THE APPLICATION OF MUTAGENESIS TO EXPLORE THE ROLE OF PRIMARY STRUCTURE IN THE FUNCTION OF A RANGE OF ENZYMES IN *E. coli*

Thesis Submitted to the University of Sheffield for the degree of Doctor of Philosophy

By

Mohammed Jasim Mohammed Al-Shamarti

Department of Molecular Biology and Biotechnology
The University of Sheffield
England, United Kingdom
July 2016

This Thesis is dedicated to my little sweet angel "Malak"

Acknowledgements

First of all, I am so grateful of Professor David Hornby who supervised this work, and I present my warm regards to him for providing precious advice and putting me on the right direction to fulfil this work.

I would like to thank the Iraqi Ministry of Higher Education and Scientific Research which is represented by the Iraqi Cultural Attaché in London for sponsoring me during the whole journey of the PhD.

Many thanks to Professor Jeffry Green for kindly giving bacterial strains and some technical advice regarding dealing with these strains.

Thanks to Doctor Lynda Partridge and Doctor Phil Mitchell who were my advisors during my PhD research.

My thanks is also extended to Linda Harris the secretary of the Molecular Biology and Biotechnology (MBB) department for her great organising and administrative effort in the department.

Finally, my warm thanks go to my laboratory colleagues and everybody who gave me their support during my studies.

Abstract

It has become possible over the last decade to compare the primary structures of proteins from many hundreds of species at the push of a button. This rich source of information on the relationship between polypeptide sequence and function is however limited since there is often no experimental evidence for protein activity. By amplifying open reading frames using an error prone DNA polymerase, I have combined the power of comparative BLAST analysis with functional screening in order to "stress test" the relationships between primary structure and function in a number of "model" enzymes which include M.HhaI 5-cytosine methyltransferase, chloramphenicol acetyltransferase (CATI), and dihydrolipoamide dehydrogenase (E3). These enzymes models represent quite different structural and functional classes, and to strengthen any general conclusions, they have been chosen here to make the evaluation of any impact on changes in primary structure on enzyme activity, as comprehensive as possible. The results showed different patterns of tolerance toward the random primary structure changes. Dihydrolipoamide dehydrogenase appeared to be the most resilient enzyme, followed by M.HhaI and CATI respectively. This suggests that enzymes of different structural and functional types will have different degrees of tolerance and/or flexibility (in term of activity maintenance) in accommodating primary structure changes. Comparing the present mutagenesis and activity screening data with the available biochemical data of the related enzymes, as well as plotting the positions of mutations on the relevant secondary and three dimensional structures, showed interesting traits in all enzymes to tolerate primary structure changes even in regions of the enzymes normally considered to be sensitive points, while some amino acid residues of no previously assigned importance could abolish enzyme activity completely.

Abbreviations

A	Adenine
Amp	Ampicillin
Amp^R	Ampicillin resistant
ATP	Adenosine triphosphate
bp	Base pair(s)
C	Cytosine
CFUs	Colony forming units
C-terminus	Carboxyl terminus
dATP	Deoxyadenosine triphosphate
dCTP	Dideoxycytidine triphosphate
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
dNTP	Deoxynucleoside triphosphate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetra acetic acid
EP	Error-prone
G	Guanine
IPTG	Isopropyl -1- thio- β-D- galactoside
Kan	Kanamycin
Kan^R	Kanamycin resistant
Kb	Kilo base pair (1000bp)
*k*cat	Overall enzymatic catalytic rate
kDa	Kilo Dalton
Km	Michaelis constant
LB	Luria Bertani Medium
mcr	Modified cytosine restriction

mRNA	Messenger ribonucleic acid
MW	Molecular weight
N-terminus	Amino terminus
OD	Optical density
Oligo	Oligodeoxynucleotide
ORF	Open reading frame
PCR	Polymerase Chain Reaction
PAGE	Polyacrylamide gel electrophoresis
Pfu	*Pyrococcus furiousus*
Pol	Polymerase
Pho	*Pyrococcus horikoshii*
RM	Random mutagenesis
SAM	S-adenosyl-L-methionine
SDS	Sodium dodecyl sulphate
S-PhoEP	Fusion error prone *Pho* polymerase
T	Thymine
TAE	Tris-acetic acid-EDTA
Taq	*Thermus aquaticus*
TEMED	Tetramethylethylenediamine
TRD	Target recognition domain
Tris	Tris-(hydroxymethyl)-methylamine
U	Uracil

CONTENTS

Chapter1 General Introduction

1.1. Introduction

The process of evolution in living organisms over the last fifty years, as first proposed coherently by Charles Darwin (Darwin, 1872), has been examined at the molecular level. This was first made possible by the development, largely thanks to Sanger, of methods for protein and nucleic acid sequencing (Sanger and Tuppy, 1951; Sanger and Coulson, 1975; Sanger *et al.*, 1977). This has been further enhanced by the last twenty years of genome science, during which many bacterial genomes and an ever-growing number of eukaryotic genomes are being sequenced at the nucleotide level. What has become clear during this period is that these studies support the key principle underlying Darwin's theory that is adaptation is driven by Natural Selection and slow genetic change. Thus, within a group of related organisms, there is a high level of conservation of protein primary structure which reflects the relatedness of organisms. The striking level of conservation of histone proteins among all eukaryotes for example, suggests that the reversible processes of genome condensation and unpacking, are clearly fundamental to all eukaryotes. The remarkable success of Nurse's group in demonstrating the cross-compatibility of the cell cycle regulator gene cdc2, illustrates again the fundamental levels of conservation of gene function amongst eukaryotes as far apart in evolutionary terms as yeast and vertebrates (Nurse and Thuriaux, 1980).

1.2. Experimental evolution

The last twenty years of experimental molecular biology has witnessed the introduction of tools to mimic the process of evolution but critically in a significantly reduced time, so that several millions of years are compressed into only a few days or even hours. Such *in vitro* generation of genetic variants is achieved by means of random mutagenesis and DNA recombination (Chen and

Arnold, 1993; Stemmer, 1994; Reetz *et al.*, 1997; Zhao *et al.*, 1998). Starting with a target gene sequence it is possible to build mutant libraries that may then be screened phenotypically within a short time scale (artificial selection) for different applications such as industrial, therapeutic, or for fundamental research (Jespers *et al.*, 1994; Coelho *et al.*, 2013; McIsaac *et al.*, 2014). This process of randomising and diversifying genes has been called ‘‘directed molecular evolution’’. The schematic diagram in Figure 1.1 shows a comparison of directed versus Darwinian evolution. The outcome of Darwinian genetic selection is achieved by directed evolution in a powerful manner especially when the screening and selection methods mimic the natural sieving and selection of new traits that often accompany spontaneous evolutionary mutations.

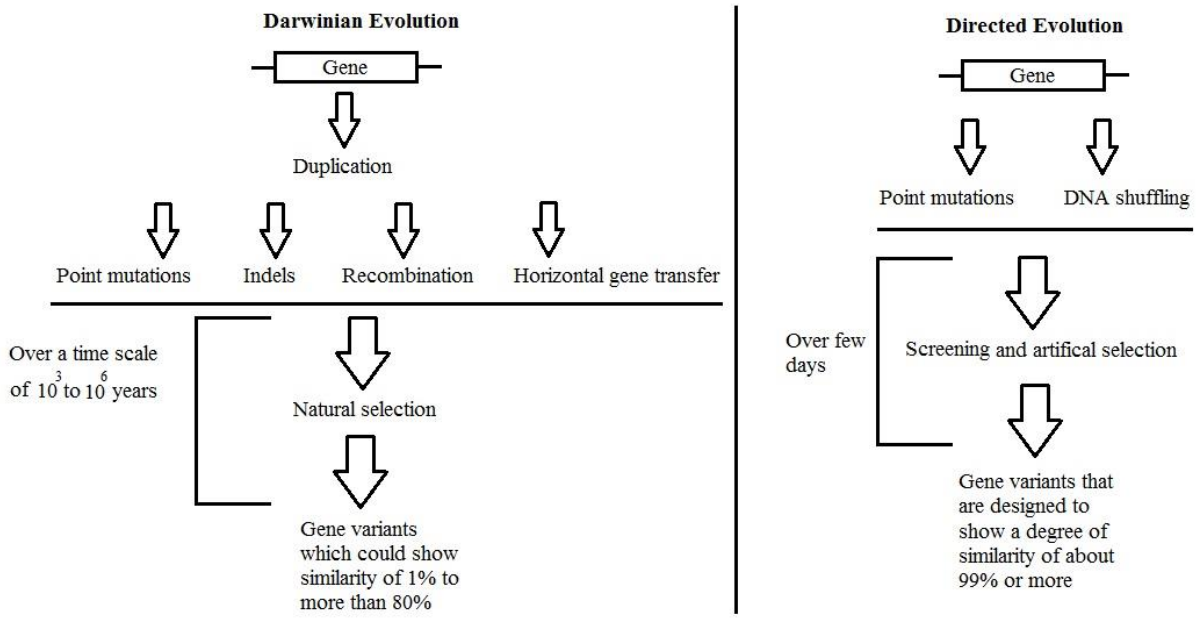


Figure 1.1: A schematic diagram comparing Darwinian evolution with the directed evolution at the molecular level.

Directed evolution has become the method of choice by which molecular biologists attempt to improve enzyme function for industrial or pharmaceutical purposes. It is therefore, an excellent strategy to generate a novel functional enzyme property or to investigate the catalytic and functional properties of

primary structures by exploring a much greater part of the protein sequence space within a greatly minimized time scale compared with the natural evolution and selection (Williams *et al.*, 2004; Packer and Liu, 2015). Such strategies offer an alternative approach to rational design of protein engineering (Dahiyat and Mayo, 1997; Root *et al.*, 2001).

Accordingly, directed evolution is a powerful strategy for studying the function of proteins where no crystal structures are available as well as proteins of known structure in order to expand knowledge regarding the relationship between the structure and activity (Wang *et al.*, 2006). The key objective of any directed evolution experiment is to generate as much genetic diversity as possible within a library of multi-sequence variants as well as developing suitable method(s) for high throughput screening of phenotype in addition to the genotype that must be linked in these experiments (Arnold, 2001; Wahler and Reymond, 2001).

The key elements of a successful directed evolution experiments are:

1- A suitable method for introducing mutation at a controllable frequency.
2- A target gene that provides a reliable screening methodology enabling scanning of large number of the generated mutant clones.
3- Ideally, a three dimensional protein structure should be available for interpretation.

It is not surprising, therefore, that researchers have developed many methods for gene randomisation each of which, has its pros and cons. Physical and chemical mutagens such as methanesulfonate (EMS); an alkylating compound, nitrous acid; a deaminating compound, 2-aminopurine; a base analogues compound, or UV light have all been used to provoke mutations by randomly damaging DNA. However, the spectrum of bias introduced by such chemical reagents and physical mutagenic agents makes them less useful in directed

evolution experiments (Freese, 1959; Bridges and Woodgate, 1985; Myers *et al.*, 1985; Lai *et al.*, 2004). Non-chemical random mutagenesis methods mostly rely on using DNA polymerases such as polymerase III as well as Tap or pfu polymerases. Polymerase III has a mutagenesis ability rate of 10^{-10} per base per replication and has been used within *E. coli* mutator strains such as mutS, mutD, and mutT where the proofreading and the enzymatic DNA repair system have been abolished leading to an enhanced mutagenesis ability for such DNA polymerase (Cox, 1976; Scheuermann *et al.*, 1983; Greener *et al.*, 1997b). For example, a mutator *E. coli* called XL-1 red strain showed a mutagenesis incidence of about 10^{-6} per base per DNA replication after transformation with plasmid carrying the gene of interest to be randomised (Greener *et al.*, 1997a). However, the disadvantage about using the mutator strains is that such strains are most likely introducing some deleterious mutations within their genomic DNA because it is vulnerable to the low fidelity DNA polymerase exactly the same as the plasmid carrying the gene of interest that is required to be mutated, therefore, some efforts have been performed to find a strategies by which only the gene of interest is targeted by random mutagenesis other than the genomic DNA in those mutator strains, such strategies of isolating the genomic DNA from the mutagenic effect of the host polymerase should reduce the limitation of *in vivo* mutagenesis especially when host is intolerant toward the accumulation of deleterious mutations within its genome. It seems that such mutagenesis isolation is hard to be applied in bacteria while in eukaryotes like yeast it is efficiently applied when *Kluyveromyces lactis* naturally-occurring cytoplasmic pGKL1/2 linear plasmids have been used as vectors to clone the genes of interest so that the genomic DNA in nucleolus is completely isolated and away from the mutagenesis effect due to the specific action of TP-DNA polymerase which is a cytoplasmic enzyme required specifically for those plasmids replication in the cell cytoplasm. Those plasmids with their TP-DNA polymerase have been used in an engineered *Saccharomyces cerevisiae* as an

efficient mutator host for gene diversification (Ravikumar *et al.*, 2014). In spite of all the enhancements for the in vivo random mutagenesis methods, they still impose limitations especially by being hard to control and by providing low rates of mutations, therefore, researchers prefer using the *in vitro* methods to introduce random mutations by DNA amplification as an artificial replication using polymerase chain reaction technique to obtain enhanced mutagenesis throughputs (Packer and Liu, 2015).

1.3. Error-prone PCR (EP PCR)

In the last decade, advances in the area of nucleic acid amplification, primarily driven by diagnostic methods employing the Polymerase Chain Reaction (PCR), have made it possible to begin to explore molecular evolution on a manageable time-scale. Much of the development of PCR enzymes has centred on their intrinsic fidelity. Fidelity is simply defined as a quality control index for the amplification reaction: how accurate is the amplification. If a single copy of a given amplicon is copied 1000000 times, are all copies identical in terms of nucleotide sequence? Many different approaches to benchmarking fidelity have been reported and the most common methods utilise microbiological screening for colour sensitive colony formation in bacterial transformation experiments that can detect changes at the frequency of approximately 1 per 1000000 events. It is clear from the last thirty years of PCR development, that some DNA Polymerases are more “faithful” than others and companies produce comparator data for applications. However, in this thesis, the focus is on a PCR enzyme, capable of efficient amplification of template DNA, in high yields, but exhibiting low fidelity. Such an enzyme has been described by Biles and Connolly (Biles and Connolly, 2004) and has also been developed by (Alharbi, 2010) . Two point mutations in either Pfu DNA Polymerase (a recombinant form of the enzyme from *Pyrococcus furiosus*), or in the closely related enzyme Pho DNA Polymerase (*from Pyrococcus*

horikoshii), produce such an error prone DNA polymerase. The characteristics of both EPPfu and EPPho have been described by and (Alharbi, 2010) respectively, and form part of the following discussion on the experimental approaches to *in vitro* evolution.

Polymerase chain reaction (PCR) is an efficient technique by which any sequence of nucleotides, especially DNA, can be propagated into billions of copies starting from relatively small amounts of that DNA (Mullis and Faloona, 1987; Erlich, 1989; Sun, 1995). Principally, PCR depends on three steps: denaturation; during which the hydrogen bonding, that stabilises the two strands of the DNA duplex, are destroyed by raising the reaction temperature so that the resulting two strands of the DNA template are available to base pair with short complementary oligonucleotides (primers) in the second step which is annealing when the temperature is lowered appropriately. The 3′ ends of the annealed primers then provide the starting point for the DNA polymerase to begin elongation, where a complementary new DNA strand is synthesised, thereby recapitulating the DNA double helix. Each of the PCR three steps (denaturation, annealing, and elongation) is performed with an appropriate temperature: generally the temperature of denaturation step is always the highest followed by the elongation and annealing respectively. Each PCR cycle is composed of these three steps, so that one DNA molecule of a certain length will be two molecules after the first cycle. The key point of PCR in amplifying a segment of DNA in a logarithmic manner is that the number of reaction cycles are repeated several times, so the final number of DNA molecules copies after a certain number of PCR cycles is calculated using the relationship $(2^n\text{-}2n)x$, where n is the number of cycles and x is the number of initial template DNA molecules and $2n$ is the number of variable length DNA fragments of undefined length that are produced initially during the first and second round of PCR (Newton and Graham, 1997).

Originally, PCR was developed for accurate amplification of DNA, but the requirements for random mutagenesis and directed evolution has extended the original purpose of PCR from accurate to error-prone DNA amplification, which was introduced for the first time by (Leung *et al.*, 1989) and is now considered to be a another invaluable application of the PCR technique.

All the known DNA polymerases act on their DNA substrates to build a new nucleotide strand accurately by incorporating the correct dNTPs into the growing DNA strand according to the sequence in the DNA template strand, so that Watson-Crick base pairing is maintained most of the time, however, the accuracy, which is commonly termed "fidelity", in amplifying the DNA without introducing mutations, is actually dependent on the DNA polymerase employed. So DNA polymerases have lower fidelity when compared among different organisms. For example, Taq DNA polymerase from *Thermus aquaticus* is of lower fidelity among all commercially available, thermostable DNA polymerases used in PCR; introducing a single error every 9000 base per PCR cycle, primarily because it lacks 3′→5′ proofreading exonuclease activity (Tindall and Kunkel, 1988). However, error-prone PCR requires a much lower fidelity than even that of wild Taq- polymerase, which is still too accurate for mutagenesis applications (Neylon, 2004), therefore, changes to the standard PCR conditions have been used to enhance the error rate of PCR as follows:

1- Manipulation of the reaction components:

Equimolar dNTPs are normally used in non-mutagenic PCR, but for epPCR, dNTPs are often included at unbalanced concentrations. For example dCTP and dTTP are supplemented in amounts exceeding the concentration of the other nucleotide triphosphates to drive Taq-pol to mutate AT to GC base pairs (Cadwell and Joyce, 1992; Cadwell and Joyce, 1994; Vartanian *et al.*, 1996b; Shafikhani *et al.*, 1997). Mutagenesis of dihydrofolate reductase (DHFR)

expressed from R67 *E. coli* plasmid by introducing biased deoxypyrimidine triphosphate concentrations in a hypermutagenic error-prone PCR reaction (where dTTP is supplied in a higher concentration than dCTP) has led to a complicated mutant DHFR gene library enriched in dATP and dTTP taking the advantage of the mismatched G:T base pairing which is stably present. While using different combinations dNTPs unbalanced concentrations such as [dTTP] > [dCTP] and [dGTP] > [dATP] unexpectedly lead to a low mutational frequency. However, a relatively efficient mutational frequency of 10% per PCR cycle, with a relatively homogenous distribution of mutations along the nucleotide sequence of the DHFR gene. A much more unfavourable situation (with respect to fidelity) has been seen with a low level bias toward transition than transversion, when Mn^{2+} cations were used in place of, or alongside Mg^{2+} ions (Vartanian *et al.*, 1996b).

In spite of the key role of manganese cations in the reduction in fidelity of the PCR (Lin-Goerke *et al.*, 1997; McCarthy *et al.*, 2004), these ions impose some limitations, especially for the subsequent cloning steps that follow the PCR mutagenesis, such as the formation of complex structures at the terminal ends of the PCR amplicons constituting a bad scenario for restriction digestion of the PCR product, as restriction sites are usually engineered into both ends of the PCR product, therefore, such complexes formed owing to the presence of Mn^{2+} could easily interfere with the restriction enzymes and therefore disrupt ligation and cloning. Indeed, ligation has been shown to be inhibited by Mn^{2+}cations, owing to their inhibitory effect on T4 DNA ligase. However, this could be overcome by incorporating a PCR product clean-up step before ligation (Wang *et al.*, 2006). Overall, manganese ions decrease the accurate recognition ability of DNA polymerase during the elongation, so that the number of mismatches in the growing DNA strand will increase subsequently

especially when unbalanced dNTP concentrations are applied (Beckman *et al.*, 1985; Leung *et al.*, 1989; Vartanian *et al.*, 1996a).

Magnesium cations are crucial for the *Taq*-pol reaction to proceed in the non-mutagenic PCR project, so this ion is always added to all PCR amplifications within optimum amounts depending on the PCR system, but excessive concentrations of Mg^{2+}have been applied in epPCR to stabilise the base pairing of uncomplimentary nucleotide bases (Eckert and Kunkel, 1991; Ling *et al.*, 1991).

2- Using unusual nucleotides:

Nucleotide analogues, exhibiting comparable kinetic characteristics and binding to conventional ones, are incorporated into the elongated DNA strand, which is stable under PCR conditions: these analogueues have been demonstrated to promote, significant levels of mutagenesis and have been used in EP-PCR for random mutagenesis. Two 5'-triphosphate compounds: 6-(2-deoxy-beta-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-C] [1,2]oxazin-7-one(dP) and 8-oxo-2' deoxyguanosine (8-oxodG) have been designed and synthesised and shown to have excellent substrate properties for the *Taq*-pol, which incorporated both of them very efficiently into the newly synthesized DNA: the dPTP analogue replaced dTTP and dCTP where the 8-oxodG analogue base pairs with adenine. The method for random mutagenesis in this case comprises two stages of PCR amplification. In the first one, both dPTP and 8-oxodG analogues or just one of them are incorporated in the presence of the all normal dNTPs. The analogues compete with the normal dNTP substrates, and are incorporated during DNA strand elongation. In the second PCR round, only the four normal DNA deoxynucleotides triphosphates are added to the reaction so that the analogues incorporated during the first reactions are replaced by normal ones during this stage to enable the subsequent cloning and

experimental steps by avoiding the *E. coli* host cell machinery of DNA repair that could exert its action on the DNA containing nucleotides analogues. After the second round of PCR the sequencing data showed, especially when both analogues were added, that a significant number of clones, containing at least 1-2 transition and/pr transversion mutations, along with extremely rare events of deletions and insertions as a result of the wobble-related replacement of the nucleotide analogues during the second PCR round were obtained (Zaccolo *et al.*, 1996).

The nucleotide dITP (2'-deoxyinosine-5'-triphosphate) containing the deoxyinosine base analogue has been used by (Spee *et al.*, 1993) to develop mutations in a random manner, since Taq-pol has been shown previously to efficiently utilise ITP (Innis *et al.*, 1988). The method involves using the four normal dNTPs in four independent PCR mixtures each of which has one of the dNTPs in a reduced concentration so the depleted pool of any dNTP will result in misincorporation of a wrong dNTP in the growing DNA strand being synthesized by Taq-pol, however, this could be enhanced when the dNTPs of the limited concentration were compensated for by dITP which is supplied simultaneously to the PCR mixture, resulting in efficient incorporation of this nucleotide analogue. A second round of PCR with only normal dNTPs in equimolar amounts, is necessary so that any position in a single DNA strand containing dITP from the first round will be possible to have any of the four dNTPs as a complementary in the opposite DNA strand being elongated by Taq-pol, therefore, gene randomisation will be the final result.

A combination of the manganese divalent cation and dITP as two factors of EP-PCR has been used in order to randomise the cutinase gene of *Fusarium solanipisi* in two PCR steps using Taq-pol. In the first PCR round, only Mn^{2+} has been added to reduce the dNTPs recognition ability of Taq-pol, then dITP is added in the second PCR step to induce the mispairing events and eliminate the

mutagenesis bias introduced during the first step by increasing the incidence of GC to CG transversions leading finally to 1-2 random mutation per clone (Bluteau *et al.*, 1999).

3- PCR sequence mutagenesis saturation:

The use of the PCR technique has been extended to include gene mutational randomisation in a saturation like manner via the mutation-saturated DNA sequence method (SeSaM). This technique was applied by (Wong *et al.*, 2004) to generate mutant libraries of green fluorescent protein (GFP) via four technical steps (Figure 1.2). The first step involves generation of DNA fragments of different lengths for the gene of interest. Using α-phosphothioate nucleotide with normal nucleotides and biotin-containing forward primer with a non-biotinylated reverse primer in PCR, will results in PCR amplicons containing α-phosphothioate nucleotides at several positions. Treating this PCR product with iodine under alkaline conditions will result in hydrolysis at every α-phosphothioate position because phosphothioate bond is readily reduced under alkaline conditions, in the presence of iodine, thereby generating several DNA fragments of a wide variety of lengths. Biotinylated single strands are then specifically isolated from the rest of the unwanted DNA species using streptavidin beads that bind the biotin moiety in every biotinylated DNA strand so that the final purified fraction of interest will be single-stranded DNA molecules of different lengths.

The second step in SeSaM is to add deoxyinosine, a universal nucleotide analogue, to the 3′ terminal ends of the previously generated single-stranded DNAs. This elongation step is catalysed enzymatically by terminal transferase. Deoxyinosine has a strong bias to complement with cytosine (Kawase *et al.*, 1986; Spee *et al.*, 1993) so it is used to replace dGTP in DNA sequencing (Dierick *et al.*, 1993; Motz *et al.*, 2000).

The third step involves extension of the DNA fragments of the terminal universal analogue to the full length double stranded gene. The full length 3′→5′ DNA strand is required to be used as a template in this step, so PCR is performed using biotinylated reverse primer for this purpose to produce full length DNA amplicons, biotinylated with respect to the 3′→5′ strandwhich is then isolated by the same methods described in the first step. Then one cycle of PCR is performed using the 3′→5′ product as a template and the DNA fragments generated previously act as primers for the elongation step, initiated from the 3′ terminus containing the deoxyinosine base analogue. The final product will be a full length double stranded DNA containing a universal bas analogue in one strand. Additional PCR cycles will lead to annealing of the reverse primer to the newly synthesised 5′→3′ DNA strand to be used as a template to synthesise a complementary 3′→5′ strand, but now with a possibility of different nucleotide being incorporated opposite to the complementary strand containing the deoxyinosine base analogue which is located in the 5′→3′ DNA template strand. The deoxyinosine within the DNA is then replace by the normal dNTPs by applying another round of PCR amplification in a final fourth step resulting in a full length double stranded mutant DNA gene library containing mutations distributed randomly throughout the whole gene length. There will also be a noticeable bias of replacements at adenine positions, due to the exclusive use of dATPαS as a nucleotide analogue that has a strong preference for base pairing with cytosine, so that the vast majority (95.8%) of adenine substitutions are of the type A→G (Kawase *et al.*, 1986; Spee *et al.*, 1993). Further improvements to the method include using different nucleotide universal analogues with different base pairing preference to provide a less biased and better distributed mutations and a better distribution

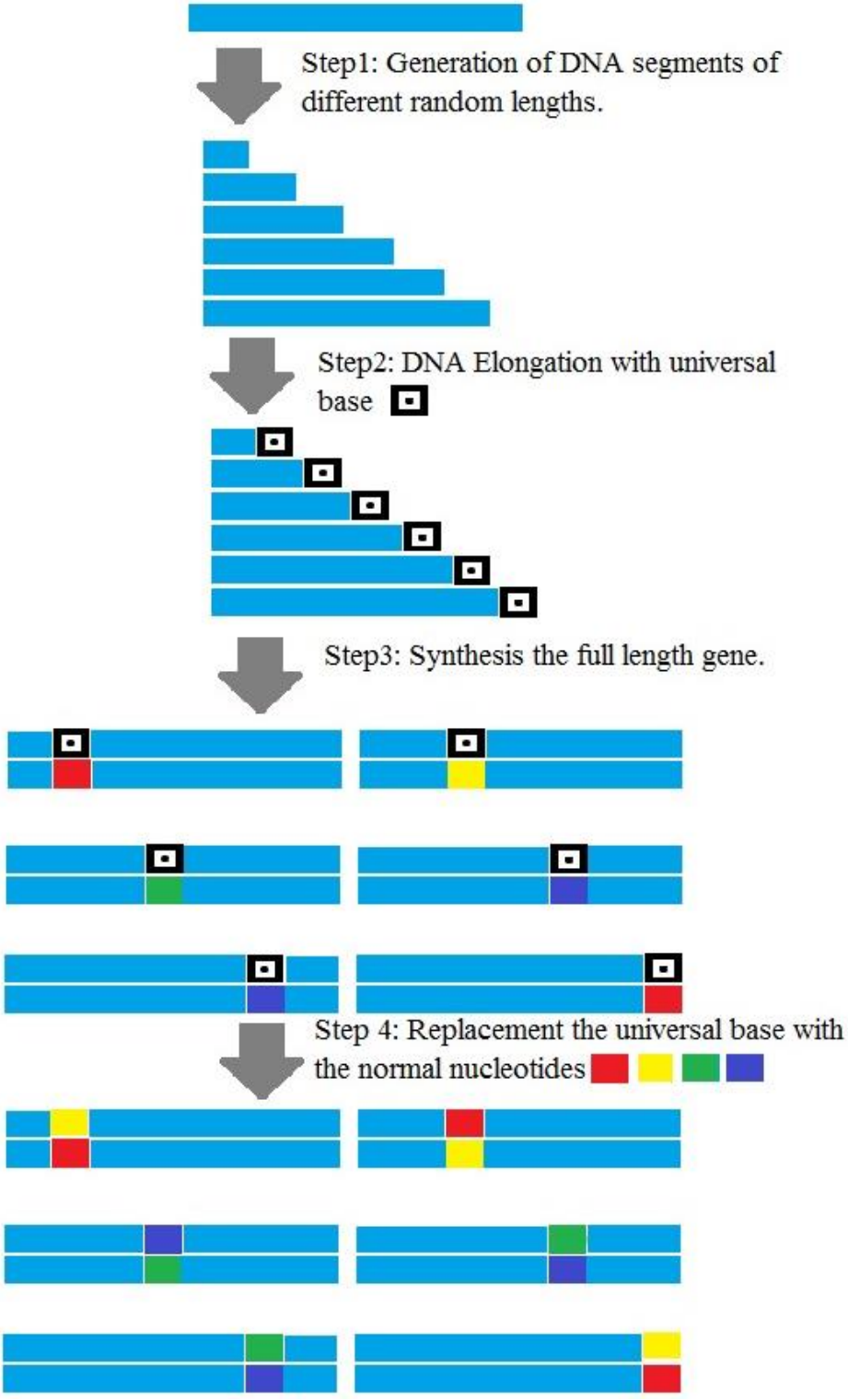


Figure 1.2: Schematic diagram showing the four main steps of the sequence saturation mutagenesis method (SeSaM). In step 1, the gene is cut to generate pools of random DNA lengths. In step 2, the DNA fragments are elongated with a universal base. In step 3, the full length gene is synthesized. Step 4 involves replacement of the universal base with the standard nucleotides. Adapted from (Wong *et al*., 2004).

so that an improved control of the mutagenesis can then be obtained using what is called “tuned” SeSaM (Wong *et al.*, 2004).

4- Mutagenic DNA polymerases:

According to the sequence similarity and phylogenetic correlations, DNA polymerases can be classified into seven major groups: A (such as *E. coli* DNA polymeraseI), B (such as *E. coli* DNA polymeraseII), C (such as *E. coli* DNA polymeraseIII), D (such as eukaryotic polymeraseII), X (such as human β-DNA polymerase), Y, and RT (such as reverse transcriptase) (Burgers *et al.*, 2001). DNA polymerases have two main roles in the cell represented by DNA replication and DNA repair varying in fidelity from error-free to error-prone (Kim *et al.*, 2008).

Understanding the contribution of the protein structure to polymerase function is key to modifying such enzymes and in controlling their functions in respect of random mutagenesis. As shown by many structural studies, polymerases are commonly multi-domain enzymes composed of three main parts: fingers, palm and thumb (Kim *et al.*, 2008). Elucidation the role of DNA polymerase from DNA substrate binding to the recognition and incorporation of the dNTPs (fidelity and processivity) has been verified by the structural studies that have shown common conformational changes occur upon polymerase binding to the DNA substrate, and in this process the enzyme fingers and thumb have been shown to be mobile and able to adjust their conformation to get closer to the palm domain during DNA recognition, thereby providing a compact, closed conformation for the enzyme to better hold on to its substrate (the DNA template and the primer), while the open conformation prevailed when the enzyme was not bound to its DNA substrate (Brautigam and Steitz, 1998; Li *et al.*, 1998). In addition to the three main domains, archaeal

polymerase possess two additional domains called exonuclease and N-terminal domains (Figure 1.3). The exonuclease part has the capacity for conformational change depending on the momentary action required by the polymerase to perform with DNA whether it is a replicative or editing role. When a wrong dNTP is incorporated to the DNA strand being synthesized, a weakness in binding between the DNA (the template and the primer) and the polymerase will emerge causing the DNA double strand to be disassembled at this vicinity so that the mismatched nucleotide will be pushed into the active site of the exonuclease domain where it is sheared. However, the steric properties of the closed DNA polymerase conformation (when the enzyme binds DNA tightly) prevent the 3′ end of the growing DNA strand from being presented into the exonuclease domain active site, so the multi-domain structure is clearly important in providing the catalytic harmony of the multifunctional DNA polymerases (Freemont *et al.*, 1988; Hopfner *et al.*, 1999; Hashimoto *et al.*, 2001a; Kim *et al.*, 2008).

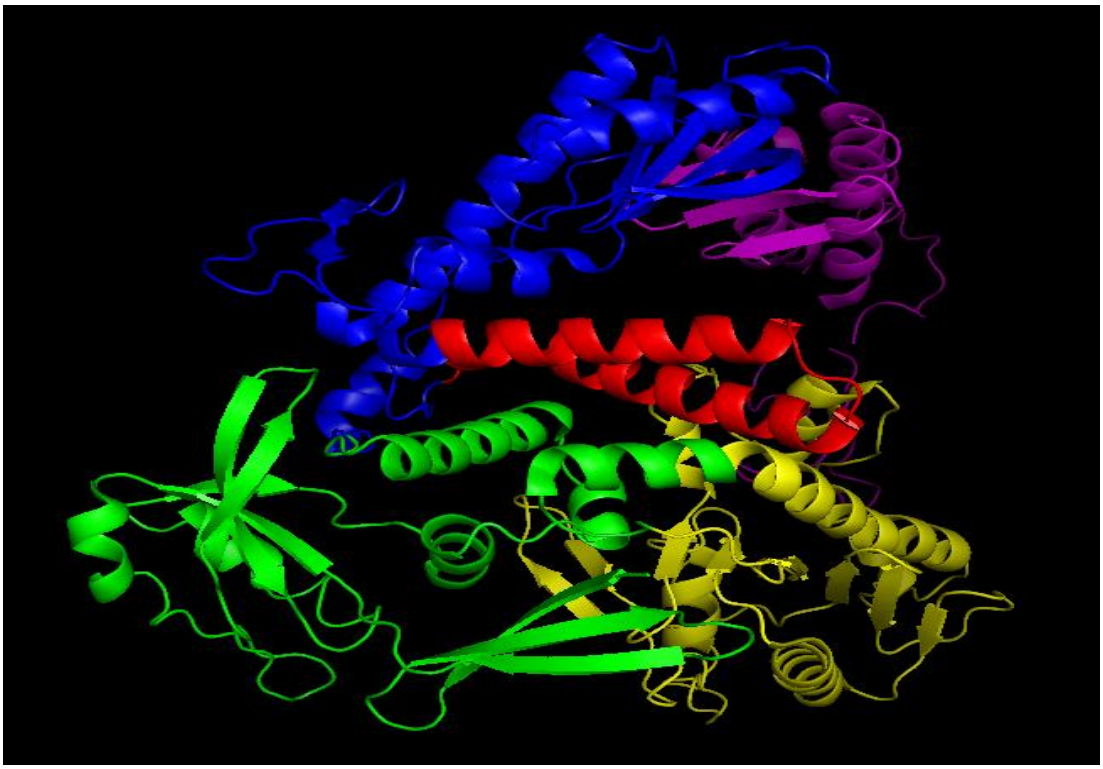


Figure 1.3: The three dimensional structure of *pfu* DNA polymerase as a canonical example to show the domain structure and organisation of the family B archaeal DNA polymerases. Green, N-terminal domain; yellow, exonuclease domain; blue, palm domain; red, fingers domain; and magenta, thumb domain. The figure is modified from (Kim *et al.*, 2008) (PDB code 2JGU).

It has been shown that the exonuclease domain of archaeal polymerases has a unique conserved loop containing a conserved His-147 residue playing important role in fixing the relatively wide gap which is required for the ssDNA 3′ end to reach the exonuclease active site. Therefore changing His-147 into Glu, which introduces a negatively-charged side chain, causes electrostatic interactions between the loop of exonuclease domain and the edge of the thumb domain, which is positively-charged, so that the thumb domain moves 1.5 °A closer to the exonuclease domain, thereby preventing access of the 3′ end of the growing DNA strand to the exonuclease domain active site. This rationalises the necessity for the wide open polymerase conformation for performing DNA editing (Wang *et al.*, 1997; Kuroita *et al.*, 2005a, b). However, crucial DNA polymerases properties such as the fidelity, processivity, robustness differences in performing the replicative and/or DNA repair role, which are all important in controlling genome stabilisation and evolution, mainly depend on minor amino acid diversity in the active site, unique loop and the edge of the thumb domain (Kim *et al.*, 2008).

As they have originally been isolated from different thermophilic organisms, many DNA polymerases provided invaluable advantages to be used in PCR technology, where high temperatures (slightly less than boiling) are applied during thermo-cycling. Furthermore, low fidelity, which is considered as a major drawback in standard PCR, is by contrast, critical to the success of

random mutagenesis by epPCR. Taq-pol from *Thermus aquaticus* and Mutazyme (Stratagene) are commonly used in epPCR (Wong *et al.*, 2004). Taq-Pol is the best known regarding the low accuracy in DNA amplification when used in PCR, even under standard reaction conditions, because it lacks the 5′→3′ proof reading exonuclease activity for removing the wrongly incorporated dNTPs so that it provides an error rate of 2.7×10^{-5} per base per amplification cycle under standard non-mutagenic PCR conditions, making it the polymerase of choice in many epPCR experiments (Neylon, 2004). It has been shown that providing increased concentrations of Taq-pol should increase the rate of errors and important to ensure unruffled extension when the dNTPs misincorporation occurs (McCabe, 1990). The disadvantage of Taq-pol in epPCR is the bias in introducing transitions rather than transversions and it has been recorded that the about 40.9% of the nucleotide mutations are of A→T and T→A replacements (Cline and Hogrefe, 2000).

The relatively poor efficiency of Taq-Pol in epPCR such as subdued yield, low mutational frequencies, and most importantly bias in mutations motivated the research wheel for looking for and engineering a better mutagenic polymerase that could provide a vehicle for random mutagenesis. The requirements for such and epPol include providing unbiased mutational patterns since a protein sequence must ideally be changed at every codon to generate all 20 possible amino acids, at least this is the goal in producing highly diversified mutational libraries.

Alongside the commonly used bacterial Taq-Pol, the DNA polymerase from *Pyrococcus furiosus* (Pfu-pol), which is related to family B polymerases, is the most commonly used thermostable archaeal enzyme in PCR (Biles and Connolly, 2004). Unlike Taq-Pol, Pfu-Pol possesses extremely high accuracy in amplifying DNA, exhibiting a very low mutational frequency of 1.3×10^{-6} mutation per base per replication (eight fold higher than Taq-Pol fidelity)

(Capson *et al.*, 1992) as it has a 3′→5′ proof reading property, so that it provides the high fidelity in standard PCR amplifications (Lundberg *et al.*, 1991; Cline *et al.*, 1996). Such high fidelity should make related polymerases unlikely candidates for use in epPCR, but engineering these enzymes to turn their high fidelity into a low one has been successfully achieved. The crystal structures of many of the family B polymerases reveal a couple of long anti-parallel helices separated by a short loop as the elements of the fingers sub-domain (Figure 1.4). This element is critical for high fidelity and recognition and incorporation of the correct dNTP during cellular DNA replication. These two helices have been named N and O respectively (Joyce and Steitz, 1994; Brautigam and Steitz, 1998; Doublie *et al.*, 1999; Zhao *et al.*, 1999; Rodriguez *et al.*, 2000; Hashimoto *et al.*, 2001b).

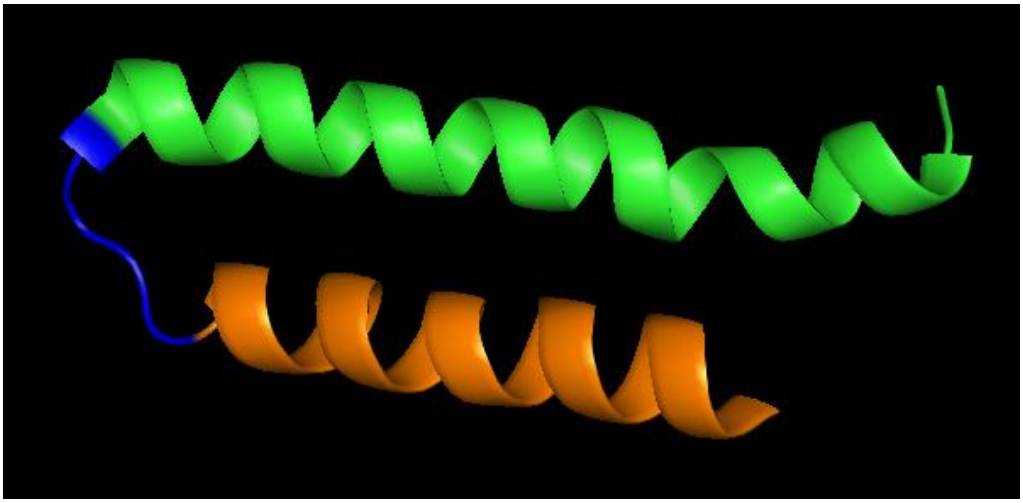


Figure 1.4: The two helices (green and orange) forming the fingers sub-domain with the short loop (blue) that links them. This is from the crystal structure of P*fu* DNA polymerase resolved by (Kim *et al.*, 2008) (PDB code 2JGU).

The two helices and the intervening short loop are actually highly conserved among the archaeal family B DNA polymerases. This suggests that the same

mechanism is followed in all these enzymes, that share the helix-loop-helix conformation, in recognizing and incorporating the correct dNTPs during DNA replication and PCR amplification (Biles and Connolly, 2004). In this manner and during the DNA polymerase catalysis in replicating the DNA, the incoming dNTP is retained by the enzyme-DNA complex only when its association is Watson-Crick competent, otherwise, it will be rejected: in this way, enzymatic fidelity is ensured. Therefore, and as shown kinetically, any structural modifications to this region of the enzyme could reduce its fidelity (Wong *et al.*, 1991; Kunkel and Bebenek, 2000; Kunkel, 2004).

According to the available kinetic and structural data, it has been concluded by (Biles and Connolly, 2004) that the orientation of N and O helices is a determinant for Pfu DNA polymerase's high fidelity; so changing their orientation should result in reduced enzyme accuracy. The three amino acids forming the short loop that links the two helices were targeted for mutagenesis: each was replaced by either Gly or Ala, which possess the smallest side chains possible and it was thought that they would make the loop junction more flexible and consequently reduce the rigidity that is believed to be behind the fixing of the long helices. The carboxyl side chain, together with the backbone carbonyl-oxygen of Asp-473 are involved in hydrogen bonding with the nearby Leu-Glu-Lys residues of the O-α-helix. These interactions are crucial in maintaining the orientation of the helices (anti-parallel) and hence the incorporation of the correct dNTPs. By changing the loop Asp-473 into Gly, the loop becomes flexible enough to change the orientation of the helices by removing hydrogen bond stabilisation which is a key part of the fidelity mechanism during DNA amplification. Unlike Asp-473, the other two loop residues (Thr-471 and Gln-472) have been not been shown to play a role in making any significant structurally determining interactions, therefore, mutating them should result in either no effect or a minor one in respect of enzyme

fidelity (Biles and Connolly, 2004). The Pfu 5′→3′ proof reading ability is required to be abolished as well by theAsp215Ala companion mutation or the wrongly incorporated dNTPs will be excised (Evans *et al.*, 2000). So the double mutant enzyme: Asp473Gly and Asp215Ala enzyme was produced as an error-prone thermostable polymerase that works efficiently under the same reaction conditions as the wild-type polymerase, generating a mutational load of about 7.2 error/kb with no need to manipulate the reaction conditions. The mutagenesis bias has been shown to be toward T→A change but it is considered as a tolerant bias with no hot spots noticed and it is much less than the biased behaviour of Taq-Pol when utilised under non-standard conditions(Biles and Connolly, 2004).

1.4. Error-prone *pho* DNA polymerase

As mentioned previously, family B archaeal DNA polymerases are extremely similar on both levels of primary and tertiary structure so that the structural data of many of them appeared common structural general features of being composed of multi domain structure with high degree of analogy in sharing sub domain structural properties especially those which are considered as a fidelity determinant features, such as the helix-loop-helix structure and the exonuclease structural determinants. Such structural analogy should provide the flexibility advantage in designing and modifying other DNA polymerase using the same mutational strategy at the same positions.

In our laboratory, another high fidelity thermostable archaeal family B DNA polymerase from *Pyrococcus horikoshii* (*Pho* polymerase) has been engineered for random mutagenesis (Alharbi, 2010). This enzyme shows a high degree of similarity to the *pfu* polymerase especially at the fidelity determinant residues of the helix-loop-helix motif, therefore, the same residues that were targeted

previously in *pfu* have been targeted in *pho* in term of two companion mutations one of them is Asp473Ala in the short loop junction between the O and N helices fingers in order to reduce enzyme fidelity by making the helical fingers more flexible and less rigid due to the reduction of the hydrogen bonding network that involves Asp-473 in the wild type enzyme. The other mutation is Asp215Ala which acts as an exonuclease activity reducer and is exactly the same mutation as that introduced into the *pfu* polymerase as mentioned previously. The result was an error-prone pho DNA polymerase (PhoEP) with a mutational load of 7 errors per kilo base with a bias towards G→A change, while C→G mutation is under represented. Transitions were more frequent than transversions accounting for 54.8% and 45.2% respectively.

It has been emphasized that reducing the fidelity of polymerases is accompanied with the simultaneous efficiency reduction in amplifying long DNA substrates and producing high yield of PCR product. The problem of low efficiency of low accuracy polymerases (especially the engineered ones) has been suggested to be related to impaired DNA binding of mutants, especially in the fingers domain. Furthermore, the misincorporated nucleotides act as barriers in the path of the polymerase while travelling along the DNA, making it less stable in holding on to the DNA, and consequently facilitating enzyme dissociation (Kelman *et al.*, 1998; Biles and Connolly, 2004; Pavlov *et al.*, 2004; Wang *et al.*, 2004; Berman *et al.*, 2007; de Vega *et al.*, 2010). Therefore, in our laboratory, enhancements have been made to the PhoEP by fusing it with a DNA binding protein Sso7d-2 to provide an additional "grip" to the DNA template. This DNA binding protein is a monomer of 7 KD (64 residues) originally expressed in the *Sulfolobus solfataricus* with a significant affinity in DNA binding to increase the DNA melting temperature of the latter thermophilic archeon (Baumann *et al.*, 1994; McAfee *et al.*, 1995; White and Bell, 2002). The idea of PhoEP fusing with Sso7d-2 in one polypeptide was

inspired by the previous work of (Wang *et al.*, 2004) that includes enhancement of the processivity and performance of polymerases by fusing with DNA binding proteins to produce novel engineered enzymes with a combined two advantages including increased robustness in amplifying long DNA strands within less time required for PCR elongation, and mutagenic ability. The polypeptide results from the fusion of PhoEP with Sso7d-2 was referred to as S-PhoEP. The results have actually demonstrated a better efficiency of S-PhoEP in amplifying the same length of DNA in a shorter elongation time compared with the not fused PhoEP. The error rate has been shown to be increased by 40% after the fusion with bias behaviour similar to that of not fused version with the exception of showing more transversions than transitions (Alharbi, 2010). In the present random mutagenesis study, S-PhoEP has been used as a mutagenic tool to generate libraries of mutations in three different enzyme models.

The aim of the study

This study aims to apply the concept of directed evolution to interrogate the relationship between primary structure and the function and activity of several enzymes starting from the following questions:

1- Is it the case that the primary structure of a given protein should be maintained for that protein to remain functional, or there is a level of flexibility that leads to different variants of the same protein performing the same function as the wild type?
2- If the answer is yes, then what level of change can the protein tolerate without loss of function?
3- How do enzymes of different functional and structural classes respond to the primary structure changes?

We know that Nature typically introduces changes in a given protein sequence over long time scales, by natural selection, in response to changes in certain environmental condition, or through natural fidelity defects. However, in the present study, the ability of the protein to perform its function is investigated while the environmental conditions are the same but the primary structure is varied experimentally *in vitro*.

The method chosen to generate mutations is error-prone PCR using the S-PhoEP mutagenic DNA polymerase as the randomising agent. By combining mutagenesis with a range of microbiological and biochemical screening methods, the impact of mutations on the activity of specific enzymes and on the growth of the host organism, was examined. Finally, attempts were made to rationalise the data using aligned primary structures and 3D models derived from available crystallographic studies. In short, this study aims to "stress-test" a group of unrelated enzymes in order to assess the relationship between Darwinian Natural Selection and the precision that chemistry teaches us, underlies catalytic mechanism.

The experimental plan is outlined below.

1- Gene design and synthesis to introduce convenient restriction sites for region-specific mutagenesis of the target gene.
2- Sub-cloning of segments of the open reading frame (orf) for error-prone PCR amplification followed by, double restriction digestion of the products, ligation to reconstitute the complete orf, and transformation.
3- Systematic "plating" of mutant libraries on selective media (where appropriate), recovery of plasmids from individual clones following transformation).
4- Functional screening (to be described for each target gene in the relevant Chapter) using microbiological and biochemical techniques.

5- DNA sequencing and data analysis to characterise the mutants recovered. Note that any clones that harbour deletions insertions or rearranged ORFs are not analysed further in this work.
6- Evaluation of the position and nature of the mutations alongside the wild-type DNA sequence and subsequent analysis of the amino acids changes predicted by the changes to the DNA sequence in silico.
7- Interpretation of the results in conjunction with published structural and biochemical data.

Chapter 2 Materials and Methods

2.1. Materials

2.1.1. Buffers, culture media, solutions, chemicals and kits

All chemicals, reagents used in this study are of a molecular biology standard manufactured by Sigma Aldrich®, Thermo Fisher®, Anala®, Difico®, Bioline®, Qiagen®, Novagen®, and New England Biolabs®. Materials are mentioned as contents of buffers, solutions, kits as well as individual chemicals and reagents as shown within the following table:

Name	Contents
LB medium (liquid/ solid)	10 g/litre tryptone; 5 g/litre yeast extract; 10 g/litre NaCl. If solidification required, 1.5% (w/v) agar is added.
SOC outgrowth medium	2% (w/v) tryptone; 0.5% (w/v) yeast extract; 10 mM NaCl; 2.5 mM KCl; 10 mM $MgCl_2$; 10 mM $MgSO_4$; 20 mM glucose.
1X M9 glucose minimal medium (1 litre) (liquid/ solid)	11.28 g of 5X minimal M9 salts (supplied by Difco™), 20 ml of 20% (v/v) glucose solution, 2 ml of 1 M $MgSO_4$ solution, 0.1 ml of 1 M $CaCl_2$ solution, and 30 ml of 10mg/ml thiamine solution. If solidification required, 1.5% (w/v) agar is added.
M9 glucose + acetate minimal medium (liquid/ solid)	The same contents of M9 glucose minimal medium supplied with 2 mM potassium acetate. If

	solidification required, 1.5% (w/v) agar is added.
M9 glucose + succinate minimal medium (liquid/ solid)	The same contents of M9 glucose minimal medium supplied with 2Mm sodium succinate. If solidification required, 1.5% (w/v) agar is added.
Ampicillin stock solution	100 mg/ml in ddH_2O
Chloramphenicol stock solution	50mg/ml in absolute ethanol
Kanamycine stock solution	25 mg/ml in ddH_2O
Glycerol solution	10% (w/v) glycerol in ddH_2O
TFB1 buffer	100 mM RbCl; 50 mM $MnCl_2$; 30 mM potassium acetate; 10 mM $CaCl_2$; 15% (w/v) glycerol. pH=8
TFB2 buffer	10 mM MoPS; 10 mM RbCl; 75 mM $CaCl_2$; 15% (w/v) glycerol. pH=6.8
BugBuster protein extraction reagent	Supplied by Novagen®
Wash buffer	50 mM NaH_2PO_4; 300 mM NaCl; 20 mM imidazole. pH=8
Elution buffer	50 mM NaH_2PO_4; 300 mM NaCl; 250 mM imidazole. pH=8
Protein storage buffer	20 mM HCl; 0.1 mM EDTA; 1 mM DTT; 100 mM KCl; 200 µg/ml BSA; 50% (w/v) glycerol. pH=7.4
2X SDS-PAGE sample loading buffer	0.09 M Tris.Cl; 20% (w/v) glycerol; 2% (w/v) SDS; 0.02% (w/v) bromophenol blue; 0.1 M DTT.

33% Acrylamide-bis acrylamide solution	Provided by Bio-rad®
Protein dialysis buffer	50 mM Tris-base; 0.1 mM EDTA; 0.1 mM DTT. pH=7.5
Ni-NTA	Provided by Qiagen®
50X TAE buffer	242 g/litre Tris base; 57.1 ml/litre glacial acetic acid; 100 ml/litre of 0.5 M EDTA. pH=8
5X Tris-glycine buffer	125 mM Tris; 1.25 M glycine; 0.4% (w/v) SDS. pH=8.3
Agarose	Provided by Bioline®
Ethidium bromide	Provided by Sigma®
DNA markers (Ladders)	Provided by Bioline®
DNA Gel loading dye blue (6X)	Provided by New England Biolabs® (NEB)
4X SDS-PAGE resolving buffer	1.5 M Tris; 0.4% (w/v) SDS. pH=8.8
4X SDS-PAGE stacking buffer	0.5 M Tris; 0.4% (w/v) SDS. pH=6.8
Ammonium persulphate solution (APS)	10% (w/v) APS in ddH_2O
NNNNTetramethylenediamine (TEMED)	Provided by Sigma®
Coomassie staining solution	0.05% (w/v) comassie brilliant blue R-250; 40% (v/v) ethanol; 10% (v/v) glacial acetic acid
De-staining solution	25% (v/v) ethanol; 5% (v/v) glacial acetic acid; 70% (v/v) ddH_2O
10% SDS-PAGE resolving gel	2.5 ml of 4X SDS-PAGE resolving buffer; 4 ml ddH_2O; 3.3 ml of 33% (w/v) acrylamide-bis acrylamide solution; 200 µl of 10% (w/v) APS solution; 40 µl TEMED.

SDS-PAGE stacking gel	1.5 ml of 4X SDS-PAGE stacking buffer; 3.4 ml ddH_2O; 1 ml of 33% (w/v) acrylamide-bis acrylamide solution; 100 µl of 10% (w/v) APS solution; 20 µl TEMED.
IPTG stock solution	1 M IPTG
Qiagen plasmid miniprep kit	Provided by Qiagen®. Contains: buffer P1; buffer P2; buffer N3; buffer PB; buffer PE; buffer EB; LysBlue reagent; RNase A; silica spin columns; and collection tubes (2 ml).
QIAquick Gel Extraction Kit	Provided by Qiagen®. Contains: buffer QG; buffer PE; buffer EB; QIAquick spin columns; loading dye; and collection tubes (2 ml).
QIAquick PCR Purification Kit	Provided by Qiagen®. Contains: buffer PB; buffer PE; buffer EB; QIAquick spin columns; loading dye; and collection tubes (2 ml).
Phusion high-fidility PCR kit	Provided by NEB®. Contains: Phusion DNA polymerase; Deoxynucleotide solution mix; 5X Phusion HF buffer; 5X Phusion GC buffer; $MgCl_2$ (50 mM); DMSO.

2.1.2. Modification and restriction enzymes

2.1.2.1. Modification enzymes

Enzyme	Source
Benzonase	Novagen®
T4 DNA ligase	NEB®
Shrimp Alkaline Phosphatase (rSAP)	NEB®
S-PhoEP DNA polymerase	Prepared during this study from a construct provided by Prof. David Hornby

2.1.2.2. Restriction enzymes

HindIII; BcII; BmtI; AflII; XbaI; NcoI; BamHI; EcoRI; SpeI; and EagI.
All restriction enzymes have been provided by New England Biolabs®.

2.1.3. Oligonucleotides

All DNA oligonucleotides were designed using clone manager software with annealing temperature of 55°C and sent for the synthesis facility of Eurofins MWG®. They are all listed in the following table:

Name of oligo	Sequence and use
1st set F New	5'-ATGCGTTCGCGTTTCTGCGGAATAC-3' Amplification the first third part (flanked by HindIII + BmtI) of M.HhaI gene.
1st set R New	5'-GGCAGTGAGCGCAACGCAATTAATG-3' Amplification the first third part of M.HhaI gene and sequencing over the mutagenesis area between HindIII and BmtI.

3rd set F New	5'-GCACATTCACGCGGATGCAGTTTAC-3' Amplification the second third part (flanked by BmtI + AflII) of M.HhaI gene.
3rd set R New	5'-CCGGCTTTCCGTGTCAGGCATTTAG-3' Amplification the second third part of M.HhaI gene and sequencing over the mutagenesis area between BmtI and AflII.
4th set F New	5'-TGTGCTGCAAGGCGATTAAGTTGGG-3' Amplification the last third part (flanked by AflII + XbaI) of M.HhaI gene.
4th set R New	5'-GCCGGATAGCGAAGTTGAACATCTG-3' Amplification the last third part of M.HhaI gene and sequencing over the mutagenesis area betweenAflII and XbaI.
1st At F	5' TCAAGGCCACGTGTCTTGTCCAGAG 3' Sequencing to check the CATI gene sequence integrity after *E. coli* transformation with pAtase2 synthetic plasmid construct (this primer covers the sequencing for the first part of the gene).
3rd At F	5' CCAGGACGTCGCATGTTATGG 3' Sequencing to check the CATI gene sequence integrity after *E. coli* transformation with pAtase2 synthetic plasmid construct (this primer covers the sequencing for the second part of the gene).
2nd At F	5' CAGACCGTTCAGCTGGATATCAC 3' Amplification the part of CATI gene targeted in mutagenesis and sequencing over the mutagenesis area betweenEcoRI and XbaI.
3rd At R	5' CAGGTTTCCCGACTGGAAAGC 3' Amplification the part of CATI gene targeted in mutagenesis (flanked by EcoRI and XbaI).

1st lpd F	5' CAAGGTAGCAAGCGCCAGAATCC 3' 1- Amplification the part of lpd gene flanked by NcoI and Af1II and sequencing over this mutagenesis area. 2- Amplification the part of lpd gene flanked by NcoI and EagI and sequencing over the first part of this mutagenesis area. 3- Covering the first part in the gene sequencing to check lpd gene sequence integrity after *E. coli* transformation with pLpd1 synthetic plasmid construct.
1st lpd R	5' CTAGTAGACGTTCCGGCACTTCC 3' Amplification the part of lpd gene flanked by NcoI and Af1II.
2nd lpd F	5' ATTGCAGCCGGTAGCCGTCCGATTC 3' 1- Amplification the part of lpd gene flanked by Af1II and XbaI and sequencing over the first part of this mutagenesis area. 2- Covering the second part of sequencing after mutating the area flanked by NcoI and EagI of the lpd gene. 3- Covering the second part in the gene sequencing to check lpd gene sequence integrity after *E. coli* transformation with pLpd1 synthetic plasmid construct.
2nd lpd R	5' CGGTCAGACCAACCCATGCAACTTC 3' Amplification the part of lpd gene flanked by NcoI and EagI.
3rd lpd F	5' CAGCTGCGTACCAATGTTCC 3' 1- Covering the second part of sequencing after mutating the area flanked by Af1II and XbaI of the lpd gene. 2- Covering the third part in the gene sequencing to check lpd gene sequence integrity after *E. coli* transformation with pLpd1 synthetic plasmid construct.

3rd lpd R	5' AGCTGGCACGACAGGTTTCC 3' Amplification the part of lpd gene flanked by Af1II and XbaI.

2.1.4. Plasmids

All the plasmids listed in the following table were either stored at -20°C as plasmid DNA minipreps in ddH_2O for a short term use or kept inside suitable *E. coli* strains which in turn were stored at -80°C with 20% (w/v) glycerol for the long term use.

Name	Source	Description
pUC19	New England Biolabs®	Cloning vector
S-PhoEP	Prof. David Hornby	Carries S-PhoEP gene
pMJA01	Developed in this work	Carries M.HhaI gene
pAtase2	Developed in this work	Carries CATI gene
pLpd1	Developed in this work	Carries *lpd* gene

2.1.5. Bacterial strains

All bacterial strains listed in the following table were stored on agar plates sealed by parafilm at 4°C for short term use or within 20% (w/v) glycerol at -80°C for the long term storage.

Name	Source	Genotype
E. coli BL21 (DE3)	Novagen®	F^- *ompT hsdS*$_B$ ($r_B^-m_B^-$) *dcm gal* λ (DE3)
E. coli DH5 alpha	New England Biolabs®	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*
E. coli K12	New England	*F- glnV44 e14-(mcrA-) rfbD1? relA1?*

ER1821	Biolabs®	*endA1 spoT1? thi-1 Δ(mcrC-mrr)114::IS10*
E. coli K12 JM109	New England Biolabs®	*F´ traD36 proA+B+ lacIq Δ(lacZ)M15/ Δ(lac-proAB) glnV44 e14- gyrA96 recA1 relA1 endA1 thi hsdR17*
E. coli K12 W3110 (JRG3503)	Prof. Jeff Green (MBB University of Sheffield)	*F^- lambda$^-$ IN(rrnD-rrnE)1 rph-1 (lpdA::kan^R)*

2.1.6. Equipment

Equipment	Provider
Eppendorf tubes of 0.5 ml and 1.5 ml	Sarstedt®
PCR tubes of 0.2 ml	Bio-Rad®
Falcon tubes	Sarstedt®
Universal tubes	Sarstedt®
Flasks, cylinders, beakers, bottles and durans of different sizes	
Protein affinity separation columns	Qiagen®
Protein dialysis tubes	Novagen®
Pipette tips	RAININ instrument Co.
Pipettes	Costar®
Micropipettes	Gilson Pipetman®
Petri dishes	Thermo Scientific®
Parafilm	Pechiney Plastic Packaging®
Autoclave tape	
Gloves	Glove plus®
Autoclave	Phoenix®
Water bath	Grant Instruments®
Deep freezer	Scientific Laboratory

	Supplies®
Refrigerators	
Incubator room (37°C)	MBB department
Manual Ultra-violate transilluminator model TM40	Genetic Research Instruments®
Computerized gel documentation	Syngene®
UV and Visible Spectrophotometer	Perkin Elmer®
Spectrophotometer 10X4X45 mm cuvettes	Sarstedt®
Magnetic stirrers and shakers	
Heat block	Grant Instruments®
Vortexer	Labnet®
Agarose gel electrophoresis unit	Bio-Rad®
SDS-PAGE electrophoresis unit	Bio-Rad®
Electrophoresis power supplies Power Pac 300	Bio-Rad®
PCR machine	Senso Quest®
Loops and spreaders	
Centrifuges: Centaur 2, and Microcentaur Beckman J-251, J-6M/E and J201 CentriconY10 centrifugal concentrator	Sanyo/MSE® Beckman® Amicon®
Syringes	Becton Dickinson®
Syringe filters	Nalgene®
Balances: 0-100g Model AJ 100 and 0-800g Model K7	Mettler®
pH meter Basic with Tris Electrode	Denver Instrument Company®

MicroPulser electroporator	Bio-Rad®
electroporation cuvettes of 0.1 cm and 0.2 cm electrode gap width	Bio-Rad®
Rotary shaker	

2.2. Methods

2.2.1. Storage of bacterial strains

Whether transformed by plasmid(s) or not, all bacterial strains were stored for both short and long term purposes. The short term storage was carried out by streaking the bacteria on an appropriate agar plate with or without antibiotic(s) and incubated overnight then the plates were sealed with parafilm and kept at 4°C to be used for no longer than two weeks. While the long term storage was performed by culturing the bacterial strain in 5ml liquid LB with or without antibiotic(s) (according to their resistance) and incubated overnight at 37°C with vigorous shaking at 250 rpm then 800 µl of the bacterial growth is mixed with 200 g of autoclaved sterile glycerol and frozen at -80°C.

2.2.2. Preparation of competent *E. coli* cells for chemical transformation

Competent cells were prepared according to (Hanahan, 1983). Bacterial cells were taken from their original glycerol stocks and streaked on LB agar then a single colony was taken and grown on LB medium overnight, a 250µl aliquot from the latter growth was used to inoculate 250ml of LB then incubated at 37°C until OD_{600nm}= 0.45-0.5. The cells were harvested by centrifugation (4000xg for 15minutes at 4°C) then resuspended in 8ml TFB1 buffer an

incubated on ice for 10-15min. The cells were precipitated again by centrifugation as above then resuspended in 2 ml TFB2 and incubated on ice for 15 min. Finally, competent cells were aliquoted in Eppendorf tubes (100 µl in each tube) and stored at -80°C until use.

2.2.3. Preparation of electrocompetent *E. coli* cells

A conical flask containing 50 ml LB and 25 µg/ml kanamycin was inoculated by 500 µl of *E. coli* K12 W3110 (JRG3503) from an overnight culture and incubated at 37°C with vigorous 250 rpm shaking until OD_{600} reached 0.5-0.7, then the cells were chilled on ice for 20 minutes and harvested by cooled centrifugation (4000xg for 15 minutes at 4°C) while LB supernatant was discarded. The cell pellet was resuspended in 50 ml of 10% (w/v) glycerol solution, mixed by vortexing and pelleted again as above then 25 ml of 10% (w/v) glycerol solution was added to the cells for a second resuspension, pelleted as before and resuspended a third time in 5 ml of 10% glycerol solution and then harvested as above. Finally, the cells were resuspended using 1-2 ml of 10% (w/v) glycerol solution, aliqouted as 50-100µl in microcentrifuge tubes and stored at -80°C until required.

2.2.4. Transformation efficiency estimation

Prior to the experimental transformation, the efficiency of competent cells in accepting new plasmid DNA was tested by transforming (following either chemical or electroporation transformation protocols as mentioned above) the bacteria with pUC19 plasmid vector of an estimated concentration (which is estimated by using DNA nanodrop) then the bacterial CFUs are counted and the

unit of efficiency is designed as the number of CFUs per 1 µg of pUC19 DNA plasmid.

2.2.5. Plasmid DNA preparation

Plasmid DNA was prepared according to Qiagen® protocol by using the Qiagen® miniprep kit. Firstly, bacteria from a single colony were grown in 5 ml LB with appropriate antibiotic(s) (100 µg/ml ampicillin, 25 µg/ml chloramphenicol, 25 µg kanamycin or combinations according to the selection required) overnight. Cells were precipitated by centrifugation (4500 rpm for 10 min at 4°C). The cell pellet was resuspended in 250 µl P1 buffer and mixed by vortex. Two hundred and fifty µl of P2 buffer was added then mixed by inverting gently until the mixture turns blue indicating cell lysis. Three hundred and fifty µl of buffer N3 was added and mixed by inversion 6 times until the colour turns white. The mixture was centrifuged at full speed 14000 rpm for 10 min. The supernatant was taken and applied into the silica spin column (left for 1min for binding) then centrifuged at full speed for 1min, the flow through was discarded and 750 µl of buffer PE was applied into the column for washing then centrifuged at full speed for 1 min, the flow through was discarded and another centrifugation was done to remove ethanol traces from buffer PE. Finally, the silica column was put into a fresh Eppendorf tube, plasmid DNA was eluted from the silica column by adding 30-100 µl ddH_2O or TE buffer to the centre of the silica then centrifuged at full speed and stored at -20°C.

2.2.6. Chemical transformation of *E. coli*

Bacterial transformation was done according to protocols provided by Novagen. The competent cells of known transformation efficiency ware defrosted on ice until the last ice particle disappeared then mixed with an estimated DNA quantity and left on ice for 30 min. Heat shock was done by

placing the transformation mixture in 42°C water bath for 90 seconds then returned back to the ice immediately and kept on it for 2 min. 500 µl of LB or SOS medium was added to the mixture and incubated for 1hour at 37°C. Finally, 100 µl of the mixture was spread on an appropriate selection agar plate containing antibiotic(s) and incubated overnight at 37°C.

2.2.7. Electroporation

Electroporation was performed according to guidance from Bio-Rad® as follows: Electrocompetent *E. coli* cells were thawed on ice for approximately 10 minutes. Electroporation cuvettes of either 0.1 or 0.2 cm gap width are kept on ice as well prior use. Once the competent cells thawed, the plasmid DNA or the ligation mixture was mixed with the *E. coli* cells in a microcentrifuge tube and kept on ice for about 1 minute. Then an appropriate amount of the cells and DNA mixture was transferred into a cooled electroporation cuvette with tapping to ensure the settlement of the mixture in the bottom of the cuvette and to avoid bubbles formation. Then the cuvette was put in the chamber slide and pushed into the electroporation chamber to be between the micropulser contacts then pulse was applied once at appropriate voltage (KV) and milliseconds time constant according to the gap width of the cuvette. However, voltage that should be used with *E. coli* is already set in the micropulser from Bio-Rad® to be 1.8 or 2.5 KV with 0.1 and 0.2 gap width cuvettes respectively, while the millisecond time constant is auto optimised by the micropulser depending on the conductivity of the mixture (i.e. the salt content). After the pulse application, 500-900 µl of SOC or LB was quickly added to the mixture while still in the cuvette then the whole mixture was transferred into a 1.5 ml microcentrifuge tube and incubated with rotation at 37°C for one hour then the transformed cell

were centrifuged at 10000 rpm for 1 minute and plated on an appropriate selection plate.

2.2.8. Protein purification

Error-prone Pho DNA polymerase is expressed from a lac T7 promoter expression vector S-PhoEP. BL21 (DE3) Rosetta strain was used for expression. The purification was done according to guidance provided by Qiagen.

2.2.8.1. Cells growth and induction

Five hundred ml of LB (containing 100µg/ml ampicillin and 34 µg/ml chloramphenicol) was inoculated with 10 ml of fresh BL21 (DE3) Rosetta growth and incubated at 37°C until OD_{600nm}=0.6. The cells were induced to express the protein by adding IPTG to a final concentration of 1mM then incubation proceeded for 4-5 hours. Cells were harvested using pre-weighted tubes by centrifugation (4000 xg for 20 min at 4°C), then cell pellet was frozen at -20°C overnight to aid cell breakage.

2.2.8.2. Cell breakage

Cell lysis was done by BugBuster (primary amine-free) Protein Extraction Reagent from Novagen® (5ml BugBuster per 1 gm cells), 25 units of benzonase were added for each 1 ml of BugBuster to get rid of the viscosity caused by nucleic acids, the cell suspension was kept on rotating mixer at a slow setting for 10–20 min at room temperature. Cell debris was removed by centrifugation (16000 xg for 20 min at 4°C), the supernatant was transferred to a fresh tube.

Since our target is heat resistant protein, the supernatant was heat treated at 85°C for 10 min (to remove heat sensitive proteins) then centrifugation (16000 xg for 20 min), the supernatant was taken in a fresh tube for affinity chromatography purification.

2.2.8.3. Nickel-NTA affinity chromatography

Since our target protein is 6xHis-tagged, Ni-NTA was used to specifically bind and purify it. The protocol was carried out according to Qiagen®. The second supernatant obtained after cell lyses was mixed with 1ml of Ni-NTA slurry and kept with rotary shaker for 60 min at 4°C for binding. The lysate-Ni-NTA mixture was loaded into a column with the bottom outlet capped. The bottom cap was removed allowing the flow-through to leave the column. The column was washed twice with 4 ml wash buffer. Finally, the protein was eluted 4 times with 0.5 ml elution buffer each time.

2.2.8.4. Protein dialysis

The elution fraction was dialyzed by putting it in a dialysis tube from Novagen which was in turn kept floating and rotating on a magnetic stirrer overnight at 4°C in 2 litters of freshly prepared protein dialysis buffer.

2.2.8.5. Protein storage

Storage buffer was added to the protein after dialysis and stored at -20°C.

2.2.9. Sodium dodecyl sulphate-polyacrylamidegel electrophoresis (SDS-PAGE)

Sodium dodecyl sulphate-polyacrylamidegel electrophoresis (Laemmli, 1970; Maniatis *et al.*, 1982) was performed for the whole cell lysate, soluble fraction before heating, soluble fraction after heating, Ni-NTA flow-through, washing fraction and elution fraction to track the protein through the purification steps. Ten µl from each sample was mixed with 10 µl of 2x protein loading buffer, heat treated (95°C for 10 min) then loaded into polyacrylamide gel (which is prepared as two layers: stacking and resolving layer) and electrophoresed under denaturing conditions at 160 V in the presence of tris-glycine buffer until the dye reached the bottom of gel. Proteins bands were visualized by keeping the gel in staining solution for 1-2 hours then de-staining overnight to remove the stain from the gel except proteins bands.

2.2.10. DNA agarose gel electrophoresis

Agarose electrophoresis was used to visualize DNA bands of: plasmid minipreps (to estimate DNA concentration and to ensure about minipreps), PCR products (to ensure about amplification) and restriction digestion. The gel was almost always prepared as 1% (w/v) agarose which is microwave heat dissolved in 1X TAE buffer until the mixture is clear then transferred to a water bath to cool to 50°C then supplied with ethidium bromide at final concentration of 0.5 µg/ml and poured in the gel template containing a comb of an appropriate size to form the required number of wells. DNA samples were mixed with 6X loading buffer, loaded into the wells and electrophoresed under 100 V for 1 hour in presence of 1x TAE buffer. After that, DNA bands were visualized by U.V. trans illuminator.

2.2.11. DNA restriction digestion

Restriction enzymes from New England Biolabs® were used to cut plasmid vectors and inserts to provide compatible ends prior ligation as well as to check the consistency of constructs after ligation. The reactions were done according to the leaflets of the enzymes supplier with some modifications. DNA concentration was always estimated prior to the restriction digestion so that an appropriate number of enzyme units was added. The reaction volume was almost always 20 µl with an incubation time between 1-2 hour at 37°C taking into consideration the star activity behaviour of some restriction enzymes which could be always avoided by using the high fidelity versions of those enzymes.

2.2.12. Removal of 5′ phosphates from the DNA

The dephosphorylation of the 5′ ends of the linearized vectors was occasionally carried out to avoid vector self-ligation. For this purpose, Shrimp Alkaline Phosphatase (rSAP) from NEB® was used and the reaction was set according to the manufacturer guidance. rSAP is working in its full activity in all NEB® restriction enzymes buffers, therefore, there was no need to purify the linearized vectors from the restriction digestion reaction prior to the dephosphorilation reaction so that rSAP was add directly just after the incubation period of the restriction digestion reaction. 1 unit of rSAP has been added for every 1 pmol of DNA ends (about 1 µg of a 3 kb plasmid) and incubate at 37°C for 30–60 minutes then heat inactivated at 65°C for 5 minutes.

2.2.13. DNA ligation

Bacteriophage T4 DNA ligase from New England Biolabs® was used to ligate plasmid vectors and inserts (PCR products) of compatible cohesive

(sticky) ends. The reaction was done according to NEB® guidance. The DNA concentration of both insert and vector was always determined before mixing the reaction components. The proportion of concentration of the vector to insert was between 1:3 to 1:10 in a total reaction volume between 20-50 µl and incubation period of 4 to 16 hours at 16°C. The ratio of the vector:insert amounts is calculated according to the formula: required mass insert (g) = desired insert/vector molar ratio X mass of vector (g) X ratio of insert to vector lengths.

2.2.14. PCR mutagenesis amplification

The mutagenesis was done by error-prone PCR using 1-3 µl of error-prone DNA polymerase S-PhoEP developed in our laboratory and purified during this study as mentioned previously. The PCR reactions (50 µl) were done following protocol of NEB® in term of contents amounts and concentrations. Contents of the NEB® PCR kit including dNTPs, $MgCl_2$, and PCR reaction buffers were used. Many oligonucleotides primers were designed by Clone Manager software so that the annealing temperature for all of them is 55°C, therefore, conditions of all reactions were the same and as follows: initial denaturation at 94°C for 2 min., denaturation at 94°C for 30 sec, annealing at 55°C for 30 second, elongation at 72°C for 45 sec, and final elongation at 72°C for 10 min. The reaction mixture was as follows: 10 µl of 5X GC phusion PCR buffer, 1 µl of 10 mM dNTPs, 2.5 µl of 10 µM forward primer, 2.5 µl of 10 µM reverse primer, 1 µl of plasmid DNA template of variable concentrations (>250 ng), 1-3 µl of S-PhoEP DNA polymerase, 1 µl of 50 mM $MgCl_2$, and ddH_2O nuclease free water to complete the volume to 50 µl.

2.2.15. PCR product clean up

The PCR product was cleaned according to the guidance from Qiagen® in order to render it pure enough for optimising the subsequent cloning reactions, such as restriction digestion and ligation. Five volumes of buffer PB was added to one volume of PCR reaction and mixed well by inverting the tube gently. The mixture was then applied to a silica spin column and left for 1 minute for binding to occur. Centrifugation was then carried out at full speed for 1 minute to remove the liquor. Seven hundred and fifty µl of PE buffer was applied to the silica column for washing, followed by a full speed centrifugation for 1 minute. Another centrifugation was done as before to remove any traces of PE buffer and the DNA was finally eluted by applying ddH_2O to the centre of the silica column.

2.2.16. DNA fragmentsrecovery from agarose gel

Fragments of interest of both vector and insert were recovered from the gel using a Qiagen® gel recovery kit and according to Qiagen® protocol. The gel was placed on U.V. trans illuminator, fragments were cut using a clean sharp surgical blade, each fragment was transferred into a pre weighted Eppendorf tube. The weight of the gel slice was taken and QG buffer was applied into the tube in a proportion of 3:1 according to the gel slice weight; for example, if the weight was 100 mg, 300 µl of QG was added and so on. After the QG addition the tube was placed in 55°C water bath for 10-15 minutes until agarose was completely melted. Isopropanol was added in a proportion of 1:1 depending on the weight of the gel slice (for example100 µl of isopropanol for 100 mg gel slice) and mixed by inverting. The mixture was then transferred on to a silica spin column and centrifuged at full speed for 1 min., the flow-through was discarded. Washing was done by adding 500 µl QG buffer and centrifugation at

full speed for 1 min., the flow-through was discarded. A second wash was done as above but with 750 µl PE buffer, ethanol traces were removed by a further spin for 1 min at full speed. Finally, the column was placed into a fresh Eppendorf tube and DNA was eluted by applying 30-50 µl of ddH_2O on to the centre of the column.

2.2.17. DNA sequencing and data analysis

The DNA sequencing facility which is situated in the Medical School at the University of Sheffield was used to sequence all the plasmid DNA samples extracted from the clones developed during this study. The sequencing chromatograms visualised and analysed using Chromas Lite software. Each nucleotide sequence was compared with the original DNA sequence to locate the nucleotide changes; for this purpose, the web NCBI BLAST nucleotide sequence comparing tool was used. The online translation tools from New England Biolabs® and ExPASy Bioinformatics Resource Portal were used to translate nucleotide sequences into their peptide chains to evaluate and allocate the amino acids changes patterns after comparing the mutant clones amino acid sequence with the wild type sequence using the web NCBI BLAST protein sequence comparing tool.

Chapter 3 M.HhaI DNA C5-methyltransferase: Random mutagenesis and functional screening

Abstract

In this chapter, the experimental random mutagenesis using error-prone Pho DNA polymerase has been applied to the M.HhaI gene. This gene is naturally translated by *Haemophilus parahaemolyticus* to express M.HhaI DNA methyl transferase that recognizes the sequence GCGC within the DNA double helix and adds a methyl group (CH_3) to the inner cytosine. Mutant clones were screened for biological and biochemical function and grouped into active, inactive and partially active clones which are then analysed according to the available three dimensional structure and the previous data related to this enzyme. The activity screening data showed that the enzyme is significantly robust and could "accept" several primary structure changes even at the conserved locations, without loss of activity. On the other hand, few mutant clones showed impaired enzyme activity unexpectedly when points of no conservation were mutated. The experimental work in this chapter provides a platform for optimising the methodology to be applied on other target genes.

3.1. Introduction

There are a number of biological modifications made to genomic DNA; one of the earliest to be identified was methylation, which involves the addition of a single methyl group (CH_3) on to either cytosine or adenine within the DNA double helix (Iqbal *et al.*, 2011). The class of enzymes responsible for this activity are termed DNA methyltransferases, abbreviated here to DNA MTases (Adams, 1990). In some eukaryotes, it has been shown that DNA methylation is involved in the regulation of gene expression (Cedar, 1988). For example, it has been shown that methylation of cytosine within the promoter region of RNA polymerase II genes in higher organisms can inhibit transcription (Besser *et al.*,

1990). Experimental investigations of methylation and its role in general gene expression, differentiation and development were clarified when Li *et al*. (1992) used gene knockout technology to demonstrate that the gene encoding DNMT1 is essential for normal embryonic development. Since this work, the sequencing of a number of genomes, including mice and human, has revealed the presence of several cytosine-C5-specific DNA MTases (Jeltsch, 2006). Several publications have described more specialised roles for cytosine-C5 DNA MTases in recombination as well as replication phenomena (Wilkinson *et al.*, 1995; Goyon, 1997; Malagnac *et al.*, 1997).

In a number of prokaryotes, in particular *E. coli*, DNA methylation is involved in a wide range of biological functions, including restriction and modification. This phenomenon, in which DNA of the host is methylated in order to "protect" it from host mediated "restriction" (sequence-specific hydrolysis) has often been described as a primitive "immune system"(Boyer, 1971; Wilson and Murray, 1991). One other role for DNA methylation in bacteria is the correction of mistakes that arise during DNA replication (Modrich, 1991). It has been found that some DNA adenine methyltransferases (Dam) are involved in the expression of certain virulence factors genes in enteric bacteria such as *Salmonella typhimurium*, therefore, Dam^- strains lack their virulence (Heithoff *et al.*, 1999).

Clearly, the biological role of restriction and modification enzymes is more complex than a primitive immune system, but for the purposes of this work, that is how it will be considered. In our methodology, we use this principle to screen mutants that are either proficient or deficient in methylation activity, using a simple restriction based method.

3.2. Classification of DNA Methyltransferases

DNA MTases have been classified by the Enzymes Commission (EC) according to the reaction that they catalyze as shown in Table 3.1.

Table 3.1: Classes of DNA methyltransferases:

DNA MTase class	**Description**
N-6 adenine-specific DNA methyltransferases (A-Mtases or m6A)	These enzymes add a methyl group to the amino group at the C-6 of adenine. They are part of several types of restriction and modification systems in bacteria. Methylation by these enzymes protects the same DNA sequence from cleavage by the cognate restriction enzyme (Loenen *et al.*, 1987).
N-4 cytosine-specific DNA methyltransferases (C4-Mtases or m4C)	These enzymes are responsible for methylation of the amino group at C-4 position of the cytosine ring. Some Type II restriction and modification systemC4-Mtases act to protect DNA against the corresponding restriction enzymes (Timinskas *et al.*, 1995).
C-5-cytosine-specificDNA methyltransferases (C5-Mtases or m5C)	Methylate the C5 position of the cytosine ring. In mammals, these enzymes are known to be involved in modifying gene expression and cell differentiation, on the other hand, they are a part of restriction and modification systems in prokaryotes (Kumar *et al.*, 1994).

3.3. Structure and architecture of HhaI methyltransferase (M.HhaI)

M.HhaI belongs to the cytosine C-5 specific group of DNA MTases (EC 2.1.1.37), abbreviated to C5-Mtases. All enzymes in this class are built from ten conserved sequence motifs and appear to share a common architecture (Yen *et al.*, 1992). There is a well characterised variable region in all C5-Mtases situated between motifs VIII and IX (Figure 3.1). This region shows the greatest level of amino acid sequence diversity among the majority of bacterial C5-MTases and is referred to as the target recognition domain (TRD): this region recognizes the DNA sequence and the cytosine to be methylated (Klimašuaskas *et al.*, 1991).

M.HhaI consists of 327 amino acids (molecular mass =37 kDa), and is one of the smallest C5-Mtases identified to date. It recognizes the sequence 5'-GCGC-3' within double stranded DNA and methylates the middle cytosine, to form 5'-GC^{CH3}GC-3'. This enzyme is part of the HhaI restriction and modification system of *Haemophilus haemolyticus* (Roberts *et al.*, 1976; Mann and Smith, 1979).

Cheng *et al.* (1993) determined the three dimensional structure of M.HhaI using X-ray crystallography. The structure is illustrated in Figure 3.2. The enzyme is composed of three main parts: a large domain, a small domain and a segment called hinge region, which connects the two domains. The large domain constitutes two thirds of the enzyme (amino acids 1-193 of the N-terminal and 304-327 of the C-terminal sequence). The two stretches of primary structure forming the large domain are separated by the TRD and the hinge region. The segment encoded by amino acids 1-193 twists to form a mixture of α and β strands, while the C-terminal segment forms α-helix. The core of the large domain is composed of six β strands numbered 4, 3, 1 and 2 which are parallel and adjacent to each other. Next to strand 4, there is a hairpin-like

structure formed by strands 5 and 6. The core is crammed between two "layers": the upper layer, which is formed by helices C and D, and the lower layer which is formed by helices A and G with strand β7. These two layers form a sandwich-like structure.

The small domain is composed of seven loops, some of which exhibit β-strand character, but this is not regular. This region is encoded by amino acids 194-275, which are arranged in a "circular" pseudo-antiparallel manner to form a propeller-like structure. The hinge region is a mixture of α and β structures joining the large and small domains, forming the lower part of the cleft. The AdoMet-binding site lies just beyond the C-terminus of β-sheet 1 on the large domain side of the cleft.

3.4. M.HhaI motifs and tertiary structure

Since the early comparative analysis of DNA MTase sequences by (Pósfai *et al.*, 1989), many more C5-Mtase genes have been sequenced. It is now clear, that Pósfai et al (1989) were essentially correct, as confirmed by a number of crystal structures of DNA MTases. Among the members of C5-Mtases; six of the conserved motifs are considered highly conserved: these are labelled I, IV, VI, VIII, IX and X.

By way of an illustration of the conservation of primary structure among methyltransferases, an alignment of 36 m5C-Mtase sequences showed that most motifs are present in all C5-Mtases, however motifs IX and X in M.SssI and M.CviJI MTases are very poorly defined. Nevertheless, in general, the conserved motifs generate the structural framework for catalytic function in this class of enzymes (Cheng *et al.*, 1993).

Analysis of the variable region between conserved motifs VIII and IX across this group of C5-Mtases reveals a considerable level of diversity in the TRD. Even though it is a variable region, a typical TRD will contain a T-L dipeptide (Lauster *et al.*, 1989). A higher resolution analysis of the variable region has led to amino acids 171-271 being labelled as the variable region and amino acids 231-253 constitute the TRD, which is found on the surface of M.HhaI, forming a "stalk-like" structure which initially crosses over the whole protein length then folds to form the major portion of the small domain (Cheng *et al.*, 1993).

Most of the conserved motifs are contained within the large domain of M.HhaI, while the major part of the variable region is included inside the small domain. The second half of motif IX and the first half of motif X form the hinge region, while the first half of motif IX is a part of the small domain and the last half of motif X is contained in the large domain. It has been noticed that M.HhaI became more active when mutants were created by exchanging sequences from other motifs adjacent to motif IX and the variable region (Klimašuaskas *et al.*, 1991; Mi and Roberts, 1992).

The core structure of M.HhaI, as described by Cheng *et al.* (1993), is formed by the most highly conserved motifs which surround the active site cleft. These motifs include three α-helices (αA, αE and αG), four β-strands (β1, β3, β4 and β6) and strand β14 in the small domain in addition to four associating loops named: I-1A (between β1 and αA), I-3C (between β3 and αC), I-4D (between β4 and αD) and I-56 (between β5 and β6). These structures are thought to be crucial for the enzyme activity.

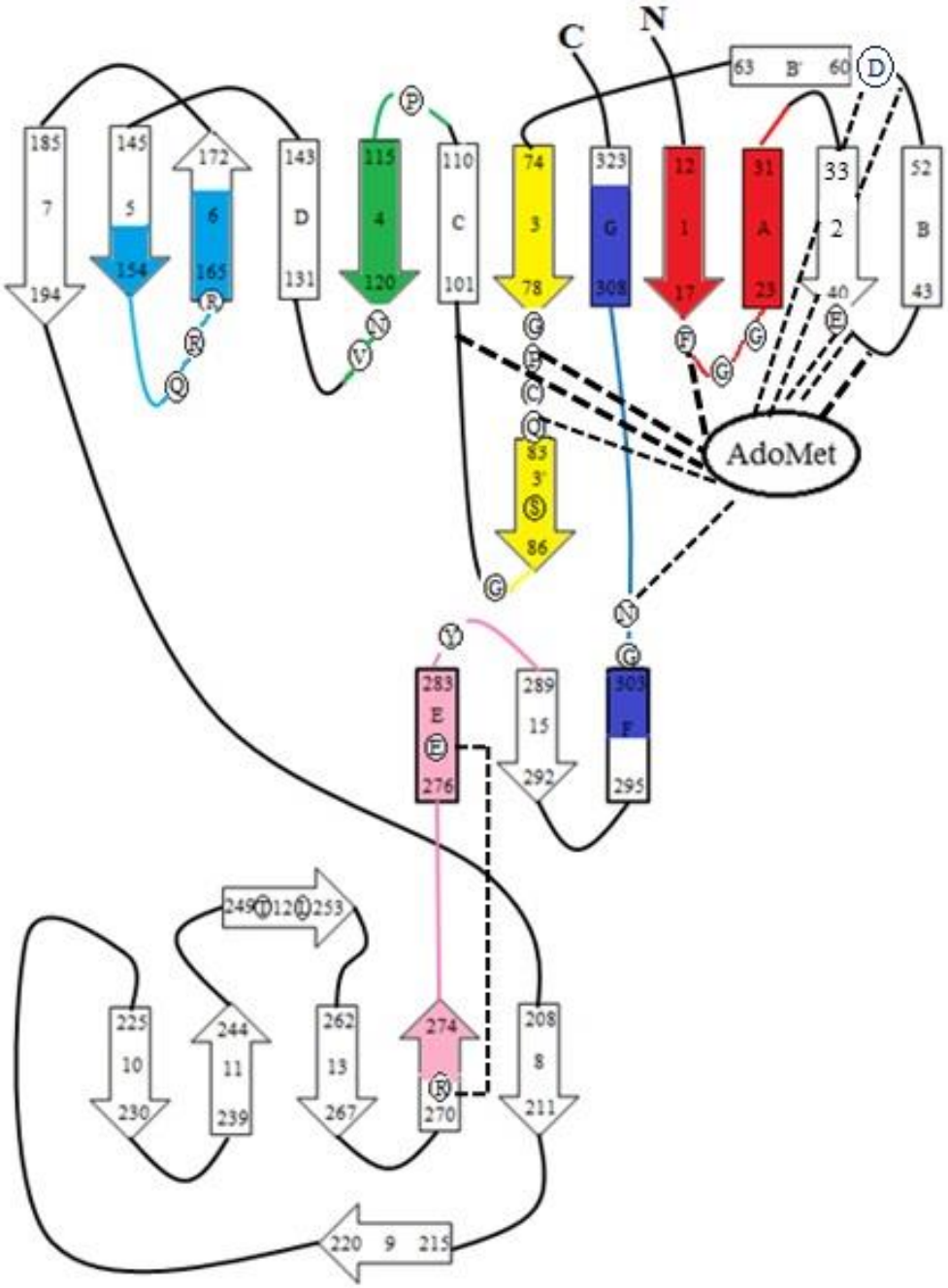


Figure 3.1.Schematic representation of the secondary structure of M.HhaI, showing the conserved motifs in different colours (red, motif I; yellow, motif IV; green motif VI; cyan, motif VIII; magenta, motif IX; dark blue, motifX).The invariant amino acids are circled. β-strands are represented as open, wide arrows labelled with numbers, while α-helices are represented as cylinders and letter labelled. Thick dashed lines represent hydrophobic interactions. Thin dashed lines represent electrostatic interactions or hydrogen bonds. This figure was taken from Cheng *et al.* (1993).

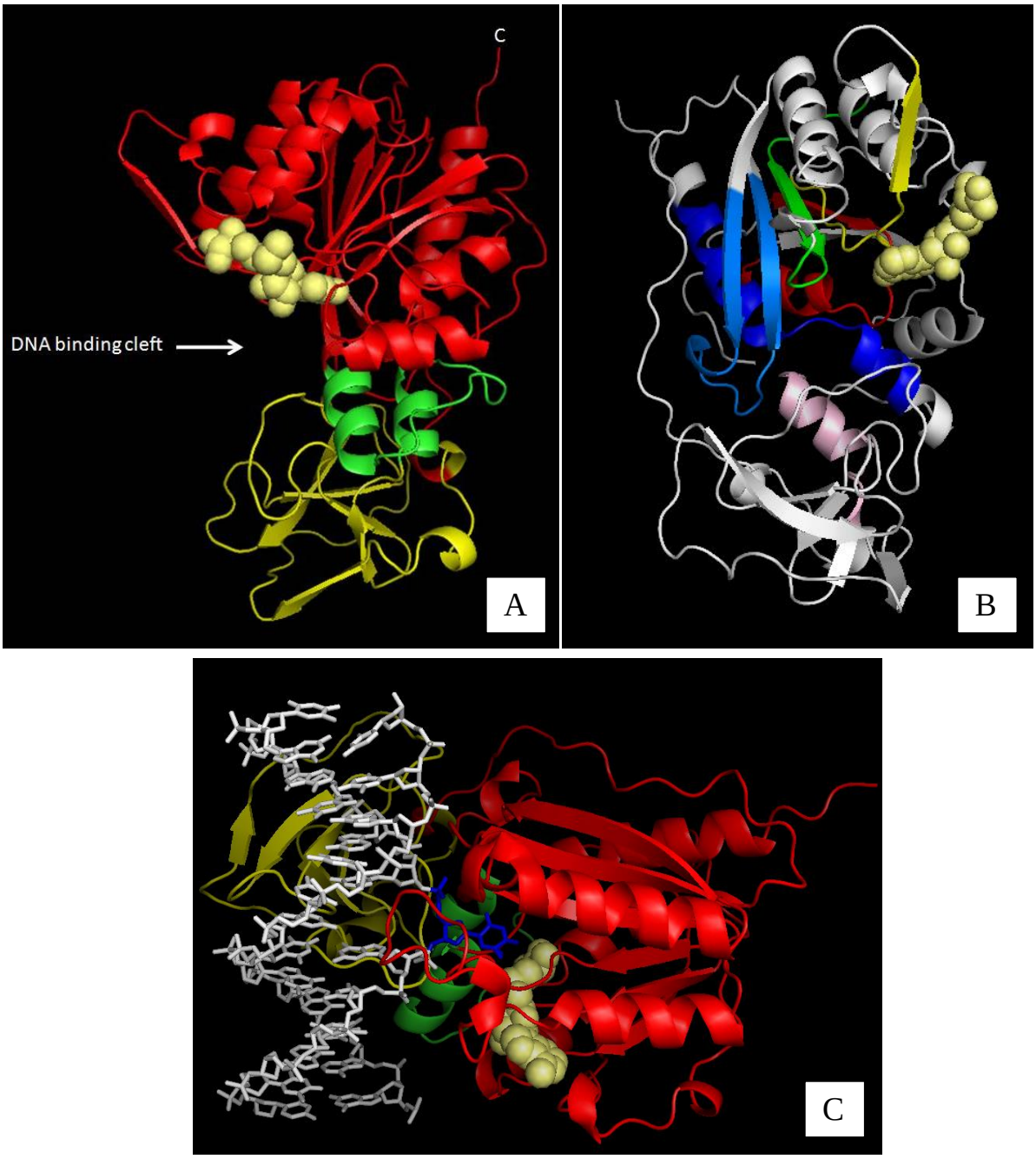


Figure 3.2: Ribbon diagrams showing: (A) large (red) and small domains (yellow), and hinge region (green). (B) Domains and motifs (coloured with the same colours of motifs which are shown in figure 3.1) organization from different angle. AdoMet is shown as pale yellow spheres. (C) Ternary complex showing M.HhaI, DNA substrate and AdoMet cofactor assembled together. The DNA double helix is shown in grey. The target cytosine is shown in blue and it appears clearly flipped out of the DNA helix to be within the enzyme active site. The AdoMet is shown as pale yellow spheres. Catalytic, TRD, and hinge

regions of the enzyme are shown in red, yellow, and green respectively. (A) and (B) were developed from the three dimensional structure resolved by (Cheng *et al.*, 1993), PDB code 1HMY, while (C) was developed from the three dimensional structure resolved by Klimasauskas *et al.* (1994), PDB code 1MHT.

3.5. The mechanism of methylation

DNA MTases catalyse DNA methylation by transferring a methyl group from the methyl donor S-adenosyl-L-methionine (AdoMet), to a cytosine, following its incorporation into DNA. In the case of C5-Mtases, the methyl group becomes attached via the C5 position of the pyrimidine ring (usually cytosine) producing modified cytosine in a series of specific stages that have been identified by a range of biochemical, chemical and biophysical approaches (Dryden, 1999).

Understanding the catalytic mechanism of MTases emerged from studies on both prokaryotic DNA MTases and mammalian enzymes in addition to the similarities between the mechanism of thymidylate synthase and C5-Mtases, led Santi's group (Wu and Santi, 1987) to propose a mechanism incorporating a transient covalent intermediate for M.HhaI based largely on steady state kinetic studies. Subsequently, Verdine's group using the suicide inhibitor, 5fluorocytosine (embedded in an oligonucleotide duplex), largely confirmed Santi's model using M.HaeIII. The first concrete evidence that the cytosine is flipped out of the double helix had to wait until the determination of the structure of a ternary complex for M.HhaI, incorporating a 5-fluorocytosine modified duplex (Klimasauskas *et al.*, 1994).

The reaction commences with the enzyme (possibly pre-loaded with AdoMet), folding around the target DNA, followed by target base rotation into the enzyme's active site. The cytosine to be methylated is firstly flipped out of

the DNA double helix; by flipping, cytosine is elegantly presented to the catalytic Cys residue and is accompanied by a significant loop displacement. The thiolate anion makes a transient covalent bond to the C-6 of the flipped ring and the cytosine is secured by several non-covalent interactions to facilitate methyl transfer. During this reaction, the orphan guanine is satisfied with compensating interactions that mimic the normal Watson Crick base pair interactions (Klimasauskas *et al.*, 1994).

The central stage in the methylation reaction involves the formation of a cytosine ring, in a Michaelis complex, predisposed to nucleophilic attack, methyl transfer and β elimination. The covalent bond is formed between the thiolate of the conserved Cys81 (in M.HhaI) and the C6 of the pyrimidine ring. This is apparently accompanied by a proton addition to the N3 position on the ring, to produce an enamine intermediate so that the C5 within the cytosine ring will be vulnerable to the methyl sulfonium centre attack of the AdoMet. The resulting intermediate of this attack is dihydro-cytosine which undergoes β-elimination to remove the C5 proton and C6 thiolate to produce the methylated cytosine (Vilkaitis *et al.*, 2001).

3.6. Design and synthesis of a modified, functional M.HhaI gene for mutagenesis

The nucleotide sequence of the gene encoding M.HhaI (gene bank accession number J02677.1) was re-configured in order to incorporate a range of unique restriction sites, without impacting on activity and minimising changes to the protein primary structure. (See Figure 3.3). This approach greatly facilitates the independent mutagenesis of segments of the target gene in a controllable manner. In designing the synthetic gene for M.HhaI, care was taken not to introduce restriction sites that are present in the chosen cloning vector,

pUC19. Prior to synthesis, codons were optimized for expression in *E. coli* and finally the sequence was submitted to a gene synthesis service (www.lifetechnologies.com).

After synthesis, the gene encoding M.HhaI was sub-cloned into pUC19 via HindIII and XbaI. The resulting recombinant plasmid was labelled pMJA01 (Figure 3.4). The success of the cloning, the sequence of the construct and the activity of the encoded product were verified and confirmed respectively (see Figures 3.4 and 3.5).

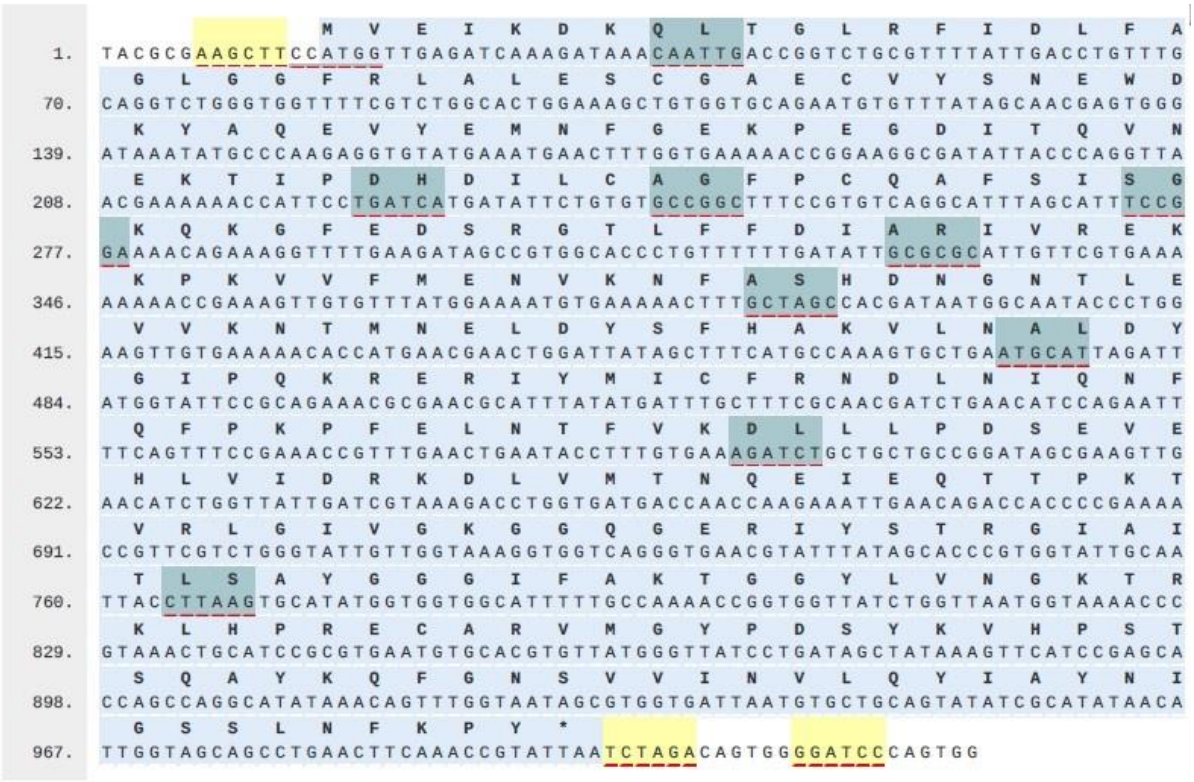


Figure 3.3: Interleaved view of the total synthetic gene sequence and primary structure of M.HhaI (327 amino acid). The ORF is cyan-shaded. The in-frame and out-of-frame restriction sites are underlined in red. (HindIII =AAGCTT, NcoI=CCATGG, MfeI=CAATTG, BclI=TGATCA, NgoMIV=GCCGGC, BspEI=TCCGGA, BssHII=GCGCGC, BmtI=GCTAGC, NsiI=ATGCAT, BglII=AGATCT, AflII=CTTAAG, XbaI=TCTAGA, BamHI=GGATCC).

3.7. Methylation activity investigation

The methylation proficiency of pMJA01was investigated by *in vitro* and *in vivo* protocols. The *in vitro* method involves digestion of minipreps of pMJA01 with the cognate HhaI restriction enzyme. If the plasmid encodes a functional GCGC specific DNA MTase, then it will be refractory to cleavage by HhaI. This is observed in all cases in Figure 3.6. In fact pMJA01 is clearly methylated at all of its 20+ HhaI sites (see Figure 3.7). The *in vivo* method involves transformation of an McrBC^{+}*E. coli* strain by pMJA01: if the plasmid encodes an active MTase, no colonies should appear since, the McrBC restriction system cleaves only at sequences in which the middle cytosine of the GCGC sequence is methylated (Raleigh, 1992).

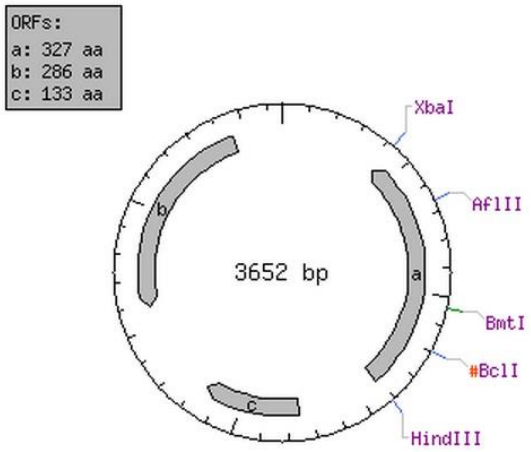


Figure 3.4: (a) Diagram of pMJA01 encoding the synthetic M.HhaI gene and its ORF (a), β-lactamase, the ampicillin resistance ORF (b) and (c) is the origin of plasmid replication.

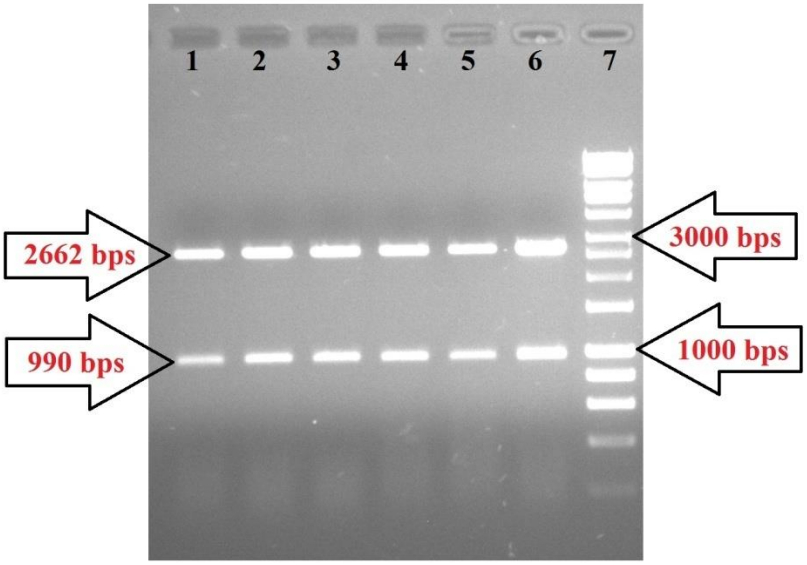


Figure 3.5: Preliminary verification of cloning the synthetic M.HhaI gene using agarose gel electrophoresis and restriction digestion. Lanes 1-6 are pMJA01 minipreps digested with HindIII and XbaI. Lane 7 is a DNA hyper ladder marker with two strategic fragment sizes shown, to illustrate consistency with predicted fragment sizes.

If the plasmid is methylation proficient, it will exhibit normal transformation efficiencies in McrBC^{-} strains, but will yield no significant numbers of transformants when a McrBC^{+} host is used. As can be seen in Figure 3.8, and this is an expected outcome.

The same *in vitro* and *in vivo* activity tests were subsequently applied to examine the methylation activity of individual mutant clones obtained from the random mutagenesis procedure. However, the *in vitro* HhaI plasmid restriction digestion was commonly the test of choice most of the time because it additionally provides an indication of very low (partial) levels of M.HhaI enzyme activity.

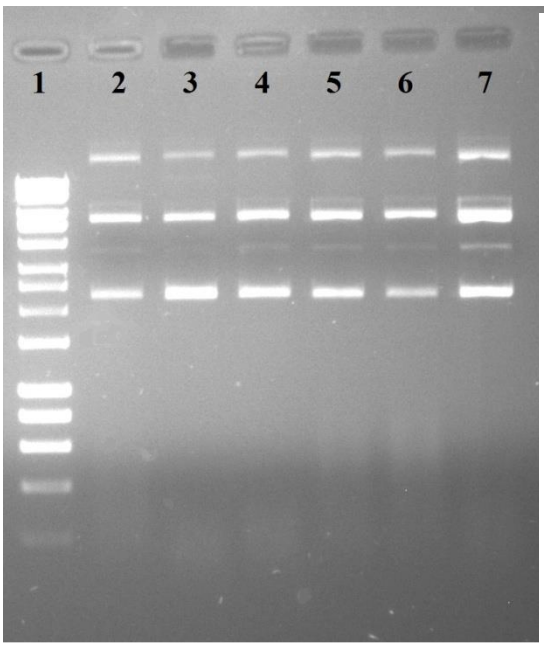


Figure 3.6: Methylation activity check of M.HhaI which is expressed from the synthetic M.HhaI gene carried on pMJA01 plasmid. Lane 2-7 show uncut pMJA01 when treated with HhaI restriction enzyme. This indicates fully active M.HhaI. Lane 1 is hyper ladder.

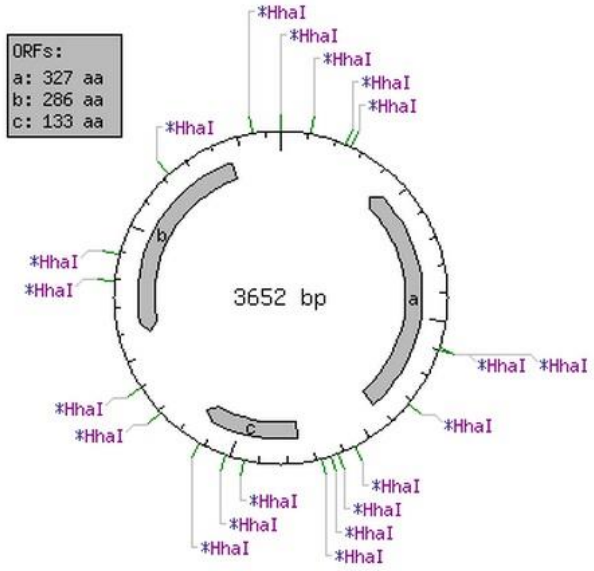


Figure 3.7: HhaI restriction sites within pMJA01.

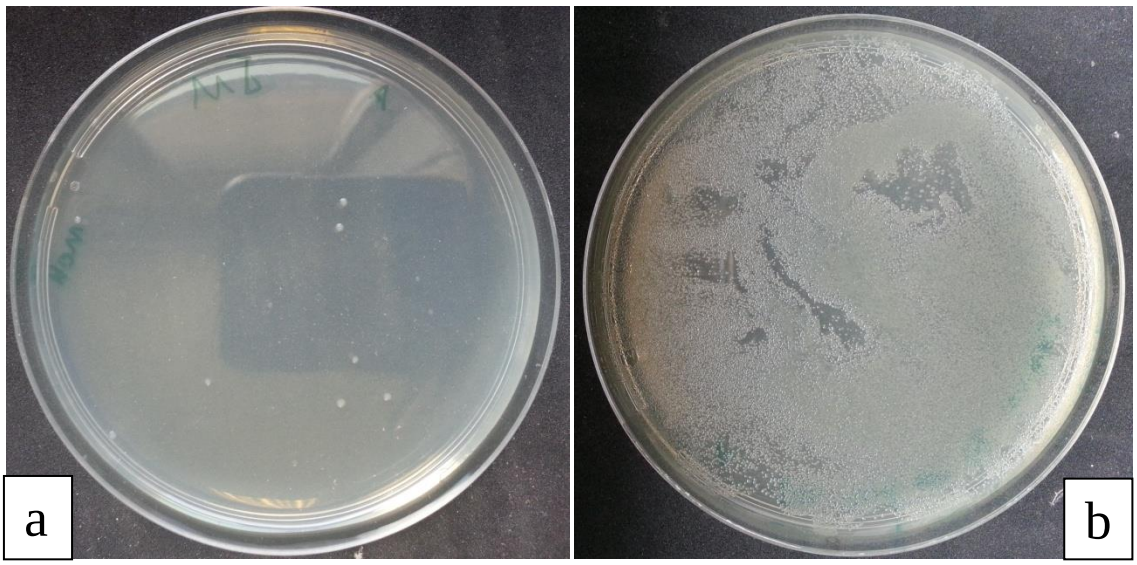


Figure 3.8: *In vivo* investigation of M.HhaI activity. (a) is McrBC^{+}*E. coli* transformed by pMJA01when few colonies appeared indicating active M.HhaI. (b) is McrBC^{-}*E. coli* transformed with pMJA01.

3.8. Error prone mutagenesis (EPM) of M.HhaI

Three sets of PCR primers were designed so that each set produces an amplified region of M.HhaI gene flanked by a pair of different restriction sites to facilitate subsequent recapitulation of the M.HhaI ORF. The general experimental strategy for all mutagenesis experiments is shown in Figure 3.9. The first phase of mutagenesis was targeted at three regions of M.HhaI orf: the first third of the orf (flanked by HindIII + BmtI), the second third (flanked by BmtI + AflII), and the last third part (flanked by AflII + XbaI) (see Figure 3.3 and 3.4).

After the EPM protocol, a library of mutant pMJA01 constructs was generated, in the form of transformed bacterial single colonies. These clones were analysed for activity as described above (see Figure 3.10). As expected, a mixture of active and inactive mutants were recovered. The sequencing data, again, as expected, showed many different mutational patterns from the many transformants as a result of EPM. (See Appendix1).

3.9. DNA sequencing

After scanning for methylation activity of 384 clones, plasmid minipreps were prepared from each colony and the plasmid DNA samples were sent for nucleotide sequencing. A small number of plasmids gave abnormal sequencing data. Some mutant clones contained truncating mutations within the open reading: these mutants were inactive. On the other hand, many mutant clones appeared with in frame TAG stop codon and methylation activity was often detected in many such mutants.

In order to rationalise the relationship between primary structure and function, mutant clones of missense nucleotide changes (leading to amino acid changes) and those of silent nucleotide changes (leading to no change in the protein primary structure), were classified according to the number of

nucleotide changes on one hand and the number of changes in the protein primary structure on the other hand (Table 3.2). Understanding the relationship between coding sequence and the protein primary structure, after random mutagenesis could demonstrate how a gene (in this case the M.HhaI gene) could tolerate mutations while retaining function at a biologically effective level.

Most of the mutant clones, their numbers are mentioned in Table 3.2, appeared with changes in DNA sequences that do not match the same number of changes in their protein primary structure reflecting the nature of the M.HhaI DNA gene sequence, which shows significant rate of mutations silencing especially when no clone among those of 8, 9 and 11 nucleotide changes occurs with the same number of amino acid changes, furthermore, high incidence of silencing appeared with the mutant clones of the three nucleotide changes where only 7 out of 34 (20.58%) of them showed the same number of amino acid changes (three sense mutations), while the rest appeared with only two, one

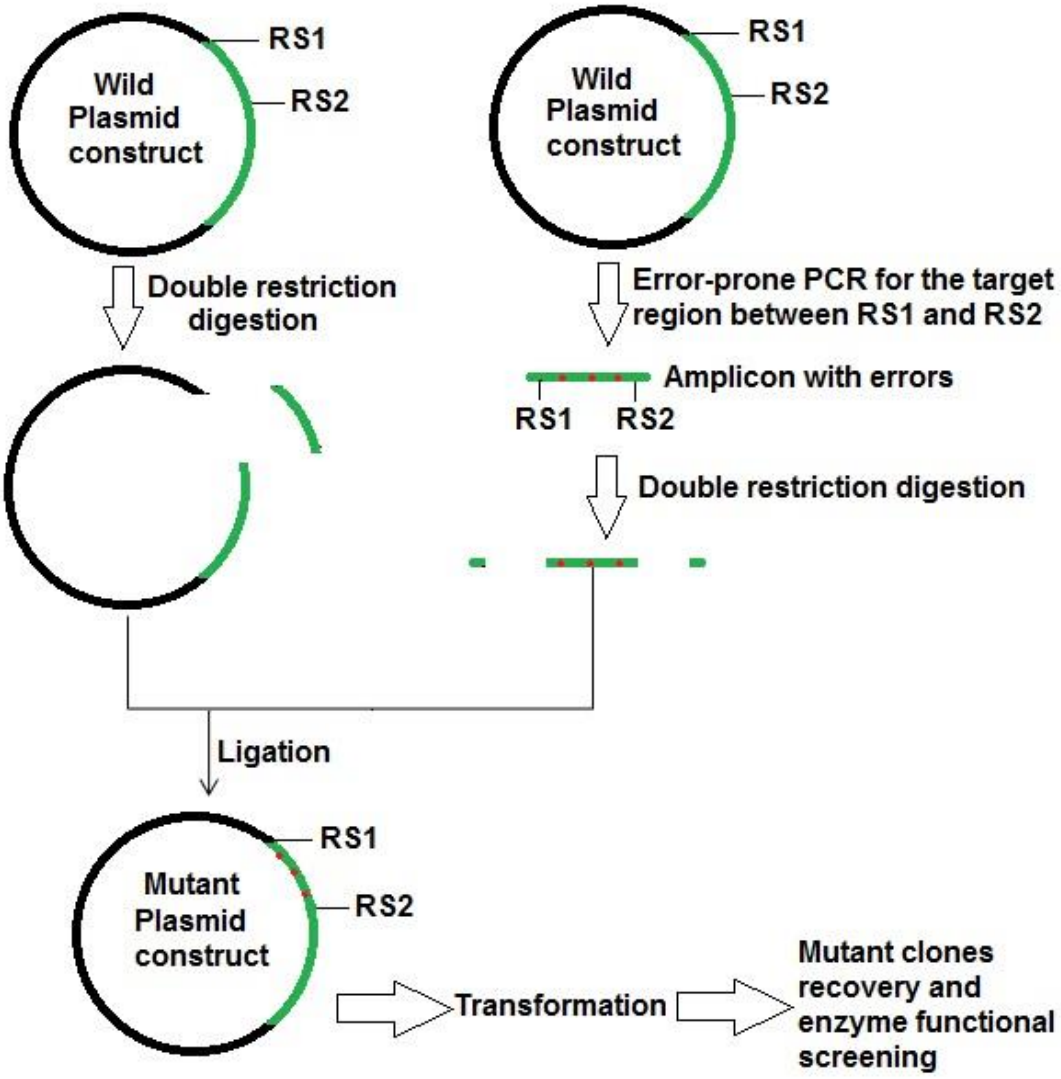


Figure 3.9: The general strategy of Error prone PCR random mutagenesis and cloning which is applied in this study on M.HhaI and the other two enzyme models (Chapter 4 and 5). The green part represents the gene ORF. The red dots represent the mutations (errors) that are generated by EPM. RS1 and RS2 refer to restriction site 1 and 2 respectively.

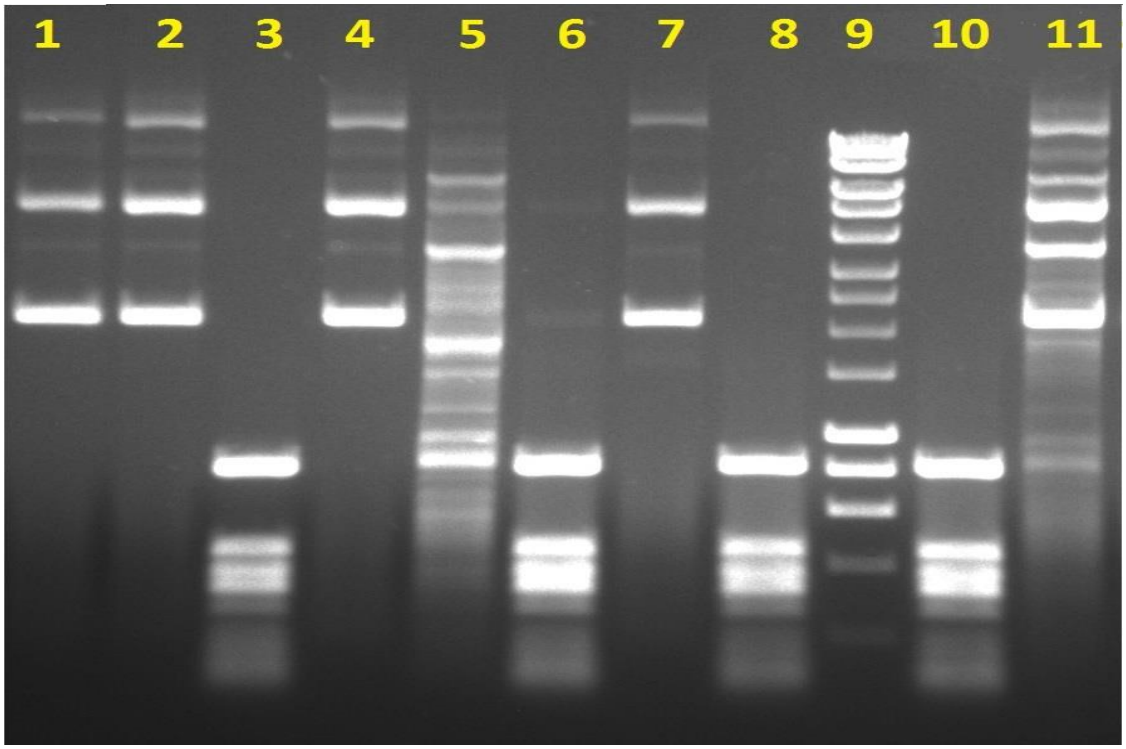


Figure 3.10: HhaI-treated pMJA01 plasmids from single bacterial colonies after mutagenesis. Lane 1, 2, 4 and 7 represent clones of active M.HhaI as plasmids resist digestion by HhaI, while lane 3, 6, 8 and 10 are clones of inactive M.HhaI which means that plasmids are not methylated and hence sensitive toward HhaI. Clones in lanes 5 and 11 exhibit different pattern of fragments arrangements which is caused by the incomplete activity of M.HhaI by which only some HhaI positions are methylated on those plasmids leaving some positions unmethylated and hence susceptible to HhaI action. Lane 9 is DNA marker (hyperladder I).

or even no amino acid change. All the remaining classes of clones that show different numbers of amino acid changes do not show the same pattern of DNA mutations but always showed a different number of nucleotide changes, so that the mutant clones of sense mutations, where the DNA changes match the protein primary structure changes tend to be a minority most of the time. However, the missense mutations of the same nucleotide changes within the clones of single and double DNA base pair changes are represented by 66.66% and 51.28%. Overall, the mutagenesis silencing incidence tends to be higher as more mutations occur in the DNA sequence and vice versa (Table 3.2).

Table 3.2: Mutagenesis silencing: correlation between the nucleotide changes and amino acid changes:

Amino acid change Per clone	Clones obtained	Nucleotide base change										
		1	2	3	4	5	6	7	8	9	10	11
0	19	14	4	1	0	0	0	0	0	0	0	0
1	54	28	15	10	1	0	0	0	0	0	0	0
2	45	-	20	16	8	0	1	0	0	0	0	0
3	29	-	-	7	12	7	2	0	0	1	0	0
4	16	-	-	-	9	5	1	0	1	0	0	0
5	9	-	-	-	-	5	2	1	1	0	0	0
6	5	-	-	-	-	-	2	1	2	0	0	0
7	1	-	-	-	-	-	-	1	0	0	0	0
8	3	-	-	-	-	-	-	-	0	1	0	2
Total	181	42	39	34	30	17	8	3	4	2	0	2

3.10. Clonal mutation frequency and enzyme tolerance

The method of mutagenesis employed here is assumed to be essentially random and the mutation frequency produced by the error prone DNA polymerase is on average 9.8 per kilo base (Alharbi, 2010). However, individual clones recovered contained mutations numbering from none at all to more than 7 (as expected). As pointed out earlier of course, many mutations had no effect on primary structure. The zero changes and silent mutation clones were not analysed any further, since this work was not aimed at evaluating the error prone polymerase itself.

In contrast, those clones carrying the “significant” mutations, that have been shown to carry one or more nucleotide changes and accompanying amino acid

changes, were classified according to the number of mutations (amino acid changes) in each clone (the frequency) and the impact on the enzymatic activity of M.HhaI (see Table 3.3). This serves to provide a simple means of evaluating how the enzyme might retain its normal activity when one or more of its amino acid are changed. That is, it provides a measure of "robustness" or "resilience" with respect to mutation.

The results showed that the mutational incidence per clone is between one to eight among the 162 significant mutant clones analysed. The most common clones (54) are those containing a single mutation and the vast majority of them (49) express a fully active M.HhaI while one and four clones exhibited partial activity and abolition of methylation respectively (see Figure 3.10 for activity phenotypes).

In surveying the forty five clones of the double mutation group, it is clear to see that the proportion of fully active M.HhaI clones is 73.33% which is lower than those found to be active from the single mutation group. As would be expected, the more mutations, the more the enzyme activity is diminished in the recovered library of mutants, regardless what the mutations are and where they are located.

The above increase in susceptibility to a loss of activity in relation to the frequency of mutations, is confirmed by the data obtained from the remaining groups of mutant clones harbouring triple, quadruple, quintuple, etc mutations. Overall, the enzyme seems to be rather resilient toward mutagenesis, as ninety seven (about 60%) out of the entire batch of the 162 clones appeared to retain activity. To map the mutations on the M.HhaI tertiary and secondary structure, see Figures 3.14 and 3.15 at the end of this chapter.

Table 3.3: Number of mutations per clone and methylation activity:

Mutation per clone	Clone frequency obtained	Methylation activity		
		Active	Inactive	Partially active
1	54	49	4	1
2	45	33	11	1
3	29	12	16	1
4	16	2	13	1
5	9	1	8	0
6	5	0	5	0
7	1	0	1	0
8	3	0	3	0
Total	162	97	61	4
%	100%	59.88%	37.65%	2.47%

3.11. Sites of mutation in relation to enzyme activity

The loss of catalytic activity of M.HhaI has been correlated with the mutations arising from amplification by EP-PCR in order to map the differential susceptibility of several specific regions of primary structure. The mutagenesis strategy above was designed to incorporate, in separate amplifications, all parts of the enzyme, by designing multiple sets of PCR primers pairs in order to obtain mutations in a controlled manner. In a landmark publication, Trautner's group (Trautner *et al.*, 1996) compared the phenotypes observed from a large scale mutagenesis study of multispecific DNA MTases, and in this way determined the relationship between catalytic residues and DNA specificity determining regions of this class of enzymes. The experiments described here bring a higher level of control to the random mutagenesis approach, enabling us to perturb regions that are expected to be sensitive to mutation: (the conserved motifs I-X, the TRD region as well as parts of the primary structure of M.HhaI whose function in folding, stability and catalysis remain unknown). Indeed in this experiment, whilst several expected mutations compromise activity a number of others also abolish activity in an unexpected manner. Whilst a

complete saturation mutagenesis of the gene encoding M.HhaI is possible, it was decided that the MHhaI experiments have demonstrated the proof of concept in a reasonably well understood system, it would be interesting to explore the methodology in respect of a wider set of genetic targets, where the results might extend our appreciation of the relationship between primary structure and function in a more sophisticated way.

We classified the 162 mutations obtained according to the domain in which they appeared and whether these mutations affect the enzyme function. Our data indicate that whilst mutations in conserved regions can lead to enzyme inactivation (as expected from structural studies), many other mutations cannot impact on activity. It is clear that our understanding of primary structure and functional relationships are limited in this class of enzyme. The data are presented in Table 3.4. See Figure 3.1 for more detailed annotations on the secondary structure.

As it constitutes the largest part of the primary structure of the enzyme, the majority (approximately two thirds) of the mutations were confined to the catalytic domain. Although it accommodates the vast majority of the conserved residues and responsible for the catalytic activity, most of the mutant clones (70 out of 106) were fully active.

Table 3.4: Position of mutation and methylation activity:

Position / Activity	C	TRD	H	TRD+H	TRD+C	H+C	C+TRD+H	Total
Active	70	8	8	0	9	2	0	97
Inactive	32	2	2	2	9	1	13	61
P. Active	4	0	0	0	0	0	0	4
Total	106	10	10	2	18	3	13	162
%	65.43%	6.17%	6.17%	1.23%	11.11%	1.86%	8.03%	100%

*C= The Catalytic region, TRD= The Target Recognition Domain, H= The Hinge region, P.Active= Partially Active.

Only ten clones were shown to have mutations within the TRD; eight of them encoded fully methylated plasmids, reflecting active M.HhaI. The same scenario was observed when the hinge region was targeted.

On the other hand the enzyme seems to be less tolerant when mutations are recovered within more than one region simultaneously. This is illustrated by the loss of activity in half of the clones (9 out of 18) that contain mutations in both the TRD and catalytic regions. Furthermore, all clones recovered in these experiments, harbouring mutations within the TRD, catalytic and hinge region lead to inactivation of M.HhaI.

From the above, we conclude that in addition to the clear functional role of those residues found in the conserved motifs and the hinge and TRD regions of M.HhaI, several unexpected residues impact upon activity so when they are combined with mutations from non-contiguous regions of primary structure, a more deleterious effect on activity is recorded. This analysis remains "a work in progress", since we are focusing on data collection in the first part of the project. However, there are some reasons why exploring the more general issues associated with the relationship between enzyme activity and primary structure and this should be extended to other enzyme encoding genes, and these are discussed in more detail in the following chapters.

3.12. Mutations in conserved regions of the M.HhaI primary structure

The analysis of data obtained from this mutagenesis study, demonstrates that mutations have been introduced into both conserved and non-conserved regions, and therefore it is very likely that enzyme activity of such mutants is on a spectrum from zero to 100%, with wild-type set at the 100% level, arbitrarily. It is of course possible to carry out a detailed kinetic comparison of all purified mutants, however this was not a primary objective of this work. The mutant sequences have been further considered as a functional BLAST comparison,

where natural variations amongst the C5-MTases can be analysed by superimposing the search data on this mutagenesis work. The results of this analysis can be seen in Table 3.5.

More than sixty bacterial C5- MTases were chosen (E.C 2.1.1.37) to build an alignment with M.HhaI to highlight the level of evolutionary conservation found in the publicly available genome data sets. This alignment was used to compare evolutionary variants with our mutagenesis data in order to further rationalise the sequence/ activity relationships that we have observed in M.HhaI. It should be pointed out that the biological activity of the majority of annotated C5- MTase genes has not been confirmed biochemically. The method of mutagenesis used here, technically extends the BLAST method and adds a powerful element of functional annotation. This alone represents a significant step forward and better than the use of bioinformatics alone to explore the relationships between coding sequences and their biological function(s).

The data shown in Table 3.5 reveal that 42 clones out of 43 of single change mutants in non-conserved residues were active. This suggests that the enzyme is relatively tolerant to point mutations in the non-conserved regions especially the single ones. Methylation activity is still apparent in the mutant clones in which one or more conserved residues are mutated. However, the most noticeable incidence of active M.HhaI could be seen in the group of single mutation clones; each mutant clone within this group contains a single amino acid change in a relatively conserved point and 7 clones out of the total 11 express full methylation activity.

As shown in Table 3.5, the three probable incidences where a conserved residue might or might not be targeted in mutant clones of double amino acid change have been obtained (0/2, 1/2, and 2/2) showing that M.HhaI is highly resistant even when two mutations arise but in the non-conserved points, the

high activity of the 0/2 double mutants is expressed by 25 out of 27 clones. The enzyme still exhibits activity even with the double mutations that include one or two conserved amino acids as about half of the total number of the 1/2 and 2/2 clones groups showed activity.

When three mutations arise, the enzyme activity is more reduced in comparison with the single and double mutations even if these three mutations happened in the non-conserved residues especially when 10 out of 18 of triple mutation samples express complete activity but when at least one mutation among those three happens to be a conserved residue, the activity will be highly reduced as shown above with only 2 active samples out of 11 in the group of 1/3.

Overall, M.HhaI tends to be less resilient as more mutations are introduced into its primary structure and this could be clearly noticed from the data presented in Table 3.5 especially when some relatively conserved residues are included amongst such multiple mutational changes.

3.13. The impact of mutations at the catalytic site of M.HhaI

The active site Cys residue (Cys81 in M.HhaI) within motif IV is the key to M.HhaI and all other C5-MTases activity (Wu and Santi, 1987). This residue is situated in the cleft near the AdoMet binding site so that its R group sulphur atom is about 10 angstrom away from the carbon atom of the transferable AdoMet methyl group. When the ternary complex of DNA, AdoMet and M.HhaI is formed, the distances between the sulphur atom and C6 of the cytosine on one side, and between cytosine ring and the AdoMet methyl group on the other, are too far apart for covalent bonding to occur. This may explain the need for the induced reorganisation of the “loop” which is formed by 20 amino acids residues located downstream of the P-C motif. This loop has been

found to be the most flexible part of the enzyme (Cheng *et al.*, 1993) so that it can promote a conformational rearrangement of the enzyme within the active site region, while DNA is bound. In this way, all members of the ternary structure are brought close enough to each other for a successful catalytic reaction to take place (Cheng *et al.*, 1993).

Table 3.5: Classification of mutations according to the incidence of changes into the conserved residues in each clone and the effect on the activity:

Mutation per clone	Single (1)		Double (2)			Triple (3)		Quadruple (4)					Quintuple (5)			
Frequency	55		45			31		17					11			
Conservation Change Incidence / Activity	0/1	1/1	0/2	1/2	2/2	0/3	1/3	0/4	1/4	2/4	3/4	4/4	0/5	1/5	2/5	3/5
Active	42	7	25	7	1	10	2	1	1	0	0	0	1	0	0	0
Inactive	1	3	2	6	3	7	9	3	8	0	1	1	2	2	3	1
Partially Active	0	1	0	0	1	1	0	1	0	0	0	0	0	0	0	0
Total	43	11	27	13	5	18	11	5	9	0	1	1	3	2	3	1

*Numerators in the ‘’conservation change incidence’’ row refers to the number of changes in the conserved or relatively conserved points, while denominators refer to the total number of amino acid changes. For example, 0/1 represents the group of mutant clones that contain single amino acid change but in a non-conserved point, 1/1represent the group of mutant clones that contain a single amino acid change in a conserved point.

*Conservation change incidences that are not mentioned=0.

Two conserved sets of amino acids (each set is formed by three residues) are located around the P-C active site: E119-N120-V121 and Q161-R163-R165. A number of interactions (hydrogen bonds and ionic interactions) have been found between these two sets and the conserved catalytic P-C residues and the commonly found T-L motif in the TRD (Cheng *et al.*, 1993).

Appendix 1 illustrates all mutations which are generated by error-prone PCR within M.HhaI gene along with functional data. Some of these mutations inactivated the M.HhaI, while the other did not (See Table 3.3). It is obvious from the positions of each residue that mutations cover both the enzyme target recognition domain and large domain. However, the effect of each mutation on the enzymatic activity is not easily rationalised from a knowledge of the tertiary structure and the chemistry of the enzymatic events.

For example, the mutation S85T appears to have an impact on the enzyme activity as residue S85 is a part of an interaction network which stabilizes the interaction between the backbone DNA sugar and the target recognition domain. These favourable Van der Waals interactions occur between the side chain hydroxyl group of residue S85, and the C4' and C5' of the deoxyribose ring, as well as between Cβ of the side chain and C5' of the sugar attached to the 3' guanine of the DNA recognition site. The importance of residue S85 is emphasised by its strong conservation in BLAST searches (Horton *et al.*, 2004), but it is not essential, or at least the BLAST data suggest that other mutations can compensate for loss of S85 (although it should be noted that many of the BLAST aligned sequences are only annotated as C5-MTases, and no biochemical validation exists).

It has been found by (Lau and Bruice, 1999) that V121 and the residues nearby are crucial for stabilizing the flipped out cytosine within the active site, therefore, mutation N123T obtained here could be a reason for enzyme inactivation (see the base flipping section below for more details). L100 together with F18 form the hydrophobic face of the AdoMet binding pocket, however, mutating L100 to S has no effect on methylation activity as mentioned by Sankpal and Rao (2002), this is in agreement with the L100M mutation obtained here after which methylation activity has been recorded in its relevant mutant clone, especially when L and M are both hydrophobic amino acids.

The flipped out cytosine needs to be retained transiently, but presumably in a suitable steric orientation at the enzyme catalytic site. In this respect, T250, which is not conserved among C5-MTases, has been demonstrated to be a key participant in the “anchoring” of the target cytosine as it appears from crystallographic studies to have rotated into the active site by following an interaction with a 5' phosphate (Vilkaitis *et al.*, 2000). Regarding the A253V mutation obtained here, even though A253 is not well conserved, it is close to T250 TRD residue, which may explain the observed abolition of activity in this mutant. However, It is notable from an analysis of an alignment of some bacterial C5-MTases that the position 253 in HhaI is largely occupied by either A or G (both are small amino acids), except for M.Eco72I which has V residue in this position. It seems that A and G are preferred, but that compensating mutations may overcome the obstruction caused by the relatively bulkier side chain of V.

The general kinetic steps of C5 DNA methylation (Figure 3.11) have largely been resolved by (Wu and Santi, 1987). The process, however, includes activation of the target cytosine at the C6 position by the nucleophilic sulfhydryl (SH) group of the C81 M.HhaI residue. As a result, this promotes the nucleophilicity of the C5 of cytosine sufficiently to facilitate its attack on the AdoMet methyl group. In this process, C81 becomes transiently bound to the C5 position of the target cytosine, therefore, the H5 proton at the C5 position has to leave. This deprotonation leads to dissociation of the enzyme and S-adenosyl-L-homocysteine (AdoHcy), and is thought to be mediated by R163 or/and a network of water molecules within the enzyme active site which facilitate the proton to transfer to the bulk solution and out of the active site.

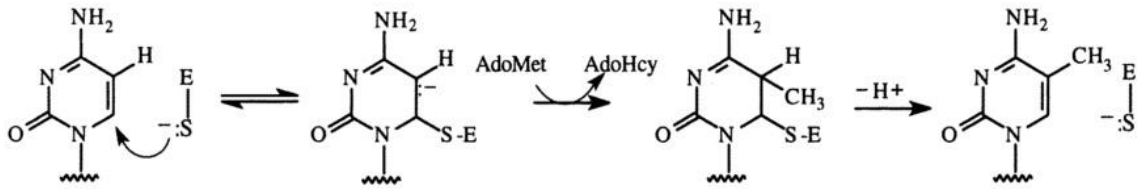


Figure 3.11: The scheme of M.HhaI-catalysed DNA methylation.

In a molecular dynamic simulation the study of M.HhaI active site (Lau and Bruice, 1999), the authors proposed that conserved residues N304 and Q82 serve as a structural channel for both protons and water molecules passage into and out of the active site. The conformational positioning of water molecules within the active site, which makes the water molecules close to the C5 of cytosine, appears to be important in making water bridges for the movement of both protons as a part of cytosine activation prior to methylation and the passage of the 5H proton from C5 of cytosine into the bulk solvent as a final step of the methylation. Hydrogen bonds are formed between water molecules inside the active site and side chains of N304 and Q82 so that water molecules are kept in their correct positions within the active site to facilitate the deprotonation. Interestingly, we have got an active mutant of Q82H, this might explain that changing Q82 with H, which is slightly larger and similar in containing NH in its R-group, does not affect the methylation.

The random mutagenesis experiments presented here have also targeted the region encoding amino acids Q301, N304,V310, Q312 and G319, all falling within the highly conserved motifX. This motif, together with motifs I to III, are primarily responsible for AdoMet binding (Cheng *et al.*, 1993; Klimasauskas *et al.*, 1994). Mutations occur at L75 and N123, which are both conserved and found within motifs IV and VI respectively. These motifs, along with VIII, V and VII are closely involved in the enzyme's catalytic mechanism, because they are responsible for active site formation (Schluckebier *et al.*, 1995).This,

however, may explain the inactivity of mutants having some of these amino acids changes (See Appendix 1).

3.14. Enzyme interaction with DNA: base flipping and the impact of random mutagenesis

M.HhaI interacts with DNA via the cleft which is between the small and large domain, it is 25-30 angstrom in width and 15-20 angstrom in depth so that it is deep enough to accommodate double stranded DNA. Its surface is positively charged which complements the negatively charged DNA. Q-R-R residues of motif VIII within the cleft are highly conserved among C5-Mtases and appear to be involved in the non-specific interactions with the DNA sugar-phosphate backbone. These interactions are assumed to be essential owing to the absolute conservation of this particular sequence motif in all C5-MTases. The specific interactions, which ensure a specific sequence of bases is recognised containing the target cytosine, are restricted to amino acid residues within the small domain. It is thought that the lower part of the cleft, which is about 40 angstrom in length (11bps accommodation capacity), is responsible for fixing the DNA helix in a correct orientation so that it makes stero-specific contact with the 15 angstrom-long upper part of the cleft (4bp accommodation capacity). The upper part of the cleft is thought to be related to specific cytosine C5 methylation (Cheng *et al.*, 1993).

The location of the M.HhaI TRD is at the side of the small domain cleft which is near T250 (Lauster *et al.*, 1989). The crystallographic study of (Cheng *et al.*, 1993) proposed that the TRD, comprising weakly conserved residues, is separated into two parts: the first one is derived from amino acids 231-240 and the second one involves amino acids 249-253.

The high resolution ternary structure of M.HhaI, DNA and S-adenosy-L-homocysteine (SAM) showed, for the first time, that the target cytosine is rotated ~180° in order to be flipped out of the DNA double helix, and situated in

the enzyme active site prior to nucleophlic attack by C81. It was revealed that such base rotation is accompanied by a movement of the catalytic loop (residues 80-89) to be closer to the flipped out cytosine (Klimasauskas *et al.*, 1994). High resolution, time-resolved NMR studies of binary complex (M.HhaI + DNA) formation showed that addition of the cofactor to the solution to form the ternary complex (M.HhaI + DNA+SAM) significantly enhances base flipping, indicating that target cytosine is not required to be bound in the enzyme active site for the flipping. This suggests that M.HhaI performs base flipping independently of the active site, and as another function beside its well-known catalysis of methylation which is of course absolutely dependent on the enzyme active site (Klimašauskas *et al.*, 1998). This was confirmed after showing that DNA substrate recognition by M.HhaI does not need a flipped out cytosine as demonstrated by a sequence mismatch study where the target cytosine was replaced by base analogues where no flipping is possible; and the enzyme still binds its DNA substrate easily and even stronger than if it contains a wild cytosine (O'Gara *et al.*, 1998; Serva *et al.*, 2004). This demonstrates that base flipping is an independent and unique function that M.HhaI is able to perform beside target recognition and methylation.

When the enzyme acts on its natural DNA substrate, it has been shown that the deoxyribose sugar and its surrounding phosphates are rotated in the same orientation of the target cytosine during the base flipping, therefore, it could be a fact that the rotation pressure is not directly on the target cytosine but rather on the sugar-phosphate backbone which simply carries the nitrogenous base with it at the same conformation when pushed and rotated along its own axis (Luo and Bruice, 2005).

It was thought that the cytosine is flipped out the DNA helix through the minor groove as the binary structure showed that the M.HhaI catalytic loop accesses its DNA substrate via the major groove so that the latter will be

blocked towards the cytosine flipping (Klimasauskas *et al.*, 1994). On the other hand, cytosine flipping from the major groove has been supported by the studies of molecular dynamics and free energy calculations upon base rotation (Huang *et al.*, 2003; Luo and Bruice, 2005). Furthermore, after the replacement of the target cytosine with a basic south-constrained sugar, crystal structures have revealed that this sugar is rotated ~90° about the flanking phosphates to be within the DNA helix major groove. This represents the mid-point of complete rotation (180°) and is in agreement with the major groove pathway (Horton *et al.*, 2004).

Flipping cytosine out of the DNA duplex requires ≥ 18kcal/mol free energy which is unfavourable for the Watson-Crick DNA structure, therefore, the role of M.HhaI in catalysis is to reduce this energy barrier by forming a catalytic complex with the DNA as substrate and AdoMet as cofactor. It has been suggested that the lowering of this free energy barrier cannot be provided by primary binding between DNA and M.HhaI to form an open binary complex, but requires the movement of the M.HhaI catalytic loop toward the target cytosine followed by the formation of a closed binary (or ternary) complex is actually required to overcome the unfavourable free energy of -5.1kcal/mol. The latter significant free energy reduction and the consequent thermodynamic stabilisation is consistent with the major groove pathway. Accordingly, the WC DNA conformation has been shown to be significantly disrupted as a result of hydrogen bonding and loss of stacking energies when a closed binary or ternary complexes are formed, especially when the hydrogen bonding distance between the target C N3 and the orphan G N1 (the guanine that complement the target cytosine) has been estimated to be significantly longer than the distance between the same N1-N3 atoms in the binary open complex. In other words, the DNA conformation around the target C has been shown to be significantly modified when the M.HhaI catalytic loop migrates to form the closed protein-

DNA complex. Such DNA conformational destabilisation has been shown to be mediated by many hydrogen bonds emanating from M.HhaI via key residues, with its DNA substrate. The number of hydrogen bonds in the closed ternary complex are greater than hydrogen bonds in the open binary complex. Residues 85-87 from the active site loop are involved in hydrogen bonding with DNA in addition to some other residue that are out of the catalytic loop such as Thr250, Tyr254 and Gly255. In the open binary complex such hydrogen bonding interactions are much less and are only recognized with Arg240 that form the most noticeable hydrogen bond when the catalytic loop is not yet moved to form the closed state complex. However, it has been suggested that such hydrogen bonding, especially in the ternary complex, is the reason behind the distortion of the DNA WC conformation and consequently facilitates base flipping. Residues Ser87 and Gln237 have been shown to play the most important role in facilitating base flipping by hydrogen bonding with target cytosine and its neighbours, however, Ser87 is involved in four hydrogen bonds; three with the orphan G and one with the target C, while G, which is next to target C from the 3` side, is involved in two hydrogen bonds with the Gln237 that is positioned directly over the target C to make stacking interactions. On binding of M.HhaI to DNA and prior to the base flipping, Ser87 has been suggested to assist base flipping by blocking the minor groove and pushing the cytosine from the minor groove side towards the major groove, from which, the cytosine is most likely flipped out, in addition, this residue has been postulated to push the guanine, which is the neighbour of the target C, to the 5` side providing more molecular space for C rotation (Luo and Bruice, 2005). It has been previously suggested that Gln237 is involved in steric clashes that cause C to be pushed out of the DNA helix from the minor groove (Cheng and Blumenthal, 1996) on the other hand, molecular dynamic simulations reported that both of the afore-mentioned stacking and hydrogen bonding, that is made by Gln237, as well as the competitive hydrogen bonding, that Ser87

makes with the orphan G, weakens the C-G wild hydrogen bond and distorts the overall local DNA conformation. All together, these events destabilise the target cytosine to be flipped out via the major groove while the original G-C base pairing is now replaced by hydrogen bonds that Gln237 makes with both the flipped cytosine and the orphan G as shown within the ternary structure, where the interactions between the protein and DNA reach an optimum which has been suggested to be related to the stabilisation of the flipped base until methylation occurs. All the aforementioned details of M.HhaI-DNA interactions and WC structure destabilisation, in term of cytosine flipping have been deduced from the molecular dynamic study of (Huang *et al.*, 2003) which are in accord with the crystallographic data for the ternary complex, regarding the final status of the completely flipped target cytosine.

The previous mutagenesis experiments showed that substitution of Gln237 by any of the other possible 19 amino acids residues significantly affects the ability of the enzyme to form a stable complex with the DNA while still the enzyme is able to recognize the DNA specific binding sequence (Mi *et al.*, 1995). The later mutagenesis study of (Serva *et al.*, 2004) found that replacement of Gln237 with short side chain residues causes a sever disruption on the base flipping so that Gln237Gly and Gln237Ala mutants efficiently bind DNA, but no base flipping is observed and, therefore, the overall rate of methylation was shown to be significantly reduced.

The present random mutagenesis work revealed three inactive M.HhaI clones of three, six and eight amino acid changes respectively; all of these include substitution of Gln237 among their mutations such as Gln237Lys and Gln237His (See Appendix 1). While Ser87 has been randomly mutated within three mutant clones; one of them is of three residue changes and two clones contain four residue changes carrying Ser87Tyr, Ser87Phe and Ser87Thr mutations respectively and showing no methylation activity. The multiple

sequence alignment (MSA) (See Appendix 4) shows that both Gln237 and Ser87 are not conserved but rather the latter residue is found as Ala in most of the C5-Mtases variants so that Ser is not only not a conserved residue, but not a common residue in that position (Ser87 in M.HhaI), while Ser85, which has been mutated to Ile within the inactive clone of double residue changes (Pro57Thr, Ser85Ile), is highly conserved and appears only to be replaced by Thr in only three variants of the aligned C5-Mtases. M.HhaI inactivation in this mutant clone could explain the importance of the S85 conservation as no Ile occurs as a replacement for the highly conserved Ser in that position according to the MSA (See Appendix 4). The MHhaI-DNA interaction study of (Luo and Bruice, 2005) has recommended the use of site directed mutagenesis to explore the impact of some residues such as S85 which is thought to interact with the DNA sugar-phosphate backbone and could be involved in rotating the nucleic acid backbone leading ultimately to the flipping of the target cytosine.

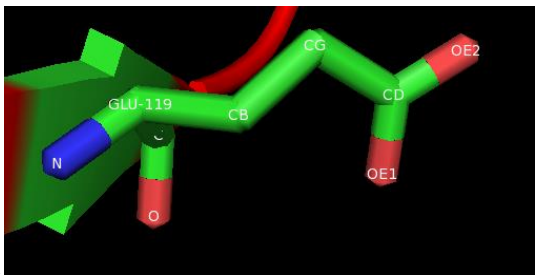
The TRD residues Arg240 and Gly255, which are not conserved but important in DNA binding as mentioned above, have been covered by the random mutagenesis within two inactive M.HhaI clones of mutational patterns Met168Ile, Arg240Leu, Ile247Thrand Gly255Asp.

NMR and protein-DNA binding electrophoresis studies (gel shift assays) both suggest the catalytic steps of methylation follow M.HhaI recognition and binding to its DNA to form the open binary complex. Remodelling of the catalytic loop towards the target cytosine facilitating its flipping and generating the closed binary complex, which is finally followed by the formation of the tertiary complex where the AdoMet cofactor is bound to form the final catalytic complex. The methyl group is then transferred to the flipped out cytosine and cofactor dissociation occurs recalling the closed binary complex again. Cofactor dissociation is thought to induce the catalytic loop to return to its original conformation in the open binary complex, where the flipped cytosine will be

less stable and, therefore, returns back to its WC status paired with the orphan G (Wu and Santi, 1987; Klimašauskas *et al.*, 1998; Lindstrom, 2000; Vilkaitis *et al.*, 2001). Such an equilibrium ordered bi-bi mechanism suggests that the enzyme would not dissociate from the DNA but rather remain bound to it and move along the DNA (after methylation and catalytic loop back movement and cofactor dissociation) to recognize and methylate another cytosine within its recognition GCGC sequence (Huang *et al.*, 2003). Although the bi-bi catalytic mechanism is highly supported as a successive step by step methylation catalysis, a random-stepped mechanism of catalysis has also been suggested (Vilkaitis *et al.*, 2001). However, both successive and random manner of M.HhaI movement in methylating its DNA substrate are proposed mechanisms and not fully proven.

The flipped-out cytosine within the ternary closed complex needs to be stabilised in the active site while the methyl group is transferred to it. This stabilisation has been suggested to be facilitated by the interactions M.HhaI makes with the flipped cytosine via residues Arg165 and Glu119. The random mutagenesis reported here has not produced mutations at Arg165, while Glu119 has been replaced by Asp within two mutant clones of Gln46His, Gln90Leu, ***Glu119Asp*** and Gly20Cys, Phe79Leu, Pro80Gln, Asp95Tyr, Ser96Ile, ***Glu119Asp***, Phe117Cys mutational patterns. The Glu119Asp mutation in the latter clone is accompanied by other mutations in sensitive parts of the enzyme that could reinforce the loss of activity in such a mutant clone, while the Glu119Asp mutation in the former mutant clone is accompanied by two additional mutations in parts that are not sensitive (Gln46 and Gln90), furthermore, these changes have been isolated within other mutant clones of fully active enzyme (See Appendix 1). Therefore, the inactivity of the mutant clone of Gln46His, Gln90Leu, ***Glu119Asp*** mutational pattern is definitely related to the Glu119Asp mutation proving the importance of the Glu119 as an

absolutely conserved residue constituting the first amino acid of the highly conserved ENV motif among the aligned C5-Mtases (See Appendix 4). Although the mutation-introduced Asp is similar to the wild Glu119 in its acidic properties, it seems that the length of the side chain, which is the only noticeable difference between the Glu and Asp, is important at this position since Asp is shorter in its R group when compared with Glu (see figure 3.12).

A 

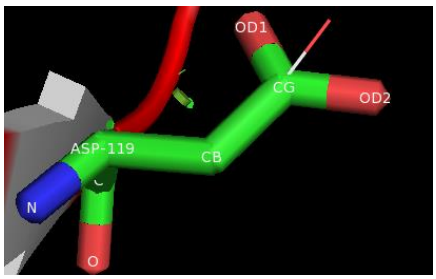
B 

Figure 3.12: 3D view showing the difference between side chains of the wild E119 and the mutant D119 which are represented in A and B respectively.

In the mutagenesis study by Estabrook *et al.* (2004), the importance of Val121 in stabilising the extrahelical cytosine has been demonstrated. The contribution of the nearby residues (His127 to Tyr132), which are close to Val121making a network of interactions with each other and with Val121, has also been investigated in the above study. All the aforementioned residues, including Val121, have been mutated into Ala in this work. Although the peptide represented by residues 127-132 including Val121 are suggested to have a role in "compression" of the active site by forming a network of interactions, as interpreted by the authors, further kinetic studies showed that all mutants, except Val121Ala, have a comparable activity to the wild M.HhaI (Estabrook *et al.*, 2004). However, the Val121Ala mutant showed a significantly reduction in the enzymatic reaction rate and DNA binding activity similar to the biochemical properties of the Cys81Gly mutant (Mi and Roberts, 1993).

Molecular dynamic simulations suggest that V121 is found in the enzyme active site in an orientation in which its side chain is close enough to the flipped cytosine and the 5` phosphate to form interactions with both of them. The positioning of V121 above the target cytosine was thought to be maintained by the hydrogen bonding between H127 and T132 but as mentioned above, there was no significant effect on the enzyme activity of both H127A and T132A mutants. Besides its direct role in interacting with and stabilisation of target cytosine, V121 may have other functions such as the positioning of some residues of catalytic importance like C81, E119, and R165 (Estabrook *et al.*, 2004). The present random mutagenesis covered V121 in four mutant clones of different mutational patterns (See Appendix 1). Interestingly, the single amino acid change V121M mutant clone appeared to show partial activity. Although V and M are different in size, they do share similar hydrophobic properties that could be related to incomplete inactivation unlike the V121A mutant, in which no detectable activity could be reported as mentioned previously. On the other hand, a mutant clone of simultaneous S87Y and V121M double amino acid change showed no M.HhaI activity indicating the definitive role of S87 in eliminating the remaining activity seen in the mentioned above mutant clone of single amino acid change (V121M).

3.15. The impact of random mutagenesis on the M.HhaI interaction with AdoMet

The region of M.HhaI which is involved in AdoMet binding contains highly conserved amino acids among C5-Mtases. Residues F-Naa-G-Naa-G which are within motif I, have been found in many AdoMet-dependent Mtases including N4-cytosine, N6-adenine, RNA and this conservation even extends across to protein Mtases (Lauster *et al.*, 1989; Pósfai *et al.*, 1989).

The close association of AdoMet with M.HhaI in the crystallographic study of (Cheng *et al.*, 1993) has given a clear view of interactions as follows: the β1-αA-β2 segment forms the binding site of AdoMet which occurs at the upper part of the cleft just beside the C-terminus of β1 strand within the large domain. The methionine part of AdoMet runs the cleft while the adenosyl part is packed inside a pocket within the cleft. When AdoMet is not available, the side chain of the conserved F18 or W41 will be inserted in the pocket. The binding of AdoMet at the right position is related to the tight loop (I-1A) formed by the glycine rich part between strand β1 and helix A so that the adenine ring of AdoMet will be as close as possible to the protein main chain. Therefore, introducing amino acids with large side chains within the glycine rich portion by mutagenesis inhibited enzymatic methylation (Wilke *et al.*, 1988) (see Figures 3.13 and 3.1).

Although motif I is the key part of AdoMet binding, there are other motifs involved in the interaction. The side chain of E40 (motif II) is contained in the hydrophobic pocket facing the N3 of the adenine ring of the bound AdoMet; this residue is hydrogen bonded with main chain nitrogen atoms of residues 20 and 42. D60 (motif III) interacts with the amino group of AdoMet methionine moiety, while Q82 is involved in the interaction with the carboxyl group of the same moiety.

Nitrogen of the side chain and oxygen of the main chain of N304 (motif X) are hydrogen bonded with N1 and N6 of the purine ring of AdoMet respectively. Hydrogen bonding occurs between amino groups nitrogens of residues 41 and 61, and O2 and O3 of AdoMet ribose ring.

Generally, AdoMet interaction-involved residues were found to be strongly conserved among C5-Mtases. F18, D60, P80, Q82, and N304, for example, are quite conserved. E40 and N304 are relatively conserved because they have only D, M, or V replacements, while others like W41 are less conserved as it could be replaced by M, I, L, Y, N, K, F or Q (Cheng *et al.*, 1993).

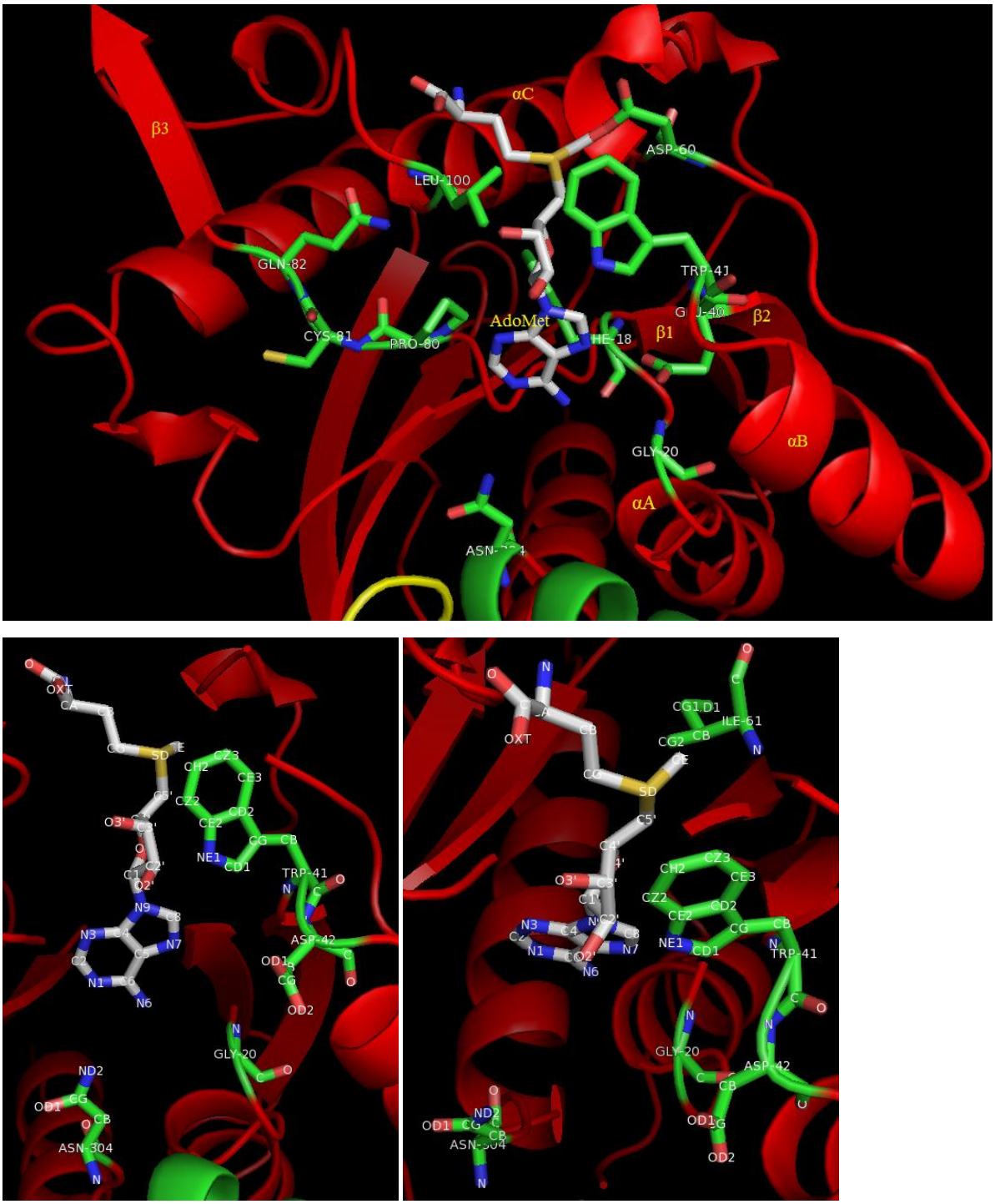


Figure 3.13: Three dimensional view showing the AdoMet in its binding pocket within M.HhaI. Some residues which are included in the interactions are labelled and represented as sticks with carbon atoms backbones in green, oxygen in red, nitrogen in blue, and sulphur in yellow colour. Secondary structure elements of α-helices and β-strands are numbered and labelled inside the figure as α and β respectively. AdoMet is represented in sticks structure with gray carbon backbone, red oxygen, blue nitrogen, and yellow sulphur atoms.

All the aforementioned residues that are involved in the AdoMet cofactor binding are conserved and lie within the large catalytic domain of M.HhaI. Most of the mutant clones which are isolated in this study appeared with several amino acid changes. It seems, therefore, irrelevant with such clones to elucidate whether the changes in some of the AdoMet-interacting residues are the reason behind the inactivation of the enzyme or the enzyme is simply inactivated by some other mutations that are not involved in the cofactor binding but they do accompany mutation(s) in residues that are involved in the AdoMet binding within the same clone. On the other hand Table 3.6 shows a couple of mutant clones that have a mutational pattern of double point mutations each clone having a mutation at residue positions known to form part of the an AdoMet-binding pocket: these mutants retain activity.

F18 and D60 are both highly conserved having few replacements as indicated from the multiple sequence alignment (MSA) (See Appendix 4). However, Y is seen as a replacement for F18 in M.HhaI as both of these residues are close to each other in structure and aromatic properties and this could explain the activity of the mutant clone containing an F18Y mutation. Although no N could be seen in MSA to replace the D at position 60, M.HhaI activity is maintained after the D60N mutation. This could be because the very close similarity in size and structure between D and N especially when the importance of the wild D60 has been shown by (Cheng *et al.*, 1993) to come from the R group oxygen atom that is maintained within the amide group of N side chain as well.

The glycine-rich motif of F-X-G-X-G which is a key motif in all C5 DNA MTases (residues 18-22 in M.HhaI) has been shown to be crucial in the maintaining the adenine moiety of the AdoMet in a suitable conformation to be close enough to the enzyme active site during the catalysis. The flexibility of this glycine-rich motif is required to form a tight loop which serves the function

of maintaining the appropriate conformational proximity of AdoMet as mentioned above (Cheng *et al.*, 1993).

Table 3.6: Mutations of AdoMet-binding residues and methylation activity:

Mutant clone ID	Mutational pattern	M.HhaI Activity
2nd54 1.10.14	***F18Y***, K112E	Active
2nd86 1.10.14	D6N, ***F18C***, Y49N, S87T	Inactive
2nd17 8.5.14	***F18S***, D42V, V116M, D103G, V121M	Inactive
2nd11 1.10.14	***E40K***	Inactive
2nd83 1.10.14	L26Q, ***E40D***, L75M, C76Y, V116L	Inactive
2nd53 8.5.14	***E40K***, ***D60N***, T62I, A77V, G92S, F101S	Inactive
2nd13 1.10.14	K5N, C35Y, ***D60V***, C76S	Inactive
2nd63 1.10.14	***D60N***, L100P	Active
2nd109 1.10.14	***G20S***, ***D60N***, D73E, D95E, R106L	Inactive
2nd6 8.5.14	***G22D***, E29D, G32C, D71V, D73V, F79L, ***P80Q***, Q90H	Inactive
2nd12 8.5.14	***G20C***, F79L, ***P80Q***, D95Y, S96I, E119D, F117C	Inactive
2nd55 8.5.14	I4F, G11I, N65H, ***P80T***, V121A	Inactive
2nd25 1.10.14	D73N, ***Q82H***, S87F, N120Y	Inactive
2nd34 1.10.14	**G20S**, A45T	Inactive
2nd98 1.10.14	***G20R***, Q63L	Inactive
2nd108 1.10.14	***G20C***, R25S, K122Q	Inactive
2nd80 1.10.14	***G20S***, D73V	Inactive
2nd12 8.5.14	***G20C***, F79L, P80Q, D95Y, S96I, E119D, F117C	Inactive
2nd42 8.5.14	***G20C***, C35F, G59D, I86F	Inactive
2nd22 8.5.14	***G22C***, F24C, R25S, T62S, V64D, F93C, E94D, F101L	Inactive

*Red-shaded numbers are of conserved residues. The AdoMet-interacting mutated residues are in **bold** *italic*.

Therefore, any mutational changes that introduce residues containing large side chains are likely to abolish the flexibility required for the AdoMet positioning and hence the methyl transfer may well be compromised. This was shown by the mutagenesis study of (Wilke *et al.*, 1988) where they described a catalytically inert M.SPR C5Mtase after the loss of the loop flexibility due to the mutations in its glycine residues. However, the current random mutagenesis has shown the same inactivation phenotype with M.HhaI when G20 has been changed into S within two double mutant clones. The companion mutations in both of these clones have been obtained separately within other clones of a recognized methylation activity. This confirms that the definitive reason behind the methylation inactivation in the aforementioned double mutants is the G20S mutation (See Appendix 1 for more information about the mutational patterns and activity data). Furthermore, G20 and G22 have been covered in the random mutagenesis in many other mutant clones, but abolition of methylation activity cannot be concluded clearly and evidently as the mutations of interest are accompanied by several other mutations that may contribute to enzyme inactivation (see the data in Table 3.6).

Finally, the importance of the acidic nature of the amino acid at position 40 is emphasised by the conservation of E60 which could only be replaced by D (the only biological alternative). This is further confirmed by the random replacement of the E60 with the oppositely charged K in a mutant clone of single amino acid change (E60K) where no methylation activity is observed.

Conclusions

The data presented in this chapter, are consistent with the general observations on the key residues that play a role in the structure and function of M.HhaI. In addition, by analysing a set of evolutionarily related C5 DNA MTases (shown in as multiple sequence alignments in Appendix 4), it is possible to place some of these residues into an evolutionary framework, assuming the major function of this class of enzymes flows from sequence specific DNA methylation. This is of course not surprising, however, such residues and regions are supported by regions of primary structure that vary considerably among this class of enzymes and our results shed light on residues that are not immediately apparent for structural studies. In the absence of a comprehensive appreciation of how a set of conserved residues can be supported by a wide range of different residues, we require strong functional correlations with mutations to complement a BLAST analysis. It is clear from these experiments that we are beginning to gain insight into less transparent patterns of amino acid distributions in proteins and the constraints imposed by primary structure on the "successful" evolution of proteins.

The enzyme continues to be more sensitive toward the mutations when the incidence of the amino acid changes is increased that might be explained that the range of the mutagenesis incidence could be expanded to include more than one functional part, for example, catalytic and target recognition domain simultaneously.

As the extent of the mutagenesis expanded, the probability of capturing highly conserved sequences increases and therefore the activity is more likely to be affected and could be abolished. This concept is clearly demonstrated by the data in Appendix 1 when active mutant clones are observed containing quadruple and quintuple mutations in the less conserved regions of primary

structure. Clones harbouring more than three mutations per clone are not necessarily in need for more than one mutation in the conserved residues to be inactive but they actually just require a “little push” to express complete inactivity by having only one mutation in a sensitive part among their mutant non-conserved residues. Clones of 6, 7 and 8 mutation, however, expressed no activity at all regardless the presence or absence of conserved residues mutations accompanying the rest of their mutations in each clone.

It is understandable to expect maintenance of an enzyme's activity when single or double mutations occur in the enzyme's non-sensitive regions. Equally, it is not surprising to abolish activity by introducing single mutations into highly conserved motifs. This is especially true when the mutations are accompanied with others which combine in a synergistic manner perhaps, to inhibit methylation activity. However, interesting inactive methylation profiles have been recorded when single amino acid change happened in a non-conserved point (See Appendix 1). Furthermore, the enzyme is shown to still express activity when multiple amino acid changes are introduced by random mutagenesis to include more than one point of conservation and in an extended manner to involve more than one domain of the primary structure (see Table 3.4, Table 3.5, and Appendix 1).

Finally, to evaluate the importance of residues to catalytic function, the mutations of single residue change with their relevant activity profiles were mapped directly into the M.HhaI three dimensional structure (Figure 3.14). Furthermore, the role of primary structure in the formation of secondary structure elements were illustrated in Figure 3.15. The latter figure could be used to map all the mutagenesis and functional screening data to evaluate the role of the mutagenesis-targeted residues in protein secondary structure.

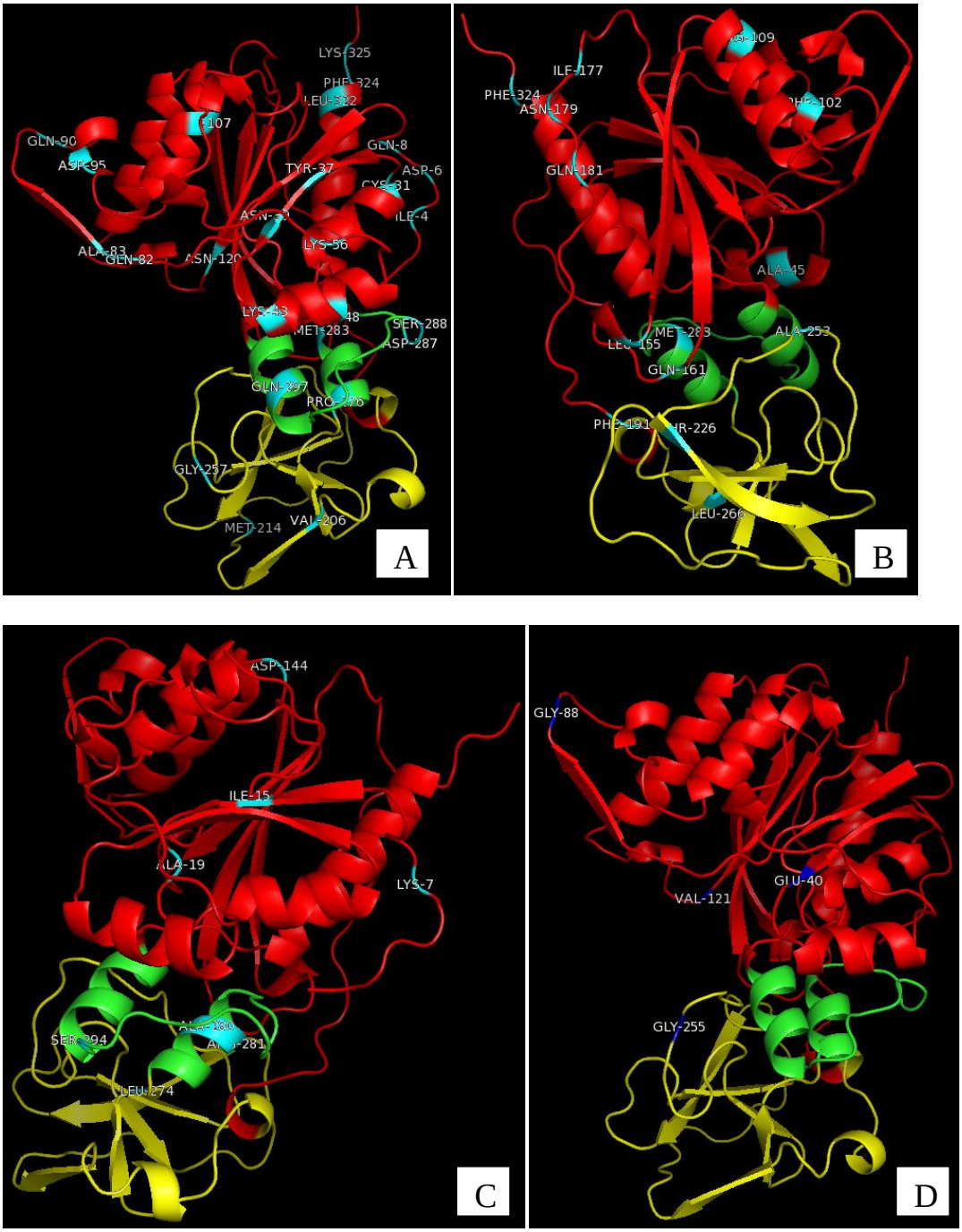


Figure 3.14: M.HhaI four different views of the three dimensional structure, highlighting mutations recovered from M.HhaI analysis: all residues highlighted represent single residue change mutants. Residues that affected the enzyme activity are labeled in blue (view D), while those which did not affect the activity are labeled in cyan (view A, B and C). Catalytic, TRD, and hinge regions are shown in red, yellow, and green respectively. (Details of the residue changes are all included in the appendices). Developed from the three dimensional structure which was resolved by Cheng *et al.* (1993), PDB code 1HMY.

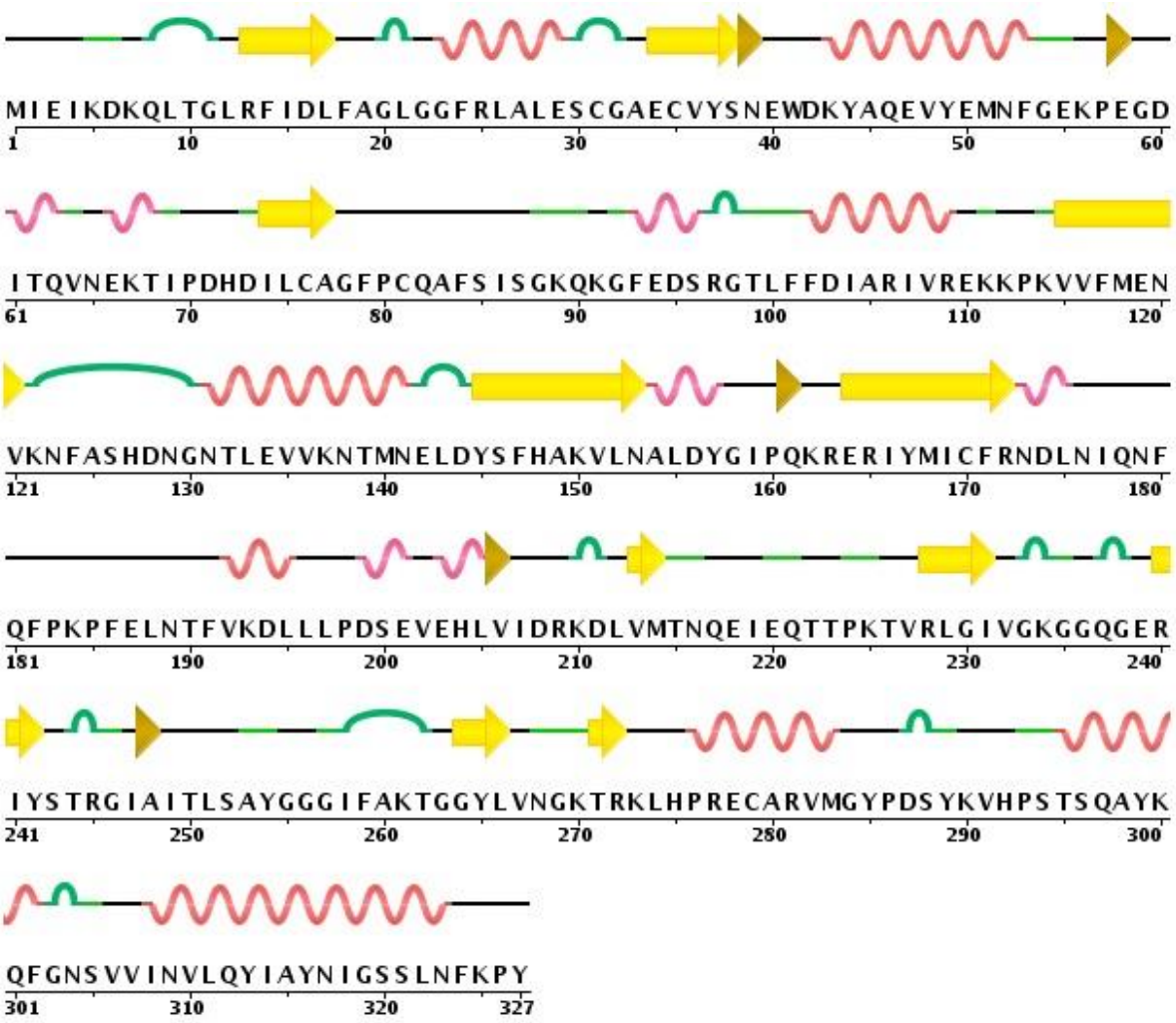


Figure 3.15: Combination of M.HhaI primary and secondary structures showing the secondary structure elements and their relevant amino acid residues. β-strands are shown in yellow arrows, α-helices are the red zigzags, loops are represented by green arcs, and the black lines are the regions of undefined secondary structure. PDB code 1HMY.

.

Chapter 4 Chloramphenicol acetyltransferase typeI (CATI): random mutagenesis and functional screening

Abstract

In this chapter, the gene encoding resistance to chloramphenicol has been subjected to random mutagenesis. Using a screening method in which agar plates are impregnated with different concentrations of the antibiotic chloramphenicol, it has been possible to investigate those amino acid residues that are essential for biological function of this enzyme. The results once again suggest that while the crystal structure allows interpretation of the role of several mutants, there are others whose role is unclear. This enzyme showed lower tolerance toward the primary structure changes when compared with M.HhaI in the previous chapter. Although mutations at absolutely conserved points inactivated the enzyme most of the time, exceptions of retained activity were noted. On the other hand, impaired activity profiles were seen when regions of no or limited primary structure conservation were targeted in mutagenesis. This is consistent with the findings of the previous chapter and emphasises the fact that it is not always possible to predict the importance of the primary structure by simply assessing natural conservation in silico. It is often the case that an activity test of the protein variants should be performed to confirm predictions.

4.1. Introduction

Bacterial antibiotic resistance has received considerable attention recently, owing to the emergence of pathogenic bacteria of clinical importance exhibiting resistance toward antibacterial agents especially when multidrug resistance appears (Wolstenholme and O'Connor, 1957). When looked at from a different view, the fact of bacterial antibiotic resistance seems to be a steadfast

in the both fields of biology and ecology. Being resistant to antibiotics is not a bacterial character derived by the pressure of natural selection or spontaneous mutations but also due to the gene transfer among bacteria. The latter mechanism has been shown not only to be mediated by plasmids and bacteriophage but also through mobile genetic elements such as transposons (Shapiro, 1983). Regarding genetically mediated antibiotic resistance, antibiotic-producing microorganisms are tolerant for the specific antibiotic that they produce as they are genetically adapted to survive in the presence of the their antimicrobial agents which are used to kill other competing microorganisms (Cundliffe, 1984).

Naturally occurring antibiotics are secondary metabolites of certain microorganism which have the ability to inhibit the growth of other microbes, therefore, the organisms, that possess the machinery of antibiotic production, process the appropriate mechanism to render them insensitive to these secondary metabolites (Bryan, 1989).

The inhibitory effect of antibiotics action is avoidable by resistant organisms by means of antibiotic exclusion from the cell interior or antibiotic target modification, but the most common resistance mechanism is the one through which the antibiotic is chemically modified by many different groups of enzymes such as β-lactamases and other enzymes like chloramphenicol acetyltransferase (CAT) which is dealt with in the current study (Shaw and Leslie, 1991).

Interestingly, the abrupt scourge of the wide spread of antibiotic resistance was first reported in the middle of the 20th century in Japan when resistance to chloramphenicol was diagnosed and detected to be caused by plasmids carrying the genetic elements of chloramphenicol resistance (Watanabe, 1963). The latter phenotype of resistance has been biochemically demonstrated to be mediated by

oxygen acetylation of the primary hydroxyl group on the C-3 position of the chloramphenicol molecule forming acetylated inactive antibiotic which is chemically modified and unable to bind to its ribosomal target in bacteria (Shaw, 1967). Despite the other mechanisms of chloramphenicol resistance, it has been shown that resistance through the CATs is the most common widespread cause of resistance among the bacterial genera (Shaw and Leslie, 1991)

4.2. Chloramphenicol

4.2.1. Chemistry

When first isolated from *Streptomyces venezuelae* in 1947, chloramphenicol was originally named as chloromycetin (Ehrlich *et al.*, 1947) but in modern nomenclature it is referred to a chloramphenicol which is named systematically as D-(-)-threo-l -p-nitrophenyl-2-dichloroacetamido 1,3-propanediol. It contains two chiral centres enabling the molecule to present in four possible forms of molecular asymmetrical diastereoisomers, owing to the position of the first and second carbon atom of the propanediol chain. The inhibitory effect of the antibiotic is only related to the D- threo isomer. Some advantages regarding chloramphenicol structure is the possibility of chemical modification such as the replacement of the C-3 hydroxyl group by fluorine to produce florfenicol and few other modifications, like the replacement of the nitro group (NO_2) with a sulfomethyl group (SO_2CH_3) in thiamphenicol, without any loss of the antimicrobial effect of the drug (Shaw, 1983). Such modifications might be advantageous in reducing or abolishing some toxic side effect of the original putative drug structure (Schwarz *et al.*, 2004). The simplicity of the molecule made it the first antibiotic to be chemically synthesised and extensively marketed since the1950s. Chloramphenicol is known for its long half-life at

room temperature and resistance to hydrolysis *in vivo* (Shaw, 1983). The amphipathic properties of chloramphenicol are explained by the presence of both hydrophilic (dihydroxypropane side chain) and hydrophobic (the p-nitrophenyl and the A'-dichloroacetyl substituent) moieties, in its molecular structure providing a reasonable ability to be soluble and therefore transferable in the body fluids (Shaw, 1984). *In vitro*, chloramphenicol molecules tend to form needle-like crystals which are completely soluble in ethanol but show a low degree of solubility in water (Maxwell and Nickel, 1954; Hahn *et al.*, 1956).

4.2.2. Spectrum

Having an inhibitory effect against both gram negative and gram positive bacteria as well as many other microorganisms such as Mycoplasma, Chlamydia and Rickettsia, chloramphenicol is classified as a broad spectrum antibiotic. Its bacteriostatic activity has been seen at a concentration of ≤ 10 µg/ml against gram positives such as Haemolytic type A and B Streptococci, *Streptococcus pneumoniae*, *Streptococcus viridians* and enterococci as well as many other pathogenic gram negatives like Neisseria, Hemophilus and Salmonella. Although chloramphenicol is considered as a bacteriostatic antibiotic, it shows a bactericidal effect against *Streptococcus pneumoniae*, *Hamophilus influenzae* and *Neisseria meningitides* (Wehrle *et al.*, 1967; Overturf *et al.*, 1975; Rahal and Simberkoff, 1979). In spite of its toxic side effects, which make it sometimes avoided (Lietman, 1979), chloramphenicol is the drug of choice to treat many infections caused by gram-negative bacteria such as typhoid fever (caused by *Salmonella typhi*) and meningitis (caused by *Neisseria meningitides*).

The poor water solubility of chloramphenicol has been overcome by some of the chloramphenicol esters, such as chloramphenicol succinate which is

significantly water soluble enabling the drug to be given parenterally. The ester groups are removed *in vivo* by esterases generating the active form of the antibiotic, since the latter is not clinically active when esterified (Simon, 2000). As a natural molecule with a good lipid solubility properties, chloramphenicol is efficiently diffusible into the central nervous system, cerebrospinal fluid and even the intracellular compartment at a significant inhibitory concentration levels (Dunkle, 1978; Friedman *et al.*, 1979; Shaw, 1984) to treat serious infections of bacterial meningitis, brain abscess and even the intracellular pathogenic bacteria (Kramer *et al.*, 1969; Black *et al.*, 1973; Shaw, 1983) especially when the causative agents, like *Neisseria meningitides* and *Hamophilus influenzae*, are β-lactam resistant or the treated patient is allergic to penicillin and other antibiotics (Turk, 1977; Feldman and Manning, 1983; Ristuccia, 1985; Mascaretti, 2003).

4.2.3. Chloramphenicol mechanism of action and interaction with bacterial ribosome

Chloramphenicol is one of the antibiotics that inhibits protein synthesis by interfering with the peptidyl transferase activity at the 50S subunit of the 70S ribosomes (Polacek and Mankin, 2005). Ribosomes are the key "machines" of protein synthesis in (prokaryotic) cells and catalyse nearly all of the steps of the RNA translation into protein (Sohmen *et al.*, 2009), while the eukaryotic 80S ribosomes are not targeted by chloramphenicol, there is a low level of chloramphenicol interference with mitochondrial ribosomal protein synthesis where it could cause malfunction to eukaryotic cells, especially bone marrow stem cells (Martelo *et al.*, 1969; Franklin and Snow, 2005).

Structural studies in which crystals of chloramphenicol-bound to the bacterial ribosome have been solved, have clarified that the drug inhibits protein biosynthesis by interfering with the position of the aminoacyl-tRNA at the

ribosomal A-site preventing polypeptide chain elongation (Celma *et al.*, 1971; Ulbrich *et al.*, 1978; Schlünzen *et al.*, 2001; Bulkley *et al.*, 2010)**.**

The interaction between chloramphenicol and the bacterial ribosome is mediated by the rRNA residues in the ribosomal peptidyl transferase centre (PTC). These rRNA residues are conserved among the chloramphenicol sensitive bacteria, so that certain mutations of these residues alter drug susceptibility and confer a chloramphenicol resistance phenotype. For example; one of the two chlorine atoms of the chloramphenicol molecule are in a position to make contact with the amine that is protruding from a specific purine ring of rRNA, A2062, a base which, when mutated, confers chloramphenicol resistance (Mankin and Garrett, 1991). The orientation of the antibiotic molecule at its binding site in the 50S subunit is maintained by the electron density associated with the coordination of rRNA residues that provide some extra stacking interactions, in particular between the nitrobenzene ring of the drug and the pyrimidine ring of C2452 facilitating the binding and stabilization of the antibiotic molecule into its site of action (Blaha *et al.*, 2008). The ribosomal RNA is arranged differently as secondary structures at the PTCs of eubacteria, archaea and eukaryotes showing different mode of sequence conservations among those three domains of organisms, but the conservation is significantly high within the same group of organisms explaining the difference in tolerance and the wide range of sensitivity towards the PTC- affecting drugs like chloramphenicol. It has been demonstrated that some residues of bacterial PTC rRNA are attributed to the selective inhibitory action of chloramphenicol in being only active against bacteria other than eukaryotes. rRNA nucleotide C2055 in bacteria (which is adenosine in eukaryotes and archaea) is of importance by rendering the secondary structure of rRNA at the peptidyl transferase centre stable in a certain conformation, such that the remaining key chloramphenicol-interacting nucleotides are kept functionally close to the

chloramphenicol molecule. Therefore, mutations of C2055 have led to displacement of four key nucleotides and subsequent chloramphenicol resistance or at least a significant change to the drug minimum inhibitory concentration (MIC) (Davidovich *et al.*, 2007; Gürel *et al.*, 2009). Such conformational arrangements are not observed in higher organisms PTC rRNA; this, however, could explain the poor binding ability of chloramphenicol to eukaryotic and archaeal ribosomes providing efficient selective therapeutic properties against pathogenic bacteria (Dunkle *et al.*, 2010).

The ribosomal chloramphenicol binding site has been studied extensively by many structural studies which proved the presence of ion(s) that play important role in fitting and binding of the drug into its target (Bayfield *et al.*, 2001; Klein *et al.*, 2004; Schwarz *et al.*, 2004; Blaha *et al.*, 2008). Potassium ion has been found to form a tetrahedral coordination with some adjacent rRNA bases (G2447, C2501, G2061) and one of the chloramphenicol's hydroxyl group (Vogel *et al.*, 1971; Bayfield *et al.*, 2001; Xaplanteri *et al.*, 2003; Klein *et al.*, 2004) so that mutagenesis experiments into the latter rRNA bases causes alteration in the conformation of the ribosomal interacting surface and therefore chloramphenicol resistance proving that these bases are in crucial contact with the chloramaphenicol molecule either directly or through the K^+ ion (Blanc *et al.*, 1981; Kearsey and Craig, 1981; Bulkley *et al.*, 2010) .

4.3. Bacterial resistance to chloramphenicol

4.3.1. Acetyltransferases (CATs)

Different mechanisms of chloramphenicol resistance have been acquired by bacteria over time. The most common mechanism is the enzymatic one in which the bacteria have developed a range genes encoding antibiotic modifying enzymes. including chloramphenicol acetyltransferases (CATs) which modify

the antibiotic by acetylation and abolish the drug inhibitory effect (Murray and Shaw, 1997). Many other Cm resistance mechanisms have been reported in bacteria such as mutational changes that lead to modifications of antibiotic target site or cell membrane, antibiotic efflux systems and resistance by phosphotransferases (Shaw, 1983; Murray and Shaw, 1997).

The acetylation is performed by CATs at the C-3 hydroxyl group of the chloramphenicol molecule and some of its derivatives like azidamfenicol and thiamphenicol but the replacement of C-3 OH group by fluorine in florfenicol renders these molecules resistant toward CAT action and, therefore, these drugs remain active against bacteria that express CATs (Cannon *et al.*, 1990).

Many CATs genes are continuously being identified in different species and genera of bacteria which makes it difficult in terms of uniformity of the enzymes nomenclature. This is reflected in the nomenclature of the identical CATs by different names, while enzymes of the virtual different identity have been designated identically. See the review by Schwarz *et al.* (2004) for more details.

Chloramphenicol acetyltransferases have been identified into two groups according to their structural and functional similarity: the classical CATs and the xenobiotic CATs; also referred to as the novel CATs (Murray and Shaw, 1997). Furthermore, the genome sequencing of some microbes such as *Rhodobacter capsulatus* (Vlček *et al.*, 1997), *Mesorhizobium loti* (Kaneko *et al.*, 2000), *Streptococcus agalactiae* strain 2306 (Tettelin *et al.*, 2002), *Bacillus cereus* (Ivanova *et al.*, 2003) and *Brucella melitensis* (DelVecchio *et al.*, 2002) have shown sequences assumed to be of acetyltransferases which are awaiting to be grouped after being fully studied catalytically as their structures reveal no relation to either the classic or the novel CATs, therefore, they might have been

sorted out temporarily as a sub-novel or xenobiotic CATs-like proteins (Schwarz *et al.*, 2004).

4.3.1.1. Classical type of chloramphenicol acetyltransferases

Classical CATs have been found in a wide range of bacteria conferring resistance toward chloramphenicol and some of its derivatives (Shaw, 1983; Murray and Shaw, 1997). In spite of the relative differences in primary structures, all CATs share the same quaternary structure, having three identical polypeptide chains of between 207 to 238 amino acid residues long assembled together to form homotrimers of between 24 to 26 kDa (Shaw, 1983; Murray and Shaw, 1997). The common properties that CATs share, render them to be structurally related so that fully functional heterotrimers could be formed when more than one gene of different, but related, CATs are expressed in the same cell (Murray and Shaw, 1997). The classical CATs have some similarities in their primary structures especially when these residues are involved in enzyme catalysis, substrate binding, cofactor binding, monomer folding and assembly of the trimer (Murray and Shaw, 1997). Some CATs have been shown to have additional properties such as providing resistance toward substrates other than chloramphenicol such as fusidic acid (Völker *et al.*, 1982; Shaw, 1983) and being inhibited by some thiol-reactive reagents (Murray *et al.*, 1990). There are at least 16 groups of the classic CATs. Multi-sequence alignments have shown an amino acid similarity of about 80% between the groups of the classic CATs (Schwarz *et al.*, 2004).

4.3.1.1.1. Chloramphenicol acetyltransferases gene distribution

The genes of CATs have been found widely prevalent in both Gram-positive and Gram-negative bacteria within plasmids, chromosomes and transposons of many species and genera. Chloramphenicol resistant wild isolates could have more than one type of CATs coded by more than one gene and the functional

trimers could be composed of three identical polypeptide chain (homotrimer) or different polypeptide chains from different genes are assembled to form the hybrid (heterotrimer) type of CATs which are as fully active and functional as the parental homotrimers (Day *et al.*, 1995).

The typical representative CAT which is related to the group 1, the first group of the 16 groups of the classical CATs (Schwarz *et al.*, 2004), is genetically assigned as CATI, its gene was originally found within Tn9 transposon of *E. coli* (Alton and Vapnek, 1979) then detected on the plasmids of many Gram-negative chloramphenicol resistant bacteria such as *Acinetobacter spp.*(Elisha and Steyn, 1991), *Photobacterium damselae* subsp. piscicida, previously named *P. piscicida* (Kim and Aoki, 1993), and *Pseudomonas putida* (Schwarz *et al.*, 2004). During the last decade, the genome sequence projects of many bacterial species recognized CATI genes in *Salmonella typhi, Serratia marcescens* and *Shigella flexneri* (Luck *et al.*, 2001; Parkhill *et al.*, 2001; Schwarz *et al.*, 2004).

The plasmids of *Haemophilus influenzae* have been found to carry more than one resistance gene but the chloramphenicol resistance gene among them is mainly CATII (Roberts *et al.*, 1982; Murray *et al.*, 1990). Genes that are similar and closely related to CATII have been found in many other bacterial species such *as E. coli, Agrobacterium tumefaciens, Photobacterium damselae* subsp. Piscicida, and *Bacteroides fragilis* (Shaw, 1983; Murray *et al.*, 1990; Morii *et al.*, 2003). Many species and genera of Enterobacteriaceae (Murray *et al.*, 1988) and Pasteurellaceae (Vassort-Bruneau *et al.*, 1996; Kehrenberg and Schwarz, 2001, 2002) have been reported to have CATIII genes in their plasmids which are usually of multidrug resistance. For more review about other CAT genes distribution see (Schwarz *et al.*, 2004).

4.3.1.1.2. Chloramphenicol acetyltransferases gene regulation

The expression of chloramphenicol acetyltransferases genes has been shown to be regulated in different ways. Gene expression of some CATs especially CAT86 as well as those which are carried on plasmids like pC221, pC223, pSCS7 and pC194 is regulated by the mechanism of attenuation by which the chloramphenicol itself plays the role of the gene inducer (Lovett, 1990). It has been shown that transcripts of the latter CAT genes have one pair of inverted repeats located upstream of the CAT open reading frame. The inverted repeats are in turn preceded by a short ORF (6-9 codon) called the stall region its sequence is highly conserved and complementary to the 16S rRNA sequence. In the absence of the chloramphenicol, the inverted repeats of the mRNA transcript are free to form a stable hairpin mRNA secondary structure, preventing the ribosome from accessing its binding site, which is located within the sequence of the second member of the inverted repeat and consequently no CAT gene expression occurs. On the other hand, when chloramphenicol is available, it binds to a ribosome causing a conformational changes to the latter so that the 16S rRNA sequence will be exposed and free to hydrogen bond with its complementary stall region, which is near the first member (5') of the inverted repeats pair. As a result of the close proximity of the stall region to the 5' part of the inverted repeats pair, ribosomal binding to the stall region hinders the formation of the hair-pin mRNA enabling a second ribosome to reach the ribosomal binding site within the 3' part of the open inverted repeats and, therefore, the translation of the CAT will happen only in the presence of the antibiotic providing high levels of chloramphenicol resistance of MICs$\geq$ 128 mg l^{-1}. Other CAT genes have been proved to be regulated by invertible promoters such as in *Pseudomonas mirabilis* strainPM13 (Charles *et al.*, 1985) or the expression could be constitutive providing different ranges of MIC according to

the promoter strength and the copy number of CAT gene in the bacterial cell (Schwarz *et al.*, 2004).

4.3.1.2. Xenobiotic chloramphenicol acetyltransferases

Some members of xenobiotic CATs have been functionally proved to carry out acetylation using chloramphenicol as a substrate and acetyl-CoA as a co-factor, but others, have been shown to transfer the acetyl group to different substrates and, therefore, referred to as XATs. The structural relationships between the members of this group itself or between the classical CATs and xenobiotic CATs/XATs may explain the difference in the catalytic properties among the xenobiotic enzymes themselves and the classical CATs.

It has been shown that these enzymes share some common structural properties with the traditional CATs especially in being homotrimers composed of three identical polypeptide chains of 209-212 amino acid residue (Murray and Shaw, 1997). On the other hand, the primary structure showed significant differences between the classic and xenobiotic CATs, providing evidence that XATs could be related to the acetylation of substrates other than chloramphenicol such as streptogramins. Genes for streptogramin A resistance have been found in staphylococci and enterococci such as VatD (previously known as SatA) (Rende-Fournier *et al.*, 1993), VatE (previously known as SatG (Werner and Witte, 1999), VatA (previously known as Vat) (Allignet and El Solh, 1995), or VatB (Allignet *et al.*, 1993). Furthermore, many other CAT-like genes have been found widely distributed throughout many of Gram-positive and Gram-negative bacterial species and genera within chromosomal DNA (such as CATB7 in the *Pseudomonas aeruginosa* strain PAO222 chromosome) (White *et al.*, 1999), plasmids (such as catB2 genes in plasmids isolated from *Salmonella enteritidis*) (Villa *et al.*, 2002), transposons (such as multi resistance transposon Tn2424 from *E. coli*) (Parent and Roy, 1992).

The expression of some xenobiotic CAT genes has been reported to be regulated by means of gene attenuation which is also a regulatory mechanism of the classical CATs genes. However, some xenobiotic CAT, such as CATB1 of *Agrobacterium tumefaciens*, it has been shown that the inverted repeats, which are key features of regulation, are four pairs instead of one regulating in more complicated mRNA secondary structures just upstream of the CAT open reading frame (Rogers *et al.*, 2002). The CATB1 gene, however, has been shown to give low resistance (<20 mgl^{-1}) to chloramphenicol (Murray and Shaw, 1997). *Agrobacterium tumefaciens* CATB1gene was first detected by (Tennigkeit and Matzura, 1991) in the chromosomal DNA, and its polypeptide chain has been shown to have significant differences when compared with the amino acids sequences of classical CATs, but the radiometric chloramphenicol acetylation assay proved it as a CAT; for all that, it has been shown to give low level of antibiotic resistance (when compared to the classical CATIII) even when its gene is cloned into multi copy in *E. coli*. This, however, could explain why CATB1 may have other physiological roles in *Agrobacterium tumefaciens* rather than chloramphenicol resistance; especially when the kinetic studies showed that the chloramphenicol Km constant for this XAT is 159 µM, while Km of the classical CATIII under the same conditions is 12 µM (Murray and Shaw, 1997).

The integron system is commonly popular in the xenobiotic CATs genes when such genes are parts of a gene cassettes regulated by a common promoter. The level of gene expression in such gene cassette integron system is dependent on the proximity of the gene's ORF to the promoter, so that the closer the gene to the promoter the stronger the gene expression would be observed. The correlation between expression levels and the gene proximity to the promoter has been demonstrated in *Vibrio cholerae.* When the CATB9 gene, which is part of a chromosomal multi-resistance integron, was placed in either of the four

proximal domains of the promoter for the septuple gene cassette, so a significant chloramphenicol resistance was observed (≥25 mgl^{-1}). However, when the CATB9 gene was placed at the distal position away from the gene cassette promoter, the bacteria are rendered nearly susceptible to the drug (<1mg l^{-1}) (Rowe-Magnus *et al.*, 2002). Multi-resistance integrons of plasmid origin have been reported to include xenobiotic CAT genes in many species of enterobacteriacea (Schwarz *et al.*, 2004) as well as *Pseudomonas aeruginosa* (Laraki *et al.*, 1999).

Multi-sequence alignments of XATs against the aforementioned XAT from *Agrobacterium tumefaciens* showed that the latter enzyme has a conserved His-78 residue, which, when mutated to alanine, leads to a loss of chloramphenicol acetylation activity. Therefore, it seems that XAT follows the same mechanism as classical CATs in performing the chloramphenicol acetylation, but the XAT efficiency in this case has been proven to be less than of classical CATs especially when the x-ray crystallographic reports revealed general differences in tertiary structures between the putative XAT and CATs (Murray and Shaw, 1997).

It has been proposed from the tertiary structures of several classical CATs such as CATIII, and the E2 component of the pyruvate dehydrogenase complex PDH, that there is an evolutionary relationship between the two enzymes suggesting the idea that CAT evolved from an E2 ancestor. The tertiary structures of some well-studied XATs, however, could not find an evolutionary link with the PDH E2 as mentioned above. The phylogenetic difference between CATs and XATs could explain the difference in their catalytic properties in spite of the similarity of their quaternary structures.(Leslie *et al.*, 1988; Leslie, 1990; Mattevi *et al.*, 1992; Mattevi *et al.*, 1993).

4.3.2. Non-enzymatic chloramphenicol resistance

It is not necessary for bacteria to rely only on enzymatic modification (especially acetylation) of chloramphenicol to escape from the antimicrobial effect of the drug: other mechanisms that are not mediated by enzymes exist. Drug efflux is one of the bacterial weapons to acquire resistance: here drugs are excreted from the cell to avoid their lethal effect. A wide range of bacteria of both clinical and environmental importance have been reported to follow chloramphenicol efflux by encoding genes that express trans-membrane proteins that control drug expulsion. Such proteins are referred to as chloramphenicol exporters which could be specific exporters that pump only chloramphenicol and its derivatives such as florfenicol or general exporters that perform many drugs and toxic compounds of unrelated chemical identity excretion out of the cell. The specificity of exporters seems effective in term of the level of the antibiotic resistance so that the chloramphenicol -specific exporters are more efficient than their general counterparts (Schwarz *et al.*, 2004).

4.4. Chloramphenicol acetyltransferase type I (CATI) random mutagenesis

There are three major classes of classical CATs; CATI, CATII and CATIII that are focused on by the experimental research especially CATIII and CATI which their three dimensional structures have been resolved in both apo and substrates bound forms. CATI is, however, less studied than CATIII, therefore, it has been considered in the current work as one of the model enzymes to be under the test of the random mutagenesis and functional analysis.

4.4.1. Gene synthesis and preliminary standardization experiments

Unlike CATIII, the CATI gene is commonly used as a chloramphenicol resistance selection marker in many cloning and expression plasmid vectors. Therefore, the wild sequence of CATI has been taken from pBR325 (the sequence is available at http://www.addgene.org/browse/sequence_vdb/1987/) a plasmid vector that provides chloramphenicol resistance due to the presence of the CATI gene. The CATI ORF is 660 bp (including the stop codon), which is translated into a 219 amino acid polypeptide. The nucleotide sequence has been modified so that many restriction sites have been introduced silently without affecting ORF integrity or the wild amino acid content (Figure 4.1 and 4.2). Restriction sites for HindIII from N-terminal side and, XbaI and BamHI from 3'(C-terminal) side have been introduced to facilitate CATI gene cloning into a wide range of plasmid vectors that could be required for many purposes such as gene expression. Both flanking and intragenic restriction sites were introduced synthetically to simplify downstream analysis. In particular selection of regions for mutagenesis is potentially made easier in this way.

Although it has been shown to be naturally present in the Tn9 *E. coli* transposon (Alton and Vapnek, 1979), when synthesized, the modified ORF nucleotide sequence has been optimized for the *E. coli* K12 (the cloning strain used in the current work) expression to facilitate mutagenesis and functional screening. Unlike the Amp^R gene which has its upstream regulatory promoter in pBR325, CATI has no promoter annotation on the same pBR325 plasmid, but the 50 bases in the pBR325, which are just upstream of the CATI ATG start codon, have been included in the synthetic gene to get the same chloramphenicol resistance phenotype as in pBR325 and to avoid any expression regulatory problem that may arise when only the CATI ORF is included. The synthetic CATI gene was provided by the gene synthesis company cloned in a cloning plasmid vector called pMA-T which has Amp^R

marker as well. The final construct of pMA-T and CATI has been named here as pAtase2 (Figure 4.3). When the gene was designed and sub-cloned, it has been taken in consideration that each introduced restriction site should not be repeated in the final construct so each new restriction enzyme site is unique.

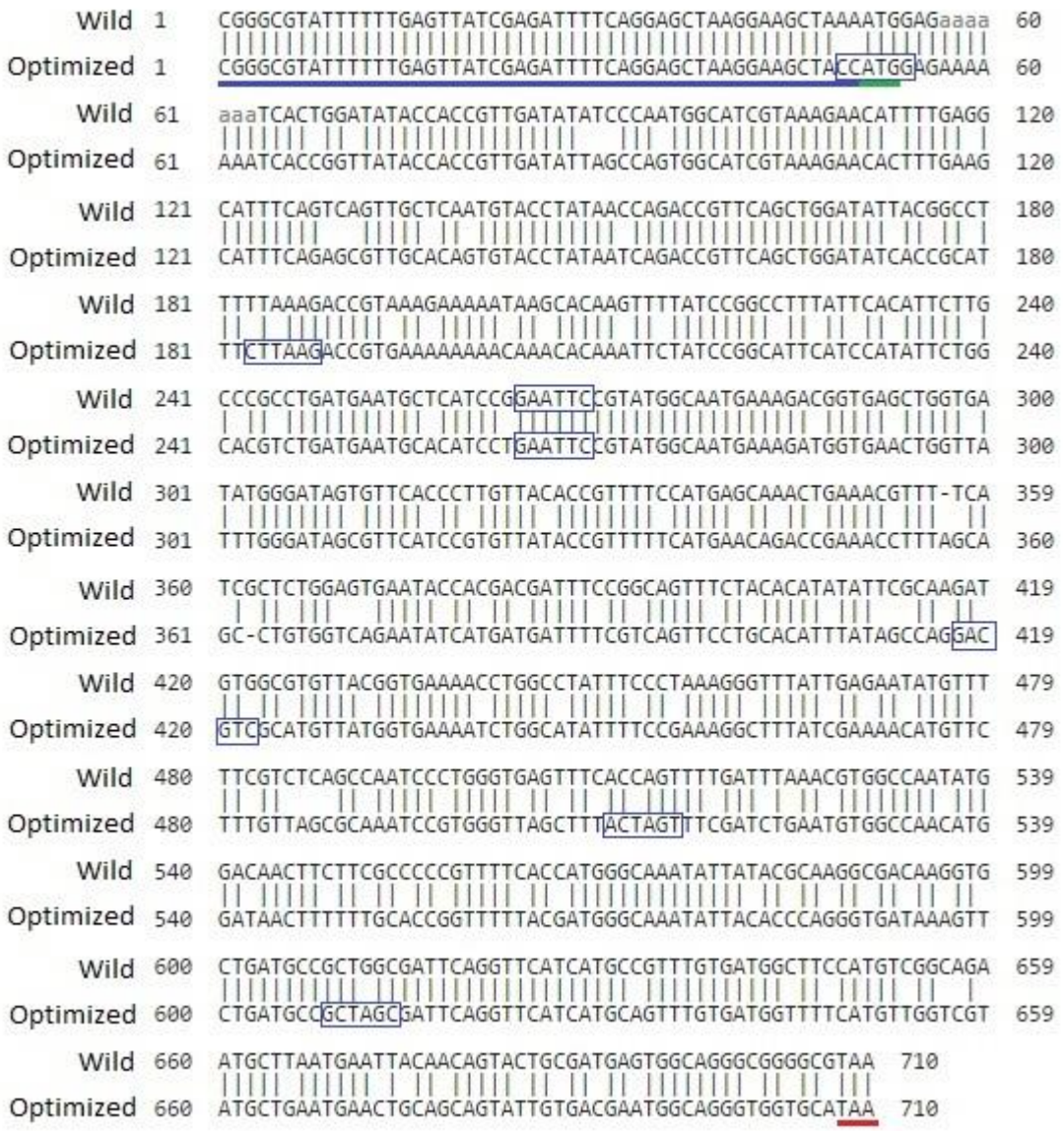


Figure 4.1: Nucleotide alignment between the wild and the optimized (synthetic) sequence of CATI gene. The upstream 50 bp are underlined by a blue line. The 49th and 50th nucleotide of the upstream sequence have been modified from AA into CC to get the NcoI restriction site at the beginning of

the ORF. The start codon ATG is underlined by a green line while the TAA stop codon is underlined by a red line. The in-frame restriction sites are boxed.

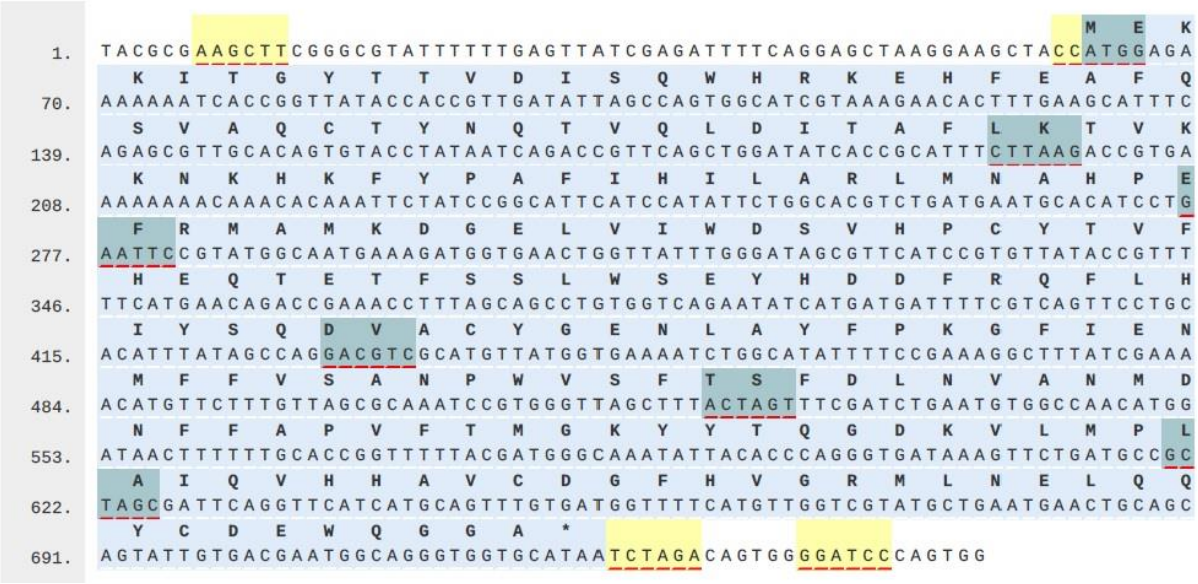


Figure 4.2: Interleaved view of the whole synthetic sequence showing the primary structure of the CATI (219 amino acid). The ORF is cyan-shaded. The in-frame and out-frame restriction sites are underlined by red dashed lines. (AAGCTT=HindIII,CCATGG=NcoI, CTTAAG=AfIII, GAATTC=EcoRI, GACGTC=AatII, ACTAGT=SpeI, GCTAGC=BmtI, TCTAGA=XbaI, GGATCC=BamHI). Each restriction enzyme has only one cutting site in the final pAtase2 construct.

4.4.2. Acetylation activity test

The synthetic gene was received lyophilized as 5µg pAtase2 construct which was rehydrated using 50 µl ddH_2O to get a DNA concentration of 100ng ml^{-1}. Once rehydrated, 50ng of the pAtase2 construct, which contains the CATI

synthetic gene, was used to transform 100µl of competent *E. coli* K12 1821using the chemical transformation method according to (Hanahan, 1983) protocol. The transformed *E. coli* then plated out on an ampicillin (100µg ml^{-1}) LB plate and incubated overnight at 37 °C.

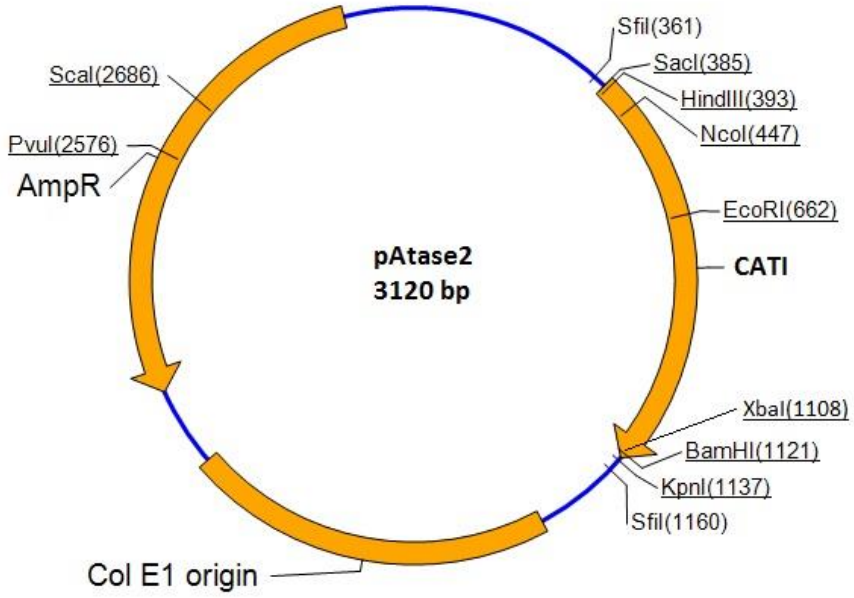


Figure 4.3: pAtase map as provided from the gene synthesis company showing ORFs of CATI and ampicillin resistance gene (AmpR). The origin of replication in this plasmid is Col E1. Many restriction sites are mentioned with their base pair cutting positions between the brackets.

Four CFUs were picked from the thousands of transformants and were sub-cultured on LB agar plates containing both ampicillin (100µgml^{-1}) and chloramphenicol (50µg ml^{-1}) to confirm that the CATI performs its role in acetylating chloramphenicol and conferring antibiotic resistance (Figure 4.4). The results showed that CATI synthetic gene is adequately expressed and fully active in *E. coli* K12 1821and, therefore, turning the latter sensitive strain into one that is chloramphenicol resistant (Figure 4.4). Higher concentrations of chloramphenicol have been used to test the minimum inhibitory concentration

MIC but the results showed that the growth of the pAtase2-transformed *E. coli* was strong even with the chloramphenicol concentration of up to 300µg ml^{-1}. The high level of resistance reached after transformation could be related to the high copy number of the pAtase2.

Plasmid minipreps were prepared from the clones that showed Cm resistance, the DNA was quantified and sent for sequencing to check the sequence integrity after transformation as recommended by the cloning standards of the gene synthesis company. The sequencing data showed 100% identity with the original sequence provided by the gene synthesis company.

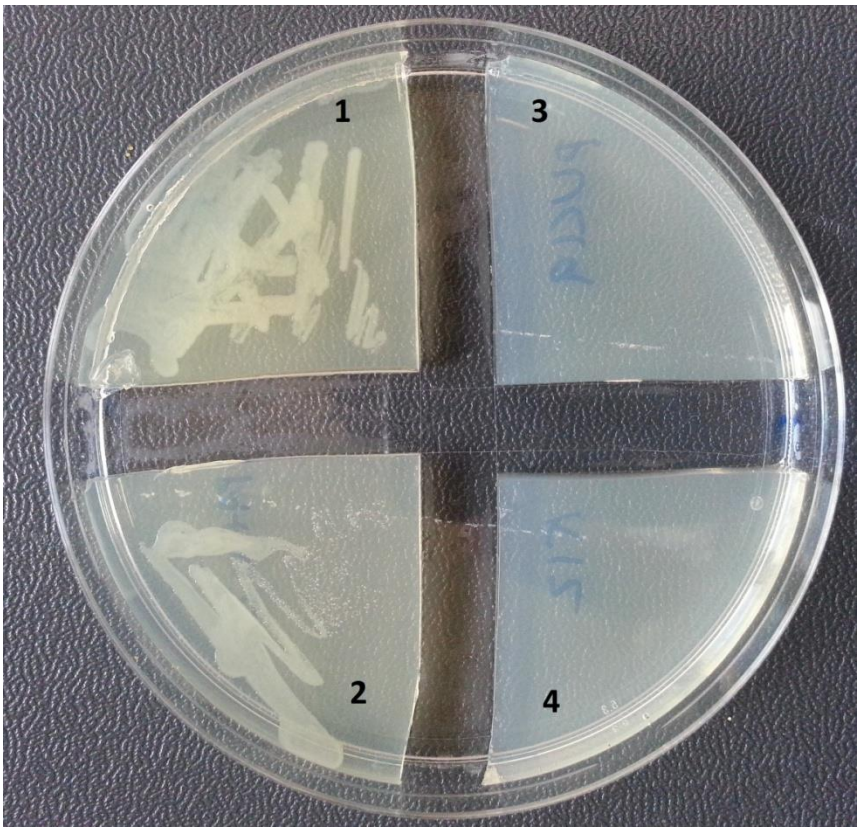


Figure 4.4: Chloramphenicol and ampicillin containing LB plate for CATI activity test. (1) pUC19-transformed *E. coli* BL21 DE3 (Rosetta) which is originally chloramphenicol resistant while its AmpR is from pUC19 plasmid. (2) *E. coli* K12 1821 transformed with pAtase2 which has both chloramphenicol

and ampicillin resistance. (3) *E. coli* K12 1821 transformed with pUC19 only therefore no growth noticed because of the presence of chloramphenicol in the media. (4) Untransformed *E. coli* K12 1821showing no growth.

4.4.3. Error-prone PCR

The mutagenesis strategy involves introducing mutations randomly into a segment of the CATI gene and returning this region as a restriction fragment library, back to the pAtase2 construct. The PCR amplification was performed using pAtase2 as a DNA template and a pair of primers (forward and reverse) that were designed so that the resulting PCR product contains EcoRI and XbaI sites; each of which has a single cutting site in pAtase2. The program and the protocol of the PCR amplification is described in chapter 2. The PCR product (651 bp) was purified according to the Qiagen® protocol and quantified prior to epPCR. DNA The purified product was double digested with EcoRI and XbaI: 1µg of pAtase was digested side by side with the PCR product using the same double digestion above. The digestion products were run on a 1% ethidium bromide-containing agarose gel for 1 hour. The bands appeared at their expected sizes (Figure 4.5).

The fragment of 447bp (lane 7 Figure 4.5), which results from the double digestion of the PCR product, and the fragment of 2673bp (lane 5 Figure 4.5), which results from the double digestion of pAtase2, have been recovered from the gel using Qiagen® gel kit. The recovered fragments were quantified and the ligation reaction was set up at 1:3 vector to insert concentration ratio. The vector is represented by the 2673bp fragment, while the insert is represented by the 447bp band. The ligation was performed at 16°C overnight using T4 DNA ligase from NEB. The ligation mixture was used to transform competent *E. coli* K12 1821 that has been prepared with a transformation efficiency of $6*10^7$ (see

chapter 2 for more protocol details). The transformed *E. coli* was then plated out on LB plate containing ampicillin only. The chloramphenicol was not added to the latter plate in order to recover all clones of both active and inactive CATI. The plate was incubated overnight at 37°C and transformation was successful (see Figure 4.6).

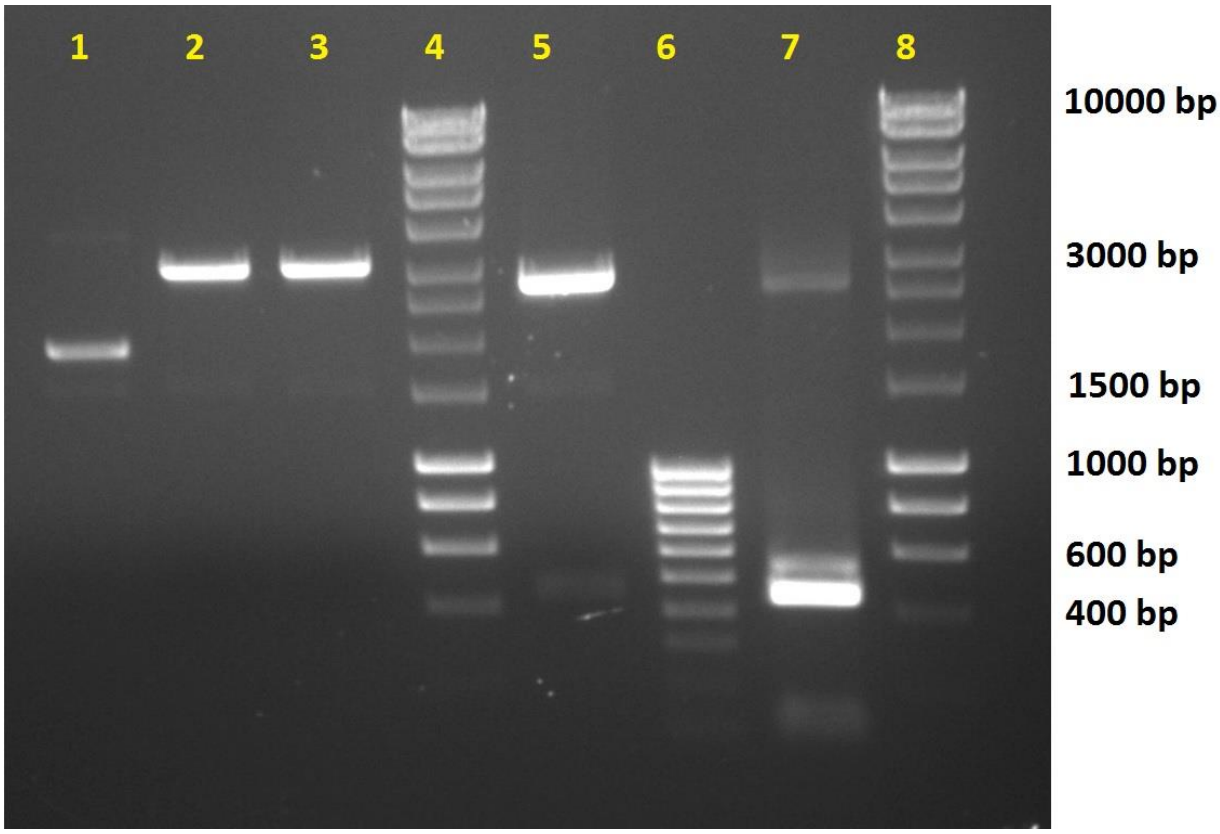


Figure 4.5: Agarose gel electrophoresis. (1) Uncut pAtase2. (2) pAtase2 digested by XbaI. (3) pAtase2 digested by EcoRI. (4) Hyperladder I. (5) pAtase2 digested by both XbaI and EcoRI. (6) Ladder IV. (7) PCR product digested by both XbaI and EcoRI. (8) Hyperladder I.

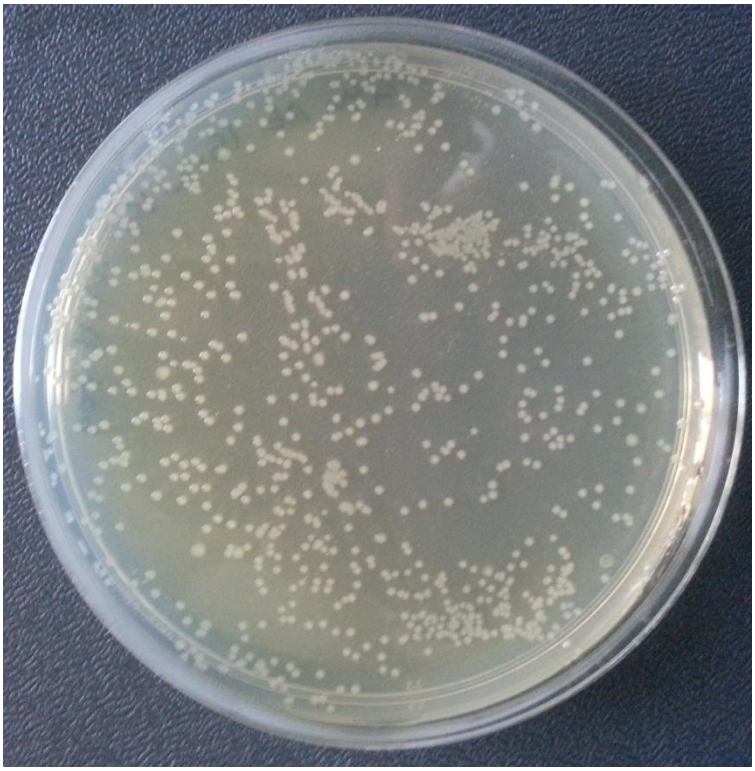


Figure 4.6: Colony forming units (CFUs) obtained after the transformation of competent *E. coli* K12 1821 by the ligation mixture described above.

4.4.4. Activity test after the mutagenesis

The mutagenesis range is represented by the 447bp which is the length of the insert (lane 7 Figure 4.5) so the DNA bases from the EcoRI position to the XbaI position (Figure 4.2) are all vulnerable to random mutation which means that the mutagenesis includes 147 amino acid residues from C-terminal side of the CATI primary structure (i.e. residue 72 to 219).

Each colony shown in Figure 4.6 represents a clone that could be inactivated by mutation(s) or remains active simply because of the mutation(s) are/is of no significant effect or no mutation has been generated so the wild DNA sequence remains intact. Investigating the activity is straightforward by picking single colony from the original ampicillin plate and sub culturing on a plate of both ampicillin and chloramphenicol so that clones of active CATI will exhibit

visible and strong growth while inactive and partially active clones will show no growth at all and weak growth (represented by few colonies) respectively (Figure 4.7).

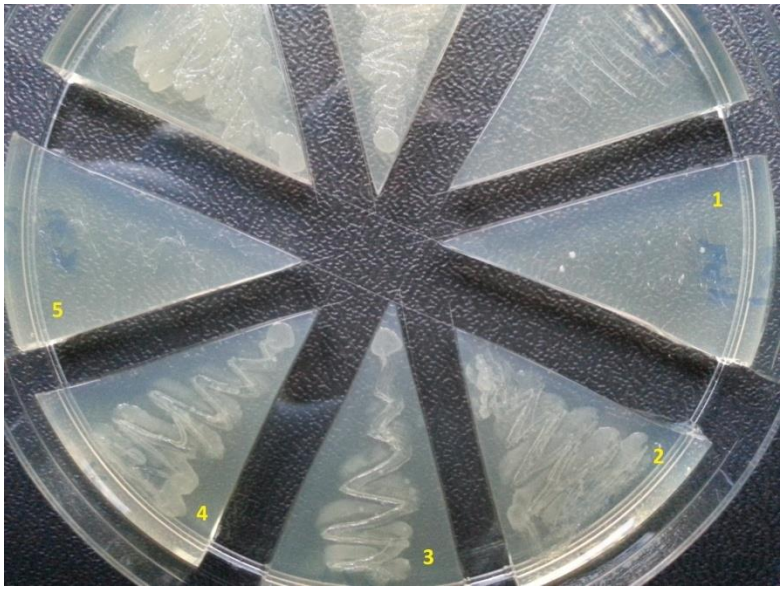


Figure 4.7: CATI Activity test. (1) clone of partially active CATI showing small and few colonies. (2), (3) and (4) clones of fully active CATI showing strong growth. (5) clone of inactive CATI showing no growth at all.

4.4.5. DNA sequencing

Two hundred clones from the original transformation plate (Figure 4.6) were randomly chosen; each of which was sub cultured in 5 ml LB containing 100µg ml^{-1} ampicillin for the sake of plasmid DNA miniprep that was being performed side by side with activity test. The two hundred plasmid samples were sent for sequencing. The results showed that 22 sample (11%) were with no change at all in nucleotide sequence, 17 sample (8.5%) have unclear sequence, 5 samples (2.5%) of silent mutation, mutants of deletions and insertions were 25 samples (12.5%) and 4 samples (2%) respectively. Mutation in the gene stop codon were 3 (1.5%), nonsense mutants or mutant clones with in-frame stop codon

were 26 (13%). On the other hand missense mutations were represented by 94 clones (47%). Since the objective was to isolate point mutations that cause amino acids changes as well as non-sense mutations, all clones that appeared with mutations of deletions, insertions and mutations in the TAA gene stop codon, which all abolish the gene ORF, have been ignored. Clones showing no change in nucleotide sequence as well as those of unclear sequence (such as over lapping chromatogram peaks or many Ns) have been ignored as well.

It has been noticed from the sequencing data that the number of amino acids changes is not necessarily caused by the same number of the nucleotide sequence point mutations. For instance, mutant clone of single amino acid change could have more than one (two or three) point mutation in its nucleotide sequence but when translated, the primary structure showed only single amino acid residue change putting the rest of the nucleotide base changes on the shelf of the silent mutations. The data regarding the correlation between the nucleotide changes and the amino acids changes and the consequent rates and incidences of the mutations silencing is provided in Table 4.1.

The distribution and arrangement of the CATI gene codons reflect the rate of silent mutations and how resistant the primary structure is toward random mutagenesis. Therefore, our mutagenesis strategy is a way by which the significance of the gene codons arrangement in term of mutational resistance or silencing could be estimated. However, mutations silencing could be explained as tool to make the enzyme able to maintain its primary structure and elude the function inhibitory mutations that may be caused by different factor in nature. The evasion of the mutational nucleotide changes is due to the presence of more than one gene codon (cognate codons) which are available to be translated into the same amino acid so that the random mutagenesis could simply change a certain codon into its cognate codon that is translated into the same amino acid and, therefore, such mutation is considered as silent.

Table 4.1: Silent mutations incidence.

Amino acid change Per clone	Clones obtained	Nucleotide base change								
		1	2	3	4	5	6	7	8	9
0	5	3	1	1	0	0	0	0	0	0
1	27	16	6	5	0	0	0	0	0	0
2	19	-	7	8	4	0	0	0	0	0
3	26	-	-	11	9	5	1	0	0	0
4	9	-	-	-	6	3	0	0	0	0
5	8	-	-	-	-	5	2	1	0	0
6	3	-	-	-	-	-	1	2	0	0
7	1	-	-	-	-	-	-	1	0	0
8	1	-	-	-	-	-	-	-	-	1
Total	99	19	14	25	19	13	4	4	0	1

The data showed some mutant clones are still expressing the wild CATI primary structure even when three DNA bases have been changed. In addition, the number of changes in DNA sequences in a significant number of mutant clones is more than the number of amino acid changes, in other words, mutations in DNA is not necessarily to be all sensed in the protein primary structure. For instance, 6 and 5 clones contain double and triple nucleotide change respectively but the primary structure showed only single amino acid change in all of them. 19 clones of double amino acid change have been obtained; 7 of them showed the same number of DNA base change but 8 and 4 clones of them showed double and triple N changes. Interestingly, one clone has been shown to have 6 DNA base changes but half of these base changes were silent so that only 3 amino acids changes were recorded. Furthermore one sample of three DNA base changes was with no amino acid change at all (see Table 4.1 for more data details).

4.4.6. Mutational frequency and CATI activity

The enzyme activity has been correlated with the number of sense mutations per clone to draw the enzyme tolerance toward the amino acids changes caused by sense point mutation (table 4.2).

Table 4.2: Correlation of the sense mutations' frequency with enzyme activity:

Mutation per clone	Clone frequency obtained	Acetylation activity		
		Active	Inactive	Partially active
1	27	17	5	5
2	19	3	10	6
3	26	3	21	2
4	9	2	7	0
5	8	1	6	1
6	3	0	3	0
7	1	0	1	0
8	1	0	1	0
Total	94	26	54	14
%	100%	27.65%	57.45%	14.9%

The results revealed that CATI tends to be more sensitive toward random mutagenesis than being resistant especially when more than half of the mutants (57.45%) showed complete loss of activity alongside with 27.65% and 14.9% that showed complete and partial activity respectively. The only noteworthy resistance toward the mutations was observed with clones of single amino acid change mutants which represent 27 clone; 17 of them are completely active while only 5 clones showed absolute inactivity and 5 appeared with weak growth (partial activity) on the chloramphenicol plate.

The remaining clones contained groups of two mutations and most gave rise to inactivation. The greater the level of mutation: the greater the probability of inactivation. Indeed, more than half of the mutant clones with the double amino acids changes were completely inactive, while only about 16% are totally active beside 32% exhibiting imperfect activity. Triple sense mutation clones, which are the second most prevalent after those of single amino acid mutation, arose with significantly affected CATI activity so that 21 clones out of 26 were 100% chloramphenicol sensitive while the minority, represented by 5 clones, divide into 3 and 2 mutants of 100% and partial chloramphenicol resistance respectively. Quadruple mutants where either 100% active or 100% inactive which constitute 2 and 7 clones respectively. The presence of one clone of quintuple sense mutation that catalysed full acetylation activity in addition to another one exhibiting partial activity reveals that the position of point mutations (which will be discussed later) in addition to their frequency are both playing a role in the enzyme activity alteration. Few clones of 6, 7 and 8 amino acid changes have been obtained showing no growth at all on the chloramphenicol LB plates indicating no significant CATI activity.

4.4.7. Mutations in conserved residues and enzyme activity

The amino acid residues conserved across the CAT family has been assessed by multiple sequence alignment of 86 chloramphenicol acetyltransferases of different types from a wide variety of bacterial sources (See Appendix 5). The minimum conservation parameter has been set arbitrarily at 60% in this analysis: which means that each residue that is present in at least 60% of the 86 sequences is considered to be conserved. According to the latter conservation parameters, 77 residues out of 219 of CATI appeared to be conserved. The random mutagenesis range incorporated 147 residue (57 of which are

conserved) formed by the polypeptide chain from E72 to A219. The mutants recovered, however, did not impact upon every single residue within this range. The incidence of conserved residues are relatively high; this may explain the observation that CATI is more sensitive to mutagenesis especially since the targeted region includes the active site and its environs. The activity of CATI has been correlated with the mutational incidence in conserved residues and the number of mutations per clone (Table 4.3).

Table 4.3: Conservation mutational changes and enzyme activity:

Mutation per clone	Single (1)		Double (2)			Triple (3)				Quadruple (4)				Quintuple (5)			
Frequency	27		19			26				9				8			
Conservation Change Incidence / Activity	0/1	1/1	0/2	1/2	2/2	0/3	1/3	2/3	3/3	0/4	1/4	2/4	3/4	0/5	1/5	2/5	3/5
Active	13	4	1	2	0	2	1	0	0	0	2	0	0	1	0	0	0
Inactive	1	4	2	6	2	5	7	8	1	1	1	4	1	0	2	2	2
Partially Active	3	2	1	5	0	0	2	0	0	0	0	0	0	0	0	0	1
Total	17	10	4	13	2	7	10	8	1	1	3	4	1	1	2	2	3

The data reveal that both the density of conserved residues and the frequency of mutations obtained clearly determine the likelihood of CATI remaining active. The mutant clones in Table 4.3 have been arranged into 5 groups according to the number of mutation per clone. It is clear from the activity data that mutations of conserved residues significantly affecting CATI acetylation activity either completely or partially. The majority of mutant clones (13 out of 17) of the single amino acid changes were fully active when mutations were not within the conserved amino acids residues, however, only one single mutation among the mentioned above 17 clones appeared to completely inactivate CATI

rendering the *E. coli* to be unable to grow in the presence of 5 µg/ml chloramphenicol. When outside of the conserved regions, single sense mutations could partially reduce CATI activity: 3 mutant clones out of 17 grew but weakly on the chloramphenicol plates.

The importance of sequence conservation in maintaining enzyme activity appeared with mutants harbouring double amino acid changes, so that only one mutation in a conserved residue is required to derive the most of clones (11 out of 13) to be affected by chloramphenicol either completely or partially.

Clones containing three missense mutations are generally inactive, even when those clones do not contain any change in the conserved positions and this is clearly shown by the data obtained here when only two clones out of seven appeared fully active. Similarly, when only one mutation happened in a conserved position of the clones of three amino acid changes, nearly the same proportion of fully active clones appeared (3 out of 10), while one more mutation in another conserved residue turns all clone to be fully inactive.

Interestingly, two mutant clones with four residue changes showed full CATI activity, even when one change happened in a conserved position, while 100% of them appeared completely inactive when two conservative mutations were introduced.

One clone was obtained carrying five amino acid changes and this grew strongly on chloramphenicol plates while the remaining clones of this group were all inactive. Furthermore, only one mutant appeared partially active when 3 out of its 5 mutations were at conserved residues.

The rest of the clones containing 6, 7 and 8 sense mutations which are all inactive, were not generally localised to conserved positions. Two clones of the three hexa-mutants include two conserved substitutions and the remaining one incorporates a single conservation change.

Finally, the role of primary structure in the formation of secondary structure elements in each CATI monomer were illustrated in Figure 4.8. In this figure, the complete set of mutagenesis and functional screening data could be mapped to evaluate the role of the mutagenesis-targeted residues in protein secondary structure. Furthermore, to evaluate the importance of conserved and non-conserved residues to the enzyme function, the mutations of single residue change with their relevant activity profiles were mapped directly into the CATI three dimensional structure (Figure 4.9).

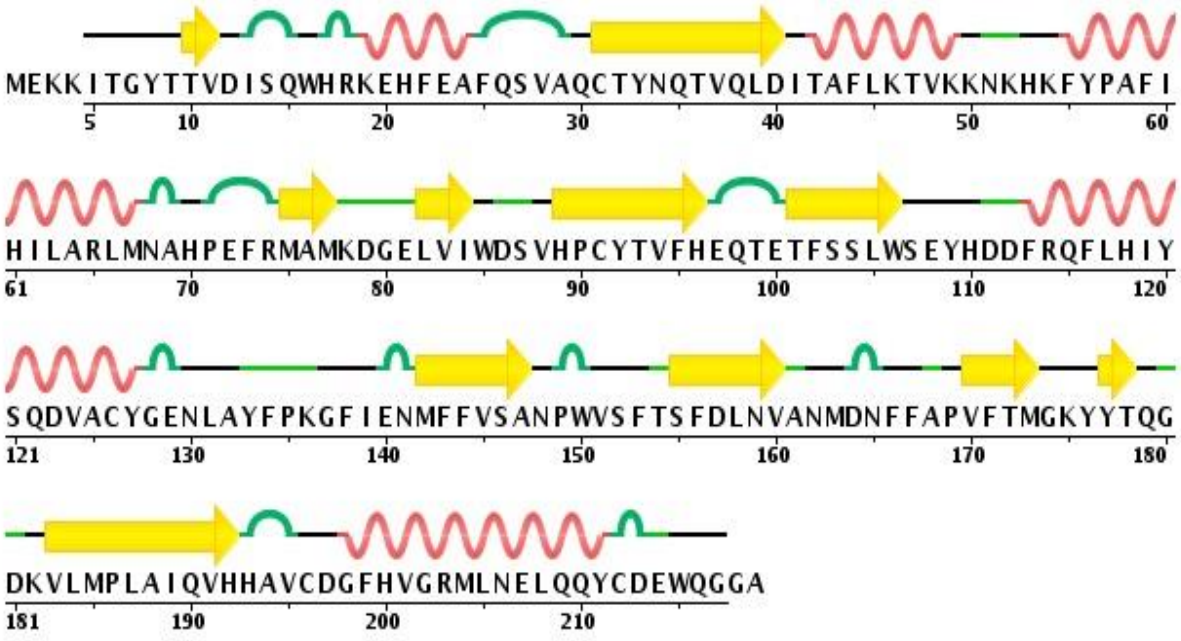


Figure 4.8: Combination of CATI primary and secondary structures showing the secondary structure elements and their relevant amino acid residues. β-strands are shown in yellow arrows, α-helices are the red zigzags, loops are represented by green arcs, and the black lines are the regions of undefined secondary structure. PDB code 3U9B.

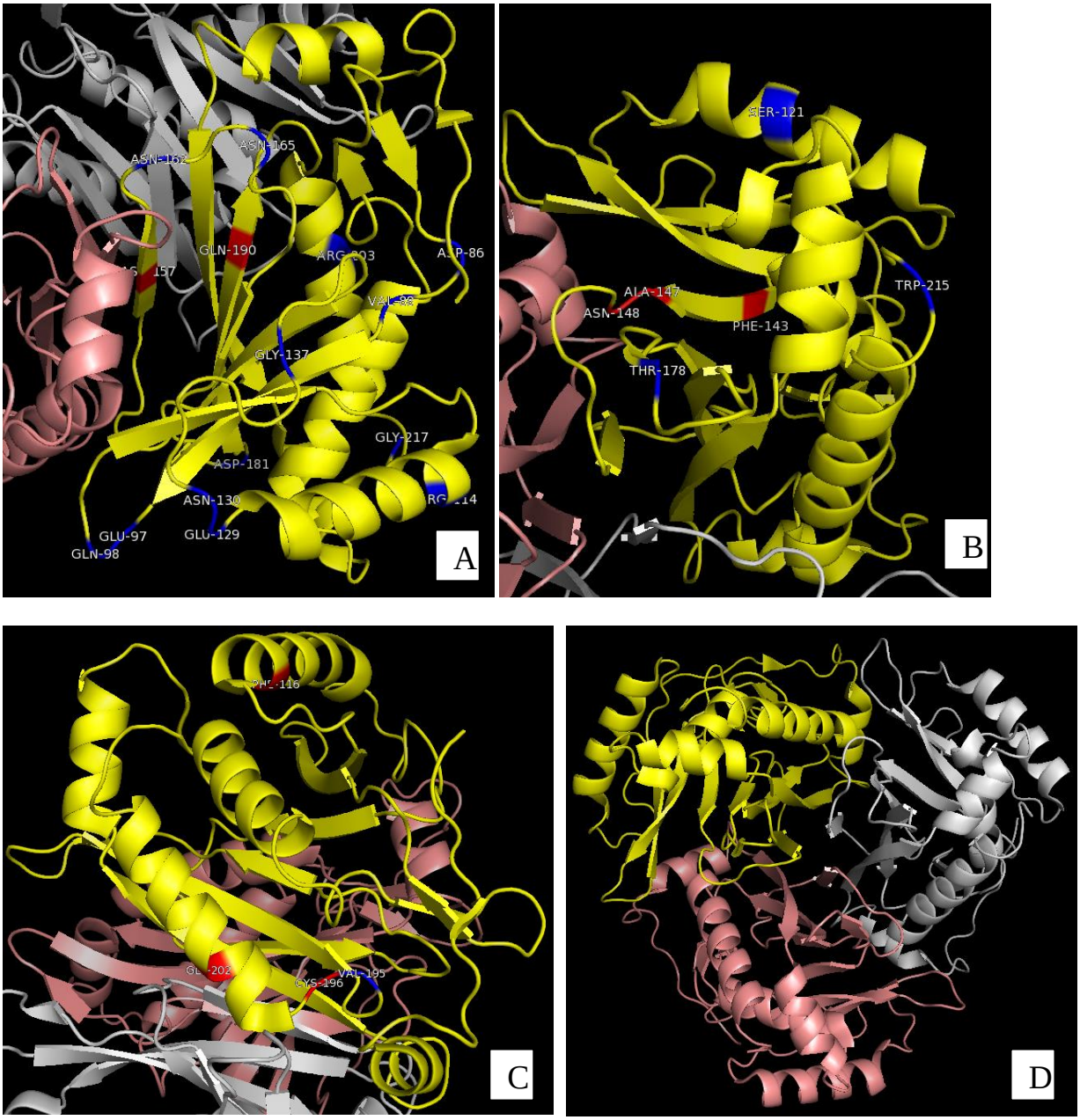


Figure 4.9: Four different views of the CATI three dimensional structure. View D shows the whole homotrimer structure. The three monomers are distinguished by colours as yellow, pink and gray. Views A, B and C highlight mutations recovered from CATI analysis: all residues highlighted represent single residue change mutants which are distributed equally within the three monomers, but here, they are shown only within a single monomer (the yellow one) for simplicity. Residues that affected the enzyme activity are labeled in red, while those which did not affect the activity are labeled in blue. (Details of the residue changes are all included in the appendices). Developed from the three dimensional structure which was resolved by Biswas *et al.* (2012), PDB code 3U9B.

4.4.8. CATI catalytic mechanism and the impact of mutations

Like all the catalytically studied classic chloramphenicol acetyltransferases, CATI has been shown to perform chloramphenicol acetylation via its catalytic histidine residue which is numbered 193 in the primary structure. The homotrimeric quaternary structure creates a triple active site formation. Each active site histidine lies in the cleft formed by the adjacent subunits. The R group of the catalytic histidine protrudes from one surface of the cleft so that the εnitrogen of the imidazole ring provides the required basic properties for the chloramphenicol 3-hydroxyl group deprotonation to produce an oxyanion intermediate which, in turn performs a nucleophilic attack on the second carbon atom of acetyl coenzyme A. This yields a tetrahedral intermediate, which then collapses into 3-acetoxy chloramphenicol and co-enzyme A. The acetyl group is then transferred to the C-1 hydroxyl of the chloramphenicol by a non-enzymatic reaction to produce 1-acetoxy chloramphenicol, therefore, the chloramphenicol 3-hydroxyl group will be free again and ready for a second enzymatic acetylation to form the final product of 1, 3-diacetoxy chloramphenicol (Figure 4.10) (Kleanthous and Shaw, 1984; Andreeva and Karamancheva, 2001; Biswas *et al.*, 2012). However, chloramphenicol acetylation at either 3-OH or 1-OH position has been proved to be sufficient to prevent the drug molecule from binding into its ribosomal target (Shaw and Unowsky, 1968; Bulkley *et al.*, 2010; Dunkle *et al.*, 2010).

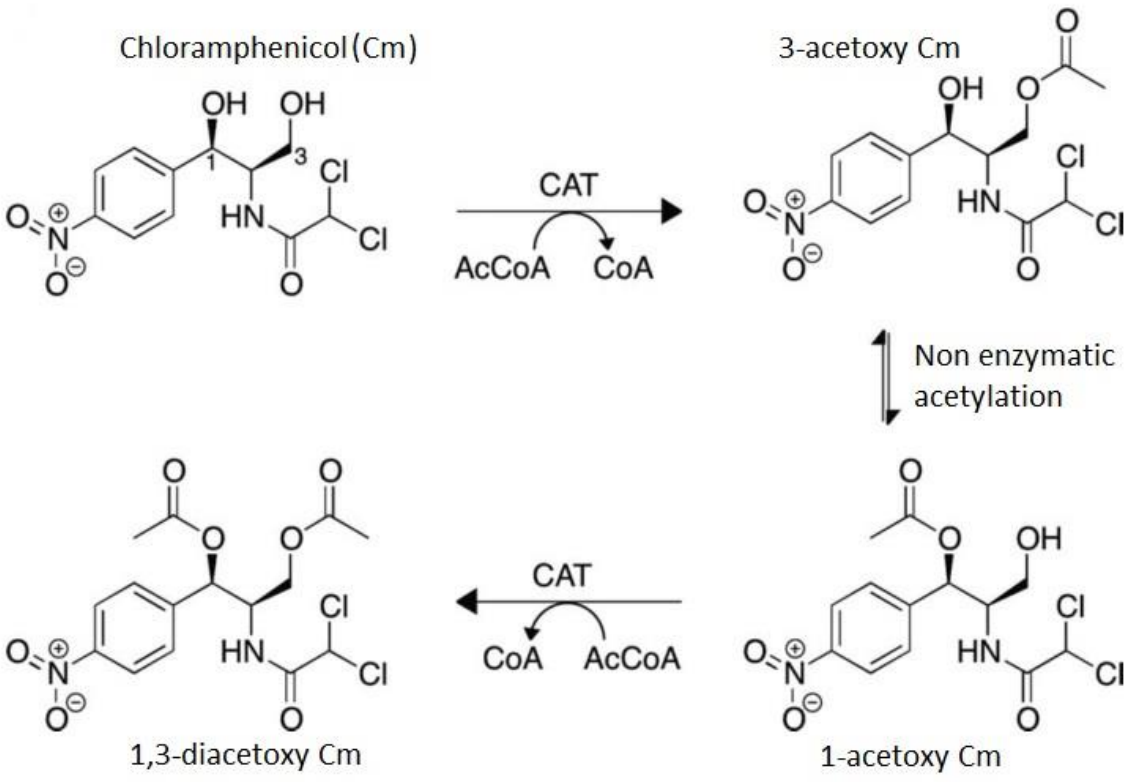


Figure 4.10: Chloramphenicol acetylation (Biswas *et al.*, 2012).

Our mutagenesis data showed three clones containing more than one mutation with the involvement of the catalytic H193 among these mutations. The mutational patterns of these clones were (1) Y120C, N130K, F171L, H193Y; (2) S103N, D111N, H193N and (3) D157V, H193L. All of these clones showed complete CATI inactivity. Although D157 is not conserved, mutation D157V occurred independently in another clone as a single sense mutation causing the CATI to be partially active but when accompanied with H193L the enzyme activity has been completely inhibited indicating the key role of H193 as a catalytic active site residue. All the aligned CATs variants (See Appendix 5) showed few replacements for the CATI D157, however, asparagine is the highly common residue which is very close in its structure and size to the aspartic acid so it is most likely for the CATI to be completely active with a mutation D157N. However, no valine is noticed to occur as one of the intended

amino acid replacements at that particular position, therefore, our mutation of D157V has been of partial acetylation activity especially when D157 participates in the trimeric structural stability of CATI.

It has been shown by the structural study of Biswas *et al.* (2012) that the overall CATI trimer is composed of three groups of beta sheets, each of which contains seven strands, six of them (β6, β5, β7, β9, β10 and β2) are from one monomer and one (β8) from the adjacent monomer. Five α-helices and three small β-sheets surround strand β8. Residue D157 is found within the β8 strand so that the aliphatic parts of R groups of all aspartate residues from the three monomers are intimately interacting with each other by hydrophobic interactions, on the other hand, the hydroxyl moieties of these R groups are forming a complex network of hydrogen bonds with the R groups of both S155 and N159. Coming together to form such hydrophobic interactions, aspartate residues in position 157 from each monomer are maintaining the conformation of the CATI trimer scaffold. The partial activity of D157V mutant could explain the incomplete loss of such interactions or the wild interactions has been replaced by other interactions which are suitable in the case of valine in the same position normally occupied by Asp. Maintaining the trimeric quaternary structure in a specific conformation, as a compact wild type CATI, is of course critical to providing the exact arrangement of atoms in the active site cleft, especially when the latter is not formed within a single monomer but between two monomers. The binding monomer (represented predominantly by residues within strands β6, β5, β7, β9, and β8) and the catalytic monomer (formed by residues of strands β2 and β10) so that CATI is only catalytically active when it is a trimer, therefore, any mutation that potentially affect the precise assembly of the three monomers could consequently cause activity loss that sometimes appears to be incomplete. Interestingly, CATIII which is very similar to CATI

in term of its quaternary structure, has been shown to be completely inactive as a monomer (Shaw *et al.*, 1970).

Mutants at position S155 have been obtained (accompanied by many other mutations) in 4 clones (See Appendix 2). All these clones showed complete CATI inactivity. S155 was mutated into threonine in two clones, while the other two clones showed S155C and S155I replacements respectively. When considering this residue in the context of the CAT multiple sequence alignments, S155 is frequently replaced by G while no T, C and I replacements are ever seen. As mentioned above, S155 participates in the maintenance of the trimer conformation: its side chain hydrogen bonds with the side chain of D157, however, the low conservation of S155 has been explained because of the D157 carboxyl group's ability to form hydrogen bond interactions with its own back bone amide (NH) disrupting the hydrogen bonding with the S155 side chain (Biswas *et al.*, 2012).

The crystal structure of a CATI-chloramphenicol complex (Biswas *et al.*, 2012) showed that the chloramphenicol binding monomer is predominantly formed by residues C91, F102, S104, F134, F144, S146, L158, and V170, while the catalytic monomer involves the most important active site residue H193 and other catalytic residues such as S146 and D197. Few residues like F25 and C31 are within the catalytic side of the active site cleft but their main role is in chloramphenicol binding. Our random mutagenesis experiment has covered both binding and catalytic parts of the enzyme. Two clones with mutation in C91 have been obtained as C91Y (with two additional mutations) and C91S (with one additional mutation). Cys-91 is not conserved but no tyrosine is observed as a replacement for the cysteine in that position while serine occurs many times throughout the CATs alignments to replace cysteine, therefore, partial loss of activity of C91S mutant is likely caused by the other accompanying mutation which is P169T especially when the P169 is conserved

residue. On the other hand, the mutant of the C91Y mutation is completely inactive.

F102 is one of the few absolutely conserved residues which has been mutated into isoleucine and tyrosine within quintuple and double mutant clones respectively. Both of these two clones showed 100% loss of activity. Within the clone of the five mutations, F102 is the only conserved among the other four mutated residues; one of them is F144 which is not conserved but it also participates in the chloramphenicol binding and mutated to Leu. Within the double mutant clone, the F102 mutation is found alongside another mutation at the conserved side chain N162. As it is absolutely conserved among the aligned CAT members, no replacements occur for F102. The absolute conservation could explain the complete CATI inactivation in two clones when F102 is mutated.

In addition to the above F144L mutation, an F144Y mutant has been seen in the only one clone of the seven sense mutations which encodes an inactive CATI. This clone involves an S146R mutation which seems inhibitory, as S146 is highly conserved and shows only one replacement represented by tyrosine in CAT from the *Clostridium botulinum* strain Eklund 17B / Type B, while all the other aligned CATs show serine in this position. The mutation at the S146 position was obtained two more times by random mutagenesis within two CATI inactive clones of triple and quintuple mutation as S146R and S146I respectively. The catalytic role of S146 in CATI has been proposed as the same of S148 in CATIII (Lewendon *et al.*, 1990) which is involved in hydrogen bonding (probably via a water molecule) and stabilization of the oxyanion intermediate during the acetylation (Biswas *et al.*, 2012). The triple mutant clone has another mutation in the highly conserved H192 as H192R which is within the catalytic face of the active site cleft. H192 is conserved in all the aligned CAT variants except CATIII from *Klebsiella pneumoniae* which has an

asparagine in this position. Combined together, S146R and H192R mutations are most likely the definite cause of the mentioned above triple mutant clone inactivation because the third mutation, which is E129D, is not in a conserved position in addition to the incidence of E129V as a single mutation in another fully active CATI clone.

The leucine at position 158 seems one of the important chloramphenicol binding residues because it appears conserved throughout the CAT alignments, with only few replacements as isoleucine which is close in structure to leucine. Histidine and glycine occur in only two and one CATs variants respectively at this position. Interestingly, random mutagenesis has reached this residue to replace L158 with proline; another hydrophobic residue but with significantly different side chain structure, to provide partially active acetylation activity in a context of quintuple mutation clone which has another four mutations; two of them are within conserved residues as well making the enzyme robust in performing its catalytic activity even under the pressure of high mutational frequency (5 simultaneous mutations in one clone) and many changes in conserved sequences.

The structural and biochemical studies have used the well-studied CATIII as a model to understand the catalytic and structural properties of CATI because they share the global conformational properties of being trimers as well as some detailed structural properties which emerged after the chloramphenicol-CATI structure was determined. This study showed a high degree of similarity between CATIII and CATI when both are complexed with chloramphenicol. The CAT structures reveal that the arrangement of the key catalytic residues of CATI can be more or less superimposed on to the structure of the intensively studied CATIII; especially the disposition of H193, S146 and D197 when the enzyme is bound to its antibiotic substrate. The relative position of the CATI catalytic H193 in relation to the bound antibiotic molecule is exactly the same

as its H195 equivalent in CATIII, sucho that the catalytic distance between the Epsilon nitrogen atom of the histidine imidazole and the 3-OH group of the bound chloramphenicol is 2.7A°. The latter distance and the appropriate conformation of the imidazole ring, enable the latter to easily abstract a proton from the drug molecule 3-OH group that will be nucleophilic enough to attack the Acetyl Co A and to" snatch" its acetyl in a mechanism similar to that of CATIII (Biswas *et al.*, 2012).

The proper catalytic positioning of the H193 is maintained by many residues within the catalytic monomer. The benzyl aromatic ring of F25 is at appropriate distance (3.6 °A) to provide face to face (π-π) stacking interaction to the H193 imidazole ring. All the aligned CAT variants presented in the Appendix 5 showed either Phe or Tyr at this position. This clearly explains the role of the Phe in CATI in making the mention above stacking interaction as Tyr is another aromatic residue with a benzyl side chain which most likely has exactly the same role in other CAT variants as its cognate Phe. Our random mutagenesis protocol, however, has not covered F25 for technical reasons.

Other residues that have been found to maintain H193 in its proper position to perform its catalytic role, D197, R18 and H193, are all involved in the formation of hydrogen bond networks with the oxygen of A194 carbonyl which is absolutely conserved in all the aligned CAT variants. The present random mutagenesis experiment has covered this residue in three mutant clones of double, triple and six amino acid changes. All these clones showed 100% loss of CATI activity. The clone of double amino acid change has the D164Yand D197C mutations; mutation in D164 residue appeared in two other clones of fully active CATI (see the clones of ID 4thAt33 15.2.15 and 4thAt175 15.2.15 in Appendix 2) proving that mutation D197C is the reason behind the CATI inactivation most likely because of the distortion in the structural arrangement

of the catalytic monomer that is required for the proper positioning of the H193 to perform the catalysis.

Multiple sequence alignments reflect the importance of some residues that are not necessarily absolutely conserved but they have only one or few replacement(s), for example, A194 in CATI is mostly conserved as Ala in all CATs but it could be replaced by Ser in many CAT variants (See Appendix 5). The mentioned above role of the A194 could have been performed by Ser otherwise, Ser wouldn't have been seen as a replacement for Ala in this particular position. Although the random mutagenesis has not hit A194, we assume that any mutation in CATI that may arise in this position to replace Ala with Ser should keep the enzyme active. Therefore, multiple sequence alignments are being used here to rationalize the activity data obtained alongside the random mutagenesis experiments.

Although the amino acid sequence alignments between CATI and CATIII revealed 9 out of 20 chloramphenicol binding residues to be different between the two enzymes (Biswas *et al.*, 2012), chloramphenicol bound structures of both CATI and CATIII showed that the antibiotic molecule is present in the same position in both complexes so that the *p*-NO_2group of the drug is facing the solvent while the aromatic ring rests on the hydrophobic surface of the active site when the antibiotic molecule is bound to the active site. The hydrophobicity on the mentioned above active site surface required for the chloramphenicol interaction is provided mainly by residuesL158 and V170. The former has been mutated into Pro and is discussed above, while V170, which is found as Ile (another hydrophobic residue) in most CATs, has not been reached by the random mutagenesis. The correct positions of the L158, V160 and F166 side chains, which line the lower part of the chloramphenicol binding site, are maintained by the assembly of the three monomers, therefore, any mutation that

could alter the correct assembly of the trimer is most likely affect the correct positioning of such residues that are enzyme's functional determinants.

The distance between the Y133 side chain hydroxyl and the carbonyl group of the chloramphenicol molecule (2.9 °A) makes it possible for strong hydrogen bonding to occur between these moieties (Biswas *et al.*, 2012). The MSA showed that Y133 is not conserved but has replacements that are hydrophobic most of the time such as Met, Leu or Ile, or of relatively similar alternative residue which is Phe. The random mutagenesis data showed mutations in Y133 within three clones; two of them are of three point mutations as Y133F, G137C, F153I and Y133D, G174S, D197E, and one clone of seven amino acids changes as E97A, T99N, C126Y, Y127C, Y133F, A168S, E207D, Y211H. All these three clones express completely inactive CATI. The reason of inactivity in clones that have Y133F could not be the latter mutation as Tyr was mutated into a very similar cognate which is Phe while the mutation Y133D, however, could be participating in the inactivity of the CATI as MSA is not showing any Asp alternative in this particular position.

4.4.9. Amino acid changes affecting CATI activity in relation to conservation

The enzyme tends to be inactive rather than being fully active after random mutagenesis, so the fully active clones are 26 (26.65%) out of the total of 94 clones (Table 4.2). All the mutagenesis and activity data for each single mutant clone related to CATI are listed in Appendix 2. Here, The full activity of the mutant clones are correlated with multiple sequence alignments to see if the new residues that are introduced into CATI primary structure by random mutagenesis are actually present throughout the MSA and therefore the enzyme activity is maintained or those mutagenesis-introduce residues are just novel

which means that they are not seen in the aligned CAT variants (See Table 4.4 below and Appendix 5).

Some mutations in some clones are discussed above in term of causing inactivation as these mutations happened in an absolutely conserved residues that could not theoretically be replaced by any other residue and, therefore, mutant clones of such mutations should express totally inactive CATI, or the mutation happened in a residue that is only possible to be replaced by another residue of the same properties to maintain the enzyme activity such as the replacement of phenylalanine by tyrosine which are both aromatic and have relatively the same size otherwise, inactive enzyme would be produced by any change in any of the absolutely conserved residue or any change in any position of conserved amino acid properties such as acidic, basic, aromatic or hydrophobic properties.

Table 4.4 Correlation of the amino acid changes in the active CATI mutant clones with the multiple sequence alignments MSA.

Amino acid changes	Changes seen in MSA	Changes not seen in MSA
R203G	-	R203G
D181V	-	D181V
E129V	-	E129V
N162K	N162K	-
A147V, D164E, A219T	D164E, A219T	A147V
F113I, E214K	F113I, E214K	-
E97V	E97V	-
D86E	D86E	-
R114S	R114S	-
S121I	S121I	-
G217S	G217S	-

G137C	-	G137C
D79Y, E81D, V191I	E81D, V191I	D79Y
M77K, A125S, M185L, L208V	M185L, L208V	M77K, A125S
N165I	-	N165I
D111E, S121G, P169S, Q179H	D111E	S121G, P169S*, Q179H
V83F, T99S, H118R, Q122H, D181N	T99S, H118R, D181N	V83F, Q122H
Q98H	-	Q98H
V88L	V88L	-
N130K	N130K	-
V195I	V195I	
D79Y, A132T	A132T	D79Y
W215R	-	W215R
T178I	T178I	-
H118R, D164E, Q210H	H118R, D164E, Q210H	-

Red shaded are conserved residues.

Red shaded with star (*) are residues that are absolutely conserved.

The CAT multiple sequence alignments show residues that are conserved in more than 60% of the aligned CATs but they still have few alternatives; if replaced with, the enzyme is assumed to still active. On the other hand, few regions of CATI appeared with absolutely conserved residues that are not replaceable with any other residue alternatives and if replaced, the enzyme is assumed to be inactive. In other words, for the CATI to be active, new residues introduced to the CATI amino acid sequence by random mutagenesis in conserved points should be seen throughout the MSA data otherwise the enzyme would not be active. While mutations in any point of not conserved residues are likely to be tolerated by the enzyme. The data in Table 4.4 have been used to test this assumption and it is obvious that the enzyme is still expressing acetylation activity even when mutational amino acid changes happened in conserved points introducing new residues (novel) that could not

be seen throughout the MSA at these particular points. Furthermore, even P169S mutation which is in an absolutely conserved point appeared in fully active quadruple mutant clone. Therefore, the general assumption of inactivating the enzyme when mutation occur in a conserved point does not always apply. This could be explained by the specific role of each residue in enzyme catalysis or maintaining the proper conformation required for the full activity. The role of some residues in relation to the activity data obtained after CATI gene random mutagenesis has been discussed previously in this chapter.

Certain amino acid change in an absolutely conserved point could cause complete loss of activity while different amino acid change at the same absolutely conserved point could give a different enzyme activity phenotype, for instance, H200 is 100% conserved residue, however, the data showed that mutant clone containing H200Y mutation among two other mutations (S107P, P135S) was completely inactive, while another mutant clone containing H200Q mutation among four other mutations (H96L, E97K, L158P, V183L) was expressing partial activity.

Certain residue characteristics (such as aromatic, acidic, basic hydrophobic, nucleophilic…etc.) are indeed required (conserved) at certain positions in the enzyme rather than a particular residue itself, therefore, in such case, amino acids of the same characteristic properties are seen throughout the MSA at any particular position of such residue properties conservation. For instance, aromatic amino acid tryptophan at position 85 in CATI has phenylalanine and tyrosine replacements most of the time; therefore, it is assumed that any mutation that replaces W85 into any residue of different characteristic should distort the enzyme activity. This could explain the complete inactivity of the double mutant clone containing W85G in addition to S155I mutations which are obtained here. Furthermore, our mutagenesis data has supported this statement from the other side when fully active CATI was obtained after V191I mutation

introduced. V191 is a conserved hydrophobic residue having leu and Ile replacements which are both hydrophobic as well.

Phenylalanine at position 145 in CATIII (F143 in CATI) has been proposed to be a key residue in initiating the protein folding (Van der Schueren *et al.*, 1998). Multiple sequence alignments showed that F143 is found in the majority of CATs with only hydrophobic residue replacements including leucine, isoleucine and valine. This could explain the absolute inactivity of our single mutant clone containing F143S mutation as serine is a hydrophilic residue while the hydrophobicity seems to be a required characteristic in this location

Sometimes certain groups in an amino acid side chain are required to be present in a proper position to serve specific function within the enzyme, for example, K136 has Q replacement most of the time. Both K and Q have amine group in their R group so this amine group seems the key moiety of the conservation. However, K136N cause partial loss of activity as N has amine group in its side chain as well but the small size of amino acid N could be the reason behind the improper interaction of the NH_2 as a result of changing the interaction appropriate distance by the overall residue small size. Another example is represented by an inactive triple mutant clone that hasT154Nmutation. T154 has only E, D and S replacements which all share the presence of OH group in their side chains which is probably the key reason for its conservation; not surprisingly then T154N is inactive .

4.4.10. C-terminal truncation and mutagenic effect on CATI activity

One interesting nonsense (TAG) mutation was recovered in one mutant clone, causing a truncation at the C-terminus, from position 216, resulting in removal of 4 residues from the CATI carboxylic end. This mutation is accompanied by another mutation which is V201I in the same clone. The CAT

activity in this clone was full. Truncation and mutagenesis study by Van der Schueren *et al.* (1996) found that even when seven C-terminal residues are removed by a deletion or nonsense truncation mutation, CATI is still expressing wild solubility and enzyme catalytic activity as long as the C-terminal α_5-helix is not included in such truncation. In other words all residues that are within the random coil following the α_5-helix have shown a possibility to be truncated leaving CATI with a full activity and solubility similar to the wild type. On the other hand, extending the truncation to be eight or more residues from the C-terminal has shown to be enzyme-inactivating and a cause behind CATI aggregation within inclusion bodies due to the loss of solubility. Residues within α_5-helix have been proposed to be involved in hydrophobic interactions crucial to maintain the correct CATI folding so that any mutation that may cause truncation or changing the required hydrophobicity of this C-terminal helix should devoid the enzyme activity not due to direct interfering with catalysis and/or chloramphenicol binding (as residues in such position are not parts of the active site) but mainly because of affecting protein folding (Shaw and Leslie, 1991). This concept has been tested when the three last residues of the C-terminal α_5-helix have been truncated and an inactive CAT was obtained demonstrating the importance of the full length presence of α_5-helix. In addition to the complete length of the α_5-helix, the amino acid content in this position is also required for correct folding. Random mutagenesis has shown the importance of some residues at the end of α_5-helix in providing the crucial interactions with other residues to enable the proper folding of the α_5-helix and consequently the global functional folding of the whole CATI scaffold(Van der Schueren *et al.*, 1996).

The hydrophobicity at the end of the α_5-helix is necessary for the folding and enzyme activity (Van der Schueren *et al.*, 1996). L208 particularly provides this characteristic property. Our data showed L208V mutation within a fully active

CATI mutant clone of quadruple mutation. The enzyme has not lost the activity here due to the replacement of L208 with another hydrophobic residue which is valine confirming the importance of the hydrophobicity in this particular position.

Chapter 5 Dihydrolipoamide dehydrogenase (E3): random mutagenesis and functional screening

Abstract

This chapter describes experiments to investigate the relationship between the primary structure and enzymatic activity of dihydrolipoamide dehydrogenase (E3). This enzyme contributes to the catalytic function of several multi enzyme complexes in *E. coli*. By applying error-prone PCR random mutagenesis to the E3 gene (*lpd*), the impact of mutations on two of these complexes (PDH and ODH) is investigated. While the catalytic activity of E3 is the same in both PDH and ODH complexes, the interaction between E3 and the other subunits within both complexes are different. The results suggest that the primary structure of E3 is remarkably adaptable and it proved challenging to obtain mutations that would impact differently in respect of these two multi enzyme complexes.

5.1. Introduction

Enzyme complexes are aggregates of several enzymes that are assembled together through noncovalent interactions to perform successive steps in an enzyme-catalysed reaction. It has been demonstrated that such catalytic functions can work better when assembled together, rather than being separated into discrete steps. Although there are many instances where Nature utilises separate enzymes, more complex organisms seem to favour multi-protein assembly over separate steps. Perhaps enzymes in such complexes provide more opportunities for regulation such as the simplicity of substrate channelling as well as intermediate stability and protection against factors, such as pH and temperature (Reed, 1974; Perham, 1975; Hammes, 1980).

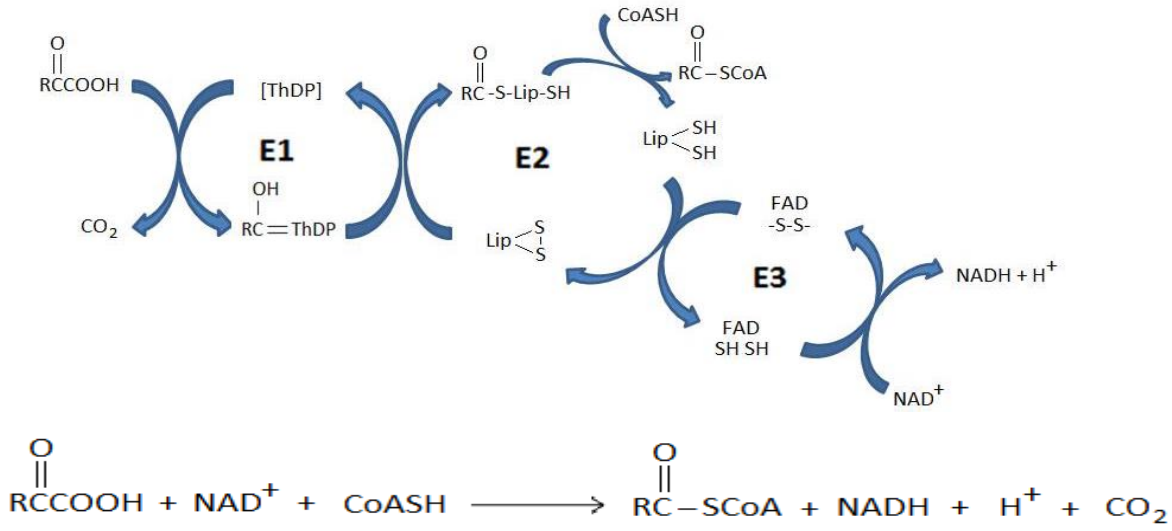
Dehydrogenase enzyme complexes that catalyse redox reactions with 2-oxo acids (Reed, 1974) are built from three enzymes, E1, E2 and E3; each of which presents as multi-copies within a complex catalysing the oxidative decarboxylation of 2-oxo acids as illustrated in Figure 5.1.

Pyruvate dehydrogenase (PDH), 2-oxoglutarate dehydrogenase (ODH) and branched-chain oxo-acid dehydrogenase (BODH) complexes have their own specific E1 and E2 components which confer the substrate specificity to a complex, while the same E3 commonly participates in all three complexes within the same organism except in *Pseudomonas putida* where three different genes of dihydrolipoamide dehydrogenase have been identified; two of them were shown to be part of puruvate dehydrogenase and branched-chain 2-oxo acid dehydrogenase complexes respectively (Burns *et al.*, 1989). The PDH complex is composed of pyruvate decarboxylase (E1p), also known as pyruvate dehydrogenase (lipoamide) (EC1.2.4.1); dihydrolipoamide acetyltransferase (E2p, EC2.3.1.12) and dihydrolipoamide dehydrogenase (E3, EC 1.8.1.4). ODH complex is composed of 2-oxoglutarate decarboxylase (E1o, EC l .2.4.2); dihydrolipoamide succnyltransferase (E2o, EC 2.3.1.61) and dihydrolipoamide dehydrogenase (E3, EC 1.8.1.4). The oxo-acids of short branched chain produced from the transamination of amino acids valine, leucine and isoleucine are decarboxilated oxidatively by BODH complex which has analogueous structure as those of PDH and ODH complexes (Randle *et al.*, 1987; Reed and Yeaman, 1987). A further enzyme reaction is the glycine reductase of *E. coli*, which also makes use of the same E3 polypeptide, but will not be discussed further here.

During catalysis, all the three enzyme components (E1, E2 and E3) of the oxo-acid dehydrogenase complexes utilse a set of coenzymes. In the PDH complex, which is the prototype of the 2-oxo acid dehydrogenase complexes, pyruvate decarboxylase is thiamine diphosphate (ThDP) dependent where the

thiazole ring of thiamine (vitamine B1) acts as the functional ring via its nucleophilic carbon atom 2 that attacks the α-carbon of the substrate

(A)



(B)

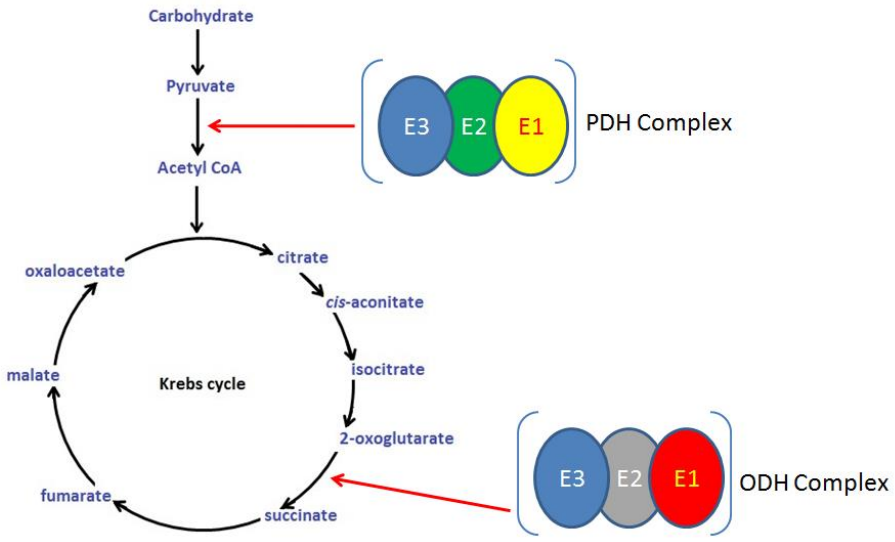


Figure 5.1: (A) Schematic and the net reactions adapted from (Reed, 1974) illustrating the oxidative decarboxylation of 2-oxo acids by 2-oxo acid dehydrogenase complexes. ThDP, thiamine diphosphate; Lip, lipoic acid. (B) The role of PDH and ODH complexes in metabolism.

pyruvate to form a 2-α-lactyl derivative that is decarboxylated into acetaldehyde and leaves the E1-ThDP. Dihydrolipoamide acetyltransferase (E2) operates in conjunction with lipoic acid, a saturated fatty acid with a tail of 8 carbon atoms: the eighth and sixth carbon atoms are joined together by a disulfide bond forming a five membered ring. Lipoic acid is tightly linked to E2 by a peptide bond formed between its carboxyl group and the amine group of a lysine from E2. The –S-S- bond of the E2-lipoyl is reduced by E1 when the acetaldehyde is transferred from ThDP to one of the sulphur atoms of the lipoic acid causing the latter to lose the ring structure and present with two thiol groups at the end of the lipoic acid tail. E2 catalyses the transfer of hydrogen atoms from acetaldehyde to the disulfide bond of the lipoyl moiety; as a result the acetaldehyde is oxidized to acetyl and the lipoic acid is reduced to its dithiol form. The acetyl group is then linked by E2 to the thiol group of coenzyme A so acetyl-coenzyme A, which is the key component linking glycolysis to tricarboxylic acid cycle, is formed and leaves the enzyme complex to react with oxaloacetate to form citrate within the tricarboxylic acid cycle (Krebs cycle). The lipoyl moiety of E2 moves freely as a swinging arm in the complex transferring the substrate between the successive active sites of the complex components (Reed, 1974; Ambrose and Perham, 1976; Grande *et al.*, 1976). The E2-lipoic acid which is left in its dithiol form is inactive unless reoxidized back to its disulfide bonded form. Lipoic acid oxidation is catalysed by the dihydrolipoamide dehydrogenase (E3) which works via FAD as a tightly bound hydrogen acceptor coenzyme abstracting the hydrogen atoms from the two thiol groups of lipoic acid to form $FADH_2$in order that the –S-S- bond cyclic form of the lipoic acid is recovered. $FADH_2$ is then re-oxidized by NAD^+ to recover the oxidative FAD again, and NADH is formed. The catalytic mechanisms of the other 2-oxo-acid dehydrogenase complexes is comparable with the mentioned above PDH mechanism (Lehninger, 1975; Perham, 1991, 2000; Patel *et al.*, 2014).

2-oxo acid dehydrogenase complexes are 30-40 nm particles as seen by electron microscopy: as such they are larger than a ribosome particle (Reed, 1974; Henderson *et al.*, 1979; Oliver and Reed, 1982). The E2 component forms a central core to which E1 and E3 components are bound strongly by non-covalent interactions. The E2 cores of ODH, branched-chain 2-oxo acid and gram negative bacterial PDH complexes are all composed of 24 copies of E2 that are arranged in an octahedral symmetry, while in PDH of eukaryotes and gram positive bacteria, 60 chains of E2 are arranged in an icosahedral symmetry (Reed, 1974; Danson *et al.*, 1979; Oliver and Reed, 1982).

In *E. coli* PDH, 24 molecule of the E1 dimers are arranged on the 12 edges of the octahedral core and 12 copies of the E3 dimers are arranged on the 6 faces of the E2 core so that the E1:E2:E3 molecules ratio is 24:24:12 reflecting an active site ratio of 1.0:1.0:0.5, while the active site ratio in ODH is 0.5:1.0:0.5 (Reed, 1974; Oliver and Reed, 1982; Yang *et al.*, 1985). The component ratios of different complexes from other sources of organisms were referred to in (Perham, 1991).

The specific feature of such enzyme complexes is that E2 subunits are arranged as a cube or dodecahedron with E1 and E3 attached to it non-covalently as peripheral proteins. The whole complex particle is up to 10 KDa (Reed, 1974). In eukaryotes, additional protein components may be added to the 2-oxo acid dehydrogenase core complexes such as kinases, phosphatases and dihydrolipoamide dehydrogenase-binding proteins which all participate in metabolic regulation (Roche *et al.*, 1989).

5.2. Dihydrolipoamide dehydrogenase random mutagenesis

The aim of the work in this chapter is to apply PCR random mutagenesis using an error-prone pho polymerase to the *lpd* gene of *E. coli* E3, to generate libraries of mutations. By combining this with functional screening of pyruvate

dehydrogenase (PDH) and 2-oxoglutarate dehydrogenase (ODH) complexes, the aim is to dissect out any differences in E3 in these two related, but distinct complexes.

5.2.1. *Lpd* gene synthesis

The sequence features of *E. coli* K12 *lpd* gene studied by Stephens *et al.* (1983b) have been taken into consideration when the gene was synthesised. The nucleotide sequence has been taken from the *Escherichia coli* strain K12 sub strain W3110 DNA complete genome (NCBI reference sequence: NC_007779.1, ORF sequence from base 127912 to 129336). Wild type *lpd* gene sequence is recognisable by two ATG start codons which are preceded by a strong promoter region within the upstream 230 bp. The first ATG is too close to the ribosomal binding site to be the initiation codon, furthermore, the rules of Stormo *et al.* (1982) do not agree with it to initiate the translation, therefore, the second ATG has been considered the actual start codon leading to the translation of *lpd* ORF into 474 amino acid residue. As first ATG is not considered the start codon, it has been omitted as a result of NcoI restriction site introduction at the beginning of the ORF leaving the second functional ATG intact (Figure 5.2).The ORF upstream sequence that includes the putative promoter region with the RNA polymerase binding site, and the -35 and-10 pribnow regions as well as the ribosomal binding site are all included in the synthetic sequence within the upstream 231bp. Many restriction sites (each site is unique in the final construct) have been introduced silently within the ORF as well as flanking the ORF to facilitate downstream processing of the error-prone PCR products (Figure 5.2). The synthetic sequence has been further optimized for *E. coli* K12 expression using the optimization option provided by gene synthesis company. The synthesized sequence has been cloned into pMA-RQ plasmid which has ampicillin resistance. The final plasmid construct of pMA-RQ+*lpd* gene has been named pLpd1 (Figure 5.3).

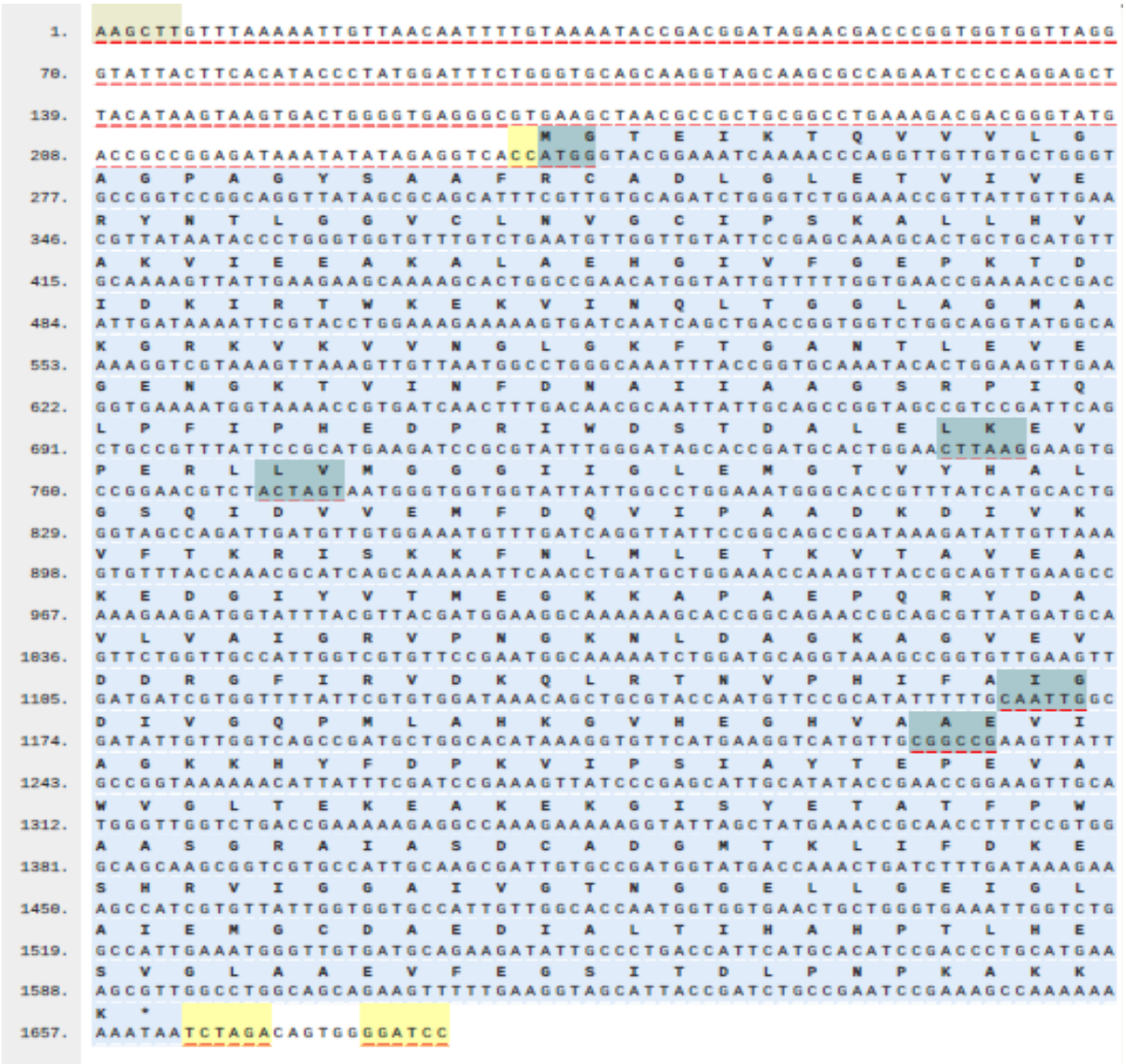


Figure 5.2: Interleaved view of the whole synthetic sequence showing the primary structure of the dihydrolipoamide dehydrogenase (474 amino acid). The ORF is cyan-shaded. The in-frame and out-frame restriction sites are shaded and underlined by red dashed lines. (AAGCTT=HindIII, CCATGG=NcoI, CTTAAG=AfIII, ACTAGT=SpeI, CAATTG=MfeI,CGGCCG=EagI,TCTAGA=XbaI, GGATCC=BamHI). Each restriction enzyme has only one cutting site in the final pLpd1 construct.

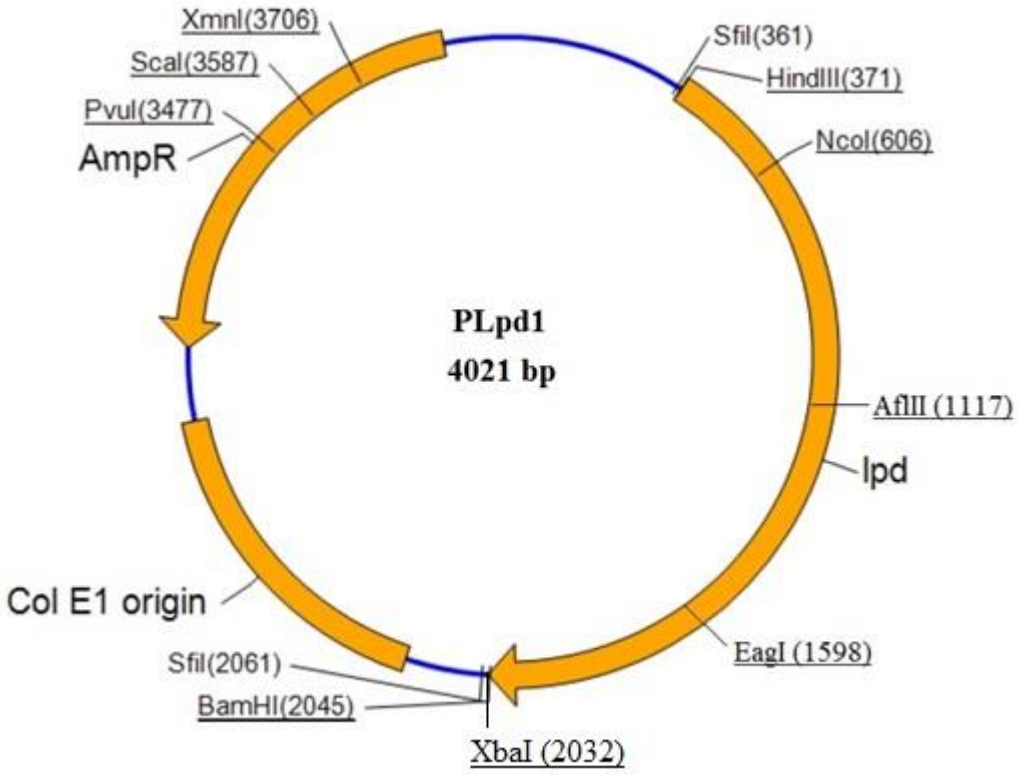


Figure 5.3: pLpd1 plasmid construct map as provided from the gene synthesis company showing *lpd* gene of the dihydrolipoamide dehydrogenase (E3) and ampicillin resistance gene (AmpR). The origin of replication in this plasmid is Col E1. Several restriction sites are indicated with their base pair cutting positions in brackets.

5.2.2. E3-knockout *E. coli* strain and *lpd* gene complementation

The activity of E3, its gene is carried on the pLpd1 plasmid, has been tested by transforming the E3 null *E. coli* K12 W3110 strain (JRG3503, *lpdA : : kan*^{R}) (kindly provided by Professor Jeff Green) in which, 661 bp segment of *lpd* gene has been replaced by a kanamycin resistance gene so that it is not able to grow on a glucose minimal media without a dual supplementation of acetate and succinate to compensate both PDH and ODH impaired functions as a result of E3 deficiency (John *et al.*, 2003). Acetate is a precursor of acetyl-coenzyme A,

therefore, it is used to overcome the disappearance of PDH function that is responsible for providing acetyl-coenzyme A from pyruvate oxidation. Succinate is supplemented to support the JRG3503 strain growth in glucose minimal media as a replacement for the succinate produced naturally from the 2-oxoglutarate oxidation carried out by ODH in the wild-type W3110 strain within the citric acid cycle.

JRG3503 strain has been shown to have no growth response at all in the non-supplemented and acetate-supplemented M9 glucose minimal media, on the other hand, succinate-supplemented M9 showed extremely slow growth turbidity while the growth was significantly visible when both acetate and succinate are added to M9. These growth results have been recorded after 48 hours incubation at 37°C and they are consistent with (John *et al.*, 2003). In LB, JRG3503 grows normally, but not similar to the wild type, even when no acetate and/or succinate is provided because *E. coli* in LB uses amino acid in its catabolism as a carbon source rather than sugars (Sezonov *et al.*, 2007).

The lyophilized 5µg pLpd1 plasmid construct, which was provided by the gene synthesis company, was rehydrated using 50 µl of ddH_2O then 1 µl from the plasmid solution was used to transform electrocompetent JRG3503 by electroporation as the chemical method of (Hanahan, 1983) did not work with it. After transformation, the bacteria have been plated out on kanamycin+ ampicillin-containing M9 glucose minimal medium plate without any supplementation of acetate and/or succinate and incubated for 48 hours at 37°C. The results showed many CFUs representing a successful transformation and E3 complementation by the multicopy pLpd1 plasmid. However, the complementation of the missing wild type chromosomal *lpd* gene by a multicopy plasmid carrying *lpd* gene has been previously shown to be possible by (John *et al.*, 2003).

A single colony from the M9 glucose minimal plate was sub-cultured in 5 ml LB then plasmid prepared prior to sequencing, to check the integrity of the plasmid after transformation, as recommended by the gene synthesis company. Three primers were used for sequencing to cover the complete *lpd* gene. The sequence after transformation was identical to the sequence received from the company.

5.2.3. *Lpd* random mutagenesis and cloning

Three sets of mutagenesis experiments were performed using three sets of primers to include three parts of *lpd* gene in random mutagenesis. The first set of mutagenesis covers the region from NcoI to AflII (Figure 5.3). The range of mutagenesis was extended to involve a more extensive section of the gene in the second set by using the same forward primer of the first set but with different reverse primer to cover the part between NcoI and EagI sites. The third mutagenesis experiment was targeted between restriction sites AflII and XbaI. Thus, the whole length of the *lpd* gene was included by these three sets of random mutagenesis experiments.

In each set of mutagenesis experiments, the forward primer was designed to initiate synthesis about 80bp upstream the 5′ restriction site while the reverse primer position is situated nearly 80bp downstream of the 3′ restriction site so that the error-prone PCR amplicon is produced with restriction sites at both 5′ and 3′ ends. This facilitated trimming of the PCR product to produce sticky ends for efficient ligation back into pLpd1 that is in turn double digested by the same restriction enzymes. Therefore, the protocol is to double digest both of the error-prone PCR amplicon and pLpd1 by the same restriction enzymes, then separate the bands by agarose gel electrophoresis (Figure 5.4) to recover the bands of interest (the mutation-containing insert and the vector) and ligate them together then transform JRG3503 (Figure 5.5).

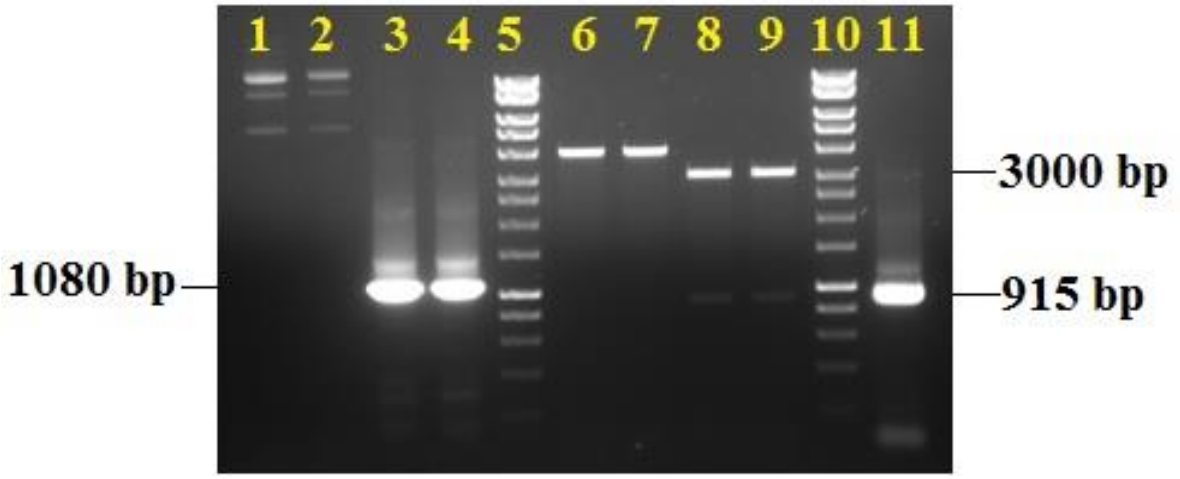


Figure 5.4: Agarose gel showing the DNA fragments related to the E3 error-prone PCR and cloning. : (1) and (2), uncut pLpd1; (3) and (4), error-prone PCR products; (5), hyperladder I; (6), pLpd1 cut with AflII; (7), pLpd1 cut with XbaI; (8) and (9), pLpd1 cut with AflII and XbaI; (10), hyperladder I; (11), error-prone PCR product cut with AflII and XbaI.

5.2.4. PDH and ODH activity test

As mentioned previously, JRG3503 is able to grow normally on LB without any additional nutritional supplements, therefore, LB plates with 100 µg/ml ampicillin and 25µg/ml kanamycin were used to plate out the transformed JRG3503 just after electroporation so that all the clones of active and inactive of both or either PDH and ODH enzyme complexes were recovered (Figure 5.5). However, testing the PDH and ODH activity was by sub-culturing each clone in 5ml LB overnight then precipitating the cells by centrifugation at 4500 rpm removing the LB then washing the cells three times with M9 glucose minimal media to remove of the LB traces completely. The activity of each clone then investigated by plating the washed bacterial cells on three M9 glucose plates all with 100 µg/ml ampicillin and 25µg/ml kanamycin. The first plate is only M9 glucose, the second is M9 glucose+acetate and the third is M9

glucose+succinate. The incubation was done at 37°C for 48 hours. The possible complexes inactivation and activity phenotypes are illustrated in table 5.1. The clones of active PDH and ODH grow normally on the first plate that has no additions with normal growth on both M9 glucose+acetate and M9 glucose+succinate plates because such clone are able to use the glucose as a carbon source. Clones of inactive PDH and ODH give no growth at all on all three plates but they grow only when both acetate and succinate are supplied to the M9 glucose medium. Clones of inactive ODH and active PDH are not able to grow on both M9 glucose and M9 glucose+acetate plates but they show visible normal growth on M9 glucose+succinate. Finally, clones of inactive PDH and active ODH would not grow on M9 glucose and M9 glucose+succinate but grow on M9 glucose+acetate (Figure 5.6).

Table 5.1: Illustration of the probable activity patterns regarding PDH and ODH with the growth behaviour on different minimal media replicates in addition to LB:

Possible activity pattern	Growth phenotype on M9 glucose media ofdifferent supplementations in addition to LB				
	M9 glucose only	M9 glucose+acetate	M9 glucose+succinate	M9 glucose+acetate+succinate	LB
Active PDH Active ODH	+	+	+	+	+
Inactive PDH Inactive ODH	-	-	-	+	+
Active PDH Inactive ODH	-	-	+	+	+
Inactive PDH Active ODH	-	+	-	+	+

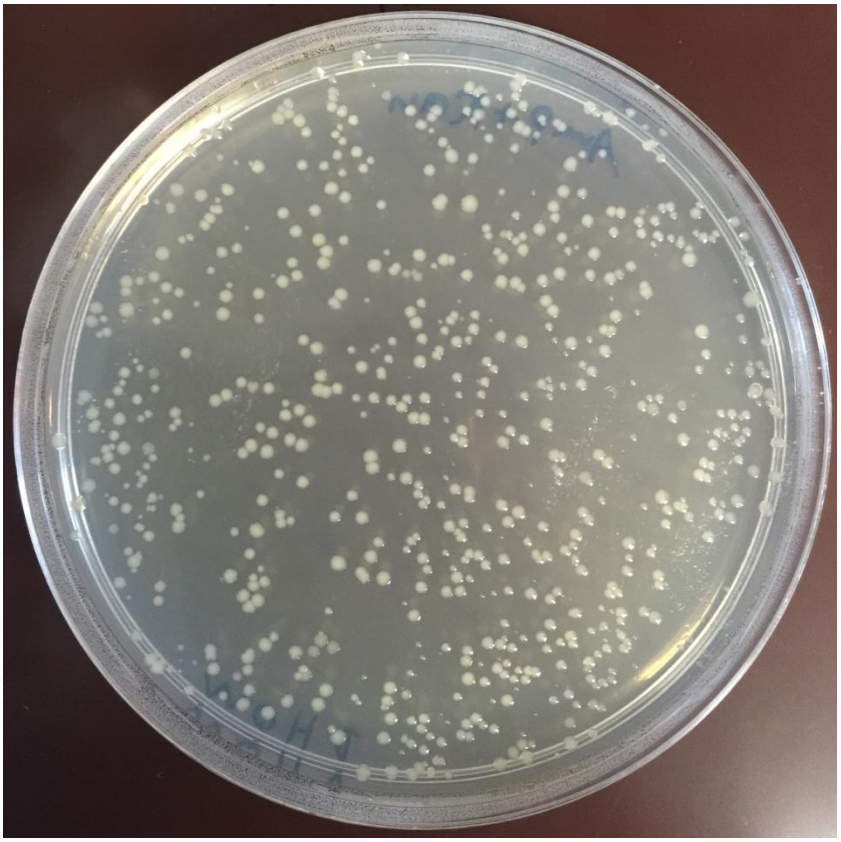


Figure 5.5: LB plate containing ampicillin and kanamycin showing the CFUs of the mutant clones after the error-prone PCR, ligation and JRG3503 electroporation transformation.

5.2.5. DNA sequencing

After the PDH and ODH activity test, DNA plasmid minipreps were prepared from randomly chosen clones obtained from each set of error-prone PCR mutagenesis and cloning then sent for sequencing to determine the nucleotide changes in *lpd* gene and the consequent amino acid change in the dihydrolipoamide dehydrogenase. The total clones that were screened (activity test and sequenced) numbered 248. The sequencing data showed that 49 clones (19.75%) showed no change in nucleotide sequence at all; most of them occur in the first set of the mutagenesis experiments as this set involves the shortest length of the DNA among the three sets. In view of the full activity of both

PDH and ODH clones obtained after this set of mutagenesis experiments it seemed sensible to extend the mutagenesis target length, to increase the probability of capturing mutations. However, the sequencing of the clones from the second and third mutagenesis sets of the DNA-extended mutagenesis range showed a significantly reduced number of clones that show no nucleotide change in addition to few clones showing enzyme inactivation.

The clones of unclear sequences (19 clones=7.66%), such as those with overlapped sequencing chromatogram peaks and Ns containing sequences, have been ignored. The deletion and insertion occurred with 11 (4.43%) and 1(0.4%) clone respectively. The point mutations that are of amino acid changes are represented by 152 clones; 3 of them appeared to have the same mutational pattern of other 3 clone. Amino acid changes of the clones of point, deletions and insertions mutations with their activity data are all illustrated in Appendix 3.

Comparisons between the sequences of DNA and amino acids for the clones of amino acid-changing point mutations showed that the number of amino acid changes are not necessarily the same number of nucleotide changes; for example, a clone of single amino acid change does not necessarily have a single nucleotide change that caused the single amino acid replacement but it could have two, three or even four nucleotide changes and only one of these point mutations is sense causing the amino acid change while the other(s) is/are silent. The statistics of the mutagenesis silencing are illustrated in table 5.2.

Table 5.2: Correlation between the nucleotide changes and amino acid changes showing the mutagenesis silencing:

Amino acid change Per clone	Clones obtained	Nucleotide base change							
		1	2	3	4	5	6	7	8
0	16	14	2	0	0	0	0	0	0
1	52	36	11	3	2	0	0	0	0
2	50	-	31	16	3	0	0	0	0
3	27	-	-	18	6	2	1	0	0
4	13	-	-	-	8	2	2	1	0
5	4	-	-	-	-	1	3	0	0
6	3	-	-	-	-	-	1	1	1
Total	165	50	44	37	19	5	7	2	1

The clones of point mutations have been statistically analysed and grouped according to the numbers of both amino acids and nucleotides changes to figure out the rate of mutagenesis silencing. When grouped according to the amino acid changes, the results showed that all the groups of mutant clones have significant incidence of silent mutations when no match is seen between the number of amino acids changes and the number of nucleotide changes. For instance, 16 clones appeared with no amino acid change at all but they do show nucleotide changes in their DNA sequences represented by 14 and 2 clone of single and double nucleotide changes respectively.

Most of the point mutations clones were of single and double amino acid change constituting 52 and 50 clones respectively. 36 of the single amino acid change mutant clones appeared with single nucleotide change while the 11, 3 and 2 clones of them showed double, triple and quadruple nucleotide changes respectively. The rate of mutational silencing in the double amino acid change

clones is slightly higher than the rate of single amino acid change clones especially when 31 of them were noticed with double nucleotide change while 19 of them are divided as 16 of triple and 3 of quadruple nucleotide changes.

18 out of 27 of the triple amino acid change mutant clones have the same number of nucleotide mutations which is three while the nucleotide changes exceeded the amino acid changes by one, two and three nucleotides in 6, 2 and 1 clones. The number of nucleotide changes in most of the quadruple amino acid change clones (8 out of 13) match the number of sense mutations while 1, 2 and 3 silent mutations appeared in 2, 2 and 1 clones respectively. The minority of the mutant clones are those which contain five and six amino acids changes which are represented by 4 and 3 clones respectively with only one clone in each group that match in its amino acid changes the number of nucleotide changes while the rest of the clones in both two groups appeared excessive changes in nucleotides to be more than the amino acid changes.

5.2.6. Mutation position and activity

The present mutagenesis work is intended to test the activity of dihydrolipoamide dehydrogenase (E3) indirectly by testing the activity of two enzyme complexes in *E. coli*: PDH and ODH in which E3 is a main component. The activity test showed that the vast majority of the mutant clones retained active PDH and ODH by their ability to grow on minimal glucose M9 medium without nutritional supplements of acetate and succinate. On the other hand, only 7 clones appeared with different activity phenotype; in all of them, ODH activity appeared to be affected. Six clones were found with ODH-affected activity beside a fully active PDH while only one clone showed affected activity of both PDH and ODH (Figure 5.6). All clones of affected ODH and PDH showed partial activity of both complexes except one clone that appeared with

completely inactive ODH beside a completely active PDH (See Appendix 3). The interaction between E3 and the other complexes components must be looked at in term of PDH and ODH activity rationalisation after E3 mutagenesis.

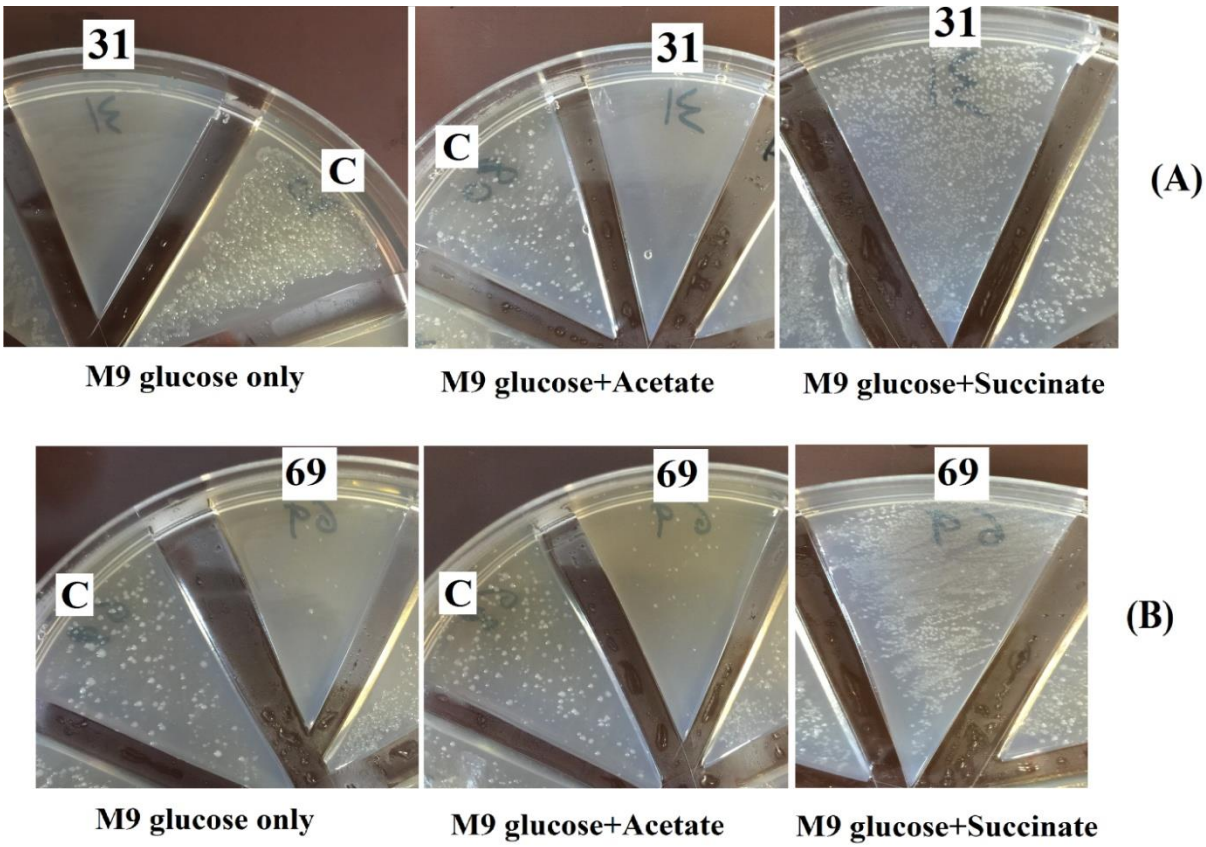


Figure 5.6: Activity test of PDH and ODH complexes after E3 mutagenesis. (A): Mutant clone 31 that has completely inactive ODH beside active PDH showing no growth at all on M9 glucose and M9 glucose+acetate while on M9 glucose+succinate, it shows normal growth. (B): Mutant clone 69 that has very weak ODH complex beside active PDH showing very weak growth on M9 glucose and M9 glucose+acetate while normal strong growth on M9 glucose+succinate. ‘’C’’ are controls which are clones expressing two active complexes of PDH and ODH showing normal growth on M9 glucose alone as well as on the other M9 variants of acetate or succinate supplements.

5.2.7. E3 interaction within PDH

In eukaryotic and prokaryotic PDH complexes, E2 is composed of a multi-domain structure of three domains linked to each other by an alanine and proline rich inter-domain hinge regions which are 20-30 amino acid long as follows: (1) lipoyl domains, to which, lipoic acid is bound. In *E. coli*, E2 contains three lipoyl domains which are ~80 amino acids each, while in human, there are two domains with the same amino acids length of those in *E. coli*. (2) The C-terminal catalytic domain, which is about 250 amino acid in size, represents the internal cubic core of the E2, in which, acetyl co A is formed. (3) Peripheral subunit-binding domain (PSBD) which is ~45 amino acids in length and represents the part of E2, to which, E3 and E1 is bound in *E. coli*. However, in human, E2 binds E1 and E3 differently by a specific E1-binding domain and E3-binding protein respectively (Bleile *et al.*, 1979; Stephens *et al.*, 1983a; Patel *et al.*, 2014).

In *E. coli* and most prokaryotic PDH complexes, there are 24 copies of E2 assembled together by interactions between their catalytic domains to form the central core of the complex which contains the PSBD where both muticopies of E1 (12 dimers) and E3 (6 dimers) are bound non-covalently via the PSBD to form an octahedral symmetrical PDH complex with a molecular mass of 4.5 MDa. Therefore, the E3 only interacts with E2 and there is no direct interactions between E1 and E3 (Reed and Hackert, 1990; Perham, 2000).

Mutagenesis studies demonstrated that the interactions of both E1 and E3 with the E2 core is electrostatically controlled by the charges of certain residues in E2 PSBD so the mutational inverse charge such as changing Arg-129 and Arg-150 into Glu strongly affects the interaction of E3 and E1 with the PSBD of E2 and therefore preventing the functional assembly of PDH complex in *E. coli*. However, E1and E3 interact with E2 by non-identical but overlapping positions

especially when some mutations, such as R129A, in E2 PSBD only prevent E3-E2 interaction with no effect on E1-E2 interaction while other E2 PSBD mutations like R150A and R150K only affect E1-E2 interaction without affecting E3-E2 interaction. This suggests different modes of interaction for E1 and E3 with E2 (Park *et al.*, 2004; Chandrasekhar *et al.*, 2013).

E3 is homodimer (50554 Da) of two identical subunits with two active sites located in the interface cleft between the two subunits. Each subunit is composed of four domains; FAD-binding domain (residue 1-149), NAD^+-binding domain (residue 150-282), the central domain (residue 283-350) and the interface domain (residue 351-474) (Figure 5.7). Each molecule of E3 tightly holds two molecules of FAD which are involved in the electrons transfer from the two thiol groups of the E2 dihydrolipoamide.

One of the E2 interaction peptides which is located in the α-helix H1 is involved in the E3 interactions in all 2-oxoacids enzyme complexes of resolved structures except the 2-oxoglutarate dehydrogenase complex (Nakai *et al.*, 2008), the Arg-333 within this peptide of 3-lip E2 (corresponds to Arg-129 in 1-lip E2) is highly conserved within the all recognised peripheral subunit binding domains of E2 and considered as the hot spot key residue of interaction with E3 (Mande *et al.*, 1996). In the whole PDH complex, E1 is also bound to the E2 PSBD with no competition between E3 and E1 on binding as they have different but, at the same time, overlapping binding loci on the E2 PSBD (Song and Jordan, 2012).

The available crystal structure of E3 helped the study and identification of the interaction loci between E3 and E2 in *E. coli* PDH by applying hydrogen/

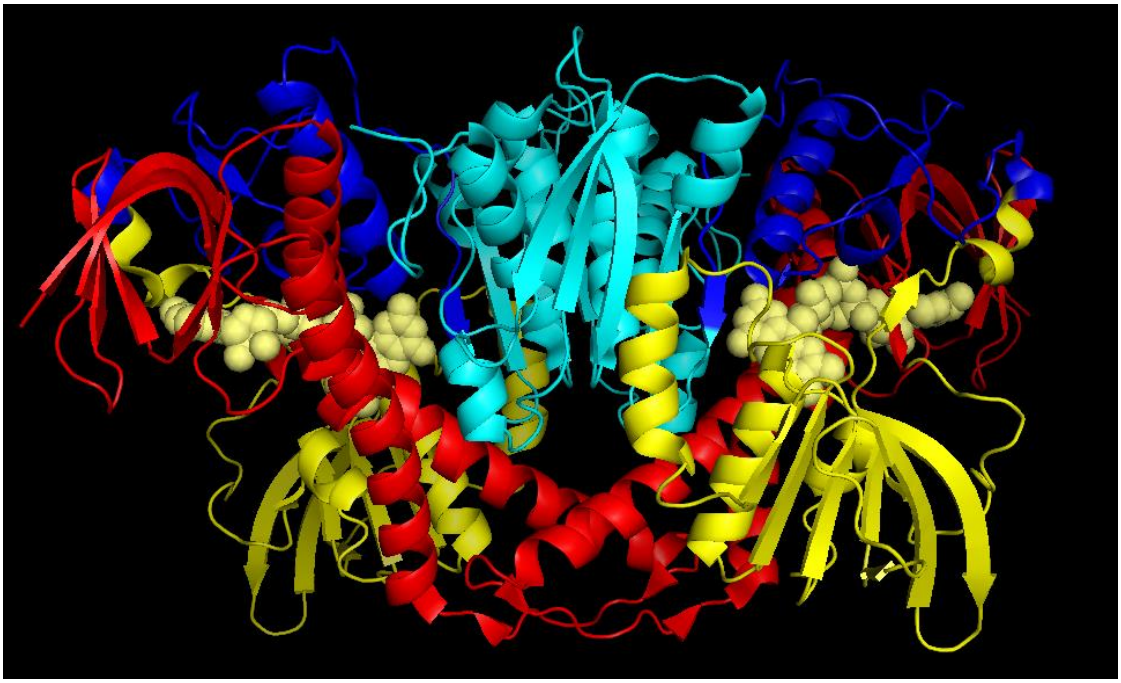


Figure 5.7: E3 homodimeric functional structure from *E. coli* showing the four domains. FAD-binding domain (red), NAD^+-binding domain (yellow), the central domain (blue) and the interface domain (cyan). FAD molecules are shown as pale yellow spheres. Developed from (Chandrasekhar *et al.*, 2013), PDB code is 4JDR.

deuterium peptide specific exchange mass spectroscopy which showed that there are significant deuterium uptakes by two peptides from E2 and three peptides from E3 when E2 and E3 are bound in a sub-complex proving the involvement of these peptides in the E2 and E3 interaction. The two peptide of E2 are both situated within the PSBD and represented by residues 325-338 in the wild 3-lip E2 (^{325}YVHATPLIRRLARE^{338}) which corresponds to residues 121-134 in 1-lip E2, and residues 363-393 of the 3-lip E2 (^{363}YVKEAIKRAEAAPAATGGGIPGMLPWPKVDF^{393}) which represent residues 159-189 of the 1-lip E2. The three interaction peptides of E3 are two peptides from the FAD-binding domain (^{33}VIVERYNTLGGVCLNVGCIPSKALL^{57} and

^{89}WKEKVINQLTGGLAGMAKGRKVKVVNGLGKFTGANTL^{125}) and one from the interface domain (^{458}VFEGSITDLPNPKAKKK^{474}). (Chandrasekhar et al., 2013).

To understand the allosteric basis of E3 and PSBD interactions within PDH complex, a structural model of E3-PSBD in *E. coli* has been postulated based on the available structure of the complexed E3-PSBD of *Thermus thermophilus* (PDB code 2EQ8) (Nakai *et al.*, 2008). According to this proposed model, the PSBD positively charged basic Arg-129 forms electrostatic interactions via its positively charged guanidinium nitrogen with the negatively charged oxygen atoms of carboxyl or carbonyl of E3 Asp-343 and Glu-430, therefore, the E2 variant that has R129E mutation shows no binding to E3 because of the charge repulsion between the same charge residues of E2 and E3. The same scenario occurs with the E2 mutant that contains R150E mutation as Arg-150 is in electrostatic interaction with the nucleophilic carbonyl oxygen of the E3 Met-431 backbone. Beside the electrostatic interactions, the main chain carbonyl oxygen and the guanidinium nitrogen of PSBD Arg-129 are involved in hydrogen bonding with hydroxyl group of E3 Tyr-341 and the carboxylate oxygen of E3 Asp-343 respectively. Such hydrogen bonds are important in strengthen the electrostatic interactions (Chandrasekhar *et al.*, 2013).

In the present random mutagenesis work, an active E3 mutant variant containing M431V mutation beside I218N was obtained. In spite of the role of Met-431 in the interaction with the E2 Arg-150, it seems that M431V E3 variant is still able to bind PSBD the same as the wild type because the wild type methionine in this position is participating in interaction with E2 by its main chain as mentioned above so there is no particular specific R-group required in such interaction, furthermore, valine has the same hydrophobic properties as methionine, therefore, M431V E3 mutant maintains the wild type binding and consequently PDH complex activity.

The present activity data showed that PDH complex is extremely resistant toward the E3 mutagenesis so only one mutant clone shows partially active PDH beside partially active ODH (See Appendix 3). This demonstrates that the E3 binding with E2 in PDH complex is challenging to be affected by E3 random mutagenesis. This could be because the presence of many residues which are outspread throughout an expanded E3 regions within three peptides as mentioned above which makes it hard to affect such binding by few point mutations within those three peptides as E3 will still has many other interacting residues to be bound to E2 and form a sub-complex within PDH and does not face any interaction distortion with the E2 PSBD. The only one mutant clone that shows a partially affected PDH has R273H, V357I, K365Q, G426D mutations which are all out of the ranges of E2-interacting residues. This clone shows a partially affected ODH activity as well which means that residues at these points are of NAD binding (R273) and catalysis (V357, K365, G426).

5.2.8. E3 interaction within ODH complex

The structural arrangement of 2-oxoglutarate dehydrogenase complex (ODH) is comparable to that of PDH complex where 24 copies of dihydrolipoamide succinyltransferase (E2 of the ODH which is distinguished as E2o from the E2p of PDH) form an octahedral core, to which, both 2-oxoglutarate decarboxylase (E1o) and dihydrolipoamide dehydrogenase (E3) bind (Reed, 1974). Proteolysis experiments on *E. coli* ODH have revealed that E2o has a multi-domain structure represented by the large domain or the C-terminal domain (~300 residue) where E1o binds and the succinyl transfer occurs , the N-terminal domain (~80 residue) which contains a single lipoate arm linked to a lysine residue, and the middle domain (~50 folded residues) which is between N and C terminal domains where E3 binds (Packman and

Perham, 1986). The latter domain is flanked by interdomain segments which are usually rich in alanine and proline as well as some residues of charged side chains. Such linkers are believed to be important in domains separation as well as providing flexibility for domains movement when required during the enzyme catalysis (Radford *et al.*, 1989; Perham, 1991).

The binding of E2o and E3 is studied but less extensively than the cognate binding of E2p and E3 in PDH. According to nuclear magnetic resonance spectroscopy data, the three-dimensional solution structure obtained by (Robien *et al.*, 1992) clarified the E3-binding positions in the dihydrolipoamide succinyltransferase (E2o) to be only within the region between residue 115-149 which are folded as two parallel helices, an irregular loop structure, two short strands, and a helix-like turn. The 115-149 peptide chain has been reproduced as a synthetic peptide in term of NMR and binding study. Compared with E2p which has two E3-binding peptides as mentioned above, E2o has only one peptide to bind E3 and this suggests that the binding of E2o-E3 sub-complex is likely to be weaker than the binding of E2p-E3 sub-complex. The E3 parts that are involved in binding with E2o have not been determined by any structural or spectroscopy study yet.

The current random mutagenesis study has revealed many clones of affected ODH complex only rather than PDH (See Appendix 3) suggesting that E3 could interact differently with E2p and E2o respectively otherwise, both PDH and ODH would have been affected if E3 interacts with E2o by the same E3 three peptides involved in E2p interaction. Evidently, two interesting mutant clones that express completely inactive and very weak ODH beside completely active PDH have mutations (Q300H, I409N, G423S, A471V) and (R302H, G328S, P380L, T397I) respectively. All the latter mutations are out of the range of the E3 three peptides which are involved in the E2p interaction suggesting the importance of all or some E3 residues among Q300, I409, G423, R302, G328,

P380, T397 and A471 in the specific sub-complex formation with ODH only rather than PDH.

5.2.9. E3 structural annotations and random mutagenesis impact

In additions to the sequence variations among the E3 from different sources which are considered as moderate variations (28-47%), the comparison of the tertiary structures of E3 from *E. coli* with many other E3 of resolved structures showed similar conformational properties and folding, however, the differences that have been reported to date (including deletions, insertions or small differences in folding) are only within the regions 36-42, 126-136, and 251-264 of *E. coli* E3. The regions mentioned do not participate in FAD or NAD cofactors binding with few exceptions that showed some of those regions to be participating in few interactions but, in general, they are constituting parts of surface loops that could be crucial in maintaining the surface interactions of E3 with the other components of 2-oxoacid complexes in their related organisms (Chandrasekhar *et al.*, 2013).

The crystal structure of *E. coli* E3 has shown that there are three proline residues (Pro-318, Pro-355, and Pro-446) arranged in a suitable positions to form peptide bonds of cis conformation. The multiple sequence alignments (MSA) shows that Pro-318 is not conserved and it is not always found in cis-conformation in the other known structures of E3, but only within E3 from *E. coli*, *Neisseria meningitides, and Thermus thermophilus*. Such cis-conformation is responsible for the proper orientation so that the Gln-317 forms a hydrogen bond between its OE1 atom and the main chain nitrogen atom of the Met-319. Prolines at the same positions from other E3 enzymes, such as *Pseudomonas putida* E3(Pro-310), is in trans-conformation so the OE1 and OE2 of preceding Glu-309 is not involved in any hydrogen bonding because such hydrogen bonding is only maintained when cis-conformation is there.

The present random mutagenesis has reached Pro-318 in two mutant clones of active E3; one of them has the mutation P318A in addition to three other mutations (K61R, L232P, K299Q), the other is a double mutant clone which appeared P318S and I425V mutations. Such mutational changes in Pro-318 abolish the cis-conformation and the hydrogen bonding built on it which seems not that important for the enzyme activity especially when both PDH and ODH remain active in both of these mutant clones.

Proline at position 355 is located in the tight β-turn of anti-parallel β-sheet of the interface domain. This residue is highly conserved and of peptide bonding of the cis-conformation which is featuring a specific hydrogen bonding between the oxygen of Thr-353 carbonyl and the nitro of the Glu-356 amide (Chandrasekhar *et al.*, 2013). The current random mutagenesis has not hit the Pro-355 but does hit both Thr-353 and Glu-356 in three clones; two of them, which are of three and five mutations, included changing Thr-353 into Ala and Ile respectively, and the third clone is of single mutation of changing Glu-356 to Asp. All these three mutant clones express fully active E3. Although Thr-353 seems to be conserved in more than 70% of the aligned E3 enzymes, T353A mutation seems to be tolerant mutation. However, when looking at MSA of some E3 variants, Ala is seen to replace Thr at 353 position many times, while Ile is not seen as a replacement at this particular position. Although the MSA does not show Asp replacement to the Glu-356, the full activity of mutant clone of E356D single mutation could be because the replacement is just between two amino acids of the same acidic properties and only one hydrocarbon unit is the difference between their side chains. As mentioned above, the hydrogen bonding between Thr-353 and Glu-356 occur between their backbone moieties and does not include any of their side chains, therefore, changing these residues (side chain change) should not affect the hydrogen bonding as long as the newly introduced side chains do not affect the proper conformation for the backbones

interactions. Such hydrogen bonding is not present in E3 of *Pisum sativum* (Faure et al., 2000) and *Bacillus stearothermophilus* (Mande et al., 1996) as their resolved crystal structures show no cis-conformation at proline corresponding to Pro-355 of *E. coli* E3 (Chandrasekhar *et al.*, 2013). Such hydrogen bonding absence in these enzymes could explain the enzyme tolerance when this interaction is abolished by mutation(s).

The proper catalytic orientation of active site residue His-445 is maintained by hydrogen bonds between the latter ND1 atom and OE1 and OE2 atoms of the Glu-450 as well as between His-445 carbonyl oxygen and the N3 atom of the Pro-446 aromatic ring (Chandrasekhar *et al.*, 2013). The MSA shows that both Glu-450 and Pro-446 are highly conserved due to their important role mentioned above, however, neither His-445 nor Glu-450 and Pro-446 have been hit by the present random mutagenesis work. This could explain the results of E3 full activity, which has been obtained most of the time, as the gene mutagenesis randomisation has been away from such sensitive points of the enzyme.

Finally, the role of primary structure in the formation of secondary structure elements in each E3 monomer are shown in Figure 5.8. In this figure, all of the mutagenesis and functional screening data are mapped to evaluate the role of the mutagenesis-targeted residues in protein secondary structure. Furthermore, to evaluate the importance of residues to the enzyme function, positions of mutations in E3 mutant clones of affected enzyme complexes activity were mapped directly into the E3 three dimensional structure (Figure 5.9).

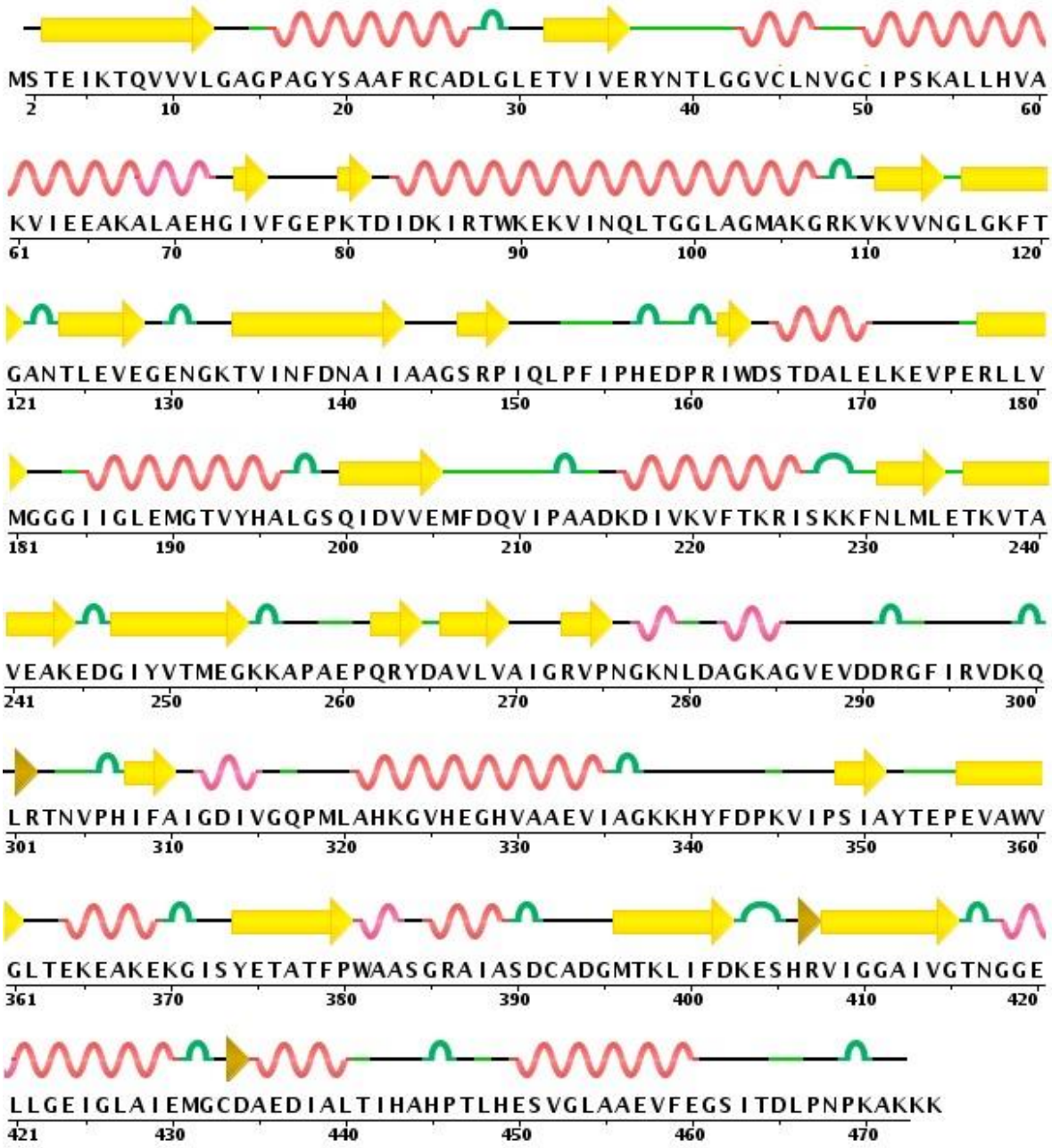
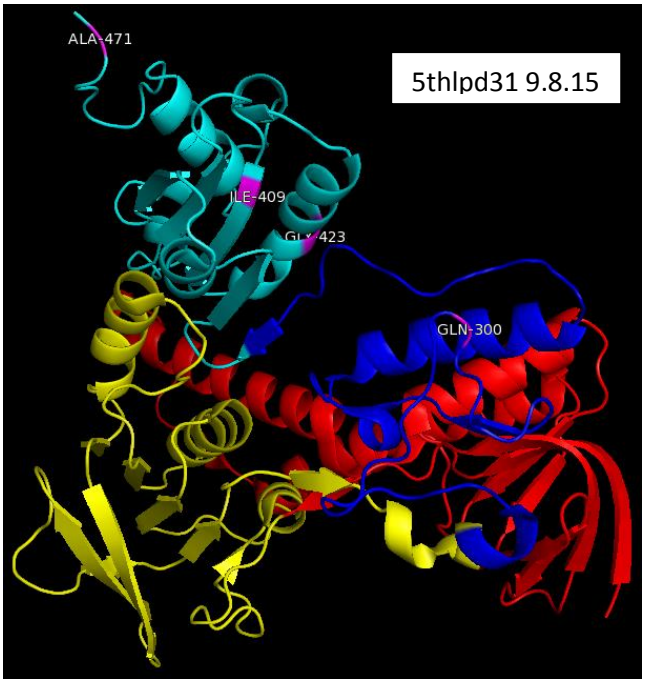
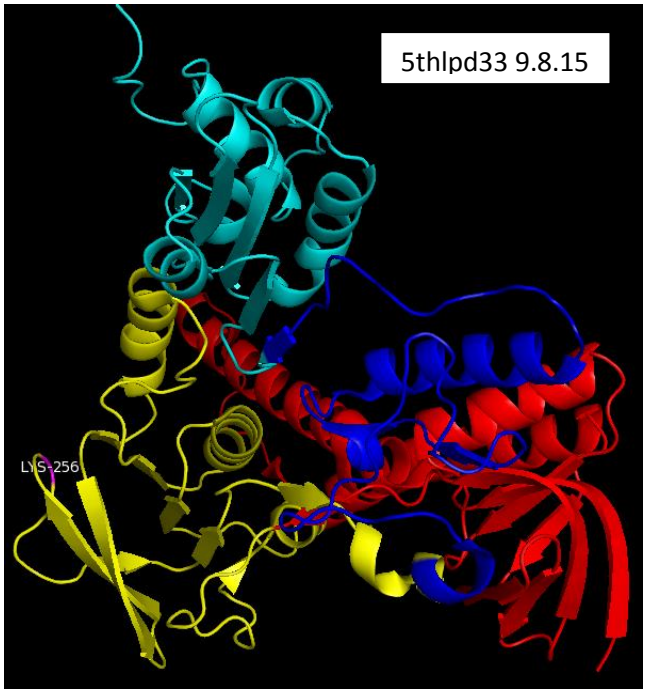
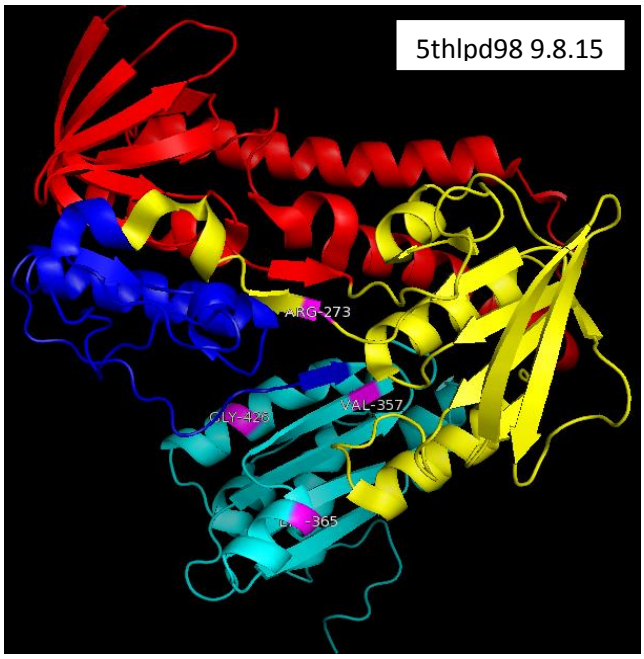
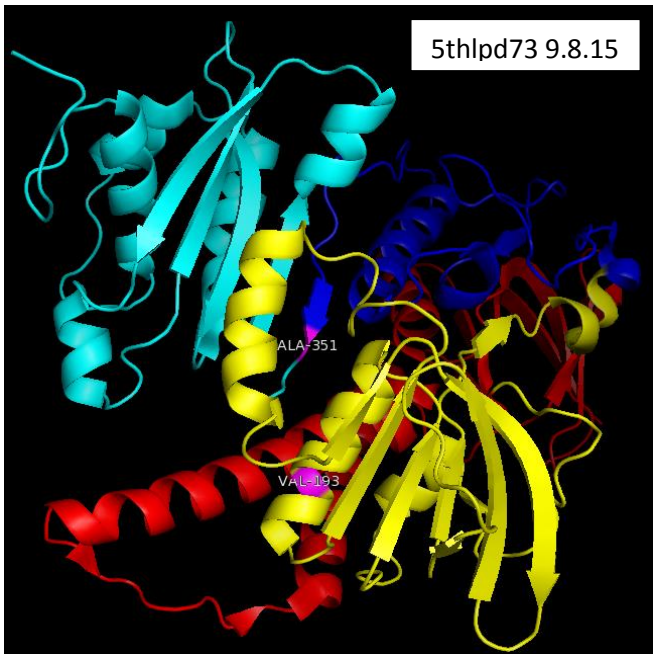


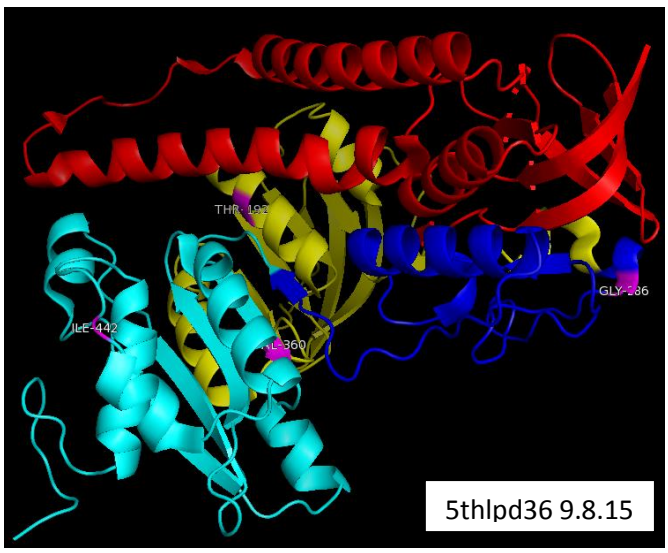
Figure 5.8: Combination of E3 primary and secondary structures showing the secondary structure elements and their relevant amino acid residues. β-strands are shown in yellow arrows, α-helices are the red zigzags, loops are represented by green arcs, and the black lines are the regions of undefined secondary structure. PDB code 4JDR.

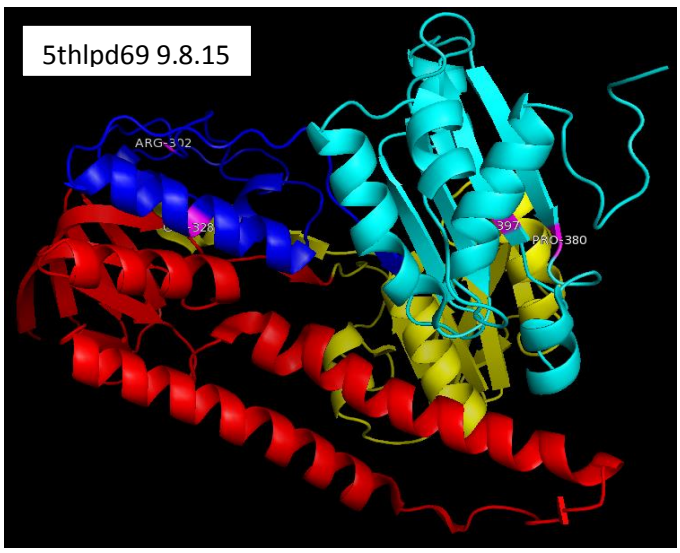

ALA-471
5thlpd31 9.8.15
ILE-409
GLY-423
GLN-300


5thlpd33 9.8.15
LYS-256


5thlpd98 9.8.15
ARG-273
VAL-357


5thlpd73 9.8.15
ALA-351
VAL-193


ILE-442
5thlpd36 9.8.15


5thlpd69 9.8.15
ARG-312

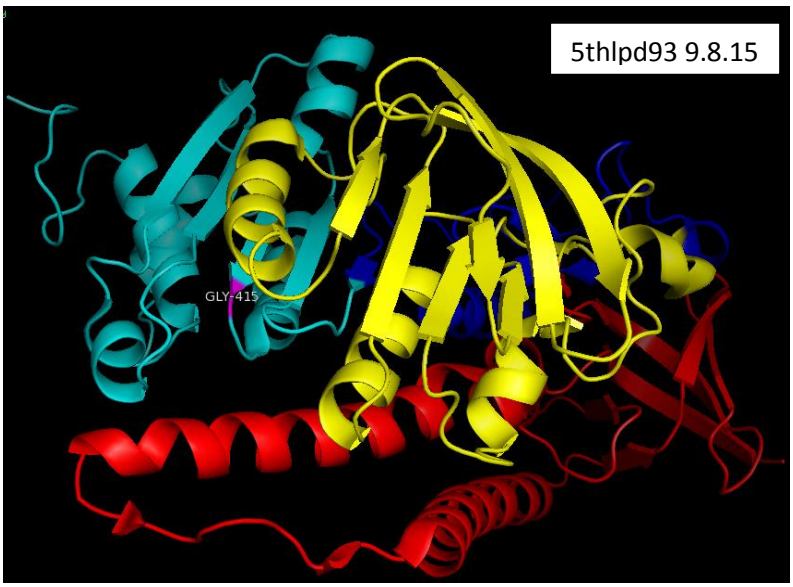


Figure 5.9: Dihydrolipoamide dehydrogenase three dimensional views highlighting (in magenta) positions of mutations within E3 mutant clones that affect the ODH and/or PDH activity. The ID of each mutant clone is written on each image. The highlighted residues are distributed equally within the two monomers of E3, but here, they are shown only within a single monomer for simplicity. Details of the residue changes and activity profiles are all included in the appendices. The four domains are distinguished by different colours: FAD-binding domain in red, NAD^+-binding domain in yellow, the central domain in blue and the interface domain in cyan. Developed from the 3D structure which was resolved by Chandrasekhar *et al.* (2013), PDB code 4JDR.

5.2.10. E3-FAD interactions and mutagenesis impact

As a homodimer, E3 molecule binds two FAD molecules which are shown with an elongated conformation with about 31.2 angstrom between the nearest points of them. The E3 FAD-binding hot spots are distributed between β-strand and α-helix of the FAD-binding domain. The latter secondary structures are joined by a characteristic sequences of small residues (GAGPAG corresponding to residue number 13-18) which are highly conserved probably because the flexibility of such hinge region is required for the FAD binding. The sequencing data showed no mutation within this hinge region. The residues that have been

found in appropriate distances for direct interactions with FAD are Gly-117, Arg-37, Glu-36, Val-44, Ala-17, Asp-313, Cys-45, Met-319, Leu-320, Ala-321, His-445, His-322, and Lys-54; six of them have been reached by the current random mutagenesis in 7 mutant clones of the following mutational patterns: *R37C*; *R37C*, G73G; *E36G*; *V44I*, A55G; K228T, Q317H, *A321S*, M396T; Q300L, *H322Q*; and M206R, R225H, A266V, V289I, *H322Y*. All these clones showed full E3 activity as each of which involves only a single mutation in one of the FAD-binding residue and this is most likely tolerated by the enzyme which is still able to bind FAD as the latter interacts with 13 residues in a manner where disrupting such interactions needs a major mutational change that causes alterations in the residue and their interacting atoms distances from the FAD ligand.

5.2.11. Subunits interaction and mutagenesis impact

E3 is only functional as a dimer so any mutation(s) that abolish the dimerization must be inhibitor for the enzyme activity, therefore, the quaternary structure is maintained by interactions between several residues from both subunits. Refer to (Chandrasekhar *et al.*, 2013) to find the list of these residue and their moieties which are involved in the two subunits interactions. The current random mutagenesis has alighted on some of these residue in 16 mutant clones (Table 5.3). Each clone has a single mutation in a residue participating in the two subunits interactions. All the mutations are not effective to disturb the E3 subunits assembly as 15 mutant clones express normal activity in both PDH and ODH complexes while one clone is of inactive ODH beside an active PDH. The inactivity of the ODH in the latter mutant clone does not mean that the interaction of the E3 subunits was affected as long the PDH complex is still active which means that E3 is still dimer to work with PDH but not with ODH. This is because of an interaction reasons that makes E3 interacts with PDH in a different way than interacting with ODH as mentioned previously. Being active

after hitting some of the inter-subunits interaction residue means that there is no hot spot or key residue among those residues covered in random mutagenesis.

Table 5.3: E3 Inter-subunit interacting residues random mutagenesis and activity data. (interacting residues of interest are in **bold***italic*).

Clone ID	Activity	Clone mutational pattern
4th lpd4 21.7.15	Active PDH and ODH	***D27N***, K216E, I248F, F288K
5th lpd59 9.8.15	=	V210A, I218T, K220N, A259E, V267I, L454M, T464P, D465I, L466C, P467R, ***N468I***, P469R, A471P, K474I + two deletions changed the ORF so that one additional amino acid (I) added to the C-terminal
1st lpd59 4.7.15	=	***H58Y***, K61E
1st lpd47 4.7.15	=	V9I, ***H72Q***, W163L
4th lpd39 21.7.15	=	R37C, ***G73G***
4th lpd55 21.7.15	=	V10I, ***G73D***, K109I, A196V, D202N, V325I
4th lpd60 21.7.15	=	K6I, ***G73D***, G286D
1st lpd3 4.7.15	=	***V75I***, V112D
4th lpd64 21.7.15	=	***R108P***
4th lpd7 21.7.15	=	***Q300L***, H322Q
5th lpd66 9.8.15	=	M206R, R225H, A266V, V289I, ***H322Y***
5th lpd9 9.8.15	=	***G419D***, G461V, P467L
5th lpd60 9.8.15	=	M206K, T353A, ***G419S***
5th lpd57 9.8.15	=	I218T, G293D, ***E420G***
5th lpd65 9.8.15	=	V204M, D291N, ***E420V***
5th lpd31 9.8.15	Active PDH, inactive ODH	Q300H, I409N, ***G423S***, A471V

However, few residues has been shown to have more than one inter-domain interaction beside the catalytic activity that they perform such as the active site residue His-445 which is found in one subunit and facing the FAD molecule situated nearby on the other counterpart subunit and makes interactions with it (Chandrasekhar *et al.*, 2013).

5.2.12. Other structural features of E3 and mutagenesis impact

The longest helix structure (Gly-43 to Ala-70) contains catalytically important intra disulfide bridge formed between the R groups sulphur atoms of Cys-45 and Cys-50. On the other hand, Met-431 and Cys-433 are in orientations which make them in a suitable conformation to form a deformed rectangle cluster of four sulphur atoms (Figure 5.10). This cluster includes two pairs of sulphur atoms from Met-431 and Cys-433 in each peptide chain (subunit).

Another sulphur cluster conformation has only been found in the lipoamide dehydrogenase of *Neisseria meningitides*, while all the E3 variants of the to date resolved structures do not show such conformation (de la Sierra *et al.*, 1997; Chandrasekhar *et al.*, 2013). The random mutagenesis has not changed Cys-45 nor Cys-50 while both Met-431 and Cys-433 have been mutated within two mutant clones of double amino acids change (I388F, ***C433S*** and I218N, ***M431V***) which are both expressing active E3 indicating that such a rectangular sulphur cluster is not important for the enzyme catalysis.

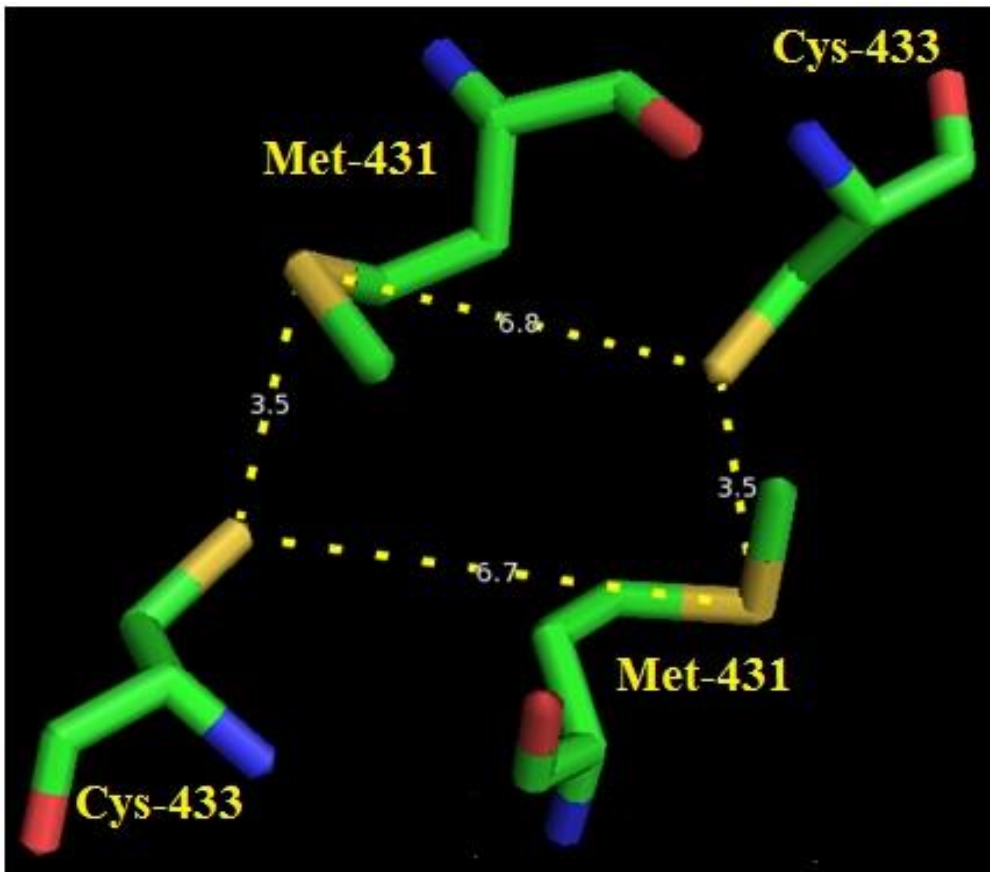


Figure 5.10: The rectangular sulphur cluster formed by the arrangement of Met-431 and Cys-433 in both *E. coli* E3 subunits. The distances between sulphur atoms are clearly indicated as 3.5 °A, 6.7 °A respectively.

Chapter 6

General discussion

Mutagenesis has been widely used in the investigation of genetics since the work of Muller and colleagues in investigating the impact of X-rays on drosophila genetics (Muller, 1928). It would take a further 50 years before the method of targeted mutagenesis, known commonly as site-directed mutagenesis would be introduced. By the end of the last century, precision approaches for site directed mutagenesis had become commonplace in the molecular biologist's experimental repertoire, while random mutagenesis using reactive chemicals such as nitrosoguanidine, hydroxyurea or controlled ionising radiation became less popular. As a result, it could be argued that progress in the understanding of gene function has become more incremental and less ground-breaking.

The development of the polymerase chain reaction (PCR) in the 1980s has undoubtedly revolutionised molecular biology research as well as providing major advances in the analysis of disease related genes, the origins of bacterial and viral infections, forensic sciences and many other areas of Biomedical Science and Biotechnology. The early emphasis was on the development of high fidelity enzymes for the efficient amplification of DNA, with Taq polymerase being replaced by enzymes such as Pfu DNA polymerase where fidelity was a priority. In fact today both types of enzymes enjoy widespread usage, with users making an informed choice based on the need for fidelity and the economics of the experiment.

While site directed mutagenesis became increasingly robust and reproducible, in particular following the introduction of the "quick change" technology by Invitrogen in the early 1990s, random methods based on PCR (in an attempt to avoid the use of hazardous reagents and radiation), were far less popular, largely owing to the effort required to optimise conditions for obtaining

reasonable yields of mutant PCR "libraries". The development of DNA sequencing technologies (and a little later, the higher throughput approaches), has also had a major impact upon the knowledge base for understanding the evolutionary relationships between genes and genomes.

Random mutagenesis is a key tool that has been used very successfully over the last decade in directed evolution experiments to provide libraries of protein variants descending form a parent protein (ancestral gene) sequence in a way that is not dissimilar to Natural Selection that underpins protein evolution. However, while there are a range of different mechanisms associated with evolution beyond single base changes, such as gene (and genome) duplication, deletion, insertion etc., random mutagenesis using error prone PCR provides a limited, but simple and efficient method to exploring the relationship between primary structure and protein function.

In contrast, site directed evolution is generally used to verify a hypothesis about the role of a small number of key residues in protein function (Wilkinson *et al.*, 1983; Fersht, 1987; Boucher *et al.*, 2016). An excellent example of the application site directed mutagenesis is provided by the work of Ptashne on the specificity determinants of bacterial repressors and by Rutter's group in the 1980s on the catalytic properties of serine proteases (Wharton and Ptashne, 1985; Craik *et al.*, 1987; Sprang *et al.*, 1987).

PCR based methods of random mutagenesis have massively compressed the time scales required by Nature to evolve protein variants. In addition, the methods applied in this work enable the analysis of "off pathway" mutants, i.e. mutants where there is a complete loss of function, many (if not most) of which are not revealed by BLAST analysis and genome sequencing projects.

The introduction of the double mutant from Connolly's laboratory (Biles and Connolly, 2004) has made it possible to derive random libraries of mutants with

limited bias, that can then be screened directly for different characteristics, depending on the reason for the experiments.

It became possible to derive a robust strategy for investigating the influence of primary structure on protein function for a “stand alone” gene as illustrated in Figure 3.9 which is reused below:

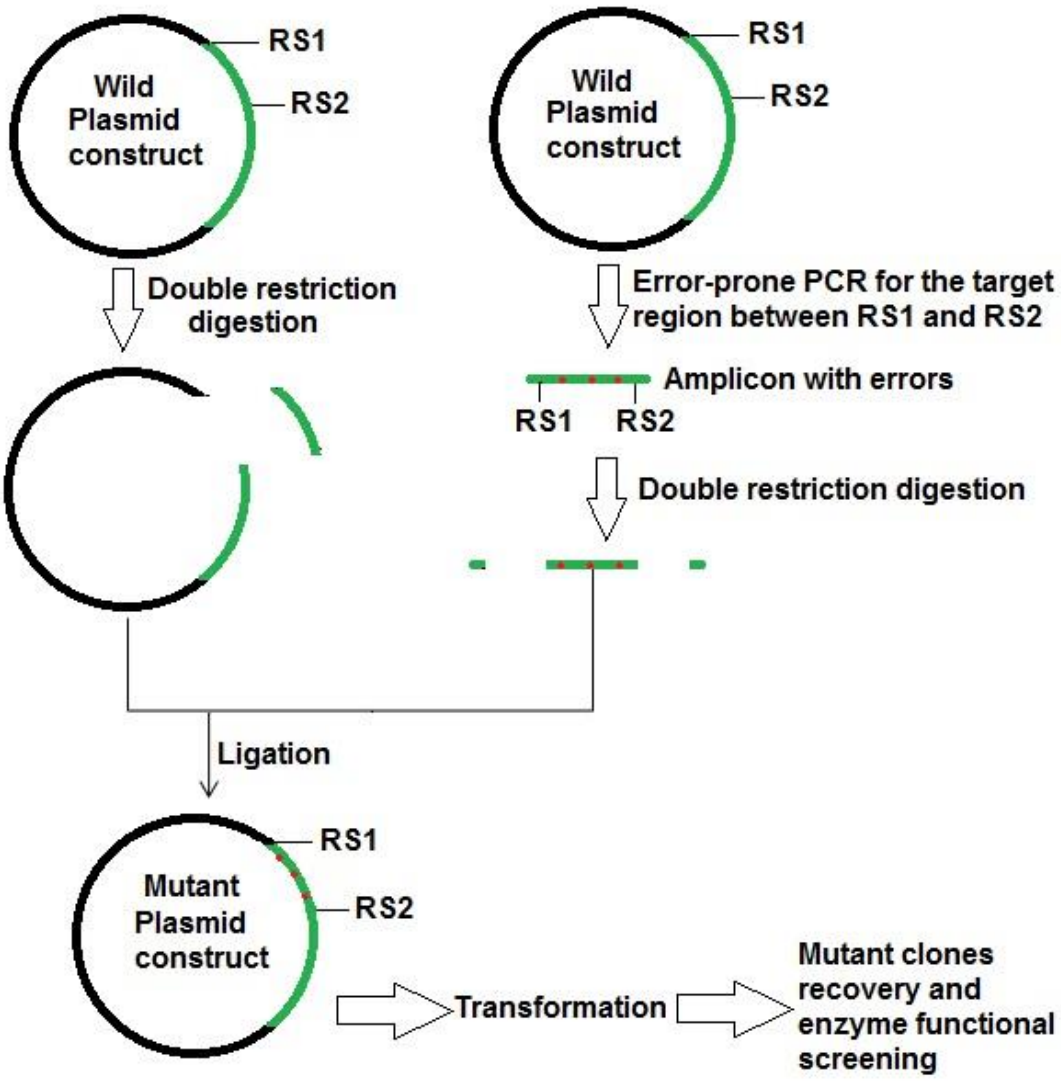


Figure 3.9: The general strategy of Error prone PCR random mutagenesis and cloning which is applied in this study on M.HhaI and the other two enzymes’ models (chapter 4 and 5). The green part represents the gene ORF. The red dots represent the mutations (errors) that are generated by EPM. RS1 and RS2 refer to restriction site 1 and 2 respectively.

In the current study, this approach has been used to stress test protein function, as a result of primary structural changes brought about by the introduction of point mutations distributed along the polypeptide chain in a random manner by error-prone PCR. This strategy will ultimately inform the rational design of proteins with potentially novel functions. This is because structural analysis of a given protein is limited in respect of dynamics that may be essential for the transitions in structure during catalysis. Moreover, high resolution structures can provide elegant explanations of enzyme mechanisms, but they possibly "over-emphasise" stereochemical precision. Moreover, having a protein crystal structure represents a single (clearly important) snap shot from the whole pathway of catalysis, but it limits understanding to these well-defined stages.

Three enzyme targets have been subjected to error-prone PCR random mutagenesis : A DNA Cytosine-5-methyl transferase, M.HhaI (Chapter 3), chloramphenicol acetyltransferase type I (CATI) (Chapter 4) and dihydrolipoamide dehydrogenase (E3) (Chapter 5), which contributes to the activity of a number of metabolic enzymes including the pyruvate dehydrogenase (PDH) and 2-oxoglutare dehydrogenase (ODH) multienzyme complexes.

The DNA sequences encoding these three enzymes have been synthesised with some modifications especially the "silent" introduction of restriction sites to aid experimental manipulation of the gene during mutagenesis. The engineered restriction sites have been designed to yield cohesive ends thus optimising downstream ligation and efficient recovery of libraries of mutant clones in conjunction with the use of high transformation efficiency *E. coli* competent cells. This strategic implementation of gene design and synthesis has proven to be of considerable value in this work, in particular when short DNA segments of the gene were amplified and cloned, but where sequencing data

showed no evidence for mutation. The presence of other restriction sites within the ORF provided a convenient level of flexibility in extending the length of the DNA to undergo mutagenic PCR amplification and cloning. In this way, the probability of increasing mutation frequency (as the S-PhoEP tends to introduce more mutations as the length of the target DNA is increased) was observed.

All of the three enzymes examples selected in this study are amenable to relatively simple screening protocols. In the case of M.HhaI and CAT (and although with some difficulties, E3), it even proved possible to distinguish low from high levels of enzyme activity (see the relevant sections).

For all three enzymes, such general activity screening does not require the enzyme to be over expressed and purified, but rather expressing a clean phenotype on agar plates. In the case of M.HhaI variants, the analysis of encoding plasmids by strategic restriction mapping adds information about intermediate levels of activity and/or enzyme stability. Hence, it was possible to explore a reasonably representative sequence diversity space within a convenient time scale. One significant observation was that functional screening indicated that different enzymes exhibit different levels of tolerance to mutational inactivation, as illustrated comparatively in Figure 6.1.

CATI tends to be the most intolerant of mutations out of the three targets. This could be because CATI is smaller in size with higher number of conserved residues which could imply that there are more “hot-spots” for inactivation, compared with the other two target enzymes. In other words, structural flexibility appears to be lower in CATI than in M.HhaI and E3.

However, the most recent random mutagenesis that compares the activity of mono-specific and multi-specific C5-methyltransferases by using the same S-PhoEP revealed that no partial activity occurs with M.HhaI (Alfageih, 2011),

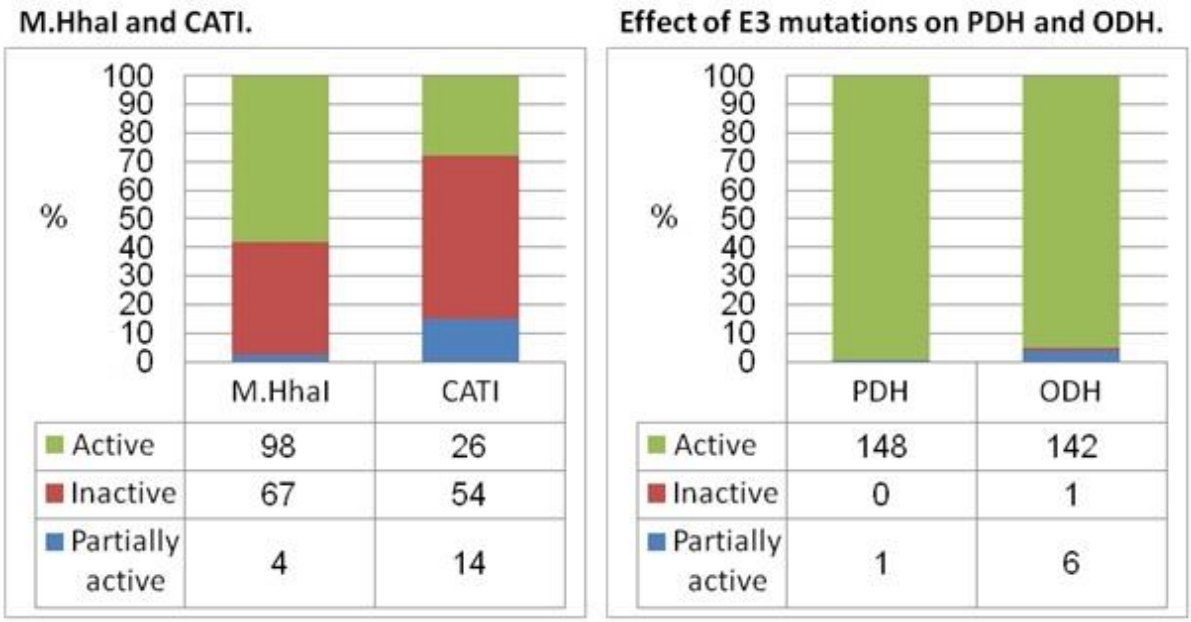


Figure 6.1: A comparative illustration for the effect of random mutagenesis on the activity of M.HhaI, CATI and E3. Active, inactive, and partially active clones are represented in numbers underneath the graphs.

While in the current study, four mutant clones have shown partial plasmid digestion by HhaI reflecting M.HhaI partial activity, (See Figure 3.10, Figure 6.1 and Appendix1). On the other hand, inactivating PDH and ODH complexes by mutating their E3 components proved more challenging in this work. This observation is similar to classic work on cytochrome c and globins from the 1970s (Stryer, 1968). This could be explained by the low chance of reaching the sensitive residues (especially the catalytic ones which are discussed in chapter 5) of E3 by random mutagenesis especially with the large size of the enzyme's monomeric subunit where such important residues are actually few in the middle of too many other residues of no or less importance.

However, alteration of key catalytic residues in E3 should abolish the enzyme's activity, and as a consequence, this would abolish the activity of all

complexes comprising E3. E3 possesses several amino acids residues that are involved in the formation of interactions network necessary to attach the enzyme to the surface of the E2 components within the complexes.

The network of interactions that form the basis of the interaction between E2 and E3 in PDH and ODH complexes has been shown to be controlled by many amino acids residues from both sides of E3 and E2 respectively (see chapter 5 for more details). It is clear from the amino acid numbering that the interface comprises widely distributed primary E3 structure elements. The second observation is that some interactions are shared by both complexes, but others are unique.

A detailed analysis of the evolution of globin genes over the last decades (Itano and Robinson, 1960; Jeffreys, 1979; Engel and Dodgson, 1980), led to a proposal for the evolutionary constraints on the development of a set of "pseudo homomeric" haemoglobin tetramers. These ideas have been largely confirmed by genomic data. Amongst the Hb variants, the variations in primary structure that have emerged through Natural Selection, to generate a diverse group of oxygen transporters with sequence in which the chains are essentially equivalent in function, but where subtle functional variants are found that can be accommodated in the tetrameric complex, without loss of function. It should be noted that whilst Hb is a sophisticated and highly regulated oxygen transporter protein, it is not an enzyme. Nevertheless, it is interesting to see how alpha polypeptides can interface successfully with beta and epsilon classes of polypeptides in such a way as to provide functional oxygen transporters that meet different cellular and physiological demands.

A quite different scenario occurs in the genome organisers, the histones. Again, this is a class of binding proteins, this time to DNA, but the four histone variants that make up the octameric core are stringently conserved across a wide

range of species. It is clear here that sequences must be preserved in order to ensure faithful packaging of the genome. The wider roles of the histones and their interaction with remodelling factors and modifying enzymes has been discussed in an evolutionary context elsewhere (Allfrey *et al.*, 1963; Littau *et al.*, 1965; Govin *et al.*, 2004; Saffarzadeh *et al.*, 2012).

Coming to the conservation of enzymes in Nature, it is clear from BLAST analysis that there are relatively few examples similar to that of the globin family (and therefore few examples of heterogeneous quaternary structures amongst enzymes exist). Most enzymes are homomeric oligomers, some are monomeric and several are heteromeric oligomers. The evolution of multi-protein complexes in relation to symmetry and structural organisation has been reviewed by Marsh and Teichmann recently (Marsh and Teichmann, 2015). The enzyme M.HhaI represents an example of a monomeric enzyme, and one that has homologues in *E. coli*, such as the product of the dcm gene and in some strains a variant of the *dcm* enzyme, EcoRII occurs (Palmer and Marinus, 1994). Extensive comparisons (summarised by the multiple sequence alignments in Appendix 4) have consistently revealed a common set of motifs and this has been reviewed in chapter 3. The data obtained here in respect of the relationship between primary structure and catalytic competence (a combination of structural integrity and catalytic activity) reveal that M.HhaI is more tolerant toward the mutagenesis active when single domain of its primary structure is targeted by mutagenesis but when the mutational range is extended to include more than one domain the activity is more reduced. This however depends on residue conservation and the number of amino acid changes (see data in chapter 3 for more details).

The enzyme CATI is a trimeric enzyme, in which activity is a function of primary structure, expressed through a combination of tertiary and quaternary structure and catalytic activity. Hence a greater set of constraints exist. On the

other hand, comparisons between CATI sequences shows more conserved points within its primary structure explaining the common inactivation phenotypes which have been observed after CATI gene mutagenesis.

In Chapter 5, the gene encoding the E3 polypeptide of the two TCA cycle enzymes' complexes (PDH and ODH) was investigated by random mutagenesis. After screening a considerable number of clones (around tenfold more than with M.HhaI or CATI) a small number of mutants were finally isolated that exhibited differential activity with the two multienzyme complexes. The experiment was designed to explore the extent of promiscuity that could be tolerated by changes to the sequence of the E3 orf.

It is clear from the results presented in Chapter 5, that the E3 sequence contains redundant elements that render the enzyme much more tolerant of changes in its primary structure. In a similar way to the C5 MTase family of enzymes, in which catalytic methyl-transfer is carried out by one module and a separate, but interactive module (the TRD) presents a cytosine from within a specific DNA sequence to the active site, E3 is able to engage its catalytic activity with several subunits from PDH, ODH and indeed other oxo-acid dehydrogenases.

The E3 experiments have enabled us to push the boundaries and extend the approach of error-prone PCR and directed evolution from testing the direct effect of mutations on the protein activity to the testing and evaluating the protein interactions. Such extended approach could be invested in the development of drugs and vaccines where the protein-protein interaction is required to be clarified and better understood especially in developing antibodies to tackle the continuously changing (mutating) viruses and bacterial antigens. Thus, we now have a promising tool to fight pathogens in the same way that they are fighting us.

Our data are consistent with the general observations in protein evolution: where there are key residues involved in catalysis or substrate recognition, mutations are likely to abolish activity. This is not surprising. However, such residues and regions are supported by regions of primary structure that vary considerably among these classes of enzymes and our results shed light on residues that are not immediately apparent for structural studies. In the absence of an understanding of how a set of conserved residues can be supported by a wide range of different residues, requires strong functional correlations with mutations to support BLAST analysis. It is clear from these experiments that we are beginning to gain insight into less transparent patterns of amino acid distributions in proteins and the constraints imposed by primary structure on the "successful" evolution of proteins.

One of the consequences of random mutation is the possibility of introducing nonsense mutations. The three canonical stop codons; Amber (TAG), Ochre (TAA), and Opal (TGA) are observed at an expected frequency in the translational reading frames of both M.HhaI and CATI genes, while no such mutants were recovered from E3 gene randomisation experiments. The activity profiles of M.HhaI and CATI mutant clones containing Ochre and Opal stop codons were always inactive, while more than 50% Amber mutants appeared with full enzyme activity (see mutations and activity data tables of M.HhaI and CATI in the appendices). These phenotypic observations emphasize the previously reported underrepresentation and weakness of the TAG stop codon compared with the higher efficiency of both TAA and TGA in terminating the poly peptide chain elongation during protein translation in bacteria (Korkmaz *et al.*, 2014). It has been shown that the difference in stop codon efficiency is closely related to the level of expression of releasing factors (RF1 and RF2), which are the key proteins responsible for ribosomal subunit dissociation and polypeptide release. Ochre is the most efficient stop codon as it can be

recognized by either releasing factors, while Amber and Opal are exclusively recognized by RF1 and RF2 respectively (Scolnick *et al.*, 1968; Brown *et al.*, 1993; Bonetti *et al.*, 1995). In *E. coli*, RF2 is five times more abundant than RF1 (Adamski *et al.*, 1994; Mora *et al.*, 2007), therefore, it is normal to expect the ribosome to pass over and read through the TAG codon occasionally. It has been demonstrated that when the Lys codon; a catalytically essential residue in firefly luciferase, is changed by mutation into TAG, the enzyme remains highly active emphasising a TAG read through, which is most likely recognized by an aminoacyl-tRNA charged with lysine (Kramer and Farabaugh, 2007).

Therefore, it seems likely that read through accounts for the residual activity of M.HhaI and CATI mutants of this type, containing in-frame Amber stop codons. In summary, the observations made here are fully consistent with the relationships between primary structure and activity in nonsense mutations analysed elsewhere.

Ever since the finding that a single amino acid substitution in haemoglobin results in the disease sickle cell anaemia (Clancy, 2008), it is clear that the network of interactions that maintain the structural and functional integrity of proteins can "be on the edge". However, the last 50 years of direct and indirect comparative analysis of primary structures of proteins has shown that single amino acid substitutions rarely abolish protein activity. In the case of M.HhaI, the substitution of the catalytic Cys at position 81 can be explained by the critical role that this side chain plays, not in stabilising the base-flipped DNA substrate, but rather inducing the necessary reactivity of the carbon atom at position 6 of cytosine, for methyl transfer (See Chapter 3). There are similar critical catalytic residues in CATI and E3 and, under appropriate growth conditions (i.e. where an absence of these activities can be compensated), such mutants are expected to arise in the random mutagenesis protocol used here. The probability of recovering such mutants is purely statistical. However, what

is clear from an evolutionary perspective, is that the vulnerability of an enzyme to inactivation by random mutagenic change, is offset by the length of most polypeptide chains. It could be argued that key residues are buffered by much of the "non-essential" nature of the majority of the primary structure of a protein. This could possibly provide a notional index for protein robustness, in which the histones would represent one extreme of such an index and the globin genes would be towards the other end of the spectrum.

It is clearly difficult to generalise in terms of evolutionary robustness in proteins from this isolated study. This is especially true in organisms where stable and transient protein interactions form the basis of such complexity (i.e. eukaryotes versus prokaryotes). Furthermore, in the spirit of Systems Biology, the degree of resilience with respect to catalytic enhancer residues or stabilising residues versus folding elements etc. which will be revealed in the differential conservation found in protein BLAST searches.

The published structural and biochemical work that has been drawn upon to rationalise the contribution of certain amino acid residues in supporting enzyme structure and function has not always been satisfactory (See Chapter 3, 4, and 5). Indeed, the role and importance of the majority of those missense mutations that abolish activity of the protein remain unclear in all three cases (Chapters 3-5). It, therefore, seems that the main outcome of this work has been to demonstrate that there are amino acids whose contribution to protein structure and function remains obscure and that the conclusions drawn from structural analysis combined with mechanistic studies and site directed mutagenesis are limited in their ability to provide a comprehensive relationship between primary structure and protein function.

On the other hand, mutations elsewhere in all three targets (M.HhaI, CATI and E3) had no observable impact on enzyme activity because they apparently

do not participate directly or indirectly in any interactions essential for catalysis, and or they play no role in structure determination or stabilisation. Furthermore, some residues that would be expected to be conserved from BLAST analysis, can be changed without effect on function, in the screens applied here. Such data reflect our limited understanding of correlating the protein structure to its function and how large or small is the "structural window" of the protein in supporting the proper folding required for the catalysis. It is also the case that the method we have employed here could be extended to identify compensatory mutations that might restore activity to a mutant where a point mutation has led to inactivation. Such patterns of synergistic mutation in which a second mutation mitigates against the deleterious effect of the first are difficult to identify by comparative sequence analysis alone (Kern and Kondrashov, 2004).

Finally, I believe that the approach which has been dealt with in this study will provide molecular enzymologists with a simple and efficient methodology to explore the hidden depths of their enzymes of interest and will throw up new ideas about enzyme mechanism. Ideally, a complementary ability to determine the structures of such mutants in real time will be developed in order to move towards more robust frameworks for predicting function from primary structure, a key element of genome annotation work.

References

Adams, R., 1990. DNA methylation. The effect of minor bases on DNA-protein interactions. Biochem. J 265, 309-320.

Adamski, F.M., McCaughan, K.K., Jorgensen, F., Kurland, C.G., Tate, W.P., 1994. The concentration of polypeptide chain release factors 1 and 2 at different growth rates of *Escherichia coli*. J Mol Biol 238, 302-308.

Alfageih, L.M., 2011. Biochemical and genetic studies of bacterial C5-DNA methyltransferases. University of Sheffield.

Alharbi, S.A., 2010. Design and analysis of DNA polymerase for the use in random mutagenesis. The University of Sheffield.

Allfrey, V., Littau, V., Mirsky, A., 1963. On the role of histones in regulating ribonucleic acid synthesis in the cell nucleus. Proceedings of the National Academy of Sciences 49, 414-421.

Allignet, J., El Solh, N., 1995. Diversity among the gram-positive acetyltransferases inactivating streptogramin A and structurally related compounds and characterization of a new staphylococcal determinant, vatB. Antimicrobial agents and chemotherapy 39, 2027-2036.

Allignet, J., Loncle, V., Simenel, C., Delepierre, M., El Solh, N., 1993. Sequence of a staphylococcal gene, vat, encoding an acetyltransferase inactivating the A-type compounds of virginiamycin-like antibiotics. Gene 130, 91-98.

Alton, N.K., Vapnek, D., 1979. Nucleotide sequence analysis of the chloramphenicol resistance transposon Tn9.

Ambrose, M.C., Perham, R.N., 1976. Spin-label study of the mobility of enzyme-bound lipoic acid in the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Biochem. J 155, 429-432.

Andreeva, A., Karamancheva, I., 2001. Insight into the secondary structure of chloramphenicol acetyltransferase type I—computer analysis and FT-IR spectroscopic characterization of the protein structure. Journal of Molecular Structure 565, 177-182.

Arnold, F.H., 2001. Combinatorial and computational challenges for biocatalyst design. Nature 409, 253-257.

Baumann, H., Knapp, S., Lundbäck, T., Ladenstein, R., Härd, T., 1994. Solution structure and DNA-binding properties of a thermostable protein from the archaeon *Sulfolobus solfataricus*. Nature Structural & Molecular Biology 1, 808-819.

Bayfield, M.A., Dahlberg, A.E., Schulmeister, U., Dorner, S., Barta, A., 2001. A conformational change in the ribosomal peptidyl transferase center upon active/inactive transition. Proceedings of the National Academy of Sciences 98, 10096-10101.

Beckman, R.A., Mildvan, A.S., Loeb, L.A., 1985. On the fidelity of DNA replication: manganese mutagenesis *in vitro*. Biochemistry 24, 5810-5817.

Berman, A.J., Kamtekar, S., Goodman, J.L., Lázaro, J.M., de Vega, M., Blanco, L., Salas, M., Steitz, T.A., 2007. Structures of phi29 DNA polymerase complexed with substrate: the mechanism of translocation in B-family polymerases. The EMBO Journal 26, 3494-3505.

Besser, D., Götz, F., Schulze-Forster, K., Wagner, H., Kröger, H., Simon, D., 1990. DNA methylation inhibits transcription by RNA polymerase III of a tRNA gene, but not of a 5S rRNA gene. FEBS letters 269, 358-362.

Biles, B.D., Connolly, B.A., 2004. Low-fidelity Pyrococcus furiosus DNA polymerase mutants useful in error-prone PCR. Nucleic Acids Research 32, e176-e176.

Biswas, T., Houghton, J.L., Garneau-Tsodikova, S., Tsodikov, O.V., 2012. The structural basis for substrate versatility of chloramphenicol acetyltransferase CATI. Protein science : a publication of the Protein Society 21, 520-530.

Black, P., Graybill, J.R., Charache, P., 1973. Penetration of brain abscess by systemically administered antibiotics. Journal of neurosurgery 38, 705-709.

Blaha, G., Gürel, G., Schroeder, S.J., Moore, P.B., Steitz, T.A., 2008. Mutations outside the anisomycin-binding site can make ribosomes drug-resistant. Journal of Molecular Biology 379, 505-519.

Blanc, H., Wright, C.T., Bibb, M.J., Wallace, D.C., Clayton, D.A., 1981. Mitochondrial DNA of chloramphenicol-resistant mouse cells contains a single nucleotide change in the region encoding the 3'end of the large ribosomal RNA. Proceedings of the National Academy of Sciences 78, 3789-3793.

Bleile, D.M., Munk, P., Oliver, R.M., Reed, L.J., 1979. Subunit structure of dihydrolipoyl transacetylase component of pyruvate dehydrogenase complex from *Escherichia coli*. Proceedings of the National Academy of Sciences 76, 4385-4389.

Bluteau, O., Legoix, P., Laurent-Puig, P., Zucman-Rossi, J., 1999. Random mutagenesis libraries: optimization and simplification by PCR. Biotechniques 27, 1102-1108.

Bonetti, B., Fu, L., Moon, J., Bedwell, D.M., 1995. The efficiency of translation termination is determined by a synergistic interplay between upstream and downstream sequences in Saccharomyces cerevisiae. Journal Molecular Biology 251, 334-345.

Boucher, J.I., Bolon, D.N., Tawfik, D.S., 2016. Quantifying and understanding the fitness effects of protein mutations: Laboratory versus nature. Protein Science 25, 1219-1226.

Boyer, H.W., 1971. DNA restriction and modification mechanisms in bacteria. Annual Reviews in Microbiology 25, 153-176.

Brautigam, C.A., Steitz, T.A., 1998. Structural and functional insights provided by crystal structures of DNA polymerases and their substrate complexes. Current opinion in structural biology 8, 54-63.

Bridges, B.A., Woodgate, R., 1985. Mutagenic repair in *Escherichia coli*: products of the recA gene and of the umuD and umuC genes act at different steps in UV-induced mutagenesis. Proceedings of the National Academy of Sciences of the United States of America 82, 4193-4197.

Brown, C.M., Dalphin, M.E., Stockwell, P.A., Tate, W.P., 1993. The translational termination signal database. Nucleic Acids Res 21, 3119-3123.

Bulkley, D., Innis, C.A., Blaha, G., Steitz, T.A., 2010. Revisiting the structures of several antibiotics bound to the bacterial ribosome. Proceedings of the National Academy of Sciences 107, 17158-17163.

Burgers, P.M., Koonin, E.V., Bruford, E., Blanco, L., Burtis, K.C., Christman, M.F., Copeland, W.C., Friedberg, E.C., Hanaoka, F., Hinkle, D.C., 2001. Eukaryotic DNA polymerases: proposal for a revised nomenclature. Journal of Biological Chemistry 276, 43487-43490.

Burns, G., Sykes, P.J., Hatter, K., Sokatch, J.R., 1989. Isolation of a third lipoamide dehydrogenase from *Pseudomonas putida*. Journal of bacteriology 171, 665-668.

Cadwell, R.C., Joyce, G.F., 1992. Randomization of genes by PCR mutagenesis. PCR methods and applications 2, 28-33.

Cadwell, R.C., Joyce, G.F., 1994. Mutagenic PCR. Genome Research 3, S136-S140.

Cannon, M., Harford, S., Davies, J., 1990. A comparative study on the inhibitory actions of chloramphenicol, thiamphenicol and some fluorinated derivatives. Journal of Antimicrobial Chemotherapy 26, 307-317.

Capson, T.L., Peliska, J.A., Kaboord, B.F., Frey, M.W., Lively, C., Dahlberg, M., Benkovic, S.J., 1992. Kinetic characterization of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage T4. Biochemistry 31, 10984-10994.

Cedar, H., 1988. DNA methylation and gene activity. Cell 53, 3-4.

Celma, M., Monro, R., Vazquez, D., 1971. Substrate and antibiotic binding sites at the peptidyl transferase centre of *E. coli* ribosomes: Binding of UACCA-Leu to 50 S subunits. FEBS letters 13, 247-251.

Chandrasekhar, K., Wang, J., Arjunan, P., Sax, M., Park, Y.-H., Nemeria, N.S., Kumaran, S., Song, J., Jordan, F., Furey, W., 2013. Insight to the interaction of the dihydrolipoamide acetyltransferase (E2) core with the peripheral components in the *Escherichia coli* pyruvate dehydrogenase complex via multifaceted structural approaches. Journal of Biological Chemistry 288, 15402-15417.

Charles, I., Keyte, J., Shaw, W., 1985. Nucleotide sequence analysis of the cat gene of *Proteus mirabilis*: comparison with the type I (Tn9) cat gene. Journal of bacteriology 164, 123-129.

Chen, K., Arnold, F.H., 1993. Tuning the activity of an enzyme for unusual environments: sequential random mutagenesis of subtilisin E for catalysis in dimethylformamide. Proceedings of the National Academy of Sciences of the United States of America 90, 5618-5622.

Cheng, X., Blumenthal, R.M., 1996. Finding a basis for flipping bases. Structure 4, 639-645.

Cheng, X., Kumar, S., Posfai, J., Pflugrath, J.W., Roberts, R.J., 1993. Crystal structure of the Hhal DNA methyltransferase complexed with S-adenosyl-L-methionine. Cell 74, 299-307.

Clancy, S., 2008. Genetic mutation. Nature Education 1, 187.
Cline, J., Braman, J.C., Hogrefe, H.H., 1996. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res 24, 3546-3551.

Cline, J., Hogrefe, H., 2000. Randomize gene sequences with new PCR mutagenesis kit. Strategies 13, 157-161.

Coelho, P.S., Brustad, E.M., Kannan, A., Arnold, F.H., 2013. Olefin cyclopropanation via carbene transfer catalyzed by engineered cytochrome P450 enzymes. Science 339, 307-310.

Cox, E.C., 1976. Bacterial mutator genes and the control of spontaneous mutation. Annu Rev Genet 10, 135-156.

Craik, C.S., Roczniak, S., Largman, C., Rutter, W.J., 1987. The catalytic role of the active site aspartic acid in serine proteases. Science 237, 909-913.

Cundliffe, E., 1984. Self defence in antibiotic-producing organisms. British Medical Bulletin 40, 61-67.

Dahiyat, B.I., Mayo, S.L., 1997. De novo protein design: fully automated sequence selection. Science 278, 82-87.

Danson, M.J., Hale, G., Johnson, P., Perham, R.N., Smith, J., Spragg, P., 1979. Molecular weight and symmetry of the pyruvate dehydrogenase multienzyme complex of *Escherichia coli*. Journal of molecular biology 129, 603-617.

Darwin, C., 1872. The origin of species. Lulu. com.

Davidovich, C., Bashan, A., Auerbach-Nevo, T., Yaggie, R.D., Gontarek, R.R., Yonath, A., 2007.

Induced-fit tightens pleuromutilins binding to ribosomes and remote interactions enable their selectivity. Proceedings of the National Academy of Sciences 104, 4291-4296.

Day, P.J., Murray, I.A., Shaw, W.V., 1995. Properties of hybrid active sites in oligomeric proteins: kinetic and ligand binding studies with chloramphenicol acetyltransferase trimers. Biochemistry 34, 6416-6422.

de la Sierra, I.L., Pernot, L., Prangé, T., Saludjian, P., Schiltz, M., Fourme, R., Padrón, G., 1997. Molecular structure of the lipoamide dehydrogenase domain of a surface antigen from *Neisseria meningitidis*. Journal of Molecular Biology 269, 129-141.

de Vega, M., Lázaro, J.M., Mencía, M., Blanco, L., Salas, M., 2010. Improvement of φ29 DNA polymerase amplification performance by fusion of DNA binding motifs. Proceedings of the National Academy of Sciences 107, 16506-16511.

DelVecchio, V.G., Kapatral, V., Redkar, R.J., Patra, G., Mujer, C., Los, T., Ivanova, N., Anderson, I., Bhattacharyya, A., Lykidis, A., 2002. The genome sequence of the facultative intracellular pathogen *Brucella melitensis*. Proceedings of the National Academy of Sciences 99, 443-448.

Dierick, H., Stul, M., De Kelver, W., Marynen, P., Cassiman, J.-J., 1993. Incorporation of dITP or 7-deaza dGTP during PCR improves sequencing of the product. Nucleic Acids Research 21, 4427- 4428.

Doublie, S., Sawaya, M.R., Ellenberger, T., 1999. An open and closed case for all polymerases. Structure 7, R31-35.

Dryden, D.T., 1999. Bacterial DNA methyltransferases. S-adenosylmethionine-dependent methyltransferases: structures and functions. World Scientific Publishing, Singapore, 283-340.

Dunkle, J.A., Xiong, L., Mankin, A.S., Cate, J.H., 2010. Structures of the *Escherichia coli* ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. Proceedings of the National Academy of Sciences 107, 17152-17157.

Dunkle, L.M., 1978. Central nervous system chloramphenicol concentration in premature infants. Antimicrobial Agents and Chemotherapy 13, 427-429.

Eckert, K.A., Kunkel, T.A., 1991. DNA polymerase fidelity and the polymerase chain reaction. Genome Research 1, 17-24.

Ehrlich, J., Bartz, Q.R., Smith, R.M., Joslyn, D.A., Burkholder, P.R., 1947. Chloromycetin, a new antibiotic from a soil actinomycete. Science (New York, NY) 106, 417-417.

Elisha, B.G., Steyn, L., 1991. Identification of an *Acinetobacter baumannii* gene region with sequence and organizational similarity to Tn2670. Plasmid 25, 96-104.

Engel, J.D., Dodgson, J.B., 1980. Analysis of the closely linked adult chicken alpha-globin genes in recombinant DNAs. Proceedings of the National Academy of Sciences 77, 2596-2600.

Erlich, H.A., 1989. PCR technology. Springer.

Estabrook, R.A., Lipson, R., Hopkins, B., Reich, N., 2004. The coupling of tight DNA binding and base flipping: identification of a conserved structural motif in base flipping enzymes. The Journal of biological chemistry 279, 31419-31428.

Evans, S.J., Fogg, M.J., Mamone, A., Davis, M., Pearl, L.H., Connolly, B.A., 2000. Improving dideoxynucleotide-triphosphate utilisation by the hyper-thermophilic DNA polymerase from the archaeon *Pyrococcus furiosus*. Nucleic Acids Res 28, 1059-1066.

Faure, M., Bourguignon, J., Neuburger, M., Macherel, D., Sieker, L., Ober, R., Kahn, R., Cohen-Addad, C., Douce, R., 2000. Interaction between the lipoamide-containing H-protein and the lipoamide dehydrogenase (L-protein) of the glycine decarboxylase multienzyme system. European Journal of Biochemistry 267, 2890-2898.

Feldman, W.E., Manning, N.S., 1983. Effect of growth phase on the bactericidal action of chloramphenicol against *Haemophilus influenzae* type b and *Escherichia coli* K-1. Antimicrobial Agents and Chemotherapy 23, 551-554.

Fersht, A.R., 1987. Dissection of the structure and activity of the tyrosyl-tRNA synthetase by site-directed mutagenesis. Biochemistry 26, 8031-8037.

Franklin, T.J., Snow, G.A., 2005. Biochemistry and molecular biology of antimicrobial drug action. Springer Science & Business Media.

Freemont, P., Friedman, J., Beese, L.S., Sanderson, M., Steitz, T., 1988. Cocrystal structure of an editing complex of Klenow fragment with DNA. Proceedings of the National Academy of Sciences 85, 8924-8928.

Freese, E., 1959. The specific mutagenic effect of base analogues on phage T4. Journal of Molecular Biology 1, 87-105.

Friedman, C.A., Lovejoy, F.C., Smith, A.L., 1979. Chloramphenicol disposition in infants and children. The Journal of Pediatrics 95, 1071-1077.

Govin, J., Caron, C., Lestrat, C., Rousseaux, S., Khochbin, S., 2004. The role of histones in chromatin remodelling during mammalian spermiogenesis. European Journal of Biochemistry 271, 3459-3469.

Goyon, C., 1997. Isolation and identification by sequence homology of a second putative C5-DNA-methyltransferase gene from *Ascobolus immersus*. DNA sequence: the journal of DNA sequencing and mapping 9, 109-112.

Grande, H.J., Telgen, H.J., Veeger, C., 1976. Symmetry and asymmetry of the pyruvate dehydrogenase complexes from *Azotobacter vinelandii* and *Escherichia coli* as reflected by fluorescence and spin-label studies. European Journal of Biochemistry 71, 509-518.

Greener, A., Callahan, M., Jerpseth, B., 1997a. An efficient random mutagenesis technique using an E. coli mutator strain. Mol Biotechnol 7, 189-195.

Greener, A., Callahan, M., Jerpseth, B., 1997b. An efficient random mutagenesis technique using an *E. coli* mutator strain. Molecular Biotechnology 7, 189-195.

Gürel, G., Blaha, G., Moore, P.B., Steitz, T.A., 2009. U2504 determines the species specificity of the A-site cleft antibiotics: the structures of tiamulin, homoharringtonine, and bruceantin bound to the ribosome. Journal of Molecular Biology 389, 146-156.

Hahn, F., Hayes, J., HOPPS, H., SMADEL, J., WISSEMAN Jr, C., 1956. Mode of action of chloramphenicol. VI. Relation between structure and activity in the chloramphenicol series. Antibiotics & Chemotherapy 6, 531.

Hammes, G., 1980. Processing of intermediates in multienzyme complexes. Biochemical Society Symposium, pp. 73-90.

Hanahan, D., 1983. Studies on transformation of *Escherichia coli* with plasmids. Journal of Molecular Biology 166, 557-580.

Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., Kai, Y., 2001a. Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. Journal of Molecular Biology 306, 469-477.

Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., Kai, Y., 2001b. Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1. Journal of Molecular Biology 306, 469-477.

Heithoff, D.M., Sinsheimer, R.L., Low, D.A., Mahan, M.J., 1999. An essential role for DNA adenine methylation in bacterial virulence. Science 284, 967-970.

Henderson, C.E., Perham, R.N., Finch, J.T., 1979. Structure and symmetry of B. stearothermophilus pyruvate dehydrogenase multienzyme complex and implications for eucaryote evolution. Cell 17, 85-93.

Hopfner, K.-P., Eichinger, A., Engh, R.A., Laue, F., Ankenbauer, W., Huber, R., Angerer, B., 1999. Crystal structure of a thermostable type B DNA polymerase from Thermococcus gorgonarius. Proceedings of the National Academy of Sciences 96, 3600-3605.

Horton, J.R., Ratner, G., Banavali, N.K., Huang, N., Choi, Y., Maier, M.A., Marquez, V.E., MacKerell, A.D., Jr., Cheng, X., 2004. Caught in the act: visualization of an intermediate in the DNA base-flipping pathway induced by HhaI methyltransferase. Nucleic Acids Research 32, 3877-3886.

Huang, N., Banavali, N.K., MacKerell, A.D., Jr., 2003. Protein-facilitated base flipping in DNA by cytosine-5-methyltransferase. Proceedings of the National Academy of Sciences of the United States of America 100, 68-73.

Innis, M.A., Myambo, K.B., Gelfand, D.H., Brow, M., 1988. DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. Proceedings of the National Academy of Sciences 85, 9436-9440.

Iqbal, K., Jin, S.-G., Pfeifer, G.P., Szabó, P.E., 2011. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proceedings of the National Academy of Sciences 108, 3642-3647.

Itano, H.A., Robinson, E.A., 1960. Genetic control of the α-and β-chains of hemoglobin. Proceedings of the National Academy of Sciences 46, 1492-1501.

Ivanova, N., Sorokin, A., Anderson, I., Galleron, N., Candelon, B., Kapatral, V., Bhattacharyya, A., Reznik, G., Mikhailova, N., Lapidus, A., 2003. Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. Nature 423, 87-91.

Jeffreys, A.J., 1979. DNA sequence variants in the Gγ-, Aγ-, δ-and β-globin genes of man. Cell 18, 1-10.

Jeltsch, A., 2006. Molecular enzymology of mammalian DNA methyltransferases. DNA Methylation: Basic Mechanisms. Springer, pp. 203-225.

Jespers, L.S., Roberts, A., Mahler, S.M., Winter, G., Hoogenboom, H.R., 1994. Guiding the selection of human antibodies from phage display repertoires to a single epitope of an antigen. Bio/technology (Nature Publishing Company) 12, 899-903.

John, R.G., Ahmed, M.A.-H., Graham, A.A., Louise, C., Robin, A.H., Rosane, S.M., Margaret, M.A., 2003. Physiological effects of replacing the PDH complex of *E. coli* by genetically engineered variants or by pyruvate oxidase. Thiamine. CRC Press.

Joyce, C.M., Steitz, T.A., 1994. Function and structure relationships in DNA polymerases. Annual Review of Biochemistry 63, 777-822.

Kaneko, T., Nakamura, Y., Sato, S., Asamizu, E., Kato, T., Sasamoto, S., Watanabe, A., Idesawa, K., Ishikawa, A., Kawashima, K., 2000. Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. DNA Research 7, 331-338.

Kawase, Y., Iwai, S., Inoue, H., Miura, K., Ohtsuka, E., 1986. Studies on nucleic acid interactions I. Stabilities of mini-duplexes (dG2A4XA4G2. dC2T4YT4C2) and self-complementary d (GGGAAXYTTCCC) containing deoxyinosine and other mismatched bases. Nucleic Acids Research 14, 7727-7736.

Kearsey, S.E., Craig, I.W., 1981. Altered ribosomal RNA genes in mitochondria from mammalian cells with chloramphenicol resistance. Nature 290, 607-608.

Kehrenberg, C., Schwarz, S., 2001. Occurrence and linkage of genes coding for resistance to sulfonamides, streptomycin and chloramphenicol in bacteria of the genera Pasteurella and Mannheimia. FEMS Microbiology Letters 205, 283-290.

Kehrenberg, C., Schwarz, S., 2002. Nucleotide sequence and organization of plasmid pMVSCS1 from *Mannheimia varigena*: identification of a multiresistance gene cluster. Journal of Antimicrobial Chemotherapy 49, 383-386.

Kelman, Z., Hurwitz, J., O'Donnell, M., 1998. Processivity of DNA polymerases: two mechanisms, one goal. Structure 6, 121-125.

Kern, A.D., Kondrashov, F.A., 2004. Mechanisms and convergence of compensatory evolution in mammalian mitochondrial tRNAs. Nature Genetics 36, 1207-1212.

Kim, E.h., Aoki, T., 1993. The structure of the chloramphenicol resistance gene on a transferable R plasmid from the fish pathogen, *Pasteurella piscicida*. Microbiology and Immunology 37, 705-712.

Kim, S.W., Kim, D.-U., Kim, J.K., Kang, L.-W., Cho, H.-S., 2008. Crystal structure of Pfu, the high fidelity DNA polymerase from *Pyrococcus furiosus*. International Journal of Biological Macromolecules 42, 356-361.

Kleanthous, C., Shaw, W., 1984. Analysis of the mechanism of chloramphenicol acetyltransferase by steady-state kinetics. Evidence for a ternary-complex mechanism. Biochemical Journal 223, 211-220.

Klein, D.J., Moore, P.B., Steitz, T.A., 2004. The contribution of metal ions to the structural stability of the large ribosomal subunit. Rna 10, 1366-1379.

Klimasauskas, S., Kumar, S., Roberts, R.J., Cheng, X., 1994. Hhal methyltransferase flips its target base out of the DNA helix. Cell 76, 357-369.

Klimašauskas, S., Szyperski, T., Serva, S., Wüthrich, K., 1998. Dynamic modes of the flipped-out cytosine during HhaI methyltransferase–DNA interactions in solution. The EMBO Journal 17, 317-324.

Klimašuaskas, S., Nelson, J.L., Roberts, R.J., 1991. The sequence specificity domain of cytosine-C5 methylases. Nucleic Acids Research 19, 6183-6190.

Korkmaz, G., Holm, M., Wiens, T., Sanyal, S., 2014. Comprehensive analysis of stop codon usage in bacteria and its correlation with release factor abundance. Journal of Biological Chemistry 289, 30334-30342.

Kramer, E.B., Farabaugh, P.J., 2007. The frequency of translational misreading errors in *E. coli* is largely determined by tRNA competition. RNA 13, 87-96.

Kramer, P.W., Griffith, R.S., Campbell, R.L., 1969. Antibiotic penetration of the brain. A comparative study. Journal of Neurosurgery 31, 295-302.

Kumar, S., Cheng, X., Klimasauskas, S., Mi, S., Posfai, J., Roberts, R.J., Wilson, G.G., 1994. The DNA (cytosine-5) methyltransferases. Nucleic Acids Research 22, 1-10.

Kunkel, T.A., 2004. DNA replication fidelity. The Journal of biological chemistry 279, 16895-16898.
Kunkel, T.A., Bebenek, K., 2000. DNA replication fidelity. Annual Review of Biochemistry 69, 497-529.

Kuroita, T., Matsumura, H., Yokota, N., Kitabayashi, M., Hashimoto, H., Inoue, T., Imanaka, T., Kai, Y., 2005a. Structural mechanism for coordination of proofreading and polymerase activities in archaeal DNA polymerases. Journal of Molecular Biology 351, 291-298.

Kuroita, T., Matsumura, H., Yokota, N., Kitabayashi, M., Hashimoto, H., Inoue, T., Imanaka, T., Kai, Y., 2005b. Structural mechanism for coordination of proofreading and polymerase activities in archaeal DNA polymerases. Journal of Molecular Biology 351, 291-298.

Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680-685.

Lai, Y.P., Huang, J., Wang, L.F., Li, J., Wu, Z.R., 2004. A new approach to random mutagenesis in vitro. Biotechnology and bioengineering 86, 622-627.

Laraki, N., Galleni, M., Thamm, I., Riccio, M.L., Amicosante, G., Frère, J.-M., Rossolini, G.M., 1999. Structure of In31, abla IMP-containing Pseudomonas aeruginosa integron phyletically related to In5, which carries an unusual array of gene cassettes. Antimicrobial Agents and Chemotherapy 43, 890-901.

Lau, E.Y., Bruice, T.C., 1999. Active site dynamics of the< i> Hha</i> I methyltransferase: insights from computer simulation. Journal of molecular biology 293, 9-18.

Lauster, R., Trautner, T.A., Noyer-Weidner, M., 1989. Cytosine-specific type II DNA methyltransferases: a conserved enzyme core with variable target-recognizing domains. Journal of Molecular Biology 206, 305-312.

Lehninger, A.L., 1975. Biochemistry: the molecular basis of cell structure and functions. Worth, New York, 659.

Leslie, A., 1990. Refined crystal structure of type III chloramphenicol acetyltransferase at 1·75 Å resolution. Journal of Molecular Biology 213, 167-186.

Leslie, A., Moody, P., Shaw, W.V., 1988. Structure of chloramphenicol acetyltransferase at 1.75-A resolution. Proceedings of the National Academy of Sciences 85, 4133-4137.

Leung, D.W., Chen, E., Goeddel, D.V., 1989. A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique 1, 11-15.

Lewendon, A., Murray, I.A., Shaw, W.V., Gibbs, M.R., Leslie, A.G.W., 1990. Evidence for transition-state stabilization by serine-148 in the catalytic mechanism of chloramphenicol acetyltransferase. Biochemistry 29, 2075-2080.

Li, Y., Korolev, S., 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. The EMBO Journal 17, 7514-7525.

Lietman, P., 1979. Chloramphenicol and the neonate--1979 view. Clinics in Perinatology 6, 151-162.

Lin-Goerke, J.L., Robbins, D.J., Burczak, J.D., 1997. PCR-based random mutagenesis using manganese and reduced dNTP concentration. Biotechniques 23, 409-412.

Lindstrom, W.M., 2000. Reconciling Structure and Function in HhaI DNA Cytosine-C-5 Methyltransferase. Journal of Biological Chemistry 275, 4912-4919.

Ling, L.L., Keohavong, P., Dias, C., Thilly, W.G., 1991. Optimization of the polymerase chain reaction with regard to fidelity: modified T7, Taq, and vent DNA polymerases. PCR Methods and Applications 1, 63-69.

Littau, V., Burdick, C., Allfrey, V., Mirsky, S., 1965. The role of histones in the maintenance of chromatin structure. Proceedings of the National Academy of Sciences 54, 1204-1212.

Loenen, W.A., Daniel, A.S., Braymer, H.D., Murray, N.E., 1987. Organization and sequence of the *hsd* genes of *Escherichia coli* K-12. Journal of Molecular Biology 198, 159-170.

Lovett, P.S., 1990. Translational attenuation as the regulator of inducible cat genes. Journal of Bacteriology 172, 1-6.

Luck, S.N., Turner, S.A., Rajakumar, K., Sakellaris, H., Adler, B., 2001. Ferric dicitrate transport system (Fec) of *Shigella flexneri* 2a YSH6000 is encoded on a novel pathogenicity island carrying multiple antibiotic resistance genes. Infection and Immunity 69, 6012-6021.

Lundberg, K.S., Shoemaker, D.D., Adams, M.W., Short, J.M., Sorge, J.A., Mathur, E.J., 1991. High-fidelity amplification using a thermostable DNA polymerase isolated from *Pyrococcus furiosus*. Gene 108, 1-6.

Luo, J., Bruice, T.C., 2005. Low-frequency normal mode in DNA HhaI methyltransferase and motions of residues involved in the base flipping. Proceedings of the National Academy of Sciences of the United States of America 102, 16194-16198.

Malagnac, F., Wendel, B., Goyon, C., Faugeron, G., Zickler, D., Rossignol, J.-L., Noyer-Weidner, M., Vollmayr, P., Trautner, T.A., Walter, J., 1997. A gene essential for de novo methylation and development in Ascobolus reveals a novel type of eukaryotic DNA methyltransferase structure. Cell 91, 281-290.

Mande, S.S., Sarfaty, S., Allen, M.D., Perham, R.N., Hol, W.G., 1996. Protein–protein interactions in the pyruvate dehydrogenase multienzyme complex: dihydrolipoamide dehydrogenase complexed with the binding domain of dihydrolipoamide acetyltransferase. Structure 4, 277-286.

Maniatis, T., Fritsch, E.F., Sambrook, J., 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Cold Spring Harbor, NY.

Mankin, A.S., Garrett, R.A., 1991. Chloramphenicol resistance mutations in the single 23S rRNA gene of the archaeon *Halobacterium halobium*. Journal of Bacteriology 173, 3559-3563.

Mann, M., Smith, H., 1979. Specificity of DNA methylases from *Haemophilus sp*. Proceedings of the Conference on Transmethylation, pp. 438-492.

Marsh, J.A., Teichmann, S.A., 2015. Structure, dynamics, assembly, and evolution of protein complexes. Annual Review of Biochemistry 84, 551-575.

Martelo, O.J., Manyan, D.R., Smith, U.S., Yunis, A.A., 1969. Chloramphenicol and bone marrow mitochondria. Journal of Laboratory and Clinical Medicine 74, 927-940.

Mascaretti, O.A., 2003. Bacteria versus Antimicrobial Agents, an Integrated Approach. ASM Press.

Mattevi, A., Obmolova, G., Kalk, K.H., Teplyakov, A., Hol, W.G., 1993. Crystallographic analysis of substrate binding and catalysis in dihydrolipoyl transacetylase (E2p). Biochemistry 32, 3887-3901.

Mattevi, A., Obmolova, G., Schulze, E., Kalk, K.H., Westphal, A.H., De Kok, A., Hol, W., 1992. Atomic structure of the cubic core of the pyruvate dehydrogenase multienzyme complex. Science 255, 1544-1550.

Maxwell, R., Nickel, V., 1954. The antibacterial activity of the isomers of chloramphenicol. Antibiotics & Chemotherapy 4, 289-295.

McAfee, J.G., Edmondson, S.P., Datta, P.K., Shriver, J.W., Gupta, R., 1995. Gene cloning, expression, and characterization of the Sac7 proteins from the hyperthermophile *Sulfolobus acidocaldarius*. Biochemistry 34, 10063-10077.

McCabe, P., 1990. A Guide to Methods and Applications in PCR Protocols. New York: Academic Press.

McCarthy, J.K., Uzelac, A., Davis, D.F., Eveleigh, D.E., 2004. Improved catalytic efficiency and active site modification of 1,4-beta-D-glucan glucohydrolase A from *Thermotoga neapolitana* by directed evolution. The Journal of Biological Chemistry 279, 11495-11502.

McIsaac, R.S., Engqvist, M.K., Wannier, T., Rosenthal, A.Z., Herwig, L., Flytzanis, N.C., Imasheva, E.S., Lanyi, J.K., Balashov, S.P., Gradinaru, V., Arnold, F.H., 2014. Directed evolution of a far-red fluorescent rhodopsin. Proceedings of the National Academy of Sciences of the United States of America 111, 13034-13039.

Mi, S., Alonso, D., Roberts, R.J., 1995. Functional analysis of Gln-237 mutants of Hhal methyltransferase. Nucleic Acids Research 23, 620-627.

Mi, S., Roberts, R.J., 1992. How M.MspI and M.HpaII decide which base to methylate. Nucleic Acids Research 20, 4811-4816.

Mi, S., Roberts, R.J., 1993. The DNA binding affinity of Hhal methylase is increased by a single amino acid substitution in the catalytic center. Nucleic acids research 21, 2459-2464.

Modrich, P., 1991. Mechanisms and biological effects of mismatch repair. Annual Review of Genetics 25, 229-253.

Mora, L., Heurgue-Hamard, V., de Zamaroczy, M., Kervestin, S., Buckingham, R.H., 2007. Methylation of bacterial release factors RF1 and RF2 is required for normal translation termination in vivo. J Biol The Journal of Biological Chemistry 282, 35638-35645.

Morii, H., Hayashi, N., Uramoto, K., 2003. Cloning and nucleotide sequence analysis of the chloramphenicol resistance gene on conjugative R plasmids from the fish pathogen. Diseases of aquatic organisms 53, 107-113.

Motz, M., Pääbo, S., Kilger, C., 2000. Improved cycle sequencing of GC-rich templates by a combination of nucleotide analogs. Biotechniques 29, 268-270.

Muller, H.J., 1928. The production of mutations by X-rays. Proceedings of the National Academy of Sciences 14, 714-726.

Mullis, K.B., Faloona, F.A., 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology 155, 335-350.

Murray, I., Hawkins, A., Keyte, J., Shaw, W., 1988. Nucleotide sequence analysis and overexpression of the gene encoding a type III chloramphenicol acetyltransferase. Biochemical Journal 252, 173-179.

Murray, I., Martinez-Suarez, J., Close, T., Shaw, W., 1990. Nucleotide sequences of genes encoding the type II chloramphenicol acetyltransferases of *Escherichia coli* and *Haemophilus influenzae*, which are sensitive to inhibition by thiol-reactive reagents. Biochemical Journal 272, 505-510.

Murray, I.A., Shaw, W.V., 1997. O-Acetyltransferases for chloramphenicol and other natural products. Antimicrobial Agents and Chemotherapy 41, 1-6.

Myers, R.M., Lerman, L.S., Maniatis, T., 1985. A general method for saturation mutagenesis of cloned DNA fragments. Science 229, 242-247.

Nakai, T., Kuramitsu, S., Kamiya, N., 2008. Structural bases for the specific interactions between the E2 and E3 components of the *Thermus thermophilus* 2-oxo acid dehydrogenase complexes. Journal of Biochemistry 143, 747-758.

Newton, C., Graham, A., 1997. PCR (Introduction to Biotechniques Series). BIOS Scientific Publishers.

Neylon, C., 2004. Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution. Nucleic Acids Research 32, 1448-1459.

Nurse, P., Thuriaux, P., 1980. Regulatory genes controlling mitosis in the fission yeast *Schizosaccharomyces pombe*. Genetics 96, 627-637.

O'Gara, M., Horton, J.R., Roberts, R.J., Cheng, X., 1998. Structures of HhaI methyltransferase complexed with substrates containing mismatches at the target base. Nature Structural & Molecular Biology 5, 872-877.

Oliver, R., Reed, L., 1982. In Electron Microscopy of Proteins, Harris JR (ed), Vol. 2. London: Academic, 1-48.

Overturf, G.D., Wilkins, J., Leedom, J.M., Ivler, D., Mathies, A.W., 1975. Susceptibility of *Hemophilus influenzae*, type b, to ampicillin at Los Angeles County/University of Southern California Medical Center: A reappraisal after ten years. The Journal of Pediatrics 87, 297-300.

Packer, M.S., Liu, D.R., 2015. Methods for the directed evolution of proteins. Nature Reviews Genetics 16, 379-394.

Packman, L.C., Perham, R.N., 1986. Chain folding in the dihydrolipoyl acyltransferase components of the 2-oxo-acid dehydrogenase complexes from *Escherichia coli*: Identification of a segment involved in binding the E3 subunit. FEBS letters 206, 193-198.

Palmer, B.R., Marinus, M.G., 1994. The *dam* and *dcm* strains of *Escherichia coli*--a review. Gene 143, 1-12.

Parent, R., Roy, P.H., 1992. The chloramphenicol acetyltransferase gene of Tn2424: a new breed of cat. Journal of Bacteriology 174, 2891-2897.

Park, Y.-H., Wei, W., Zhou, L., Nemeria, N., Jordan, F., 2004. Amino-terminal residues 1-45 of the *Escherichia coli* pyruvate dehydrogenase complex E1 subunit interact with the E2 subunit and are required for activity of the complex but not for reductive acetylation of the E2 subunit. Biochemistry 43, 14037-14046.

Parkhill, J., Dougan, G., James, K., Thomson, N., Pickard, D., Wain, J., Churcher, C., Mungall, K., Bentley, S., Holden, M., 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature 413, 848-852.

Patel, M.S., Nemeria, N.S., Furey, W., Jordan, F., 2014. The pyruvate dehydrogenase complexes: structure-based function and regulation. Journal of Biological Chemistry 289, 16615-16623.

Pavlov, A.R., Pavlova, N.V., Kozyavkin, S.A., Slesarev, A.I., 2004. Recent developments in the optimization of thermostable DNA polymerases for efficient applications. Trends in biotechnology 22, 253-260.

Perham, R., 1975. Self-assembly of biological macromolecules. Philosophical Transactions of the Royal Society B: Biological Sciences 272, 123-136.

Perham, R.N., 1991. Domains, motifs, and linkers in 2-oxo acid dehydrogenase multienzyme complexes: a paradigm in the design of a multifunctional protein. Biochemistry 30, 8501-8512.

Perham, R.N., 2000. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. Annual Review of Biochemistry 69, 961-1004.

Polacek, N., Mankin, A.S., 2005. The ribosomal peptidyl transferase center: structure, function, evolution, inhibition. Critical Reviews in Biochemistry and Molecular Biology 40, 285-311.

Pósfai, J., Bhagwat, A.S., Pósfai, G., Roberts, R.J., 1989. Predictive motifs derived from cytosine methyltransferases. Nucleic Acids Research 17, 2421-2435.

Radford, S.E., Perham, R.N., Ullrich, S.J., Appella, E., 1989. Antibodies against an inter-domain segment of polypeptide chain inhibit active-site coupling in the pyruvate dehydrogenase multienzyme complex. FEBS Letters 250, 336-340.

Rahal, J.J., Simberkoff, M.S., 1979. Bactericidal and bacteriostatic action of chloramphenicol against meningeal pathogens. Antimicrobial agents and chemotherapy 16, 13-18.

Raleigh, E., 1992. Organization and function of the mcrBC genes of *Escherichia coli* K-12. Molecular Microbiology 6, 1079-1086.

Randle, P.J., Patston, P.A., Espinal, J., 1987. 4 Branched-Chain Ketoacid Dehydrogenase. In: Paul, D.B., Edwin, G.K. (Eds.), The Enzymes. Academic Press, pp. 97-121.

Ravikumar, A., Arrieta, A., Liu, C.C., 2014. An orthogonal DNA replication system in yeast. Nature Chemical Biology 10, 175-177.

Reed, L.J., 1974. Multienzyme complexes. Accounts of Chemical Research 7, 40-46.

Reed, L.J., Hackert, M.L., 1990. Structure-function relationships in dihydrolipoamide acyltransferases. The Journal of Biological Chemistry 265, 8971-8974.

Reed, L.J., Yeaman, S.J., 1987. Pyruvate dehydrogenase. The Enzymes 18, 77-95.

Reetz, M.T., Zonta, A., Schimossek, K., Jaeger, K.E., Liebeton, K., 1997. Creation of enantioselective biocatalysts for organic chemistry by in vitro evolution. Angewandte Chemie International edition in english 36, 2830-2832.

Rende-Fournier, R., Leclercq, R., Galimand, M., Duval, J., Courvalin, P., 1993. Identification of the satA gene encoding a streptogramin A acetyltransferase in *Enterococcus faecium* BM4145. Antimicrobial Agents and Chemotherapy 37, 2119-2125.

Ristuccia, A.M., 1985. Chloramphenicol: clinical pharmacology in pediatrics. Therapeutic Drug Monitoring 7, 159-167.

Roberts, M., Corney, A., Shaw, W., 1982. Molecular characterization of three chloramphenicol acetyltransferases isolated from *Haemophilus influenzae*. Journal of Bacteriology 151, 737-741.

Roberts, R.J., Myers, P.A., Morrison, A., Murray, K., 1976. A specific endonuclease from *Haemophilus haemolyticus*. Journal of Molecular Biology 103, 199-208.

Robien, M.A., Clore, G.M., Omichinski, J.G., Perham, R.N., Appella, E., Sakaguchi, K., Gronenborn, A.M., 1992. Three-dimensional solution structure of the E3-binding domain of the dihydrolipoamide succinyltransferase core from the 2-oxoglutarate dehydrogenase multienzyme complex of *Escherichia coli*. Biochemistry 31, 3463-3471.

Roche, T.E., Reed, L.J., Patel, M.S., 1989. Alpha-keto acid dehydrogenase complexes: organization, regulation, and biomedical ramifications: a tribute to Lester J. Reed. New York Academy of Sciences.

Rodriguez, A.C., Park, H.W., Mao, C., Beese, L.S., 2000. Crystal structure of a pol alpha family DNA polymerase from the hyperthermophilic archaeon *Thermococcus sp*. 9 degrees N-7. Journal of Molecular Biology 299, 447-462.

Rogers, E.J., Rahman, M.S., Hill, R.T., Lovett, P.S., 2002. The chloramphenicol-inducible catB gene in *Agrobacterium tumefaciens* is regulated by translation attenuation. Journal of Bacteriology 184, 4296-4300.

Root, M.J., Kay, M.S., Kim, P.S., 2001. Protein design of an HIV-1 entry inhibitor. Science 291, 884-888.

Rowe-Magnus, D.A., Guerout, A.M., Mazel, D., 2002. Bacterial resistance evolution by recruitment of super-integron gene cassettes. Molecular Microbiology 43, 1657-1669.

Saffarzadeh, M., Juenemann, C., Queisser, M.A., Lochnit, G., Barreto, G., Galuska, S.P., Lohmeyer, J., Preissner, K.T., 2012. Neutrophil extracellular traps directly induce epithelial and endothelial cell death: a predominant role of histones. PloS one 7, e32366.

Sanger, F., Coulson, A.R., 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology 94, 441-448.

Sanger, F., Nicklen, S., Coulson, A.R., 1977. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences 74, 5463-5467.

Sanger, F., Tuppy, H., 1951. The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. Biochemical Journal 49, 481-490.

Sankpal, U.T., Rao, D.N., 2002. Mutational analysis of conserved residues in HhaI DNA methyltransferase. Nucleic Acids Research 30, 2628-2638.

Scheuermann, R., Tam, S., Burgers, P.M., Lu, C., Echols, H., 1983. Identification of the epsilon-subunit of *Escherichia coli* DNA polymerase III holoenzyme as the dnaQ gene product: a fidelity subunit for DNA replication.

Proceedings of the National Academy of Sciences of the United States of America 80, 7085-7089.

Schluckebier, G., O'Gara, M., Saenger, W., Cheng, X., 1995. Universal catalytic domain structure of AdoMet-dependent methyltransferases. Journal of molecular biology 247, 16-20.

Schlünzen, F., Zarivach, R., Harms, J., Bashan, A., Tocilj, A., Albrecht, R., Yonath, A., Franceschi, F., 2001. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. Nature 413, 814-821.

Schwarz, S., Kehrenberg, C., Doublet, B., Cloeckaert, A., 2004. Molecular basis of bacterial resistance to chloramphenicol and florfenicol. FEMS Microbiology Reviews 28, 519-542.

Scolnick, E., Tompkins, R., Caskey, T., Nirenberg, M., 1968. Release factors differing in specificity for terminator codons. Proceedings of the National Academy of Sciences of the United States of America 61, 768-774.

Serva, S., Vilkaitis, G., Venclovas, Č., Klimašauskas, S., 2004. HhaI DNA methyltransferase uses the protruding Gln237 for active flipping of its target cytosine. Structure 12, 1047-1055.

Sezonov, G., Joseleau-Petit, D., D'Ari, R., 2007. *Escherichia coli* physiology in Luria-Bertani broth. Journal of Bacteriology 189, 8746-8749.

Shafikhani, S., Siegel, R., Ferrari, E., Schellenberger, V., 1997. Generation of large libraries of random mutants in *Bacillus subtilis* by PCR-based plasmid multimerization. Biotechniques 23, 304-311.

Shapiro, J., 1983. Mobile Genetic Elements (Academic, New York). ShapiroMobile Genetic Elements 1983.

Shaw, W., 1967. The enzymatic acetylation of chloramphenicol by extracts of R factor-resistant *Escherichia coli*. Journal of Biological Chemistry 242, 687-693.

Shaw, W., Bentley, D., Sands, L., 1970. Mechanism of chloramphenicol resistance in *Staphylococcus epidermidis*. Journal of Bacteriology 104, 1095-1105.

Shaw, W., Leslie, A., 1991. Chloramphenicol acetyltransferase. Annual Review of Biophysics and Biophysical Chemistry 20, 363-386.

Shaw, W., Unowsky, J., 1968. Mechanism of R factor-mediated chloramphenicol resistance. Journal of Bacteriology 95, 1976-1978.

Shaw, W.V., 1983. Chloramphenicol acetyltransferase: enzymology and molecular biology. CRC Critical Reviews in Biochemistry 14, 1-46.

Shaw, W.V., 1984. Bacterial resistance to chloramphenicol. British Medical Bulletin 40, 36-41.

Simon, W.S., 2000. Chloramphenicol. Therapie in Klinik und Praxis (10th. edn.), Schattauer, Stuttgart, New York (2000), , pp. 172–175.

Sohmen, D., Harms, J.M., Schlünzen, F., Wilson, D.N., 2009. SnapShot: Antibiotic Inhibition of Protein. Cell 139, 212-212.

Song, J., Jordan, F., 2012. Interchain acetyl transfer in the E2 component of bacterial pyruvate dehydrogenase suggests a model with different roles for each chain in a trimer of the homooligomeric component. Biochemistry 51, 2795-2803.

Spee, J.H., de Vos, W.M., Kuipers, O.P., 1993. Efficient random mutagenesis method with adjustable mutation frequency by use of PCR and dITP. Nucleic Acids Research 21, 777-778.

Sprang, S., Standing, T., Fletterick, R., Stroud, R., Finer-Moore, J., Xuong, N., Hamlin, R., Rutter, W., Craik, C., 1987. The three-dimensional structure of Asn102 mutant of trypsin: role of Asp102 in serine protease catalysis. Science 237, 905-909.

Stemmer, W.P., 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature 370, 389-391.

Stephens, P.E., Darlison, M.G., Lewis, H.M., Guest, J., 1983a. The pyruvate dehydrogenase complex of *Escherichia coli* K12. European Journal of Biochemistry 133, 155-162.

Stephens, P.E., Lewis, H.M., Darlison, M.G., Guest, J.R., 1983b. Nucleotide sequence of the lipoamide dehydrogenase gene of *Escherichia coli* K12. European Journal of Biochemistry 135, 519-527.

Stormo, G.D., Schneider, T.D., Gold, L.M., 1982. Characterization of translational initiation sites in *E. coli*. Nucleic Acids Research 10, 2971-2996.

Stryer, L., 1968. Implications of X-ray crystallographic studies of protein structure. Annual Review of Biochemistry 37, 25-50.

Sun, F., 1995. The polymerase chain reaction and branching processes. Journal of Computational Biology 2, 63-86.

Tennigkeit, J., Matzura, H., 1991. Nucleotide sequence analysis of a chloramphenicol-resistance determinant from *Agrobacterium tumefaciens* and identification of its gene product. Gene 98, 113-116.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proceedings of the National Academy of Sciences 99, 12391-12396.

Timinskas, A., Butkus, V., Janulaitis, A., 1995. Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. Gene 157, 3-11.

Tindall, K.R., Kunkel, T.A., 1988. Fidelity of DNA synthesis by the *Thermus aquaticus* DNA polymerase. Biochemistry 27, 6008-6013.

Trautner, T., Pawlek, B., Behrens, B., Willert, J., 1996. Exact size and organization of DNA target-recognizing domains of multispecific DNA-(cytosine-C5)-methyltransferases. The EMBO Journal 15, 1434-1442.

Turk, D., 1977. A comparison of chloramphenicol and ampicillin as bactericidal agents for *Haemophilus influenzae* type B. Journal of Medical Microbiology 10, 127-131.

Ulbrich, B., Mertens, G., Nierhaus, K.H., 1978. Cooperative binding of 3′-fragments of transfer ribonucleic acid to the peptidyltransferase center of *Escherichia coli* ribosomes. Archives of biochemistry and biophysics 190, 149-154.

Van der Schueren, J., Robben, J., Goossens, K., Heremans, K., Volckaert, G., 1996. Identification of local carboxy-terminal hydrophobic interactions essential for folding or stability of chloramphenicol acetyltransferase. Journal of molecular biology 256, 878-888.

Van der Schueren, J., Robben, J., Volckaert, G., 1998. Misfolding of chloramphenicol acetyltransferase due to carboxy-terminal truncation can be corrected by second-site mutations. Protein Engineering 11, 1211-1217.

Vartanian, J.-P., Henry, M., Wain-Hobson, S., 1996a. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. Nucleic Acids Research 24, 2627-2631.

Vartanian, J.P., Henry, M., Wain-Hobson, S., 1996b. Hypermutagenic PCR involving all four transitions and a sizeable proportion of transversions. Nucleic Acids Research 24, 2627-2631.

Vassort-Bruneau, C., Lesage-Descauses, M.-C., Martel, J.-L., Lafont, J.-P., Chaslus-Dancla, E., 1996. CAT III chloramphenicol resistance in *Pasteurella haemolytica* and Pasteurella multocida isolated from calves. Journal of Antimicrobial Chemotherapy 38, 205-213.

Vilkaitis, G., Dong, A., Weinhold, E., Cheng, X., Klimašauskas, S., 2000. Functional roles of the conserved threonine 250 in the target recognition domain of HhaI DNA methyltransferase. Journal of Biological Chemistry 275, 38722-38730.

Vilkaitis, G., Merkiene, E., Serva, S., Weinhold, E., Klimasauskas, S., 2001. The mechanism of DNA cytosine-5 methylation. Kinetic and mutational dissection of Hhai methyltransferase. The Journal of Biological Chemistry 276, 20924-20934.

Villa, L., Mammina, C., Miriagou, V., Tzouvelekis, L.S., Tassios, P.T., Nastasi, A., Carattoli, A., 2002. Multidrug and broad-spectrum cephalosporin resistance among *Salmonella enterica* serotype Enteritidis clinical isolates in southern Italy. Journal of Clinical Microbiology 40, 2662-2665.

Vlček, Č., Paces, V., Maltsev, N., Pačes, J., Haselkorn, R., Fonstein, M., 1997. Sequence of a 189-kb segment of the chromosome of *Rhodobacter capsulatus* SB1003. Proceedings of the National Academy of Sciences 94, 9384-9388.

Vogel, Z., Vogel, T., Zamir, A., Elson, D., 1971. Correlation between the peptidyl transferase activity of the 50 s ribosomal subunit and the ability of the subunit to interact with antibiotics. Journal of Molecular Biology 60, 339-346.

Völker, T.A., Iida, S., Bickle, T.A., 1982. A single gene coding for resistance to both fusidic acid and chloramphenicol. Journal of Molecular Biology 154, 417-425.

Wahler, D., Reymond, J.L., 2001. High-throughput screening for biocatalysts. Current Opinion in Biotechnology 12, 535-544.

Wang, J., Sattar, A.A., Wang, C., Karam, J., Konigsberg, W., Steitz, T., 1997. Crystal structure of a pol α family replication DNA polymerase from bacteriophage RB69. Cell 89, 1087-1099.

Wang, T.-W., Zhu, H., Ma, X.-Y., Zhang, T., Ma, Y.-S., Wei, D.-Z., 2006. Mutant library construction in directed molecular evolution. Molecular Biotechnology 34, 55-68.

Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., Vander Horn, P.B., 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. Nucleic Acids Research 32, 1197-1207.

Watanabe, T., 1963. Infective heredity of multiple drug resistance in bacteria. Bacteriological Reviews 27, 87.

Wehrle, P.F., Mathies, A.W., Leedom, J.M., Ivler, D., 1967. Bacterial meningitis. Annals of the New York Academy of Sciences 145, 488-498.

Werner, G., Witte, W., 1999. Characterization of a new enterococcal gene, satG, encoding a putative acetyltransferase conferring resistance to streptogramin A compounds. Antimicrobial Agents and Chemotherapy 43, 1813-1814.

Wharton, R.P., Ptashne, M., 1985. Changing the binding specificity of a repressor by redesigning an alpha-helix. Nature 316, 601-605.

White, M.F., Bell, S.D., 2002. Holding it together: chromatin in the Archaea. Trends in Genetics 18, 621-626.

White, P.A., Stokes, H., Bunny, K.L., Hall, R.M., 1999. Characterisation of a chloramphenicol acetyltransferase determinant found in the chromosome of *Pseudomonas aeruginosa*. FEMS Microbiology Letters 175, 27-35.

Wilke, K., Rauhut, E., Noyer-Weidner, M., Lauster, R., Pawlek, B., Behrens, B., Trautner, T., 1988. Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. The EMBO Journal 7, 2601-2609.

Wilkinson, A.J., Fersht, A.R., Blow, D.M., Winter, G., 1983. Site-directed mutagenesis as a probe of enzyme structure and catalysis: tyrosyl-tRNA synthetase cysteine-35 to glycine-35 mutation. Biochemistry 22, 3581-3586.

Wilkinson, C.R., Bartlett, R., Nurse, P., Bird, A.P., 1995. The fission yeast gene pmt1+ encodes a DNA methyltransferase homologue. Nucleic Acids Research 23, 203-210.

Williams, G.J., Nelson, A.S., Berry, A., 2004. Directed evolution of enzymes for biocatalysis and the life sciences. Cellular and Molecular Life Sciences : CMLS 61, 3034-3046.

Wilson, G.G., Murray, N.E., 1991. Restriction and modification systems. Annual Review of Genetics 25, 585-627.

Wolstenholme, G., O'Connor, C.M., 1957. Drug resistance in micro-organisms. Symposium on Drug Resistance in Micro-organisms (1957: London). Little, Brown and Co.

Wong, I., Patel, S.S., Johnson, K.A., 1991. An induced-fit kinetic mechanism for DNA replication fidelity: direct measurement by single-turnover kinetics. Biochemistry 30, 526-537.

Wong, T.S., Tee, K.L., Hauer, B., Schwaneberg, U., 2004. Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution. Nucleic Acids Research 32, e26-e26.

Wu, J.C., Santi, D., 1987. Kinetic and catalytic mechanism of HhaI methyltransferase. Journal of Biological Chemistry 262, 4778-4786.

Xaplanteri, M.A., Andreou, A., Dinos, G.P., Kalpaxis, D.L., 2003. Effect of polyamines on the inhibition of peptidyltransferase by antibiotics: revisiting the mechanism of chloramphenicol action. Nucleic Acids Research 31, 5074-5083.

Yang, H., Hainfeld, J., Wall, J., Frey, P.A., 1985. Quaternary structure of pyruvate dehydrogenase complex from *Escherichia coli*. Journal of Biological Chemistry 260, 16049-16051.

Yen, R.W., Vertino, P.M., Nelkin, B.D., Yu, J.J., el-Deiry, W., Cumaraswamy, A., Lennon, G.G., Trask, B.J., Celano, P., Baylin, S.B., 1992. Isolation and characterization of the cDNA encoding human DNA methyltransferase. Nucleic Acids Research 20, 2287-2291.

Zaccolo, M., Williams, D.M., Brown, D.M., Gherardi, E., 1996. An approach to random mutagenesis of DNA using mixtures of triphosphate derivatives of nucleoside analogues. Journal of Molecular Biology 255, 589-603.

Zhao, H., Giver, L., Shao, Z., Affholter, J.A., Arnold, F.H., 1998. Molecular evolution by staggered extension process (StEP) in vitro recombination. Nature Biotechnology 16, 258-261.

Zhao, Y., Jeruzalmi, D., Moarefi, I., Leighton, L., Lasken, R., Kuriyan, J., 1999. Crystal structure of an archaebacterial DNA polymerase. Structure 7, 1189-1199.

Appendices

Appendix 1: Amino acid changes and activity of M.HhaI mutants:

Sample ID	Activity	Amino acids changes
2nd1 1.10.14	PA	V36I, G92S, T99N, V108L
2nd3 1.10.14	IA	K7Q, V48F, P57S
2nd4 1.10.14	A	K56I
2nd5 1.10.14	A	V2I, L9S, R13C
2nd11 1.10.14	IA	E40K
2nd13 1.10.14	IA	K5N, C35Y, D60V, C76S
2nd14 1.10.14	A	T10I, F93I
2nd16 1.10.14	A	V48M
2nd19 1.10.14	A	K56Q
2nd22 1.10.14	A	Q8E
2nd25 1.10.14	IA	D73N, Q82H, S87F, N120Y
2nd28 1.10.14	PA	A45T, F84L
2nd29 1.10.14	A	I15N, E66A
2nd31 1.10.14	A	F14Y, G59D, V115I
2nd32 1.10.14	A	E47K, G54D
2nd33 1.10.14	A	V48L, G92S
2nd34 1.10.14	IA	G20S, A45T
2nd35 1.10.14	A	E3K, E58D
2nd36 1.10.14	A	N39D
2nd38 1.10.14	A	F102Y
2nd40 1.10.14	A	A45P
2nd41 1.10.14	A	R109H
2nd42 1.10.14	A	V48A
2nd44 1.10.14	IA	Q46H, Q90L, E119D
2nd45 1.10.14	IA	G23V, E47D, K67E, P70H, A83T, F102Y
2nd46 1.10.14	A	L17M, N65Y, A83V, V116L
2nd51 1.10.14	A	Y37H
2nd52 1.10.14	A	I4N, R13L, K91I, I104V, R109H
2nd54 1.10.14	A	F18Y, K112E
2nd57 1.10.14	IA	A19V, G88E
2nd58 1.10.14	A	E50G, A34V
2nd63 1.10.14	A	D60N, L100P
2nd71 1.10.14	IA	G11C, L21R, M51V, T99N
2nd73 1.10.14	A	N52K, V64I
2nd75 1.10.14	IA	G88V
2nd76 1.10.14	IA	A19V, G88E
2nd78 1.10.14	A	A27T, K89E
2nd80 1.10.14	IA	G20S, D73V
2nd81 1.10.14	A	K5N, M51L, D73V
2nd83 1.10.14	IA	L26Q, E40D, L75M, C76Y, V116L
2nd85 1.10.14	A	D95Y
2nd86 1.10.14	IA	D6N, F18C, Y49N, S87T
2nd87 1.10.14	A	T10A, L100M
2nd88 1.10.14	A	I4L
2nd92 1.10.14	A	C31Y
2nd95 1.10.14	A	D103V, R106L
2nd97 1.10.14	IA	S87Y, V121M

2nd98 1.10.14	IA	G20R, Q63L
2nd99 1.10.14	IA	G88E
2nd100 1.10.14	A	K43R
2nd104 1.10.14	A	I107F
2nd108 1.10.14	IA	G20C, R25S, K122Q
2nd109 1.10.14	IA	G20S, D60N, D73E, D95E, R106L
2nd110 1.10.14	IA	K7I, F14Y, I15N, L21M, N65S, R109C
2nd111 1.10.14	A	A83T
2nd113 1.10.14	A	E47K, G54D
2nd114 1.10.14	A	I15N
2nd115 1.10.14	IA	S30C, A33S, E40-, K43I, A105T
2nd116 1.10.14	A	I4F, F84Y
2nd117 1.10.14	A	A19V
2nd119 1.10.14	IA	C35-, I86M, L100M
2nd120 1.10.14	A	E3D, I15V, W41-
2nd121 1.10.14	A	F14I, I107V
2nd122 1.10.14	A	I61M, P70H
2nd124 1.10.14	A	Q90-
2nd125 1.10.14	IA	A45G, E50I, F53Y, R106H
2nd126 1.10.14	PA	F24Y, A83T, V108F
2nd127 1.10.14	PA	V121M
2nd6 8.5.14	IA	G22D, E29D, G32C, D71V, D73V, F79L, P80Q, Q90H
2nd11 8.5.14	A	Y37F, K89R, R97C
2nd12 8.5.14	IA	G20C, F79L, P80Q, D95Y, S96I, E119D, F117C
2nd14 8.5.14	IA	G59V, K91N, M118T, F124C
2nd16 8.5.14	IA	F24L, A45T, K89E
2nd17 8.5.14	IA	F18S, D42V, V116M, D103G, V121M
2nd19 8.5.14	IA	G23D, E29V, K91N
2nd20 8.5.14	IA	G23S, K56I, I104N, K111K
2nd22 8.5.14	IA	G22C, F24C, R25S, T62S, V64D, F93C, E94D, F101L
2nd24 8.5.14	A	K5N, E29D, F79I
2nd25 8.5.14	IA	P57T, S85I
2nd28 8.5.14	A	G11S, S38G, M51V
2nd29 8.5.14	IA	I4F, L9S, L17M, G32C, K91R, V115I
2nd31 8.5.14	A	D6H
2nd32 8.5.14	IA	Q46F, N52S, T99I
2nd37 8.5.14	A	N120I
2nd42 8.5.14	IA	G20C, C35F, G59D, I86F
2nd45 8.5.14	A	Q46H, Q90H
2nd51 8.5.14	A	V2I, E55G, F53S
2nd53 8.5.14	IA	E40K, D60N, T62I, A77V, G92S, F101S
2nd54 8.5.14	A	E3G, F53Y, R109C
2nd55 8.5.14	IA	I4F, G11I, N65H, P80T, V121A
2nd59 8.5.14	A	S96N, I104N
3rd2 8.8.14	A	D128F, R245H
3rd3 8.8.14	IA	E142K, F171L, F186C
3rd4 8.8.14	IA	M168I, R240L, I247T
3rd6 8.8.14	A	I177T
3rd9 8.8.14	A	T222I, R245S
3rd10 8.8.14	A	Q181K
3rd12 8.8.14	A	G158S, I219V

3rd13 8.8.14	A	D156V, N189K
3rd16 8.8.14	A	L155F
3rd17 8.8.14	A	L143P, S146N
3rd19 8.8.14	IA	V136A, P183L, N189S
3rd22 8.8.14	IA	L205V, G230D, K234T
3rd23 8.8.14	A	N138I, V192M, S200G
3rd24 8.8.14	A	Q161H
3rd26 8.8.14	A	V206I
3rd27 8.8.14	A	A149G, I177T
3rd28 8.8.14	A	H148Y, H204Q
3rd29 8.8.14	A	F191I
3rd30 8.8.14	A	N176D, F171I
3rd32 8.8.14	IA	N131K, I207F, Q237K
3rd34 8.8.14	A	M139I, V213A, T222I
3rd35 8.8.14	IA	E142D, G158D, H204L
3rd37 8.8.14	A	T226I
3rd38 8.8.14	IA	P160T, I207T, K234T
3rd43 8.8.14	A	E164V, I231F
3rd45 8.8.14	A	H127Y, L229Q
3rd47 8.8.14	IA	H127L, R163G, I169L, L188Q, P198S, T226I, I231V, Q237H
3rd49 8.8.14	A	I247F
3rd50 8.8.14	A	D144Y
3rd51 8.8.14	A	N138T, V206I, Q217H, T226I
3rd53 8.8.14	A	N173I, H204L
3rd55 8.8.14	A	N179Y
3rd57 8.8.14	IA	A154T, Y157H, D174V, N189D, V227D, Q237K
3rd58 8.8.14	A	M214I
3rd59 8.8.14	A	M140I, A149T
3rd60 8.8.14	A	K137E, T139S
3rd61 8.8.14	IA	L143M, K150I, G230D, I247F
4th7 23.6.14	IA	P286S, C279Y
4th9 23.6.14	A	Q297K
4th15 23.6.14	A	A253V
4th16 23.6.14	IA	A253P, V282D, D287E, S294N, N323Y
4th19 23.6.14	A	P276Q
4th23 23.6.14	IA	G256R, H275N, S294N, Q301R, G319S
4th24 23.6.14	A	F324S
4th25 23.6.14	IA	S296N, K300E, V307M
4th39 23.6.14	IA	S252R, P286S, Y316C, L322M
4th42 23.6.14	A	A253V
4th5 1.7.14	A	Y313C, N317D
4th6 1.7.14	A	C279G, Y313C
4th7 1.7.14	A	L322P
4th8 1.7.14	A	G257S
4th10 1.7.14	A	M283I
4th11 1.7.14	A	S294G
4th15 1.7.14	A	S288R
4th16 1.7.14	A	R281C
4th17 1.7.14	IA	E278G, G303D
4th19 1.7.14	IA	T271I, C279R, S294N, P326S
4th23 1.7.14	IA	F259V, S288I, Q301H, K325T
4th26 1.7.14	A	L266M

4th28 1.7.14	IA	G255D
4th32 1.7.14	A	D287N
4th36 1.7.14	A	K325R
4th38 1.7.14	A	A280G
4th42 1.7.14	IA	G256D, V310M
4th44 1.7.14	A	P293H, Q297R, A315V
2nd2 15.4.14	A	Q82H
2nd1 16.4.14	A	K7Q
4th4 15.4.14	IA	K261Q, K273I, R281C, Q312L
4th1 25.4.14	IA	G263D, G264S, S288N, T295I, G319D
4th5 25.4.14	IA	H275R, V291D
4th6 25.4.14	IA	K273I, R281C, Q312L

*Red-shaded are mutations within highly conserved and relatively conserved residues.

*Yellow-shaded are mutations within the TRD.

*Green-shaded are mutations within the hinge region.

*Not-shaded are mutations of not conserved residues within catalytic domain.

*Hyphens (-) indicate that the codons of the mentioned residue were changed by mutations into TAG stop codon.

* A= Active, IA= Inactive, PA= Partially active.

Appendix 2: Amino acid changes and activity of CATI mutants:

Sample ID	Activity	Amino acids changes
4thAt1 15.2.15	IA	F95S, T101I, T172K
4thAt5 15.2.15	A	R203G
4thAt6 15.2.15	IA	V124A, S155T, G180S, Q190R, G202D
4thAt10 15.2.15	IA	Y133F, G137C, F153I
4thAt13 15.2.15	PA	C196Y
4thAt15 15.2.15	IA	H110Q, W215C, L208P, V201D
4thAt16 15.2.15	IA	E129D, S146R, H192R
4thAt17 15.2.15	IA	V88D, R114C, Y120F, F166I, Q190K
4thAt19 15.2.15	PA	R203H
4thAt23 15.2.15	IA	S155C, F167V, F171L
4thAt24 15.2.15	IA	E81V, F171L, F167Y, D181V
4thAt27 15.2.15	A	D181V
4thAt28 15.2.15	A	E129V
4thAt30 15.2.15	A	N162K
4thAt33 15.2.15	A	A147V, D164E, A219T
4thAt36 15.2.15	IA	A132E, Y176N, V191D
4thAt37 15.2.15	A	F113I, E214K
4thAt43 15.2.15	IA	E97V, I139T, C196R
4thAt47 15.2.15	IA	Q190R
4thAt49 15.2.15	IA	Y120C, N130K, F171L, H193Y
4thAt53 15.2.15	IA	A161V, M204V
4thAt 5515.2.15	IA	F116L, F143I, S146I, F166I, L117P
4thAt56 15.2.15	IA	V94G, Y127F, D164V, W215R, G217A, A219T
4thAt57 15.2.15	IA	W85G, S155I
4thAt58 15.2.15	IA	T93I, E108D, F167L, Q190H
4thAt60 15.2.15	IA	S103N, D111N, H193N
4thAt63 15.2.15	IA	R74C, C91Y, L205M
4thAt65 15.2.15	PA	A125V, K136N
4thAt67 15.2.15	IA	Q115H, T154N, M185K
4thAt71 15.2.15	IA	S107P, P135S, H200Y
4thAt82 15.2.15	A	E97V
4thAt84 15.2.15	IA	A132V, T154I, M163L
4thAt87 15.2.15	IA	I84F, A161V
4thAt88 15.2.15	A	D86E
4thAt91 15.2.15	A	R114S
4thAt93 15.2.15	IA	S103T, F113Y, I189T, C196Y
4thAt94 15.2.15	PA	D111N, N162S
4thAt95 15.2.15	IA	F167L, Y176C, K175I
4thAt96 15.2.15	IA	V83D, N141T, A168T
4thAt97 15.2.15	A	S121I
4thAt98 15.2.15	IA	E97A, T99N, C126Y, Y127C, Y133F, A168S, E207D, Y211H
4thAt100 15.2.15	IA	T101I, S155T, M204T
4thAt103 15.2.15	PA	D79N, A147T, C196S
4thAt105 15.2.15	A	G217S
4thAt106 15.2.15	A	G137C
4thAt107 15.2.15	IA	G202D
4thAt109 15.2.15	PA	H96L, E97K, L158P, V183L, H200Q
4thAt110 15.2.15	IA	F95L, G198C
4thAt111 15.2.15	PA	A76G, C212Y

4thAt112 15.2.15	PA	S152R, H100Q
4thAt113 15.2.15	PA	G128A, N130S
4thAt115 15.2.15	A	D79Y, E81D, V191I
4thAt118 15.2.15	IA	F116L*
4thAt119 15.2.15	IA	D164Y, D197C
4thAt120 15.2.15	A	M77K, A125S, M185L, L208V
4thAt121 15.2.15	IA	V88D, Q98H, S103R, Q122P
4thAt122 15.2.15	IA	C196R
4thAt127 15.2.15	IA	V88D, V183F, M204I
4thAt129 15.2.15	IA	F143S
4thAt130 15.2.15	PA	N148D
4thAt132 15.2.15	A	N165I
4thAt137 15.2.15	A	D111E, S121G, P169S, Q179H
4thAt139 15.2.15	IA	F167L, V195I
4thAt141 15.2.15	A	V83F, T99S, H118R, Q122H, D181N
4thAt144 15.2.15	A	Q98H
4thAt145 15.2.15	IA	V83L, S103R, N130H, M142V, M173K, Q216L
4thAt146 15.2.15	IA	A76V, F102I, G128D, F144L, V160M
4thAt147 15.2.15	IA	F102Y, N162H
4thAt148 15.2.15	PA	C91S, P169T
4thAt153 15.2.15	A	M204L, W215C
4thAt154 15.2.15	PA	A147T
4thAt156 15.2.15	A	V88L
4thAt157 15.2.15	IA	V94G, R114S, F171L
4thAt159 15.2.15	IA	K78E, F171L, G118S
4thAt161 15.2.15	A	N130K
4thAt164 15.2.15	A	V195I
4thAt165 15.2.15	A	D79Y, A132T
4thAt166 15.2.15	IA	D157V, H193L
4thAt167 15.2.15	IA	H96R, L117V, F153L, M185V
4thAt168 15.2.15	IA	D112V, F153Y, F156L
4thAt171 15.2.15	A	W215R
4thAt172 15.2.15	IA	W106R, D213E
4thAt173 15.2.15	A	T178I
4thAt174 15.2.15	IA	F171L, T172M, R203H
4thAt175 15.2.15	A	H118R, D164E, Q210H
4thAt177 15.2.15	PA	D157V
4thAt178 15.2.15	IA	R74S, G128D, L131R, E140D, F144Y, S146R, F153L
4thAt179 15.2.15	IA	D79V, G80D, V83A, D111E, D197G, M204I
4thAt186 15.2.15	IA	Y109C, Q115R, P149S, F171L, G217D
4thAt187 15.2.15	PA	H89L, A168S, K182I
4thAt192 15.2.15	A	V201I, Q216TAG, G217D
4thAt193 15.2.15	IA	M77V, W85R, Q98H, T154I, P169S
4thAt196 15.2.15	IA	Y133D, G174S, D197E
4thAt198 15.2.15	IA	N162Y, F166L, G174C
4thAt199 15.2.15	IA	M173K, G198S

*Red-shaded are mutations within highly conserved and relatively conserved residues.

*A=Active, IA=Inactive, PA=Partially active.

Appendix 3: Amino acid changes in E3 mutant clones and activity impact on PDH and ODH enzymes complexes:

Sample ID	Activity	Amino acidschanges and comments
1st lpd7 22.6.15	Active PDH and ODH	V62D
1st lpd8 22.6.15	=	E65D
1st lpd3 4.7.15	=	V75I, V112D
1st lpd4 4.7.15	=	G121R, G132C
1st lpd6 4.7.15	=	I63V, M104K
1st lpd9 4.7.15	=	G99D, I136T
1st lpd10 4.7.15	=	G29S
1st lpd15 4.7.15	=	V11M, F76L
1st lpd18 4.7.15	=	L116R
1st lpd19 4.7.15	=	R37C
1st lpd22 4.7.15	=	E78K
1st lpd23 4.7.15	=	V44I, A55G
1st lpd25 4.7.15	=	E31D
1st lpd26 4.7.15	=	R161H
1st lpd29 4.7.15	=	G77S, K133R
1st lpd31 4.7.15	=	V62I, G107D
1st lpd33 4.7.15	=	V93L
1st lpd34 4.7.15	=	N140S, F154L
1st lpd35 4.7.15	=	L28M, L69P, N140D
1st lpd36 4.7.15	=	K85T
1st lpd38 4.7.15	=	T81P
1st lpd39 4.7.15	=	V59A, T166I
1st lpd40 4.7.15	=	H157Y, L169M
1st lpd41 4.7.15	=	E78V
1st lpd44 4.7.15	=	E31D, L46M
1st lpd45 4.7.15	=	A102T, Q151R, D164G
1st lpd47 4.7.15	=	V9I, H72Q, W163L
1st lpd53 4.7.15	=	E71K, K118I, P160A
1st lpd56 4.7.15	=	K67N
1st lpd57 4.7.15	=	G103D, T134I
1st lpd59 4.7.15	=	H58Y, K61E
4th lpd1 21.7.15	=	K61R, L232P, K299Q, P318A
4th lpd2 21.7.15	=	I5T, N114H
4th lpd3 21.7.15	=	E36G
4th lpd4 21.7.15	=	D27N, K216E, I248F, F288K
4th lpd5 21.7.15	=	V62A, K229N
4th lpd6 21.7.15	=	V241L
4th lpd7 21.7.15	=	Q300L, H322Q
4th lpd8 21.7.15	=	H326Y
4th lpd9 21.7.15	=	G121D, N131S, I150N
4th lpd10 21.7.15	=	T88I, N276K
4th lpd11 21.7.15	=	V241I
4th lpd13 21.7.15	=	G129S
4th lpd18 21.7.15	=	Q96H, A145V, E288D
4th lpd20 21.7.15	=	I94N
4th lpd21 21.7.15	=	K237R
4th lpd22 21.7.15	=	R148H
4th lpd23 21.7.15	=	G77C

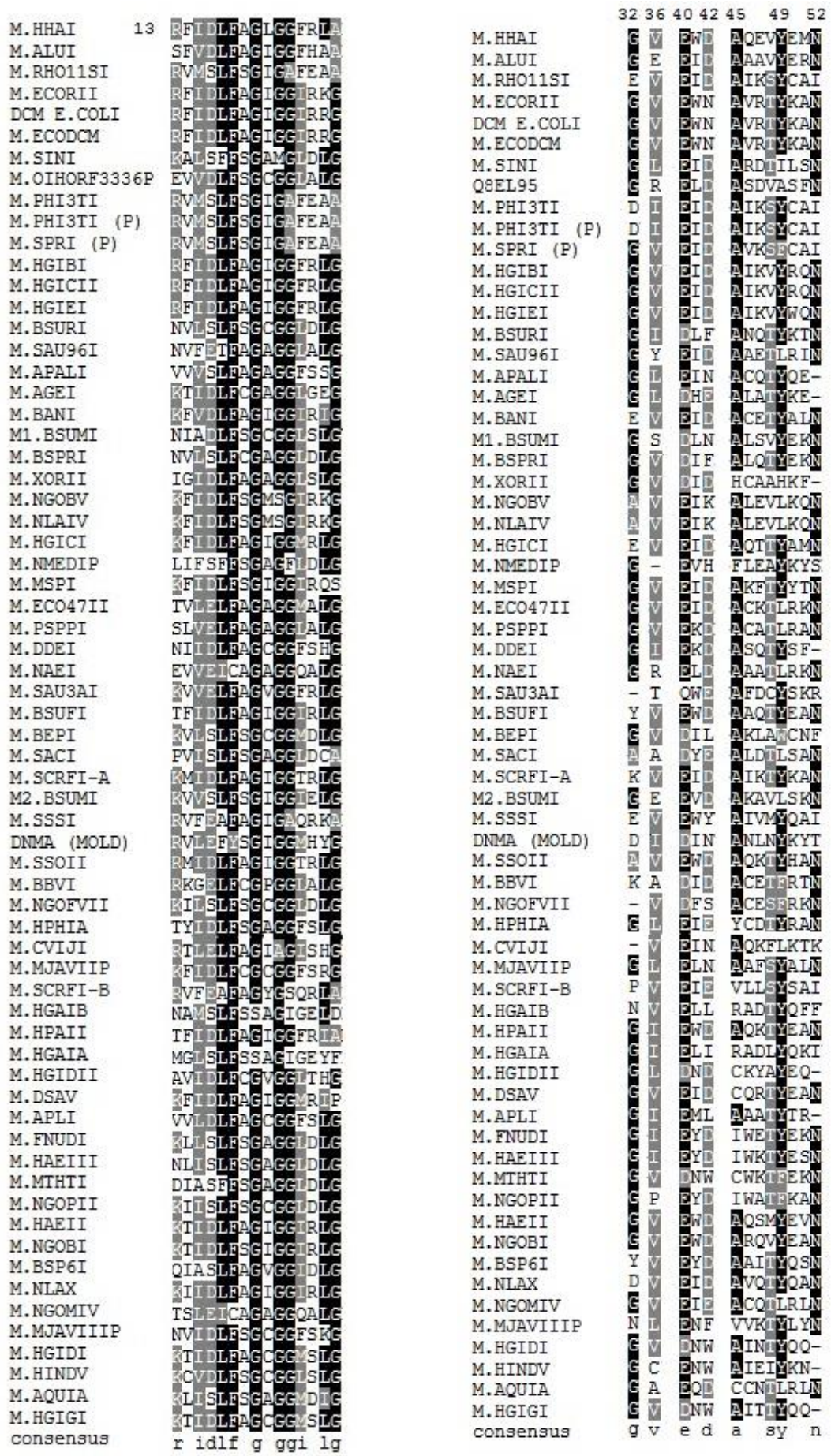
4th lpd24 21.7.15	Active PDH and ODH	E65A, G77D, T223I, K256N, N279I, R296P
4th lpd26 21.7.15	=	E78K
4th lpd27 21.7.15	=	V210I
4th lpd28 21.7.15	=	I51F, S227R
4th lpd29 21.7.15	=	E176D, A196V, G254D
4th lpd30 21.7.15	=	G77D, G283R, K299E
4th lpd31 21.7.15	=	K85I
4th lpd32 21.7.15	=	Q96L
4th lpd33 21.7.15	=	K118N, R296H
4th lpd35 21.7.15	=	L178P, G272S
4th lpd39 21.7.15	=	R37C, G73G
4th lpd42 21.7.15	=	V180L, V193I
4th lpd45 21.7.15	=	V193I
4th lpd46 21.7.15	=	A22S
4th lpd47 21.7.15	=	T98A
4th lpd50 21.7.15	=	V10I, H157Y, K237I
4th lpd51 21.7.15	=	V62I, M252V
4th lpd52 21.7.15	=	V9I, R24P, L46Q, T192A
4th lpd54 21.7.15	=	V180E
4th lpd55 21.7.15	=	V10I, G73D, K109I, A196V, D202N, V325I
4th lpd56 21.7.15	=	T134I
4th lpd57 21.7.15	=	A22V, V33F
4th lpd59 21.7.15	=	S53G, F222S
4th lpd60 21.7.15	=	K6I, G73D, G286D
4th lpd62 21.7.15	=	N123Y, N140I, V267F
4th lpd63 21.7.15	=	L188M, V193A
4th lpd64 21.7.15	=	R108P
5th lpd4 9.8.15	=	G254D, I311S
5th lpd5 9.8.15	=	T441I
5th lpd6 9.8.15	=	E327Q, M396I
5th lpd7 9.8.15	=	K323N, K365R
5th lpd8 9.8.15	=	L362P
5th lpd9 9.8.15	=	G419D, G461V, P467L
5th lpd10 9.8.15	=	V204A, V221L, R263P, K403E
5th lpd11 9.8.15	=	S349R
5th lpd12 9.8.15	=	K224I, G293D, A351T
5th lpd14 9.8.15	=	D202N + A deletion has changed the ORF so that 29 additional amino acids (NLDSGDPWASWAFRSLPAFQSGNLSCQLH) added to the C-terminal
5th lpd15 9.8.15	=	L280R, Q300E
5th lpd17 9.8.15	=	G198A, I271T, R302C
5th lpd18 9.8.15	=	A184A, P261S, V458L
5th lpd20 9.8.15	=	D291E, A387T
5th lpd21 9.8.15	=	F294S
5th lpd22 9.8.15	=	K244N, S349I
5th lpd23 9.8.15	=	K228N, G272S, R302H
5th lpd24 9.8.15	=	A243T, A257S, I409N, L422M
5th lpd25 9.8.15	=	G272A
5th lpd26 9.8.15	=	A310S
5th lpd27 9.8.15	=	I425T, S462C
5th lpd28 9.8.15	=	G254C, L454M, F459Y
5th lpd29 9.8.15	=	P318S, I425V

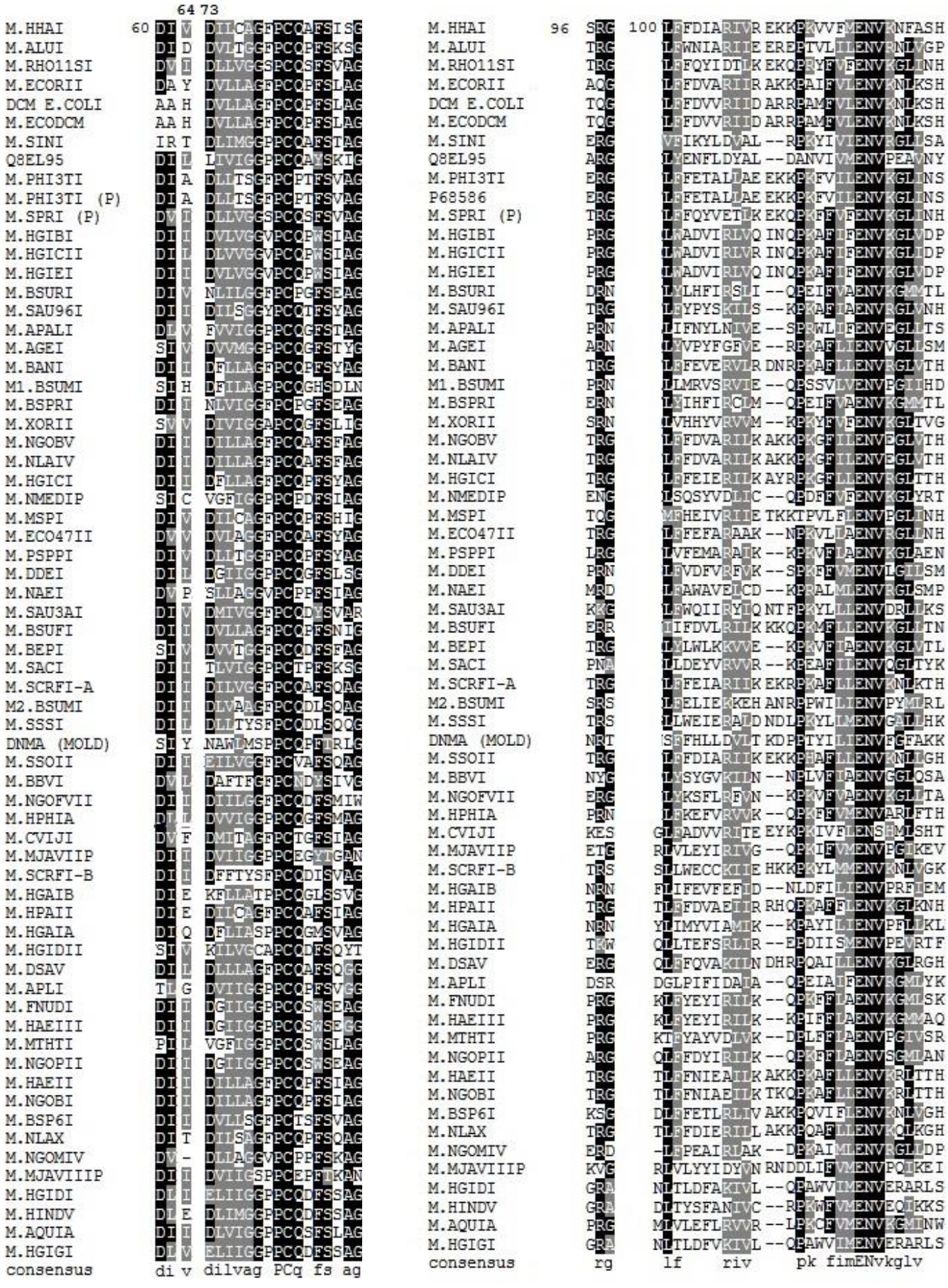
5th lpd31 9.8.15	Active PDH, inactive ODH	Q300H, I409N, G423S, A471V
5th lpd32 9.8.15	Active PDH and ODH	A389G
5th lpd33 9.8.15	Active PDH, partially active ODH	K256I
5th lpd35 9.8.15	Active PDH and ODH	L190M, Q205E, V240A, I289V
5th lpd36 9.8.15	Active PDH, partially active ODH	T192N, G286D, V360I, I442F
5th lpd41 9.8.15	Active PDH and ODH	Y264S
5th lpd42 9.8.15	=	Q200H, G418S, E457V, A471S, K472Q + A deletion has changed the ORF so that 29 additional amino acids (NLDSGDPWASWAFRSLPAFQSGNLSCQLH) added to the C-terminal
5th lpd43 9.8.15	=	T363I, T378I + A deletion changed the ORF so that 29 additional amino acids (NLDSGDPWASWAFRSLPAFQSGNLSCQLH) added to the C-terminal
5th lpd44 9.8.15	=	Q200H, D402G, G461C + A deletion changed the ORF so that 29 additional amino acids (NLDSGDPWASWAFRSLPAFQSGNLSCQLH) added to the C-terminal
5th lpd46 9.8.15	=	I186T, V274I
5th lpd47 9.8.15	=	A257S, G337D
5th lpd49 9.8.15	=	V219I, I226N
5th lpd50 9.8.15	=	A213P, R296H, I438T
5th lpd52 9.8.15	=	V174M, V204L, E366D
5th lpd54 9.8.15	=	T303I
5th lpd55 9.8.15	=	T397I, I429V
5th lpd57 9.8.15	=	I218T, G293D, E420G
5th lpd59 9.8.15	=	V210A, I218T, K220N, A259E, V267I, L454M, T464P, D465I, L466C, P467R, N468I, P469R, A471P, K474I + two deletions changed the ORF so that one additional amino acid (I) added to the C-terminal
5th lpd60 9.8.15	=	M206K, T353A, G419S
5th lpd61 9.8.15	=	V180I, I185F, P261S, V287F, N304K, T441S
5th lpd62 9.8.15	=	A358T, A214V
5th lpd64 9.8.15	=	I388F, C433S
5th lpd65 9.8.15	=	V204M, D291N, E420V
5th lpd66 9.8.15	=	M206R, R225H, A266V, V289I, H322Y
5th lpd67 9.8.15	=	P258S, I388T
5th lpd68 9.8.15	=	D208E, V241L, G337S, E369K, K472Q
5th lpd69 9.8.15	Active PDH and very weak ODH	R302H, G328S, P380L, T397I
5th lpd70 9.8.15	Active PDH and ODH	T441S
5th lpd71 9.8.15	=	P258S
5th lpd72 9.8.15	=	R292S, A439V + A deletion has changed the ORF so that 29 additional amino acids (NLDSGDPWASWAFRSLPAFQSGNLSCQLH) added to the C-terminal
5th lpd73 9.8.15	Active PDH and partially active ODH	V193I, A351E
5th lpd74 9.8.15	Active PDH and ODH	I218T

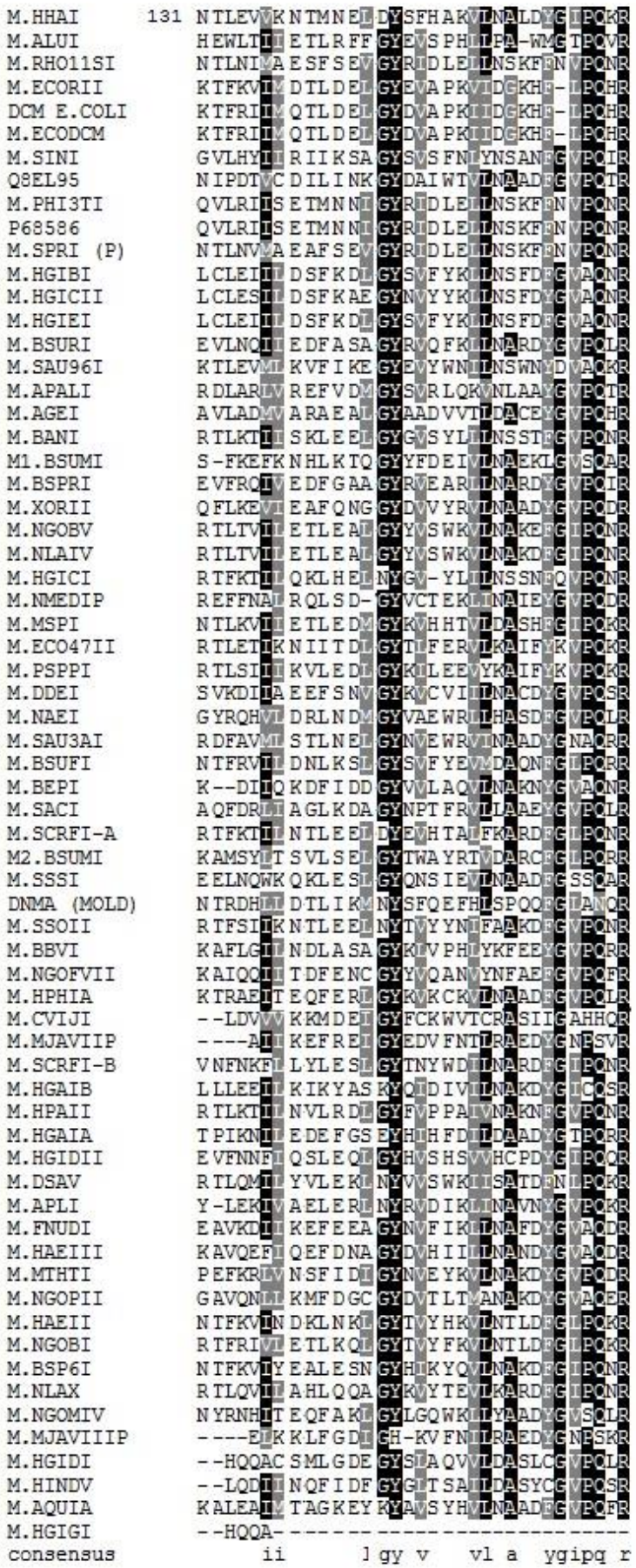
5th lpd75 9.8.15	Active PDH and ODH	R302H, V360G + an insertion has changed the ORF so that two additional amino acids (II) added to the C-terminal
5th lpd76 9.8.15	=	M181I, R292H, H340D
5th lpd78 9.8.15	=	M190I, V287F
5th lpd79 9.8.15	=	M206T, V357I, I335T, T378S
5th lpd80 9.8.15	=	V203A, K368E, E404G, A456S
5th lpd81 9.8.15	=	D437N + A deletion has changed the ORF so that 29 additional amino acids (NLDSGDPWASWAFRSLPAFQSGNLSCQLH) added to the C-terminal
5th lpd83 9.8.15	=	E205Q
5th lpd84 9.8.15	=	Y194F, V221A, H326Y, T353I, T447A
5th lpd85 9.8.15	=	A439T
5th lpd86 9.8.15	=	A214V, K224T, K368N + A deletion has changed the ORF so that 29 additional amino acids (NLDSGDPWASWAFRSLPAFQSGNLSCQLH) added to the C-terminal
5th lpd87 9.8.15	=	K228T, Q317H, A321S, M396T
5th lpd88 9.8.15	=	I211T, D291Y, R302C
5th lpd90 9.8.15	=	A240T, T397S
5th lpd93 9.8.15	Active PDH and partially active ODH	G415D
5th lpd94 9.8.15	Active PDH and ODH	A336V, S405N
5th lpd96 9.8.15	=	P461L, Q317H, H326Q
5th lpd97 9.8.15	=	G385S, T397I
5th lpd98 9.8.15	partially active PDH and partially active ODH	R273H, V357I, K365Q, G426D
5th lpd99 9.8.15	Active PDH and ODH	H307R, F379V, G432D
5th lpd100 9.8.15	=	V193I
5th lpd102 9.8.15	=	V210I, H326Y, K368E
5th lpd103 9.8.15	=	A214T, M233L, M252T, K278I, I335V
5th lpd104 9.8.15	=	I218N, M431V
5th lpd105 9.8.15	=	E356D
5th lpd106 9.8.15	=	L427M
5th lpd107 9.8.15	=	A deletion has changed the ORF so that 29 additional amino acids (NLDSGDPWASWAFRSLPAFQSGNLSCQLH) added to the C-terminal
5th lpd108 9.8.15	=	V250I, R263C
5th lpd109 9.8.15	=	P261L
5th lpd110 9.8.15	=	V346I, G385C

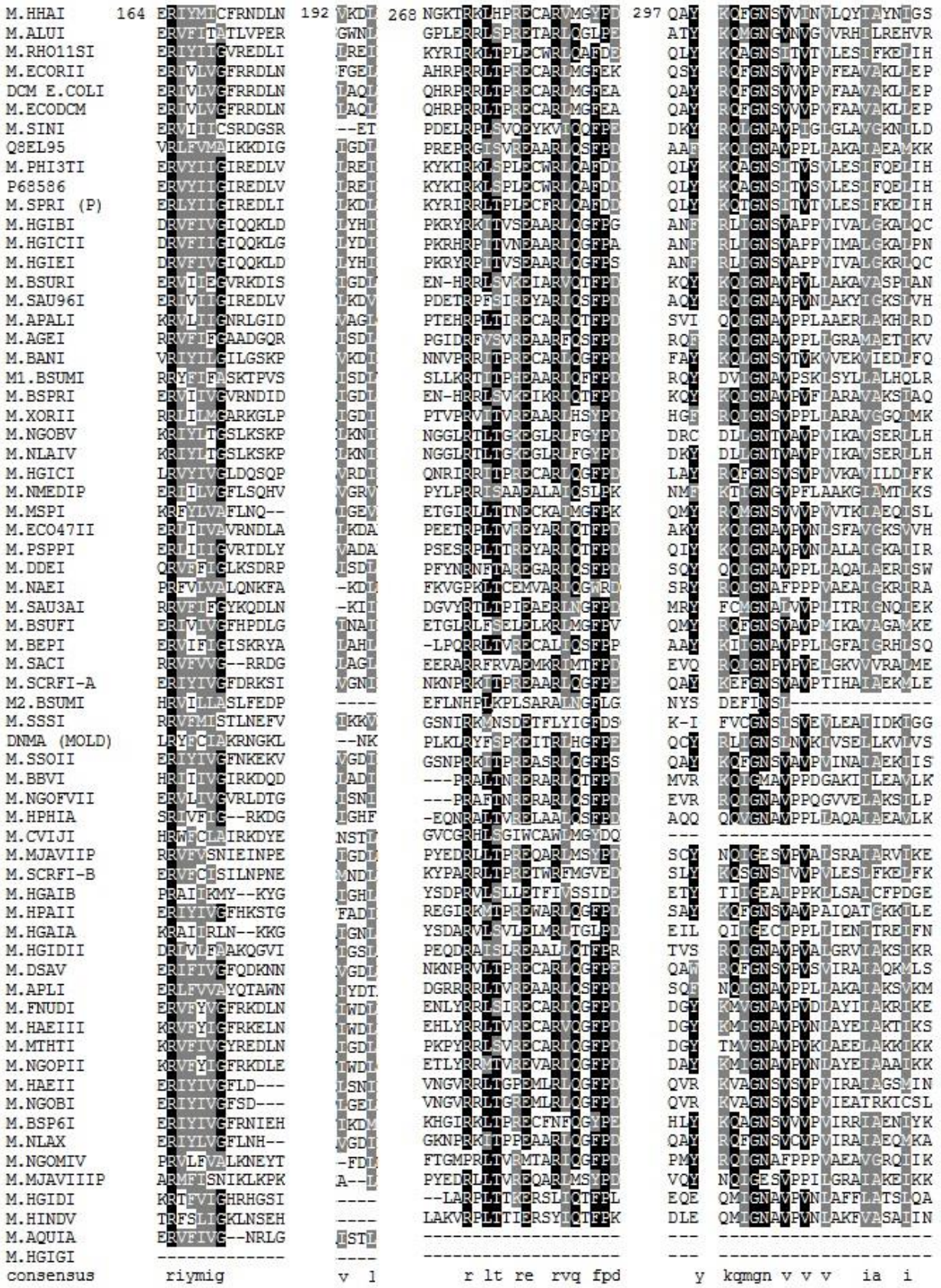
*****E3 mutant clones of affected ODH and/or PDH are yellow shaded.

Appendix 4: Multi sequence amino acids alignments of some bacterial C5-Mtases.

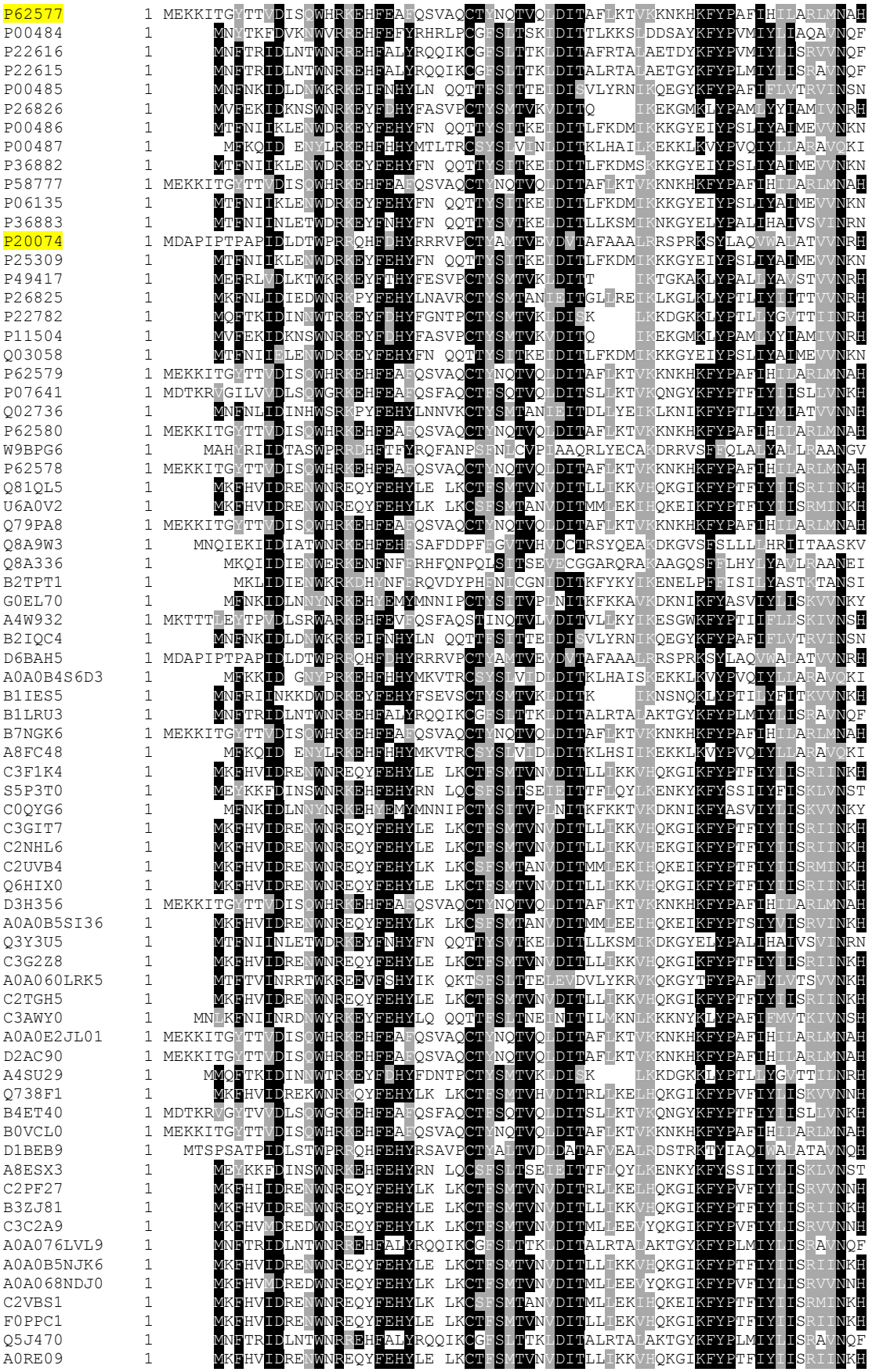


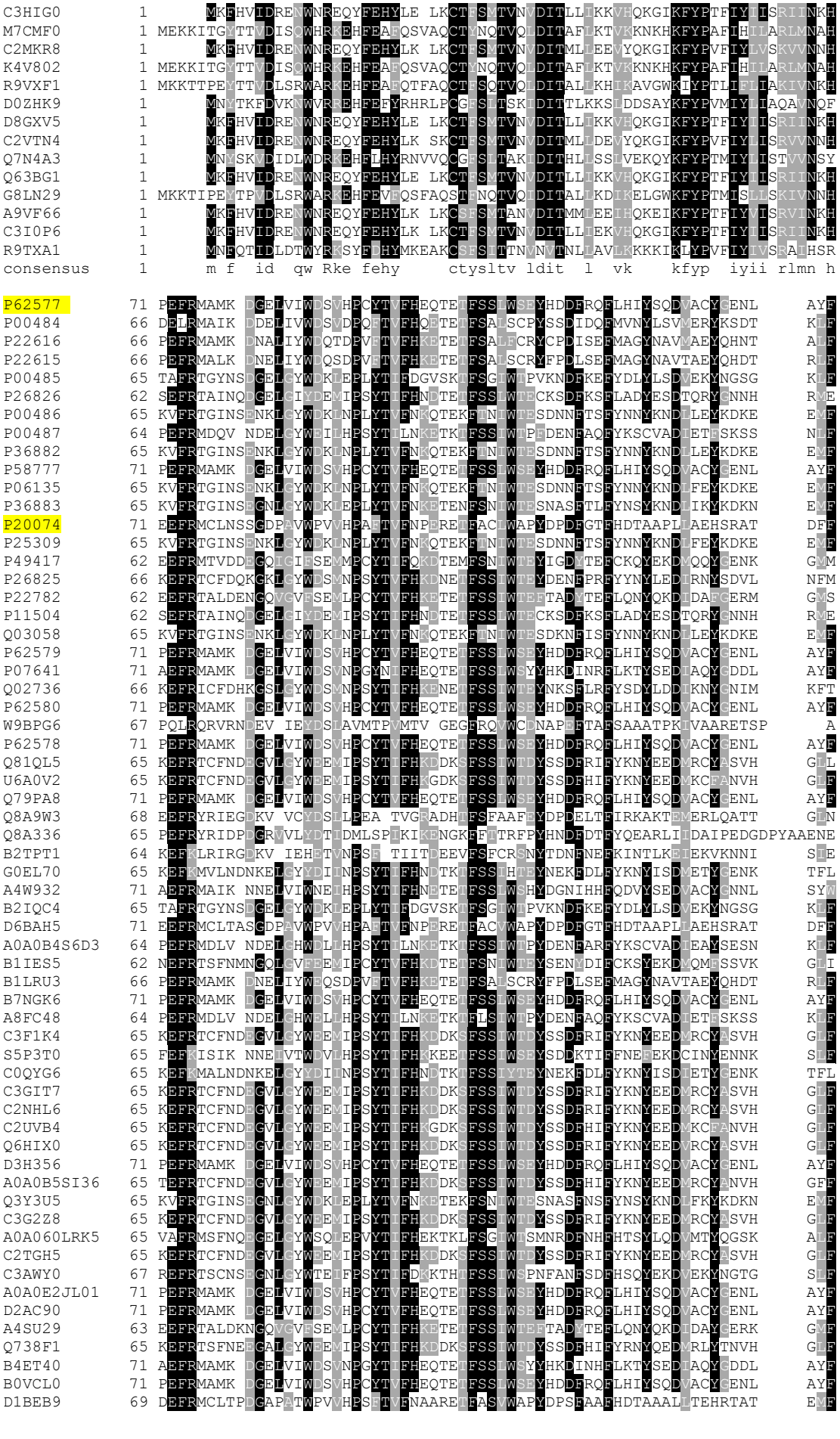

64 73
M.HHAI 60 DI V DILCAGFPCQAFSISG
M.ALUI DI D DVLTGGFPCQPFSKSG
M.RHO11SI DV I DLLVGGSPCQSFSVAG
M.ECORII DA Y DVLLAGFPCQPFSLAG
DCM E.COLI AA H DVLLAGFPCQPFSLAG
M.ECODCM AA H DVLLAGFPCQPFSLAG
M.SINI IR T DLIMGGPPCQAFSTAG
Q8EL95 DI L LIVIGGPPCQAWSKIG
M.PHI3TI DI A DLLTSGFPCPTFSVAG
M.PHI3TI (P) DI A DLLTSGFPCPTFSVAG
M.SPRI (P) DV I DLLVGGSPCQSFSVAG
M.HGIBI DI I DVLVGGVPCQPWSIAG
M.HGICII DI L DLVVGGVPCQPWSIAG
M.HGIEI DI I DVLVGGVPCQPWSIAG
M.BSURI DI V NLILGGFPCPGFSEAG
M.SAU96I DI I DILSGGYPCQTFSYAG
M.APALI DL V FVVIGGPPCQGFSTAG
M.AGEI SI V DVVMGGPPCQGFSTYG
M.BANI DI I DFLLAGFPCQPFSYAG
M1.BSUMI SI H DFILAGPPCQGHSDLN
M.BSPRI DI I NLVIGGFPCPGFSEAG
M.XORII SV V DIVIGGAPCQGFSLIG
M.NGOBV DI I DILLAGFPCQAFSFAG
M.NLAIV DI I DILLAGFPCQAFSFAG
M.HGICI DI I DFLLAGFPCQPFSYAG
M.NMEDIP SI C VGFIGGPPCPDFSIAG
M.MSPI DI V DILCAGFPCQPFSHIG
M.ECO47II DV V DVLAGGFPCQAFSYAG
M.PSPPI DI V DLLIGGFPCQPFSYAG
M.DDEI DI L DGIIGGPPCQGFSLSG
M.NAEI DV P SLIAGGVPCPPFSIAG
M.SAU3AI DI V DMIVGGFPCQDYSVAR
M.BSUFI DI I DVLLAGFPCQPFSNIG
M.BEPI SI V DVVTGGFPCQDFSFAG
M.SACI DI I TLVIGGPPCQTFSFKSG
M.SCRFI-A DI I DILVGGFPCQAFSQAG
M2.BSUMI DI I DLVAAGFPCQDLSQAG
M.SSSI DI L DLLTYSFPCQDLSQQG
DNMA (MOLD) SI Y NAWDMSPPCQPFTRLG
M.SSOII DI I EILVGGFPCVAFSQAG
M.BBVI DV L DAFTFGFPCNDYSIVG
M.NGOFVII DI I DIILGGFPCQDFSMIW
M.HPHIA DL I DVVIGGPPCQGFSMAG
M.CVIJI DV F DMITAGFPCTGFSIAG
M.MJAVIIP DI I DVIIGGPPCEGYIGAN
M.SCRFI-B DI I DFFTYSFPCQDISVAG
M.HGAIB DI E KFLLATPPCQGLSSVG
M.HPAII DI E DILCAGFPCQAFSIAG
M.HGAIA DI Q DFLIASPPCQGMSVAG
M.HGIDII SI V KILVGCAPCQDFSQYT
M.DSAV DI L DLLLAGFPCQAFSQGG
M.APLI TL G DVIIGGPPCQPFSVGG
M.FNUDI DI I DGIIGGPPCQSWSEAG
M.HAEIII DI I DGIIGGPPCQSWSEGG
M.MTHTI PI L VGFIGGPPCQSWSLAG
M.NGOPII DI I DGIIGGPPCQSWSEAG
M.HAEII DI I DILLAGFPCQPFSIAG
M.NGOBI DI I DILLAGFPCQPFSIAG
M.BSP6I DI I DVLLSGFPCTSFSVAG
M.NLAX DI T DIISAGFPCQPFSQAG
M.NGOMIV DV - DLIAGGVPCPPFSKAG
M.MJAVIIIP DI I DVIIGSPPCEPFTKAN
M.HGIDI DL I ELIIGGPPCQDFSSAG
M.HINDV DL E DLIMGGPPCQDFSSAG
M.AQUIA DI I DLVIGGPPCQSFSLAG
M.HGIGI DL V ELIIGGPPCQDFSSAG
consensus di v dilvag PCq fs ag
M.HHAI 96 SRG 100 LFFDIARIVRKKKPKVVFMENVKNFASH
M.ALUI TRG LFWNIARIIEERERPTVLIIENVRNLVGP
M.RHO11SI TRG LFFQYIDTLKEKQPRYFVFENVKGLINH
M.ECORII AQG LFFDVARIIRAKKRAIFVLENVKNIKSH
DCM E.COLI TQG LFFDVVRIIDARRRAMFVLENVKNIKSH
M.ECODCM TQG LFFDVVRIIDARRRAMFVLENVKNIKSH
M.SINI ERG VFIKYLDVAL--RPKYIVIENVRGLLSA
Q8EL95 ARG LYENFLDYAL--DANVIVMENVPEAMNY
M.PHI3TI ERG LFFETALIAEEKKPKFVILENVKGLINS
P68586 ERG LFFETALIAEEKKPKFVILENVKGLINS
M.SPRI (P) TRG LFFQYVETLKEKQPKFFVFENVKGLINH
M.HGIBI PRG LWADVIRLVQINQPKAFIFENVKGLVDP
M.HGICII PRG LWADVIRLVRINQPKAFIFENVKGLIDP
M.HGIEI PRG LWADVIRLVQINQPKAFIFENVKGLVDP
M.BSURI DRN LYLHFIRSLI--QEEIFVAENVKGMMTL
M.SAU96I TRG LFYPYSKILS--KPKAFIAENVSGLVNH
M.APALI PRN LIFNYLNIVE--SPRWLIFENVEGLLTS
M.AGEI ARN LYVPYFGFVE--RPKAFLLENVVGLLSM
M.BANI TRG LFFEVERVLRDNRPKAFLLENVRGLVTH
M1.BSUMI PRN LLMRVSRVIE--QESSVLVENVPGIIHD
M.BSPRI ERN LYIHFIRCLM--QEEIFVAENVKGMMTL
M.XORII SRN LVHHYVRVVM--KPKYFVFENVKGLTVG
M.NGOBV TRG LFFDVARILKAKKPKGFLIENVEGLVTH
M.NLAIV TRG LFFDVARILKAKKPKGFLIENVEGLVTH
M.HGICI TRG LFFEIERILKAYRPKGFLLENVRGLTTH
M.NMEDIP ENG LSQSYVDLIC--QEDFFVFENVKGLYRT
M.MSPI TQG MFHEIVRIIETKKTPVLFLENVEGLLNH
M.ECO47II TRG LFFEFARAAK--NPRVLIAENVRGLLNH
M.PSPPI LRG LVFEMARAIK--KPWVFIAENVKGLAEN
M.DDEI PRN LFVDFVRFVK--SPKFFVMENVLGLLSM
M.NAEI MRD LFAWAVELCD--KPRALMLENVEGLSMP
M.SAU3AI KKG LFWQIIRYLQNTFPPKYLLLENVDRLLKS
M.BSUFI ERR IIFDVLRILKKKQPKMFLLENVKGLLTN
M.BEPI TRG LYLWLKKVVE--KPKVFIAENVKGLVTL
M.SACI PNA LLDEYVRVVR--KPEAFIIENVQGLTYK
M.SCRFI-A TRG LFFEIARIIKEKRPKAFLLENVKNLKTH
M2.BSUMI SRS LFELIEKKEHANRPPWILIENVPYMVRL
M.SSSI TRS LLWEIERALDNDLPKYLLMENVGALLHK
DNMA (MOLD) NRT SFFHLLDVLTKDPPTYILIENVFGFAKK
M.SSOII TRG LFFDIARIIKEKKPHAFLLENVKNLIGH
M.BBVI NYG LYSYGVKILN--NPLVFIAENVGGLQSA
M.NGOFVII ERG LYKSFLRFVN--KPKVFVAENVKGLLTA
M.HPHIA PRN LFKEFVRVVK--QPKFFVMENVARIFTH
M.CVIJI KES GLFADVVRITEEYKPKIVFLENSHMISHT
M.MJAVIIP ETG RLVLEYIRIVG--QPKIFVMENVPGLKEV
M.SCRFI-B TRS SLLWECCKIIEHKKPKYLMMENVKNLVGK
M.HGAIB NRN FLIFEVFEFID--NLDFILIENVPRFIEM
M.HPAII TRG TLFFDVAEIIRRHQPKAFLLENVKGLKNH
M.HGAIA NRN YLIMYVIAMIK--KPAYILIENVPFLLKL
M.HGIDII TKW QLLTEFSRLIR--EPDIISMENVPEVRTF
M.DSAV ERG QLFFQVAKILNDHRPQAILLENVKGLRGH
M.APLI DSR DGLPIFIDAIA--QPEIAIFENVRGMLYK
M.FNUDI PRG KLFYEYIRILK--QPKFFLAENVKGMLSK
M.HAEIII PRG KLFYEYIRILK--KPEIFLAENVKGMMAQ
M.MTHTI PRG KTFYAYVDLVK--DPLFFLAENVPGIVSR
M.NGOPII ARG QLFFDYIRILK--QPKFFLAENVSGMLAN
M.HAEII TRG TLFFNIEAILKAKKPKAFLLENVKRLTTH
M.NGOBI TRG TLFFNIAEILKTKQPKAFLLENVKRLTTH
M.BSP6I KSG DLFFETLRLIVAKKPQVIFLENVKNLVGH
M.NLAX TRG TLFFDIERILLAKKPQAFLLENVKQLKGH
M.NGOMIV ERD -LFPEAIRLAK--DPKAIMLENVRGLLDP
M.MJAVIIIP KVG RLVLYYIDYVNRNDDLIFVMENVPQLKEI
M.HGIDI GRA NLTLDFAKIVL--QPAWVIMENVERARLS
M.HINDV GRA DLTYSFANIVC--RPKWFVMENVEQLKKS
M.AQUIA PRG MLVLEFLRVVR--DPKCFVMENVKGMINW
M.HGIGI GRA NLTLDFVKIVL--QPAWVIMENVERARLS
consensus rg lf riv pk fimENvkglv

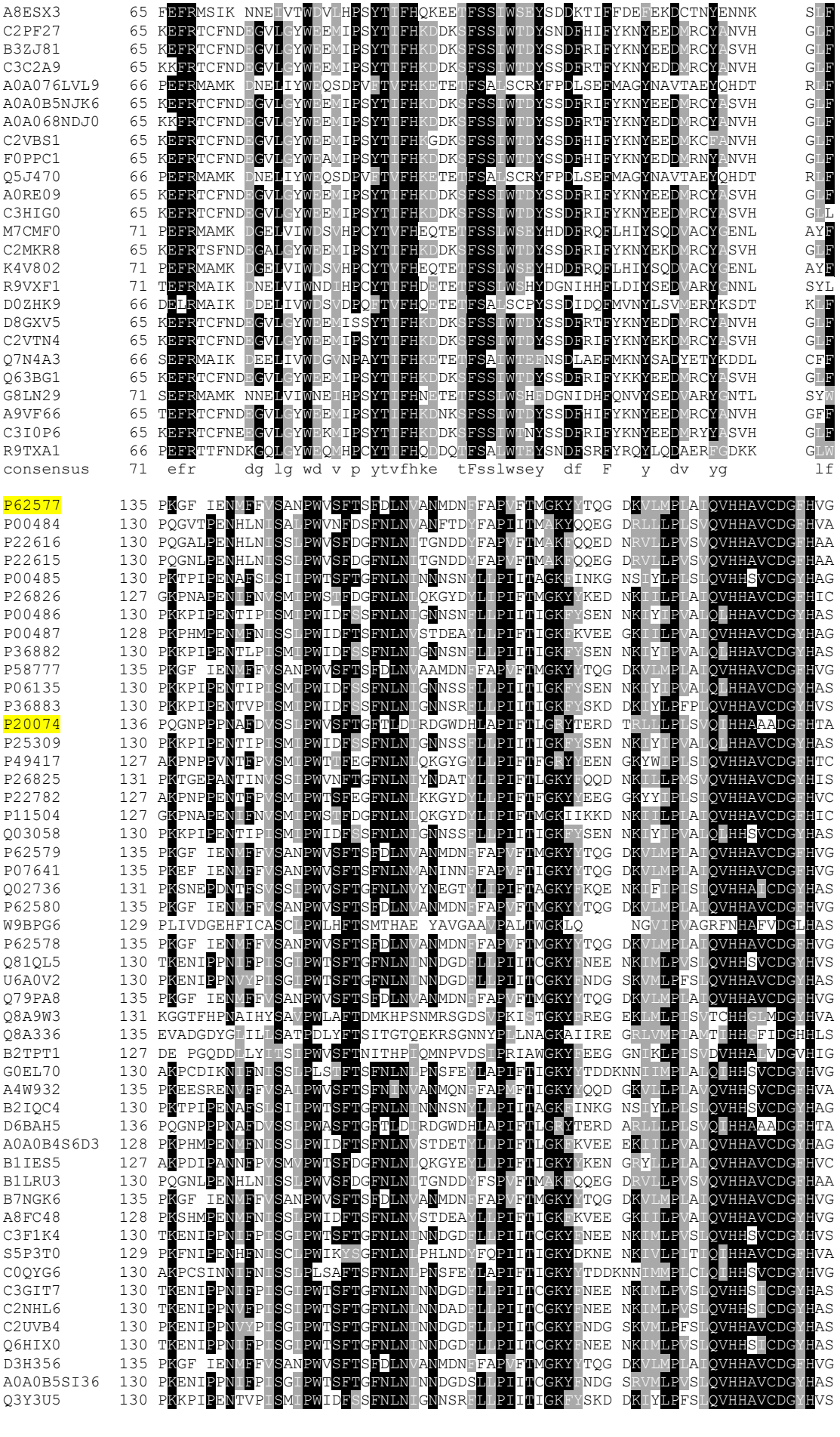

M.HHAI 131 NTLEVVKNTMNELDYSFHAKVLNALDYGIPQKR
M.ALUI HEWLTIIETLRFFGYEVSPHLLPA-WMGTPQVR
M.RHO11SI NTLNIVAESFSEVGYRIDLELLNSKFPNVPQNR
M.ECORII KTFKVIVDTLDELGYEVAPKVIDGKHP-LPQHR
DCM E.COLI KTFRIIVQTLDELGYDVAPKIIDGKHP-LPQHR
M.ECODCM KTFRIIVQTLDELGYDVAPKIIDGKHP-LPQHR
M.SINI GVLHYIIRIIKSAGYSVSFNLYNSANFGVPQIR
Q8EL95 NIPDTVCDILINKGYDAIWTVLNAADFGVPQIR
M.PHI3TI QVLRIISETMNNIGYRIDLELLNSKFPNVPQNR
P68586 QVLRIISETMNNIGYRIDLELLNSKFPNVPQNR
M.SPRI (P) NTLNVVAEAFSEVGYRIDLELLNSKFPNVPQNR
M.HGIBI LCLEIILDSFKDLGYSVFYKLLNSFDFGVAQNR
M.HGICII LCLESILDSFKAEGYNVYYKLLNSFDYGVAQNR
M.HGIEI LCLEIILDSFKDLGYSVFYKLLNSFDFGVAQNR
M.BSURI EVLNQIIEDFASAGYRVQFKLLNARDYGVPQLR
M.SAU96I KTLEVMLKVFIKEGYEVYWNLLNSWNYDVAQKR
M.APALI RDLARLVREFVDMGYSVRLQKVNLAAYGVPQTR
M.AGEI AVLADMVARAEALGYAADVVTLDACEYGVPQHR
M.BANI RTLKTIISKLEELGYGVSYLLLNSSTFGVPQNR
M1.BSUMI S-FKEFKNHLKTQGYYFDEIVLNAEKLGVSQAR
M.BSPRI EVFRQIVEDFGAAGYRVEARLLNARDYGVPQIR
M.XORII QFLKEVIEAFQNGGYDVVYRVLNAADYGVPQDR
M.NGOBV RTLTVILETLEALGYYVSWKVLNAKEFGIPQNR
M.NLAIV RTLTVILETLEALGYYVSWKVLNAKDFGIPQNR
M.HGICI RTFKTILQKLHELNYGV-YLLLNSSNFQVPQNR
M.NMEDIP REFFNALRQLSD-GYVCTEKLLNAIEYGVPQDR
M.MSPI NTLKVILETLEDMGYKVHHTVLDASHFGIPQKR
M.ECO47II RTLETIKNIITDLGYTLFERVLKAIFYKVPQKR
M.PSPPI RTLSIIIKVLEDLGYKILEEVYKAIFYKVPQKR
M.DDEI SVKDIIAEEFSNVGYKVCVIILNACDYGVPQSR
M.NAEI GYRQHVLDRLNDMGYVAEWRLLHASDFGVPQLR
M.SAU3AI RDFAVMLSTLNELGYNVEWRVINAADYGNAQRR
M.BSUFI NTFRVILDNLKSLGYSVFYEVVDAQNFGLPQRR
M.BEPI K--DIIQKDFIDDGYVVLAQVLNAKNYGVAQNR
M.SACI AQFDRLIAGLKDAGYNPTFRVLLAAEKGVPQLR
M.SCRFI-A RTFKTILNTLEELDYEVHTALFKARDFGLPQNR
M2.BSUMI KAMSYLTSVLSELGYTWAYRIVDARCFGLPQRR
M.SSSI EELNQWKQKLESLGYQNSIEVLNAADFGSSQAR
DNMA (MOLD) NTRDHLLDTLIKMNYSFQEFHLSPQQFGLANQR
M.SSOII RTFSIIKNTLEELNYTVYYNIFAAKDFGVPQNR
M.BBVI KAFLGIINDLASAGYKLVPHLYKFEEYGVPQRR
M.NGOFVII KAIQQIITDFENCGYYVQANVYNFAEFGVPQFR
M.HPHIA KTRAEITEQFERLGYKVKCKVLNAADFGVPQLR
M.CVIJI --LDVVVKKMDEIGYFCKWVICRASIIGAHHQR
M.MJAVIIP ----AIIKEFREIGYEDVFNIIRAEDYGNPSVR
M.SCRFI-B VNFNKELLYLESLGYTNYWDILNARDFGIPQNR
M.HGAIB LLLEEILKIKYASKYQIDIVLLNAKDYGICQSR
M.HPAII RTLKTILNVLRDLGYFVPPAIVNAKNFGVPQNR
M.HGAIA TPIKNILEDEFGSEYHIHFDILDAADYGTPQRR
M.HGIDII EVFNNFLQSLEQLGYHVSHSVVHCPDYGIPQQR
M.DSAV RTLQMILYVLEKLNYVVSWKIISATDFNLPQKR
M.APLI Y-LEKIVAELERLNYRVDIKLLNAVNYGVPQKR
M.FNUDI EAVKDILKEFEEAGYNVFIKLLNAFDYGVAQDR
M.HAEIII KAVQEFLQEFDNAGYDVHIILLNANDYGVAQDR
M.MTHTI PEFKRLVNSFIDIGYNVEYKVLNAKDYGVPQDR
M.NGOPII GAVQNLLKMFDGCGYDVTLTVANAKDYGVAQER
M.HAEII NTFKVINDKLNKLGYTVYHKVLNTLDFGLPQKR
M.NGOBI RTFRIVLETLKQLGYTVYFKVLNTLDFGLPQKR
M.BSP6I NTFKVIYEALESNGYHIKYQVLNAKDFGIPQNR
M.NLAX RTLQVILAHLQQAGYKVYTEVLKARDFGIPQNR
M.NGOMIV NYRNHLTEQFAKLGYLGQWKLIYAADYGVSQLR
M.MJAVIIIP ----ELKKLFGDIGH-KVFNIIRAEDYGNPSKR
M.HGIDI --HQQACSMLGDEGYSLAQVVLDASLCGVPQLR
M.HINDV --LQDTINQFIDFGYGLTSAILDASYCGVPQSR
M.AQUIA KALEAIVTAGKEYKYAVSYHVLNAADFGVPQFR
M.HGIGI --HQQA----------
consensus ii]gy v vl a ygipq r

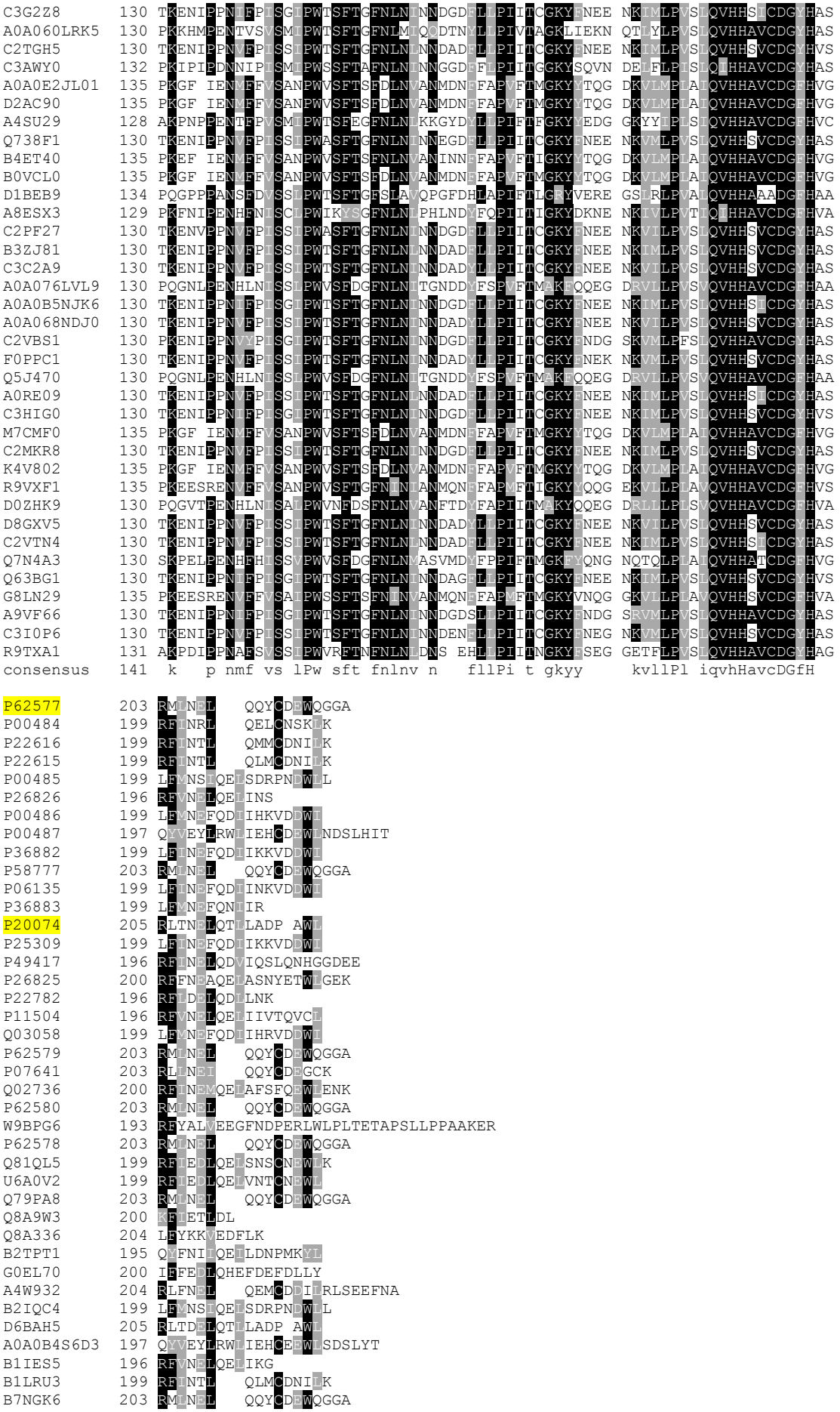

M.HHAI 164 ERIYMICFRNDLN 192 VKDL 268 NGKTRKTHPRECARVMGYPD 297 QAY KQFGNSVVINVLQYIAYNIGS
M.ALUI ERVFITATLVPER GWNL GPLERRLSPRETARLQGLPE ATY KQMGNGVNVGVVRHILREHVR
M.RHO11SI ERIYIIGVREDLI IREI KYRIRKLTPLECWRLQAPDD QLY KQAGNSITVTVLESIFKEIIH
M.ECORII ERIVLVGFRRDLN FGEL AHRPRRLTPRECARLMGPEK QSY RQFGNSVVVPVFEAVAKLVEP
DCM E.COLI ERIVLVGFRRDLN IAQL QHRPRRLTPRECARLMGPEA QAY RQFGNSVVVPVFAAVAKLVEP
M.ECODCM ERIVLVGFRRDLN IAQL QHRPRRLTPRECARLMGPEA QAY RQFGNSVVVPVFAAVAKLVEP
M.SINI ERVIIICSRDGSR --ET PDELRPLSVQEYKVIQQFPD DKY RQLGNAVPIGLGLAVGKNILD
Q8EL95 VRLFVMAIKKDIG IGDI PREPRGISVREAARLQSFPD AAF KQIGNAVPPLLAKAIAEAVKK
M.PHI3TI ERVYIIGIREDLV IREI KYKIRKLSPLECWRLQAPDD QLY KQAGNSITVSVLESIFQEIIH
P68586 ERVYIIGIREDLV IREI KYKIRKLSPLECWRLQAPDD QLY KQAGNSITVSVLESIFQEIIH
M.SPRI (P) ERLYIIGIREDLI IKDL KYRIRRLTPLECFRLQAPDL QLY KQIGNSITVTVLESIFKEIIH
M.HGIBI DRVFIVGIQQKLD IYHI PKRYRKITVSEAARLQGFPG ANF RLIGNSVAPPVIVALGKAIQC
M.HGICII DRVFIIGIQQKLG IYDI PKRHRPITVNEAARIQGFPA ANF RLIGNSVAPPVIMALGKAIPN
M.HGIEI DRVFIVGIQQKLD IYHI PKRYRPITVSEAARLQGFES ANF RLIGNSVAPPVIVALGKRIQC
M.BSURI ERVIIEGVRKDIS IGDI EN-HRRLSVKEIARVQTFPD KQY KQIGNAVPVLLAKAVASPIAN
M.SAU96I ERIVIIGIREDLV IKDV PDETRPFSIREYARIQSFPD AQY RQIGNAVPVNLAKYIGKSIVH
M.APALI KRVLIIGNRLGID VAGL PTEHRPLTIRECARIQTFPD SVI QQIGNAVPPLAAERLAKHIRD
M.AGEI RRVFIFGAADGQR ISDL PGIDRRFVSVREAARFQSFPD RQF RQIGNAVPPLLGRAMAETIKV
M.BANI VRIYIILGILGSKP VKDI NNVPRRITPRECARLQGFPD FAY KQLGNSVTVKVVEKVIEDIFQ
M1.BSUMI RRYFIFASKTPVS ISDI SLLKRTITPHEAARIQTFPD RQY DVIGNAVPSKLSYLLALHQLR
M.BSPRI ERVIIVGVRNDID IGDI EN-HRRLSVKEIKRIQTFPD KQY KQLGNAVPVFLARAVAKSIAQ
M.XORII RRLILMGARKGLP IGDI PTVPRVITVREAARIHSVPD HGF RQLGNSVPPLLARAVGGQIMK
M.NGOBV KRIYLTGSLKSKP LKNI NGGLRTLTGKEGLRLFGVPD DRC DLLGNTVAVPVIKAVSERILH
M.NLAIV KRIYLTGSLKSKP LKNI NGGLRTLTGKEGLRLFGVPD DKY DLLGNTVAVPVIKAVSERILH
M.HGICI LRVYIVGLDQSQP VRDI QNRIRRITPRECARLQGFPD LAY RQFGNSVSVPVVKAVILDIFK
M.NMEDIP ERIILVGFLSQHV VGRV PYLPRRISAADALAIQSLPK NMF KIIGNGVPFLAAKGIAMTIKS
M.MSPI KRFYLVAFLNQ-- IGEV ETGIRLLTINECKAIMGFPK QMY RQMGNSVVVPVVTKIAEQISL
M.ECO47II ERLIIVAVRNDLA IKDA PEETRPLTVREYARIQTFPD AKY KQIGNAVPVNLSFAVGKSVVH
M.PSPPI ERLIIIGVRTDLY VADA PSESRPLTIREYARIQTFPD QIY KQIGNAVPVNLALAIGKAIIR
M.DDEI QRVFFIGLKSDRP ISDL PFYNRNFTAREGARIQSFPD SQY QQLGNAVPPLLAQALAERISW
M.NAEI PRFVLVALQNKFA -KDL FKVGPKLTCEMVARIQGVRD SRY RQIGNAFPPPVAEAIGKRIRA
M.SAU3AI RRVFIFGYKQDLN -KII DGVYRRLTIPIEAERLNGFEV MRY FQMGNALVVPLITRIGNQIEK
M.BSUFI ERIVIVGFHPDLG INAI ETGLRLPSELELKRLMGFEV QMY RQFGNSVAVPMIKAVAGAVKE
M.BEPI ERVIFIGISKRYA IAHL -LPQRRLTVRECALIQSFEP AAY KIIGNAVPPLLGFAIGRHISQ
M.SACI RRVFVVG--RRDG IAGL EERARRFRVAEMKRIMIFPD EVQ RQIGNPVPVELGKVVRAIME
M.SCRFI-A ERIYIVGFDRKSI VGNI NKNPRKITPREAARLQGFPE QAY KEFGNSVAVPTIHAIAEKVLE
M2.BSUMI HRVILLASLFEDP ---- EFLNHEIKPLSARALNGELG NYS DEFINSL----------
M.SSSI RRVFMISTLNEFV IKKV GSNIRKVNSDETFLYIGEDS K-I FVCGNSISVEVLEAIIDKIGG
DNMA (MOLD) LRYFCIAKRNGKL --NK PLKLRYFSPKEITRLHGFPE QCY RLIGNSLNVKIVSELLKVLVS
M.SSOII ERIYIVGFNKEKV VGDI GSNPRKITPREASRLQGFES QAY KQFGNSVAVPVINAIAEKIIS
M.BBVI HRIIIVGIRKDQD IADI ---PRALTNRERARLQTFPD MVR KQIGMAVPPDGAKIILEAVLK
M.NGOFVII ERVLIVGVRLDTG ISNI ---PRAFTNRERARLQSFPD EVR RQIGNAVPPQGVVELAKSILP
M.HPHIA SRIVFIG--RKDG IGHF -EQNRALTVRELAALQSFPD AQQ QQVGNAVPPLIAQAAIAEAVLK
M.CVIJI HRWFCLAIRKDYE NSTI GVCGRRLSGIWCAWIMGVDQ --- ----------
M.MJAVIIP RRVFVSNIEINPE IGDI PYEDRLLTPREQARLMSVPD SCY NQIGESVPVAISRAIARVIKE
M.SCRFI-B ERVFCIISILNPNE VNDI KYPARRLTPRETWRFMGVED SLY KQSGNSIVVPVLESLFKEIFK
M.HGAIB PRAITKMY--KYG IGHL YSDPRVLSLLETFIVSSIDE ETY TIIGEAIPPKLLSAICFPDGE
M.HPAII ERIYIVGFHKSTG FADI REGIRKMTPREWARLQGFPD SAY KQFGNSVAVPAIQATGKKILE
M.HGAIA KRAIIRLN--KKG IGNI YSDARVLSVLELMRLTGLPD EIL QIIGECIPPLIIENTTREIFN
M.HGIDII DRIVLFPAKQGVI TGSL PEQDRAISLREAALLQTFER TVS RQIGNAVPVALGRVIAKSIKR
M.DSAV ERIFIVGFQDKNN VGDI NKNPRVLTPRECARLQGFPE QAV RQFGNSVPVSVIRAIAQKVLS
M.APLI ERLFVVAYQTAWN IYDT DGRRRRLTVRECAARLQSFPD SQF NQIGNAVPPLLAKAIAKSVKM
M.FNUDI ERVFYVGFRKDLN IWDI ENLYRRLSIRECARIQGFPD DGY KMVGNAVPVDLAYIIAKRIKE
M.HAEIII KRVFYIGFRKELN IWDI EHLYRRLTVRECARVQGFPD DGY KMIGNAVPVNLAYEIAKTIKS
M.MTHTI KRVFIVGYREDLN IGDI PKPYRRLSVRECARIQGFPD DGY TMVGNAVPVKLAEELAKKIKK
M.NGOPII KRVFYIGFRKDLE IWDI ETLYRRLTVREVARIQGFPD DAY KMIGNAVPVNLAYEIAAAIKK
M.HAEII ERIYIVGFLD--- ISNI VNGVRRLTGPEMLRLQGFPD QVR KVAGNSVSVPVIRAIAGSVIN
M.NGOBI ERIYIVGFSD--- IGEI VNGVRRLTGREMLRLQGFPD QVR KVAGNSVSVPVIEATRKICSL
M.BSP6I ERIYIVGFRNIEH IKDV KHGIRKLTPRECFNFQGVPE HLY KQAGNSVVVPVIRRIAENIYK
M.NLAX ERIYLVGFLNH-- VGDI GKNPRRITPPEAARLQGFPD QAY RQFGNSVCVPVIRAIAEQVKA
M.NGOMIV PRVLFVALKNEYT -FDI FTGMPRLTVRMTARIQGFPD PMY RQIGNAFPPPVAEAVGRQIIK
M.MJAVIIIP ARMFISNIKLKPK A--I PYEDRLLTVREQARLMSVPD VQY NQLGESVPPILGRAIAKEIKK
M.HGIDI KRTFVIGHRHGSI ---- --LARPLTTKERSLIQTFEL EQE QMIGNAVPVNLAFFLATSIQA
M.HINDV TRFSLIGKLNSEH ---- LAKVRPLTTIERSYIQTFEK DLE QMIGNAVPVNLAKFVASAIIN
M.AQUIA ERVFIVG--NRLG IGST ---------- --- ----------
M.HGIGI ---------- ---- ---------- --- ----------
consensus riymig v l r lt re rvq fpd y kqmgn v v v ia i

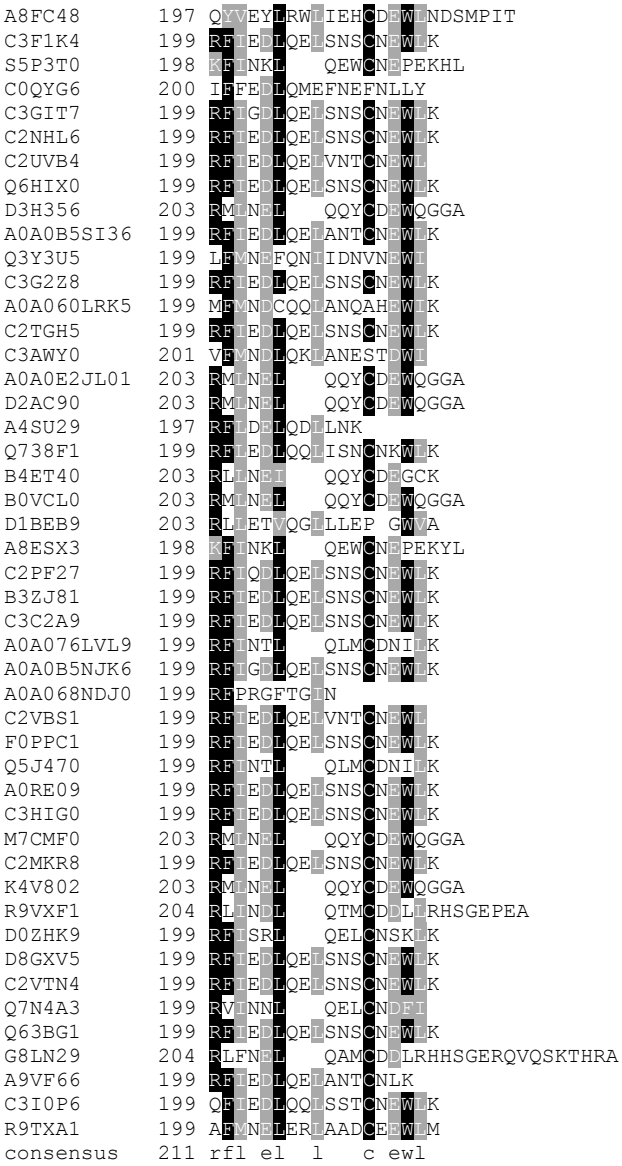
Appendix 5: Multi sequence amino acids alignments of some CATs:




C3HIG0
M7CMF0
C2MKR8
K4V802
R9VXF1
D0ZHK9
D8GXV5
C2VTN4
Q7N4A3
Q63BG1
G8LN29
A9VF66
C3I0P6
R9TXA1
consensus
P62577
P00484
P22616
P22615
P00485
P26826
P00486
P00487
P36882
P58777
P06135
P36883
P20074
P25309
P49417
P26825
P22782
P11504
Q03058
P62579
P07641
Q02736
P62580
W9BPG6
P62578
Q81QL5
U6A0V2
Q79PA8
Q8A9W3
Q8A336
B2TPT1
G0EL70
A4W932
B2IQC4
D6BAH5
A0A0B4S6D3
B1IES5
B1LRU3
B7NGK6
A8FC48
C3F1K4
S5P3T0
C0QYG6
C3GIT7
C2NHL6
C2UVB4
Q6HIX0
D3H356
A0A0B5SI36
Q3Y3U5
C3G2Z8
A0A060LRK5
C2TGH5
C3AWY0
A0A0E2JL01
D2AC90
A4SU29
Q738F1
B4ET40
B0VCL0
D1BEB9


A8ESX3 65
C2PF27 65
B3ZJ81 65
C3C2A9 65
A0A076LVL9 66
A0A0B5NJK6 65
A0A068NDJ0 65
C2VBS1 65
F0PPC1 65
Q5J470 66
A0RE09 65
C3HIG0 65
M7CMF0 71
C2MKR8 65
K4V802 71
R9VXF1 71
D0ZHK9 66
D8GXV5 65
C2VTN4 65
Q7N4A3 66
Q63BG1 65
G8LN29 71
A9VF66 65
C3I0P6 65
R9TXA1 66
consensus 71
P62577 135
P00484 130
P22616 130
P22615 130
P00485 130
P26826 127
P00486 130
P00487 128
P36882 130
P58777 135
P06135 130
P36883 130
P20074 136
P25309 130
P49417 127
P26825 131
P22782 127
P11504 127
Q03058 130
P62579 135
P07641 135
Q02736 131
P62580 135
W9BPG6 129
P62578 135
Q81QL5 130
U6A0V2 130
Q79PA8 135
Q8A9W3 131
Q8A336 135
B2TPT1 127
G0EL70 130
A4W932 135
B2IQC4 130
D6BAH5 136
A0A0B4S6D3 128
B1IES5 127
B1LRU3 130
B7NGK6 135
A8FC48 128
C3F1K4 130
S5P3T0 129
C0QYG6 130
C3GIT7 130
C2NHL6 130
C2UVB4 130
Q6HIX0 130
D3H356 135
A0A0B5SI36 130
Q3Y3U5 130


C3G2Z8 130 TKENIPPNIFPISGIPWTSFTGFNLNINNDGDLLPIITCGKYGNEE NKIMLPVSLQVHHSICDGYHAS
A0A060LRK5 130 PKKHMPENTVSVSMIPWTSFTGFNIMIQQDTNYLLPIVTAGKLIEKN QTLYLPVSLQVHHAVCDGYHAS
C2TGH5 130 TKENIPPNVFPISSIPWTSFTGFNLNLNNDADLLLPIITCGKYGNEE NKIMLPVSLQVHHSVCDGYHVS
C3AWY0 132 PKIPIPDNNIPISMIPWSSFTAFNLNINNGGDGFLPIITGGKYSQVN DELFLPISLQIHHAVCDGYHAS
A0A0E2JL01 135 PKGF IENVFFVSANPWVSFTSFDLNVANMDNGFAPVFTMGKYYTQG DKVLMPLAIQVHHAVCDGFHVG
D2AC90 135 PKGF IENVFFVSANPWVSFTSFDLNVANMDNGFAPVFTMGKYYTQG DKVLMPLAIQVHHAVCDGFHVG
A4SU29 128 AKPNPPENTFPVSMIPWTSFEGFNLNLKKGYDYLLPIFTFGKYYEDG GKYYIPLSLQVHHAVCDGFHVC
Q738F1 130 TKENIPPNVFPISSIPWASFTGFNLNINNEGDGLLPIITCGKYGNEE NKVMLPVSLQVHHSVCDGYHAS
B4ET40 135 PKEF IENVFFVSANPWVSFTSFNLNVANINNGFAPVFTIGKYYTQG DKVLMPLAIQVHHAVCDGFHVG
B0VCL0 135 PKGF IENVFFVSANPWVSFTSFDLNVANMDNGFAPVFTMGKYYTQG DKVLMPLAIQVHHAVCDGFHVG
D1BEB9 134 PQGPPPANSFDVSSLPWTSFTGFSLAVQPGFDHIAPIFTLGRYVERE GSLRLPVAIQVHHAAADGFHAA
A8ESX3 129 PKFNIPENHFNVSCLPWIKYSGFNLNLPHLNDYFQPIITIGKYDKNE NKIVLPVTIQIHHAVCDGFHVA
C2PF27 130 TKENVPPNVFPISSIPWASFTGFNLNINNDGDLLPIITCGKYGNEE NKVILPVSLQVHHSVCDGYHAS
B3ZJ81 130 TKENIPPNVFPISSIPWTSFTGFNLNLNNDADLLLPIITCGKYGNEE NKIMLPVSLQVHHSVCDGYHAS
C3C2A9 130 TKENIPPNVFPISSIPWTSFTGFNLNINNDADYLLPIITCGKYGNEE NKVILPVSLQVHHSVCDGYHAS
A0A076LVL9 130 PQGNLPENHLNISSLPWVSFDGFNLNITGNDDYFSPVFTMAKFQQEG DRVLLPVSVQVHHAVCDGFHAA
A0A0B5NJK6 130 TKENIPPNIFPISGIPWTSFTGFNLNINNDGDLLPIITCGKYGNEE NKIMLPVSLQVHHSICDGYHAS
A0A068NDJ0 130 TKENIPPNVFPISSIPWTSFTGFNLNINNDADYLLPIITCGKYGNEE NKVILPVSLQVHHSVCDGYHAS
C2VBS1 130 PKENIPPNVYPISGIPWTSFTGFNLNINNDGDFLLPIITCGKYGNDG SKVMLPFSLQVHHAVCDGYHAS
F0PPC1 130 TKENIPPNVFPISSIPWTSFTGFNLNINNDADYLLPIITCGKYGNEK NKVMLPVSLQVHHSVCDGYHAS
Q5J470 130 PQGNLPENHLNISSLPWVSFDGFNLNITGNDDYFSPVFTMAKFQQEG DRVLLPVSVQVHHAVCDGFHAA
A0RE09 130 TKENIPPNVFPISSIPWTSFTGFNLNLNNDADFLLPIITCGKYGNEE NKIMLPVSLQVHHSICDGYHAS
C3HIG0 130 TKENIPPNIFPISGIPWTSFTGFNLNINNDGDLLPIITCGKYGNEE NKIMLPVSLQVHHSVCDGYHVS
M7CMF0 135 PKGF IENMFFVSANPWVSFTSFDLNVANMDNGFAPVFTMGKYYTQG DKVLMPLAIQVHHAVCDGFHVG
C2MKR8 130 TKENIPPNVFPISSIPWTSFTGFNLNINNDGDLLPIITCGKYGNEE NKIMLPVSLQVHHSVCDGYHAS
K4V802 135 PKGF IENMFFVSANPWVSFTSFDLNVANMDNGFAPVFTMGKYYTQG DKVLMPLAIQVHHAVCDGFHVG
R9VXF1 135 PKEESRENVFFVSANPWVSFTGFNINIANMQNGFAPMFTIGKYYQQG EKVLLPLAVQVHHAVCDGFHVS
D0ZHK9 130 PQGVTPENHLNISALPWVNFDSFNLNVANFTDYFAPIITMAKYQQEG DRLLLPLSVQVHHAVCDGFHVA
D8GXV5 130 TKENIPPNVFPISSIPWTSFTGFNLNINNDADYLLPIITCGKYGNEE NKVILPVSLQVHHSVCDGYHAS
C2VTN4 130 TKENIPPNVFPISSIPWTSFTGFNLNLNNDADFLLPIITCGKYGNEE NKIMLPVSLQVHHSICDGYHAS
Q7N4A3 130 SKPELPENHFHISSVPWVSFDGFNLNMASVMDYFPPIFTMGKFYQNG NQTQLPLAIQVHHATCDGFHVG
Q63BG1 130 TKENIPPNIFPISGIPWTSFTGFNLNINNDAGFLLPIITCGKYGNEE NKIMLPVSLQVHHSVCDGYHVS
G8LN29 135 PKEESRENVFFVSAIPWSSFTSFNLNVANMQNGFAPVFTMGKYVNQG GKVLLPLAVQVHHSVCDGFHVA
A9VF66 130 TKENIPPNIFPISGIPWTSFTGFNLNINNDGDSLLPIITCGKYGNDG SRVMLPVSLQVHHAVCDGYHAS
C3I0P6 130 TKENIPPNVFPISSIPWTSFTGFNLNINNDENFLLPIITCGKYGNEG NKVMLPVSLQVHHSVCDGYHAS
R9TXA1 131 AKPDIPPNAFSVSSIPWVRFTNFNLNLDNS EHLLPIITNGKYYSEG GETFLPVSLQVHHAVCDGYHAG
consensus 141 k p nmf vs lPw sft fnlnv n fllPi t gkyy kvllPl iqvhHavcDGfH
P62577 203 RMLNELQQYCDEWQGGA
P00484 199 RFINRLQELCNSKLK
P22616 199 RFINTLQMMCDNILK
P22615 199 RFINTLQLMCDNILK
P00485 199 LFMNSIQELSDRPNDWLL
P26826 196 RFVNELQELINS
P00486 199 LFMNEFQDIIHKVDDWI
P00487 197 QYVEYLRWLIEHCDEWLNDSLHIT
P36882 199 LFINEFQDIIKKVDDWI
P58777 203 RMLNELQQYCDEWQGGA
P06135 199 LFINEFQDIINKVDDWI
P36883 199 LFMNEFQNIIR
P20074 205 RLTNELQTILADP AWL
P25309 199 LFINEFQDIIKKVDDWI
P49417 196 RFINELQDVIQSLQNHGGDEE
P26825 200 RFFNEAQELASNYETWLGEK
P22782 196 RFIDELQDLLNK
P11504 196 RFVNELQELIIVTQVCL
Q03058 199 LFMNEFQDIIHRVDDWI
P62579 203 RMLNELQQYCDEWQGGA
P07641 203 RLLNEIQQYCDEGCK
Q02736 200 RFINEMQELAFSFQEWLENK
P62580 203 RMLNELQQYCDEWQGGA
W9BPG6 193 RFYALVEEGFNDPERLWLPLTETAPSLLPPAAKER
P62578 203 RMLNELQQYCDEWQGGA
Q81QL5 199 RFIEDLQELSNSCNEWLK
U6A0V2 199 RFIEDLQELVNTCNEWL
Q79PA8 203 RMLNELQQYCDEWQGGA
Q8A9W3 200 KFIETLDL
Q8A336 204 LFYKKVEDFLK
B2TPT1 195 QYFNIIQEILDNPMKYL
G0EL70 200 IFFEDLQHEFDEFDLLY
A4W932 204 RLFNELQEMCDDILRLSEEFNA
B2IQC4 199 LFMNSIQELSDRPNDWLL
D6BAH5 205 RLTDELQTILADP AWL
A0A0B4S6D3 197 QYVEYLRWLIEHCEEWLSDSLYT
B1IES5 196 RFVNELQELIKG
B1LRU3 199 RFINTLQLMCDNILK
B7NGK6 203 RMLNELQQYCDEWQGGA



* All CATs are represented by their accession numbers on http://www.uniprot.org/; for example, P62577=CATI, P20074=CATIII which are yellow-shaded.

Appendix 6: pMJA01 whole nucleotide sequence:

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCT
GTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGG
CTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATGCGGTGTGAAATACCGCACAGAT
GCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATC
GGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTA
ACGCCAGGGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCGAGCTCGGTACCCGGGGAT
CCTCTAGATTAATACGGTTTGAAGTTCAGGCTGCTACCAATGTTATATGCGATATACTGCAGCACATTAA
TCACCACGCTATTACCAAACTGTTTATATGCCTGGCTGGTGCTCGGATGAACTTTATAGCTATCAGGATA
ACCCATAACACGTGCACATTCACGCGGATGCAGTTTACGGGTTTTACCATTAACCAGATAACCACCGGTT
TTGGCAAAAATGCCACCACCATATGCACTTAAGGTAATTGCAATACCACGGGTGCTATAAATACGTTCAC
CCTGACCACCTTTACCAACAATACCCAGACGAACGGTTTTCGGGGTGGTCTGTTCAATTTCTTGGTTGGT
CATCACCAGGTCTTTACGATCAATAACCAGATGTTCAACTTCGCTATCCGGCAGCAGCAGATCTTTCACA
AAGGTATTCAGTTCAAACGGTTTCGGAAACTGAAAATTCTGGATGTTCAGATCGTTGCGAAAGCAAATCA
TATAAATGCGTTCGCGTTTCTGCGGAATACCATAATCTAATGCATTCAGCACTTTGGCATGAAAGCTATA
ATCCAGTTCGTTCATGGTGTTTTTCACAACTTCCAGGGTATTGCCATTATCGTGGCTAGCAAAGTTTTTC
ACATTTTCCATAAACACAACTTTCGGTTTTTTTTCACGAACAATGCGCGCAATATCAAAAAACAGGGTGC
CACGGCTATCTTCAAAACCTTTCTGTTTTCCGGAAATGCTAAATGCCTGACACGGAAAGCCGGCACACAG
AATATCATGATCAGGAATGGTTTTTTCGTTAACCTGGGTAATATCGCCTTCCGGTTTTTCACCAAAGTTC
ATTTCATACACCTCTTGGGCATATTTATCCCACTCGTTGCTATAAACACATTCTGCACCACAGCTTTCCA
GTGCCAGACGAAAACCACCCAGACCTGCAAACAGGTCAATAAAACGCAGACCGGTCAATTGTTTATCTTT
GATCTCAACCATGGAAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCA
CAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACT
CACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGA
ATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGC
TGCGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGA
ATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCC
GCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGA
GGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCC
TGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAT
AGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCC
CCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTT
ATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTC
TTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAG
TTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTT
TGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGG
TCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCA
CCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGA
CAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCC
TGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATAC
CGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAG
AAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGT
TCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTG
GTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAA
AGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTT
ATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACT
CAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAA
TACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCA
AGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTT
TTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGC
GACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGT
CTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCC
GAAAAGTGCCACCTGACGTCTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCAC
GAGGCCCTTTCGTC

*M.HhaI ORF is from ATG to TAA which are green and red-shaded respectively.

Appendix 7: pAtase2 whole nucleotide sequence:

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAC
CAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCGCT
ACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCTTCG
CTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCC
AGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATTGGCGG
AAGGCCGTCAAGGCCACGTGTCTTGTCCAGAGCTCTACGCGAAGCTTCGGGCGTATTTTTTGAGTTATCG
AGATTTTCAGGAGCTAAGGAAGCTACCATGGAGAAAAAAATCACCGGTTATACCACCGTTGATATTAGCC
AGTGGCATCGTAAAGAACACTTTGAAGCATTTCAGAGCGTTGCACAGTGTACCTATAATCAGACCGTTCA
GCTGGATATCACCGCATTTCTTAAGACCGTGAAAAAAAACAAACACAAATTCTATCCGGCATTCATCCAT
ATTCTGGCACGTCTGATGAATGCACATCCTGAATTCCGTATGGCAATGAAAGATGGTGAACTGGTTATTT
GGGATAGCGTTCATCCGTGTTATACCGTTTTTCATGAACAGACCGAAACCTTTAGCAGCCTGTGGTCAGA
ATATCATGATGATTTTCGTCAGTTCCTGCACATTTATAGCCAGGACGTCGCATGTTATGGTGAAAATCTG
GCATATTTTCCGAAAGGCTTTATCGAAAACATGTTCTTTGTTAGCGCAAATCCGTGGGTTAGCTTTACTA
GTTTCGATCTGAATGTGGCCAACATGGATAACTTTTTTGCACCGGTTTTTACGATGGGCAAATATTACAC
CCAGGGTGATAAAGTTCTGATGCCGCTAGCGATTCAGGTTCATCATGCAGTTTGTGATGGTTTTCATGTT
GGTCGTATGCTGAATGAACTGCAGCAGTATTGTGACGAATGGCAGGGTGGTGCATAATCTAGACAGTGGG
GATCCCAGTGGGGTACCTGGAGCACAAGACTGGCCTCATGGGCCTTCCGCTCACTGCCCGCTTTCCAGTC
GGGAAACCTGTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCT
TCCTCGCTCACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCA
GCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAG
CATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTC
CCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCT
CCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGC
TCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTC
TTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGC
GAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTA
TTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAAC
AAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAGGATCTCA
AGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTG
GTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCT
AAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGAT
CTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTA
CCATCTGGCCCCAGTGCTGCAATGATACCGCGAGAACCACGCTCACCGGCTCCAGATTTATCAGCAATAA
ACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAA
TTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACA
GGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAG
TTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAA
GTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTA
AGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTT
GCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGG
AAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACT
CGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGC
AAAATGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATA
TTATTGAAGCATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAA
CAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCAC

*CATI ORF is from ATG to TAA which are green and red-shaded respectively.

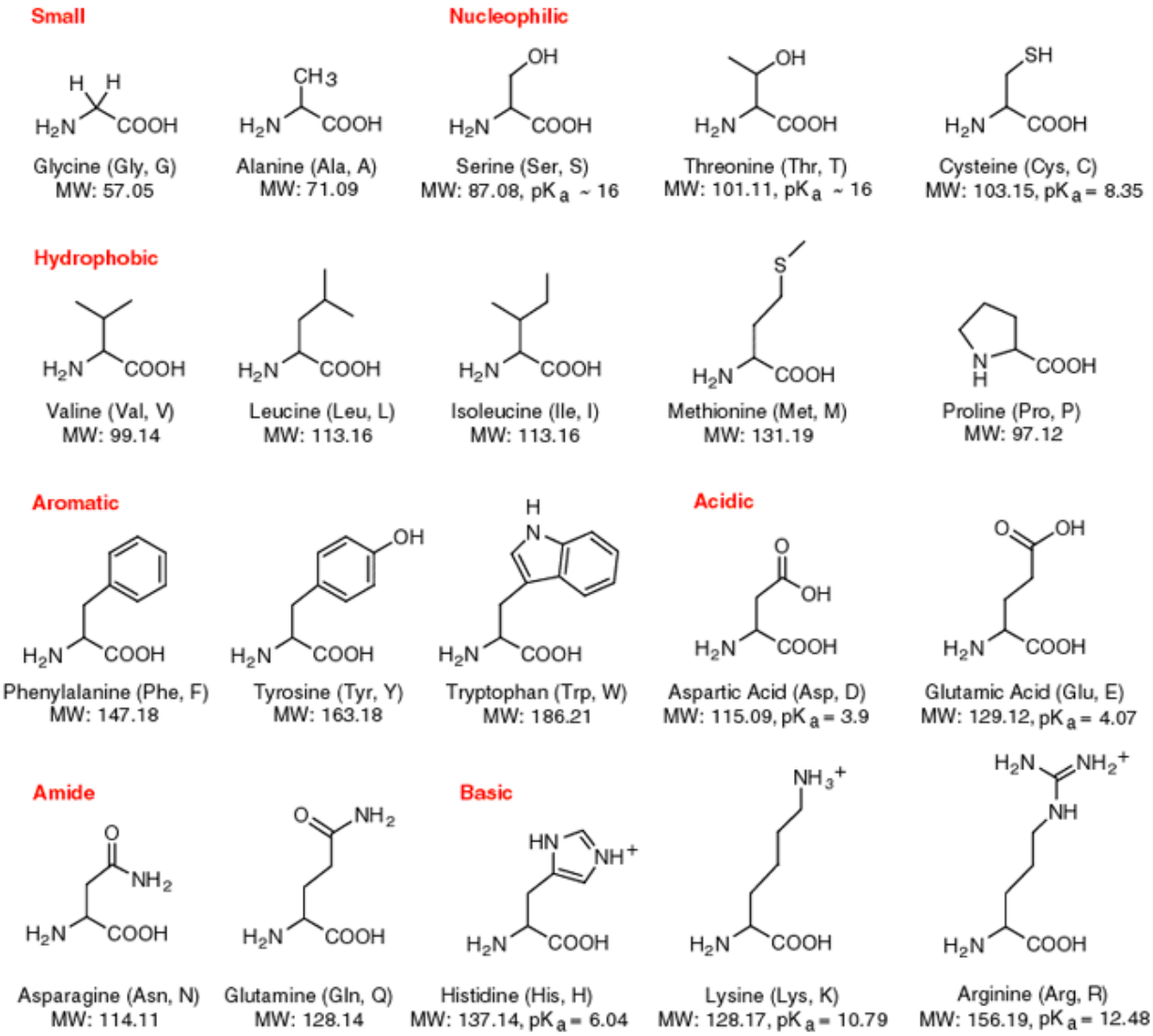
Appendix 8: pLpd1 whole nucleotide sequence:

CTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAAC
CAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGGCCGCT
ACAGGGCGCTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGTTTCGGTGCGGGCCTCTTCG
CTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCC
AGTCACGACGTTGTAAAACGACGGCCAGTGAGCGCGACGTAATACGACTCACTATAGGGCGAATTGGCGG
AAGGCCGTCAAGGCCGCATAAGCTTGTTTAAAAATTGTTAACAATTTTGTAAAATACCGACGGATAGAAC
GACCCGGTGGTGGTTAGGGTATTACTTCACATACCCTATGGATTTCTGGGTGCAGCAAGGTAGCAAGCGC
CAGAATCCCCAGGAGCTTACATAAGTAAGTGACTGGGGTGAGGGCGTGAAGCTAACGCCGCTGCGGCCTG
AAAGACGACGGGTATGACCGCCGGAGATAAATATATAGAGGTCACCATGGGTACGGAAATCAAAACCCAG
GTTGTTGTGCTGGGTGCCGGTCCGGCAGGTTATAGCGCAGCATTTCGTTGTGCAGATCTGGGTCTGGAAA
CCGTTATTGTTGAACGTTATAATACCCTGGGTGGTGTTTGTCTGAATGTTGGTTGTATTCCGAGCAAAGC
ACTGCTGCATGTTGCAAAAGTTATTGAAGAAGCAAAAGCACTGGCCGAACATGGTATTGTTTTTGGTGAA
CCGAAAACCGACATTGATAAAATTCGTACCTGGAAAGAAAAAGTGATCAATCAGCTGACCGGTGGTCTGG
CAGGTATGGCAAAAGGTCGTAAAGTTAAAGTTGTTAATGGCCTGGGCAAATTTACCGGTGCAAATACACT
GGAAGTTGAAGGTGAAAATGGTAAAACCGTGATCAACTTTGACAACGCAATTATTGCAGCCGGTAGCCGT
CCGATTCAGCTGCCGTTTATTCCGCATGAAGATCCGCGTATTTGGGATAGCACCGATGCACTGGAACTTA
AGGAAGTGCCGGAACGTCTACTAGTAATGGGTGGTGGTATTATTGGCCTGGAAATGGGCACCGTTTATCA
TGCACTGGGTAGCCAGATTGATGTTGTGGAAATGTTTGATCAGGTTATTCCGGCAGCCGATAAAGATATT
GTTAAAGTGTTTACCAAACGCATCAGCAAAAAATTCAACCTGATGCTGGAAACCAAAGTTACCGCAGTTG
AAGCCAAAGAAGATGGTATTTACGTTACGATGGAAGGCAAAAAAGCACCGGCAGAACCGCAGCGTTATGA
TGCAGTTCTGGTTGCCATTGGTCGTGTTCCGAATGGCAAAAATCTGGATGCAGGTAAAGCCGGTGTTGAA
GTTGATGATCGTGGTTTTATTCGTGTGGATAAACAGCTGCGTACCAATGTTCCGCATATTTTTGCAATTG
GCGATATTGTTGGTCAGCCGATGCTGGCACATAAAGGTGTTCATGAAGGTCATGTTGCGGCCGAAGTTAT
TGCCGGTAAAAAACATTATTTCGATCCGAAAGTTATCCCGAGCATTGCATATACCGAACCGGAAGTTGCA
TGGGTTGGTCTGACCGAAAAAGAGGCCAAAGAAAAAGGTATTAGCTATGAAACCGCAACCTTTCCGTGGG
CAGCAAGCGGTCGTGCCATTGCAAGCGATTGTGCCGATGGTATGACCAAACTGATCTTTGATAAAGAAAG
CCATCGTGTTATTGGTGGTGCCATTGTTGGCACCAATGGTGGTGAACTGCTGGGTGAAATTGGTCTGGCC
ATTGAAATGGGTTGTGATGCAGAAGATATTGCCCTGACCATTCATGCACATCCGACCCTGCATGAAAGCG
TTGGCCTGGCAGCAGAAGTTTTTGAAGGTAGCATTACCGATCTGCCGAATCCGAAAGCCAAAAAAAAATA
ATCTAGACAGTGGGGATCCCTGGGCCTCATGGGCCTTTCCGCTCACTGCCCGCTTTCCAGTCGGGAAACCT
GTCGTGCCAGCTGCATTAACATGGTCATAGCTGTTTCCTTGCGTATTGGGCGCTCTCCGCTTCCTCGCTC
ACTGACTCGCTGCGCTCGGTCGTTCGGGTAAAGCCTGGGGTGCCTAATGAGCAAAAGGCCAGCAAAAGGC
CAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCATCACAAA
AATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAA
GCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGG
AAGCGTGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTG
GGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCA
ACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGT
AGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATC
TGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCG
CTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCC
TTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGA
TTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATAT
ATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATT
TCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGC
CCCAGTGCTGCAATGATACCGCGAGAACCACGCTCACCGGCTCCAGATTTATCAGCAATAAACCAGCCAG
CCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCG
GGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTG
GTGTCACGCTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGAT
CCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGC
AGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTT
TCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCC
CGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTC
TTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCC
AACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCG

CAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAG
CATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGG
GTTCCGCGCACATTTCCCCGAAAAGTGCCAC

*E3 ORF is from ATG to TAA which are green and red-shaded respectively.

Appendix 9: Amino acid structures:



Appendix 10: Amino acids codons:

Amino Acid	Amino Acid Abbr.	Codon		Amino Acid	Amino Acid Abbr.	Codon
Alanine	A	GCA		Proline	P	CCA
	A	GCC			P	CCC
	A	GCG			P	CCG
	A	GCU			P	CCU
Cysteine	C	UGC		Glutamine	Q	CAA
	C	UGU			Q	CAG
Aspartate	D	GAC		Arginine	R	AGA
	D	GAU			R	AGG
Glutamate	E	GAA			R	CGA
	E	GAG			R	CGC
Phenylalanine	F	UUC			R	CGG
	F	UUU			R	CGU
Glycine	G	GGA		Serine	S	AGC
	G	GGC			S	AGU
	G	GGG			S	UCA
	G	GGU			S	UCC
Histidine	H	CAC			S	UCG
	H	CAU			S	UCU
Isoleucine	I	AUA		Threonine	T	ACA
	I	AUC			T	ACC
	I	AUU			T	ACG
Lysine	K	AAA			T	ACU
	K	AAG		Valine	V	GUA
Leucine	L	CUA			V	GUC
	L	CUC			V	GUG
	L	CUG			V	GUU