precipitation. Ideally one would want growth, nucleation and back to growth phase until the largest possible crystals are obtained (Rhodes 2006).

2.2.10.2 Crystal soaking

For the purpose of a soaking experiment a suitable cryoprotectant was made and supplemented with the desired molecule or substrate to be soaked into the crystal. It is also important to note that as a rule of thumb when soaking old crystals where

precipitant concentration may have increased a fraction, it is important to make adjustments to the percentage of precipitant by 0.5 % of the precipitant concentration i.e 40 % PEG 8000 would be increased to 40.2 % PEG 8000. Failure to match the conditions correctly may result in disordering and cracking of the crystal. This rule should also be followed in the case when preparing cryoprotectants for older crystals when freezing.

To soak a crystal a small volume (5 µl) of the supplemented cryoprotectant was transferred onto a clean glass slide, the crystal was looped out of its drop using a wand and litho loop (both Hampton Research) and deposited into the supplemented cryoprotectant for some time (see section 4.4.2 and 5.5.2 for detailed cryoprotectant and soak conditions). Once the crystal had been soaked for the desired amount of time the crystal was looped out of the drop using a litho loop of the correct size and placed into liquid nitrogen to freeze.

2.2.10.3 Crystal freezing

Commonly used cryoprotectants include glycerol, ethyleneglycol, PEG 400 and propylene glycol. A good cryoprotectant will provide protection and prevent damage caused by freezing and at the same time will not hinder the diffraction potential of the crystal (McFerrin 2002). Typically the composition of a cryoprotectant is the reservoir condition which the crystal was grown in with 20 to 25 % (v/v) of the desired cryoprotectant. Crystals were frozen for ease of transportation and time saving purposes at synchrotrons. Crystals were looped out and briefly pulled through a drop of suitable cryoprotectant before dipping into liquid nitrogen vitrifying the crystal. Once the crystals were frozen they were transferred onto canes and into a pre-cooled storage dewar.

2.2.10.4 Data collection and reduction

A diffraction spot or Bragg reflection is observed when the angle of incidence of the incoming X-ray beam is such that the path difference between rays reflected from successive layers in the crystal, separated by spacing d, equals an integer number of multiples of the X-ray wavelength (λ) so the reflected waves interfere constructively (see *Figure* 2.2.3). This condition for diffraction is Braggs Law *(a)*:

(a) $$2d \sin \theta = n\lambda$$

d is the interplanar distance, θ is the angle of incidence, n is an integer and λ is the wavelength of the X-rays

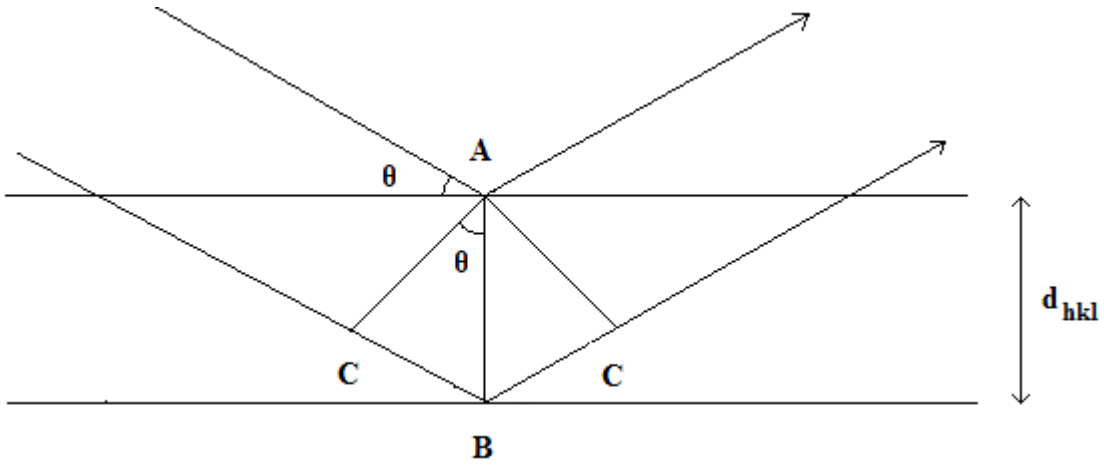


Figure 2.2.3 Illustration of Bragg's law and the conditions needed in order to produce diffracted rays. d_{hkl} is the interplanar distance.

$$\sin \theta = BC/AB$$

$$BC = AB \sin \theta = d_{hkl} \sin \theta$$

Experiments requiring data collection took place at three different synchrotrons: SRS Daresbury, ESRF Grenoble and Diamond Light Source. Crystals were mounted onto a goniometer where a constant cold stream of nitrogen gas at 100 K keeps the crystal vitrified. Data was processed using MOSFLM (Leslie 1992) and reduced using SCALA (Evans 2005). Processing refers to the indexing of the diffraction pattern and the integration of the intensities image by image and reduction is the process whereby equivalent reflections, the same reflection measured twice of crystallographically equivalent reflections, are merged together to form the unique data set. At this stage the intensities are converted to structure factor amplitudes.

2.2.10.5 Fourier syntheses

Processing the diffraction data produces a list of reflections, or more precisely structure factor amplitudes |F(hkl)|, which are identified by their Miller indices h, k, and l. A Fourier synthesis can be used to convert the structure factor amplitudes |F(hkl)| and phase angles α(hkl) into electron density once the phase angles are known, *(b)*.

(b) $$\rho\,(x,y,z) = 1/V\ \Sigma_{hkl}\ |F(hkl)|\ \exp\,[-2\pi i(hx+ky+lz) + i\alpha\,(hkl)]$$

In the absence of phase information a Fourier summation with the intensities as coefficients and without phase angles, or rather with phase angles all set to zero, can be calculated. This is called the Patterson function, after its originator, *(c)*.

(c) $$P(u,v,w) = 1/V\ \Sigma_{hkl}\ |F(hkl)|^2\ \cos\,[2\pi(hu+kv+lw)]$$

To avoid confusion with real space (x,y,z), u,v,w are used in the Patterson cell. The Patterson function can be calculated without any previous knowledge of the structure or the phases.

The phase angles may be obtained experimentally by isomorphous replacement, by exploiting anomalous scattering of an intrinsic or introduced metal, or can be calculated from an existing structure by a process called molecular replacement.

Molecular replacement was the method used in this work and involves relating a model structure, derived from an existing crystal structure, to the molecule (or molecules) in the unknown crystal. The calculation can be separated into two problems, the rotation problem and the translation problem. The Patterson function contains all the information from the amplitudes and none from the phase and represents the collection of vectors between the elements of scattering density. The basic idea of the rotation function is that the Patterson function of a molecule will have a characteristic distribution of densities so that the Patterson function of the known molecule can be rotated over that of the unknown molecule and the greatest overlap will correspond to the orientation of the unknown molecule in the unknown crystal. Once the rotational relationship between the two molecules has been established, the translation of the molecule can be established. One method is to simply compute the crystallographic Rfactor as a function of the position of the correctly rotated molecule in the unit cell. Molecular replacement calculations were done using MOLREP (Vagin 1997).

2.2.10.6 Refinement and model building

Successive cycles of refinement and model building were carried out using REFMAC (Vagin 2004) and COOT (Emsley 2004) to improve the agreement of the model and experimentally derived structure factor amplitudes. Agreement was measured as the Rfactor and along with this value Rfree was also calculated as a control, using the 5 % of reflections consistently left out of the refinement. After refinement the resultant electron density map was examined to see if there are any parts of the structure that could be improved manually using COOT (Emsley 2004). The Rfree and Rfactor indicate the quality of the refined data. Pymol was used to draw structure figures throughout (The Pymol Molecular Graphics System, Version 0.99, Schrödinger, LLC).

2.2.10.7 Validation

For validation, ADIT Validation Server was used (RCSB Protein Data Bank) this provided a validation report listing close contacts, chirality errors, bond angle and deviations, missing and extra atoms or residues and distant waters. The stereochemistry of the model was checked using a Ramachandran plot highlighting the φ and ψ angles of residues that lie outside of the allowed regions.

Chapter Three:

Cloning and expression of *Caenorhabtidis elegans* telomerase reverse transcripase

3.1 *Caenorhabtidis elegans* reverse transcriptase subunit

Telomerase is a ribonucleoprotein found in all eukaryotes, it plays a vital role in maintaining telomere length, preventing end-to-end chromosome fusion and circumventing the end-replication problem. This dilemma is found in all eukaryotes due to the linear nature of eukaryotic chromosomal DNA and the 5'-to-3' uni-directional action of DNA polymerase. During replication the linear chromosomes in eukaryotes separate into replication forks a leading strand and a lagging strand. The lagging strand runs in a 5'-to-3' direction causing the formation of small Okazaki fragments that require ligation after the removal of RNA primer. At its most 3' end of the newly formed strand, DNA is lost or not copied due to the inability of RNA priming (Watson 1972; Lingner 1995). However with telomerase action, sequences are added to the ends to allow for appropriate RNA priming and complete DNA replication, this was first identified in ciliate *Tetrahymena thermophila* (Greider 1985). Telomerase has the capability of synthesising a repetitive sequence that is used to cap and maintain the ends of telomeres, without such an ability, telomeres would become successively shorter with each cell division and lead to premature cell senescence (Yu 2001). Typically the enzyme is found in highly proliferative cells and interestingly in most cancerous cells. Telomerase appears to provide cancerous cells immortality allowing continual replication (Shay 2002).

Telomerase minimally requires two subunits to function in *vitro*, the catalytic component telomerase reverse transcriptase (TERT) and an intrinsic RNA template (TR). The possession of this intrinsic RNA template makes telomerase a unique reverse transcriptase enabling the synthesis of iterative telomeric sequences. It is generally accepted that most TERT proteins identified are made up of seven conserved reverse transcriptase (RT) motifs which are found universally across other RTs. These seven motifs are located in the central region of TERT, commonly the other RTs are divided into fingers and palm and a thumb region (see *Figure* 3.1.1a). The difference between

telomerase and other RTs, apart from the intrinsic TR template, is an insertion between motifs A and B' called IFD (insertion in fingers domain) which is thought to be involved in RNA recognition and telomere lengthening (Lue 2003). The TERT subunit also consists of a large N-terminal extension which is approximately 400 amino acids long and can be separated into two regions, distal and proximal and a shorter 150 amino acid C-terminal extension, see *Figure* 3.1.1b (Lingner 1997). The function of N- and C-terminal domains in humans and *T. thermophila* were demonstrated to be critical for activity, but in yeast TERT the C-terminal domain appeared to be dispensable. Within the N- and C-terminal regions are motifs which are essential for TR recognition; binding and assembling of template region and repeat addition processivity (see *Table* 1.1.1 in section 1.1.3).

The C-terminal of TERT was suggested to be involved in promoting telomerase processivity and regulating telomerase localisation. In humans this domain has been shown to maintain telomere length by repeat addition processivity, a technique used to elongate telomere ends which is a unique biochemical characteristic of telomerase (Huard 2003).

Little is known about the TR template that is used intrinsically to reverse transcribe short tandem G-rich repeats used as a protective cap on telomeres. TRs vary widely in length and in primary sequence homology making it difficult to identify potential TRs. However studies have shown TRs share common secondary structures such as a putative pseudoknot, a loop-closing helix, a 5' template boundary region and a template region. Since there are conserved core structural elements it has been suggested that telomerase also shares a common mechanism for telomere lengthening among organisms (Chen 2004). TR structures have been identified in a range of organisms including ciliate, vertebrate and yeast (Romero 1991; Chen 2000; Dandjinou 2004).

During this work the crystal structure of an N-terminal domain of TERT termed TEN in *T. thermophila* was solved with a groove on its surface harbouring conserved amino acid residues. Site-directed mutagenesis studies have confirmed a number of crucial residues required for catalytic activity. Mutations of residues W187A, Q168A and F178A which all lie in the groove structure showed highly decreased levels or no measurable catalytic activity at all. It was proposed that the TEN domain is necessary for proper assembly and recognition of telomeric DNA (Jacobs 2006). The crystal structure of the RNA binding domain (TRBD) has also been solved, the domain is made up of conserved motifs CP and T. These motifs are exclusive to telomerase only and no other reverse transcriptases. CP and T motif are found in the proximal N-terminal domain (Rouda 2007).

The first breakthrough on the structure of TERT was achieved by the over expression and crystallisation of *Tribolium castaneum* TERT. The structure of full length TERT was solved to 2.71 Å and consisted of three conserved domains; an RNA binding domain, the reverse transcriptase domain and the C-terminal domain which are organised in a ring-like structure. The N-terminal domain does not exist in this organism. The structure revealed motifs involved in RNA binding that were located on

the interior of the ring that is able to accommodate double stranded nucleic acid of up to eight bases (Gillis 2008).

The main focus of this chapter is *Caenorhabditis elegans* and its TERT domain, the gene encoding for this protein is termed DY3.4. *C.elegans* possesses a very small TERT component at 66 kDa compared to other TERTs (see *Table* 3.1.1) and this provided the motivation for cloning, expressing and ultimately crystallisation of the protein since it may be easier to overproduce and crystallise. DY3.4 lacks the N- and C-terminal domains (Autexier 2006). Evidence suggests the C-terminal is responsible for promoting telomerase processivity and in regulating telomerase localisation and the N-terminal domain mediates recognition of TR structures in different organisms (Bosoy 2003). To compensate for the loss of these domains, auxillary proteins would need to be utilised to replace the functions that otherwise would be missing, however little is known about *C. elegans* TERT.

Organism	TERT Size (kDa)
Tetrahymena thermophila	133
Saccharomyces cerevisiae	103
Plasmodium falciparum	280
Homo sapiens	127
Tribolium castaneum	70
Caenorrhabitidis elegans	66

Table 3.1.1 A summary of a few organisms and their corresponding TERT component sizes, *C. elegans* TERT is the smallest TERT in comparison to other named TERTs.

Past studies have predicted a transcribed telomerase-like sequence, tts-1 (see Appendix 6) to be the TR component of *C. elegans* telomerase (Jones 2001). The sequence has common characteristics which would indicate the presence of a possible TR such as a one-and-a half repeat telomeric template reading TAAGCCTAA allowing the synthesis

of telomeres found in *C. elegans* and also a predicted pseudoknot. It is thought that a one-and-a-half telomeric template would allow easier translocation of the synthesised DNA by repositioning the most 3'-end to the start of the telomeric template.

Since little is known about the structure and function of telomerase from this frequently used model organism *C. elegans* and because of the advantages offered by its small size there was motivation to work on this small TERT component, DY3.4.

3.2 Cloning and site-directed mutagenesis

cDNA for the gene DY3.4 encoding for *C. elegans* telomerase was purchased from WormBase. The aim was to clone the gene into a vector for protein production, purification and ultimately crystallisation of the protein.

3.2.1 Primer design

A number of primers were designed for cloning and site-directed mutagenesis experiments, the method used for the design of primers can be found in section 2.2.1.1. TERT was successfully cloned into pET14b using direct cloning (see section 2.2.1.3) and pET41 vector which was obtained via Ek/LIC cloning system which was digestion and ligation independent (see section 2.2.1.4). Site-directed mutagenesis was carried out to correct point mutations after PCR cloning, for a list of all the primers used for this results chapter see *Table* 3.2.1.

Primer pair	Usage	Restriction site
3′ CGAGCAGTTAAAAACTATTTCAACGTTATTCGACTCGTGGC 5′ 5′ GCCGCTCTCGAGATGGCACCAACGATTAAGAGTTCACTTAC 3′	Direct cloning	BlpI XhoI
5′ GACGACGACAAGATggcaccaacgattaag 3′ 3′ tatttcaacgttATTTGGCCCGAAGAGGAG 5′	Ek-LIC cloning	—
5′ GTGGCACCAACGATTAAGAGTTCAC 3′ 3′ CACCGTGGTTGCTAATTCTCAAGTG 5′	site-directed mutagenesis	—
5′ GGCTCTGATGGCTTTACGACAAG 3′ 3′ CCGAGACTACCGAAATGCTGTTC 5′	site-directed mutagenesis	—
5′ GGCTAATCCGAAAAAGTTCAAG 3′ 3′ CCGATTAGGCTTTTTCAAGTTC 5′	site-directed mutagenesis	—

Table 3.2.1 The primers designed and used throughout this chapter for cloning and site-directed mutagenesis are listed above detailing the restriction sites introduced if any. The nucleotides highlighted in grey are the restriction sites introduced or a single base mutation for site-directed mutagenesis. In the case of the primers used for Ek-LIC cloning strategy the highlighted ATT is the stop codon.

3.2.2 Digestion and ligation independent cloning

To achieve fast effective cloning an Ek/LIC system from Novagen was used. The kit allows digestion and ligation independent cloning (see section 2.2.1.4). The PCR reaction was successful, see *Figure* 3.2.1, however confirmation of the gene sequence was needed via sequencing analysis (see section 3.2.4). The cycling parameters used are listed in *Table* 3.2.2. The PCR reaction mixture set-up was 1 µl insert template, 5 µl 10× reaction buffer, 1.25 µl forward Ek/LIC primer, 1.25 µl reverse Ek/LIC primer, 1 µl dNTP, 39.5 µl ddH_2O and 1 µl Pfu Turbo polymerase.

Step	Temperature (°C)	Duration
1. Activation	95	2 min
2. Denaturation	95	20 sec
3. Annealing	60	10 sec
4. Elongation	70	40 sec
5. Final elongation	72	3 min
Steps 2 to 4 cycled 30 times.		

Table 3.2.2 Summarised cycling parameters used for cloning TERT into pET41 vector using the Ek/LIC cloning system.

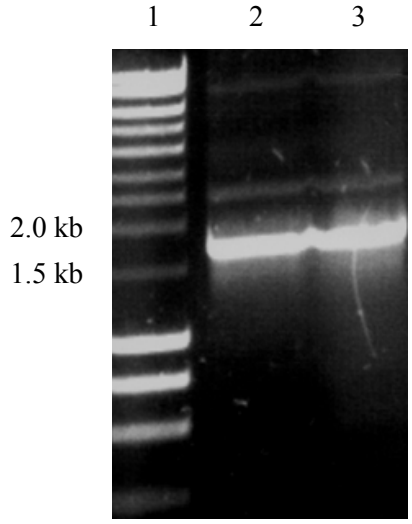


Figure 3.2.1 Lane **(1)** DNA marker hyperladder **(2)** and **(3)** bands showing the successful PCR reaction and presence of DNA. The size of insert was 1.6kb.

The insert was subsequently gel purified and T4 treated followed by annealing to pET41 vector (see section 2.2.1.5). The ligated product was transformed into NovaBlue competent cells and single colonies were grown and isolated in a plasmid preparation before sending off for sequencing.

3.2.3 Direct cloning

TERT was cloned into pET14b using primers designed to insert specific restriction sites, XhoI and BlpI on either end of the gene. The cycling parameters used for the PCR reaction are listed in *Table* 3.2.3. The PCR reaction set-up was 1 µl insert template, 5 µl 10× reaction buffer, 1 µl Q-solution, 2 µl $MgCl_2$, 1 µl forward primer (XhoI restriction site), 1 µl reverse primer (BlpI restriction site), 2 µl dNTP, 36.5 µl ddH_2O and 0.5 µl Hotstar Taq polymerase. The resultant insert from the PCR reaction was then ligated to pET14b which had been double digested at the specific restriction sites. A diagnostic test was carried out using XhoI and BlpI restriction enzymes to confirm the presence of the insert and vector (see *Figure* 3.2.2). Sequencing analysis was carried out to confirm the successful cloning, see later section for sequencing results.

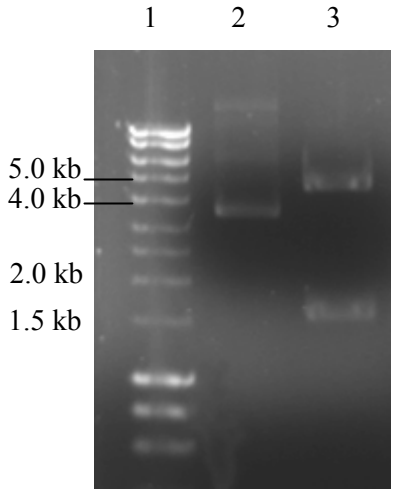


Figure 3.2.2 Lane **(1)** Hyperladder DNA ladder **(2)** band showing insert ligated to vector **(3)** Double digested pET14b-TERT after ligation showing two bands one at around 1.5 kb indicating the presence of TERT (1.6 kb) and a band at around 4.0 kb showing pET14b (4.6 kb).

Step	Temperature (°C)	Duration
1. Activation	95	15 min
2. Denaturation	94	30 sec
3. Annealing	60	30 sec
4. Elongation	72	3 min
5. Final elongation	72	10 min
Steps 2 to 4 cycled 30 times.		

Table 3.2.3 Lists the temperature and duration for each step used in the PCR reaction for the cloning of DY3.4 into pET14b.

3.2.4 Sequencing results

Purified plasmids and pre-designed primers were sent off for sequencing analysis at MWG Biotech. The primers were used to ensure the entire gene sequence was covered in the analysis (see *Table* 3.2.4). The results from sequencing showed TERT had been successfully cloned into pET41 vector with no mutations present in the gene (see Appendix 2 for alignment).

Primer	Primer name	Direction of primer	Base pair covered
5′ CGTTATATAGTTCAAAAAGCTCCG 3′	Int2	Forward	780-1454
5′ CCGAGAGATGAGTTCCTCTACAACA 3′	Int3	Forward	1261-1686
5′ CGCGATACCGTTGTTTTACTGG 3′	Int4	Reverse	1-184
5′ GGTTGAATTCGATGCTCTC 3′	Int5	Reverse	1-838
5′ CACAGGGACATCCTATATCTTC 3′	Int6	Reverse	430-1407

Table 3.2.4 A summary of the primers used for sequence analysis of pET41-DY3.4 and pET14b-DY3.4 construct. The regions of the gene sequence which provided a good strong read by each primer are also noted.

Sequencing results for pET14b-DY3.4 construct showed the gene had been successfully cloned with no insertions or deletions; however six point mutations were evident upon examination of the sequence (see Appendix 3). This may be due to the polymerase, Hotstar Taq, used in the cloning process. Hotstar Taq does not have 3' to 5' exonuclease activity. Also thirty cycles were used in the PCR cycling parameters which could have increased the rate or likelihood of mutations. The mutations are summarised in *Table* 3.2.5. The rationale for using HotStar Taq was because DY3.4 was a particularly difficult gene to clone and a number of polymerases were used including Pfu Turbo polymerase and KOD Hotstart polymerase and none were successful. The Hotstar Taq from Qiagen had an optional PCR additive, Q solution; which was useful for getting difficult PCRs to work as it changes the melting behaviour of the DNA template.

Point mutation	Residue mutation
T to G	Ile to Ser
C to T	Ala to Val
G to A	Ala to Thr
T to C	Trp to Gly
A to G	Lys to Arg
A to G	Silent mutation (Gly)

Table 3.2.5 Mutations revealed in pET14b-DY3.4 construct after sequence analysis showed six point mutations of which one was a silent mutation (Gly).

In order to correct the point mutations site-directed mutagenesis was carried out for all the point mutations apart from the silent Gly mutation to convert them back to the original base. The cycling parameters used for site-directed mutagenesis are listed in *Table* 3.2.6, the polymerase used was KOD Hotstar polymerase and the reaction set-up was 5 µl 10× reaction buffer, 3 µl 25 mM $MgSO_4$, 5 µl dNTPs, 1.5 µl forward primer, 1.5 µl reverse primer (primers for each mutagenesis can be found in *Table* 3.2.1), 1 µl template DNA, 32 µl ddH_2O and 1 µl KOD Hotstart polymerase. Parental DNA was digested with 1 µl of DpnI at 37 ºC for one hour subsequent newly formed DNA were

transformed into DNA replication competent cells XL-1 Blue (see section 2.2.3 and *Table* 2.11).

Step	Temperature (°C)	Duration
1. Activation	95	2 min
2. Denaturation	95	20 sec
3. Annealing	Lowest primer T_m	10 sec
4. Elongation	70	3 min
Steps 2 to 4 cycled 25 times.		

Table 3.2.6 (T_m is the melting temperature) Cycling parameters used for correction of point mutations in pET14b-DY3.4 using KOD Hotstar polymerase from Novagen.

Plasmids were sent to MWG Biotech for sequencing to check for corrections of mutations and the results showed all mutations which required amendment were successfully mutated back to the original base (see Appendix 4 for sequence alignment). The plasmid was then used for protein expression studies. The primers designed and used for sequencing purposes are listed in *Table* 3.2.7, the primers used enabled the entire DY3.4 gene to be covered in the sequence analysis.

Primer	Primer name	Direction of primer	Regions of DY3.4 confirmed via sequencing (bp)
5′ ATGGCACCAACGATTAAG 3′	Int1	Forward	56-759
5′ CGTTATATAGTTCAAAAAGCTCCG 3′	Int2	Forward	780-1454
5′ CCGAGAGATGAGTTCCTCTACAACA 3′	Int3	Forward	1261-1686
5′ GATACATTTGTTGCATGCTCC 3′	Int10	Reverse	1-307

Table 3.2.7 Primers used for sequencing of DY3.4 are listed above with details of the base pair regions covered in the TERT sequencing results.

3.3 Protein production

3.3.1 Rare codon analysis

Analysis of the amino acids of TERT in *C.elegans* revealed there were a total of 29 rare codons (RaCC: Rare Codon Calculator by NIH MBI Laboratory for Structural Genomics and Proteomics). These rare codons included Arg AGG, AGA, CGA, Leu CTA, Ile ATA and Pro CCC (see Appendix 5 for analysis). To supplement these codons BL21 codonplus (DE3)-ril *E.coli* competent cells from Stratagene were used.

3.3.2 Expression of pET41-DY3.4

Protein expression required transformation of pET41-DY3.4 into BL21 codonplus (DE3)-ril cells. The protocol for general transformations can be found in section 2.2.3. Cells were grown in one litre LB media supplemented with chloramphenicol and kanamycin (see section 2.1.9 for working concentrations) at 37 ºC until an $O.D_{600}$ of 0.6 was reached. Cells were then induced using 0.5 mM IPTG and grown at 37 ºC for three hours. Cells were lysed using a French press (see section 2.2.4.2) and purified on a GST column (see section 2.2.5.2 for protocol). SDS-PAGE gel analysis of fractions collected from the GST column showed that there did not appear to be a large amount of protein expressed which was evident in the pre- and post-induction fractions. However since bands at around 98 kDa were seen on the gel in the eluted fractions and were of approximately the size of DY3.4 (63.8 kDa) and GST fusion tag (26 kDa) these fractions were pooled and purified on an S200 size exclusion column (see *Figure* 3.3.1).

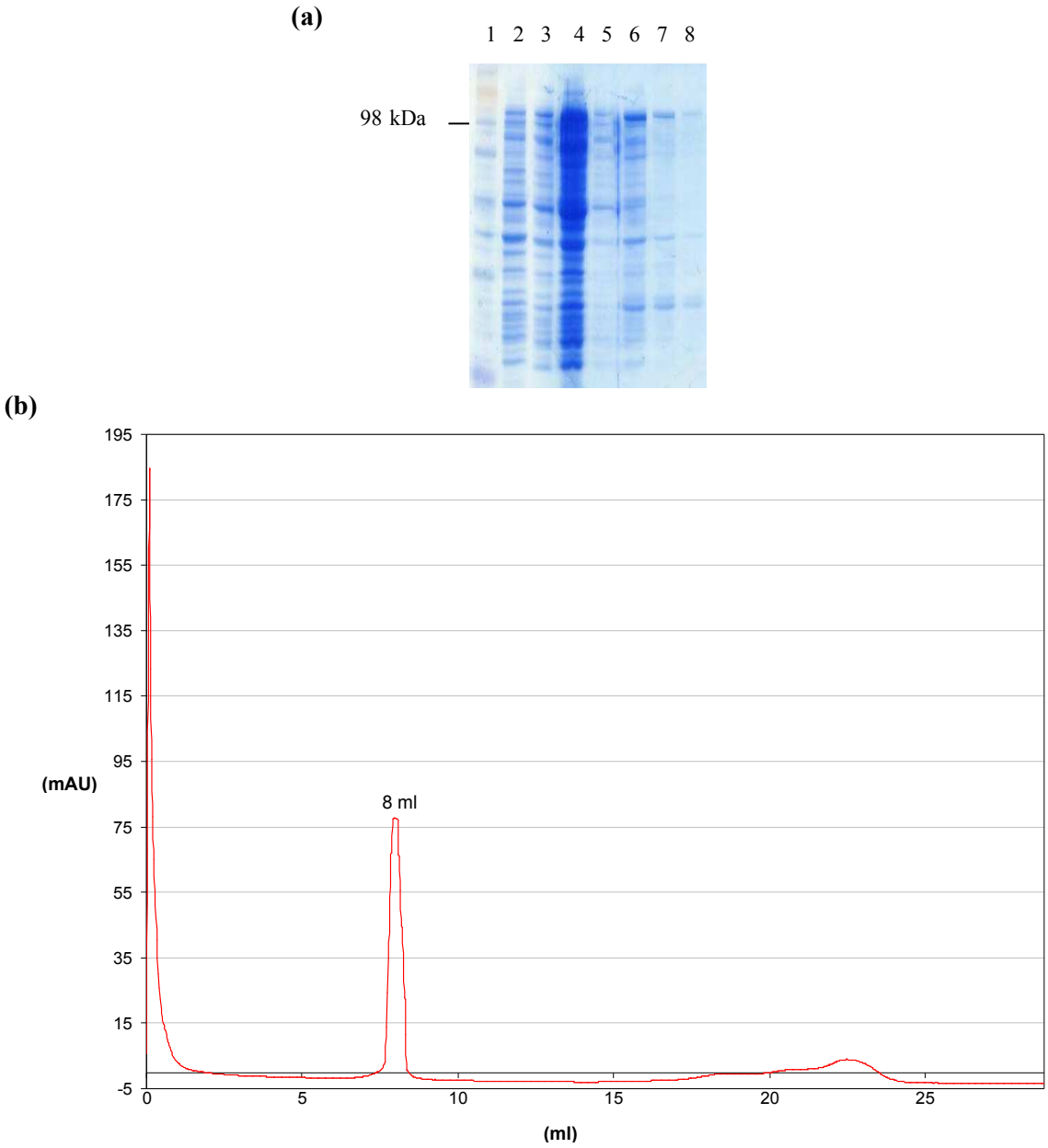


Figure 3.3.1 **(a)** SDS-PAGE gel analysis of pET41-DY3.4 construct. Lanes (1) SeeBlue Plus 2 protein ladder, (2) pre-induction sample, (3) post-induction sample, (4) and (5) PBS wash and (6) to (8) fractions obtained after elution with reduced glutathione sepharose buffer. **(b)** S200 size exclusion UV trace showing a large peak in the void volume at 8 ml identifying the presence of possible pET41-DY3.4 aggregated protein.

3.3.3 Expression and protein identification of pET14b-DY3.4

pET14b-DY3.4 was transformed into BL21 codonplus (DE3)-ril cells following the same procedure used for protein expression of pET41-DY3.4. Cells were lysed and centrifuged (see section 2.2.4.2) separating cell lysate and cell debris. To identify the DY3.4 protein a western blot was carried out (see section 2.2.7) using anti-His antibodies since pET14b is N-terminally His-tagged. The results showed that pET14b-DY3.4 was detected in the insoluble pellet with a band at around 64 kDa (see *Figure* 3.3.2). Given that TERT was insoluble alternative methods were adopted in an attempt to obtain soluble TERT.

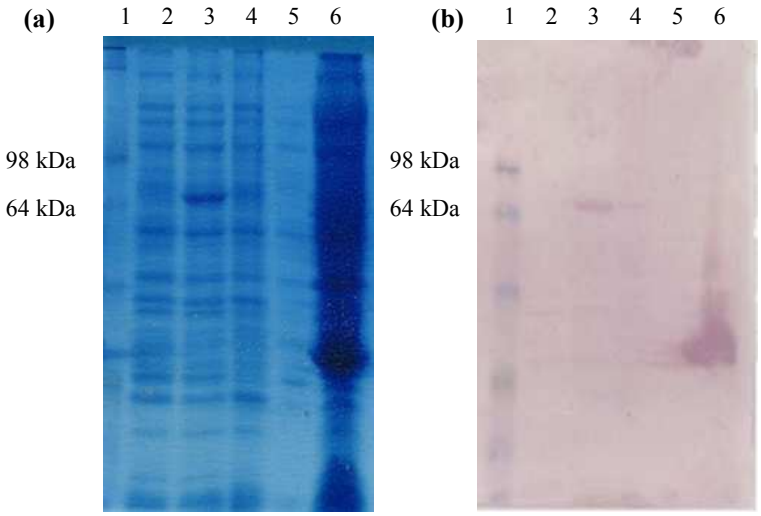


Figure 3.3.2 **(a)** SDS-PAGE gel of pET14b-DY3.4 expressed in BL21 codonplus (DE3)-ril cells Lane **(b)** western blot analysis. (1) SeeBlue Plus 2 protein ladder, (2) pre-induction sample, (3) post-induction sample, (4) pellet after cell lysis and centrifugation and (5) supernatant after cell lysis and centrifugation (6) His-tagged protein positive control.

3.4 Solubilising pET14b-DY3.4

The following sections of this chapter focus on different methods used to try and obtain soluble TERT protein from *C. elegans*.

3.4.1 Expression trials

Expression trials were carried out using various *E. coli* cells and growth and induction conditions. The cells used in these trials were BL21 codonplus (DE3)-ril, Origami (DE3), Rosetta Blue (DE3) pLysS, C43 (DE3) pLysS and BL21 (DE3) pLysS the expression trial conditions are listed in *Table* 3.4.1.

Growth temperature (ºC)	Induction temperature (ºC)	Induction duration	IPTG concentration (mM)
37	37	4 hrs	1
37	37	4 hrs	0.5
37	30	4 hrs	1
37	30	4 hrs	0.5
30	30	4 hrs	1
30	30	4 hrs	0.5
37	18	Over night	1
37	18	Over night	0.5

Table 3.4.1 Summarises the expression trial conditions used for various competent cells as listed previously.

Unfortunately none of the conditions yielded any soluble protein as indicated by western blot analysis.

3.4.2 GroEL and GroES chaperones

GroEL and GroES are chaperones from *E. coli*, the two proteins form a chaperonin complex that aids the folding and or refolding of proteins. GroEL is a 60 kDa protein which forms two stacked rings of seven subunits, together with 10 kDa heptamer cochaperonin GroES which forms the "lid" they work in tandem to mediate correct protein folding (Sigler 1998).

DY3.4 protein contains 14 cysteine residues, improper disulfide bond formation may have accelerated the misfolding of protein and hence forming insoluble aggregates. In an attempt to try and help folding and prevent aggregation a plasmid (pBAD) containing GroEL and GroES chaperones was used in a cotransformation with pET14b-DY3.4. The pBAD plasmid was courtesy of Dr Syeed Hussain. The pBAD system required L-arabinose to induce the growth of chaperones with chloramphenicol resistance. Co-transformation of pBAD and pET14b-DY3.4 was carried out using Origami (DE3) cells. Cells were grown in LB media supplemented with L-arabinose 0.05 %, chloramphenicol and ampicillin at 37 ºC until $O.D_{600}$ of 0.6 was reached followed by induction at 37 ºC for four hours using 0.5 mM IPTG. Cells were lysed using a sonicator and centrifuged to separate insoluble pellet from soluble supernatant. Fractions were analysed using western blot, initial results appeared to show pBAD-GroES/GroEL had been successful in solubilising DY3.4, with bright band in supernatant fraction at approximately 64 kDa indicating the presence of soluble protein. However further analysis showed that pBAD plasmid transformed into Origami (DE3) cells as a control also showed bands at around 64 kDa (see *Figure* 3.4.1).

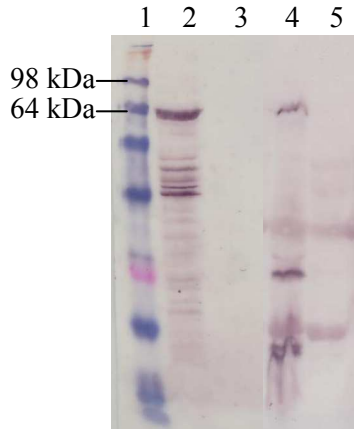


Figure 3.4.1 Lane **(1)** SeeBlue Plus2 protein ladder, **(2)** and **(3)** supernatant and pellet samples of pET14b-DY3.4 co-transformed with pBAD-GroES/GroEL into Origami (DE3) cells respectively. Lane 2 shows a bright band at around 64 kDa indicating possible soluble TERT. **(4)** and **(5)** supernatant and pellet samples of pBAD-GroEL/GroES transformed into Origami (DE3) cells respectively. A band can also be seen at around 64 kDa, confirming the chaperones were unsuccessful in solubilising TERT.

3.4.3 Solubilising TERT using urea

Since soluble TERT could not be obtained by expression trials or co-transforming with chaperones GroES and GroEL an alternative was to solubilise insoluble TERT pellet with high concentrations of urea. Cells were grown and induced as described in section 3.3.3. The cell pellet was harvested and resuspended in binding buffer 5 mM imidazole, 500 mM NaCl and 20 mM Tris pH 8.0. Cells were freeze thawed three times followed by sonication (5 sec pulse with 5 sec intervals for ten minutes). The lysed cells were centrifuged for 30 minutes at 10,000 g. The pellet was resuspended in binding buffer containing successively increased urea concentration ranging from 2 M, 3 M and 4 M plus 1 % Triton X-100. Each time the resuspended pellet was placed under agitation for one hour followed by centrifugation for 30 minutes at 10,000 g. The final buffer used to resuspend the inclusion bodies pellet contained 6 M urea plus 1 % Triton X-100 and left overnight under agitation at room temperature. Pellet and supernatant were separated via

centrifugation for 30 minutes at 10,000 g samples of supernatant and pellet were retained for western blot analysis (Lemercier 2003).

Results from western blot analysis revealed that the solubilising of TERT using 6 M urea only achieved partial solubilisation see *Figure* 3.4.2. Very small amounts of protein was recovered from the supernatant fraction compared to the insoluble pellet due to the small amount of soluble protein that could be easily made and the difficult nature of the protein this method was not pursued.

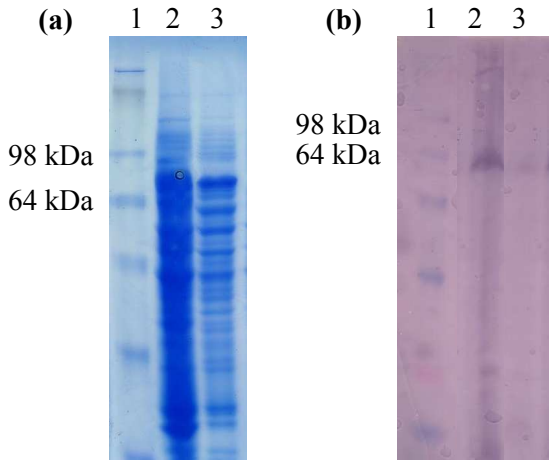


Figure 3.4.2 **(a)** SDS-PAGE gel. Lane (1) SeeBlue Plus 2 protein ladder. (2) Pellet after treatment with 6 M urea. (3) Supernatant after treatment with 6 M urea. **(b)** Western blot results. Lane (1) SeeBlue Plus 2. (2) Pellet portion after treatment with urea, band can be seen at around 64 kDa indicating the majority of DY3.4 was still present in the insoluble pellet. (3) Supernatant sample after resuspension in urea, only a small amount of soluble DY3.4 is present indicative of a faint band at approximately 64 kDa.

3.5 Conclusion and future work

TERT from *C. elegans* has proved to be a difficult protein to clone and over-express. Two constructs were successfully made pET41-DY3.4 GST-tagged and pET14b-DY3.4 His-tagged using the EK/LIC ligation and digestion independent system and direct cloning respectively. The former construct produced protein, but the protein was aggregated as judged from its elution in the void volume. The pET14b-DY3.4 construct produced inclusion bodies and was located in the insoluble pellet as identified via western blotting experiments. Attempts at solubilising the protein were not productive with no soluble protein obtained from the co-tranformation with pBAD-GroES/GroEL chaperones and only partial solubilisation was achieved using urea.

More work is required on the expression of soluble DY3.4. An alternative expression system could be used to help obtain non-aggregated soluble protein such as a yeast expression system or insect cells. This work was brought to a conclusion when the first full length TERT structure from *T. castaneum* was published (Gillis 2008).

The TR component for *C.elegans* TERT catalytic subunit is unknown. Possible future work could be the identification of the telomerase RNA component. There have been suggestions that this component may be tts-1 (Jones 2001), with possible similar secondary structure and common characteristics such as a pseudoknot, a telomeric template sequence and region of base pairing upstream and downstream of the telomeric template. A possible strategy for identifying TR is to synthesise tts-1 as DNA and use *in vitro* reconstitution (Xie 2008).

In vitro transcription/translation in rabbit reticulocyte lysate might be used for *in vitro* reconstitution of telomerase. It utilises the cellular components necessary for protein synthesis such as tRNA, ribosomes, amino acids, initiation, elongation and termination factors and chaperones needed for correct folding of proteins. Cloned TERT synthesised

in rabbit reticulocyte and *in vitro* synthesised TR could be added to the reaction to reconstitute telomerase (Xie 2008). The advantage of using such a system is that correct folding and TERT/TR interaction could be obtained with the help of correct chaperones that are already present in rabbit reticulocyte lysate. Past studies involving reconstitution of telomerase have used rabbit reticulocyte lysate successfully and produced active telomerase.

In the event that tts-1 is not the TR sequence an alternative method to hunt for the sequence requires using TERT protein as bait to isolate TR from *C.elegans* total RNA. Assuming TERT and TR binds tightly, total RNA added to His-tagged TERT protein synthesised in rabbit reticulocyte would allow selection of TERT protein bound with TR on a nickel column. The TERT bound RNA would then be used in RT PCR.

Chapter Four:
High resolution structure and substrate binding to CobJ

4.1 CobJ and ring contraction

Cobalamin (vitamin B12) has been described as the most complex small molecule because of the number of biosynthetic steps involved in its synthesis. There are two distinct, but related, pathways; an aerobic and an anaerobic route. The differences between the two lie in the requirement for molecular oxygen and at which enzyme mediated step cobalt chelation occurs. For the anaerobic pathway cobalt insertion takes place early on and requires no oxygen whereas the aerobic pathway requires molecular oxygen and adopts a late onset cobalt insertion (Battersby 1994; Raux 1996; Scott 2003).

Conversion of uroporphyrinogen III (uro'gen III) to cobalamin involves a number of types of reactions including methylation, amidation, removal of a ring carbon, insertion and reduction of a cobalt atom, addition of adenosyl moiety and the aminopropanol side chain. Of these reactions the focus will be on methylation at C17 and removal of a ring carbon in the pathway. In total there are eight methylations and six methyltransferases, in summary these include CobA (CysG) (anaerobic equivalents in brackets) which methylates uro'gen III at C2 and C7, CobI (CbiL) which methylates at C20, CobJ (CbiH) which methylates at C17, CobM (CbiF) which methylates at C11, CobF (CbiD) which methylates at C1 and CobL (CbiE and CbiT) which methylates at C5 and C15 (Crouzet 1990; Roth 1993). Several of these enzymes have auxiliary reactions, for example, CobJ catalyses ring contraction.

CobA is an S-adenosyl-L-methionine (SAM) dependent uro'gen III methyltransferase (SUMT) which catalyses the methylation of uro'gen III at C2 and C7 forming precorrin-2, only from this point does the pathway diverge into an anaerobic one. In the anaerobic pathway precorrin-2 is oxidised to sirohydrochlorin (oxidised form of precorrin-2) via SirC before insertion of a cobalt ion via a cobalt chelatase, CbiX or CbiK (Leech 2003; Raux 2003). In the aerobic pathway cobalt insertion occurs at hydrogenobyrinic acid

a,c-diamide via cobalt chelatase CobNST complex to generate cobyrinic acid a,c-diamide (Blanche 1993).

In this chapter the focus is on the ring contraction step of the bacterium *Rhodobacter capsulatus*. A contracted macrocycle is a unique characteristic of cobalamin. In the aerobic pathway; ring contraction is catalysed by the SAM dependent C17 methyltransferase, CobJ. In the past elaborate NMR labelling studies deduced a pinacol type arrangement as the mechanism for the ring contraction step, where CobG, a monooxygense uses molecular oxygen and sets up the ring contraction reaction by hydroxylating at C20 of the macrocycle producing a lactone ring and a methyl ketone pendant. CobJ subsequently methylates at C17 followed by a spontaneous ring contraction with the extrusion of C20 and bond formation between C1 and C19, this a reaction unprecedented in nature (see *Figure* 4.1.1) (Debussche 1993; Scott 1993).

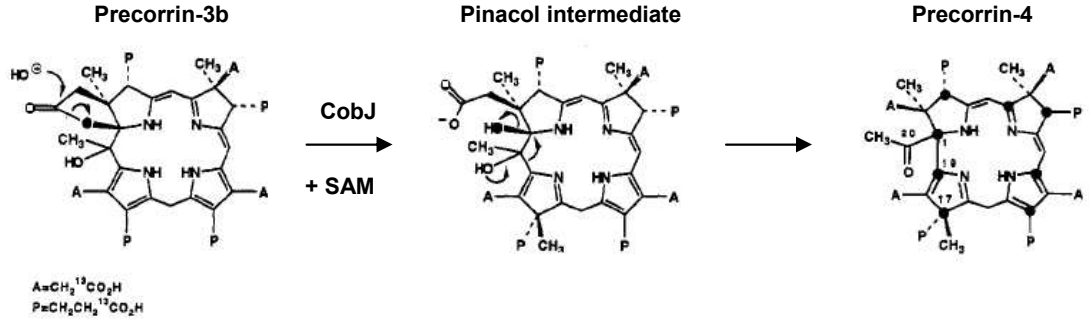


Figure 4.1.1 A schematic diagram showing the pinacol type arrangement proposed by Scott *et al* 1993 where precorrin-3b is synthesised via the hydroxylation of precorrin-3a (not shown) to form a lactone ring and a methyl ketone pendant.

However in *R. capsulatus* this CobG gene is missing and instead contains a gene CobZ which is able to compliment a CobG deficient strain (McGoldrick 2005). Other methyltransferases in the aerobic vitamin B_{12} pathway include CobA, CobI, CobM, CobF and CobL. Ring contraction in the anaerobic pathway is catalysed by CbiH and

mediates the formation of δ-lactone and ring contraction and methylation of a metallated macrocycle ring. (Roth 1993; Santander 1997)

The genes for cobalamin synthesis have been structurally and functionally characterised for *Pseudomonas denitrificans* and from this it was possible to compare with *Salmonella typhimurium* genes in the cob operon and assign the function of each. A high degree of similarity between the two shows that these genes had evolved from a common ancestor (Roth 1993). Similarly 80 ORFs were functionally assigned for *R. capsulatus* of which 26 of them matched with high similarity scores to the genes for synthesis of cobalamin in *P. denitrificans* and *S. typhimurium* (Vlcek 1997).

A number of structures for methyltransferases have been solved and these are summarised in *Table* 4.1.1. Obtaining an enzyme-substrate or -product complex has proved to be more challenging, some past studies have achieved such a feat including hydrogenobyrinic acid (HBA) bound to CobH and cobalt factor II bound to CbiL.

CobH is a methyl mutase that catayses the migration of a methyl group from C11 of precorrin-8x to C12 forming HBA. A crystal structure of CobH from *P. denitrificans* bound to its product HBA revealed a 1:1 stoichiometry with the observed product bound to near full occupancy in the active site. The binding pocket nearly encapsulates the entire tetrapyrrole with rings A and B near the opening of the active site exposed to solvent and rings C and D situated in the inner most part of the cavity with many hydrophobic residues making hydrogen bonds to the product. The opening of the pocket appeared too small to allow product to move in freely, which led to the conclusion of a C-terminal loop region that changed conformation allowing the entry of HBA. This was substantiated by the weak density observed in the residues making up this loop marking its flexibility (Shipman 2001).

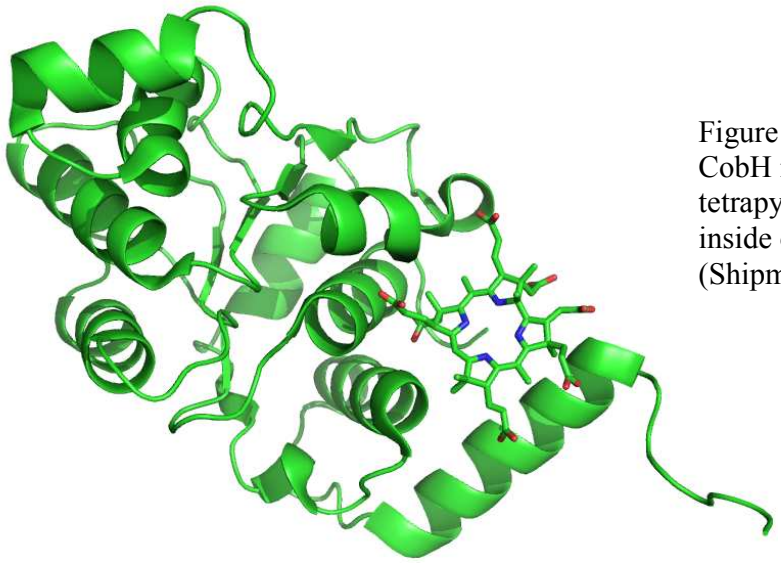


Figure 4.1.2 Ribbon diagram of CobH in complex with HBA. The tetrapyrrole is nearly encapsulated inside of the catalytic pocket (Shipman 2001).

A structure containing a doughnut-like ring of density near the catalytic pocket of CbiL has been reported. CbiL is a methyltransferase that catalyses the methylation of C20 via the methyl donor group SAM (SAH in the crystalline form) in the anaerobic pathway. Although a structure of cobalt factor II substrate bound to CbiL has been achieved, the density quality was poor for the substrate and therefore was not concisive in the residues responsible for binding of tetrpyrrole and resiudes involved in methylation and ring contraction. What was deducible from the result was that the density for cobalt factor II substrate would place C20 close to the sulphur group of SAH at 5.5 Å appropriate for methyl transfer (Frank 2007).

In order to elucidate the mechanism of ring contraction, thereby solving the puzzle of the unusual step of a contracting macrocycle, extensive work has been carried out to try and obtain an enzyme substrate complex through methods of crystallisation, crystal soaking and co-crystallisation with tetrapyrroles and small molecules. More probing was required since it is unclear whether the contraction of the macrocycle occurs before or after methylation, or indeed if these events are concerted.

Methylates	Enzyme	Pathway	Catalyses the formation of:	Structural data	Organism	Reference
C2 and C7	CobA	aerobic	uro'gen III into precorrin-2	2.70 Å	*Pseudomonas denitrificans*	(Vevodova 2004)
C2 and C7	CysG	anaerobic	uro'gen III into precorrin-2	2.21 Å	*Salmonella enterica*	(Stroupe 2003)
C20	CbiL	anaerobic	cobalt precorrin-2 into cobalt precorrin-3	2.10 Å	*Methanothermobacter thermautotrophicus.*	(Frank 2007)
C11	CbiF	anaerobic	cobalt precorrin-4 into cobalt precorrin-5	2.4 Å	*Bacillus megaterium*	(Schubert 1998)
C1	CobF	aerobic	precorrin-5 into precorrin-6	1.60 Å	*Corynebacterium diphtheriae*	(Nocek)
C1	CbiD	anaerobic	cobalt precorrin-5 into cobalt precorrin-6	1.90 Å	*Archaeoglobus fulgidus*	(Zhang)
C5	CbiE	anaerobic	Dihydro cobalt precorrin-6 into cobalt precorrin-7	2.27 Å	*Archaeoglobus fulgidus*	(Kim)
C15	CbiT	anaerobic	Cobalt precorrin-7 into cobalt precorrin-8	2.30 Å	*Methanocaldococcus jannaschii dsm*	(Padmanabhan)

Table 4.1.1 Details of all the methyltransferase structures solved from both aerobic and anaerobic pathways.

4.2 Protein production and purification of CobJ and CobG

4.2.1 CobJ native and mutant protein production

CobJ from *R. capsulatus* was cloned into an N-terminal His-tagged pET14b vector courtesy of Dr Evelyne Deery and Prof Martin Warren of Kent University additionally CobJ mutants pET14b-D82A and pET14b-H129A were kindly provided by Susanne Schroeder also of Kent University.

For protein over production of pET14b-CobJ and mutants they were transformed into BL-21 (DE3) pLysS cells (the protocol for transformation can be found in section 2.15 and 2.2.4). CobJ and mutant protein required 37 °C growth temperature until an $O.D_{600}$ of 0.6 was reached and induced at 18 °C overnight using 0.4 mM IPTG. Harvested cells were lysed using a sonnicator (see section 2.2.4.1).

4.2.2 CobJ native and mutant protein purification

His-tagged native CobJ and mutants were firstly purified on a nickel immobilised column (see section 2.2.5.1) fractions were collected at 400 mM imidazole and at 300 mM imidazole (Figure 4.1.3) and concentrated ready for size exclusion purification (see section 2.2.5.5).

Concentrated samples were purified on S200 size exclusion column as a final buffering step where buffer exchange into low concentration salt took place and removal of any aggregated proteins. CobJ eluted at approximately 16 ml, see Figure 4.1.4. Typically from one litre of cells 10 mg/ml of CobJ protein was produced.

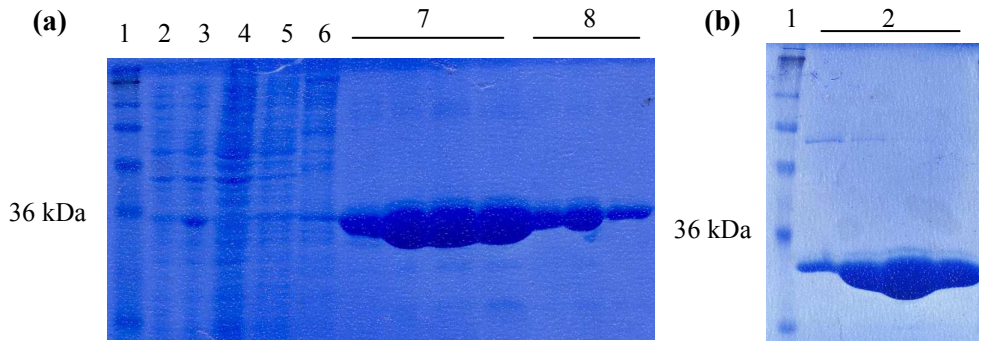


Figure 4.1.3 (**a)** CobJ SDS-PAGE gel showing a prominent band at approximately 36 kDa in the post induction and eluted fractions from a nickel immobilised column. (1) Protein ladder SeeBlue Plus (Invitrogen), (2) pre-induction, (3) post- induction, (4) flow-through, (5) wash I, (6) wash II, (7) wash III (collected as fractions for size exclusion chromatography), (8) eluted fractions. **(b)** SDS-PAGE gel showing fractions obtained after size exclusion chromatography on the S200 column. (1) Protein ladder SeeBlue Plus, (2) CobJ fractions from S200.

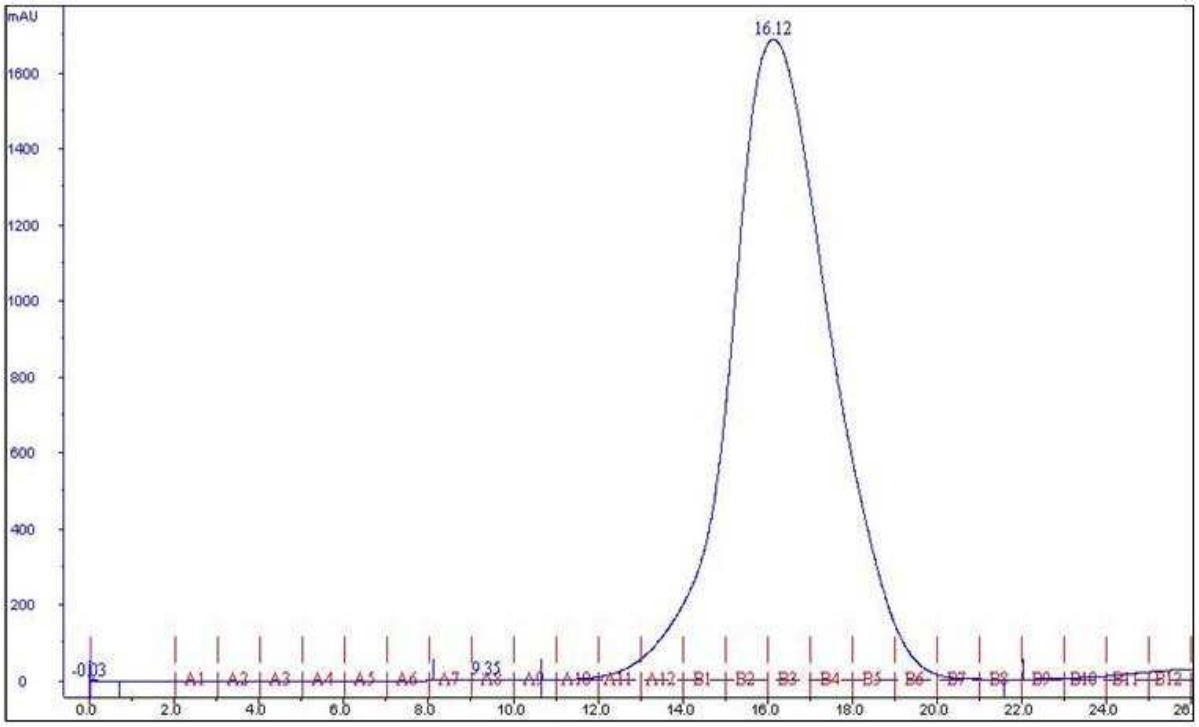


Figure 4.1.4 An FPLC trace at 280 nm showing a peak at 16.12 ml from a size exclusion S200 column where CobJ protein was eluted, fractions under the peak were checked on an SDS-PAGE gel pooled and then concentrated.

4.2.3 CobG protein production

CobG from *Synechococcus elongatus* was cloned into an N-terminal His-tagged pET14b vector courtesy of Dr Evelyne Deery and Prof Martin Warren from Kent University.

Protein over expression was achieved by transforming the construct pET14b- CobG into BL21 (DE3) pLysS cells (see section 2.2.4 for the protocol) and grown at 37 °C until an $O.D_{600}$ of 0.6 was reached followed by induction at 18 °C overnight, subsequent cells were harvested and lysed using a sonnicator (see section 2.2.4.1). The crude lysate was then transferred into the anaerobic chamber ready for protein purification since CobG has a 4Fe-4S cluster and is oxygen sensitive.

4.2.4 CobG protein purification

As described in section 2.2.8.6 all buffers and columns were degassed and placed into the anaerobic chamber the night before for equilibration. Since CobG posessess an iron-sulphur cluster and is oxygen sensitive the lysate was purified on a nickel immobilised column anaerobically (see section 2.2.5.1). Fractions were collected from 300 mM imidazole (wash III) and 400 mM imidazole see *Figure* 4.1.5 the fractions were pooled, buffer exchanged and purified on a pd-10 column (see section 2.2.5.4 for pd-10 column protocol).

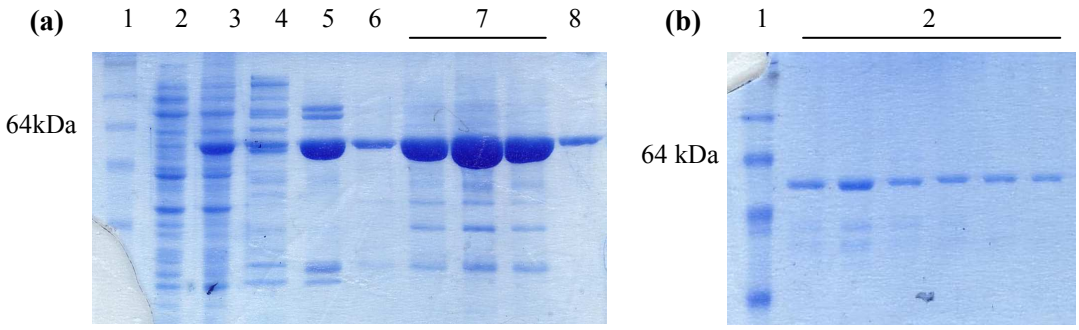


Figure 4.1.5 **(a)** CobG SDS-PAGE gel with a prominent band at around 64 kDa indicative of the size of pET14b-CobG in the post-induction, elution I and elution II lanes. Lane (1) Protein ladder SeeBlue Plus, (2) pre-induction, (3) post-induction, (4) flow-through, (5) wash I, (6) wash II, (7) elution I, (8) elution II. Fractions from lanes 7 and 8 were pooled and applied to a pd-10 column. **(b)** CobG SDS-PAGE gel showing fractions collected from a pd-10 column. Lane (1) Protein ladder SeeBlue plus, (2) fractions collected from pd-10 column.

4.3 Substrate production

4.3.1 Precorrin-3a

Precorrin-3a (PC-3a hereafter) is a tetrapyrrole synthesised in the vitamin B_{12} pathway by the methylation of Precorrin-2 at C20 via the methyltransferase CobI. In this thesis the production of PC-3a is described and was readily made. However PC-3a is not the correct substrate required for CobJ to carry out its ring-contraction step, ideally Precorrin-3b (PC-3b hereafter) was needed. Production of PC-3b requires CobG to hydroxylate C20 of PC-3a generating a gamma lactone structure and from this CobJ is then able to methylate at C17 and exude C20 from the macrocycle and form a bond between C1 and C19 (Schroeder 2009).

4.3.1.1 Multi-enzyme reaction

Production of substrate PC-3a required a multi-enzyme reaction containing CobI, CobA, HemB, HemC and HemD enzymes (see *Figure* 4.1.6) in an anaerobic glovebox (BelleTechnology) details of the method can be found in section 2.2.8.

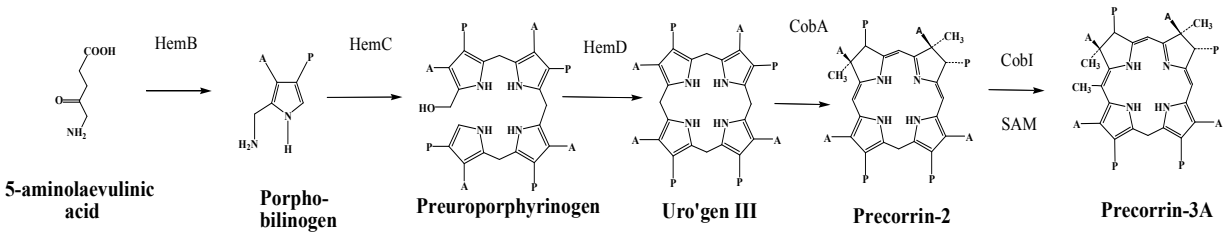


Figure 4.1.6 A schematic diagram depicting the multi-enzyme assay, which converts ALA into PC-3a via the enzymes HemB, HemC, HemD, CobA and CobI.

The products of the multi-enzyme assay were measured using a UV-3010 spectrophotmeter and a UV-visible trace was recorded over a range of 300 to 700 nm for three reactions A, B and C (see *Figure* 4.1.7). PC-3a produces a characteristic peak at around 420 nm coupled with a yellow coloured product under anaerobic conditions and fluoresces yellow-orange when exposed to UV light (see *Figure* 4.1.8). The product of reaction A produced precorrin-2 using enzymes HemB, HemC, HemD and CobA. Tetrapyrroles have a conjugated π-electron system and the arrangements of the electrons in the macrocycle ring give the compound its characteristic colour. PC-3a is light and oxygen sensitive, after long exposure to oxygen a peak maxima was observed at 378 nm with a shoulder at 411 nm and an increase in absorption at 593 nm indicating the oxidation of PC-3a to sirohydrochlorin. Sirohydrochlorin is characterised by absorption maximas at 376 nm and 590 nm (Raux 2003).

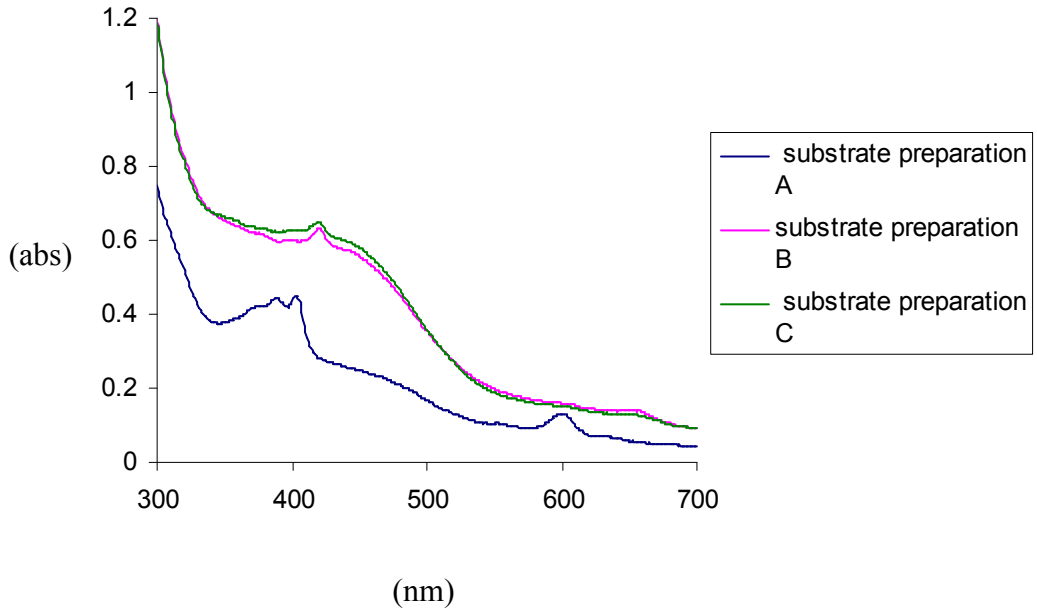


Figure 4.1.6 Shows a UV trace between 300 to 700 nm of each vial A, B and C where A was a negative control producing up to the tetrapyrrole precorrin-2 in the absence of all enzymes apart from CobI, reaction A and B both produced PC-3a, however the quantity of CobI in reaction B was half of that present in reaction A.

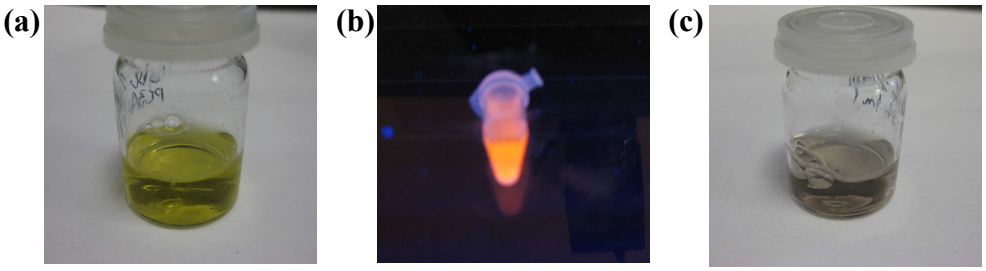


Figure 4.1.8 **(a)** PC-3a was a yellow product under anaerobic conditions and **(b)** fluoresces yellow-orange under UV light, the fluorescence changes to pink when exposed to oxygen **(c)** PC-3a was pink coloured under aerobic conditions.

4.3.2 Precorrin-3b

PC-3b as mentioned earlier is the tetrapyrrole CobJ methylates at C17 and ring contracts. The synthesis of PC-3b enzymatically uses the same multi-enzyme assay (see section 2.2.8.6) with the addition of purified CobG, all under anaerobic conditions. To activate CobG, molecular oxygen was introduced by taking the mixture out of the glovebox.

4.3.2.1 Production of Precorrin-3b

To start the production of PC-3b 2 ml of 0.2 mg/ml CobG was added to 500 µl of PC-3a inside the glovebox, to activate the reaction; oxygen was introduced by taking the mixture out of the anaerobic chamber. To begin with the colour of the mixture was green, as expected for CobG under anaerobic conditions (Schroeder 2009), however over time under oxygen exposure the green colour changed to a brown colour, coupled with this observation the fluorescence changed from a strong pink to a lighter pink fluorescence see *Figure* 4.1.9.

Further analysis by obtaining UV-visible traces after mixing PC-3a with CobG immediately after and 3.5 hours later showed a gradual change in the trace. Initial absorption maxima was recorded at 377 nm, a shoulder peak at 411 nm and two small absorption peaks at 590 and 630 nm. Gradually overtime the peaks flatten out and become more reminiscent of PC-3a when exposed to oxygen (*Figure* 4.1.10).

It appeared that the reaction of PC-3a with CobG did not produce PC-3b since the product absorption maximas were very similar to that of PC-3a under aerobic conditions. The disappearance of fluorescence was caused by the possible subsequent metallation that occurs overtime from metal ions. Fluorescence is quenched by proximity of a coordinating metal.

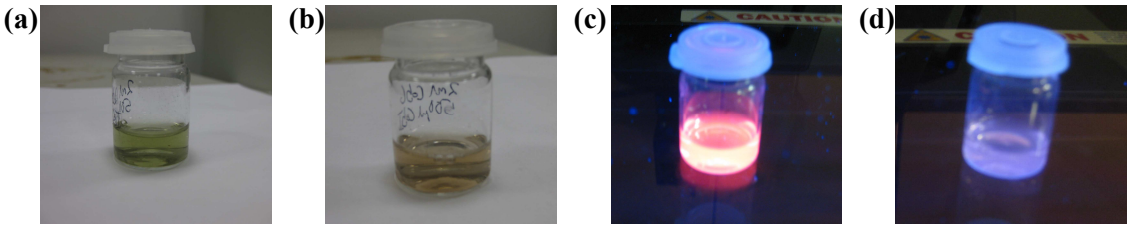


Figure 4.1.9 The reaction of CobG with PC-3a changes from **(a)** green to **(b)** brown over time. **(c)** The green substance fluoresces pink **(d)** gradually the pink fluorescence decreases.

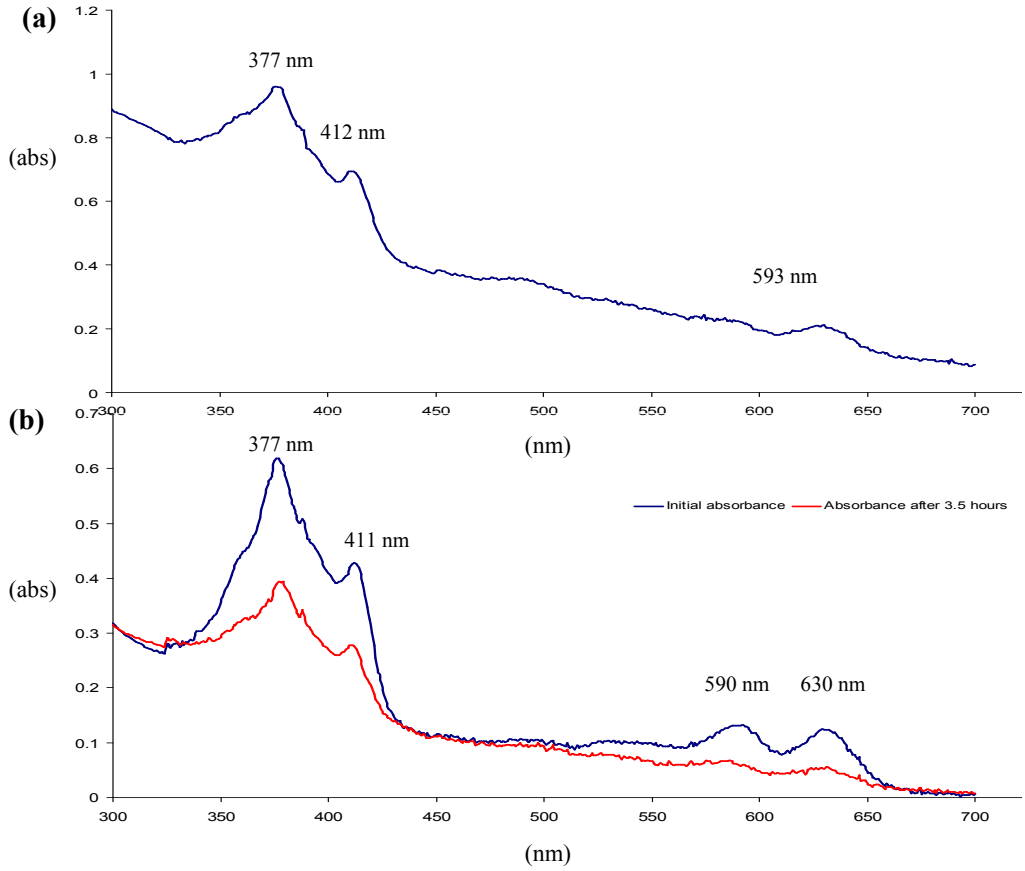


Figure 4.1.10 **(a)** UV- visible trace for PC-3a after exposure to oxygen with peak maxima at 375 nm, a shoulder peak at 412 nm and small bump at 593 nm. **(b)** Two UV-visible traces from the same product after mixing PC-3a with CobG under aerobic conditions. After 3.5 hours the trace flattened out considerably and resembles PC-3a under aerobic conditions with similar UV-visible trace characteristics.

4.4 Native and mutant CobJ crystallography

A selection of CobJ mutants were designed based on analysis of a 2.7 Å native CobJ structure solved by Dr Thomas Hutchison. *In vitro* experiments were carried out to test the enzyme activity by using a *Salmonella* strain deficient in cobalamin synthesis and complementation with various mutant CobJ plasmids revealed three mutants D82A and H129A with lowest activities at 1.5 % and 4.3 % respectively. This formed the basis of

motivation behind crystallising and working with these particular mutants. Other mutants were also tested and the relative activity can be seen in *Figure* 4.1.11. Activity assay results are courtesy of our collaborators at Kent University: Susanne Schroeder and Prof Martin Warren.

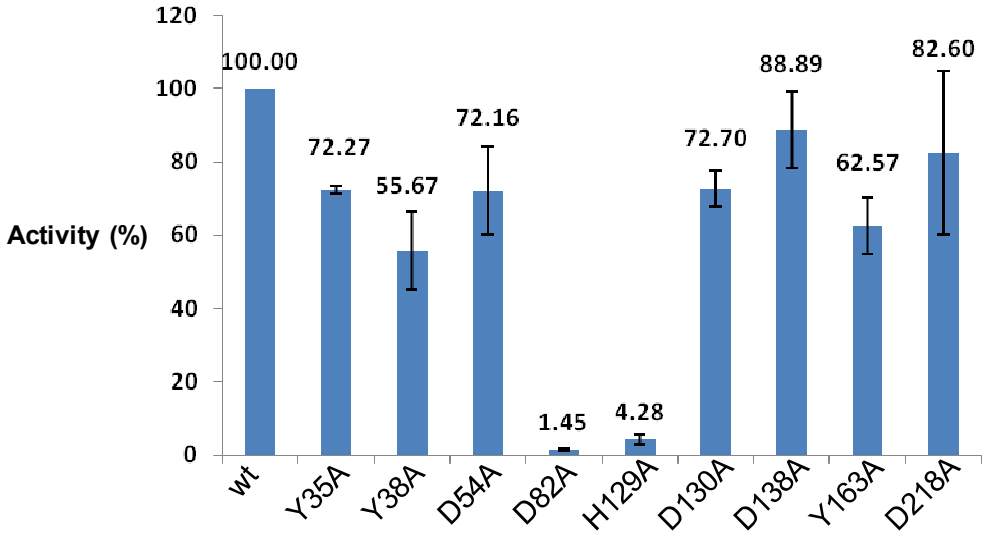


Figure 4.1.11 Summary of the relative activities of various mutants, courtesy of Susanne Schroeder of Kent University.

Purified and concentrated native and mutant (D82A and H129A) CobJs were used to set up crystallisation trays as detailed in section 2.2.10.1. For CobJ protein all crystallisation reservoirs were optimised conditions from Hampton research crystal screen 1 (CSI) condition 46, consisting of 0.2 M calcium acetate hydrate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 18 % w/v PEG 8000.

4.4.1 Crystallisation of native and mutant CobJ

Optimised crystallisation conditions used for native and mutant CobJs are listed in *Table* 4.1.2. A hanging drop set-up was used throughout with 2 µl reservoir mixed with 2 µl protein at 7 mg/ml which was supplemented with fresh SAM at varying concentrations see *Figure* 4.1.12.

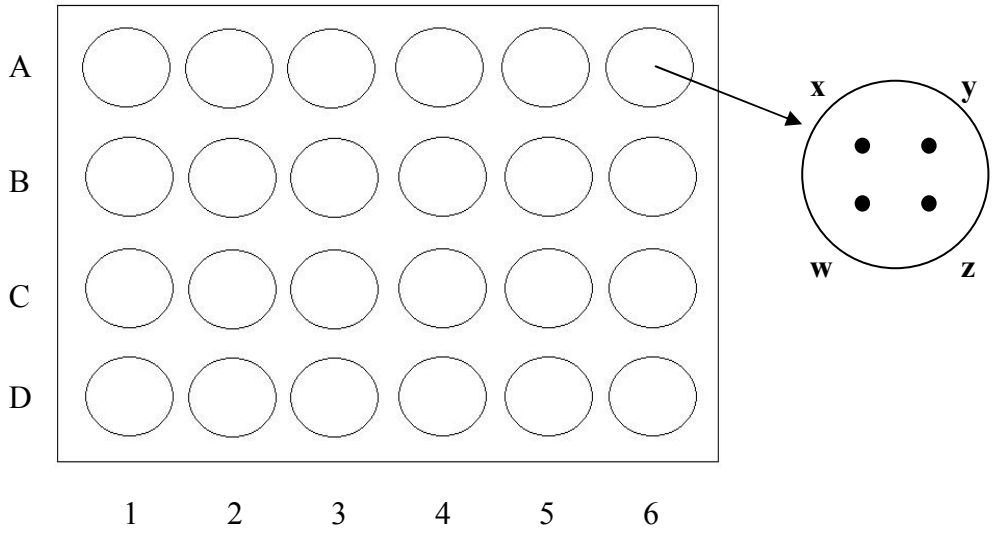


Figure 4.1.12 A schematic diagram of a 24-well hanging drop plate from Hampton research. Each greased well contained 500 µl of reservoir covered with a siliconated glass slide forming an air-tight seal. Each slide had four drops with reservoir and protein supplemented with SAM at different concentrations. (X) CobJ with no SAM, (Y) CobJ with 0.6 mM SAM, (Z) CobJ with 1.2 mM SAM, (W) CobJ with 1.8 mM SAM.

Reservoir	**Sodium cacodylate trihydrate pH 6.5 (mM)**	**Calcium acetate hydrate (mM)**	**PEG 8000 (%)**	**ddH_2O (µl)**
A1	0.1	0	14	320
A2		100		270
A3		200		220
A4		300		170
A5	0.1	0	15	307.5
A6		100		257.5
B1		200		207.5
B2		300		157.5
B3	0.1	0	16	295
B4		100		245
B5		200		195
B6		300		145
C1	0.1	0	17	282.5
C2		100		232.5
C3		200		182.5
C4		300		132.5
C5	0.1	0	18	270
C6		100		220
D1		200		170
D2		300		120

Table 4.1.2 Lists the optimised reservoir conditions used for the crystallisation of native CobJ and mutant CobJs. Stock concentrations used include 10 mM sodium cacodylate, 1 M calcium acetate and 40 % PEG 8000. The final volume of all reservoirs were 500 µl.

Native CobJ protein was readily crystallised as was mutant H129A, however D82A proved harder to produce good quality crystals see *Figure* 4.1.13. Native CobJ crystals measured 0.1 mm wide by 0.2 mm long and were produced from optimised conditions from CSI condition 46 as did D82A and H129A. H129A formed crystals which were approximately 0.2 mm in length. Crystal trays set up for D82A proved unsuccessful in many cases with precipitation forming in the drops. In order to combat this problem precipitant (PEG 8000) was reduced by half for each of the original reservoir conditions for example 14 % PEG 8000 was reduced to 7 % PEG 8000 (see *Table* 4.1.2 for crystallisation conditions) and in addition protein concentration was also reduced to 4 mg/ml. This enabled formation of small crystals of approximately 0.05 mm.

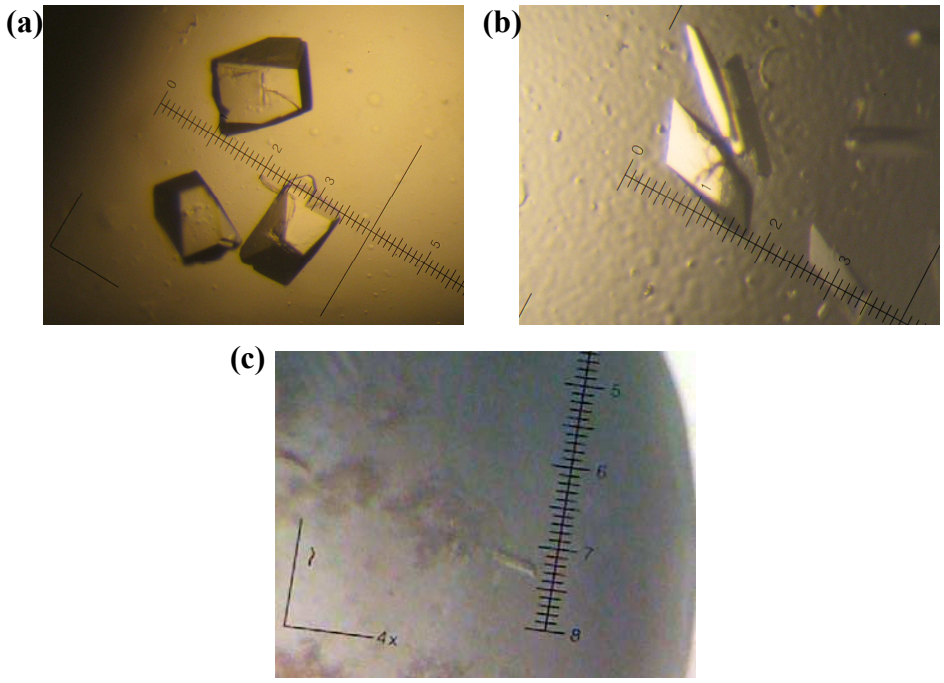


Figure 4.1.13 *R.capsulatus* CobJ crystals obtained from hanging drop vapour diffusion trays crystallization trials. **(a)** Native CobJ crystals from condition B4, **(b)** H129A crystals from condition A6, **(c)** D82A crystal from condition C3. For reservoir conditions see *Table* 4.1.2. Every ten small divisions on the scale bar was equivalent to 0.1 mm.

4.4.2 Crystal soaking

Crystal soaking experiments were carried out aerobically and anaerobically in an anaerobic chamber, a general soaking method can be found in section 2.2.10.2. For more specific native and mutant CobJ crystal soaks in a variety of environments see *Table* 4.1.3, which details the type of crystal (reservoir the crystal was crystallised from and SAM concentration used), soak formula, soak time and any observations made during the time. The majority of the soaks were supplemented with pyrrole-2-carboxylic acid and PC-3a.

	Soak condition	**Crystal type**	**Soak time (mins)**	**Observations**
1	PC-3a, 20 % glycerol anaerobic	WT A1 1.2 mM SAH	N/a	Crystal cracked
		H129A A1 1.2 mM SAH		Crystal cracked
2	40 mM P2CA, 20 % glycerol	WT B1 0.6 mM SAH	10	Clear drop
		WT A1 1.2 mM SAH		
		H129A A1 1.2 mM SAH		
4	40 mM P2CA, 20 % glycerol	WT A3 1.8 mM SAH	60	For all crystals a hard crusty skin had formed after 60 minutes.
		H129A A1 1.2 mM SAH		
5	100 mM P2CA in DMSO, 20 % glycerol	WT A3 1.8 mM SAH	30	For all crystals a hard crusty skin had formed after 30 minutes.
		H129A A1 1.2 mM SAH		
6	70 mM P2CA in DMSO, 20 % glycerol	WT A3 1.8 mM SAH	10	Clear drop
		H129A A1 1.2 mM SAH		

Table 4.1.3 **PC-3A = precorrin-3a and P2CA = pyrrole-2-carboxylic acid**
Conditions used for soak experiments for wild type (WT) CobJ and mutants are listed. Crystal type lists the reservoir the crystal originates from e.g. A1, details of the composition can be found in *Table* 4.1.2 in section 4.4.1 and the concentration of SAH.

Crystal soaking using PC-3a proved to be unsuccessful as any crystal subjected to a soaking experiment with the tetrapyrrole caused the crystal to crack instantly as seen in *Figure* 4.1.14. The cryoprotectant consisted of the reservoir condition of the crystal supplemented with 20 % glycerol and the remaining volume made up with PC-3a (see section 2.2.10.2).

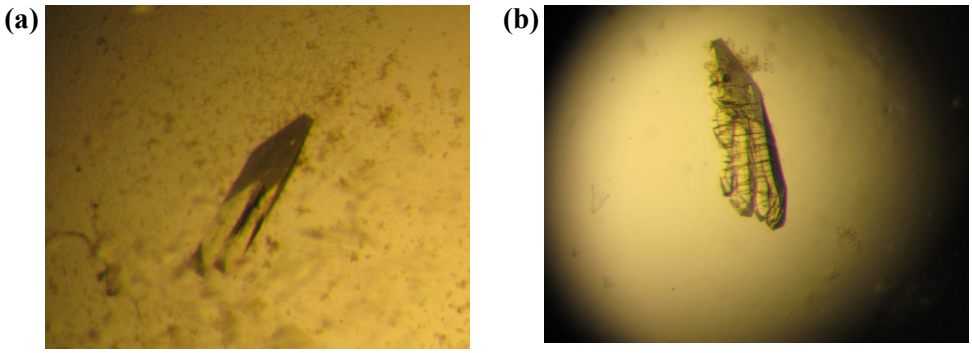


Figure 4.1.14 **(a)** A native CobJ crystal grown inside of an anaerobic chamber. **(b)** The same crystal after being soaked in cryoprotectant supplemented with PC-3a.

4.4.3 Co-crystallisation

Soaking experiments with PC-3a caused immediate cracking of the crystal suggesting that the substrate was causing some change in conformation of the enzyme which in turn caused disruption of the crystal lattice. SAM is hydrolysed to SAH during crystallisation. An alternative method to crystal soaking was to use co-crystallisation of a pre-formed enzyme-tetrapyrrole complex. This circumvents any problems caused by the enzyme moving in response to the tetrapyrrole binding. However this method does require the complex to crystallise rather than exploiting a pre-grown crystal.

For co-crystallisation a selection of molecules were used including tetrapyrroles PC-3a and hydrogenobyrinic acid known as HBA hereafter (HBA courtesy of Dr Arefeh Seyedarabi), in addition one other tetrapyrrole was used which was precorrin-2. This was produced as a negative control in the multi-enzyme assay during the production of PC-3a (section 4.3.1.1). The product was purified in the same way as PC-3a, however no testing was carried out to ensure that this was in fact precorrin-2 produced, but was used in co-crystallisation anyway in the hope of a binding tetrapyrrole.

Other small molecules used in co-crystallisation experiments were pyrrole-2-carboxylic acid and porphobilinogen both chosen for its small size and presence of a pyrrole ring, the basis of a tetrapyrrole. Conditions used are detailed in *Table* 4.1.4 and co-crystallised crystals successfully obtained can be seen in *Figure* 4.1.15.

Crystal type	**Co-crystallised with**
CobJ WT A2 1.6 mM SAH	PC-3a
CobJ WT A6 1.6 mM SAH	PC-3a
CobJ WT B5 1.2 mM SAH	PC-3a
CobJ WT B4 1.6 mM SAH	PC-3a
CobJ WT A3 0.6 mM SAH	P2Ca 10 mM
CobJ WT B1 0.6 mM SAH	P2Ca 10 mM
CobJ WT D2 1.8 mM SAH	PC-2
CobJ WT D1 0.6, 1.2 and 1.8 mM SAH	PC-2
CobJ WT C6 no SAH	PC-2
CobJ WT	HBA
CobJ H129A A4 1.8 mM SAH	P2Ca 5mM
CobJ WT A1 0.6 mM SAH	PBG 2 mM

Table 4.1.4 **PC-3a = precorrin-3a, PC-2 = precorrin-2, P2Ca = pyrrole-2-carboxylic acid, HBA = hydrogenobyrinic acid, PBG = porphobilinogen**

Details of the types of crystals obtained from co-crystallisation experiments with various small molecules and tetrapyrroles. Glycerol was used as a cryoprotectant in all cases at 20 % (v/v).

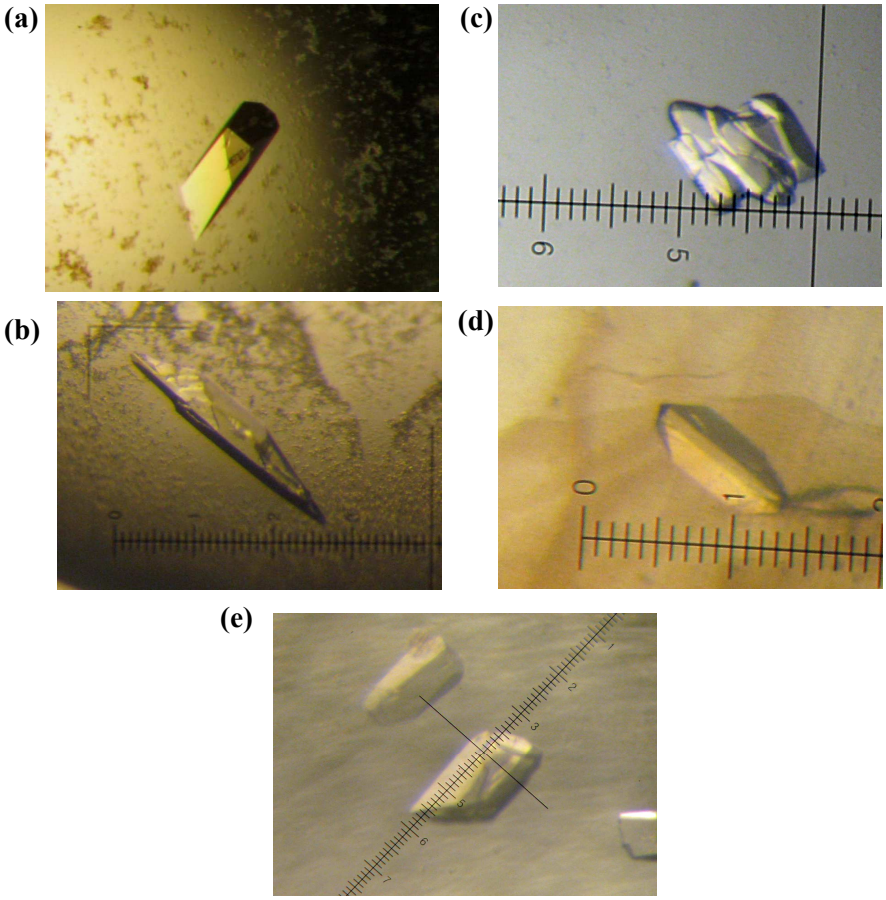


Figure 4.1.15 Examples of CobJ native crystals successfully formed from co-crystallisation experiments. **(a)** co-crystallised with PC-3a from condition B5, **(b)** co-crystallised with PBG from condition A1, **(c)** co-crystallised with PC-2 from condition D2, **(d)** co-crystallised with P2Ca from condition A4, **(e)** co-crystallised with HBA from condition D2. For reservoir conditions see *Table* 4.1.2. Every ten small divisions on the scale bar is equivalent to 0.1 mm.

4.4.4 pET14b-CobJ from pull-down assay

Aside from crystal soaking and co-crystallisation experiments, a pull-down experiment was used in an attempt to trap a complex. Courtesy of our collaborators from Kent University, Dr Evelyn Deery and Prof Martin Warren provided crystals of His-tagged CobJ grown from 0.1 M sodium cacodylate pH 6.9, 0.2 M calcium acetate and 22-26 % PEG 8000. Two separate constructs were used pET14b-CobJ and pETcoco2-ABCD, CobI and CobG. This approach enables the pull-down and crystallisation of His-tagged CobJ and any bound product.

The construct produces CobJ protein of a strong blue hue aerobically. However when purified anaerobically CobJ was yellow in colour but once exposed to molecular oxygen the colour changed to blue (Debussche 1993). This observation would appear to suggest that precorrin-4 is present with CobJ enzyme anaerobically but once exposed to aerobic conditions the tetrapyrrole oxidises to factor IV (blue in colour). These simple experiments suggest that CobJ is being purified along with its product as had previously been seen with CobH and HBA (Prof Martin Warren, Kent University).

Product binding may be a method of protecting oxygen-sensitive intermediates in the cobalamin biosynthetic pathway and may support the idea of 'retrograde evolution' as suggested by (Horowitz 1945).

4.4.5 High resolution data collection of native CobJ

A 2.7 Å structure of CobJ was solved by Dr Thomas Hutchison using Se-Met enriched protein and single anomalous dispersion (SAD). Crystals grew in clusters and a single crystal required dissecting from the rest before data collection, however single straight edged crystals typically no smaller than 0.1 mm (see *Figure* 4.1.13) were successfully obtained for native CobJ from reservoir condition B4 (see *Table* 4.1.2) with 1.2 mM SAM. Crystals were supplemented with 20 % glycerol prior to freezing as a cryoprotectant. From these crystals a 2.2 Å high resolution dataset was collected at Diamond Light Source synchrotron beamline IO3. Pre-frozen crystals were automatically mounted via a Rigaku ACTOR sample changer and a total of 180 images were collected in 1° increments at a wavelength of 0.9808 Å. The detector used was an ADSC Q315r detector. These native CobJ crystals belonged to spacegroup of $P2_12_12_1$ with cell dimensions of a= 85.64 Å, b= 110.99 Å, c= 116.19 Å and with four molecules (two dimers) in the asymmetric unit. The crystals were the same as previously described, but diffracted to a higher resolution.

Collected data were processed using iMOSFLM (Leslie 1992) and reduced using SCALA from the CCP4i suite (Dodson 1997) see *Table* 4.1.5 for further crystallographic statistics. Structures were viewed using COOT (Emsley 2004). The structure is a closely associated homodimer with a buried area of 4780 $Å^2$, but it crystallises with four chains forming a dimer of dimers with only 750 $Å^2$ buried area between the authentic dimers, values were calculated using EBI-PISA (Krissinel 2007). One SAM molecule is bound to each chain in the hydrolysed form SAH (Figure 4.1.16).

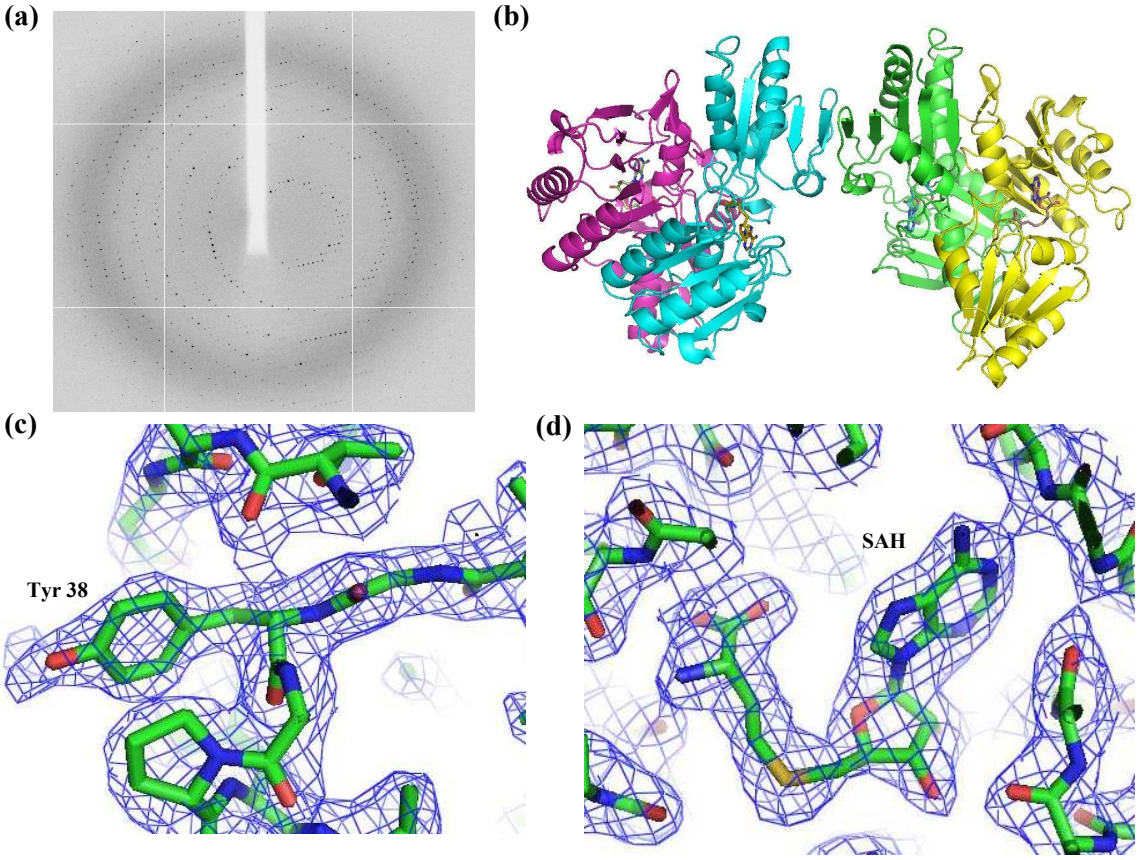


Figure 4.1.16 **(a)** A diffraction pattern obtained for a native CobJ crystal diffracting to 2.2 Å **(b)** A ribbon diagram depicting four chains in the asymmetric unit in native CobJ, the adjoining homodimers have a small contact surface area with 750 square Å. **(c)** and **(d)** A 2.2 Å structure provided strong electron density for example residue Tyr 38 and molecule SAH. Figures were drawn using Pymol.

	Overall	Inner shell	Outer shell
Low resolution limit (Å)	58.56	58.56	2.34
High resolution limit (Å)	2.22	7.03	2.22
[a]Rmerge (%)	0.081	0.026	0.525
[b]Rmeas (within I+/I-) (%)	0.094	0.031	0.613
[b]Rmeas (all I+ & I-) (%)	0.094	0.031	0.613
[c]Rpim (within I+/I-) (%)	0.046	0.016	0.309
[c]Rpim (all I+ & I-) (%)	0.046	0.016	0.309
Fractional partial bias	-0.016	-0.023	-0.068
Total number of observations	215361	6705	31449
Total number unique	54689	1821	7910
Mean [(I)/sd(I)] (%)	12.3	51.8	2.6
Completeness (%)	99.4	96.3	99.9
Multiplicity	3.9	3.7	4.0
Wilson B-factor ($Å^2$)	38.27		

Table 4.1.5 Details of the crystallographic statistics for native CobJ.
[a] Rmerge = ($\Sigma hkl\ \Sigma j\ |\ Ihkl - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl,j$)
[b] R*meas* = ($\Sigma hkl \sqrt{(n/n-1)}\ \Sigma^{n}_{j=1} |\ Ihkl,j - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl, j$)
[c] Rpim = ($\Sigma hkl \sqrt{(1/n-1)}\ \Sigma^{n}_{j=1} |\ Ihkl,j - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl, j$)

4.4.5.1 Data collection for native CobJ co-crystallised with PC-3a

Data was collected to 2.75 Å at the European Synchrotron Radiation Facility (ESRF) using beamline ID 14-1 for native CobJ co-crystallised with PC-3a (see *Figure* 4.1.15 panel a). This particular crystal was from reservoir condition B5 with 0.6 mM SAM added (*Table* 4.1.2) and the cryoprotectant used was 20 % glycerol. The detector was a ADSC Q210 CCD detector and the wavelength used was 0.9334 Å. In total 180 images were collected using a 1 ° oscillation range. The spacegroup was identified as $C222_1$ using POINTLESS (Evans 2005) with cell edges of a= 49.73 Å, b= 112.85 Å and c= 91.25 Å. Images were processed using iMOSFLM (Leslie 1992) and reduced using SCALA from the CCP4i suite (Dodson 1997) see *Table* 4.1.6 for further crystallographic statistics. Matthews coefficient (Kantardjieff 2003) analysis shows there is one molecule in the asymmetric unit with 47.5 % solvent content (see *Table* 4.1.7).

	Overall	Inner shell	Outer shell
Low resolution limit (Å)	48.09	48.09	2.90
High resolution limit (Å)	2.75	8.70	2.75
[a]Rmerge (%)	0.103	0.042	0.631
[b]Rmeas (within I+/I-) (%)	0.112	0.047	0.682
[b]Rmeas (all I+ & I-) (%)	0.112	0.047	0.682
[c]Rpim (within I+/I-) (%)	0.043	0.021	0.256
[c]Rpim (all I+ & I-) (%)	0.043	0.021	0.256
Fractional partial bias	-0.052	-0.064	-0.172
Total number of observations	47126	1208	6962
Total number unique	7019	241	1007
Mean [(I)/sd(I)] (%)	12.2	32.4	2.7
Completeness (%)	99.7	92.6	100.0
Multiplicity	6.7	5.0	6.9
Wilson B-factor ($Å^2$)	66.509		

Table 4.1.6 Details of the crystallographic statistics obtained for CobJ co-crystallised with PC-3a.

[a] Rmerge = ($\Sigma hkl\, \Sigma j \mid Ihkl - \langle Ihkl \rangle \mid$) / ($\Sigma hkl\, \Sigma j\, Ihkl,j$)

[b] R*meas* = ($\Sigma hkl \sqrt{(n/n-1)}\, \Sigma^{n}_{j=1} \mid Ihkl,j - \langle Ihkl \rangle \mid$) / ($\Sigma hkl\, \Sigma j\, Ihkl,j$)

[c] Rpim = ($\Sigma hkl \sqrt{(1/n-1)}\, \Sigma^{n}_{j=1} \mid Ihkl,j - \langle Ihkl \rangle \mid$) / ($\Sigma hkl\, \Sigma j\, Ihkl,j$)

Nmol/asym	Matthews coeff	% solvent
1	2.34	47.5

Table 4.1.7 Statistics from Matthews coefficient suggests one molecule in the asymmetric unit and corresponding solvent content.

4.4.5.2 Data collection for native CobJ co-crystallised with porphobilinogen

Data was collected for native CobJ co-crystallised with porphobilinogen to a resolution of 1.97 Å, the crystal (see *Figure* 4.1.15 panel c) was from condition A1 (see *Table* 4.1.2) supplemented with 0.6 mM SAM and 2 mM porphobilinogen. The cryoprotectant used was 20 % glycerol. The beamline for data collection was the same for native CobJ co-crystallised with PC-3a and a total of 180 images were collected. The crystal had cell dimensions a= 68.31 Å, b= 69.43 and c= 72.65 and β= 118.16. The program POINTLESS (Evans 2005) suggested the spacegroup $P2_1$.

Collected images were processed using iMOSFLM (Leslie 1992) and reduced using SCALA from the CCP4i suite (Dodson 1997) see *Table* 4.1.8 for further crystallographic statistics. Matthews coefficient (Kantardjieff 2003) analysis suggests there are two molecules in the asymmetric unit with 53.4 % of solvent (see *Table* 4.1.9).

Although CobJ was co-crystallised with porphobilinogen none was found to bind in the catalytic pocket, but bound elsewhere non-specifically. A total of four molecules of porphobilinogen were bound non-specifically.

	Overall	Inner shell	Outer shell
Low resolution limit (Å)	34.72	34.72	2.08
High resolution limit (Å)	1.97	6.23	1.97
[a]Rmerge (%)	0.195	0.039	0.792
[b]Rmeas (within I+/I-) (%)	0.232	0.047	0.927
[b]Rmeas (all I+ & I-) (%)	0.232	0.047	0.927
[c]Rpim (within I+/I-) (%)	0.124	0.026	0.479
[c]Rpim (all I+ & I-) (%)	0.124	0.026	0.479
Fractional partial bias	-0.081	-0.037	-0.315
Total number of observations	151141	4343	22774
Total number unique	41959	1319	6172
Mean [(I)/sd(I)]	7.3	14.0	2.7
Completeness (%)	98.8	94.0	100
Multiplicity	3.6	3.3	3.7
Wilson B-factor ($Å^2$)	16.99		

Table 4.1.8 Details of crystallographic statistics for native CobJ co-crystallised with porphobilinogen.
[a] Rmerge = ($\Sigma hkl\ \Sigma j\ |\ Ihkl - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl,j$)
[b] R*meas* = ($\Sigma hkl \sqrt{(n/n-1)}\ \Sigma^{n}_{j=1}\ |\ Ihkl,j - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl, j$)
[c] Rpim = ($\Sigma hkl \sqrt{(1/n-1)}\ \Sigma^{n}_{j=1}\ |\ Ihkl,j - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl, j$)

Nmol/asym	Matthews coeff	% solvent
1	5.51	77.69
2	2.76	55.39
3	1.84	33.08
4	1.38	10.77

Table 4.1.9 Statistics from Matthews coefficient indicating two molecules in the asymmetric unit.

4.4.6 Data collection of mutant CobJ

Data was collected for the H129A mutant at Diamond Light Source (DLS) beamline IO2 using an ADSC Q315 CCD detector and 0.979 Å wavelength. The sample was frozen using 20 % glycerol as cryoprotectant prior to data collection; a Rigaku ACTOR automatic sample changer was used.

4.4.6.1 Data collection for H129A soaked in pyrrole-2-carboxylic acid

Mutant CobJ H129A was crystallised from reservoir condition A6 (see *Table* 4.1.2) supplemented with 0.6 mM SAM. The crystal was soaked in 70 mM pyrrole-2-carboxylic acid for ten minutes with 20 % glycerol as cryoprotectant and frozen in liquid nitrogen (see section 2.2.10.2). The data was collected using a 1 ° oscillation and a total of 230 images were recorded at 2.5 Å resolution. The spacegroup was $P2_12_12_1$ with cell dimensions a= 84.96 Å, b= 109.43 and c= 114.65 Å.

The structure had four molecules in the asymmetric unit as suggested by the Matthews coefficient (Kantardjieff 2003) with 50.9 % solvent (see *Table* 4.1.11). For more detailed crystallographic statistics collected for H129A mutant see *Table* 4.1.10. Although the crystal was soaked in pyrrole-2-carboxylic acid no small molecule was seen in the map bound near the catalytic pocket or non-specifically elsewhere.

	Overall	Inner shell	Outer shell
Low resolution limit (Å)	57.2	57.2	2.64
High resolution limit (Å)	2.50	7.91	2.50
[a]Rmerge (%)	0.110	0.040	0.611
[b]Rmeas (within I+/I-) (%)	0.122	0.045	0.681
[b]Rmeas (all I+ & I-) (%)	0.122	0.045	0.681
[c]Rpim (within I+/I-) (%)	0.053	0.020	0.296
[c]Rpim (all I+ & I-) (%)	0.053	0.020	0.296
Fractional partial bias	-0.002	-0.041	-0.103
Total number of observations	190643	6313	27721
Total number unique	37611	1312	5384
Mean [(I)/sd(I)]	8.7	17.3	2.4
Completeness (%)	99.8	98.8	99.9
Multiplicity	5.1	4.8	5.1
Wilson B-factor (Å^2)	53.73		

Table 4.1.10 Crystallographic data statistics for CobJ mutant H129A, data collected at 2.50Å.

[a] Rmerge = ($\Sigma hkl\ \Sigma j\ |\ Ihkl - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl,j$)

[b] R*meas* = ($\Sigma hkl \sqrt{(n/n-1)}\ \Sigma^{n}_{j=1}\ |\ Ihkl,j - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl, j$)

[c] Rpim = ($\Sigma hkl \sqrt{(1/n-1)}\ \Sigma^{n}_{j=1}\ |\ Ihkl,j - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl, j$)

Nmol/asym	Matthews coeff	% solvent
1	10.01	87.71
2	5.00	75.43
3	3.34	63.14
4	2.50	50.86
5	2.00	38.57
6	1.67	26.29
7	1.43	14.00
8	1.25	1.72

Table 4.1.11 Statistics from Matthews coefficient shows there are four molecules in the asymmetric unit for H129A mutant structure with 50.86 % solvent.

The absence of the histidine side chain (H129A) was confirmed by inspection of the map, where electron density for the imidazole group was missing see *Figure* 4.1.17. His 129 from one subunit is close to the SAH binding centre of the other subunit in the homodimer (*Figure* 4.1.18).

Once the structure was refined it was superimposed onto native CobJ using SUPERPOSE (Krissinel 2004) for analysis of any structural changes caused by the mutation. From the superimposition no major structural differences can be seen between native and mutant CobJ (rmsd length 0.019, rmsd angle 1.842 and rmsd length 0.016, rmsd angle 1.723 respectively) therefore any reduced enzyme activity observed was not due to structural changes of the enzyme as a whole, but from the individual residue mutation H129A, which has caused the decrease in enzyme function. It is possible that H129 facilitates ring contraction.

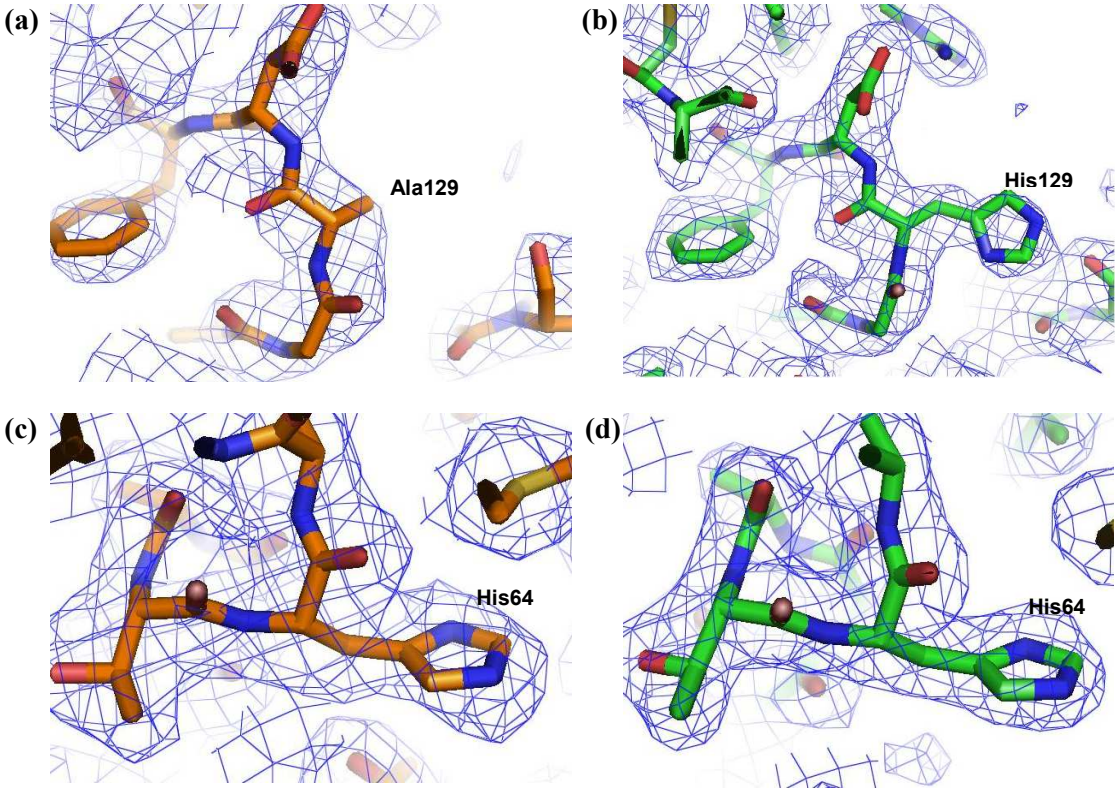


Figure 4.1.17 Comparison of H129A and native CobJ structures highlighting the absence and presence of His 129 respectively. **(a)** CobJ mutant H129A can be seen with no density corresponding to His 129, **(b)** native CobJ showing His 129, **(c)** as a control His 64 is shown with strong density in the H129A structure, **(d)** and similarly native CobJ structure and His 64 shows clear density.

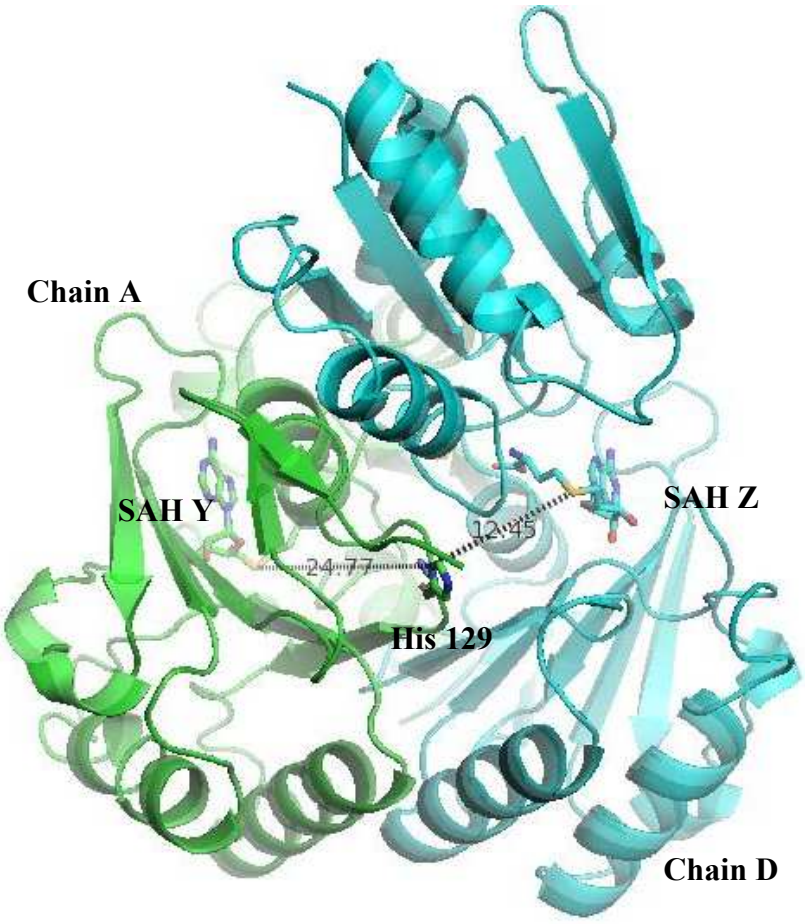


Figure 4.1.18 A dimer consisting of chain A and chain D with its corresponding SAH Y and Z respectively. Highlighted residue His 129 belongs to chain A, however it is closer to SAH at the active centre of chain D (12.5 Å) compared to that of chain A (24.8 Å).

4.4.7 Data collection for pET14b-CobJ from pull-down assay

The data was collected to a resolution of 2.5 Å using a 1 ° oscillation at DLS, beamline IO4. The spacegroup was $P2_12_12_1$ with cell dimensions a= 84.96 Å, b= 109.270 and c= 113.330 Å for more crystallographic statistics see *Table* 4.1.12. On examination of the density map precorrin-4 was not bound to the active site.

	Overall	Inner shell	Outer shell
Low resolution limit (Å)	67.98	67.98	2.65
High resolution limit (Å)	2.58	11.55	2.58
[a]Rmerge (%)	0.08	0.023	0.733
[b]Rmeas (within I+/I-) (%)	0.094	0.027	0.854
[b]Rmeas (all I+ & I-) (%)	0.094	0.027	0.854
[c]Rpim (within I+/I-) (%)	0.035	0.011	0.312
[c]Rpim (all I+ & I-) (%)	0.048	0.014	0.435
Fractional partial bias	0.0	0.0	0.0
Total number of observations	242365	2680	17987
Total number unique	33659	446	2427
Mean [(I)/sd(I)]	18.9	49.7	2.6
Completeness (%)	99.6	99.1	99.5
Multiplicity	7.2	6.0	7.4
Wilson B-factor ($Å^2$)	57.97		

Table 4.1.12 Crystallographic data statistics for native pET14b-CobJ pull-down experiment, data collected to 2.58Å.
[a] Rmerge = $(\Sigma hkl\, \Sigma j \mid Ihkl - \langle Ihkl \rangle \mid) / (\Sigma hkl\, \Sigma j\, Ihkl,j)$
[b] R*meas* = $(\Sigma hkl \sqrt{(n/n-1)}\, \Sigma^n_{j=1} \mid Ihkl,j - \langle Ihkl \rangle \mid) / (\Sigma hkl\, \Sigma j\, Ihkl,j)$
[c] Rpim = $(\Sigma hkl \sqrt{(1/n-1)}\, \Sigma^n_{j=1} \mid Ihkl,j - \langle Ihkl \rangle \mid) / (\Sigma hkl\, \Sigma j\, Ihkl,j)$

4.4.8 Refinement

For all structures obtained refinement was carried out using REFMAC5 (Vagin 2004); see *Table* 4.1.13 for the Rfactor and Rfree values to which each structure was refined. Of the datasets collected the ones that were pursued in refinement processes included native CobJ and mutant CobJ H129A structure. A number of other structures that were co-crystallised with tetrapyrrole or small molecules were not refined due to the lack of bound molecules in the active site.

Native CobJ co-crystallised with PC-3a was not further refined because no density resembling PC-3a was found bound near the pocket nor was there any density to indicate non-specific binding. Therefore refinement was not continued especially because a 2.2 Å native CobJ model has been obtained already. The same goes for native CobJ co-crystallised with porphobilinogen since none was found at the active site, but found only bound non-specifically elsewhere in the protein. Refinement was not pursued since there were many poorly defined blobs of electron density.

The native CobJ structure at 2.2 Å resolution was deposited into the RCSB protein data bank under the entry 3NUT.

	Structure	**Rfactor (%)**	**Rfree (%)**	**Resolution (Å)**
1	Native CobJ	20.63	27.36	2.20
2	Native CobJ co-crystallised with PC-3a	31.27	40.08	2.75
3	Native CobJ co-crystallised with porphobilinogen	25.48	31.95	1.97
4	H129A co-crystallised with pyrrole-2-carboxylic acid	20.48	27.15	2.50
5	pET14b-CobJ from pull-down experiment	21.59	29.40	2.50

Table 4.1.13 Summary of Rfree and Rfactor values for various native and mutant CobJ structures.

4.4.9 Validation

Native CobJ and mutant H129A were the two structures further refined and validated. All other structures were not pursued since no tetrapyrrole was evident in the electron density map they were not further pursued.

To ensure a refined model is of a good quality it is important to validate the structure. As a model is refined Rfree, Rfactor and root mean square (rms) deviations of the model's bond lengths, angles and conformational angles are used to monitor the convergence to a final refined model. Generally rms deviations of no more than 0.02 Å for bond lengths and 4 ° for bond angles are accepted (Rhodes 2006). The two structures which were fully refined and validated were native CobJ and H129A mutant CobJ for the corresponding rmsd values see Table 4.1.14.

Structure	Rms length	Rms bond
Native CobJ	0.019	1.842
H129A CobJ	0.016	1.723

Table 4.1.14 The Rmsd length and bond values for native and mutant H129A CobJ structures.

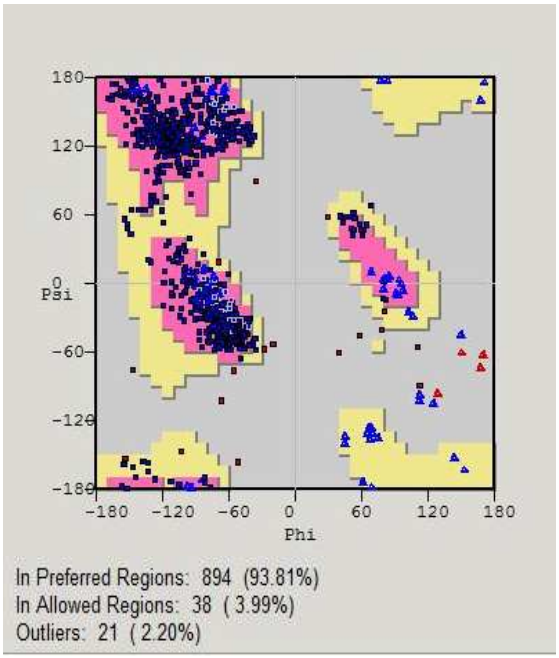
A Ramachandran plot allows visualisation of dihedral angles ψ against φ enabling quick detection of conformationally unrealistic regions of the structure. Residues with the most favourable ψ and φ angles are situated in core areas highlighted with pink colour. The higher the percentage of residues in these core areas, the higher the stereochemical quality of the model. For native CobJ and mutant CobJ H129A models the Ramachandran plots show both models have less then 3 % of residues in disallowed regions with the majority (more than 90 %) in preferred regions (see *Figure* 4.1.19).

Validation of models were carried out using PROCHECK V3.4.4 (Morris 1992; Laskowski 1993) from RCSB Protein Data Bank. Stereochemical and geometrical

features were calculated highlighting any close contacts with distances less than 2.2 Å, bond angles and distances, covalent bond lengths and angles, torsion angles and chirality were also scrutinized.

An overall summary of the validation for native CobJ and mutant CobJ H129A can be seen in *Figures* 4.1.20, 4.1.21, 4.1.22 and 4.1.23 showing main chain and side chain parameters where most residues are situated in favoured regions.

(a)



(b)

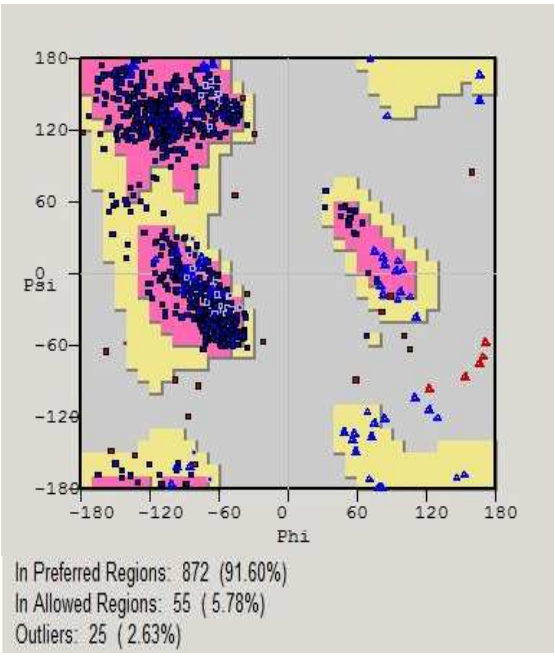


Figure 4.1.19 **(a)** Ramachandran plot for native CobJ model and **(b)** Ramachandran plot for mutant CobJ H129A model.

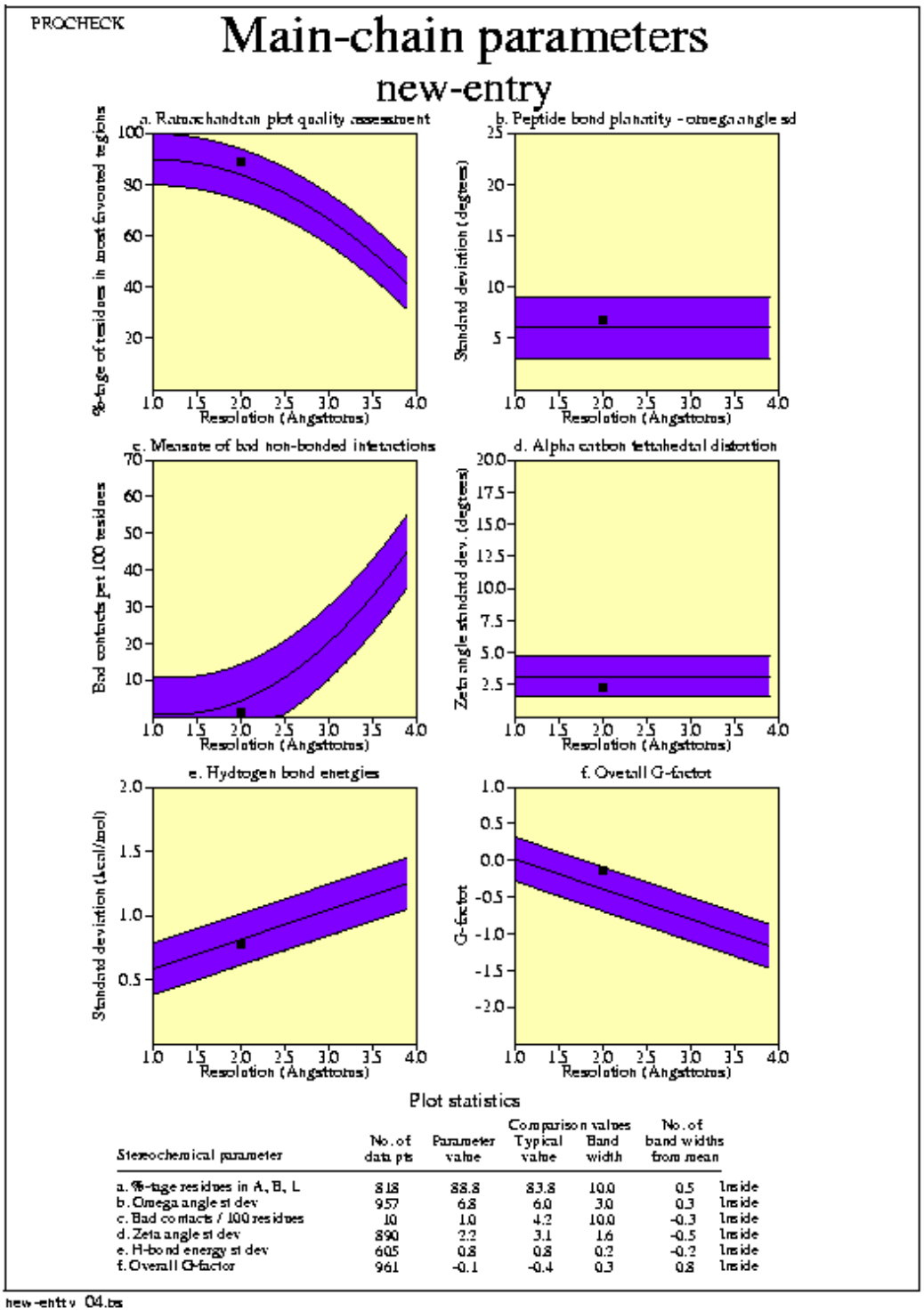


Figure 4.1.20 Main chain analysis of native CobJ model showing all criteria lie in the acceptable regions highlighted in purple.

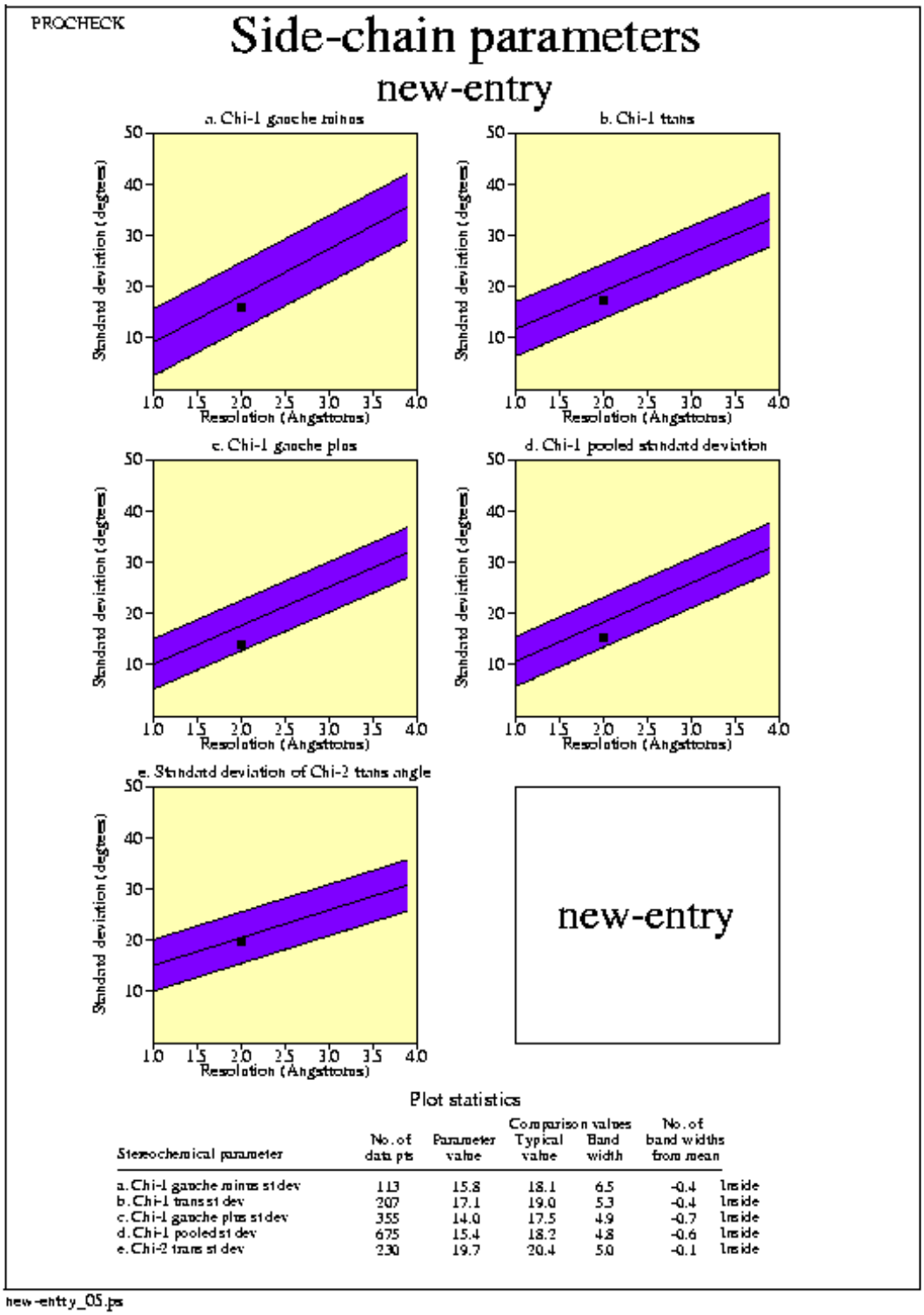


Figure 4.1.21 Side chain analysis of native CobJ model showing all criteria lie in the acceptable regions highlighted in purple.

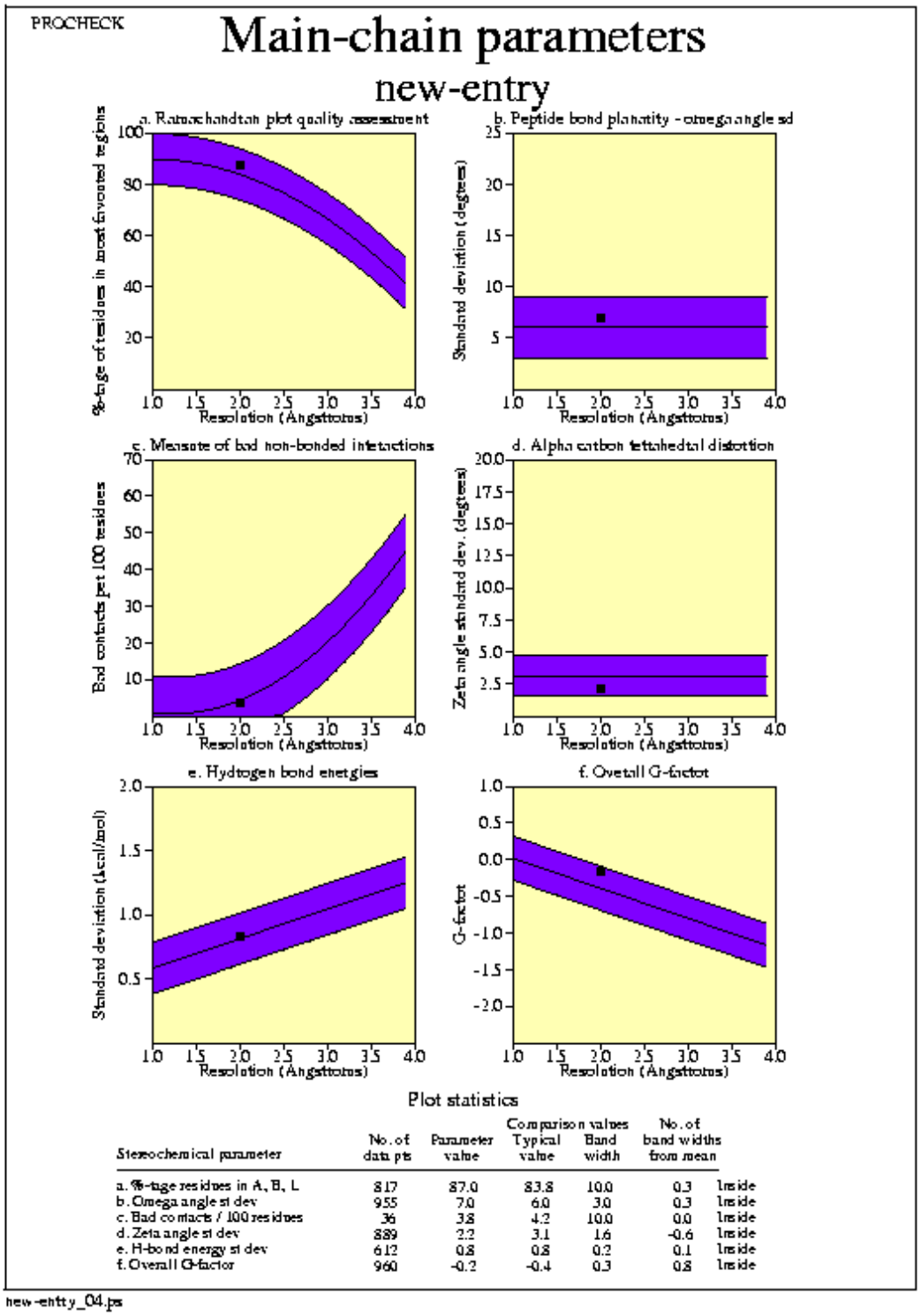


Figure 4.1.22 Main chain analysis of mutant CobJ H129A model showing all criteria lie in the acceptable regions highlighted in purple.

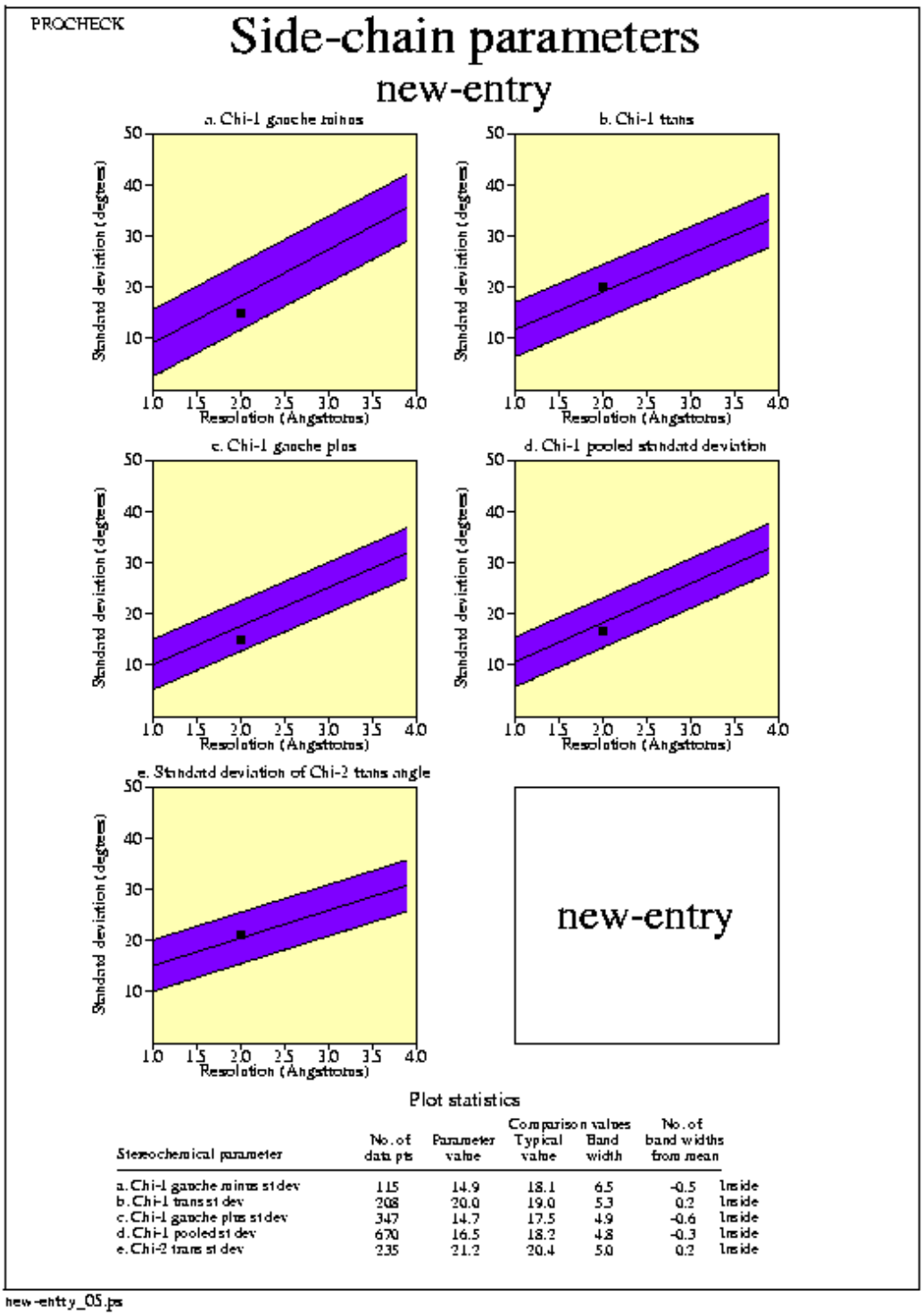


Figure 4.1.23 Side chain analysis of mutant CobJ H129A model showing all criteria lie in the acceptable regions highlighted in purple.

4.5 Analysis of CobJ active site

4.5.1 Alignment of CobJ sequences

In order to identify conserved residues across organisms expressing CobJ which posseses a C17 methyltransferase activity, a BLAST search was carried out using NCBI-BLAST2 from EBI server (Shpaer 1996). The protein-potein BLAST search using Swissprot database, identified organisms with similar sequences to *R. capsulatus* CobJ (see *Figure* 4.1.24).

There are as many as 32 conserved residues identified in the BLAST search and of the 32 residues some interesting ones that may play a role in catalysis include Tyr 35, Tyr 38, His129 and Asp138; these residues are highlighted in *Figure* 4.1.25 showing the positions of residues relative to SAH. All these residues have the ability to act as a base and accept a proton; these residues could therefore play a role in ring-contraction and methylation. They are also well positioned to do either of the reactions catalysed by CobJ.

Some other conserved residues identified are well positioned for SAH binding including Asp82, Gly84, Gly112 and Tyr163 or are structural residues such as Pro240 and Arg241 which are positioned far away (more than 15 Å) from the SAH and cleft region. Only the conserved residues with a possible role in catalysis are highlighted in *Figure* 4.1.25

```
A0Q339    MNN------KGKLYVIGIGPGSLDEMSIRAKKAIEESEIIVGYTKYIKLIEPLIEGKEIF   54
Q73Q21    MS---------KLFVVGIGPGGTEYMSAQAVEALKQSEIIVGYSGYIEYIKPFIEGKEVF   51
A5TUV4    MNN-------GKIYVVGIGPGNMEDISIRAYNILKNINVIAGYTTYVDLVKDEFPDKEFL   53
Q24Q37    MERGHGGHQQGKIYVVGIGPGDAEHMTQRAQSVWNEVDVVAGYKTYIDLIRPWLADKKVV   60
Q720N0    -----------MIYVIGIGPGDKRLMTGEALQAIEDAEVIVGYVTYIKLIKELIKDKEVV   49
Q4EKZ3    --------------------------MTGEALQAIEDAEVIVGYVTYIKLIKELIKDKEVV   35
Q4EV42    -----------MIYVIGIGPGDKRLMTGEALQAIEDAEVIVGYVTYIKLIKELIKDKEVV   49
A2SNT3    ---MRTTQAAGKIMLVGIGPGSVGHMTQRAREAIAEADVVVGYVTYIKLVADLIEGKEVV   57
Q6ARS2    MEKYNNGNGPGTLYVIGTGPGARDHIIPAAIEAIERADTIIGYSTYLDLIEELLEGKEVI   60
Q3J2J7    --------MSGWLVVAGLGPGAEHLVTPEVSVALAEATDVVGYIPYVARVAERP-GLTLH   51
Q9R9K1    --------MSGWVAVAGLGPGAEDLVTPEVRAALAEATDVVGYIPYVARVAPRP-GLALH   51
Rc-CobJ   --------MSGWVTVAGLGPGREDLVTPEVTAALAEATDIVGYIPYVARIAPRE-GLTLH   51
Q5LPI1    --------MSGWLRIVGLGPGSEALVTPEVTAVLEEATDVVGYIPYVARVAERP-GLTLH   51
Q89Q63    --------MTGTLTIAGLGPGSDALVTPEVSAALGSATDILGYAPYVARVPPRP-GLTLH   51
Q6N816    --------MTGSVVVAGLGPGAQELITPEVSAALALATDIVGYAPYVARVAERD-GLQRH   51
P21640    --------MTGTLYVVGTGPGSAKQMTPETAEAVAAAQEFYGYFPYLDRLNLRP-DQIRV   51
Q92LV0    --------MTGKLFIVGTGPGNPAQMTPEAEDAIAVATEFFGYGPYLDRLHLRA-DQRRI   51
Q2JZY5    --------MSGRLFVIGTGPGNPEQMTPEALAVVEAATDFFGYGPYLDRLQLRS-DQLRH   51
Q1M5A7    --------MSGRLFVIGTGPGNPEQMTPEALAAVDAATDFFGYGPYLDRLQLRH-DQLRH   51
                                  :   .          . **  *:  :     .

A0Q339    STGMRGELERVEYALNESK-SKTVSIISTGDAGIYGMAGPILEMATN--------EEVIV  105
Q73Q21    QTGMTGEIERCKYAVSKAKEGKTVSIISTGDAGLYGMAGPILELAPD--------LNVEI  103
A5TUV4    VSGMKREIERCREVLEVAKTGKDVALISSGDAGIYGMAGIMLEVAMES------GIEVEV  107
Q24Q37    ATGMRQEIDRCREVVEIALEGQSIALVSSGDAGVYGMAGILIECLEEKDAL---DISLEI  117
Q720N0    KTGMRREIDRCQEAVDIALTGKKVAVVSSGDAGIYGMAGLVLELAEKSNP----DLEVKV  105
Q4EKZ3    KTGMRREIDRCQEAVDIALTGKKVAVVSSGDAGIYGMAGLVLELAEKSNP----DLEVKV  91
Q4EV42    KTGMRREIDRCQEAVDIALTGKKVAVVSSGDAGIYGMAGLVLELAEKSNP----DLEVKV  105
A2SNT3    RKGMTEELDRAVSALEAARAGKKVALISSGDAGVYGMAGPTYEVLFQAGWTPEDAVQVEI  117
Q6ARS2    SSSMMKEVDRCRLSLEMAESGKSVALVSGGDAGIYAMAGLVLEMASQN----DFQAEIKI  116
Q3J2J7    ASDNRVEVERAAHALQMAAEGRRVVVVSSGDPGVFAMASALFEALEARPEWQ--ALDIRI  109
Q9R9K1    PSDNRVELERARLALDLAARGRRVVVVSSGDPGVFAMASALFEALEGCEN----TPDIRI  107
Rc-CobJ   PTDNRVELDRATHALEMAAEGRRVVVVSSGDPGVFAMASALFEALEAHPEHA--GTEIRI  109
Q5LPI1    ASDNRVEIDRSRHALELAADGHRVVVVSSGDPGVFAMASAVFEAVEAGPRAWR-DLDIQV  110
Q89Q63    PSDNREELARASEALRLAAEGGQVVVVSSGDPGVFAMASAVFEALEQAPQYR--ELPIRV  109
Q6N816    ASDNREELDRAGFALRLATEGRHVVIVSSGDPGVFAMAAALFEAIEAGDPSWR-ELDIRV  110
P21640    ASDNREELDRAQVALTRAAAGVKVCMVSGGDPGVFAMAAAVCEAIDKGPAEWK-SVELVI  110
Q92LV0    ASDNREELDRAHAALARAAAGADVCVVSGGDPGIFAMAAAVCEAIDKGPQEWR-EVDLTI  110
Q2JZY5    ASDNREELDRAGAALSMAADGANVCIVSGGDPGVFAMAAAVCEAIENGPAAWR-AVDLTV  110
Q1M5A7    ASDNREELDRAGAALAMAADGAKVCVVSGGDPGIFAMAAAVCEAIENGPAAWR-AVDLTI  110
           ..   *: *    :  :  .  : ::* **.*::.**.   *              : :

A0Q339    IPGITASSSAASLLGAPLMHDNCNISLSDLLTPYEVIKNRVECAAKGDFIISLYNPKSKG  165
Q73Q21    IPGISAAFAAASRLGAPLMHDTALISLSDRLTDYEVIKKRVGLAAEGDFVIALYNPKSKT  163
A5TUV4    VPGITSTIAGAALVGAPLMHDQAIISLSDLLTDWEVIKKRIDCASQGDFVISLYNPKSKG  167
Q24Q37    IPGVSAANAASSLLGAPLMHDFAVISLSDLLTPWEVIQKRVKLAAEGDFVMAIYNPKSKG  177
Q720N0    IPGITASIGAAAVLGAPIMHDFCHISLSDLMTPWEVIEKRLTHAAMADFVVCFYNPRSKG  165
Q4EKZ3    IPGITASIGAAAVLGAPIMHDFCHISLSDLMTPWEVIEKRLTHAAMADFVVCFYNPRSKG  151
Q4EV42    IPGITASIGAAAVLGAPIMHDFCHISLSDLMTPWEVIEKRLTHAAMADFVVCFYNPRSKG  165
A2SNT3    VPGASALNACAALVGAPLTHDFCAISLSDLLTPWPVIARRLDAVAMADFVVALYNPKSGR  177
Q6ARS2    IPGIAAVNACAARLGAPLMHDFAAISLSDLLTPWETIVARLEATASTDFVVALYNPKSKR  176
Q3J2J7    LPGITAMLAAAAAAGAPLGHDFCAINLSDNLKPWALIEKRLRLAAEADLAMAFYNPRSKS  169
Q9R9K1    LPGITAMLAASARLGAPLGHDFCAINLSDNLKPWALIEKRLRLAAEADFAMAFYNPRSKA  167
Rc-CobJ   LPGITAMLAAAAAAGAPLGHDFCAINLSDNLKPFEILEKRLRHAARGDFAMAFYNPRSKS  169
Q5LPI1    LPGITAMLAAAARAGAPLGHDFCCINLSDNLKPWPLIERRLRLAAQADFAMAFYNPRSKS  170
Q89Q63    LPGITAMLAAAARAGAPLGHDFCAINLSDNLKPWALIEKRLRLAAEADFAIALYNPRSAS  169
Q6N816    LPGISAMFATAARIGAPLGHDFCAINLSDNLKPWETVEKRLRAAAEADFVIALYNPISKA  170
P21640    TPGVTAMLAVAARIGAPLGHDFCAISLSDNLKPWEVITRRLRLAAEAGFVIALYNPISKA  170
Q92LV0    TPGVTAMLAVAARIGAPLGHDFCAMSLSDNLKPWDVITRRLRLAAEAGLVIALYNPISKA  170
Q2JZY5    LPGVTAMLAVAARAGAPLGHDFCAISLSDNLKPWNVIENRLELAARAGFVMALYNPISRA  170
Q1M5A7    LPGITAMLAVAARAGAPLGHDFCAISLSDNLKPWNIIETRLVLAAKAGFVIALYNPISRA  170
           ** ::  . ::  ***: ** . :.*** :. :  :  *:  .:  .: :.:*** *
```

```
A0Q339    RPHYLKECLNIIKKYRNDNTPIGVVKNALREGQEITVTTIGNFNDDIVDMMSIVVVGNSK  225
Q73Q21    RSDYIEEAVNIILKFRAPQTPVGIVKNACRNNEKITVTELQKINYDEIDMFSLIIIGNSN  223
A5TUV4    RIEQIVEAREIMLKHKLPTTPVALLRHIGRKEENYTLTTLEDFLNFEIDMFTIVLVGNSN  227
Q24Q37    RQDQIEWVRQAVLDYRDPQTPVGIVREALRGESQVIITTLAEFTQSPIDMLTTVVIGNSH  237
Q720N0    RANHLANAFQKMMEHKSGDTVVGIVKDVGRKEERKIITTMRDIDYELVDMTTMVIVGNKE  225
Q4EKZ3    RANHLANAFQKMMEHKSGDTVVGIVKDVGRKEERKIITTMRDIDYELVDMTTMVIVGNKE  211
Q4EV42    RANHLANAFQKMMEHKSGDTVVGIVKDVGRKEERKIITTMQGIDYELVDMTTMVIVGNKE  225
A2SNT3    RTRQIVEAQRLFLRHRRADTPVAIVKSAYRRRQHIEFRTLETMCEADIGMLSTVLIGNSH  237
Q6ARS2    RTAQIAQAREIFLKHRDPANVVGIVTGATRENEVITLTTLDKMLECEIGMQSTVIIGNSK  236
Q3J2J7    RPEGFVRALDLLRETCGPDRLVTFARAVSTPEQHLRTVTLGEATPEMADMRTVVIVGNSA  229
Q9R9K1    RPEGFARVLEILREACGPDRLISFARAVSTPDEALNTVTLAEARPEMADMRTVVILGNSA  227
Rc-CobJ   RPHQFTRVLEILREECEPGRLILFARAVTTPEQAISVVELRDATPEMADMRTVVLVGNAA  229
Q5LPI1    RPEGFEKTLQILREECEPERLILFARAVSRPDEALRIAQLAEATPEMADMQTVVLVGSSR  230
Q89Q63    RPEGFGRALAVLKDAGCGERLVIFARAISASDETIETVTLNAARPEMADMRTLVIVGNSQ  229
Q6N816    RPWQLGRAFELLRSIHPASVPVIFATAISDPRERIDVAPLGEAVPQRADMRTLVMIGSSQ  230
P21640    RPWQLGEAFELLRSVLPASVPVIFGRAAGRPDERIAVMPLGEADANRADMATCVIIGSPE  230
Q92LV0    RPWQLGEAFEVLRQVLPAQVPVIFGRAAGRPDERIAVMPLGEADAGRADMATCVIIGSPE  230
Q2JZY5    RPWQLGEAFKLLRHHLPATTPVIFGRAAGRVNEHIAVQPLSQADASIADMATCIIIGSAE  230
Q1M5A7    RPWQLGEAFKLLRDHLPAATPVIFGRAAGRPDERIAVQQLSQADASIADMATCIIIGSAE  230
          *   :       .          : .         .       :         .* : :::*.

A0Q339    SYIKN-----NKFITPRGYENKAKGEVK--------------------------------  248
Q73Q21    TYIQN-----GKIITPRGYKIK----------------------------------------  240
A5TUV4    TYVKD-----GKMITPRGYEKKSNWGK--------------------------------  249
Q24Q37    TRVIG-----PYMVTPRGYIL--------------------------------------  253
Q720N0    TYVKN-----GKMITPRGYTL--------------------------------------  241
Q4EKZ3    TYVKN-----GKMITPRGYTL--------------------------------------  227
Q4EV42    TYVKN-----GKMITPRGYTL--------------------------------------  241
A2SNT3    TFVRD-----GLMVTPRGYANKYDIEQGGETREGERPGRSLSTGLNGWLENLLDDHAGGE  292
Q6ARS2    TFVWR-----DKMITPRGYGEKYQL----------------------------------  256
Q3J2J7    TRRVG-----RWVYTPRSAG---------------------------------------  244
Q9R9K1    TRRAG-----AWVYTPRSVP---------------------------------------  242
Rc-CobJ   TRRVG-----PWVYTPRGVAP--------------------------------------  245
Q5LPI1    TRVIERD-GTPIVYTPRFTPE--------------------------------------  250
Q89Q63    TRRVG-----RWIYAPRQVR---------------------------------------  244
Q6N816    TRIIPRSSGADFVYTPRFSGAVS------------------------------------  253
P21640    TRIVERDGQPDLVYTPRFYAGASQ-----------------------------------  254
Q92LV0    TRIIAREGKHDLVYTPRSFTGESH-----------------------------------  254
Q2JZY5    TRIVTRPGKPDLVYTPRFMAGGKR-----------------------------------  254
Q1M5A7    TRIVSRPGRLDLVYTPRFMAGGNR-----------------------------------  254
          :           . :**
```

Figure 4.1.24 BLAST search of *R.capsulatus* CobJ revealing conserved residues across different organisms. A0Q339: *Clostridium novyi,* Q73Q21: *Treponema denticola*, A5TUV4: *Fusobacterium nucleatum*, Q24Q37: *Desulfitobacterium psychrophila*, Q720N0: *Listeria monocytogenes*, Q4EKZ3: *Listeria monocytogenes* Str 4b, Q4EV42: *Listeria monocytogenes* Str 1/2a, A2SNT3: *Methylibium petroleiphilum*, Q6ARS2: *Desulfotalea psychrophila*, Q3J2J7: *Rhodobacter sphaeroides*, Q9R9K1: *Paracoccus denitrificans*, Q5LPI1: *Silicibacter pomeroyi*, Q89Q63: *Bradyrhizobium japonicum*, Q6N816: *Rhodopseudomonas palastris*, P21640: *Pseudomonas dentrificans*, Q92LV0: *Rhizobium meliloti*, Q2JZY5: *Rhizobium etli*, Q1M5A7: *Rhizobium leguminosarum*, Rc-CobJ: *Rhodobacter capsulatus*. Conserved residues which may play a role in catalysis are highlighted in yellow.

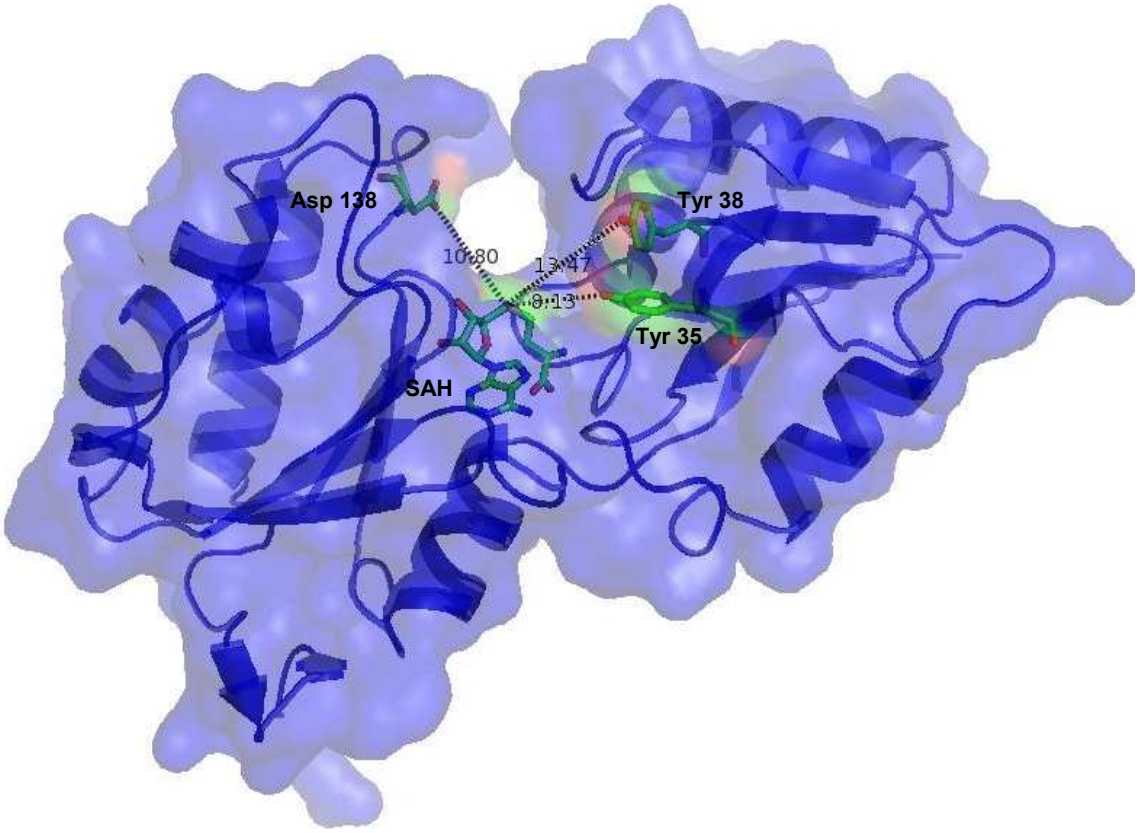


Figure 4.1.25 Conserved residues Tyr 35, Tyr 38 and Asp 138 highlighted in chain A of CobJ at 8.1 Å, 13.5 Å and 10.8 Å respectively from SAH.

4.5.2 Superimposition of four molecules in CobJ structure

Since native CobJ crystallises with four molecules in the asymmetric unit a simple superimposition of chains B, C and D onto chain A separately were carried out using SUPERPOSE (Krissinel 2004) to analyse any structural differences between chains (see *Figure* 4.1.26).

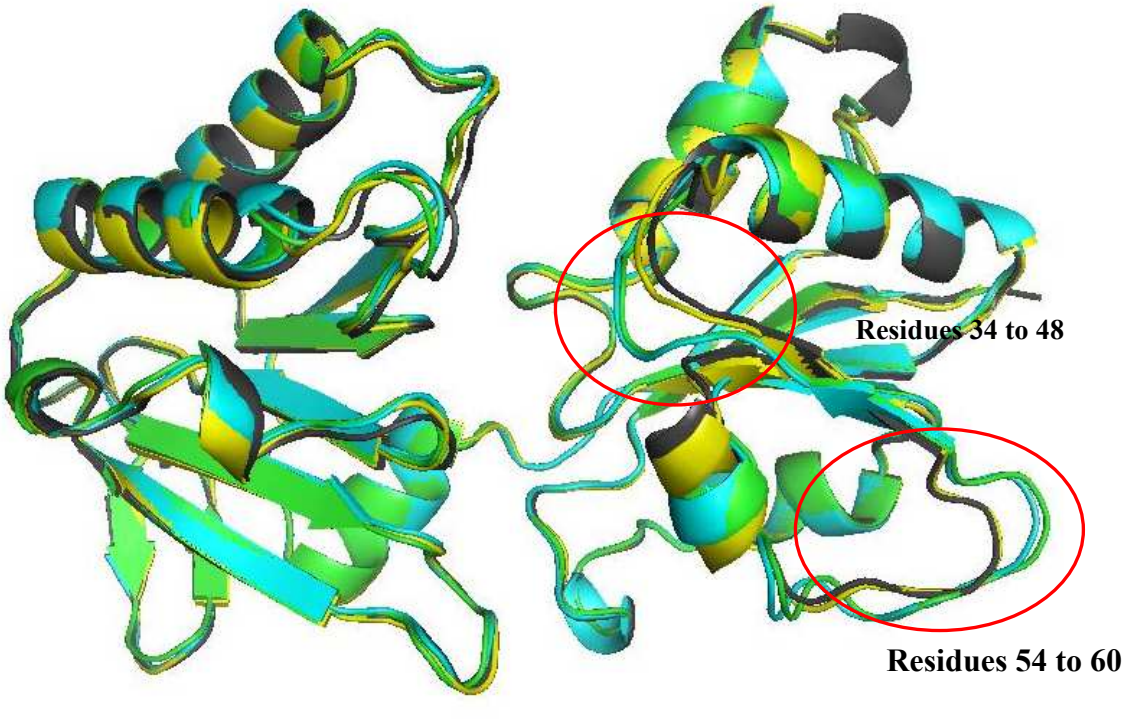


Figure 4.1.26 Ribbon diagram depicting the main chain (Cα trace) of chains A (grey), B (yellow), C (blue) and D (green), from native CobJ structure superimposed onto each other with areas of change highlighted.

The superimposition revealed similarities structurally between chains A and B and similarities between chains C and D. Since chains A/D and B/C form dimers we can say native CobJ assembles as a homodimer, with two dissimilar chains forming a single dimer (see *Figure* 4.1.27). From the superimposition, the residues 34 to 48 and 54 to 60, formed loop regions, which carry different conformations across the four chains. The

loop consisting of residues 34 to 48 is close to the cleft pocket. In this loop conserved residues including Gly 34, Tyr35 and Tyr38, are present and these are likely to be important since they are conserved and are situated near the catalytic site where they may play a role in the auxillary reaction. The fact that they are in a loop showing structural differences may be explained by the possibility of flexible residues allowing the uptake of a tetrapyrrole and assisting its binding and then releasing the tetrapyrrole once ring-contraction and methylation has taken place.

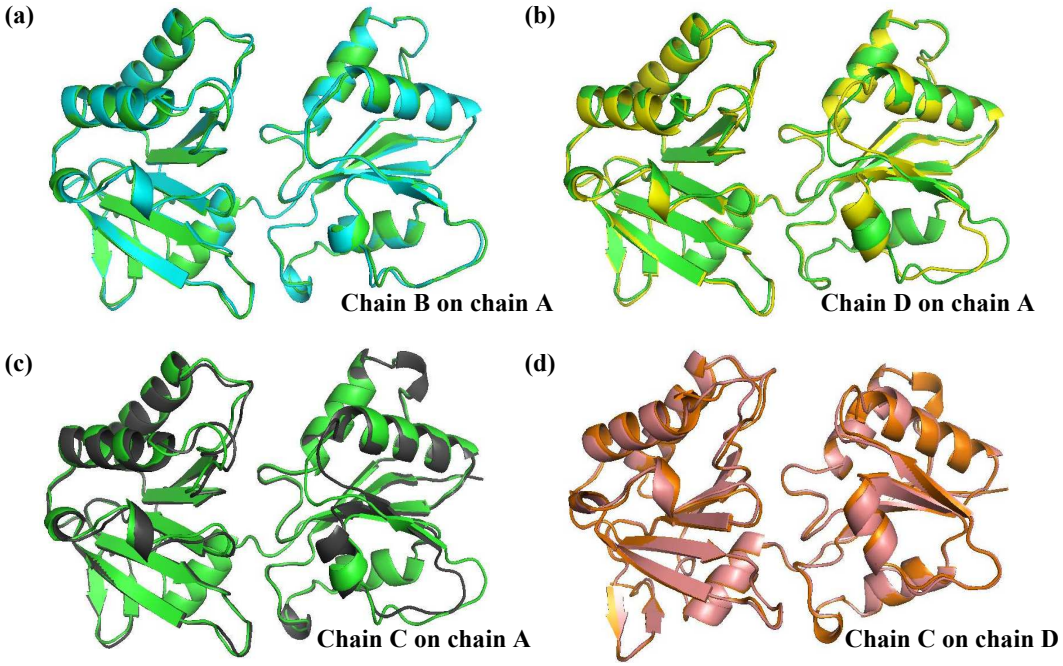


Figure 4.1.27 Native CobJ molecule. **(a)** Chains B and A have no major structural differences. **(b)** Chain D and A superimposed onto each other show loop regions that differ structurally; these chains form a dimer. **(c)** Chain C and A differ from each other in some loop regions. **(d)** With chain C superimposed onto chain D, no differences in the structures can be seen.

B factor analysis shows regions of high flexibility (*Figure* 4.1.28) with most flexible regions in red and least flexible in blue, it would seem the C terminal is more flexible than N terminal region.

From *Figure* 4.1.26 we can see the loop regions from residue 34 to 48 and residues 54 to 60 as identified earlier from superimposition of chains also correspond to areas of high flexibility as seen in the B factor analysis. Closer inspection shows the residues surrounding the cleft of the catalytic site are allowed greater movement compared to other residues in the structure, most probably to accommodate a tetrapyrrole by widening the cleft area possibly.

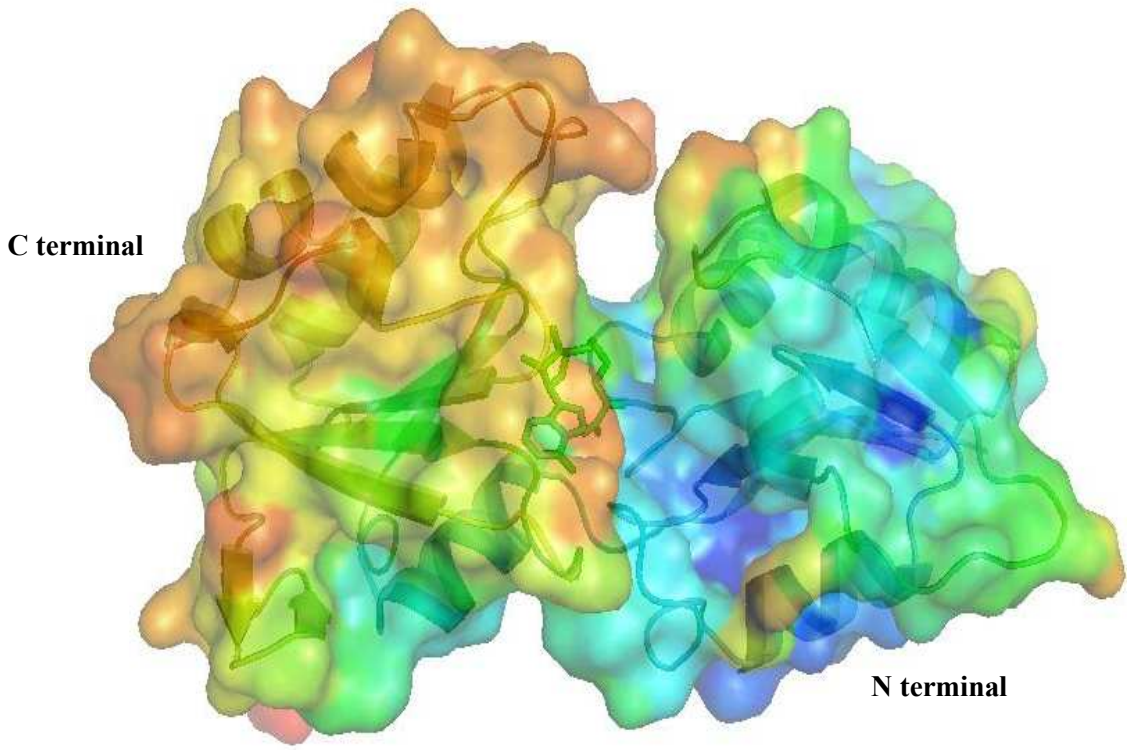


Figure 4.1.28 B factor analysis shows the surface of native CobJ structure chain A with residues highlighted in red for highly flexible regions and areas of lesser flexibility highlighted in blue colour.the figure was created using Pymol.

4.5.3 Proposed ring contraction mechanism

The precise mechanism of the ring-contraction reaction is unknown. It is unclear whether ring-contraction occurs before or after methylation or indeed if these events are concerted.

A proposed mechanism for the CobJ ring contraction step is a pinacol type arrangement (see *Figure* 4.1.29). In preparation for the reaction CobG acts as a monooxygenase and adds a hydroxyl group to C20 forming precorrin-3b, hence setting up the macrocycle for ring contraction and methylation at C17 by CobJ (Scott 1993).

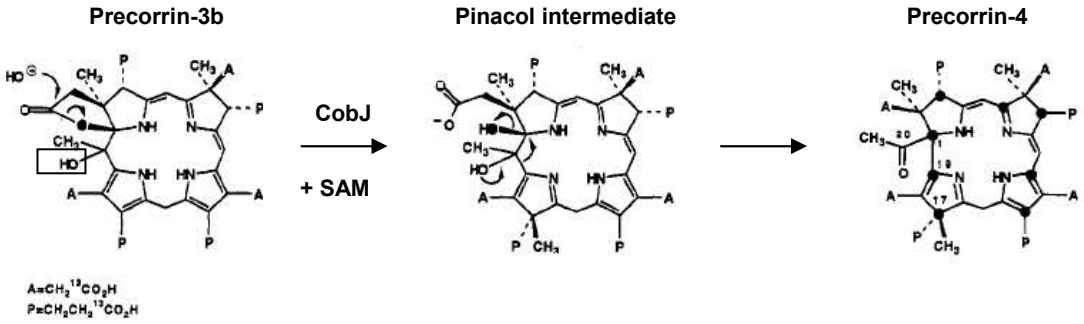


Figure 4.1.29 Schematic diagram depicting the proposed mechanism for the ring-contraction and methylation step catalysed by CobJ (Scott 1993). The boxed OH identifies the H which is abstracted later on in the mechanism.

Two other possible mechanisms could be employed to carry out the ring contraction of the macrocycle and methyation of the ring including Scheme I where methylation occurs after ring contraction and Scheme II where methylation occurs before ring contraction see *Figure* 4.1.30 and *Figure* 4.1.31. The proposed mechanisms require a base to facilitate the proton abstraction from the hydroxyl group and a second base to catalyse the methylation at C17, similar to Scotts mechanism (see *Figure* 4.1.29).

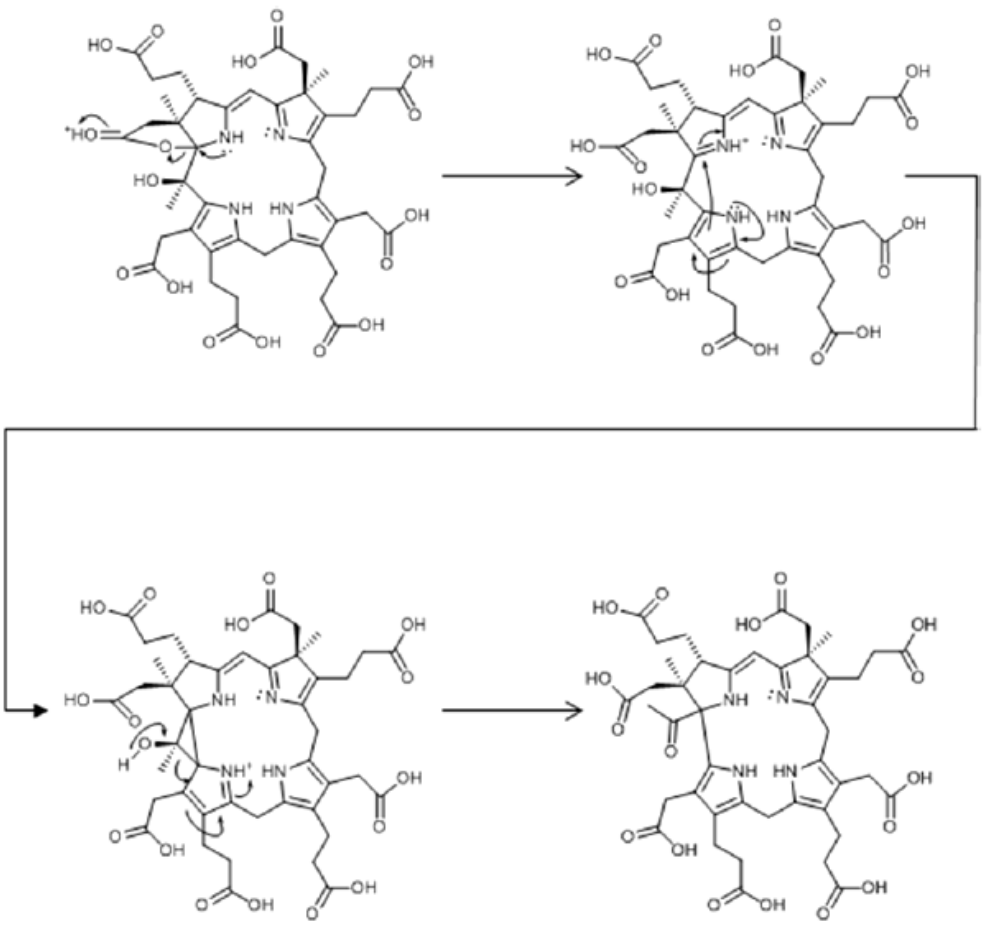


Figure 4.1.30 Scheme I, a possible reaction mechanism for ring contraction in the absence of C17 methylation.

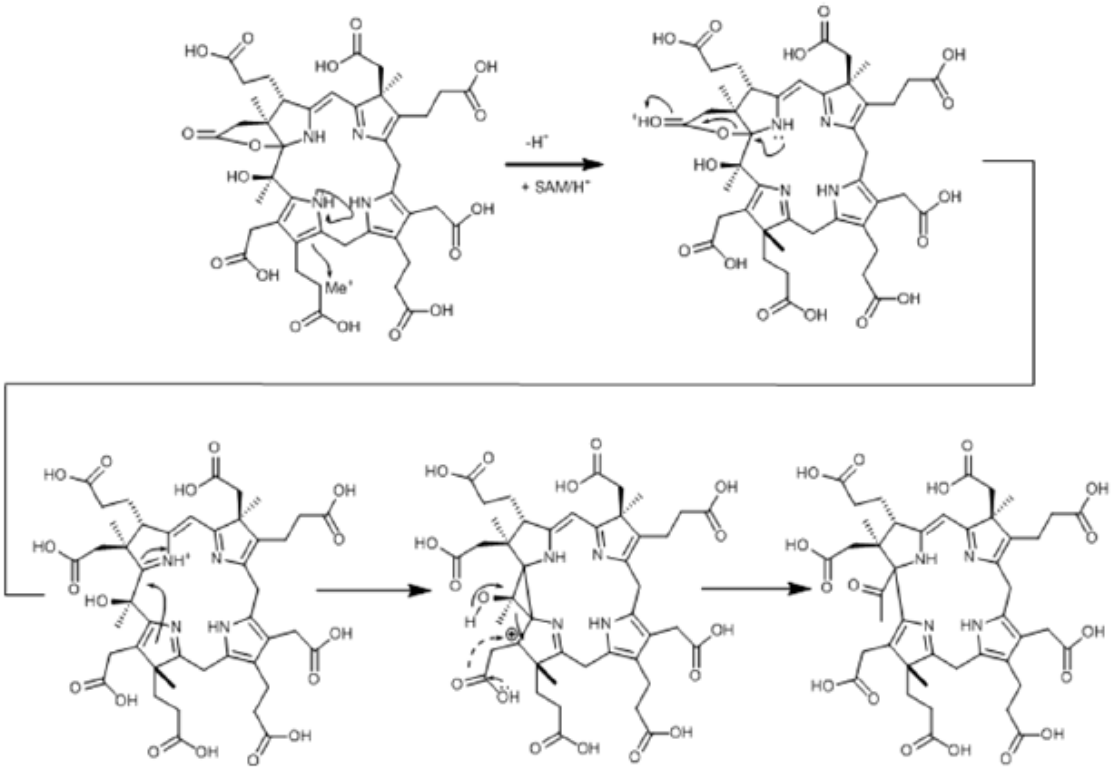


Figure 4.1.31 Scheme II depicts a possible ring contraction mechanism with the presence of C17 methylation.

4.6 Conclusion: Structural implications of CobJ

From the work carried out in this chapter we can establish that a higher resolution structure of native CobJ at 2.2 Å has been obtained improving from a previous 2.7 Å structure. The structure reveals residues of interest lining the catalytic pocket that have been highlighted using searches of the swissprot database using BLAST which identified important conserved residues. These residues include Tyr35, Tyr38, His129 and Asp138, which are found near the catalytic pocket and have been recognised as potentially important for the methylation step.

Tyr38 is 5.47 Å away from the sulphur group of SAH and could be involved in methylation at C17 of the macrocycle since it is appropriately positioned for facilitating this reaction as a proton acceptor. Possible candidates for assisting the ring contraction reaction are Tyr35 and Asp138, which are 13.47 Å and 10.80 Å away from the sulphur group of SAH respectively, this may be too far to promote methylation at C17, but could assist ring contraction.

Superimposition experiments showed Tyr35 and Tyr38 are both found on a loop that carries some flexibility which could be involved in proton abstraction. They also sit in a region where the highest temperature factors are found, suggesting that there could be structural flexibility. Since the opening of the cavity is small the ability for this loop region to change conformation in order to accommodate incoming tetrapyrroles, in the case of CobJ, precorrin-3b, would be ideal. This is indicative of an induced fit mechanism.

His129 may be the best candidate for the residue that catalyses ring contraction, it is more distant from the SAH site and mutation of it has a profound effect on activity. It is plausible that residues close to the SAM are involved or provide the framework for SAM binding and methylation e.g. mutation of Asp82 will disrupt SAM binding. On the other hand, His129 is over 12 Å from the SAM and may be involved in the ring

contraction process. It would be interesting to see what the products of the reaction of these variants with precorrin-3b would be using mass-spectrometry and NMR analysis.

4.7 Future work

In order to help elucidate the mechanism of the ring contraction step catalysed by CobJ it would be immensely useful to obtain an enzyme-substrate or -product complex with its natural tetrapyrrole, precorrin-3b or precorrin-4 bound. This would reveal the exact residues involved in tetrapyrrole binding, and residues involved in methylation and ring contraction reactions.

Two mutants appear to be inactive for different reasons, Asp82 (SAM binding) and His129 (catalysis), it would be interesting to characterise the products of the reaction of these with precorrin-3b. If His129 is involved in the ring contraction reaction it may be possible to confirm the order of the mechanism and the role of His129 in terms of whether methylation occurs before or after ring contraction.

Recently, it has been proved possible to make product complexes in several of the biosynthetic enzymes using the cloned biosynthetic pathway in *E.coli* up to and including the enzyme to be studied e.g. CobH. If CobH is His-tagged then it can be pulled out and crystallized along with its bound product in this case, HBA. This method has been used to obtain the structure of the CobH/HBA complex and has been applied to CobJ however further investigation is needed. CobJ apparently also binds product, but it is not yet clear if the binding will be sufficiently tight to allow the structure of the complex to be resolved.

Chapter Five:
Trapping the Michaelis complex in pectate lyase

5.1 Introduction- Design of mutants

Plant pathogens produce a number of enzymes capable of initiating soft-rot disease in plants and fruits by breaking down polygalacturonate which is a major plant cell wall component. These enzymes randomly cleave the α-1,4-glycosidic linkages between galacturonate residues using an anti β-elimination chemistry generating products with a 4,5-unsaturated galacturonosyl residue at the non-reducing end (Collmer 1986) see *Figure* 5.1.1. Pectate lyase, pectin lyase and exopolygalacturonate lyase all utilise such reaction chemistry, in this Thesis the main focus is on pectate lyase (BsPel) from the bacterium *Bacillus subtilis*.

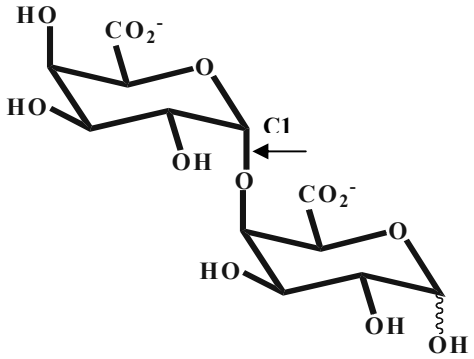


Figure 5.1.1 Two galacturosyl sugars depicted with one glycosidic linkage between the two. Pectate lyase cleaves at α-1,4-glycosidic bonds as highlighted by the arrow.

Pectate lyases play a pivotal role in recycling and remodelling of plant material which is important for maintaining the biosphere (Scavetta 1999). There are 22 families of polysaccharide lyases and pectate lyase belongs to five of the 22 including PL-1, PL-2, PL-3, PL-9 and PL-10. The allocation of families is based on the amino acid sequence similarities reflecting their structural features (Cantarel 2009). From the three dimensional structures reported there have been three types of topologies including the parallel β-helix, $(\alpha/\alpha)_7$ barrel and the $(\alpha/\alpha)_3$ barrel, a common feature in polysaccharide lyases.

Pectate lyase activity was first discovered in 1962 by Starr and Morán in *Erwinia carotovora* (Starr 1962) and since then the three dimensional structures of pectate lyases PelC, PelE from *Erwinia chrysanthemi* and BsPel from *B. subtilis* have been solved (Yoder 1993; Lietzke 1994; Pickersgill 1994). Each of the structures exhibited a parallel β-helix with β-sheets folded into a large right handed helix with calcium bound, all belonging to PL-1 family.

A complex between mutant R218K pectate lyase C (PelC) and pentagalacturonate sugar substrate has been solved to a resolution of 2.2 Å from *Erwinia chrysanthemi* (Scavetta 1999). The substrate binds in a cleft interacting with positively charged groups arginine and lysine. The structure represents the intermediate in the reaction pathway since PelC-R218K is inactive and therefore reveals some important mechanistic details of the β-elimination reaction such that an arginine is responsible for proton abstraction at C5 of pentagalacturonate during the cleavage of α-1,4-glycosidic bonds (Scavetta 1999). The converging geometry of the active site between families PL-1 and PL-10 family lyase structures suggests that the mechanism involving the use of an arginine as a base is common.

An inactive mutant D389A pectate lyase (Pel10Acm) from *Cellvibrio japonicas* in complex with a trisaccharide sugar substrate has been solved to 2.15 Å. This enzyme belongs to family PL-10, although the structure is unrelated to polygalacturonate lyases from family PL-1 there are similaritites in the catalytic machinery used by the two families, this may be due to the effect of convergent evolution. The residue Arg524 is well positioned at 2.5 Å away from C5 of the galacturonic residue. On comparison with the structure of PL-1 family lyase PelC1 complexed to tetragalacturonate substrate, both structures have different topologies, however in both cases the substrate carboxylates are coordinated to the putative -1 and +1 subsites with a Ca^{2+} ion bound. Surrounding the metal ion are ligands that are conserved. Along with these observations is the close vicinity of Arg254 in Pel10Acm in place to act as a catalytic base in the same location as

the putative arginine base in PL-1 enzymes, (Charnock 2002). Therefore this mechanism seems like a plausible assumption.

The structure of BsPel from *B. subtilis* in complex with calcium at position one (see Figure 5.1.2 for the positioning of calcium ions) was solved to a resolution of 1.80 Å (Pickersgill 1994). The structure of the lyase consisted of a parallel β-helix domain and a loop region forming the catalytic site. A calcium ion was bound to three aspartate residues including Asp184, Asp223 and Asp227. In total there were seven ligands making contacts between the calcium and BsPel including one carboxy oxygen from Asp223 and Asp227, both carboxy oxygens from Asp184 and three oxygens from three clearly defined water molecules.

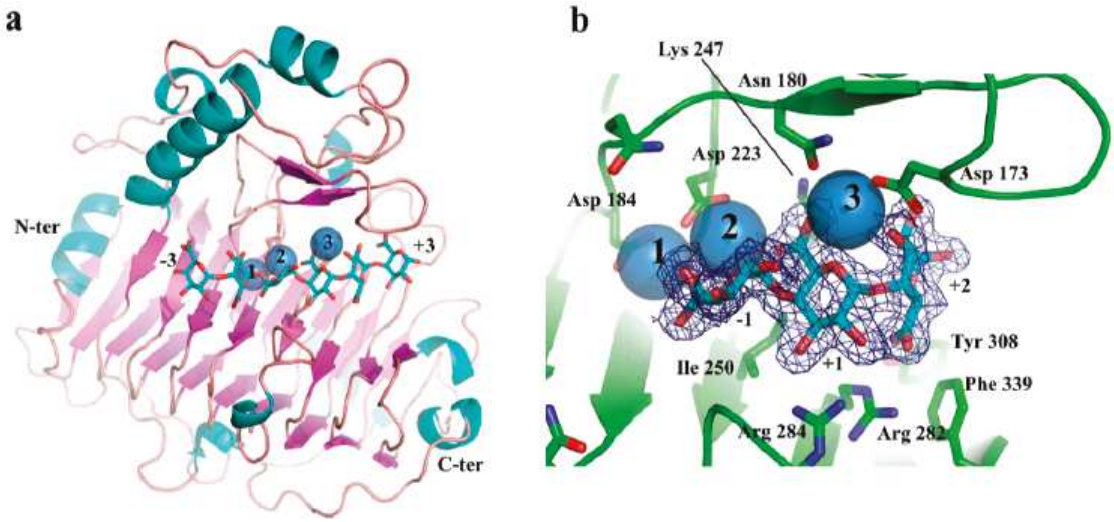


Figure 5.1.2 (a) Cartoon representation of BsPel showing bound sugar substrate and three calciums (blue spheres labeled 1, 2 and 3) commonly seen bound to BsPel. (b) Closer look at the contacts the calcium ions make with the protein and sugar substrate (Seyedarabi 2010b).

Pectate lyase (BsPel) from *B. subtilis* is the subject of this results chapter, the aim was to trap the Michaelis complex in an inactive enzyme in order to reveal the mechanistic details of the *anti* β-elimination reaction cleaving the α-1,4-glycosidic bond. In order to gain an enzyme-substrate complex, an inactive enzyme was needed and site-directed mutagenesis experiments mutating the possible catalytic base Arg279 to an Ala279

proved to provide an inactive enzyme with no turnover of sugar substrate (Seyedarabi 2010b). However in order to confirm the role of Arg279 the residue should be present in the structure.

In this chapter the residues that have been subjected to site-directed mutagenesis include Asn180 and Asp173. These residues are positioned to bind the third calcium binding site important for catalysis. Lys247 is thought to protonate substrate and stabilise the intermediate therefore this was also a candidate for mutation to an alanine. Site-directed mutagenesis of these three residues should provide an inactive enzyme to allow for successful binding and trapping of sugar substrate. In effect the third calcium binding site is removed.

5.2 Site-directed mutagenesis

In this chapter site-directed mutagenesis was used to probe the activity of BsPel by mutating residues that are involved in catalysis, once these residues were mutated crystallisation of the protein was carried out followed by substrate soaking. In order to trap an enzyme-substrate complex an inactive enzyme must be formed first. Site-directed mutagenesis is a molecular technique where a point mutation is introduced to the DNA via pre-designed primers.

5.2.1 Primer design

Primers were designed as detailed in section 2.2.1.1, the target mutation for BsPel was a lysine residue at position 247 mutated to an alanine, details of the primer can be found in *Table* 5.1.1.

	Primer pair	
3’	GATAATAGTGCTAGTACTACGTTCGAGGTAAAAGCCTAG	5’
5’	CTATTATCACGATCATGATGCAAGCTCCATTTTCGGATC	3’

Table 5.1.1 Primers used for site-directed mutagenesis mutating lysine to an alanine at position 247, the bases that were changed to allow for an alanine mutation are highlighted in grey.

5.2.2 PCR based site-directed mutagenesis

BsPel was cloned into pET3d vector by Dr Katherine Worboys and the plasmid which was used for site-directed mutagenesis was already mutated with the mutations D173A and N180A courtesy of Salyha Ali. To synthesise a triple mutant site-directed mutagenesis was carried out using PfuTurbo polymerase, the general cycling parameters for this polymerase and reaction set up can be found in section 2.2.1.2 and *Table* 2.10. For specific cycling parameters used in the formation of K247A mutation (see *Table* 5.1.2) and PCR reaction set up see below.

Step	Temperature (°C)	Duration
1. Activation	95	30 sec
2. Denaturation	95	30 sec
3. Annealing	55	1 min
4. Elongation	68	6 min
Steps 2 to 4 cycled 16 times.		

Table 5.1.2 Specific cycling parameters used for site-directed mutagenesis of BsPel.

PCR reaction set up consisted of:

2 µl DNA template (already with mutations D173A and N180A)

5 µl 10× reaction buffer

1.5 µl forward K247A primer

1.5 µl reverse K247A primer

1 µl dNTP

38 µl ddH_2O

1 µl Pfu Turbo polymerase

Once the PCR reaction was completed the product was DpnI treated (see section 2.2.2) and then checked on a 1 % agarose gel followed by plasmid preparation (see *Figure* 5.1.3) the subsequent plasmids (expected size 5.8 kb) were ready for sequence analysis.

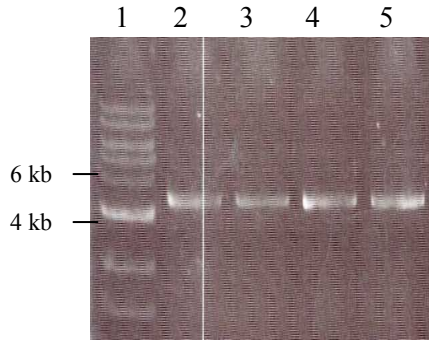


Figure 5.1.3 Plasmids sent for sequencing were first checked on an agarose gel to ensure DNA is present and PCR reaction had been successful. Lane **(1)** 1 kb ladder from novagen. Lanes **(2-4)** are plasmids from site-directed mutagenesis PCR reaction.

5.2.3 Sequencing

Plasmids were sent off to MWG Biotech for sequencing to ensure the site-directed mutagenesis was successful. The results showed lysine at position 247 was mutated to alanine, see Appendix 7 for alignment.

5.3 Expression and purification of mutant protein

5.3.1 Expression of triple mutant BsPel

Triple mutant (D173A, N180A and K247A) pET3d-BsPel was transformed into XL-1 Blue cells for cloning, for protein expression BL21 (DE3) pLysS cells were used, for transformation protocol see section 2.2.3 and *Table* 2.11. For protein over-production, pET3d-BsPel cells were grown at 37 °C until an $O.D_{600}$ of 0.6 was reached, followed by induction at 30°C overnight with 0.1 mM IPTG. Harvested cells were resuspended in 20 mM MES pH 6 and lysed using a French press (see section 2.2.4.2).

5.3.2 Purification of triple mutant BsPel

Once cells had been lysed and centrifuged to remove cell debris the lysate was loaded onto a cation exchanger HiTrap SpHp column. Buffers for this particular column can be found in section 2.1.6, *Table* 2.4. A clean peak was eluted from the cation exchanger and the protein was pooled and checked on an SDS-PAGE gel (see *Figure* 5.1.4 and 5.1.5). Fractions were subsequently dialysed into 20 mM Tris pH 7.0 and 100 mM NaCl. Typically approximately 28 mg/ml of protein was produced from one litre of cell culture.

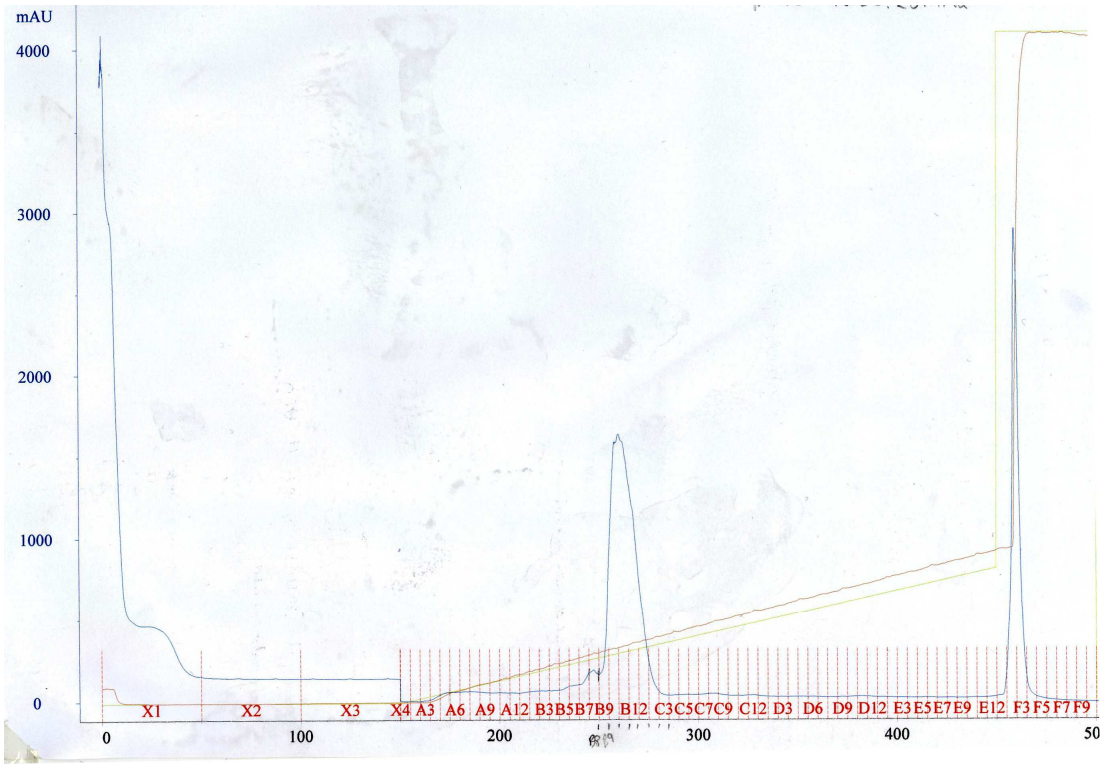


Figure 5.1.4 Protein was eluted on a gradient elution from a cation exchanger.

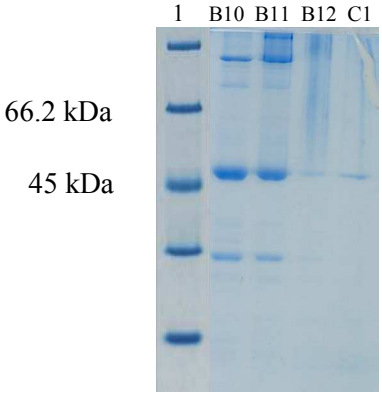


Figure 5.1.5 SDS-PAGE gel of pET3d-BsPel after purification on a cation exchanger, bands at around 45 kDa indicate the presence of BsPel protein. Fractions B10, B11, B12 and C1 were checked on the SDS-PAGE gel.

5.4 Spectroscopic enzyme assay

The aim of this project was to trap substrate in the enzyme allowing examination of the structure with the supposed catalytic base Arg 279 in place. In the past a number of mutagenesis experiments have been carried out on BsPel in an attempt to produce an inactive enzyme, these mutants and its relative activities are summarised in *Table* 5.1.3.

Enzyme	Role of residue substituted	Relative activity
Native BsPel	___	100 %
R279A	Catalytic base	<0.01 %
R279K	Catalytic base	1 %
D184A	Binds calcium 1	8 %
D223A	Binds calcium 1 and 2	<0.01 %
D227A	Binds calcium 1	0.5 %
D173A	Binds calcium 3	40 %
N180A	Binds calcium 3	30 %
D173A/ N180A/K247A	Above, plus catalytic acid substituted	0.2 %

Table 5.1.3 A summary of the relative activities and the role of various BsPel mutants using polygalacturonate as substrate (Seyedarabi 2010b).

To determine the enzyme catalytic activity of triple mutant BsPel a spectroscopic enzyme assay was used to identify the conditions producing the lowest activity i.e lowest ability to turnover the sugar substrate and allowing trapping. The assay measures the formation of carbon-carbon double bonds at 235 nm. The formation of the bond is the result of BsPel breaking down α-1,4-glycosidic bonds in the sugar provided in the assay. Relative enzyme activity was calculated as a percentage using native BsPel activity. Details of the assay protocol can be found in section 2.2.9.

Triple mutant BsPel had 0.2 % activity, which was too high and meant that oligosaccharide substrate would be turned over too quickly; therefore the conditions of the assay were modified in order to decrease the catalytic activity. In place of $CaCl_2$, $CoCl_2$ was used at varying concentrations also the pH of the assay was reduced from pH 8.5 to pH 4.6 and then pH 3.0. The results showed the final enzyme assay condition which had no recorded enzyme activity consisted of 1 mM $CoCl_2$ at pH 3.0. This was the condition used for trapping of sugar substrate in BsPel with Arg 279 present.

5.5 Crystallography

5.5.1 Crystallisation of triple mutant BsPel

Purified and concentrated protein was used in crystallisation experiments in a hanging drop set up see section 2.2.10.1. Target concentration for BsPel crystallisation was 25 mg/ml and the well conditions are detailed in *Table* 5.1.4, all stock concentrations can be found in section 2.1.10.1, *Table* 2.9.

Reservoir	Tris HCL pH 8.5 (µl)	Ammonium acetate pH 4.6 (µl)	Sodium acetate (µl)	PEG 4000		ddH_2O
				(µl)	(%)	
A1	50	0	100	190	19	160
A2	50	0	100	200	20	150
A3	50	0	100	210	21	140
A4	50	0	100	220	22	130
A5	50	0	100	230	23	120
A6	50	0	100	240	24	110
B1	0	100	50	200	20	150
B2	0	100	50	210	21	140
B3	0	100	50	220	22	130
B4	0	100	50	230	23	120
B5	0	100	50	240	24	110
B6	0	100	50	250	25	100

Table 5.1.4 Summary of the reservoir conditions used for crystallisation of BsPel.

Crystals were successfully grown from a number of the reservoirs however the largest single crystals (see *Figure* 5.1.6) were obtained from condition B6 (see *Table* 5.1.4) consisting of 0.2 M ammonium acetate pH 4.6, 0.1 M sodium acetate and 25 % PEG 4000.

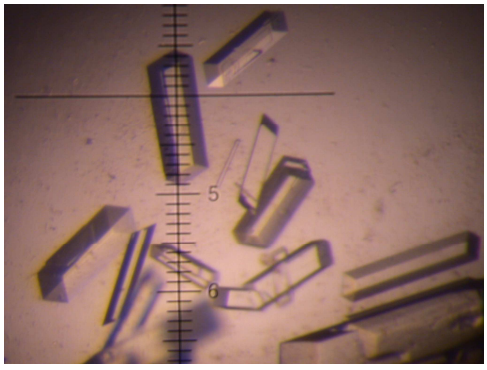


Figure 5.1.6 Large singular, straight edged triple mutant BsPel crystals were obtained from reservoir condition B6 (see *Table* 5.1.4) they are 0.15 mm in length. Every ten small divisions on the scale bar was equivalent to 0.1 mm.

5.5.2 Crystal soaking

In order to obtain a Michaelis complex with oligosaccharide substrate hexagalacturonate (DP6), kindly provided by Alistair MacDougall, bound to BsPel enzyme the crystal was soaked in various solutions under various conditions and these are summarised in *Table* 5.1.5. For crystal soaking protocol see section 2.2.10.2, the cryoprotectant used was 20 % glycerol. Once the crystals had been soaked they were immediately mounted onto the goniometer ready for data collection.

2 mM $CoCl_2$ PH 3	
DP6 (mM)	Soak time (mins)
2	5
16	10
20	10
20	55
33	12
42	55
100	10

Table 5.1.5 Lists the concentration of sugar substrate DP6 used in the soak composition together with the duration of individual soaks.

5.5.3 Data collection and structure solution

A high resolution dataset of a calcium-free triple mutant BsPel with DP6 bound was collected to a resolution of 1.90 Å at SRS Daresbury Synchrotron at workstation 10.1. The detector used was a MARMOSAIC 225 mm CCD with 320 images collected using a 1 ° oscillation. The crystal had cell dimensions; a= 50.6 Å, b= 69.0Å, c= 59.2 Å and β= 113.6º. The spacegroup was $P2_1$ with one molecule in the asymmetric. The triple

mutant BsPel crystal was soaked in 100 mM DP6 for 10 minutes. Collected images were processed using iMOSFLM (Leslie 1992) and reduced using SCALA from the CCP4i suite (Dodson 1997). For further crystallographic data statistics see *Table* 5.1.6.

	Overall	Inner shell	Outer shell
Low resolution limit (Å)	69.01	69.01	2.00
High resolution limit (Å)	1.90	6.01	1.90
[a]Rmerge (%)	0.057	0.057	0.067
[b]Rmeas (within I+/I-) (%)	0.067	0.067	0.079
[b]Rmeas (all I+ & I-) (%)	0.067	0.067	0.079
[c]Rpim (within I+/I-) (%)	0.035	0.036	0.041
[c]Rpim (all I+ & I-) (%)	0.035	0.036	0.041
Fractional partial bias	-0.028	-0.035	-0.025
Total number of observations	106755	3450	15382
Total number unique	28827	967	4138
Mean [(I)/sd(I)]	23.8	27.9	18.3
Completeness (%)	97.2	98.5	96.1
Multiplicity	3.7	3.6	3.7
Wilson B-factor (Å^2)	8.608		

Table 5.1.6 Crystallographic data statistics for BsPel triple mutant N180A, D173A and K247A with calcium substituted with cobalt, data collected to 1.90Å.
[a] Rmerge = ($\Sigma hkl\ \Sigma j\ |\ Ihkl - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl,j$)
[b] R*meas* = ($\Sigma hkl \sqrt{(n/n-1)}\ \Sigma^{n}_{j=1}\ |\ Ihkl,j - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl,j$)
[c] Rpim = ($\Sigma hkl \sqrt{(1/n-1)}\ \Sigma^{n}_{j=1}\ |\ Ihkl,j - \langle Ihkl \rangle\ |$) / ($\Sigma hkl\ \Sigma j\ Ihkl,j$)

5.5.4 Refinement

The structure was refined using REFMAC5 (Vagin 2004) to an Rfactor and Rfree of 13.1 % and 19.6 % respectively. The Rfactor was calculated using 95 % of the data included in the refinement and Rfree was calculated from the 5 % of reflections not

included. The triple mutant structure in complex with hexagalacturonate and Co^{2+} was deposited in Protein Data Bank under the entry 3KRG.

5.5.5 Validation

Validation of the structure took place to ensure the model quality was high. As a model is refined Rfree, Rfactor and root mean square (rms) deviations including bond lengths, angles and conformational angles are used to monitor the convergence to a final refined model. Generally rms deviations of no more than 0.02 Å for bond lengths and 4 ° for bond angles are accepted (Rhodes 2006), for the BsPel triple mutant final structure the rms length and bond values were 0.018 and 1.648 repsectively. For visualisation of dihedral angles ψ against φ enabling quick detection of conformationally unrealistic regions of the structure a Ramachandran plot was used (see *Figure* 5.1.7). Most residues were situated in preferred regions with only six outliers.

Validation of the model was carried out using PROCHECK V3.4.4 (Morris 2003) from RCSB Protein Data Bank (see *Figures* 5.1.8 and 5.1.9).

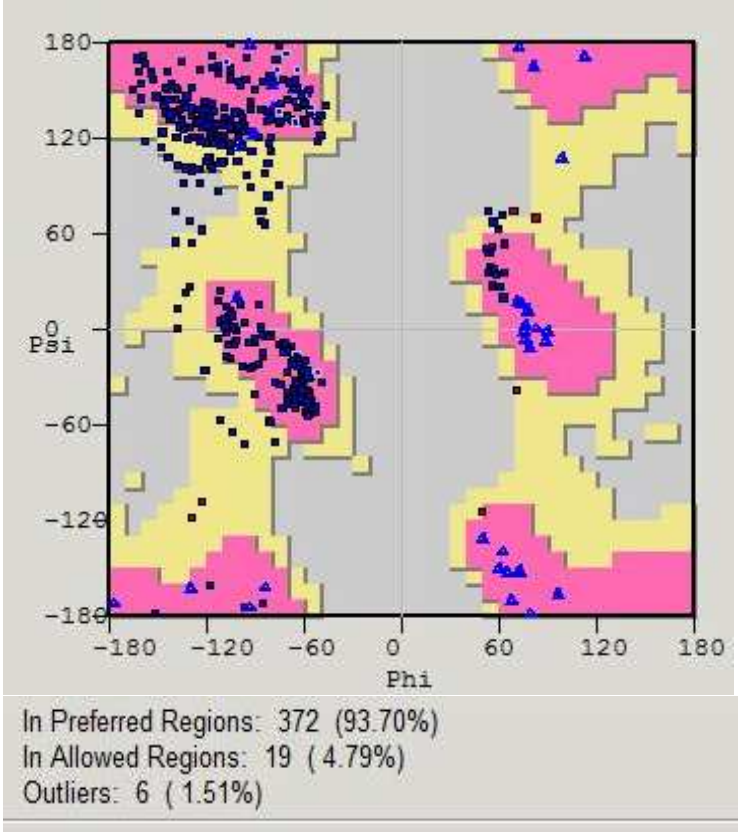


Figure 5.1.7 Ramachandran plot showing the majority of residues are in conformationally allowed regions with only 1.51 % outliers.

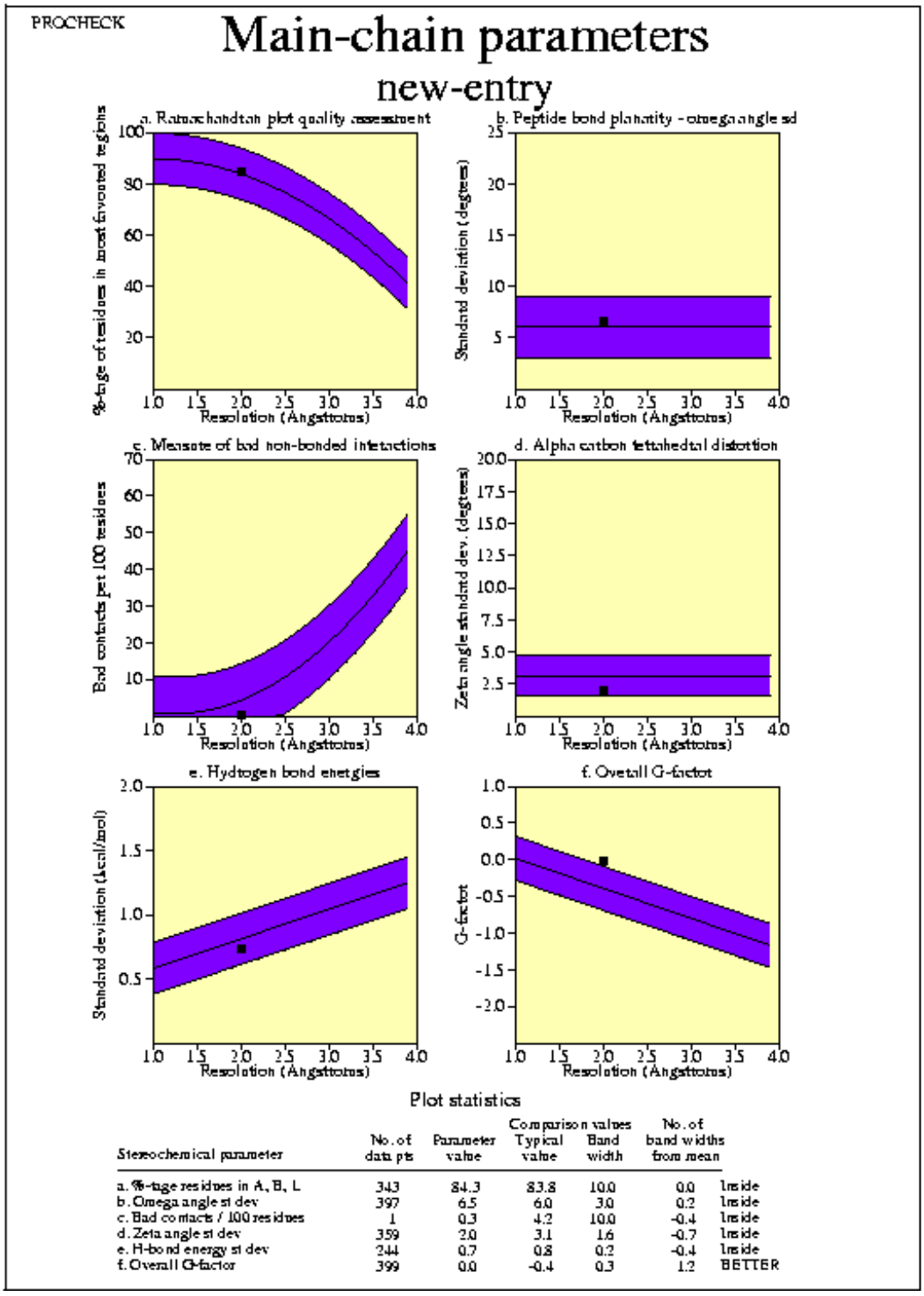


Figure 5.1.8 Main chain parameter analysis from PROCHECK V3.4.4 for BsPel bound to hexagalacturonate with a cobalt ion.

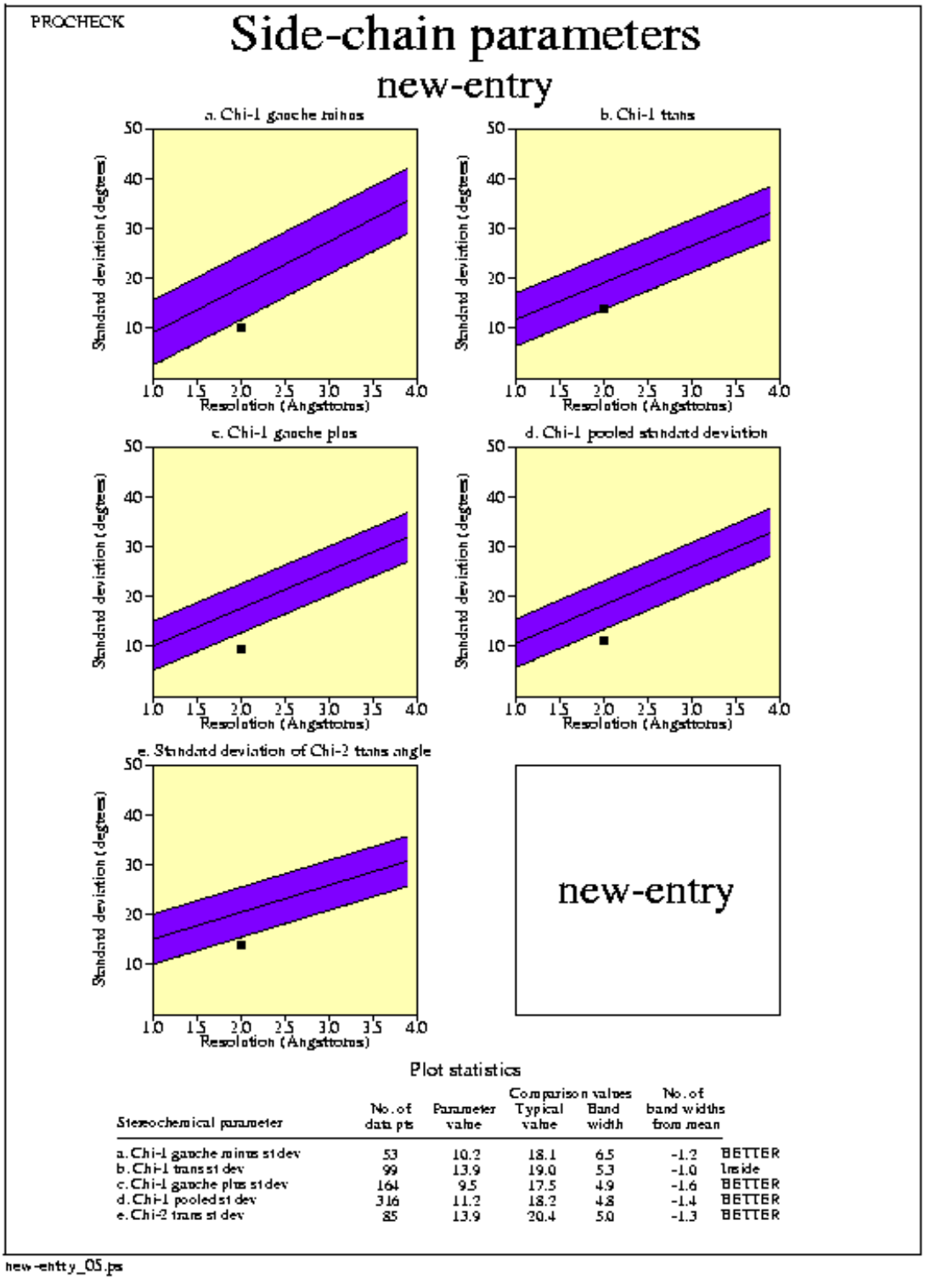


Figure 5.1.9 Side chain parameter analysis from PROCHECK V3.4.4 4 for BsPel bound to hexagalacturonate with a cobalt ion.

5.5.6 Michaelis complex of BsPel

The structure of BsPel-D173A, D180A and K247A mutant in complex with hexagalacturonate with a Co^{2+} ion bound at the first calcium binding site replacing the original Ca^{2+} ion was readily solved by molecular replacement using MOLREP (Vagin 1997) and the native structure (protein data bank code: 1BN8) as the search model. Refinement and model building made use of REFMAC5 (Vagin 2004) and COOT (Emsley 2004), respectively. Arg279 can be seen in an appropriate position for abstracting the proton from C5 of hexagalacturonate, see *Figure* 5.1.10. The presence of the three mutations was confirmed by examination of the electron density map (see *Figure* 5.1.11).

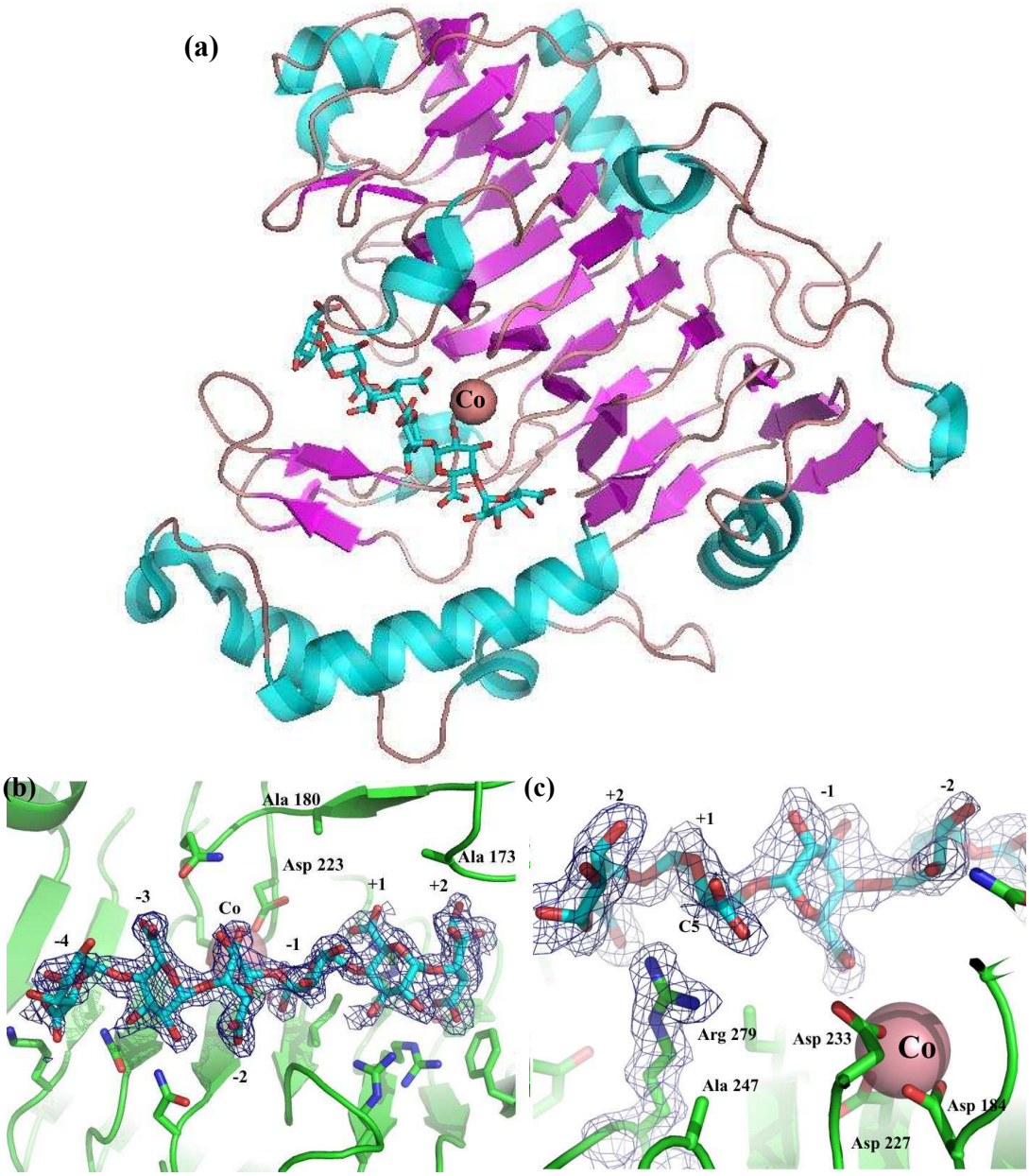


Figure 5.1.10 **(a)** Diagram illustrating the parallel β-helix structure of triple mutant inactive BsPel in complex with hexagalacturonate and a Co^{2+} ion bound. **(b)** Density showing bound DP6 with mutated residues N180A and D173A, along with the catalytic base. **(c)** Residue Arg 279 is situated in close proximity to C5 of sugar occupying +1 subsite ready for proton abstraction. The density contour level was at 1σ for panels (b) and (c) (figures from Seyedarabi 2010b).

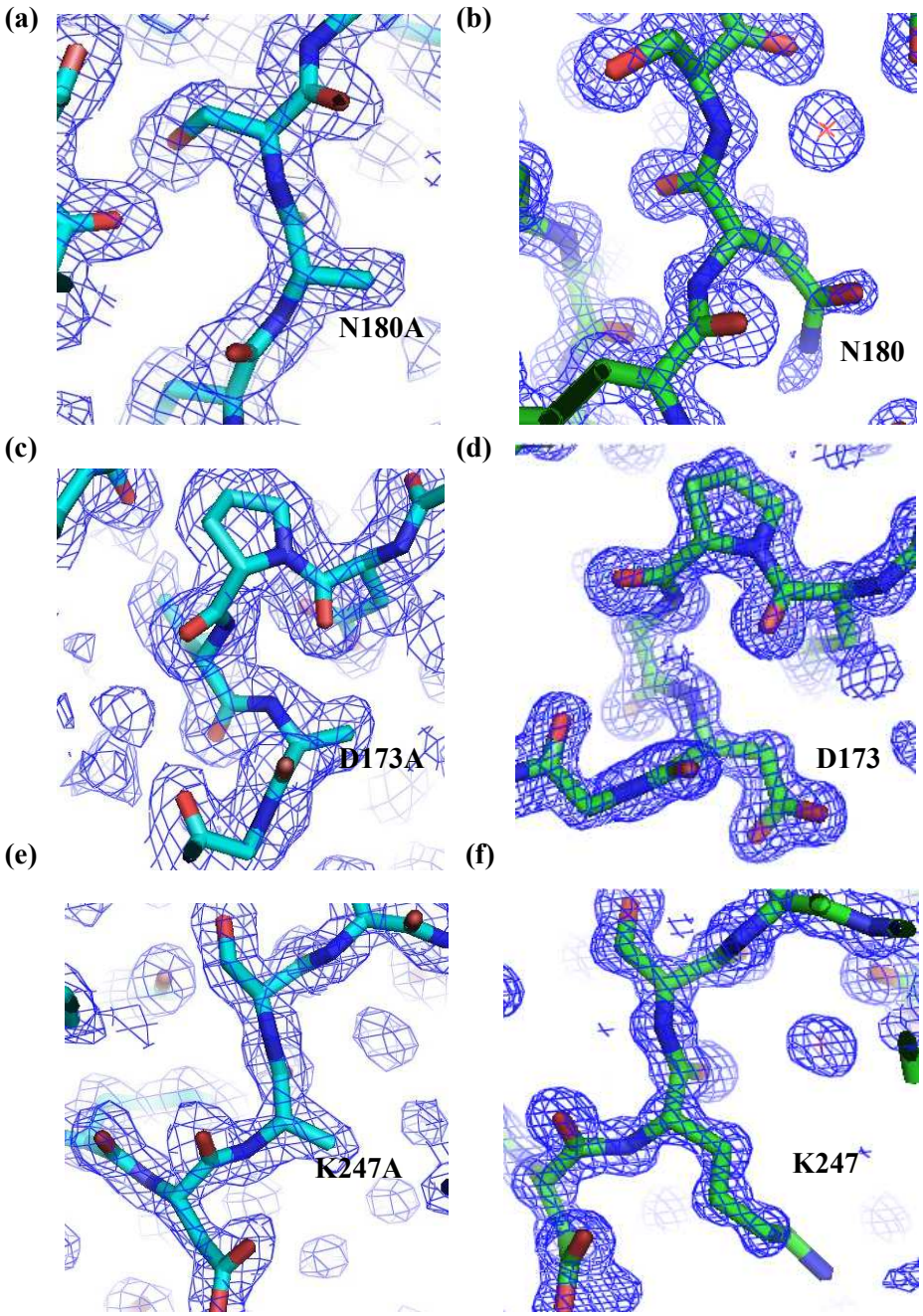


Figure 5.1.11 **(a)** Mutant N180A showing density for alanine confirming the mutation from asparagine. **(b)** N180 from wildtype BsPel. **(c)** Successful mutation from aspartate to alanine, no density is present for asapartate. **(d)** Control showing clear density for and D173 residue from wildtype BsPel structure. **(e)** Density shown for alanine in place of lysine confirming the mutation K247A. **(f)** Density for K247 from wild type BsPel.

5.6 Conclusion

The structure of triple mutant (D173A, N180A and K247A) BsPel in complex with hexagalacturonate and a cobalt ion has been solved to a resolution of 1.90 Å. Hexagalacturonate binds from subsites +2 through to -4 with weak density at -4 subsite (0.5 σ contour level). In order for an inactive enzyme to be obtained, mutations of three residues surrounding the third calcium were made and calcium was replaced with cobalt along with lowering of pH to pH 3. In doing so a Michaelis-complex was successfully trapped with the residue Arg279 present in the structure. This enabled viewing of the enzyme with arginine in place and allowing the confirmation of an appropriately positioned residue for proton abstraction at C5 (see *Figure* 5.1.10 panel c). This validates the Arg279 as the catalytic base as proposed previously (Charnock 2002) by the observation of converging geometry in the active site as seen in the comparison between PL-10 and PL-1 family lyase structures.

Pectate lyase uses *anti* β-elimination chemistry cleaving α-1,4-glycosidic bonds in polygalacturonate. The enzyme requires essential catalytic residues together with calcium ion coordination and a bound sugar substrate in order for catalysis to take place. This mechanism is depicted in *Figure* 5.1.12 showing Arg279 abstracting the proton from C5 and cleaving the glycosidic bond separating R_1 (additional α-1,4 linked galacturonate residues) as the leaving group. During this process Lys247 and two calcium ions binding between the enzyme and substrate carboxylates at the +1 subsite stabilise the reaction intermediate.

This work forms part of a recent publication (Seyedarabi 2010b) see Appendix 9.

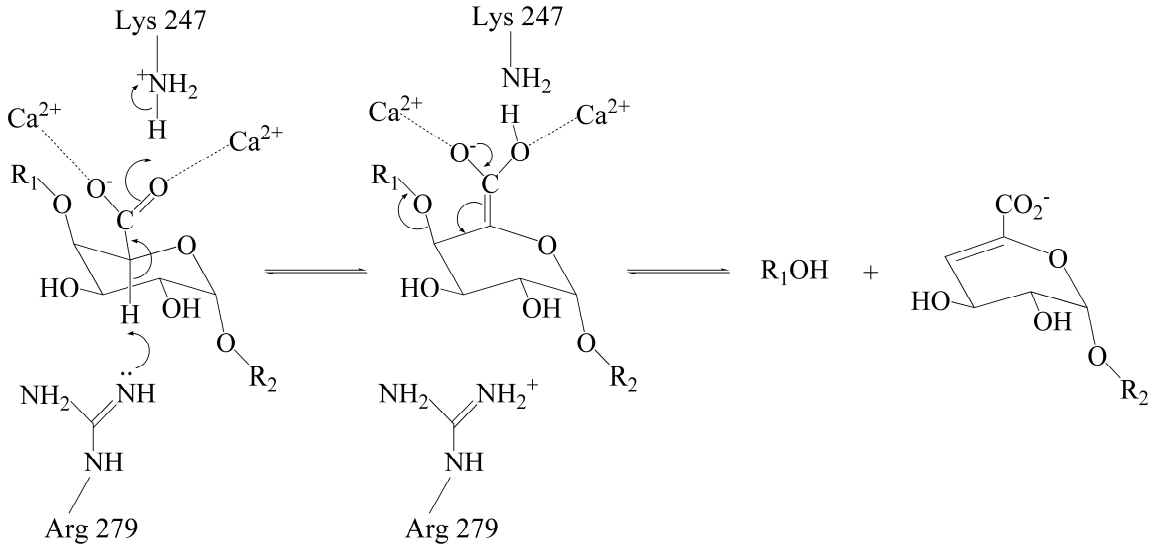


Figure 5.1.12 Reaction scheme showing the mechanism used for cleaving of glycosidic bonds with Arg279 acting as the catalytic base and Lys247 together with two calcium ions stabilising the reaction intermediate (Figure from Seyedarabi 2010b). See Figure 5.1.2 for the role and contacts made by D173 and N180.

5.7 Future work

The original motivation for this project was to trap the Michaelis complex in the prescence of the catalytic base by mutating the third calcium-binding site. This proved to be insufficient to inactivate the enzyme and in addition lysine 247 was mutated, forming a triple mutant BsPel. To further suppress the enzyme activity cobalt was used in place of calcium and the pH was lowered.

It would be worth investigating if the Michaelis complex is formed using wild type enzyme by simply lowering the pH. Another possibility is just the substitution of cobalt for calcium may be sufficient to inactivate the enzyme, or this substitution in combination with low pH may be needed. It is plausible that both cobalt substitution and low pH are needed to inactivate the enzyme, but this could be tested experimentally by assaying the activity in solution and soaking oligosaccharide in the crystal.

It would also be interesting to know the pKa of the catalytic arginine, this could be investigated both theoretically and experimentally. Calculations may suggest the depression of pKa caused by the presence of oligosaccharide substrate and the proximal calcium ions. Experimentally it might be possible to measure the pKa of the arginine in the complex formed using a non-cleavable substrate analogue, but to date no such analogue, for example with sulphur replacing the glycosidic oxygen, has been made. If a suitable non-cleavable substrate were available, then the pKa measurement could be made using NMR spectroscopy or possibly infrared spectroscopy.

Chapter Six:
Final Conclusions

In this thesis I have studied the structure and function of three enzymes: telomerase, the cobalamin methyltransferase CobJ, and pectate lyase. The three projects were at different stages of their development, there was no structure of any telomerase, there was no structure of CobJ although members of the cobalamin methyltransferase family had been solved, and there were structures and complexes formed in pectate lyase but not without mutation of the catalytic base. A thread that runs through this thesis is the medical and biotechnological impact that knowledge of these enzymes may provide. Inhibiting the action of telomerase has potential application in cancer therapy since there is evidence that shows the enzyme is active and provides immortality in cancerous cells. A vital enzyme required for the living of eukaryotes, but also the key to the sustainability of the biggest killer in humans (Hahn 1999).

Results chapter three covered results obtained from what proved to be a challenging protein, telomerase. Two constructs His-tagged and GST-tagged failed to yield soluble and non-aggregated protein. Knowing the insoluble fate of *C.elegans* TERT expressed in *E.coli*, further action was taken to try and produce soluble protein using expression trials, folding chaperones GroES and GroEL and solubilisation using high concentrations of urea. Clearly there is much more that could be done with *C. elegans* TERT including screening a large number of constructs for solubility, discovering the TR component and co-expression to name just two. Publication of the first telomerase structure combined with the expression difficulties signalled the time to concentrate on more soluble proteins.

Chapter four focused on the results obtained from CobJ the methyltransferase from *R.capsulatus*. CobJ is one of the enzymes from the vitamin B_{12} pathway responsible for synthesising cobalamin an essential cofactor required by humans who are otherwise unable to make it. This provides much biotechnological interest in the synthesis of vitamin B_{12} supplements since a lack of vitamin B_{12} can lead to a variety of health problems such as pernicious anaemia. This work was important because CobJ catalyses

the important ring-contraction reaction, a contracted macrocycle sets cobalamin apart from all other tetrapyrroles. Crystallisation of CobJ was explored (Seyedarabi 2010a) see Appendix 8, and a high resolution structure of CobJ at 2.2 Å was refined and validated alongside a mutant H129A CobJ structure. The main aim was to obtain an enzyme complex with a tetrapyrrole to help elucidate the ring contraction mechanism enabling the identification of amino acid residues involved in methylation and ring contraction. This has also proved to be a challenging goal. Enzymatically the production of precorrin-3a was achievable under anaerobic conditions, but ideally precorrin-3b or precorrin-4 was a more suitable candidate for complexing with CobJ since it is the direct substrate/product. A number of methods were employed in an attempt to trap a complex including soaking experiments and co-crystallisation with a variety of tetrapyrroles and derivatives and using pull-down experiments using a multi-enzyme assay and exploiting the His-tag pET14b-CobJ.

I was able to extend the resolution of the CobJ structure to 2.2 Å and flexible regions were identified by B-factor analysis showing a region of residues lining the opening of the pocket with flexibility which could move to accommodate a large incoming tetrapyrrole. This is backed by the low density seen for these residues in the structure and the instant cracking of the crystal during soaking experiments with precorrin-3a. Analysis of the CobJ structure highlighted conserved residues in the vicinity of the active site pocket which could act as a proton acceptor in the ring-contraction step and His 129 appears to be an important residue and possibly plays a role in ring contraction.

All enzymes in the vitamin B_{12} pathway have highly specific preferences to their "correct" tetrapyrroles this is highlighted by the fact that CobJ complexing with "incorrect" tetrapyrroles or fragments of a tetrapyrrole was not seen indicating that the "correct" tetrapyrrole precorrin-3b or precorrin-4 is needed in order to form a complex. *In silico* studies may help to further elucidate the mechanism or alternatively further work is required on the pull-down experiment since the technique was successful for

CobH pulled-down with HBA product bound and for cobalt chelatases CbiX and CbiK complexes.

Finally, in chapter five I discuss trapping a complex in *B. subtilis* pectate lyase. Pectate lyases have been long used in the food, healthcare and textile industries and phytopathogenic bacteria secreting pectate lyase continue to challenge global food security. Further understanding of the mechanism of pectate lyases may help to develop mechanism-based inhibitors useful for the production of pesticides. Manipulation of pectate lyase specificity might be useful for producing novel polysaccharide products. Previous mutagenesis and crystallographic studies have identified an arginine as the likely candidate for the catalytic base that abstracts the C5 proton from the sugar residue in the +1 subsite.

I successfully trapped a Michaelis complex revealing the structure of an inactive BsPel with Arg279 present in complex with hexagalacturonate and a cobalt ion. Upon examination of the structure it was clear to see the Arg279 was positioned in an ideal place to abstract a proton from C5 and catalyse the *anti* β-elimination reaction cleaving the α-1,4 glycosidic bond (Seyedarabi 2010b), see Appendix 9.

Bibliography

Akita, M., Suzuki, A., Kobayashi, T., Ito, S., and Yamane, T. (2001). "The first structure of pectate lyase belonging to polysaccharide lyase family 3." Acta. Crystallogr. Sect. D. **57**: 1786-1792.

Alkorta, I., Garbisu, C., Llama, M.J., and Serra, J.L. (1998). "Industrial applications of pectic enzymes: a review." Process Biochemistry **33**: 21-28.

Autexier, C., and Lue, N.F. (2006). "The structure and function of telomerase reverse transcriptase." Annu. Rev. Biochem. **75**: 493-517.

Allen, R. H., Stabler, S.P., Savage, D.G., and Lindenbaum, J. (1993). "Metabolic abnormalities in cobalamin (vitamin B_{12}) and folate deficiency." FASEB J **7**(14): 1344-1353.

Battersby, A. R. (1994). "How nature builds the pigments of life: the conquest of vitamin B_{12}." Science **264**(5156): 1551-1557.

Blackburn, E. H., and Szostak, J.W. (1984). "The molecular structure of centromeres and telomeres." Annu. Rev. Biochem. **53**: 163-194.

Blanche, F., Couder, M., Debussche, L., Thibaut, D., Cameron, B., and Crouzet, J. (1991). "Biosynthesis of vitamin B_{12}: stepwise amidation of carboxyl groups b,d,e, and g of cobyrinic acid a,c-diamide is catalyzed by one enzyme in *Pseudomonas denitrificans*." J. Bacteriol **173**: 6046-6051.

Blanche, F., Famechon, A., Thibaut, D., Debussche, L., Cameron, B., and Crouzet, J. (1992). "Biosynthesis of vitamin B12 in Pseudomonas denitrificans: the biosynthetic sequence from precorrin-6y to precorrin-8x is catalyzed by the cobL gene product." J. Bacteriol **174**: 1050-1052.

Blanche, F., Thibaut, D., Debussche, L., Hertle, R., Zipfel, F, and Muller, G. (1993). "Parallels and Decisive Differences in Vitamin B12 Biosyntheses." Angew. Chem. Int. Ed. Engl. **32**: 1651.

Bosoy, D. P. Y., Milan, I.S., and Lue, N.F. (2003). "Conserved N-terminal motifs of telomerase reverse transcriptase required for ribonucleoprotein assembly in vivo." J. Biol. Chem **6**: 3882-3890.

Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." Analytical Biochem. **72**: 248-254.

Bryan, T. M., goodrich, K.J., and Cech, T.R. (2000). "Telomerase RNA bound by protein motifs specific to telomerase reverse transcriptase." Mol. Cell. **6**: 493-499.

Cameron, B., Blanche, F., Rouyez, M.C., Bisch, D., Famechon, A., Couder, M., Cauchois, L., Thibaut, D., Debussche, L., and Crouzet, J. (1991). "Genetic analysis, nucleotide sequence, and products of two *Psuedomonas denitrificans* cob genes encoding nicotinate nucleotide: dimethylbenzimidazole phosphoribosyltransferase and cobalamin (5'-phosphate) synthase." American Society for Microbiology **173**: 6066-6073.

Cantarel, B. L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics." Nucleic Acids Res. **37**: D233-D238.

Carpita, N. C., and Gibeaut, D.M. (1993). "Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth." The Plant Journal **3**: 1-30.

Charnock, S. J., Brown, I.E., Turkenburg, J.P., Black, G.W., and Davies, G.J. (2002). "Convergent evolution sheds light on the anti-β-elimination mechanism common to family 1 and 10 polysaccharide lyases." Proc. Natl. Acad. Sci. U.S.A. **99**: 12067-12072.

Chen, J. L., Blasco, M.A., and Greider, C.W. (2000). "Secondary structure of vertebrate telomerase RNA." Cell **100**: 503-514.

Chen, J. L., and Greider, C.W. (2003a). "Template boundary definition in mammalian telomerase." Genes. Dev. **17**: 2747-2752.

Chen, J. L., and Greider, C.W. (2003b). "Determinants in mammalian telomerase RNA that mediate enzyme processivity and cross-species incompatibility." EMBO J **22**: 304-314.

Chen, J. L., and Greider, C.W. (2004). "An emerging consensus for telomerase RNA structure." Proc. Natl. Acad. Sci. USA **101**: 14683-14684.

Collins, K. (1999). "Ciliate telomerase biochemistry." Annu. Rev. Biochem. **68**: 187-218.

Collins, K. (2000). "Mammalian telomeres and telomerase." Curr. Opin. Cell. Biol. **12**: 378-383.

Collmer, A., and Keen, N.T. (1986). "The role of pectic enzymes in plant pathogenesis." Ann. Rev. Phytopathol. **24**: 383-409.

Counter, C. M., Hahn. W.C., Wei, W., Dickinson Caddle, S., Beijersbergen, R.L., Lansdorp, P.M., Sedivy, J.M., and Weinberg, R.A. (1998). "Dissociation among in *vitro* telomerase activity, telomere maintenance, and cellular immortilization." Proc. Natl. Acad. Sci. USA **95**: 14723-14728.

Creze, C., Castang, S., Derivery, E., Haser, R., Hugouvieux-Cotte-Pattat, N., Schevchik, V.E., and Gouet, P. (2008). "The crystal structure of pectate lyase PelI from soft rot pathogen *Erwinia chrysanthemi* in complex with its substrate." J. Biol. Chem. **283**: 18260-18268.

Crouzet, J., Cauchois, F., Blanche, L., Debussche, D., Thibaut, M., Rouyez, C., Rigault, S., Mayaux, J.F., and Cameron, B. (1990). "Nucleotide sequence of a *Pseudomonas denitrificans* 5.4-kilobase DNA fragment containing five cob genes and identification of structural genes encoding S-adenosyl-L-methionine: uroporphyrinogen III methyltransferase and cobyrinic acid a,c-diamide synthase." J. Bacteriol **172**: 5968.

Crouzet, J., Levy-Schil, S., Cameron. B., Cauchois, L., Rigault, S., Rouyez, M.C., Blanche, F., Deussche, L., and Thibaut, D. (1991). "Nucleotide sequence and genetic analysis of a 13.1-kilobase-pair *Pseudomonas denitrificans* DNA fragment containing five cob genes and identification of structural genes encoding Cob(I)alamin adenosyltransferase, cobyric acid synthase, and bifunctional cobinamide kinase-cobinamide phosphate guanylyltransferase." J. Bacteriol **173**: 6074-6087.

Dandjinou, A. T., Levesque, N., Larose, S., Lucier, J.F., AbouElela, S., and Wellinger, R.J. (2004). "A phylogenetically based secondary structure for the yeast telomerase RNA." Curr. Biol. **14**: 1148-1158.

Davies, G., and Henrissat, B. (1995). "Structures and mechanisms of glycosyl hydrolases." Structure **3**: 853-859.

Debussche, L., Crouder, M., Thibaut, D., Cameron, B., Crouzet, J., and Blanche, F. (1991). "Purification and partial characterization of Cob(I)alamin adenosyltransferase from *Psuedomonas denitrificans*." J. Bacteriol **173**: 6300-6302.

Debussche, L., Thibaut, D., Cameron, B., Crouzet, J., and Blanche, F. (1990). "Purification and characterization of cobyrinic acid a,c-diamide synthase from *Pseudomonas denitrificans.*" J. Bacteriol **172**: 6239-6244.

Debussche, L., Thibaut, D., Cameron, B., Crouzet, J., and Blanche, F. (1993). "Biosynthesis of the corrin macrocycle of coenzyme B_{12} in *Pseudomonas denitrificans*." J. Bacteriol **175**(22): 7430-7440.

Debussche., C., M., Thibaut, D., Cameron, B., Crouzet, J., and Blanche, F. (1992). "Assay, purification, and characterization of cobaltchelatase, a unique complex enzyme

catalyzing cobalt insertion in hydrogenobyrinic acid a,c-diamide during coenzyme B_{12} biosynthesis in *Pseudomonas denitrificans*." J. Bacteriol **174**: 7445-7451.
Dodson, E. J., Winn, M., and Ralph, A. (1997). "Collaborative computational project, number 4: providing programs for protein crystallography." Methods Enzymol **277**: 620-633.

Evans, P. R. (2005). "Scaling and assessment of data quality." Acta Cryst D62: 72-82

Emsley, P., and Cowtan, K. (2004). "Coot: model- building tools for molecular graphics." Acta Cryst D Biol Crystallography **60**: 2136-32.

Forsyth, N. R., Wright, W.E., and Shay, J.W. (2002). "Telomerase and differenciation in multicellular organisms: turn it off, turn it on, and turn it off again." Differenciation **69**: 188-197.

Frank, S., Brindley, A.A., Deery, E., Heathcote, P., Lawrence, A.D., leech, H.K., Pickersgill, R.W., and Warren, M.J. (2005). "Anaerobic synthesis of vitamin B_{12}: characterization of the early steps in the pathway." Bio. Soc. Trans. **33**: 811-814.

Frank, S., Deery, E., Brindley, A.A., Leech, H.K., Lawrence, A., Heathcote, P., Schubert, H.L., Brocklehurst, K., Rigby, S.E., Warren, M.J., and Pickersgill, R.W. (2007). "Elucidation of substrate specificity in the cobalamin (vitamin B12) biosynthetic methyltransferases. Structure and function of the C20 methyltransferase (CbiL) from *Methanothermobacter thermoautotrophicus*." J. Biol. Chem **282**: 23957-23969.

Friedman, K. L., and Cech, T.R. (1999). "Essential functions of amino-terminal domains in the yeast telomerase catalytic subunit revealed by selection for viable mutants." Genes. Dev. **13**: 2863-2874.

Friedman, K. L., Heit, J.J., Long, D.M., and Cech, T.R. (2003). "N-terminal domain of yeast telomerase reverse transcriptase: recruitment of Est3p to the telomerase complex." Mol. Biol. Cell. **14**: 1-13.

Gillis, A. J., Schuller, A.P., and Skordalakes, E. (2008). "Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT." Nature **455**: 633-638.

Greider, C. W., and Blackburn, E.H. (1985). "Identification of a specific telomere terminal transferase activity in Tetrahymena extracts." Cell **43**: 405-413.

Hahn, W. C., Stewart, S.A., Brooks, M.W., York, S.G., Eaton, E., Kurachi, A., Beijersbergen, R.L., Knoll, J.H.M., Meyerson, M., and Weinberg, R.A. (1999). "Inhibition of telomerase limits the growth of human cancer cells." Nature Medicine **5**: 1164-1170.

Harrington, L. A., and Greider, C. W. (1991). "Telomerase primer specificity and chromosome healing." Nature **353**: 451-454.

Hayflick, L., and Moorhead, P.S. (1961). "The serial cultivation of human diploid cell strains." Exp. Cell. Res. **25**: 585-621.

Heldt, D., Lawrence, A.D., Lindenmeyer, M., Deery, E., Heathcote, P., Rigby, S.E., and Warren, M.J. (2005). "Aerobic synthesis of vitain B_{12}: ring contraction and cobalt chelation." Bio. Soc. Trans. **33**: 815-819.

Herron, S. R., Benen, J. A. E., Scavetta, R. D., Visser, J. and Jurnak, F. (2000). "Colloquium Paper: Structure and function of pectic enzymes: Virulence factors of plant pathogens." Proc. Natl. Acad. Sci. USA **97**: 8762-8769.

Huard, S., Moriarty, T.J., and Autexier, C. (2003). "The C terminus of the human telomerase reverse transcriptase is a determinant of enzyme processivity." Nucleic Acids Research **31**: 4059-4070.

Horowitz, N. H. (1945). "On the evolution of biochemical synthesis." Proc. Natl. Acad. Sci. USA **31**: 153-157.

Huard, S., Moriarty, T.J., and Autexier, C. (2003). "The C terminus of the human telomerase reverse transcriptase is a determinant of enzyme processivity." Nucleic Acids Research **31**: 4059-4070.

Jacobs, S. A., Podell, E.R., and Cech, T.R. (2006). "Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase." Nat. Struct. Mol. Biol **13**: 218-225.

Jenkins, J., Mayans, O., Smith, D., Worboys, K., Pickersgill, R. (2001). "Three-dimensional structure of *Erwinia chrysanthemi* pectin methylesterase reveals a novel esterase active site." J. Mol. Biol. **305**: 951-960.

Jenkins, J., Schevchik, V.E., Hugouvieux-Cotte-Pattat, N., and Pickersgill, R.W. (2004). "The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*." J. Biol. Chem. **279**: 9139-9145.

Jones, S. J. M., Riddle, D.L., Pouzyrev, A.T., Velculescu, V.E., Hillier, L., Eddy, S.R., Stricklin, S.L., Baillie, D.L., Waterston, R., and Marra, M.A. (2001). " Changes in gene expression associated with developmental arrest and longevity in *Caenorhabtidis elegans*." Genome Research **11**: 1346-1352.

Joyce, C. M., and Steitz, T.A. (1994). "Function and structure relationships in DNA polymerases." Annu. Rev. Biochem. **63**: 777-822.

Kadner, R. J., and Bassford, P.J. Jr. (1977). "Relation of cell growth and colicin tolerance to vitamin B_{12} uptake in *Eschericia coli*." J. Bacteriol. **129**: 254-264.
Lai, C. K., Mitchell, J.R., and Collins, K. (2001). "RNA binding domain of telomerase reverse transcriptase." Mol. Cell. Biol **21**: 990-1000.

Kantardjieff, K. A., and Rupp, B. (2003). "Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein-nucleic acid complex crystals." Protein Science **12**: 1865-1871.

Kashyap, D. R., Vohra, P.K., Chopra, S., and Tewari, R. (2001). "Applications of pectinases in the commercial sector: a review." Bioresource Technology **77**: 215-227.

Kim, Y., Xu, X., Savchenko, A., Edwards, A., and Joachimiak, A. "Crystal Structure of Cobalamin Biosynthesis Precorrin-6Y Methylase (cbiE) from *Archaeoglobus fulgidus.*"

Krissinel, E., and Henrick K. (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Cryst **D60**: 2256-2268.

Krissinel, E., and Henrick K. (2007). "Inference of macromolecular assemblies from crystalline state." J. Mol. Biol. **372**: 774-797.

Laine, R. A. (1994). "A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05x 10^{12} structures for a reducing hexasaccharide: the isomer barrier to development of single-method saccharide sequencing or synthesis systems." Glycobiology **4**: 759-767.

Laskowski, R. A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Cryst. **26**: 283.

Lawrence, A. D., Deery, E., McLean, A., Munro, A.W., Pickersgill, R.W., Rigby, S.E., and Warren, M.J. (2008). "Identification, characterization, and structure/function analysis of a corrin reductase involved in adenosylcobalamin biosynthesis." J. Biol. Chem **283**: 10813-10821.

Leech, H. K., Raux, E., McLean, K.J., Munro, A.W., Robinson, N.J., Borrelly, G.P.M., Malten, M., Jahn, D., Rigby, E.J., Heathcote, P., and Warren, M.J. (2003). "Characterization of the cobaltchelatase CbiX." J. Bio. Chem. **278**: 41900-41907.

Ledley, F. D. (1990). "Perspectives on methylmalonic acidemia resulting from molecular cloning of methylmalonyl CoA mutase." Bioessays **12**(7): 335-340.

Lemercier, G., Bakalara, N., and Santarelli, X. (2003). "On-column refolding of an insoluble histidine tag recombinant exopolyphosphatase from *Trypanosoma brucei* overexpressed in *Eschericia coli*." J. Chromatography. B. **786**: 305-309.

Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and plate data." Joint CCP4 and ESF-EAMCB Newsletter on protein crystallography 26.

Lietzke, S. E., Yoder, M.D., Keen, N.T., and Jurnal, F. (1994). "The three-dimensional structure of pectate lyase E, a plant virulence factor from *Erwinia chrysanthemi*." Plant Physiology **106**: 849-862.

Lingner, J., and Cech, T.R. (1996). "Purification of telomerase from *Euplotes aediculatus*: requirement of a primer 3' overhang." Proc. Natl. Acad. Sci. USA **93**: 10712-10717.

Lingner, J., Cooper, J.P., and Cech, T.R. (1995). "Telomerase and DNA end replication: No longer a lagging strand problem." Science **269**(1533-34).

Lingner, J., Hughes, T.R., Schevchenko, A., Mann, M., Lundblad, V., and Cech, T.R. (1997). "Reverse transcriptase motifs in the catalytic subunit of telomerase." Science **276**: 561-67.

Lue, N. F., Lin, Y.C., and Mian, I.S. (2003). "A Conserved Telomerase Motif within the Catalytic Domain of Telomerase Reverse Transcriptase Is Specifically Required for Repeat Addition Processivity." Mol. Cell. Biol **23**: 8440-8449.

Ludwig, M. L., and Matthews, R.G. (1997). "Sructure-based perspectives on B_{12}-dependant enzymes." Annu. Rev. Biochem **66**: 269-313.

Lundblad, V. (2003). "Telomere replication: an Est fest." Curr. Biol. **13**: R439-R441.

Lundblad, V., and Blackburn, E.H. (1993). "An alternative pathway for yeast telomere maintenance rescues est-1 senescence." Cell **73**: 347-360.

Matthews, B. W. (1968). "Solvent content of crystals." J Mol Biol **2**: 491-497.

Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R., Jenkins, J. (1997). "Two crystal structures of pectinlyase A from *Aspergillus niger* reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases." Structure **5**: 677-689.

McFerrin, M. B., and Snell, E.H. (2002). "The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images." J. Appl. Cryst. 35: 538-545.

McGoldrick, H. M., Roessner, C.A., Raux, E., Lawrence, A.D., McLean, K.J., Munro, A.W., Santabarbara, S., Rigby, S.E.J., Heathcote, P., Scott, A.J., and Warren, M.J. (2005). "Identification and characterization of a novel vitamin B_{12} (cobalamin) biosynthetic enzyme (CobZ) from *Rhodobacter capsulatus*, containing flavin, heme, and Fe-S cofactors." J. Bio. Chem. **280**: 1086-1094.

Miller, M. C., and Collins, K. (2002). "Telomerase recognizes its template by using an adjacent RNA motif." Proc. Natl. Acad. Sci. USA **99**: 3585-6590.

Mohnen, D. (1999). Biosynthesis of pectins and galactomannans. Amsterdam, Elsevier Science.

Moriarty, T. J., Huard, S., Dupuis, S., and Autexier, C. (2002). "Functional multimerization of human telomerase requires an RNA interaction domain in the N-terminus of the catalytic subunit." Mol. Cell. Biol **22**: 1253-1265.

Moriarty, T. J., Marie-Egyptienne, D.T., and Autexier, C. (2004). "Functional organization of repeat addition processivity and DNA synthesis determinants in the human telomerase multimer ." Mol. Cell. Biol **24**: 3720-3733.

Morin, G. B. (1991). "Recognition of a chromosome truncation site associated with α-thalassaemia by human telomerase." Nature **353**: 454-456.

Morris, R. J., Perrakis, A., and Lamzin, V.S. (2003). "ARP/wARP and automatic interpretation of protein electron density maps." Methods Enzymol **374**: 229-244.

Morris, A. L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. (1992). "Stereochemical quality of protein structure coordinates." Proteins **12**: 345.

Muller, G., Zipfel, F., Hilineny, K., Savvidis, E., Hertle, R., and Traub-Eberhard, U. (1991). "Timing of cobalt insertion in vitamin B_{12} biosynthesis." J. Am.Chem. Soc. **113**: 9893-9895.

Nakamura, T. M., Morin, G.B., Chapman, K.B., Weinrich, S.L., Andrews, W.H., Lingner, J., Harley, C.B., and Cech, T.R. (1997). "Telomerase catalytic subunit homologs from fission yeast amd human." Science **277**: 955-959.

Nocek, B., Zhou, M., Clancy, S., and Joachimiak, A. "Crystal structure of cobalamin synthesis related protein (CobF) from Corynebacterium diphtheriae."

O'Neill, M. A., Albersheim, P., and Darvill, A. (1990). The pectic polysaccharides of primary cell walls. London, Academic Press.

Padmanabhan, B., Bessho, Y., and Yokoyama, S. "Crystal Structure of Cobalamin biosynthesis precorrin 8W decarboxylase (cbiT)."

Panchev, I. N., Kirtchev, N.A., Kratchanov, C.G., and Proichev, T. (1988). "On the molecular weight of pectic substances and its relation to their gel strengths." Carbohydrate Polymers **8**: 257-269.

Patrick, N., Stamford, J., Crouzet, J., Cameron, B., Alanine, A.I.D., Pitt, A.R., Yeliseev, A.A., and Battersby, A.R. (1996). "Biosynthesis of vitamin B_{12}: the preparative multi-enzyme synthesis of precorrin-3a and 20-methylsirohydrochlorin (a 2,7,20-trimethylisobacteriochlorin)." Biochem. J. **313**: 335-342.

Peng, Y., Mian, S.I., and Lue, N.F. (2001). "Analysis of telomerase processivity: mechanistic similarity to HIV-1 reverse transcriptase and role in telomere maintenance." Mol. Cell. **7**: 1201-1211.

Petersen, T. N., Kauppinen, S., Larsen, S. (1997). "The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatu*s: a right-handed parallel beta-helix." Structure **5**: 533-544.

Peterson, S. E., Stellwagen, A.E., diede, S.J., Singer, M.S., Haimberger, Z.W., Johnson, C.O., Tzoneva, M., and Gottschling, D.E. (2001). "The function of a stem-loop in telomerase RNA is linked to the DNA repair protein Ku." Nat. Genet. **27**: 64-67.

Pickersgill, R., Jenkins, J., Harris, G., Nasser, W., and Robert-Baudouy, J. (1994). "The structure of *Bacillus subtilis* pectate lyase in complex with calcium." Nat. Struct. Biol **1**: 717-723.

Pickersgill, R., Smith, D., Worboys, K., Jenkins, J. (1998). "Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. carotovora." J. Biol. Chem. **273**: 24660-24664.

Raux, E., Lanois, A., Levillayer, F., Warren, M.J., Brody, E., Rambach, A., and Thermes, C. (1996). "*Salmonella typhimurium* Cobalamin (Vitamin B12) Biosynthetic Genes:Functional Studies in *S. typhimurium* and *Escherichia coli*." J. Bacteriol **178**(3): 753.

Raux, E., Leech, H.K., Beck, R., Schubert, H.L., Santander, P.J., Roessner, C.A., Scott, A.I., Martens, J.H., Jahns, D., Thermes, C., Rambach, A., and Warren, M.J. (2003). "Identification and functional analysis of enzymes required for precorrin-2 dehydrogenation and metal ion insertion in the biosynthesis of sirohaem and cobalamin in *Bacillus megaterium*." Biochem. J **370**: 505-516.

Rhodes, G. (2006). Crystallogrgaphy made crystal clear, Elsevier.

Richards, R. J., Theimer, C.A., Finger, L.D., and Feigon, J. (2006a). "Structure of the *Tetrahymena thermophila* telomerase RNA helix II template boundary element." Nucleic Acids Research **34**: 816-825.

Rexova-Benkova, L., and Marcovic, O. (1976). "Pectic enzymes." Adv. Carbohydr. Chem. Biochem. **33**: 323-385.

Richards, R. J., Wu, H., Trantirek, L., O'Connor, C.M., collins, K., and Feigon, J. (2006b). "Structural study of elements of Tetrahymena telomerase RNA stem-loop IV domain important for function." RNA **12**: 1475-1485.

Rickes, E. L., Brink, N.G., Koniuszy, F.R., Wood, T.R., and Folkers, K. (1948). "Crystalline Vitamin B_{12}." Science **107**: 396-397.

Romero, D. P., and Blackburn, E.H. (1991). "A conserved secondary structure for telomerase RNA." Cell **67**: 343-353.

Roth, J. R., Lawrence, J.G., and Bobik, T.A. (1996). "Cobalamin (coenzyme B_{12}): Synthesis and biological significance." Annu. Rev. Microbiol. **50**: 137-181.

Roth, J. R., Lawrence, J.G., Rubenfield, M., Kieffer-Higgins, S., and Church, G.M. (1993). "Characterization of the cobalamin (Vitamin B_{12}) biosynthetic genes of *Salmonella typhimurium*." J. Bacteriol **175**(11): 3303-3316.

Rouda, S., and Skordalakes, E. (2007). "Structure and implications for RNA recognition and binding." Structure. **15**: 1403-1412.

Santander, P. J., Roessner, C.A., Stolowich, N.J., Holderman, M.T., and Scott, A.I. (1997). "How corrinoids are synthesized without oxygen: nature's first pathway to vitamin B_{12}." Chemistry and Biology **4**: 659-666.

Scavetta, R. D., Herron, S.R., Hotchkiss, A.T., Kita, N., Keen, N.T., Benen, J.A.E., Kester, H.C.M., Visser, J., and Jurnak, F. (1999). "Structure of a plant cell wall fragment complexed to pectate lyase C." The Plant Cell **11**: 1081-1092.

Schneider, Z., and Stroinski, A. (1987). Methylcobamide-dependent reactions., Berlin: Gruyter.

Schroeder, S., Lawrence, A.D., Biendendieck, R., Rose, R.S., Deery, E., Graham, R.M., McLean, K.J., Munro, A.W., Rigby, S.E.J., and Warren, M.J. (2009). "Demonstration that CobG, the monooxygenase associated with the ring contraction process of the aerobic cobalamin (vitamin B_{12}) biosynthetic pathway, contains an Fe-S center and a mononuclear non-heme iron center." The Journal of Biological Chemistry **284**(8): 4796-4805.

Schubert, H. L., Wilson, K.S., Raux, E., Woodcock, S.C., and Warren, M.J. (1998). "The X-ray structure of a cobalamin biosynthetic enzyme, cobalt-precorrin-4 methyltransferase." Nat. Struct. Biol **5**: 585-592.

Scott, A. I. (2003). "Discovering nature's diverse pathways to Vitamin B_{12}: A 35-year odyssey." J.Org.Chem **68**: 2529-2539.

Scott, A. I., and Roessner, C.A. (2002). "Biosynthesis of cobalamin (vitamin B_{12})." Bio. Soc. Trans. **30**: 613-620.

Scott, A. I., Roessner, C.A.,Stolowich, N. J., Spencer, J. B., Min, C., and Ozaki, S. I. (1993). "Biosynthesis of vitamin $B_{12.}$ Discovery of the enzymes for oxidative ring contraction and insertion of the fourth methyl group." FEBS Lett **331**(12): 105-108.

Seyedarabi, A., Hutchison, T., To, T.T., Deery, E., Brindley, A., Warren, M.J., Pickersgill, R.W. (2010a). "Cloning, purification and preliminary crystallographic analysis of cobalamin methyltransferases from *Rhodobacter capsulatus*." Acta Cryst.

Seyedarabi, A., To, T.T., Ali, S., Hussain, S., Fries, M., Madsen, R., Clausen, M.H., Teixteira, S., Brocklehurst, K., and Pickersgill, R.W. (2010b). "Structural insights into substrate specificity and the *anti* β-elimination mechansim of pectate lyase." Biochemistry **49**: 539-546.

Shay, J. W., and Bacchetti, S. (1997). "A survey of telomerase activity in human cancer." Eur. J. Cancer. **33**: 787-791.

Shay, J. W., and Wright, W.E. (2002). "Telomerase: a target for cancer therapeutics." Cancer Cell **2**: 257-265.

Shipman, L. W., Li, D., Roessner, C.A., Scott, A.I., and Sacchettini, J.C. (2001). "Crystal structure of Precorrin-8x methyl mutase." Structure **9**: 587-596.

Shpaer, E. G., Robinson, M., Yee, D., Candlin, J.D., Mines, R., and Hunkapiller, T. (1996). "Sensitivity and selectivity in protein similarity searches: Comparison of Smith-Waterman in hardware." Genomics **38**: 179-191.

Sigler, P. B., Xu, Z., Rye, H.S., Burston, S.B., Fenton, W.A., and Horwich, A.L. (1998). "Structure and function in GroEL-mediated protein folding." Annu. Rev. Biochem. **67**: 581-608.

Smith, E. L. (1948). "Purification of anti-pernicious anaemia factors from liver." Nature **161**: 638-639.

Sperger, J. M., and Cech, T.R. (2001). "A stem-loop of Tetrahymena telomerase RNA distant from the template potentiates RNA folding and telomerase activity." Biochemistry **40**: 7005-7016.

Starr, M. E., and Morán, P. (1962). "Eliminative split of pectic substances by phytopathogenic soft-rot bacteria." Science **135**(920-921).

Stroupe, M. E., Leech, H.K., Daniels, D.S., Warren, M.J., and Getzoff, E.D. (2003). "CysG structure reveals tetrapyrrole-binding features and novel regulation of siroheme biosynthesis." Nat. Struct. Biol **12**: 1064-1073.

Tesmer, V. M., Ford, L.P., Holt, S.E., Frank, B.C., Yi, X., Aisner, D.L.,Ouellette, M., Shay, J.W., and Wright, W.E. (1999). "Two inactive fragments of the integral RNA cooperate to assemble active telomerase with the human protein catalytic subunit (hTERT) *in vitro*." Mol. Cell. Biol **19**: 6207-6216.

Theimer, C. A., Blois, C.A., and Feigon, J. (2005). "Structure of the human telomerase RNA pseudoknot reveals conderved tertiary interactions essential for function." Mol. Cell. **17**: 671-682.

Thibaut, D., Couder, M., Crouzet, J., Debussche, L., Cameron, B., and Blanche, F. (1990). "Assay and purfication of S-adenosyl-L-methionine: precorrin-2 methyltransferase from *Pseudomonas denitrificans*." J. Bacteriol **172**: 6245-6251.

Thomas, L. M., Doan, C.N., Oliver, R.L., and Yoder, M.D. (2002). "Structure of pectate lyase A: comparison to other isoforms." Acta. Crystallogr. Sect. D. **58**: 1008-1015.

Vagin, A. A., Steiner, R.S., Lebedev, A.A., Potterton, L., McNicholas, S., Long, F., and Murshudov, G.N. (2004). "REFMAC5 dictionary: organisation of prior chemical knowledge and guidelines for its use." Acta Cryst 60: 2284-2295.

Vagin, A., and Teplyakov, A. (1997). "MOLREP: an automated program for molecular replacement." J. Appl. Cryst. **30**: 1022-1025.

Van Santen, Y., Benen, J.A.E., Schroter, K-H., Kalk, K.H., Armand, S., Visser, J., Dijksta, B.W. (1999). "1.68 Å crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis." J. Biol. Chem. **274**: 30474-30480.

Van steensel, B., Smogorzewska, A., and de Lange, T. (1998). "TRF2 protects human telomeres from end-to-end fusions." Cell **92**: 401-413.

Vevodova, J., Graham, R.M., Raux, E., Schubert, H.L., Roper, D.L., Brindley, A.A., Scott, A.I., Roessner, C.A., Stamford, N.P., Stroupe, M.E., Getzoff, E.D., Warren, M.J.,

and Wilson, K.S. (2004). "Structure/Function Studies on a S-Adenosyl-l-methionine-dependent Uroporphyrinogen III C Methyltransferase (SUMT), a Key Regulatory Enzyme of Tetrapyrrole Biosynthesis." J. Mol. Biol. **344**: 419-433.

Vidal, S., Doco, T., Williams, P., Pellerin, P., York, W.S., O'Neill, M.A., (2000). "Structural characterization of the pectic polysaccharide rhamnogalacturonan II: evidence for the backbone location of the acid-containing oligoglycosyl side chain." Carbohydrate Res. **326**: 277-294.

Vincken, J. P., Schols, H.A., Oomen, R.J.F.J., McCann, M.C., Ulvskov, P., Voragen, A.G.J., and Visser, R.G.F. (2003). "If homogalacturonan were a side chain of rhamnogalacturonan I. Implications for cell wall architecture." Plant Physiology **132**: 1781-1789.

Vitali, J., Schick, B., Kester, H.C.M., Visser, J., Jurnak, F. (1998). "The three-dimensional structure of *Aspergillus niger* pectinlyase B at 1.7 (A resolution." Plant Physiol **116**: 69-80.

Vlcek, C., Paces, V., Maltsev, N., Paces, J., Haselkorn, R., and Fonstein, M. (1997). "Sequence of a 189-kb segment of the chromosome of *Rhodobacter capsulatus* SB1003." Proc. Natl. Acad. Sci. **94**: 9384-9388.

Wang, H., Giley, D., Blackburn, E.H. (1998). "A novel specificity for the primer-template pairing requirement in *Tetrahymena* telomerase." EMBO **17**: 1152-1160.

Warren, M. J., and Scott, A.I. (1990). "Tetrapyrrole assembly and modification into the ligands of biologically functional cofactors." Trends Biochem. Sci **15**: 486-491.

Warren, M. J., Raux, E., Schubert, H.L., and Escalante-Semerena, J.C. (2002). "The biosynthesis of adenosylcobalamin (vitamin B_{12})." Nat. Prod. Rep. **19**: 390-412.

Watson, J. D. (1972). "Origin of concatemeric T7 DNA." Nat. New. Biol. **94**: 197-201.

Williams, S. (2009). "Glycoside hydrolases." From http://www.cazypedia.org/index.php/File:Inverting_glucosidase_mechanism.png.

Wright, W. E., Piatyszek, M.A., Rainey, W.E., Byrd, W., and Shay, J.W. (1996). "Telomerase activity in human germline and embryonic tissues." Dev. Genet. **18**: 173-179.

Xie, M., Mosig, A., Qi, X., Li, Y., Stadler, P.F., and Chen, J.L. (2008). "Structure and function of the smallest vertebrate telomerase RNA from teleost fish." J. Biol. Chem **283**(4).

Yoder, M. D., Keen, N.T., and Jurnak, F. (1993). "New domain motif: The structure of pectate lyase C, a secreted plant virulence factor." Science **260**: 1503-1507.

Yu, G. L., Bradley, J.D., Attardi, L.D., and Blackburn, E.H. (2001). "In *vivo* alteration of telomere sequences and senescence caused by mutated *Tetrahymena* telomerase RNAs." Nature **344**: 126-132.

Zapulla, D. C., Goodrich, K., and Cech, T.R. (2005). "A miniature yeast telomerase RNA functions in *vivo* and reconstitutes activity in *vitro*." Nat. Struct. Mol. Biol **12**: 1072-1077.

Zhan, D., Janssen, P., and Mort, A.J. (1998). "Scarcity or complete lack of single rhamnose residues interspersed within the homogalacturonan regions of citrus pectin." Carbohydrate Res. **308**: 373-380.

Zhang, R., Skarina, T., Savchenko, A., Edwards, A., and Joachimiak, A. "Structural Genomics, 1.9A crystal structure of cobalamin biosynthesis protein (cbiD) from *Archaeoglobus fulgidus*."

Appendices

Appendix 1

The vectors; pET3d, pET14b, pET41 and pETcoco2 used throughout this thesis are shown in this Appendix, all figures were from Novagen.

pET-3a-d Vectors

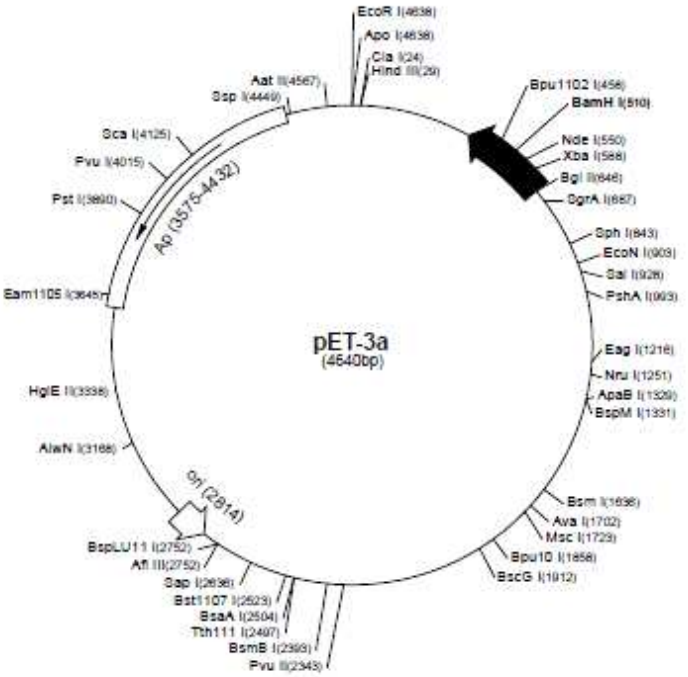
TB026 12/98

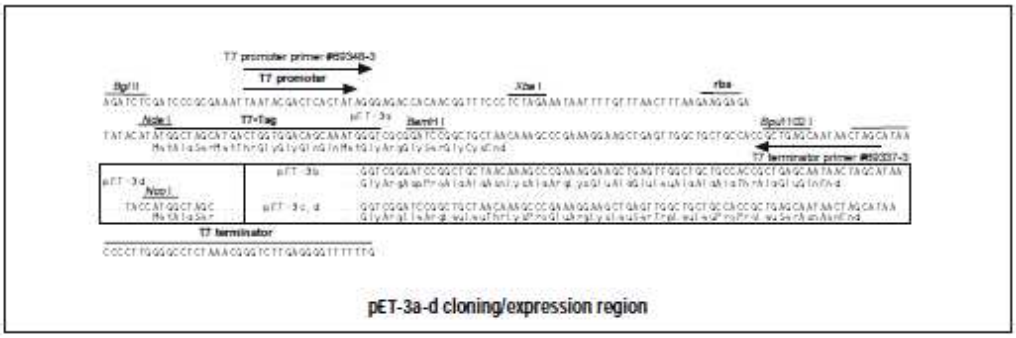
	Cat. No.
pET-3a DNA	69418-3
pET-3b DNA	69419-3
pET-3c DNA	69420-3
pET-3d DNA	69421-3

The pET-3a-d vectors carry an N-terminal T7•Tag® sequence and *Bam*H I cloning site. These vectors are the precursors to many pET family vectors; the pET-23a-d(+) series corresponds to pET-3a-d but incorporates several additional features. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below.

pET-3a sequence landmarks	
T7 promoter	615-631
T7 transcription start	614
T7•Tag coding sequence	519-551
T7 terminator	404-450
pBR322 origin	2814
bla coding sequence	3575-4432

The maps for pET-3b, pET-3c and pET-3d are the same as pET-3a (shown) with the following exceptions: pET-3b is a 4639bp plasmid; subtract 1bp from each site beyond *Bam*H I at 510. pET-3c is a 4638bp plasmid; subtract 2bp from each site beyond *Bam*H I at 510. pET-3d is a 4637bp plasmid; the *Bam*H I site is in the same reading frame as in pET-3c. An *Nco* I site is substituted for the *Nde* I site with a net 1bp deletion at position 550 of pET-3c. As a result, *Nco* I cuts pET-3d at 546. For the rest of the sites, subtract 3bp from each site beyond position 551 in pET-3a. *Nde* I does not cut pET-3d.





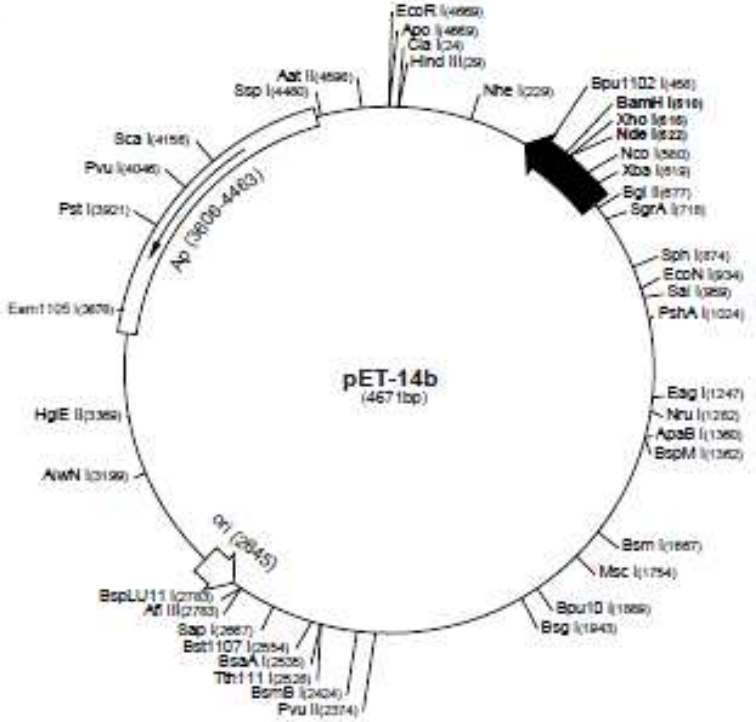
pET-3a-d cloning/expression region

pET-14b Vector

TB044 12/98

The pET-14b vector (Cat. No. 69660-3) carries an N-terminal His•Tag® sequence followed by a thrombin site and three cloning sites. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circular map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below.

pET-14b sequence landmarks	
T7 promoter	646-662
T7 transcription start	645
His•Tag coding sequence	554-571
Multiple cloning sites	
(*Nde* I-*BamH* I)	510-526
T7 terminator	404-450
pBR322 origin	2845
Amp coding sequence	3606-4463



pET-14b cloning/expression region

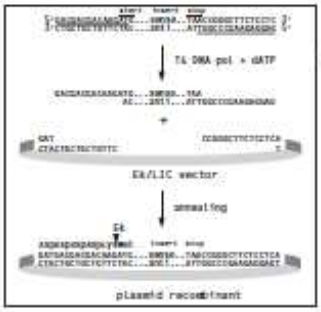
pET–41 Ek/LIC Vector

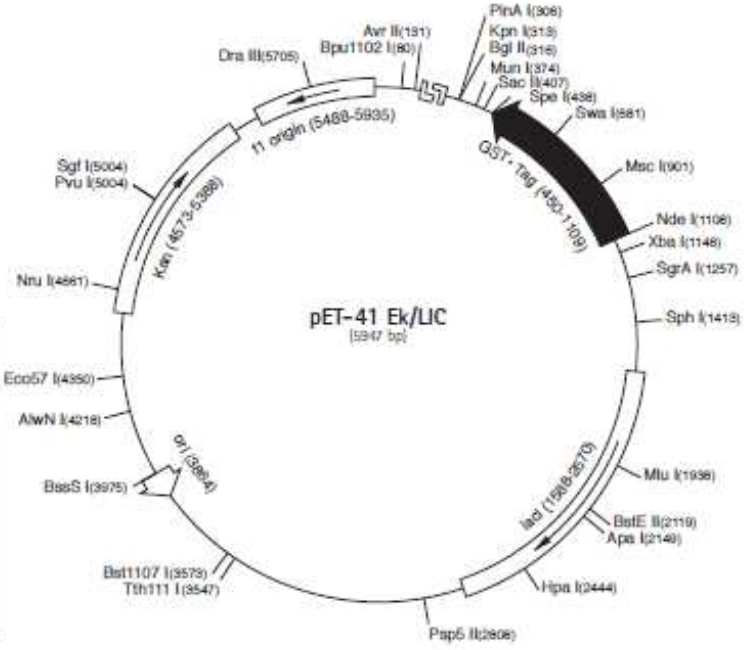
TB317 0806

	Cat No.
pET–41 Ek/LIC Vector Kit	71071-3

pET–41 Ek/LIC sequence landmarks	
T7 promoter	1181–1197
T7 transcription start	1180
GST•Tag coding sequence	450–1109
His•Tag coding sequence	411–428
S•Tag coding sequence	324–368
Multiple cloning sites (BseR I-Xho I)	174–277
His•Tag coding sequence	150–173
T7 terminator	26–72
lacI coding sequence	1588–2670
pBR322 origin	3864
Kan coding sequence	4573–5388
f1 origin	5935–5488

The pET–41 Ek/LIC vector is prepared for rapid, directional cloning of PCR-amplified DNA for high-level expression of polypeptides fused with N-terminal GST•Tag™, His•Tag® and S•Tag™ sequences. Using specifically designed primers for amplification and the pET–41 Ek/LIC Vector Kit (Cat. No. 71071-3), inserts can be efficiently cloned without the need for restriction digestion or ligation. Unique sites are shown on the circle map. Note that the sequence is numbered by the pBR322 convention, so the T7 expression region is reversed on the circle map. The cloning/expression region of the coding strand transcribed by T7 RNA polymerase is shown below. The f1 origin is oriented so that infection with helper phage will produce virions containing single stranded DNA that corresponds to the coding strand. Therefore, single stranded sequencing should be performed using the T7 terminator primer (Cat. No. 69337-3). Vector encoded sequence can be completely removed when cloning into the Ek/LIC site (as shown below left) by cleaving the fusion protein with enterokinase.



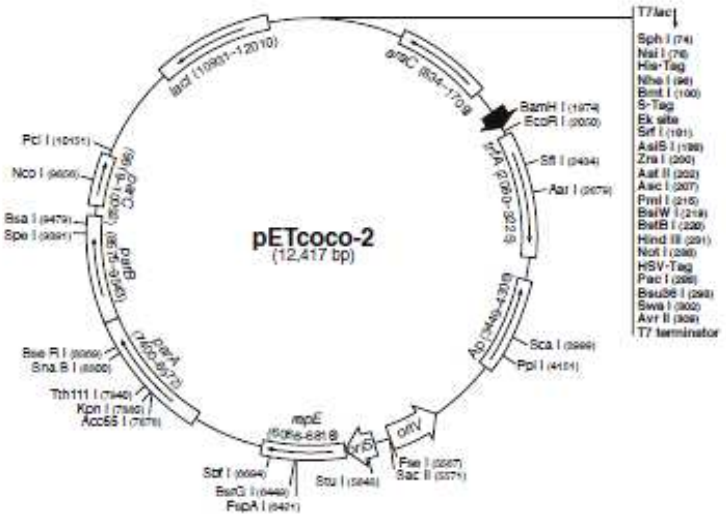


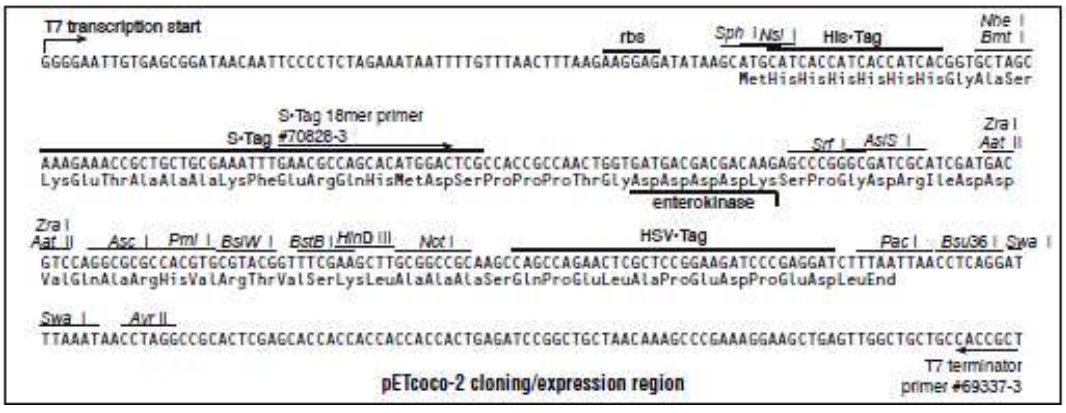
pET–41 Ek/LIC cloning/expression regions

pETcoco-2 Vector

	Cat. No.
pETcoco-2 DNA	71148-3
pETcoco-2 sequence landmarks	
T7 promoter	12401–12417
lac operator	5–25
T7 transcription start	1
Multiple cloning sites (*Sph* I–*Avr* II)	74–309
His•Tag® coding sequence	75–92
S•Tag™ coding sequence	102–146
HSV•Tag® coding sequence	246–281
T7 terminator	410–457
ampC coding sequence	834–1709
trfA coding sequence	2080–3225
Ap (*bla*) coding sequence	3449–4306
oriV	5040–5565
oriS	5756–5972
repE coding sequence	6066–6818
parA coding sequence	7400–8572
parB coding sequence	8575–8543
parC region	9619–10092
lacI coding sequence	10031–12010

The pETcoco™ expression vectors are designed to allow "on command" amplification of vector copy number from single copy (1–2 plasmid copies per cell) to medium copy (20–50 per cell). The vector is maintained in the single copy state by utilizing the *oriS* and *repE* elements of the F episome together with the *parABC* partition determinants. The medium copy state is obtained by employing the RK-2 derived *trfA* replicator acting at *oriV*. The switch from single copy to medium copy status is achieved by inducing the expression of the *trfA* gene with arabinose. The dual replicon nature of these vectors promotes enhanced stability of recombinant plasmids and extremely low basal expression levels in the single copy state, while allowing robust target protein expression typical of the pET vectors. Unique restriction sites are shown on the circle map (below). The sequence is numbered relative to the start of T7 transcription at +1.





Appendix 2

Sequence alignment of pET41-DY3.4 construct using five different primers Int2, Int3 forward primers, Int4, Int5 and Int6 reverse primers to ensure the entire gene sequence was covered in the analysis. Below shows the sequence alignment using Int2 primer.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   785        795        805        815        825        835
DY3-4-Int2      GCTCTCAATT TCTCGTCAAA AATTGCGACC TTTGTTCAAA AAAAAAGCAA TCGATAAAAA
DY3.4           gctctcaatt tctcgtcaaa aattgcgacc tttgttcaaa agaaaagcaa tcgataaaaa

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   845        855        865        875        885        895
DY3-4-Int2      AACAGAGACA ATGCAATGGA AAAGGTTGAA TTCGATGCTC TCATGGTGTT TGGAAAGAAG
DY3.4           aacagagaca atgcaatgga aaaggttgaa ttcgatgctc tcatggtgtt tggaaagaag

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   905        915        925        935        945        955
DY3-4-Int2      TGGAGTTTAC CGACATACAA TCAGAGATTC GTGTAAAAAA GTGTCAGATT TTCTAAAGAA
DY3.4           tggagtttac cgacatacaa tcagagattc gtgtaaaaaa gtgtcagatt ttctaaagaa

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   965        975        985        995       1005       1015
DY3-4-Int2      GAACTCACAA AACTCAAAAA TAATTGGATA CACAGCAGAT GTTAGCAAAT TTTTCTCAAC
DY3.4           gaactcacaa aactcaaaaa taattggata cacagcagat gttagcaaat gtttctcaac

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1025       1035       1045       1055       1065       1075
DY3-4-Int2      TGTTAATCAT GACGTGTTGA TATCAATTAT TGACCGACTT TTTTCTCAAG AGCACGATAT
DY3.4           tgttaatcat gacgtgttga tatcaattat tgaccgactt ttttctcaag agcacgatat

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1085       1095       1105       1115       1125       1135
DY3-4-Int2      TTATACGGTA TGTGGAAAAG GAAGAAATCA CGGAGGATTT CATAAATTGA TATTCTGTTC
DY3.4           ttatacggta tgtggaaaag gaagaaatca cggaggattt cataaattga tattctgttc

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1145       1155       1165       1175       1185       1195
DY3-4-Int2      AGCTGGAACT GAGTTGAATG CACATGAGGC GCTTCGTCGA AAAATGGAAT TGAAAGGAGT
DY3.4           agctggaact gagttgaatg cacatgaggc gcttcgtcga aaaatggaat tgaaaggagt

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1205       1215       1225       1235       1245       1255
DY3-4-Int2      ATTCAATTTT GAAGTTTGCC ACCGAGAGAT GAGTTCCTCT ACAACACTCT ACAGTGTTAT
DY3.4           attcaatttt gaagtttgct accgagagat gagttcctct acaacactct acagtgttat

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1265       1275       1285       1295       1305       1315
DY3-4-Int2      TCGTACCACA CTTTCAACGT ACTACTACAA GCGTGGCCCA ACCTCGGGGG AGAATTACAA
DY3.4           tcgtaccaca ctttcaacgt actactacaa gcgtggccca acatcgtgg- agaattacaa

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1325       1335       1345       1355       1365       1375
DY3-4-Int2      AAGGAGTTCC ACAGGGACAT CCTATATCTT CAAATTTAGC ACATATGCAC CTCAATAACT
DY3.4           aaggagttcc acagggacat cctatatctt caaatttagc acatatgtac ctcaataact

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1385       1395       1405       1415       1425       1435
DY3-4-Int2      TTGAGCAGAA ATATTGGAGC AAAGAAAAAG AGGATTCGAG AATTGTTTTC TGCAGATATG
DY3.4           ttgagcagaa atattggagc aacgaaaaag aggattcgag aattgttttc tgcagatatg

                ....|....|
                  1445

DY3-4-Int2      AGGATGATTT
DY3.4           aggatgattt
```

Sequence analysis of pET41-DY3.4 construct using Int3 forward primer.

```
             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                1205       1215       1225       1235       1245       1255
DY3-4-Int3   ---------- ---------- ---------- ---------- ----AATCC- CGGTGTTATT
DY3.4        ttcaattttg aagtttgcta ccgagagatg agttcctcta caacactcta cagtgttatt

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                1265       1275       1285       1295       1305       1315
DY3-4-Int3   TCGTACCACA CTTTCA-CGT ACTACTACAA GCGTGGCCCA ACATCGTGGA GAATTACAAA
DY3.4        -cgtaccaca ctttcaacgt actactacaa gcgtggccca acatcgtgga gaattacaaa

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                1325       1335       1345       1355       1365       1375
DY3-4-Int3   AGGAGTTCCA CAGGGACATC CTATATCTTC AAATTTAGCA CATATGTACC TCAATAACTT
DY3.4        aggagttcca cagggacatc ctatatcttc aaatttagca catatgtacc tcaataactt

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                1385       1395       1405       1415       1425       1435
DY3-4-Int3   TGAGCAGAAA TATTGGAGCA ACGAAAAAGA GGATTCGAGA ATTGTTTTCT GCAGATATGA
DY3.4        tgagcagaaa tattggagca acgaaaaaga ggattcgaga attgttttct gcagatatga

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                1445       1455       1465       1475       1485       1495
DY3-4-Int3   GGATGATTTC ATTTTCATTA CAACTGAAAA TTCTTTATTC GAGAAGATGA TGAAACCATT
DY3.4        ggatgatttc attttcatta caactgaaaa ttctttattc gagaagatga tgaaaccatt

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                1505       1515       1525       1535       1545       1555
DY3-4-Int3   ATCTACTGGC AATAACACTC ATTTTTTGAC GGCTAATCCG AAAAAGTTCA AGAAATCAGA
DY3.4        atctactggc aataacactc atttttgac ggctaatccg aaaaagttca agaaatcaga

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                1565       1575       1585       1595       1605       1615
DY3-4-Int3   GCGATGTGGA GCATCACAAG TCCTTCAATG GTGTGGAGTG AAACTGGATT TTCAATCGGG
DY3.4        gcgatgtgga gcatcacaag tccttcaatg gtgtggagtg aaactggatt ttcaatcggg

             ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                1625       1635       1645       1655       1665       1675
DY3-4-Int3   AAATTGCTTT ATTCGACGAA GATGCAAAGA CGGTGTGGCT CGTCAATTTT TGATAAAGTT
DY3.4        aaattgcttt attcgacgaa gatgcaaaga cggtgtggct cgtcaatttt tgataaagtt

             ....|....|
                1685
DY3-4-Int3   GCAATAAACC
DY3.4        gcaataa---
```

Sequence analysis of pET41-DY3.4 construct using Int4 reverse primer.

```
                210       220       230       240       250       260
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      CCACACCGTCTTTGCATCTTCGTCGAATAAAGCAATTTCCCGATTGAAAATCCAGTTTCA
DY3.4           CCACACCGTCTTTGCATCTTCGTCGAATAAAGCAATTTCCCGATTGAAAATCCAGTTTCA

                270       280       290       300       310       320
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      CTCCACACCATTGAAGGACTTGTGATGCTCCACATCGCTCTGATTTCTTGAACTTTTTCG
DY3.4           CTCCACACCATTGAAGGACTTGTGATGCTCCACATCGCTCTGATTTCTTGAACTTTTTCG

                330       340       350       360       370       380
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      GATTAGCCGTCAAAAAATGAGTGTTATTGCCAGTAGATAATGGTTTCATCATCTTCTCGA
DY3.4           GATTAGCCGTCAAAAAATGAGTGTTATTGCCAGTAGATAATGGTTTCATCATCTTCTCGA

                390       400       410       420       430       440
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      ATAAAGAATTTTCAGTTGTAATGAAAATGAAATCATCCTCATATCTGCAGAAAACAATTC
DY3.4           ATAAAGAATTTTCAGTTGTAATGAAAATGAAATCATCCTCATATCTGCAGAAAACAATTC

                450       460       470       480       490       500
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      TCGAATCCTCTTTTTCGTTGCTCCAATATTTCTGCTCAAAGTTATTGAGGTACATATGTG
DY3.4           TCGAATCCTCTTTTTCGTTGCTCCAATATTTCTGCTCAAAGTTATTGAGGTACATATGTG

                510       520       530       540       550       560
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      CTAAATTTGAAGATATAGGATGTCCCTGTGGAACTCCTTTTGTAATTCTCCACGATGTTG
DY3.4           CTAAATTTGAAGATATAGGATGTCCCTGTGGAACTCCTTTTGTAATTCTCCACGATGTTG

                570       580       590       600       610       620
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      GGCCACGCTTGTAGTAGTACGTTGAAAGTGTGGTACGAATAACACTGTAGAGTGTTGTAG
DY3.4           GGCCACGCTTGTAGTAGTACGTTGAAAGTGTGGTACGAATAACACTGTAGAGTGTTGTAG

                630       640       650       660       670       680
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      AGGAACTCATCTCTCGGTAGCAAACTTCAAAATTGAATACTCCTTTCAATTCCATTTTTC
DY3.4           AGGAACTCATCTCTCGGTAGCAAACTTCAAAATTGAATACTCCTTTCAATTCCATTTTTC

                690       700       710       720       730       740
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      GACGAAGCGCCTCATGTGCATTCAACTCAGTTCCAGCTGAACAGAATATCAATTTATGAA
DY3.4           GACGAAGCGCCTCATGTGCATTCAACTCAGTTCCAGCTGAACAGAATATCAATTTATGAA

                750       760       770       780       790       800
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      ATCCTCCGTGATTTCTTCCTTTTCCACATACCGTATAAATATCGTGCTCTTGAGAAAAAA
DY3.4           ATCCTCCGTGATTTCTTCCTTTTCCACATACCGTATAAATATCGTGCTCTTGAGAAAAAA

                810       820       830       840       850       860
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      GTCGGTCAATAATTGATATCAACACGTCATGATTAACAGTTGAGAAACATTTGCTAACAT
DY3.4           GTCGGTCAATAATTGATATCAACACGTCATGATTAACAGTTGAGAAACATTTGCTAACAT

                870       880       890       900       910       920
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      CTGCTGTGTATCCAATTATTTTTGAGTTTTGTGAGTTCTTCTTTAGAAAATCTGACACTT
DY3.4           CTGCTGTGTATCCAATTATTTTTGAGTTTTGTGAGTTCTTCTTTAGAAAATCTGACACTT

                930       940       950       960       970       980
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-4-Int4      TTTTACACGAATCTCTGATTGTATGTCGGTAAACTCCACTTCTTTCCAAACACCATGAGA
DY3.4           TTTTACACGAATCTCTGATTGTATGTCGGTAAACTCCACTTCTTTCCAAACACCATGAGA

                990       1000      1010
                ....|....|....|....|....|....|
DY3-4-Int4      GCATCGAATTCAACCTTTTCCATTGCATTG
DY3.4           GCATCGAATTCAACCTTTTCCATTGCATTG
```

Sequence analysis of pET41-DY3.4 construct using Int5 reverse primer.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   905        915        925        935        945        955
DY3-4-Int5      GAGAAATTGA GAGCTTGAAG GTGGCAACAT TCGGTCTTAT AAAATGCGGA GCTTTTTGAA
DY3.4 rev       GAGAAATTGA GAGCTTGAAG GTGGCAACAT TCGGTCTTAT AAAATGCGGA GCTTTTTGAA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   965        975        985        995       1005       1015
DY3-4-Int5      CTATATAACG TTGTTTGAAA TCCTTTATTT CTTTTGTTAA AATATTTAAA TATCCATCTC
DY3.4 rev       CTATATAACG TTGTTTGAAA TCCTTTATTT CTTTTGTTAA AATATTTAAA TATCCATCTC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1025       1035       1045       1055       1065       1075
DY3-4-Int5      TCCACAGCAA AACTCTTTCT TCTTTAATTA CAATTGGAAT CATAACTTGT CGTAAAGCCA
DY3.4 rev       TCCACAGCAA AACTCTTTCT TCTTTAATTA CAATTGGAAT CATAACTTGT CGTAAAGCCA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1085       1095       1105       1115       1125       1135
DY3-4-Int5      TCAGAGCCCA CCAGTGTACT CCTGATTTTA AATTATTCAC AATTGAAGTT TGAATATTCC
DY3.4 rev       TCAGAGCCCA CCAGTGTACT CCTGATTTTA AATTATTCAC AATTGAAGTT TGAATATTCC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1145       1155       1165       1175       1185       1195
DY3-4-Int5      TCTTCACATT TGGATCCACA TTTTTCATAA CTTTAGATGC ACGAAATGCG CGAGTGAAAG
DY3.4 rev       TCTTCACATT TGGATCCACA TTTTTCATAA CTTTAGATGC ACGAAATGCG CGAGTGAAAG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1205       1215       1225       1235       1245       1255
DY3-4-Int5      GAGCTTCTTT GTAGACAAAT GGCTTCATTT TATGATAACC ACCATTGTTT AACATTTTTG
DY3.4 rev       GAGCTTCTTT GTAGACAAAT GGCTTCATTT TATGATAACC ACCATTGTTT AACATTTTTG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1265       1275       1285       1295       1305       1315
DY3-4-Int5      TAATCCACTC CAGTATATTT GCACAATTAG TTGATCCAAT ATATTTTTCC AGATGAAACA
DY3.4 rev       TAATCCACTC CAGTATATTT GCACAATTAG TTGATCCAAT ATATTTTTCC AGATGAAACA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1325       1335       1345       1355       1365       1375
DY3-4-Int5      TTGAAATTAA TTTGGAGCAT GCAACAAATG TATCATTTTT AGACACAAAA CAATGTTCGA
DY3.4 rev       TTGAAATTAA TTTGGAGCAT GCAACAAATG TATCATTTTT AGACACAAAA CAATGTTCGA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1385       1395       1405       1415       1425       1435
DY3-4-Int5      TTTCAACCAT TTTTCGTGTT GGTTTTAGTT TTCTCAGCTC AAAAATATTG CGAGAAGACA
DY3.4 rev       TTTCAACCAT TTTTCGTGTT GGTTTTAGTT TTCTCAGCTC AAAAATATTG CGAGAAGACA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1445       1455       1465       1475       1485       1495
DY3-4-Int5      GGATACTCTG CGATAATTTG TTGATTTCCA GTAAAACAAC GGTATCGCGA ATATTTGAAA
DY3.4 rev       GGATACTCTG CGATAATTTG TTGATTTCCA GTAAAACAAC GGTATCGCGA ATATTTGAAA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1505       1515       1525       1535       1545       1555
DY3-4-Int5      TGTTTTCAGA AGAAGCCAAA AAAGTGGACA TATGGCGGTA GTTTAATTTT TGTCTTCGTT
DY3.4 rev       TGTTTTCAGA AGAAGCCAAA AAAGTGGACA TATGGCGGTA GTTTAATTTT TGTCTTCGTT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1565       1575       1585       1595       1605       1615
DY3-4-Int5      TCTTTTTGAA AATATTTCTG ACGAGACTTA AATGAGCACT TTTGCGAATA ATTGTGTTCC
DY3.4 rev       TCTTTTTGAA AATATTTCTG ACGAGACTTA AATGAGCACT TTTGCGAATA ATTGTGTTCC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1625       1635       1645       1655       1665       1675
DY3-4-Int5      AATGACGTCT TTTATCGTAT TTTCGTATTT CTTCATGAAG TGTTGTAAGT GAACTCTTAA
DY3.4 rev       AATGACGTCT TTTATCGTAT TTTCGTATTT CTTCATGAAG TGTTGTAAGT GAACTCTTAA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1685       1695       1705       1715       1725       1735
DY3-4-Int5      TCGTTGGTGC CATCTTGTCG TCGTCATCAC CAGAACCACC ACCGGTACCC AGATCTGGGC
DY3.4 rev       TCGTTGGTGC CAT------- ---------- ---------- ---------- ----------
```

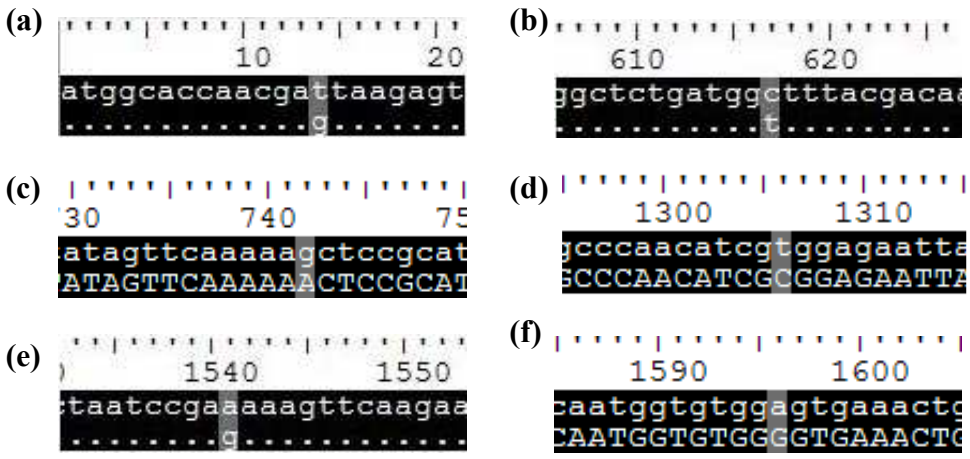
Sequence analysis of pET41-DY3.4 construct using Int6 reverse primer.

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   375        385        395        405        415        425
DY3-4-Int6      GATGGGGATT CTC-ACGATG TTGGGC-ACG CTTGTAGTAG TACGTTGAAA GTGTGGTACG
DY3.4 rev       TTTTGTAATT CTCCACGATG TTGGGCCACG CTTGTAGTAG TACGTTGAAA GTGTGGTACG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   435        445        455        465        475        485
DY3-4-Int6      AATAACACTG TAGAGTGTTG TAGAGGAACT CATCTCTCGG TAGCAAACTT CAAAATTGAA
DY3.4 rev       AATAACACTG TAGAGTGTTG TAGAGGAACT CATCTCTCGG TAGCAAACTT CAAAATTGAA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   495        505        515        525        535        545
DY3-4-Int6      TACTCCTTTC AATTCCATTT TTCGACGAAG CGCCTCATGT GCATTCAACT CAGTTCCAGC
DY3.4 rev       TACTCCTTTC AATTCCATTT TTCGACGAAG CGCCTCATGT GCATTCAACT CAGTTCCAGC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   555        565        575        585        595        605
DY3-4-Int6      TGAACAGAAT ATCAATTTAT GAAATCCTCC GTGATTTCTT CCTTTTCCAC ATACCGTATA
DY3.4 rev       TGAACAGAAT ATCAATTTAT GAAATCCTCC GTGATTTCTT CCTTTTCCAC ATACCGTATA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   615        625        635        645        655        665
DY3-4-Int6      AATATCGTGC TCTTGAGAAA AAAGTCGGTC AATAATTGAT ATCAACACGT CATGATTAAC
DY3.4 rev       AATATCGTGC TCTTGAGAAA AAAGTCGGTC AATAATTGAT ATCAACACGT CATGATTAAC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   675        685        695        705        715        725
DY3-4-Int6      AGTTGAGAAA CATTTGCTAA CATCTGCTGT GTATCCAATT ATTTTTGAGT TTTGTGAGTT
DY3.4 rev       AGTTGAGAAA CATTTGCTAA CATCTGCTGT GTATCCAATT ATTTTTGAGT TTTGTGAGTT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   735        745        755        765        775        785
DY3-4-Int6      CTTCTTTAGA AAATCTGACA CTTTTTTACA CGAATCTCTG ATTGTATGTC GGTAAACTCC
DY3.4 rev       CTTCTTTAGA AAATCTGACA CTTTTTTACA CGAATCTCTG ATTGTATGTC GGTAAACTCC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   795        805        815        825        835        845
DY3-4-Int6      ACTTCTTTCC AAACACCATG AGAGCATCGA ATTCAACCTT TTCCATTGCA TTGTCTCTGT
DY3.4 rev       ACTTCTTTCC AAACACCATG AGAGCATCGA ATTCAACCTT TTCCATTGCA TTGTCTCTGT

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   855        865        875        885        895        905
DY3-4-Int6      TTTTTTATCG ATTGCTTTTC TTTTGAACAA AGGTCGCAAT TTTTGACGAG AAATTGAGAG
DY3.4 rev       TTTTTTATCG ATTGCTTTTC TTTTGAACAA AGGTCGCAAT TTTTGACGAG AAATTGAGAG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   915        925        935        945        955        965
DY3-4-Int6      CTTGAAGGTG GCAACATTCG GTCTTATAAA ATGCGGAGCT TTTTGAACTA TATAACGTTG
DY3.4 rev       CTTGAAGGTG GCAACATTCG GTCTTATAAA ATGCGGAGCT TTTTGAACTA TATAACGTTG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   975        985        995       1005       1015       1025
DY3-4-Int6      TTTGAAATCC TTTATTTCTT TTGTTAAAAT ATTTAAATAT CCATCTCTCC ACAGCAAAAC
DY3.4 rev       TTTGAAATCC TTTATTTCTT TTGTTAAAAT ATTTAAATAT CCATCTCTCC ACAGCAAAAC

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1035       1045       1055       1065       1075       1085
DY3-4-Int6      TCTTTCTTCT TTAATTACAA TTGGAATCAT AACTTGTCGT AAAGCCATCA GAGCCCACCA
DY3.4 rev       TCTTTCTTCT TTAATTACAA TTGGAATCAT AACTTGTCGT AAAGCCATCA GAGCCCACCA

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1095       1105       1115       1125       1135       1145
DY3-4-Int6      GTGTACTCCT GATTTTAAAT TATTCACAAT TGAAGTTTGA ATATTCCTCT TCACATTTGG
DY3.4 rev       GTGTACTCCT GATTTTAAAT TATTCACAAT TGAAGTTTGA ATATTCCTCT TCACATTTGG

                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                  1155       1165       1175       1185       1195       1205
DY3-4-Int6      ATCCACATTT TTCATAACTT TAGATGCACG AAATGCGCGA GTGAAAGGAG CTTCTTTGTA
DY3.4 rev       ATCCACATTT TTCATAACTT TAGATGCACG AAATGCGCGA GTGAAAGGAG CTTCTTTGTA
```

```
                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                   1215       1225       1235       1245       1255       1265
DY3-4-Int6      GACAAATGGC TTCATTTTAT GATAACCACC ATTGTTTAAC ATTTTTGTAA TCCACTCCAG
DY3.4 rev       GACAAATGGC TTCATTTTAT GATAACCACC ATTGTTTAAC ATTTTTGTAA TCCACTCCAG

                ....|....| ....|....| ....|....|
                   1275       1285       1295
DY3-4-Int6      TATATTTGCA CAATTAGTTG ATCCAATATA
DY3.4 rev       TATATTTGCA CAATTAGTTG ATCCAATATA
```

Appendix 3

Sequence alignment of pET14b-DY3.4 construct highlighting the point mutations found.



Portions of alignment are shown highlighting the point mutation that has occurred in each position. **(a)** is mutation one with base change from t to g (Ile to Ser) **(b)** is mutation two with base change c to t (Ala to Val) **(c)** base change g to a is mutation number three (Ala to Thr) **(d)** mutation to t to c (Trp to Gly) is mutation number four **(e)** mutation number five is an a to g (Lys to Arg) **(f)** last mutation number six is a silent Gly mutation.

Appendix 4

Sequence analysis of pET14b-DY3.4 construct using Int1 forward primer. Alignment shows successful correction of mutations T back to C (Val to Ala) and A back to G (Thr to Ala) corrected bases are highlighted in grey.

```
                  70        80        90       100       110       120
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   AAAAGACGTCATTGGAACACAATTATTCGCAAAAGTGCTCATTTAAGTCTCGTCAGAAAT
DY3.4       ............................................................

                 130       140       150       160       170       180
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   ATTTTCAAAAAGAAACGAAGACAAAAATTAAACTACCGCCATATGTCCACTTTTTTGGCT
DY3.4       ............................................................

                 190       200       210       220       230       240
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   TCTTCTGAAAACATTTCAAATATTCGCGATACCGTTGTTTTACTGGAAATCAACAAATTA
DY3.4       ............................................................

                 250       260       270       280       290       300
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   TCGCAGAGTATCCTGTCTTCTCGCAATATTTTTGAGCTGAGAAAACTAAAACCAACACGA
DY3.4       ............................................................

                 310       320       330       340       350       360
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   AAAATGGTTGAAATCGAACATTGTTTTGTGTCTAAAAATGATACATTTGTTGCATGCTCC
DY3.4       ............................................................

                 370       380       390       400       410       420
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   AAATTAATTTCAATGTTTCATCTGGAAAAATATATTGGATCAACTAATTGTGCAAATATA
DY3.4       ............................................................

                 430       440       450       460       470       480
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   CTGGAGTGGATTACAAAAATGTTAAACAATGGTGGTTATCATAAAATGAAGCCATTTGTC
DY3.4       ............................................................

                 490       500       510       520       530       540
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   TACAAAGAAGCTCCTTTCACTCGCGCATTTCGTGCATCTAAAGTTATGAAAAATGTGGAT
DY3.4       ............................................................

                 550       560       570       580       590       600
            ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1   CCAAATGTGAAGAGGAATATTCAAACTTCAATTGTGAATAATTTAAAATCAGGAGTACAC
DY3.4       ............................................................
```

```
                    610       620       630       640       650       660
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1       TGGTGGGCTCTGATGGCTTTACGACAAGTTATGATTCCAATTGTAATTAAAGAAGAAAGA
DY3.4           ................C.............................................

                    670       680       690       700       710       720
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1       GTTTTGCTGTGGAGAGATGGATATTTAAATATTTTAACAAAAGAAATAAAGGATTTCAAA
DY3.4           ............................................................

                    730       740       750       760       770       780
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1       CAACGTTATATAGTTCAAAAAGCTCCGCATTTTATAAGACCGAATGTTGCCACCTTCAAG
DY3.4           ......................G.....................................

                    790       800       810       820       830       840
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1       CTCTCAATTTCTCGTCAAAAATTGCGACCTTTGTTCAAAAAAAAAGCAATCGATAAAAAA
DY3.4           ............................................................

                    850       860       870       880       890       900
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1       ACAGAGACAATGCAATGGAAAAGGTTGAATTCGATGCTCTCATGGTGTTTGGAAAGAAGT
DY3.4           ............................................................

                    910       920       930       940       950       960
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1       GGAGTTTACCGACATACAATCAGAGATTCGTGTAAAAAAGTGTCAGATTTTCTAAAGAAG
DY3.4           ............................................................

                    970       980       990       1000      1010      1020
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1       AACTCACAAAACTCAAAAATAATTGGATACACAGCAGATGTTAGCAAATGTTTCTCAACT
DY3.4           ............................................................

                    1030      1040      1050      1060      1070      1080
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-Int 1       GTTAATCATGACGTGTTGATATCAATTATTGACCGACTTTTTTCTCA
DY3.4           ...........................................
```

Sequence analysis of pET14b-DY3.4 construct using Int3 forward primer. Alignment shows successful correction of mutations, shaded, C back to T (Gly to Trp); G back to A (Arg to Lys) and the final mutation in this sequence alignment shows the Gly silent mutation.

```
                  1270      1280      1290      1300      1310      1320
                  ....|....|....|....|....|....|....|....|....|....|....|....|
tel4-int3for      CGTACCACACTTTCAACGTACTACTACAAGCGTGGCCCAACATCGTGGAGAATTACAAAA
DY3.4             ..............................................T.............

                  1330      1340      1350      1360      1370      1380
                  ....|....|....|....|....|....|....|....|....|....|....|....|
tel4-int3for      GGAGTTCCACAGGGACATCCTATATCTTCAAATTTAGCACATATGTACCTCAATAACTTT
DY3.4             ............................................................

                  1390      1400      1410      1420      1430      1440
                  ....|....|....|....|....|....|....|....|....|....|....|....|
tel4-int3for      GAGCAGAAATATTGGAGCAACGAAAAAGAGGATTCGAGAATTGTTTTCTGCAGATATGAG
DY3.4             ............................................................

                  1450      1460      1470      1480      1490      1500
                  ....|....|....|....|....|....|....|....|....|....|....|....|
tel4-int3for      GATGATTTCATTTTCATTACAACTGAAAATTCTTTATTCGAGAAGATGATGAAACCATTA
DY3.4             ............................................................

                  1510      1520      1530      1540      1550      1560
                  ....|....|....|....|....|....|....|....|....|....|....|....|
tel4-int3for      TCTACTGGCAATAACACTCATTTTTTGACGGCTAATCCGAAAAAGTTCAAGAAATCAGAG
DY3.4             ........................................A...................

                  1570      1580      1590      1600      1610      1620
                  ....|....|....|....|....|....|....|....|....|....|....|....|
tel4-int3for      CGATGTGGAGCATCACAAGTCCTTCAATGGTGTGGAGTGAAACTGGATTTTCAATCGGGA
DY3.4             ....................................G.......................

                  1630      1640      1650      1660      1670      1680
                  ....|....|....|....|....|....|....|....|....|....|....|....|
tel4-int3for      AATTGCTTTATTCGACGAAGATGCAAAGACGGTGTGGCTCGTCAATTTTTGATAAAGTTG
DY3.4             ............................................................

                  1690      1700      1710      1720      1730      1740
                  ....|....|....|....|....|....|....|....|....|....|....|....|
tel4-int3for      CAATAAGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAACGGGTCTTGAGGGGT
DY3.4             ......------------------------------------------------------
```

Sequence analysis of pET14b-DY3.4 construct using Int10 reverse primer. Since a reverse primer was used for the sequencing analysis the alignment shows the reverse compliment of DY3.4. Successful correction of mutations C back to T (sense direction) or G back to A (antisense direction) Ser to Ile was achieved.

```
                  1390      1400      1410      1420      1430      1440
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-rev10_In    CCATTTTTCGTGTTGGTTTTAGTTTTCTCAGCTCAAAAATATTGCGAGAAGACAGGATAC
DY3.4 rev       ............................................................

                  1450      1460      1470      1480      1490      1500
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-rev10_In    TCTGCGATAATTTGTTGATTTCCAGTAAAACAACGGTATCGCGAATATTTGAAATGTTTT
DY3.4 rev       ............................................................

                  1510      1520      1530      1540      1550      1560
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-rev10_In    CAGAAGAAGCCAAAAAAGTGGACATATGGCGGTAGTTTAATTTTTGTCTTCGTTTCTTTT
DY3.4 rev       ............................................................

                  1570      1580      1590      1600      1610      1620
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-rev10_In    TGAAAATATTTCTGACGAGACTTAAATGAGCACTTTTGCGAATAATTGTGTTCCAATGAC
DY3.4 rev       ............................................................

                  1630      1640      1650      1660      1670      1680
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-rev10_In    GTCTTTTATCGTATTTTCGTATTTCTTCATGAAGTGTTGTAAGTGAACTCTTAATCGTTG
DY3.4 rev       ...........................................................A......

                  1690      1700      1710      1720      1730      1740
                ....|....|....|....|....|....|....|....|....|....|....|....|
DY3-rev10_In    GTGCCAT
DY3.4 rev       .......
```

Appendix 5

Rare codons identified by Rare codon calculator (*RaCC)* present in DY3.4 TERT gene sequence from *C. elegans.*

Red = rare Arg codons **AGG, AGA, CGA**
Green = rare Leu codon **CTA**
Blue = rare Ile codon **ATA**
Orange = rare Pro codon **CCC**

For the following input sequence:

atg gca cca acg att aag agt tca ctt aca aca ctt cat gaa gaa ATA cga aaa
tac gat aaa aga cgt cat tgg aac aca att att cgc aaa agt gct cat tta agt
ctc gtc aga aat att ttc aaa aag aaa cga aga caa aaa tta aac tac cgc cat
atg tcc act ttt ttg gct tct tct gaa aac att tca aat att cgc gat acc gtt
gtt tta ctg gaa atc aac aaa tta tcg cag agt atc ctg tct tct cgc aat att
ttt gag ctg aga aaa CTA aaa cca aca cga aaa atg gtt gaa atc gaa cat tgt
ttt gtg tct aaa aat gat aca ttt gtt gca tgc tcc aaa tta att tca atg ttt
cat ctg gaa aaa tat att gga tca act aat tgt gca aat ATA ctg gag tgg att
aca aaa atg tta aac aat ggt ggt tat cat aaa atg aag cca ttt gtc tac aaa
gaa gct cct ttc act cgc gca ttt cgt gca tct aaa gtt atg aaa aat gtg gat
cca aat gtg aag agg aat att caa act tca att gtg aat aat tta aaa tca gga
gta cac tgg tgg gct ctg atg gct tta cga caa gtt atg att cca att gta att
aaa gaa gaa aga gtt ttg ctg tgg aga gat gga tat tta aat att tta aca aaa
gaa ATA aag gat ttc aaa caa cgt tat ATA gtt caa aaa gct ccg cat ttt ATA
aga ccg aat gtt gcc acc ttc aag ctc tca att tct cgt caa aaa ttg cga cct
ttg ttc aaa aga aaa gca atc gat aaa aaa aca gag aca atg caa tgg aaa agg
ttg aat tcg atg ctc tca tgg tgt ttg gaa aga agt gga gtt tac cga cat aca
atc aga gat tcg tgt aaa aaa gtg tca gat ttt CTA aag aag aac tca caa aac
tca aaa ATA att gga tac aca gca gat gtt agc aaa tgt ttc tca act gtt aat
cat gac gtg ttg ATA tca att att gac cga ctt ttt tct caa gag cac gat att
tat acg gta tgt gga aaa gga aga aat cac gga gga ttt cat aaa ttg ATA ttc
tgt tca gct gga act gag ttg aat gca cat gag gcg ctt cgt cga aaa atg gaa
ttg aaa gga gta ttc aat ttt gaa gtt tgc tac cga gag atg agt tcc tct aca
aca ctc tac agt gtt att cgt acc aca ctt tca acg tac tac tac aag cgt ggc
cca aca tcg tgg aga att aca aaa gga gtt cca cag gga cat cct ATA tct tca
aat tta gca cat atg tac ctc aat aac ttt gag cag aaa tat tgg agc aac gaa
aaa gag gat tcg aga att gtt ttc tgc aga tat gag gat gat ttc att ttc att
aca act gaa aat tct tta ttc gag aag atg atg aaa cca tta tct act ggc aat
aac act cat ttt ttg acg gct aat ccg aaa aag ttc aag aaa tca gag cga tgt
gga gca tca caa gtc ctt caa tgg tgt gga gtg aaa ctg gat ttt caa tcg gga
aat tgc ttt att cga cga aga tgc aaa gac ggt gtg gct cgt caa ttt ttg ATA
aag ttg caa taa

Appendix 6

Possible predicted sequence of TR named tts-1 (transcribed telomerase–like sequence) from *C.elegans.* Highlighted in grey are the possible one-and-a half telomeric template for telomere synthesis.

agtcagtattctttttttgagctagctataccttgagttaaaggttccgcttttctataaatatggctttgatgtaggtggaaattggcaa
gtagtacctaagcctaaacctaggccgaagtctaggaataagcctaatcctaagcccttacctaacttgcctgcttccatctatatc
aaagcccaatttccttacctcatttctctccaattccagaaaacttgaccggctcaaccatgtgacttctgacaattttcctcaatcct
ccgggcacaggaaccctgtaacaattctcgacttctcgtaaatggaagcacttcttcaaatgctctgcaacacaaaaatccaaaa
aaatgtatataaatgacacacaaaaatgacaaaaaacaaacgctttttctgatgatggttctctgttttttttttcaagtaaattgtttcgt
gcgacgacgggtttttgatacaattatcggaattttattggttgagtcaatgggtaaaacctacttttttaatttttttttgttttggtatttt
gtaaaaaatttttcaagcttgtaatttgtaattcaaattgtaaaagtaattgtaaaaatttgatttttttcttctttttttttctttttttgttttaagttt
tgaataaatttttttaatgctatt

Appendix 7

Sequence alignment of BsPel with the successful mutation of amino acid lysine 247 to alanine as highlighted in grey.

```
Query64    GCTGATTTAGGCCACCAGACGTTGGGATCCAATGATGGCTGGGGCGCGTACTCGACCGGC  123
Sbjct41    ............................................................  100

Query124   ACGACAGGCGGATCAAAAGCATCCTCCTCAAATGTGTATACCGTCAGCAACAGAAACCAG  183
Sbjct101   ............................................................  160

Query184   CTTGTCTCGGCATTAGGGAAGGAAACGAACACAACGCCAAAAATCATTTATATCAAGGGA  243
Sbjct161   ............................................................  220

Query244   ACGATTGACATGAACGTGGATGACAATCTGAAGCCGCTTGGCCTAAATGACTATAAAGAT  303
Sbjct221   ............................................................  280

Query304   CCGGAGTATGATTTGGACAAATATTTGAAAGCCTATGATCCTAGCACATGGGGCAAAAAA  363
Sbjct281   ............................................................  340

Query364   GAGCCGTCGGGAACACAAGAAGAAGCGAGAGCACGCTCTCAGAAAAACCAAAAAGCACGG  423
Sbjct341   ............................................................  400

Query424   GTCATGGTGGATATCCCTGCAAACACGACGATCGTCGGTTCAGGGACTAACGCTAAAGTC  483
Sbjct401   ............................................................  460

Query484   GTGGGAGGAAACTTCCAAATCAAGAGTGATAACGTCATTATTCGCAACATTGAATTCCAG  543
Sbjct461   ............................................................  520

Query544   GATGCCTATGACTATTTTCCGCAATGGGATCCGACTGACGGAAGCTCAGGGAACTGGAAC  603
Sbjct521   ............................................................  580

Query604   TCACAATACGACAACATCACGATAAACGGCGGCACACACATCTGGATTGATCACTGTACA  663
Sbjct581   ............................................................  640

Query664   TTTAATGACGGTTCGCGTCCGGACAGCACATCACCGAAATATTATGGAAGAAAATATCAG  723
Sbjct641   ............................................................  700

Query724   CACCATGACGGCCAAACGGATGCTTCCAACGGTGCTAACTATATCACGATGTCCTACAAC  783
Sbjct701   ............................................................  760

Query784   TATTATCACGATCATGATAAAAGCTCCATTTTCGGATCAAGTGACAGCAAAACCTCCGAT  843
Sbjct761   ..................GCG.....................................  820

Query844   GACGGCAAATTAAAAATTACGCTGCATCATAACCGCTATAAAAATATTGTCCAGCGCGCG  903
Sbjct821   ............................................................  880

Query904   CCGAGAGTCCGCTTCGGGCAAGTGCACGTATACAACAACTATTATGAAGGAAGCACAAGC  963
Sbjct881   ............................................................  940

Query964   TCTTCAAGTTATCCTTTTAGCTATGCATGGGGAATCGGAAAGTCATCTAAAATCTATGCC  1023
Sbjct941   ............................................................  1000

Query1024  CAAAACAATGTCATTGACGTACCG  1047
Sbjct1001  ........................  1024
```

Appendix 8

Paper published in Acta Crystallographica Section F:

Cloning, purification and preliminary crystallographic analysis of cobalamin methyltransferases from *Rhodobacter capsulatus*

Arefeh Seyedarabi, Thomas Hutchison, Teng Teng To, Evelyn Deery, Amanda Brindley, Martin J. Warren and Richard W. Pickersgill.

Cloning, purification and preliminary crystallographic analysis of cobalamin methyltransferases from *Rhodobacter capsulatus*

Arefeh Seyedarabi[1], Thomas Hutchison[1], Teng Teng To[1], Evelyn Deery[2], Amanda Brindley[2], Martin J. Warren[2] and Richard W. Pickersgill[1,*]

[1]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS [2]Centre for Molecular Processing, School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent CT2 7NJ, UK

*Correspondence should be addressed to: r.w.pickersgill@qmul.ac.uk

Synopsis

We report diffraction from crystals of the cobalamin methyltransferases CobJ, CobM, CobF and CobL (full-length and C-terminal domain) from *Rhodobacter capsulatus;* three of these enzymes also catalyse the auxiliary reactions of ring-contraction (CobJ), decarboxylation (CobL) and deacylation (CobF).

Abstract

Of the thirty biosynthetic steps necessary for the production of cobalamin (vitamin B_{12}), eight involve the addition of S-adenosyl-methionine derived methyl groups to the tetrapyrrole framework. The number of methylations highlights their importance in orchestrating the chemistry of cobalamin biosynthesis by changing the reactivity of the groups to which they attach and by causing prototrophic arrangement favourable for subsequent chemistries. In fact, three of the four methyltransferases for which we report diffracting crystals also catalyse auxillary reactions: ring contraction (CobJ), deacylation (CobF) and decarboxyation (CobL).

1. Introduction

There are seven methyltransferases that add eight S-adenosyl-methionine (SAM)-derived methyl groups to the tetrapyrrole framework that display a high level of regional

and temporal specificity. The discrepancy between the number of methyltransferases and the number of methylations is due to the fact that the first methyltransferase in the sequence, CobA, carries out two methylations. The remaining enzymes catalyse a single methylation reaction. To complicate things further, there are two similar though distinct routes for vitamin B_{12} biosynthesis, which differ in a requirement for molecular oxygen and the timing of cobalt insertion. These two routes are referred to as the aerobic and anaerobic pathways. However, both pathways require the same temporal methylation pattern although the intermediates differ, mainly due to the presence or absence of cobalt. Table 1 lists the methyltransferases associated with the aerobic late cobalt insertion pathway of B_{12} biosynthesis. Within this pathway, 6 of the 7 methyltransferases display sequence similarity to one another, clearly indicating that these enzymes have evolved from a common ancestor. Indeed, most of the methyltransferases of the aerobic route have a homologue in the anaerobic pathway, indicating that the aerobic and anaerobic routes split after the primordial B_{12} pathway had evolved (Raux, Schubert et al. 2000; Warren, Raux et al. 2002). The structure of the canonical cobalamin methyltransferase class III domain has been solved (Schubert, Wilson et al. 1998), as has the structure of the non-canonical fold (Keller, Smith et al. 2002). However, our understanding of the specificity and chemistry of the auxillary reactions as well as the evolutionary significance is unresolved.

Our current understanding of the evolution of complex metabolic pathways is based upon the idea that pathways evolved from a collection of broad specificity enzymes, which subsequently refined their binding pattern to become particular for one substrate. This theory is known as the patchwork evolution model and infers that the recruitment of single enzymes from different pathways is the driving force for pathway evolution, where enzymes evolve to catalyse different reactions on the same structure scaffold. This explains the appearance of enzyme superfamilies, in which enzymes retain a common structural or mechanistic strategy for catalysis within a shared protein framework.

The alternative theory is the retrograde model of pathway evolution, where the pathway evolves backwards from a key metabolite. Moreover, it was suggested that the enzymes of a pathway were descended from a common ancestral gene, by a process of gene duplication followed by functional differentiation, driven by substrate depletion. By this model, enzymes in a pathway would have overlapping substrate specificities with an inherent molecular memory of the original key metabolite. Such a theory was used to explain the existence of operons and end product inhibition of the first enzyme of the pathway and could be used to explain the prevalence of class III methyltransferases in the cobalamin biosynthetic pathway.

2. Material and Methods

2.1. Protein expression and purification

The DNA fragments corresponding to amino acid residues 1-245, 1-255, 1-261 and 1-396 encoded by the *cobJ*, *cobF*, *cobM* and *cobL* genes, respectively, were amplified from *Rhodobacter capsulatus* genomic DNA by polymerase chain reaction (PCR) using the FastStart High Fidelity PCR system (Roche). The PCR products were initially cloned into a modified pET3a (Novagen), which contains one *Spe*I site in 5' of *Bam*HI, by *Nde*I and *Spe*I digests (Table 2). The same restriction enzymes were used to subclone them into pET14b (Novagen), which had also been modified by adding the *Spe*I site in 5' of *Bam*HI. The DNA fragment corresponding to amino acids 214-396 encoded by the C-terminal domain of the *cobL* gene was amplified by PCR using the *pET3a::cobL(1-396)* construct as template and cloned directly into the modified pET14b. The pET14b vector added an N-terminal His_6-tag and a thrombin cleavage site to all the five enzymes. Recombinant clones were selected for by ampicillin resistance. The expression constructs were transformed into competent *E. coli* BL21 star (DE3) plysS cells. A small 10 ml overnight culture grown at 310 K in Luria-Bertani broth containing 100 μg ml^{-1} ampicillin and 34 μg ml^{-1} of chloramphenicol was used to inoculate a 1 litre culture. Cells were grown to an OD_{600} of 0.6. To express the recombinant proteins, the cells

were then induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. After overnight induction at 291 K or 303 K, the cells were harvested by centrifugation at 4950 g for 20 minutes at 277 K. Subsequently, the cell pellets were resuspended in 25 mM Tris-HCl buffer pH 8.0 containing 10 mM imidazole and 400-500 mM NaCl (buffer A) and frozen at 193 K. For purification of the over-produced proteins, the frozen cell pellets were thawed and sonicated, followed by centrifugation at 26900 g for 30 minutes at 277 K to remove cell debris. A 5 ml HisTrap nickel immobilised metal affinity chromatography column (GE Healthcare) equilibrated with buffer A was attached to an ÄKTA FPLC system (GE Healthcare). The soluble fraction was applied onto the column and then the column was washed with 40 ml of buffer A before the target protein was eluted with a linear gradient of imidazole (from 10 to 500 mM in 20 column volumes) in buffer A. The fractions containing the His_6-tagged Cobs (CobJ, CobF, CobM, CobL, CobL-C) were then identified by SDS-PAGE analysis. His_6-tagged Cobs were concentrated using 10 kDa molecular weight cutoff vivaspin concentrators (Sartorius) and applied to Superdex-200 size exclusion column (GE Healthcare). The CobL-C protein was eluted in 25 mM MES pH 6.0 and 250 mM NaCl, while all the other Cobs were eluted in 50 mM Tris- HCl pH 8.0 and 100 mM NaCl. The protein fractions from the Superdex column were further analysed by SDS-PAGE analysis and the protein fractions concentrated. CobJ protein was concentrated to 10-20 mg ml^{-1}; CobF to 25-50 mg ml^{-1}; CobM to 20-40 mg ml^{-1}; CobL to 8-10 mg ml^{-1} and CobL-C to 2-6 mg ml^{-1}. The protein concentration was estimated from absorption values at 280 nm using ProtPram analysis and the ExPASy server.

His_6-tagged CobM and CobL protein at 7 mg ml^{-1} and 10 mg ml^{-1}, respectively, were treated with thrombin (Novagen) to cleave off the His_6-tag. 1 unit of thrombin was sufficient to cleave 1 mg of protein in thrombin cleavage buffer consisting of 20 mM Tris-HCl pH 8.4, 150 mM NaCl, and 2.5 mM $CaCl_2$, over the course of 18-20 hours at room temperature. A final concentration of 1 mM PMSF was added to stop the thrombin

activity after the incubation period. The thrombin was then removed by using size exclusion chromatography (Superdex-200).

2.2. Crystallisation

Crystallisation was at 292 K using the hanging-drop vapour diffusion method in 24-well plates. The methyl donor, S-adenosyl-methionine (SAM; Sigma), was used to aid crystallisation of the proteins. Two crystallisation screens exploiting sparse-matrix sampling (Jancarik and Kim 1991), Hampton Research Crystal Screens I and II (Hampton Research) with a total of 98 different conditions, were used to screen for crystallisation conditions. SAM was varied in the range 0-5 mM. 2 µl drops containing 1:1 ratio of protein and reservoir solution were equilibrated against 500 µl of each reservoir solution.

His_6-tagged CobJ crystals grew in 2-4 days from Crystal Screen I condition 46 (0.2 M calcium acetate hydrate, 0.1 M sodium cacodylate trihydrate pH 6.5 and 18 % PEG 8000). Three crystal forms were characterised, slightly varying calcium acetate and PEG concentrations may have contributed to these different forms. His_6-tagged CobF crystals (type I, II, III) were obtained within 2-5 days in Crystal Screen I condition 37 (0.1 M sodium acetate trihydrate pH 4.6 and 8% PEG 4000). His_6-tagged CobM crystals were obtained within 2-5 days in Crystal Screen I condition 17 (0.2 M Lithium sulfate monohydrate, 0.1 M Tris pH 8.5 and 30 % PEG 4000) and Crystal Screen II condition 38 (0.1 M Hepes pH 7.5 and 20 % PEG 10,000), while crystals of His_6-cleaved CobM were obtained within 5-7 days in Crystal Screen I condition 41 (0.1 M Hepes pH 7.5, 10% isopropanol and 20% PEG 4000). His_6-tagged CobL-C crystals were obtained within 1-2 days in numerous conditions including for types I and II respectively, Crystal Screen I condition 9 (0.2 M Ammonium acetate, 0.1 M tri-sodium citrate dihydrate pH 5.6 and 30% PEG 4000) and condition 40 (0.1 M tri-sodium citrate dihydrate pH 5.6, 20% isopropanol and 30% PEG 4000). His_6-tagged CobL crystals grew after 1-2 weeks

from CSI condition 14 (0.2 M calcium chloride, 0.1 M sodium HEPES pH 7.5, 28% PEG 400). Figure 1 shows a crystal of each of the four methyltransferases.

2.3. Cryoprotection and data collection

Augmentation with 20 % glycerol was successful for CobJ and CobL-C crystals and augmentation with 25 % MPD was successful with CobF and CobM crystals. Diffraction data from eleven crystals were collected at 100 K and processed with MOSFLM (Leslie 2006). POINTLESS was used to identify the correct space group for each dataset before scaling using SCALA (Evans 2006).

3. Results and discussion

Four methyltransferases from *Rhodobacter capsulatus* CobJ, CobF, CobM and CobL, were successfully cloned, overproduced and crystallized (Figure 1). The C-terminal domain of CobL (CobL-C) was also successfully cloned, overproduced and crystallized. In each case, the authenticity of the cloned DNA was confirmed by sequencing. All five constructs expressed large quantities of protein and well diffracting crystals were grown of all, with the possible exception of CobL which gave reasonable diffraction (diffraction limit 3.2 Å). CobL has two domains, a canonical and a non-canonical methyltransferase (see Table 1) and there may be some flexibility between the two domains which might contribute to the lower diffraction limit for these crystals. The inclusion of SAM was crucially important in obtaining well-diffracting crystals of these SAM-binding methyltransferases. Crystals were mostly grown with the hexahistidine tag intact, though removal of the tag from CobM did give slightly better diffracting crystals. The proteins had good stability with the exception of the C-terminal domain of CobL (CobL-C) which was stabilized by the addition of sodium chloride. The quality of the diffraction data is presented in Table 3. It is immediately clear from molecular replacement calculations that CobF, CobM, and CobL-C can be readily solved using existing methyltransferase structures and refinement and validation of these structures is

underway. Experimental phases are being determined for CobJ and CobL from Se-Met labelled proteins.

Acknowledgements

We acknowledge support from the BBSRC (grant number BB/E00213) and the Diamond Light Source (Oxford), ESRF (Grenoble) and SRS (Daresbury).

Table 1 Methyltransferases the cobalamin biosynthesis pathway

Methylates:	**Enzyme**	**Pathway**	**Catalyses the transformation**	**Reference**
C2 & C7	CobA	Both	uroporphrinogen III into precorrin-2	(Vevodova, Graham et al. 2004)
C20	CobI	Aerobic	precorrin-2 into precorrin-3A	No data
C20	CbiL	Anaerobic	Cobalt-factor II into cobalt factor III	(Frank, Deery et al. 2007; Wada, Harada et al. 2007)
C17	CobJ	Aerobic	precorrin3B into precorrin-4	This work
C17	CbiH	Anaerobic	Cobalt-factor III into cobalt factor IV	No data
C11	CobM	Aerobic	precorrin 4 into precorrin-5	This work
C11	CbiF	Anaerobic	Cobalt-factor IV into cobalt factor V	(Schubert, Wilson et al. 1998)
C1	CobF	Aerobic	precorrin 5 into precorrin-6X	This work and PDB code: 2NPN
C1	CbiD	Anaerobic	Cobalt-factor V into cobalt factor VI	Not canonical. PDB code: 1SR8
C5	CobL-N	Aerobic	precorrin 6Y into precorrin-7	This work
C5	CbiE	Anaerobic	Cobalt-factor VI into cobalt factor VII	PDB code: 2BB3
C15	CobL-C	Aerobic	precorrin 7 into hydrogenobyrinic acid	This work
C15	CbiT	Anaerobic	Cobalt-factor VII into hydrogenobyrinoic acid	(Keller, Smith et al. 2002)

The enzymes are given in order according to their position on the biosynthetic pathway (with the exception of CobA which is common to both) the methyltransferase of the aerobic (early cobalt insertion) pathway is given followed by the equivalent enzyme of the anaerobic (late cobalt insertion) pathway. When a PDB code is given, the structure is the result of a structural genomics consortium without other citation. CobL-N and CobL-C refer to the N- and C-terminal domains of CobL.

Table 2 List of primers used for PCR amplification of the methyltransferases

Gene	Primers	Restriction sites
cobJ	Forward: CAT**CATATG**AGCGGTTGGGTCACG Reverse: CAT**ACTAGT**GGCTTGCCGATCACGGC	*Nde*I *Spe*I
cobF	Forward: CAT**CATATG**ATCGAGCTTTCCCTG Reverse: CAT**ACTAGT**CACGCCGCGGGCACATTG	*Nde*I *Spe*I
cobM	Forward: CAT**CATATG**ACGGTGCATTTCATCG Reverse: CAT**ACTAGT**CATTCGCTGCCCTCCGG	*Nde*I *Spe*I
cobL	Forward: CTA**CATATG**TCTGATCCGTGGTTG Reverse: CAT**ACTAGT**CATCTTTGCCCGCTCC	*Nde*I *Spe*I
CobL-C	Forward: CAT**CATATG**GGCACGGGGCTTTCGCAG Reverse: as for *cobL* above	*Nde*I

Table 3: Crystallographic data statistics

Enzyme	CobJ (I)	CobJ (II)	CobJ (III)	CobF (I)	CobF (II)	CobF (III)	CobM (I)	CobM (II)	CobL-C (I)	CobL-C (II)	CobL
Synchrotron	DLS IO3	ESRF ID14-1	ESRF ID14-1	SRS PX10.1	ESRF ID14-1	ESRF ID29	SRS PX10.1	ESRF ID29	ESRF ID29	DLS IO2	ESRF ID14-4
Wavelength (Å)	0.9808	0.933	0.933	1.117	0.933	1.011	1.117	1.011	1.011	0.9795	0.939
Detector	ADSC Q315R	ADSC Q210	ADSC Q210	MAR MOSAIC	ADSC Q210	ADSC Q315R	MAR MOSAIC	ADSC Q315R	ADSC Q315R	ADSC Q315	ADSC Q315R
Space group	$P2_12_12_1$	$P2_1$	$C222_1$	$P2_12_12_1$	$P2_1$	$C222_1$	$P6_5$	$P2_122_1$	C2	$P2_12_12_1$	I222
Unit-cell parameters (Å)	a=85.64 b=110.99 c=116.19	a= 68.31 b= 69.43 c=72.65 β=118.16	a= 49.73 b=112.85 c=91.25	a=61.75 b=89.42 c=107.14	a=54.15 b= 84.78 c= 58.61 β=98.65 °	a=61.38 b= 89.28 c=107.99	a=84.52 b= 84.52 c= 144.64	a=51.05 b=54.50 c=175.85	a=114.35 b=44.80 c=84.11 β=119.75	a=62.67 b=86.39 c=149.96	a=103.99 b=112.33 c=373.69
Unique reflections	54689 (7910)	41959 (6172)	7019 (1007)	95539 (13791)	41043 (5940)	22829 (3264)	39472 (5720)	17755 (2525)	10402 (1514)	21312 (3045)	36000 (5188)
Multiplicity	3.9 (4.0)	3.6 (3.7)	6.7 (6.9)	6.9 (6.4)	3.6 (3.5)	4.5 (4.3)	9.8 (8.5)	7.0 (7.2)	3.6 (3.7)	5.5 (5.6)	14.4 (14.8)
Resolution (Å)	58.56-2.22 (2.34-2.22)	34.72-1.97 (2.08-1.97)	48.09-2.75 (2.90-2.75)	26.78-1.50 (1.58-1.50)	33.23-1.90 (2.00-1.90)	53.99-1.93 (2.03-1.93)	42.26-2.00 (2.11-2.00)	87.93-2.50 (2.64-2.50)	72.93-2.70 (2.85-2.70)	74.98-2.78 (2.93-2.78)	107.57-3.2 (3.37-3.2)
Completeness (%)	99.4 (99.9)	98.8 (100.0)	99.7 (100.0)	99.9 (100.0)	99.5 (98.9)	99.9 (99.7)	100.0 (100.0)	100.0 (100.0)	99.9 (100.0)	99.9 (100.0)	98.6 (98.6)
R_{merge} (%) [a]	8.1 (52.5)	19.5 (79.2)	10.3 (63.1)	6.4 (41.3)	8.1 (40.7)	11.6 (27.5)	6.8 (62.0)	14.2 (30.5)	6.0 (21.4)	13.2 (63.2)	17.3 (65.2)
$R_{p.i.m.}$ (%) [b]	4.6 (30.9)	12.4 (47.9)	4.3 (25.6)	2.6 (17.6)	4.9 (25.0)	6.3 (15.1)	2.3 (22.3)	5.8 (12.2)	3.7 (13.1)	6.3 (30.0)	4.6 (17.4)
Mean I/σ(I)	12.3 (2.6)	7.3 (2.7)	12.2 (2.7)	15.2 (3.7)	13.1 (3.1)	9.0 (4.6)	19.2 (3.3)	9.9 (5.6)	14.6 (5.3)	9.4 (3.0)	17.3 (5.4)
Wilson B-factor ($Å^2$)	38.3	16.9	66.5	16.3	18.7	20.30	33.2	35.1	56.5	63.4	66.8
Mol. per au/ fraction solvent	4/0.51	2/0.55	1/0.47	2/0.48	2/0.43	1/0.49	2/0.51	2/0.36	2/0.40	4/0.44	4/0.58

Values in parentheses are for the highest resolution shell. [a]$R_{merge} = (\Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle|) / (\Sigma_{hkl} \Sigma_i I(hkl))$, where $I_i(hkl)$ is the intensity measured for a each unique Bragg reflection with indices (hkl), $\langle I(hkl)\rangle$ is the average intensity for multiple measurements of this reflection. [b]$R_{meas} = \Sigma_{hkl} [n/(n-1)]^{1/2} \Sigma_i |I_i(hkl) - \langle I(hkl)\rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$, where n is the multiplicity, other variables as defined for R_{merge}. V_m is Matthew's coefficient and au represents asymmetric unit.

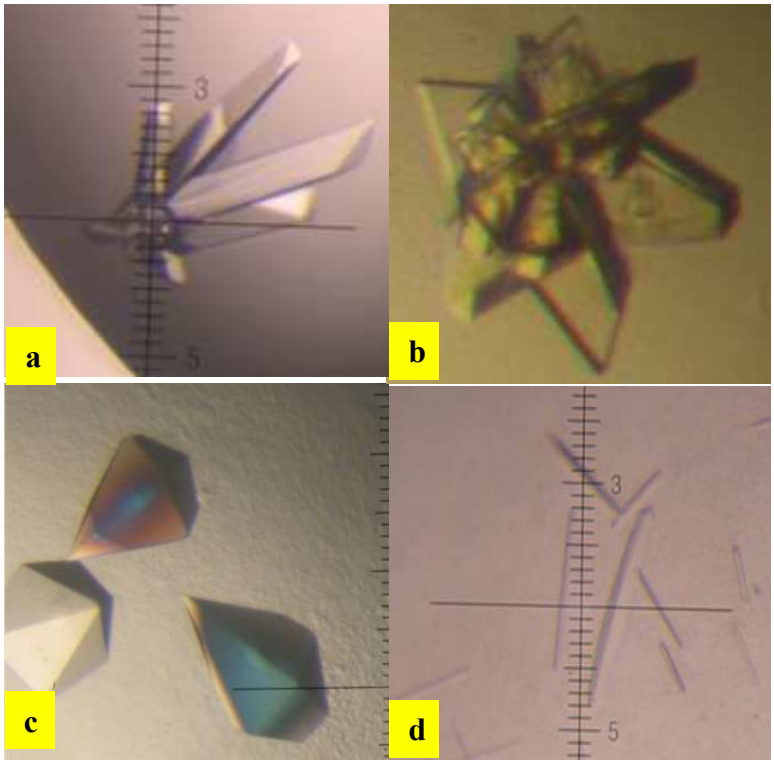


Figure 1 Crystals of *Rhodobacter capsulatus* methyltransferases: (a) CobJ (type I), (b) CobF (type I), (c) CobM (type I) and (d) CobL.

References

1. Collmer, A., and Keen, N. T. (1986) The Role of Pectic Enzymes in Plant Pathogenesis, Annual Review of Phytopathology 24, 383-409.

2. Wolfenden, R., and Snider, M. J. (2001) The depth of chemical time and the power of enzymes as catalysts, Accounts of chemical research 34, 938-945.

3. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics, Nucleic Acids Res. 37, D233-D238.

4. Yoder, M. D., Keen N, T., and Jurank, F. (1993) New domain motif: the structure of pectate lyase C, a secreted plant virulence factor, Science 260, 1503-1507.

5. Pickersgill, R. W., Jenkins, J., Harris, G., Nasser, W., and Robert-Baudouy, J. (1994) The structure of Bacillus subtilis pectate lyase in complex with calcium, Nat Struct Biol 1, 717-723.

6. Akita, M., Suzuki, A., Kobayashi, T., Ito, S., and Yamane, T. (2001) The first structure of pectate lyase belonging to polysaccharide lyase family 3, Acta Crystallogr. Sect. D-Biol. Crystallogr. 57, 1786-1792.

7. Jenkins, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N., and Pickersgill, R. W. (2004) The crystal structure of pectate lyase Pel9A from Erwinia chrysanthemi, J Biol Chem 279, 9139-9145.

8. Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R., and Jenkins, J. (1997) Two crystal structures of pectin lyase A from Aspergillus reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases, Structure 5, 677-689.

9. Vitali, J., Schick, B., Kester, H. C. M., Visser, J., and Jurnak, F. (1998) The three dimensional structure of Aspergillus niger pectin lyase B at 1.7-angstrom resolution, Plant Physiol. 116, 69-80.

10. Petersen, T. N., Kauppinen, S., and Larsen, S. (1997) The crystal structure of rhamnogalacturonase A from Aspergillus aculeatus: A right-handed parallel beta helix, Structure 5, 533-544.

11. Pickersgill, R., Smith, D., Worboys, K., and Jenkins, J. (1998) Crystal structure of polygalacturonase from Erwinia carotovora ssp. carotovora, J. Biol. Chem. 273, 24660-24664.

12. Jenkins, J., Mayans, O., Smith, D., Worboys, K., and Pickersgill, R. W. (2001) Three-dimensional structure of Erwinia chrysanthemi pectin methylesterase reveals a novel esterase active site, J Mol Biol 305, 951-960.

13. Johansson, K., El-Ahmad, M., Friemann, R., Jornvall, H., Markovic, O., and Eklund, H. (2002) Crystal structure of plant pectin methylesterase, FEBS Lett 514, 243-249.

14. Fries, M., Ihrig, J., Brocklehurst, K., Shevchik, V. E., and Pickersgill, R. W. (2007) Molecular basis of the activity of the phytopathogen pectin methylesterase, EMBO J. 26, 3879-3887.

15. Jenkins, J., and Pickersgill, R. (2001) The architecture of parallel beta-helices and related folds, Prog. Biophys. Mol. Biol. 77, 111-175.

16. Abbott, D. W., and Boraston, A. B. (2007) A family 2 pectate lyase displays a rare fold and transition metal-assisted beta-elimination, J. Biol. Chem. 282, 35328-35336.

17. Charnock, S. J., Brown, I. E., Turkenburg, J. P., Black, G. W., and Davies, G. J. (2001) Characterization of a novel pectate lyase, Pel10A, from Pseudomonas cellulosa, Proc Natl Acad Sci 99, 12067-12072.

18. Scavetta, R. D., Herron, S., R.,, Hotchkiss, A. T., Kita, N., Keen, N. T., Benen, J. A., Kester H.C., Visser, J., and Jurank, F. (1999) Structure of a plant cell wall fragment complexed to pectate lyase C, Plant Cell 11, 1081-1092.

19. Herron, S., Benen, J. A., Scavetta, R. D., Visser, J., and Jurank, F. (2000) Structure and function of pectic enzymes: virulence factors of plant pathogens, Proc Natl Acad Sci 97, 8762-8769.

20. Leslie, A. G. (1999) Integration of macromolecular diffraction data, Acta Crystallogr D Biol Crystallogr 55, 240-255.

21. Evans, P. (2006) Scaling and assessment of data quality, Acta Crystallogr. Sect. D-Biol. Crystallogr. 62, 72-82.

22. Vagin, A. A., and Teplyakov, A. (1997) MOLREP: An automated program for molecular replacement, J Appl Crystallogr 30, 1022-1025.

23. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr D53, 240-255.

24. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model

building combined with iterative structure refinement, Nat Struct Biol 6, 458-463.

25. Jones, T. A., Zou, J.-Y., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models, Acta Crystallogr A47, 110-119.

26. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, Acta Crystallogr D Biol Crystallogr 60, 2126-2132.

27. Rigby, N. M., MacDougall, A. J., Ring, S. G., Cairns, P., Morris, R., and Gunning, P. A. (2000) Observations on the crystallization of oligogalacturonates, Carbohydr Res 328, 235-239.

28. Clausen, M. H., Jorgensen, M. R., Thorsen, J., and Madsen, R. (2001) A strategy for chemical synthesis of selectively methyl-esterified oligomers of galacturonic acid, Journal of the Chemical Society-Perkin Transactions 1, 543-551.

29. Clausen, M., and Madsen, R. (2004) Synthesis of oligogalacturonates conjugated to BSA, Carbohydr Res 339, 2159-2169.

30. Clausen, M. H., and Madsen, R. (2003) Synthesis of hexasaccharide fragments of pectin, Chem.-Eur. J. 9, 3821-3832.

31. Petersen, B. O., Meier, S., Duus, J. O., and Clausen, M. H. (2008) Structural characterization of homogalacturonan by NMR spectroscopy-assignment of reference compounds, Carbohydr. Res. 343, 2830-2833.

32. Herron, S., R.,,, Scavetta, R. D., Garrett, M., Legner, M., and Jurank, F. (2003) Characterization and implications of Ca2+ binding to pectate lyase C, J Biol Chem 278, 12271-12277.

33. Morris, E. R., Powell, D. A., Gidley, M. J., and Rees, D. A. (1982) Conformations and interactions of pectins 1. Polymorphism between gel and solid states of calcium polygalacturonate, J. Mol. Biol. 155, 507-516.

34. Powell, D. A., Morris, E. R., Gidley, M. J., and Rees, D. A. (1982) Conformations and interactions of pectins 2. Influence of residue sequence on chain association in calcium pectate gels, J. Mol. Biol. 155, 517-531.

35. Walkinshaw, M. D., and Arnott, S. (1981) Conformations and interactions of pectins .1. X-ray-diffraction analyses of sodium pectate in neutral and acidified forms, J. Mol. Biol. 153, 1055-1073.

36. Walkinshaw, M. D., and Arnott, S. (1981) Conformations and interactions of

pectins models for junction zones in pectinic acid and calcium pectate gels, J. Mol. Biol. 153, 1075-1085.

37. Jarvis, M. C., and Apperley, D. C. (1995) Chain conformation in concentrated pectic gels - evidence from C-13 NMR, Carbohydr. Res. 275, 131-145.

38. Hricovini, M., Bystricky, S., and Malovikova, A. (1991) Conformations of (1-4)-linked alpha-D-galacturono-di-saccharides and alpha-D-galacturono-trisaccharides in solution analyzed by NMR measurements and theoretical calculations, Carbohydr. Res. 220, 23-31.

39. Gouvion, C., Mazeau, K., Heyraud, A., Taravel, F. R., and Tvaroska, I. (1994) Conformational study of digalacturonic acid and sodium digalacturonate in solution, Carbohydr. Res. 261, 187-202.

40. Dinola, A., Fabrizi, G., Lamba, D., and Segre, A. L. (1994) SOLUTION Conformation of a pectic acid fragment by H-1-NMR and molecular-dynamics, Biopolymers 34, 457-462.

41. Gerlt, J. A., and Gassman, P. G. (1992) Understanding enzyme-catalyzed proton abstraction from carbon acids: details of stepwise mechanisms for betaelimination reactions, J Am Chem Soc 114, 5928-5934.

42. Gerlt, J. A., and Gassman, P. G. (1993) Understanding the rates of certain enzyme-catalyzed reactions: proton abstraction from carbon acids, acyl-transfer reactions, and displacement reactions of phosphodiesters, J Am Chem Soc 115, 11552-11568.

43. Jenkins, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N., and Pickersgill, R. W. (2004) The crystal structure of pectate lyase Pel9A from Erwinia chrysanthemi, J. Biol. Chem. 279, 9139-9145.

44. Atkins, E. D. T., Nieduszy.Ia, Parker, K. D., and Smolko, E. E. (1973) Structural components of alginic acid .2. Crystalline-structure of poly-alpha-L-guluronic acid - results of X-ray-diffraction and polarized infrared studies, Biopolymers 12, 1879-1887.

45. Perez, S., Mazeau, K., and du Penhoat, C. H. (2000) The three-dimensional structures of the pectic polysaccharides, Plant Physiol. Biochem. 38, 37-55.

46. Braccini, I., Grasso, R. P., and Perez, S. (1999) Conformational and configurational features of acidic polysaccharides and their interactions with calcium ions: a molecular modeling investigation, Carbohydr. Res. 317, 119-130.

47. DeLano, W. L., (Ed.) (2002) The PyMOL molecular graphics system on World Wide Web. 0.80 version. The PyMol User's Manual, DeLano Scientific, San Carlos, CA, USA.

48. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT - A program to generate schematic diagrams of protein ligand interactions, Protein engineering 8, 127-134.

Appendix 9

Paper published in Biochemistry journal:

Structural insights into substrate specificity and the *anti* β-elimination mechanism of pectate lyase

Arefeh Seyedarabi, Teng Teng To, Salyha Ali, Syeed Hussain, Markus Fries, Robert Madsen, Mads H. Clausen, Susana Teixteira, Keith Brocklehurst and Richard W. Pickersgill.

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 539–546 539
DOI: 10.1021/bi901503g

Structural Insights into Substrate Specificity and the *anti* β-Elimination Mechanism of Pectate Lyase[†,‡]

Arefeh Seyedarabi,[§] Teng Teng To,[§] Salyha Ali,[§,⊥] Syeed Hussain,[§,▽] Markus Fries,[§,#] Robert Madsen,[∥] Mads H. Clausen,[∥] Susana Teixteira,[⊥,@] Keith Brocklehurst,[§] and Richard W. Pickersgill*[,§]

[§]*School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, United Kingdom,* [∥]*Department of Chemistry, Technical University of Denmark, DK-2800 Lyngby, Denmark,* [⊥]*Institut Laue Langevin, 6 rue Jules Horowitz, 38000 Grenoble, France, and* [@]*EPSAM, Keele University, Keele, Staffordshire ST5 5BG, United Kingdom* [▽]*Current address: AstraZeneca, Alderly Park, Cheshire SK10 4TF, United Kingdom.* [#]*Current address: Baxter BioScience, Biomedical Research Center, Uferstrasse 15, A-2304 Orth an der Donau, Austria.*

Received August 27, 2009; Revised Manuscript Received November 19, 2009

ABSTRACT: Pectate lyases harness *anti* β-elimination chemistry to cleave the α-1,4 linkage in the homogalacturonan region of plant cell wall pectin. We have studied the binding of five pectic oligosaccharides to *Bacillus subtilis* pectate lyase in crystals of the inactive enzyme in which the catalytic base is substituted with alanine (R279A). We discover that the three central subsites (−1, +1, and +2) have a profound preference for galacturonate but that the distal subsites can accommodate methylated galacturonate. It is reasonable to assume therefore that pectate lyase can cleave pectin with three consecutive galacturonate residues. The enzyme in the absence of substrate binds a single calcium ion, and we show that two additional calcium ions bind between enzyme and substrate carboxylates occupying the +1 subsite in the Michaelis complex. The substrate binds less intimately to the enzyme in a complex made with a catalytic base in place but in the absence of the calcium ions and an adjacent lysine. In this complex, the catalytic base is correctly positioned to abstract the C5 proton, but there are no calcium ions binding the carboxylate at the +1 subsite. It is clear, therefore, that the catalytic calcium ions and adjacent lysine promote catalysis by acidifying the α-proton, facilitating its abstraction by the base. There is also clear evidence that binding distorts the relaxed 2_1 or 3_1 helical conformation of the oligosaccharides in the region of the scissile bond.

Pectate lyases are carbon−oxygen lyases that harness *anti* β-elimination chemistry to cleave the α-1,4 glycosidic linkage between D-galacturonate (GalA) residues in the homogalacturonan region of the plant polysaccharide pectin (*1*). In the reaction scheme below, R and R′ are additional α-1,4-linked GalA residues.

Pectate lyases play a pivotal role in remodeling and recycling the pectin polysaccharides present as insoluble composites in plant cell walls, accelerating rates of reaction by factors exceeding 10^{17}-fold and cleaving one of the most stable bonds in nature (*2*). The bacterial pectate lyases loosen the adhesion between plants cells, allowing the bacteria to invade causing crop spoilage. Pectate lyases (EC 4.2.2.2) occur in five of the 21 families of polysaccharide lyases (*3*): PL1, PL2, PL3, PL9, and PL10. *Bacillus subtilis* pectate lyase (BsPel),[1] the subject of this paper, belongs to PL1 and adopts the parallel β-helix fold (*4*, *5*); PL3 (*6*) and PL9 (*7*) pectate lyases also have this protein architecture as do fungal pectin lyases A and B (*8*, *9*), rhamnogalacturonase (GH28) (*10*), polygalacturonase (GH28) (*11*), and pectin methylesterase (CE8) (*12*−*15*). Pectate lyases in families PL2 and PL10 form $(\alpha,\alpha)_7$ and $(\alpha,\alpha)_3$ barrels, respectively (*16*, *17*).

The structure of pentagalacturonate in complex with *Erwinia chrysanthemi* pectate lyase C (PelC) has been reported previously (*18*), and a catalytic mechanism featuring abstraction of a proton from C5 of the GalA residue binding the +1 subsite with the elimination of the leaving group from C4 is reasonable (*19*). This reaction mechanism featuring proton abstraction by arginine acting as a base is supported by the convergence in active site geometry seen upon comparison of the PL10 and PL1 polysaccharide lyase structures (*17*). In this paper, we probe the mechanism and substrate specificity of BsPel in depth. We confirm the role of the catalytic base by observing its presence in the correct place for proton abstraction. We show that two

[†]This work was supported by a grant from the Biotechnological and Biology Sciences Research Council (Grant BBS/B/07896). The Higher Education Funding Council of England also supported this work.

[‡]Protein coordinates and structure factor amplitudes are available from the Protein Data Bank as entries 2NZM, 2O04, 2O0V, 2O17, 2O1D, and 3KRG.

*To whom correspondence should be addressed. Phone: ++44 20 7882 8444. Fax: +44 20 8983 0973. E-mail: r.w.pickersgill@qmul.ac.uk.

[1]Abbreviations: BsPel, *B. subtilis* pectate lyase; EDTA, ethylenediaminetetraacetic acid; Gal, galacturonate; GalMe, C6-methylated galacturonate; GalA3, trigalacturonate; GalA6, hexagalacturonate; HPLC, high-performance liquid chromatography; IPTG, isopropyl β-D-1-thiogalactopyranoside.

Published on Web 12/15/2009

Table 1: Crystallographic Data Statistics for the Six Oligosaccharide Soaks

	GalA3·3Ca	GalA6·3Ca	I·3Ca	II·3Ca	III·3Ca	GalA6·Co
resolution (Å)	2.0	2.3	1.8	1.7	1.9	1.9
R_{merge}[a,b] (%)	7.6 (13.8)	9.7 (22.0)	5.4 (11.7)	6.7 (24.1)	5.1 (14.6)	5.7 (6.7)
completeness[a] (%)	99.1 (97.6)	92.1 (90.3)	98.9 (98.4)	96.3 (83.4)	100.0 (100.0)	97.2 (96.1)
mean $I/\sigma(I)$[a]	15.7 (8.8)	11.0 (4.2)	19.7 (11.1)	15.0 (5.0)	18.4 (8.8)	23.8 (18.3)
R factor (R_{free})[c]	14.9 (19.7)	14.4 (22.3)	15.5 (18.5)	16.6 (20.1)	15.2 (19.2)	13.1 (19.6)

[a]Values in parentheses are for the highest-resolution shell. [b]$R_{merge} = (\sum_{hkl}\sum_i |I_i - \langle I \rangle|)/\sum_{hkl}\sum_i I_i$, where I_i is the intensity of the ith observation, $\langle I \rangle$ is the mean intensity of the reflection, and the summations extend over all unique reflections (hkl) and all equivalents (i). [c]R factor $= \sum_{hkl}|F_o - F_c|/\sum_{hkl}|F_o|$, where F_o and F_c represent the observed and calculated structure factors, respectively. The R factor is calculated using 95% of the data included in the refinement and R_{free} the 5% excluded. GalA3 and GalA6 are trigalacturonate and hexagalacturonate, respectively. The specifically methylated hexasacharrides (I, II, and III) are described in Materials and Methods. All complexes were formed in R279A crystals (catalytic base mutant) in the presence of calcium, with the exception of the last complex, GalA6·Co, formed in the triple mutant enzyme (N180A/D173A/K147A), in the absence of calcium (see the text for more details).

calcium ions that bind between the enzyme and substrate carboxylates at the +1 subsite together with an adjacent lysine residue stabilize the reaction intermediate. We also provide evidence that supports the view that substrate distortion is a characteristic of catalysis by pectate lyase. We show that the three central subsites (+2, +1, and −1) have a strong preference for galacturonate and that methylated galacturonate residues can be accommodated in the more distal subsites. Combining the results of crystallography, use of mutant enzymes, and isothermal titration calorimetry, we reveal the importance of entropy in binding and gain insight into the calcium binding characteristics of the free enzyme and the enzyme in the presence of substrate. The process of binding essential catalytic components, two calcium ions, together with substrate is uncommon in enzyme catalysis and is reminiscent of substrate-assisted catalysis.

MATERIALS AND METHODS

Production of Native and Mutant BsPel Proteins. A DNA fragment encoding the mature BsPel protein was amplified from *B. subtilis* genomic DNA using PCR and oligonucleotide primers that included NcoI restriction sites immediately flanking the coding sequence. The PCR product was subcloned into expression vector pET3d and the plasmid transformed into *Escherichia coli* BL21(DE3) pLysS. The sequence was confirmed before protein production. Protein overproduction was induced using IPTG (0.1 mM), and purification was conducted by ion exchange chromatography and size exclusion chromatography. Mutations were introduced using appropriate mutagenic primers and the QuickChange kit (Stratagene).

Kinetic Analyses. Enzyme assays using polygalacturonate were conducted at 42 °C in 50 mM Tris (pH 8.5) in the presence of 5 mM $CaCl_2$; the pH was checked for consistency before and after the assay. The formation of product was assessed at 235 nm (extinction coefficient of 5200 M^{-1} cm^{-1}). Acids were separated by HPLC using a calibrated Microsorb amino column. Unsaturated galacturonic acids were detected from their UV absorption at 232 nm.

Crystallography. Crystals of the R279A mutant were grown by streak seeding pre-equilibrated protein drops (20 mg/mL) using wild-type crystals as seeds. The soak conditions for the trigalacturonate complex were the reservoir conditions: 0.1 M sodium acetate (pH 4.6), 0.2 M ammonium acetate, and 18% PEG 4000, supplemented with 20 mM trigalacturonate, 7 mM calcium chloride, and 15% glycerol as a cryoprotectant. The conditions used for the hexasaccharide complexes were identical except that the hexasaccharide and calcium concentrations were 5 and 2 mM, respectively. The product complex used native crystals soaked under similar crystallization conditions supplemented with 20 mM GalA3, 5 mM sodium chloride, and 15% glycerol. The calcium-free complex was formed using the triple mutant which crystallized in the same space group and under soak conditions similar to those above with the exception that 2 mM cobalt chloride was used in place of calcium chloride, the hexagalacturonate concentration was increased to 100 mM, and the pH of the sodium acetate was lowered 0.6 unit to pH 4.0. The crystals, of space group $P2_1$ (a = 50.6 Å, b = 69.0 Å, c = 59.2 Å, and β = 113.6°) with one molecule in the asymmetric unit, were soaked in substrate for 10 min prior to being cryocooled and data collection.

Diffraction data were recorded at SRS Daresbury and ESRF Grenoble, and the crystallographic data are summarized in Table 1. Data were processed and reduced using MOSFLM (*20*) and SCALA (*21*), solved using MOLREP (*22*), and refined and visualized using REFMAC5/ARP (*23*, *24*) and O/COOT (*25*, *26*), respectively. In each case, calculated anomalous difference maps allowed the metal ions to be assigned unambiguously. Hexagalacturonate was produced as described previously (*27*). Three specifically methylated hexasaccharides (*28*−*31*) were also used: compound I, MMCCCM; compound II, CCMMMC; and compound III, MCCCCM (where M represents a methylated and C a nonmethylated galacturonate residue and the sequence is given starting from the nonreducing end of the hexasaccharide).

Isothermal Titration Calorimetry. Calcium ion binding to R279A BsPel and to R279A BsPel in the presence of 2 mM hexagalacturonate was assessed in 25 mM Tris-HCl (pH 7.5) and 100 mM NaCl at 25 °C using a MicroCal VP-ITC device. For the calcium titrations, the protein and hexasaccharide were dialyzed extensively against buffer supplemented with 1 mM EDTA and then against EDTA-free buffer before degassing and use. Heats of dilution were negligible for the titration of calcium into buffer and relatively small for the titration of calcium into hexasaccharide. The BsPel concentration used in these experiments was 0.13 mM.

RESULTS

Characterization of Mutant Proteins. We initially produced five mutants of BsPel to find an inactive but structurally unperturbed variant in which we could form enzyme−substrate complexes in the crystal (Table S1 of the Supporting Information). The residues chosen for mutation were the putative catalytic base, R279, and the three aspartates, 184, 223, and 227, that bind the single calcium seen in native BsPel. The R279A mutant had virtually no activity and only a minor perturbation of the active center, an increase in the mobility of Ile 250, and was

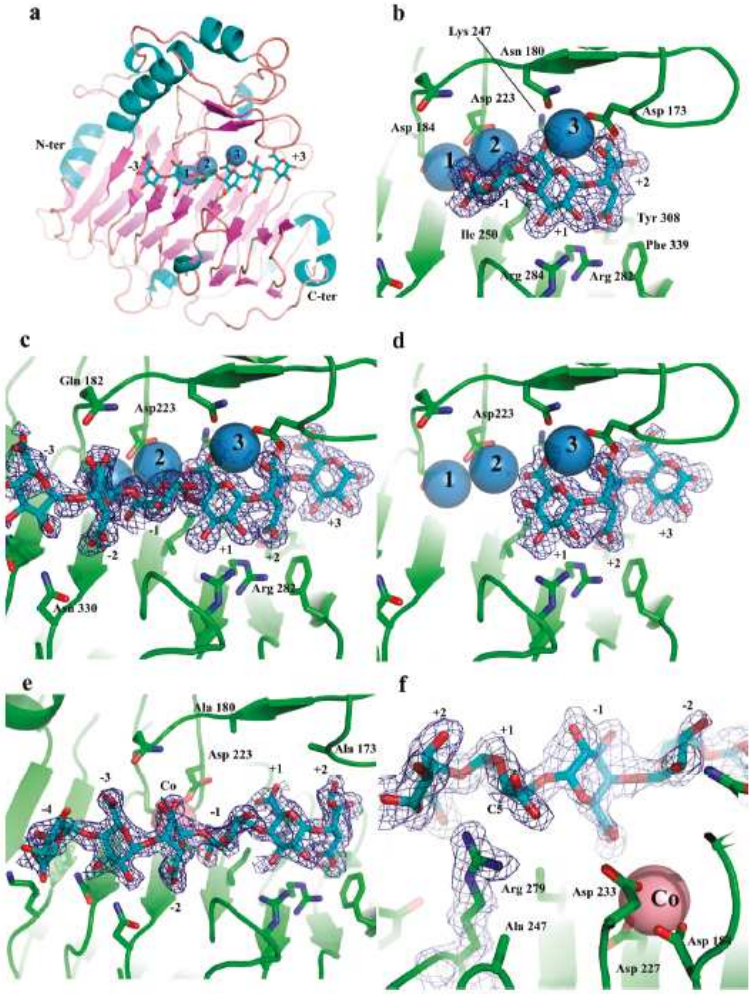


FIGURE 1: Pectin oligosaccharide binding to *B. subtilis* pectate lyase. Complexes shown in panels b–d were formed in crystals of the R279A (catalytic base mutant) form of BsPel. The complex shown in panels e and f was formed using the triple mutant (N180A/D173A/K247A) and cobalt in place of calcium. The maximum-likehood/σ_A-weighted $2F_o - F_c$ syntheses are contoured at 0.5σ for all apart from the GalA3 complex which is contoured at 1.0σ. All are shown as blue wire mesh. Calcium ions are shown as blue spheres, and cobalt is shown as the magenta sphere. (a) Cartoon representation of the parallel β-helix architecture of BsPel with β-strands and α-helices represented as magenta arrows and cyan helices, respectively. The cyan/red liquorice bonds represent bound hexasaccharide (pectin I). The oligosaccharide binds to the surface of the β-sheet known as PB1. The reducing end of the hexasaccharide binds toward the C-terminal end of the parallel β-helix. (b) GalA3, shown as cyan/red liquorice bonds, occupies subsites −1 to +2. Key side chains involved in binding are shown in stick representation. (c) Pectin I, a hexasaccharide with three consecutive carboxylates, binds to subsites −3 to +3. (d) Pectin II, hexasaccharide with only two consecutive carboxylates (see Materials and Methods), binds nonproductively to subsites +1 to +3. (e) GalA6 binding to the triple mutant (N180A/D173A/K277A) in the presence of cobalt (magenta sphere) which binds in place of calcium 1. The oligosaccharide binds to subsite +2 tightly but more loosely to subsites −4 to +1. The cobalt does not bind the −1 carboxylate directly, unlike calcium 1. (f) Enzyme's eye view of the complex shown in panel e. The experimental evidence for the location of the catalytic arginine and substrate is shown. Arginine 279 is ideally positioned to abstract the C5 proton from GalA binding in subsite +1. See the text for more details. This figure was produced using PyMOL (*47*).

therefore chosen for these studies. The R279K and D227A mutants had too much residual activity for preparation of complexes in the crystal. D223A had low activity and is a further candidate for preparation of the complexes in the crystal. Kinetic analysis of the wild-type enzyme showed that trigalacturonate was cleaved uniquely into unsaturated digalacturonate and saturated galacturonate, suggesting that trigalacturonate binds productively to subsites +2, +1, and −1. Digalacturonate was not cleaved at an appreciable rate and therefore presumably binds nonproductively or not at all. At pH 8.5 and in the presence of 5 mM Ca^{2+}, K_M and k_{cat} for trigalacturonate were 1.2 mM and 340 s^{-1}, respectively. K_M and k_{cat} for tetragalacturonate were 0.19 mM and 1000 s^{-1}, respectively; the k_{cat} value is similar to that measured using polygalacturonate (1200 s^{-1}). If the stability of pectin is on the order of 10 million years (*2*), then the rate enhancement is at least 10^{17}-fold.

Enzyme–Substrate Complexes Formed in the R279A Crystals. The inactive but structurally unperturbed mutant of

BsPel (R279A) was used to produce five enzyme−substrate complexes in the crystal (crystallographic details are given in Table 1). In each of these complexes, the substrates bind in the cleft formed by the surface of parallel β-sheet 1 and the adjacent loops, and two additional calcium ions are bound between the enzyme and oligogalacturonate (Figure 1a). Trigalacturonate binds to subsites −1, +1, and +2 (Figure 1b), in agreement with the measured kinetic data that showed that the product was an unsaturated digalacturonate and saturated galacturonate. Hexagalacturonate and pectins I and III bound subsites −3 to +3 (Figure 1c). Pectin II with only two consecutive GalA residues does not bind productively and occupied subsites +1 to +3 (Figure 1d). Continuous electron density was seen in maximum-likehood/σ_A-weighted $2F_o - F_c$ synthesis for GalA3. Five residues of the six were seen for GalA6 at 1.0σ with the sixth (subsite +3) weaker (mean *B* factors are given for each of the subsites in Table 2); five of six residues were seen for both pectins I and III at 1.0σ with weaker electron density for subsites away from the catalytic center. Maximum-likehood/σ_A-weighted $2F_o - F_c$ syntheses contoured at 1.0σ are shown for GalA3 and 0.5σ for pectin I and pectin II binding in Figure 1. The pectins superimpose on each other closely so that a description of the binding of subsites −1, +1, and +2 of pectin I is appropriate for all four productive complexes. The GalA carboxylate occupying the +1 subsite is bound by two calcium ions and K247 (Figures 1b and 2a). Both C2 and C3 hydroxyl groups form hydrogen bonds to R284 with the C2 hydroxyl close to the carboxyl binding subsite +2. At subsite +2, O3 binds the third calcium and the carboxylate oxygen atoms form hydrogen bonds to the NE and NH2 atoms of R282. At subsite −1 (leaving group), O5 binds the second calcium ion and the carboxylic O6A atoms bind to both the first and second calcium ions (Figures 1 and 2a and Table S2 of the Supporting Information).

Hexagalacturonate binds similarly to trigalacturonate, but in addition, subsites −2, −3, and +3 are occupied. Subsites −2 and −3 make fewer contacts with the enzyme than central sites +2, +1, and −1, and this is reflected in the mean *B* factors (Table 2). The partially methylated pectins I and III bind in a manner similar to that described for hexagalacturonate. Only subsites +3 to +1 are occupied for compound II, revealing that three consecutive carboxylates are needed for productive binding. No methyl groups were seen in the electron density for any of the specifically methylated compounds, probably due to rotation of

Table 2: Summary of the Binding of Pectic Oligosaccharides and Calcium to Pectate Lyase[a]

enzyme	substrate	−3	−2	−1	+1	+2	+3	metal
R279A	GalA3			C	C	C		3 Ca^{2+}
	B factor			9.7	8.8	9.3		
R279A	GalA6	C	C	C	C	C	C	3 Ca^{2+}
	B factor	41.3	28.2	13.1	12.4	13.9	30.0	
R279A	pectin I	M	M	C	C	C	M	3 Ca^{2+}
	B factor	47.9	37.5	17.2	10.9	9.4	22.3	
R279A	pectin II				C	C	M	3 Ca^{2+}
	B factor				28.6	25.5	34.7	
R279A	pectin III	M	C	C	C	C	M	3 Ca^{2+}
	B factor	46.1	34.9	17.7	10.4	8.9	24.9	
triple mutant	GalA6	C	C	C	C	C		1 Co^{2+}
	B factor	37.7	44.3	38.9	34.7	22.1		

[a]C and M indicate GalA and GalAMe, respectively. There is a strong preference for GalA at the +2, +1, and −1 subsites. Binding to the −3 subsite is seen only at low contour levels in the electron density maps. The *B* factors given here are the mean *B* factors of the atoms in the stated subsite in square angstroms. The isotropically refined *B* factor (*B*) is related to the mean square displacement of the atoms ($\langle u \rangle$) by the equation $B = 8\pi^2\langle u^2 \rangle$. The sixth GalA of GalA6 bound to the complex formed in the absence of calcium (Triple mutant) occupies the −4 subsite (see Figures 1e and 2b).

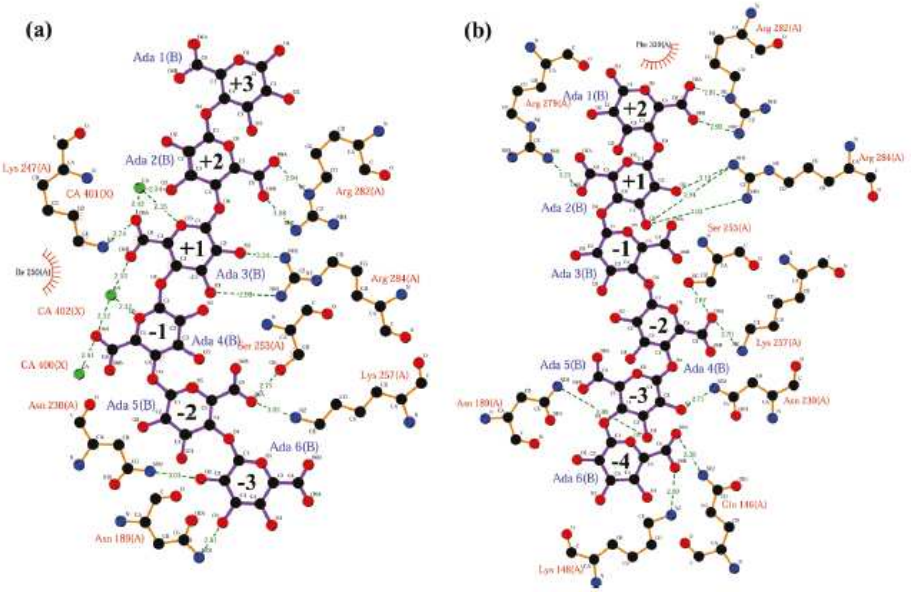


FIGURE 2: Schematic representation of the binding of hexaglacturonate to the R279A mutant and triple mutant in the absence of calcium (a and b, respectively). The molecular structures with corresponding density are shown in panels c and d of Figure 1, respectively. This figure was produced using LIGPLOT (*48*). Ligand bonds are colored blue and protein bonds brown; hydrogen bonds are drawn as dotted lines, and protein side chains involved in hydrophobic interactions are colored brown (for more information, see Figures S1−S4 of the Supporting Information).

Table 3: Torsion and Bond Angles about Selected Glycosidic Bonds[a]

oligosaccharide	subsites	ϕ (deg)	ψ (deg)	τ (deg)
GalA3	+1 to +2	115	153	124
	−1 to +1	63	58	130
pectin I	+1 to +2	118	151	118
	−1 to +1	75	45	140
pectin III	+1 to +2	119	150	120
	−1 to +1	70	52	137
PelC·GalA5	+1 to +2	117	157	117
PelC·GalA5	−1 to +1	73	51	120

[a]Torsion angles ϕ and ψ are about the C1−O4 and O4−C4 bonds, respectively. They are defined as O5′−C1′−O4−C4 (ϕ) and C1′−O4−C4−C5 (ψ), where the prime indicates atoms in the sugar residue toward the nonreducing end of the oligosaccharide. τ is the C1−O4−C4 bond angle. The ϕ and ψ angles of approximately 120° and 150°, respectively, correspond a 2_1 helical conformation of pectin (*18*, *44*−*46*). The ϕ and ψ angles of approximately 70° and 50°, respectively, correspond to a less-relaxed 6_1 helical conformation, if the angles are repeated. The values are shown for the complexes at 2.0 Å resolution, or better, and for the glycosidic bonds between the best-defined subsites −1, +1, and +2 as determined from analysis of *B* factors (Table 3). The values for PelC·GalA5 were taken from Table 1 of ref *18*, where the structure was refined at 2.2 Å resolution.

the methyl groups about the ester bond. The presence of the methyl groups is therefore inferred from the chemical structure of the specifically methylated hexasaccharides rather than by direct observation. LIGPLOT diagrams for the two GalA6 complexes appear in Figure 2 and for the other four complexes in the Supporting Information.

Calcium Binding Sites. Anomalous difference maps clearly revealed the presence of three calcium ions in the five enzyme−substrate complexes and a single calcium ion bound to the enzyme in absence of oligosaccharide. The first calcium binding site makes contacts with three aspartates (D184, D223, and D227), one water, and the carboxylate O6A atom of the GalA binding to subsite −1, although this is not a tight contact (Figure 1 and Table S2). D223 is the major contributor to the second calcium-binding site which binds tightly to the ring O5 and O6A atoms of the GalA in the −1 subsite and to the O6B atom of the GalA in the +1 subsite. The third calcium binds tightly to the O6A atom of the GalA in the +1 subsite and to the ring oxygen and carboxylate oxygen of the sugar residue in the +2 subsite. The stoichiometric enzyme:substrate:calcium ratio is 1:1:3 for these productive binding substrate complexes, and the structural and thermodynamic results together reveal that the three calcium ions have clearly defined roles in binding and catalysis. The mutational results suggest that changes to the environment of calcium 1 have significantly less impact on catalysis than those that affect calcium 2 (Table S1 and Figure 1) and that can be understood in terms of the number of tight contacts with calcium 2 in the enzyme−substrate complexes.

Conformation of the Pectic Oligosaccharides. The conformation of the pectic oligosaccharides bound in the substrate binding cleft adopts the expected relaxed 2_1 or 3_1 helical torsion angles with the important exception of those in the vicinity of the scissile bond (Table 3). If these torsion angles were repeated, the pectin helix would be approximately a 6_1 helix because when one looks down the chain the adjacent carboxylates in subsites −1 and +1 are rotated by approximately 60° with respect to one another rather than the 180° of the 2_1 helix or the 120° of the 3_1 helix. Evidence for distortion of the scissile bond comes first from a consideration of torsion angle ψ (C1′−O4−C4−C5, where prime indicates the GalA to the nonreducing side) which is within 45° of *cis* for the scissile bond as opposed to within 30° of *trans* for the other bonds and in the relaxed conformation of pectin (Table 3). Additionally, bond angle τ about the glycosidic oxygen (C1′−O−C4) is distorted from the anticipated value of 110−120° seen for the nonscissile bonds to 130−140° for the scissile bond (Table 3); this distortion is consistent with bringing the C1′ and C5 atoms close as a result of the 45° ψ angle. One must be careful in interpreting distortion observed in 1.8−2.0 Å ligand complex structures, but the conformation of the pectic oligosacharide is unusual and could result in distortion of the scissile bond.

Complex Formed in the Absence of the Calcium Ions. In an attempt to form a complex in the presence of the catalytic base (R279), the side chains forming the third calcium site (D173 and N180) were removed. The idea was that in the absence of one of the catalytic calcium ions the rate might be sufficiently low to enable the complex to be seen in the crystal with a catalytic base in place. The single mutants D173A and N180A and the double mutant D173A/N180N resulted in approximately 40, 30, and 5% of wild-type activity, respectively (Figure 1 and Table S1). K247 is also close to the substrate carboxylate bound by the third calcium. When the first two mutations were combined with a third mutation, K247A, to form the triple mutant, the activity dropped to 0.2% (Table S1). This residual activity still proved to be too great to produce complexes in the crystal. However, in the presence of 2 mM colbalt chloride in place of 1 mM calcium chloride, the complex could be trapped. This complex spans subsites −4 to +2 but is not as intimately bound to the active center as the other enzyme−substrate complexes (Figure 2), most obviously because the cobalt does not bind the −1 carboxylate and the two calcium ions bridging the substrate and enzyme carboxylates are absent (Figures 1 and 2). The conformation of the bound hexagalacturonate is closely similar to that seen in complexes in the presence of the two additional calcium ions, so the binding cleft is still distorting the pectin from the relaxed conformation in the region of the scissile bond despite the lack of oligosaccharide−calcium interactions. R279, clearly seen in the electron density map, is ideally positioned to abstract the C5 proton (Figure 1f).

Thermodynamic Measurements. Calcium binding to the first site of BsPel is stoichiometric and enthalpically driven with a K_d of 2.3 ± 0.1 μM (Figure S5 of the Supporting Information). Calcium binding to BsPel in the presence of a substrate gives a more complex binding curve, which cannot be fitted as a single process. As the crystal structure of the resulting complex is known (Figure 2a), the events can be interpreted in terms of the crystal structure. The binding can be modeled as a two-step process with first one calcium and then two additional calcium ions binding. This would correspond to the binding of calcium to site 1 first, followed by additional calcium ions binding to sites 2 and 3 as the enzyme−substrate complex is formed. Although the affinities are not accurately defined by the data, the first step is enthalpically driven and has an affinity comparable to that of the binding of calcium to the enzyme in the absence of substrate, and the second shows lower-affinity binding driven by entropic changes. (Figure S5). The thermodynamic parameters obtained from fitting the experimental data are listed in Table S3 of the Supporting Information.

DISCUSSION

Substitution of the catalytic arginine with alanine resulted in an enzyme with virtually no catalytic activity. Binding of pectic

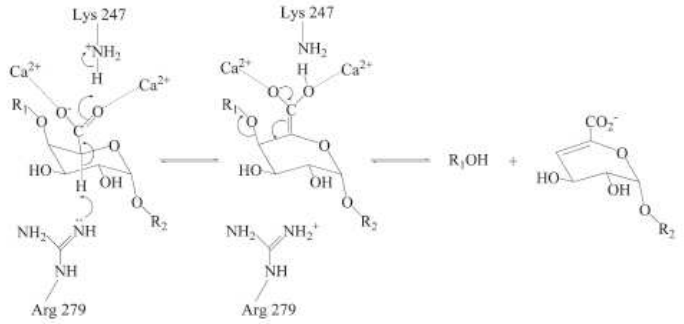


FIGURE 3: Plausible two-step mechanism for BsPel: (1) initial concerted abstraction of the α-proton and protonation of the carboxylic acid to generate an enol intermediate followed by (2) elimination of the β-leaving group from the enol. The intermediate is stabilized by the two catalytic calcium ions (calcium ions 2 and 3 which bind between the enzyme and substrate). One of these calcium ions may facilitate protonation of the leaving group.

oligosaccharides to crystals of the inactive enzyme showed that three consecutive substrate GalA residues are essential for productive binding to the calcium ions at subsites −1 and +1 and to the arginine at subsite +2. Outside these central subsites, methylated GalA residues can be accommodated. It is reasonable therefore to conclude that pectate lyase can cleave a pectin molecule with three of more consecutive GalA residues. The complex formed in crystals of the triple mutant protein lacking the calcium ions highlights the importance of the +2 subsite as the key substrate anchor. In fact, in every complex examined, the +2 subsite is occupied by GalA and the mean *B* factor for the +2 subsite is the lowest except for that of GalA3 where subsites −1 to +2 have similar values (Table 2). Pectin II binds to subsites +1, +2, and +3 yet still binds the catalytic calcium ions despite not binding to subsite −1. This binding mode is dictated by the preference of subsite +2 for GalA and the lack of three consecutive carboxylates in the oligosacharide. The importance of subsite +2 as the substrate anchor is further underlined by the observation that the same interaction between substrate carboxylate and arginine is seen in the family PL10 and PL3 complexes.

BsPel binds a single calcium ion, calcium 1, with an affinity (K_d) of 2.3 μM. This is substantially tighter than that measured indirectly for binding of the corresponding calcium to PelC (*32*). When calcium is titrated into the enzyme in the presence of substrate, two processes can be distinguished. The first is enthalpically driven with an affinity similar to that of the binding of the first calcium to BsPel; it is entirely plausible that this corresponds to the binding of the first calcium to BsPel. It is also plausible that the second process corresponds to the binding of the two additional calcium ions to form the Michealis complex. Remarkably, this second process is entropically driven, presumably due to the release of water from the surface of the hexasaccharide, calcium ions, and enzyme as the enzyme−substrate complex forms. Unlike most protein−carbohydrate interactions where there are aromatic residues involved in hydrophobic interactions, here there is an anionic substrate, a charged substrate binding cleft harboring a calcium ion, and two additional calcium ions that mediate the interactions between the enzyme and substrate. There is a single aliphatic hydrophobic residue, Ile 250, in the central part of the substrate-binding cleft of BsPel. Water molecules are released from the sugars binding the central subsites (+2, +1, and −1) and from the surface of the enzyme, and these are expected to contribute to the favorable entropic interaction. Water molecules are also released from the two calcium ions that are bound in the complex. The other binding sites are more waterlogged, and presumably favorable entropic changes are less significant here.

In the presence of calcium, the homogalacturonan region of pectin assumes a 2_1 helical conformation in dilute polymer solutions (*33*, *34*) and a 3_1 helix at higher concentrations in the gel or solid phase (*35*, *36*). More recent studies using nuclear magnetic resonance suggest the calcium−polygalacturonate complex in the plant cell wall contains both 2_1 and 3_1 helices as well as intermediate conformational states (*37*). Studies on di- and trisaccharides reveal similar helical conformations (*38*−*40*). There is clear evidence that the conformation of pectin is perturbed from relaxed 2_1 or 3_1 helical conformation by binding across the −1 and +1 subsites. The distortion from a helical 3_1 or 2_1 conformation is not simply a consequence of binding the second calcium because in the complex formed in the absence of calcium the conformation of the oligosaccharide is similar, showing the conformation is determined by the overall shape and characteristics of the binding cleft.

Catalysis by *B. subtilis* pectate lyase and by inference the family 1 polysaccharide lyases can be rationalized by a stepwise mechanism involving (i) initial concerted α-proton abstraction and protonation of the carboxylic acid to generate an enol intermediate followed by (ii) elimination of the β-leaving group from the enol (Figure 3). This is precisely as proposed by Gerlt and Gassman from consideration of the factors that would decrease the pK_a value of the α-proton and account for the observed rates of proton abstraction from a carbon adjacent to a carboxylate group (*41*, *42*). The crystallographic evidence for general base catalysis by R279 and general acid catalysis by R247 is strong, with both groups exquisitely poised for their respective functions (Figures 1 and 3).

The α-glycosidic linkage of the substrate and the C5 proton has ideal *anti* periplanar geometry for the *anti* β-elimination reaction catalyzed by pectate lyase to proceed. The conformation of the sugar in the +1 subsite must change from relaxed $^4C_1$ chair in the Michaelis complex to partially planar in the unsaturated product. A migration of C4 into the plane of C2, C3, and C5 would be the simplest change in conformation necessary for the reaction and would also be in accord with the principal of least nuclear motion during catalysis. Acidification of the α-proton by the catalytic calcium ions, especially the second calcium bound to both −1 and +1 carboxylates and D223, will lead to efficient transfer of a proton from C5 of the sugar in the +1 subsite to

R279. K247 appears to contribute to rate enhancement in the family 1 polysaccharide lyases by generating the more stable enol−enolate intermediate as opposed to the enolate−enolate intermediate, though it should be noted that the PL family 9 and 10 enzymes are active, albeit less so, with hydrogen bonding to the carboxylate from adjacent asparagines (*17*, *43*). A neutral lysine 247, neutral after protonation of the substrate carboxylate, proximal to the calcium ions, will also tend to stabilize the intermediate with respect to the Michaelis complex, thereby promoting catalysis.

The absence of an obvious group that protonates the glycosidic oxygen of the scissile bond is a concern because the alkoxide is a poor leaving group. The lone pair on the ring oxygen of the leaving group is *anti* periplanar to the C1−O4 bond, and the −1 GalA is intimately associated with calcium 2; these are both factors that will help to stabilize the leaving group. The closest group to either of the lone pairs of the glycosidic oxygen is the O3 hydroxyl from the sugar in the +1 subsite at a distance of 2.8 Å. The O3 hydroxyl might shuffle a proton from R284 and protonate the leaving group in a substrate-assisted protonation step that may be conserved between the family 1 and 10 lyases (*17*). However, given the intimate association of calcium and the GalA residue in the −1 subsite, protonation of the leaving group by solvent facilitated by calcium is also plausible.

ACKNOWLEDGMENT

We acknowledge the help given by Kathryn Worboys, Drummond Smith, Nicola Cummings, Ian Connerton, and John Jenkins in the early stages of this research and thank Alistair MacDougall for preparing the hexasaccharide used for the ITC studies. We acknowledge use of the ESRF Grenoble and SRS Daresbury for X-ray data collection.

SUPPORTING INFORMATION AVAILABLE

Additional results and analyses, Tables S1−S3, and Figures S1−S5. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Collmer, A., and Keen, N. T. (1986) The Role of Pectic Enzymes in Plant Pathogenesis. *Annu. Rev. Phytopathol. 24*, 383–409.
2. Wolfenden, R., and Snider, M. J. (2001) The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res. 34*, 938–945.
3. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): An expert resource for Glycogenomics. *Nucleic Acids Res. 37*, D233–D238.
4. Yoder, M. D., Keen, N. T., and Jurank, F. (1993) New domain motif: The structure of pectate lyase C, a secreted plant virulence factor. *Science 260*, 1503–1507.
5. Pickersgill, R. W., Jenkins, J., Harris, G., Nasser, W., and Robert-Baudouy, J. (1994) The structure of *Bacillus subtilis* pectate lyase in complex with calcium. *Nat. Struct. Biol. 1*, 717–723.
6. Akita, M., Suzuki, A., Kobayashi, T., Ito, S., and Yamane, T. (2001) The first structure of pectate lyase belonging to polysaccharide lyase family 3. *Acta Crystallogr. D57*, 1786–1792.
7. Jenkins, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N., and Pickersgill, R. W. (2004) The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*. *J. Biol. Chem. 279*, 9139–9145.
8. Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R., and Jenkins, J. (1997) Two crystal structures of pectin lyase A from *Aspergillus* reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyases. *Structure 5*, 677–689.
9. Vitali, J., Schick, B., Kester, H. C. M., Visser, J., and Jurnak, F. (1998) The three-dimensional structure of *Aspergillus niger* pectin lyase B at 1.7-angstrom resolution. *Plant Physiol. 116*, 69–80.
10. Petersen, T. N., Kauppinen, S., and Larsen, S. (1997) The crystal structure of rhamnogalacturonase A from *Aspergillus aculeatus*: A right-handed parallel β helix. *Structure 5*, 533–544.
11. Pickersgill, R., Smith, D., Worboys, K., and Jenkins, J. (1998) Crystal structure of polygalacturonase from *Erwinia carotovora ssp. carotovora*. *J. Biol. Chem. 273*, 24660–24664.
12. Jenkins, J., Mayans, O., Smith, D., Worboys, K., and Pickersgill, R. W. (2001) Three-dimensional structure of *Erwinia chrysanthemi* pectin methylesterase reveals a novel esterase active site. *J. Mol. Biol. 305*, 951–960.
13. Johansson, K., El-Ahmad, M., Friemann, R., Jornvall, H., Markovic, O., and Eklund, H. (2002) Crystal structure of plant pectin methylesterase. *FEBS Lett. 514*, 243–249.
14. Fries, M., Ihrig, J., Brocklehurst, K., Shevchik, V. E., and Pickersgill, R. W. (2007) Molecular basis of the activity of the phytopathogen pectin methylesterase. *EMBO J. 26*, 3879–3887.
15. Jenkins, J., and Pickersgill, R. (2001) The architecture of parallel β-helices and related folds. *Prog. Biophys. Mol. Biol. 77*, 111–175.
16. Abbott, D. W., and Boraston, A. B. (2007) A family 2 pectate lyase displays a rare fold and transition metal-assisted β-elimination. *J. Biol. Chem. 282*, 35328–35336.
17. Charnock, S. J., Brown, I. E., Turkenburg, J. P., Black, G. W., and Davies, G. J. (2001) Characterization of a novel pectate lyase, Pel10A, from *Pseudomonas cellulosa*. *Proc. Natl. Acad. Sci. U.S.A. 99*, 12067–12072.
18. Scavetta, R. D., Herron, S. R., Hotchkiss, A. T., Kita, N., Keen, N. T., Benen, J. A., Kester, H. C., Visser, J., and Jurank, F. (1999) Structure of a plant cell wall fragment complexed to pectate lyase C. *Plant Cell 11*, 1081–1092.
19. Herron, S., Benen, J. A., Scavetta, R. D., Visser, J., and Jurank, F. (2000) Structure and function of pectic enzymes: Virulence factors of plant pathogens. *Proc. Natl. Acad. Sci. U.S.A. 97*, 8762–8769.
20. Leslie, A. G. (1999) Integration of macromolecular diffraction data. *Acta Crystallogr. D55*, 240–255.
21. Evans, P. (2006) Scaling and assessment of data quality. *Acta Crystallogr. D62*, 72–82.
22. Vagin, A. A., and Teplyakov, A. (1997) MOLREP: An automated program for molecular replacement. *J. Appl. Crystallogr. 30*, 1022–1025.
23. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D53*, 240–255.
24. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol. 6*, 458–463.
25. Jones, T. A., Zou, J.-Y., and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A47*, 110–119.
26. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D60*, 2126–2132.
27. Rigby, N. M., MacDougall, A. J., Ring, S. G., Cairns, P., Morris, R., and Gunning, P. A. (2000) Observations on the crystallization of oligogalacturonates. *Carbohydr. Res. 328*, 235–239.
28. Clausen, M. H., Jorgensen, M. R., Thorsen, J., and Madsen, R. (2001) A strategy for chemical synthesis of selectively methylesterified oligomers of galacturonic acid. *J. Chem. Soc., Perkin Trans. 1*, 543–551.
29. Clausen, M., and Madsen, R. (2004) Synthesis of oligogalacturonates conjugated to BSA. *Carbohydr. Res. 339*, 2159–2169.
30. Clausen, M. H., and Madsen, R. (2003) Synthesis of hexasaccharide fragments of pectin. *Chem.—Eur. J. 9*, 3821–3832.
31. Petersen, B. O., Meier, S., Duus, J. O., and Clausen, M. H. (2008) Structural characterization of homogalacturonan by NMR spectroscopy-assignment of reference compounds. *Carbohydr. Res. 343*, 2830–2833.
32. Herron, S. R., Scavetta, R. D., Garrett, M., Legner, M., and Jurank, F. (2003) Characterization and implications of Ca^{2+} binding to pectate lyase C. *J. Biol. Chem. 278*, 12271–12277.
33. Morris, E. R., Powell, D. A., Gidley, M. J., and Rees, D. A. (1982) Conformations and interactions of pectins 1. Polymorphism between gel and solid states of calcium polygalacturonate. *J. Mol. Biol. 155*, 507–516.
34. Powell, D. A., Morris, E. R., Gidley, M. J., and Rees, D. A. (1982) Conformations and interactions of pectins 2. Influence of residue sequence on chain association in calcium pectate gels. *J. Mol. Biol. 155*, 517–531.
35. Walkinshaw, M. D., and Arnott, S. (1981) Conformations and interactions of pectins. 1. X-ray-diffraction analyses of sodium pectate in neutral and acidified forms. *J. Mol. Biol. 153*, 1055–1073.

36. Walkinshaw, M. D., and Arnott, S. (1981) Conformations and interactions of pectins. 2. Models for junction zones in pectinic acid and calcium pectate gels. *J. Mol. Biol. 153*, 1075–1085.
37. Jarvis, M. C., and Apperley, D. C. (1995) Chain conformation in concentrated pectic gels: Evidence from C-13 NMR. *Carbohydr. Res. 275*, 131–145.
38. Hricovini, M., Bystricky, S., and Malovikova, A. (1991) Conformations of (1−4)-linked α-D-galacturono-di-saccharides and α-D-galacturono-tri-saccharides in solution analyzed by NMR measurements and theoretical calculations. *Carbohydr. Res. 220*, 23–31.
39. Gouvion, C., Mazeau, K., Heyraud, A., Taravel, F. R., and Tvaroska, I. (1994) Conformational study of digalacturonic acid and sodium digalacturonate in solution. *Carbohydr. Res. 261*, 187–202.
40. Dinola, A., Fabrizi, G., Lamba, D., and Segre, A. L. (1994) Solution conformation of a pectic acid fragment by ^{1}H-NMR and molecular dynamics. *Biopolymers 34*, 457–462.
41. Gerlt, J. A., and Gassman, P. G. (1992) Understanding enzyme-catalyzed proton abstraction from carbon acids: Details of stepwise mechanisms for β-elimination reactions. *J. Am. Chem. Soc. 114*, 5928–5934.
42. Gerlt, J. A., and Gassman, P. G. (1993) Understanding the rates of certain enzyme-catalyzed reactions: Proton abstraction from carbon acids, acyl-transfer reactions, and displacement reactions of phosphodiesters. *J. Am. Chem. Soc. 115*, 11552–11568.
43. Jenkins, J., Shevchik, V. E., Hugouvieux-Cotte-Pattat, N., and Pickersgill, R. W. (2004) The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*. *J. Biol. Chem. 279*, 9139–9145.
44. Atkins, E. D. T., Nieduszy.Ia, Parker, K. D., and Smolko, E. E. (1973) Structural components of alginic acid. 2. Crystalline-structure of poly-α-L-guluronic acid: Results of X-ray-diffraction and polarized infrared studies. *Biopolymers 12*, 1879–1887.
45. Perez, S., Mazeau, K., and du Penhoat, C. H. (2000) The three-dimensional structures of the pectic polysaccharides. *Plant Physiol. Biochem. 38*, 37–55.
46. Braccini, I., Grasso, R. P., and Perez, S. (1999) Conformational and configurational features of acidic polysaccharides and their interactions with calcium ions: A molecular modeling investigation. *Carbohydr. Res. 317*, 119–130.
47. DeLano, W. L., Ed. (2002) The PyMOL molecular graphics system. 0.80 version. The PyMOL User's Manual, DeLano Scientific, San Carlos, CA
48. Wallace, A. C., Laskowski, R. A., and Thornton, J. M. (1995) LIGPLOT: A program to generate schematic diagrams of protein ligand interactions. *Protein Eng. 8*, 127–134.