

University of Bath



PHD

Directed evolution of *Thermus aquaticus* DNA polymerase by compartmentalised self-replication

Lamble, Sarah

Award date:
2009

Awarding institution:
University of Bath

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 22. May. 2019

**Directed Evolution of *Thermus aquaticus* DNA Polymerase by
Compartmentalised Self-Replication**

Submitted by

Sarah Ann Lamble

for the degree of PhD
of the University of Bath
2009

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with its author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and they must not copy it or use material from it except as permitted by law or with the consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purpose of consultation.

Table of Contents

Summary	1
Abbreviations.....	2
Chapter One - Introduction.....	3
Polymerase Chain Reaction (PCR)	3
History	3
Applications: PCR-based Diagnostics and DNA Sequencing	3
Recent Developments	6
DNA Polymerases	7
Natural Function of DNA Polymerases.....	7
Classification and Structural Overview.....	9
Properties.....	10
Thermophilic DNA Polymerases	11
Taq DNA Polymerase	13
Discovery and Characterisation.....	13
Applications.....	14
Structure	15
Related <i>Thermus</i> DNA Polymerases	16
PCR Inhibitors	17
Enzyme Engineering	19
Approaches to Engineering	19
Compartmentalised Self-Replication (CSR)	23
Aims of Project.....	25
Chapter Two - Compartmentalised Self Replication: Developing a System for the Directed Evolution of <i>Taq</i> DNA Polymerase	26
Introduction	26
Materials and Methods	30
Results	36
Expression and Activity Test of Sac7-TaqPol	36
SacB Cloning.....	37
Auto Induction.....	41
Emulsion PCR	41
Discussion	44
Chapter Three - Testing the CSR system by Evolving <i>Taq</i> DNA polymerase Variants with Resistance to Heparin	46
Introduction	46
Materials and Methods	49
Results	56
Random Mutagenesis	56
Production of the Mutant Library.....	58
CSR Heparin Screen.....	62
Selecting Heparin Resistant Polymerase Variants	64
Characterisation of Heparin Resistant Variants	69

Discussion	76
Chapter Four – Engineering <i>Taq</i> DNA polymerase resistance to Phytate by CSR.....	79
Introduction	79
Material and Methods.....	82
Results	86
CSR Phytate Screen.....	86
Selecting Phytate Resistant Polymerase Variants	88
Re-diversification and Production of Mutant Library Two.....	90
2nd Round CSR – Screening and Selection	91
Characterisation of Phytate Resistant Variants	96
Discussion	103
Chapter Five – Investigation of the Mechanism of phytate Inhibition of PCR 107	107
Introduction	107
Methods	110
Results	113
Investigation of Mg ²⁺ Chelation in Phytate PCRs.....	113
Investigation of Phytate/Enzyme Binding by X-ray Crystallography.....	115
Investigation of Phytate Inhibition by Site-Directed Mutagenesis	117
Discussion	123
Chapter Six – Perspectives and Future Work.....	127
Overview	127
Exploring the Applied Use of Novel <i>Taq</i> DNA polymerase Variants	128
Directed Evolution Studies.....	128
Biophysical Approaches to Investigate Phytate Interactions	130
References	131
APPENDIX A	143
Translation of the DNA sequence of Sac7- <i>Taq</i> Polymerase:.....	143
APPENDIX B.....	146
pTTQ18NAK-Sac7 <i>Taq</i> Pol plasmid construct.....	146
APPENDIX C	147
<i>Taq</i> DNA polymerase sequence alignment	147
Appendix D	148
1kb DNA ladder	148

Summary

The thermophilic enzyme, *Thermus aquaticus* (*Taq*) DNA polymerase, is an essential tool in molecular biology because of its ability to synthesis DNA *in vitro* and its inherent thermal stability. *Taq* DNA polymerase is widely used in the polymerase chain reaction (PCR), an essential technique in a broad range of different fields from academic research to clinical diagnostics. The use of PCR-based tests in diagnostic testing is ever increasing; however, many of the samples being tested contain substances that inhibit PCR and prevent target amplification. Many attempts have been made to engineer polymerases not only to increase resistance to overcome the problem of inhibition, but also to enhance other characteristics such as fidelity, processivity and thermostability.

Heparin, found in blood samples, and phytate, found in faecal samples, are two examples from a number of known PCR inhibitors. The mode of action of most PCR inhibitors is not well understood, but inhibition is thought to occur by enzyme binding or through the chelation of Mg^{2+} ions essential for PCR. In this project, a system of directed evolution by compartmentalised self-replication (CSR) was established and successfully employed to screen a mutant library for *Taq* DNA polymerase variants with enhanced resistance to the inhibitors heparin and phytate.

CSR is a recently-established high-throughput method for the creation of novel polymerases, based on a feedback loop whereby polymerase variants replicate their own encoding gene. A mutant library of 10^6 variants was produced by random mutagenesis error-prone PCR, in which only the polymerase domain of *Taq* was mutagenised. Firstly, the CSR system was established and tested by performing a screen in the presence of heparin to select for heparin-resistant variants. Characterisation of selected variants revealed that a single round of CSR had produced a *Taq* variant (P550S, T588S) with a 4-fold increase in heparin resistance. The IC_{50} was increased from 0.012U/ml heparin to 0.050U/ml heparin.

The study with heparin was followed by a phytate screen, in which two rounds of CSR were performed with an initial round of error-prone PCR followed by re-diversification (recombination) of the mutant library using the staggered extension process (StEP). The two rounds of CSR yielded a *Taq* variant with a 2-fold increase in phytate-resistance compared to the wild-type, with IC_{50} increased from 360 μ M phytate to 700 μ M phytate. The best phytate mutant (P685S, M761V, A814T) was further characterised and it was found that the catalytic activity, thermostability and fidelity of the mutant were comparable to the wild-type enzyme.

The position of resistance-conferring mutations of the novel *Taq* variants evolved in this study provided some evidence for the inhibitors' predicted modes of action in the case

of both phytate and heparin. As phytate's mode of action is poorly understood, further investigations were performed to elucidate its role in PCR inhibition. A thorough investigation into the importance of relative phytate and Mg^{2+} levels on PCR was conducted and revealed for the first time convincing evidence that the primary mode of phytate-mediated PCR inhibition is by chelation. Further work led to the successful crystallisation of *Taq* in the presence of phytate, although subsequent X-ray diffraction data to 2.5Å did not reveal phytate bound within the enzyme structure. Site-directed mutagenesis studies were used to probe cross-over between heparin and phytate-conferring mutations. Thus, in addition to providing valuable information for novel *Taq* variants with a potential application in fecal-based PCR diagnostic tests, this project has begun to provide insight into the fundamental aspects of the mode of action of phytate as a polymerase and PCR inhibitor.

Abbreviations

CSR – Compartmentalised self-replication

dNTP – Deoxynucleotide triphosphate

DMSO – Dimethyl sulfoxide

emPCR – Emulsion PCR

IPTG – Isopropyl β -D-1-thiogalactopyranoside

IVC – *In vitro* compartmentalisation

Kan – Kanamycin

LB – Luria broth

PCR – Polymerase chain reaction

PEI - Polyethyleneimine

PEG - Polyethylene glycol

Pol – Polymerase

SAP – Shrimp alkaline phosphatase

SDM - Site-directed mutagenesis

SDS-PAGE – Sodium dodecyl sulphate polyacrylamide gel electrophoresis

Taq DNA Pol – *Thermus aquaticus* DNA polymerase

Wt - Wild-type (Sac7-*Taq*Polymerase)

Chapter One - Introduction

Polymerase Chain Reaction (PCR)

History

The Polymerase Chain Reaction (PCR) was first developed by Kary Mullis and colleagues in the 1980s. When first introduced, PCR was a rapid and highly sensitive method of *in vitro* DNA amplification using the Klenow fragment of *Escherichia coli* DNA polymerase (Saiki *et al.*, 1985). It enabled for the first time the amplification of specific DNA sequences, producing billions of copies of the molecule in just hours. PCR is comprised of 3 steps: first the double-stranded DNA is separated or denatured at temperatures above 90°C, second the primers anneal to the DNA template at temperatures around 50-60°C and lastly primer extension occurs at around 72°C. Initially, PCR was hindered by the use of a polymerase that could not withstand the repeated high temperature incubations needed for the denaturation of the DNA duplex. This meant that fresh enzyme had to be added after every cycle, which was both time consuming and costly. Further investigation led to the use of the thermostable *Thermus aquaticus* (*Taq*) DNA polymerase in the reaction, which allowed continuous thermocycling and the development of PCR as we know it today (Saiki *et al.*, 1988). From this moment on, the use of thermostable DNA polymerases became key to PCR-based molecular biology techniques such as cloning and sequencing.

Applications: PCR-based Diagnostics and DNA Sequencing

Taq DNA polymerase (*Taq* DNA pol) revolutionised not only PCR but all research in biological sciences and medicine and has had a particularly huge impact on clinical diagnostics. PCR is routinely used in the diagnosis of disease, whether it is for the detection of genetic mutations or microbial or viral infectious agents. Further uses of PCR include the detection of antibiotic resistance genes (Dutka-Malen *et al.* 1995) and biothreat agents (Makino *et al.*, 2001), both of which are becoming more prevalent.

One of the first diagnostic uses of PCR was in a prenatal diagnosis test for sickle cell anaemia. The use of PCR for detection of the sickle cell mutation was more rapid and sensitive than previous methods (Saiki *et al.*, 1985). More PCR-based diagnostics followed, including tests for the detection of low-copy number viral targets such as HIV (Guatelli *et*

al., 1989), tests for the diagnosis of pulmonary tuberculosis through the detection of *Mycobacterium tuberculosis* (D'Amato *et al.*, 1995), and for the detection of different isolates of the gastrointestinal bacterium *Helicobacter pylori* (Akopyanz *et al.*, 1992). A selection of commercially available PCR-based diagnostic tests is listed in Table 1.1. Compared to traditional culture methods, PCR-based assays have shown increased speed, specificity and sensitivity in detection. This is particularly important in clinical diagnostics, where accurate and rapid answers are often needed in such places as hospital or veterinary laboratories.

The introduction of *Taq* DNA polymerase to PCR also led to its use in DNA sequencing (Innis *et al.*, 1988). The thermostability of *Taq* DNA pol facilitated the automated cycling of sequencing reactions, which simplified the procedure. It also allowed for higher annealing and extension temperatures which increased the specificity, yield and length of targets that could be amplified or sequenced (Innis *et al.*, 1988). The traditional Sanger dideoxynucleotide method (Sanger *et al.*, 1977) of DNA sequencing benefited from the use of *Taq* DNA pol because less DNA template was needed per reaction thus allowing the detection and sequencing of rare targets. DNA sequencing continued to advance over the next decade and the increase in available sequences led to a growth in the understanding of structure function relationships. In 1995 Tabor and Richardson found a single residue mutation that altered *Taq* DNA pol so that it could more efficiently incorporate the dideoxynucleotides needed in sequencing reactions (Tabor and Richardson, 1995). This discovery came after the molecular comparison of a modified bacteriophage T7 DNA polymerase to *Taq* DNA pol. The T7 polymerase was known to incorporate dideoxynucleotides more efficiently than other pol I type polymerases such as *E. coli* and *Taq* DNA polymerase (Tabor and Richardson, 1987). The critical residue in T7 polymerase that was needed to distinguish between deoxy- and dideoxynucleotides was transferred to the homologous position in *Taq* DNA pol, thus engineering an enzyme with enhanced properties for DNA sequencing.

Due to the development of automated sequencing, the number of sequenced genomes is increasing significantly every year, adding to the vast sequence knowledge that is already available. The first genome sequence of a pathogenic bacterium was published in 1995 and was that of *Haemophilus influenza* (Fleischmann *et al.*, 1995). With improvements in DNA sequencing techniques, the number of known genomes of infectious pathogens and catalogued genes is increasing, which is key to the growth of PCR in diagnostic testing. This wealth of sequence data is fundamental to the design of accurate targets in clinical diagnostic tests, as the use of PCR relies on the availability of at least a partial sequence of the target DNA, in order to design specific primers. Over the past decade, the increase in

available sequence data has been essential to the development of new PCR-based diagnostic tests.

Table 1.1 - Selected Commercial PCR-based Diagnostic tests.

NAME	DETECTION OF	COMPANY
Amplicor CT/NG Test	<i>Chlamydia trachomatis</i> <i>Neisseria gonorrhoeae</i>	Roche
Lightcycler Septifast	Species ID of pathogens causing blood stream infections	Roche
COBAS Ampliscreen HBV	Hepatitis B Virus (HBV)	Roche
COBAS Ampliscreen HCV	Hepatitis C virus (HCV)	Roche
COBAS Ampliscreen HIV-1	Human-immunodeficiency Virus type I	Roche
Quantifiler Human DNA Quantification Kit	Quantifies human DNA in sample	Applied Biosystems
AccuPower Toxoplasma Kit	Species of <i>Toxoplasma</i>	Krypton
AccuPower Rubella Kit	Rubella virus	Krypton
Mycosensor QPCR Assay Kit	Specific detection of most common <i>Mycoplasma</i> species	Stratagene

Recent Developments

Although PCR had become widely used, it still had some limitations as an analytical technique because traditionally it was difficult to quantify the DNA due to the fact that only the end product was analysed. This problem was solved when Higuchi *et al.* (1992) developed real-time PCR. This enhanced method of PCR detects the amount of product formed during the course of the reaction through the use of fluorescent stain (e.g. SYBR green), which is specific for dsDNA, or by fluorescent resonance energy transfer (FRET) probes. Three main FRET probes exist; 5' nuclease probes (TaqMan), molecular beacons and FRET hybridization probes. Of these TaqMan (Livak *et al.*, 1995) is probably the most well known and used. The TaqMan probe takes advantage of the 5'-3' nuclease activity of *Taq* DNA polymerase. The TaqMan fluorescent probe binds to the target sequence downstream of one of the primers; it is then cleaved by the 5' nuclease activity of *Taq* DNA polymerase as it extends along the DNA. The cleavage of the probe causes a fluorescent signal which is then measured. Real-time PCR makes it possible to calculate the starting amount of DNA in the reaction and the number of copies produced throughout the course of the PCR. This quantitative method of PCR has become useful in gene expression analysis and for monitoring viral loads (Mackay, 2004).

Optimization of real-time PCR reagents and equipments has enabled detection in as little as one hour (Barken *et al.*, 2007). There are constant modifications being made to PCR in order to try to increase the speed and sensitivity of the reaction even further. Such things as multiplex PCR, where multiple primer sets are used in a single reaction (Monstein and Ellnebo-Svedlund, 2002), and improvements to thermocyclers so that separate PCR reactions can be run simultaneously with different cycling conditions, are examples of some of the developments revolutionizing PCR and clinical diagnostics (Yang and Rothman, 2004). The PCR market is constantly developing; as genome databases expand the number of available PCR assays increases. As PCR proves itself to be a valuable method in diagnostics, its use is constantly expanding, developing and taking over from more traditional methods of diagnosis.

DNA Polymerases

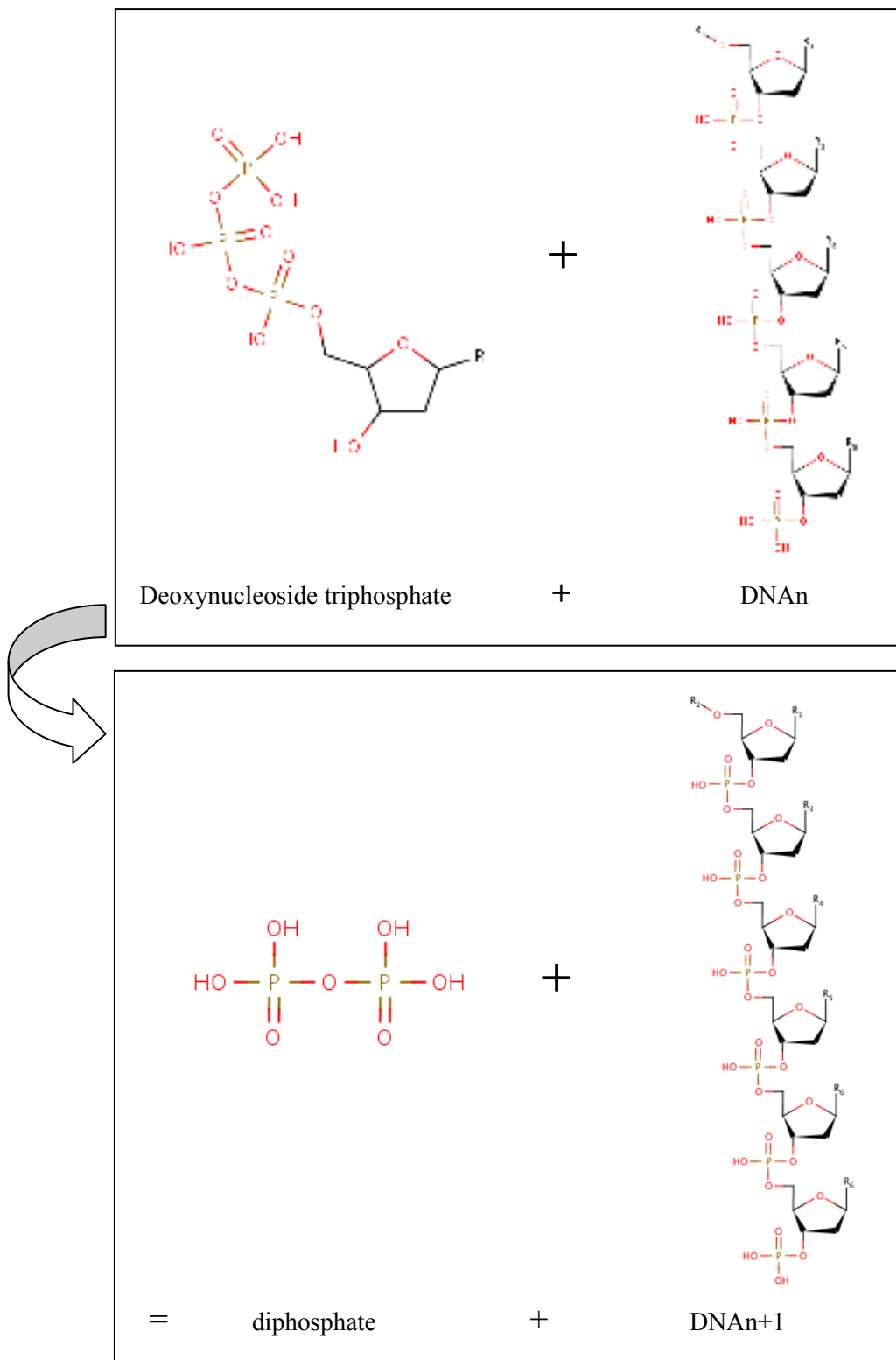
Natural Function of DNA Polymerases

In addition to being essential to PCR, in nature DNA polymerase (E.C. 2.7.7.7) has a fundamental role in the survival and preservation of all species. The ability of DNA polymerases to replicate DNA allows genetic information to be passed from one generation to the next and for DNA to be repaired. DNA polymerase catalyses the addition of a nucleotide onto the 3' end of a DNA primer (Figure 1.1), creating a new DNA strand complementary to the template strand. A two-metal ion mechanism has been proposed for polymerase-catalyzed nucleotidyl transfer reactions (Beese and Steitz, 1991). The two-metal ions, which are normally Mg^{2+} , act with two conserved acidic residues, usually aspartate, during catalysis (Steitz, 1999). All classes of polymerase use the same five-step scheme for nucleotide incorporation: (1) enzyme-nucleic acid complex binds the nucleoside triphosphate forming the ternary complex, (2) there is a conformational change, which orients the triphosphate ready for catalysis, (3) nucleotidyl transfer occurs, (4) a second conformational change occurs, which is followed by the final step (5) pyrophosphate release (Castro *et al.*, 2007). In addition to the catalytic polymerase domain, DNA polymerases often have other domains that are essential for editing, excision or interaction with other proteins (Alba, 2001).

In 1958, Lehman and colleagues were the first to report the enzymatic synthesis of DNA with a partially purified enzyme from *E. coli*. The enzyme responsible for deoxyribonucleotide incorporation was then named as a 'polymerase' (Lehman *et al.*, 1958). Since then the natural function of polymerase in DNA replication has been exploited in the laboratory. *In vitro* DNA polymerase is capable of primer-mediated enzymatic amplification of a target DNA sequence. Its capacity for template-directed DNA synthesis *in vitro* is what has led DNA polymerase to become a crucial tool in molecular biology.

Figure 1.1 – Reaction catalysed by DNA-dependent DNA polymerase¹

DNA polymerase catalyses the addition of a nucleotide onto the 3' OH of a growing DNA strand, which is followed by pyrophosphate release.



¹ Reproduced from BRENDA database <http://www.brenda-enzymes.info/>

Classification and Structural Overview

Numerous DNA polymerases have been isolated from a range of organisms including Archaea, Bacteria, bacteriophage, mitochondria, Eukaryotes and viruses (Braithwaite and Ito, 1993). There are six different families of DNA polymerases based on sequence similarity: A, B, C, D, X and Y. Thermophilic enzymes can be found in all six of the families. Viral DNA polymerases belong to family A or B and those involved in chromosome replication belong to B, C or D (Pavlov *et al.*, 2004). In Bacteria there are three main families, A, B and C, whilst most archaeal DNA polymerases belong to family B (Cann and Ishino, 1999).

In Bacteria there are three main types of polymerase, polI (*polA*), polII (*polB*) and polIII (*polC*), which are categorized by their function and homology to the DNA polymerases found in *E. coli* (Braithwaite and Ito, 1991). PolI represents Family A and is used in nucleotide excision repair and the processing of Okazaki fragments generated during DNA replication (Joyce *et al.*, 1982). Family A polymerases include *E. coli* polI, *Thermus aquaticus* DNA polymerase I and *Thermus flavus* DNA polymerase I. Some replicative DNA polymerases from bacteriophages (T3, T5 and T7) are included in this family as they have homology to Pol I (Alba, 2001). PolII represents Family B and is involved in the replication of damaged DNA (including cross-links caused by UV radiation). Family B includes archaeal polymerases from *Pyrococcus furiosus* (*Pfu*) and *Thermococcus litoralis* (vent). Lastly, polIII represents Family C and is the main replicative polymerase; it is made up of many subunits (Alba, 2001).

Most family A and B Bacterial polymerases are composed of a single subunit made up of three domains (Figure 1.2): the 3-5' exonuclease (proof-reading) domain, the 5-3' exonuclease (nick translation) domain and the polymerase domain. Although the polymerase domain is essential for catalytic activity, polymerases may also contain a 3-5' and/or a 5-3' exonuclease domain. The combination of domains will affect a polymerase's proof-reading ability. Despite differences in sequences, X-ray crystal structures have shown that all polymerases have a similar overall structure. In particular, the architecture of the polymerase domain is said to resemble a hand with 'thumb', 'finger' and 'palm' sub-domains as shown in Figure 1.2b (Steitz, 1999). Structural studies show that each of the sub-domains plays a different role in the nucleotidyl-transfer reaction. The active site, which typically contains three acidic residues, is located in the palm domain where catalysis of the phosphoryl transfer reaction occurs. The palm domain is the most conserved region among polymerase families (Pavlov *et al.*, 2004). The finger domain interacts with the incoming nucleoside triphosphate and the template base. Lastly, the thumb domain plays a role in positioning the duplex DNA (Steitz, 1999).

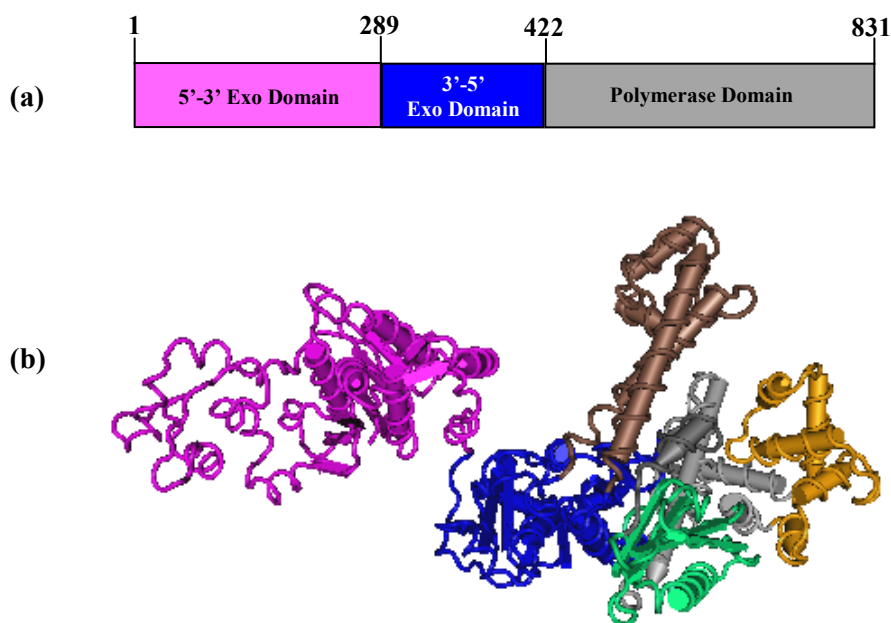


Figure 1.2. Organisation of *Taq* DNA Polymerase.

(a) Schematic diagram of domain layout of *Taq* DNA polymerase. Numbers correspond to amino acids. (b) *Taq* DNA polymerase domain organisation: 5-3' exonuclease domain (pink), proof-reading 3-5' exonuclease domain (blue), which is inactive in *Taq*, and the polymerase domain, which consists of palm (green/grey), finger (orange) and thumb (brown) subdomains (PDB ID: 1TAQ picture produced in Cn3D program).

Properties

Although DNA polymerases typically catalyse the same reaction of template mediated DNA extension, the properties of each polymerase can vary; some of these key properties are thermostability, processivity, fidelity and the rate of extension. Thermostability is characterised by the half-life of a DNA polymerase at a certain temperature. The importance of thermostability for PCR applications has previously been discussed. The fidelity of a DNA polymerase is the error rate of the enzyme and is based on the frequency of incorporation of incorrect nucleotides. Fidelity is important for techniques such as cloning where exact copies of the DNA template are essential. Processivity is defined as the probability of an enzyme not to dissociate from the DNA while it translocates to the next position after the addition of the nucleotide. Highly processive enzymes add more nucleotides in a single binding event compared to those with low processivity. A similar feature is the extension rate, which is characterised by the number of nucleotides added per second per molecule of DNA polymerase (Pavlov *et al.*, 2004). Since naturally

occurring polymerases have different thermostability, fidelity and processivity, their overall properties and therefore their applications are limited (Henry and Romesberg, 2005). The demand for enzymes to be tailored to specific applications has led to the rise of engineered polymerases with advantages over their naturally occurring equivalents. Methods of enzyme engineering will be discussed later in the chapter.

Thermophilic DNA Polymerases

Thermostability is a key polymerase property needed for the high temperature cycling involved in PCR. Thermostable enzymes are isolated from 'thermophiles' or thermophilic organisms. Thermophiles typically grow at temperatures between 55°C and 80°C, and hyperthermophiles grow above 80°C and can exceed growth temperatures of 110°C (Hough and Danson, 1999). The enzymes isolated from these organisms are extremely thermostable. Thermophiles, fall under the larger category of 'extremophiles', which are organisms that exist in extreme environments. There are many types of extremophiles that live in a wide range of habitats, such as acidophiles (low pH), halophiles (high salt concentration) and psychrophiles (low temperature) (Demirjian, *et al.*, 2001). In 1974 Macelroy first coined the term 'extremophiles' and since then interest in extremophiles has grown considerably and investigations have led to the discovery of many novel extremophiles in all three domains of life (Rothschild and Manicinelli, 2001). Enzymes isolated from extremophiles have many industrial applications as a result of their unique properties. Their inherent stability and capacity to function under conditions in which their mesophilic (moderate) equivalents could not, makes these enzymes a key part of the growing biotechnology industry.

Thermophilic DNA polymerases have played a key role in modern biotechnology ever since *Thermus aquaticus* (*Taq*) DNA polymerase, the first of its kind to be characterised, was discovered in 1976 (Chien *et al.*, 1976). The use of *Taq* DNA pol in PCR led to a rise in the isolation and cloning of further thermophilic DNA polymerases from a range of both archaeal and bacterial species. Discoveries are still being made to this day and recently a new highly thermostable DNA polymerase has been characterised from *Thermococcus* sp.NA1, a new hyperthermophilic strain isolated from a deep-sea hydrothermal vent (Jae *et al.*, 2007). *Pyrococcus furiosus* (*Pfu*) DNA polymerase (Uemori *et al.*, 1993), *Thermococcus litoralis* (Vent) DNA polymerase (Kong *et al.*, 1993) and *Pyrococcus* sp. strain KOD1 (KOD) DNA polymerase (Takagi *et al.*, 1997) are examples of other thermostable polymerases that are now frequently used in PCR and related applications. Less well known thermophilic polymerases have been isolated from the

archaeon *Sulfolobus solfataricus* (Pisani *et al.*, 1992) as well as from numerous *Geobacillus* species including *sterothophilus* and *caldotenax* (Sellmann *et al.*, 1992; Burrows and Goward, 1992). A list of selected thermophilic DNA polymerases and their properties can be seen in Table 1.2.

Table 1.2 – Selected thermophilic DNA polymerases and their properties

Thermophilic organism	DNA polymerase	Family	5'-3' exonuclease activity	3'-5' exonuclease activity	Half-life	Fidelity ¹
<i>Thermus aquaticus</i>	<i>Taq</i>	A	+	-	9min at 97.5°C ^A	2.1x10 ⁻⁴ ^B
<i>Thermus aquaticus</i> large fragment	Stoffel		-	-	21min at 97.5°C ^A	
<i>Thermus aquaticus</i> N-terminal deletion	Klentaq		-	-	Dependent on N-terminal deletion ^C	2.0x10 ⁻⁴ ^D
<i>Thermus thermophilus</i> HB8	<i>Tth</i>	A	+	-		3.0x10 ⁻⁵ ^E
<i>Thermococcus litoralis</i>	<i>Tli</i> or <i>Vent</i>	B	-	+	8h at 95°C ^F	2.8x10 ⁻⁶ ^G
<i>Pyrococcus woesei</i>	<i>Pwo</i>	B	-	+	>2h at 100°C ^H	3.2x10 ⁻⁶ ^E
<i>Pyrococcus kodakaraensis</i> *	KOD1	B	-	+	12h at 95°C ^I	3.5x10 ⁻³ ^I
<i>Pyrococcus furiosus</i>	<i>Pfu</i>	B	-	+	6h at 95°C ^I	1.3x10 ⁻⁶ ^G
<i>Rhodothermus marinus</i>	<i>Rma</i>	A	+	+	2min at 90°C ^J	
<i>Rhodothermus marinus</i> Large fragment			-	-	11min at 90°C ^J	
<i>Thermococcus</i> sp. NA1	TNA1	B			12.5h at 95°C ^K	2.2x10 ⁻⁴ ^K
<i>Sulfolobus solfataricus</i>	<i>Sso</i>	B	-	+	6min at 90°C ^L	

¹ errors per base duplication

* now *Thermococcus kodakaraensis* (Kuroita *et al.*, 2005).

^A (Lawyer *et al.*, 1993)

^B (Keohavong and Thilly, 1989)

^C (Barnes, 1995)

^D (Barnes, 1992)

^E (Frey and Suppmann, 1995)

^F (Kong *et al.*, 1993)

^G (Cline *et al.*, 1996)

^H Data derived from commercial sources (Roche)

^I (Takagi *et al.*, 1997)

^J (Blondal *et al.*, 2001)

^K (Jae *et al.*, 2007)

^L (Pisani *et al.*, 1998)

Taq DNA Polymerase

Discovery and Characterisation

The thermophilic Gram-negative bacterium, *Thermus aquaticus*, was first isolated from a hot spring in Yellowstone National Park (Brock and Freeze, 1969). *Thermus aquaticus* belongs to the Deinococcus-Thermus group (Figure 1.3) and has an optimum growth temperature of 70°C. It has since been found in other similar thermal environments. Over the years *Thermus aquaticus* has become the source of many different thermostable enzymes with potential applications including an aldolase (Freeze and Brock, 1970), an RNA polymerase (Air and Harris, 1974) and most importantly *Taq* DNA polymerase (Chien *et al.*, 1976).

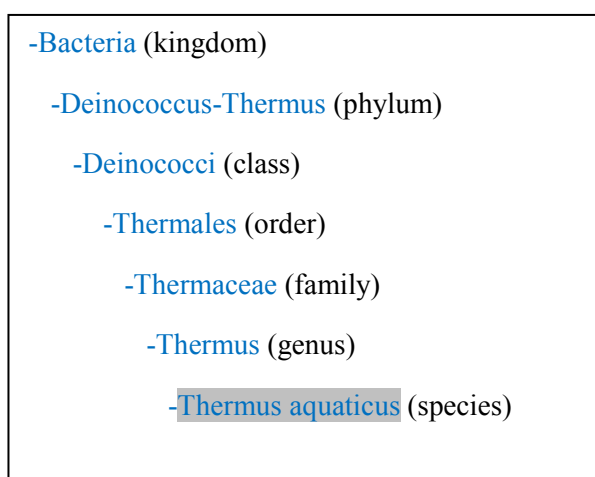


Figure 1.3 – Phylogeny of *Taq*

Thermus aquaticus (*Taq*) DNA polymerase was first isolated, purified and characterised by Chien *et al.* in 1976. It was the first thermophilic polymerase to be discovered and it was found to have a temperature optimum of 80°C (Chien *et al.*, 1976). In 1989 the *Taq* DNA polymerase gene was cloned and expressed in *E. coli*, with the pure, active form of the protein having a molecular mass of 94kDa (Lawyer *et al.*, 1989). *Taq* DNA pol was classified as a type I pol based on its homology to *E. coli* polI.

E. coli pol I was the first polymerase to be discovered in 1956 by Arthur Kornberg and colleagues (Kornberg *et al.*, 1956). Discovery of *E. coli* pol II and pol III followed shortly after, and it was then realised that Pol I was not the main replicative polymerase of the cell, a function performed by pol III (Kornberg and Gefter, 1971). Friedberg (2006)

outlined the discovery of DNA polymerase and the timeline that led these three enzymes (pol I, II and III) to form the basis for classification of family A, B and C polymerases that we use today. After the discovery of *Taq* DNA pol, many studies were carried out comparing it to *E. coli* pol I. Research showed that unlike *E. coli* pol I, portions of *Taq* DNA pol were not essential for DNA polymerase activity and it appeared that natural mutations had destroyed the 3-5' exonuclease activity leaving *Taq DNA pol* with no proof-reading ability (Lawyer *et al.*, 1989). Further studies showed that like *E. coli* pol I, *Taq* DNA pol could be cleaved into two fragments. In *E. coli* pol I the larger of the two fragments contains the polymerase and 3-5' exonuclease domain and the smaller fragment is made up of the 5-3' exonuclease domain (Klenow and Henningsen, 1970). The large fragment is known as the Klenow fragment and retains polymerase activity. The *Taq* DNA polymerase equivalent is the 61kDa Stoffel fragment, which has its 5-3' exonuclease domain removed and is the shortest form of an active *Taq* DNA polymerase (Lawyer *et al.*, 1993). An N-terminal deletion of *Taq* DNA polymerase can produce a slightly larger 67kDa protein called KlenTaq, which retains polymerase activity but lacks the 5-3' exonuclease domain (Barnes, 1992). Studies by Barnes (1992) showed that KlenTaq had a reduced tendency towards error compared to the full length *Taq* DNA polymerase.

Applications

Taq DNA polymerase has a major advantage over its equivalent *E. coli* pol I as it is thermostable. Whereas most Bacterial DNA polymerases are inactivated by high temperature, *Taq* DNA polymerase has the ability to withstand temperatures of up to ~100°C (Lawyer *et al.*, 1989), which lead to its pivotal role in PCR. *Taq* DNA pol is stable whereas *E. coli* pol I is inactivated at 93-95°C, which is the temperature required to denature duplex DNA as required in a PCR reaction. *Taq* DNA pol's inherent thermal stability was crucial for its application in the repeated rounds of PCR thermocycling.

Over the past twenty years, the high demand for thermostable DNA polymerases in the biotech industry has led to a great increase in their availability (Pavlov *et al.*, 2004). Selecting the correct polymerase variant for different PCR applications is an important part of a successful reaction, a property that is advantageous in one situation may be unfavourable in another. An example of this is the use of *Taq* DNA pol in error-prone PCR, which is used for the random mutagenesis of genes. Leung *et al.* (1989) reported that adding Mn^{2+} to a PCR instead of Mg^{2+} increased the error-rate of *Taq* DNA pol and could therefore be used to introduce random point mutations into genes. In this case, increasing the error rate of *Taq* DNA pol, which has been reported to be 2×10^{-4} mutations per base duplication

(Keohavong and Thilly, 1989), had its advantages. Later on a mutant of *Taq* DNA pol, with a single amino acid substitution, was designed to incorporate dNTPs less efficiently and used for mutagenic PCR (Patel *et al.*, 2001). Mixtures of non-proof-reading (e.g. *Taq*) and proof-reading (e.g. *Pfu*) polymerases are often used in PCR as *Taq* DNA pol is beneficial when amplifying longer templates but *Pfu* DNA pol solves the problem of fidelity by exhibiting proofreading activity (Barnes, 1994). Although 3-5' exonuclease activity increases fidelity, it can be a disadvantage as the 3-5' activity has been reported to attack unannealed primers causing a low yield in PCR (Barnes, 1992). *Taq* DNA polymerase, in its natural form or one of its many engineered counterparts, is still currently widely used despite the availability of numerous alternative thermophilic polymerases.

Structure

The crystal structure of full length *Taq* DNA polymerase (Kim *et al.*, 1995) and KlenTaq (Korolev *et al.*, 1995) have been known for some time and have greatly increased our knowledge of the enzyme. The three domains of full length *Taq* DNA pol form a strung-out arrangement as seen in Figure 1.2(b). Structural data revealed that the inactive 3-5' exonuclease domain of *Taq* DNA pol lacked 4 important carboxylate residues that are known to be essential for divalent metal binding and therefore 3-5' exonuclease activity in the *E. coli* Klenow fragment (Derbyshire *et al.*, 1988; Kim *et al.*, 1995). Further crystallisation studies demonstrated key DNA interactions in the active site (Eom *et al.*, 1996), and led to the discovery of conformational changes made by the enzyme that established the structural basis for nucleotide incorporation (Li *et al.*, 1998).

Crystal structures of KlenTaq bound to primer/template DNA and ddCTP revealed key interactions between the enzyme and its substrates. An important conformational change in the enzyme during nucleotide incorporation, which affects the tip of the finger domain, was also discovered. The enzyme was observed to have a closed form and an open form, in which a crevice was clearly visible. The transition between open and closed forms affects the orientation of the finger domain and specifically residue Tyr671, which is thought to act as a positioning device for the DNA (Li *et al.*, 1998). The DNA forms an S shape with one bend interacting with the palm domain at the active site and the second interacting with the thumb domain (Li *et al.*, 1998; Doublet *et al.*, 1998). The incoming nucleotide becomes buried within the crevice in the closed form of the enzyme. The base of the nucleotide stacks against Phe667 in close proximity to the three active site carboxylates (Asp785, Glu 786 and Asp610). In the closed form, two Mg²⁺ ions are octahedrally coordinated by the triphosphate moiety of the incoming nucleotide and carboxylate side

chains in the active site. Residues Asp610 and Asp785 are key to the metal ion coordination. Some of the key enzyme/substrate interactions are mentioned above; however, many more residues make important interactions in both the open and closed forms of the enzyme (Appendix A). The determination of such substrate/enzyme interactions provided important information for protein engineering and fundamental studies of *Taq* DNA polymerase.

Related *Thermus* DNA Polymerases

A number of polymerases have been isolated from other related *Thermus* species such as *Thermus thermophilus* and *Thermus flavus*. The isolation of various *Thermus thermophilus* strains was followed by the discovery of a thermostable polymerase in each of them. The *Tte* DNA polymerase was isolated from strain B35, which was found in hot spring water (Rechkunova *et al.*, 1998), and is very similar to the *Tth* DNA polymerase which was isolated from strain HB8 (Carballeira *et al.*, 1990). A third *Thermus thermophilus* strain, HB27, yielded a gene encoding Top DNA polymerase, which shares 87% amino acid identity with *Taq* DNA polymerase (Kim *et al.*, 1998). The polymerases isolated from *Thermus thermophilus* strains performed under similar PCR conditions to that of *Taq* DNA pol and were found to lack 3-5' exonuclease activity (Bechtereva *et al.*, 1989). Like *Taq*, *Tth* DNA polymerase has been shown to have detectable reverse transcriptase (RT) activity in the presence of MnCl₂ and has been reported to be 100-fold more efficient in a coupled RT/PCR than *Taq* DNA pol (Jones and Foulkes, 1989; Myers and Gelfand, 1991). As the two enzymes have significant sequence similarity it is not known why *Tth* DNA pol's RT ability is greater than that of *Taq* DNA pol's (Shandilya *et al.*, 2004).

Although the *Thermus* DNA polymerases share significant sequence homology and properties, their capacity for DNA amplification in the presence of certain PCR-inhibiting samples has been reported to vary. *Taq* DNA pol has been shown to be very sensitive to blood samples, whereas *Tth* and *Tfl* from *Thermus flavus* (Akmetzjanov and Vakhitov, 1992) were less affected. *Taq* and *Tfl* DNA polymerase were observed to be extremely sensitive to faecal-samples whereas *Tth* DNA polymerase showed a degree of resistance to faecal-inhibitors (Al-Soud and Radstrom, 1998). Another study reported *Tth* DNA polymerase to be more resistant to inhibitors than *Taq* DNA polymerase in both reverse transcription and PCR reactions (Poddar *et al.*, 1998). Despite being closely related, different thermophilic DNA polymerases can have subtle differences that enhance their capacity for DNA amplification under various conditions.

PCR Inhibitors

As discussed previously, PCR is widely used in diagnostic testing. However, many of the samples being tested contain substances that are inhibitory and reduce the efficiency of the PCR. Inhibitors can reduce the sensitivity of the PCR as well as preventing amplification completely. In general, biological samples cause a problem at three key stages in diagnostic testing; (1) they interfere with cell lysis during DNA extraction, (2) they degrade the sample DNA or prevent its capture, and (3) they inhibit polymerase activity, which prevents target DNA amplification (e.g. binding to the polymerase or chelation of essential Mg^{2+} ions). Many components have been reported to inhibit PCR but the mode of action for most is unclear (Wilson, 1997). Common inhibitors are found in a variety of samples, including body fluids, food components and environmental soil samples. Known PCR inhibitors include bile salts, phytic acid, haemoglobin, heparin, urea, phenolic compounds and humic acids (Al-Soud and Radstrom, 1998).

Different techniques have been developed to try to remove the inhibitors from samples but excess processing is not ideal. Some disadvantages of inhibitor removal techniques include the fact that they are time consuming, costly, and there is the potential for loss of target DNA during processing, further, they are sample specific so different methods are needed depending on the sample type. A better understanding of the modes of action of the inhibitors would allow further development of sample treatments prior to use in PCR (Al-Soud and Radstrom, 2001). One inhibitor removal technique is the use of the enzyme heparinase, which is frequently added to samples containing heparin before DNA extraction. Although the enzyme successfully eliminates heparin inhibition, it represents a costly procedure (Taylor, 1997). The addition of PCR facilitators or enhancers to the reaction has also been used as a means of overcoming inhibition. BSA has been shown to prevent inhibition of PCR amplification by hemin, iron chloride, tannic acids, fulvic acids, extracts from feces and freshwater or marine water (Kreader, 1996). Other PCR facilitators such as betaine and glycerol have also been used with mixed results (Al-Soud and Radstrom, 2000). Although PCR facilitators do in part solve the problem of inhibition, some optimization of facilitator concentration is always needed, which is time-consuming. Many inhibitors work by interfering with the DNA-polymerase interaction. There is therefore also a strong case for the development of inhibitor resistant polymerases as opposed to pre-treatment of samples or the addition of PCR facilitators.

Two possible modes of action for inhibitors that work by interference with the DNA-polymerase are either by competition with the DNA template (e.g. enzyme binding) or chelation of the Mg^{2+} ions, which are essential for polymerase activity. A common inhibitor, heparin, which is found in blood samples, is suggested to compete with the DNA template

for binding to the polymerase (Al-Soud and Radstrom, 2001). Studies indicate that heparin reversibly binds the DNA polymerases (DiCioccio and Srivastava, 1978). Phytate, an inhibitor found in fecal samples, is thought to inhibit by chelation of Mg^{2+} ions (Thornton and Passen, 2004) but may also act by competition with the DNA template; the true mode of action of phytate is unknown.

Heparin and phytate are the two main PCR inhibitors focused on in this project. Both compounds cause a significant problem in diagnostic testing. Heparin is an anticoagulant found in blood samples. Many clinical diagnostic tests that are used to diagnose infectious and genetic diseases are carried out using DNA or RNA extracted from heparinised whole blood samples. The presence of heparin in these samples severely interferes with the PCR reaction (Yokota *et al.*, 1999). The extent to which the PCR is affected depends on the dilution of the sample being tested, the concentration of the inhibitor present in the sample and the type of DNA polymerase used in the reaction. The amplification capacity of some polymerases varies based on the enzyme's sensitivity to the inhibitor (Al-Soud and Radstrom, 1998).

The same principles apply to the inhibitor phytate, which is found in fecal samples. Phytate is the main store of phosphorous in plants and is therefore found in most seeds. It makes up a key component of animal feed, but is also highly abundant in human diets and human cells (Raboy, 2007). The presence of phytate in fecal samples causes a problem in veterinary and medical testing. It has been shown to inhibit PCR on bovine fecal specimens (Thornton and Passen, 2004). Phytate is also likely to cause a problem in human fecal-based PCR testing, such as the colorectal cancer screening test from EXACT Sciences where DNA is amplified from a stool sample to detect mutations. Other PCR-based fecal tests that suffer from the presence of inhibitors include the detection of *Helicobacter pylori* in humans (Kabir, 2004) and the detection of the enteric bacterium *Lawsonia intracellularis* in pigs (Jacobson, *et al.*, 2004). For efficient PCR to occur, the inhibitors have to be removed from the sample or treated with enzymes such as phytase (Thornton and Passen, 2004).

The removal of inhibitors from samples before diagnostic testing is time consuming and labour intensive as well as adding an extra cost to the procedure. An alternative approach to solving the problem of PCR inhibition is to study the target of the inhibitors, which in many cases is thought to be the polymerase. These days there are many engineered DNA polymerases on the market with specific advantages over naturally occurring polymerases. One such example is Phusion[®] Blood Direct, an enzyme variant based on a DNA polymerase from a *Pyrococcus* strain, with an additional dsDNA binding domain, which functions to enhance the processivity of the enzyme. Phusion[®] DNA polymerase is reported to be resistant to inhibitors found in blood samples and therefore maintains its high

standard of activity even in blood-based PCRs (New England Biolabs²). The development of inhibitor-resistant polymerases like Phusion[®] Blood Direct, could aid in reducing the time spent preparing samples prior to diagnostic tests thus cutting overall costs. The improvement of biological sample screening could be essential as the use and availability of diagnostic tests increase.

Enzyme Engineering

Approaches to Engineering

The growth of the biotechnology industry has meant there is always a demand for enzymes with specific properties. Although the list of characterised enzymes is constantly increasing, these days there is a need for particular properties not found in enzymes isolated from their natural environments. Enzymes are used in many applications where they need to have high stability and catalytic activity or even display activity with non-natural substrates. With the increase in DNA manipulation techniques, enzyme engineering quickly becomes the most efficient approach to the production of enzymes with adapted characteristics.

Enzyme engineering can be performed in a number of ways using rational design, semi-rational design or directed evolution. Rational design uses functional and structural knowledge of an enzyme to predict the outcome of alterations to the amino acid sequence. Site-directed mutagenesis is then used to introduce these specific changes to the protein sequence. In addition to a comprehensive knowledge of the enzyme, this method involves an element of speculation and is often unsuccessful. Despite continuous efforts of structural molecular biologists and other experts, it is still not possible to reliably relate protein sequence changes to altered properties. However, one successful application of rational design was in the creation of a mutant *Taq* DNA pol used in automated DNA sequencing. One amino acid change produced a modified *Taq* DNA pol that was able to incorporate ddNTPs more efficiently than the wild-type *Taq* DNA pol (Tabor and Richardson, 1995).

Another approach to enzyme engineering is directed evolution, which has become the most widely used method. Directed evolution aims to mimic natural evolution through sequential rounds of random mutagenesis or recombination (Chen and Arnold, 1993). In this technique, a mutant library with a low level of mutations is formed and mutants are screened for selected properties (Figure 1.4). Mutant libraries can be formed in a number of ways including by combinatorial cassette mutagenesis (Black *et al.*, 1996), error-prone PCR

² Phusion DNA polymerase is developed and manufactured by Finnzymes Oy (Espoo, Finland) and distributed by New England Biolabs, Inc. (Hitchin, UK) <http://www.neb.com/nebecomm/products/productF-547.asp>

(Fromant *et al.*, 1995), staggered extension process (StEP) (Zhao *et al.*, 1998) and gene shuffling (Stemmer, 1994). Two of the techniques, error-prone PCR and StEP, were used in this project and will thus be described in further detail below. Error-prone PCR involves the misincorporation of nucleotides by a polymerase to create random mutations in the target gene. A number of things can be done to increase the rate of misincorporation: (1) the use of a polymerase without proofreading ability, (2) the use of unequal concentrations of the four deoxynucleoside triphosphates (dNTPs), (3) the use of Mn^{2+} instead of Mg^{2+} in the PCR and (4) increased PCR cycle number so that misincorporations accumulate (Fromant *et al.*, 1995). StEP is a method of *in vitro* recombination that uses multiple primed templates that undergo repeated cycles of denaturation and extremely short annealing/polymerase-catalysed extension steps (Figure 1.5). Multiple full-length templates are primed and then, due to abbreviated extension times, short fragments are created rather than full length genes. In each cycle, these growing fragments can anneal to different templates based on sequence complementarity and extend further to create recombined genes. This ‘template switching’ results in the growing DNA fragment to comprise a recombined ‘hybrid’ from multiple parental templates. Numerous cycles occur and StEP continues until a full-length gene product is formed; the end result is a recombined gene (Zhao *et al.*, 1998). epPCR and StEP offer distinct but powerful techniques for the production of variants and creation of mutant libraries.

The main challenge of directed evolution comes from the high quantity of mutants produced during library formation, which then need to be screened. Performing a high-throughput method of screening and selection is a key factor in successful enzyme evolution. The technique of directed evolution is now routinely used to improve existing enzymes, change enzyme selectivity and to evolve completely new enzyme function (Tao and Cornish, 2002).

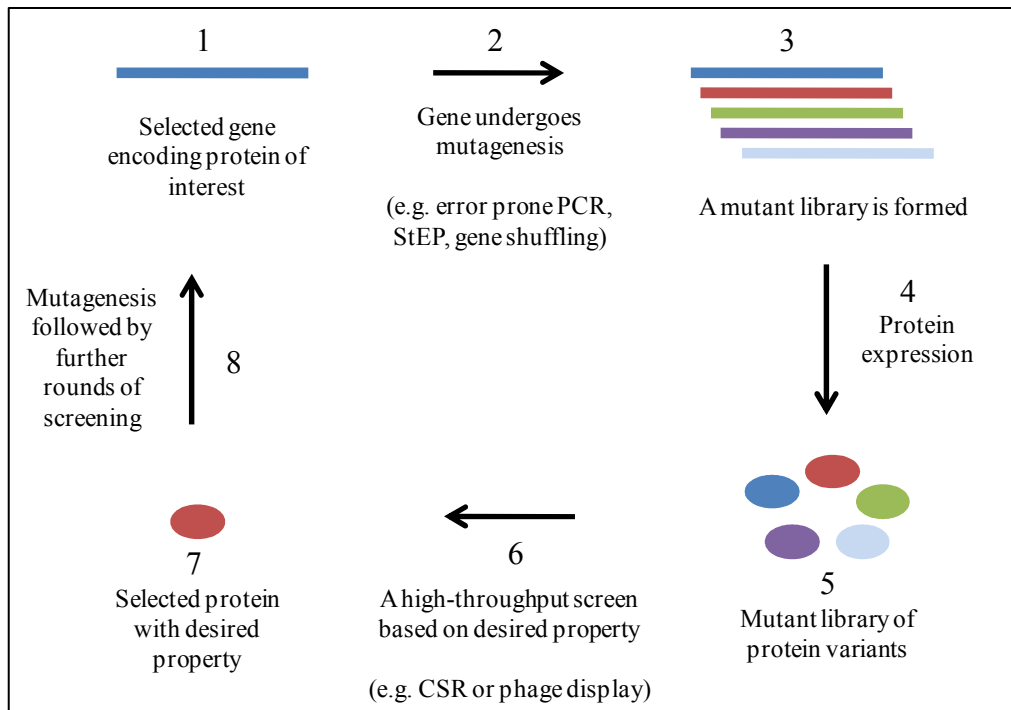


Figure 1.4 – Directed Evolution.

The gene coding for the protein of interest is selected (1) and the gene is mutated (2). A mutant library is formed (3). The mutants are then expressed (4), forming a library of protein variants (5). The protein variants are then screened based on a desired property (6). The variants containing enhanced properties are selected (7) and then carried forward for further rounds of mutagenesis and selection (8).

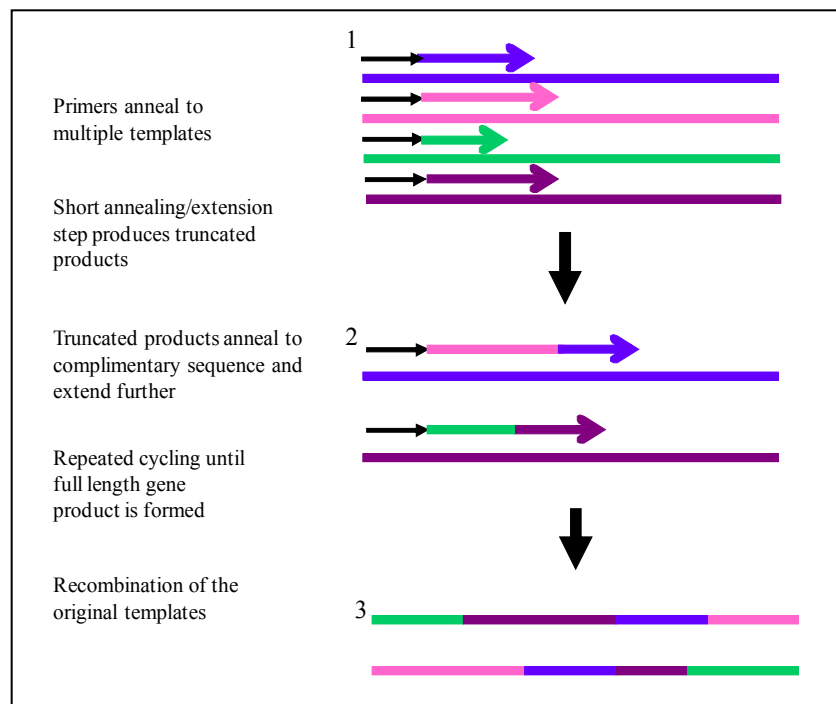


Figure 1.5 – Staggered extension process (StEP)

- 1) Primers anneal and polymerase produces short truncated products of varying length due to a short annealing/extension step
- 2) In the next cycle, these short products are used as primers and can anneal to complementary sequence in any of the templates
- 3) Cycling continues until full-length 'hybrid' templates are formed. The end result is millions of variants of recombined/ shuffled templates.

A combination of rational design and directed evolution, ‘semi-rational’ design, forms a third approach to enzyme engineering. In this method, structural and functional knowledge may be used to target a certain region of an enzyme, such as its active site (Chica *et al.*, 2005). Mutant libraries are created and screened but there is a focus on mutating specific residues, which have been chosen using all available knowledge of the enzyme. Saturation mutagenesis is often used to introduce all possible amino acid substitutions at the selected sites (Hayashi *et al.*, 1994; Miyazaki and Arnold, 1999), which is not always achieved through random mutagenesis. Saturation mutagenesis and a semi-rational approach produce a more focused library of mutants and can be useful when a high-throughput screen is not available.

When larger modifications to an enzyme structure are needed to alter function, then ‘domain swapping’ or ‘shuffling’ can be performed (Pavlov *et al.*, 2004). In this manner hybrid or chimeric DNA polymerases can be created. Davidson (2003) combined domains from two different polymerases to produce an enzyme with increased processivity. In this example, the T3 DNA polymerase thioredoxin-binding domain, which is known to play a key role in processivity, was inserted into *Taq* DNA polymerase to create a highly processive enzyme (Davidson *et al.*, 2003). Attempts have also been made to introduce the 3-5’ exonuclease function into *Taq* DNA pol by domain exchange. Two chimeric DNA polymerases were made by inserting the 3-5’ exonuclease domain of *E. coli* pol I and thermophilic *Thermotoga neapolitana* (*Tne*) polymerase in place of the inactive 3-5’ exonuclease domain of *Taq* DNA pol (Villbrandt *et al.*, 2000). Although the 3-5’ exonuclease function was introduced successfully, there was a reduction in thermostability despite the use of the thermostable *Tne* polymerase to construct the chimera. This highlights a problem with large domain swapping, which is that protein structure and interactions are likely to be disturbed when large segments of a protein are altered (Villbrandt *et al.*, 2000).

Two particular methods that have become popular for the *in vitro* selection of enhanced polymerases during directed evolution include phage display (Xia *et al.*, 2002) and compartmentalised self-replication (Ghadessy *et al.*, 2001). *In vitro* selection is challenging, as the association of phenotype and genotype must be maintained by some means. Phage display is a technique that allows the selection of active polymerases displayed on phage particles. This method has been used to evolve a variant of the Stoffel fragment of *Taq* DNA polymerase to have RNA polymerase activity and therefore incorporate ribonucleoside triphosphates (Xia *et al.*, 2002). In this example, selection of desired mutants relied on enzyme-catalysed labelling of the attached substrate. The labelled substrate is attached to the polymerase creating a phenotype-genotype linkage that is key to the procedure and phage particles displaying active polymerases are then isolated.

Compartmentalised self-replication (CSR) is a technique based on the compartmentalization of variant polymerases and a feedback loop whereby a polymerase can only replicate its own encoding gene. Only the polymerases that are active under the set conditions are able to replicate and are therefore selected. This method has been used to evolve a polymerase with an 11-fold higher thermostability than wild-type *Taq* DNA pol. The selected mutant had a half-life of 16.4 minutes at 97.5°C compared to only 1.5 minutes for the wild-type at the same temperature (Ghadessy *et al.*, 2001). This method was also used to screen for heparin-resistant *Taq* DNA pol variants, and was successfully used to isolate a variant that was reported to have a 130-fold increase in heparin resistance (Ghadessy *et al.*, 2001). The technique of compartmentalised self-replication is the chosen screening method used in the directed evolution of *Taq* DNA polymerase in this project.

Compartmentalised Self-Replication (CSR)

Directed evolution by compartmentalised self-replication (CSR) combines high-throughput screening with selection and is fast becoming a key procedure in enzyme engineering. This technology works particularly well with RNA and DNA polymerases. It can also be used for the evolution of other enzymes, such as nucleoside diphosphate kinases, although an extra step in CSR is often needed (Ghadessy *et al.*, 2001). CSR (Figure 1.6) is based on a feedback loop whereby each polymerase variant replicates its own encoding gene (Ghadessy *et al.*, 2001). After random mutagenesis to create a library, polymerase variants are isolated in compartments by the formation of a water-in-oil emulsion (Tawfik and Griffiths, 1998). Each compartment typically contains a single expressor *E. coli* cell containing a polymerase variant and this provides the link between phenotype and genotype that is so key to the procedure. PCR reagents are also contained within the compartment and combine with expressed polymerase enzyme when thermocycling causes cell lysis to occur. If the polymerase variant is expressed as active enzyme, or is active under the desired condition, then it will replicate its own gene, which increases the copy number of genes encoding an active polymerase during thermocycling (Ong *et al.*, 2006). Inactive polymerases that are incapable of self-replication will be lost from the gene pool and in this way the ‘sorting’ of variants becomes an inherent part of the procedure.

Along with the examples described above, CSR has been used to screen for DNA polymerases with the ability to extend mismatched bases, which therefore allows amplification of damaged DNA to occur (d’Abbadie *et al.*, 2007). A mutant library was created using the shuffling technique of ‘molecular breeding’ (Cramer *et al.*, 1998) in which fragments of the polymerase genes from three species, *Thermus aquaticus*, *Thermus thermophilus* and *Thermus flavus*, were recombined to form chimeric polymerases. CSR

was then used to screen the library of chimeric DNA polymerases and select for those variants that were able to extend primers with up to four mismatched bases. The end result was a DNA polymerase variant that was able to outperform wild-type *Taq* DNA pol in the amplification of ancient damaged DNA (d'Abbadie *et al.*, 2007). CSR has also been used to isolate *Taq* DNA pol variants that are able to incorporate not only dNTPs but also ribonucleotides (NTPs). In this study, short-patch CSR (spCSR) was used to diversify and replicate only a short region of the *Taq* DNA pol gene. Two specific motifs involved in substrate recognition were targeted and mutagenised and following screening by CSR, *Taq* DNA pol variants with apparent dual substrate specificity were selected (Ong *et al.*, 2006). These examples illustrate how directed evolution by CSR has successfully become a key technique in DNA polymerase engineering.

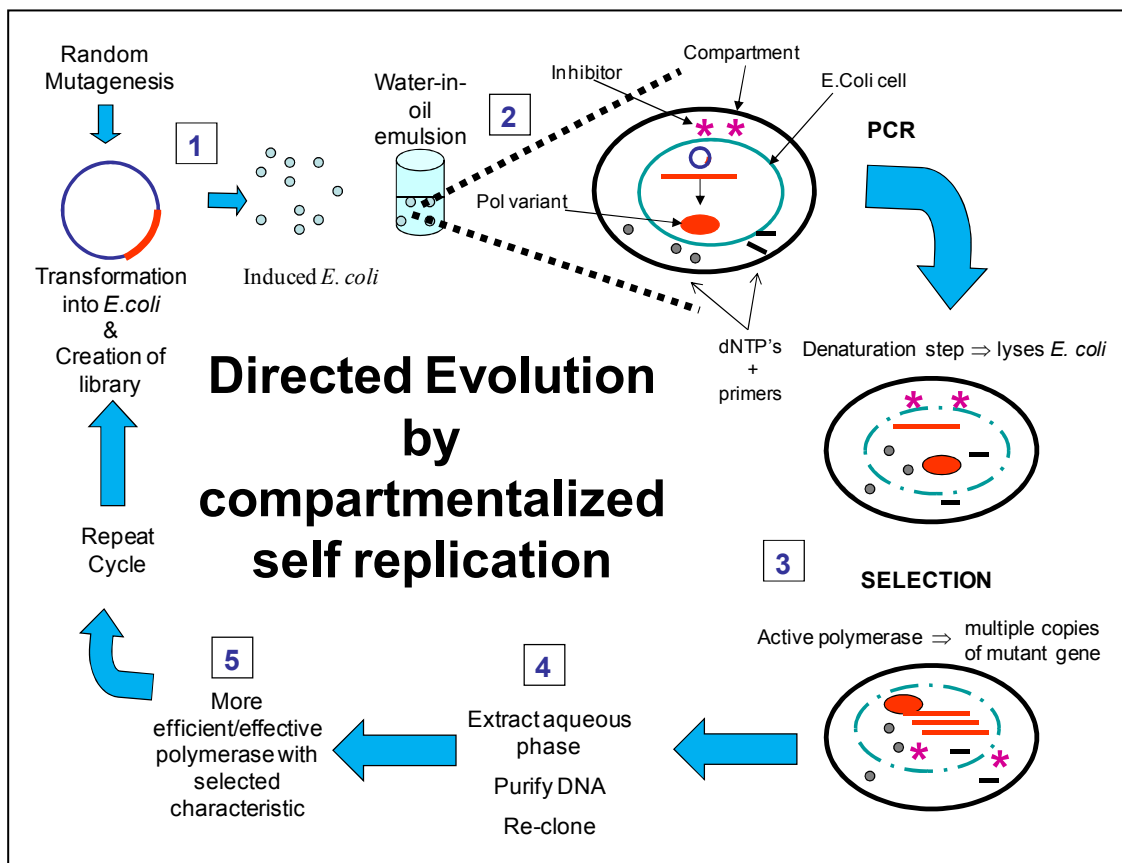


Figure 1.6 – A schematic diagram of compartmentalised self-replication

- 1) A mutant library is produced by error-prone PCR and variants are expressed in *E. coli*
- 2) *E. coli* cells are resuspended in PCR reaction buffer and PCR components, including primers to amplify the encoded polymerase gene, and are then separated into aqueous compartments by an oil-in-water emulsion. An inhibitor is also added to the compartment to apply a selective pressure to the evolution process.
- 3) Thermocycling lyses the *E. coli* cell and releases the polymerase and its encoding gene, allowing self-replication to take place. Poorly active polymerases are unable to amplify their encoding genes so are lost from the gene pool.
- 4) The amplified gene products are released and re-cloned for another cycle of CSR.

Aims of Project

The use of PCR-based assays in clinical diagnostics is expanding. PCR often suffers because of inhibitors found in an array of biological samples used in diagnostic testing. One method to overcome the problem of inhibition is to engineer polymerases with increased resistance to known PCR inhibitors such as heparin and phytate. The technique of directed evolution by compartmentalised self-replication (CSR) has recently been established as an efficient high-throughput method for the creation of novel polymerases. The aims of this project were: (1) to establish a CSR system to screen a mutant library, (2) use the system to select for novel heparin and phytate resistant *Taq* DNA polymerase variants, (3) to characterise the selected variants, and (4) explore the fundamental aspects of the inhibitors' mode of action.

Chapter Two - Compartmentalised Self Replication: Developing a System for the Directed Evolution of *Taq* DNA Polymerase

Introduction

In order to enhance the efficiency of the directed evolution of *Taq* DNA polymerase by compartmentalised self replication (CSR), it was deemed to be advantageous to start from an engineered fusion enzyme with enhanced processivity. Processivity is defined as the probability of an enzyme not to dissociate from the DNA while it translocates to the next position after the addition of a nucleotide (Pavlov *et al.*, 2004). Natural replicative polymerases often rely on accessory proteins to form stable complexes during replication and to achieve high processivity. The processivity of DNA polymerases is related to their natural role *in vivo* (Davidson *et al.*, 2003). *Taq* DNA pol belongs to the Pol A family, which function mainly as repair enzymes. These enzymes have evolved to fill in short stretches of DNA rather than to replicate entire genomes, and are therefore not very processive (Wang *et al.*, 2004).

Sac7 is a hyperthermostable, 7kDa chromatin protein from *Sulfolobus acidocaldarius* (Figure 2.1; Ko *et al.*, 2004) that is thought to have a DNA-stabilizing function *in vivo* (McAfee *et al.*, 1995). The Sac7 family of proteins are the most abundant DNA-binding proteins found in *S. acidocaldarius* and have been seen to form compact complexes with DNA (Lurz *et al.*, 1986). When bound, Sac7 has been shown to bend and induce a kink in the DNA duplex. Although their function is not fully understood, the Sac7 family of proteins appear to be unique to *Sulfolobus* species (Bedell *et al.*, 2005).

Since the discovery of these small non-specific dsDNA-binding proteins, further work has been carried out using Sac7 in protein engineering. When fused to the N-terminal region of *Taq*DNA pol, Sac7 has been shown to increase the processivity of *Taq* DNA pol without altering its catalytic activity (Wang *et al.*, 2004). In the fusion protein, the increase in the processivity of the enzyme is thought to occur via stabilisation of the DNA-polymerase complex, keeping *Taq* DNA polymerase tightly bound to the DNA during extension. Wang *et al.* (2004) reported that Sac7 was shown to increase the performance of wild-type *Taq* DNA pol by enhancing its processivity from an average primer extension length of 22 nucleotides to 104 nucleotides.

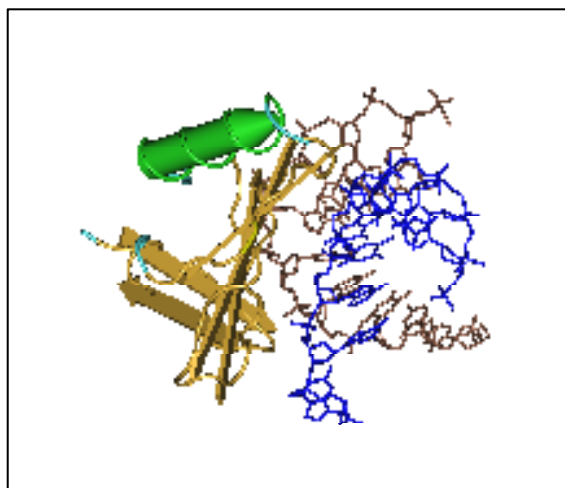


Figure 2.1 – Crystal structure of Sac7 from *Sulfolobus acidocaldarius* complexed with DNA. Two DNA decamers shown in brown and blue. Produced using Cdn3 (Ko *et al.*, 2004; PDB ID: 1WD1)

Other attempts at increasing polymerase processivity include the addition of thioredoxin binding domains (TBD) (Davidson *et al.*, 2003) and helix-hairpin-helix (HhH) motifs (Pavlov *et al.*, 2002). The T3 DNA polymerase TBD was successfully inserted into *Taq* DNA polymerase in an analogous position to its location in its native T3 polymerase by Davidson *et al.* (2003). In the presence of thioredoxin, the hybrid polymerase was found to become tightly bound to the DNA as the thioredoxin acts as a clamp to increase processivity. In an alternative approach, HhH motifs have also been used to enhance polymerase processivity. HhH motifs are found in many regulatory proteins and allow the proteins to bind to DNA in a non-sequence specific manner by interacting with the sugar-phosphate backbone. The fusion of HhH motifs to a number of polymerases has been investigated with varying results. While HhH motifs increased the processivity of *Pfu*, they were found to have no effect on *Taq* DNA polymerase (Pavlov *et al.*, 2002).

Compared to the examples described in the previous paragraph, the fusion of Sac7 to a polymerase appears to have a broader effect as it has been shown to enhance processivity with both Pol A (*Taq*) and Pol B (*Pfu*) polymerases (Wang *et al.*, 2004). For this reason, a clone of the Sac7-*Taq* polymerase gene inserted in a pUC18-derived vector called pTTQ18NAK (Appendix A) for the high-level expression of the fusion protein was selected as the starting point for modification by directed evolution in this project.

Another crucial part of setting up and preparing the CSR system to be employed in this project was the choice of a system to select efficiently the positive clones after each round of CSR. The chosen method was the use of SacB as a selectable marker as developed by Schweizer and Hoang (1995). The SacB gene from *Bacillus subtilis* encodes the enzyme levan sucrase and confers sucrose lethality (Bramucci *et al.*, 1996). The build up of levans

in cells containing SacB causes cellular osmosis when cells are plated on media containing sucrose (Recorbet *et al.*, 1992). Inserting SacB into a vector in the cloning site creates a selection system based on sucrose lethality. When subsequently employed to select positive clones during CSR, cloning the desired gene into the SacB vector and plating transformants on sucrose plates provides a direct screen and eliminates any colonies that still contain SacB. In this way, SacB is a selectable marker that can be used for screening by reducing background and therefore increasing mutant library size during directed evolution.

Since the focus of this project is on the evolution of variants that are resistant to inhibitors of the *Taq* DNA polymerase enzyme, it was deemed to be preferable only to mutate the polymerase domain prior to CSR screening. Therefore the SacB gene was designed to replace just the polymerase domain rather than the whole Sac7-Taq polymerase fusion. To allow removal or exchange of just the polymerase domain to occur, site-directed mutagenesis was employed to introduce an extra restriction site in the appropriate site in the construct. A *Bgl*II site was designed just upstream of the *Taq* DNA polymerase linker region joining the exonuclease and the polymerase domains (Figure 2.2). The same restriction site was also designed into the primers for amplification of the SacB gene thus allowing exchange of the two fragments by digestion with *Bgl*II.

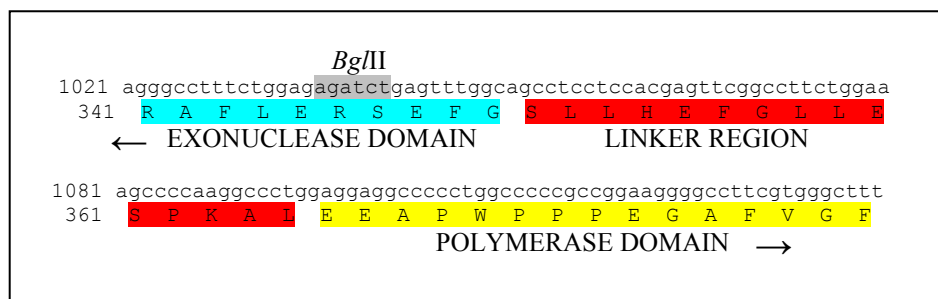


Figure 2.2 – The Linker region between exonuclease and polymerase domains in Sac7-Taq polymerase.

In addition to the use of a processivity-enhancing domain and the SacB selection system described above, a number of other aspects of the CSR system had to be investigated to ensure that the process could be set up and conducted efficiently. One of these was the optimisation of emulsion PCR for use in the system.

The use of emPCRs or *in vitro* compartmentalization (IVC) has increased considerably over the last few years. It is the basis of the revolutionary Roche 454 DNA sequencing system, in which individual DNA-capture beads are compartmentalised within a

water-in-oil emulsion droplet (Margulies *et al.*, 2005). Each bead captures an individual fragment of DNA which allows parallel amplification and sequencing reactions to take place in millions of compartments within the emulsion. IVC has also been widely used in the directed evolution of a range of different proteins. The simplest application of IVC is use with DNA modifying enzymes such as DNA methyltransferase, restriction endonucleases and polymerases. However, IVC has also been used to produce new ribozyme variants and peptides with specific antibody selectivity as reviewed by Griffiths and Tawfik (2006).

Although emPCR is well characterised, the use of it and the CSR system had to be set up and verified under the chosen conditions for this project. Before carrying out directed evolution on the polymerase using CSR, several things had to be investigated and optimised: (1) droplet formation, (2) PCR amplification within the emulsion droplets, (3) PCR amplification using *E. coli* cells containing induced polymerase as the only source of enzyme in the emulsion droplets and (4) extraction of the DNA product from the emulsion for further use. The investigation and optimisation of each of the steps (1)-(4) above are reported in this chapter.

Materials and Methods

Standard Activity and Expression Testing

10ml LB medium [10g/L tryptone, 5g/L yeast extract, 5g/L NaCl (all Sigma, Gillingham, UK)] containing Kanamycin (50µg/ml) was inoculated from a freshly streaked plate or an overnight starter culture and incubated at 37°C until an OD₆₀₀ of >0.6 was reached. The culture was induced with 0.5mM IPTG and incubated for a further 4 hours. Cells were harvested by centrifugation and resuspended in 500µl of TMN Lysis buffer [50mM Tris-HCl (pH8.5), 10mM MgCl₂, 16mM (NH₄)₂SO₄, 0.1% (v/v) Tween 20]. Samples were sonicated for 30 s at 15 micron and centrifuged. 200µl of lysate was then heat treated at 70°C for 20 min. Debris was removed by centrifugation and the heat-treated lysate was used as enzyme stock for an activity PCR.

Protein samples were analysed by 10% (w/v) SDS-PAGE gels with standard buffers (Sambrook & Russell, 2001) and BioRad Mini Protean equipment. Samples in loading buffer were denatured at 95°C for 5min before loading. Proteins were visualised with Coomassie Brilliant Blue R 250 and molecular weights determined using the Broad Range Marker (BioRad, Hemel Hempstead, UK) containing: Myosin (200kDa), β-galactosidase (116kDa), phosphorylase b (97kDa), bovine serum albumin (66kDa), ovalbumin (45kDa), carbonic anhydrase (31kDa), soybean trypsin inhibitor (22kDa), lysozyme (14kDa) and aprotinin (7kDa). Protein concentrations were measured using the Bradford Assay (BioRad, Hemel Hempstead, UK).

An M13 or 500bp Lambda activity PCR was set up with final concentrations as follows: 1X Taq buffer (GeneSys, Camberley, UK) containing 1.5mM MgCl₂, 200µM dNTPs (GeneSys, Camberley, UK), M13 (or 500bp λ) forward and reverse primers at 25pmol, 1-10ng of pTTQ18NHK Sac7-TaqPol plasmid (or λ DNA), 1µl lysate, and ddH₂O up to 50µl. 1.25U Taq (GeneSys, Camberley, UK) was used as a control and reactions were run under standard M13 or 500bp λ conditions depending on the primers used. M13 reactions were cycled as follows: 94°C for 3min, 30 cycles of 94°C for 15s, 60°C for 15s and 72°C 3min. 500bp λ reactions were cycled as follows: 94°C for 3min, 30 cycles of 94°C for 15s, 55°C for 15s and 72°C 30s. 5% DMSO was used with the M13 primers. 5µl of each PCR reaction was run on a 1% (w/v) agarose gel with a 1kb DNA ladder [New England Biolabs, Hitchin, UK (see Appendix D)].

500bp λ primers:

Common (Forward) 5' - GAT GAG TTC GTG TCC GTA CAA CTG G

500bp (Reverse) 5' - GGT TAT CGA AAT CAG CCA CAG CGC C

Plasmid Prep

5ml LB [10g/L tryptone, 5g/L yeast extract, 5g/L NaCl (all Sigma, Gillingham, UK)] containing Kan (50 μ g/ml) was inoculated with a single colony from a freshly streaked plate and incubated overnight at 37°C. The plasmid DNA was purified from the culture using Mini Prep kit (Promega, Southampton, UK) following the standard protocol. The plasmid prep was checked by running 1 μ l of plasmid on a 1% (w/v) agarose gel.

Site-Directed Mutagenesis

An additional restriction site, *Bgl*III, needed to be added to the pSac7-TaqPol construct (Appendix B) and this was done using the QuikChange[®] Site-Directed Mutagenesis kit (Stratagene, Cheshire, UK). QuikChange mutagenic primers were designed as per Stratagene's guidelines. The original *Taq* DNA polymerase sequence was altered from AGGCTT to AGATCT, which gave one amino acid change (L347S). QuikChange primers were as follows:

Forward primer:

Arg→Arg Leu→Ser

AGG CTT

5'AGGGCCTTTCTGGAG AGATCTGAGTTTGGCAGCCTCCTC

Reverse primer:

5' GAGGAGGCTGCCAAACTCAGATCTCT CCAGAAAGGCCCT

The mutagenesis reaction was set up using 1X Pfu Ultra reaction buffer (Stratagene, Cheshire, UK), 200 μ M dNTPs (GeneSys, Camberley, UK), 125ng of each primer, 10ng pSac7-TaqPol vector template, 2.5U Pfu Ultra (Stratagene, Cheshire, UK) and made up to

50µl with water. The reactions were run on an Eppendorf (Cambridge, UK) Mastercycler under QuikChange conditions as follows: 95°C for 30s, 18 cycles of 95°C for 30s, 60°C for 1min and 68°C for 8min. After thermocycling, the reactions were treated with *DpnI* before being transformed into XL1 Blue chemically competent cells (Stratagene, Cheshire, UK). 1µl of *DpnI* treated reaction was added to 50µl XL1 Blue cells and incubated on ice for 30 min. The cells were heat shocked at 42°C for 30s and placed on ice for 2 min. 0.5ml of 42°C pre-heated SOC [0.2g/L Bacto-tryptone (BD, Oxford, UK), 0.05g/L bacto-yeast extract (BD, Oxford, UK), 10mM NaCl, 2.5mM KCl, once autoclaved 10mM MgCl₂, 10mM MgSO₄ and 20mM glucose were added (all Sigma, Gillingham, UK)] was added to the cells and they were incubated at 37°C for 1h before being spread onto LB/KAN plates (50µg/ml). After overnight incubation at 37°C, colonies were picked and screened for the *Bg/II* site mutation.

Screening

Individual colonies from the transformation plate were picked using sterile toothpicks and gridded onto a fresh LB/KAN plate (50µg/ml) to keep as a reference plate. Cells from each colony were lysed in 20µl water at 100°C for 5 min before being used as template in a screening PCR. *Bg/II* screening primers were designed to amplify 500bp sequence around the *Bg/II* site (F-5' CACCCCGCCTGGCTTTG and R-5' CAGGGCCAGA ACGCTCAGGTCTT). PCR reactions were run as follows (final concentrations): 1X Mastermix (GeneSys, Camberley, UK), 25pmol F' and R' primers, 1µl lysed template and water up to 50µl. Reactions were cycled under screening PCR conditions: 94°C for 3min, 30 cycles of 94°C for 10s, 55°C for 10s and 72°C for 45s. 5µl of PCR product was run on a 2% (w/v) agarose gel. PCR products were then digested with *Bg/II* to screen for correctly mutated clones. Digests were as follows: 1X buffer 3 (New England Biolabs), 7µl PCR product, 10U *Bg/II* and water up to 30µl. Digests were left at 37°C for 1h and then the entire reaction was run on a 2% (w/v) agarose gel.

SacB Cloning

Primers were designed to amplify the *Bacillus subtilis* sucrase gene (SacB) from the vector pDNR-LIB (Clontech, St. Germain-en Laye, France). SacB was amplified using Phusion polymerase (New England Biolabs, Hitchin, UK) and a 2X dilution series of the pDNR-LIB template vector ranging from 50ng to 1ng. The PCR was set up as follows: 1X

HF phusion buffer, 200µM dNTPs (GeneSys, Camberley, UK), 25pmol forward and reverse primer, pDNR-LIB template, 1U Phusion, and water up to 50µl. Reactions were cycled following Phusion guidelines.

All successful PCRs were treated with *DpnI* and purified using the Wizard SV gel and PCR clean-up system (Promega, Southampton, UK). SacB PCR product and pSac7-TaqPol (*Bg/II* Mutant 8) were digested with *Bg/II* and *SaII* (New England Biolabs, Hitchin, UK) and the vector was phosphatased with 1U Shrimp Alkaline Phosphatase (SAP) (New England Biolabs, Hitchin, UK) per pmol of DNA. SAP was inactivated at 65°C for 15min. Both digests were run on a 1% (w/v) agarose gel and the large fragment of vector and the SacB were gel extracted and purified using Perfectprep Gel Clean up kit (Eppendorf, Cambridge, UK).

The digested pSac7-Taq and SacB were ligated using T4 Ligase (Promega, Southampton, UK) at room temperature for 3h. The ligation reaction was ethanol precipitated and 1µl was transformed into TOP 10F' electrocompetent cells (prepared as below) using a standard protocol.

Ethanol Precipitation

The volume of aqueous phase was measured and 0.1 volumes of 3M sodium acetate (pH5.2) and 2 volumes of cold ethanol were added. The sample was mixed and left to precipitate at -20°C for 30 min. The sample was centrifuged at 13,000g, the ethanol was removed and the pellet was washed in 70% (v/v) ethanol. The pellet was left to air dry for approximately 10 min before being resuspended in sterile ddH₂O.

Transformation

20-40µl of electrocompetent cells were thawed on ice. 1µl of ligation reaction was added to the cells and they were left on ice for 30 min before being transferred to a chilled Gene Pulser 0.1cm cuvette (BioRad, Hemel Hempstead, UK). The cells were pulsed at 1.8kV, immediately resuspended in 1ml of SOC [0.2g/L Bacto-tryptone (BD, Oxford, UK), 0.05g/L bacto-yeast extract (BD, Oxford, UK), 10mM NaCl, 2.5mM KCl, once autoclaved 10mM MgCl₂, 10mM MgSO₄ and 20mM glucose were added (all Sigma, Gillingham, UK)] and were left shaking at 37°C for 1h. Typically 50µl and 100µl aliquots of transformed cells were spread onto an LB/Kan plate (50µg/ml) and incubated overnight at 37°C.

Preparation of Electrocompetent Cells

10ml LB [10g/L tryptone, 5g/L yeast extract, 5g/L NaCl (all Sigma, Gillingham, UK)] was inoculated with a single colony of fresh TOP10F' *E. coli* (F' {*lacI^q*, Tn10(Tet^R)} *mcrA* (*mrr-hsdRMS-mcrBC*) ϕ 80 *lacZ* M15 *lacX74* *deoR* *recA1* *araD139* (*ara-leu*)7679 *galU* *galK* *rpsL*(str^R) *endA1* *nupG*) and incubated overnight at 37°C with shaking. 2x 250ml LB was inoculated with 2.5ml of the overnight culture and grown to an OD_{600nm} ~1.0 at 37°C. The cultures were chilled on ice for 5min before centrifuging in pre-chilled bottles at 5,000g for 10min. The supernatant was discarded and the pellets were gently resuspended in 25ml ice-cold ddH₂O before being pooled together. The cells were centrifuged as before, the supernatant was discarded and the pellet gently resuspended in 5ml ice-cold 10% (v/v) glycerol. The cells were centrifuged again and supernatant discarded. The pellet was resuspended in ice-cold 10% (v/v) glycerol to a final volume of less than 1.5ml. 40 μ l aliquots were pipetted into chilled tubes, frozen on dry ice/methanol and stored at -80°C.

Auto Induction

A 10ml culture of LB [10g/L tryptone, 5g/L yeast extract, 5g/L NaCl (all Sigma, Gillingham, UK)] containing Kan (50 μ g/ml) was inoculated with a single colony and grown with shaking at 37°C until an OD_{600nm} of 0.5 was reached. 30ml of autoinduction Overnight Express medium (Novagen, Nottingham, UK) in a 250ml shake-flask was then inoculated with 5% (v/v) of the starter culture. The autoinduction culture was left to grow for a further 24h before cells were harvested by centrifugation.

Emulsion PCR

2x10⁸ induced *E. coli* cells expressing polymerase variants were centrifuged to remove media and then washed in 200 μ l 1X Taq buffer. The cells were resuspended in 20 μ l 1X Taq buffer before being added to the compartmentalised self replication (CSR) mix. 200 μ l of CSR mix was made up as follows: 1x Taq Buffer containing 1.5mM MgCl₂, 0.25mM dNTP, 1 μ M of pol domain forward and reverse primer, *E. coli* cells expressing Sac7-TaqPol and water to final volume. The oil phase was made up in light mineral oil (Sigma, Gillingham, UK) and consisted of 4.5% (v/v) Span 80 (Fluka), 0.4% (v/v) Tween 80 (Sigma, Gillingham, UK), 0.05% (v/v) Triton X-100 (Sigma, Gillingham, UK), as described by Ghadessy *et al.*(2001). A small magnetic stirrer was placed into a 5ml cryotube (Corning) and 0.2ml of CSR reaction (water phase) was mixed with 0.4ml of oil phase to

form an emulsion. The emulsion was stirred at ~1100rpm for 5 min to create an average droplet size of 3-5µm as determined by light microscopy. The emulsion PCR (emPCR) was aliquoted into 0.5ml thin-walled PCR tubes in volumes of less than 100µl and temperature cycled as follows: 94°C for 5min, 25 cycles of 94°C for 1min, 55°C for 1min and 72°C for 5min.

Pol domain primers:
Forward 5' AGGGCCTTTCTGGAGAGATCTGAGTTTGGCAGC CTCC
(*BglII*)
(*SalI*)
Reverse 5' GCCTGCAGGTCGACTCTAGAGGATCTATCACTCCTTGG

Aqueous Phase Extraction

The emPCRs were centrifuged at 3000g for 5 min and the top oil layer was removed. 0.1ml of quenching buffer [1M NaCl, 10mM Tris/HCl (pH7.4), 1mM EDTA] was added to each tube and the emulsion was resuspended. The reactions were pooled together in a 1.5ml tube and 1ml of water-saturated ether was added. The samples were vortexed and centrifuged briefly. The ether (top layer) was removed leaving the aqueous layer. 1ml water-saturated ether was added and the previous step repeated. Finally the aqueous phase was washed with 1ml ether, vortexed and centrifuged. The ether layer was removed and the aqueous phase was processed further. The aqueous phase was ethanol precipitated to concentrate the DNA from the sample. The pellet was then resuspended in 30µl sterile water.

Results

Expression and Activity Test of Sac7-TaqPol

An agar stab stock of Sac7-TaqPol (GeneSys) in TOP10F' was used to streak a fresh LB Kan plate, which was incubated at 37°C overnight. A small scale expression of the clone was performed to check expression and activity levels. Cells were harvested from a 10ml LB culture with a 4h IPTG induction. The cell pellet was resuspended, sonicated and the cell extract was heat treated to remove the bulk of the *E. coli* proteins. Protein expression was analysed by 10% SDS-PAGE (Figure 2.3). 25U of purified *Taq* was run as a size marker.

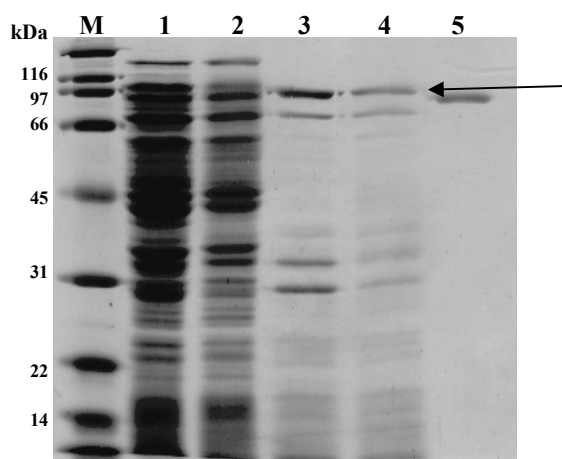


Figure 2.3 – Expression of Sac7-TaqPol. This SDS-PAGE gel shows various samples of Sac7-TaqPol soluble cell extract. (M) marker, (1) uninduced sample, (2) induced sample, (3) uninduced, heat treated sample, (4) induced, heat treated sample, (5) 25U *Taq* polymerase (92kDa). Sac7-TaqPol (99kDa) is indicated by the arrow).

The SDS-PAGE gel shows that both the uninduced sample and the induced sample produced protein of the correct size (99kDa). The uninduced culture grew to a higher OD and also appeared to contain a higher expression of Sac7-TaqPol compared to the induced culture.

To check that the expressed Sac7-TaqPol was active, a standard activity PCR with M13 primers was run for 30 cycles using the heat-treated cell extract in place of purified *Taq*DNA pol. A 2X dilution series of the extract was prepared and used in the PCR. 5µl of each PCR was run on a 1% (w/v) agarose gel (Figure 2.4).

From a 5ml LB Kan culture of pSac7-Taq Pol in TOP10F', plasmid was extracted using the Promega Mini Prep kit. To check the size of the gene, the plasmid was digested

with *NdeI* and *SalI*. When digested the plasmid should produce 3 fragments: a 212bp *Sac7* fragment (not visible in picture), a 2521bp *TaqPol* fragment and a large 5.4kb vector fragment (Figure 2.5).

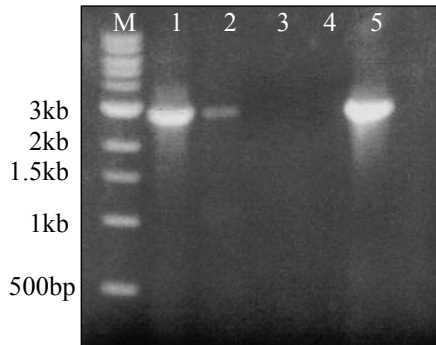


Figure 2.4 – Activity test PCR. Varying levels of cell-free extract were used as enzyme in the 30 cycle activity PCR. (M) 1kb DNA ladder (NEB), (1) 1 μ l extract, (2) 0.5 μ l extract, (3) 0.25 μ l extract, (4) blank lane, (5) 1.25U *Taq* control.

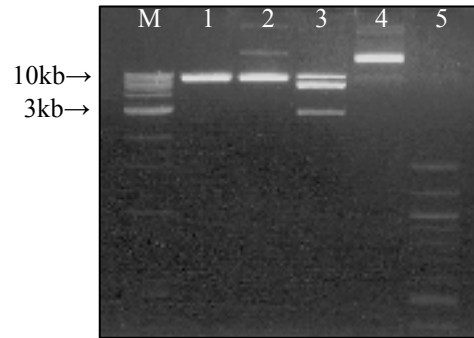


Figure 2.5 - p*Sac7*-*TaqPol* Digest. (M)1kb ladder, (1) *NdeI* single digest, (2) *SalI* single digest, (3) *NdeI* and *SalI* double digest, (4) no enzyme control digest, (5) 100bp ladder.

Figure 2.4 shows that the *Sac7*-*TaqPol* clone was active in a standard PCR using 1-0.5 μ l of cell-free extract. At a lower concentration of extract the PCR failed to produce any visible product. The agarose gel shows that 1 μ l of extract is approximately equivalent to 1.25U of purified *Taq* DNA pol. Figure 2.5 indicates that the *Sac7*-*TaqPol* clone contains inserts of the expected size. Lane 3 shows the *TaqPol* gene band at 2521bp, with the larger vector fragment above it. The small 212bp *Sac7* fragment was visualised early on in the electrophoresis run; however, further separation of the other bands was needed and this meant the *Sac7* band ran off the gel at the time off the photo. The control lanes show the vector after single digests (1) and (2), compared to double digestion (3) and uncut circular plasmid (4). As expected, the plasmid ran in a different way in all four lanes due to its conformational changes after single, double or no digestion.

SacB Cloning

The polymerase domain of *Taq* DNA pol was replaced with *SacB* in the pTTQ18NAK-*Sac7*-*TaqPol* expression vector to create a direct screen for insert-containing transformants. The *SacB* gene was amplified from the vector pDNR-LIB (Clontech) using primers designed to incorporate *BglII* and *SalI* restriction sites. To enable the cloning to

take place, a *Bgl*II site also had to be added to the Sac7-TaqPol clone, which was done using site directed mutagenesis (Figure 2.6).

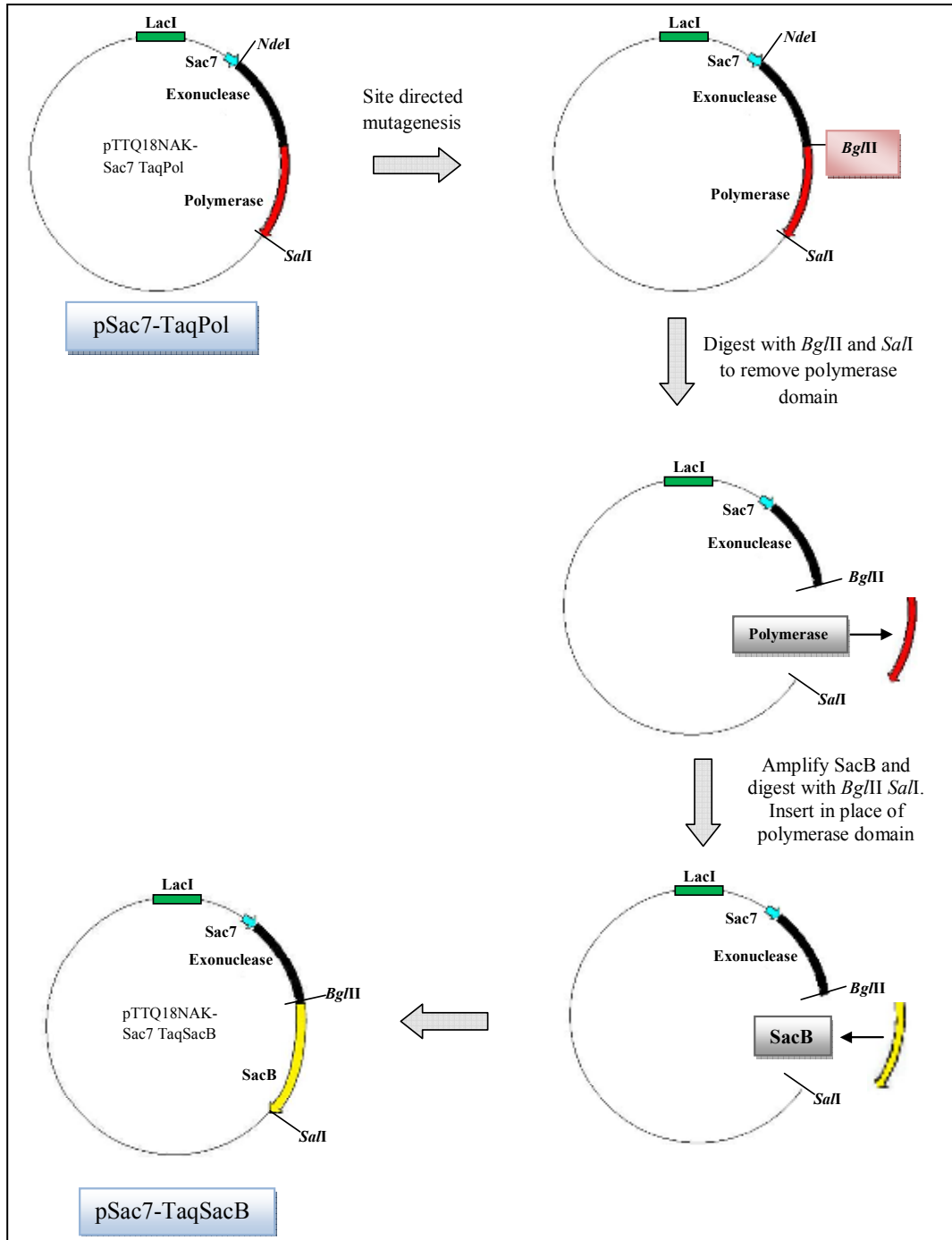


Figure 2.6 - A schematic diagram of polymerase domain replacement with the selectable marker SacB. (1) starting vector - pSac7-TaqPol, (2) site directed mutagenesis was used to incorporate a *Bgl*II site, (3) removal of the polymerase domain by digestion with *Bgl*II and *SalI*, (4) *SacB* was cloned and digested with *Bgl*II and *SalI* and then inserted in place of the polymerase domain (5) final vector – pSac7-TaqSacB.

Transformants produced from the site directed mutagenesis were then screened for the *Bgl*III site by PCR followed by a digest with *Bgl*III. A ~570bp fragment surrounding the *Bgl*III site was amplified and digested. The resulting digested PCR product produced two fragments of ~331bp and ~241bp, demonstrating that the mutagenesis was successful. The fragments were visualised on an agarose gel (Figure 2.7).

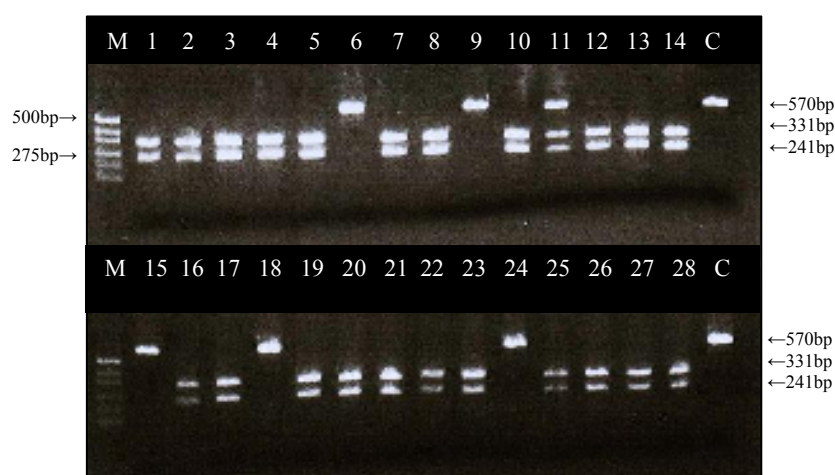


Figure 2.7 – *Bgl*III Digest of PCR products. An agarose gel showing the digested PCR fragment (surrounding the *Bgl*III site). (M) marker, (C) negative control, (1-28) numbered lanes refer to colony screened.

The clones that were cut into two fragments had undergone successful mutagenesis and were thus demonstrated to contain the *Bgl*III site. Since the introduction of a *Bgl*III site had caused one amino acid change in the original sequence, the *Bgl*III clones were carried forward for small scale expression and activity tests to check that the amino acid change had had no impact on protein function. The heat-treated extract was analysed by SDS-PAGE (Figure 2.8a), and the results showed that the majority of the clones expressed protein of the correct size. The extract was then used in a 500bp PCR to check that polymerase activity was still present. PCRs were conducted with 1µl and 0.5µl of extract from each mutant. When analysed on an agarose gel, the PCRs indicated that all but one of the mutants was active (Figure 2.8b).

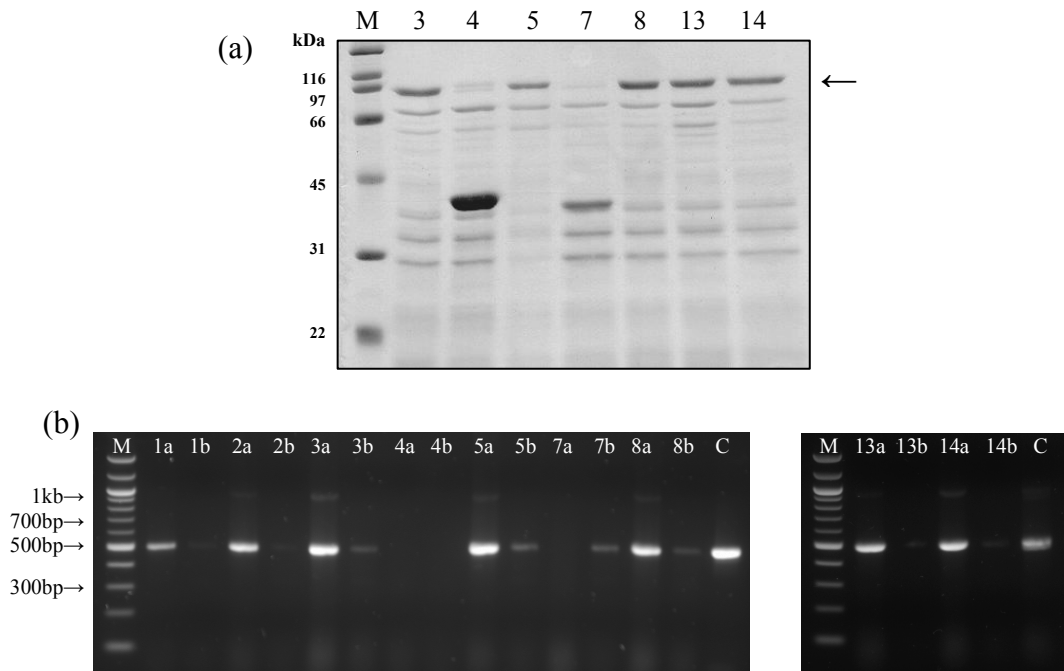


Figure 2.8 – Expression and activity PCR of *BgIII* Mutants. (a) SDS-PAGE gel showing the heat-treated extract of several *BgIII* mutants (lane numbers correspond to mutants). First lane contains marker. Arrow indicates expression of Sac7-TaqPol. (b) Agarose gel showing the activity of several mutants (lane numbers correspond to mutants). (M) 100bp NEB marker, (a) 1 μ l extract, (b) 0.5 μ l extract, (C) control PCR with 1.25U *Taq*.

Sac7-TaqPol *BgIII* mutant 8 (M8) was chosen for use in all further work. M8 and the SacB gene, previously amplified, were digested with *BgIII* and *SalI*. This removed the polymerase domain from the TaqPol gene of M8. SacB was then ligated into Sac7-TaqPol M8 in place of the polymerase domain. Transformants were plated on both Kan and Kan/Sucrose plates and left overnight at 37°C. As expected no colonies grew on the Kan/Sucrose plates, which indicated that the ligation was successful and all transformants contained the SacB gene.

The reverse cloning procedure was then performed to show that the polymerase domain could be successfully ligated back into Sac7-Taq(SacB). The original Sac7-TaqPol and the SacB containing vector were digested with *SalI* and *BgIII*. After agarose gel separation, the vector and pol domain DNA were gel purified and the products were ligated. The transformants were plated on Kan and Kan/Sucrose, and as expected colonies were seen on both plates. Several transformants were picked from the Kan/Sucrose plate and were expressed on a small scale. The cell extracts were then tested in an M13 activity PCR. 100% of the clones tested were active, which showed that the SacB screening process was successful.

Auto Induction

To increase the expression of Sac7-TaqPol, and therefore increase the yield in future CSR reactions, the Overnight Express (Novagen) auto-induction system was used. A small scale expression test was carried out using auto-induced culture and a 4h IPTG-induced culture. Cells were harvested, resuspended and processed as standard. Equivalent volumes of cell-lysate from the auto-induced and IPTG-induced samples were analysed by SDS-PAGE (Figure 2.9). The overnight express system increased the expression of Sac7-TaqPol and was used in all further expression.

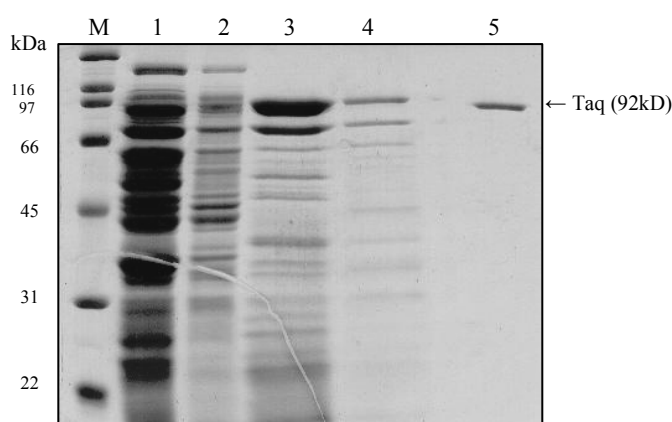


Figure 2.9 – Comparison of Auto and IPTG induction. This SDS-PAGE gel shows cell extract from auto induction and IPTG-induced cultures. (1) auto-induced sample, (2) 4h IPTG-induced sample, (3) heat-treated auto-induced sample, (4) heat-treated IPTG-induced sample, (5) 25U purified *Taq* polymerase.

Emulsion PCR

The oil phase (400 μ l) was made according to Ghadessy *et al.* (2001) and mixed with 200 μ l of 1X Taq Buffer to form an emulsion following the standard protocol. The emulsion was viewed under a light microscope and the average size of the droplets was found to be 5.1 μ m (\pm 0.2). In this work, the number of droplets in a typical emulsion was estimated to be 9.7×10^8 per ml, by cell-counting under a light microscope. *E. coli* cells at 3.3×10^8 per ml were then mixed with buffer and used to form an emulsion, resulting in an approximate 3-fold excess of emulsion droplets to *E. coli* cells. When viewed under the light microscope, the *E. coli* cells in individual droplets could not be visualised.

In order to visualise the *E. coli* in the droplets, pGFP (Clontech), a vector carrying the green fluorescent protein (GFP) coding sequence, was used in the emulsion. The vector

was transformed into *E. coli* and then expressed. The induced cells were used in an emulsion and the droplets were then viewed using a fluorescence microscope. The *E. coli* containing GFP were photographed using a blue filter so the GFP could be visualised and then a second picture was taken of the emulsion droplets without the filter (Figure 2.10). Whilst it was challenging to superimpose the light microscopy and fluorescence images, the data obtained were consistent with the predictions that approximately one third of droplets contained an *E. coli* cell and that the vast majority of droplets that contained *E. coli*, contained a single cell.

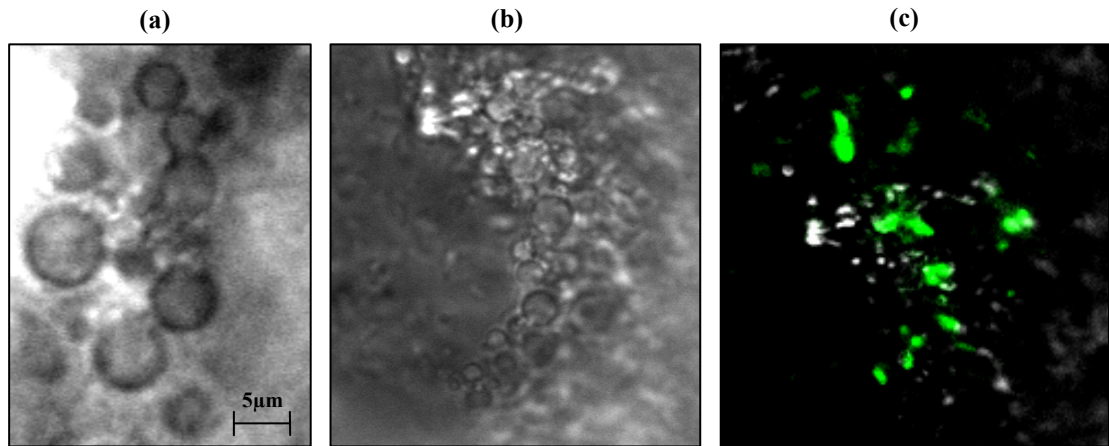


Figure 2.10 – A typical water-in-oil emulsion. (a) A light micrograph displaying a typical view of the water-in-oil emulsion with an average droplet size of 5µm (indicated by scale bar). (b) and (c) are two microscope images of the same section of emulsion without and with fluorescence. In (b) the emulsion droplets can be visualised and in (c) the *E. coli* cells containing GFP tagged vector can be visualised.

To test that amplification could be achieved within the emulsion droplets, a standard PCR using purified *Taq* DNA pol was run in the emulsion. This proved to be successful and the PCR product was extracted from the emulsion. Next the use of *E. coli* cells and the activity of Sac7-TaqPol were tested in an emPCR. Standard *Taq* DNA pol was replaced by *E. coli* cells containing induced polymerase. After auto induction, the OD₆₀₀ of the culture was measured and aliquots of $\sim 2 \times 10^8$ cells (3.3×10^8 per ml) were spun down and washed. A cell extract was made as standard from the remaining culture. An aliquot of induced cells was added to the CSR mix and an emulsion made as standard. To optimize the MgCl₂ concentration of the reaction, three emPCRs with M13 primers were run with varying amounts of Mg²⁺ ranging from 1.5mM to 2.5mM. After thermocycling, the aqueous phase of the reaction was extracted. Each emPCR was resuspended in 30µl of sterile water and 10µl was run on a 1% (w/v) agarose gel for comparison (Figure 2.11). The 2.7kb Sac7-

TaqPol was amplified under all three conditions but the optimum $MgCl_2$ concentration was found to be between 1.5-2mM. Standard Taq buffer contains 1.5mM $MgCl_2$ and was used in all further emPCRs.

The product yield of an emPCR was much lower than that of a standard PCR. To check that the emulsion PCR was working efficiently, a set of emulsion PCRs were run with varying amounts of induced *E. coli* cells ranging from double the stated amount of 2×10^8 cells in 200 μ l, down to only 25% of the cells in 200 μ l. The emPCRs were extracted and analysed on a 1% (w/v) agarose gel (Figure 2.12). In this particular PCR, the primers used amplified a 1.9kb fragment of the Sac7-TaqPol gene as seen on the gel below.

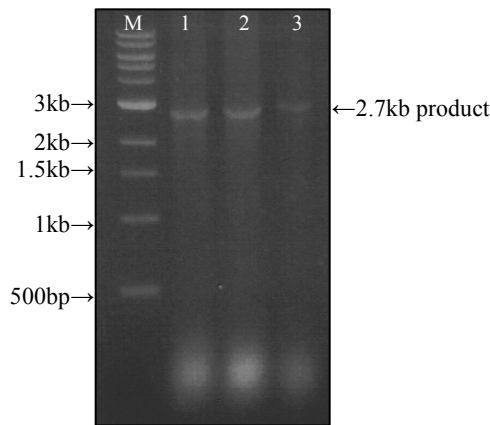


Figure 2.11 – Emulsion PCR with varying Mg^{2+} concentrations. (M) 1kb DNA ladder (NEB), (1) 1.5mM $MgCl_2$, (2) 2.0mM $MgCl_2$, (3) 2.5mM $MgCl_2$.

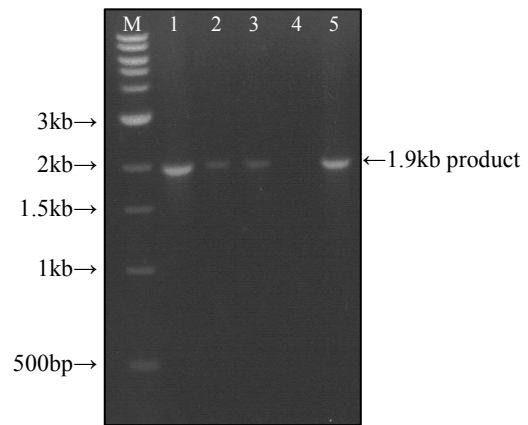


Figure 2.12 – Emulsion PCR with varying amounts of *E. coli* cells per 200 μ l reaction. (M) 1kb DNA ladder (NEB), (1) 2×10^8 cells, (2) 1.5×10^8 cells, (3) 1×10^8 cells, (4) 5×10^7 cells, (5) 4×10^8 cells.

An emPCR was also set up using a cell extract instead of *E. coli* cells as the source of polymerase. It was run under emulsion conditions and extracted as standard. When run on an agarose gel, the cell extract emPCR showed increased yield compared to standard *E. coli* emPCR. This indicated that it was not the emulsion conditions that were inhibiting the emPCRs but rather the level of polymerase per compartment derived from the single induced *E. coli* cell.

Discussion

This chapter presents the work carried out in preparation for the directed evolution by compartmentalised self-replication of the polymerase domain of a Sac7-TaqPol fusion protein. It presents the results of successful cloning experiments required to set up an efficient CSR system and also the results of the optimisation of the emulsion PCR (emPCR). Crucially, the Sac7-TaqPol protein was shown to be active in an emPCR which is essential for successful directed evolution by CSR that is the focus of this project.

The Sac7-TaqPol clone provided by GeneSys that was used as a starting point for this project was tested and confirmed to express the polymerase fusion protein. A heat-treated lysate of Sac7-TaqPol was shown to be active in a standard PCR. The starting clone was then modified to help create a more efficient CSR system.

Site directed mutagenesis was used to incorporate a *Bgl*III site in the linker region between the exonuclease and the polymerase domain. The addition of this restriction site allowed the removal and further manipulation of the polymerase domain. SacB was cloned and inserted into Sac7-TaqPol in place of the polymerase domain. SacB confers sucrose lethality and will be used as a selectable marker to reduce background and increase positive selection in the CSR procedure described in Chapter 1. Sac7-TaqSacB was used as the starting vector in mutant library construction for directed evolution experiments which are described in subsequent Chapters. The use of the SacB system ensured that cloning during library construction was efficient, which was crucial to the whole CSR procedure.

The correct set up and testing of the CSR system is essential to the success of future screening. One key element was the formation of the emulsion; the stirring speed at which the emulsion was formed and the components in the oil phase both contribute to droplet size. With a stirring speed of ~1100rpm, the emulsion was shown to form droplets of an average size of 5µm. The work of Tawfik and Griffiths (1998) and Ghadessy *et al.* (2001) were used as guidelines for setting up the emulsion. In their study, Tawfik and Griffiths (1998) showed that droplet size varied from 0.5 to 7µm with the average droplet size being 2.5µm.

The droplet size of 5µm in this study was the correct size for each droplet to contain a single *E. coli* cell, which has a typical size range of 1 to 5µm (Campbell *et al.*, 1999; Alberts *et al.*, 1994) It was estimated that a typical emulsion contained 9.7×10^8 droplets per ml, indicating that droplet number was in excess of the number of *E. coli* cells in the mixture and therefore it would be unlikely to have a significant proportion of droplets containing more than one *E. coli* cell. The vast majority of compartments that contained *E. coli* contained a single *E. coli* cell, but due to the variability of droplet size there may have been a low proportion of compartments that contained two *E. coli* cells. This was not considered

to represent a serious impediment to the success of CSR because two *E. coli* cell-droplets are in such low abundance. This point is important as it is the separation of variants into individual replication compartments that ensures the key phenotype-genotype linkage, which is imperative for a highly efficient CSR screen. As the vast majority of compartments will contain only one variant, the low abundance of two-cell droplets should not have an effect on the screening process.

As well as looking at emulsion droplet formation, standard PCR was achieved within the droplets and the PCR product was extracted from the emulsion for further use. Producing a working emulsion PCR containing *E. coli* cells was slightly more challenging. Many attempts were made before emPCR with *E. coli* produced enough DNA amplification to be visualised on a gel. It was found that the yield from an emulsion PCR containing *E. coli* cells, was lower than that of an emulsion PCR using purified *Taq* DNA pol, when run under the same conditions. This is likely due to a suboptimal amount of polymerase contained within each self-replication compartment upon lysis of a single *E. coli* cell. Increasing the number of cells, above 3.3×10^8 per ml used in the emulsion, did not have a noticeable effect on the yield. However, decreasing the cell number did have an effect and when only 25% of the stated number was used then no visible product formed. This was probably due to the reduced proportion of droplets containing an active polymerase-producing cell. As increasing the cell number above 3.3×10^8 cells per ml had no effect there may be another limiting factor such as the concentration of PCR reagents at this level. An auto-induction expression system was used to increase the expression level of Sac7-TaqPol in order to increase the yield of emPCRs. Optimisation of the emPCR will be discussed further in chapter 3.

Having successfully established and tested the required *Taq* DNA pol clones and conditions to set up a functioning CSR process, the next step was to use the system as a whole to evolve polymerase variants. This involved the creation and screening of a mutant library and selection of polymerases with improved properties. In the first instance, the CSR approach was used to evolve and screen for heparin resistant polymerase mutants, as described in Chapter 3.

Chapter Three - Testing the CSR system by Evolving *Taq* DNA polymerase Variants with Resistance to Heparin

Introduction

The widespread use of PCR-based diagnostic testing has led to an increase in the study of PCR inhibitors, which are abundant in biological samples. Many diagnostic tests, for both microbial infections and genetic disease, are carried out on blood samples, which often contain the inhibitor heparin. As early as 1978, heparin was discovered to be an inhibitor of cellular DNA polymerases and reverse transcriptase (DiCioccio and Srivastava, 1978). More recently, heparin was shown to inhibit *Taq* DNA polymerase amplification of the HIV genes in a detection test (Holodniy *et al.*, 1991). The effect of heparin on PCR using white blood cells was further studied by Yokota and it was found that as little as 0.0016 to 0.1U of heparin³ in a 50µl PCR was enough to suppress the reaction so that there was no visible amplification (Yokota *et al.*, 1999).

Heparin has wide therapeutic use as an anticoagulant agent as it prevents the formation of blood clots, and it is therefore found in many blood samples. Heparin binds to the plasma protease inhibitor antithrombin, which leads to an increase in the rate of inhibition of coagulation proteases (Jordan *et al.*, 1979). It has been found that between 20 and 50U of heparin per millilitre is needed to prevent coagulation in whole blood.⁴ Heparin is a highly sulphated glycosaminoglycan (GAG) or polysaccharide (Mulloy and Forster, 2000) which falls under the broad category of polyanions (Jones *et al.*, 2004). It is a polymer made up of repeating disaccharide units (Figure 3.1) and can have a molecular weight ranging from 3 to 30kDa (Hirsch *et al.*, 2001). Heparin occurs naturally in the form of many different salts and can be found in vertebrate cells such as liver, lung and mast cells. There are many available commercial preparations of heparin with various molecular weights although most are around 17-19kDa (Sigma). At physiological pH, heparin is a highly negatively charged compound, by virtue of the multiple sulphate groups on the repeating sugar unit.

³ A heparin unit (U) is a measure of the anticoagulant properties of a given heparin product as it acts on antithrombin III (AT-III). One unit of heparin is the quantity required to keep 1ml of cat's blood fluid for 24h at 0°C.

⁴ Sigma technical bulletin #870.

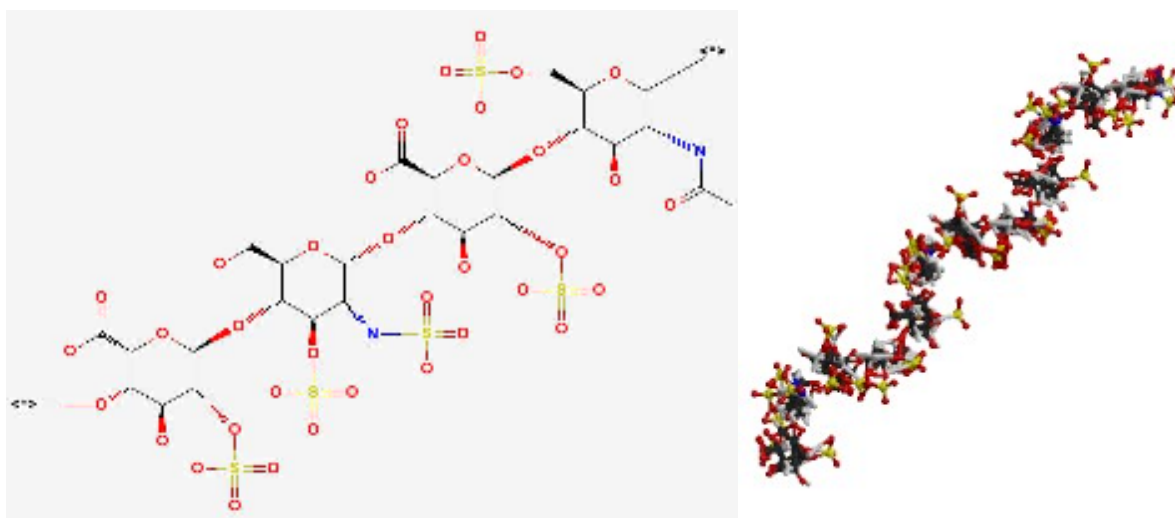


Figure 3.1 – The structure of heparin. (a) An example of the repeating disaccharide units that make up Heparin (PubChem SID3664) (b) The helical conformation of a heparin molecule (Mulloy *et al.*, 1993; PDB-1hpn).

Studies suggest that heparin competitively inhibits DNA polymerase by reversibly binding to the enzyme and interfering with DNA template binding (DiCioccio and Srivastava, 1978). Characterisation studies of heparin show that it forms a helical structure and that many of its protein interactions are dominated by its highly-acidic sulphate groups (Mulloy *et al.*, 1993). Although heparin is known to form high-affinity interactions with proteins such as the antithrombin complex, it also forms less specific interactions with many proteins (Mulloy *et al.*, 1993). The complementary electrostatic interactions between polyanions (e.g. heparin) and proteins have been studied, which has led to a number of proteins being identified and classified as ‘Heparin-binding proteins’ (Jones *et al.*, 2004). These heparin binding sites have even been predicted to be ‘helical’ in structure (Cardin and Weintraub, 1989). All these findings suggest that heparin may inhibit DNA polymerase by mimicking the helical and negatively-charged structure of DNA.

The technique of directed evolution by compartmentalised self-replication (CSR) has been used to create DNA polymerase variants with increased resistance to heparin (Ghadessy *et al.*, 2001). After three rounds of CSR screening in the presence of increasing amounts of heparin, a variant of *Taq* DNA polymerase was isolated that was estimated to have a 130-fold increase in resistance compared to wild-type. There were six mutated residues in total in the heparin-resistant *Taq* DNA pol variant and four of the residues are known through structural studies to directly contact either the template or primer strand. The mutations clustered in the base of the finger and thumb polymerase sub-domains. These

results suggested that DNA and heparin bind in the same place in the polymerase active site (Ghadessy *et al.*, 2001) and also support the view that heparin mimics and competes with the DNA template.

With the Ghadessy study in mind, it was decided that concentrating the mutations to the polymerase domain alone would be beneficial in the selection of inhibitor resistant polymerases described in this chapter. A mutant library of polymerase variants was constructed using random mutagenesis but where only the polymerase domain was exposed to mutations. The GeneMorph II Random Mutagenesis Kit (Stratagene) was used to introduce low levels of mutation into the polymerase domain of the Sac7-TaqPol clone resulting from the work described in Chapter 2, by error-prone PCR.

Once a mutant library comprising many thousands of mutants was constructed, it was screened by CSR in the presence of heparin, the aim being that polymerase variants that are able to replicate their own gene under the selective pressure of heparin inhibition would come through the screen. These heparin resistant *Taq* DNA pol variants were then characterised and assayed to determine their level of heparin resistance compared to the wild-type Sac7-TaqPol after a single round of directed evolution. It was deemed that setting up the CSR system (described in this chapter) for the evolution of *Taq* DNA polymerase variants resistant to heparin, as had been conducted previously (Ghadessy *et al.*, 2001), would provide a powerful verification of the system.

Materials and Methods

Random Mutagenesis

Random mutagenesis of the polymerase domain of Sac7-TaqPol clone (M8 Chapter 2) was carried out using the GeneMorph II Random Mutagenesis Kit (Stratagene, Cheshire, UK). Error prone PCR was performed using Mutazyme II DNA polymerase (Stratagene, Cheshire, UK) under conditions designed to achieve a mutation frequency of 0-4.5 nucleotide changes per kb. The mutagenesis reaction was set up as follows (final concentrations): 1X Mutazyme buffer (Stratagene, Cheshire, UK), 200 μ M dNTPs (Stratagene, Cheshire, UK), 125ng of each polymerase domain primer (GeneSys, Camberley, UK), ~100ng initial target template⁵ (480ng Sac7-TaqPol M8), 5U Mutazyme II (Stratagene, Cheshire, UK) and ddH₂O to 50 μ l. The reactions were run on an Eppendorf Mastercycler under Gene Morph II recommended temperature cycling conditions as follows: 95°C for 3min, 25cycles of 95°C for 30s, 65°C for 30s, 72°C for 2min. Original Pol domain primers (GeneSys, Camberley, UK) were used in initial reactions (restriction sites in grey): Forward 5' AGGGCCTTTCTGGAGAGATCTGAGTTTGGCAGCCTCC and Reverse 5' GCCTGCAGGTCGACTCTAGAGG ATCTATCACTCC TTGG. After cloning problems, a second set of primers were designed outside the original pol primers and used in further random mutagenesis reactions. Outer pol domain primers (MWG, Ebersberg, Germany): Forward 5' ATCCTGGCCCACATGGACGA and Reverse 5' CCGCCAGGCAAATTCT GTTTTATCA. 10 μ l of PCR product was run on a 1% (w/v) agarose gel with a 1kb DNA ladder (New England Biolabs) and a 1.1kb standard (Stratagene, Cheshire, UK). PCR products were pooled and purified using Perfect Prep Gel Cleanup Kit (Eppendorf, Cambridge, UK).

Vector and PCR Product Digest

pTTQ18NHK Sac7-TaqSacB (Chapter 2, Figure 2.6) and the mutant polymerase domain PCR product were digested with *Bgl*II and *Sal*I (New England Biolabs) following the recommended protocol. 10 μ l of each digest was run on a 1% (w/v) agarose gel with

⁵ Initial target DNA refers to the amount of target DNA to be amplified, not total amount of plasmid DNA template to add to the reaction (as per GeneMorph II instruction manual).

undigested controls to check proper digestion had occurred. The vector digest (pSac7-TaqSacB) was then treated with Shrimp Alkaline Phosphatase (New England Biolabs) as recommended. Digests were purified using Perfect Prep Gel Cleanup Kit (Eppendorf, Cambridge, UK). 1µl of purified, digested product was run on a 1% (w/v) agarose gel with 1kb DNA ladder (New England Biolabs).

When sequential digests were run, the vector and PCR product were first digested with *Bgl*II, followed by a clean-up step and digestion with *Sal*I.

Ligation

In general, ligations were set up with a 1:4 molar ratio of vector to insert. Each ligation contained 1X T4 Ligase Buffer (Promega, Southampton, UK), vector (240ng), insert (180ng), 1µl (1-3U) T4 Ligase (Promega, Southampton, UK) and ddH₂O to 10-30µl. Ligations were left at room temperature for 3h. After incubation 20µl of ddH₂O was added to the reaction and the ligase was heat killed at 70°C for 20min. The ligation product was precipitated either in ethanol or butanol before further use.

Butanol Precipitation

The ligation reaction was made up to 50µl total volume with ddH₂O. 500µl of butanol was added, the sample was vortexed briefly and then centrifuged at 16,000g for 10min. The butanol was removed and the DNA pellet was dried in a SpeedVac for 10min or until all residual traces of butanol had been removed. Pellets were resuspended in 5µl ddH₂O.

AKTA Plasmid Purification

A 2L LB [10g/L tryptone, 5g/L yeast extract, 5g/L NaCl (all Sigma, Gillingham, UK)] culture was inoculated from a freshly streaked plate of Sac7-TaqSacB in TOP10F' and grown up overnight at 37°C. Cells were harvested by centrifugation and the cell pellet was resuspended in 60ml S1 [50mM Tris-HCl (pH7.5), 10mM EDTA, no RNase added]. 60ml S2 (0.2M NaOH, 1% SDS) was added and the tube was gently inverted. After 5min 60ml S3 (1.25M Kacetate, 14% glacial acid) was added, the tube was mixed and left on ice for a further 5min. To remove cell debris the sample was centrifuged for 1h at 14,000g. The lysate was removed and clarified by running through sterile filter wool. An equal volume of

isopropanol was added and the sample was centrifuged for 30min. The supernatant was discarded and the pellet was resuspended in 10ml TE (Tris pH 8, 0.1mM EDTA). 0.5µl of the sample was run on a 1% (w/v) agarose gel with EcoRI/HindIII DNA ladder (GeneSys, Camberley, UK).

After clarification of the cell lysate the sample was loaded onto a Sepharose 6 Fast Flow column (GE Life Sciences, Little Chalfont, UK) for RNA removal and buffer exchange, followed by capture of the supercoiled plasmid on a PlasmidSelect column. Plasmid purification chromatography was performed using an AKTA (GE Life Sciences, Little Chalfont, UK) and following the PlasmidSelect Plasmid Purification Starter Kit protocol (GE Life Sciences, Little Chalfont, UK).

Second Amplification of Mutant PCR Product

A 1/10 dilution of mutagenic pol domain PCR product (using outer pol primers) was used as template in a second PCR. This PCR used the original pol domain primers (see random mutagenesis) which contain restriction sites. PCR was set up as follows (final concentrations): 1X Taq Buffer containing 1.5mM MgCl₂ (GeneSys, Camberley, UK), 200µM dNTPs (GeneSys, Camberley, UK), 25pmol of each pol domain primer (GeneSys, Camberley, UK), 1µl 1/10 dilution of template, 5% (v/v) DMSO (Sigma, Gillingham, UK), 1.25U *Taq* DNA polymerase (GeneSys, Camberley, UK) and ddH₂O up to 50µl. Reactions were run on an Eppendorf Mastercycler for 10 cycles under pol domain PCR conditions: 94°C for 3min, 25 cycles of 94°C for 15s, 60°C for 15s, 72°C for 2min. 5µl of PCR product was run on a 1% (w/v) agarose gel with 1kb DNA ladder (New England Biolabs). Multiple PCRs were pooled and purified with the Perfect Clean Up kit (Eppendorf, Cambridge, UK) before further use.

Screening for Pol Domain Insert

Colonies were picked and lysed as described in the Chapter 2 screening method. PCR reactions were run as follows (final concentrations): 1X MasterMix (GeneSys, Camberley, UK), 25pmol pol domain F and R primers, 1µl template and water up to 50µl. Reactions were cycled under screening PCR conditions: 94°C for 3min, 30 cycles of 94°C for 10s, 60°C for 10s and 72°C for 2min. Sac7-TaqPol M8 was used as control template. 5µl of each PCR was run on a 1% (w/v) agarose gel.

Glycerol Stocks of Mutant Library

Mutant library transformations were spread onto large LB Kan/Sucrose (5%) plates and left overnight at 37°C. The colonies were then resuspended on the plate using LB and gentle agitation with a spreader. The resuspended cells were pipetted and pooled into a cryotube and 0.5 vol of 50% (v/v) glycerol was added. The tube was gently mixed before being stored at -80°C.

***DpnI* Treatment and Second Amplification of emPCR Product**

After initial ether extraction, the aqueous phase of the emulsion PCR (emPCR) was treated with 20U of *DpnI* (New England Biolabs) for 2h at 37°C to remove parental plasmid DNA. The sample was then concentrated using a SpeedVac before being purified using a spin column PCR Purification Kit (Qiagen, Crawley, UK) and eluted in 50µl of ddH₂O. The *DpnI* treated emPCR product was then used as template in a second PCR. The PCR was run as follows (final concentrations): 1X HF Phusion Buffer (New England Biolabs, Hitchin, UK), 200µM dNTPs (GeneSys, Camberley, UK), 25pmol of each pol domain primer, 1-5µl emPCR template, 5% DMSO (Sigma, Gillingham, UK), 1U Phusion (New England Biolabs, Hitchin, UK), and ddH₂O to 50µl. Reactions were run for 25 cycles following Phusion guidelines on an Eppendorf (Cambridge, UK) Mastercycler.

Protein Expression and Polyethyleneimine Precipitation

250ml auto-induction medium Overnight Express (Novagen, Nottingham, UK) containing kan (50µg/ml) was inoculated from a freshly streaked plate. The culture was left shaking at 37°C for 24h and then the cells were harvested by centrifugation. Cell pellets were weighed and resuspended in 5ml TE [50mM Tris (pH7.5), 1mM EDTA] per gram of cells. Ammonium sulphate was added to a final concentration of 0.2M and samples were left at 80°C for 30min for lysis and precipitation to occur. The samples were cooled to 4°C and 100µl aliquots of the supernatant were tested with varying concentrations of polyethyleneimine (PEI) ranging from 0-1%. Samples were left overnight at 4°C and centrifuged at 14,000g for 10min. Samples were run on a 1% (w/v) agarose gel and an 8% (w/v) SDS-PAGE gel. 0.4% (w/v) PEI was found to be optimal and was added to the remainder of the sample, which was then left at 4°C overnight. The sample was spun at 14,000g for 20min and the supernatant was kept for further purification.

Buffer Exchange and Sample Concentration

PEI-treated cell lysate was filtered through a 0.22µm pore filter (Millipore, Watford, UK) before being transferred to a 10kDa (NMWL) Amicon Ultra-15 centrifugal filter device (Millipore, Watford, UK). The device was centrifuged at 6,000g until the sample was concentrated to about 2ml. The sample was then resuspended in 12ml Buffer A [150mM KCl, 50mM Tris (pH 7.5), 0.2% (v/v) Tween 20, 0.1mM EDTA] and centrifuged again to concentrate the sample. Buffer exchange and sample concentration continued until all ammonium sulphate and PEI were removed. The final sample volume was approximately 2ml.

Protein Purification (Affinity Chromatography)

Concentrated protein in Buffer A was filtered through a 0.22µm pore filter (Millipore, Watford, UK) before being loaded onto a 5ml Heparin-Sepharose High Performance column (GE Life Sciences, Little Chalfont, UK). Protein was purified using affinity chromatography with a salt gradient elution from 0% - 100% Buffer B [600mM KCl, 50mM Tris (pH 7.5), 0.2% (v/v) Tween 20, 0.1mM EDTA] over 50min at 1ml/min flow rate. Protein eluted at approximately 300mM KCl. 1ml fractions were collected and samples were analysed on an 8% (w/v) SDS-PAGE gel.

Dialysis and Protein Concentration using Polyethelene glycol

Dialysis tubing was prepared as follows: tubing was treated with 2% sodium bicarbonate and 1mM EDTA and boiled for 10min; it was then rinsed with ddH₂O and either used immediately or stored in 20% (v/v) ethanol. Key fractions that were collected during protein purification were pooled together and pipetted into a sterile dialysis tube. The dialysis tube was left in a beaker of Polyethylene glycol (PEG) 6000 until the volume of the protein sample had decreased by half. The dialysis tubing was then transferred to a beaker of *Taq* Storage Buffer [100mM KCl, 20mM Tris (pH 8), 0.1mM EDTA, 1mM Dithiothreitol (DTT), 0.5% (v/v) Tween 20, 0.5% (v/v) Igepal CA630 (Sigma, Gillingham, UK)] and left for a 4h before being transferred to a fresh beaker of *Taq* Storage Buffer and left overnight at 4°C to dialyse. Samples were removed from the dialysis tube and stored at 4°C.

Polymerase Activity Assay

1 μ M M13 Oligonucleotide (5'CGCCAGGGTTTTCCCAGTCACGACGTTGTA AAACGACGGCC) was pre-annealed to 1 μ g M13mp18 ssDNA template (New England Biolabs) in a 1X Taq Buffer (GeneSys, Camberley, UK) total volume 10 μ l. Reagents were mixed and heated to 95°C for 2min and then cooled (slow ramp speed on PCR machine) to 72°C for 5min to allow annealing. Annealed oligo/template was cooled to 4°C before other reaction components were added. Final reaction concentrations were as follows: 1X Taq Buffer (GeneSys, Camberley, UK) containing 1.5mM MgCl₂, 200 μ M dNTPs (GeneSys, Camberley, UK), 1 μ g M13 ssDNA, 0.04 μ M M13 oligo and water up to 45 μ l. The reaction was kept on ice, 5 μ l of enzyme was added and the reaction was split into two aliquots of 24 μ l. One aliquot was kept on ice and used as a control/time zero, the other was heated at 72°C for 10min and then placed immediately back on ice. 75 μ l of TE [10mM Tris (pH 7.5), 1mM EDTA] was added to each aliquot and after mixing 50 μ l of each reaction was transferred to a black flat bottomed 96-well plate (Thermo Scientific, Basingstoke, UK). A 1:400 dilution of PicoGreen[®] (Invitrogen, Paisley, UK) in TE was made and 50 μ l was mixed with each reaction in the 96-well plate. The plate was incubated at room temperature for approximately 5min. The amount of DNA synthesised was quantified using a PHERAstar fluorescent plate reader (BMG Labtech, Aylesbury, UK). Graphical and statistical analysis of the data were performed using FigSys (Biosoft, Cambridge, UK).

Taq DNA polymerase control assays were performed with commercially available *Taq* (GeneSys, Camberley, UK) at a known Unit⁶/ μ l concentration. These data were used for unit estimation of the mutant polymerases.

General Methods

Standard protein expression was performed as in Chapter 2 but Overnight Express auto-induction medium (Novagen, Nottingham, UK) was used instead of LB and cultures were left for 24h. Due to discontinuation of the original electrocompetent cell strain, TOP10F' (Invitrogen, Paisley, UK), NEB 10- β strain was used in all further work. SDS-PAGE analysis was performed as described in Chapter 2. Protein concentrations were measured using the Bradford Assay (BioRad, Hemel Hempstead, UK).

⁶ One Taq unit is defined as the amount of enzyme that will incorporate 10nmol of dNTP into acid-insoluble material in 30min at 75°C.

Standard activity testing using lysate in M13 and 500bp λ PCRs was as described in Chapter 2.

All DNA sequencing used plasmid DNA as template and was carried out by GeneServices (Cambridge, UK).

Results

Random Mutagenesis

Random mutagenesis was carried out on the polymerase (pol) domain of the Sac7-TaqPol M8 clone using the GeneMorph[®] II Random Mutagenesis Kit (Stratagene). The GeneMorph protocol was followed and the initial target amount of 100ng (equivalent to 480ng plasmid template) with 25 PCR cycles was used to achieve a low mutation frequency (0-4.5 nucleotides/kb). The product was analysed on an agarose gel (Figure 3.2) with 100ng of 1.1kb DNA standard (Stratagene).

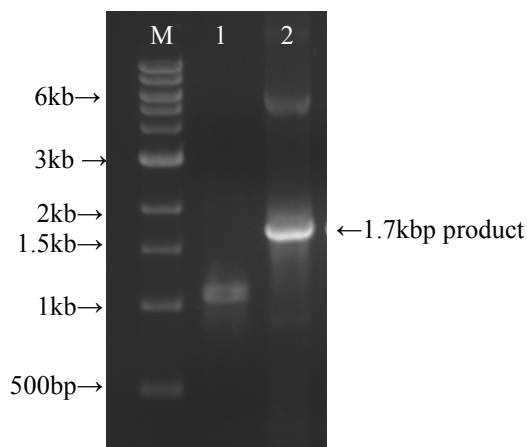


Figure 3.2 – Random Mutagenesis PCR. This gel shows the PCR product of the random mutagenesis reaction. (M) 1kb DNA ladder, (1) 1.1kb DNA standard (100ng), (2) PCR product.

The random mutagenesis reaction produced product of the correct size (1.7kb) and had a good yield. The 1.7kb corresponds to the expected size of the *Taq* Pol domain PCR insert. The mutant polymerase domain PCR product and the vector pTTQ18NHK Sac7-TaqSacB were digested with *SalI* and *BglII*. Digestion of the vector removed the SacB domain selective marker (as discussed in Chapter 2) and allowed insertion of the mutant Pol domain during cloning.

Following initial unsuccessful attempts to ligate and clone the mutated PCR product, a large scale purification of pTTQ18NHK Sac7-TaqSacB was carried out using PlasmidSelect to produce a good, clean sample of vector. New pol domain primers were also designed outside the original Pol domain primers. Investigation of the unsuccessful cloning attempts suggested that disruption of the restriction sites might be causing a problem. To

combat this, random mutagenesis was performed using the outer pol domain primers and the PCR product was then subjected to a second amplification with the inner pol domain primers to make sure the restriction sites were intact (as they are contained in the primer sequence). The amplification was successful and multiple PCR reactions were pooled and purified. Both vector and mutant Pol domain were then sequentially digested with *Bgl*II and *Sal*I before ligation and transformation into Top10F' cells. Colonies were screened for pol domain insert by PCR as displayed on the agarose gel in Figure 3.3.

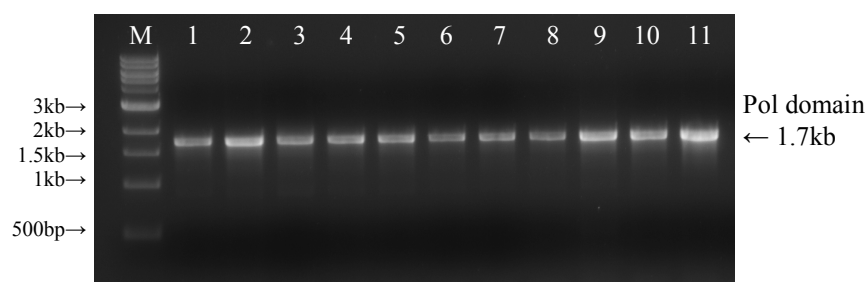


Figure 3.3 – Polymerase domain screening PCR. Ten colonies were screened for the presence of polymerase domain insert. (M) 1kb DNA ladder, colonies numbered 1-10, (11) pol domain control.

The screening PCR showed that 10/10 of the colonies screened had the correct size insert and the ligation of the mutant Pol domain had been successful. The ligation was then repeated but on a larger scale (600ng total DNA). Transformants were screened for insert and the positive clones were grown up for expression and activity testing. Although the majority of the clones expressed protein of the correct size, none of them were active in a standard control PCR. Plasmid DNA was extracted from randomly selected clones and sequenced. The mutant polymerase domain sequences were aligned using CLUSTALW and the mutation frequency was determined. Mutations ranged from 6-13 nucleotides per Pol domain with an average mutation frequency of 5.6 per kb. This was slightly higher than expected which may explain the low proportion of active polymerase variants observed. It was decided to adjust the mutagenesis rate before proceeding to library construction.

Random Mutagenesis was repeated as before with the outer pol domain primers but with more initial template (580ng Sac7-TaqPol M8) in order to lower the mutation frequency. Once again all the colonies that were screened contained the correct size insert. Sequencing and alignment showed the mutation frequency to be an average of 6.5 nucleotide changes per Pol domain which corresponds to a mutation frequency of 3.8 nucleotides per kb. These mutations led to an average of 5 amino acid changes per Pol domain. Standard

expression and activity testing revealed that most of the selected clones expressed the correct size protein and that 20% showed some activity in standard PCR. The percent of active clones had increased as desired. This lower mutation frequency mutagenesis reaction was used to construct the polymerase variant library used in screening.

Production of the Mutant Library

In order to create a large mutant library the transformation was extensively optimised to achieve the highest efficiency possible. The voltage at which the cells were electroporated was investigated and 1.8kV was found to be optimal. The concentration of DNA in the electroporation had an effect on the number of transformant colonies produced. It was found that transformations with 100ng DNA produced more colonies than if only 1ng of DNA was transformed. The pre-transformation treatment of the ligation reaction was also found to be a critical variable. During optimisation it was found that both ethanol- and butanol-precipitated samples showed an approximate 20-25% loss of DNA but that the butanol-precipitated ligation had a 10-fold higher efficiency than that of the ethanol-precipitated ligation. Butanol precipitation followed by drying of the DNA pellet using a SpeedVac was selected for use during library construction.

Multiple ligation and transformation reactions were conducted in order to construct the mutant library. Colonies were counted and used to estimate total variant number or library size. A single transformation produced 1×10^5 colonies on average. A total of 10 transformations were carried out to produce a final library size of 1×10^6 variants. The mutant library was then pooled, aliquoted and stored at -80°C .

A final characterisation of the library was performed to make sure the ligations were successful and the estimated mutation frequency was still accurate. It was confirmed that 100% of the 20 clones screened were positive for the Pol domain insert. Sequencing confirmed that the mutation frequency of the final library was 4.8 nucleotide changes per kb. There was an average of 8.2 nucleotide changes and 6 amino acid changes per pol domain. Details of the final library are listed in the table below (Table 3.1). Standard activity and expression tests showed that, while most of the mutant library clones expressed protein of the correct size, less than 20% of them showed polymerase activity.

Table 3.1 – Characterisation of Final Mutant Library

Range of nucleotide changes seen in mutants (per pol domain)	1-13
Average nucleotide changes per pol domain	8.2
Average nucleotide changes per kb	4.8
Average amino acid changes per pol domain	6.0
% Transitions	48
% Transversions	53
Ts/Tv	0.9
% Silent Mutations	27
% Conservative Amino Acid Changes	21

Optimization of Emulsion PCR

Further optimisation of the emulsion PCR (emPCR) was conducted prior to screening the above generated mutant library in the system since it was expected that its activity would be much lower than that of wild-type polymerase expressing culture. A series of emPCRs were carried out using not only the mutant library but also different percentages of cells expressing the wild-type polymerase. Under the same conditions it was found that no amplification was observed by agarose gel electrophoresis using the mutant library cells, whereas a high yield was observed with the wild-type cells (Figure 3.4). The wild-type cells were found to still produce a good yield even when the amount of cells was considerably decreased (Figure 3.4).

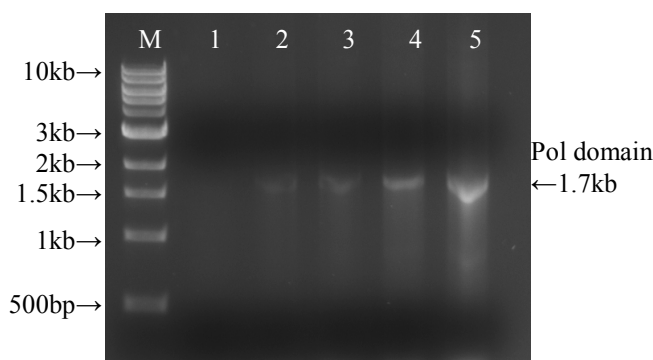


Figure 3.4 – emPCR with varying amounts of *E. coli* cells.

The standard amount of cells in an emPCR is 2×10^8 and this was varied in this experiment to see the effect on amplification. (M) 1kb DNA Ladder, (1) 2×10^8 (100%) mutant library cells, (2) 0.2×10^8 (10%) wt cells, (3) 0.5×10^8 (25%) wt cells, (4) 1×10^8 (50%) wt cells, (5) 2×10^8 (100%) wt cells.

The emPCR with the mutant library produced no visible amplification. Using as little as 10% of wild-type polymerase cells in an emPCR produced some amplification, which could indicate that less than 10% of the mutant library is made up of polymerases that are active under the emPCR conditions. The above experiment indicated that the system required further optimisation. Firstly, to test that the mutant library culture had expressed at least a low level of active polymerases, a cell lysate was made and used in a standard activity PCR. The activity of the wild-type and mutant polymerases was compared by agarose gel (Figure 3.5).

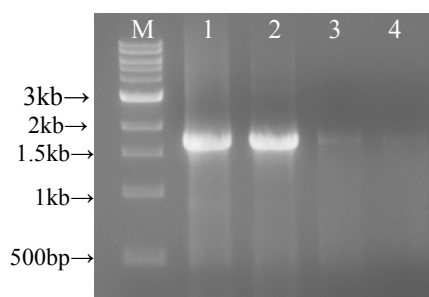


Figure 3.5 – Lysate activity PCR

Agarose gel displaying the activity PCR using lysate from a mutant library and a wild-type culture (M) 1kb DNA ladder (NEB), (1) wt, (2) wt, (3) mutant library, (4) mutant library

The gel showed that the mutant library was producing some active polymerases as amplification could be seen in lanes 3-4 (Figure 3.5). However, as expected the activity from the mutant polymerase lysate was much lower than that of the wild-type polymerase lysate, presumably due to the relatively low proportion of active variants in the mutant library.

Further optimisation of the emPCR was done to try to improve overall yield. Five identical emPCRs were run using cells from a wild-type polymerase culture and each was run for a different number of PCR cycles ranging from the standard 20 up to 40. Analysis of the emPCR showed that increasing the cycle number did not increase the yield of the PCR and in fact, 20-25 cycles was found to produce the most DNA. This suggested that there may be a limiting reagent preventing further amplification beyond 25 cycles. Several more emPCRs were set up and run under standard conditions but with varying either the Mg^{2+} or primer concentration to see if an increase in either improved the PCR yield. Analysis of the emPCRs showed that increasing neither the Mg^{2+} or primer concentration had a positive effect on the yield of the PCR. However, the addition of 5% DMSO to the PCR, which is known to aid in dissociation of the DNA strands in GC-rich templates, was found to decrease non-specific binding of the primers and increase overall yield of the PCR. Next,

varying the dNTP concentration was tested and, as dNTPs are known to chelate Mg^{2+} , this too was varied alongside (Figure 3.6).

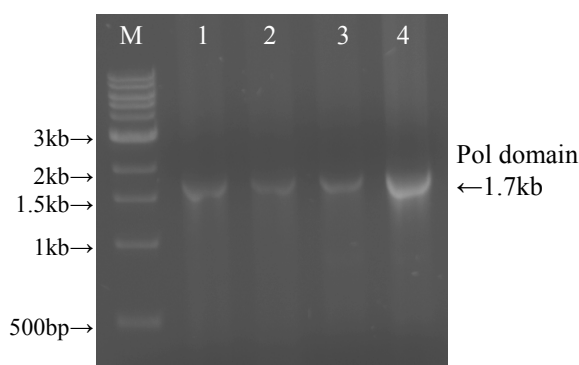


Figure 3.6 – Optimisation of emPCR.

Agarose gel displaying emPCR products under various dNTP and $MgCl_2$ concentrations (M) 1kb DNA ladder (NEB), (1) standard 250 μ M dNTPs, 1.5mM $MgCl_2$, (2) 500 μ M dNTPs, 1.5mM $MgCl_2$ (3) 250 μ M dNTPs, 2mM $MgCl_2$, (4) 500 μ M dNTPs, 2mM $MgCl_2$.

Analysis of the emPCRs showed that just doubling the dNTP concentration had little effect on the PCR; however, when both the dNTPs and Mg^{2+} were increased the yield of the PCR improved significantly. From this point on it was decided that 500 μ M dNTPs, 2mM $MgCl_2$ and 5% DMSO would be used in the standard emPCR protocol.

Two final aspects of the emPCR were tested and they were PCR volume and inhibition by oil components. A standard polymerase domain PCR was run and all the components of the oil phase (mineral oil, Span 80, Tween 20 and Triton X-100) were spiked into the PCR to see if they had an effect. None of the components had an inhibitory effect on the PCR when they were used at equivalent concentrations to those used in an emPCR. When three different volume PCRs (50 μ l, 100 μ l and 150 μ l) were run, agarose gel analysis showed that their yields differed greatly. The Eppendorf MasterCycler PCR block was not designed to heat larger volumes, so a significant decrease in relative yield was seen when a 150 μ l PCR was run. The relative yield of a 100 μ l PCR was slightly less than that of a standard 50 μ l PCR. In order to get optimal temperature cycling the emPCRs should be aliquoted into volumes less than 100 μ l. This information was important because of the large volume of each emPCR (600 μ l) that is initially mixed and is required to be aliquoted prior to thermocycling. Despite numerous attempts at optimisation, the use of mutant library cells in an emPCR still gave very little, if any, visible yield. It was decided that a second amplification using the emPCR product as template would be needed to produce enough DNA to enable cloning after a CSR screen.

CSR Heparin Screen

All naturally occurring polymerases that have been characterised have different levels of sensitivity to heparin. As the amount of polymerase in each *E. coli* cell was unknown it was important to calculate the sensitivity of Sac7-TaqPol to heparin in an emPCR in order to derive the level of heparin to be used in the CSR screen. First, a 10x dilution series of heparin was tested in an emulsion PCR with wild-type polymerase expressing *E. coli* cells. The products were analysed on an agarose gel and it was found that between 5U/ml and 0.5U/ml of heparin inhibited the PCR (Figure 3.7). A 2x dilution series between the two values was then carried out to find the exact heparin sensitivity threshold (Figure 3.8).

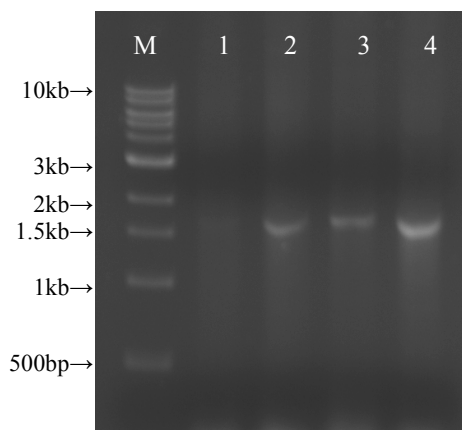


Figure 3.7 – emPCRs with 10x dilution series of heparin. Agarose gel displaying the emPCRs at different levels of heparin. (M) 1kb DNA ladder (NEB), (1) emPCR with 5U/ml heparin, (2) emPCR with 0.5U/ml heparin, (3) emPCR with 0.005U/ml heparin, (4) no heparin control.

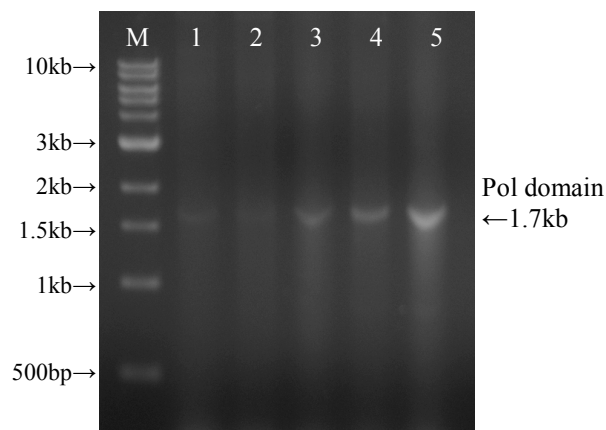


Figure 3.8 – emPCR with 2x dilution series of heparin. (M) 1kb DNA ladder (NEB), (1) emPCR with 5U/ml heparin, (2) emPCR with 2.5U/ml heparin, (3) emPCR with 1.25U/ml heparin, (4) emPCR with 0.625U/ml heparin, (5) no heparin control.

Analysis of the emPCRs showed that PCR within the emulsion containing cells expressing wild-type Sac7-TaqPol was inhibited, to the extent that no product could be observed by agarose gel electrophoresis, by any concentration greater than 1.25U/ml of heparin. Amplification appeared to be completely inhibited at 5 and 2.5U/ml of heparin. At lower concentrations of 1.25 and 0.625U/ml of heparin there was still some inhibition as the amplification yield was significantly decreased compared to the no heparin control. The heparin sensitivity results obtained from wild-type polymerase were used as a guideline to pick the level of heparin in which to screen the mutant library. It was decided that the CSR

heparin screen using cells expressing mutant polymerase variants would be run under two different concentrations of heparin: 2.5U/ml and 1.25U/ml.

Despite the optimisation carried out in the previous section, the yield of an emPCR containing the mutant library was very low and the addition of heparin to the emPCR produced even further inhibition and reduction in yield of the PCR. A second amplification of the emPCR product was needed to produce enough DNA to enable cloning. Multiple emPCRs for each heparin concentration (2.5 and 1.25U/ml) were set up using the mutant library and the Outer Pol domain primers. After thermocycling and extraction, the reactions were concentrated and treated with *DpnI* to make sure any parental DNA was removed. The emPCRs were pooled together and purified but, even after combining multiple emPCRs, no product was visible when the samples were run on an agarose gel. The purified emPCR product was concentrated to prepare template for a second standard Phusion PCR using the original Pol domain primers. The PCR products from this second amplification were analysed on an agarose gel (Figure 3.9).

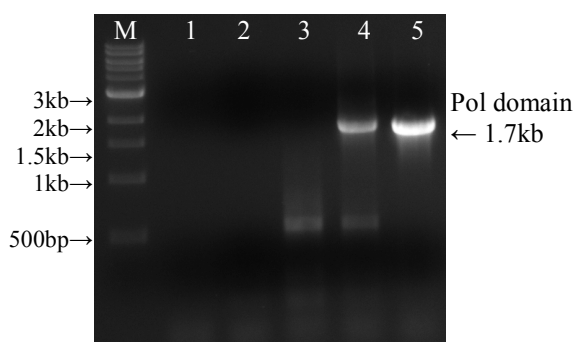


Figure 3.9 – Second amplification of the heparin emPCR.

Different amounts of the two concentrated heparin emPCR were used as template to re-amplify the polymerase domain. (M) 1kb marker, (1) 5µl of 2.5U/ml template, (2) 1µl of 2.5U/ml template, (3) 5µl of 1.25U/ml, (4) 1µl of 1.25U/ml, (5) control template.

Product was only observed on the gel in the 1.25U/ml heparin emPCR. In multiple attempts, no amplification could be obtained using the 2.5U/ml heparin emPCR as template. The second amplification Phusion PCR was repeated with the 1.25U/ml heparin emPCR template to produce enough DNA for cloning.

Selecting Heparin Resistant Polymerase Variants

The mutant polymerase domains that were amplified during the CSR heparin screen belonged to those polymerases that were still active under the inhibitory heparin conditions of the screen. In order to further select for the polymerase with the highest resistance to heparin, the polymerase domain had to be inserted into a full-length Sac7-TaqPol clone so that the whole protein variant could be expressed. The amplified mutant polymerase domain PCR product (from 1.25U/ml of heparin reaction) was cloned into the *Bgl*II and *Sal*I site of the pTTQ18NHK Sac7-TaqSacB vector following excision of the SacB domain as described in Chapter 2. Subsequent transformation into NEB10- β cells produced the first round heparin resistant mutants.

A standard expression and activity test was done using 25 of the first round mutants. A cell extract was prepared using heat precipitation as standard. SDS-PAGE gel analysis of the heat-treated cell extracts revealed that 80% of the mutants showed strong expression of the correct size protein while the remaining 20% showed very little expression of polymerase. Standard 500bp λ activity PCRs with the extract revealed that 80% of the mutants had polymerase activity, which correlated with the expression results. The activity level observed was significantly higher than that of the starting mutant library, which indicated that the screen had successfully selected for active mutants.

Next the heparin sensitivity of these mutants was tested in a PCR alongside the wild-type. A cell extract of Sac7-TaqPol was prepared using heat treatment to provide a wild-type control. The extract was then used in an activity PCR using Pol domain primers and a range of heparin concentrations. The PCR reactions were analysed and it was found that under standard activity PCR conditions, the reaction was slightly inhibited by 0.625U/ml of heparin and was completely inhibited by 1.25U/ml of heparin. The sensitivity of the polymerase first round mutants was then tested at similar heparin levels. A comparable level of mutant polymerase extract was used in an activity PCR with four different levels of heparin: 0, 0.3125U/ml, 0.625U/ml and 1.25U/ml. Sac7-TaqPol extract was used as a control so that the activity of the mutants could be compared to wild-type. The activity of selected mutants in the presence of heparin is shown on the agarose gel below (Figure 3.10).

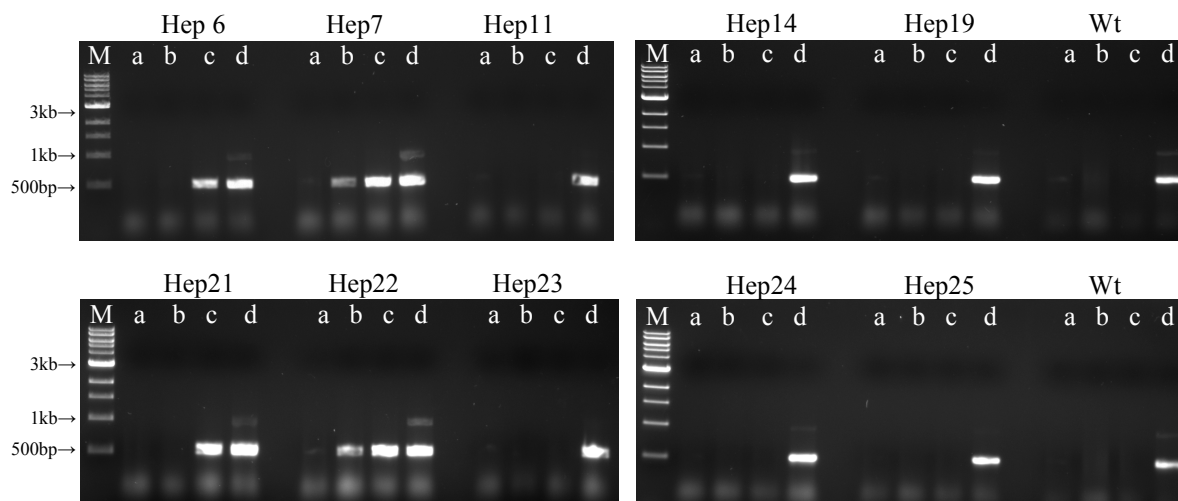


Figure 3.10 – Heparin mutant activity in PCR with different concentrations of heparin compared to wt. This agarose gel shows a 500bp λ activity PCR with different levels of heparin (a) 1.25U/ml (b) 0.625U/ml (c) 0.3125U/ml (d) 0 control. The results from various Heparin resistant mutants are displayed alongside wild-type controls.

Figure 3.10 shows that mutants 6, 7, 21 and 22 demonstrate clearly an increased resistance to heparin. These four mutants were able to produce amplification at levels of heparin that completely inhibited the wild-type polymerase. Although the remaining mutants in the above gel (Figure 3.10) showed polymerase activity in the no heparin control, they did not seem to have any increased resistance to heparin compared to the wild-type.

A total of forty polymerase variants were tested for activity and increased resistance to heparin. Ten of the active variants were selected for sequencing (GeneServices). Table 3.2 shows the amino acid changes of selected heparin resistant mutants isolated in this study as well as changes seen in the heparin resistant mutants isolated by Ghadessy *et al.* (2001). Using CLUSTALW 10 sequences were aligned (Figure 3.11) and it was found that the mutants contained from 1 to 10 amino acid changes, with an average of two amino acid changes per mutant.

Table 3.2 – The amino acid changes of selected heparin resistant mutants.**(Taq DNA polymerase numbering⁷)**

Mutant	Amino Acid Changes	Total Amino Acid Changes	Rounds of CSR
Hep 7	R630L	1	1
Hep11	V529A	1	1
Hep21	D355G, Y430N, V453T, Y456*, F482L, L490Q, L519M, K540M, S612I, E713G	10	1
Hep22	P550S, T588S	2	1
Hep35	S409T, G809A	2	1
Ghaddesy H32	K340E, Q534R, T539A, V703A, R778K	5	1
Ghaddesy H94 ⁸	K225E, L294P, A454S, L461R, D578G, N583S	6	2
Ghaddesy H15 ⁹	K225E, E388V, K540R, D578G, N583S, M747R	6	3

* stop codon

⁷ See AppendixC for Sac7-TaqPol and *Taq* alignment⁸ Patent application WO 2002/022869 A2⁹ Patent application WO 2002/022869 A2

TaqPol	270	RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep6		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep7		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep11		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep14		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep19		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep21		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep22		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep32		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep34		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD
Hep35		RAFSEERLEFGSLLHEFGLLSPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLEKARGLLAKD

		← Palm
TaqPol	356	LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep6		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep7		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep11		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep14		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep19		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep21		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep22		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep32		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep34		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL
Hep35		LSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALSERL PANLWGRLEGEERLLWL YREVERPLSAVL

		Palm → ← Thumb
TaqPol	442	AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep6		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep7		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep11		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep14		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep19		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep21		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep22		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep32		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep34		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH
Hep35		AHMEATGVRLDVAYLRALSLEVAEIEARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTKRSTSAAVLEALREAH

		Thumb → ← Palm
TaqPol	528	IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep6		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep7		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep11		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep14		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep19		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep21		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep22		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep32		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep34		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ
Hep35		IVEKILQYRELTCLKSTYIDPLDPLIHPTGRHLHTRFNQTATATGRLSSSDPNLQNI PVRTPLGQIRRAFAIEEGWLLVALDYSQ

		← Fingers
TaqPol	614	IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep6		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep7		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep11		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep14		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep19		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep21		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep22		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep32		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep34		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS
Hep35		IELRVLAHLSGDNELIRVQEGRDIHTETASWMFVGPVREAVDPLMRAAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQS

		Fingers → ← Palm
TaqPol	700	FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep6		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep7		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep11		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep14		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep19		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep21		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep22		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep32		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep34		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD
Hep35		FPKVRAWIEKTLLEEGRRRGYVETLFGRRRYVPDLEARVKSVEAAERMAFNMPVQGTADLMKLAAMVKLFPRLPEEMGARMLLQVHD

		Palm →
TaqPol	786	ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep6		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep7		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep11		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep14		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep19		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep21		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep22		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep32		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep34		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----
Hep35		ELVLEAPKERAEEAVARLAKEVMGVPVPLAVPLEVEVIGEDWLSAKE----

Figure 3.11 – Alignment of the polymerase domain of wild-type *Taq* and selected heparin mutants. Alignment of *Taq* and 10 selected 1st round heparin mutants was performed using ClustalW. Mutations are highlighted in red. * = single fully conserved residue, : = conservation of strong groups, . = conservation of weak groups. Thumb, finger and palm sub-domains are labelled.

Purification of Selected Heparin Resistant Mutants

The five polymerase mutants, which displayed the most heparin resistance in initial PCR tests were grown up, expressed and purified alongside wild-type Sac7-TaqPol.

An ammonium sulphate precipitation and heat lysis step were performed on the cell lysate to remove unwanted *E. coli* proteins and a PEI precipitation was performed to remove nucleic acid from the protein preparation. 100µl aliquots of extract were tested with various concentrations of PEI to determine the optimum level for complete removal of the nucleic acids while still leaving the protein in solution. Following this study, 0.4% (w/v) PEI was chosen as the optimal level for further PEI treatment.

After PEI precipitation, the protein sample underwent buffer exchange to remove the PEI and ammonium sulphate that had been previously added during the first steps of purification. The existing buffer was replaced with Buffer A (p.51), which meant that the protein sample was ready to be loaded onto a Heparin Sepharose column (GE LifeSciences) pre-equilibrated with Buffer A. All protein samples were purified using affinity chromatography with a salt gradient (150-600mM KCl). Samples that were collected during protein purification were analysed on an 8% SDS-PAGE gel. The purification of heparin mutant number 7 is shown in Figure 3.12.

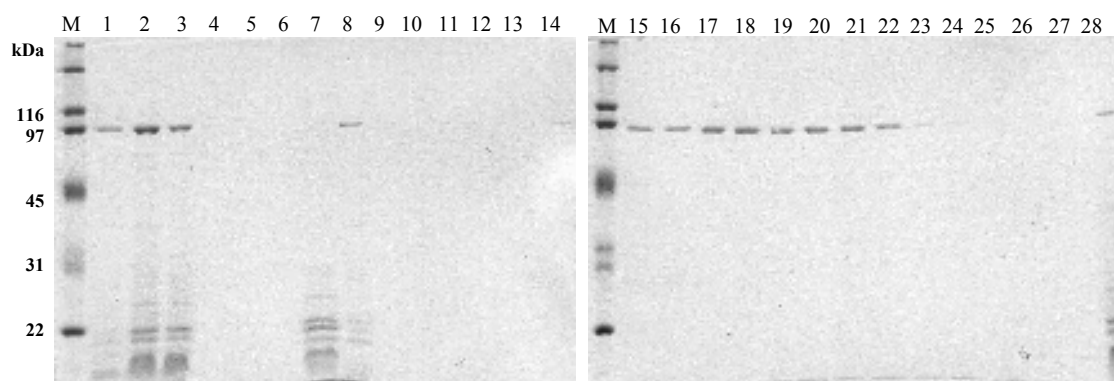


Figure 3.12 – SDS-PAGE gel of the protein purification of first round heparin mutant 7. (M) marker, (1) after $(\text{NH}_4)_2\text{SO}_4$ /PEI treatment, (2) unfiltered sample, (3) sample loaded on column, (5) buffer exchange flow through (FT) 1, (6) buffer exchange FT 2, (7) purification FT 1, (8) purification FT 2, (9)-(27) fractions collected sequentially over increasing salt gradient, (28) sample loaded on column.

Analysis of protein fraction showed that the protein eluted off the Heparin Sepharose column over a wide range of salt concentrations but yielded very pure protein at the expected molecular weight of 99kDa. Mutant Heparin 7 eluted at approximately 270-315mM KCl, which can be seen in Figure 3.12 lanes 14-23. The wild- type Sac7-TaqPolymerase and all the mutants that were purified eluted at around 300mM KCl. The peak fraction (containing the most protein) from each purification was run on an 8% SDS-

PAGE gel with whole cell extract and the sample that was loaded onto the column for purification (Figure 3.13).

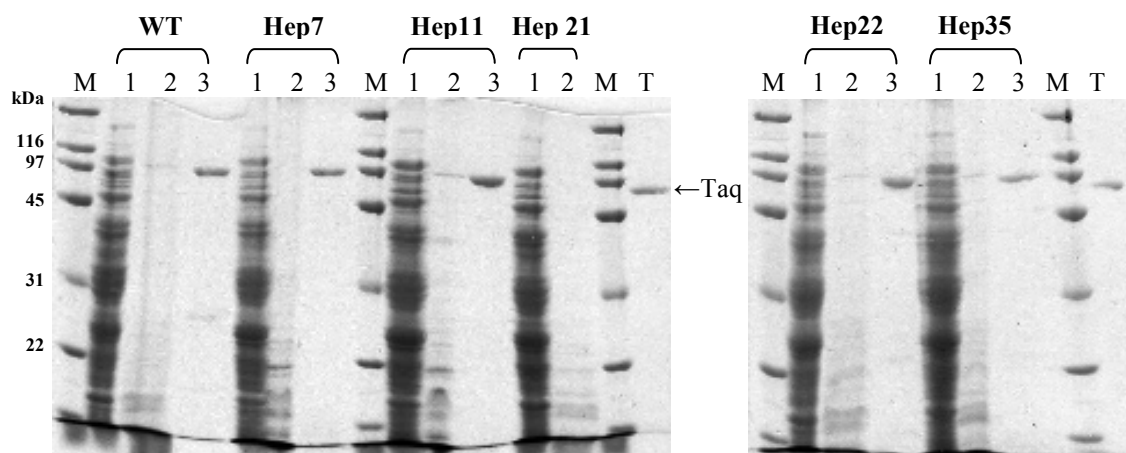


Figure 3.13 – SDS-PAGE gel of the purification of all polymerase variants and wt. Protein samples from key steps in the purification and the final peak fraction of purified protein were analysed by 8% SDS-PAGE. Wt Sac7-TaqPol and each variant is labelled above the gel. The samples from each purification are: (1) whole cell extract, (2) loaded sample, (3) peak fraction, except in the case of Hep21 where the pure fraction could not be visualised by SDS-PAGE. (M) marker. (T) Pure Taq (no Sac7).

The protein samples that were loaded onto the Heparin Sepharose column did not have a very high concentration of protein as seen in Figure 3.13. However, after purification, the eluted fractions showed concentrated protein of the correct size which was estimated to be over 90% pure. The protein-containing fractions from each polymerase variant were pooled together, concentrated and dialysed. Hep21 had an extremely low yield and could not be visualised by SDS-PAGE gel.

Characterisation of Heparin Resistant Variants

Following protein purification, the protein concentration of the polymerase variants was estimated using a Bradford Assay (Table 3.3). The first-round heparin variants were then characterised using a standard polymerase assay and the specific activity of the wild-type Sac7-TaqPol and the selected heparin mutants was determined (Table 3.3). No activity was detectable with Heparin mutant 21.

Table 3.3 – Calculated activity, protein concentration and specific activity of wild-type Sac7-TaqPol and first-round heparin variants.

	Activity (U/ml)	Protein Concentration (mg/ml)	Specific Activity (U/mg)
Wild-Type	1310	0.066	19,850±580
Hep7	2270	0.212	10,700±1250
Hep11	7100	0.394	18,010±270
Hep22	7450	0.502	14,870±620
Hep35	2670	0.450	5,940±1890

Further characterisation was performed and the polymerase activity of the variants was compared to that of wild-type Sac7-TaqPol in the presence of the inhibitor heparin. Initially, different dilutions of the enzymes were assayed in order to determine the optimal concentration for the fluorescent signal to be in range and to ensure that doubling the amount of enzyme in the assay gave double the activity. Once the optimal enzyme concentration was established, each variant polymerase was assayed in the presence of heparin. The percent activity with heparin was then calculated based on the 100% activity control¹⁰. Preliminary data were collected for each mutant at 0.01U/ml heparin (Table 3.4).

Name	% activity at 0.01U/ml heparin
Wild-type	50%
Hep7	89%
Hep11	74%
Hep22	90%
Hep35	100%

The initial assays revealed that all the polymerase variants that were tested showed increased resistance to heparin compared to wild-type Sac7-TaqPol. Mutants Hep7, Hep22 and Hep35 all showed very little inhibition at 0.01U/ml heparin, which was sufficient to inhibit wild-type enzyme by 50%. Due to its specific activity and the results of the earlier

¹⁰ 100% activity refers to each enzyme in the assay without the presence of inhibitor (zero heparin control).

PCR experiments (Figure 3.10), mutant Heparin22 was selected for more extensive characterisation.

A full characterisation of Heparin22 using the polymerase assay over a range of heparin concentrations was carried out alongside the wild-type. Graphical analysis of the assay data was used to determine the level of heparin resistance of each polymerase (Figure 3.14).

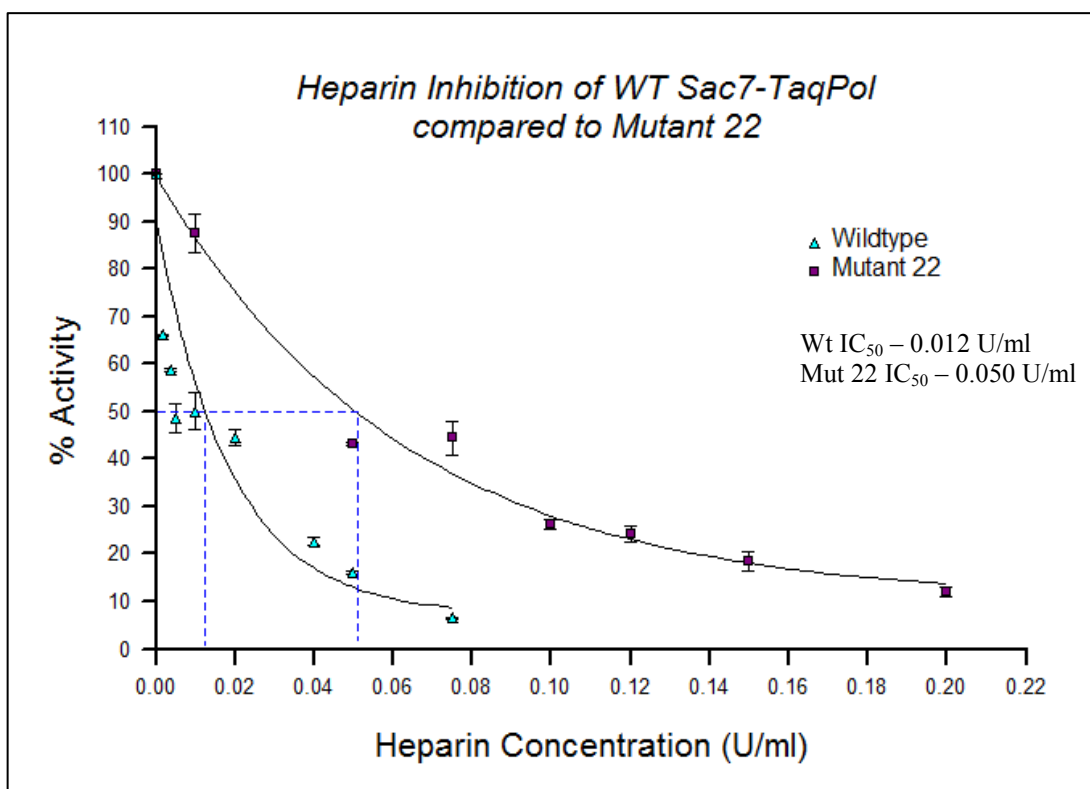


Figure 3.14 – A graph showing the polymerase activity of wild-type Sac7-TaqPol compared to Heparin Mutant 22 in the presence of heparin.

Graphical analysis of the polymerase assay data over a range of heparin concentrations shows that there is a significant difference in the sensitivity to heparin of the wild-type compared to mutant 22. The IC_{50} (inhibitory concentration at 50% activity) for Mutant 22 was 0.050 U/ml, which is approximately four times higher than that of the wild-type IC_{50} of 0.012 U/ml, indicating a substantial increase in resistance to heparin.

As well as using PicoGreen to analyse activity, the polymerase assays were run on an agarose gel to see if the change in ssDNA to dsDNA could be visualised, thus confirming DNA extension and polymerase activity. An assay using the Heparin 22 variant over a range of heparin concentrations was run as standard. Instead of mixing the assay with PicoGreen, it was loaded straight onto a 0.7% (w/v) agarose gel and the DNA was visualised. Figure

3.15 shows the expected change in DNA conformation during the polymerase assay. An assay with commercially available *Taq* DNA polymerase (GeneSys) was also run on the gel; the assay was run as standard for 10 min with various concentrations of *Taq*.

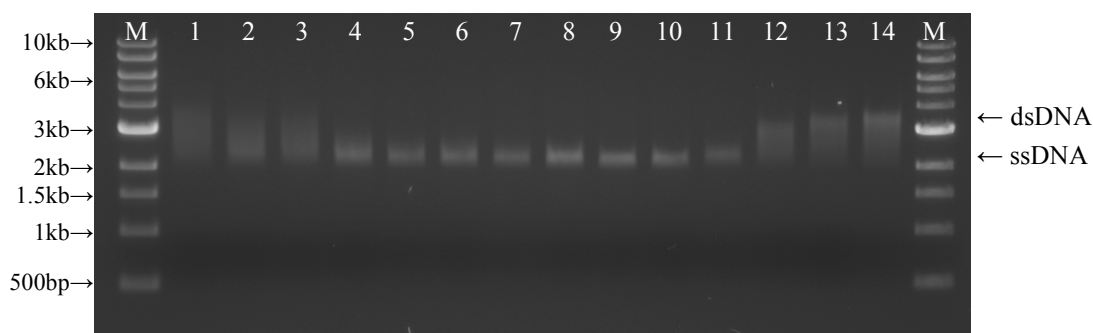


Figure 3.15 – Agarose gel of polymerase assay.

10 min polymerase assays with Heparin 22 over a range of heparin concentrations and purified *Taq* over a range of enzyme concentrations. (M) 1kb marker, Hep22 with heparin at the following concentrations (U/ml): (1) 0, (2) 0.005, (3) 0.01, (4) 0.05, (5) 0.075, (6) 0.1, (7) 0.15, (8) 0.2. (9) Hep22 time 0 control. (10) *Taq* time 0 control, (11) 0.125U *Taq*, (12) 0.25U *Taq*, (13) 0.35U *Taq*, (14) 0.4U *Taq*.

Analysis of the polymerase assay on an agarose gel showed that there was a shift in migration in the DNA as it went from single stranded to double stranded. In the *Taq* DNA polymerase assay the effect of a higher concentration of enzyme can be seen; as more enzyme is added, more double stranded DNA is produced over the 10min time period. In the Heparin 22 assay, the affect of heparin can be seen; as the heparin concentration increases the amount of dsDNA produced decreases. At 0.2U/ml of heparin there is almost complete polymerase inhibition. There is only a slight shift in the band in lane 9 (0.2U/ml heparin) compared to the time zero control (M13 ssDNA) in lane 10. These data correspond to that found using fluorescent readings of the assay, where only approximately 10% activity was seen at 0.2U/ml heparin, and presents a powerful verification of the fluorescence-based assay system used.

Further characterisation was done by analysing the amino acid changes in each of the mutants using the sequencing data previously obtained. There was a low mutation frequency in all the mutants that came through the first-round heparin screen; most only had 1 or 2 amino acid changes (Table 3.2). The exception was Heparin21, which showed activity in initial PCR tests but had a low yield in purification, showed no detectable activity, and had a total of 10 amino acid mutations. One of the mutations in Hep21 (K540M) was very interesting as it had also been mutated in the Ghadessy heparin resistant mutant and is found in the DNA binding pocket. In general, when compared to Heparin resistant mutants

that were evolved by Ghadessy (2001) over three rounds of CSR, the first round mutants in this study had fewer amino acid changes and no identical mutations were seen amongst the mutants.

The mutated residues were modelled onto the published structure of *Taq* DNA polymerase (Li *et al.*, 1998: PDB ID-2KTQ), using the Cn3D program, to see if any of them came into contact with the DNA or potentially influenced heparin/DNA binding in the DNA binding pocket. The mutated residues are highlighted in Figure 3.16. To follow up the extensive polymerase assays carried out with Heparin 22, a closer look at one of the amino acid changes in this mutant was performed. In order to do this, the *Taq* DNA polymerase structure (Li *et al.*, 1998: PDB ID-2KTQ) was visualised using the Cn3D program. Heparin mutant 22 had a residue change (T588S) that was very close to the DNA binding pocket. Although T588 is not predicted to bond directly with the DNA template based upon X-ray crystallographic structures, it is possible that Heparin has a slightly different binding site compared to DNA. Its location with respect to the location of bound DNA can be seen in Figure 3.17.



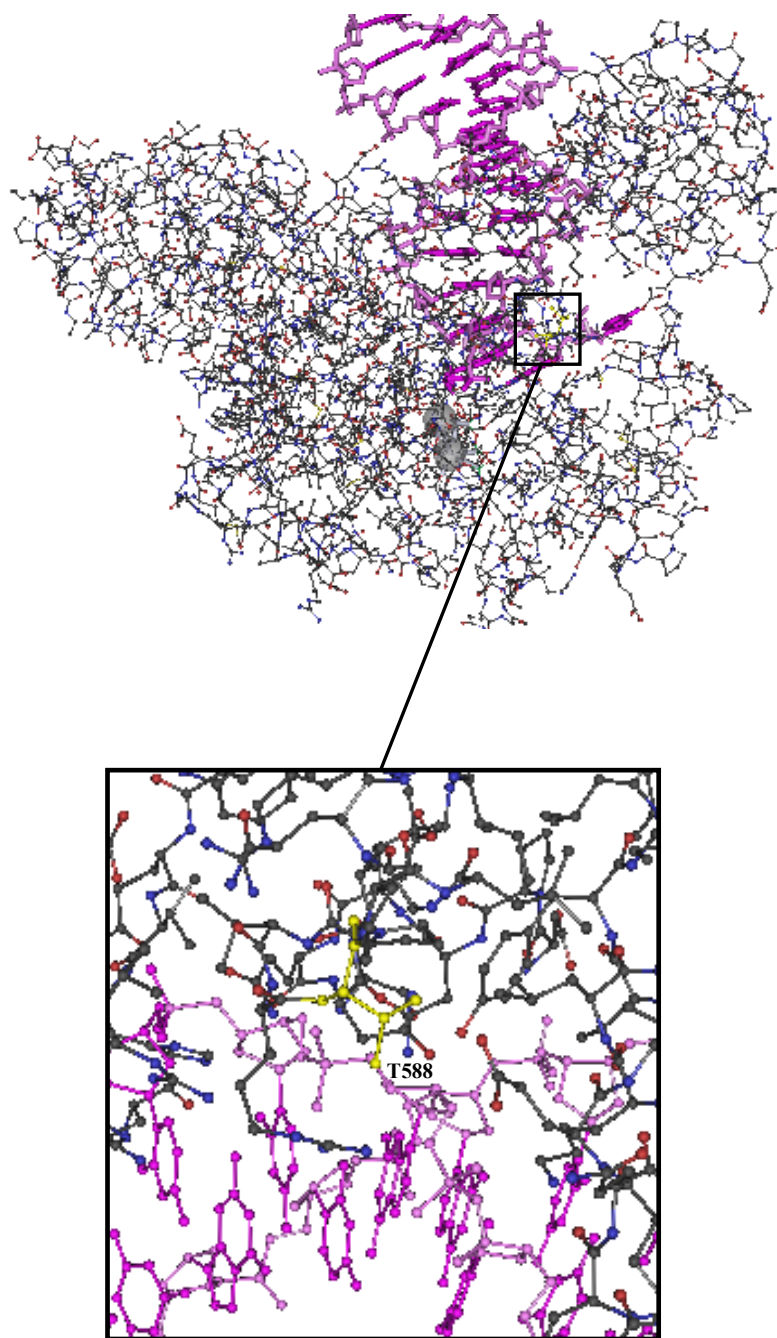


Figure 3.17 – Taq Polymerase structure showing the location of T588S (one of the mutated residues of Heparin 22) (a) The whole *Taq* Polymerase structure with the mutated residue highlighted in yellow and the DNA shown in pink. The image shows that the residue is located in the DNA binding pocket. (b) A close up of the residue highlights its close proximity to the DNA strand.

Discussion

The data presented in this chapter demonstrate that a system of directed evolution by compartmentalised self-replication has been successfully established and employed to produce heparin resistant DNA polymerases. The CSR system, set up in Chapter 2, was used to screen a mutant library to demonstrate that it was capable of selecting active heparin resistant variants from a pool of mainly inactive polymerases. The process was verified by the characterisation of selected *Taq* DNA pol variants in the presence of heparin to reveal the extent of resistance achieved by just a single round of directed evolution.

The GeneMorph[®] II Random Mutagenesis Kit (Stratagene) was used to create *Taq* DNA polymerase mutants by error-prone PCR. A low mutation frequency was decided on based on the fact that most directed evolution experiments are carried out to produce 1-4 amino acid changes per polypeptide (You and Arnold, 1994 and Vichier-Guerre *et al.*, 2006), although libraries have been screened containing mutants with as many as 20 mutations per gene (Daugherty *et al.*, 1999). In general, a low error rate is expected to accumulate mostly adaptive mutations, as opposed to a high error rate which will generate deleterious mutations (Arnold *et al.*, 2001). A final mutant library was produced containing approximately 1×10^6 mutants with an average of 6 amino acid changes (8.2 nucleotide changes) per polymerase molecule. Analysis of the final library showed that almost 100% of the mutant library clones were expressing polymerase but that less than 20% of them were active. These final figures are comparable to the work carried out by Ghadessy (2001) where a 10^7 mutant clone library had an average of 4-5 amino acid mutations per polymerase variant and only 1-5% of the resulting mutants were said to be active.

In directed evolution, the size of the library to be screened is ultimately limited by the screen itself. In high-throughput *in vitro* screens such as CSR and phage display, libraries of 10^7 or 10^8 can be screened successfully (Holmberg *et al.*, 2005). Genetic complementation screening or screens that require 96-well plate activity assays are limited to library sizes of 10^4 or 10^5 (Henry and Romesberg, 2005). In this project, optimisation of ligation and transformation reactions was performed to increase efficiency and aid in library construction. Through numerous transformations a library size of 10^6 was achieved. The use of the CSR high-throughput screen allowed for the relatively easy screening of such a large library.

Before screening of the library could occur, the emPCR was optimised to account for the low level of active polymerases that would be present when the CSR was run with the mutant library and in the presence of heparin. Despite extensive optimisation, a secondary amplification of the emPCR was needed to produce enough DNA for cloning.

Another key step in the screening of the mutant library was to determine the level of heparin to be used in CSR. It was necessary to use a level of heparin that did not completely inhibit the reaction but still allowed some amplification to occur otherwise nothing would come through the screen. *E. coli* cells containing wild-type Sac7-TaqPol were used in an emPCR with various concentrations of heparin to determine the level of heparin inhibition. It was important to run this test in the emPCR with 2×10^8 cells, exactly as in a CSR screen, because different conditions and levels of enzyme may affect the heparin sensitivity. It was concluded that wild-type enzyme was slightly inhibited by 1.25U/ml of heparin and completely inhibited by 2.5U/ml of heparin under emPCR conditions. The first-round CSR screen using the mutant library was run in the presence of 1.25U/ml of heparin. If further rounds of directed evolution and CSR were to be undertaken then the level of heparin used in the screen would increase each round.

The polymerases that were able to replicate their polymerase domain in the presence of 1.25U/ml of heparin came through the first round of screening. Standard activity testing of the first-round heparin mutants showed that the percent active clones had increased to 80% compared to less than 20% in the original mutant library. This was a good indication that the CSR screen was able to pull out the active polymerases from the pool of largely inactive polymerases that made up the mutant library.

The next step in the verification of the CSR screen came from the characterisation of the first-round mutants. Forty mutants were tested in a set of PCRs over a range of heparin concentrations. The mutants' sensitivity to heparin was compared to that of wild-type Sac7-TaqPol and the mutants with the highest heparin resistance were chosen for further characterisation. After scale-up, expression and purification using ammonium sulphate precipitation, heat precipitation and PEI precipitation followed by Heparin Sepharose affinity chromatography, preliminary polymerase assays were used to investigate the activity of each mutant in the presence of 0.01U/ml of heparin. All the mutants tested out-performed the wild-type enzyme at 0.01U/ml of heparin, indicating that the CSR screen had been successful.

Heparin mutant 22 was then taken forward for further study with assays over a range of heparin concentrations to determine the percent activity compared to the wild-type. Graphical analysis of the data revealed that Heparin22 had a 4-fold increase in resistance to heparin compared to the wild-type. This is a substantial change to observe in a single round of directed evolution. In the similar experiment performed by Ghadessy (2001), it was said that an 8-fold increase in resistance had been achieved after one round of CSR, a comparable change (Patent application WO 2002/022869). Had further rounds of CSR been carried out, then the 4-fold increase in resistance of the first round mutants would have been built on and even more significant changes could have been achieved. Ghadessy observed further

changes of 4-fold in round two and a further 4-fold in round three, leading to a final mutant with 130-fold increased resistance to Heparin.

When the specific mutations found in the first-round heparin mutants were studied further, a large proportion of them were seen to cluster around the DNA binding pocket (Figure 3.16), as was also seen with the Ghadessy heparin resistant mutants (Ghadessy *et al.*, 2001). In particular, the position of the amino acid change (T588S) in the most well characterized mutant, Heparin22, was of interest as it was in the DNA binding pocket of *Taq*DNA pol and is therefore likely to have some influence on Heparin binding in the DNA binding site. Whilst the substitution of Threonine 588 with Serine is a conservative substitution, it is envisaged that this alteration in such close vicinity of the predicted heparin binding site could have a significant effect on the relative binding of heparin and DNA. Although residue T588 has not been shown to have direct interactions with the DNA, its close proximity to other residues that do interact with the DNA may have led to a small change, which caused a regional affect on DNA/heparin binding. Although there were no identical residue changes in any of the mutants, K540 was mutated in both Heparin21 from this study and Ghadessy15 (Table 3.2); it is also situated in the DNA binding pocket, which suggests it may be an important residue in conferring heparin resistance.

The work presented in this chapter provides clear evidence that the CSR system is successfully functioning as a screen for improved *Taq* DNA pol variants. Starting from a mutant library containing a low proportion of active variants, CSR enabled the screening and selection of a *Taq*DNA pol variant with a 4-fold increase in resistance to heparin. The optimisation and verification of the CSR system described in this chapter provide the groundwork for the technique to be applied to a novel inhibitor of *Taq* DNA pol. The inhibitor selected for this work is phytate and the next Chapter describes the attempts to evolve phytate resistant mutants of *Taq* DNA pol by CSR.

Chapter Four – Engineering *Taq* DNA polymerase resistance to Phytate by CSR

Introduction

Fecal-based PCR assays for the detection of bacterial infection and disease are commonly used on veterinary samples (Rajeev *et al.*, 2005) and are becoming more widespread in human clinical analysis (Al Amri *et al.*, 2007). Traditional methods of detection require culture-based techniques, which are labour-intensive and far less rapid than PCR-based detection. Although PCR-based diagnostics have obvious advantages, there is one key difficulty in fecal-based PCR that stems from the problematic and time-consuming removal of PCR inhibitors with the use of phytase or sample processing kits.

Phytic acid, also known as inositol hexakisphosphate (IP₆), or phytate in its salt form, is the main PCR inhibitor found in fecal samples. It is the main store of phosphorus in many plant tissues, especially seeds, and is thus highly abundant in soil and animal manure (Graf, 1983). Phytic acid makes up a key component of animal feed and is also highly abundant in human diets and human cells (Raboy, 2007); it therefore causes a problem in both veterinary and medical fecal-based diagnostics.

Phytic acid is a small molecule that contains a 6-C ring with one phosphate group attached to each carbon atom. Each of the 6 phosphate groups has 2 replaceable hydrogens, which leads to interactions with various metal ions in the environment and the formation of phytate salts (He *et al.*, 2006), which are typically found to form amorphous powders with non-stoichiometric atomic ratios (Evans and Pierce, 1982). An example of a phytate sodium salt can be seen in Figure 4.1 along with the most well accepted structure of phytic acid, which was proposed by Anderson in 1914. As a result of its structure and the ionisable nature of its six phosphate groups, phytic acid is known to be a highly potent chelator of divalent and trivalent metal ions, including Cu²⁺, Zn²⁺, Ni²⁺, Co²⁺, Mn²⁺, Ca²⁺, Mg²⁺, Fe²⁺ and Fe³⁺ (Cheryan, 1980; Maenz *et al.*, 1999). The binding affinity of phytate with the different ions and chemical properties of the resulting phytate-metal chelates vary significantly and are influenced by a number of factors including pH (Cheryan *et al.*, 1983; Crea *et al.*, 2006). Various researchers have studied the formation of phytate complexes. For example, Luttrell (1993) reported a binding affinity of phytate for Ca²⁺ under physiological conditions of 1.9 x 10⁵ liters/mol. In a comprehensive study Crea *et al.* (2006) compared the binding characteristics of phytate- Ca²⁺ and phytate-Mg²⁺ complexes and found them to have broadly similar binding parameters. In each case they observed 10 different stable species to form.

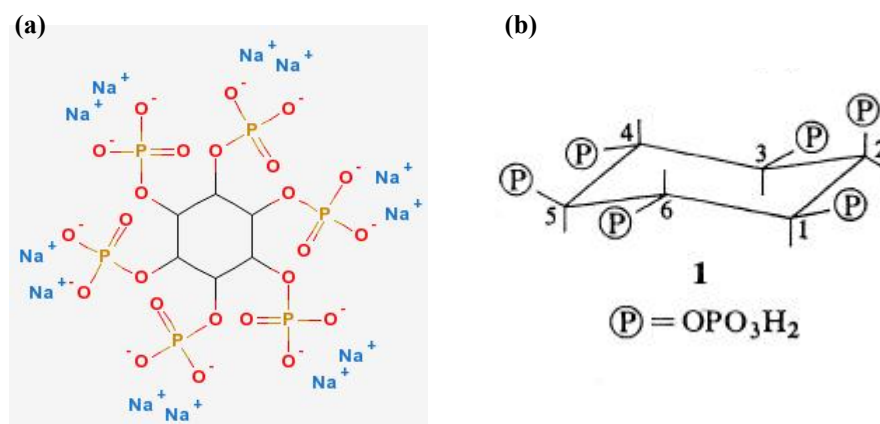


Figure 4.1 – The structure of phytic acid. (a) Dodecasodium phytate salt (Pub Chem:CID 66391). (b) One of the first published structures of phytic acid by Anderson, 1914 (Johnson and Tate, 1969)

In nature, as well as its role as a phosphate store in plants, phytate (IP_6) is a regulator of many cellular functions in eukaryotes (Shears, 2001). For example, IP_6 plays a role in regulating endocytosis by binding to the α subunit of the tetrameric clathrin assembly protein AP-2 complex (Beck and Keen, 1991) and has also been observed to bind other proteins involved in protein trafficking (Shears, 2001). Furthermore, it is an activator of enzymes conducting DNA repair, is a cellular antioxidant and has been reported to have anticancer activity (Vucenik *et al.*, 2004). It is thought that phytate may have therapeutic effects because of its ability to inhibit proteins through metal ion chelation and its antioxidant properties. Among other properties, phytate's potent chelation of Fe^{3+} is thought to decrease iron absorption in humans and protect against free radical formation (Hawkins *et al.*, 1993 and Reddy *et al.*, 1996).

The conformation of IP_6 has been suggested to mimic the pyranose ring structure of heparin and therefore to cause inhibition of heparin-binding proteins through competitive inhibition (Vucenik *et al.*, 2004). Given that heparin is known to bind to polymerases and interfere with template/primer interaction (DiCioccio and Srivastava, 1978), it is reasonable to hypothesise that phytate may also inhibit in this way, particularly as phytate is known to bind a range of other proteins as mentioned above.

Although phytate has many important cellular functions, its presence in the human diet and its potential negative effects have been open to debate. Phytate has been shown to have an inhibitory effect on the iron absorption or bioavailability of key minerals in the body due to its chelating potential (Layrisse *et al.*, 2000, Glahn and Wortley, 2002). Phytate has

been said to exacerbate mineral deficiency in developing countries where diets are cereal and legume based; this has led to the development of low-phytate crops as some scientists believe that a reduction in phytate content of grains could be beneficial (Raboy, 2007). However, others argue that in most diets phytate does not cause a significant problem and its health benefits out-weigh any negative effects (Shamsuddin, 2008).

Although much is known about phytate and its natural role *in vivo*, its mode of action as a PCR inhibitor is still not clear. The affect of phytate on PCR amplification of fecal samples has been studied by Thornton and Passen (2004). It was found that at levels above 0.3mM phytate, PCR was completely inhibited and at lower levels phytate could cause non-specific amplification in the PCR. Although processing samples with phytase (50U/ml) prior to PCR was found to reduce inhibition, this did not represent a very economic solution to the problem. Thornton and Passen proposed two ways in which phytate may inhibit PCR (1) removing Mg^{2+} by chelation and (2) changing the ionic composition of the PCR through the addition of potentially inhibitory counter-ions from the phytate salts. Based on their investigations with different counter-ions, they concluded that phytate inhibition was produced by chelation, which means that phytate, not its counter-ions, causes inhibition (Thornton and Passen, 2004), although no direct evidence was presented. As other PCR inhibitors such as heparin (DiCioccio and Srivastava, 1978) and components of soil (Watson and Blackwell, 2000) are known to inhibit polymerase by binding and competing with the DNA template, it is a possibility that phytate may also inhibit in this way.

One thing that is certain is that phytate interferes with amplification and causes a problem in PCR-based diagnostics. This chapter will present the results of attempts to evolve resistance to PCR inhibition by phytate. Using the technique of directed evolution by compartmentalised self-replication, which has previously been used to evolve polymerase resistance to heparin, it was hoped that phytate resistant variants may be evolved to combat the problem of phytate inhibition in PCR.

Following on from the work in Chapter 3, a mutant library was screened by CSR in the presence of the novel PCR inhibitor phytate, the aim being that polymerase variants that are able to replicate their own gene under the selective pressure of phytate inhibition come through the screen. A total of two rounds of CSR were carried out with the use of a recombination technique called staggered extension process (StEP), described in Chapter 1, employed to rediversify the library in between rounds of screening. After two rounds of CSR, the phytate resistant *Taq* DNA pol variants were then characterised and assayed to determine the level of phytate resistance compared to the wild-type Sac7-TaqPol.

Material and Methods

Emulsion PCR

emPCRs were performed as described in Chapter 2, except for the following changes to the final concentrations of some of the PCR reagents: 0.5mM dNTPs, 1 μ M of outer pol domain forward and reverse primers (Chapter 3), and 5% (v/v) DMSO. emPCRs were extracted as standard (Chapter 2).

General Methods

DpnI treatment and amplification of emPCR products were as standard (Chapter 3). Ligation reactions were performed as standard and butanol precipitated (Chapter 3). Screening for pol domain insert was conducted as standard (Chapter 3). Standard expression and activity tests were carried out as described in Chapter 2, but with the following changes: 30ml of Overnight Express auto-induction medium (Novagen, Nottingham, UK) containing kan (50 μ g/ml) was inoculated and left shaking at 37°C for 24h. 10ml pellets were harvested by centrifugation and a cell lysate was prepared as standard. Glycerol stocks of the 2nd mutant library were made as in Chapter 3. Purification and dialysis was carried out as in Chapter 3.

Staggered Extension Process (StEP)

StEP was carried out using first-round round phytate mutants as template. 50ng of each of 9 mutants were combined to give 450ng total template. The StEP reaction was set up with the following final concentrations: 1x Taq Buffer (GeneSys, Camberley, UK) containing 1.5mM MgCl₂, 200 μ M dNTPs (GeneSys, Camberley, UK), 25pmol of Pol domain forward and reverse primers (Chapter 3), 500ng template, 2.5U *Taq* DNA polymerase (GeneSys, Camberley, UK) and ddH₂O to 100 μ l. The reaction was thermocycled as follows: 94°C for 5min, and 80 cycles of 94°C for 30s and 60°C 5s. 5 μ l of the reaction was removed for agarose gel electrophoresis analysis after 20, 40, 60 and 80 cycles of the StEP.

DpnI Digestion and Secondary Amplification of StEP

PCR products from the StEP reaction were treated with 10U of *DpnI* for 1 h at 37°C and then used as template in a secondary PCR. A Phusion PCR was set up with the following final concentrations: 1x Phusion HF Buffer (New England Biolabs, Hitchin, UK), 200µM dNTPs (GeneSys, Camberley, UK), 25pmol of original pol forward and reverse primer (Chapter 3), 1-5µl StEP PCR product template, 5% (v/v) DMSO (New England Biolabs, Hitchin, UK), 1U Phusion (New England Biolabs, Hitchin, UK) and ddH₂O up to 50µl. Reactions were thermocycled as follows: 98°C for 30s, followed by 25 cycles of 98°C for 30s, 60°C for 15s, and 72°C for 45s. 5µl of the PCR was analysed on a 1% (w/v) agarose gel.

Secondary Amplification of 2nd Round CSR emPCR Product

The emPCR product was *DpnI* treated as standard. New polymerase domain primers were designed: Forward 5' GCCCCTGGAGGTGGACTTC and Reverse 5' GGATGTGCTGCAAGGCG ATTA. A PCR using the 2nd round emPCR as template was set up with final concentrations as follows: 1x Phusion HF Buffer (New England Biolabs, Hitchin, UK), 400µM dNTPs (GeneSys, Camberley, UK), 25pmol of forward and reverse primers (GeneSys, Camberley, UK), 1µl template, 5% (v/v) DMSO (New England Biolabs, Hitchin, UK), 1U Phusion (New England Biolabs, Hitchin, UK), 1.5mM MgCl₂ and ddH₂O up to 50µl. Reactions were run following the recommended protocol for 30 cycles.

Polymerase Assay

M13 oligonucleotide was pre-annealed to M13mp18ssDNA template (New England Biolabs, Hitchin, UK) as described in Chapter 3. Annealed oligo/template was cooled to 4°C before other reaction components were added. Final reaction concentrations were as described in Chapter 3. 5µl of enzyme was added to the 45µl of final reaction mixture. After mixing, the reaction was divided into 3 aliquots of 15µl. One aliquot was kept on ice and used as a control/time zero, the other two replicates were heated at 72°C for 10min and then placed immediately back on ice. 35µl of TE [10mM Tris (pH7.5), 1mM EDTA] was added to each aliquot and they were transferred to a black flat-bottomed 96-well plate (Thermo Scientific, Basingstoke, UK). 50µl of a 1:400 dilution of Quant-iT PicoGreen[®] dsDNA reagent (Invitrogen/Molecular Probes, Paisley, UK) in TE was then added to each well and

mixed with the reaction. The plate was incubated at room temperature for approximately 5min and the amount of DNA synthesized was quantified and analysed as in Chapter 3.

Thermal Inactivation Assay

20µl aliquots of enzyme in 1x Taq Buffer (GeneSys, Camberley, UK) were heated on a PCR block at 90°C for set time intervals (0 to 90min) and quenched by immediate transfer to ice. 10µl of each aliquot was added to a standard polymerase assay (as above) to a final volume of 50µl. The reaction was split into 3 aliquots of 15µl and polymerase assays were run as standard for 10min at 72°C. Reactions were treated and quantified as in the polymerase assay (as above).

Fidelity Assay

pFidelity plasmid (based on pUCIQ17 from Frey and Suppmann, 1995) pre-linearised by digestion with *Dra*II was supplied by GeneSys (Camberley, UK). A fidelity PCR was set up with the following final concentrations: 1x Taq Buffer (GeneSys, Camberley, UK) with 1.5mM MgCl₂, 200µM dNTPs, 25pmol forward and reverse *Cla* primers (*Cla* Forward 5' AGCTTATCG ATGGCACTTTTCG GGGAA ATGTGCG, *Cla* Reverse 5' AGCTTATCGATAAGCG ATGCCGGGAGCAGACAAGC), 10ng linearised pUCIQ17, 1µl of mutant or wild-type Sac7-TaqPol enzyme and ddH₂O up to 50µl. Reactions were thermocycled as follows: 94°C for 3min, followed by 25 cycles of 94°C for 10s, 55°C for 30s and 72°C for 5min. 5µl of the fidelity PCR was analysed on a 1% (w/v) agarose gel. After pooling multiple PCRs, an aliquot was run on a 1% (w/v) agarose gel next to a 3.6kb PCR product Marker (GeneSys, Camberley, UK) to estimate DNA concentration. Fidelity PCR products were digested with *Cla*I (New England Biolabs, Hitchin, UK) as recommended and then gel purified using the NucleoSpin Extract II kit (Machery-Nagel). Purified products were checked on a 1% (w/v) agarose gel before approximately 100ng of the *Cla*I digested product was ligated (to form circular plasmid) with T4 ligase (Promega, Southampton, UK) as recommended. Ligations were ethanol precipitated, resuspended in 5µl, and 1µl was transformed into NEB 5α electrocompetent cells (New England Biolabs, Hitchin, UK) as described in Chapter 2. Transformations were spread onto Met:Amp:Xgal (40ug/ml) LB agar plates and incubated overnight at 37°C before blue and white colonies were counted. The error rate was calculated using the following formulae (Frey and Suppmann, 1995):

- Duplication = $\log_{10}(\text{yield}/\text{template})/\log_{10}(2)$
- Fraction white = white colonies/total
- Calculated error rate = $-\ln(\text{fraction white})/(\text{duplications} \times 349^{11})$

¹¹ LacI gene is 1080bp so there are 349bp phenotypically identified (by blue-white screening) single-base substitutions possible within the LacI gene.

Results

CSR Phytate Screen

An initial test to gauge polymerase sensitivity to phytate was performed using a standard 500bp λ PCR with *Taq* DNA polymerase (GeneSys) and purified Sac7-TaqPol (from Chapter 3) over a range of phytate concentrations. The sensitivity to phytate in the PCRs was analysed by gel electrophoresis (Figure 4.2).

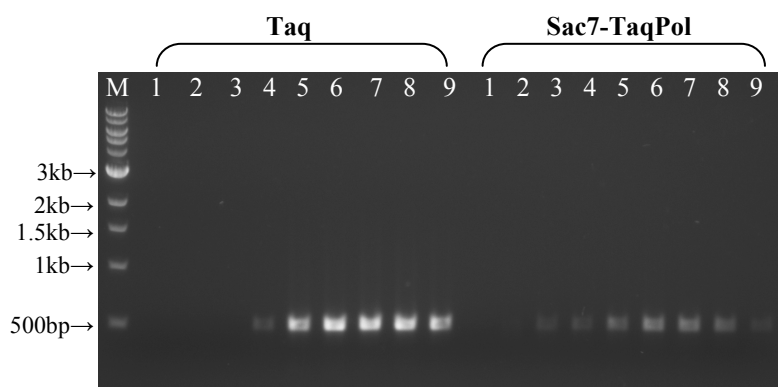


Figure 4.2 – A Phytate dilution series with *Taq* and Sac7-TaqPol. An agarose gel showing the results of a 500bp PCR with *Taq* and Sac7-TaqPol over a range of phytate concentrations. (M) 1kb DNA ladder (NEB), 1 to 9 indicate phytate concentrations: (1) 2.5mM, (2) 1mM, (3) 500μM, (4) 250μM, (5) 100μM, (6) 50μM, (7) 10μM, (8) 5μM, (9) zero phytate control.

Analysis revealed that in this PCR, *Taq* DNA pol appeared significantly inhibited by 250μM phytate and completely inhibited by 500μM. Although the Sac7-TaqPol PCR yield was lower than that of *Taq* DNA pol, it was only slightly inhibited by 500μM whereas no amplification could be seen at this level in the *Taq* DNA pol PCR. The results of this PCR were used as an initial guideline to select appropriate concentrations for testing phytate levels in emPCR. Four levels of phytate, 500, 250, 100 and 50μM, were selected and used in a standard emPCR with *E. coli* cells. After extraction of the emPCR, products were analysed on an agarose gel alongside a zero phytate control (Figure 4.3).

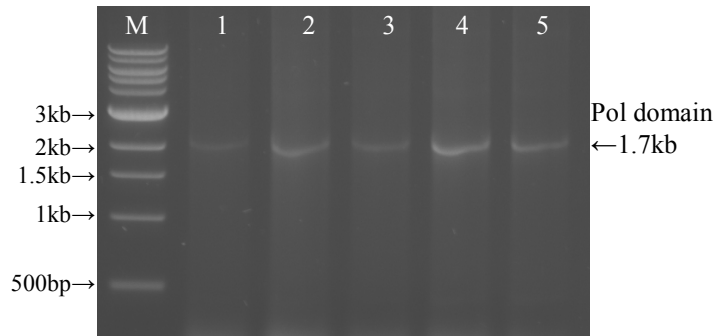


Figure 4.3 – emPCRs over a range of phytate concentrations. (M) 1kb Marker (NEB), (1) 500µM phytate, (2) 250µM phytate, (3) 100µM phytate, (4) 50µM phytate, (5) zero phytate control.

Analysis of the emPCRs over a range of phytate concentrations showed a slight inhibition and a decrease in product yield at 500µM phytate. Although further inhibition would have been seen at higher concentrations, 500µM phytate was chosen as the starting concentration for the CSR screen due to the fact that producing amplification in the emPCR with the mutant library and inhibitor had been difficult during the previous heparin CSR (Chapter 3). It was decided that 500µM phytate was a good level at which to start the CSR screen and still achieve some amplification.

The mutant library that was produced in Chapter 3 and initially screened for heparin resistant *Taq* DNA pol variants was now used to screen for phytate resistant variants. To run the CSR screen, two sets of emPCRs using cells from the mutant library were run in the presence of 500µM phytate. The first contained 1.5mM MgCl₂, which is standard in *Taq* Buffer, and the second contained 2mM MgCl₂, as this was seen to improve the PCR yield in previous tests (Chapter 3). After extraction and *DpnI* treatment, the emPCR was analysed on an agarose gel (Figure 4.4).

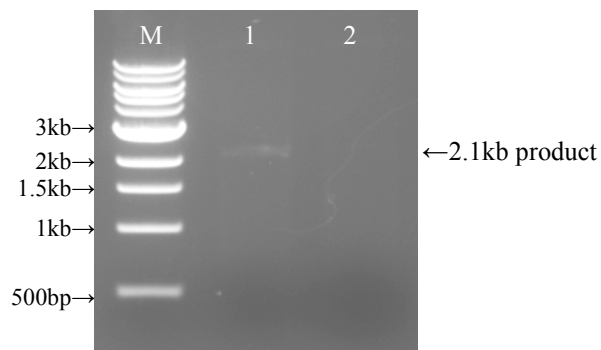


Figure 4.4 – emPCR CSR screen with 500µM phytate. An agarose gel displaying the results of a CSR screen at 500µM phytate with the mutant library. The resulting emPCR, with outer pol domain primers, produced a 2.1kb product. (M) 1kb DNA ladder, (1) 500µM emPCR with 1.5mM MgCl₂, (2) 500µM emPCR with 2mM MgCl₂.

Gel electrophoresis of the emPCR revealed that a small amount of product at 2.1kb was visible in lane 1 (Figure 4.4). This result was unexpected as all previous attempts using the mutant library in an emPCR had produced no visible product (as described in Chapter 3). Although the 1.5mM MgCl₂ emPCR had yielded visible product, the level of DNA was still too low for cloning purposes so a secondary amplification of the emPCR was performed. The resulting pol domain PCR product was processed and digested as standard and cloned into the *SalI* and *BglIII* site of the pTTQ18NHK Sac7-TaqSacB vector (as described in Chapter 2). Transformation into NEB 10-β cells produced the first-round phytate resistant mutants.

Selecting Phytate Resistant Polymerase Variants

The mutant polymerase domains that were amplified during the CSR phytate screen belonged to those polymerases that were the most active under the inhibitory phytate conditions of the screen. Once the pol domain was inserted into the full length Sac7-TaqPol clone, the whole protein was expressed and further selection of the first-round mutants was performed.

A standard expression test using 45 mutants revealed that 60% of the mutants showed strong expression of the correct size protein with the remainder displaying poor or zero expression; some mutants appeared to produce a truncated protein product. A standard activity test revealed similar results, with 50% of the mutants showing polymerase activity, which was an increase on the 20% activity seen in the original mutant library. Further selection was performed by PCR to select mutants to progress to the second round of directed evolution. Active mutants were tested in a PCR in the presence of phytate. The sensitivity of the mutants to phytate was compared to that of the wild-type Sac7-TaqPol. One mutant that performed well in the PCRs was Phytate10, which is shown in Figure 4.5 alongside wild-type.

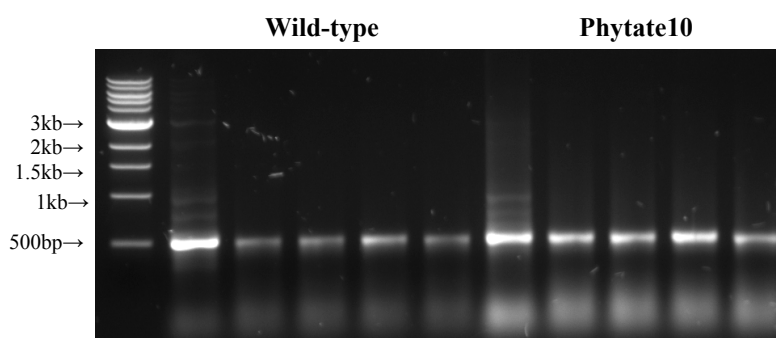


Figure 4.5 – Phytate PCR comparing sensitivity of wild-type to mutant Phytate 10. A 500bp λ activity PCR with wt and Phytate 10 over a range of phytate concentrations. (M) 1kb DNA ladder, (1) zero phytate control, (2) 500μM phytate, (3) 550μM phytate, (4) 600μM phytate, (5) 650μM phytate.

Analysis of the phytate activity PCRs revealed that whilst wild-type was still active over all four of the phytate concentrations tested, it was seen to be inhibited by all levels of phytate compared to the no phytate control. Phytate 10, however, performed strongly at all phytate concentrations and only showed signs of slight inhibition at the highest concentration of 650µM.

Mutants that performed well in the phytate PCRs were selected for sequencing and further study. Sequencing revealed that mutations ranged from 4-14 nucleotides per pol domain with an average of 6.3 changes. This led to an average of 4.8 amino acid changes per mutant. A few of the residues were mutated in more than one mutant, and these same residue changes are highlighted in Table 4.1. In particular, K540 was mutated in two of the selected mutants and had previously been found to be mutated in one of the heparin first-round mutants as well as the Ghadessy heparin resistant mutant (Table 3.2).

Table 4.1 – Residue changes of first-round phytate mutants
(*Taq* DNA Polymerase numbering. Coloured boxes indicate residues mutated in more than one mutant. * = stop codon)

Mutant	Nt changes	AA changes	Residue Changes
Phy1	14	8	W299C, K 314M, L365R, L459F, E473V, L477Q, K508*, K767T
Phy3	5	4	R476L, L733Q, L781P, V810E
Phy5	5	5	L288M, W318*, M444T, E524G, I594V
Phy9	6	6	W299R, F482L, K540E, E734V, K804M, V821A
Phy10	4	3	K340I, M761V, A814T
Phy11	6	4	R425W, S575C, R636W, M658T
Phy26	4	4	P373S, A454V, R492T, S515I,
Phy36	9	8	L324M, D371E, K540M, T557R, A675P, E721G, A748S, L781V
Phy42	4	2	P685S, A791T
Average	6.3 nt	4.8 AA	

Re-diversification and Production of Mutant Library Two

After initial selection of mutants from the first-round phytate CSR screen, the selected clones were rediversified by the staggered extension process (StEP), a PCR shuffling technique. Rediversification of the pol domain was performed so that a new mutant library could be constructed and screened in a second round of CSR.

The 9 mutants (Table 4.1) were combined and used as template in a StEP PCR. Aliquots of the StEP PCR were analysed on an agarose gel to determine if the full-length gene product had been produced (Figure 4.6).

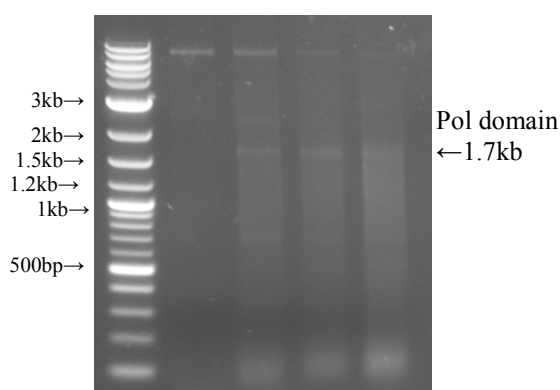


Figure 4.6 – StEP PCR over 80 cycles.

An agarose gel displaying aliquots of the StEP PCR after 20 cycle increments. The production of full-length gene product can be visualised. (M) 2-log ladder (NEB), (1) after 20 cycles, (2) after 40 cycles, (3) after 60 cycles, (4) after 80 cycles.

Analysis of the StEP PCR by gel electrophoresis showed that full-length pol domain product could be visualized after 40 cycles. DNA smears can be seen in lanes 2-4 due to the large amount of short fragments that were being produced during each cycle. The yield from StEP was low as the process requires multiple cycles to fully extend the gene product and therefore growth is not exponential as in standard PCR.

Although the StEP had successfully produced full-length product, the yield from the StEP PCR was low so a secondary amplification of the product was performed after *DpnI* digestion to remove parental plasmid DNA. The recombinant pol domain StEP PCR product was then cloned into the *SalI* and *BglIII* site of the pTTQ18NHK Sac7-TaqSacB vector (as described in Chapter 2). A large ligation and multiple transformations into NEB 10- β cells were performed. The resulting transformant colonies produced mutant library two, just as the original mutant library was produced in Chapter 3. The second mutant library produced from the re-diversification of first-round phytate mutants had a final library size of 5×10^4 .

Mutants from the library were screened for the pol domain insert and it was found that not all the colonies tested contained the correct size insert. Despite multiple gel purifications of the pol domain PCR product following StEP and during the cloning

procedure, it appeared that some fragments of the wrong size had been cloned. This was due to the fact that the StEP PCR would have produced many different recombination fragments which may not all have been a full-length polymerase domain. The process of purification and cloning of the StEP pol domain was repeated but 100% insertion of the correct size pol domain could not be achieved. It was decided that the small percentage of clones that contained the wrong size insert would not cause a significant problem as they would most likely not produce active polymerases and would therefore be lost during the next round of CSR.

Characterisation of the second mutant library using standard expression and activity tests of 15 mutants revealed that 75% of the mutants expressed protein of the correct size and 50% showed polymerase activity. Sequencing data confirmed that the StEP PCR reaction had successfully recombined first-round phytate mutants. All the mutants that were sequenced contained combinations of the amino acid changes that were seen in the phytate mutants selected for StEP (Table 4.1). A small number of new mutations were also introduced possibly because of DNA template switching. There was an average of 4.6 nucleotide changes and 3.4 amino acid changes per mutant in the recombined library (Table 4.2). It was concluded that a second mutant library had been successfully produced by re-diversification of selected phytate mutants from the first-round CSR screen.

Table 4.2 – Mutation frequency of selected clones from Mutant Library Two produced by StEP

Mutant	Nt changes	AA changes
ML2 (1)	6	3
ML2 (2)	4	3
ML2 (3)	3	3
ML2 (4)	4	2
ML2 (5)	6	6
Average	4.6 nt	3.4 AA

2nd Round CSR – Screening and Selection

A second round of CSR was performed to screen the re-diversified mutant library. In this round, the level of phytate was increased from that used in the first round screen. Two concentrations, 600µM and 750µM, of phytate were tested in an emPCR; neither

produced visible amplification when analysed by gel electrophoresis, so a secondary amplification was carried out. The secondary amplification had to be optimised to achieve any visible yield. New primers around the polymerase domain were designed, the dNTP concentration was increased and extra $MgCl_2$ was also added. After these changes were made, the secondary amplification of the emPCR was performed and analysed on an agarose gel (Figure 4.7)

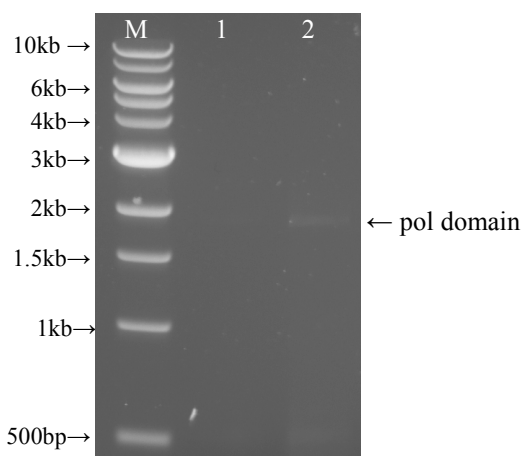


Figure 4.7 – Secondary amplification of emPCR products during the 2nd round of CSR. 600 μ M and 750 μ M emPCR product was used as template in a secondary PCR. (M) 1kb DNA ladder (NEB), (1) 750 μ M PCR, (2) 600 μ M PCR product.

Following optimisation, the secondary amplification produced enough product to be visualised by gel electrophoresis but only when the 600 μ M emPCR screen was used as template. 750 μ M phytate proved too high a concentration to use in the CSR screen. The yield of the 600 μ M PCR was still quite low, so multiple PCRs were run and pooled together to produce enough DNA for cloning. The pol domain was then cloned into the *SalI* and *BglII* site of the pTTQ18NHK Sac7-TaqSacB vector. Following cloning and transformation, the resulting colonies were screened for the pol domain insert. Approximately 1000 transformant colonies were produced and 30 of these were screened for insert; 100% of the selected colonies had the correct size insert. These clones represented the second-round CSR mutants that had come through the 600 μ M phytate screen.

Standard activity and expression tests on the second-round CSR mutants revealed that 80% of the mutants that were tested expressed the correct size protein and showed polymerase activity. Next the mutants' sensitivity to phytate was compared to that of wild-type in a range of standard 500bp λ PCRs (Figure 4.8).

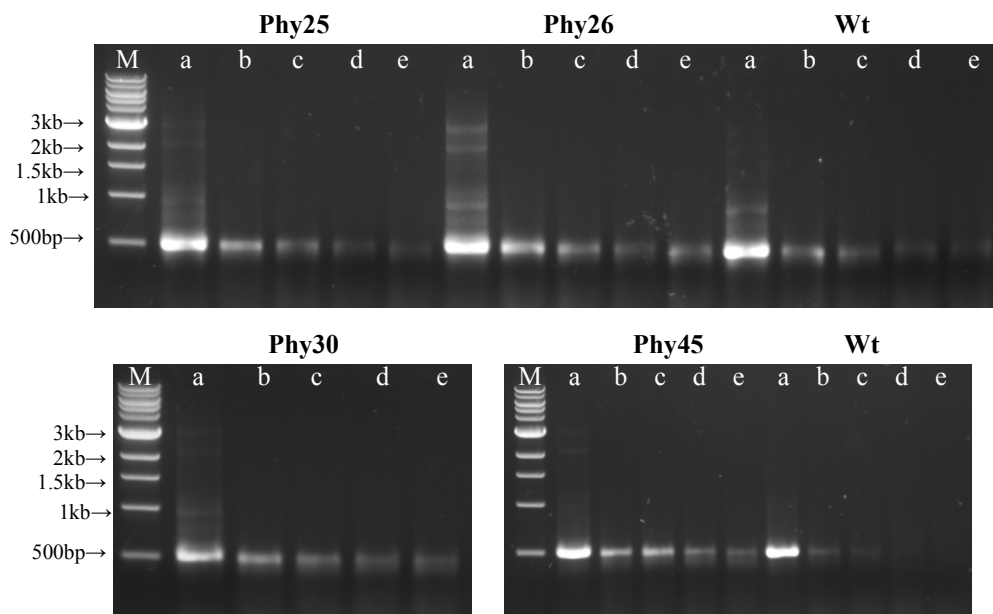


Figure 4.8 – Selected Round Two mutant PCRs with Phytate.

Agarose gels displaying the results of PCRs with round two mutants run over a range of phytate concentrations. (M) 1kb DNA ladder (NEB), (a) zero phytate control, (b) 400 μ M phytate, (c) 600 μ M phytate, (d) 800 μ M phytate, (e) 1000 μ M phytate. The mutant number is indicated above each set of phytate PCRs.

Analysis of phytate PCRs revealed that whilst wild-type was almost completely inhibited at 800-1000 μ M, the selected mutants shown in Figure 4.8 were all able to produce amplification at all the phytate levels tested. These initial PCRs demonstrated that selected second-round mutants had an increased resistance to phytate compared to the wild-type Sac7-TaqPol.

A total of 50 mutants from the second-round CSR screen were tested for activity and increased resistance to phytate. Fifteen of the second-round mutants were selected for sequencing and it was found that there was an average of 5.2 nucleotide and 3.6 amino acid changes per mutant. Two of the mutants were identical to the first-round phytate mutants used in the StEP re-diversification procedure; however, as expected, the majority of the mutants that were sequenced contained combinations of amino acid changes that were found in the first-round mutants that were used to create the second StEP re-diversified library. Following phytate PCRs and sequencing, five mutants (Table 4.3) were selected for further characterisation. Figure 4.9 shows the alignment of the polymerase domain of selected mutants and wild-type Sac7-TaqPol.

Table 4.3 – Residue changes of selected second-round mutants (*Taq* numbering).

Mutant	Nt changes	AA changes	Residue changes
Phy2(2)	5	2	P685S, M761V
Phy2(19)	6	6	R405Q, R469C, S475F, F495Y, T557R, P792Q
Phy2(28)	6	5	K340R, P382T, P685S, M761V, A814T
Phy2(30)	5	3	P685S, M761V, A814T
Phy2(45)	5	4	K340I, E694D, M761V, A814T

The mutated residues from the 5 selected mutants were modelled onto the *Taq* DNA polymerase structure (Li *et al.*, 1998; PDB ID 2KTQ) using Molscript to see if any of them came into contact with the DNA or potentially influenced DNA binding. The mutations did not appear to cluster at any one particular location on the *Taq* DNA polymerase structure (Figure 4.10).

```

Taq          270  RAFLERSEFGSLLHEFGLLESFKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARG
Phy2 (45)    RAFLERSEFGSLLHEFGLLESFKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARG
Phy2 (28)    RAFLERSEFGSLLHEFGLLESFKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARG
Phy2 (30)    RAFLERSEFGSLLHEFGLLESFKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARG
Phy2 (2)     RAFLERSEFGSLLHEFGLLESFKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARG
Phy2 (19)    RAFLERSEFGSLLHEFGLLESFKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARG
*****

Taq          330  GRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEG
Phy (2) 45   GRVHRAPEPYI*ALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEG
Phy2 (28)   GRVHRAPEPYI*ALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLD*TSNTTPEG
Phy2 (30)   GRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEG
Phy2 (2)     GRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEG
Phy2 (19)   GRVHRAPEPYKALRDLKEARGLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEG
*****

                ← Palm                                Palm →
Taq          390  VARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGV
Phy2 (45)    VARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGV
Phy2 (28)    VARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGV
Phy2 (30)    VARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGV
Phy2 (2)     VARRYGGEWTEEAGERAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGV
Phy2 (19)    VARRYGGEWTEEAGEQAALSERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGV
*****

                ← Thumb
Taq          450  RLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKT
Phy2 (45)    RLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKT
Phy2 (28)    RLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKT
Phy2 (30)    RLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKT
Phy2 (2)     RLDVAYLRALSLEVAEEIARLEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKT
Phy2 (19)    RLDVAYLRALSLEVAEEIACLEAEVSRLAGHPFNLSRDQLERLVYDELGLPAIGKTEKT
*****

                Thumb → ← Palm
Taq          510  GKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLIHPRTGRLHTRFNQTAT
Phy2 (45)    GKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLIHPRTGRLHTRFNQTAT
Phy2 (28)    GKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLIHPRTGRLHTRFNQTAT
Phy2 (30)    GKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLIHPRTGRLHTRFNQTAT
Phy2 (2)     GKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLIHPRTGRLHTRFNQTAT
Phy2 (19)    GKRSTSAAVLEALREAHPIVEKILQYRELTKLKSTYIDPLDLIHPRTGRLHTRFNQTAT
*****

                Palm → ← Thumb → ← Palm → ← Fingers
Taq          570  ATGRLSSSDPNLQNI PVRTPLGQRIRAFAIAEGWLLVALDYSQIELRVLAHLSGDNENLI
Phy2 (45)    ATGRLSSSDPNLQNI PVRTPLGQRIRAFAIAEGWLLVALDYSQIELRVLAHLSGDNENLI
Phy2 (28)    ATGRLSSSDPNLQNI PVRTPLGQRIRAFAIAEGWLLVALDYSQIELRVLAHLSGDNENLI
Phy2 (30)    ATGRLSSSDPNLQNI PVRTPLGQRIRAFAIAEGWLLVALDYSQIELRVLAHLSGDNENLI
Phy2 (2)     ATGRLSSSDPNLQNI PVRTPLGQRIRAFAIAEGWLLVALDYSQIELRVLAHLSGDNENLI
Phy2 (19)    ATGRLSSSDPNLQNI PVRTPLGQRIRAFAIAEGWLLVALDYSQIELRVLAHLSGDNENLI
*****

Taq          630  RVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEEA
Phy2 (45)    RVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEEA
Phy2 (28)    RVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAISYEEEA
Phy2 (30)    RVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAISYEEEA
Phy2 (2)     RVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAISYEEEA
Phy2 (19)    RVFQEGRDIHTETASWMFGVPREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEEA
*****

Taq          690  QAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAF
Phy2 (45)    QAFIDRYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAF
Phy2 (28)    QAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAF
Phy2 (30)    QAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAF
Phy2 (2)     QAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAF
Phy2 (19)    QAFIERYFQSFPKVRAWIEKTLEEGRRRGYVETLFGRRRYVPDLEARVKSVREAAERMAF
*****

                Fingers → ← Palm
Taq          750  NMPVQGTAADLVKLAMVKLPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEG
Phy2 (45)    NMPVQGTAADLVKLAMVKLPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEG
Phy2 (28)    NMPVQGTAADLVKLAMVKLPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEG
Phy2 (30)    NMPVQGTAADLVKLAMVKLPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEG
Phy2 (2)     NMPVQGTAADLVKLAMVKLPRLEEMGARMLLQVHDELVLEAPKERAEAVARLAKEVMEG
Phy2 (19)    NMPVQGTAADLVKLAMVKLPRLEEMGARMLLQVHDELVLEAQKERAEAVARLAKEVMEG
*****

                Palm →
Taq          810  VYPLAVPLEVEVGIGEDWLSAKE
Phy2 (45)    VYPLTVPLEVEVGIGEDWLSAKE
Phy2 (28)    VYPLTVPLEVEVGIGEDWLSAKE
Phy2 (30)    VYPLTVPLEVEVGIGEDWLSAKE
Phy2 (2)     VYPLAVPLEVEVGIGEDWLSAKE
Phy2 (19)    VYPLAVPLEVEVGIGEDWLSAKE
*****

```

Figure 4.9 – Alignment of the polymerase domain of wild-type *Taq* and selected phytate round 2 mutants. Alignment of *Taq*, Phy2(2), Phy2(19), Phy2(28), Phy2(30) and Phy2(45) were performed using ClustalW. Mutations are highlighted in red. * = single fully conserved residue, : = conservation of strong groups, . = conservation of weak groups.

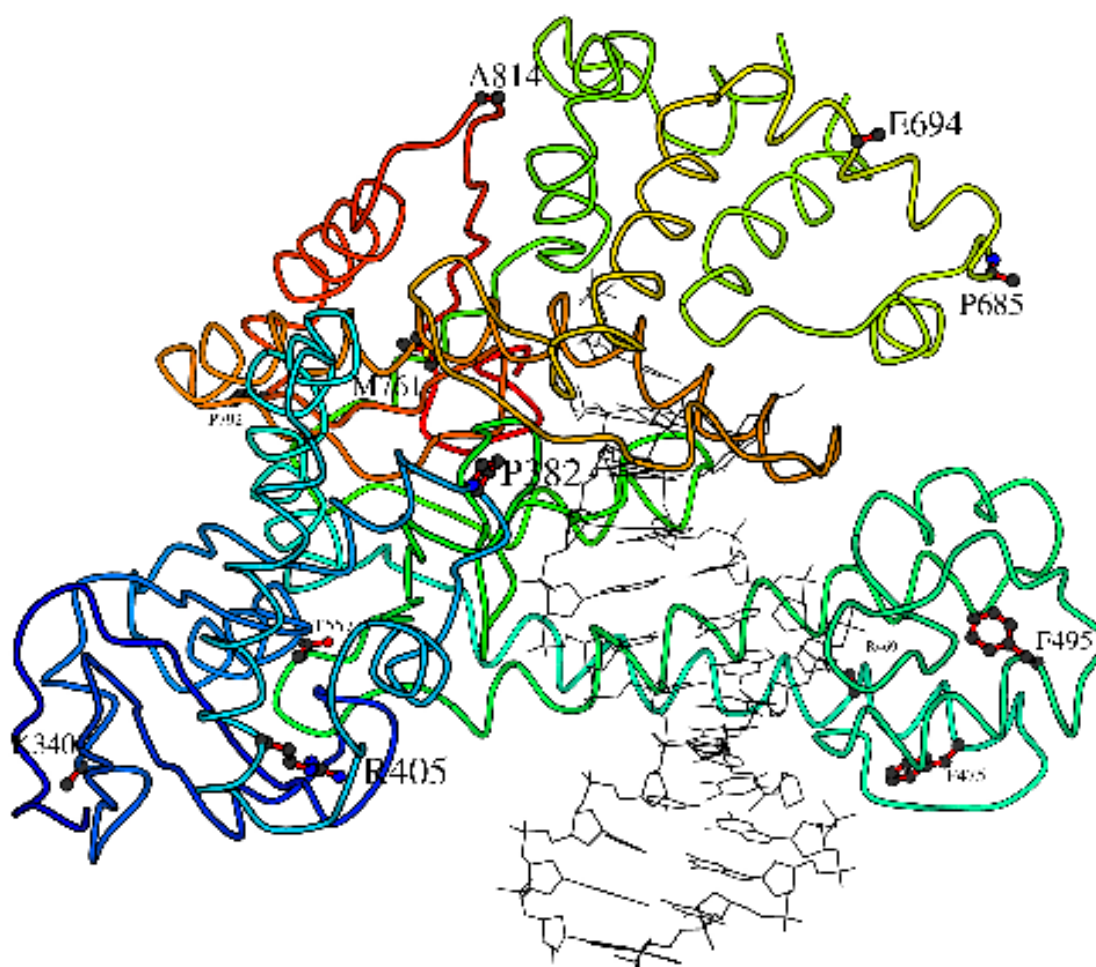


Figure 4.10 – A structural diagram of the polymerase domain of *Taq* DNA polymerase complexed with DNA. Mutated residues from selected second-round variants are labelled.

Characterisation of Phytate Resistant Variants

Five second-round polymerase mutants selected for further analysis were purified and further characterised alongside the wild-type Sac7-TaqPol. A few mutants from the first-round screen were also selected for purification; they had not been characterised initially but were important as they made up the basis of the StEP re-diversified second library as their amino acid changes were carried through to second-round mutants.

Mutants were purified following the standard procedure used during Heparin mutant purification as described in Chapter 3. Following initial treatment and buffer exchange, the protein samples were purified using Heparin Sepharose affinity chromatography, with elution over a salt gradient (100-600mM KCl). Samples collected during the protein

purification were analysed by SDS-PAGE to determine which fractions contained pure protein of the correct size. The purification of Phy2(30) is shown in Figure 4.11 as an example.

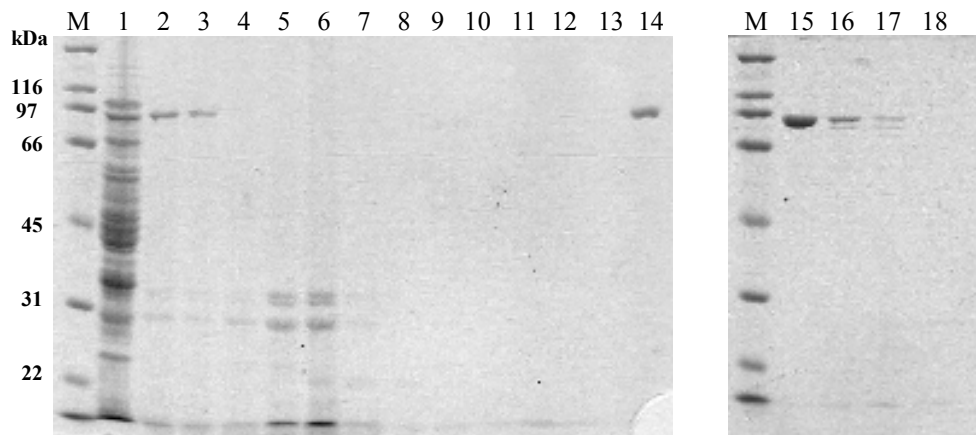


Figure 4.11 – SDS-PAGE gel of the purification of Phy2(30) round two mutant. This gel shows the samples collected over the course of the purification starting from the initial sample (1) to the pure polymerase mutant peak fraction (15). (1)-(2) sample after initial heat lysis and PEI precipitation, (4)-(6) initial wash/ flow through before salt elution gradient, (7)-(18) fractions from elution gradient.

Analysis of the protein fractions by SDS-PAGE revealed that the peak fraction for this particular purification was 15 (Figure 4.11) and the mutant eluted off the Heparin column at around 270mM KCl. All mutants that were purified eluted at around 270-300mM KCl. The peak fraction from each purification, along with the whole cell extract and the sample that was loaded onto the column for purification, were analysed by SDS-PAGE (Figure 4.12).

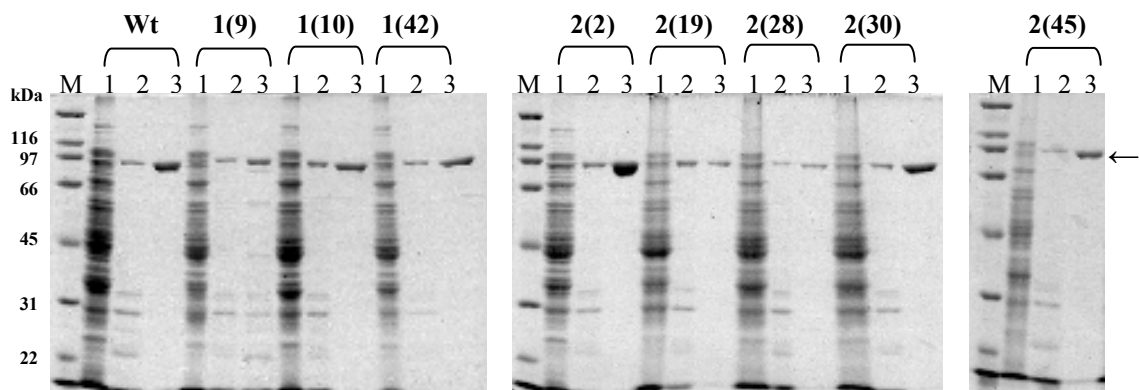


Figure 4.12 – SDS-PAGE gel of the purification of wt polymerase and variants from rounds 1 and 2 of CSR. Protein samples from key steps in the purification and the final peak fraction of purified protein were analysed by 8% SDS-PAGE. Each variant is labelled above the gel; numbers 1 and 2 indicate the round of CSR in which the mutant was obtained followed by the individual mutant number. The samples from each purification are: (1) whole cell extract, (2) loaded sample, (3) peak fraction. (M) marker. Sac7-TaqPolymerase is indicated by the arrow.

Following purification, the concentration of the protein samples was determined using the Bradford Assay. Then, using the standard polymerase assay and samples of a known activity, the specific activities of the mutants were determined (Table 4.4). No detectable activity was found in mutants Phy1(9) and Phy2(19). Two of the second-round mutants were selected, based on their performance in preliminary assays, and they were characterised further in the presence of phytate. After initial polymerase assays were performed to determine the optimal enzyme level and fluorescent signal, assays were run over a range of phytate concentrations. The percent activity at each phytate concentration was then calculated based on 100% activity controls. Graphical analysis of the data, performed as in Chapter 3 for the Heparin mutant Hep22, was used to determine the level of phytate resistance of Phy2(30) and Phy2(45) (Figure 4.13).

Table 4.4 – Calculated activity, protein concentration and specific activity of the purified polymerase variants.

	Activity (U/ml)	Protein Concentration (mg/ml)	Specific Activity (U/mg)
Wild-type	2380	0.120	19,850±580
Phy10	5670	0.343	16,530±1900
Phy42	1700	0.122	13,970±1400
Phy2(2)	4350	0.204	21,340±2300
Phy2(28)	3840	0.130	29,560±2000
Phy2(30)	3610	0.153	23,620±2000
Phy2(45)	2800	0.132	21,180±3100

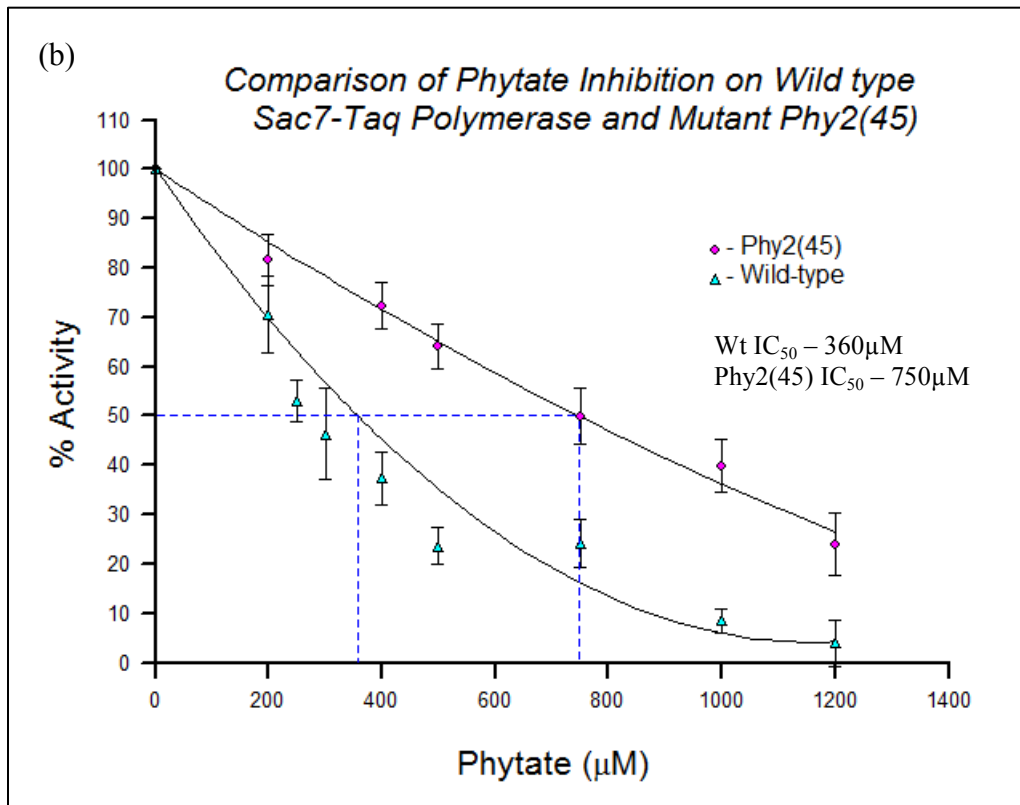
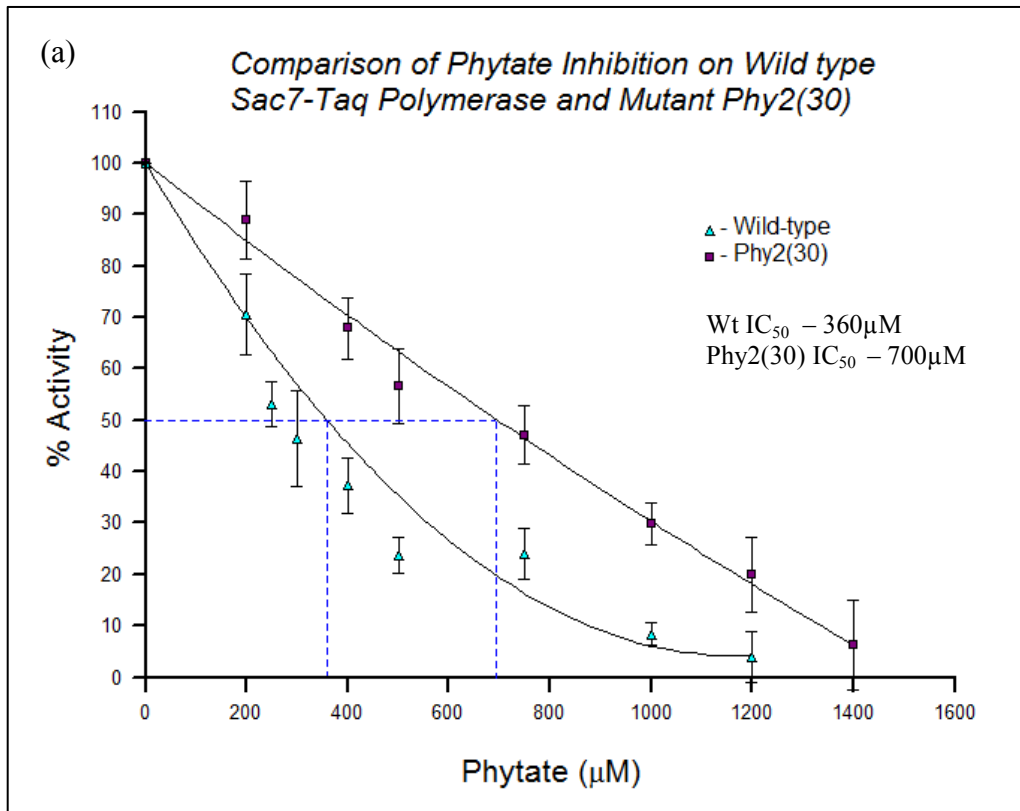


Figure 4.13 Graphs showing the polymerase activity of (a) mutant Phy2(30) and (b) Phy2(45) at varying concentrations of phytate. A polynomial regression curve fit of the data was performed.

Graphical analysis of the polymerase assay data over a range of phytate concentrations revealed a significant difference in the sensitivity to phytate of the wild-type compared to the both Phy2(30) and Phy2(45). The IC_{50} (inhibitory concentration at 50% activity) for Phy2(30) was $700\mu\text{M}$ and Phy2(45) was $750\mu\text{M}$, both of which are approximately 2 times higher than that of the wild-type enzyme ($IC_{50}=360\mu\text{M}$), indicating a substantial 2-fold increase in resistance to phytate. The increased resistance of Phy2(30) to phytate was also demonstrated in a 500bp λ PCR over a range of phytate concentrations (Figure 4.14).

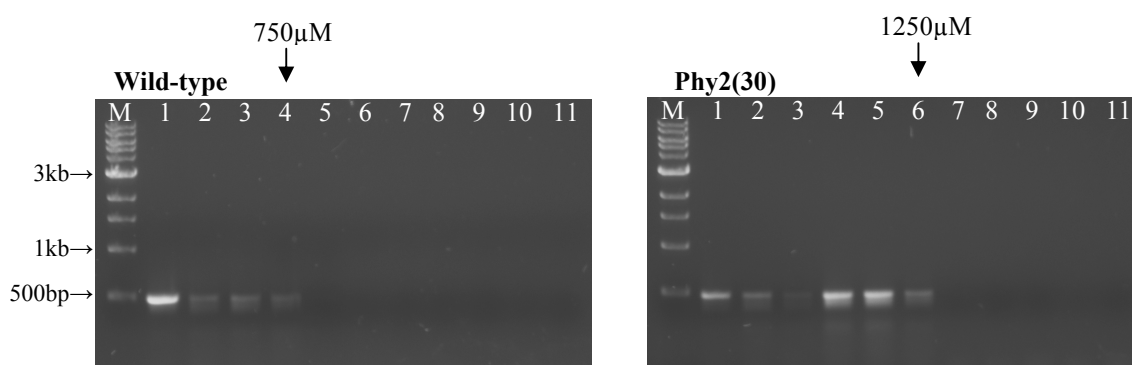


Figure 4.14 – Wt and Mutant Phy2(30) in a phytate PCR.

Agarose gels displaying the ability of wt and Phy2(30) to amplify DNA in the presence of phytate. PCRs were run over a range of phytate concentrations: (1) zero phytate control, (2) $250\mu\text{M}$, (3) $500\mu\text{M}$, (4) $750\mu\text{M}$, (5) $1000\mu\text{M}$, (6) $1250\mu\text{M}$, (7) $1500\mu\text{M}$, (8) $1750\mu\text{M}$, (9) $2000\mu\text{M}$, (10) $2250\mu\text{M}$, (11) $2500\mu\text{M}$. (M) 1kb DNA ladder (NEB).

Figure 4.14 clearly demonstrates that Phy2(30) was able to amplify a 500bp DNA template in the presence of phytate at concentrations up to $1250\mu\text{M}$, whereas the wild-type was inhibited at levels above $750\mu\text{M}$. Increased levels of phytate may even have had an enhancing effect on the Phy2(30) mutant PCRs, as seen in lanes 4 and 5 (Figure 4.14). These PCRs were repeated numerous times and the same effects were seen in all cases.

To characterise mutant Phy2(30) further, a thermal inactivation assay was performed. Enzyme aliquots were heated at 90°C for set time periods and the enzyme was then tested in a standard polymerase assay. The percent activity after heating at each time point was then calculated; graphical analysis of the data can be seen in Figure 4.15. The wild-type enzyme was slightly more thermostable than the mutant with a $t_{1/2}$ of 37 min compared to a $t_{1/2}$ of 30 min for mutant Phy2(30). This indicated that the mutations in Phy2(30) did not significantly affect the thermostability of the enzyme.

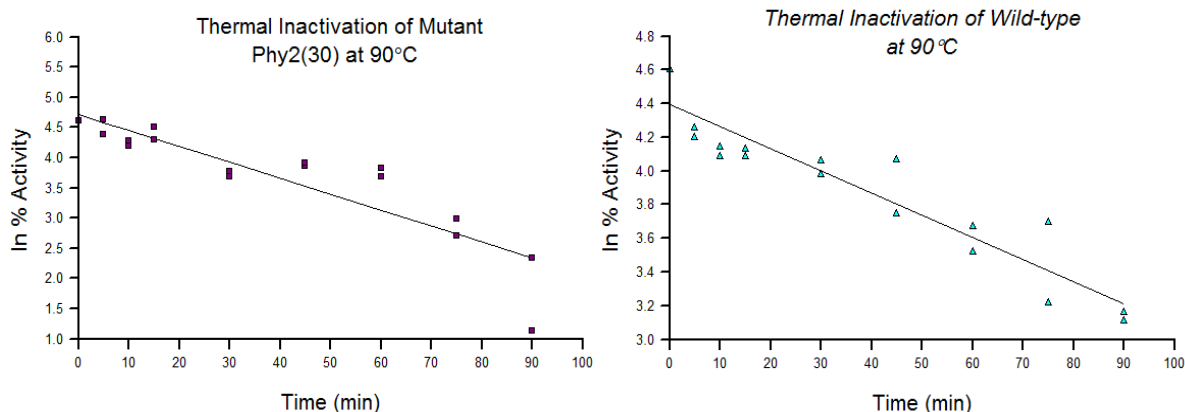


Figure 4.15 – Thermal inactivation of wild-type and Phy2(30).
 $t_{1/2}$ of wild-type was 36 min compared to $t_{1/2}$ of 30 min for mutant Phy2(30) .

Lastly, a fidelity assay was run comparing the wild-type Sac7-TaqPol and mutant Phy2(30). In the LacI PCR-based fidelity assay (Frey and Suppmann, 1995), a linearised fidelity plasmid was amplified by both enzymes. Following re-ligation of the plasmid and transformation, the number of blue and white colonies were counted and used to estimate error rate. Disruption of the LacI gene, caused by errors introduced during PCR, led to the LacI⁻ phenotype (blue colonies) on X-gal plates. Error rates (Table 4.5) were calculated using the formulae provided in the methods section. It was found that the error rates of the two enzymes differed slightly; the mutant was found to be more error prone than the wild-type enzyme.

Table 4. 5 – Fidelity assay data and calculated error rates of wild-type and Phy2(30).

Enzyme	Amount of template(ng)	Yield (ng)	Duplications	White colonies	Total Colonies	Fraction of white colonies	Calculated error rate
Wt	20	3000	7.23	727	1027	0.7079	1.37×10^{-4}
Phy2(30)	20	300	3.91	2300	3423	0.6719	2.92×10^{-4}

Following characterisation of mutant Phy2(30), the positions of its 3 amino acid residue changes (P685S, M761V, A814T) were modelled onto the *Taq* DNA polymerase structure (Li *et al.*, 1998) which was viewed using Cn3D (PDB ID: 3KTQ). Unlike with the mutations seen in the heparin resistant mutants, none of the 3 amino acid changes were close

to the DNA binding pocket of *Taq* DNA polymerase, although M761V was in close proximity to one of the Mg^{2+} binding sites of the enzyme (Figure 4.16).

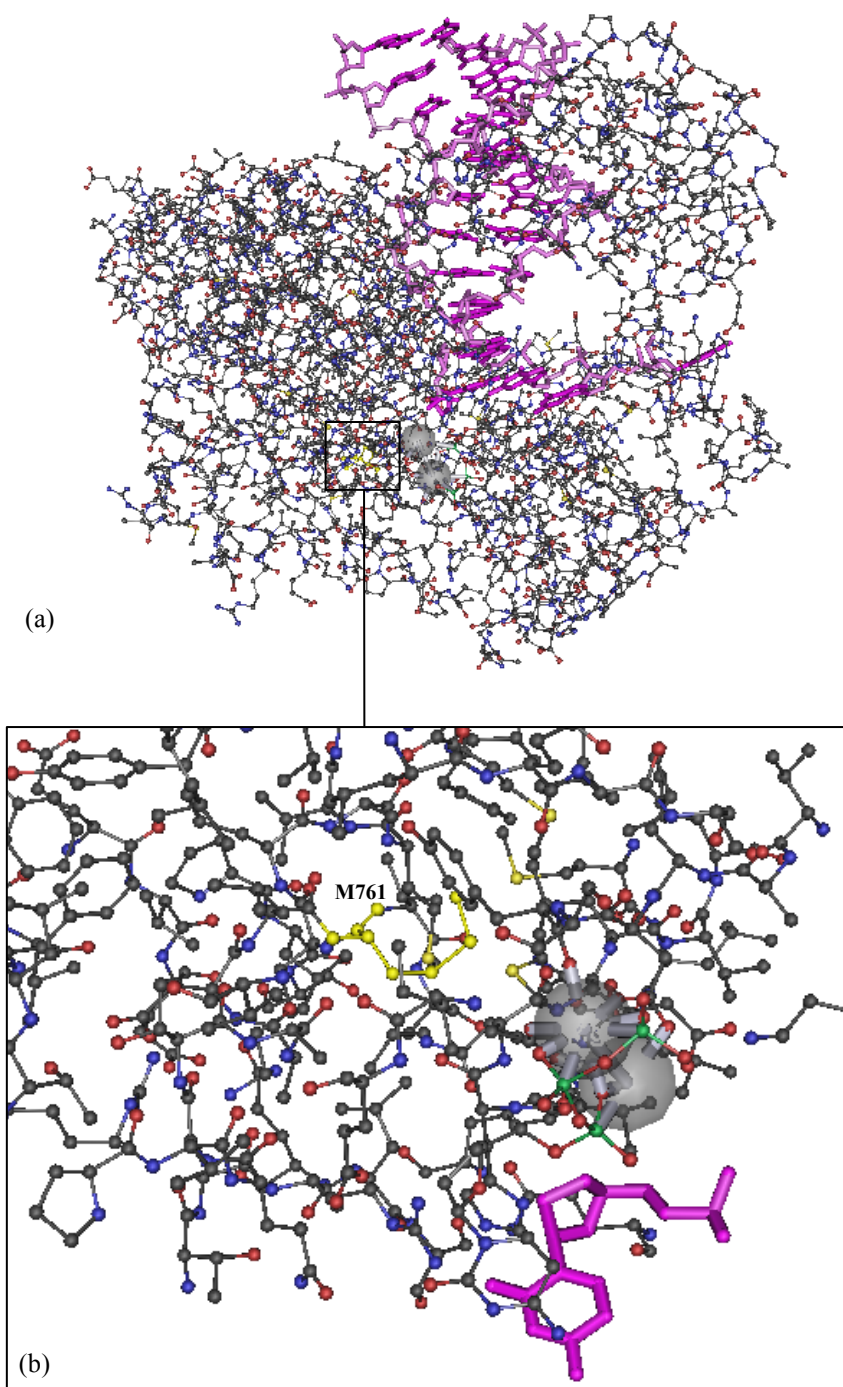


Figure 4.16 – M761V the mutated residue of Phy2(30) shown on the *Taq* Polymerase structure (PDB ID: 2KTQ). (a) The large fragment of *Taq* with the mutated residue highlighted in yellow, the DNA shown in pink and grey spheres representing the Mg^{2+} . (b) A close up of the highlighted residue and the Mg^{2+} .

Discussion

In this chapter, two rounds of directed evolution by compartmentalised self-replication (CSR) were performed to produce phytate resistant DNA polymerases. The CSR system, which was set up and verified in the previous two chapters, was used with the inhibitor, phytate. After one round of screening against the error-prone PCR library produced in Chapter 3, a re-diversified mutant library was produced using staggered extension process (StEP), followed by a second round of CSR. Selected mutants were characterised and the extent of phytate resistance achieved after two rounds of CSR was determined using polymerase assays.

The mutant library that was produced in Chapter 3 was used and screened again in this chapter. Instead of performing CSR in the presence of heparin, as in Chapter 3, the CSR screen was run with the addition of phytate. Initial emPCR experiments testing the sensitivity of wild-type Sac7-TaqPol to phytate revealed that 500 μ M phytate caused slight inhibition in the PCR and this was selected as an appropriate concentration to use in the first round of CSR.

The CSR screen was run with the mutant library in the presence of 500 μ M phytate. The polymerases able to replicate their polymerase domain under the set conditions came through the screen and were known as the first-round mutants. Standard activity testing of the first-round mutants showed that the percentage of active clones had increased to 50% compared to less than 20% in the original mutant library. This figure of 50% active clones was slightly lower than expected, as in the first round of heparin screening the percent of active clones increased to 80%. Nevertheless, the screen had increased the level of active clones compared to the pool of largely inactive polymerases that made up the original mutant library and thus demonstrated that active variants had been enriched by the process.

Sequencing of selected first-round mutants revealed that there was an average of 4.8 amino acid changes per mutant. It was found that one amino acid, K540, had been mutated in two of the mutants and had previously been found to be mutated in one of the first-round heparin mutants as well as a heparin-resistant mutant described by Ghaddessy *et al.* (2001) (Table 3.2). K540 can be found in the DNA binding pocket of *Taq* DNA Polymerase and has been shown to interact with the DNA template (Li *et al.*, 1998). It was interesting that this residue appeared mutated in a number of selected mutants in independent experiments, which suggests that it may be an important residue in conferring inhibitor resistance.

Forty-five mutants were tested in PCRs over a range of phytate concentrations and their sensitivity to phytate was compared to that of the wild-type. The mutants that performed well in initial PCR screens in the presence of phytate were selected to be taken

forward and used in StEP. The PCR shuffling technique, StEP, was performed to re-diversify the mutant library through the recombination of the first-round selected clones. Shuffling of the first-round selected clones combined multiple mutations that were thought to confer phytate resistance and also allowed for different combinations of these positive mutations to be explored. In directed evolution, selection from a random mutagenesis library followed by recombination is often employed, as recombination allows for the rapid accumulation of beneficial mutations as well as the removal of deleterious mutations from the population (Moore *et al.*, 1997; Arnold *et al.*, 2001).

A new library of re-diversified mutants was constructed with a final size of 5×10^4 . The size of the second library was smaller than that of the original mutant library. The use of the re-diversification technique rather than random mutagenesis limited the number of possible mutant outcomes so it was therefore deemed not necessary to make such a large library. Sequencing of the mutant library revealed that the clones contained mutations from the first-round phytate mutants as expected, which verified that StEP had been performed successfully. Activity testing revealed that the new library contained 50% of active mutants.

Following re-diversification of the library, a second round of CSR was performed using the new library and a higher level of phytate. The concentration of phytate was increased from $500 \mu\text{M}$ to $600 \mu\text{M}$. Although the higher level of $750 \mu\text{M}$ phytate was tried in the screen, no amplification was achieved as the inhibition must have been too strong for any significant portion of the mutant library to withstand and to come through the screen. After the $600 \mu\text{M}$ CSR screen, 80% of the second-round mutants showed polymerase activity, which was an increase on the 50% activity seen in the second library. Sequencing of 15 mutants revealed there to be an average of 3.6 amino acid changes per mutant.

Second-round mutants were tested by PCR over a range of phytate concentrations. The mutants' sensitivity to phytate was compared to that of wild-type by PCR and the mutants with the highest phytate resistance were chosen for further characterisation. After purification using Heparin Sepharose affinity chromatography, preliminary polymerase assays were performed and the specific activities of the mutants were determined; no significant changes to the specific activities were observed. Two mutants, Phy2(30) and Phy2(45), were characterised further using polymerase assays over a range of phytate concentrations. Graphical analysis of the assay data revealed that both mutants had a 2-fold increase in resistance to phytate compared to the wild-type. Although a substantial increase in resistance was seen, it was not as significant an increase as was seen previously in the heparin screen (Chapter 3) after one round of CSR. In this chapter, two rounds of CSR were performed, which was anticipated to produce even more substantial changes in resistance.

PCR experiments confirmed the results of the polymerase assays by demonstrating that Phy2(30) could withstand phytate concentrations 2-fold higher than wild-type enzyme

before amplification was inhibited. Phy2(30) was able to amplify 500bp DNA at concentrations up to 1250 μ M phytate whereas the wild-type was inhibited at levels above 750 μ M. Increased levels of phytate may even have had an enhancing effect on the Phy2(30) mutant PCRs; this could be the result of evolution performed under the selective pressure of phytate. Mutants may have evolved not only to withstand the inhibitory effects of phytate but also to require it for efficient PCR. Phy2(30) was characterised further using thermal inactivation and fidelity assays.

Thermal inactivation assays under PCR buffer conditions at 90°C revealed the wild-type to be slightly more thermostable with a $t_{1/2}$ of 36.8 min compared to a $t_{1/2}$ of 30.3 min for the mutant Phy2(30). As a comparison, standard published figures for the activity half-life of *Taq* DNA pol is 9 min at 97.5°C (Lawyer *et al.*, 1993). The fusion of the Sac7 protein onto *Taq* DNA pol has been shown not to alter the stability of the enzyme (Wang *et al.*, 2004). Given that the CSR selection was performed at high temperatures, it would not be anticipated that thermostability would be selected against. However, thermal assays were an important verification to ensure that the unnatural amino acid changes of the mutant had not affected this crucial enzyme property. CSR has been used for the evolution of increased thermostability in *Taq* DNA polymerase simply by increasing the duration of the initial denaturation step in the PCR reaction, which allowed for the selection of thermostable variants (Ghadessy *et al.*, 2001). A *Taq* DNA pol variant was isolated with an 11-fold longer half life ($t_{1/2}$ 16.5min) at 97.5°C compared to the wild-type ($t_{1/2}$ 1.5min).

Fidelity is another property that could affect the potential utility of the mutant enzyme. Fidelity assays showed the error rate of the mutant to be higher than that of the wild-type Sac7-TaqPol enzyme. This can be expected when mutations are introduced into an enzyme; however, the fidelity of the mutant enzyme was found to be only slightly higher than that of standard *Taq* DNA polymerase, which has a stated error rate of 2×10^{-4} (Saiki *et al.*, 1988). The slight increase in the error rate observed is not likely to cause a problem with further use of this mutant. Subsequent rounds of mutagenesis and screening may serve to decrease the error rate again.

Wild-type Sac7-TaqPol was found to have a lower error rate than that reported for standard *Taq* DNA pol, which was not predicted. Fidelity assays with a Sac7-Pfu polymerase fusion protein have shown that Sac7-Pfu maintains the same fidelity as the standard *Pfu* enzyme in PCR (Wang *et al.*, 2004). Similar assays have not previously been conducted using Sac7-Taq, but it has been reported that processivity, which is known to be enhanced by the Sac7 protein, may be inversely correlated with its accuracy in nucleotide incorporation (Barnes, 1992).

Sequencing results revealed Phy2(30) to contain 3 amino acid changes. When these mutations were mapped onto the *Taq* DNA polymerase structure (Figure 4.10) none of them

fell in the DNA binding pocket, as had been the case with the Heparin mutations (Chapter 3). It was thus not immediately possible to rationalise how the observed mutations had affected the enzyme or to make further conclusions regarding the mechanism of phytate inhibition. However, M761V was in close proximity to the Mg^{2+} binding site of the enzyme (Figure 4.16) so it is possible that this mutation may have had an effect on the Mg^{2+} binding of the enzyme by subtly changing the environment in the vicinity of the ion binding site instead of affecting phytate-enzyme binding.

The structural location of mutation M761V, along with the fact that two rounds of CSR only yielded polymerase variants with a 2-fold increase in phytate resistance, may indicate that phytate inhibits by metal ion chelation and not competitive inhibition with the DNA template. If this is the case, then in directed evolution by CSR one would expect to produce an enzyme with increased metal binding affinity, therefore giving increased resistance to metal chelators such as phytate. The 2-fold increase in phytate resistance achieved during two rounds of CSR may have been due to the fact that it is difficult to evolve increased metal binding as the residues that have direct contact with metal ions are highly conserved (Ador *et al.*, 2004).

The work presented in this chapter provides some evidence that the CSR system can be used to screen for phytate resistant *Taq* DNA pol variants. CSR enabled the screening and selection of active variants with a 2-fold increase in resistance to phytate. Whilst the screen was able to select active variants, only a two-fold increase in resistance was observed after two rounds of CSR. Despite this result it is possible that the final characterised mutant enzyme Phy2(30) (P685S, M761V, A814T) may have potential application in PCR amplification in natural phytate-containing samples, such as faecal samples. The actual level of phytate in faecal samples is hard to predict as it is dependent on both diet and the digestion process. Ruminant animals, such as cows and sheep, do not completely digest phytate (He and Honeycutt, 2001) and animals such as pigs and chickens cannot digest phytate at all (Wodzinski and Ullah, 1996); therefore the level of phytate in faecal samples is very variable but given its abundance in most animal and human diets its presence in fecal samples is certain.

In light of the results described in this chapter, the small improvement in resistance achieved may have been due in part to the mode of action of the phytate; however, further rounds of CSR would need to be carried out before firm conclusions could be made. Work in the next chapter investigates the possible enzyme binding sites of phytate and its chelating potential in an attempt to further elucidate the role of phytate in PCR inhibition.

Chapter Five – Investigation of the Mechanism of phytate Inhibition of PCR

Introduction

As previously discussed in Chapter 4, the mechanism of phytate inhibition of PCR is not clearly understood. Although phytate as a compound has been well known and studied for years, investigations into its role as a PCR inhibitor have been limited. One study of phytate inhibition in PCR led to the proposal that inhibition occurs through the chelation of Mg^{2+} ions (Thorton and Passen, 2004), which are essential for the polymerase-catalysed nucleotidyl transfer reaction (Steitz, 1999). As phytate is known to be a potent chelator of metal ions (Maenz *et al.*, 1999), then chelation is definitely one possible mode of inhibition; however, other facts suggest that phytate may inhibit by binding to the enzyme, which must be considered as an alternative mode of action.

In solution at physiological pH, phytate will carry a strong negative charge due to multiple phosphate groups. As well as forming complexes with metal ions, phytate is able to bind directly to proteins through electrostatic charges (Angel *et al.*, 2002). Although phytate is a small molecule, it can self-aggregate into multivalent, macromolecular complexes that are proposed to have higher apparent binding affinity to proteins than individual phytate molecules (Shears, 2001). These facts may imply a potential similarity in the mode of action between phytate and heparin, the PCR inhibitor discussed in Chapter 3. We know that heparin is negatively charged, forms a helical structure and can form interactions with many proteins (Mulloy *et al.*, 1993). Therefore, it is thought that heparin may inhibit DNA polymerase by mimicking the helical and negatively-charged structure of DNA. The conformation of phytate has been suggested to mimic the pyranose ring structure of heparin and therefore to cause inhibition of heparin-binding proteins through competitive inhibition (Vucenik *et al.*, 2004). The reverse situation has also been reported, where heparin inhibits binding of inositol phosphates causing the inactivation of Ca^{2+} channels and signal transduction events (Tones *et al.*, 1989). In conclusion, both phytate and heparin may inhibit DNA polymerase activity and therefore PCR through competitive binding at the DNA primer/template binding site.

In work by Ghadessy and colleagues (2001), a *Taq* DNA pol variant was produced that had a 130-fold increase in resistance to heparin compared to the wild-type. This heparin-resistant variant had 6 residue changes that clustered in the base of the finger and

thumb subdomains, which are regions involved in binding the primer-template duplex. Four of the mutated residues actually come in contact with either the primer or template strand in the open and closed ternary polymerase-template-primer complex (Li *et al.*, 1998). As heparin is thought to compete with the DNA for binding to the polymerase, the results of this study suggest that the heparin-resistance conferring mutations reduced affinity for heparin whilst not altering the polymerases binding of DNA (Ghadessy *et al.*, 2001). The potential similarity in the mode of action between phytate and heparin suggests that heparin-conferring mutations may also confer phytate resistance.

Another PCR inhibitor that has been recently investigated is humic acid, which is found in soil. Humic acid is a poorly characterised heterogeneous mixture of phenolic compounds, which like phytate, has the ability to complex with metal ions (Zipper *et al.*, 2003). A recent investigation by Kermekchiev and colleagues (2009) not only involved an investigation into the mode of action of humic acid but also the production of a *Taq* DNA polymerase mutant with resistance to humic acid. It was proposed that soil could exhibit two modes of inhibition, one by affecting the catalytic activity of the DNA polymerase, and two by reducing enzyme processivity/speed either directly or by blocking the DNA template. In the above study the humic acid resistant mutants showed a stronger affinity to DNA than the wild-type *Taq* DNA pol, which suggests the mutants had a stronger interaction with the template and an increased ability to overcome competitive inhibition caused by inhibitors (Kermekchiev *et al.*, 2009). In research of other soil components, inhibitors have been observed to bind the DNA polymerase, therefore suggesting polymerase to be the target for inhibition (Watson and Blackwell, 2000). These studies indicate further that binding to the DNA polymerase is a key mechanism of inhibition for many naturally occurring PCR inhibitory compounds.

Kermekchiev and colleagues (2009) showed that the effect of inhibitors, found in soil and blood samples, was primarily on the *Taq* DNA polymerase as inhibition could be overcome by the introduction of mutations into the polymerase. The key residue 708, which was mutated to confer resistance, is located on the surface of the protein 40 residues from the finger domain which interacts with the incoming template and dNTPs (Li *et al.*, 1998). This key residue falls in a 'hinge' region that allows movement of the finger domain during conformational changes of the enzyme. Kermekchiev proposes that some PCR inhibitors may interact with the surface of the enzyme and slow down the movement of the fingers and by mutating residue 708 the effect of the inhibitor can be reduced.

Although there are many published crystal structures of KlenTaq and *Taq* DNA polymerase (Korolev *et al.*, 1995; Kim *et al.*, 1995), to date no structures have been published with a PCR inhibitor bound, which leaves us to speculate on their apparent binding sites. The crystal structures of the open and closed forms of the large fragment of

Taq DNA pol provide crucial information regarding the residues that contact the primer/template and dNTP polymerase substrates (Eom *et al.*, 1996; Li *et al.*, 1998). The crystal structures of a phytase-phytate complex (Lim *et al.*, 2000) and the antithrombin-thrombin-heparin complex (Li *et al.*, 2004), present the natural binding interactions of these compounds with their associated enzymes. Figure 5.1 shows a phytate molecule in the active site of *E. coli* Phytase (Lim *et al.*, 2000). The available *Taq* DNA polymerase crystal structures, along with the structures of PCR inhibitors, allow some speculation as to inhibitor modes of action or possible enzyme binding sites; however, docking studies with inhibitory compounds such as heparin and phytate would provide valuable information and further our understanding of the mode of action of PCR inhibitors.

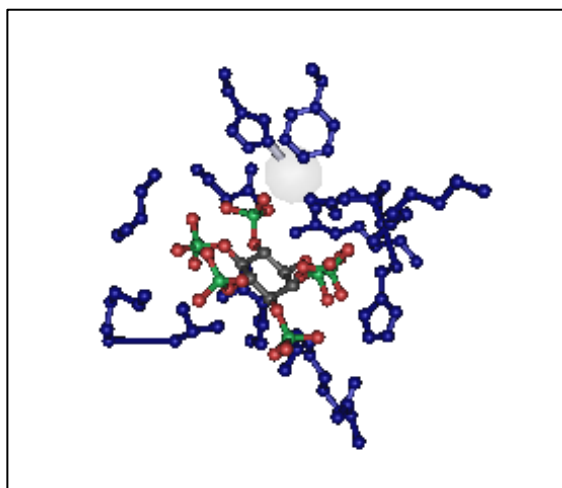


Figure 5.1 – Crystal structure of *E. coli* Phytase complexed with phytate. This picture shows a phytate molecule in the active site of the Phytase enzyme (PDB ID: 1DKM). Phytate molecule shown in grey, red and green. Phytase residues are shown in blue. Hg^{2+} depicted by white sphere.

This chapter presents the results of some experiments that attempted to elucidate further the potential mode of action of phytate's inhibition of PCR. Three possible hypotheses were considered: (1) chelation of Mg^{2+} , (2) competitive inhibition by enzyme binding, or (3) a combination of (1) and (2). Mg^{2+} titration PCRs were carried out to investigate phytate chelation in PCR and X-ray crystallography studies were performed in an attempt to determine whether phytate binds the DNA polymerase. Furthermore, as phytate also shows some similarities to heparin, site-directed mutagenesis was used to investigate whether mutations known to confer heparin-resistance could also confer phytate resistance.

Methods

Magnesium Titration Phytate PCRs

Five sets of 500bp λ PCRs were set up with 0, 100, 250, 500 and 1000 μ M phytate (phytic acid dodecasodium salt hydrate supplied by Sigma, Gillingham, UK). Each set of PCRs was run over a range of Mg^{2+} concentrations from 0 to 6mM increasing in 0.5mM increments. Final PCR concentrations were as follows: 1x Taq Buffer (GeneSys, Camberley, UK) with no added $MgCl_2$, 200 μ M dNTPs, 25pmol of forward and reverse 500bp λ primers, 1ng λ DNA, 1 μ l Sac7-TaqPol enzyme (purified in Chapter 4), XmM $MgCl_2$, Y μ M phytate and ddH₂O up to 50 μ l. Cycling conditions: 94°C for 3min, followed by 30 cycles of 94°C for 15s, 55°C for 15s and 72°C for 45s. 5 μ l was analysed on a 1% (w/v) agarose gel.

Buffer Exchange and Analysis of KlenTaq Sample

For crystallisation, pure KlenTaq protein sample was supplied by GeneSys (Camberley, UK). The purity of the sample was analysed by 8% SDS-PAGE as standard (Chapter 2). To remove glycerol from the original enzyme storage buffer, a Microcon 30kDa filter device (Millipore, Watford, UK) was used to perform buffer exchange and also protein concentration. The final protein buffer was 50mM KCl, 20mM Tris (pH 8), 0.1mM EDTA, 1mM DTT, 0.5% (v/v) Tween 20, and 0.5% (v/v) Nonidet 40. The Bradford assay (BioRad, Hemel Hempstead, UK) was used to estimate the protein concentration of the sample.

Crystallisation

The Hampton Research pre-crystallisation test was used as recommended with ~20mg/ml KlenTaq sample to check correct protein concentration and sample homogeneity. A crystallisation screen was set up using the hanging drop method with 1 μ l of protein to 1 μ l of well/reservoir buffer solution; crystals were grown at 16°C. Final crystallisation conditions were as follows: 13% (w/v) polyethylene glycol (PEG) 4000, 100mM Tris (pH 8.5), 200mM $MgCl_2$, and 1mM phytic acid dodecasodium salt hydrate (Sigma, Gillingham, UK). Standard crystallisation buffer for initial crystal growth was as above but without the addition of phytate. Crystals were seeded using seed bead (Hampton Research, Aliso Viejo, CA, USA) in order to produce larger single crystals.

Crystal Data Collection and Structural Analysis

The selected crystal was frozen in a beam of nitrogen at 100K before x-ray diffraction, which took place at the Diamond Light Source synchrotron, Oxfordshire, using a 3 Giga electron volt light beam. 180 images were taken while rotating the crystal by 1° for each image collected. The data collection was performed by Dr.Susan Crennell. The structure was resolved to 2.45Å and refined to 2.5Å. All structural determination was performed by Dr.Susan Crennell and Lisette Meerstein at the University of Bath.

Site-Directed Mutagenesis

Site-directed mutagenesis was performed using the QuikChange® Multi Site-Directed Mutagenesis kit (Stratagene, Cheshire, UK) to introduce multiple residue changes into Sac7-TaqPol. The following mutants of Sac7-TaqPol were created: (1) Ghadessy, K611R, D649G, N654S and M781R, (2) Heparin, V600A, T659S, R701L. QuikChange mutagenic primers were designed as per Stratagene's guidelines.

Heparin mutant primers (Bioneer, Alameda, CA, USA):

HepV600A - 5' GAGGCCACCCATCGCGGAGAAGATC

Hep T659S – 5' AACATCCCCGTCCGCAGCCCGCTTGGG

Hep R701L – 5' GAACCTGATCTGGTCTTCCAGGAGGG

Ghadessy Mutant primers (Bioneer, Alameda, CA, USA):

Ghad K611R – 5' CGGGAGCTCACCAAGCTGAAGAGCACCTAC

Ghad D649G and N654S – 5'

AGTAGTCCGTCCCAACCTCCAGAGCATCCCCGTCC

Ghad M781R – 5' GCCGAGCGCAGGGCCTTCAACATGCC

The mutagenesis reactions were performed following the recommended protocol. After thermocycling, the reactions were treated with 10U *DpnI* (Stratagene, Cheshire, UK) before being transformed into XL10 gold cells (Stratagene, Cheshire, UK) following the Stratagene guidelines. Transformation reactions were spread onto LB/Kan (50µg/ml) agar plates and incubated at 37°C overnight. Ten colonies were selected for culture and plasmid prep, and were then sent for sequencing as standard. If all mutations were not present following sequencing, the mutagenised plasmid was used in another mutagenesis reaction

with the appropriate primers until all mutations were added. Mutations from each sequential mutagenesis reaction were confirmed by sequencing.

Purification of Site Directed Mutants

Protein expression, PEI precipitation and buffer exchange to prepare the protein samples for purification were performed as in Chapter 3. Affinity chromatography using a heparin Sepharose column (GE Life Sciences, Little Chalfont, UK) was performed followed by dialysis into *Taq* storage buffer, as described in Chapter 3.

Polymerase Activity Assay

The DNA polymerase assays were run and analysed as described in Chapter 4. Graphical analysis was performed using the program FigSys (BioSoft, Cambridge, UK).

Results

Investigation of Mg^{2+} Chelation in Phytate PCRs

As phytate is known to be a potent chelator of divalent metal ions, a set of phytate PCRs were performed with the aim to investigate whether phytate chelation of Mg^{2+} leads to inhibition in PCR and to see if this inhibition could be overcome with the addition of extra Mg^{2+} . Five different phytate concentrations were tested, including a zero phytate control, over a range of Mg^{2+} concentrations. The wild-type Sac7-TaqPol, purified in Chapter 4, was used in a range of PCRs as described above and each set of phytate PCRs was analysed by gel electrophoresis (Figure 5.2).

Analysis of the Mg^{2+} /phytate PCRs revealed that in the zero phytate control PCR, amplification could be visualised at all levels of Mg^{2+} (1-6mM) and as expected no amplification was seen when zero Mg^{2+} was added as it is essential for DNA polymerase activity. In the remaining 4 PCRs there was a general trend, as the level of phytate increased so did the level of Mg^{2+} needed in order to get visible amplification. Unexpectedly, at all levels of phytate, amplification could only be visualised over a certain range of Mg^{2+} concentrations. At 100 μ M phytate, amplification was seen at 1-2.5mM Mg^{2+} but above this level of Mg^{2+} it appeared that the PCR was inhibited. Similar results were obtained in the 250 μ M phytate PCRs where amplification was only visualised between 1-3.5mM Mg^{2+} , and in the 500 μ M phytate PCRs where amplification was only visualised at 1.5-4mM Mg^{2+} .

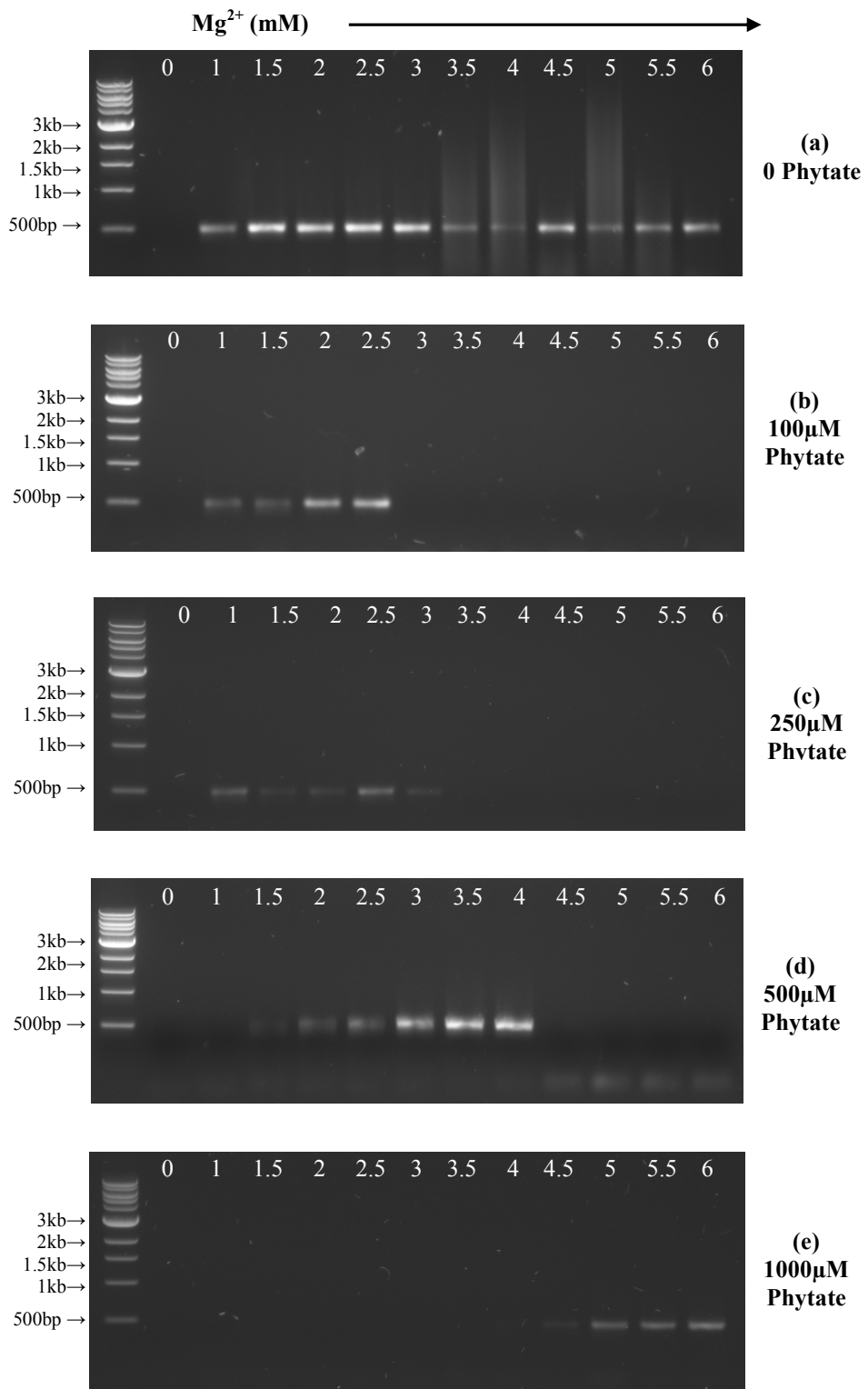


Figure 5.2 – Phytate PCRs over a range of Mg²⁺ concentrations. 500bp λ PCRs were run with Sac7-TaqPol with 5 different levels of phytate: (a) zero, (b) 100µM, (c) 250µM, (d) 500µM, (e) 1000µM. Each level of phytate was tested over a range of Mg²⁺ concentrations (the amount in mM is marked above each lane).

Investigation of Phytate/Enzyme Binding by X-ray Crystallography

Crystallisation studies were performed in order to investigate whether phytate inhibition of PCR was caused by phytate binding to the DNA polymerase. The enzyme polymerase domain, in the form of KlenTaq, was crystallised in the presence of phytate to attempt to dock the molecule into any binding site(s) within the enzyme structure prior to X-ray diffraction and polymerase structure determination.

A purified KlenTaq DNA polymerase (62.5kDa) sample was provided by GeneSys (Appendix C). To check the sample purity and homogeneity, a dilution series of the protein was analysed by SDS PAGE (Figure 5.3).

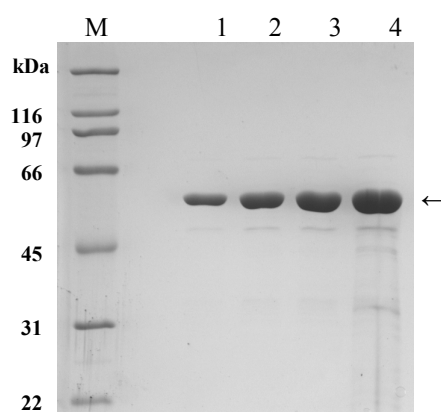


Figure 5.3 – Purified KlenTaq DNA polymerase
A dilution series of the protein sample was analysed by SDS-PAGE. (M) Marker, (1) ~0.5µg, (2) ~1µg, (3) ~2µg, (4) ~3µg. 62.5kDa KlenTaq indicated by arrow.

The KlenTaq sample was estimated to be >95% pure when it was analysed by SDS PAGE. A Bradford assay was performed to check the concentration of the sample which was calculated as 16mg/ml. Additives, such as glycerol, that are present in storage buffers can hinder crystallisation trials so were removed from the protein sample. The KlenTaq sample was transferred to a glycerol-free storage buffer through buffer exchange to prepare it for crystallisation.

The Hampton Research pre-crystallisation test was carried out on the KlenTaq sample to check for homogeneity and that the sample was in the correct protein concentration range needed for crystallisation. The test indicated that the protein concentration may be too low so the sample was concentrated slightly from 16mg/ml to 20mg/ml. The pre-crystallisation test was repeated and this time crystals started to form after a couple of days, which indicated that the protein sample was at a good concentration for

crystallisation. As crystals had grown in the pre-crystallisation test buffer B2, it was decided to start crystal trials under similar buffer conditions: 0.1M Tris (pH8.5), 0.2M MgCl₂, 15% (w/v) PEG 4K. Crystallisation screens (Figure 5.4a) were set up using the hanging drop method and through the course of the crystal trials many aspects of the buffer such as pH, % PEG and concentration of MgCl₂ were adjusted to promote crystal growth.

Once the optimal buffer conditions for crystal growth had been established, phytate was then added to the buffer and a new set of crystals were grown in the presence of 1mM phytate. Precipitation was observed to occur readily in certain samples. The crystals that formed were fairly thin and flat and were either rectangular or leaf shaped. A few needle clusters were also seen. It was decided to try to seed the crystals to promote the growth of larger crystals. The results of the seeding experiments were very positive with more well-defined and larger crystal formations. The final crystals used in X-ray diffraction were thin and rectangular in shape with an approximate size of 2.1µm by 5.0µm (Figure 5.4b).

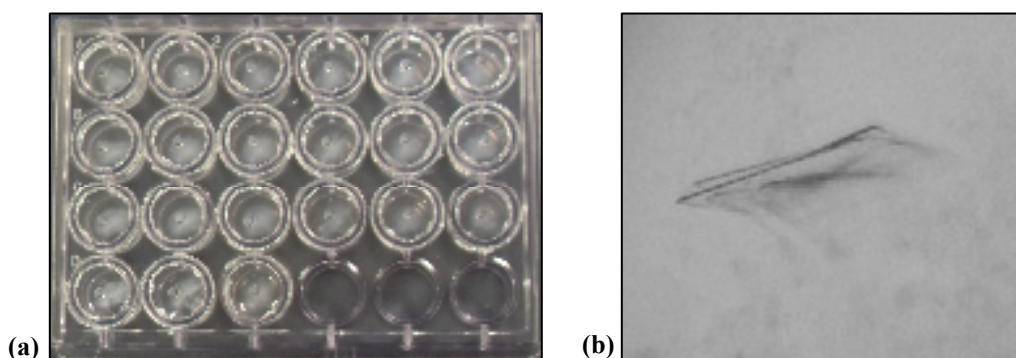


Figure 5.4 – (a) Photograph of a crystallisation screen plate with hanging drops and (b) Klentaq crystal grown in the presence of phytate.

A crystal grown in the presence of phytate was selected for X-ray diffraction. During diffraction, 180 images were taken while rotating the crystal by 1° each image. Diffraction images, such as the one in Figure 5.5a, were processed using the position and intensity of each spot, or reflection from the crystal. The final structure of Klentaq DNA polymerase, crystallised in the presence of phytate, was resolved to 2.5Å and can be seen in Figure 5.5b. Unfortunately, no phytate was visible in the structural data which suggested that phytate had not bound to the enzyme.

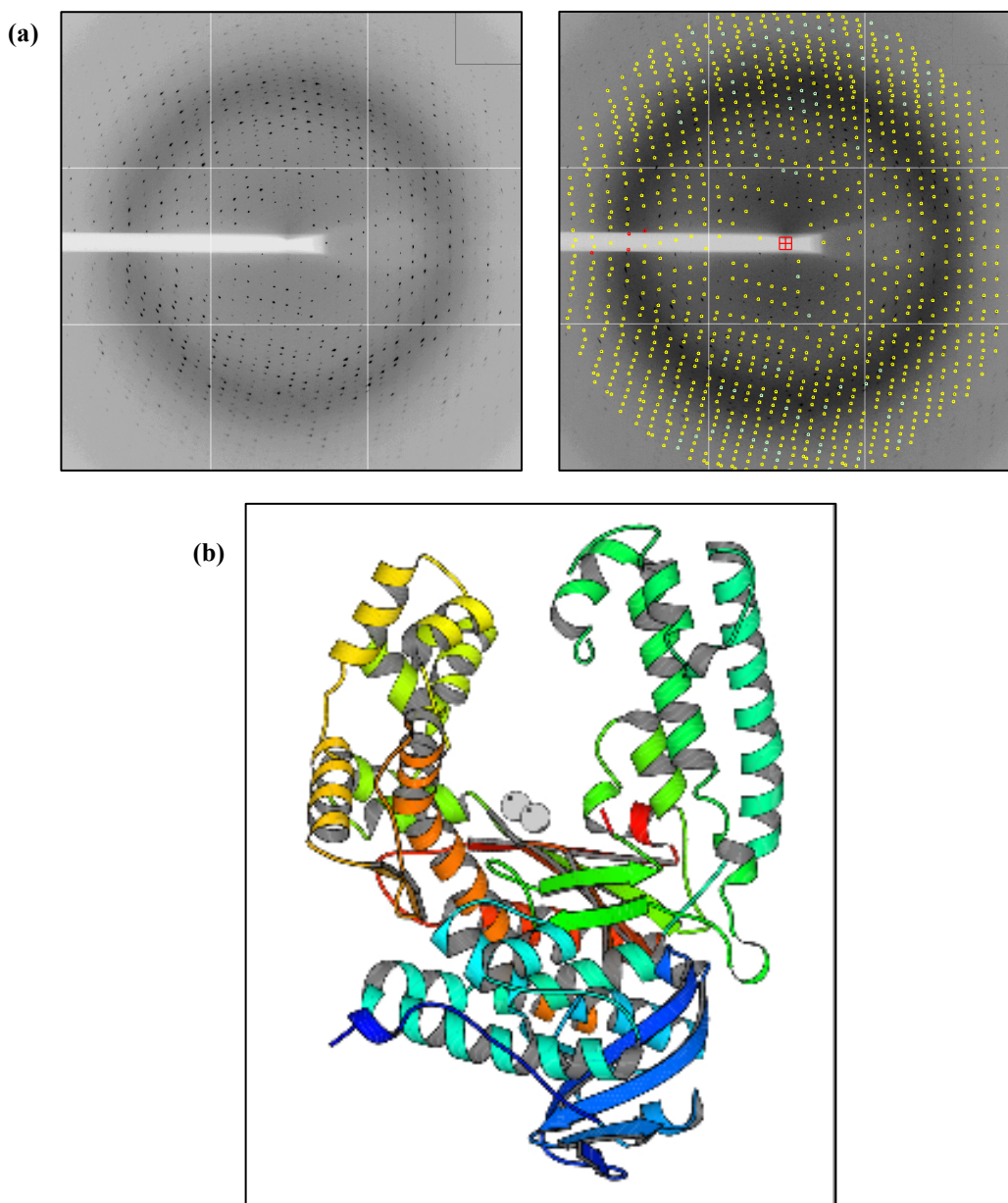


Figure 5.5 – Diffraction images and final Klentaq structure.

(a) Examples of images obtained from X-ray diffraction. Black dots or Spots were produced by reflections from the crystal during diffraction and were used in preliminary data analysis.

(b) The structure of Klentaq DNA polymerase determined by X-ray crystallography and resolved to 2.5Å. Grey spheres indicate Mg^{2+} ions.

Investigation of Phytate Inhibition by Site-Directed Mutagenesis

In an attempt to figure out whether the mutations that conferred heparin-resistance could also confer phytate-resistance, a set of mutants were produced using the QuikChange[®] Multi Site-Directed Mutagenesis kit. The first of these was the Ghadessy mutant (K540R,

D578G, N583S, M747R), which was based on a heparin-resistant *Taq* DNA polymerase variant that Ghadessy and colleagues (2001) produced and which incorporated four amino acid changes into the polymerase domain. The original variant produced by Ghadessy and colleagues contained 6 mutations but only those in the polymerase domain were selected for use in this project. The second mutant was created using a combination of amino acid changes from heparin mutants in chapter 3 that were located close to the DNA binding pocket. The triple heparin mutant (V529A, T588S, R630L) combined mutations found in Hep7, Hep11 and Hep22 (Chapter 3).

The site-directed mutants (SDM) were expressed and purified as standard (Chapter 3) using a Heparin-Sepharose column and affinity chromatography. Purification samples were analysed by SDS-PAGE and it was found that the mutants eluted off the column, as expected, at around 300mM KCl. Key stages in the purification, including the peak elution fraction, are shown in Figure 5.6.

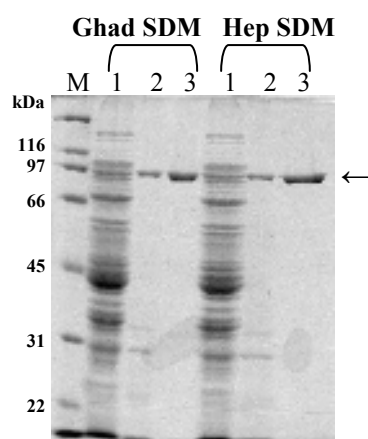


Figure 5.6 – SDS-PAGE gel of key stages in the purification of site-directed mutants. Protein samples from key steps in the purification and the final peak fraction of purified protein were analysed by 8% SDS-PAGE. The samples from each purification are: (1) whole cell extract, (2) loaded sample, (3) peak fraction. (M) markers. Sac7-TaqPolymerase is indicated by the arrow.

From SDS-PAGE analysis, the protein was estimated to be over 90% pure. Following purification, the protein concentration of the site-directed mutants was determined using a Bradford assay. The two mutants were then characterised further using the polymerase assay. The specific activities of the mutants were determined using the standard polymerase assay and samples of a known activity (Table 5.1).

Table 5.1 – The calculated specific activities of site-directed mutants.

	Activity (U/ml)	Protein Concentration (mg/ml)	Specific Activity (U/mg)
Ghadessy SDM	2563	0.142	18,050
Heparin SDM	4815	0.159	30,290

After initial polymerase assays were performed to determine the optimal enzyme level and fluorescent signal, assays were run over a range of phytate concentrations to determine if these mutants, despite being evolved for heparin-resistance, may also exhibit phytate-resistance. The percent activity at each phytate concentration was calculated based on 100% activity controls. Graphical analysis of the data, performed as in Chapters 3 and 4 for the previous mutants, was used to determine the level of phytate resistance of the Ghadessy (Figure 5.7a) and Heparin site-directed mutants (Figure 5.7b). The wild-type data from Chapter 4 were used as a comparison.

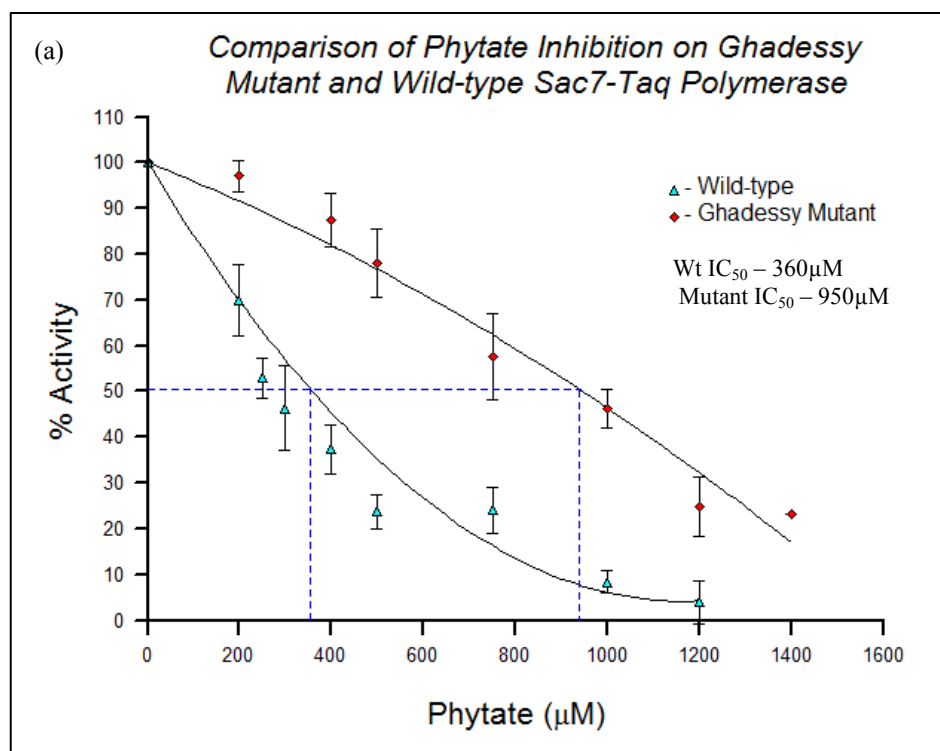


Figure 5.7(a) – Graph showing the polymerase activity of Ghadessy site-directed mutant at varying concentrations of phytate. A polynomial regression curve fit was performed on the data.

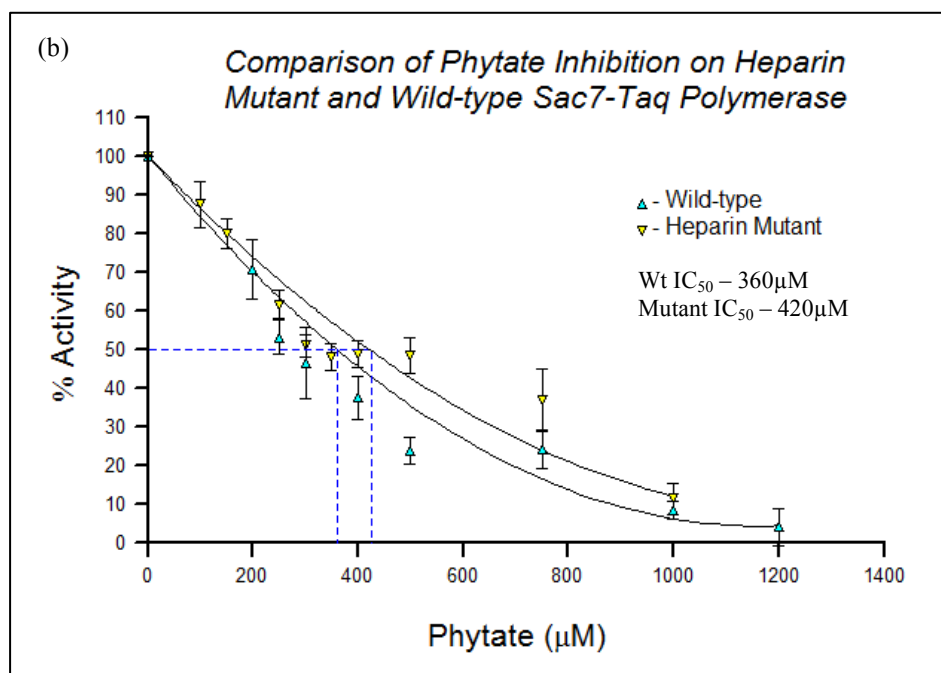


Figure 5.7(b) – Graphs showing the polymerase activity of Heparin site-directed mutant at varying concentrations of phytate. A polynomial regression curve fit was performed on the data.

Graphical analysis of the polymerase assays over a range of phytate concentrations revealed that there was a difference in the sensitivity to phytate of the wild-type compared to the Ghadessy mutant. The IC₅₀ (inhibitory concentration at 50% activity) for Ghadessy was 950µM, which was almost 3 times higher than that of the wild-type (IC₅₀ of 360µM), indicating a substantial 2.6-fold increase in resistance to phytate. These data indicated that the Ghadessy mutations, which had previously been shown to confer heparin-resistance, also confers phytate-resistance to some extent. Analysis of the Heparin site-directed mutant revealed that there was a slight difference in its sensitivity to phytate compared to the wild-type, albeit not statistically significant. Both the wild-type and the heparin mutant had a similar IC₅₀; 360µM and 420µM, respectively. This revealed that the combination of amino acid changes, which were selected from different heparin-resistant *Taq* DNA polymerase variants produced in Chapter 3, did not confer a significant increase in phytate resistance. The mutants' relative sensitivity to phytate was also demonstrated in a 500bp λ PCR over a range of phytate concentrations with standard 1.5mM Mg²⁺ (Figure 5.8).

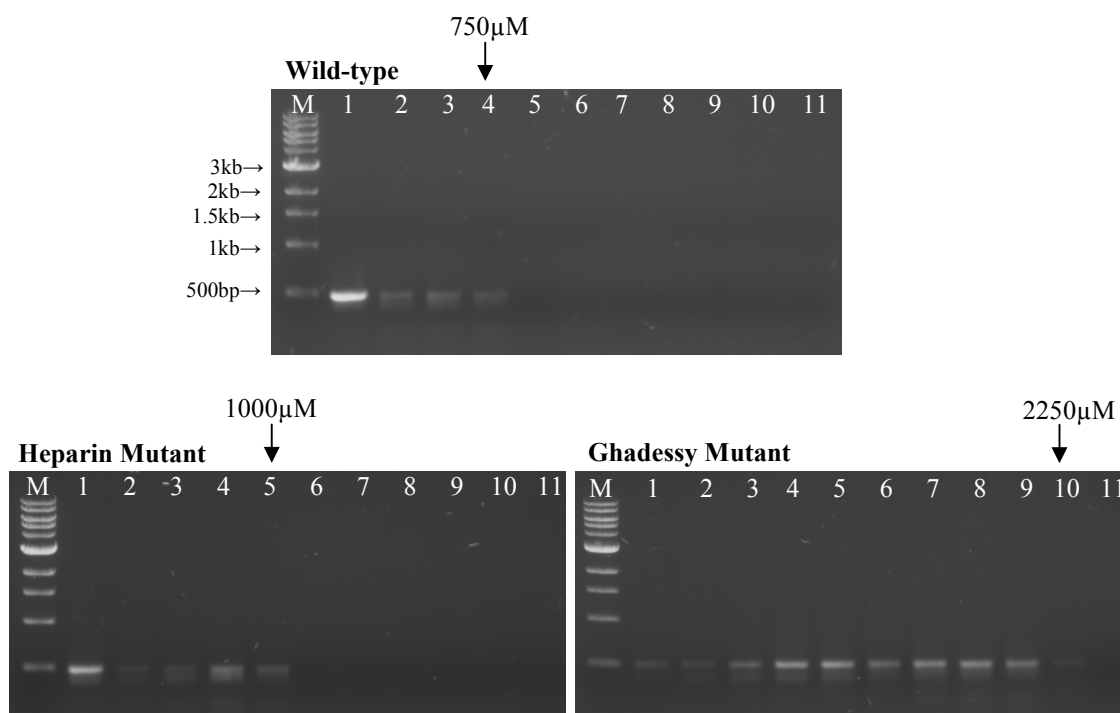


Figure 5.8 – Wt and site-directed Heparin and Ghadessy mutants in a phytate PCR.

Agarose gel displaying the ability of the wt and the mutants to amplify DNA in the presence of phytate. PCRs were run with a standard 1.5mM Mg^{2+} over a range of phytate concentrations: (1) zero phytate, (2) 250 μ M, (3) 500 μ M, (4) 750 μ M, (5) 1000 μ M, (6) 1250 μ M, (7) 1500 μ M, (8) 1750 μ M, (9) 2000 μ M, (10) 2250 μ M, (11) 2500 μ M. (M) 1kb DNA ladder (NEB).

Figure 5.8 clearly demonstrates that the Ghadessy mutant was able amplify 500bp DNA at phytate concentrations up to 2250 μ M whereas the wild-type was inhibited at levels above 750 μ M. These results corresponded to the polymerase assay data, where a 3-fold increase in phytate-resistance was observed with the Ghadessy mutant. The Heparin mutant was able to amplify 500bp DNA in concentrations up to 1000 μ M phytate, which was a slight improvement on the wild-type, but no substantial change in resistance was observed, which once again corresponded to polymerase assay data.

To characterise the Heparin site-directed mutant further, a series of heparin polymerase assays was performed to see if the 3 mutations (V600A, T659S, R701L) selected from different heparin-resistant *Taq* DNA pol variants (Chapter 3) demonstrated increased heparin-resistance when all the amino acid changes were combined in one mutant. Although the mutant had been previously tested with phytate and only demonstrated a minor increase in phytate-resistance, the mutations were originally selected based on their ability to confer heparin-resistance and therefore the mutant's sensitivity to heparin was tested. Polymerase assays were run as standard over a range of heparin concentrations and graphical analysis of the data was performed as in Chapter 3 (Figure 5.9).

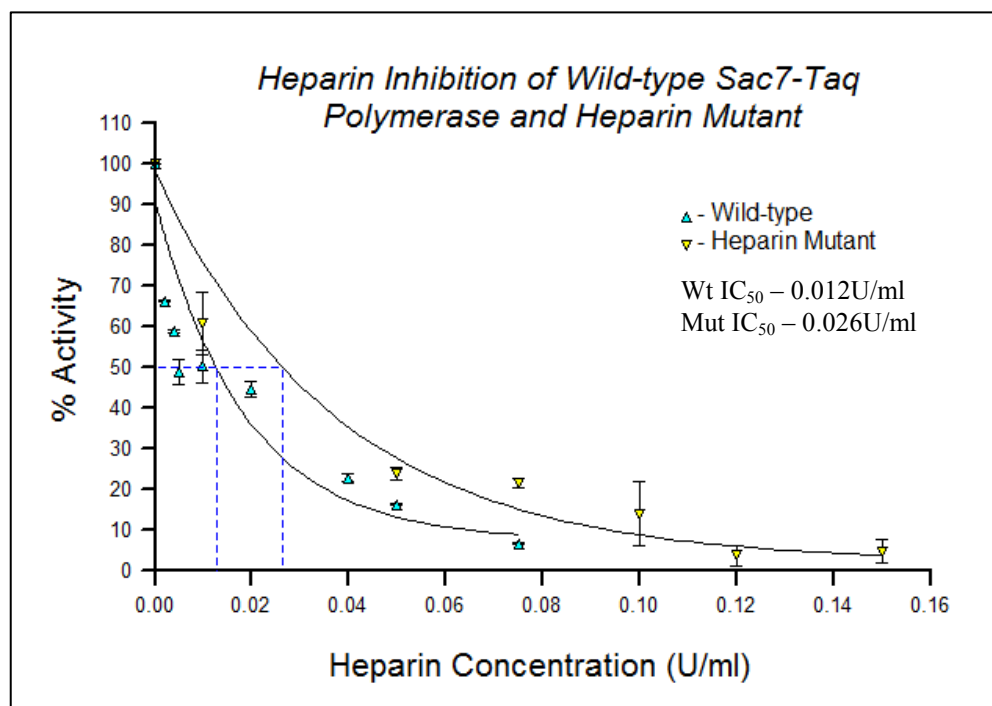


Figure 5.9 – A graph showing the polymerase activity of wild-type Sac7-Taq Polymerase and the site-directed Heparin mutant over a range of heparin concentrations. A first-order decay with residual curve fit was performed on both sets of data.

Graphical analysis of the polymerase assay data over a range of heparin concentrations revealed that there was a difference in the sensitivity to heparin of the wild-type compared to the Heparin mutant. The IC_{50} of the heparin mutant, 0.026U/ml, was double that of the wild-type IC_{50} of 0.012U/ml. This revealed that the combination of amino acid changes selected from different heparin-resistant *Taq* DNA pol variants, conferred a 2-fold increase in heparin resistance; however, the increase in resistance was not as significant as the 4-fold resistance achieved in Heparin mutant 22 (Chapter 3).

Discussion

In this chapter a number of experiments are described which attempted to elucidate the mode of action of phytate inhibition of PCR, which was initially hypothesised to occur in three ways: (1) chelation of Mg^{2+} , (2) competitive inhibition by enzyme binding, or (3) a combination of (1) and (2). Experiments were performed in three areas to investigate phytate inhibition: (a) the chelation of Mg^{2+} ions, which are essential for *Taq* DNA polymerase activity, (b) binding of phytate to the polymerase as evidence of competitive and/or non competitive inhibition, and (c) the ability of heparin-conferring mutations to also confer phytate resistance.

Mg^{2+} titration experiments over a range of phytate concentrations revealed some important findings. They provided strong evidence that chelation of Mg^{2+} ions by phytate leads to a decrease in the availability of Mg^{2+} to the extent that polymerase activity and thus PCR amplification are inhibited. It was found that at each level of phytate tested, addition of Mg^{2+} to a certain critical level was able to overcome inhibition and thus to permit amplification. As the level of phytate increased, so did the level of Mg^{2+} needed in order to achieve PCR amplification. PCRs at all different levels of phytate tested, started working at a critical point (Figure 5.2), where the relative level of Mg^{2+} was in excess of the predicted number of ion binding sites within each phytate molecule, assuming that approximately 5-6 Mg^{2+} ions bind to each phytate molecule. The ability of phytate to inhibit the PCRs below the critical level of Mg^{2+} provides strong evidence that the phytate- Mg^{2+} binding affinity is stronger than that of the enzyme- Mg^{2+} binding affinity, therefore allowing phytate to bind Mg^{2+} preferentially over the enzyme.

Thorton and Passen (2004) speculated that phytate's mode of inhibition was primarily by chelation of Mg^{2+} ; however, the investigation reported herein presents the first experimental evidence that this is the case. The findings imply that the problem of phytate inhibition in experimental samples (e.g. for diagnostic applications) can be overcome to some extent by addition of Mg^{2+} at a level in excess of the critical point, at least over the range of phytate concentrations tested (0-1mM).

Unexpectedly, it was found that as the Mg^{2+} level was increased significantly above the relative number of phytate ion binding sites the PCRs were again inhibited. This contrasted with the zero phytate control PCRs which worked over all the Mg^{2+} levels tested, with an optimal concentration of 1.5-3mM. At 100 μ M phytate, amplification could be seen at 1-2.5mM Mg^{2+} but above this level of Mg^{2+} PCR was inhibited (Figure 5.2b). As Mg^{2+} -phytate complexes are known to exist in a complex mix of up to ten different chelate species in solution (Crea *et al.*, 2006), it is possible that one or more of these species may be a

potent inhibitor of the enzyme and at a certain ratio of Mg^{2+} to phytate. If this is the case at this point, the mode of inhibition moves from chelation to enzyme binding, which would explain the drop off in activity seen in the phytate PCRs above a certain Mg^{2+} level (Figure 5.2 b,c and d). The makeup of the different phytate- Mg^{2+} species in solution is known to be affected by variables such as pH and the presence of other ions in solution (Cheryan *et al.*, 1983; Crea *et al.*, 2006). The results of the investigation reported in this chapter indicate the complexity of the phytate- Mg^{2+} interaction and have important implications for phytates' role in PCR inhibition.

Following Mg^{2+} titration PCRs, the potential mechanism of phytate PCR inhibition in the form of phytate-enzyme binding was investigated. Crystallisation screens were carried out in the presence of phytate in an attempt to dock the molecule into the enzyme KlenTaq, prior to X-ray diffraction. Although the structure of KlenTaq has been published (Korolev *et al.*, 1995), it has never been crystallised in the presence of an inhibitor to date. Flat rectangular crystals were successfully produced in phytate containing buffer and X-ray diffraction data were collected and analysed. The KlenTaq structure was resolved to 2.5Å but no phytate was found bound to the enzyme.

There are a number of reasons why phytate may not have been visible in the structural data. The first is that phytate does not bind the *Taq* DNA pol polymerase domain (KlenTaq) and therefore does not inhibit PCR by competing with the template for binding to the enzyme. The second explanation is that phytate may not bind very strongly to the enzyme as it has no high affinity binding site. In this case, phytate may have been bound in a number of different places on the enzyme but not always in the same orientation. Structural data would not have picked up the 'low occupancy' of phytate if it bound in numerous places on the enzyme at a low frequency. The third possibility is that all available phytate formed a chelate with the Mg^{2+} (Cheryan *et al.*, 1983; Crea *et al.*, 2006) present in the crystallisation buffer and was therefore not accessible for enzyme binding. As discussed previously, the formation of phytate- Mg^{2+} complexes is extremely complicated and is affected by a number of variables. Further studies involving the 'soaking' of crystals in different phytate concentrations to increase the chances of phytate, or solution chelates of Mg^{2+} -phytate, docking to the enzyme should be performed in order to give a definitive answer as to whether phytate does or does not bind the *Taq* DNA pol polymerase domain (Klentag).

This chapter also described an investigation into the ability of heparin-resistance conferring mutations to also confer phytate resistance. Using site-directed mutagenesis, two mutants were successfully produced which contained mutations previously found to confer heparin-resistance. The two mutants were purified and characterised in polymerase assays. The Ghadessy mutant (K540R, D578G, N583S, M747R), based on a heparin-

resistant *Taq* DNA pol variant produced by Ghadessy (2001), incorporated 4 amino acid changes into the polymerase domain. These 4 mutations were chosen, out of 6 in the original mutant, as they fell in the polymerase domain, which was the main focus of directed evolution in this project; they were also thought to confer most of the heparin-resistance as they were clustered in the DNA binding region of the polymerase (Ghadessy *et al.*, 2001). Polymerase assays over a range of phytate concentrations revealed that the Ghadessy mutant had a significant 2.6-fold increase in phytate-resistance compared to the wild-type (Figure 5.7). This was interesting as, despite being evolved for heparin-resistance, the Ghadessy mutant also exhibited phytate-resistance. The results of the polymerase assay were supported by a phytate PCR (with 1.5mM Mg²⁺) in which the Ghadessy mutant was able to amplify at concentrations of 2250µM phytate compared to the wild-type which was inhibited at concentrations above 750µM (Figure 5.8).

Although a small but significant increase in phytate-resistance was seen with the Ghadessy mutant, it was not as significant as the 130-fold increase in heparin resistance that Ghadessy and colleagues (2001) managed to produce over 3 rounds of evolution with their mutant (Table 3.2). This suggests that phytate probably does not bind the enzyme specifically at the heparin binding site within the DNA binding pocket. The observed increase in phytate resistance may be due to an increased metal binding affinity which evolved in part with the heparin resistance. In some cases heparin has been observed to chelate metal ions (Shin *et al.*, 2006; Whitfield and Sarkar, 1992) and it may be this mechanism that is shared by heparin and phytate, although to different extents, that has caused evolved heparin-resistance mutations to confer phytate resistance. Recent work by Kermekchiev *et al.*(2009) revealed that a single amino acid change in *Taq* DNA pol can confer resistance to inhibition by both blood and soil, which may indicate a common mechanism between certain inhibitors found in soil and blood.

This experiment provides some evidence that the heparin evolution of the Ghadessy mutant has increased ion binding affinity to some extent. The results of the Ghadessy phytate PCR (Figure 5.8) show that the mutant was able to produce amplification in levels of up to 2250µM phytate with only 1.5mM Mg²⁺. However, in the initial Mg-titration experiments (Figure 5.2), when only 1.5mM Mg²⁺ was added to the PCR, the wild-type enzyme was inhibited at 1000µM phytate due to a lack of available Mg²⁺. The fact that the Ghadessy mutant can produce amplification at such high relative levels of phytate implies that the enzymes binding affinity for Mg²⁺ has increased above the affinity of phytate-Mg²⁺ and the mutant is therefore able to overcome the chelating potential of phytate.

The second site-directed mutant was created using a combination of 3 amino acid changes selected from the heparin mutants in Chapter 3. The triple heparin mutant (V529A, T588S, R630L) was tested in a polymerase assay over a range of phytate and heparin

concentrations to see if (a) the heparin mutations conferred phytate resistance and (b) the combined effect of the selected heparin mutations had enhanced the heparin-resistance of the enzyme in comparison to the original heparin mutants (Chapter 3). Polymerase assays over a range of phytate concentrations revealed that there was a slight difference in the sensitivity to phytate of the heparin site-directed mutant compared to the wild-type (Figure 5.7). The difference in sensitivity to phytate was not statistically significant, however, and this was supported by results of phytate PCRs. The heparin mutant was able to amplify in up to 1000 μ M phytate whereas the wild-type was only able to produce amplification in concentrations up to 750 μ M phytate (Figure 5.8). These results demonstrate that in this case the selected heparin-resistance conferring mutations did not confer any significant phytate resistance.

The heparin mutant was then tested in polymerase assays over a range of heparin concentrations. The assays revealed that the mutant had a 2-fold increase in heparin resistance compared to the wild-type. Although an increase in resistance was observed, it was not as significant as the 4-fold increase in resistance achieved with mutant Heparin 22 in Chapter 3. These results indicated that the particular set of amino acid changes selected for the site-directed heparin mutant did not increase the heparin-resistance of the mutant beyond that which had already been achieved by evolution in Chapter 3. These results also suggest that the mutation P550S, or the combination of P550S and T588S, was key to conferring heparin resistance in heparin 22 since P550S was not selected for the site-directed mutant; this led to a smaller increase in heparin-resistance being observed in the site-directed mutant compared to the directed evolution Hep 22 mutant. As it is extremely difficult to predict the effects of amino acid changes, the results of this site-directed experiment were not entirely surprising.

The work in this chapter provides some evidence that phytate most likely causes PCR inhibition through the chelation of essential Mg²⁺ ions. Mg-titration PCRs showed that the negative effects of phytate inhibition of PCR can be overcome by the addition of extra Mg²⁺ to the PCR; they also served to illustrate the true complexity of the Mg²⁺-phytate interaction. KlenTaq-phytate docking studies did not provide any evidence that phytate or its Mg²⁺ chelation species bind directly to the enzyme, at least not with high affinity or with high specificity. Finally, site-directed mutagenesis studies showed that heparin-resistance evolved mutations can in some part also confer phytate resistance, which may be a result of increased Mg²⁺ binding affinity overcoming the chelating potential of both heparin and phytate. Further studies would be needed to clarify and fully understand the mode of action of phytate inhibition.

Chapter Six – Perspectives and Future Work

Overview

This project reports the establishment of a high-throughput system for the directed evolution of DNA polymerase variants by compartmentalised self-replication (CSR) (Chapter 2). The system was successfully employed to evolve *Taq* DNA polymerase variants with enhanced resistance to heparin (Chapter 3) and phytate (Chapter 4) over one or two rounds of CSR, respectively. Further work (Chapter 5) established the importance of the relative Mg^{2+} and phytate levels in PCR and their effect on PCR inhibition. Site-directed mutagenesis studies also revealed some crossover between mutations that confer heparin and phytate resistance. An overview of the properties of a selection of novel variants produced and studied in this project is presented in Table 6.1.

Table 6.1 – Properties of novel *Taq* DNA polymerase variants evolved in this project.

		AA Mutations (<i>Taq</i> Numbering)	Specific Activity (U/mg)	IC₅₀ Heparin (U/ml)	IC₅₀ Phytate (μM)
	Wild-type	n/a	19,850	0.012	360
Heparin	Hep22	P550S, T588S	14,870	0.050	
	Ghadessy SDM¹²	K540R, D578G, N583S, M747R	18,050	n/d ¹³	950
	Hep SDM	V529A, T588S, R630L	30,290	0.026	420
Phytate	Phy2(30)	P685S, M761V, A814T	23,620	n/d	700
	Phy2(45)	K340I, E694D, M761V, A814T	21,180	n/d	750

Whilst this project reported some success in evolving novel variants and understanding the mode of phytate inhibition, there are a number of different ways in which

¹² Based on a heparin-resistant *Taq* variant produced by Ghadessy *et al.*, 2001.

¹³ Not determined.

this work could be extended. Some potential future work based on the results reported in this project will now be discussed.

Exploring the Applied Use of Novel *Taq* DNA polymerase Variants

The main aim of this project was to produce a DNA polymerase variant with increased resistance to phytate with the intention being that the enzyme could then be used in fecal-based PCRs, thus eliminating the need for pre-treatment of the sample to remove inhibitory phytate. The experiments performed in this project used a commercial source of phytate in all characterisation assays; however, it would be interesting to see how robust the phytate-resistant variants evolved in this project are when used in clinical samples. Studies into the potential application of phytate-resistant variants would first require an exploration of the phytate levels present in natural samples that are subjected to diagnostic testing. The level of phytate in faecal samples is hard to predict as it is dependent on both diet and the digestion process (Joung *et al.*, 2007; He and Honeycutt, 2001). Other aspects such as the dilution of the sample and the nature of the diagnostic test being performed should also be considered when testing novel *Taq* DNA polymerase variants.

The findings in Chapter 5 revealed that if the phytate and Mg^{2+} levels are at certain relative concentrations in a PCR then inhibition can be overcome. These results should be explored further and tested in an applied system, as it is possible that the simple addition of Mg^{2+} to an appropriate level is sufficient to overcome the problem of phytate inhibition of PCR. This would mean that extensive further attempts to engineer enzyme variants with phytate resistance would not be warranted from an applied perspective. On the other hand, if the novel variants produced through directed evolution by CSR were able to completely overcome the problem of PCR inhibition in an applied system, then further rounds of directed evolution may still prove valuable.

Directed Evolution Studies

The work presented in Chapters 3 and 4 demonstrates that the technique of directed evolution by CSR can be used to evolve *Taq* DNA polymerase variants with increased resistance to inhibitors. However, the number of rounds of directed evolution that were performed were low, which restricted the overall outcome of the experiment. If further rounds of directed evolution by CSR were performed, each round of evolution may contribute an additional increase in resistance leading to a final *Taq* DNA polymerase variant with substantially increased resistance compared to the wild-type enzyme.

The CSR heparin screen performed in Chapter 3, produced a mutant with a substantial 4-fold increase in heparin resistance in a single round of CSR, which was similar to the 8-fold increase in resistance seen by Ghadessy *et al.* (2001) after one round of directed evolution. However, Ghadessy *et al.* (2001) went on to perform multiple rounds of CSR with a final outcome of an improved variant with 130-fold increased resistance after 3 rounds of CSR. These results suggest that continuing work with the heparin-resistant mutant produced in Chapter 3, and performing further rounds of mutagenesis followed by CSR screening or StEP recombination, would be beneficial. Continuing the directed evolution of the heparin-resistant mutant would provide insight into whether the mutations gained in this study overlap with the mutations that conferred heparin resistance in the Ghadessy *et al.* (2001) final *Taq* DNA polymerase variant or whether it is possible to produce heparin resistance through entirely different mutation combinations.

The mutations found in heparin-resistant variants in this study and subsequent work involving further rounds of CSR could provide enhanced resistance in addition to providing further evidence for heparin's mode of action. Follow-up work could involve saturation mutagenesis of key residues or further recombination techniques such as StEP. Selected mutations could also potentially result in variants with resistance to other inhibitors that have a shared binding site.

The CSR phytate screen performed in Chapter 4 could also be extended by further rounds of directed evolution but the situation with phytate PCR inhibition is more complicated than that of heparin. Phytate's mode of action is not known although the work completed in Chapter 5 suggests that the main mode of PCR inhibition by phytate is through the chelation of essential Mg^{2+} ions. This implies that in order to achieve an increase in phytate-resistance by CSR, the metal binding affinity of the enzyme must be increased to overcome the effects of phytate. Overall, the CSR phytate investigation led to a lower relative increase in resistance than that of the heparin screen, which may suggest it is more challenging to evolve phytate resistance or increase enzyme metal binding affinity. Further rounds of CSR would provide some evidence as to whether or not it is possible to evolve phytate resistance significantly beyond the level achieved in this study. The number of rounds of directed evolution conducted varies with each experiment but in general an average of at least 3-5 rounds of evolution are performed to achieve a significantly enhanced variant (Bolt *et al.*, 2008; Johannes *et al.*, 2005). If further rounds of CSR were successful, then the subsequent study of phytate-conferring mutations would be key to providing further insight into phytate's mode of PCR inhibition. Furthermore, the ability of heparin-conferring mutations to also confer phytate resistance (as in Chapter 5) and vice versa could be explored.

Biophysical Approaches to Investigate Phytate Interactions

The work presented in Chapter 5 began to elucidate the mode of action of phytate in PCR inhibition. Although it appears that phytate mainly inhibits PCR through the chelation of essential Mg^{2+} ions, inhibition by binding to the enzyme still cannot be ruled out. The initial crystallisation of Klentaq in the presence of phytate indicated that phytate did not bind the enzyme; however, the evidence was not conclusive. This initial experiment provides a useful starting point for further attempts at phytate docking with the enzyme, which could involve growing crystals under various combinations of Mg^{2+} and phytate or soaking crystals in different levels of phytate. As performed in this study, X-ray crystallography could then be used to detect potential phytate or phytate- Mg^{2+} chelate binding sites in the *Taq* DNA pol polymerase domain. Docking studies would provide additional evidence as to whether phytate does or does not bind the enzyme with high affinity or specificity.

Other approaches to investigate phytate's interaction with DNA polymerase could also be used to attempt to establish the mode of phytate inhibition in PCR. The technique of isothermal titration calorimetry (ITC) is one possible method for measuring biological interactions through the detection of heat changes. The heat that is either generated or absorbed upon binding of a protein-ligand complex is measured to determine binding constants, reaction stoichiometry, and the enthalpy and entropy of the interaction (Perozzo *et al.*, 2004). ITC has previously been used to study the thermodynamics of the binding of *Taq* DNA polymerase with primed template DNA (Datta and LiCata, 2003). Surface Plasmon Resonance (SPR) is another method for measuring biological interactions. SPR measures small changes caused by the interaction of a protein-ligand complex, which allows the determination of binding constants. The target molecule is immobilised on a solid surface, which is then treated with a ligand in a mobile phase. If binding occurs then there is a change in the intensity of reflected light, which is used to analyse the interaction (van Der Merwe, 2001).

Understanding the interaction of phytate and its Mg^{2+} chelates with DNA polymerase, if there is one, would also be key to determining phytate's mode of action and may provide valuable information that can be applied to other inhibitors. It would also fuel rational and semi-rational mutagenesis attempts to further engineer phytate resistance in DNA polymerases. Further knowledge of the formation of complexes between Mg^{2+} and phytate may also contribute to understanding phytate's inhibition of PCR through chelation.

All the above methods would be key to development of an efficient phytate-resistant *Taq* DNA polymerase variant, which could be used in clinical or veterinary diagnostic tests with PCR-inhibitory samples.

References

- Ador, L., Jaeger, S., Geslain, R., Martin, F., Cavarelli, J. and Eriani, G. (2004) Mutation and evolution of the magnesium-binding site of a class II aminoacyl-tRNA synthetase. *Biochemistry* **43**, 7028-7037.
- Air, G.M. and Harris, J.I (1974) DNA-Dependent RNA Polymerase From the Thermophilic Bacterium *Thermus aquaticus*. *FEBS Letters* **38**, 277-281.
- Akhmetzjanov, A. and Vakhitov, V. (1992) Molecular cloning and nucleotide sequence of the DNA polymerase gene from *Thermus flavus*. *Nucleic Acids Research* **20**, 5839.
- Akopyanz, N., Bukanov, N., Westblom, T. and Berg, D. (1992) PCR-based RFLP analysis of DNA sequence diversity in the gastric pathogen *Helicobacter pylori*. *Nucleic Acids Research* **20**, 6221-6225.
- Alba, M. (2001) Replicative DNA Polymerases. *Genome Biology* **2**, 1-4.
- Al Amir, A., Senok, A., Ismaeel, A., Al-Mahmeed, A. and Botta, G. (2007) Multiplex PCR for direct detection of *Campylobacter* spp. in human and chicken stools. *Journal of Medical Microbiology* **10**, 1350-1355.
- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J. (1994) *Molecular Biology of the Cell*, 3rd edition, Garland Publishing, New York.
- Al-Soud, W. And Radstrom, P. (1998) Capacity of nine thermostable DNA polymerases to mediate DNA amplification in the presence of PCR-inhibiting samples. *Applied and Environmental Microbiology* **64**, 3748-3753.
- Al-Soud, W. and Radstrom, P. (2000) Effects of amplification facilitators on diagnostic PCR in the presence of blood, feces and meat. *Journal of Clinical Microbiology* **38**, 4463-4470.
- Al-Soud, W. and Radstrom, P. (2001) Purification and characterization of PCR-inhibitory components in blood cells. *Journal of Clinical Microbiology* **39**, 485-493.
- Anderson, R. (1914) Concerning the organic phosphoric acid of cottonseed meal II. *Journal of Biological Chemistry* **17**, 141-150.
- Angel, R., Taminm, N., Applegate, T., Dhandu, A. and Ellestad, L. (2002) Phytic acid chemistry: influence on phytin-phosphorus availability and phytase efficacy. *Journal of Applied Poultry Research* **11**, 471-480.
- Arnold, F., Wintrode, P., Miyazaki, K. and Gershenson, A. (2001) How enzymes adapt: lessons from directed evolution. *TRENDS in Biochemical Sciences* **26**, 100-106.
- Barken, K., Haagensen, J. and Tolker-Nielsen, T. (2007) Advances in nucleic acid-based diagnostic of bacterial infections. *Clinica Chimica Acta* **384**, 1-11.
- Barnes, W. (1992) The fidelity of *taq* polymerase catalyzing PCR is improved by an N-terminal deletion. *Gene* **112**, 29-35.
- Barnes, W. (1994) PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 2216-2220.
- Barnes, W. (1995) Thermostable DNA polymerase with enhanced thermostability and enhanced length and efficiency of primer extension. US patent 5436149.

- Bechtereva, T., Pavlov, Y., Kramorov, V., Migunova, B. and Kiselev, O. (1989) DNA sequencing with thermostable Tet DNA polymerase from *Thermus thermophilus*. *Nucleic Acids Research* **17**, 10507.
- Beck, K. and Keen, J. (1991) Interaction of phosphoinositide cycle intermediates with the plasma membrane-associated clathrin assembly protein AP-2. *Journal of Biological Chemistry* **266**, 4442-4447.
- Bedell, J., Edmonson, S. and Shriver J. (2005) Role of a surface tryptophan in defining the structure, stability, and DNA binding of the hyperthermophile protein Sac7d. *Biochemistry* **44**, 915-25
- Beese, L. And Steitz, T. (1991) Structural basis for the 3-5' exonuclease activity of *Eschericia coli* DNA polymerase I: a two metal ion mechanism. *EMBO* **10**, 25-33.
- Black, M., Newcomb, t., Wilson, H. and Loeb, L. (1996) Creation of drug-specific herpes simplex virus type 1 thymidine kinase mutants for gene therapy. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3525-3529.
- Blondal, T., Thorbjarnardottir, S., Kieleczawa, J., Hjorleifsdottir, S., Kristjansson, J., Einarsson, J. and Eggertsson, G. (2001) Cloning, sequence analysis and functional characterisation of DNA polymerase I from the thermophilic eubacterium *Rhodothermus marinus*. *Biotechnology and Applied Biochemistry* **34**, 37-45.
- Bolt, A., Berry, A. and Nelson, A. (2008) Directed evolution of aldolases for exploitation in synthetic organic chemistry. *Archives of Biochemistry and Biophysics* **474**, 318-330.
- Braithwaite, D. and Ito, J. (1993) Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Research* **21**, 787-802.
- Bramuccci, M. and Nagarajan, V. (1996) Direct Selection of Cloned DNA in *Bacillus subtilis* Based on Sucrose-Induced Lethality. *Applied and Environmental Microbiology* **62**, 3948-3953.
- BRENDA: The comprehensive enzyme information system, Reaction catalyzed by DNA-directed DNA polymerase (2.7.7.7), <http://www.brenda-enzymes.info/Mol/reaction-popup.php?id=2.7.7.7>, 20.1.2009.
- Brock, T. and Freeze, H. (1969) *Thermus aquaticus* gen. n. and sp. N., a Non-sporulating extreme thermophile. *Journal of Bacteriology* **98**, 289-297.
- Burrows, J. and Goward, C. (1992) Purification and properties of DNA polymerase from *Bacillus caldotenax*. *Biochemical Journal* **287**, 971-977.
- Campbell, N., Reece, J. and Mitchell, L. (1999) *Biology*, 5th edition, Published by Benjamin Cummings, CA, USA.
- Cann, I. and Ishino, Y. (1999) Archaeal DNA replication: identifying the pieces to solve a puzzle. *Genetics* **152**, 1249-1267.
- Carballeira, N., Nazabal, M., Brito, J. and Garcia, O. (1990) Purification of a thermostable DNA polymerase from *Thermus thermophilus* HB8, useful in the polymerase chain reaction. *Biotechniques* **9**, 276-81.
- Cardin, A. and Weintraub, H. (1989) Molecular modelling of protein-glycosaminoglycan interactions. *Arterioscler. Thromb. Vasc. Biol.* **9**, 21-32.

- Castro, C., Smidansky, E., Maksimchuk, K., Arnold, J., Korneeva, V., Gotte, M., Konigsberg, W. And Cameron, C. (2007) Two proton transfers in the transition state for nucleotidyl transfer catalyzed by RNA- and DNA-dependent RNA and DNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4267-4272.
- Chen, K. and Arnold, F. (1993) Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. *Biochemistry* **90**, 5618-5622.
- Cheryan, M. (1980) Phytic acid interactions in food systems. *Critical Reviews in Food Science and Nutrition* **13**, 297-335.
- Cheryan, M., Anderson, F. And Grynspan, F. (1983) Magnesium-phytate complexes: effect of pH and molar ratio on solubility characteristics. *Cereal Chemistry* **60**, 235-237.
- Chica, R., Doucet, N. And Pelletier, J. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. *Current Opinion in Biotechnology* **16**, 378-384.
- Chien, A., Edgar, D. and Trela, J. (1976) Deoxyribonucleic Acid Polymerase from Extreme Thermophile *Thermus aquaticus*. *Journal of Bacteriology* **27**, 1550-1557.
- Cline, J., Braman, J. and Hogrefe, H. (1996) PCR fidelity of Pfu DNA polymerase and other thermostable DNA polymerases. *Nucleic Acids Research* **24**, 3546-3551.
- Cramer, A., Raillard, S., Bermudez, E. and Stemmer, W. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. *Nature* **391**, 288-291.
- Crea, P., de Robertis, A., de Stefano, C. and Sammartano, S. (2006) Speciation of phytate ion in aqueous solution. Sequestration of magnesium and calcium by phytate at different temperatures and ionic strengths, in NaCl. *Biophysical Chemistry* **124**, 18-26.
- d'Abbadie, M., Hofreiter, M., Vaisman, A., Loakes, D., Gasparutto, D., Cadet, J., Woodgate, R., Paabo, S. and Holliger, P. (2007) Molecular breeding of polymerases for amplification of ancient DNA. *Nature Biotechnology* **25**, 939-943.
- D'Amato, R., Wallman, A., Hochstein, L., Colaninno, P., Scardamaglia, M., Ardila, E., Ghouri, M., Kim, K., Patel, R. and Miller, A. (1995) Rapid Diagnosis of Pulmonary Tuberculosis by using Roche AMPLICOR *Mycobacterium tuberculosis* PCR test. *Journal of Clinical Microbiology* **33**, 1832-1834.
- Datta, K. and LiCata, V. (2003) Thermodynamics of the binding of *Thermus aquaticus* DNA polymerase to primed-template DNA. *Nucleic Acids Research* **31**, 590-5597.
- Daugherty, P., Chen, G., Iverson, B. and Georgiou, G. (1999) Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies. *Proc. Natl. Acad. Sci. USA.* **97**, 2029-2034.
- Davidson, J., Fox, R., Harris, D., Lyons-Abbott, S. and Loeb, L. (2003) Insertion of the T3 DNA polymerase thioredoxin binding domain enhances the processivity and fidelity of Taq DNA polymerase. *Nucleic Acids Research*, **31**, 4702-4709.
- Demirjian, D., Moris-Varas, F. and Cassidy, C. (2001) Enzymes from extremophiles. *Current Opinion in Chemical Biology* **5**, 144-151.

- Derbyshire, V., Freemont, P., Sanderson, M., Beese, L., Friedman, J., Joyce, C. and Steitz, T. (1988) Genetic and crystallographic studies of the 3-5'-exonucleolytic site of DNA polymerase I. *Science* **240**, 199-201.
- DiCioccio, R. and Srivastava, S. (1978) Inhibition of deoxynucleotide-polymerizing enzyme activities of human cells and of simian sarcoma virus by heparin. *Cancer Research* **38**, 2401-2407.
- Doublet, S., Tabor, S., Long, A., Richardson, C. and Ellenberger, T. (1998) Crystal structure of bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature* **391**, 251-8.
- Dutka-Malen, S., Evers, S. and Courvalin, P. (1995) Detection of glycopeptide resistance genotypes and identification to the species level of clinically relevant enterococci by PCR. *Journal of Clinical Microbiology* **33**, 24-27.
- Eom, S., Wang, J. and Steitz, T. (1996) Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature* **382**, 278-281.
- Evans, W. and Pierce, A. (1982) Interaction of phytic acid with the metal ions copper (II), cobalt (II), iron (II), magnesium (II) and manganese (II). *Journal of Food Science* **47**, 1014-1015.
- Fleischmann, R. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* **269**, 496.
- Friedberg, E. (2006) The eureka enzyme: the discovery of DNA polymerase. *Nature Reviews Molecular Cellular Biology* **7**, 143-147.
- Freeze, H. and Brock, T.D. (1970) Thermostable Aldolase from *Thermus aquaticus*. *Journal of Bacteriology* **101**, 541-550.
- Frey, B. and Suppmann, B. (1995) Demonstration of the Expand™ PCR system's greater fidelity and higher yields with a lacI-based PCR fidelity assay. *Biochemica* **2**, 8-9.
- Fromant, M., Blanquet, S. and Plateau, P. (1995) Direct random mutagenesis of gene-sized DNA fragments using polymerase chain reaction. *Analytical Biochemistry* **224**, 347-353.
- Ghadessy, F.J., Ong, J.L. and Holliger, P. (2001) Directed evolution of polymerase function by compartmentalized self-replication. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4552-4557.
- Glahn, R., Wortley, G., South, P. and Miller, D. (2002) Inhibition of iron uptake by phytic acid, tannic acid, and ZnCl₂: studies using an in vitro digestion/Caco-2 cell model. *Journal of Agricultural and Food Chemistry* **50**, 390-395.
- Graf, E. (1983) Applications of phytic acid. *Journal of the American Oil Chemistry Society* **60**, 1861-1868.
- Griffiths, A. and Tawfik, D. (2006) Miniaturising the laboratory in emulsion droplets. *TRENDS in Biotechnology* **24**, 395-402.
- Guatelli, J., Gingeras, T. and Richman, D. (1989) Nucleic acid amplification in vitro: detection of sequences with low copy numbers and application to diagnosis of human immunodeficiency virus type 1 infection. *Clinical Microbiology Reviews* **2**, 217-226.
- Hawkins, P., Poyner, D., Jackson, T., Letcher, A., Lander, D. and Irvine, R. (1993) Inhibition of iron-catalysed hydroxyl radical formation by inositol polyphosphates: a

- possible physiological function for myo-inositol hexakisphosphate. *Biochemistry Journal* **294**, 929-934.
- Hayashi, N., Welschhof, M., Zewe, M., Braunagel, M., Dubel, S., Breitling, F., and Little, M. (1994) Simultaneous mutagenesis of antibody CDR regions by overlap extension and PCR. *Biotechniques* **17**, 310-315.
- He, Z. and Honeycutt, C. (2001) Enzymatic characteristics of organic phosphorus in animal manure. *Journal of Environmental Quality* **30**, 1685-1692.
- He, Z., Honeycutt, C., Zhang, T. and Bertsch, P. (2006) Preparation and FT-IR characterisation of metal phytate compounds *Journal of Environmental Quality* **35**, 1319-1328.
- Henry, A. and Romesberg, F. (2005) The evolution of DNA polymerases with novel activities. *Current Opinion in Biotechnology* **16**, 370-377.
- Higuchi, R., Dollinger, G., Walsh, P. and Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. *Biotechnology* **10**, 413-417.
- Hirsch, J., Warkentin, T., Shaughnessy, S., Anand, S., Halperin, J., Raschke, R., Granger, C., Ohman, E. and Dalen, J. (2001) Heparin and Low-Molecular-Weight Heparin. *CHEST*, **119**, 645-945.
- Holmberg, R., Henry, A. and Romesberg, F. (2005) Directed evolution of novel polymerases. *Biomolecular Engineering* **22**, 39-49.
- Holodniy, M., Kim, S., Katzenstein, D., Konrad, M., Groves, E. and Merigan, T. (1991) Inhibition of Human Immunodeficiency Virus Gene Amplification by Heparin. *Journal of Clinical Microbiology* **29**, 676-679.
- Hough, D. and Danson, M. (1999) Extremozymes. *Current Opinion in Chemical Biology* **3**, 39-46.
- Innis, M., Myambo, K., Gelfand, D. and Brow, M. (1988) DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9436-9440.
- Ito, J. and Braithwaite, D. (1991) Compilation and alignment of DNA polymerase sequences. *Nucleic Acids Research* **19**, 4045-4057.
- Jae, K., Lee, H., Bae, S., Jeon, J., Lim, J., Cho, Y., Nam, K., Kang, S., Kim, S., Kwon, S. and Lee, J. (2007) Cloning, purification, and characterisation of a new DNA polymerase from a hyperthermophilic archaeon, *Thermococcus* sp. NA1. *Journal of Microbiology and Biotechnology* **17**, 1090-1097.
- Jacobson, M., Aspan, A., Konigsson, M., Segerstad, C., Wallgren, P., Fellstrom, C., Jensen-Waern, M. and Gunnarson, A. (2004) Routine diagnosis of *Lawsonia intracellularis* performed by PCR, serological and post mortem examination, with special emphasis on sample preparation methods for PCR. *Veterinary Microbiology* **102**, 189-201.

- Johannes, T., Woodyer, R. and Zhao, H. (2005) Directed evolution of a thermostable phosphate dehydrogenase for NAD(P)H regeneration. *Applied and Environmental Microbiology* **71**, 5728-5734.
- Johnson, L. and Tate, M. (1969) Structure of 'phytic acids'. *Canadian Journal of Chemistry* **47**, 63-73.
- Jones, L., Yazzie, B. and Middaugh, R. (2004) Polyanions and the proteome. *Molecular and Cellular Proteomics* **3**, 746-69.
- Jones, M. and Foulkes, N. (1989) Reverse transcription of mRNA by *Thermus aquaticus* DNA polymerase. *Nucleic Acids Research* **17**, 8387.
- Jordan, R., Beeler, D. and Rosenberg, R. (1979) Fractionation of low molecular weight heparin species and their interaction with antithrombin. *Journal of Biological Chemistry* **254**, 2902-2913.
- Joung, H., Jeun, B., Li, S., Kim, J., Woodhouse, L. King, J., Welch, R. and Paik, H. (2007) Fecal phytate excretion varies with dietary phytate and age in women. *Journal of the American College of Nutrition* **26**, 295-302.
- Joyce, C., Kelley, W. and Grindley, N. (1982) Nucleotide sequence of the *Escherichia coli* *polA* gene and primary structure of DNA polymerase I. *The Journal of Biological Chemistry* **257**, 1958-1964.
- Kabir, S. (2004) Detection of *Helicobacter pylori* DNA in feces and saliva by polymerase chain reaction. *Helicobacter* **9**, 115-123.
- Keohavong, P. and Thilly, W. (1989) Fidelity of DNA polymerases in DNA amplification. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9253-9257.
- Kermekchiev, M., Kirilova, L., Vail, E. and Barnes, W. (2009) Mutants of *Taq* DNA polymerase resistant to PCR inhibitors allow DNA amplification from whole blood and crude soil samples. *Nucleic Acids Research* **37**, e40.
- Kim, J., Koh, S., Kim, J., Kwon, S. and Lee, D. (1998) Top DNA polymerase from *Thermus thermophilus* HB27: gene cloning, sequenc determination, and physiochemical properties. *Molecules and Cells* **8**, 157-161.
- Kim, Y., Eom, S., Wang, J., Lee, D., Suh, S. and Steitz, T. (1995) Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature* **376**, 612-616.
- Klenow, H. and Henningsen, I. (1970) Selective elimination of the exonuclease activity of deoxyribonucleic acid polymerase from *Escherichia coli* B by limited proteolysis. *Proc. Nat. Acad. Sci. U.S.A.* **65**, 168-175.
- Ko, T, Chu, H., Chen, C., Chou, C. and Wang, A. (2004) Structures of the hyperthermophilic chromosomal protein Sac7d in complex with DNA decamers. *Acta Crystallography D Biol.* **60**, 1381-1387.
- Kong, H., Kucera, R. and Jack, W. (1993) Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litorallis*. Vent DNA polymerase, steady state

kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *Journal of Biological Chemistry* **268**, 1965-1975.

Kornberg, A., Lehman, I. and Simms, E. (1956) Polydesoxyribonucleotide synthesis by enzymes from *Escherichia coli*. *Federation Proceedings* **15**, 291.

Kornberg, A. and Geftter, M. (1971) Purification and DNA synthesis in cell-free extracts: properties of DNA polymerase II. *Proc. Nat. Acad. Sci. U.S.A.* **68**, 761-764.

Korolev, S., Nayal, M., Wayne, B., Di Cera, E. And Waksman, G. (1995) Crystal structure of the large fragment of thermus aquaticus DNA polymerase I at 2.5-Å resolution: structural basis for thermostability. *Biochemistry* **92**, 9264-9268.

Kreader, C. (1996) Relief of amplification inhibition in PCR with bovine serum albumin or T4 gene 32 protein. *Applied and Environmental Microbiology* **62**, 1102-1106.

Lawyer, F., Stoffel, S., Saiki, R., Myambo, K., Drummond, R. and Gelfand, D. (1989) Isolation, characterization and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *Journal of Biological Chemistry* **264**, 6427-6437.

Lawyer, F., Stoffel, S., Saiki, R., Chang, S., Landre, P., Abramson, R. and Gelfand, D. (1993) High-level expression, purification, and enzymatic characterization of full-length *Thermus aquaticus* DNA polymerase and a truncated form deficient in 5' to 3' exonuclease activity. *PCR Methods and Applications* **2**, 275-287.

Layrisse, M., Garcia-Casal, M., Solano, L., Baron, M., Arguello, F., Llovera, D., Ramirez, J., Leets, I. and Tropper, E. (2000) Iron bioavailability in humans from breakfast enriched with iron bis-glycine chelate, phytates and polyphenols. *Journal of Nutrition* **130**, 2195-2199.

Lehman, I., Bessman, M., Simms, E. and Kornberg, A. (1958) Enzymatic synthesis of deoxyribonucleic acid. *Journal of Biological Chemistry* **233**, 163-170.

Leung, D., Chen, E. and Goeddal, D. (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. *Technique* **1**, 11-15.

Li, Y., Korolev, S. and Waksman, G. (1998) Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO* **17**, 7514-7525.

Li, W., Johnson, D., Esmon, C. and Huntington, J. (2004) Structure of the antithrombin-thrombin-heparin ternary complex reveals the antithrombotic mechanism of heparin. *Nature Structural Molecular Biology* **11**, 857-862.

Lim, D., Serguei, G., Forsberg, C. and Jia, Z. (2000) Crystal structures of *Escherichia coli* phytase and its complex with phytate. *Nature Structural Molecular Biology* **7**, 108-113.

Livak, K., Flood, S., Marmaro, J., Giusti, W. And Deetz, K. (1995) Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *PCR Methods and Applications* **4**, 357-362

- Lurz, R., Grote, M., Dijk, J., Reinhardt, R. and Dobrinski. (1986) Electron microscopic study of DNA complexes with proteins from Archaeobacterium *Sulfolobus acidocaldarius*. *EMBO* **5**, 3715-3721.
- Luttrell, B. (1993) The biological relevance of the binding of calcium ions by inositol phosphates. *Journal of Biological Chemistry* **268**, 1521-1524.
- Macelroy, R. (1974) Some comments on the evolution of extremophiles. *Biosystems*, **6**, 74-75.
- Mackay, I. (2004) Real-time PCR in the microbiology laboratory. *Clinical Microbiology and Infectious Diseases* **10**, 190-212.
- Maenz, D., Engele-Schaan, C., Newkirk, R. and Classen, H. (1999) The effect of minerals and mineral chelators on the formation of phytase-resistant and phytase-susceptible forms of phytic acid in solution and in slurry of canola meal. *Animal Food Science and Technology*, **81**, 177-192.
- Makino, S., Cheun, H., Watarai, M., Uchida, I. and Takeshi, K. (2001) Detection of anthrax spores from the air by real-time PCR. *Letter in Applied Microbiology* **33**, 237-240.
- Margulies, M., Egholm, M., Altman, W., Attiya, S., Bader, J., Bemben, L., Berka, J., Braverman, M., Chen, Y., Chen, Z., Dewell, S., Du, L., Fierro, J., Gomes, X., Godwin, B., He, W., Helgesen, S., Ho, C., Irzyk, G., Jando, S., Alenquer, M., Jarvie, J., Jirage, B., Kim, J., Knight, J., Lanza, J., Leamon, J., Lefkowitz, S., Lei, M., Li, J., Lohman, K., Lu, H., Makhijani, V., McDade, K., McKenna, M., Myers, E., Nickerson, E., Nobile, J., Plant, R., Puc, B., Ronan, M., Roth, G., Tomasz, A., Vogt, K., Volkmer, G., Wang, S., Wang, Y., Weiner, M., Yu, P., Begley, R. and Rothberg, J. (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376-80.
- McAfee, J.G., Edmondson, S.P., Datta, P.K., Shriver, J.W. and Gupta, R. (1995) Gene Cloning, Expression, and Characterisation of the Sac7 Proteins from the Hyperthermophile *Sulfolobus acidocaldarius*. *Biochemistry* **34**, 10063-10077.
- Miyazaki, K. and Arnold, F. (1999) Exploring nonnatural evolutionary pathways by saturation mutagenesis: rapid improvement of protein function. *Journal of Molecular Evolution* **49**, 716-720.
- Monstein, H. and Ellnebo-Svedlund. (2002) Molecular typing of *Helicobacter pylori* by virulence-gene based multiplex PCR and RT-PCR analysis. *Helicobacter* **7**, 287-96.
- Moore, J., Jin, H., Kuchner, O. and Arnold, F. (1997) Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. *Journal of Molecular Biology* **272**, 336-347.
- Mulloy, B., Forster, M., Jones, C., Davies, D. (1993) N.m.r. and molecular-modelling studies of the solution conformation of heparin. *Biochemical.Journal* **293**, 849-858
- Mulloy, B. and Forster, M. (2000) Conformation and dynamics of heparin and haparan sulphate. *Glycobiology* **10**, 1147-1156.

Myers, T. and Gelfand, D. (1991) Reverse transcription and DNA amplification by *Thermus thermophilus* DNA polymerase. *Biochemistry* **30**, 7661-7666.

Ong, J., Loakes, D., Jaroslowski, S., Too, K. and Holliger, P. (2006) Directed evolution of DNA polymerase, RNA polymerase and reverse transcriptase activity in a single polypeptide. *Journal of Molecular Biology* **361**, 537-550.

Patel, P., Kawate, H., Adman, E., Ashbach, M. and Loeb, L. (2001) A single highly mutable catalytic site amino acid is critical for DNA polymerase fidelity. *The Journal of Biological Chemistry* **276**, 5044-5051.

Patent Lens. Directed Evolution Method Patent Application WO 2002/022869.
http://www.patentlens.net/patentlens/quick.cgi?patnum=WO_2002/022869_A2#show.
3.2.2009

Pavlov, A., Belova, G., Kozayavkin, S. and Slesarev, A. (2002) Helix-hairpin-helix motifs confer salt resistance and processivity on chimeric DNA polymerases. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 13510-13515.

Pavlov, A., Pavlova, N., Kozayavkin, S. and Slesarev, A. (2004) Recent developments in the optimisation of thermostable DNA polymerases for efficient applications. *TRENDS in Biotechnology* **22**, 253-260.

Perozzo, R., Folkers, G. and Scapozza, L. (2004) Thermodynamics of protein-ligand interactions: history, presence, and future aspects. *Journal of Receptor and Signal Transduction Research* **24**, 1-52

Pisani, F., De Martino, C. and Rossi, M. (1992) A DNA polymerase from the archaeon *Sulfolobus solfataricus* shows sequence similarity to family B DNA polymerases. *Nucleic Acids Research* **20**, 2711-2716.

Pisani, F., De Felice, M., Manco, G. and Rossi, M. (1998) Domain organisation and biochemical features of *Sulfolobus solfataricus* DNA polymerase. *Extremophiles* **2**, 171-177.

Poddar, S., Sawyer, M., and Connor, J. (1998) Effect of inhibitors in clinical specimens on Taq and Tth DNA polymerase-based PCR amplification of influenza A virus. *Journal of Medical Microbiology* **47**, 1131-1135.

Protein Data Base (PDB). N.M.R. and Molecular-Modelling Studies of the Solution Conformation of Heparin. Mulloy, B and Forester, M.
<http://www.rcsb.org/pdb/explore.do?structureId=1HPN>. 3.2.2009

Raboy, V. (2007) The ABCs of low-phytate crops. *Nature Biotechnology* **25**, 874-875.

Rajeev, S., Zhang, Y., Sreevastan, S., Motiwala, A. and Byrum, B. (2005) Evaluation of multiple genomic targets for identification and confirmation of *Mycobacterium avium* subsp. *Paratuberculosis* isolates using real-time PCR. *Veterinary Microbiology* **105**, 215-221.

Rechkunova, N., Akishev, A., Lebedeva, N., Khodyreva, S., Degtyarev, S. and Lavrik, O. (1998) Thermostable DNA-polymerase from *Thermus thermophilus* B35: isolation and characterisation of some properties. *Biochemistry (Mosc.)* **63**, 1266-1270.

- Recorbet, G., Robert, C., Givaudan, A., Kudla, B., Normand, P. and Faurie, G. (1993) Conditional Suicide System of *Escherichia coli* Released into Soil That Uses the *Bacillus subtilis* sacB Gene. *Applied and Environmental Microbiology* **59**, 1361-1366.
- Reddy, M., Hurrell, R., Juillerat, M. and Cook, J. (1996) The influence of different protein sources on phytate inhibition of nonheme-iron absorption in humans. *Journal of Clinical Nutrition* **63**, 203-207.
- Rothschild, L. and Mancinelli, R. (2001) Life in extreme environments. *Nature* **409**, 1092-1101.
- Saiki, R., Scharf, S., Faloona, F., Mullis, K., Horn, G., Erlich, H. and Arnheim, N. (1985) Enzymatic Amplification of β -Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science* **230**, 1350-1354.
- Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. And Erlich, H. (1988) Primer-directed Enzymatic Amplification of DNA with a Thermostable DNA polymerase. *Science* **239**, 487-489.
- Sambrook, J. and Russell, D. (2001) Molecular Cloning: a laboratory manual. 3rd edition, Published by CSHL, New York.
- Sanger, F., Nicklen, S. and Coulson, A. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A* **74**, 5463-5467.
- Schweizer, H.P. and Hoang, T.T. (1995) An improved system for gene replacement and *xylE* fusion analysis in *Pseudomonas aeruginosa*. *Gene* **158**, 15-22.
- Sellmann, E., Schroder, K., Knoblich, I. and Westermann, P. (1992) Purification and characterisation of DNA polymerases from *Bacillus* species. *Journal of Bacteriology* **174**, 4350-4355.
- Shamsuddin, A. (2008) Demonizing Phytate. *Nature Biotechnology* **26**, 496-497.
- Shandilya, H., Griffiths, K., Flynn, E., Astatke, M., Shih, P., Lee, J., Gerard, G., Gibbs, M. and Berquist, P. (2004) Thermophilic bacterial DNA polymerases with reverse-transcriptase activity. *Extremophiles* **8**, 243-251.
- Shears, S. (2001) Assessing the omnipotence of inositol hexakisphosphate. *Cellular Signalling* **13**, 151-158.
- Shin, C., Chang, C. and Kim, J. (2006) Liquid heparin anticoagulant produces more negative bias in the determination of ionized magnesium than ionized calcium. *Yonsei Medical Journal* **47**, 191-195.
- Steitz, T., (1999) DNA polymerases: structural diversity and common mechanisms. *The Journal of Biological Chemistry* **274**, 17395-17398.
- Stemmer, W. (1994) DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10747-10751.

- Tabor, S. and Ricahrdson, C. (1987) DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 4767-4771.
- Tabor, S. and Ricahrdson, C. (1995) A single reside in DNA polymerase of the *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxyribonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6339-6343.
- Takagi, M., Nishioka, M., Kakihara, H., Kitabayashi, M., Inoue, H., Kawakami, B., Oka, M. and Imanaka, T. (1997) Characterisation of DNA polymerase from *Pyrococcus* sp. strain KOD1 and its application to PCR. *Applied and Environmental Microbiology* **63**, 4504-4510.
- Tao, H. and Cornish, V. (2002) Milestones in directed enzyme evolution. *Current opinion in Chemical Biology* **6**, 858-864.
- Tawfik, D. And Griffiths, A. (1998) Man-made cell-like compartments for molecular evolution. *Nature Biotechnology* **16**, 652-656.
- Taylor, A. (1997). Titration of heparinise for removal of the PCR-inhibitory effect of heparin in DNA samples. *MolecularEcology* **6**, 383-385.
- Thornton, C. And Passen, S. (2004). Inhibition of PCR amplification by phytic acid, and treatment of bovine fecal specimens with phytase to reduce inhibition. *Journal of Microbiological Methods* **59**, 43-52.
- Tones, M., Bootman, M., Higgins, B., Lane, D., Pay, G. and Lindahl, U. (1989) The effect of heparin on the inositol 1,4, 5-triphosphate reseptor in rat liver microsomes. *FEBS* **252**, 105-108.
- Uemori, T., Ishino, Y., Toh, H., Asada, K. and Kato, I. (1993) Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Research* **21**, 259-65.
- van der Merwe, P. (2001) Surface Plasmon Resonance in Protein-Ligand interactions: hydrodynamics and calorimetry, edited by S Harding and P Z Chowdhry. Practical Approach series, Oxford University Press, 137-170.
- Vichier-Guerre, S., Ferris, S., Auberger, N., Mahiddine, K. and Jestin, J. (2006) A population of thermostable reverse transcriptases evolved from thermus aquaticus DNA polymerase I by phage display. *Angew. Chem. Int. Ed.* **45**, 6133-6137.
- Villbrandt, B., Sobek, H., Frey, B. and Schomburg, D. (2000) Domain exchange: chimeras of *Thermus aquaticus* DNA polymerase, *Escherichia coli* DNA polymerase I and *Thermotoga neapolitana* DNA polymerase. *Protein Engineering* **13**, 645-654.
- Vucenik, I., Passaniti, A., Vitolo, M., Tantivejkul, K., Eggleton, P. and Shamsuddin, A. (2004) Anti-angiogenic activity of inositol hexaphosphate (IP₆). *Carcinogenesis* **25**, 2115-2123.
- Wang, Y., Prosen, D., Mei, L., Sullivan, J., Finney, M. and Vander Horn, P. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Research* **32**, 1197-1207.

- Watson, R. and Blackwell, B. (2000) Purification and characterisation of a common soil component which inhibits the polymerase chain reaction. *Canadian Journal of Microbiology* **46**, 633-642.
- Whitfield, D. and Sarkar, B. (1992) Heavy metal binding to heparin disaccharides. II First evidence for zinc chelation. *Biopolymers* **6**, 597-619.
- Wilson, I. (1997). Inhibition and facilitation of nucleic acid amplification. *Applied and Environmental Microbiology* **63**, 3741-3751.
- Wodzinski, R. and Ullah, A. (1996) Phytase. *Advances in Applied Microbiology* **42**, 263-302.
- Xia, G., Chen, L., Sera, T., Fa, M., Schultz, P. And Romesberg, F. (2002) Directed evolution of novel polymerase activities: mutation of a DNA polymerase into an efficient RNA polymerase. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 6597-6602.
- Yang, S. and Rothman, R. (2004). PCR-based diagnostics for infectious diseases: uses, limitations, and future applications in acute-care settings. *The Lancet Infectious Diseases* **4**, 337-348.
- Yokota, M., Tatsumi, N., Nathalang, O., Yamada, T. And Tsuda, I. (1999). Effects of heparin on polymerase chain reaction for blood white cells. *Journal of Clinical Laboratory Analysis* **13**, 133-140.
- You, L. and Arnold, F. (1994) Directed evolution of subtilisin E in *Bacillus subtilis* to enhance total activity in aqueous dimethylformamide. *Protein Engineering* **9**, 77-83.
- Zhao, H., Giver, L., Shao, Z., Affholter, J. and Arnold, F. (1998) Molecular evolution by staggered extension process (StEP) in vitro recombination. *Nature Biotechnology* **16**, 258-261.
- Zipper, H., Buta, C., Lammle, K., Brunner, H., Bernhagen, J. and Vitzthum, F. (2003) Mechanisms underlying the impact of humic acids on DNA quantification by SYBR Green I and consequences for the analysis of soils and aquatic sediments. *Nucleic Acids Research* **31**, e39.

APPENDIX A

Translation of the DNA sequence of Sac7-TaqPolymerase:

```
1 atggtgaaagtgaaatTTAAATACAAAGGTGAAGAGAAAGAAGTTGACACTTCTAAAATT
1 M V K V K F K Y K G E E K E V D T S K I
61 aaaaaagtgtggcgtgtgggtaaaatggtttctttcacctatgatgataacggcaaaacc
21 K K V W R V G K M V S F T Y D D N G K T
121 ggtcgcggtgcggtgtctgagaagacgcgccgaaagaactgctggatagctggcgcgt
41 G R G A V S E K D A P K E L L D M L A R
181 gcggaacgcgagaaaaaagggtggtgttaccatgaattcggggatgctgcccctctttgag
61 A E R E K K G G V T M N S G M L P L F E
241 cccaaggcgggtcctcctggtggacggccaccacctggcctaccgcaccttccacgcc
81 P K G R V L L V D G H H L A Y R T F H A
301 ctgaaggcctcaccaccagccgggggagccggtgcagggcgtctacggcttcgccaag
101 L K G L T T S R G E P V Q A V Y G F A K
361 agcctcctcaaggccctcaaggaggacggggacgcggtgatcgtggtctttgacgccaag
121 S L L K A L K E D G D A V I V V F D A K
421 gccccctccttcgccaagcagggcctacgggggtacaaggcgggcccggccccacgccg
141 A P S F R H E A Y G G Y K A G R A P T P
481 gaggactttccccggcaactcgcctcatcaaggagctggtggacctctggggctggcg
161 E D F P R Q L A L I K E L V D L L G L A
541 cgctcagaggtcccgggctacgaggggacgacgtcctggccagcctggccaagaaggcg
181 R L E V P G Y E A D D V L A S L A K K A
601 gaaaaggagggtacgaggtccgcacctcaccgccgacaaagacctttaccagctcctt
201 E K E G Y E V R I L T A D K D L Y Q L L
661 tccgaccgatccacgtcctccaccccaggggtacctcatcaccggcctggctttgg
221 S D R I H V L H P E G Y L I T P A W L W
721 gaaaagtacggcctgagcccagaccagtgggccgactaccgggcccctgaccgggacgag
241 E K Y G L R P D Q W A D Y R A L T G D E
781 tccgacaaccttcccgggtcaaggcatcgggggagaagacggcgaggaagcttctggag
261 S D N L P G V K G I G E K T A R K L L E
841 gagtgggggagcctggaagccctcctcaagaacctggaccggctgaagcccgccatccgg
281 E W G S L E A L L K N L D R L K P A I R
901 gagaagatcctggcccacatggacgatctgaagctctcctgggacctggccaaggtgctc
301 E K I L A H M D D L K L S W D L A K V R
961 accgacctgcccctggagggtgacttcgcaaaaaggcgggagcccgaccgggagaggctt
321 T D L P L E V D F A K R R E P D R E R L
1021 agggcctttctggagagatctgagtttggcagcctcctccacgagttcggccttctggaa
341 R A F L E R S E F G S L L H E F G L L E
1081 agccccaaaggcctggaggaggccccctggccccgcgggaaggggcttctggtggcttt
361 S P K A L E E A P W P P P E G A F V G F
1141 gtgctttcccgaaggagcccatgtgggcccgatcttctggccctggccgccagggg
381 V L S R K E P M W A D L L A L A A A R G
1201 ggccgggtccaccgggcccccgagccttataaagccctcagggacctgaaggaggcggcg
401 G R V H R A P E P Y K A L R D L K E A R
1261 gggcttctcgcaaaagacctgagcgttctggccctgaggggaaggccttggcctccccgcc
```

421 G L L A K D L S V L A L R E G L G L P P
1321 ggcgacgaccccatgctcctcgccctacctcctggacccttccaacaccaccccgaggg
441 G D D P M L L A Y L L D P S N T T P E G

1381 gtggcccgcgctacggcgggagtgacggaggagggcggggagcgggcccctttcc
461 V A R R Y G G E W T E E A G E R A A L S

← Palm

1441 gagaggctcttcgccaacctgtgggggagccttgagggggaggagagcctctttggctt
481 E R L F A N L W G R L E G E E R L L W L

Palm →

1501 taccgggaggtggagaggccctttccgctgtcctggccacatggaggccacgggggtg
501 Y R E V E R P L S A V L A H M E A T G V

← Thumb

1561 cgcctggacgtggcctatctcaggccttgcctggaggtggccgaggagatcgcccg
521 R L D V A Y L R A L S L E V A E E I A R

1621 ctgaggccgaggtcttccgctggccggccacccttcaacctcaactcccgggaccag
541 L E A E V F R L A G H P F N L N S R D Q

1681 ctgaaaggtcctctttgacgagctaggccttcccgccatcggaagacggagaagacc
561 L E R V L F D E L G L P A I G K H H X H

1741 ggcaagcgtccaccagcggcgcgtcctggaggcctccgagggccaccccatcggtg
581 G K R S H S A A V L E A L R E A H P I V

1801 gagaagatcctgcagtaccgggagctcaccaagctgaagagcacctacattgacccttg
601 E K I L Q Y H E L T X L K S H Y I D P L

Thumb → ← Palm

1861 ccggacctcatccacccagcggcggcctccacaccgcttcaaccagacggccacg
621 P D L I H P R T G R L H T R F N Q T A H

Palm → ← Thumb

1921 gccacgggaggtaagtagctccgatcccaacctccagaacatccccgctccgaccccg
641 A T G R L S S S D P N L Q N L P V R T P

Thumb → ← Palm

1981 cttggcagaggatccgcccggccttcatcgccgaggaggggtggctattggcgcctg
661 L G Q R I R R A F I A E E G W L L V A L

Palm → ← Fingers

2041 gactatagccagatagagctcagggtgctggcccacctctccggcagcagaaacctgatc
681 D Y S Q H E L R V L A H L S G D E N L I

2101 cgggtcttccaggagggcgggacatccacacggagaccgacagctggatgttcggcgtc
701 R V F Q E G R D I H T E T A S W M F G V

2161 ccccgaggcgcgtggacccctgatgcccggcggccaagaccatcaacttcggggtc
721 P R E A V D P L M R S A A X H I N H C V

2221 ctctacggcatgtcggcccaccgctctccaggagctagccatcccttacgaggaggcc
741 L Y G M S A H R L S Q E L A I P Y E E A

2281 caggccttcattgagcgtactttcagagcttcccaaggtgcccctggattgagaag
761 Q A F I E R Y F Q S F P K V R A W I E K

2341 accctggaggagggcaggaggggggtacgtggagaccctcttcggccgcccgcctac
781 T L E E G R R R G Y V E T L F G R R R Y

2401 gtgccagacctagaggcccgggtgaagagcgtgcccggaggcggccagcgcgatggccttc
801 V P D L E A R V X S V R H A A E R V A F

2461 aacatgcccgtccaggcaccgcccagcctcatgaagctggctatggatgaagctcttc
821 N M P V Q G T A A D L M K L A M V K L F

Fingers → ← Palm

2521 cccaggctggaggaatggggccaggatgctcctcaggtccacgacgagctggtcctc
841 P R L E E M G A R M L L Q V H D E L V L

2581 gaggccccaaaagagagggcggaggccgtggcccggctggccaaggaggtcatggagggg

```

861 E A P K E R A E A V A R L A K E V M E G
2641 gtgtatccctggccgtgccctggaggtggaggtgggatagggaggactggctctcc
881 V Y P L A V P L E V E V G I G E D W L S
      Palm →
2701 gccaaaggagtgatagatcctctagagtcgac
901 A K E * * I L * S R X

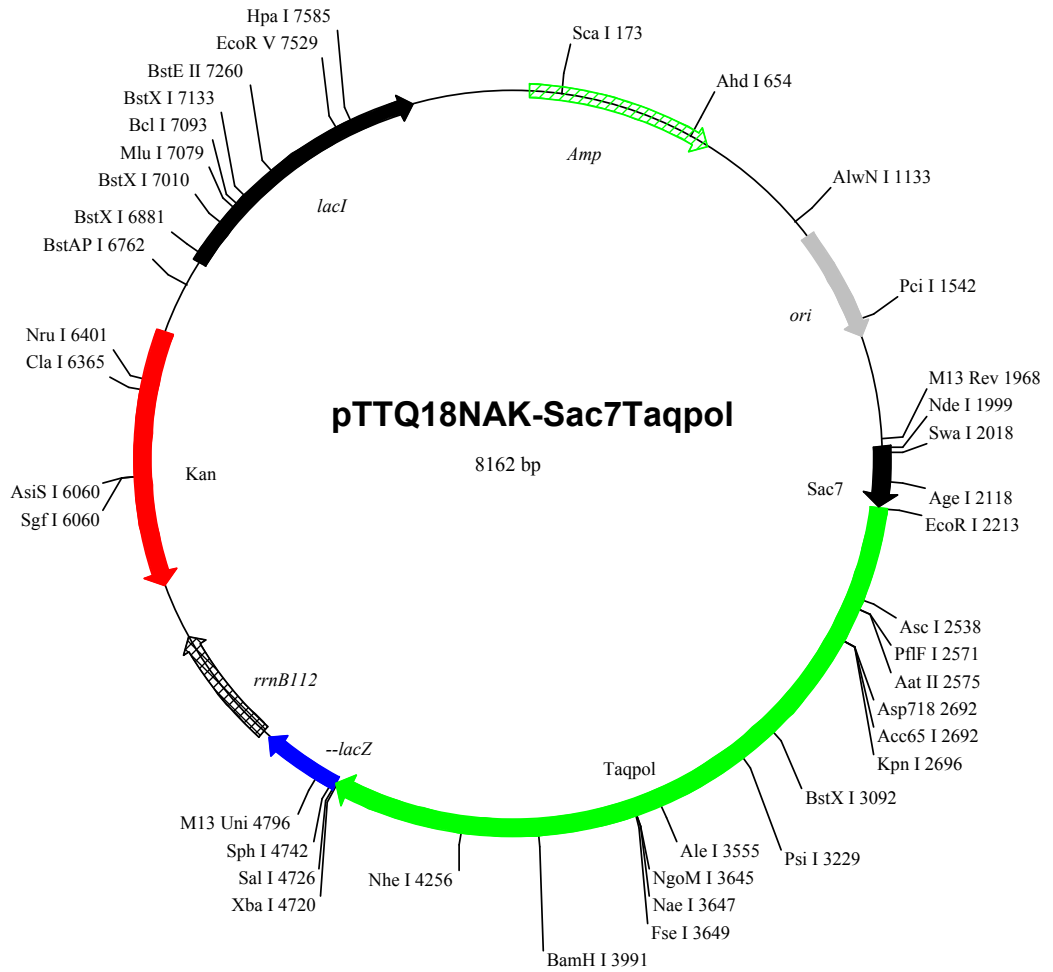
```

Residues that interact with the DNA template or dNTPs (green) and the catalytic triad (yellow) are indicated. Finger, thumb and palm sub-domains of the polymerase domain are labelled.

APPENDIX B

pTTQ18NAK-Sac7TaqPol plasmid construct

(GeneSys, Camberley, UK)



APPENDIX C

Taq DNA polymerase sequence alignment

```

TaqPol -----
Sac7-Taqpol 1 MVKVKFKYKGEEKEVDTSKIKKVVRVGMVSTFYDDNGKTGRGAVSEKDPKELLDMLAR
Klentaq -----

TaqPol -----MNSGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFKA
Sac7-Taqpol 61 AEREKKGVTMNSGMLPLFEPKGRVLLVDGHHLAYRTFHALKGLTTSRGEPVQAVYGFKA
Klentaq -----

TaqPol 51 SLLKALKEDGDAIVVFDKAPSPFRHEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLA
Sac7-Taqpol 121 SLLKALKEDGDAIVVFDKAPSPFRHEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLA
Klentaq -----

TaqPol 111 RLEVPGYEADDVSLASLAKKAEKEGYEVRILTADKDYQLLSDRIHVLHPEGYLITPAWLW
Sac7-Taqpol 181 RLEVPGYEADDVSLASLAKKAEKEGYEVRILTADKDYQLLSDRIHVLHPEGYLITPAWLW
Klentaq -----

TaqPol 171 EKYGLRPDQWADYRALTGDESNDLPGVKGIGEK TARKLLEEWGSLEALLKNLDRKPAIR
Sac7-Taqpol 241 EKYGLRPDQWADYRALTGDESNDLPGVKGIGEK TARKLLEEWGSLEALLKNLDRKPAIR
Klentaq -----

TaqPol 231 EKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLELERLEFGSLLHEFGGLE
Sac7-Taqpol 301 EKILAHMDDLKLSWDLAKVRTDLPLEVDFAKRREPDRERLRAFLELERLEFGSLLHEFGGLE
Klentaq -----MG-LLHEFGGLE
: * *****

TaqPol 291 SPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEAR
Sac7-Taqpol 361 SPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEAR
Klentaq 12 SPKALEEAPWPPPEGAFVGFVLSRKEPMWADLLALAAARGGRVHRAPEPYKALRDLKEAR
*****

TaqPol 351 GLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALS
Sac7-Taqpol 421 GLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALS
Klentaq 72 GLLAKDLSVLALREGLGLPPGDDPMLLAYLLDPSNTTPEGVARRYGGEWTEEAGERAAALS
*****

TaqPol 411 ERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEIIAR
Sac7-Taqpol 481 ERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEIIAR
Klentaq 132 ERLFANLWGRLEGEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEIIAR
*****

TaqPol 471 LEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIV
Sac7-Taqpol 541 LEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIV
Klentaq 192 LEAEVFRLAGHPFNLSRDQLERLVFDELGLPAIGKTEKTGKRSTSAAVLEALREAHPIV
*****

TaqPol 531 EKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNI PVRTP
Sac7-Taqpol 601 EKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNI PVRTP
Klentaq 252 EKILQYRELTKLKSTYIDPLPDLIHPRTGRLHTRFNQTATATGRLSSSDPNLQNI PVRTP
*****

TaqPol 591 LGQRIRRAFIAEEGWLLVALDYSQIELRVL AHLSGDENLIRVFQEGRDIHTETASWMMFGV
Sac7-Taqpol 661 LGQRIRRAFIAEEGWLLVALDYSQIELRVL AHLSGDENLIRVFQEGRDIHTETASWMMFGV
Klentaq 312 LGQRIRRAFIAEEGWLLVALDYSQIELRVL AHLSGDENLIRVFQEGRDIHTETASWMMFGV
*****

TaqPol 651 PREAVDPLMRRRAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQSFPKVRAWIEK
Sac7-Taqpol 721 PREAVDPLMRRRAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQSFPKVRAWIEK
Klentaq 372 PREAVDPLMRRRAKTINFGVLYGMSAHRLSQELAI PYEEAQAFIERYFQSFPKVRAWIEK
*****

TaqPol 711 TLEEGRRRGYVETLFGRRRYVPDLEARVKS VREAAERMAFNMPVQGTAADLMKLAMVKLF
Sac7-Taqpol 781 TLEEGRRRGYVETLFGRRRYVPDLEARVKS VREAAERMAFNMPVQGTAADLMKLAMVKLF
Klentaq 432 TLEEGRRRGYVETLFGRRRYVPDLEARVKS VREAAERMAFNMPVQGTAADLMKLAMVKLF
*****

TaqPol 771 PRLEEMGARMLLQVHDELVL EAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGI GEDWLS
Sac7-Taqpol 841 PRLEEMGARMLLQVHDELVL EAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGI GEDWLS
Klentaq 492 PRLEEMGARMLLQVHDELVL EAPKERAEAVARLAKEVMEGVYPLAVPLEVEVGI GEDWLS
*****

TaqPol 831 AKEILSRX
Sac7-Taqpol 901 AKEILSRX
Klentaq 552 AKEX----
***

```

Sequence alignment of Sac7-TaqPolymerase, *Taq* DNA polymerase, and Klentaq (the large fragment of *Taq*, which comprises the polymerase domain). * - single, fully conserved residue, - conservation of strong groups, . - conservation of weak groups.

Appendix D

1kb DNA ladder

(New England Biolabs, Hitchin, UK)

