THE UNIVERSITY *of* EDINBURGH

Mapping charge to function relationships of the DNA mimic protein Ocr

Nisha Kanwar

Thesis presented for the degree of

Doctor of Philosophy

University of Edinburgh

2014

Declaration

I hereby declare that this thesis and the work presented in it are my own and have been generated by me as a result of my own original research.

Nisha Kanwar

Aug 2013

For my grandparents

Mr. Nand Singh Kanwar and Mrs. Laxmi Devi Kanwar

and my parents

Mr. Shiv Singh Kanwar and Mrs. Ramesh Kanwar

Their courage and strength in life an inspiration,

Their good humour and loving nature a precious gift

Their limitless faith…encourage-able

Acknowledgements

There are a great many people, who in many ways have contributed to the success of my last four years in Edinburgh. It is not possible for me to mention them all here however amongst them there have those who have been invaluable,

Most importantly I would like to give a special thank you to Dr David Dryden and his band of merry men! David's support, guidance and loyalty have been unquestionable and he has been a great supervisor over the past four years. Not to mention his unrelenting eye for detail.

To the merry men; Dr Gareth Roberts and Mr. Laurie Cooper who have offered been higly supportive throughout particularly in protein purification and biophysical techniques. Both are in possession of a great breadth of knowledge on all manner of skills from Scientific to Historical to Celtic FC. In addition, Dr AS Stephanou contributed a great deal to the early work in my thesis and with Gareth provided me with many samples and informative discussions and ideas. Other members of the lab I wish to thank for their support, discussion and conversation include Dr John White, Dr Kai Chen, Edward Bower and Matt Tilling.

I am indebted to Dr Angela Dawson for samples, useful discussions but more importantly for her training in all things biological and bacterial in the early stages of my PhD. Angela is a gifted teacher; her patience, experience and ability to explain things simply have meant that she has been a constant source of knowledge throughout my PhD. In addition her kindness and that of her supervisor Prof Wilson Poon in regards to lab space and lab resources and equipment.

I would like to thank Dr Garry Blakely and his team for the use of their microscope and the use of Garry's seemingly boundless knowledge of bacteria. He has provided great insight into the subtleties of microbiology.

Other collaborators to thank include Dr Ryan Morris for use of the plate reader in Physics and Liz from the PPF unit for training in TDA and ITC. I would like to thank my brother Vikram who went out of his way to design me a very clever spreadsheet that saved me a great deal of time and sanity.

In addition I would like to thank many of my fellow PhD students, who shared many late night shift both social and scientific. In particular I would like to thank my neighbouring biochemist Dr Martin Wilkinson who never should have been subjected to the Ocr protein in such detail but who bore it well.

It remains to thank my friends and family, to my friends who have tolerated my inconsistencies in the last four yrs. To my brothers Aalok and Vikram, who have been sure to keep me grounded and offer perspective and i would like to thank my parents. It goes without saying that their support and encouragement is irreplaceable and their deep pockets usually unquestionable.

My PhD was sponsored by the DTC of Cell and Proteomic Technologies from the University of Glasgow.

Abstract

This thesis investigates the functional consequences of neutralising the negative charges on the bacteriophage T7 antirestriction protein ocr. The ocr molecule is a small highly negatively charged, protein homodimer that mimics a short DNA duplex upon binding to the Type I Restriction Modification (RM) system. Thus, ocr facilitates phage infection by binding to and inactivating the host RM system.

The aim of this study was to analyse the effect of reducing the negative charge on the ocr molecule by mutating the acidic residues of the protein. The ocr molecule (117 residues) is replete with Asp and Glu residues; each monomer of the homodimer contains 34 acidic residues. Our strategy was to begin with a synthetic gene in which all the acidic residues of ocr had been neutralised. This so called 'positive ocr' (or pocr) was used as a template to gradually reintroduce codons for acidic residues by adapting the ISOR strategy proposed by D.S.Tawfik. After each round of mutagenesis an average of 4-6 acidic residues were incorporated into pocr. In this fashion a series of mutant libraries in which acidic residues were progressively introduced into pocr was generated.

A high-throughput *in vivo* selection assay was developed and validated by assessing the antirestriction behaviour of a number of mutants of the DNA mimic proteins wtOcr and Orf18 ArdA. Further to this, selective screening of the libraries allowed us to select clones that displayed antirestriction activity. These mutants were purified and *in vitro* characterisation confirmed these mutants as displaying the minimum number of acidic residues deemed critical for the activity of ocr. This *in vitro* process effectively simulated the evolution of the charge mimicry of ocr.

Moreover, we were able to tune the high-throughput assay to different selection criteria in order to elucidate various levels of functionality and unexpected changes in phenotype. This approach enables us to map the "*in vitro"* evolution of ocr to identify acidic residues that are required for protein expression, solubility and function proceeding to a fully functional antirestriction protein.

Contents

Chapter 1. Introduction

1.1 Restriction Modification (RM) enzymes

In 1962 Arber and Dussiox discovered a family of proteins that could provide host controlled restriction and modification of DNA. These enzymes were later named restriction modification systems. Since their early discovery, investigations have elucidated several key roles of Restriction Modification (RM) systems within a cell; a defensive role, the acquisition of beneficial sequence codes and/or propagation of selfish genetic elements.[1] RM systems bind to DNA providing two different enzymatic processes; DNA cleavage and restriction or DNA modification. The RM restriction and modification complexes are able to recognise specifically modified sequences of DNA or "recognition sites". In the event of an unmodified recognition site, the restriction process is initiated, resulting in the cleavage of the DNA duplex by the restriction endonuclease (REase). However, hemimethylated recognition sites initiate the modification process and so are methylated by the methyltransferase (MTase).

As a defense mechanism RM systems have the ability to distinguish between the host cells genome and 'foreign' DNA. Each organism has its own signature DNA distinguishable by differential modification patterns (i.e. methylation sites) that are present on the DNA. In a bacterial cell the chromosomal (host) DNA is methylated at specified recognition sites which can contain methylated cytosine or adenine basepairs. During each round of replication newly synthesised DNA will lack this signature generating hemimethylated sites which are identified by RM enzymes and duly modified. However if the RM enzymes encounter unmodified DNA, it will trigger the restriction reaction of the REase. Figures 1.1 and 1.2 demonstrate the general steps involved in the restriction and modification processes.[1,2,3]

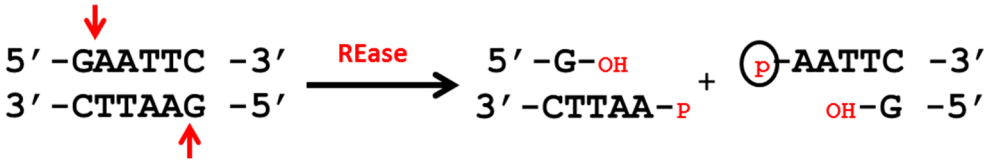


Figure 1.1: Hydrolysis of phosphodiester bonds and cleavage of dsDNA by a REase. The reaction generally requires Mg^{2+} ions and for some RM enzymes ATP.

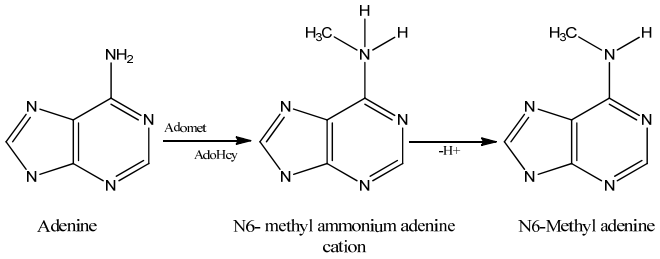


Figure 1.2: Reaction pathway for the methylation of adenine by a MTase. The reaction includes the use of the cofactor S-adenosyl-methionine (SAM) which is converted to S-adenosylhomocysteine (SAH) upon donation of a methyl group to the DNA.[4]

Several groups of RM systems have been identified over the years. These groups are distinguishable from each other by differences in enzyme composition, cofactor requirements, target site properties, the site of DNA cleavage and are named Type I-IV, see table 1.1.[1,3,5]

	Type I	Type II	Type III	Type IV
Example	EcoKI	EcoRI	EcoP1I	EcoMcrBC
Genes	*hsd R,M,S*	*ecorIR, ecorIM*	*mod res*	*mcrB, mcrC*
Subunits	MTase: M_2S_1 REase: $R_2M_2S_1$	MTase: M^2 REase: R^2	mod_2res_2 mod_2res_1	McrB McrC
NTPase	ATPase	ATPase	ATPase	GTPase
Co-factors for Restriction	ATP, SAM, Mg^{2+}	Mg^{2+}	ATP, Mg^{2+}, (SAM)	GTP, Mg^{2+}
Co-factors for methylation	SAM	SAM	SAM	no methylation
Recognition Site	Asymmetric and bipartite	Symmetric	Asymmetric	Bipartite and methylated
Cleavage Site	1000bp from recognition site (variable)	Fixed location at or near recognition site	Fixed location 25-27bp from recognition site	Between two methylation sites at multiple sites
DNA Translocation	Y	N	Y	Y

Table 1.1: The different characteristics of Type I-IV RM Enzymes.[2]

1.2 Type II, III and IV enzymes

The Type II RM systems are the most well-known family of RM systems as they paved the way for bioengineering and experimental cloning. The Type II systems generally consist of two separate enzymes which perform the restriction (REase) and modification (MTase) processes, with both enzymes targeting the same recognition site on DNA. Upon binding to DNA, the MTase ensures that specific bases on both the sense and antisense strands of the target site are modified. If not, the MTase acts to modify the bases in the target site. Type II enzymes can modify the target site in various ways modifications include; N^6-methyladenine

(m_6A), N^4- and N^5-methylcytosine (m_4C and m_5C). A methyl group is donated to DNA from the cofactor S-adenosylmethionine (SAM). If the target site is unmodified then it is targeted by the restriction enzyme using an Mg^{2+} cofactor. The enzyme cleaves the DNA at specified sites that are either near or very close to the target site. The specific restriction sites of the Type II systems is a key feature that has been utilised within protein recombinant technology.[2,3]

Type III systems consist of two subunits, the mod (recognition and modification) and res (restriction). The modification subunit can act independently from the restriction subunit and requires the cofactor SAM. To carry out restriction the enzyme can form one of two complexes either mod_2res_2 or mod_2res_1. The enzyme interacts with two copies of the recognition sequence in an inverse orientation, typically target sites are asymmetric and 5-6bp long. For restriction activity the cofactor SAM is required in addition to Mg^{2+} and ATP. The restriction process is dependent on ATP hydrolysis which drives the translocation process of the DNA until the enzyme collides with another entity, usually a second RM enzyme, which then stimulates DNA cleavage. Translocation for Type III enzymes is a relatively short process and so cleavage by Type III enzymes occurs on the 3' side of the recognition sequence, not far from the original binding site.[2,6,7]

Type IV enzymes are counterintuitive as they are methylation-dependent restriction systems. Therefore they require a methylated substrate to restrict the DNA. Due to this reason these systems act as REase enzymes only, seeking out foreign DNA methylated at a particular recognition site and subsequently cleaving the DNA. [2,7,8]

1.3 Type I RM Systems

Type I RM systems produce multifunctional enzymes comprised of three different subunits; HsdS, HsdM and HsdR. These subunits assemble into two active forms which perform as an MTase or REase to provide methylation and restriction activity.[1-5] The unique feature of Type I RM enzymes is the presence of the HsdS subunit also known as the specificity subunit. Upon binding the HsdS subunit recognises a specific recognition site on dsDNA[9] and is present in both the MTase and REase complexes. Whereas the MTase complex facilitates the methylation reaction using the cofactor SAM, the REase complex translocates and cleaves the DNA in an ATP-dependent process. The restriction reaction is facilitated by the presence of the cofactors Mg^{2+} and SAM. DNA cleavage is initiated by an obstruction of

the translocation process, either by collision with another translocating complex or the 'topology' of the DNA.[3,7,10,11]

1.3.1 Subunit composition

The molecular weight of the HsdS and HsdM subunits is 50-60kDa whereas the HsdR subunit is substantially larger with a molecular weight of ~140kDa.[12,13]

The complete RM complex which is responsible for the restriction process is assembled with a stoichiometry of $R_2M_2S_1$ producing a typical mass of 440kDa.[7] However, the MTase complex which modifies the DNA consists of the trimeric core enzyme only; M_2S_1. This shows that the restriction subunits are not necessary for the methylation reaction of Type I enzymes. [14] Furthermore, the *hsdM* and *hsdS* genes have the same promoter but the *hsdR* gene is located with a separate promoter downstream.[15] The resulting populations of both complexes in the cell show that there is a higher ratio of MTase to REase.[1,16-17]

Differences between Type I RM systems are based on genetic complementation of the *hsd* genes, antibody cross-reactivity and DNA hybridisation from different systems producing five subfamilies of the Type I enzymes. These subfamilies are named Type IA-IE, however this thesis considers the Type IA RM enzyme, EcoKI in more detail.

1.3.2 HsdS, the specificity subunit

The role of the S subunit is to recognise specific DNA target sites. Typically these target sites are nonpalindromic bipartite sequences recognising two different regions of specificity separated by a number of non-specific DNA basepairs, for example, the recognition site of EcoKI is 5' AAC(N_6)GTGC 3', where N signifies any nucleotide.[17] A number of HsdS subunits have been crystallised and the solved structures show four distinct regions on the HsdS subunit, see figure 1.7.[18,19,20] These include two globular domains designated target recognition domains (TRD) 1 and 2 which are separated by a highly conserved coiled coil region and a secondary conserved region located on the N-terminus. Sequence alignment studies show that the coiled coil region is the result of a highly conserved number of residues that are present among many HsdS subunits. The length of the coiled coil and alignment studies show that this coiled coil differs with the length of the non-specific region of the recognition site. Sequence analysis has shown that the HsdS coiled coil is very similar between Type IA RM systems.[1,7,21]

As the name suggests the two globular TRD regions recognise and bind to the specific target sequence of the DNA recognition site. The overall folds of both TRD domains are near

identical not only with each other but also with HsdS subunits from other Type I enzymes, although sequence similarity is only ~40%. Any difference in sequence appears to be concentrated around the TRD domains which affects the local secondary structure and the binding interface of the HsdS subunits with DNA.[22] The binding interface was identified by the presence of distinct DNA binding clefts present on each TRD.[22] The clefts possess a number of polar residues which bind to the negatively charged phosphate backbone of DNA.[7] However interestingly, a number of hydrophobic residues were also conserved within these clefts, which are thought to facilitate binding to DNA bases through van der Waals interactions.[10, 21,23]

Models of the interaction between DNA and the HsdS subunit show that in order for the DNA target sites to align with the designated TRD sites of the HsdS subunit, the DNA needs to bend and kink and this has been shown to also partly unravel the DNA. This reveals the methylation target sites on the adenine nucleotide.[19,24,25]

The HsdS subunit of the Type I RM enzyme from *Thermoanaerobacter tengcongensis* (TTE) was solved in 2011 by Gao *et al*.[19] The structure identified differences in the orientation of the TRD domains relative to previously solved HsdS structures. The TRD domains exhibit a rotational motion as a result of the bending and twisting of the conserved coiled coil regions.[19] Modeling of the HsdM subunits onto the structure of the HsdS revealed a potential open/close hinge mechanism for the MTase complex as a result of the domain motion, figure 1.3.

A conserved region among all Type I TRD domains is an enriched proline hinge which has been identified as a linking region between the TRD domains with the coiled coils, allowing rotational freedom between the two domains. On the other hand the highly conserved coiled coil region which joins the two TRD domains together is held tightly together by a number of hydrophobic residues. These have been pinpointed by a number of mutagenesis studies that have been carried out using information gained from the results of the crystal structure.[19]

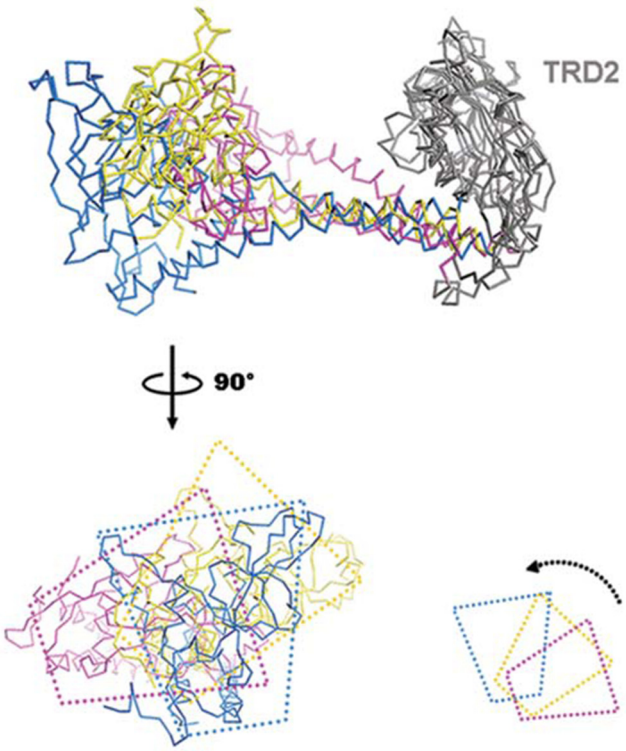


Figure 1.3: The alignment of three HsdS subunits have been overlaid to show the effect of the twisting by the conserved coiled coil region and its contribution to the hinge mechanism of the MTase complex of Type I enzymes. TTE-HsdS (blue), Mja-HsdS (yellow) and Mge-HsdS (red). This shows clear differences in the orientation of the three HsdS subunits.[19]

1.3.3 HsdM subunit

The C-terminal domain of HsdM subunit is highly conserved in Type IA enzymes, so much so that these HsdM subunits are interchangeable. As such it is thought that the C-terminal domain may act as an interface between the HsdS and the HsdM subunits.

Several conserved regions have been identified from comparisons of HsdM subunits to Type II MTases. The regions concerned are located in the central part of the Type I HsdM subunit. Extensive studies have revealed that these motifs are important for the recognition of the methylation status, the docking of the SAM cofactor and the adenine target site which is situated in close proximity to the HsdS TRD sites.[17,26] The partial unravelling of the DNA upon binding results in the insertion of the adenine target sites into the binding pocket of the HsdM subunit. This mechanism is a phenomenon termed nucleotide base flipping, figure 1.4. [10,15,27] Co-crystal structures of a number of Type II MTases interacting with DNA have captured base flipping. The structures show that unmodified nucleotide target rotates out of the DNA helix, by the twisting of the sugar phosphate backbone, resulting in the capture of these targets in the catalytic pocket of the MTase enzyme.[23,28]

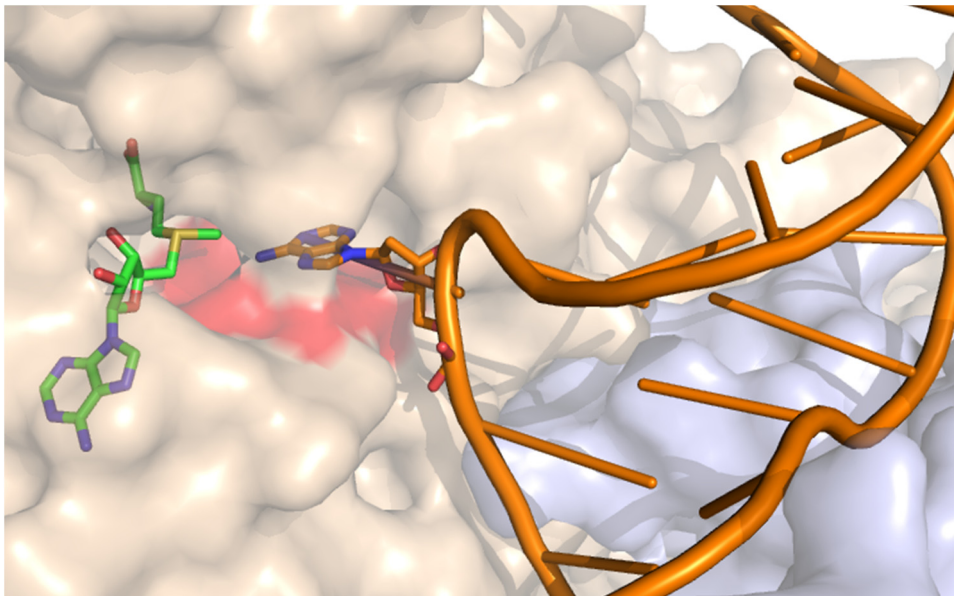


Figure 1.4: Model of DNA bound to MEcoKI, where the HsdM subunit is wheat coloured and the HsdS subunit is pale blue (pdb: 2Y2C). The DNA is highlighted in orange, an adenine base has flipped into the catalytic domain of the HsdM subunit (wheat colour). The catalytic domain contains a binding pocket for the cofactor SAM (green). Highlighted in red is a conserved region found on all MTase enzymes (NPPF).

The HsdM subunit for EcoKI was successfully crystallised in 2005 by K M Rajashankar *et al*. (pdb: 2AR0)

1.3.4 HsdR subunit

HsdS and HsdM subunits that belong to the same family of Type I RM systems show high levels of sequence similarity and the HsdR subunit is no exception.

In fact sequence alignments of HsdR subunits have revealed a conserved region located in Type I and Type II RM systems. This domain is located near the N-terminal region of Type I systems and studies have shown that this region is responsible for DNA cleavage and is an endonuclease domain. The central region of the HsdR subunit appears to be a domain that is unique to Type I and Type III enzymes. This region appears to be responsible for the translocation of DNA as it contains the conserved motifs of DEAD box helicases which are typically prevalent in many RNA and DNA helicases.[15]

The first crystal structure of a near complete HsdR subunit was solved in 2008. Lapkouski *et al* revealed that the subunit consisted of four globular domains, see figure 1.5.[29] These domains assemble to form a square planar arrangement with the formation of a groove through the middle, which is enriched with positive amino acid residues.[10] The ATP and Mg^{2+} binding sites were identified in the centre of the structure by extensive mutagenesis. It

was found that the absence/inhibition of DNA translocation also prevents the DNA cleavage process by the HsdR subunit.[29,30]

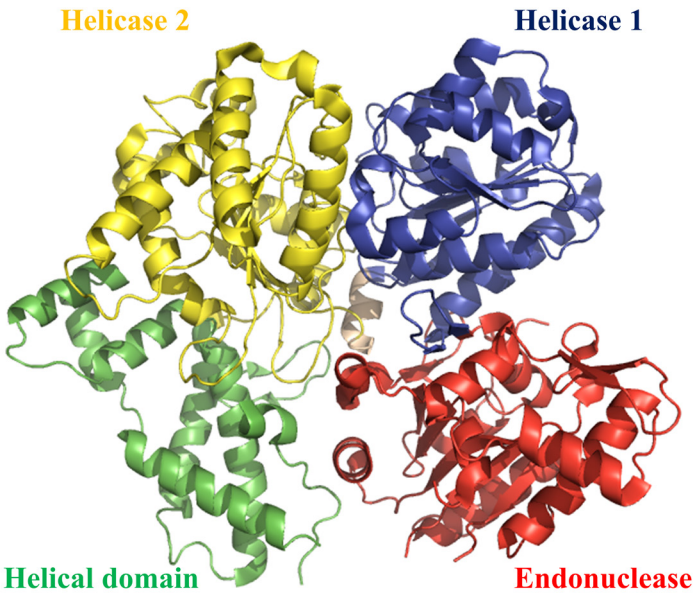


Figure 1.5: The crystal structure of HsdR subunit of the Type IC RM enzyme, EcoR124I in a planar domain arrangement. The domains have been highlighted in red for the endonuclease (aa13-260), blue and yellow for Helicase 1 and 2 (aa261-461 and 470-731) and a fourth helical domain in green (aa732-892) as interpreted by Lapkouski *et al*, 2008.[29]

1.3.5 Translocation mechanism

Upon binding to the recognition site of a DNA duplex, the HsdR subunit translocates the DNA, by "pulling" DNA towards the enzyme. This simultaneous movement both left and right is known as bidirectional translocation. If one HsdR subunit is inactive then dsDNA cannot be cleaved efficiently, although translocation is not affected, which implies that two proficient HsdR complexes are necessary to cleave double stranded breaks on a DNA duplex.[11,31,32]

The first suggestion of translocation motion was by Shulman in 1974.[33,34] Since then several models have been proposed in the literature. The current model is by Neaves *et al*, 2009, which was inspired from the previous model of Studier and Bandyopadhyay, 1988 and is outlined in Figure 1.6.[35] In 1988 Studier and Bandyopadhyay had established that EcoKI enzymes individually bind to and translocate the target site on a DNA duplex. This is regardless of other RM enzymes binding to target sites located on the same piece of DNA. Cleavage occurs due to a collision between the two Type I enzymes.[35] However, the AFM

results outlined in Neaves *et al*, 2009 have shown a cleavage mechanism similar to the Type II nuclease, FokI, which only cleaves DNA after dimerisation of the enzyme. They found that EcoKI can dimerise on the DNA and find the recognition site by linear diffusion up and down the DNA from a non-specific site.[36] DNA cleavage in the presence of the dimer is caused by 'DNA stalled translocation', when the loops formed through translocation are fully contracted.[37,38,11]

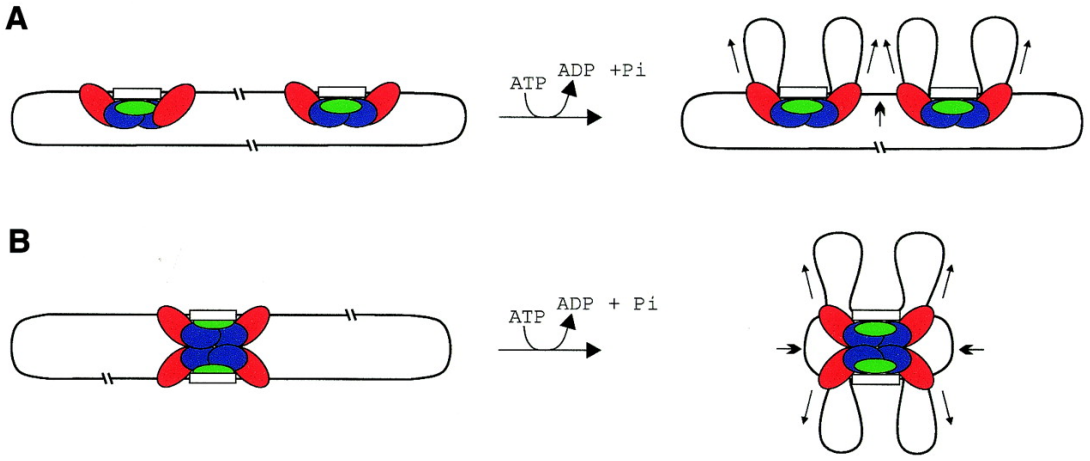


Figure 1.6: Diagram of EcoKI binding, translocation and restriction on DNA. Panel (a) shows the previous model (Studier and Bandyopadhyay, 1988). EcoKI monomers bind individually to each site and translocation occurs independently from both occupied sites. When the monomers meet, restriction occurs due to stalled translocation. Panel (b) shows a new model.[34,37,38] EcoKI binds to one site and then a second "monomer" binds to form a dimerised complex. The EcoKI dimer then forms diffusive loops with non-specific regions of DNA until it is stabilised by contact with the secondary EcoKI site. Translocation is then initiated from both sides of both monomers and restriction occurs when the translocation process is stalled. For both panels DNA is represented as a line; specific EcoKI sites are represented by dots on the DNA molecules, and EcoKI monomers are represented as individual spherical objects.[38]

Although evidence for dimerisation is strong, it is not an absolute requirement for translocation, as cleavage can occur in the presence of only one enzyme and one site.[32,39] Instead dimerisation seems to enhance restriction. The cooperation between two Type I complexes allows the enzymes to remain stationary, without having to pull another RM complex through a crowded solution of macromolecules.

The translocation process has been recorded to distances of up to 50,000 basepairs. Various translocation rates have been recorded for different nucleases with the fastest translocation rate recorded at 1kb/s .[40,31,41] The translocation rate of EcoKI was measured by introducing the enzyme into the genome of T7 phage. Translocation was monitored by the progressive

methylation of Sau3A1 restriction sites and estimated the translocation rate of EcoKI to be 100bp/s.[42] Further experimentation by Seidel and Szczelkun, 2008, showed that each translocation step is a movement of 1-2bp of DNA with an ATP expenditure of ~1ATP molecule per basepair.[43]

1.4 EcoKI

EcoKI was the first Type I RM system identified and purified (1968)[44]. However, attempts to successfully crystallise any Type I RM systems have as yet been futile. It is only recently that researchers have managed to crystallise each of the subunits.

However, in 2009, Kennaway *et al* had resolved an 18Å resolution model of the MTase bound to a DNA mimic protein, wtOcr, using EM microscopy.[45] Using the structure of recently crystallized HsdS, HsdM and HsdR alongside the extensive biochemical/biophysical data on M.EcoKI, Kennaway and colleagues managed to successfully produce a model of the M.EcoKI complex that fitted the EM structure. The subsequent models produced were of unbound M.EcoKI, DNA bound M.EcoKI and M.EcoKI bound with the DNA mimic Ocr. EcoKI appears to preferentially favour a closed state and the dimensions of the complex are ~120Å by 100Å. The models show that upon binding to DNA or the DNA mimic Ocr the M.EcoKI enzyme produces a clamp-like structure which allows the enzyme to engulf Ocr/DNA. [14] The kink that is produced when Ocr dimerises and when DNA binds to M.EcoKI is accommodated by the enzyme by a kink that is formed in the coiled coil region of the HsdS subunit. [27]

1.4.1 Interactions of HsdM and HsdS

Although the location of the C-terminal domains remains speculative and its structure ambiguous, the models designed by Kennaway *et al* have suggested that these domains are in close proximity to the HsdS subunit. Supposedly the C-terminal domain of the M subunit (aa470-529) interacts with the highly conserved coiled-coil region of the HsdS subunit. Previous biochemical data have shown that the removal of ~43 residues from the C-terminal region of the HsdM subunit still produced a folded and stable subunit but this was unable to bind to the complex M_1S_1. In addition the domains that are involved in this potential interface (C-terminal domain on the HsdM subunit and the coiled coil region of HsdS) are highly conserved in Type IA RM families, so much so that they are thought to be

interchangeable. Their high conservation among the enzyme family supports the suggestion that this domain on the HsdM subunits plays a role in subunit assembly.

A secondary interface was also resolved which involved the two loop regions located on the catalytic domain of the HsdM subunit (aa430-433 and aa464-470) and an area close to the TRD domains of the HsdS subunit (N133-A144). However, the TRD domains are highly variable depending on the specific DNA site the HsdS subunit recognizes, therefore this interface is thought to be of secondary importance compared to the C-terminal interface. [45,24]

1.4.2 HsdS and DNA

The TRD domains of the HsdS subunit on EcoKI are known to interact with the specified recognition site on dsDNA. The Kennaway M.EcoKI model confirmed the identity of two loop regions on the TRD 1 domain (located at the N-terminal region of the HsdS subunit) that were in contact with DNA; aa83-91 and aa96-118. Previous mutagenesis studies of these sites revealed that these amino acids bind to the complementary strand of the recognition site which contains the thymine basepair to the targeted adenine sites, upon substitution of these residues DNA was still able to bind to the enzyme but no activity was observed. Further mutagenesis studies identified further DNA-binding regions in the TRD domains; S149-K148 in TRD 1 and T354-K363 in TRD 2 which bind to the major groove of dsDNA which is where the DNA recognition site is located, figure 1.7. Notably the model revealed that the amino acids that are in contact with the adenine nucleotides targeted for methylation were N145 and Q357.

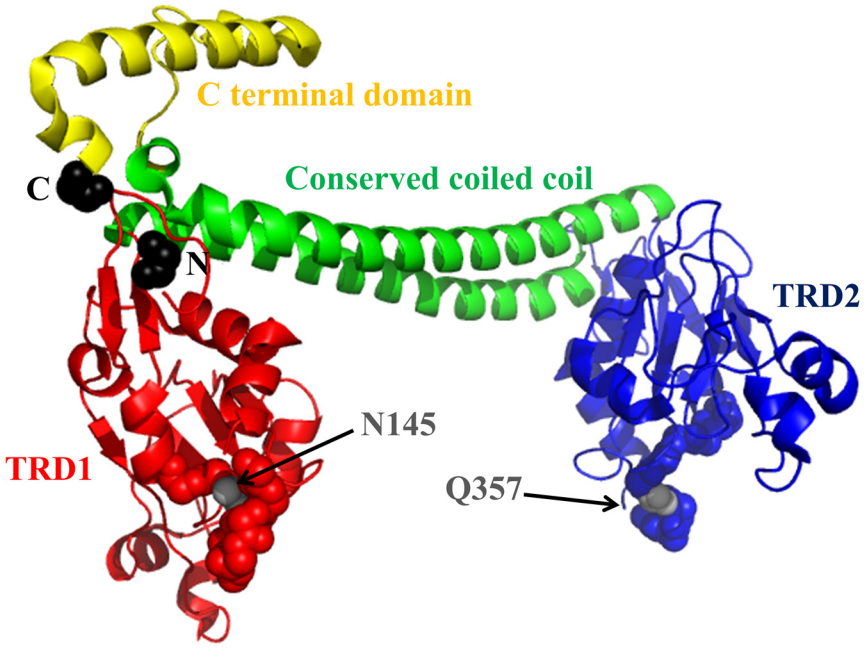


Figure 1.7: The structure of the HsdS subunit obtained from the M.EcoKI model by Kennaway (pdb:2Y7H). Each domain has been highlighted in red TRD1 and blue TRD2 which bind to the EcoKI recognition site on DNA (5'-AAC-$(N)_6$-GATC-3'). The residues thought to be directly involved are represented as spheres; for TRD 1 (S139-K148) and TRD 2 (T354-K363). The grey spheres pinpoint the two residues N145 and Q357 which target the adenine basepairs for methylation. In green is the conserved coiled coil region that determines the length of the non-specific residues.[45]

1.4.3 HsdM with DNA

It is thought that the HsdM subunits play two roles in its interaction with DNA. Firstly the N-terminal domain acts as the interface for the hinge mechanism of M.EcoKI, which allows the flexing of the open and closed state of the enzyme that allows the DNA to bind. Mutations termed the m* mutations targeted residues within this N-terminal region (aa1-153) and on helix G64-L73 which is located at the HsdM:HsdM interface. The m* mutations are so called because the enzyme has lost its preference to bind to hemimethylated sites and so the enzyme is able to bind and modify both hemi-methylated and unmethylated DNA sites. Methylation and restriction processes are not affected therefore these mutations allow the enzyme to produce an r^+m^* phenotype (restriction proficient and methylation proficient enzymes for unmethylated and hemi-methylated sites). Additional mutations are located in the central catalytic domain of the HsdM subunit producing long distance conformation changes that alter the local N-terminal region. Mutations within the m* domain have been shown to induce an r^-m^* (restriction deficient and methylation proficient enzymes for

unmethylated and hemi-methylated sites) suggesting the possibility of an interaction between the HsdM N-terminal domain and the HsdR subunits, see figure 1.8.

The DNA recognition site not only binds to the HsdS subunit but also the HsdM subunit where adenine base flipping is initiated.The adenine base is flipped into the catalytic domain of the HsdM subunit where it is modified with a methyl group donated by the cofactor SAM. The catalytic domain of the HsdM subunit consists of the AdoMet binding motif I for cofactor binding. However, mutations at N266 and F269 from catalytic domain IV reduce the methylation activity of the MTase, figure 1.4 shows that these residues are located next to the adenine target sites.[45]

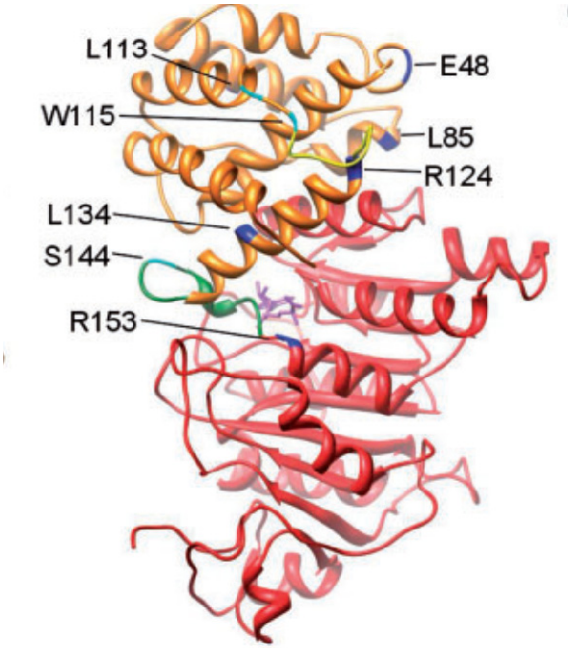


Figure 1.8: The m*mutations on the HsdM subunit taken from the Kennaway model of M.EcoKI (pdb:2Y7H) The central domain is highlighted in red and the N-terminal domains in orange the C-terminal domain has be omitted for clarity. Mutations that cause the r+m* are shown in blue (E48, L85, R124, L134 and R153) and r-m* are in cyan (S144, W115, L113). Loops that were disordered in the crystal structure are shown in green and yellow and the cofactor Adomet is in purple located in the catalytic domain of the HsdM subunit.[25]

1.5 REase Complex

After revealing the M.EcoKI structure, Kennaway *et al* (2012) resolved the structure of the REase complex for the EcoKI enzyme with and without DNA.[25] The structure confirmed the location and assembly of the HsdS and HsdM subunits of the M.EcoKI enzyme revealed on

the M.EcoKI model by Kennaway et al (2009).on the earlier M.EcoKI model. However the 2012 model also revealed the location of the HsdR subunits and the pathway of DNA through the complete enzyme. Although the EM models produced are low resolution in comparison to a crystal structure; extensive SAXS and SANS analysis, the known structures of crystallised subunits and the large amount of biophysical and biochemical data on the enzyme were used to produce a model which orientated the subunit structures to fit the EM model provided, figure 1.9.

Although the HsdR subunits were well-defined, their orientation in the 2012 model remained ambiguous and the path of the DNA through the complex was obscured. However the crystal structure of the RecA-like motor domains of the chromatin remodeling translocase SW12/SNF2, which acts with a similar mechanism, was used to align the location and direction of the DNA path, by overlaying the crystal structure of the translocase with the HsdR subunits. The DNA path forms an S-like shape through the EcoKI complex with the M.EcoKI core flanked by the two HsdR subunits. DNA footprinting experiments and AFM studies on the translocation mechanism of EcoKI indicate that the pathway of the DNA through the enzyme is kinked.[46,38,43] The orientation of the HsdR subunits allows for the formation of supercoiled DNA structures which occur as a result of HsdR subunits translocating the DNA.[29]

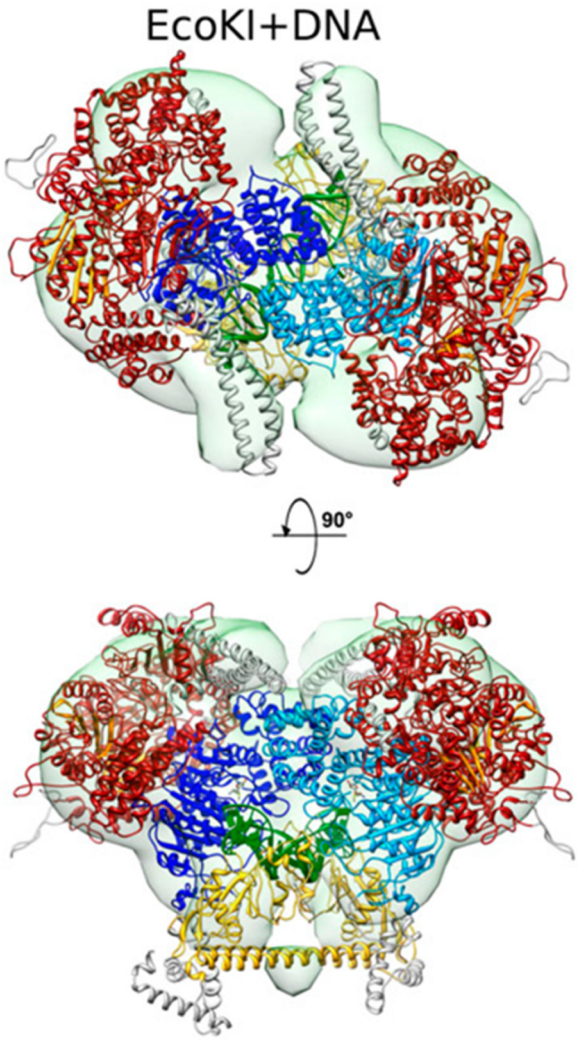


Figure 1.9: Extracted from Kennaway *et al* 2012.[25] Atomic model of EcoKI+DNA DNA is coloured green (but is not shown bound to each HsdR for clarity), HsdM subunits are light and dark blue, the HsdS is yellow and the two HsdR subunits are shown in red, with the β sheets of the recA-like motor domains in orange and residues modelled *de novo* are shown in grey. The HsdS and HsdM have been modelled from the MTase structure (pdb: 2Y2C) and the HsdR modelled on those from EcoR124I (pdb: 2W00).[25]

1.6 Mobile genetic elements

Genetic diversity and evolution among the prokaryotic community is driven by a phenomenon known as horizontal gene transfer. Horizontal gene transfer (HGT) occurs due to the transmission and acquisition of genes from across different species as a result of mobile genetic elements. The acquisition of these genes occurs through the transfer of mobile elements between different species by means of transformation, conjugation or transduction.[31]

As a result HGT is responsible for the spread of pathogenic and environmentally hazardous genes and is a factor in the spread of antibiotic resistance. However, HGT is prevented, in part, by the presence of Type I RM systems. Type I RM systems are widespread among bacteria and archaea and thought to act as an innate immune response in bacteria against the introduction of foreign DNA in the host cell. Nevertheless HGT is rife and therefore the transmission of genes among bacteria is high, suggesting that RM systems do not appear to be efficient protectors of the cell. This is because mobile genetic elements have evolved to produce a number of antirestriction strategies which encourage the transmission of the genes across species.[2] Most of the strategies adopted either modify the phage genome itself or produce accessory components which help the phage DNA to evade Type I RM systems.

For instance, phage DNA can avoid having DNA recognition sites. Counter selection upon phage and plasmids has reduced the number of target sites on the phage genome. Phage can also modify basepairs. Examples of basepair substitutions include thymine to 5-hydroxymethyluracil in *B. subtilis* phage, T-even phage substitutes hydroxylmethylcytosine for cytosine and the *mom* gene found in Mu phage modifies adenine to N6-(1-acetamido)adenine.[2]

In addition, a number of accessory proteins have been identified, which when co-injected upon conjugational infection, bind to the phage DNA to occlude the recognition sites from the endonuclease. This allows the subsequent modification of the phage DNA. Examples include the DarA and DarB proteins from phage P1.[47] Other strategies include the hydrolysis of cofactors or interaction with the RM systems; phage lambda produces the Ral protein which stimulates the host MTase and suppresses the REase activity.[48] lambda prophage found on the *E. coli* genome also encodes the Lar protein which has an equivalent role towards the Type IA RM system EcoKI.[2,49]

There are also a number of antirestriction proteins that take a more direct approach; these proteins act as competitive inhibitors of Type I restriction enzymes by mimicking the size and shape of the B-form bent DNA. The most well studied antirestriction proteins include the Ocr protein from T7 bacteriophage and the Orf18 ArdA protein from the conjugative transposon Tn916. Both of these antirestriction proteins provide non-specific inhibition of ATP dependent Type I RM systems.[1,13]

1.7 Ocr

In 1973, Studiers laboratory began publishing a series of investigations characterising the bacteriophage T7. By 1975, Studier had shown that phage T7 and T3 were susceptible to restriction from the RM systems upon mutations in gene 0.3.[50,51] However it was investigations led by Kruger *et al* that showed that gene 0.3 protected phage T3 and T7 from a number of 'classical' restriction systems, namely *E. coli* Type I RM enzymes; O, B and K. The exact mechanism of antirestriction would not be known for a number of years, however; in 1977 gene 0.3 was labelled the 'overcome classical restriction' or Ocr gene function.[52,53] A similar antirestriction function had been discovered previously which resides in gene 0.3 of T3 phage. For T3 phage, this gene also translated a SAMase protein to cleave SAM.[54,55] Gene 0.3 is located early in the viral genome and is one of the first phage related genes to be induced upon infection. By 1981 the Ocr protein expressed from gene 0.3 of T7 phage was successfully purified and its sequence determined.[56] The Ocr protein consists of 117 amino acids of which ~30% are glutamic acids or aspartic acids conferring net negative charge on the protein.

It was not until 2002 that the structure of the protein was solved, figure 1.10.[46] The structure shows that the Ocr protein produces a stable homodimer in solution and this dimer mimics 24bps of bent B-form DNA. The structure is dominated by three helices; helix A (aa7-24), helix B (aa34-44) and helix D (aa78-106). A fourth helix, helix C (aa49-57) is located at the dimer interface.[46] The structure confirmed much of the biochemical data that had been collected previously. DNA mimicry was produced by a constellation of acidic residues upon the surface of the protein that mimicked the placement of a DNA phosphate backbone as demonstrated in figure 1.11. A central kink at the point of the dimer interface was observed. This bend in the homodimer was substantial, producing an angle of ~34°.[46] The kink is produced upon formation of the homodimer and provides the protein with a banana shape that is indicative of the shape that is induced by DNA upon binding to Type I RM systems.[57,2] AFM has been used to determine that the bend induced by DNA upon its interaction to the Type IA enzyme EcoKI is on average 46°.[57,46]

The antirestriction activity of Ocr inhibits all of the RM families, Type IA-D. This suggests that the binding of Ocr does not target the specific recognition domains but the electrostatic surfaces of the RM enzymes.[46]

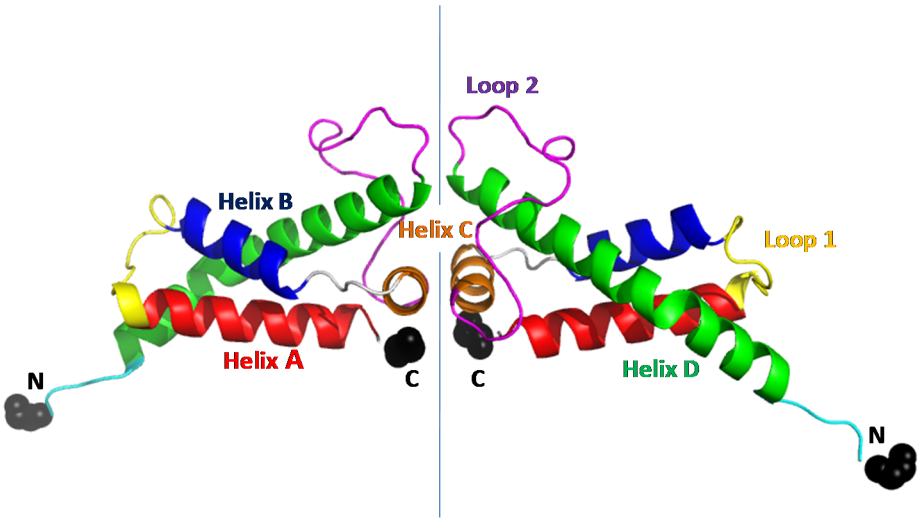


Figure 1.10: The structure of the Ocr homodimer. Each of the domains on the structure has been differentiated by a different colour and the N and C terminal residues are highlighted by black spheres.

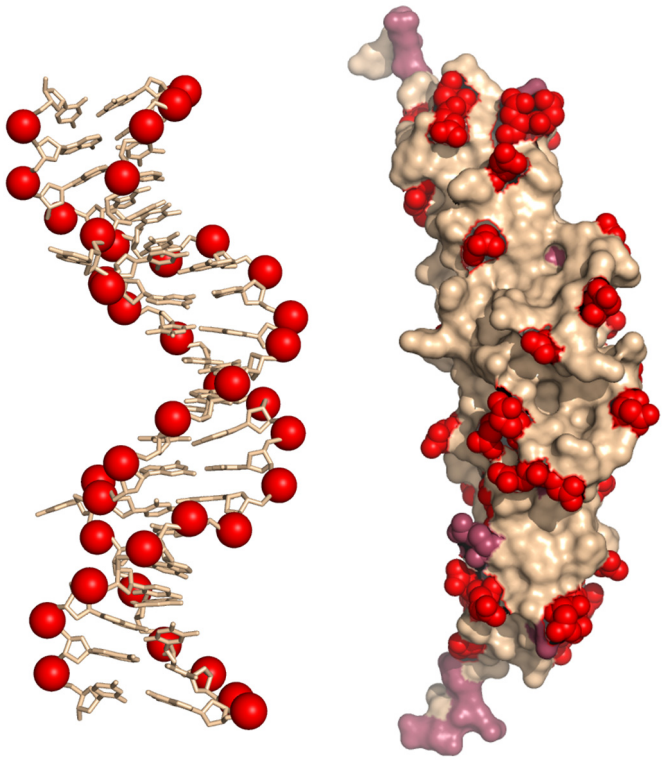


Figure 1.11: The DNA mimicry of the Ocr protein compared to the phosphate backbone of DNA bound to M. EcoKI, from the model by Kennaway *et al*, 2012. The acidic residues have been highlighted as red spheres on the structure of the Ocr homodimer (pdb:1S7Z) as determined by Putnam and Tainer in 2005.[62,58] Dark red spheres indicate additional acidic residues that are present on the structure. The red spheres that mimic the phosphate backbone of DNA clearly display similarities to the major and minor groove of DNA.

Isothermal titration calorimetry (ITC) analysis of the interaction between the MTase of EcoKI and the wtOcr protein shows that the interaction is a highly favorable exothermic reaction with a ΔH of -85.8kJ/mol and a stoichiometry of 1:1 dimer: M.EcoKI. However, the entropy of the interaction was unfavorable, thought to be due to the loss of rotational freedom of the Ocr protein upon binding. The binding affinity of the interaction was too tight to be determined by ITC data, but fluorescent anisotropy experiments estimated that the binding affinity of the interaction was ≤100pM.[57]

The same series of experiments showed that the Ocr protein also interacts with other derivatives of the EcoKI complex. Tryptophan fluorescence was used along with fluorescein tagging of the molecule to determine the binding of the Ocr towards the R.EcoKI nuclease. This interaction showed tight binding with a stoichiometry of $2Ocr_2$(dimer): EcoKI. In addition, the inactive M_1S_1 derivative of the MTase complex produced a tight interaction with Ocr, whereas the Ocr dimer with the M subunit alone exhibited a much weaker interaction.[59,60,61] Fluorescein tags that were distributed across the whole surface of the Ocr protein were quenched upon binding of the Ocr protein to the M.EcoKI enzyme. This indicated that the binding interface covered a large surface area of Ocr.[60].

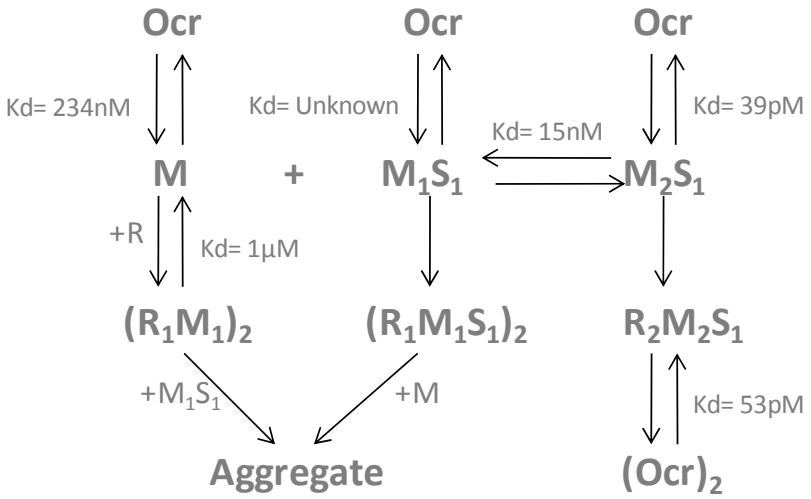


Figure 1.11: An equilibrium diagram for the *in vitro* assembly of the EcoKI and its interaction with the protein Ocr. Proposed binding affinities for the assembly of the EcoKI from its constituent subunits were a result of *in vitro* gel filtration measurements. Whereas, the interaction between Ocr and each of these complexes was measured by *in vitro* tryptophan fluorescence and fluorescent anisotropy. Although no quantitative measurement was measured for the binding affinity of M_1S_1 and Ocr they produce a tight interaction according to results from Atanasiu *et al*, 2002.[59]

A total of 34 acidic residues are present on the Ocr protein, many of which reside in the C-terminal/tail region. Truncations have shown that the last fourteen acidic residues located on

the C-terminal tail are nonessential for activity of the protein and even deleting the last 18 acidic residues only showed a slight reduction in the stability of the protein while retaining full activity.[57,60] In fact, alignment of the acidic residues from the Ocr homodimer with the DNA bound by the EcoKI enzyme, demonstrated that the recognition site corresponds to the amino acid residues D43 and D93 on the Ocr protein, Figure 1.12.

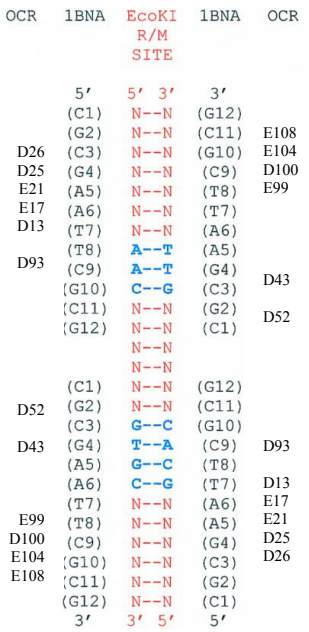


Figure 1.12: Alignment of the primary structure for the target recognition site of EcoKI and wtOcr from T7. The alignments highlight that D43 and D92 appear to align with the target adenine basepair. Atanasiu *et al*, 2001.[65]

Clearly, many of the acidic residues that reside on the Ocr protein, particularly within the region D43-D93, are important to the interaction of Ocr with the EcoKI enzyme. As a result, studies on the importance of DNA mimicry and the charge of the Ocr protein were investigated by A. Stephanou and colleagues to ascertain which acidic residues are important for the interaction between EcoKI and Ocr.[62] A number of mutagenesis studies were carried out; initially single mutants were targeted for a number of acidic residues that were distributed across the whole Ocr sequence.[62] Activity assays *in vivo* and *in vitro* showed that each of these mutants were active and only very minor, if any, differences were detected compared to the activity of the wtOcr protein. The largest difference was demonstrated by the mutants E16, D25 or D26, however these differences were barely detectable when accounting for the margin of error.[62]

As a result, a larger study was carried out to chemically modify the acidic residues on the surface of the Ocr protein. The chemical 1-ethyl-3-(3-dimethylaminopropyl)carboiimide hydrochloride (EDC) was used to modify the carboxyl groups present on the side chains of the two acidic residues Asp and Glu, producing a neutral and stable amide in each case. A series of reactions exhibited an increasing number of modifications. Analysis demonstrated that the removal of 46% (~16 amino acids) of the 34 acidic amino acid residues decreased the binding affinity by 50-fold.[57] This binding affinity was equivalent to the binding interaction between M.EcoKI and dsDNA. Analysis showed that protein stability and the protein fold were maintained. Binding between Ocr and the enzyme was still apparent even with an 86% loss of surface charge. The protein was able to bind although the interaction was much weaker; this binding was thought to be a result of the shape mimicry by Ocr of dsDNA.

1.8 ArdA

The ArdA genes were first discovered in the late 1980s. By the 1990s the first *ardA* genes were sequenced and the primary sequence showed that the protein was ~170 amino acids long with a total negative charge of -10 to -30. Sequence homology among conjugative plasmids and transposons demonstrated the widespread nature of the *ardA* gene among the community.[63] Like the Ocr protein, the ArdA protein also inhibits all of the Type I RM families. The activity of the ArdA protein and the Ocr protein were compared using a controlled promoter which controlled the expression of both antirestriction proteins.[64] The results showed that the antirestriction activity of the ArdA protein was 1700 times weaker than the Ocr protein, comparison of K_d values of both antirestriction proteins with M.EcoKI shows that Ocr has a much tighter interaction, at 1×10^{-10}M whereas the K_d for ArdA is 1.7×10^{-7}M.[65,66]

ArdA is one of the first genes transferred into the host cell upon infection. However unlike Ocr, Orf18 ArdA is found on conjugative plasmids/transposons not lytic bacteriophage. Where conjugative plasmids aim to produce a symbiotic relationship with the host cell upon infection, lytic phage do not require the presence of the host cell after replication of the phage virion. The lifetime of lytic phage infection is short with the one aim of producing more bacteriophage, even at a cost to the host cell.[63]

The analysis of the antirestriction and antimodification activity of several different ArdA proteins shows that the ArdA proteins exhibit different antirestriction and antimodification properties towards different Type I families.[67] For instance, ArdA proteins possess weak antimodification activity when interacting with Type IC enzymes in comparison to the activity demonstrated for other Type I families. Such differences could be for a number of reasons; the number of active complexes that reside in the cell at any one time differs for each Type I RM system or the expression of each ArdA protein may vary.[67]

In 2009 McMahon and colleagues solved the first structure revealing a rod-like structure for the Orf18 ArdA homodimer, as seen in figure 1.13.[68] The protein was a mimic of B-form DNA and the large number of acidic residues upon the surface of the protein corresponded to the sugar phosphate backbone of a 42bp unit of dsDNA. This elongated structure is composed of three domains.[68]

The third domain contains the dimer interface which is small and rich in hydrophobic residues. Upon dimerisation the structure produces the pre bent structure of the 42bp dsDNA identified from the crystal structure.[68] The central kink in the protein structure is reminiscent of the kink exhibited by DNA upon binding to the TRD sites on the HsdS subunits and that exhibited on the Ocr protein.

The structure of the ArdA protein has been aligned to a model of dsDNA bound to the MTase complex on EcoKI from Kennaway *et al*, 2009. The analysis carried out by McMahon *et al*, 2009, showed that domain 2 is located on the outer reaches of the TRD domains of the HsdS subunits and within the catalytic domain of the HsdM subunits.[68] Using the EM model proposed for DNA, the DNA bound to MTase was used to superimpose the DNA mimic, replacing the DNA in the model. This showed that the first domain did not interact with the MTase core complex as this domain 'juts' out from the MTase complex. In fact a loop was identified that linked domain 2 and domain 1. This link suggested a rotational freedom for the first domain; as such it is thought that this domain is likely to be in contact with the HsdR subunit. The binding interaction of the ArdA protein with the EcoKI enzyme has been measured at <1nM *in vitro* and 170nM *in vivo*. Although it is a tight binding interaction, it is not as tight as the Ocr: MTase interaction.

Truncations of the ArdA protein at the N and C terminal appear to have little effect on the antirestriction activity.[66] However, the hydrophobic residues V163A and F164A severely affected the antirestriction activity of the protein; this important region on the ArdA protein

is identified as a conserved motif among several homologues of the ArdA protein and is defined as the VF motif.

Single mutations for a number of polar residues thought to be important for the binding interaction of the ArdA protein were also examined by site directed mutagenesis. Many of these mutants reside in the C-terminal domain and their substitution with alanine caused no effect on the antirestriction activity of ArdA.

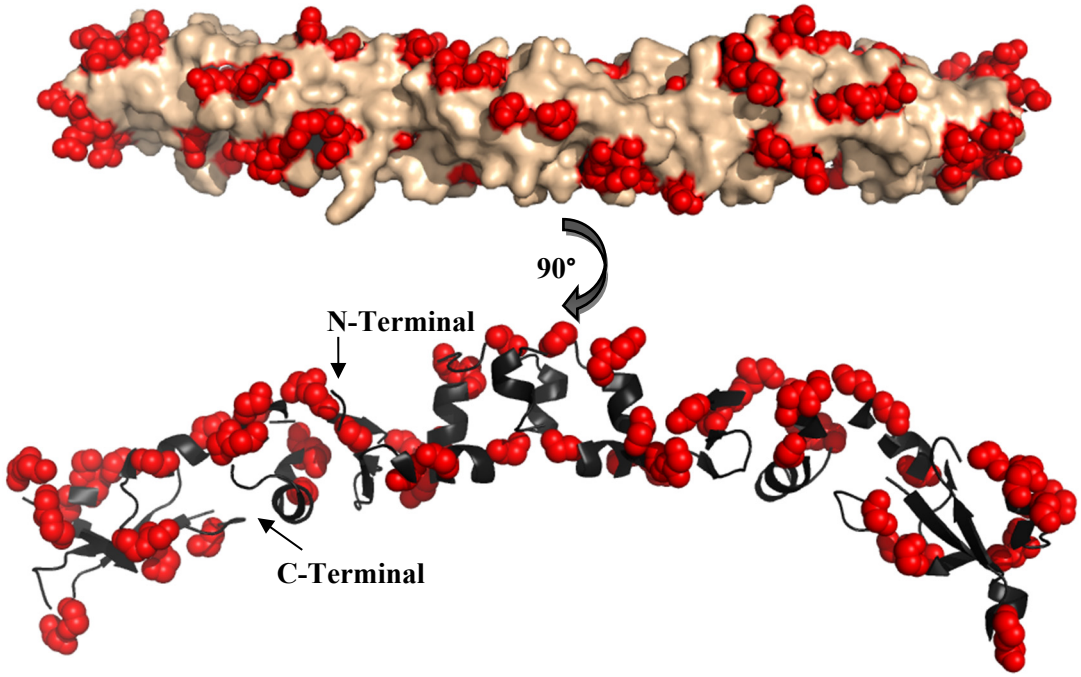


Figure 1.13: The crystal structure of the Orf 18, ArdA homodimer from the transposon Tn916. The acidic residues have been highlighted in red spheres and resemble the double stranded helix of B-form DNA.[68]

DNA mimicry is not just restricted to the antirestriction proteins, many other DNA mimics exist among the prokaryotic and eukaryotic communities.[69] The protein DinI mimics single stranded DNA and in doing so, inhibits the interaction of RecA with DNA. RecA is a protein which is involved in the DNA repair pathway. The inhibition of this protein could lead to the induction of the SOS response in the cell with deleterious consequences. Another example is the protein UGI which mimics kinked DNA like Ocr, with the inhibition of the uracil glycolylase. It seems that these DNA mimics provide a regulatory role.[69]

1.9 Acquisition of *hsd* genes and regulation of Type I RM systems

Hsd genes that specify Type I systems are readily transferable. Successful transfer has been carried out by conjugation, transformation and by P1 transduction with only a slight reduction in cell count observed.[70,71,72,73] Expression levels of each subunit were measured after conjugational transfer and no differences were observed for the expression of each subunit. The promoter for p_{mod} is slightly stronger then p_{res}, however both are weak.[30,72] As there appears to be no difference in the expression state or any regulation associated with the plasmid and the *hsd* genes, it suggests that the bacterial strain itself enforces a regulatory mechanism upon the acquisition of a new RM system.[1,64,74]

The conjugational transfer of both Type IA and Type IB to the recipient bacterial strain can cause a reduction in the cell count but overall the host cell survives. The transfer of a Type IC system appears to generate no loss in the host cell population.[75,76] This difference is thought to be associated with differential binding of the R subunits to the MTase core.[43,29,77,73]

Studies on the archetypal Type IA enzyme EcoKI show that more MTase is present in the cell then REase. However, as mentioned previously, the promoters are equally active and this reinforces the suggestion that the regulation may come from the recipient bacterial host.[78] The ability of the host cell to be unaffected by the introduction of new *hsd* genes is known as a state of restriction alleviation. This has been witnessed as a result of infection by phage, conjugation and transformation.[70]

1.10 Restriction alleviation mechanism

In 1985 researchers proposed a regulation mechanism for Type I enzymes. Extensive research carried out in the late 1980s by GB Zavilgelsky and AA Belogurov suggested that restriction alleviation was activated by the SOS response, upon exposure to DNA damaging agents like UV light or high doses of nalidixic acid. These DNA damaging agents lead to the production of unmethylated DNA which induces restriction alleviation.[79] This protects the host cell DNA from the indiscriminate restriction of unmodified chromosomal target sites caused by the repair of the damaged chromosome.[101,5,80]

1.11 2-Aminopurine

Belogurov *et al*. investigated the phenomenon of restriction alleviation using 2-aminopurine (2AP).[79] 2AP renders the chromosomal DNA unmethylated. 2AP induces damage in two ways; firstly, as an adenine analogue 2AP can be inserted as a replacement for adenine, and secondly as an analogue of guanine, see figure 1.14. Incorporation of the 2AP instead of adenine during DNA replication inhibits methylation of the adenine target sites as the RM enzymes are able to bind but unable to methylate the 2AP, therefore the DNA remains unmethylated and upon further replication the chromosomal DNA becomes unmethylated.[81]

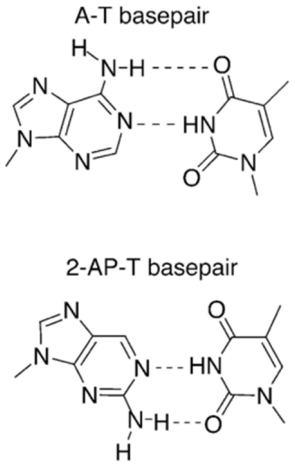


Figure 1.14: Comparing the Hydrogen bond interaction of; adenine with thymine and 2-aminopurine with thymine.

2AP can also act as an analogue for guanine and can induce mismatches in the chromosomal sequence. This correlates with the onset of restriction alleviation and the generation of unmodified target sites, which is only reaching its maximum at 1-1.5 hours for 2AP and 2-3hrs for UV damage.[81]

In the mid-nineties Bickle *et al*, (1996) showed that the restriction subunit, HsdR, was only present in small quantities in the absence of modification by 2AP.[8] However, it was two years later that a series of experiments by S. Makovets and N.M. Murray, identified the proteins ClpX and ClpP as regulating the transmission of the *hsd* genes.[82] In 1999 Makovets *et al*, demonstrated that the HsdR subunit of Type IA enzymes was a target for proteolysis by the protease ClpXP, and a necessary component for the restriction alleviation response, described in figure 1.15.[82] Makovets used phage to infect Clp^+ and Clp^- *E. coli* cells in to the presence of DNA damage caused by 2AP. Cell death followed treatments of the Clp^- strain

with 2AP and the formation of filaments. Cell filamentation is induced by the SOS response of the bacterial cells.[83,77,80] This is a direct response to the DNA damage inflicted on the cell and the filaments dissipate upon recovery of the cells from reversible DNA damage such as UV irradiation.

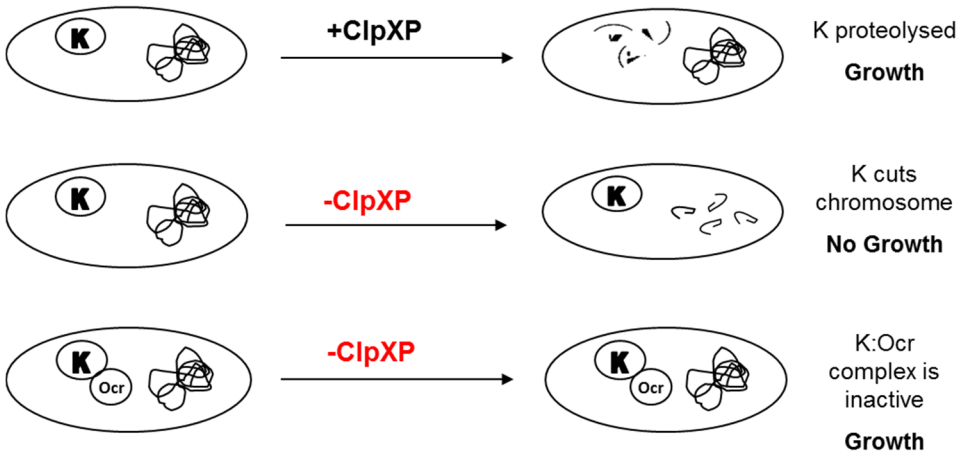


Figure 1.15: A schematic displaying the effects on cell growth when cells are exposed to 2AP in the presence and absence of the clp gene. *E. coli* cells when exposed to 2AP activate the ClpXP proteolysis of the EcoKI enzyme for continued cell growth. In the absence of the clp gene the protease cannot breakdown the EcoKI enzyme which targets the unmethylated chromosomal DNA upon exposure to 2AP causing cell death. However the presence of the Ocr protein forms a complex with EcoKI inhibiting its restriction activity allowing the cell to survive.

These studies showed that the presence of the ClpXP protease induced the alleviation of restriction. This was corroborated by experiments which showed that the concentration of the HsdR subunit was reduced while the quantities of the modification and specificity subunits remained the same.[82,3,80] Further mutagenesis of the HsdR subunit reveals that a ClpXP dependent degradation was active upon HsdR mutants that inhibit the DNA cleavage mechanism of the R subunit. However, ClpXP activity ceases if the HsdR subunits are unable to translocate DNA.[10,84]

Therefore, ClpXP degradation is initiated during the translocation of chromosomal DNA. However, how the ClpXP protease can differentiate between RM enzymes translocating on chromosomal and foreign DNA is still unknown.[64,24,78]

1.12 ClpXP protease

The ClpXP protease is compromise of the CpX ATPase and the ClpP protease.[85] The ClpP protease assembles into a complex heptameric ring that consists of 14 subunits which produce a peptidase chamber containing 14 peptidase sites. The site of entry to the proteolytic centre (ClpP peptidase chamber) is through the ClpX subunit.[86,87] The structure of ClpX has been solved and shows that the ATPase consists of two large hexameric rings that bind to the core of the ClpP peptidase heptameric ring, figure 1.16.[88] The hexameric rings of the ClpX component contain ATP and Mg^{2+} binding sites. Therefore, upon recognition of a substrate for proteolysis, the substrate is translocated though the ClpX rings into the ClpP peptidase chamber where the many peptidase sites degrade the protein into small peptides of 5-10 amino acids in length. These peptides then diffuse out into the cell after completion of the reaction.[89] The roles of the protease can be varied and suggests the regulation of a number of pathways. For instance, the ClpXP protease regulates the amount of HsdR present in the cell upon DNA damage.[15,78]

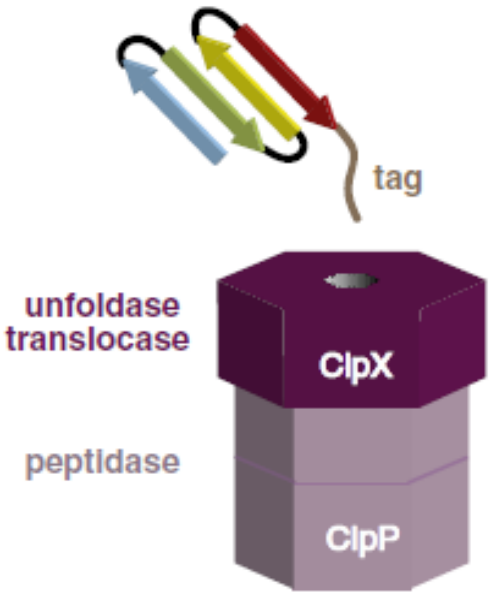


Figure 1.4 : Extracted from Sauer *et al*, 2012[90]a cartoon representation of the ClpXP protease. In dark purple the two hexameric rings of the ClpX ATPase stacked onto the large ClpP heptameric structure in lilac.

1.13 SOS response

In the event of DNA damage to bacterial cells, many bacteria induce a sophisticated regulon known as the SOS response. This pathway was first discovered over 100 years ago, however many components in its pathway still remain unresolved. The exposure of DNA to damaging agents such as UV irradiation and 2AP, which have been detailed previously, induce the SOS response which triggers the expression of numerous genes to repair the damage inflicted on the DNA.[91] Upon DNA damage, the repressor protein LexA is cleaved and this reduces it affinity for the operator sites of several genes which possess the 'SOS box'. These genes are distributed across the whole of the genome.[74] The removal of the repressor protein induces the expression of over 30 known genes and produces at least 16 known proteins.[92,93] Upon DNA damage the protein RecA binds to ssDNA and stimulates the autoproteolysis of the LexA repressor from the operator sites of the genes. Once the repairs have been completed by the cell, the RecA protein releases the ssDNA which decreases the proteolysis of the LexA protein and increases the amount of LexA which is then able to bind back onto to the operator sites. [80, 91,94,95]

The genes that are controlled by the SOS regulon are not expressed at the same time or to the same extent. In fact the SOS response of bacterial cells occurs as a cascade, each stage attempting more elaborate DNA repair designs. Firstly, UvrABD and the endonuclease UvrC catalyse the nucleotide excision repair (NER) for damaged nucleotides from dsDNA. This includes the targeted excision of single basepair mismatches. The second stage attempts homologous recombination. At the sites of damage, proteins are released that create double strand breaks (DSB) using the NER system, and the introduction of the DSB on the DNA means that homologous recombination can be performed to repair the DNA.[80,96] If damage is still present after 40mins then the SOS response initiates the induction of PolV (a DNA repair polymerase). These polymerases along with others that are induced by the SOS response repair the DNA at the risk of introducing errors onto the genome. LexA is continuously produced during the SOS response thereby enabling the tight regulatory inhibition of the SOS response as soon as repair has been managed.[91,97]

The regulation of the SOS response is determined by the protein DinI. DinI is a DNA mimic protein that mimics ssDNA and interacts with the RecA protein, thereby inhibiting the ssDNA: RecA filament.[94] Although more recent evidence suggests that instead of DinI interacting as a competitive inhibitor, the protein binds and distorts the RecA:ssDNA protein filament. Whichever mechanism is used, the DinI protein is part of a classical feedback loop

where the formation of dimers and tetramers of the protein bind nonspecifically to RecA/RecA: ssDNA, preventing its ability to induce the SOS response.[96,93,98]

1.14 Cell division and formation of filaments

So that bacterial cells have time to repair DNA, the induction of the SOS response not only releases genes that are appropriate for DNA repair but it also induces genes and proteins that inhibit cell division. By inhibiting cell division the undamaged portion of the genome is continually replicated, therefore the cells become multinucleated elongated cells.[96]

Conventionally in *E. coli*, cell division is initiated once the appropriate size is reached. Indentations are formed at the midpoint of the cell which marks the point of septation and the bacteria split into two daughter cells. Production of the septum is initiated by a series of *fts* genes, however mutations in the genes *fts* E, F, H, Q, Z and the protein FtsD can inhibit the production of the septum and as a result filaments are formed.[99]

Upon DNA damage, the SOS response stimulates expression from the lexA gene which expresses the protein SfiA. The SfiA protein acts as a structural inhibitor of the ftsZ gene which is involved in the early stages of septum formation. Upon completion of repair, the RecA protein is deactivated and SfiA is subsequently degraded by a protease. Providing that the cells successfully repair the DNA, the process of filamentation is reversible. However, the formation of the septum can shift from the midpoint and division occurs with the outer most cells of the filament.[100,101,77,94]

There are many examples of how the induction of the SOS response can inhibit cell division.[92] The inhibition of cell division and the formation of filaments can be a result of FtsZ inhibition. Not only by the proteins SfiA and SfiC but also YneA can delay the septation and separation of cells and inhibit division, by transiently binding to peptidoglycan.[102,101] The protein SulA has been shown to inhibit ring formation of the FtsZ protein, which inhibits division and produces cell filaments.[100] Furthermore, the interaction of proteins towards genes and proteins that are present in the SOS regulon could induce the SOS response and in turn stimulate the production of filaments via the inhibition of cell division.

The formation of filaments has also been commonly attributed to the overproduction of recombinant proteins in *E. coli*.[103] Such an accumulation of protein in the cell can cause the

cell to filament. This is usually because the accumulation of protein, especially thermosensitive proteins, can lead to the suppression of the FtsZ protein in the cell. The FtsZ protein is very sensitive to changes in the cell, as its interaction with FtsA is tightly regulated at a specified ratio.[99] The reduction in FtsZ activity appears to correlate with an increase in the number of inclusion bodies in the cells. Inclusion bodies are pockets in the cell which contain insoluble protein, usually as a result of the overproduction of an insoluble recombinant protein or a designed mutant.[103]

1.15 Protein libraries

1.15.1 Types of libraries

The production of libraries to gain an insight into the design of proteins has become a major research area. Understanding the sequence space that proteins can occupy and their ability to evolve new functions has been stimulated by the technique of directed evolution and the production of tailor-made protein and DNA libraries.[104] Libraries allow the screening and selection of a great number of proteins/mutants. This is advantageous if the researcher is analysing the global effects of mutations on a protein's structure, stability and function.[105] However, it is said that libraries are only as good as the selection assays that are designed to elucidate activity. Therefore, the selection method needs to be thought through and many of the limitations in the design of libraries resides in the capability of the screening assays.[106,107,108]

Generally the strategy when designing protein libraries is to explore as much sequence space as possible while minimising the number of deleterious mutants. The focus is generally to be able to explore sequence space by producing the number of mutations that confer a desired function.[105] However, the diversity of the library needs to be balanced with a minimal number of deleterious mutants. Generally the frequency of the beneficial mutants tends to include only 0.001% compared to >30% of all mutations inducing deleterious effects upon the introduction of mutations.[109]

Each library design has its own set of advantages and disadvantages which need to be considered. Therefore, the most difficult job is usually to select the appropriate library in the first instance. Library designs can be categorised into two strategies; rational design or random mutagenesis.[108,110]

The production of small libraries is usually a result of rational design; this is the production of a library which considers the location of the target sites on the proteins structure in combination with the functional outcome desired. These libraries target specific regions or positions with the aid of extensive structural knowledge of the protein and complex. This leads to a decrease in the number of assumptions and predictions that need to be addressed by the researcher and therefore the smaller the library size.[110,107]

Generalisations can be made about the structure-function relationship of many proteins however; the ability to predict changes in structure is limited. In this case, the production of large libraries using random mutagenesis or saturation mutagenesis is more advantageous. Such libraries are required when the structure of the proteins are not well defined.[111] An example is an adaptation of conventional alanine scanning of a protein; ordinarily shotgun phage display has been used to mutate single positions to alanine to study the effects on the activity of a protein. However Sidhu and co-workers designed a library which expanded alanine scanning into a saturation analysis, which included permutations of all 20 amino acids but at only a few selected positions. This was applied to the interaction between human growth hormone and the human growth hormone receptor, and revealed that the interaction was less tolerant of chemically conserved residues.[112]

1.15.2 Analysis of libraries

Selection methods are ideal for large libraries; large libraries are designed to investigate whether the desired function of the protein remains active or changes upon mutation. As such, only mutants that exhibit the desired function are of interest. Selection methods can screen larger numbers by being able to select for the desired function of interest only.[109] In such cases the design of the selection assay is very important as assays may not consider changes in the expression level, the solubility and the induction capability of each protein. Such properties of the library variants will affect their selection capabilities, causing the selection of artifacts rather than selecting those proteins/mutants that possess the greatest enzymatic activity.[105]

In contrast, smaller libraries are generally designed to investigate the effect of particular mutations to the functionality of the protein.[110] In this case the researcher is interested in the changes that are induced on the protein as a result of the mutation. Therefore selections of isolates taken at random from the library are analysed. As a result, screens can provide an assessment as to the diversity of the library by screening each mutant regardless of activity and monitoring a number of activity parameters of the protein.[108] A study to analyse tumor

derived missense mutations on the function of the protein p53 was used to correlate the function of the protein with structure and tumor derived mutations. To do this, PCR techniques were used to incorporate 2314 basepair mismatches into wt cDNA; this covered all possible single point mutations.[113]

Advances in screening capability have been driven by recent developments in the production of microfluidic devices and droplet manipulation which allow for high throughput screening assays. However, at the current stage of development and design, the capacity of many selection and screening methods is incapable of screening the diversity of the library under investigation.[107] This issue is common and a number of options are available to researchers.

1.15.3 Library design considerations

The screening of only a small fraction of the library is only advantageous if functionally active proteins/mutants can be selected from this small fraction of the library. The success of screening is dependent on the number of active mutants. If the probability is high, then screening a small fraction from the library can successfully elucidate a functionally active protein to analyse. Hecht and colleagues have designed a number of *de novo* protein libraries.[114] These libraries contain non-natural proteins which are designed in a fashion which allows them to possess secondary structure.[114] These libraries allow an exploration of the potential of sequence space beyond natural proteins to gain a greater understanding of the evolve-ability and adaptability of protein structure and functions.[105,115,116] The complete screening of such libraries itself would be a huge task. Due to the nature of the protein, only a small fraction of the library needs to be screened to elucidate a number of mutants that elicit peroxidise activity, inhibit cell growth and encourage amyloidal formation. [109,117,118]

Another option is to target each of the target positions in an iterative manner. The idea is that a small group of mutations will be targeted. Those which appear to be advantageous or neutral are then kept for another round of mutations and in this way selection will be carried out after each cycle. However, this strategy excludes ‘epistatic combinations of mutations’; mutations that are dependent on the presence of another mutation to provide activity. This is particularly difficult to predict and can usually only be resolved by the production of a random library[107].

With ever more breakthroughs and techniques in the field of protein engineering, a number of new strategies are emerging that bridge the gap. The exploration of a large number of residues is balanced by minimising the number of mutations that are present and therefore

the size of the library. An example of such a library is the creation of ISOR (Incorporating Small Oligonucleotides via Reassembly).[108,119]

1.16 ISOR library design

The ISOR library methodology was created to offer parsimonious mutagenesis; this is the partial diversification of each residue that is targeted. Rather than creating a saturated library where the effect of all amino acids are investigated, by using a restricted codon set, a smaller number of residues are substituted in each position.[119,108]

ISOR relies on a simple gene shuffling technique; the library can introduce insertions, deletions or substitutions by designing specific oligonucleotides for each position of interest with the appropriate codon. These oligonucleotides are then shuffled with fragments of the wild type gene. Upon reassembly, the gene incorporates the oligonucleotides and this produces libraries where specified positions have been mutated in a random manner. The libraries should represent each of the mutations that were targeted, but each variant of the library produces a random sequence of the gene with an entirely unique subset of mutations. The ISOR libraries created by Herman, Tawfik and colleagues have shown how the libraries have been implemented in the design of a series of libraries which target the DNA MTase *M*.HaeIII and the protein serum paraoxonase (PON1).[119,107]

A series of libraries were produced to target 45 base pairs for substitution on the M.HaeIII protein. By using different oligonucleotide concentrations the average mutation rate of the library could be varied. This created a series of libraries of which the average number of mutations ranged from 1-6 per library variant at each of the 45 possible target sites. This produced a huge library, however the diversity is limited compared to saturation mutagenesis and combinatorial libraries. Therefore a selection method was used to analyse the M.HaeIII activity of the library. 10^{10} library variants were screened by selecting for mutants with reduced methylation activity. Isolates of the library were then screened to identify library variants which exhibit a range of methylation activities.[119]

ISOR has also been used to generate a series of libraries that introduce indels of insertions and deletions to investigate the structural features of the PON1 active site. The library was designed in much the same way as the M.HaeIII library. The library was designed to introduce oligonucleotides with indels and therefore the oligonucleotides were designed with deleted and inserted codons to a particular region on the protein shown to contribute to the

activity of the paraoxonase. A number of the variants from the library were isolated and tested. From the small number of mutations screened, each provided variable activity. However, mutations in one particular domain (Helix 3) appeared to correlate with a reduction in activity. This highlighted the importance of this domain to the esterase activity of the serum paraoxonase.[119,107,108]

Thesis Aims and Objectives

This project will attempt to answer the following questions:

1. What is the minimal number of acidic residues the Ocr protein needs to perform as a fully functional antirestriction protein?
2. What, if any, is the importance of these residues and what can they tell us about the interaction of the Ocr protein with the Type I RM system, EcoKI?
3. How does the decrease in negative charge affect the protein's stability and fold and its impact on the host-pathogen interaction?
4. Can differential antirestriction and antimodification activity be designed into the Ocr protein by reducing the charge?
5. Can one explain the evolution from an inactive protein to a fully functional Ocr protein by examining the charge to function relationships?

It is hoped that the answers to these questions will increase our understanding of the relationship between RM systems and HGT and provide a window into the 'game of bluff' played out between bacteria and mobile genetic elements in the protection and the evolution of bacterial species.

To answer these questions, this project had the following objectives:

1. Design and create a library of Ocr multimutants where each mutant possesses a unique sequence and surface charge.
2. To develop a selection assay that can be used to screen the library for mutants that exhibit antirestriction activity and obtain the minimum number of residues that are required for activity.
3. To purify and characterise mutants with minimal surface charge that exhibit antirestriction activity by assessing their protein fold, stability and function.
4. To screen additional mutants from the library at random, to assess the effect, if any, of negative charge in every aspect of the proteins existence. Therefore its effect on function, stability, fold, solubility and expression.

Chapter 2. Materials and Methods

2.1 Bacterial strains, phage and plasmids

E. coli Strain	Relevant Genotype	Reference or Origin
NM1261	MG1655 *ΔhsdR*$_{EcoK1}$.*:cat*	
NM1041	MG1655 *clpX ::kan*	Blakely *et al*. (2006)
NM1049	MG1655 *lar.:cat*	Blakely *et al*. (2006)
NM1057	MG1655 *clpX ::kan, ΔhsdR*$_{EcoK1}$	Prepared byDr A. Dawson
NM1056	MG1655 *clpX ::kan Δ(hsdS*$_{EcoK1}$*-mcr)*	
MG1655	*F$^-$ λ$^-$ ilvG rfb-50 rph1*	Blattner *et al*. (1997)
Phage		
$\lambda_{v.0}$	λ*vir, (unmodified DNA)*	Donated by
$\lambda_{v.k}$	λ*vir, (modified DNA)*	Prof. N. Murray
Plasmids		
pTrc99A	Carbenicillin Resistant	Donated by Dr. A. Dawson

2.2 DNA Techniques

2.2.1 Subcloning of Ocr multimutants into pTrc99A

The ocr multimutants that were donated by Dr A. Stephanou were spliced out of pET24b and inserted into pTrc99A.

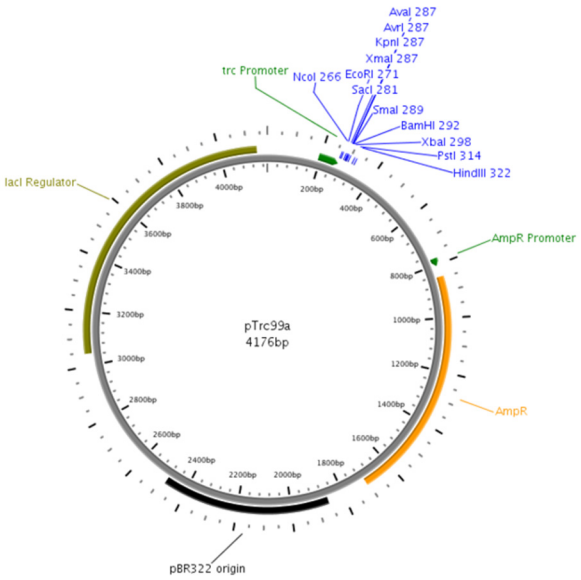


Figure 2.1: A plasmid map of the pTrc99A. This vector was used to insert Ocr multimutants using the restriction sites NcoI and BamHI. To note, in order to insert the POcr plasmid library into pTrc99A the POcr gene was inserted using the restriction sites NcoI and HindIII details in section 2.4.

Linearised pTrc99A plasmid vector

For preparation of the linearised pTrc99A plasmid; *E. coli* DH5α cells transformed with pTrc99A were used to inoculate 100ml of LB broth and incubated at 37°C, 200rpm for 12hrs. Plasmid DNA was extracted from the resulting cell pellets of pTrc99A using a QIAgen Midi prep kit and the DNA was quantified using UV-Vis spectrophotometry. The pTrc99A plasmid was linearised by carrying out a double digestion reaction using the restriction enzymes NcoI and BamHI, the digest reaction mix (below) was incubated at 37°C for 4hrs.

Digest Mix	
10(×)NEB Buffer 4	10µl
MQ Water	Make to 100µl
0.1M BSA	10µl
pTrc99A DNA	~4µg
NcoI	5U
BamHI	5U
Total	100µl

The linearised plasmid was purified by extracting the relevant band from a 1% Agarose DNA gel using a QIAgen DNA extraction kit.

Preparation of the Ocr mulitmutant gene inserts

To splice the ocr multimutant genes from pET24b to pTrc99A, the genes were amplified with a simple amplification PCR reaction, using primers with the NcoI and BamHI sites. These primers were obtained from invitrogen.

Reaction Cycle		
Temp (°C)	**Time**	
94	2mins	
94	30secs	30 Cycles
55	30secs	
72	1min	
72	2mins	
4	store	

PCR Reaction Mix	
Mq Water	Make to 100µl
10(×)pfu Buffer	10µl
Ocr DNA Template	2µl (at ~50ng/µl)
Primer Fwd	1ng/µl
Primer Rev	1ng/µl
10mM dNTP Mix	3µl
pfu DNA polymerase	2.5U
Total	100 µl

The amplified genes were purified using the QIAgen PCR Purification KitA double digest reaction was carried out as per the manufacturer's guidelines (New England Biolabs, Ipswich, MA). Typically, 5µg of an amplified gene was digested with the restriction enzymes NcoI and BamHI, the 50µl reactions were incubated at 37°C for 2.5hrs. The digested DNA was purified using a 1% DNA agarose gel and the QIAgen gel extraction kit.

Ligation Reaction of pTrc99A and ocr multimutant gene insert

Each ocr multimutant gene was ligated to previously linearised pTrc99A. These reactions were setup at room temperature for 2.5hrs and a typical reaction mixture is shown below.

Ligation Mix	
	Vol (µl)
MQ Water	Make to 20µl
pTrc99A	~100ng
DNA Insert	~500ng
10(×)Ligase buffer	2µl
T4 DNA ligase	10U
Total	20 µl

Transformations of the ocr multimutants in the pTrc99A vector were performed with *E. coli* DH5α competent cells using a similar procedure as outline in section 2.2.3. Successful colonies were isolated and overnight cell cultures were grown to isolate the plasmid DNA using a Qiagen Mini Prep kit. To ensure that each ocr multimutant had been successfully transferred to pTrc99A, there sequences were verified by Sanger DNA sequencing carried out by Genepool, Edinburgh.

2.2.2 Chemically competent cells

An overnight culture of *E. coli* NM1041 cells was subcultured (1/200) into LB broth and incubated at 37°C, 200rpm. At an OD_{600} of 0.4-0.6, the 50ml culture was centrifuged (13,000rpm, 10mins) and the supernatant removed, the resulting cell pellets were placed on ice. The cells were resuspended with ice-cold 0.1M $CaCl_2$ (20ml) and the cells were incubated on ice for 40mins and centrifuged (13,000rpm, 3mins). The pellets were resuspended with ice-cold 0.1M $CaCl_2$ (10ml), a second time (10ml) and incubated for a further 45mins on ice. The competent cells were stored at -80°C with glycerol (~20%) until required for transformations.

The same protocol was carried out for all the competent bacterial cell strains used in this thesis. Furthermore, for large quantities of transformations, the competent cells (200µl) were transferred into sterile, 2ml deep, 96-well plates (USA scientific) before storage at -80°C.

2.2.3 Bacterial transformation

Aliquots (200µl) of *E. coli* competent cells were thawed on ice. Each of the newly ligated constructs (2µl) was added to an aliquot of cells in a microcentrifuge tube and left to incubate on ice for 30mins. The microcentrifuge tubes were heat shocked at 42°C for 30s. The cells were incubated with (0.8ml) LB broth at 37°C, 200rpm for 1hr. The transformation mix was plated out onto (200µl) LB-Agar plates supplemented with carbenicillin (50µg/ml). Once the plates were dry they were inverted and incubated overnight (~12hrs) at 37°C. Single colonies were restreaked, on fresh LB-Agar plates and overnight cultures were used to purify the plasmid DNA using a QIAgen mini prep kit.

To verify the successful ligation of the ocr multimutant inserts into the plasmid, pTrc99A, sanger DNA sequencing was carried out by Genepool, School of Biological Sciences, Edinburgh.

2.2.4 *In vivo* restriction alleviation assay

E. coli NM1041 cells were transformed with each of the ocr multimutants in pTrc99A, in order to perform the restriction alleviation assay. Transformations were carried out as outlined in section 2.2.3. However, due to the low transformation frequency of *E. coli* NM1041 cells heat shock was carried out for 1min 42secs. A single colony was used to inoculate an overnight culture. Overnight cultures (100µl) were diluted to 10-fold increments from 10^{-1} to 10^{-8} and viable counts were spotted in triplicate onto LB-agar plates containing different concentrations of 2AP: 0, 20, 40 or 80µg/ml. The plates were supplemented with, kanamycin (30µg/ml) and carbenicillin (50µg/ml) or kanamycin (30µg/ml), carbenicillin (50µg/ml), IPTG (50µg/ml) and 2AP. Once dry the plates were inverted, sealed in an air-tight container and incubated at 37°C for 18hrs, after which colony numbers were counted and the number of viable cells per ml calculated.

Viable counts were compared to a negative and positive control; the empty vector pTrc99A and either wild type Ocr or wild type ArdA Orf18, which werealso tested alongside each restriction alleviation assay performed.

2.2.5 Cell growth plate assay +/- hsdR gene

E. coli NM1041(r^+m^+, ClpXP$^-$) and *E. coli* NM1057 (r^-m^+ HsdR$^-$,ClpXP$^-$) were transformed with various plasmids expressing Ocr or its mutants. Cultures of the bacterial strains were grown in L Broth with aeration at 37°C and shaking at 200 rpm for 18hrs. Ten-fold serial dilutions of the cultures were spotted onto LB-agar plates, supplemented with Kanamycin (30 mg/ml), Carbenicillin (50 mg/ml) and IPTG (50 mg/ml). The plates were incubated for 18hrs at 37°C and the resulting colonies (viable counts) were calculated.

2.2.6 Bacterial growth curves

Overnight cultures of *E. coli* NM1041 (r^+m^+, *clpX*$^-$) transformed with each of the plasmids expressing Ocr or its mutants were incubated at 37°C. Samples (0.2µl) were taken from each culture and used to inoculate 1ml L Broth, supplemented with or without IPTG and increasing concentrations of 2AP: 0, 20, 40 or 80µg/ml. All of the cultures were in the lag phase. Sterile covered microplates (96-well plates) and lids were obtained from Greiner. 200µl samples from the prepared cultures were pipetted, in triplicate, into the 96-well plates. Bacterial growth curves were obtained by measuring the optical density (λ= 600nm) of the cultures as a function of time. The optical density measurements were carried out by a Fluorostar OPTIMA plate reader (BMG Labtech). The reader was programmed to measure the samples on the plate every 5min for 14hrs at a wavelength of 600nm. The reader

maintained a temperature of 37°C throughout the experiment and a medium shaking speed of 200rpm between readings.

The same protocol and conditions were used to carry out all of the different types of bacterial growth curves covered in this thesis. The bacterial cell strain, and the chemicals/antibiotics used to supplement the cultures were dependent on the experiment taking place.

In addition, large scale bacterial growth was also carried out for overnight cultures. 2ml deep, 96-well plates were used to aliquot LB broth (96×1ml) which was subsequently inoculated with the appropriate antibiotics and a swab from a single colony, the plate was incubated overnight at 37°C, 100rpm.

2.2.7 Microscopy

Overnight cultures of *E. coli* NM1041 (r^+m^+, $ClpX^-$) cells, transformed with wild-type Orf18 ArdA or the mutant derivatives, were used to inoculate 10ml of L Broth (50µl). The cultures were grown to an optical density of 0.2 at 600nm and subsequently induced with IPTG (50µg/ml) and incubated for a further 2.5hrs (the OD_{600} was measured to ensure each culture was still in the logarithmic phase of growth). To condense the nucleoid, chloramphenicol (10µl) was added to each culture and incubation at 37°C continued for 30min. The bacterial cells were fixed with the addition of methanol at a 2:1 ratio and a further incubation on ice for 10min. The cells were collected by centrifugation and resuspended in buffer; 10mM Tris–HCl (pH 7.5) 10mM $MgSO_4$ and observed under the microscope. For DAPI-stained samples 1µl of a DAPI stock of 1mg/ml in EtOH was added to the fixed cells and the cells were observed under a fluorescent microscope.

2.3 Phage techniques

2.3.1 Phage ($\lambda_{V.0.}$) amplification

To amplify phage $\lambda_{V.0}/\lambda_{V.k}$ BBL soft agar (3ml) was equilibrated at 45-46°C, the phage was diluted with phage buffer to give a moi (multiplicity of infection) of 10^3 infectious virus particles per cell. 100µl of the diluted phage was combined with 100µl from an overnight culture of *E. coli* NM1056 cells and incubated at 37°C for 15mins. The mixture was added to the soft agar, vortexed briefly, and poured onto (wet) LB-Agar plates. After the plate was incubated overnight at 37°C, 3mls of phage buffer was added to the confluent lawn, on the plate, and incubated at room temperature for 30mins. Subsequently the soft agar top and residual buffer was scrapped off using a sterile spatula and placed in a glass universal bottle with 1ml of chloroform. The mixture was vortexed for 30 seconds and left on the bench for

20mins and then centrifuged at 5000g for 10-15mins.The top layer of supernatant contained the amplified phage was transferred to a clean glass universal bottle. A phage titre was used to quantify the fresh stock, which was stored at 4°C, above chloroform (200µl to 1ml phage stock).

2.3.2 Phage antirestriction Assay

Overnight cultures were prepared for each of the mutants in *E. coli* strains NM1041 and NM1057, in addition, restriction proficient and restriction deficient cells were grown as (NM1041 and NM1057) negative and positive controls for restriction alleviation.

For each mutant assayed, 3ml of BBL top agar was equilibrated at 45-46°C and vortexed with 100µl of an overnight culture creating a bacterial lawn upon which infection of phage can be carried out. The plates were appropriately labelled and left to dry at 37°C for 20-30mins). A phage titre was performed, in which amplified $\lambda_{V.0}$ and $\lambda_{V.k}$ with a similar concentration were diluted with 10 fold serial dilutions from 10^{-2} to 10^{-7}, the dilutions were spotted in triplicate onto each bacterial lawn, the plates were left to dry, inverted and incubated overnight at 37°C. If the presence of the mutant alleviates restriction the number of plaques formed from the unmodified phage would not change on restriction proficient or restriction deficient cells.

2.3.3 Phage antimethylation assay

Single plaques from the phage restriction assay were picked into 200µl of phage buffer with 10µl of chloroform, vortexed and incubated overnight at 4°C. The phage stocks from unmodified phage which have infected the cells expressing each of the mutants were titred on plates with restriction proficient and restriction efficient cells as described above. If the number of plaques on each plate is the same, the phage has been modified and not susceptible to restriction by the R subunit.

2.4 Generation of ISOR Library

POcr is a synthetic gene that shares the same DNA sequence as the wtOcr gene from T7 phage except that each of the codons that translate the acidic residues D/E were substituted with neutral residues N/Q respectively. The POcr gene was synthesised by Geneart (Invitrogen). ISOR was used to introduce mutation on the POcr gene that targeted the neutral residues on the POcr protein and substituted them to acidic residues. The ISOR procedure consists of seven stages which are depicted in figure 2.2.

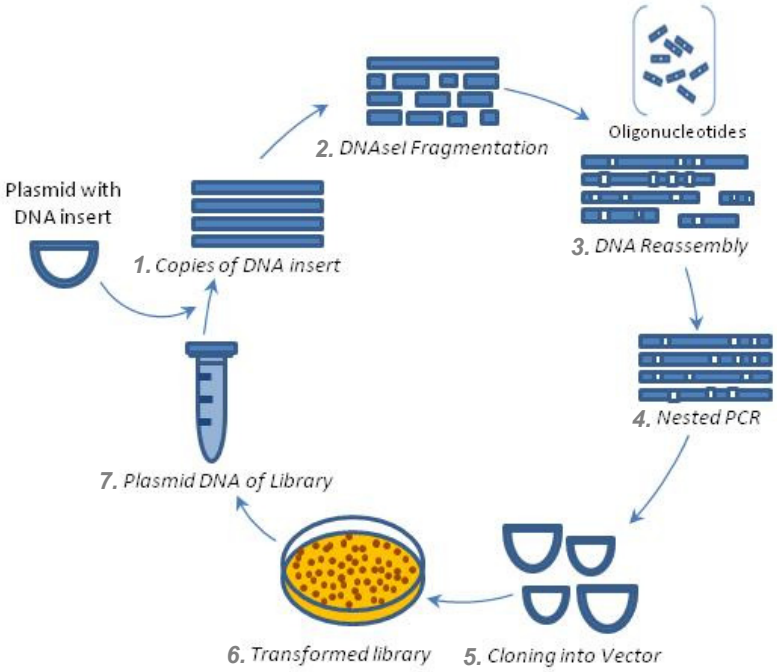


Figure 2.2: A flowchart describing the main stages involved in the development of the library. (1) Initially the POcr gene was amplified; (2) the insert was then fragmented using the nuclease, DNAseI. (3) The fragments underwent a recombination reaction with DNA primers which contained, mutations at specified positions and (4) this was followed by a Nested PCR reaction, which trimmed each of the newly combined inserts (5) allowing them to be cloned back into a plasmid vector. (6) Finally the library was transformed onto LB-Agar plates, some of the colonies were sent to sequencing to ascertain the diversity of the library. Once known, transformations were performed to adequately cover the estimated diversity by 120%. The library was stored as Plasmid DNA at -20˚C.

1. POcr Amplification

To obtain a large quantity of the POcr DNA sequence, the insert was amplified from the original pET24b vector by a PCR amplification reaction, primers were designed that ensured the incorporation of a generous flanking region at either end of the POcr gene (PCR parameters and mixtures are displayed below). Several reactions (×20) were carried out to achieve a high concentration of the POcr gene, these PCR reactions were subsequently pooled and purified using a Qiagen PCR purification kit. The concentration of the DNA insert was quantified using a Nanodrop UV Spectrophotometer.

Reaction Cycle	
Temp (°C)	**Time**
98	30secs
98	30secs
58	20secs
72	20secs
72	10mins
4	Store

30 Cycles

PCR Reaction Mix	
Mq Water	36µl
5(×)Phusion Buffer	10µl
Ocr DNA Template	0.5µl (at ~50ng/µl)
Primer Rev/Fwd	1ng/µl
10mM dNTP Mix	1µl
Phusion DNA polymerase	2.5U
Total	50µl

2. DNaseI Digestion

The amplified pOcr insert was fragmented using the DNaseI enzyme. 6µg of amplified insert was mixed with 0.05 Units of DNaseI enzyme in a 100µl reaction volume. The digest mixture was incubated for 8mins at 37°C. The reaction was quenched using EDTA (0.5M, 15µl) and mixture incubated at 90°C for 10mins. The resulting fragments from the digest were purified by DNA agarose gel. Fragments were expected to be between 0-100bp, therefore a 1.6% agarose gel was cast to capture the small fragments. The gel band of fragments was excised and the DNA extracted using the Qiagen Gel extraction kit.

3. Reassembly reaction

A second PCR reaction was used to reassemble the POcr gene fragments with specifically designed primers that were designed with specific mutations, the rationale and the design behind these primers are detailed in Chapter 4 and were obtained from Invitrogen. The PCR conditions and the PCR reaction mix are detailed in the tables below.

Reaction cycle	
Temp (°C)	**Time**
96	1.5mins
94	30secs
65	1.5mins
63	1.5mins
60	1.5mins
57	1.5mins
54	1.5mins
51	1.5mins
48	1.5mins
45	1.5mins
41	1.5mins
72	1.5mins
72	7mins
4	Store

35 Cycles

PCR Reaction Mix	
Mq Water	Make to 50µl
10(×)pfu Buffer	10µl
pOcr DNA fragments	100ng
Primers	200nM/per primer
dNTP's	0.4mM
Phusion DNA polymerase	2.5Units
Total	50µl

4/5/6/7. Nested PCR and subcloning of POcr gene into the pTrc99A plasmid vector.

The DNA from the reassembly reaction was used directly in the nested PCR reaction, primers were designed to amplify the reassembled POcr gene and insert as well as introducing the restriction sites, NcoI and HindIII at the start and stop codons of the gene. The PCR reaction followed the same condition as the POcr amplification PCR reaction detailed in step1. The resulting products from the nested PCR products were purified further with the Qiagen PCR purification kit. A small aliquot of the DNA product was run on a 1% Agarose gel (100V) to verify the successfully reassembly of the POcr gene.

The modified POcr gene was inserted into the plasmid pTrc99A using the same general procedures carried out for the Ocr multimutants gene that are detailed in section 2.2.1. The modified insert underwent a double digest using the restriction enzymes NcoI and HindIII and the ligation reaction inserted the POcr gene into pre-linearised pTrc99A vector.

Intially, XL-Gold Supercompetent cells were transformed with the POcr library, as per the manufacturer's guidelines (Stratagene). The complete transformation mix was plated out onto 20×LB-agar plates supplemented with carbenicillin and incubated overnight at 37°C. The resulting colonies from all of the plates were pooled using 3ml of LB-broth per plate, the resulting mixture was centrifuged (15000rpm, 10mins at 4°C) and a Qiagen Maxi prep kit was used to isolate the plasmid DNA library. The library was subsequently quantified using the Nanodrop spectrophotometer.

In the event of producing more than one library and increasing the diversity or increase the number of sites targeted for mutagenesis the whole library procedure was repeated however instead of amplifying the original POcr template an aliquot of the newly modified POcr gene was used as the template gene.

2.4.1 2AP Selective screening assay

The expression of the antirestriction protein wtOcr can rescue the growth of *E. coli* NM1041 cells that have been exposed to the DNA mutagen 2-aminopurine (2AP). The 2AP resistance of *E. coli* NM1041 cells acquired in the presence of the wtOcr protein is indirectly associated with the antirestriction activity of the Ocr protein. Hence if *E. coli* NM1041 cells acquire resistance to 2AP in the presence of a POcr mutant it suggests that the POcr mutant has antirestriction properties that are equal to wtOcr.

This was designed into a plate based selection method where the POcr libraries were transformed with *E. coli* NM1041 cells and plated onto plates exposed with high levels of 2AP (80µg/ml). By selecting for 2AP resistance the growth of *E.coli* NM1041 cells is dependent onthe expression of POcr multimutants that share the same antirestriction activity as wtOcr.

Transformations of *E. coli* NM1041 cells with the POcr library, LibDg were carried out as previously described and LB-agar plates were prepared with 2AP (IPTG, 50µg/ml and 2AP, 80µg/ml) in addition to the appropriate antibiotics. The plates were incubated at 37°C overnight and then left on the bench for a further two days; the number of colonies present each day was recorded.

The resulting colonies from the initial screen were restreaked onto fresh LB-agar plates supplemented with Km, Carb, IPTG and 2AP. After overnight incubation at 37°C, if cell growth was homogeneous, a swab of the colony was used to inoculate 5mls of LB broth to produce small scale bacterial cultures. The Qiagen mini prep kit was used to isolate the DNA from each overnight culture and *E. coli* NM1041 cells were transformed with the resulting POcr mutant gene and cells were selected for 2AP resistance as above. The POcr genes that were expressed in *E. coli* cells which were positive for growth were subsequently sequenced by Genepool.

2.5 Protein techniques

2.5.1 Protein purification

E. coli BL21 (DE3) pLysS cells were transformed with each Ocr mutant plasmid. Single colonies were picked to grow cells overnight which were subsequently used to inoculate 2L of LB Broth (2×1L in 2L conical flask) supplemented with 50µg/ml of carbenicillin. Cells were grown at 37°C with shaking (~200rpm), when the cell density OD_{600} measurement was ~0.6 gene expression was induced by the addition of IPTG (final concentration of 1mM) growth was continued for 2.5 hours before harvesting the cells by centrifugation (8000g for 10mins at 4°C). Cells pellets were weighed and stored at -20°C.

Ion exchange chromatography

Wild type Ocr and its mutants exhibit negatively charged surfaces. Therefore, the first step in purification involved an ion-exchange column. The cells were resuspended in ice-cold buffer A (20mM Tris-HCl, 300mM NH_4Cl, pH8) in the presence of a protease inhibitor cocktail (Roche, Basel, Switzerland). The cells were lysed on ice by sonication using a Soniprep 150 Sonicator (Sanyo, Tokyo, Japan) fitted with a 9mm probe (1min/g of cell paste). Cell debris was removed by high speed centrifugation (17,000g for 1hr at 4°C) and the supernatant was filtered using a 0.22µM filter.

The supernatant was loaded onto a 20 x 1.6cm diameter DEAE-Sepharose fast flow anion-exchange column (GE Healthcare, Piscataway, NJ), which had been pre-equilibrated in buffer A, at a flow rate of 48ml/hr. Upon loading the supernatant the column was extensively washed with buffer A and then a 500ml gradient from 0.3-1.0M NH_4Cl in buffer A was run at 24ml/h. Using an SDS-PAGE gel, fractions containing Ocr multimutants were identified and subsequently pooled. UV-spectrophotometry indicated that the samples were contaminated with nucleic acid. Therefore the second purification step was TCA precipitation.

Trichloroacetic Acid (TCA) Precipitation

The crude Ocr preparation was precipitated by the addition of 1.2 volumes of 10% (w/v) trichloroacetic acid (TCA) and incubated on ice for ~10min. The precipitate was collected by centrifugation at 15,000g for 10min at 4°C and the pellet was resuspended in 95% ethanol with gentle mixing for ~10min. After centrifugation (15,000g for 15 min at 4°C) the supernatant, which contained the Ocr protein, was transferred to a clean tube. This cycle of TCA precipitation followed by resuspension in 95% ethanol was then repeated a further two times except the final precipitate was resuspended in 20mM Tris HCl pH 8.0, instead of 95% ethanol.

Dialysis

If the sample remained as a precipitate, dialysis was performed against ~4L 20 mM Tris-HCl, 10mM NaCl pH 8.0 for 16hr at 4°C and concentrated by centrifugation using a Vivaspin concentrator (10,000 MWCO; VivaScience AG, Hannover, Germany). UV spectrophotometry was used to determine the concentration of the protein and to indicate the removal of nucleic acid contamination. Finally, an equal volume of glycerol was added to the sample, which was then stored at -20°C until required.

2.5.2 UV Spectrophotometry

All protein and DNA concentrations were calculated by measuring the UV-vis spectrum of the sample using a HITACHI U2900/2910 Double Beam Spectrometer. Samples were filtered through a 0.2µM filter (Sartorius) before checking for UV absorption. Using the buffer, the baseline was set to zero. The sample was added to an air-dried 1 cm pathlength quartz cuvette (Starna). The spectrum was recorded in the 210–340nm wavelength range, with a scan speed of 400nm/min and a slit width of 2nm. The molar extinction coefficient of the protein and the Beer-Lambert law were used to calculate the final protein concentration. However, to calculate the concentration of DNA the assumption was taken that for dsDNA an absorption value of 1 at 260 nm corresponded to 50 µg/ml.

2.5.3 CD analysis

Circular dichroism (CD) measurements were carried out on a Jasco Model J-180 spectropolarimeter (Jasco Corporation, Tokyo, Japan). Protein samples were prepared in 10mM NaH_2PO_4 pH 8.0, 10mM NaF. Far-UV CD spectra (190–260 nm) were performed at a protein concentration of 30µM using a 1mm path length cell. A bandwidth of 1nm was used with a scan speed of 10nm/min. All CD measurements were made at 25°C and each spectrum was an accumulation of four individual scans. The spectra were corrected for buffer contribution by subtracting the CD absorption signal of the buffer from each of the data sets.

2.5.4 Unfolding studies

Equilibrium unfolding as a function of guanidinium hydrochloride (GdmCl) concentration was monitored by fluorescence spectroscopy (excitation 280nm; emission scan from 330-450nm; bandwidth 1nm). A stock solution of GdmCl was freshly prepared from ultrapure reagents and the precise concentration determined using a refractometer. The protein samples in 20mM Tris-HCl pH 8.0 were incubated overnight at 4°C in the presence of various concentrations of GdmCl. Fluorescent data was then measured and the data was fitted to a two-state unfolding model assuming a linear relationship between ΔG of unfolding and the concentration of GdmCl.

2.5.5 Cross linking

Cross linking studies were carried out using the cross linking agent glutaraldehyde. Concentrated protein solutions were diluted to 25µg/250µl using a 10mM HEPES, pH7.0 buffer. 1% of a 25%w/v glutaraldehyde solution was added and incubated for 2mins, at room temperature. The reaction was quenched using freshly prepared 2M $NaBH_4$ (in 0.1M NaOH) and incubated for a further 20mins at room temperature. 10%w/v sodium deoxycholate was added. TCA precipitation was carried out using 78%w/v trifluoroacetic acid and the resulting precipitate was spun down (13k for 10mins). The pellet was resuspended in 12µl of SDS loading buffer and small amounts of Tris-HCl were added to neutralise the buffer. The samples were heated for 5mins at 90°C and the pellets spun down. The supernatant was loaded onto an SDS-PAGE gel; the samples ran at 100volts for ~1hr. The gel was subsequently stained with Coomassie blue. After destaining, the gel was analysed on a light box.

2.5.6 ITC

Isothermal titration calorimetry (ITC) was carried out using an Auto-iTC200 (Microcal, Northampton, MA). The stocks of Ocr and M.EcoKI were buffer exchanged into 20 mM HEPES, pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol using a PD-10 gel filtration column (GE Healthcare). The concentration of the protein solution was then adjusted appropriately by dilution into the same HEPES buffer. To avoid degradation, SAM was added to the final protein solutions (100 µM). The ratio of Ocr:MEcoKI was 10:1 in each case with 30µM:3µM concentrations. All titrations were carried out at 25°C. The heat of dilution was obtained by injecting Ocr (0.2µl×19) into buffer. The calorimetric data were converted into differential binding curves by integration of the resultant peaks. The data were fitted into a single-site binding model. The automated protocol degassed all solutions before each run and carried out an automated wash cycle between each sample investigated.

2.5.7 *In vitro* nuclease assay

The *in vitro* assay indirectly measured antirestriction activity by the cleavage of unmethylated pBRsk1 using purified EcoKI in the absence or presence of the Ocr protein. Reactions were performed in a 50 µl in 10 mM Tris-acetate pH 7.5, 10 mM magnesium acetate, 7 mM 2-mercaptoethanol, 50 µg/ml bovine serum albumin, 2 mM ATP, 0.1 mM SAM, 2 nM pBRsk1 and 10 nM EcoKI at 37°C. Ocr (~100 nM for example) was briefly preincubated with EcoKI in the reaction mix prior and reactions were initiated by the addition of DNA substrate. The reactions were incubated for 10 min before quenching by incubation at 68°C for 10 min. Samples were mixed with gel loading buffer and loaded onto a 1% agarose gel containing 0.1 µg/ml ethidium bromide and subjected to electrophoresis in

TAE buffer (40 mM Tris acetate, 2 mM EDTA) at 100 V/h. DNA was then visualised under UV illumination.

Chapter 3. Ocr multimutants

3.1 Introduction

The protein Ocr mimics the charge and shape of bent B-form DNA. Charge mimicry is produced by a negatively charged protein surface which results in an isoelectric point, pI, of 3.9, for the Ocr monomer. To understand the importance and necessity of the large number of acidic residues on the surface of Ocr, which are responsible for the extremely low pI, an investigation on the importance of charge for antirestriction activity was carried out by producing a plasmid library. The aim was to produce a library of Ocr multimutants possessing a range of isoelectric points and assess their activity with a sensitive *in vivo* assay. However, the value of such libraries is a reflection of the sensitivity of the screening assay.

This chapter describes the development of an *in vivo* screening method which has the ability to select for mutants with antirestriction properties similar to wtOcr. Previous studies of the *in vivo* anti-restriction activity of DNA mimics and their mutants have relied on phage lambda infection experiments; although highly sensitive this technique is laborious and costly in consumables. In addition subtle differences in antirestriction activity cannot be easily identified by phage lambda experiments; such differences have only been elucidated from the *in vitro* characterisation of the purified proteins. The development of an *in vivo* selection assay that is sensitive enough to differentiate between the anti-restriction activity of DNA mimics can eliminate the time spent characterising purified proteins, which are biologically irrelevant to the purposes of the researcher.[120,61]

3.2 Specific aims

In this chapter an *in vivo assay* was developed to exploit the natural restriction alleviation system residing in *E. coli* NM1041 cells, in response to the DNA mutagen, 2-aminopurine (2AP). By selecting for 2AP resistance, the assay can assess the anti-restriction behavior of DNA mimic proteins like wtOcr and Orf18 ArdA. A series of Ocr multimutants that targeted clusters of acidic residues on the surface of Ocr were also investigated. These Ocr multimutants have been extensively investigated with *in vitro* experiments, therefore the characterisation of these mutants can act as a validation of the 2AP assay and allow for comparisons between the *in vivo* and *in vitro* results.

3.3 Proof of principle

The antirestriction proteins, wtOcr from T7 bacteriophage and Orf18, ArdA from the conjugative transposon Tn916, mimic DNA and function as competitive inhibitors of the EcoKI enzyme. The introduction of these proteins into a *ΔClpXP*$^{-}$ bacterial strain, treated with 2AP, alleviates restriction allowing the continued growth of the cells as described in chapter 1.

E. coli cells of the bacterial *E. coli* strain, NM1041 Δ*ClpXP*$^{-}$ (a kind gift from Dr. Angela Dawson, School of Physics and Astronomy, Edinburgh) were transformed with wtOcr and wtOrf18 ArdA along with the empty plasmid vector pTrc99A (each of the antirestriction genes, wtOcr and wtOrf18 ArdA were inserted into the plasmid pTrc99A to maintain consistency between results). *E. coli* NM1041 cells expressing either wtOcr, wtOrf18 ArdA or pTrc99A, were grown on two types of LB-agar plates; a plate supplemented with antibiotics and a plate containing the antibiotics, IPTG, as an inducing agent and 80µg/ml of 2AP, (80µg/ml was the concentration needed to induce restriction alleviation from unpublished work carried out with Dr. Angela Dawson).

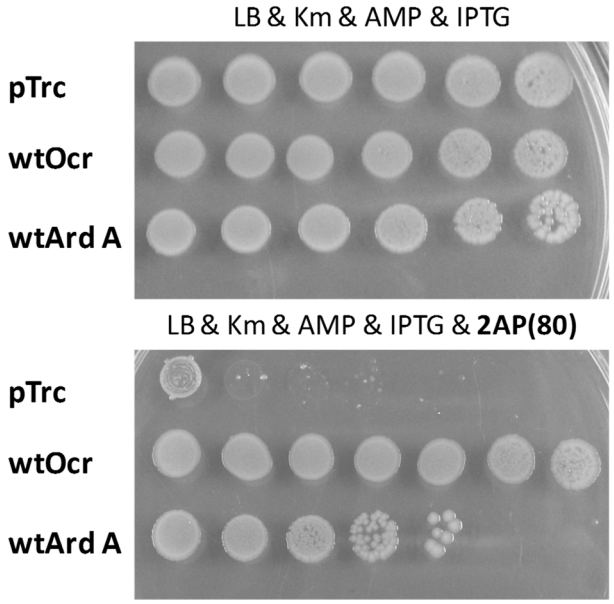


Figure 3.1: Viable counts of 10-fold serial dilutions from left to right of *E. coli* NM1041 (ClpXP^{-}) cells were plated onto two plates. Above the plate is supplemented with antibiotics and the inducing agent IPTG whereas the plate below has an additional supplement of 2AP, at 80µgml-1 per plate. Cells grown in the presence of the either antirestriction protein; wtOcr or ArdA, are resistant to 2AP.

The resulting plates in figure 3.1 suggest that the presence of either antirestriction protein rescued the viable count of *E. coli* NM1041 cells from 2AP treatment, whereas *E. coli* NM1041 cells transformed with the empty vector pTrc99A did not survive 2AP treatment. The presence of the ArdA protein in *E. coli* NM1041 cells, treated with 2AP, maintains ~70% of the cells. This is increased to a 100% survival rate for *E. coli* cells transformed with wtOcr. This implies that the protein ArdA does not completely negate the effects of 2AP treatment. In fact, *in vitro* data published on both antirestriction proteins does show that, the wtOcr protein has a tighter binding affinity to M.EcoKI compared to the ArdA protein.[121]

3.4 Ocr multimutants

The ability to differentiate between the antirestriction activity of wtOcr and wtArdA by the *in vivo* RA assay was unexpected. It suggests that the screen could be sensitive enough to elucidate changes in activity induced by mutants of the proteins. To investigate this, the assay was adapted to perform a titration of 2AP to enhance the sensitivity of the assay to distinguish between the antirestriction activity of a number of Ocr multimutants.

Overnight *E. coli* NM1041 cell cultures, expressing either wtOcr or a mutant of Ocr, were plated onto LB-agar plates supplemented with increasing concentrations of 2AP. The reduction in 2AP concentration could lead to a decrease in the number of unmodified sites, resulting in fewer sites of targeted restriction. Although this should have no effect on the viable count for wtOcr, the decrease in unmodified sites might cause weak binding mutants of Ocr to display resistance at decreased levels of 2AP.

To assess this hypothesis and determine the capability of the RA assay, a number of multimutants of the T7 wtOcr protein were accepted as a kind gift from Dr. Augoustinos Stephanou (Dryden Lab, Edinburgh). The mutants from A. Stephanou neutralised clusters of negative charge by substituting residues of aspartic acid and glutamic acid with asparagine and glutamine, according to their position in the crystal structure of wtOcr (figure 3.2). The mutants were created to investigate the effect of negative surface charge on antirestriction activity and elucidate regions of importance for the interaction of the protein with Type I restriction enzymes.

The genes for each of the mutants were ligated into the plasmid vector pTrc99A to allow a direct comparison to be made between the Ocr mutants.

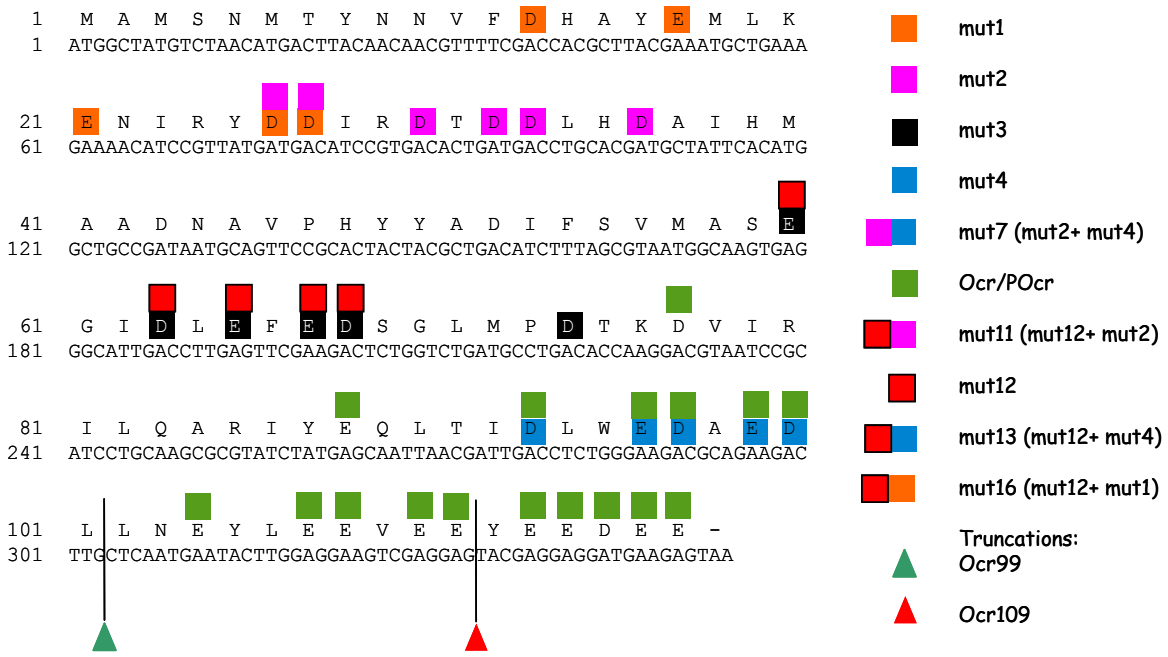


Figure 3.2: Color coding of the amino acids targeted for mutagenesis for each Ocr multimutant. The protein and corresponding DNA sequence of wild-type Ocr is shown. The panel on the right shows the color coding used to highlight the amino acids targeted for mutagenesis. Extracted from A. Stephanou. PhD Thesis, 2010.[122]

3.5 Restriction alleviation titration assay

The restriction alleviation (RA) titration assay screened *E. coli* NM1041 cells expressing each of the Ocr multimutants exposed to 2AP concentrations ranging from 0-80μg/ml, figure 3.3. Viable cells were recorded from several plates; a control plate (containing antibiotics only) and plates supplemented with IPTG and 0, 5, 10, 20, 40 and 80μg/ml of 2AP.

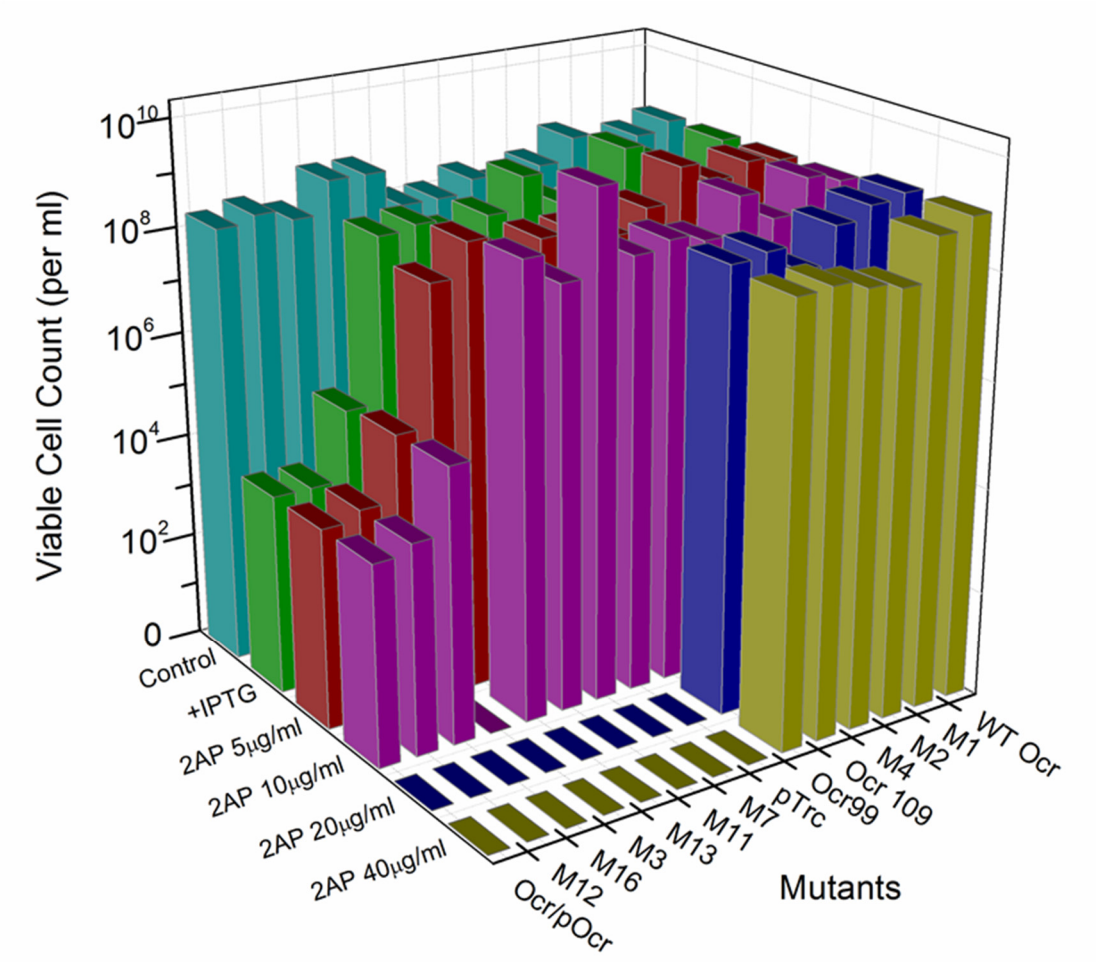


Figure 3.3: The RA assay measuring cell survival of *E. coli* NM1041 (ΔClpXP$^-$, r$^+$m$^+$) when transformed with various plasmids expressing Ocr or its mutant forms. The graph shows growth on LB-agar plates supplemented with antibiotics (cyan) only, antibiotics and IPTG (green) and, in the presence of antibiotics, IPTG plus increasing concentrations of 2AP (5, 10, 20 and 40 µg/ml are red, magenta, blue and olive, respectively).

Cell survival was calculated for *E. coli* NM1041 cells transformed with each of the mutants, upon increasing levels of 2AP and compared to the survival rate of cells grown with pTrc99A, wtOcr and the control plates, which were used to determine the concentration of each of the mutants concerned. Viable cells counts in the presence of wtOcr survived to a 2AP concentration of 80µg/ml, this level of 2AP resistance represents a fully active antirestriction protein. In contrast, cells grown with the empty vector pTrc99A did not survive concentrations of 2AP that exceeded 20µg/ml. The 2AP resistance demonstrated by cells grown with pTrc99A is equivalent to an inactive antirestriction protein; therefore mutants that have a resistance to 2AP that is greater than 20µg/ml are exhibiting some level of antirestriction activity and if 2AP resistance is ≥80µg/ml then the mutant protein is demonstrating full antirestriction behavior.

As such, the RA titration assay in figure 3.3 indicates that the mutants; M1, M2, M4, Ocr99 and Ocr109 retained the same 2AP resistance as wtOcr, whereas the mutants M7, M11 and M13 shared the same resistance as pTrc99A. This suggests that these mutants retain full activity or no activity respectively.

The titration assay revealed a third phenotype induced by the Ocr multimutants M3, M12, M16 and Ocr/pOcr. The mutant M3 appears to maintain 2AP resistance to a concentration of 10µg/ml. This level of resistance is lower than the pTrc99A vector and indicates that the mutant is actively reducing cell growth. Moreover *E. coli* NM1041 cells grown in the presence of mutants M12, M16 and Ocr/pOcr had a reduced viable cell count of 10^4 cells per ml when grown on plates with the inducing agent IPTG. This implies that the expression of these mutants also effects cell growth without the addition of 2AP. Incidentally upon exposure to 2AP, these mutants exhibit the same 2AP resistance as pTrc99A, which suggests that they possess limited, if any, antirestriction activity.

3.6 Colony appearance

Upon exposure to increasing concentrations of 2AP the viable cell count does not result in a gradual reduction in *E. coli* titre, instead the quantitative data from the assay is limited to observing either complete cell survival or total cell death. Further analysis of the assay shows that 2AP concentration causes a reduction in the cell density of the colonies produced, colonies become fainter with increasing levels of 2AP. This reduction in the density of the colonies is demonstrated in figure 3.4, where initially the density of colonies is uniform between *E. coli* cells expressing either pTrc99A or wtOcr. However, at a 2AP concentration of 10µg/ml the contrast and colour between the colonies is noticeably fainter for cells grown with pTrc99A, upon 20µg/ml the cells are misty in appearance until eventually at 40µg/ml there is no growth. This reduction in the size of colonies indicates that the cell growth has been compromised by the expression of the particular Ocr multimutant.

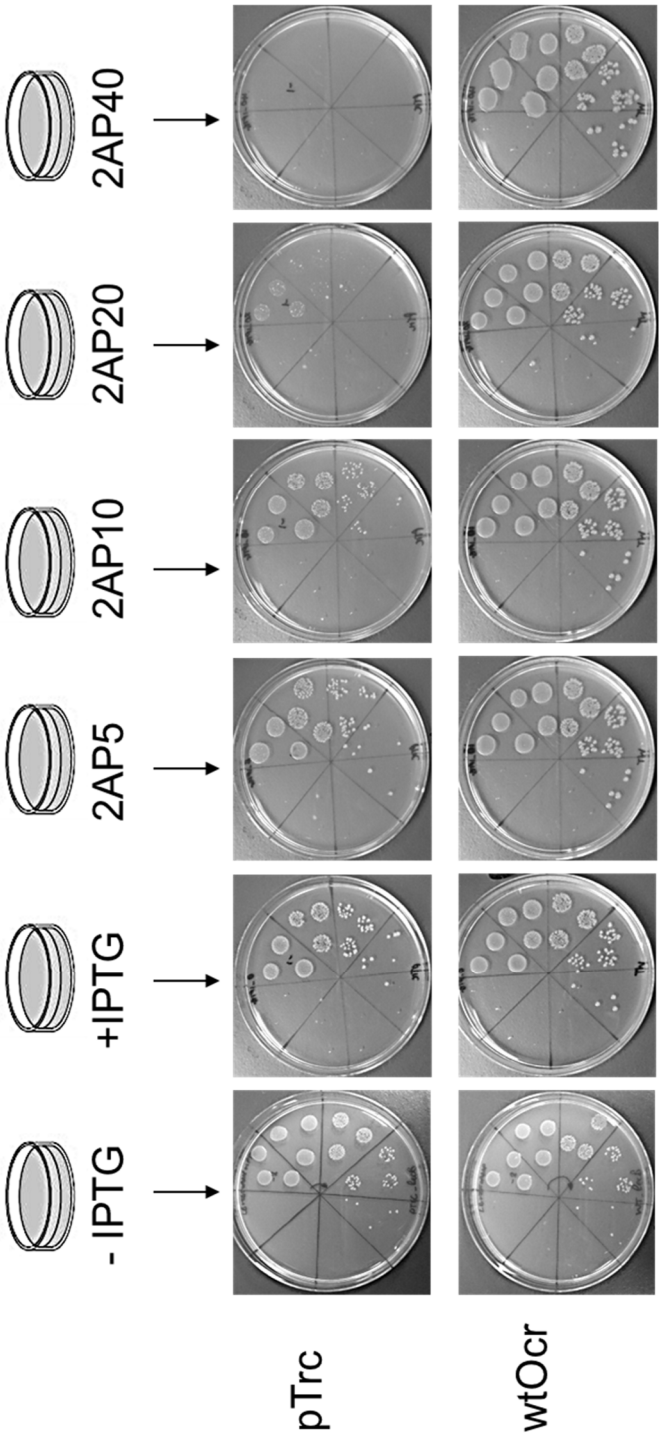


Figure 3.4: A representation of the 2AP plate assay. *E. coli* NM1041 cells transformed with either pTrc99A or wtOcr was grown overnight. Serial 10 fold dilutions were plated in a clockwise rotation on LB agar plates supplemented with increasing concentrations of 2AP and with and without induction (IPTG) of pTrc or wtOcr.

Changes in colony morphology were also observed for *E. coli* NM1041 cells in the presence of the Ocr mutant, M3. Figure 3.5 vividly shows that the expression (+IPTG) of the Ocr mutant M3 causes a change in the size of the colony. The total viable cell count of *E. coli* NM1041 cells transformed with mutant M3, +/-IPTG, demonstrates that cell count is unaffected but the size of the colonies in the presence of IPTG and M3 are visibly smaller.

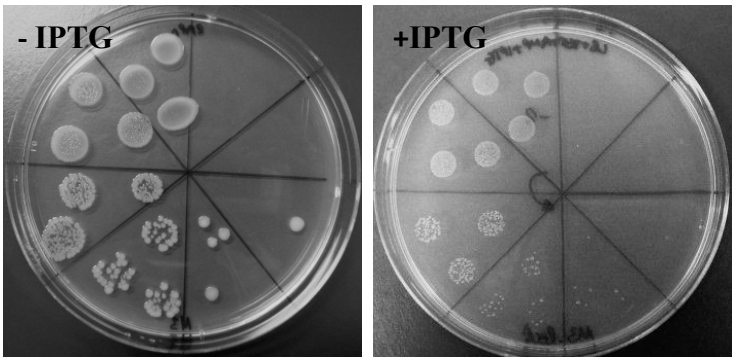


Figure 3.5: Serial 10 fold dilutions of an overnight cell culture of *E. coli* NM1041 cells transformed with the Ocr multimutant M3, were plated out in an anticlockwise direction in triplicate. The left plate is supplemented with antibiotics (Km & Carb) only whereas the right plate is supplemented with antibiotics and the inducing agent, IPTG.

The assessment of colony morphology supplements the quantitative data. For example, the quantitative data suggests that M3 induces a 2AP resistance of 10μg/ml of 2AP in *E. coli* NM1041 cells and the morphology shows that the cells are sensitive to the expression of the mutant M3. In some ways the phenotype of M3 is similar to the phenotype induced by the mutants M12, M16 and Ocr/pOcr. Both phenotypes show that expression of the Ocr mutant is affecting the growth of *E. coli* NM1041 cells, however the extent of this effect is greater for the Ocr mutants; M12, Ocr/pOcr and M16.

3.7 Bacterial growth curves

The expression of the Ocr multimutants M3, M12, Ocr/pOcr and M16 seem to be deleterious to the cells compromising cell growth or causing cell death. To investigate the apparent toxicity caused by these Ocr mutants, bacterial growth curves were monitored for each of the mutants. The use of LB-agar plates provides a result from a fixed time point (18 hours of growth) whereas bacterial growth curves observe the growth of *E. coli* cells in real time.

Overnight cell cultures of *E. coli* NM1041 cells transformed with each Ocr multimutant, pTrc99A and wtOcr were used to inoculate fresh LB-medium, with and without the inducing

agent IPTG. Subsequently the *E. coli* cells were incubated at 37°C and 200rpm, cell growth was monitored by measuring the optical density at a wavelength of 600nm, with readings taken every 5 minutes for 500mins.

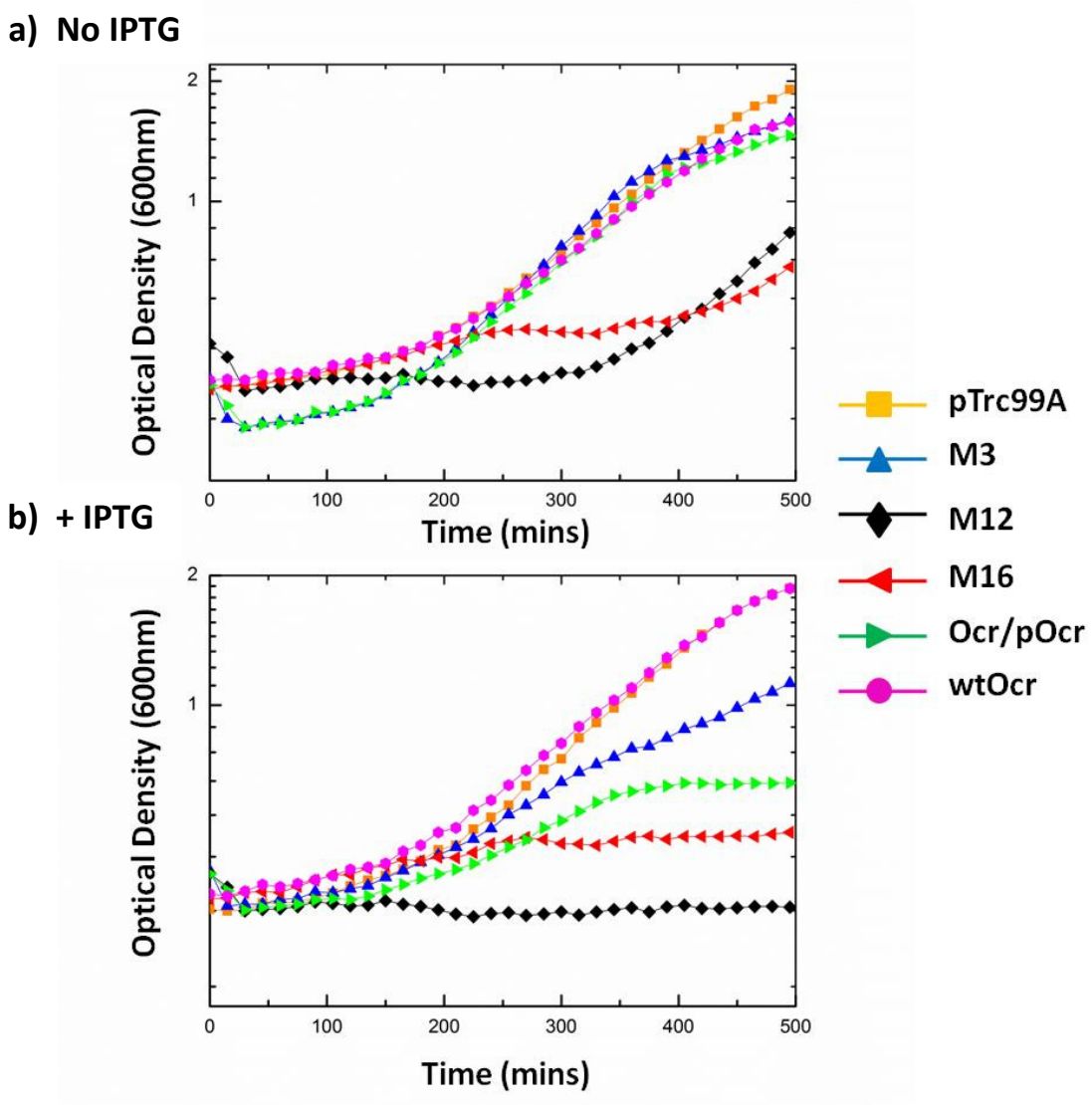


Figure 3.6: Cell growth with the Ocr multimutants expressed in *E. coli*, NM1041 (*ClpXP*-, r^+m^+) cells in the absence of 2AP. It is clear that cells expressing different Ocr multimutants grow at different rates. pTrc99A is the vector alone, all other data are for cells expressing Ocr wild-type (wt) or Ocr mutants. OD600 measurements were recorded every 15min for 500min a) cells growth without the inducing agent IPTG, b) cells grown with the inducing agent IPTG. Errors are ±7% for triplicate measurements (error bars not shown for clarity). To maintain clarity those mutants which shared the same growth pattern as wtOcr and pTrc99A (M1, M2, M4, M7, M13) can be found in the appendix (Figure A.1).

The bacterial growth curves show that the expression of wtOcr or pTrc99A has no effect on cell growth which complements the 2AP RA assay above. Indeed cell growth reflects an

archetypal growth curve, producing a doubling time of ~130mins. Typically, bacterial growth curves occur in four stages; the lag phase is the initial slow growing phase where cells adapt to the their new environment, the log phase is where cells grow exponentially in nutrient rich environments, the stationary phase where growth is reduced almost completely due to depleted nutrients, and the death phase where nutrients are scarce. The expression (addition of IPTG) of the mutants M1, M2, M4, M7, M11 and M13 produced the same growth profile as wtOcr and pTrc99A, a lag phase of ~150mins followed by a doubling rate of ~130mins (Appendix, figure A1).

However, comparisons between *E. coli* NM1041 cells with and without expression (+/- IPTG) of the Ocr multimutants M3, M12, M16 and Ocr/pOcr displayed different growth profiles. *E. coli* NM1041 cells expressing the Ocr multimutants M3 or Ocr/pOcr showed an increase in the doubling time to ~150mins and ~200mins causing a reduction in the growth rate. Furthermore, the OD_{600} absorbance was ~0.7 at stationary phase; this is much lower than the OD_{600} of *E. coli* cells expressing wtOcr which reached an absorbance value of 2, see figure 3.6a for cell growth patterns without IPTG and figure 3.6b for changes in cell growth upon expression (+IPTG) of the Ocr multimutants. The results are consistent with the appearance of small 'sick' colonies observed with M3, in Figure3.5, and the reduced cell count observed for cells in the presence of the mutant Ocr/pOcr in Figure 3.3.

Cells grown with mutants M12 and M16 showed slow growth rates in the presence and absence of IPTG. For cells grown with either mutant, in the absence of IPTG a lag time of ~300 mins and a doubling time of ~150 mins were observed. In the presence of IPTG, growth was inhibited, the expression of Ocr mutant M12 in *E. coli* NM1041 cells inhibited growth completely and the expression of M16 reduces growth rate and induces an early stationary phase after ~200 mins. Like the mutant Ocr/pOcr, the mutants M12 and M16 reached at an absorbance OD_{600} no greater than ~0.6 at 500 mins, figure 3.6b.

The bacterial growth curves were able to differentiate between the deleterious effects induced by the mutants M3, M12, M16 and Ocr/pOcr. The Ocr multimutants M3 and Ocr/pOcr are not as deleterious as the mutants M12 and M16. Such a distinction could not be ascertained from the results of the RA plate assay alone.

3.8 Cell toxicity and the restriction pathway

The reduction in cell viability caused by the mutants M12, M16 and Ocr/pOcr in *E. coli* NM1041 cells indicates toxicity, an intolerance of the proteins by the cell. To pinpoint whether this toxicity is specific to the restriction pathway, plate assays were performed which measured cell viability upon expression of the mutants in restriction proficient and restriction deficient *E. coli* cells. *E. coli* NM1041 and NM1057 cells respectively were grown in the presence of the Ocr multimutants M3, Ocr/pOcr, M16 and M12. Viable counts were determined using LB-agar plates supplemented with IPTG. For comparisons, the plate assay was carried out for *E. coli* cells expressing wtOcr and pTrc99A both of which do not induce deleterious effects in *E. coli* cells (+/-HsdR).

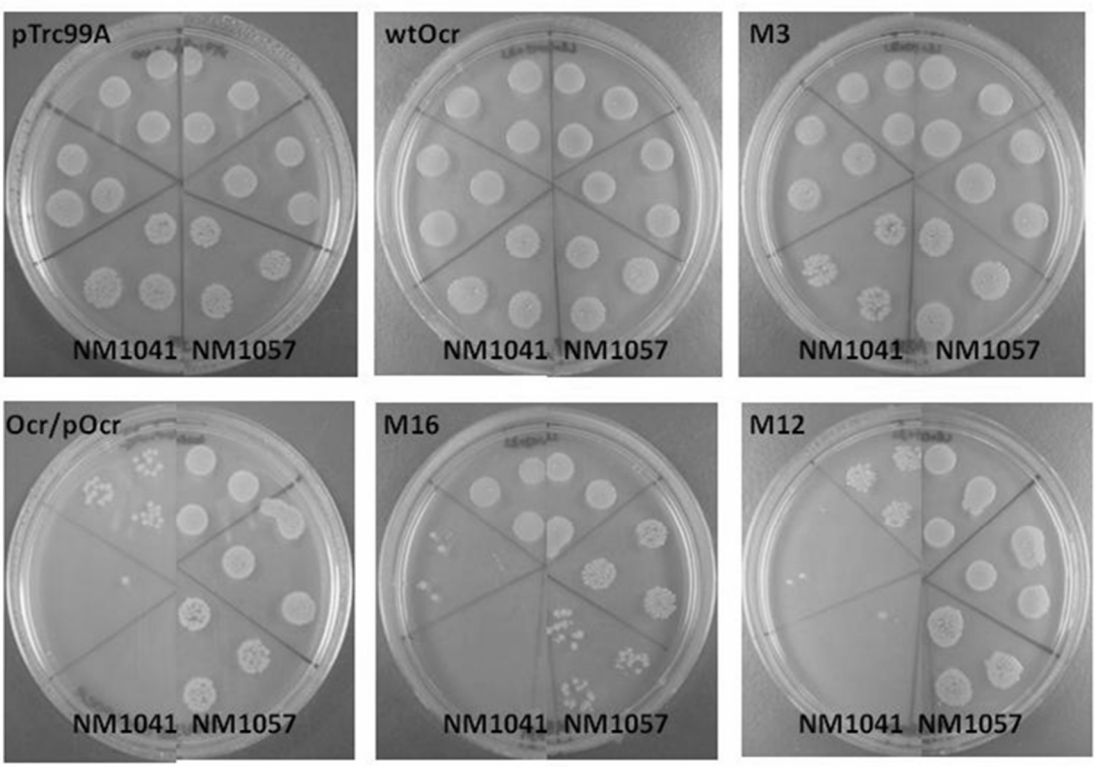


Figure 3.7: Loss of cell growth of *E. coli* NM1041 (*ClpXP*-, r^+m^+) and NM1057 (*ClpXP*-, r^-m+) expressing the mutants mut3 (m3), OCR/POCR, mut16 (m16) and mut12 (m12) compared to cells expressing wild-type Ocr (wtOcr) or transformed with vector alone (pTrc99A). Serial dilutions of cultures were spotted onto LB-Agar plates supplemented with appropriate antibiotics. The dilutions of 10^{-2}, 10^{-3} and 10^{-4} fold of the liquid culture are spotted in triplicate clockwise for *E. coli* NM1041 and anticlockwise for *E. coli* NM1057 from the top of the plate.

Serial dilutions show that viable cells grew to a dilution of 10^{-4} for mutants grown in restriction deficient *E. coli* NM1057 cells. This is identical to the viable counts observed for wtOcr and pTrc99A. In comparison, the viable count of restriction proficient *E. coli*

NM1041 cells was much reduced for the mutants M12, M16, Ocr/pOcr and to a lesser extent M3, as shown previously. This indicates that the toxicity induced by the expression of these mutants is a result of the restriction pathway.

3.9 Evidence of cell filamentation

Figure 3.7 shows, that the interaction between EcoKI and the Ocr multimutants M3, M12, M16 and Ocr/pOcr is deleterious to *E. coli* cells. As toxicity only affects *E. coli* in the presence of EcoKI it suggests that the toxicity is a result of restriction activity on the host chromosome. This suggests that the expression of the Ocr mutant induces a phenotype that inhibits the methylation but not the restriction process of EcoKI. If so, expression of the Ocr mutant could induce the SOS response in *E. coli* cells preventing cell division and producing filamentous cells.

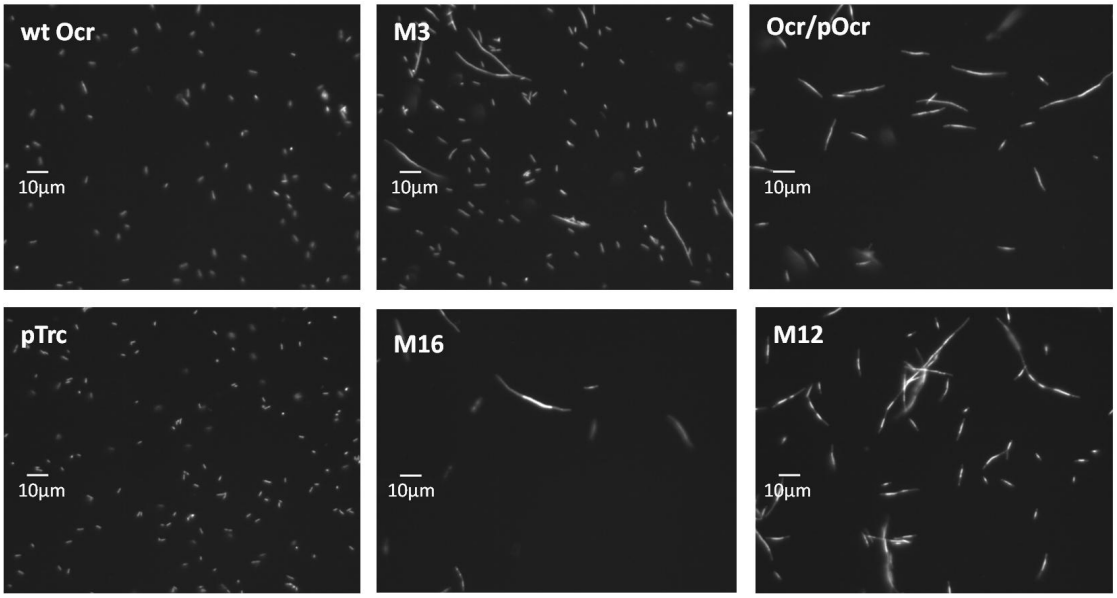


Figure 3.8: Fluorescence microscopy images of cell morphology of bacterial cultures of *E. coli* NM1041 ($clpX^-$, r^+m^+) cells transformed with plasmids expressing wtOcr or its variants. Control cells were transformed harboring the empty plasmid vector, pTrc99A. Cell nucleoids were revealed with DAPI staining. Scale bar is 10 microns.

This phenomenon was observed by inducing *E. coli* NM1041 cells, at an OD_{600} of 0.3, transformed with the expression constructs pTrc99A, wtOcr or an Ocr mutant. After a further 2hr incubation the cells were fixed with methanol, stored in Tris buffer, stained with DAPI and examined by fluorescent microscopy.

Micrographs displaying *E. coli* NM1041 cells expressing the mutants M1, M2, M4, M7, M11, M13, Ocr99 and Ocr109 contained intact nucleoids upon DAPI staining with the cell morphology identical to *E. coli* (NM1041) cells expressing either wtOcr and pTrc99A, as shown in figure 3.8. However, those mutants which appear to induce deleterious effects i.e. M3, M12, M16 and Ocr/pOcr, displayed extensive filamentation which is indicative of the SOS response in *E. coli* which has prevented cell division. The extent of filamentation seems more severe in the presence of the Ocr multimutants M12, M16 and Ocr/pOcr, compared to M3. It suggests that cells in the presence of these mutants appears to inhibit the methylation activity but not the restriction activity of EcoKI, resulting in unmodified sites on the chromosome susceptible to restriction.

3.10 *In vivo* versus *in vitro*

By adapting the 2AP screen from Makovets (2003) and Serfiotis-Mitsa (2010)[123,121] to include a titration of increasing 2AP concentrations the screen can assess the antirestriction activity of several Ocr multimutants *in vivo*. The assay has elucidated several phenotypes suggesting that different antirestriction behaviors have been induced. As such active, inactive and partially active Ocr multimutants have been identified. In addition, analysis has revealed that the level of antirestriction activity produced by the partially active mutants affects the growth of *E. coli* cells in the presence of the restriction pathway. To be sure of this interpretation the results in this chapter were compared with an *in vitro* study of the same Ocr multimutants carried out by Dr. A. Stephanou.

	Antirestriction Phage Assay	Inhibition of nuclease *in vitro*	Binding to M.EcoKI *in vitro*
WtOcr	Y	Uncut	Too tight
M1	Y	Uncut	Too tight
M2	Y	Uncut	Too tight
M4	Y	Uncut	Too tight
Ocr99	Y	Uncut	Too tight
Ocr109	Y	Uncut	Too tight
M7	Y	Partial Cut	Too tight
Ocr/pOcr	Y	Cut	Too tight
M12	Y	Cut	20+/-9 nM
M16	Y	Cut	71+/-30 nM
M3	Y	Cut	79+/-21 nM
M13	N	Cut	41+/-16 nM
M11	N	Cut	No binding

Table 3.1: A table extracting results of *in vitro* characterisation of the Ocr multimutants from the thesis of Dr. A Stephanou.[122]

The results in Table 3.1 are an extract of previous data kindly donated by Dr. A Stephanou. It shows extensive investigations of the interaction between each of the Ocr multimutants with the EcoKI enzyme. Antirestriction activity was investigated *in vivo* by comparing the phage titre of unmodified $\lambda_{V.0.}$ phage, on restriction proficient and restriction deficient cells, in the presence of each mutant. This showed that all of the mutants appeared to be actively inhibiting the nuclease except M11 and M13, these both appear to possess no antirestriction activity from the phage assay, however the phage assay relies on leaky gene expression and production of the protein or Ocr mutant.

In vitro data provided more direct measurements of the interaction between the mutants and the EcoKI enzyme. The interaction between the mutants and the endonuclease were investigated by carrying out an *in vitro* nuclease assay. This assay incubates the unmodified DNA plasmid, pBrsKI, which possesses a single recognition site for EcoKI, with the wtOcr protein or an Ocr mutant protein. The reaction mix is introduced to the EcoKI nuclease in a 10:1 ratio of mutant: nuclease. A decrease in the restriction activity of the nuclease signifies an increase in the antirestriction activity exhibited by the Ocr mutant protein. The results show that the mutants Ocr/pOcr, M12, M16, M3, M13 and M11 failed to inhibit the restriction activity of the nuclease, with mutant M7 producing partial protection.

Finally antimodification activity was assessed through ITC experiments investigating the binding interaction between M.EcoKI and each of the mutants. The results showed that no binding is observed for the mutant M11, whereas the mutants M12, M16, M3 and M13 bind weakly to M.EcoKI, with binding affinities ranging from 20nM-79nM. The remaining mutants seem to bind tightly to M.EcoKI.

The results from the separate investigations were compared. Comparisons between the RA assay and the ITC data showed that the results from both investigations shared a great number of similarities. These correlations provide an idea of the relationship between the phenotype observed from the RA assay and the interaction of the mutant with the MTase and/or REase, figure 3.9.

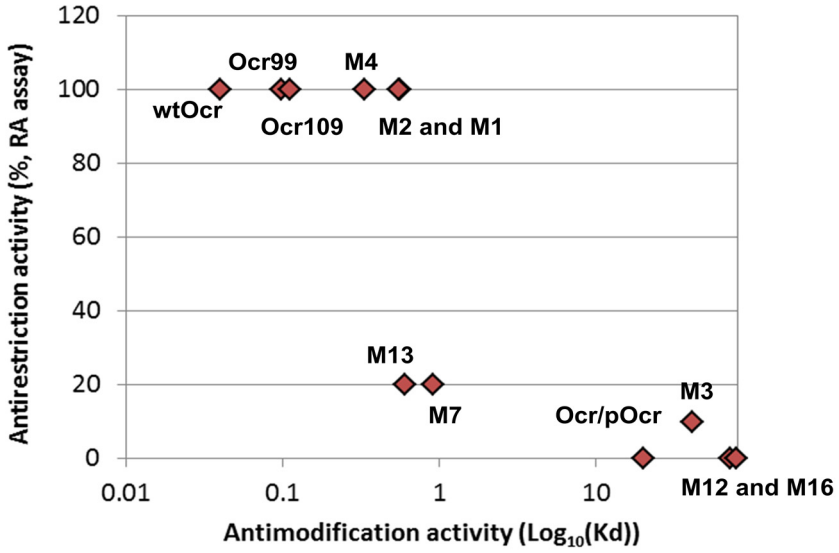


Figure 3.9: A scatter graph which looks at the relationship between the K_d values obtained for the interaction of the Ocr multimutants with the M.EcoKI complex *in vitro*, as a measure of antimodification activity compared to the restriction alleviation that each mutant exhibited as a value of antirestriction activity *in vivo*. The binding interaction between M.EcoKI with MTase complex was measured by ITC for weak binding mutants; M3, M12, M13 and M16 but for tight interactions between the mutants; M1, M2, M4, M7, Ocr/pOcr, Ocr99 and Ocr109 with M.EcoKI fluorescent anisotropy was performed. A clear relationship shows that an increase in the K_d value leads to a reduction in 2AP resistance. Ocr mutant M11 has not been shown because a binding interaction between the mutant and the M.EcoKI complex could not be measured, M11 appears to be inactive and appears unable to bind to the M.EcoKI enzyme.

The *in vitro* investigations measured the interaction of the Ocr multimutants with either the M.EcoKI or EcoKI active complexes separately whereas the *in vivo* studies measured the activity of the mutants within the whole cellular environment which contains a mixture of

the two active complexes of EcoKI. Nevertheless correlations were observed between the data sets *in vivo* and *in vitro* comparing the level of 2AP resistance induced by an Ocr mutant with the strength of its interaction with the M.EcoKI/EcoKI active complexes. The mutants M1, M2, M4, Ocr99 and Ocr109 induced a high level of 2AP resistance and possessed tight binding affinities towards the M.EcoKI complex similar to the wtOcr protein. However the mutant M11 exhibited no resistance to 2AP and no binding interaction with M.EcoKI, both studies indicate that mutant M11 does not exhibit any antirestriction or antimethylation activity. Furthermore it was revealed that the Ocr mutants; M12, M16, M3 and Ocr/pOcr that induced a deleterious effect in *E. coli* NM1041 cells exhibited intermediate binding affinity towards M.EcoKI. This confirms that the intermediate binding affinity of the protein to M.EcoKI/EcoKI can induce the phenotype $m^{+/-}r^{+/-}$ which effects the growth of cells in the presence of a Type I RM enzyme.

However there were some inconsistencies between the *in vivo* and *in vitro* data. For example, the mutant M7, presumed inactive from the 2AP assay, exhibited poor antirestriction and antimethylation activity *in vitro*. In addition the mutant M13, which appeared inactive from the 2AP assay, phage assay and *in vitro* nuclease inhibition, appeared to successfully bind to the M.EcoKI enzyme by ITC, albeit with reduced affinity.

In conclusion the 2AP RA assay is able to differentiate between mutants which possess full, partial and zero antirestriction activity. Multimutants that exhibit full antirestriction activity possess a 2AP resistance >80µg/ml, whereas multimutants with no activity possess a 2AP resistance of 20µg/ml. In addition those mutants which appear to be toxic to *E. coli* NM1041 cells upon induction are partially active displaying weak binding affinity towards the MTase and nuclease.

3.11 Discussion

3.11.1 Development of a robust *in vivo* assay

The importance of an informative screening method when assessing possibly hundreds of mutants is paramount. Not only should the method provide fast, reliable and reproducible results but the results should provide as much detail as possible to justify the investment in time for construction of the library. Measuring the 2AP resistance of *E. coli* NM1041 cells without the *clp* gene, provides a robust technique to assess the inhibitory activity of

antirestriction proteins such as wtOcr. The assay relies on the ability of the antirestriction protein to rescue cell growth and by doing so provides a highly visible phenotype.

This chapter has shown that the presence of the antirestriction proteins wtOcr and ArdA, rescue *E. coli* NM1041 cells treated with 2AP. The assay was enhanced to select for resistance to different concentrations of 2AP and then used to assess mutants of Ocr known to have variable antirestriction activities. Screening of the Ocr multimutants revealed several phenotypes, the mutants M1, M2, M4, Ocr99 and Ocr109 exhibited 2AP resistance at the maximum concentration of 2AP, 80µg/ml, indicating fully active antirestriction properties. Mutants M7, M11 and M13 provided no antirestriction activity, while a third phenotype, reduce the growth rate of *E. coli* NM1041 cells.

The third set of Ocr multimutants that induced a deleterious effect on the growth of *E. coli* NM1041 cells were examined further. Bacterial growth curves of *E. coli* NM1041 cells expressing each of the Ocr multimutants showed that the cellular growth rate differed for each of the deleterious mutants. Cultures grown with the mutants M12 and M16 possessed a reduced growth rate. Upon over expression this rate was further reduced to no growth. The mutants Ocr/POcr and M3 have a lesser effect on the growth rate and only when over expressed. Therefore the activity of the intermediate mutants has been further differentiated; the mutant M3 inducing minor decreases in cellular growth, this negative effect on the growth of cells increases upon expression of the mutants Ocr/POcr, M16 and M12. M12 reduces growth to the greatest extent.

3.11.2 Partially active phenotype

To pinpoint the cause of the deleterious effect, viable cell counts in the presence and absence of the restriction subunit, HsdR, were compared. If viable cell counts were reduced in both bacterial strains it would suggest that the mutants are interacting with another component of the cell. However, viable cell counts show that the cells were rescued in the absence of the restriction subunit for each of the mutants (M3, M12, M16, Ocr/pOcr). This implies that the reduction in cell count is dependent on the HsdR subunit suggesting that the restriction enzyme EcoKI is targeting the host DNA. Microscopy was carried out to see if cells grown with these particular mutants had become susceptible to restriction, from the HsdR subunits. These results show that expression of the deleterious mutants caused cells to filament. Upon damage to DNA the host cell initiates the SOS response and, as part of this response, cell division is inhibited but replication continues. This indicates that the deleterious mutants are not inhibiting restriction efficiently in agreement with the *in vitro* studies carried out by A.

Stephanou which showed that these mutants possess a poor binding affinity towards the M.EcoKI enzyme. The deleterious mutants M3, M12, M16 and Ocr/pOcr induce an $M^{+/-}R^{+/-}$ phenotype which is detrimental to the growth of *E. coli* cells in the presence of a Type I RM enzyme.

Such a phenotype could be caused by differential binding of the Ocr multimutants to the methylase or nuclease complex although it seems unlikely that the Ocr multimutants could inhibit the methylase completely and not the nuclease, or *vice versa*, given the nature of the Ocr binding pocket on EcoKI. The difference in location of the two active complexes and their accessibility could cause differences. The nuclease is thought to mostly reside outside the nucleoid compared to the methylase which needs to be in close proximity to the hemi-methylated product of DNA replication to modify the DNA accordingly.[40] Furthermore the quantity of M.EcoKI in the cell is slightly greater than the amount of R.EcoKI. In the presence of weak binding inhibitors, the efficiency of the methylation and restriction activity will not be equally affected and as such the unmodified chromosomal target sites could become susceptible to restriction.

3.11.3 Charge-function relationships

The Ocr multimutants targeted specific clusters of acidic residues on the surface of the Ocr protein, these acidic residues mimic the charge displayed by the phosphate backbone of B-form DNA. The mutations targeted groups and clusters of acidic residues from different domains from the structure of the Ocr monomer and the acidic residues were substituted for analogous neutral amino acid residues; hence D/E residues were substituted with N/Q respectively. Examination of the activity of each Ocr multimutants and the location of its mutations on the Ocr protein can identify regions and clusters of negative charge that are important in the interaction between Ocr and the EcoKI Type I RM enzyme.

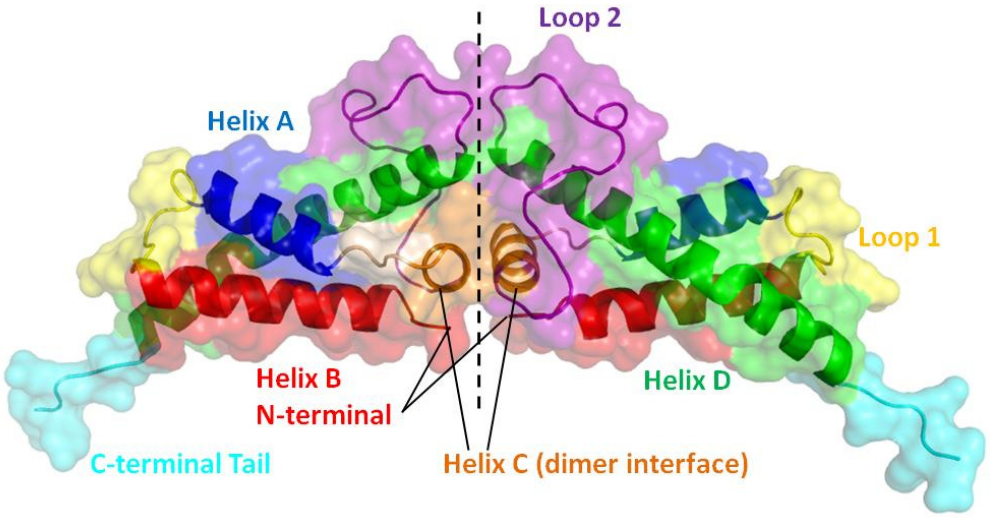
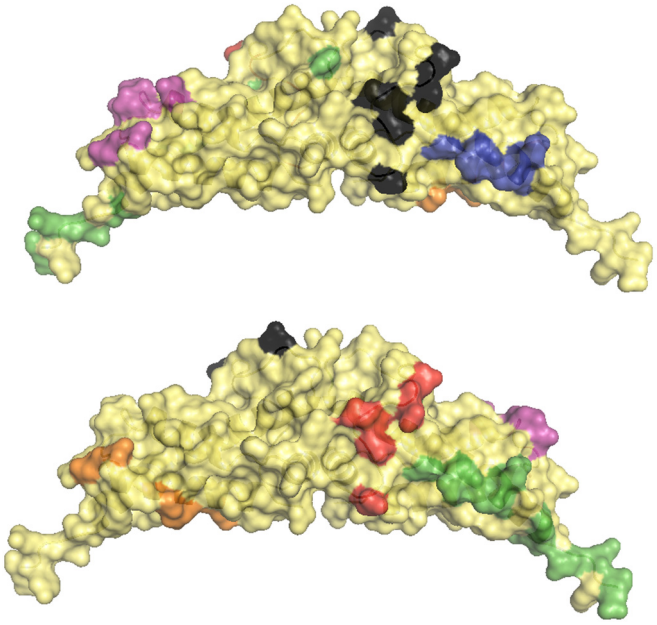


Figure 3.10: The Ocr dimer crystal structure. Each of the domains that have targeted in the mulitmutants has been represented in a different colour, M1-Helix A, M2-Loop 1, M3 and M12-Loop2, M4-Helix D, M7-Loop1+HelixD, M11-Loop1+Loop2, M13-Loop2+HelixD, M16-Loop2+Helix A and Ocr/POcr-Helix D+Tail.



Mutant name	Amino Acid Changes	Total mutated residues	Domain
pTrc99A	n/a	n/a	n/a
wtOcr	0	0	n/a
M1	D12N, E16Q, E20Q, D25N, D26N	5	HA
M2	D25C, D26N, D29N, D31N, D32N, D35N	6	Lp1 + HB
M3	E60Q, D62N, E64Q, E66Q, D67N, D73C	6	Lp2
M4	D92N, E95Q, D76N, E98Q, D99N	5	HD
M7	M2+M4	11	Lp1+HD
M11	M2+M12	11	Lp1 + Lp2
M12	E60Q, D62N, E64Q, E66Q, D67N	5	Lp2
M13	M4+M12	10	HD + Lp2
M16	M1+M12	10	HA + Lp2
O/P	D76N, E87Q, D92N, E95Q, D96N, E98Q, D99N, E103Q, E106Q, E107Q, E109Q, E110Q, E112Q, E113Q, D114N, E115Q, E116Q	17	HD + Tail
Ocr99	Truncation at L99	16	HD + Tail
Ocr109	Truncation at E109	6	Tail

Figure 3.11: The Ocr homodimer structure with the location of the mutations for each of the multimutants, the colours coordinate with the table below which outlines the specific mutations carried out and the domain these mutations reside in from the Ocr crystal structure, Walkinshaw et al, 2002.(46) Domains include; Helix A, aa7-24 (HA), HelixB, aa34-44 (HB), Helix D, aa78-106 (HD), C-terminal region, aa107-117 (Tail), Loop1, aa25-33 (Lp1) and Loop2, aa58-77 (Lp2).

The mutants M1, M2 and M4 appear to have no effect on the antirestriction or antimodification activity of the protein. However, the combination of M2 and M4 to make mutant M7, seems to reduce binding between the Ocr mutant and the nuclease complex. On the other hand the mutants M3 and M12 produce deleterious effects in *E. coli* cells in the presence of the HsdR subunit. The residues targeted by M12 and M3 dominate the Loop 2 region of the Ocr structure, the mutations are identical with the exception of D73C in mutant M3. This single amino acid change reduces the deleterious effect induced by the mutant M3 compared to M12, suggesting an increase in the stability of the interaction with the Type I RM enzyme. The loop region is centrally located on the homodimer and the Ocr:M.EcoKI models show that this region is in close proximity to the HsdS subunit, in particular the TRD domains. The combination of the mutant M12 with either M2 or M4 appears to inactivate the Ocr protein and no binding is exhibited by the mutants M11 (M12+M2) and M13 (M12+M4). This confirms that the acidic residues mutated in M2 and M4 have a smaller impact on the activity of wtOcr. The residues on mutant M2 are located on Loop1, which is also shown to be in close proximity to the TRD domains of the HsdS subunit and possibly the HsdM subunit.

In contrast the acidic residues targeted on mutant M4 reside on the elongated Helix D, see figure 3.10. Models of the EcoKI and Ocr homodimer indicate that this domain may interact with the HsdR subunit, the models show that this helix domain does not appear to contact the HsdS or HsdM subunits (Figure 3.12). This could account for the difference in binding affinity that is exhibited by the mutant M13 upon binding to the M.EcoKI (weak binding) and the EcoKI enzyme (no activity) and the reduced antirestriction activity of mutant M7. Both of these mutants contain the mutations targeted in M4 in combination with other regions of interest. However it is the acidic residues located on the loop regions of the Ocr homodimer which appear important for its interaction with M.EcoKI with the Loop 2 region being of primary importance.

Overall the results show that the central loop region represented by mutant M12 contains acidic residues that are important for the interaction with the EcoKI enzyme. The residues represented in M2 (loop1) and M4 (Helix D) are also important but to a lesser extent. Early indications suggest that the residues in M2 bind to the M.EcoKI complex whereas the residues in M4 may interact with the HsdR subunit and contribute by stabilising the interaction of the EcoKI nuclease with the Ocr protein. However it is important to note that on their own these regions have no noticeable effect on the antirestriction or antimodification activity of the protein.

Another series of mutations measured the effect of C-terminal truncations on the antirestriction activity of the Ocr proteins. Previous investigations of the negatively-enriched C-terminal region showed that it appeared to be of little importance to the overall antirestriction function of the protein. Analysis of the truncations Ocr99 and Ocr109 confirm this as each induced no effect in the 2AP assay and only nominal effects were found *in vitro*. However mutations that combine the M4 region and the C-terminal region with the addition of D76N (OCR/POCR) appeared to cause a deleterious effect in the 2AP assay even though the *in vitro* data showed a strong interaction with M.EcoKI. This suggests that the residues in mutant M4 are important for the interaction of Ocr with the nuclease complex.

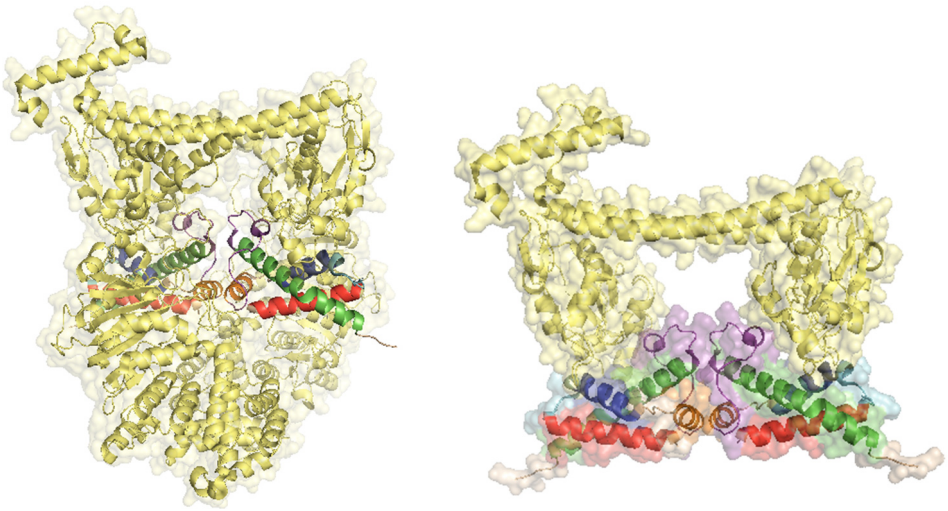


Figure 3.12: Left- Structure of M.EcoKI with Ocr. Structural domains on the Ocr homodimer have been highlighted as shown in figure 3.10. Right- A closer look at the HsdS interaction with the Ocr homodimer, demonstrating the importance of loop 1 and 2 towards the interaction. The domains of Ocr have been highlighted using the same colours as shown in figure 3.10.

To conclude, this chapter has outlined the development of an assay that is relatively fast and reproducible and displays a strong phenotype to select for active antirestriction proteins. By extending the assay to include a titration of 2AP, the assay can select for inactive and partially active antirestriction proteins, these phenotypes correlate with *in vitro* binding affinity measurements. The results have shown that mutants that possess a weak binding affinity also affect the viability of *E. coli* cells in the presence of the HsdR subunit and display a phenotype of $m^{+/-}r^{+/-}$. Overall, the results show that acidic residues that reside in the loop 2 region and, to a lesser extent, the loop 1 region of the Ocr homodimer make important electrostatic interactions to EcoKI. Furthermore the acidic residues in M4, (Helix D) may

interact with the HsdR subunit of EcoKI and this could indicate a secondary binding site on the Ocr homodimer.

Chapter 4. Development of a POcr library

4.1 Introduction

The surface of the protein Ocr is replete with acidic residues, many of which mimic the phosphate backbone of DNA. Initial studies in chapter 3 have shown that these acidic residues are important for the interaction of Ocr with the restriction endonuclease EcoKI. However, it has also been shown that many of these acidic residues appear to possess little or no functional advantage towards the inhibitory activity of Ocr, which suggests that the positional context of the acidic residues as well as the number of acidic residues could both play an important role in the antirestriction activity of the protein Ocr.

This chapter attempts to produce a plasmid library which targets each of the acidic residues in a random manner, to produce mutants of Ocr with differential pI values in order to determine the minimum number of acidic residues needed to provide full antirestriction activity. To achieve this, the gene POcr, 'positive Ocr', was synthesised by Geneart (now part of Invitrogen) using their GeneAssembler process. When the gene encoding POcr was engineered for expression no recombinant protein could be detected. This was not surprising, given that 34 out of the 117 residues (~30%) of Ocr had been substituted. Such a large number of mutations were likely to adversely affect the overall structural integrity of the protein leading to folding/stability issues. Nonetheless, the synthetic gene encoding POcr could be used as a template, in a PCR-based approach, to gradually reintroduce the acidic residues of wtOcr. It was essential that the reintroduction of acidic residues at the specified positions was a random process without bias. Although POcr displays no antirestriction activity, it was envisaged that the gradual introduction of acidic residues would at some stage generate an active protein.

4.2 Aims

This chapter will describe how a series of libraries were developed that could target specified positions in the protein sequence of POcr and reintroduce acidic residues randomly with a controlled mutation rate. Mutations targeted neutral residues located on the POcr protein sequence, which correspond to the acidic residues on wtOcr. By slowly reintroducing the negative charge of wild type Ocr onto the POcr protein, nine libraries were generated producing Ocr mutants which were gradual mutated from POcr back to wtOcr.

4.3 Strategy

To introduce each acidic residue in an unbiased manner, ISOR was used to generate a series of libraries. Each library introduced an increasing number of acidic residues, by manipulating the POcr gene. Due to the close proximity of some target sites the oligonucleotides developed that substituted Q/N residues to E/D residues were separated into four groups and this produced four initial libraries; Lib1, Lib2, Lib3 and Lib4. Collectively these libraries targeted each of the neutral amino acid residues on the POcr protein, which are acidic on wtOcr. As a result, the four libraries were subsequently combined to produce a fifth library, the Shuffle Lib, which targeted all of the appropriate sites on POcr.

The recombination of fragments from the libraries Lib1-4 to produce the Shuffle Lib produced POcr mutants with a range of only 1-8 acidic residues out of a possible 34 (explained in detail below). As such, additional oligos were redesigned with degeneracy. However, because of overlapping target sites these oligos were separated into two groups generating a further two libraries; 1D and 2D. These two libraries were subsequently combined using ISOR to produce the library 1D+2D. The final library, LibDg, was created to increase the number of acidic residues further by introducing a third group of degenerate oligos to library 1D+2D.

Hence nine libraries were generated; the rest of this chapter will describe in detail the efforts in producing each of these libraries.

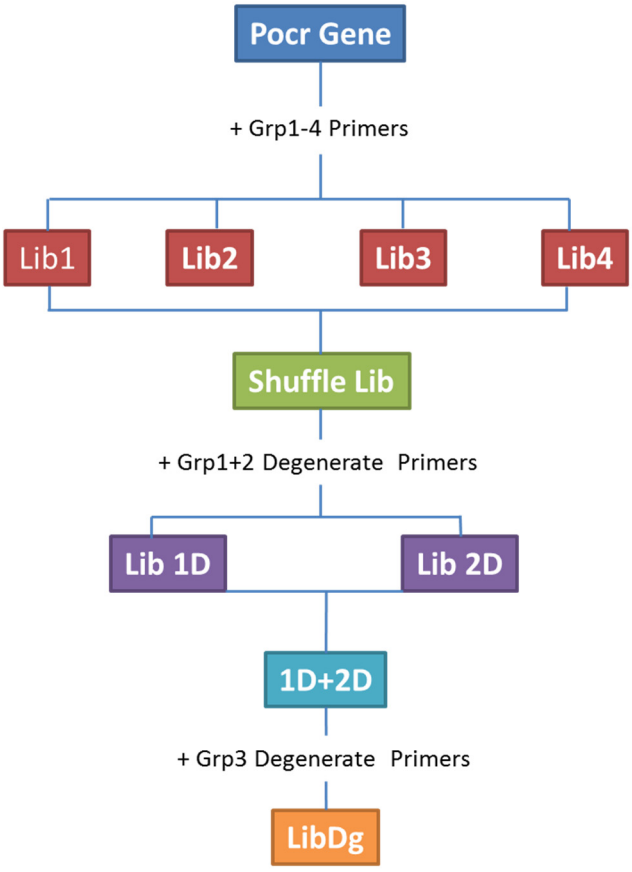


Figure 4.1: A flowchart that shows how the nine POcr libraries were developed. Staring from the original template of the POcr gene, a group of specifically designed primers that targeted neutral residues across the surface of the POcr protein were each reassembled with fragments of the POcr gene producing four libraries, each of which contained a different set of target sites. Each of the four libraries was recombined to produce the Shuffle Lib. In turn the Shuffle Lib was exposed to two different sets of degenerate primers which targeted groups/clusters of neutral residues this generated the libraries Lib 1D and Lib2D which were subsequently combined to produce the library 1D+2D. A final set of degenerate primers were then exposed to fragments of the library 1D+2D and thus produced LibDg.

4.4 POcr Libraries: Lib1-4

The ISOR method for producing plasmid mutant libraries is described in chapter 2, figure 2.2. The ISOR cycle consists of a series of stages, each stage requires a large amount of DNA. Therefore, stage 1 consists of a PCR amplification step of the POcr DNA sequence followed by a digestion reaction. The POcr insert was digested to produce fragments, which upon recombination with specifically designed primers introduced the appropriate mutations. For non-specific digestion the DNAseI nuclease was used. DNAseI cleaves DNA at phosphodiester sites adjacent to a pyrimidine nucleotide (C or T). The POcr insert is a small gene, 351bps long and as such the digest reaction was developed to yield fragments that were ≤100bps in length. This was regulated by varying the digestion time and optimal fragments were observed with an 8min digestion, (Figure 4.2).

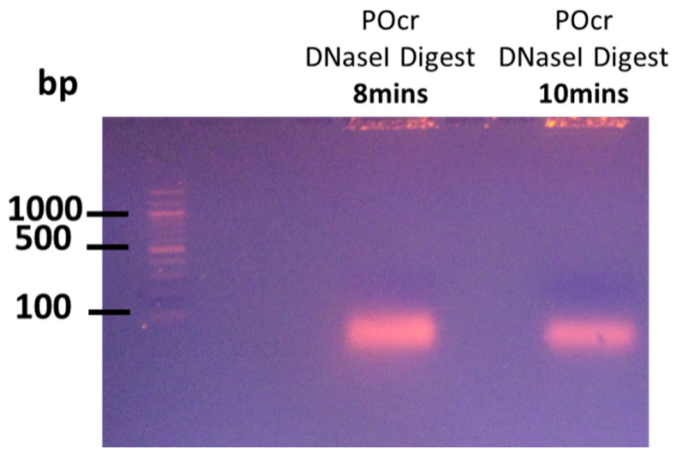


Figure 4.2: A 1.5% DNA agarose gel of the POCR amplification product digested with the enzyme DNaseI for 8mins and 10mins. Both digest reactions have yielded fragments of the insert which are 100bp or smaller. However the 8min digest appears to yield a greater number of fragments, the 10min digest may lose more fragments in the agarose gel.

The POcr fragments were recombined with the addition of primers, however many of the target sites on POcr are in close proximity to each other. An effort was made to design oligos with close target sites on opposing sense and antisense strands of the POcr gene. However this was balanced with the need to design oligos that were the same length and with a similar GC content in order to reduce variability in the Tm temperature. As a result, many of the designed oligonucleotides contained flanking regions that overlapped other target sites. As such the oligos were split into four groups, to avoid primers with overlapping regions competing for positions upon the reassembly of POcr gene fragments, see figure 4.3.

Subsequently the four groups of oligos were reassembled with fragments of the POcr gene to generate four mini libraries.

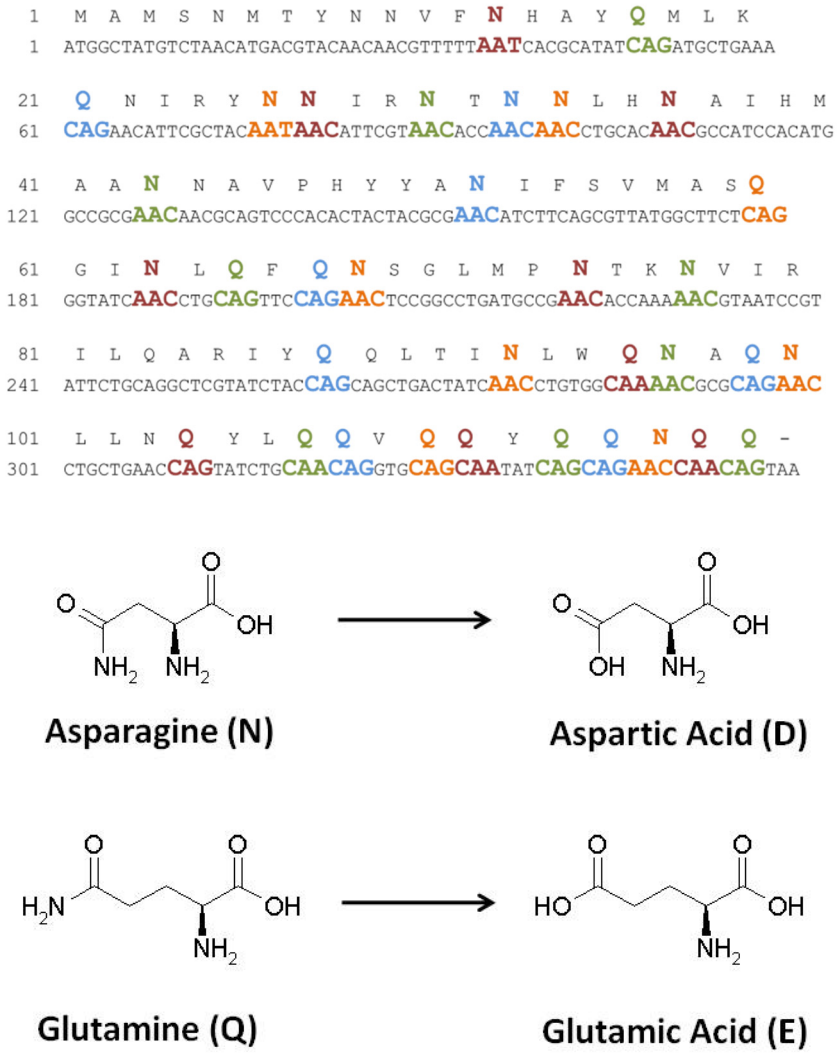


Figure 4.3: Above, the POcr DNA and resulting protein sequence. Highlighted in colour are of all the neutralised residues that were targeted for mutagenesis using the ISOR library methodology. Each color represents the targets for each of the four libraries where; Red is Lib1, Green Lib2, Blue Lib3 and Yellow Lib4. Below, the chemical structures of the amino acid residues involved in the mutagenesis, the neutral residues; asparagine and glutamine were targeted for mutagenesis to aspartic acid and glutamic acid, respectively.

From previous work carried out by the Tawfik lab, it was found that the mutation rate of an ISOR library could be controlled by varying the oligonucleotide concentration.[119,107] A total oligonucleotide concentration of 144nM could give an average of 6 mutations per library variant, however if the concentration exceeded 360nM it would inhibit the reassembly

reaction.[119] To produce the four libraries Lib1-4 a total primer concentration of 200nM was used in the reassembly reaction, with the aim of achieving the maximum mutation rate (8 mutations per gene). Each of the four groups contained 8/9 oligos resulting in a concentration of 25/22nM for each oligo.

The reassembled POcr genes were subsequently amplified using a nested PCR reaction. This is achieved by designing primers which amplify a target within the original template, in this case within the POcr gene insert. Oligos were designed to overlap the start and stop codons of the gene insert, incorporating the restriction sites NcoI and HindIII to insert the gene into the pTr99A plasmid. The nested PCR products were purified using a 1% DNA agarose gel resulting in only one DNA species which was between 300-400bp in size (Figure 4.5), the same size as the original POcr gene (379bp).

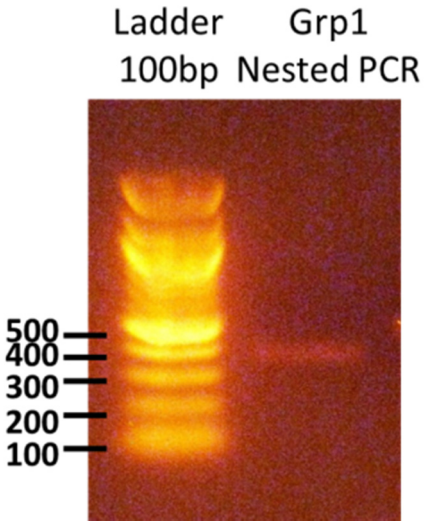


Figure 4.4: A small amount of the nested PCR reaction of the group 1 library was run on a 1% DNA agarose gel. By comparing the gel band present in the Group 1 PCR product with a DNA ladder the DNA species is between 300-400bp long.

To test the success of the four initial libraries each was subcloned into the plasmid vector, pTrc99A. In order to plate out the complete diversity of the library onto LB-agar plates, XL-gold supercompetent cells were used to carry out the initial transformation. Supercompetent cells have a high transformation frequency of 5 x 10^9 transformants/µg therefore, these cells are ideal when plasmid product is limited such as for ligation reactions and for obtaining the full capacity of the libraries from using a small sample. Each library targets 8 sites providing two potential residues per site. The maximum number of variants for each of the libraries is 2^8 = 256 different mutants. In addition transformation frequencies of the empty pTrc99A

vector and the library showed that ~1% of the transformant mix of the library could potentially be background colonies of the empty plasmid.

To ensure that enough of the library is represented, the transformation mix was plated out to accommodate 10-fold more colonies (2560) than the maximum 256 variants, this was to accommodate for repeating sequences, background colonies and any potential non-targeted point mutations incurred through the ISOR process. To assess the success of the initial libraries ~10 colonies were picked from each library; the plasmids were isolated and sent for DNA sequencing.

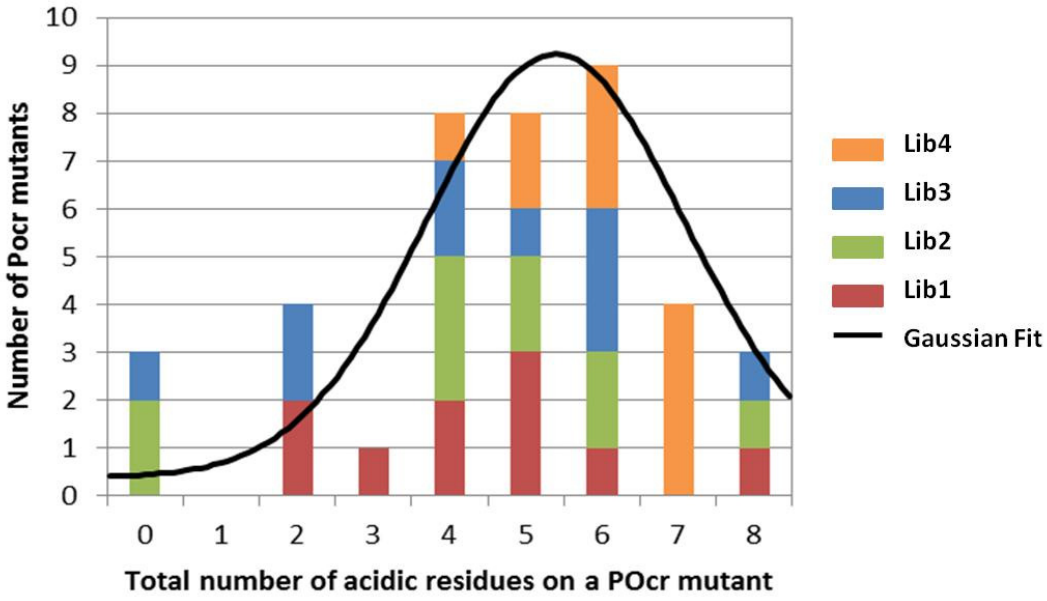


Figure 4.5: Ten POcr multimutants were isolated from each library. The bar chart is a histogram representing how many acidic residues were introduced to POcr multimutants belonging to Lib1, Lib2, Lib3 and Lib4. The distribution of the number of acidic residues that were introduced to all 40 mutants that were isolated from the libraries Lib1-4 was fitted to a Gaussian curve.

From the sequencing results 37 unique POcr mutants were isolated, each targeting different combinations of the eight target sites presented in the libraries Lib1-4. Each of the libraries produced POcr mutants with a range of 0-8 acidic residues, therefore a high diversity appears to be present in each of the four libraries. A Gaussian fit (Figure 4.5) shows that on average 4-6 acidic residues were present on POcr mutants isolated from each of the libraries.

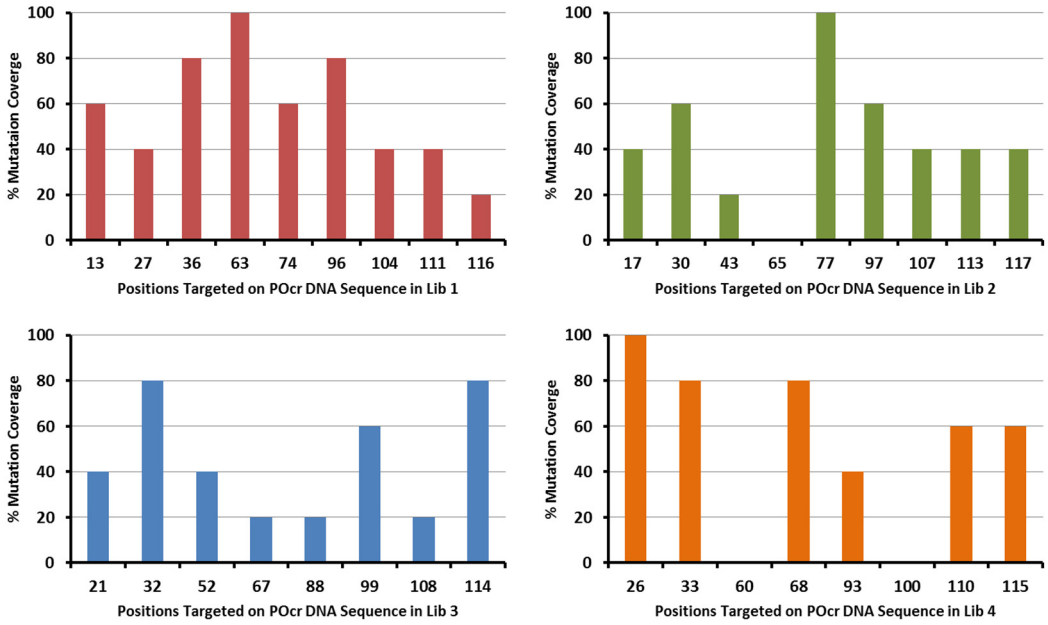


Figure 4.6: From the 10 colonies that were sequenced from each of the libraries 1-4, each graph above represents the mutation coverage for each target site covered by these sequences, as a representation of the diversity and mutational coverage of the sites in each library. Where; Red is Lib1, Green Lib2, Blue Lib3 and Yellow Lib4.

A closer look at sequences of POcr mutants from each library, shows that some mutations were more abundant compared to others. Figure 4.6 shows that the sequences from Lib1 and Lib3 appear to mutate all of the target sites with a mutational coverage of 20-100% for each site. However, the libraries Lib2 and Lib4 both contain target sites which have not been mutated at all. However this difference is slight when considering the sample size, the 10 sequences used to assess each of the libraries barely represents 3% of their total population.

The production of non-specific mutations and indels within the libraries were measured. These represent the introduction of mutations; deletions, insertions and substitutions on the POcr insert that have not been targeted. These could arise from multiple rounds of PCR amplification or upon reassembly to the wrong fragment. The table below accounts for all these anomalies, these have been kept to a minimum representing on average 1 in 10 sequences, although this is something to be wary of during repeated rounds of the ISOR methodology.

Library	Number of Indels	Anomaly type
1	1	Truncation
2	1	Unable to Sequence
3	1	Unable to Sequence
4	0	

Table 4.1: A table displaying the number of sequences that contain anomalous mutations or deletions from each of the libraries 1-4 and the type of mutations that were exhibited.

As randomisation of the mutations in the ISOR method relies on the reassembly of fragments from a non-specific digestion, it is impossible to know the full diversity of the libraries, which is why any conclusions that these sequences give us are a rough guide. Therefore to compensate for any bias a 10 fold increase is added to any predicted values in the population size and diversity of the libraries.

4.5 Shuffle Library

The four separate libraries provide a broad and diverse range of mutations, although in each library only a quarter of the desired sites are targeted. Therefore, the four individual libraries were combined in order to obtain a single library in which all potential sites are targeted. Fragments of each library from a DNAseI digest were combined and reassembled using ISOR. After the reassembly of the fragments this 'shuffle' library was cloned back into pTrc99A as previously described, 15 colonies were isolated and sequenced from the subsequent transformation in order to ascertain the potential diversity of the Shuffle library.

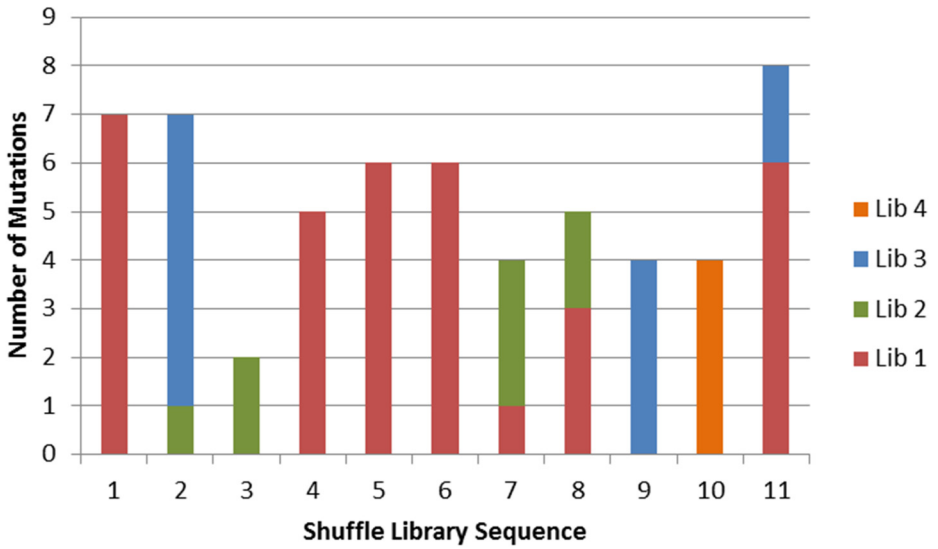


Figure 4.7: A bar chart representing the total number of mutations and whether they were targets from libraries 1, 2, 3 or 4. This highlights the extent of shuffling present in each of the sequences from the shuffle library (NB, 11 out of the 15 colonies isolated were successfully sequenced.

Fifteen POcr mutants were sequenced producing 11 unique sequences and 4 sequences with non-specific mutations/indels. Disregarding those mutants with non-specific mutations the average number of acidic residues per gene remained at ~5. Analysis of each of the sequences from the shuffle library showed that sequences contained only small mixtures of target sites from each of the libraries, for example >60% of the sequences had target sites from only one of the four original library groups. Those that did show mixtures were limited, no greater than two groups were shuffled in a gene and the distribution of the sequences was limited, see figure 4.7. This is highlighted by the fact that almost none of the sequences had mutations on target sites that were next to each other as shown in figure 4.8.

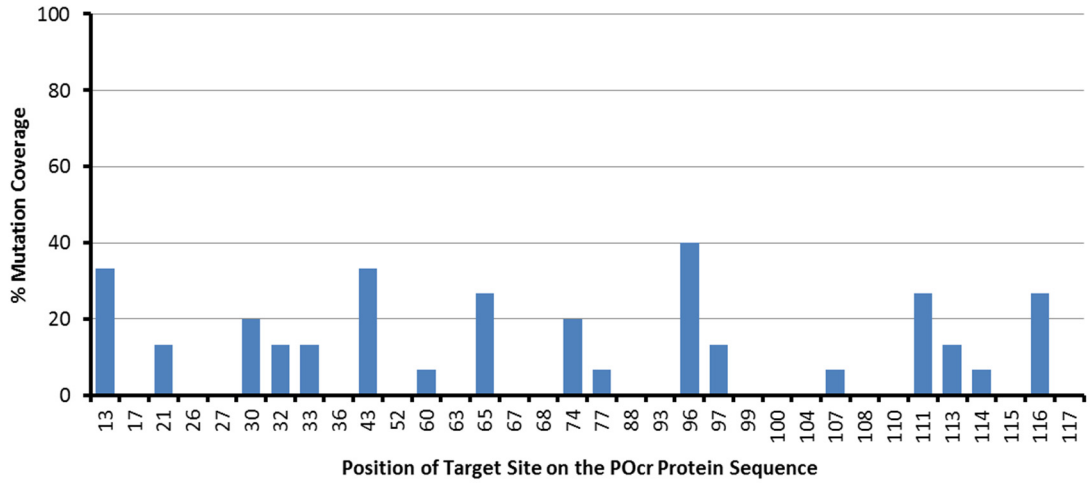


Figure 4.8: A graph depicting the percentage mutation coverage of each of the target site residues. The mutations coverage is the percentage of mutations from neutral to acidic residues that have been identified at each target site from the sequences that have been solved for the shuffle library.

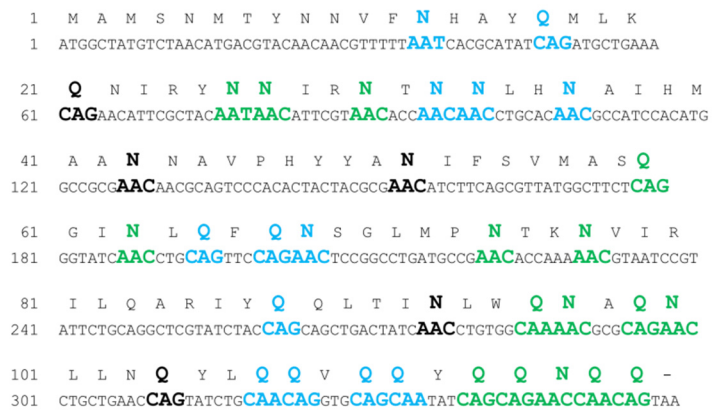
In fact, sequences isolated from the shuffle library showed that the mutational coverage ranged from 0-40% for each acidic residue. Acidic residues had not been introduced to 17 different target sites. A closer look at the mutational coverage in figure 4.8 does show a correlation between those mutations underrepresented and target residues which are in close proximity to each other. The shuffle library showed that residues next to each other were especially difficult to target; for example, residues at positions 26-27 and 67-68. In order to reduce bias oligonucleotides were designed to be the same size and therefore possessed overlapping target sites. Due to this and insufficient fragmentation, it was realised that the shuffling could inadvertently mutate the acidic residues that were present in one library, back to neutral residues once shuffled. This would be most prominent for target sites in close proximity to each other.

4.6 POcr Libraries: Lib1D and Lib2D

To increase the number of acidic residues on the POcr gene, degenerate oligonucleotides were used to target cluster regions in close proximity on the POcr gene. The results of the shuffle library showed that gaining mutations from residues next to each other was difficult using separate oligonucleotides so the use of degenerate sequences can overcome this, by targeting a series of neutral residues at the same time with one oligonucleotide. However the design of these primers needed to be carefully considered; each of the residues targeted must have adequate flanking regions to anneal; they must be the same, contain the same GC

content, with no overlapping target sequences and to ensure equal representation of each residue, the degenerate codons should translate to a 1:1 ratio of acidic: neutral residues for all mutations concerned.

Due to the size of POcr, overlap between non-targeted regions of the wild type POcr protein could not be avoided in the design of the primers. Therefore the newly designed degenerate oligonucleotides were separated, this time into two groups to avoid any overlapping wild type sequence in the same reassembly reaction and prevent the primers from competing to anneal to the POcr gene.



Primer Name	Target Sites	Primer Sequence
N113-117	113,114,115,116,117	CGCCAAGCTTTTACTSTTSGTYCTSCTSATA
N96-100	96,97,99,100	CTGTGGSAARACGCGSAGRACCTGCTGAAC
N74-77	74,77	GAATACGGATTACGTCTTTGGTGTCCGGCATC
N60-63	60,63	GCGTTATGGCTTCTGAGGGTATCGACCTG
N26_27	26,27	CATTCGCTACGATGACATTCGTAATACC
N26_30	26,30	CATTCGCTACGATAACATTCGTGATACC
N27_30	27,30	CATTCGCTACAATGACATTCGTGATACC
N107-111	107,108,110,111	TATCTGSAASAGGTGSAGSAATAT
N88	88	CTGCAGGCTCGTATCTACGAGCAGCTGAC
N63-68	63,65,67,68	GGTATCRACCTGSAGTTCSAGRACTCCGGC
N13-17	13,17	CGTTTTTAATCACGCATATGAGATGCTG
N17	17	CGTTTTTAATCACGCATATGAGATGCTG
N30-36	30,32,33,36	ATTCGTRATACCRACRACCTGCACRACGCC

Figure 4.9: Above, the POcr DNA and complementary protein sequence. Highlighted in color are of all the residues that were targeted for degenerate oligonucleotides. Each color represents the targets for each of the two libraries; green refers to Lib1D oligos and in blue the oligos in 2D. In black are stray residues that have been targeted previously but are not present in clusters therefore not targeted in this round of ISOR.

Using ISOR, the degenerate primers 1D and 2D were incorporated into fragments of the shuffle library, producing two different libraries, Lib1D and Lib2D. The agarose gel which displayed the products of the nested POcr reaction for both libraries showed that the POcr gene had successfully reassembled at ~385bp. However, additional bands were present on the agarose gel, showing the presence of DNA species which were 200bp long for both libraries and an additional band at ~300bp for library 1D, (Figure 4.10), which could be a result of mispriming events. Fortunately, the truncations observed in figure 4.10 could be

easily removed upon purification of the nested PCR products, by excising the gel band that corresponds to the size of the fully reassembled gene, at 385bp.

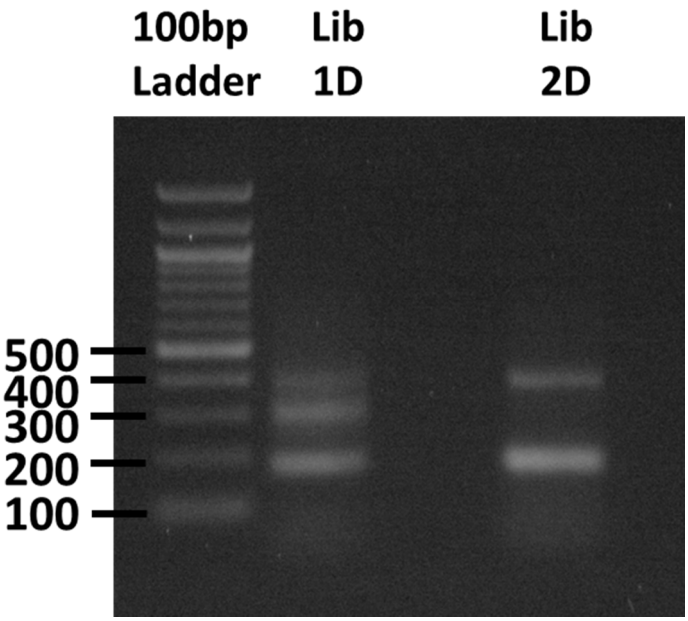


Figure 4.20: The DNA species present in the PCR products of the reassembled libraries 1D and 2D.Lib 1D contains DNA products of sizes 200bp, ~300 and 400bp long. The weakest band represents the gel band of interest (400bp). Similarly the library 2D has multiple species however only two different species a band at ~400bp long appears to be the same size as the pOcr gene and the band at 200bp could represent a mispriming event which produces a truncated pOcr sequence.

4.7 POcr library: 1D+2D

The libraries Lib1D and Lib2D were shuffled using ISOR in the same way as described previously. The diversity and successful introduction of acidic residues in close proximity was assessed by, isolating colonies to be sequenced before and after the shuffle of the libraries (Figure 4.11).

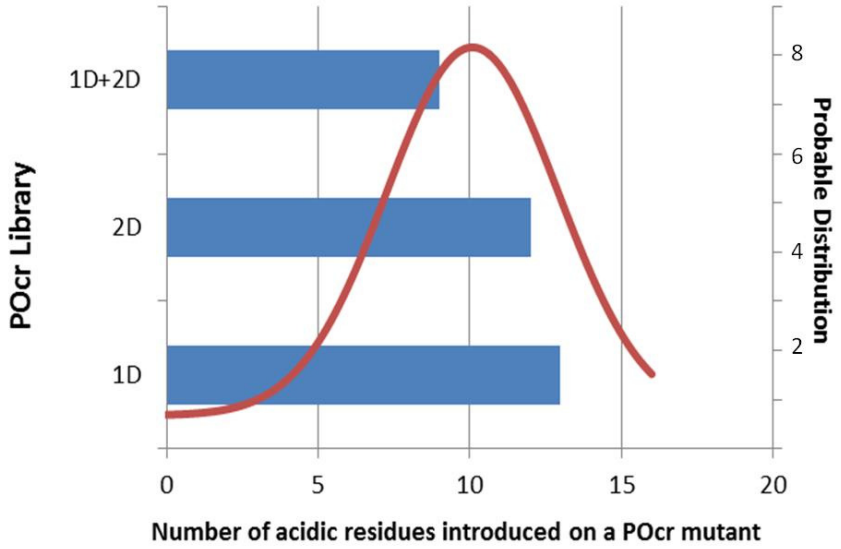


Figure 4.31: The bar chart shows the average number of mutations per gene from the colonies sequenced (3 × 10) for the libraries 1D and 2D and the resulting shuffle library 1D and 2D. The probable distribution is a Gaussian fit of the mutation rate of all the sequences from each of the three libraries; it gives an estimation of the potential diversity of the libraries.

The average mutation rate increased to 10 acidic residues per POcr going up to a maximum of 15 acidic residues reintroduced, this is a marked increase from an average of 5 acidic residues per POcr mutant from the shuffle library. This suggests that the incorporation of the degenerate primers had been successful.

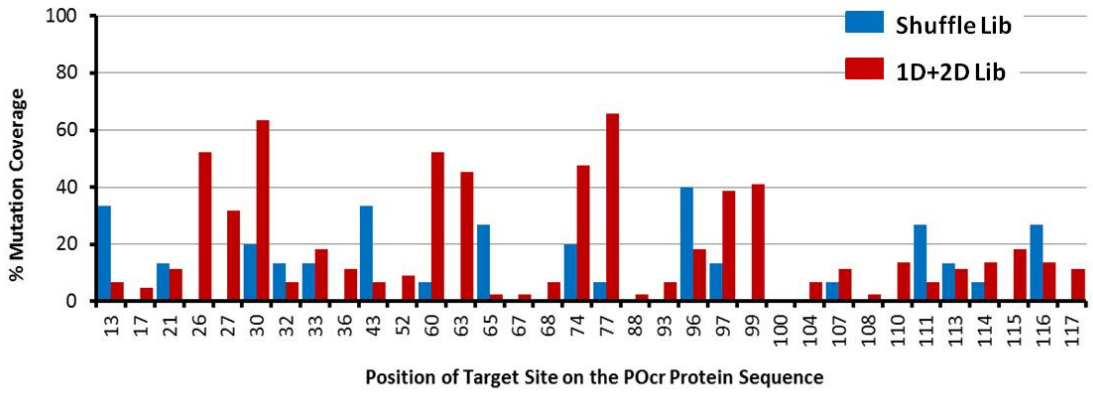


Figure 4.42: A graph depicting the percentage mutation coverage of each of the target site residues. The percentage mutation coverage has been compared between sequences isolated from the library 1D and 2D and the Shuffle library.

Upon closer examination of the sequences isolated from the 1D and 2D library, the mutational frequency of each target site was quantified for 30 sequences isolated from the library. Although this is relying on the results from a small subpopulation of the library, the values obtained provide a rough average of the mutational coverage. It is likely that if an oligonucleotide has not annealed successfully to POcr in the reassembly reaction, then the percentage coverage should be extremely low.

Figure 4.12 shows that from the 30 POcr multimutants sequenced that at least 40% of the neutral amino acids at the target sites 26, 30, 60, 63, 74 and 77 have been mutated to acidic residues. This is a large mutational coverage in such a small population therefore, a good indicator that the oligonucleotides that possessed these target sites were successfully incorporated into the library. Positions 27, 99 and 100 possess 30% mutation frequency. This also highlights that primers targeting these sites were successfully incorporated into the 1D and 2D library.

The percentage of acidic residues introduced at other target sites remained low; for example, the percentage of acidic residues introduced at the target sites 65, 67 and 68 were below 5%. It is difficult to discern whether the low frequency of mutations is because the sampling size is small or whether the oligos designed failed to anneal. However, due to the relatively low frequency exhibited in contrast to other target sites it was assumed that the oligonucleotide containing these positions failed to anneal to the POcr sequence. As such the small percentage of acidic residues at these target sites are thought to be a result of residues incorporated previously from the shuffle library.

4.8 POcr Library: LibDg

To ensure that acidic residues were incorporated at each target site to the same extent, those target sites that possessed a mutational frequency that was less than 5% in the library 1D+2D (figure 4.13) were assumed to be present on primers that did not successfully reassemble, to the POcr library. Therefore a third set of primers were designed for these designated positions (mainly by designing the reverse complement primer) and were incorporated into the library 1D+ 2D by the ISOR cycle (Table 4.2).

Primer Name	Target Sites	Primer Sequence
N32-36_dg	32,33,36	GATGGC**GTY**GTGCAG**GTYGTY**GGT**ATC**ACG
Q65-68_dg	65,67,68	CGGA**GTYCTS**GAA**CTS**CAG**GTC**GATACC
N88+93_dg	88,93	CACAG**GTY**GATAGTCAGCTG**CTS**GTAG
N43D	43	GGGACTGCGTT**GTC**CGCGGC
N52D	52	CTACGCG**GAC**ATCTTCAGCG
Q21E	21	GCTGAAA**GAG**AACATTCGCTAC

Table 4.2: The table displays the oligonucleotides that were used to target the residues underrepresented in Lib 1D and 2D. Those target sites with less than 5% mutation coverage have been represented in primers in this table. This included the addition of a number of single mutations.

No attempt was made to increase the mutation rate of the target sites 100-117, for two reasons. Firstly, experiments have shown that the C-terminal domain appears to have no effect on the antirestriction activity of the Ocr protein. Secondly, and more importantly, the number of residues to target on the C-terminal tail led to the design of primers with large amounts of degeneracy (Table 4.1). The large degree of degeneracy led to a higher rate of mis-priming events during the reassembly PCR, resulting in the mixed DNA species found in the nested PCR products of Lib1D and Lib2D (Figure 4.10). Not only this, the introduction of largely degenerate primers appeared to correlate with an increased number of non-specific mutations/indels from the later libraries. As such, primers targeting residues in the C-terminal tail region of POcr were omitted from further rounds of the ISOR. Therefore, the primers shown in Table 4.2 were used to reassemble fragments of the library 1D+2D in order to generate another library, LibDg. As before, 15 colonies were isolated from the library and sequenced to assess the success of the oligonucleotides incorporated into the library.

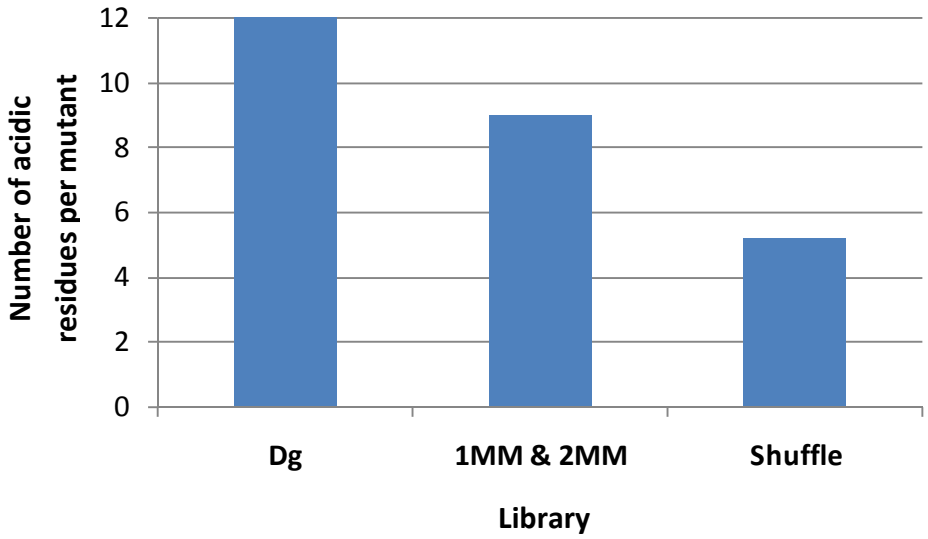


Figure 4.53: The graph compares the average number of mutations per gene, calculated from isolates sequenced from each of the libraries.

A closer analysis of the average number of acidic residues introduced from the 15 POcr mutants sequenced from library LibDg were compared to the average from the libraries 1D+2D and the Shuffle library. The average number of acidic residues introduced into POcr mutants from LibDg is 12, with a range of 8-17 acidic residues per gene. This is a slight increase of 3 mutations per gene, from the previous 1D +2D library.

From the graph below (Figure 4.14) it can be seen that the POcr mutants isolated from the library LibDg possess acidic residues at each of the target sites. The only exceptions are observed for the target sites, at the C-terminal, 100-117. Omitting these positions and assuming that the control parameters enforced in the design of each of the primers provides equal representation, it is assumed that each of the target sites have been exposed to mutations in an essentially random but unbiased fashion.

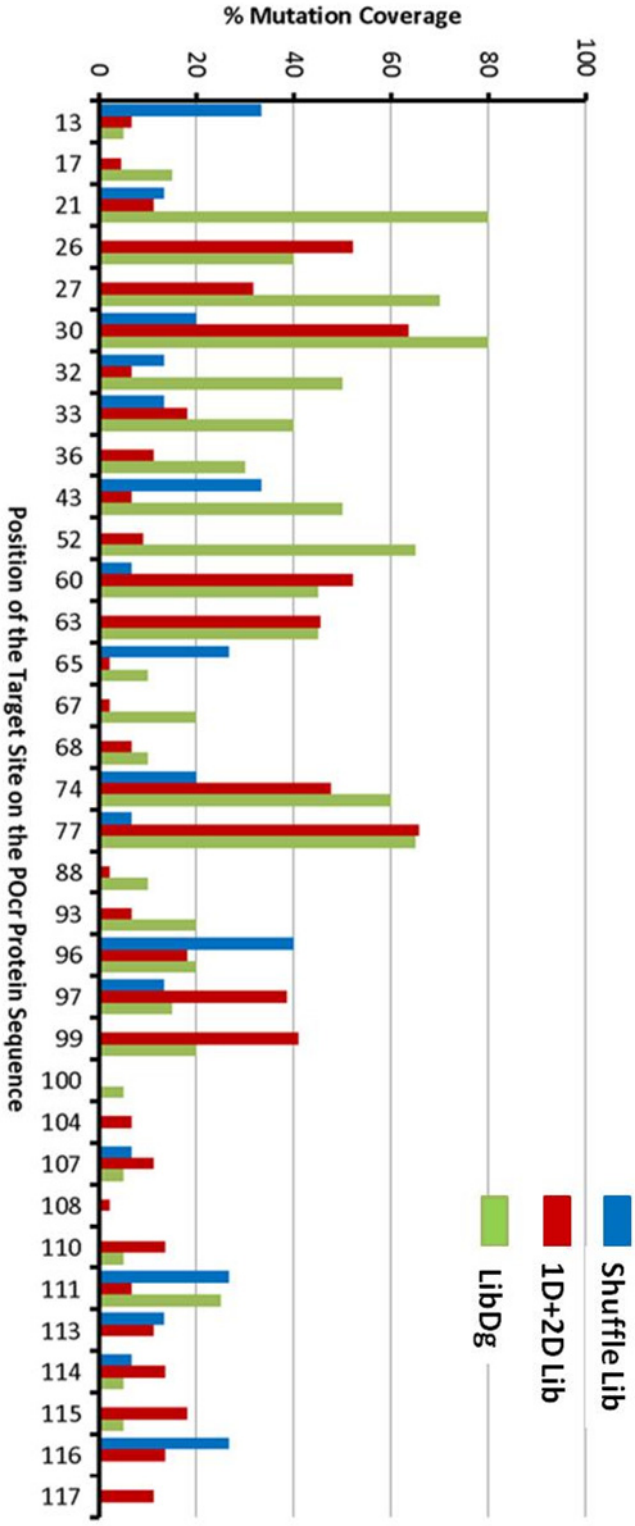


Figure 4.14: Comparison of the mutation coverage, the % of mutants that had acidic residues present at each of the target sites. From the libraries LibDg (green), Lib1D and 2D (red) and the Shuffle Lib (blue).

The estimated diversity of library LibDg can be calculated from the sum of the binomial coefficient for the range of acidic residues introduced. For instance, for LibDg the range of mutations introduced is ~8-16 acidic residues from a total of 34. Even if the target sites from the C-terminal region of POcr were omitted as they were not targeted in the later libraries estimations suggest that potentially 7.8 million POcr mutants are present in LibDg. This number is incredibly large and reflects the need for robust screening techniques.

4.9 Discussion

In summary, a series of nine libraries have been developed that slowly reintroduce codons that translate to acidic residues at specified sites on the POcr protein, slowly converting the POcr gene to the wtOcr gene. Table 4.3 shows that mutagenesis using ISOR has introduced an increasing number of acidic residues to POcr from an average of 3-13 acidic residues per gene. Furthermore, each library contains POcr mutants possessing a range of acidic residues producing diverse libraries with large numbers of unique POcr mutants.

Libraries				
Name	**Range of Mutations**	**Average number of mutations per gene**	**Number of genes sequenced**	**Number of anomalies**
Lib1	3-9	5	10	1
Lib2	0-7	4	10	1
Lib3	2-6	3	10	1
Lib4	5-8	6	10	0
Shuffle	2-8	5	15	4
1D	7-14	12	15	1
2D	2-14	13	10	1
1D and 2D	6-14	9	15	4
Lib Dg	8-16	12	15	5

Table 4.3: A table of the libraries resulting from multiple rounds of the ISOR library, the table summarises the vital statistics from each of the libraries, Their diversity, the average number of mutation per gene, the number of sequences with random point mutations /anomalous sequences and finally the number of sequences that were sequenced from each library.

In the following chapters the libraries will be used to carry out two different screening procedures.

In Chapter 5, the newly developed 2AP assay from chapter 3 will be used as a selection method to screen one million POcr mutants from the LibDg library. POcr mutants that exhibit a high level of 2AP resistance and therefore exhibit antirestriction properties similar

to the wtOcr protein will be isolated, examined *in vivo* and *in vitro* and the minimum number of acidic residues that allows the POcr protein to function like wtOcr will be ascertained.

The second screening method is a large scale screen of 100 POcr mutants that were selected at random from each of the nine POcr libraries produced in this chapter. The Table 4.3 shows that each library contains POcr mutants that exhibit a wide range of pI values. By screening 100 POcr proteins at random the charge-function relationship can be further understood. These POcr mutants were characterised using the 2AP titration assay developed in chapter 3 and possessed variable antirestriction activities and surprising phenotypes.

In summary the production of the POcr libraries successfully reintroduced the acidic residues of wtOcr onto the POcr gene. The gradual reintroduction has resulted in the generation of nine POcr libraries that can be used to select for antirestriction properties. The development of libraries that have incrementally decreased the pI value of POcr allows the investigation of not only the functional advantages of the acidic residues on Ocr but also to investigate the role of these residues towards the general fitness of the protein, therefore how charge might facilitate the structure, fold, solubility and expression of the Ocr protein.

Chapter 5. Maximum activity with minimal charge

5.1 Introduction

In chapter 4 a series of libraries were successfully produced. Using a template of the positive Ocr protein POcr, acidic residues were reintroduced onto the surface of the protein to gradually convert the positive POcr protein into the wtOcr DNA mimic. This chapter describes how the restriction alleviation assay developed in chapter 3 was used as a selective screen to isolate POcr multimutants, from the POcr library LibDg that exhibited antirestriction properties comparable to wtOcr. Library screening for 2AP resistance gives us the means to not only highlight the relative importance of each acidic residue towards the function of the wtOcr protein but also discern the minimum number of acidic residues for antirestriction activity.

5.2 Aims

a) The largest library, Lib Dg, was screened to select for 2AP resistance of ≥80μg/ml. Initial hits from the library screen were further verified using phage antirestriction and antimethylation assays to identify whether these mutants induce the same antirestriction and antimethylation activity as wtOcr. POcr mutants that display activity that is equal to that of the wild type protein were sequenced to discern any commonality and to identify key acidic residues important in the interaction and activity of the wtOcr protein.

b) To ensure that the POcr mutants selected from the 2AP screen contain the absolute minimum number of acidic residues for activity, a further three libraries were generated from the active POcr mutant M2.1. These libraries targeted the enriched number of acidic resides that reside on each of the loop regions. The aim was to decipher whether the acidic residues that are present on each loop region can interchange whilst maintaining full antirestriction activity and whether the acidic residues that are conserved among all four active mutants are essential for the antirestriction activity exhibited by the mutant M2.1.

5.3 Selective screening

To select for antirestriction activity with the 2AP assay *E. coli* NM1041 cells, transformed with the POcr library, were spread onto LB-Agar plates supplemented with carbenicillin, kanamycin and 2AP at a concentration of 80µg/ml. In order to carry out a plate screen, it is important to ensure that the number of library variants per plate is not excessive compared to the concentration of 2AP. If there is an over abundance of cells the 2AP will not be effective, allowing cells to survive and generating a large number of negative results. To avoid this, previous investigations by Dr. Angela Dawson showed that 80µg/ml of 2AP would be effective on two thousand colonies per LB-agar plate (15cm diameter petri dish). It has been shown from previous studies that the maximum number of cells that can be plated is ten thousand per plate in order to allow for; (i) sufficient area for each colony and (ii) full representation of the library. This value of 2000 colonies per plate considers the number of transformed and non-transformed cells which can sequester the 2AP, as well as the number of colonies that can be efficiently plated. To be able to control the number of colonies plated onto each plate the transformation frequency of the *E. coli* NM1041 cells with the library LibDg was determined by plating out viable counts of the transformant mix.

The transformation frequency was determined by plating viable counts of *E. coli* NM1041 cells transformed with 1µl of the plasmid library, at concentrations of 0.99ng/µl, 3.97ng/µl and 39.7ng/µl. The transformant mix for each concentration was plated out onto two LB-agar plates supplemented with the appropriate antibiotics (Km and Carb) with a volume of 100µl and 10µl each. The plates were incubated at 37°C overnight and the number of colonies produced by each transformation is displayed in Table 5.1 below.

Concentration of plasmid DNA (ng/µl)	**Volume of transformant mix (µl)**	**Number of colonies**	**Transformation Frequency (per µg of DNA)**
0.99	100	21	2.12×10^5
0.99	10	1	1.01×10^4
3.97	100	100	2.52×10^5
3.97	10	10	2.52×10^4
39.7	100	267	6.73×10^4
39.7	10	37	9.32×10^3

Table 5.1: The number of colonies observed by transforming *E. coli* NM1041 cells with various concentrations of plasmid DNA from the library LibDg DNA.

Table 5.1 shows that the transformation frequency of *E. coli* NM1041 cells is very low compared to cloning strains such as DH5 alpha. The parent strain for NM1041 is the bacterial strain MG1655, which is also not readily transformed. The cells were made chemically competent using $CaCl_2$ because NM1041 cells have proved difficult to handle with electroporation techniques (unpublished work by Dr Angela Dawson). The table also shows that a similar transformation frequency was observed with 1ng or 3ng of DNA (2×10^5 transformants/ per µg of plasmid DNA).

Therefore, to obtain enough DNA to represent at least 1 million mutants of the library, approximately 6.132µg of the plasmid DNA from LibDg needed to be screened. Therefore if 1ml of transformant mix which uses ~1ng of plasmid DNA produces a total of 210 colonies then 94ng is needed to plate ~20,000 colonies per 1ml which would mean that a total of 20,000 variants of the library would be screened per transformation, this would use 10 LB agar plates supplemented with antibiotics and 2AP at a concentration of 80µg/ml. This results in ~2000 colonies per plate using 100µl of the transformant mix. Therefore to screen 1million POcr mutants from LibDg, a total of 64 transformations need to be performed and so 640 plates need to be poured for the full screen.

The screening process was carried out in three batches. Each batch screened 20 transformations each, which is a total of 400,000 variants of the library. Upon transformation, the plates are incubated for three days. Initially, the transformations are incubated overnight at 37°C however the plates are then left on the bench for a total of 2-3 days and the number of colonies that emerge with each day was recorded.

Batch Number	**1**	**2**	**3**
Day1	9	5	10
Day2	13	14	12
Day3	30	10	33
Day4	21		
Total	**73**	**29**	**55**

Table 5.2: The number of colonies that were produced after three days of incubation for the plates that were screened in batch 1 to 3. Each batch screened 400000 mutants of the library with 200 LB-agar plates with 2AP (80µg/ml).

Each of the three screens that were carried out for LibDg produced $2AP^R$ colonies. The colonies that were produced were streaked onto fresh LB-agar plates supplemented with and

without 2AP for a second screening process; this ensured that $2AP^R$ is due to the POcr mutant expressed.

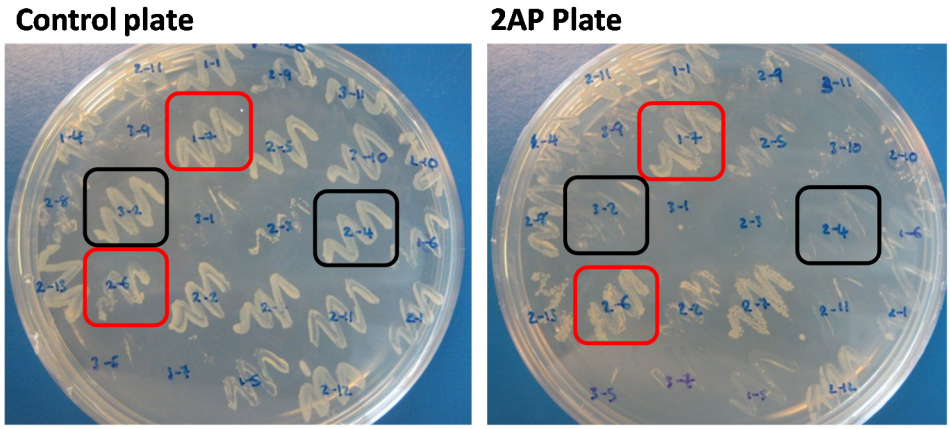


Figure 5.1: A selection of the colonies that were $2AP^R$ for batch 1. Colonies that were $2AP^R$ from the initial screen were re-streaked onto a control plate supplemented with antibiotics only and a plate with 2AP 80µg/ml). Both plates were used to compare the growth of the cells, if cells were truly resistant then homogenous growth would be apparent on both plates. For ease of comparison the colonies were streaked in corresponding positions on the two plates.

POcr mutants that retained their $2AP^R$ grew colonies on both the control plate and the 2AP enriched plate (Figure 5.1, red squares). However many of the initial hits from the screen did not retain their resistance and so did not grow in the presence of 2AP (Figure 5.1, black squares). For those colonies that did appear to retain their resistance to 2AP, the plasmid DNA of the POcr mutant expressed was isolated. These plasmids were retransformed onto 2AP-enriched plates as a triple assurance of their resistant properties.

A total of 73 library mutants were retransformed with *E. coli* NM1041 cells and examined for $2AP^R$ by comparing the viable counts of the mutants in the presence and absence of 2AP. From this only seven out of the 73 mutants that were retransformed appeared to maintain 2AP resistance. Two of these mutants seemed to be partially active displaying a reduced cell count upon 2AP exposure (Figure 5.2, M2.8 and M1.1). Plasmid DNA from each of the seven colonies was prepared and the POcr gene and protein sequence determined.

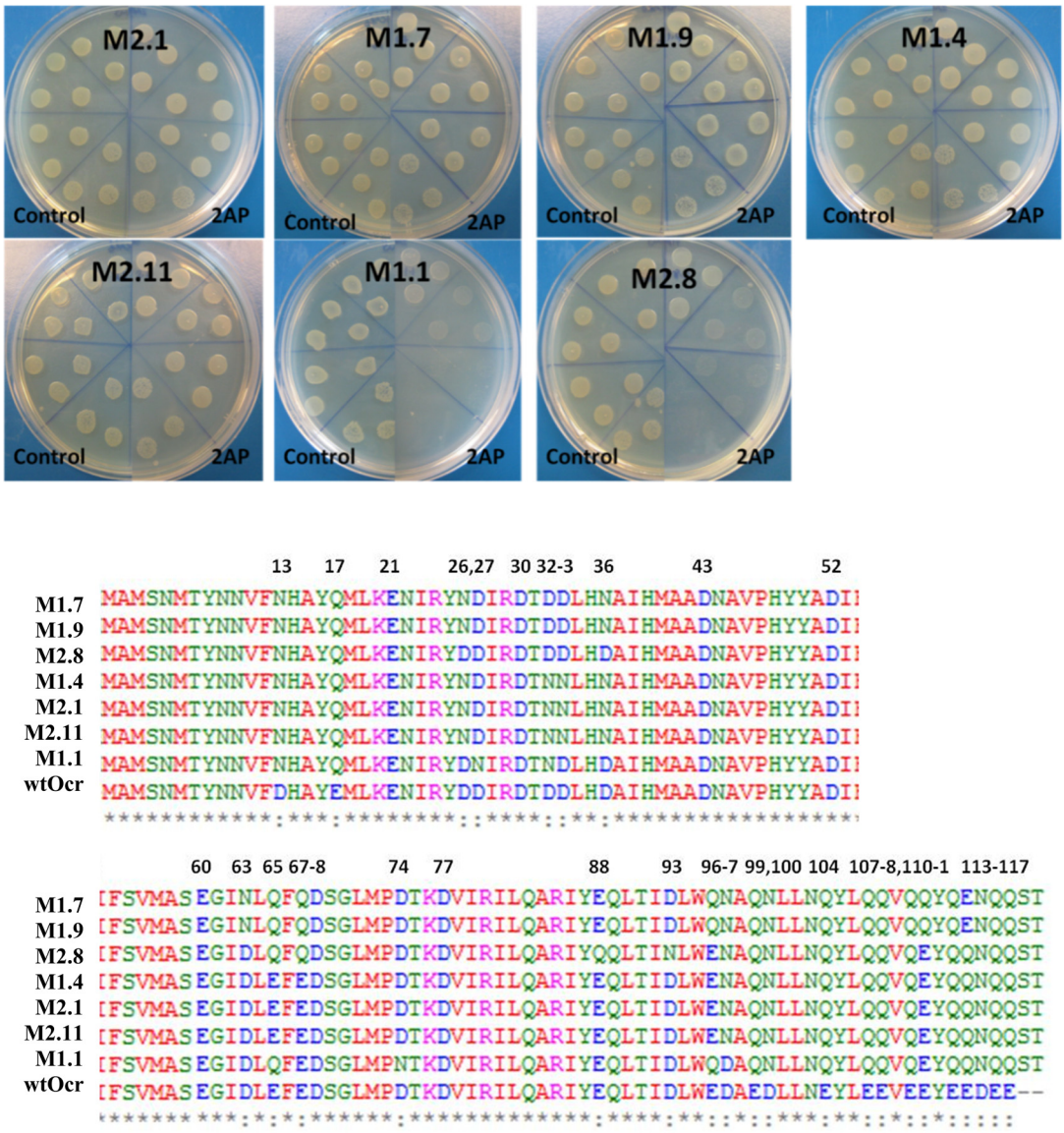


Figure 5.2: Viable cell counts on plates with and without 2AP for mutants screened from the Lib Dg plasmid library of Ocr mutants, which retained 2AP resistance upon retransformation. The viable cell counts show that five out of the seven selected clones retained the total cell count. Mutant clones M1.1 and M2.8 showed reduced cell count. Below: protein sequence of each mutant translated from the corresponding DNA chromatographs. Each mutant is compared to wtOcr.

The viable cell count is much reduced for the POcr mutants M1.1 and M2.8, when exposed to 2AP compared to the other mutants. Therefore, because the resistance of these two mutants to 2AP is reduced in comparison to the other five mutants, one might expect these mutants to display reduced antirestriction activity relative to the other mutants and wtOcr. Thus, the library seems to have selected for partially active and fully active mutants.

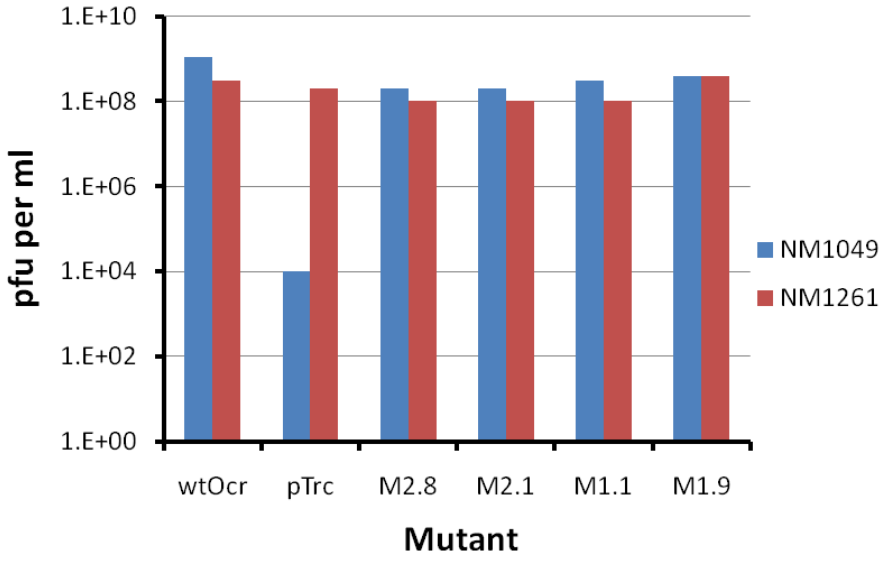
5.4 Active POcr mutants

The protein sequences of the seven POcr mutants isolated from LibDg possessed between 13-15 acidic residues. Even though the number of acidic residues has been reduced by ~60%, from wtOcr, these POcr proteins retain an activity that is almost/if not completely equivalent to the wtOcr protein.

Interestingly, closer inspection of the isolated clones reveals that a number of the sequences that were isolated were in fact the same. Mutants M1.7 and M1.9 had the same gene sequence as did the mutants M1.4, M2.1 and M2.11. Therefore of the seven sequences that have been selected from the screen four of the protein sequences are unique. The acquisition of repeated sequences is however encouraging because it strongly suggests that the particular array of acidic residues on these POcr proteins are important in producing antirestriction activity.

5.4.1 Phage infection assays

To be sure that the 2AP resistance exhibited by the selected POcr proteins is a result of their antirestriction properties, the proteins were examined using phage infection assays which are designed to measure antirestriction and antimethylation activity. These phage assays are similar to those that have been performed previously in chapter 3 for the Ocr multimutants. Restriction proficient and restriction deficient cells are transformed and then unmodified phage lambda, $\lambda_{v.o.}$, is used to infect the cells. The phage titre was calculated upon overnight incubation of the cells at 37°C. If a reduction in the phage titre was demonstrated in restriction proficient cells, then the presence of the POcr mutant did not alleviate the restriction enzyme. As previously mentioned this assay does rely on leaky expression but if the mutants exhibit a 2AP resistance that is equal to 80μg/ml, then these mutants should still maintain full antirestriction activity within this assay.



	wtOcr	pTrc	M2.8	M2.1	M1.1	M1.9
NM1049	1.10×10^{9}	1.00×10^{4}	2.00×10^{8}	2.00×10^{8}	3.00×10^{8}	4.00×10^{8}
NM1261	3.00×10^{8}	2.00×10^{8}	1.00×10^{8}	1.00×10^{8}	1.00×10^{8}	4.00×10^{8}
eop of phage λ	4	0.00005	2	2	3	1
Antirestriction	73333	1	40000	40000	60000	20000

Figure 5.3: The phage titre of the *in vivo* phage infection restriction assay. Restriction proficient (NM1049) and restriction deficient (NM1261) *E. coli* cells were infected with unmodified phage lambda. The number of plaques per ml were calculated from viable plaque counts for each bacterial strain and displayed in the graph above. Below: A table which compares the phage titre and the efficiency of plating of the phage (eop) in the presence of each of the mutants compared to wtOcr and pTrc99A. The antirestriction value is a ratio of the eop in the presence of the mutant/protein compared to the absence of any protein.

Comparisons of the phage titre show that each of the POcr proteins demonstrates full antirestriction activity. There is no difference between the number of plaques present on restriction proficient and deficient cells in the presence of wtOcr and each of the mutants. By contrast, the phage titre is much reduced for restriction proficient cells harboring the empty pTrc99A vector. Therefore each of the POcr proteins successfully inhibits the restriction activity of the EcoKI enzyme. The ratio between the phage titre of restriction proficient and deficient cells is known as the eop (efficiency of plating) of phage and is used to calculate a value for antirestriction. The two mutants that demonstrated reduced activity in the 2AP assay, M2.8 and M1.1 appear to possess the same antirestriction activity as the other POcr proteins from the phage *in vivo* assay. However, this reduction in activity could be a result of reduced antimodification behavior, as witnessed for a number of the Ocr multimutants in chapter 3. Therefore the phage plaques that successfully infected restriction proficient cells in the restriction phage assay above can be used to investigate the antimodification activity of each of the mutants.

The unmodified phage that had successfully infected cells, transformed with each of the POcr mutants, in the antirestriction assay, can be extracted and used to infect restriction-proficient and deficient cells. If a POcr mutant inhibits the activity of M.EcoKI in the initial infection, then the phage DNA would have remained unmodified, and upon infection of restriction proficient cells the phage will be recognised by the restriction enzymes as foreign DNA and destroyed. If the mutants did not inhibit the M.EcoKI enzyme during the initial infection, then the phage is methylated. Upon a secondary infection of this phage into restriction proficient cells, the restriction enzyme will not differentiate between its own DNA and the phage, as both are modified. Thus, the phage will be immune to restriction, allowing successful phage infection and producing a large phage titre.

	wtOcr	**pTrc**	**M2.8**	**M2.1**	**M1.1**	**M1.9**
NM1049	1.40×10^{4}	1.00×10^{8}	9.00×10^{3}	8.00×10^{3}	1.40×10^{4}	5.00×10^{3}
NM1261	3.00×10^{8}	3.00×10^{8}	1.00×10^{8}	1.40×10^{8}	2.00×10^{8}	9.00×10^{7}
eop of phage λ	21429	3	11111	17500	14286	18000
Antimodification	7143	1	3704	5833	4762	6000

TTable 5.3: The calculated values for the phage titre, eop of antimodification activity of the mutants selected from the Library LibDg compared to the antimodification activity of the wtOcr protein and the empty vector pTrc99A.

The phage titre showed a large difference when infecting bacterial cells in the presence and absence of the HsdR subunit. This indicates that the phage has remained unmodified from the initial infection; therefore the POcr mutants have successfully inhibited the M.EcoKI enzyme. This is in contrast to the phage titre retrieved from cells in the presence of the empty pTrc99A vector. Furthermore the POcr proteins M2.8 and M1.1 both exhibit antimodification activity comparable to wtOcr and the other POcr proteins, which suggests that all four of these mutants inhibit the methylase successfully.

5.5 Minimum charge for maximum activity

The phage assays have verified that the four unique POcr mutants isolated from LibDg all induce the same antirestriction and antimodification activity as the wild type protein, even though the number of acidic residues present on these mutants is considerably lower compared to wtOcr.

To identify any commonality between the POcr mutants, their protein sequences were aligned in Figure 5.4 below. Each of the POcr mutants has a similar number of acidic residues ranging from 13-15 acidic residues. In addition eight acidic residues were conserved; E21, D30, D43, D52, E60, D68, D77 and E88. In chapter 3 the analysis of the multimutants showed that the presence of acidic residues at the two loop regions appeared to be particularly important for antirestriction activity, However this accounts for only three of the conserved residues, D30, E60 and D68, on the four POcr mutants, whereas most of the conserved residues are located at sites just outside the loop regions, suggesting that these residues may be more important than the supposed loop regions.

```
wtOcr  MAMSNMTYNNVFDHAYEMLKENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIFSVMASE
1.1    MAMSNMTYNNVFNHAYQMLKENIRYDNIRDTNDLHDAIHMAADNAVPHYYADIFSVMASE
1.9    MAMSNMTYNNVFNHAYQMLKENIRYNDIRDTDDLHNAIHMAADNAVPHYYADIFSVMASE
2.1    MAMSNMTYNNVFNHAYQMLKENIRYNDIRDTNNLHNAIHMAADNAVPHYYADIFSVMASE
2.8    MAMSNMTYNNVFNHAYQMLKENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIFSVMASE

wtOcr  GIDLEFEDSGLMPDTKDVIRILQARIYEQLTIDLWEDAEDLLNEYLEEVEEYEEDEE
1.1    GIDLQFEDSGLMPNTKDVIRILQARIYEQLTIDLWQDAQNLLNQYLQQVQEYQQNQQ
1.9    GINLQFQDSGLMPDTKDVIRILQARIYEQLTIDLWQNAQNLLNQYLQQVQQYQENQQ
2.1    GIDLEFEDSGLMPDTKDVIRILQARIYEQLTIDLWENAQNLLNQYLQQVQEYQQNQQ
2.8    GIDLQFQDSGLMPDTKDVIRILQARIYQQLTINLWENAQNLLNQYLQQVQEYQQNQQ
```

Figure 5.4: Sequence alignment for each of the Ocr mutants. Each of the 34 target sites were either highlighted in red or yellow, red if the residue was acidic and yellow if the residue was the neutral amino acid analogue. The acidic residues have also been highlighted in red on the wtOcr protein, of which all have been targeted on POcr proteins, from the library LibDg.

The residues that were found to be common among all of the proteins accounts for about 60% of the acidic residues. Although the other 40% of residues are not common throughout the sequence of each mutant, they could be located in regions that are similar in each mutant. A number of these regions comprise of clusters of acidic residues. As a result it is possible that the representation of 1 or 2 acidic residues from each of these clusters may contribute to the activity of the protein, rather than individual residues.

5.6 Further analysis of M2.1 mutant.

Table 5.4 outlines the residues that were targeted on M2.1, highlighting whether these residues are already acidic on the mutant M2.1 and if these acidic residues are conserved among all four active POcr mutants selected from the original screen.

Target Site	Group	Acidic on M2.1	Conserved on active mutants	% Mutated in Lib Dg
D26N	I	N	N	40
D27N	I	Y	N	70
D30N	I	Y	Y	80
D32N	I	N	N	50
D33N	I	N	N	40
D36N	I	N	N	30
E60Q	II	Y	Y	45
D63N	II	Y	N	45
E65Q	II	Y	N	10
E67Q	II	Y	N	20
D68N	II	Y	Y	10
E21Q	III	Y	Y	80
D43N	III	Y	Y	50
D52N	III	Y	Y	65
D74N	III	Y	N	60
D77N	III	Y	Y	65

Table 5.4: A table which highlights which mutants will be targeted in each group and therefore each of the three libraries produced. Comparisons have been made to demonstrate the conserved residues from all four active mutants and those acidic residues present on the mutant M2.1. In addition the percentage coverage of acidic residues at each of the positions from a sample population of the library Lib Dg has been recorded.

The table shows that the acidic residues that belong to group3 are almost all conserved on the four POcr proteins and are present in the POcr library LibDg in high abundance. The residues that belong to group 1 and group 2 are located in the loop1 and loop2 regions of wtOcr. Results in chapter 3 highlighted that acidic residues belonging to these loop regions are likely important to the activity of the Ocr. However, only three out of the ten acidic residues that dominate these regions are conserved on the active POcr proteins. Closer inspection of the sequences above implied that both loop regions are highly represented in each of the POcr proteins, however the acidic residues present differed with each protein.

To determine whether; i) the acidic residues present on the loop regions of the POcr protein M2.1 are interchangeable with no effect on activity, ii) M2.1 contains the minimum number of acidic residues to induce wild type antirestriction activity and iii) the conserved residues highlighted on each of the POcr proteins are essential towards the activity of the POcr protein M2.1, three small combinatorial libraries were generated that targeted acidic residues on the POcr mutant M2.1 and substituted them with neutral amino acid species. If the activity diminished or reduced with the absence of any acidic residues that were otherwise

present on M2.1, then it would show that all the acidic residues present on the sequence of M2.1 are essential for its activity.

The multisite mutagenesis kit from Stratagene was used to generate the libraries. A degenerate primer was designed for targeting each of the loop regions as the residues are all in close proximity to each other, whereas a series of primers were designed to target those sites highlighted in group III which are scattered throughout the sequence of the M2.1 gene sequence.

ATGGCTATGTCTAACATGACGTACAATAACGTTTTTAATCACGCATATCAGATGCTGAAAGAGAACAT
TCGCTACAATGACATTCGTGATACCAACAACCTGCACAACGCCATCCACATGGCCGCGGACAACGCAG
TCCCACACTACTACGCGGACATCTTCAGCGTTATGGCTTCTGAGGGTATCGACCTGGAGTTCGAGGAC
TCCGGCCTGATGCCGGACACCAAAGACGTAATCCGTATTCTGCAGGCTCGTATCTACGAGCAGCTGAC
TATCGACCTGTGGGAAAACGCGCAGAACCTGCTGAACCAGTATCTGCAACAGGTGCAGGAATATCAGC
AGAACCAACAG**TAA**

Group	Primer Design
I	CGCTAC**RATRAC**ATTCGT**RAT**ACC**RACRAC**CTGCAC**RAC**GCCATCC
II	GGCTTCT**SAG**GGTATC**RAC**CTG**SAG**TTC**SAGRAC**TCCGGCC
III	GCATATCAGATGCTGAAA**CAG**AACATTCGCTAC
	CCACATGGCCGCG**AAC**AACGCAGTC
	CACTACTACGCG**AAC**ATCTTCAGCG
	GCCTGATGCCG**AAC**ACCAAA**AAC**GTAATCCG

Figure 5.5: Above is the DNA sequence of the mutant M2.1, highlighted in green, pink and blue are the target sites for each of the three libraries. The remaining codons that are highlighted in yellow signify other acidic residues on the sequence of M2.1 which will remain untouched. Below the primers designed to mutate the groups belonging to each of the target sites.

The multisite kit contains three steps, the PCR round anneals the specifically designed primers to the template DNA (M2.1) and generates dsDNA, then DpnI nuclease is used to digest the parental DNA template. Super competent cells are transformed with the DNA mixture producing a library of mutants from either the use of a degenerate primer for groups 1 and 2 or a multitude of primers in group 3. As a result three libraries were generated.

Instead of isolating individual samples from the library, which is very laborious, costly and could throw back many isolates of the original M2.1 template, the successful incorporation of each primer was assessed by sequencing an aliquot of the plasmid library mixture. The electropherogram of the sequence from the subsequent library would remain identical pre

and post multisite mutagenesis, with the exception of the target sites where the signal would be noisy on account of a mixture of codons which correspond to D /N or Q/E accordingly. This was made easier because each pair of residues shared similar DNA codons that differ only by one basepair. If the library has a mixture of the two residues then the two different base pairs should show up as a mixture on the DNA electropherogram at specific target sites, with predictable noisy signals. Although not definitive this method can provide a quick assessment as to the relative importance of acidic residues at each of the target sites.

Amino Acid (Acidic)	DNA codon	Amino Acid (Neutral)	DNA codon
Glu/D	GAA GAG	Gln/N	CAA CAG
Asp/E	GAT GAC	Asn/Q	AAT AAC

Table 5.5: The DNA codons that translate to the acidic and neutral amino acids D/E and N/Q residues, involved in the mutagenesis. The codons show that there is only one bp change between the acidic and neutral amino acid.

5.6.1 Library 1: group 1-loop1

To target each of the residues that are found in Group 1 (aa26-36) which includes the target sites that are located on the Loop 1 region, a degenerate primer was introduced to POcr M2.1, by multisite mutagenesis. The resulting library that was formed was sequenced in figure 5.6 below, which compares the DNA electropherogram from an aliquot of the newly generated library 1 and the original template of the mutant M2.1.

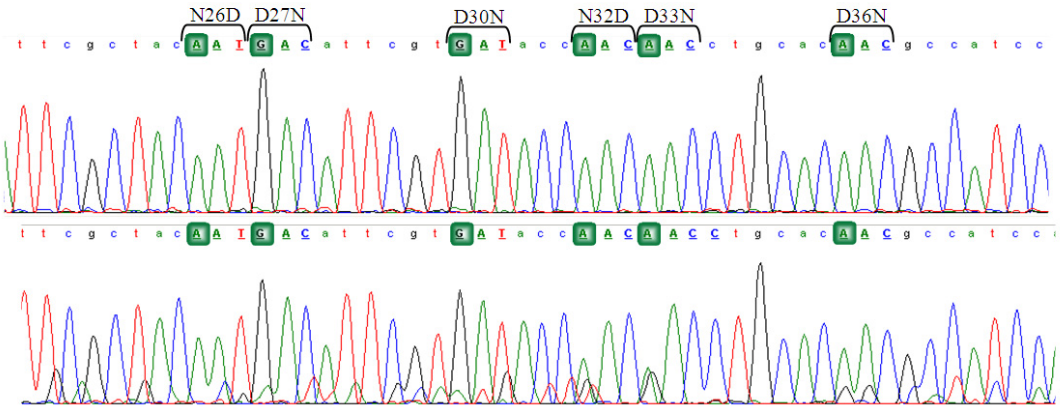


Figure 5.6: Electropherograms showing mutations introduced into the DNA sequence of mutant M2.1, by a degenerate primer. The original template sequence of M2.1 is shown above. Whereas the electropherogram underneath is the aliquot from the generated library. The sites of mutagenesis are labeled and the specific basepairs involved in the mutagenesis are indicated by green boxes.

The incorporation of the primer was successful, the codons that translate to each of the target sites are captalised and labeled accordingly. The individual nucleotides that changed the codon from D/N or E/Q are underlined in Figure 5.6. The signal for each basepair is clearly reduced and in its place a mixture of two signals for two different basepairs is present on the sequence of the DNA electropherogram. The intensity of the signals for either basepair appears to be near equal for positions 32, 33 and 36 given the equal intensity exhibited by each of the basepairs concerned. However the intensity of one basepair appears to dominate for residues 26, 27, and 30, nevertheless an almost equal distribution is assumed because the mutations were introduced using a degenerate primer with equimolar concentrations.

To see if the mutations on loop 1 affect the antirestriction activity of the mutants the library was screened using the 2AP assay developed in chapter 3. The screening process was similar to the screening process carried out for the library LibDg but on a much smaller scale. In this case 6 target sites were designed and one of two residues could be present on each site giving a total number of 2^6, therefore 64 possible sequences.

E. coli NM1041 cells were transformed with the plasmid DNA from group 1 and library mutants were selected from LB agar plates supplemented with 80µg/ml of 2AP. A plasmid preparation was carried out to extract the remaining library that exhibited 2AP resistance, an aliquot was set aside for sequencing and the process was repeated for a second round of selection. The plasmid DNA was extracted from this second selection and sequenced to

identify if there were any changes to the number of the acidic residues that are required to induce the antirestriction activity of M2.1.

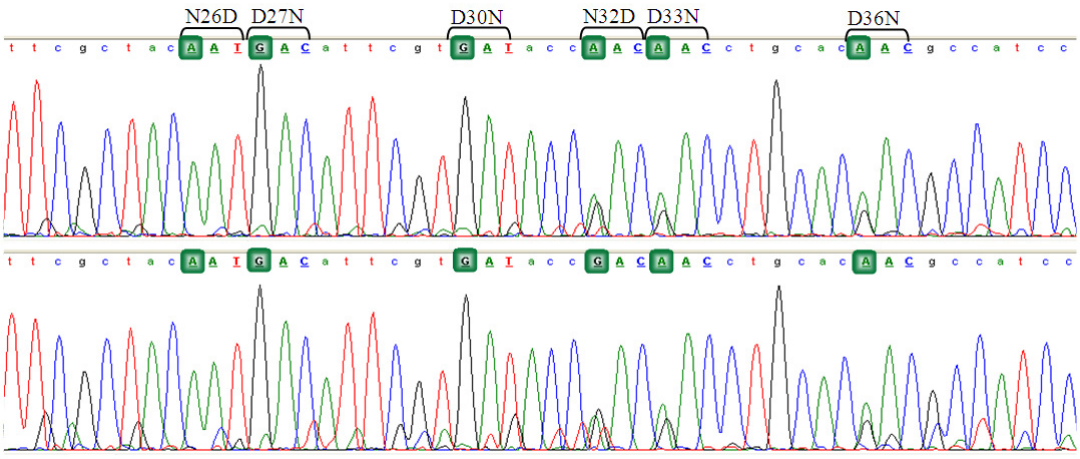


Figure 5.7: Electropherograms showing the ratio of basepairs which correspond to a change in the amino acid residue translate, therefore a ratio of acidic: neutral amino acids upon section of 2APR. The sequence of the colonies that were selected from the first round of screening is shown above and the sequence for the second round of screening is shown below. The sites of interest are labelled and the specific basepairs involved in translating either an acidic or neutral amino acid at this position are indicated by green boxes. Comparison of these electropherograms indicates that none of the target sites have maintained mixtures of both the bp indicating the loss of one residue upon selection.

By sequencing each round of the screening process, the sequences above in Figure 5.7 represent the remainder of the library that appears to be resistant to 2AP and consists of M2.1 mutants which retain the antirestriction activity of M2.1 template.

The sequences show that no change is exhibited for the residues 32, 33 and 36; at each target site mixtures of two bps are present at near equal intensity therefore these mutants contain both acidic and neutral amino acid residues at each of these positions. This implies that these acidic residues are not essential in any of these three positions. On the other hand the target sites 27 and 30 are dominated by a high intensity peak of the basepair guanine which translates to the acidic residue D. This suggests that the acidic residues in these two positions; 27 and 30 are absolutely necessary for M2.1 to function as an active antirestriction protein, whereas the absence or presence of acidic residues at the positions 32, 33 and 36 can be either acidic or neutral. Notable is the slight peak signaling the basepair adenine for the equivalent neutral amino acid at position D27N, comparisons of all the electropherograms identifies the presence of a small signal of the opposing basepair; however it is hard to

distinguish whether this is an artifact or the emission of a real fluorescent signal. This could only be confirmed from extensive sequencing of the library.

5.6.2 Library 2: group 2-loop2

The entire loop2 region was mutated to acidic amino acids in the POcr mutant M2.1. Due to the close proximity of these target sites, a degenerate primer was designed to introduce neutral residues onto M2.1.

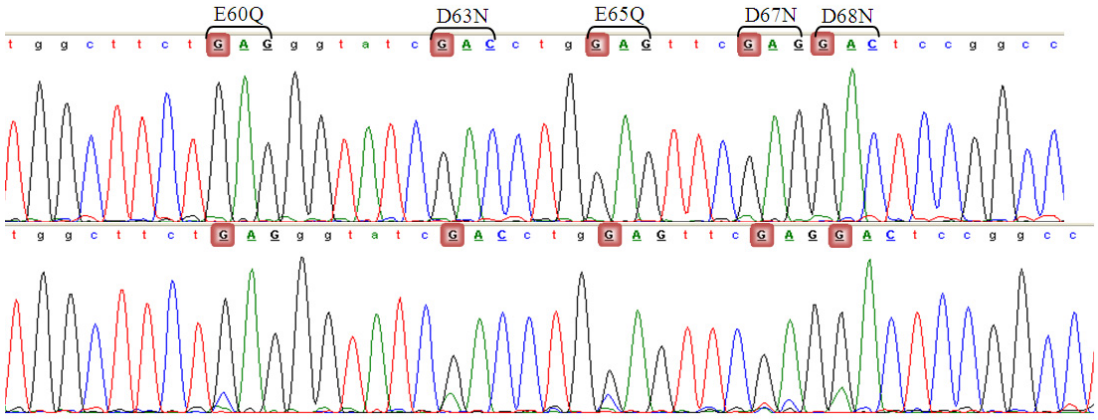


Figure 5.8: Electropherograms showing mutations introduced into the DNA sequence of mutant M2.1, by a degenerate primer. The original template sequence of M2.1 is shown above. Whereas the electropherogram underneath is the aliquot from the generated library. The sites of mutagenesis are labelled and the specific base pairs involved in the mutagenesis are indicated by pink boxes.

The DNA sequence of the resulting library was compared to the original M2.1 template in figure 5.8. Each of the target sites were originally acidic residues however, mixtures of two basepair signals were clearly displayed for the codons targeting the residues E63, E65 and D68. Although the signals expressed for the target sites E60 and D67 were dominated by the parental codon each of the target sites were present on a single degenerate primer, therefore all of the target sites were assumed to be successfully targeted.

Library 2 targets 5 sites with the possibility of two different residues and so there is the possibility of 32 different samples and therefore over 320 sequences were screened for 2AP resistance. The colonies from the screen were collated and rescreened as previously seen with library 1. The plasmid DNA from the collated colonies of screen 1 and 2 were extracted and sequenced, to identify whether some acidic residues were more selectively advantageous then others for antirestriction activity.

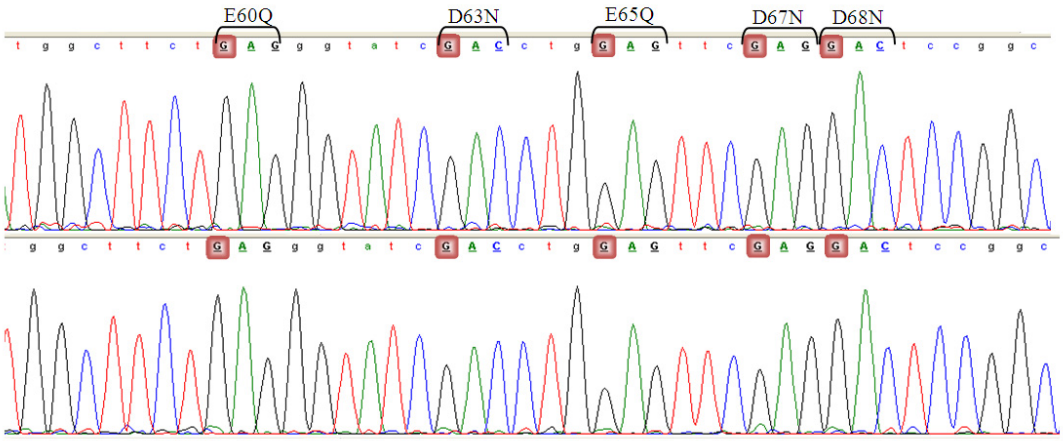


Figure 5.9: Electropherograms showing the ratio of base pairs which correspond to a change in the amino acid residue translate, therefore a ratio of acidic: neutral amino acids upon section of 2APR. The sequence of the colonies that were selected from the first round of screening is shown above and the sequence for the second round of screening is shown below. The sites of interest are labelled and the specific base pairs involved in translating either a acidic or neutral bp at this position are indicated by pink boxes. Comparison of these electropherograms indicates that none of the target sites have maintained mixtures of both the bp indicating the loss of one residue upon selection.

The analysis of the sequences seems to suggest that all of the acidic residues that were present on the mutant M2.1 are essential for antirestriction activity. However, a closer look at the sequence from the second screen does show a small signal for the opposing adenine basepair for the residues D67 and possibly E60. The sequence of library 2 does appear to have varied signals for each basepair but the mutagenesis was carried out with a degenerate primer. This makes it hard to decide whether residues such as D67 are in fact present as a mixture of D67 and N67. The high intensity peaks exhibited for each of the codons that translate the acidic amino acid residues at each of these positions seems to dominate, therefore the mutant M2.1 seems to require all of the acidic residues in Loop 2 to function as a fully active antirestriction protein.

5.6.3 Library 3: multiple target library

The target sites for this library are present across the whole DNA/protein sequence of the mutant M2.1; therefore the use of one degenerate primer would not be adequate. The residues that were targeted included E21, D43, D52, D74 and D77, these residues reside on helices A, B, and C, and on loop2 of the Ocr crystal structure. Therefore to produce this library separate oligonucleotides were designed for each target site except for the residues D74 and D77, where a primer with a double mutant was designed. The residues that were

targeted for mutagenesis in this library were acidic residues that were conserved in all four of the active mutants selected from Lib Dg.

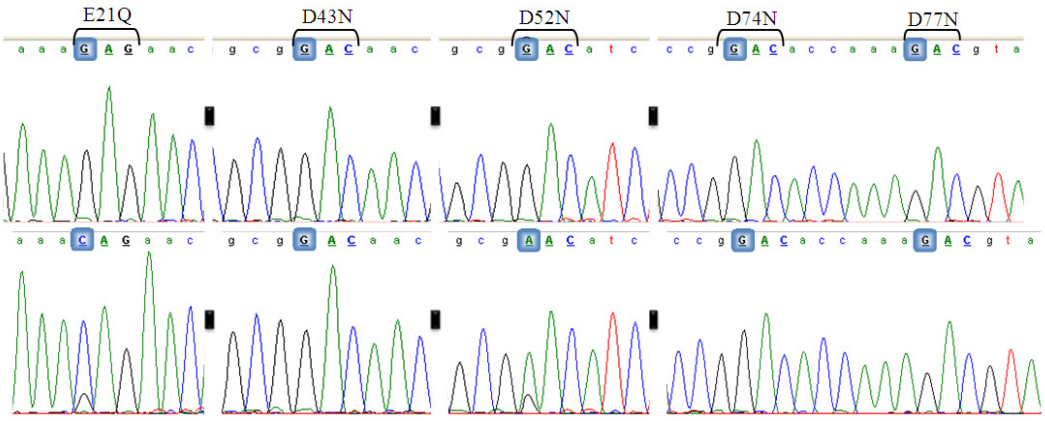


Figure 5.10: Electropherograms showing mutations introduced into the DNA seqyuence of mutant M2.1, by a number of oligonucleotides. The original template sequence of M2.1 is shown above. Whereas the electropherogram underneath is the aliquot from the generated library. The sites of mutagenesis are labelled and the specific base pairs involved in the mutagenesis are indicated by blue boxes.

The results showed that only two of the target sites obtained mixtures of both acidic and neutral codons. The target sites E21Q and D52N have in fact been superseded by a larger proportion of neutral residues when compared to the original codon. Mixtures of acidic and neutral amino acid residues were not successfully introduced to the target sites D43N, D74N and D77N. Although there is the possibility of a signal for D43N, the mutagenesis appeared to be unsuccessful for D74N and D77N. Nevertheless the screening process was carried out for this library and the results are shown in figure 5.11.

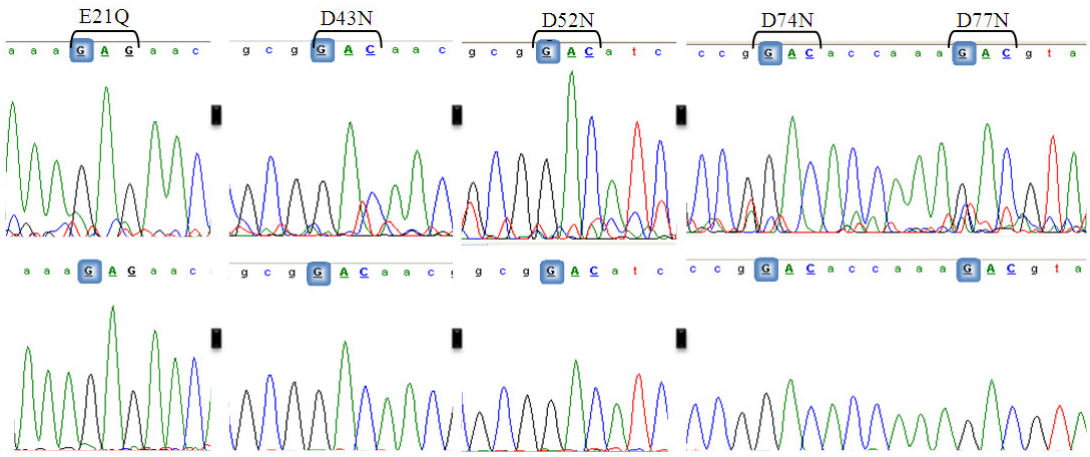


Figure 5.11: Electropherograms showing the ratio of base pairs which correspond to a change in the amino acid residue translate, therefore a ratio of acidic: neutral amino acids upon section of $2AP^R$. The sequence of the colonies that were selected from the first round of screening is shown above and the sequence for the second round of screening is shown below. The sites of interest are labelled and the specific base pairs involved in translating either an acidic or neutral bp at this position are indicated by pink boxes. Comparison of these electropherograms indicates that not of the target sites have maintained mixtures of both the bp indicating the loss of one residue upon selection.

The sequence from the first screen is much noisier than the sequence from the second screen. However both clearly show that only codons that translate acidic amino acids are present upon selection of 2AP resistance. This confirms that the each of these residues is essential to the activity of mutant M2.1.

In summary, it appears that all of the acidic residues that are present on the mutant M2.1 appear to be essential to the antirestriction activity of that POcr mutant. This confirms that for POcr M2.1, the minimum number of acidic amino acid residues needed to retain full antirestriction activity has been obtained. Furthermore, if this is the case for all four active mutants then the minimum number of acidic residues that are essential for antirestriction activity differs depending on the unique sequence of the POcr protein.

5.7 Discussion

5.7.1 Screening

The largest library, LibDg, was screened using a plate-based assay which selects for $2AP^R$ at a concentration of 80µg/ml. LibDg was initially screened not only because it contained the greatest mutation rate of all the libraries but it also successfully targeted each of the mutation

sites, whereas two of the oligonucleotides used in library 1D and 2D failed to anneal successfully.

Batch	Mutants screened from library	$2AP^R$ colonies	Growth upon restreaking	Successfully retransformed
1	400000	73		
2	400000	29	38	7
3	400000	55	35	0
Total	**1200000**	**157**	**73**	**7**

Table 5.6: A table that shows the screening process of each of the mutants, the number of mutants that were screened from each batch, those that were initially screened with $2AP^R$, those mutants that grew once plated onto fresh 2AP exposed LB-agar plates and the number of these that successfully retransformed with $2AP^R$ in the same *E. coli* NM1041 cells.

The screening was carried out in three batches. In total, 10^6 clones were screened. Each screen produced a large number of initial hits from the library. However, upon replating the number of positive hits was much reduced as many of the colonies did not grow on fresh 2AP plates. A rather liberal estimate of homogeneous cell growth was considered and carried forward for retransformation. This panning process resulted in a total of 7 mutants that maintained $2AP^R$. From these seven clones, four unique sequences were found.

5.7.2 Active POcr mutants

Four unique sequences were selected from the library LibDg and two of these mutants, M2.1 and M1.9, displayed a resistance to 2AP that exceeded 80μg/ml. Like the Ocr protein both of these mutants displayed a high degree of antirestriction and antimodification activity upon phage infection comparable to wtOcr. On the other hand, the POcr mutants M2.8 and M1.1 were also selected from the library exhibiting a slightly reduced resistance. However, these two POcr proteins also exhibited high levels of antirestriction and antimodification activity upon phage DNA again comparable to the activity of wtOcr. Each of the sequences contained 13-15 acidic residues of which seven are conserved. This results in a loss of 60% of the acidic residues that are present on wtOcr. This suggests that ~40% of the acidic residues on the protein wtOcr are nonessential for the activity of the protein *in vivo*. The seven conserved residues are distributed throughout the whole of the POcr sequence and correspond to acidic residues that are located across the surface of Ocr. Most notably only 3 of the conserved residues D30, D60 and D68 are present in either of the two loop regions that were deemed important in chapter 3. The alignment of the active POcr proteins below is complemented with a table that shows the number of acidic residues that were introduced to

each POcr mutant and their location on the sequence and the structure of the protein. The actual number of acidic residues has been recorded as well as the percentage number of acidic residues, to gauge the distribution of the acidic residues across the structure of Ocr.

```
wtOcr MAMSNMTYNNVFDHAYEMLKENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIFSVMASE
M2.1  MAMSNMTYNNVFNHAYQMLKENIRYNDIRDTNNLHNAIHMAADNAVPHYYADIFSVMASE
M1.1  MAMSNMTYNNVFNHAYQMLKENIRYDNIRDTNDLHDAIHMAADNAVPHYYADIFSVMASE
M1.9  MAMSNMTYNNVFNHAYQMLKENIRYNDIRDTDDLHNAIHMAADNAVPHYYADIFSVMASE
M2.8  MAMSNMTYNNVFNHAYQMLKENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIFSVMASE
      ************:***:********::****::**:************************

wtOcr GIDLEFEDSGLMPDTKDVIRILQARIYEQLTIDLWEDAEDLLNEYLEEVEEYEEDEE
M2.1  GIDLEFEDSGLMPDTKDVIRILQARIYEQLTIDLWENAQNLLNQYLQQVQEYQQNQQ
M1.1  GIDLQFEDSGLMPNTKDVIRILQARIYEQLTIDLWQDAQNLLNQYLQQVQEYQQNQQ
M1.9  GINLQFQDSGLMPDTKDVIRILQARIYEQLTIDLWQNAQNLLNQYLQQVQQYQENQQ
M2.8  GIDLQFQDSGLMPDTKDVIRILQARIYEQLTINLWENAQNLLNQYLQQVQEYQQNQQ
      **:*:*:******:*****************:**::*::***:**::*::*::::::
```

Mutant	Nº of Mutations (34)	Helix A (3)		Helix B (2)		Helix C (1)		Helix D (7)		Loop 1 (5)		Loop 2 (7)		Tail (9)	
M2.1	**14**	**1**	33%	**1**	50%	**1**	100%	**3**	43%	**2**	40%	**7**	100%	0	0%
M1.1	**15**	**1**	33%	**2**	100%	**1**	100%	**3**	43%	**3**	60%	**5**	71%	1	11%
M1.9	**13**	**1**	33%	**1**	50%	**1**	100%	**2**	29%	**4**	80%	**4**	57%	1	11%
M2.8	**15**	**1**	33%	**2**	100%	**1**	100%	**2**	29%	**5**	100%	**5**	71%	1	11%

Table 5.7: Above is an alignment of the sequences for each of the active POcr mutants that demonstrate wtOcr functionality. Comparisons in the number and position of acidic residues (highlighted in red) on each POcr protein were compared to each other and to wtOcr. The four different domains present on the Ocr monomer were highlighted in colour on the wtOcr protein sequence. Below the table looks at the distribution of acidic residues across each of the five domains on the Ocr monomer structure. In brackets is the possible number of acidic residues that could be targeted within each domain.

The table 5.7 shows that on Helix A, the acidic residue E21 is conserved. In addition the acidic residue D52 located on Helix C is also conserved. Both acidic residues are located in areas of the protein isolated from the cluster regions and it suggests that these residues may provide a structural rather than a functional role.

However, the majority of acidic residues are located on Helix D, Loop1 and Loop2. At least 57% of each loop region is represented in each of the mutants. Generally if poor representation is found in one of the loop regions, it is compensated by a high representation in the other loop region. These mutants appear to mutually compensate for any lack of acidic residues found in the other. So even though the number of conserved acidic residues found among the two loop regions is low, the overall representation of each loop is high and so there is a disproportionate number of acidic residues that have been reintroduced into the POcr scaffold in order to generate an active antirestriction protein.

5.7.3 Comparisons between the active POcr mutants and Ocr homologues

Alignment studies were carried out with a BLAST search of Ocr homologues found on other phage entities. The BLAST searched showed a similar pattern of acidic residues on each of the different proteins. There were eleven amino acids that were conserved completely on each of the proteins and an additional five residues that possessed conserved acidic residues (E or D). The specific residues these refer to are; E21, D26, D32, D33, D36, D65, D68, D77, E88, D93, E103 and E106 which are conserved and the following amino acids which display a negative charge are; D52, E60, E67, D74 and E108. A comparison with the conserved residues of the active POcr proteins selected from LibDg showed that the residues E21, D68, D77 and E88 were commonly conserved across the mutants and the Ocr proteins, suggesting that each of these residues are important to its function.

Residue	*M2.1*		*M1.1*		*M1.9*		*M2.8*	
E21	Y		Y		Y		Y	
D26	N		Y		N		Y	
D32	N		N		Y		Y	
D33	N		Y		Y		Y	
D36	N		Y		N		Y	
E65	Y		N		N		N	
D68	Y		Y		Y		Y	
D77	Y		Y		Y		Y	
E88	Y		Y		Y		Y	
D93	Y		Y		Y		N	
E103	N		N		N		N	
E106	Y		N		N		Y	
Total	**77**	58%	**88**	67%	**77**	58%	**99**	75%

Table 5.8: Residues that were commonly conserved on each of the Ocr proteins from the different phage entities and are also present on the selected mutants of the POcr library LibD.The last row highlights the total number/percentage of commonly conserved acidic residues found on each POcr mutant that are also conserved on Ocr proteins from different phage entities.

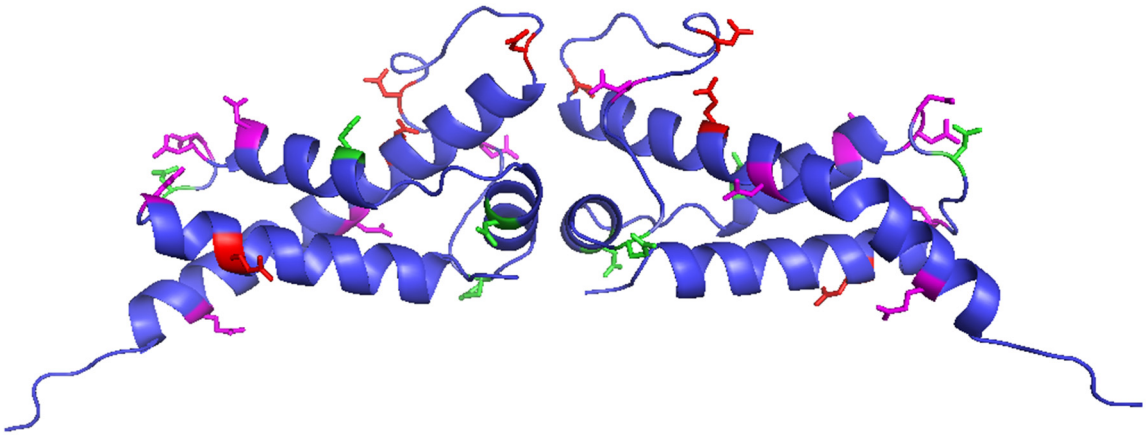


Figure 5.12: The structure of Ocr from phage T7, highlighted in red are residues that are conserved on the Ocr homologues and the active POcr mutants, in magenta are residues that are acidic (either D or E) on Ocr protein and the active POcr proteins. Finally, in green are residues conserved on the active POcr mutants but not on Ocr homologues.

5.7.4 Maximum activity with minimum charge

The results from the mutagenesis of the mutant M2.1 showed that further mutation of M2.1 to reduce the number of acidic residues knocked out any antirestriction or antimethylation activity that the mutant possessed. This seems to stress the importance of each acidic residue that is present on mutant M2.1. The only exception appears to be E21Q, the electropherogram shows that a small signal is present for the adenine base alongside the large signal of the Guanine base that encodes the acidic residue. The total number and position of the acidic residues on M2.1 appears to be finely balanced to provide antirestriction activity.

In summary, the 2AP screening assay developed in chapter 3 was used to select for POcr mutants that display antirestriction properties from the plasmid POcr library, LibDg generated in chapter 4. One million POcr mutants were screened from LibDg using 2AP and four unique mutants were isolated that possessed a high level of antirestriction and antimodification activity. Sequence alignment showed that there were a number of conserved residues among the four mutants. The targeting of these acidic residues on the mutant M2.1 showed that each one was important for its antirestriction activity; the substitution of even one of these residues could knockout the antirestriction activity of the mutant. As a result this also shows that each of the mutants are likely to exhibit the absolute minimum number of residues to allow for antirestriction activity and that the position of the residues and their combinations are equally important.

Overall the conserved residues appear to be distributed across the whole of the POcr protein. The loop regions were previously deemed important for the interaction of Ocr with EcoKI in chapter 3. The results in this chapter highlight this, by selecting for POcr proteins with a large number of acidic residues within the loop regions, but these were not conserved. Alignment of the conserved residues with the phosphate backbone of dsDNA indicates that many of the conserved acidic residues overlap with the phosphate groups on DNA. However, there are a number of acidic residues that don't overlap with the phosphate backbone and this suggests a different reason for their conservation, for instance these residues may be important for structure/stability. The sequences of the POcr proteins showed that the minimum number of acidic amino acids needed for POcr to be fully functional *in vivo* is a total of 13. This is a reduction of over 60% of the acidic residues that are present on the wtOcr protein. These findings indicate that many of the acidic residues of wtOcr appear to play no direct role in the activity of the protein. The reason why Ocr contains such an

overabundance of acidic residues that have apparently been conserved through the evolutionary process, is not immediately apparent.

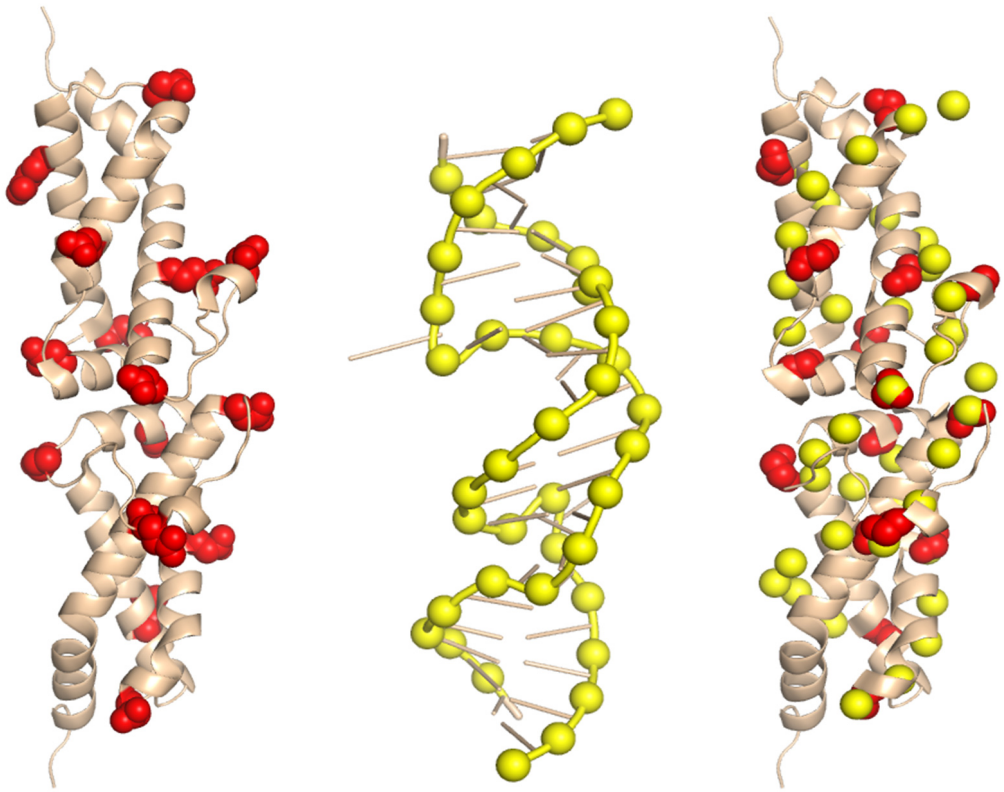


Figure 5.13: Alignment of conserved acidic residues on the Ocr homodimer with DNA phosphate backbone of dsDNA. Left: A cartoon representation of the Ocr homodimer, highlighted in red spheres are the acidic residues that are conserved across each of the four mutants. Middle: The structure of dsDNA as it binds to EcoKI, taken from the M.EcoKI model (2yc2) yellow spheres indicate phosphate groups on the structure. Right: Alignment of the Ocr homodimer and the dsDNA structures to align the conserved acidic residues with the phosphate groups.[58]

Chapter 6. *In vitro* study of the active POcr mutants

6.1 Introduction

The synthetic gene POcr encodes a protein that is a multiple mutant of the antirestriction protein wtOcr in which all the acidic residues have been substituted for neutral amino acids i.e., aspartic acid for asparagine and glutamic acid for glutamine. The POcr gene is not easily expressed in *E. coli* and the resulting protein possesses no antirestriction or antimodification activity. Therefore, in order to ascertain the number of acidic residues that are needed to produce a fully functional antirestriction protein, the POcr gene was used as a template to create a series of libraries, where acidic residues were slowly reintroduced onto the POcr gene.

The largest library, LibDg, was screened using a 2AP assay to obtain POcr mutants that exhibited full antirestriction behavior. Out of the one million POcr mutants that were screened only four unique POcr proteins have been selected that regained full antirestriction and antimethylation properties. Sequencing results showed that the four active mutants possessed less than 60% of the acidic residues that reside on wtOcr. This suggests that over half of the acidic residues on wtOcr are redundant in its functionality as an antirestriction protein.

Why does wtOcr have an excessive number of acidic residues that appear to impart no apparent selective advantage? Such a large number of acidic residues cannot be favorable to the stability of such a small protein.

6.2 Aims

The four active POcr proteins selected from the library LibDg were purified for a series of *in vitro* investigations to probe their structure, stability and functionality. These investigations were carried out alongside a purified sample of wtOcr protein kindly donated by Mr. Laurie Cooper.

6.3 Protein expression

Before "large scale" protein expression was carried out, a small expression test in *E. coli* BL21 (DE3) cells (10ml of LB medium) was used to confirm the expression of soluble protein. The resulting SDS PAGE gel (figure 6.1) displays the soluble cell lysate generated after induction of protein expression with IPTG.

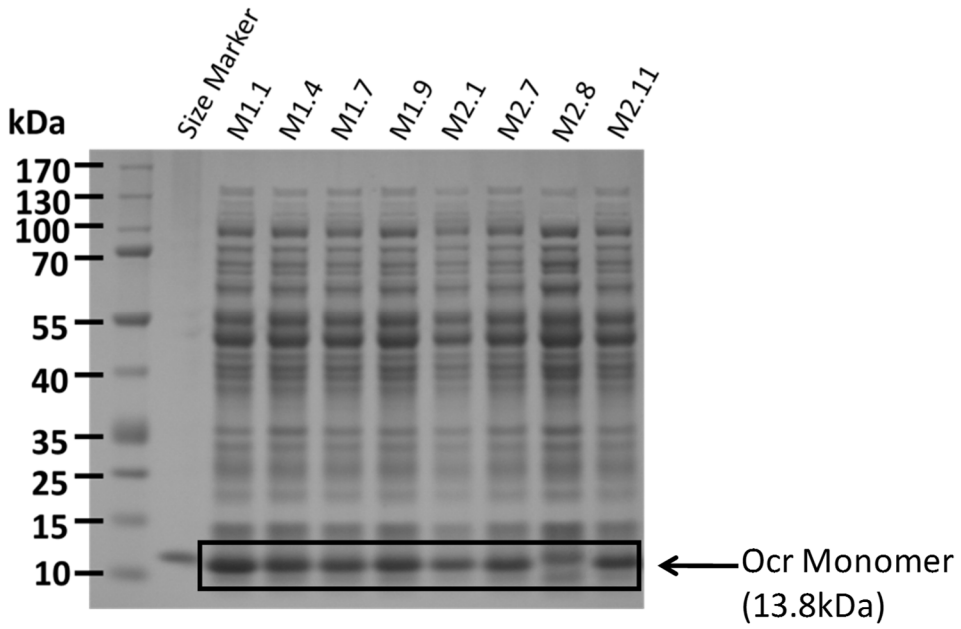


Figure 6.1: An SDS-PAGE gel of the soluble extract of *E. coli* BL21 (DE3) cells which produce each of the mutants. The size marker is a sample of previously purified wtOcr protein (kind gift from Mr. Laurie Cooper). Except for M2.8, soluble protein that co-migrates with the wtOcr was detected for each of the mutants.

SDS-PAGE analysis confirms, with the possible exception of M2.8, successful expression of each POcr protein in a soluble form.

6.4 Protein purification

The protein expression was repeated at a larger scale (~3L of LB medium) producing a cell pellet of 10-15g. The purification protocol consists of two steps; a DEAE (diethylaminoethyl)–Sepharose fast flow anion exchange column, followed by TCA precipitation of the protein solution to remove contaminating nucleic acids and other residual protein contaminants.

Initially, the cell pellets in ice cold buffer A (20mM Tris, 300mM NH_4Cl, pH8) were lysed by sonication and the clarified cell-free extract was loaded onto a DEAE ion exchange column. The column was thoroughly washed with buffer A and bound proteins were subsequently eluted on a salt gradient (0.3-1.0M NH_4Cl in buffer A). A typical UV 280nm transmission profile is shown below. The protein of interest was located by selecting fractions from across the elution profile to be analysed by SDS-PAGE gel electrophoresis. As anticipated, each of the mutant proteins eluted earlier in the salt gradient than wtOcr.

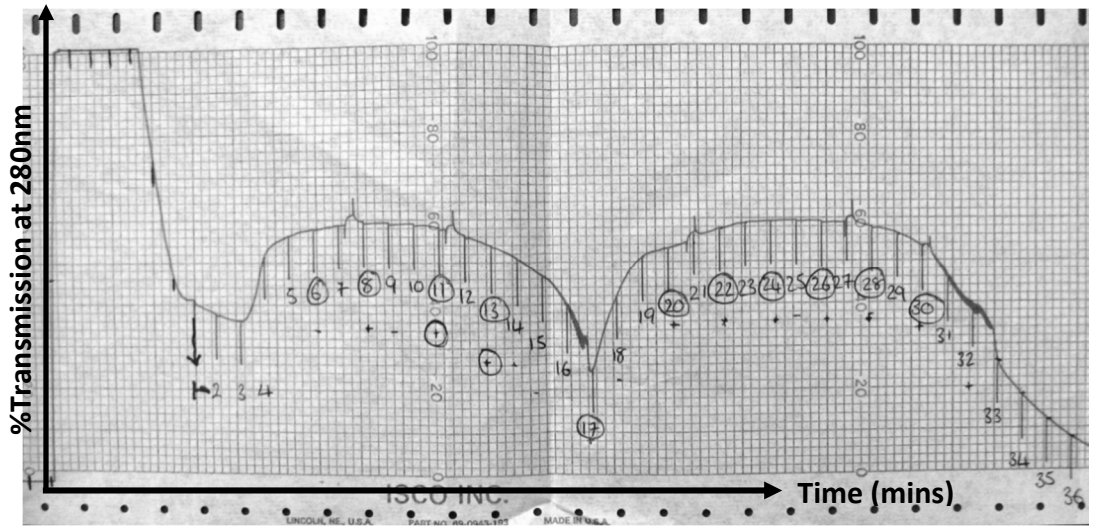


Figure 6.2: Elution profile from a DEAE column with a salt gradient for the purification of Ocr. The 5 ml fraction numbers are indicated on the x-axis. The y-axis shows the transmission % at 280 nm of the eluate.

Contaminating DNA was removed by treating the sample with trichloroacetic acid (TCA). The POcr proteins were successfully purified however the proteins failed to resuspend in buffer B (10mM Tris-HCl, pH8.0). The protein samples were subsquently dialysed with ~4 L of buffer B for 16hrs at 4°C. This resuspended the proteins and removed any traces of TCA. The samples were subsequently concentrated and each POcr protein was stored at -20°C in the presence of 50% glycerol (v/v).

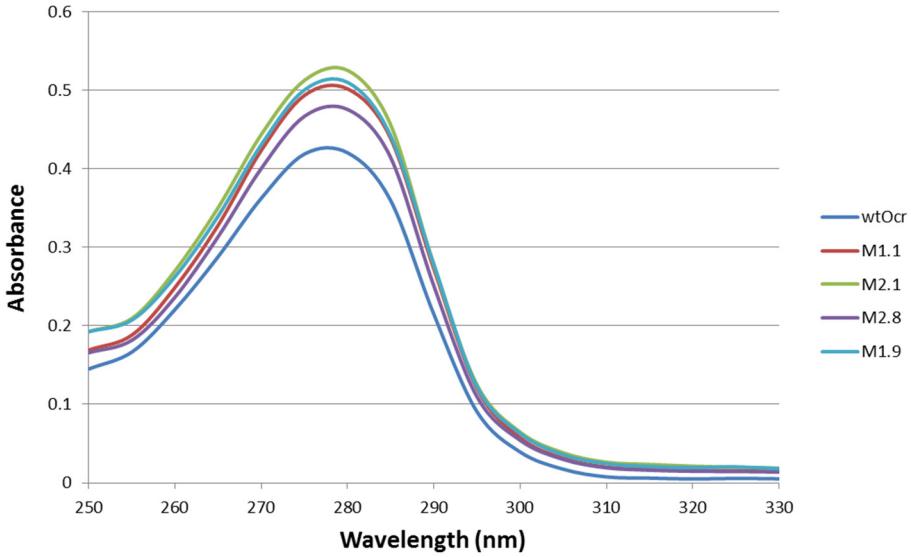


Figure 6.3: A UV A_{280} trace of the purified active mutants that were selected by the 2AP assay and wtOcr that were used for the preceding experiments. Each trace is a diluted solution of the protein.

From unique protein sequences were successfully selected from the library LibDg therefore, the mutant proteins that were successfully purified were M2.1, M1.9, M1.1 and M2.8.

6.5 Protein structure

The secondary structure of each POcr protein was examined by crosslinking studies and circular dichroism (CD) analysis.

6.5.1 Dimerisation of the mutant

Crosslinking studies were carried out to investigate whether the active POcr proteins exist as a homodimer in solution, like wtOcr. The crosslinking agent used was glutaraldehyde, which is a commonly used homobifunctional reagent that reacts with the ε-amino groups of lysine residues. The reaction proceeds *via* a poorly stable Schiff base, which upon reduction with $NaBH_4$ gives a hydrazone linkage. Hence, exposure of a protein to glutaraldehyde will cause covalent crosslinks between closely spaced lysines that can then be detected by SDS-PAGE analysis. The mutant proteins and wtOcr were exchanged into 10mM HEPES buffer (pH 7.0) because the Tris-HCl buffer used to store the proteins would quench the glutaraldehyde reaction. The reaction between each POcr protein (25µg in 25ml) and 1% glutaraldehyde was quenched with $NaBH_4$ and the proteins concentrated by TCA precipitation (full protocol

in chapter 2). If the POcr proteins do exist as homodimers then analysis by an SDS-PAGE gel should show the existence of the homodimer, as well as the monomer for each of proteins.

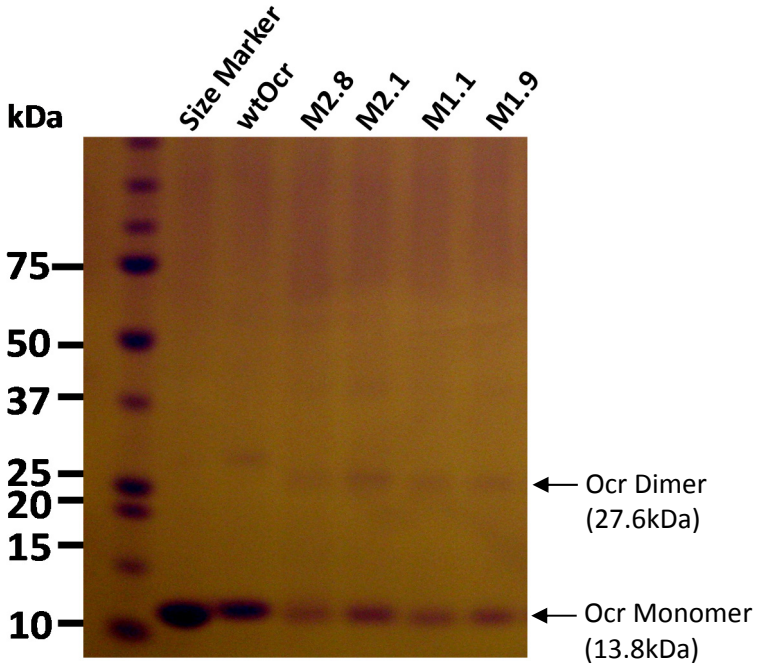


Figure 6.4: SDS PAGE gel of the glutaraldehyde cross linked products. There are two species present in each lane, corresponding to Ocr monomer at 13kDa and another species that appears to be approximately double the size i.e. just above 26kDa.

The SDS-PAGE gel in figure 6.4, shows that each of the mutants exist as two species, the first and most distinct band is between 10-15kDa and is indicative of the Ocr monomer protein. This is confirmed by the size marker, which is a sample of the Ocr monomer. However a second species exists that appears to be double the monomer size suggesting the presence of a cross linked homodimer. This higher molecular weight species is very faint but is present for each of the mutant proteins and wtOcr. Notably, the wtOcr cross linked homodimer appears to run slower than the homodimer of the mutant proteins, the reason for this observed difference in migration speed is thought to be due to the difference in charge between the mutant proteins and wtOcr.

6.6 Secondary structure

Although the mutant proteins exist as homodimers in solution, the loss of so many acidic residues could affect the local secondary structure and the fold of the protein compared to wtOcr. A common technique used to examine conformational changes between proteins and their mutants is circular dichroism (CD). CD analysis can investigate conformational changes to the secondary structure of the whole protein, by measuring the differential absorption patterns of left and right circularly polarised light. Protein secondary structure (i.e. α-helices, β-sheets and random coils) gives a distinctive CD spectrum in the far UV region.

Each of the proteins was exchanged into a buffer containing 10mM sodium phosphate, 10mM sodium fluoride (pH8.0). The concentration of each protein was then determined accurately by UV spectrophotometry by carrying out an absorption scan (210-340nm) on each sample at a concentration of 14.5µM +/-0.5µM.

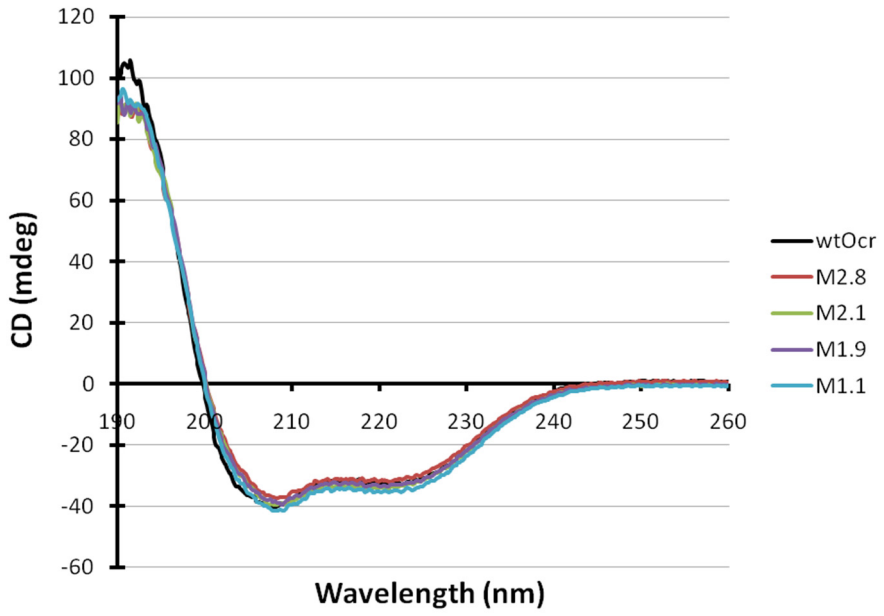


Figure 6.5: The Far UV CD spectra of the active mutants selected from LibDg compared to the spectrum of the wtOcr protein.

The spectra obtained for each of the POcr proteins were very similar to each other and wtOcr, see figure 6.5. This suggests that there were only very slight changes, if any to the secondary structure of the POcr proteins compared to wtOcr.

6.7 Protein stability

The removal of more than half of the acidic residues appears to have no effect on protein fold. Therefore the focus moved to investigating the stability of the POcr proteins. Protein stability was examined by thermal and chemical stability studies.

6.7.1 Thermal stability assay (TDA)

Thermal stability of the protein was measured by a thermal denaturation assay (TDA). This assay measures the transition melting temperature (T_m) of the protein of interest. The thermal denaturation assay measures the fluorescence emitted by the dye SYPRO orange, which is an environmentally sensitive dye. This dye is quenched in an aqueous environment. However, denaturation of the protein exposes the proteins hydrophobic core which can then bind to the dye and cause an increase in the fluorescence signal. The assay is carried out using a thermal cycler measuring the emission of the SYPRO orange dye (emλ: 575nm) as a function of temperature. TDA is greatly affected by the buffer. Therefore all the experiments were conducted using the same buffer system (10mM HEPES, pH8.0).

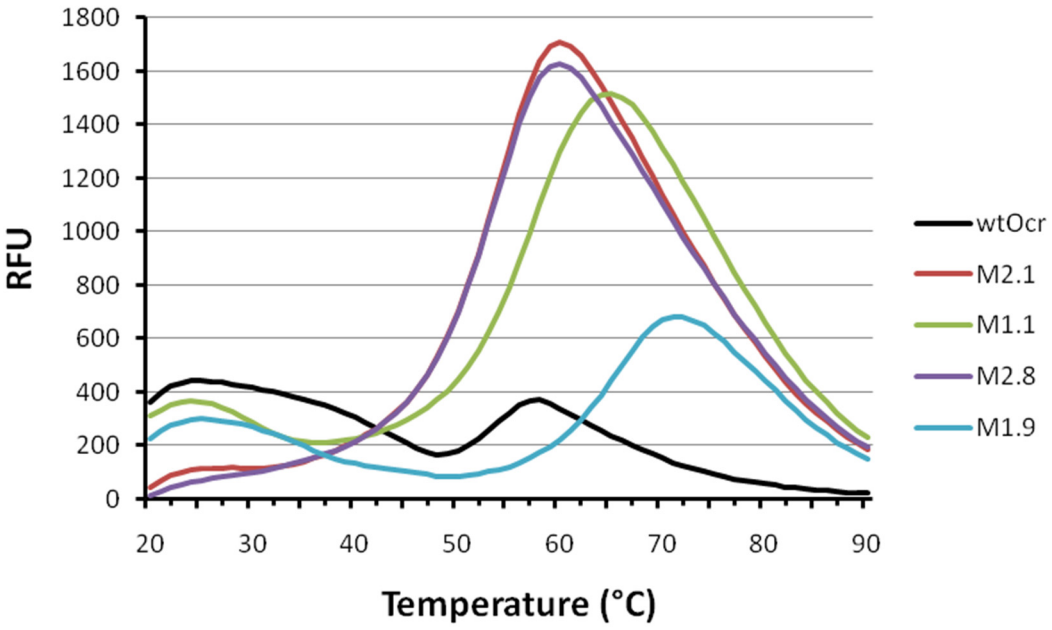


Figure 6.6: The peaks that result from the emission of fluorescence from the SYPRO orange dye as a function of temperature. The intensity of the emission peak is thought to be dependent on the concentration of the protein, which appears to be relatively reduced for wtOcr and M1.9. The peaks are similar for proteins wtOcr, M2.1 and M2.8 whereas the peaks for proteins M1.9 and M1.1 have shifted to a higher temperature.

The fluorescent signal produces an emission signal corresponding to the unfolding transition state of the protein. The transition state of wtOcr, M2.1 and M2.8 is simlar with a T_m of 54°C. However the transition state has shifted to higher temperatures for the mutants M1.1 and

M1.9. M1.1 has a T_m temperature of 57°C and M1.9 a T_m of 67°C, figure 6.6. The increase in temperature suggests that more heat energy is needed to denature the mutants M1.1 and M1.9, hence these proteins are more stable. The mutants M2.1 and M2.8 appear to possess the same thermal stability as wtOcr which suggests that the charge state is not the only factor which could affect the stability of the active POcr proteins.

6.7.2 Chemical stability

The salt, guanidine hydrochloride (GdmCl) can act as a chemical denaturant of peptides and proteins. Most proteins are completely denatured in the presence of 6M GdmCl. However, there is still a debate as to how GdmCl breaks down the secondary structure of proteins. The unfolding process of the denaturant is usually measured either by changes in the intrinsic fluorescence of the protein or changes in an extrinsic fluorescent probe. The aromatic residue with the greatest intrinsic fluorescence is the amino acid residue tryptophan. Tryptophan (Trp) has an excitation wavelength of 280nm and an emission wavelength range of 300-350nm. The intensity of the fluorescence emitted from Trp residues depends on the number of Trp residues and the local microenvironment in which each Trp residue resides. The fluorescence of Trp is quenched in hydrophobic environments. Hence the residue is an ideal candidate for monitoring the unfolding of proteins. The unfolding process of the active POcr mutants, in the presence of increasing concentrations of GdnHCl, was monitored by measuring the intrinsic fluorescence of the only Trp residue on the Ocr monomer (W94). W94 is located on Helix D of the Ocr protein, which is towards the C-terminal end of the protein.

Each of the POcr proteins and wtOcr (3µm) were incubated overnight at 4°C with increasing concentrations of GdmHCl, from 0-6M (see chapter 2 for details). Trp florescence was measured using an excitation wavelength of 295nm and emission from 300-450nm.

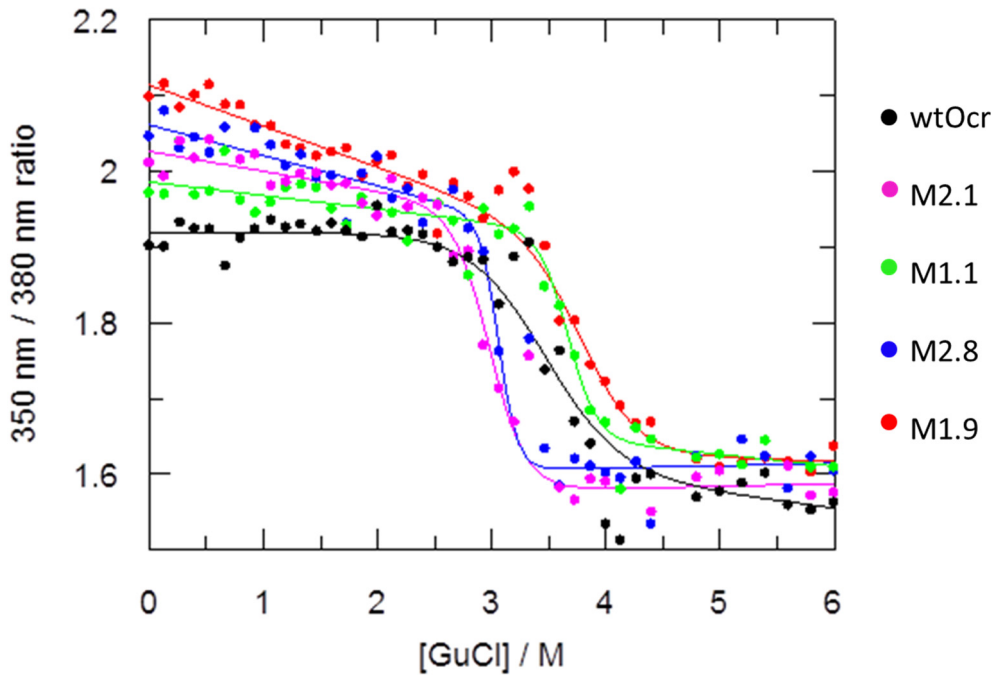


Figure 6.7: Denaturation curves induced by increasing the concentration of the denaturant GdmHCl for wtOcr and each of the mutants. Measurements were taken at an excitation λ=295nm and an emission λ340-380nm was measured for changes in the Trp fluorescence. All data points were fitted using a two state model (Equation 1) which calculated the midpoint value (D50%) and the ΔG of unfolding.

Denaturation curves were plotted using the emission ratio of 350:380nm, this took into account any discrepancies in protein concentration, figure 6.7. The denaturation curves were fitted to a two-state model using Equation 1, a linear relationship was assumed between the free energy of unfolding and the concentration of GdmCl.

$$F = \frac{(\alpha_F+\beta_F[D])+(\alpha_U+\beta_U[D])\exp[(m[D]-[D]_{50\%})/RT]}{1+\exp[(m[D]-[D]_{50\%})/RT]}$$

Equation 1

F – fluorescence at a given denaturant concentration

D – molar concentration of denaturant

R – gas constant = $8.314 Jmol^{-1}$

T – Temperature = 293 K

Realistic estimates were used for the parameters below for each POcr mutant and wtOcr

α_F - native signal

α_U - denatured signal

β_F - native slope

β_U - denatured slope

The fitted data was used to obtain the midpoint value ($D_{50\%}$) and the transition slope (m) Furthermore the free energy of unfolding at 0M GdmCl was calculated using Equation 2, see table 6.1 below.

$$\Delta G_{unfolding} = m\, D_{50\%}$$

Equation 2

Pocr Mutant	Unfolding midpoint (M)	Transition slope (m)	Free energy of unfolding (ΔG in $kJmol^{-1}$)
M2.1	3	15.2	45.7
M2.8	3.1	29.5	91.5
wtOcr	3.4	7.4	25.1
M1.1	3.7	8.9	33
M1.9	3.8	9.7	36.9

Table 6.1: The midpoint value, the gradient of the transition slope and the free energy of unfolding calculated from fitting the experimental chemical denaturation data of each POcr protein/wtOcr to Equation 1 and 2.

The midpoint value is the concentration at which 50% of the protein molecules are unfolded. Comparisons between the midpoint value of wtOcr and the POcr proteins indicate that M2.1 and M2.8 are less stable then wtOcr whereas, M1.1 and M1.9 are more stable. However this is not reflected for the values of free energy that were calculated for each of the POcr mutants. The ΔG values suggest that all of the POcr proteins are more stable than wtOcr. In fact the mutants M2.1 and M2.8 appear to be more stable with greater ΔG values.

The discrepancy between the values is a result of the transition slope. The gradient for the transition slope is notably steeper for the POcr mutants, the slope is dependent on changes to the surface area, in particular the accessibility of surface area between the native and unfolded state. For the POcr proteins accessibility increases compared to wtOcr due to the large reduction in the number of acidic residues on each POcr protein. However, additional differences between the transition state of each POcr protein could be due to the specific microenvironment of the tryptophan, W94. The tryptophan is located on Helix D of the Ocr monomer, located in a hydrophobic pocket flanked by several acidic residues. This could explain the differences in the local stability of the proteins, indicated by the differences in the initial slope.

Protein Stability Tests				
	TDA	**GdnHCl**		
Pocr Mutant	**Tm Temperature (°C)**	**Unfolding midpoint (M)**	**Transition slope (m)**	**Free energy of unfolding (ΔG in kJmol^{-1})**
M2.1	54	3	15.2	45.7
M2.8	54	3.1	29.5	91.5
wtOcr	54	3.4	7.4	25.1
M1.1	57	3.7	8.9	33.0
M1.9	67	3.8	9.7	36.9

Figure 6.8: The transition midpoints for the chemical and thermal stability studies on the wtOcr protein and the mutant proteins selected from the library LibDg. Right is the midpoint concentration of the step transition from the chemical denaturation curves and left is the Tm temperature calculated from the TDA assay.

The thermal and chemical stability assays showed that the mutants M1.9 and M1.1 consistently show slight increases in stability compared to wtOcr. However, the mutants M2.1 and M2.8 appear to maintain the same stability as wtOcr from the thermal studies. However, the free energy of unfolding calculated from the chemical denaturation studies indicates that wtOcr is less stable then all of the POcr proteins with POcr mutant M2.1 exhibiting the greatest stability. The values obtained for wtOcr are in agreement with previous results obtained and literature values. [122]

6.8 Protein function

The ability of each of the selected mutant proteins to prevent restriction and modification was assayed by measuring the interactions of the POcr proteins with the two active complexes of EcoKI; the M.EcoKI methylase complex and the complete EcoKI endonuclease.

6.8.1 Interaction with the M.EcoKI complex

ITC measures the heat produced or expended upon titration of a ligand (i.e., Ocr) into a cell containing an interacting partner (i.e., M.EcoKI). The heat changes are measured against a reference cell that is maintained at a constant temperature. Sensitive thermopiles produce a feedback circuit to the sample cell containing the interacting species. Any fluctuations in temperature from the sample cell are equilibrated and accounted for against the conditions of the reference cell. Therefore the signal is a measure of the time it takes for the sample cell to return to the equilibrated conditions of the reference cell and the heat energy evolved in that time.

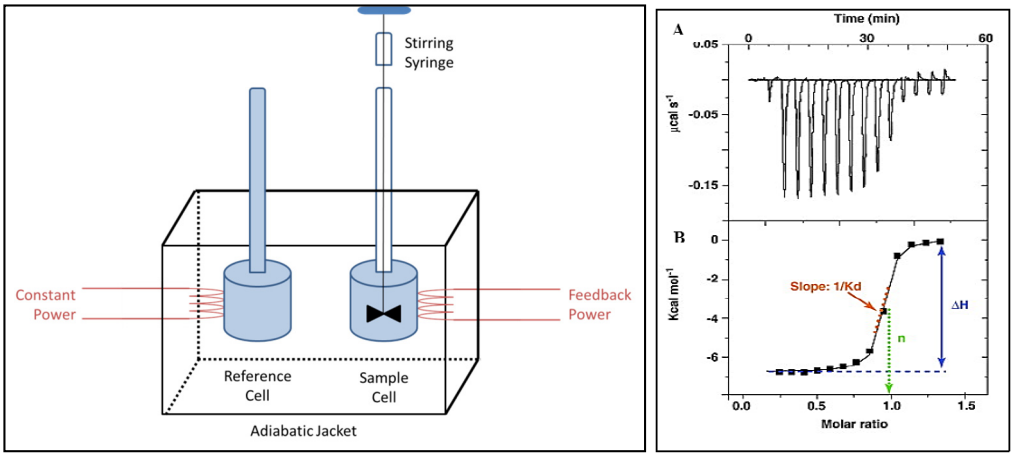


Figure 6.9: LEFT: A schematic of the ITC instrumentation. IT consists of two cells a reference cell and a sample cell (enzyme). Syringe contains the ligand of interest, which is titrated into the sample cell and mixed using the stirrer; any fluctuations in temperature are equilibrated by the feedback power linked to the reference cell. RIGHT: The heat change is recorded as a function of time as shown in graph A, the integral of this data produces a curve exhibited in graph B, where the gradient is the binding affinity value (K_d), midpoint of transition the stoichiometry (n) of the reaction and the initial heat change equals the enthalpy (ΔH) of the reaction measured.

To perform ITC experiments the proteins were exchanged into 20mM HEPES, 6mM $MgCl_2$ and 7mM 2-mercaptoethanol (pH8.0). The ligand proteins were diluted to a concentration of ~30µM (wtOcr/ POcr mutant) and the enzyme, M.EcoKI, to 3µM for a 1:10 fold ratio of

enzyme: ligand. The concentrations of the various proteins were determined from their UV absorbance. Samples were subsequently analysed using an Auto-iTC200 instrument.

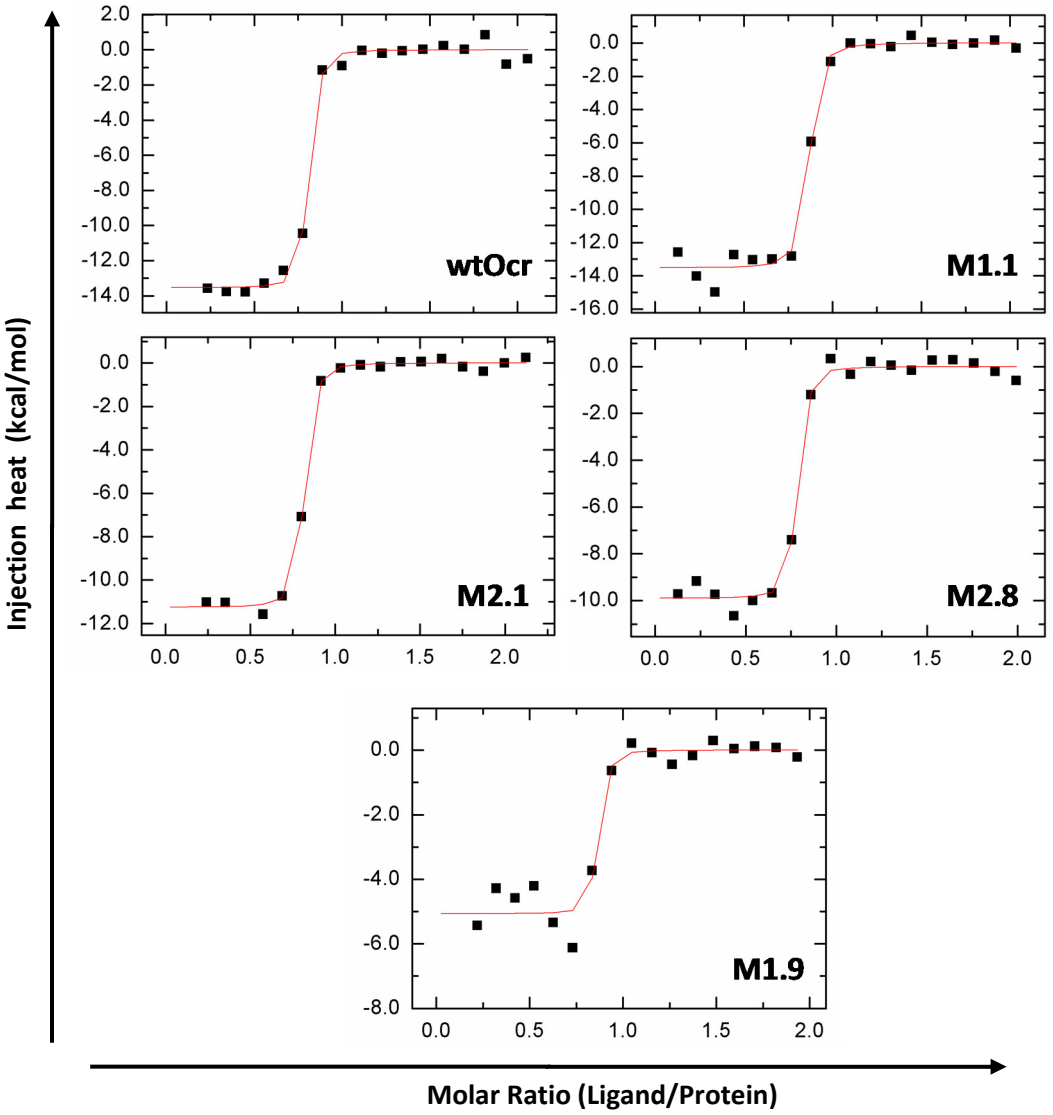


Figure 6.10: The ITC traces which follow the energy dynamics of the reaction between M.EcoKI with wtOcr and each of the four mutants selected from the library LibDg. The "Sample Cell" contained the M.EcoKI, while the "Syringe" contained corresponding Ocr protein.

The ITC traces for each for the mutants and wtOcr are displayed in figure 6.10. The graphs monitor heat changes as a function of time with each of the spikes in heat energy representing an injection of the ligand species into the cell. The baseline for each of the mutants appears to show a good signal to noise ratio. Anomalous spikes are thought to be the result of micro gas bubbles although all solutions were degassed, (the spikes of interest are

only minor errors in the baseline they appear to be large because the scale of the heat change is very small).

By representing the heat change per injection as a function of the molar ratio, transition curves for each POcr protein and wtOcr were produced. Each curve in Figure 6.11 displays a steep step wise transition which is indicative of a tight binding interaction. A comparison between the transition curve of wtOcr and the POcr mutants does not distinguish much difference between the slopes of each curve. The experimental data highlighted in table 6.2 shows that each POcr protein has a subnanomolar binding affinity towards the M.EcoKI enzyme. Although, the sensitivity of the ITC instrument means the technique can only accurately determine the binding affinity of interactions that are 10nM or weaker.

ITC Properties	wtOcr	M2.8	M2.1	M1.1	M1.9
N	0.76 +/-0.01	0.74 +/-0.01	0.77 +/-0.003	0.8 +/-0.01	0.82 +/-0.02
K_d (nM)	0.54 +/-0.26	0.49 +/-0.25	0.4 +/-0.12	0.34 +/-0.22	0.75 +/-0.01
ΔH (kcal/mol)	-13.5 +/-0.22	-9.9 +/-0.17	-11.3 +/-0.12	-13.5 +/-0.25	-5.71 +/-0.22
ΔS (cal/mol/Deg)	-5.45	6.58	1.63	-6.34	23.6
Temperature (K)	298.16	298.16	298.16	298.16	298.15

Table 6.2: A summary of the thermodynamic data extracted from the ITC traces for the interaction of M.EcoKI with each of the mutants and wtOcr.

Furthermore, the stoichiometry for each of the mutants and wtOcr appears to be the same at 0.7-0.8 of dimer:M.EcoKI. The discrepancy from a 1:1 ratio allows for errors in the active concentration of the enzyme and protein but it would appear that a 1:1 ratio is observed for each M.EcoKI:Ocr/mutant interaction.

The striking difference between each POcr protein is the ΔH value displayed in Figure 6.13. The enthalpy change upon binding, between wtOcr or POcr mutants with M.EcoKI is exothermic. Wild type Ocr the mutants M2.1 and M1.1 display similar enthalpies of -13.5kcal/mol and -11.3kcal/mol. However, the enthalpy of the proteins M2.8 and M1.9 is greatly reduced to -9.9kcal/mol and -5.6kcal/mol, respectively. From the table above the decrease in the exothermic enthalpy appears to loosely correlate with an increase in positive entropy value as would be expected.

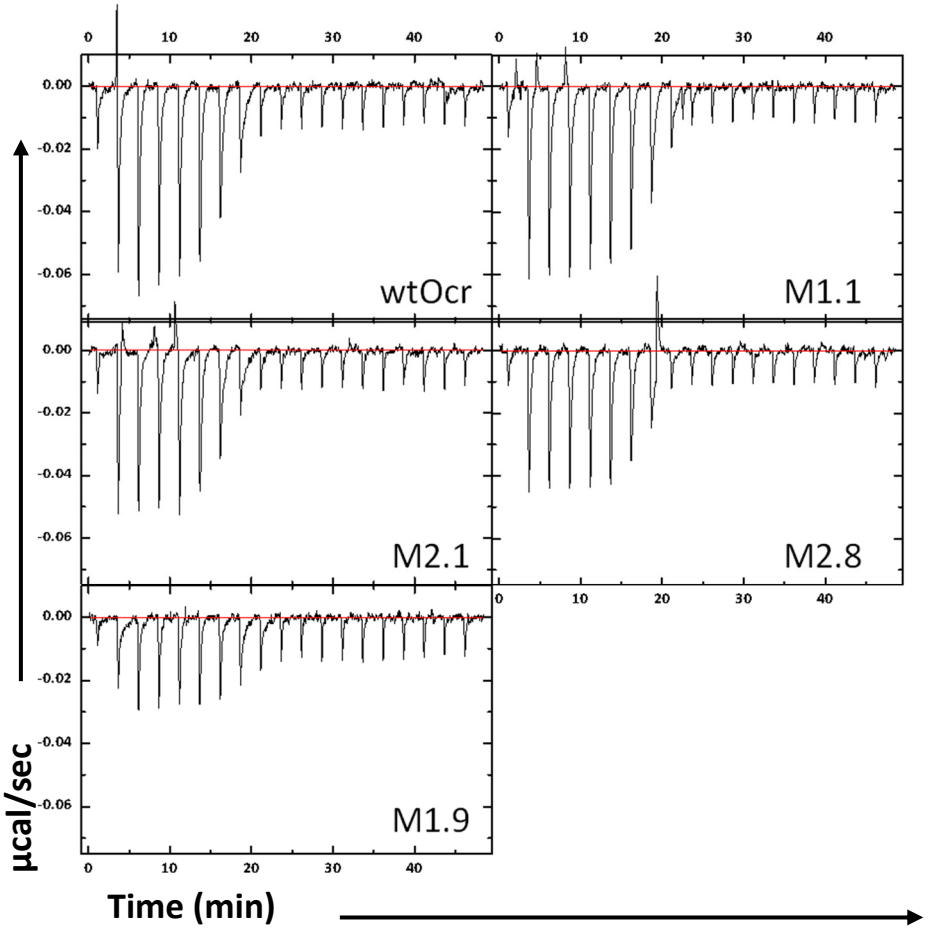


Figure 6.11: A comparison of the ITC heat energy exchange traces. Each trace has been altered to the same scale as the wtOcr trace to make the traces directly comparable. Each spike represents the injection of the ligand, Ocr/mutant protein in the cell sample, M.EcoKI.

6.8.2 Interaction with the nuclease

The interaction of each mutant with EcoKI was measured by performing a nuclease assay. This assay measures the linearisation of the unmethylated plasmid pBrsk1, which contains a unique recognition site for the EcoKI nuclease. In each case, wtOcr or the POcr proteins were preincubated with EcoKI in the presence of other cofactors required for digestion (i.e., ATP and SAM) and then the reaction was initiated by the addition of plasmid DNA. The reaction was stopped after 8mins and the extent of digestion was analysed by agarose gel electrophoresis. The EcoKI enzyme was initially incubated with a 20-fold or 10-fold excess of each mutant protein and wtOcr for comparison.

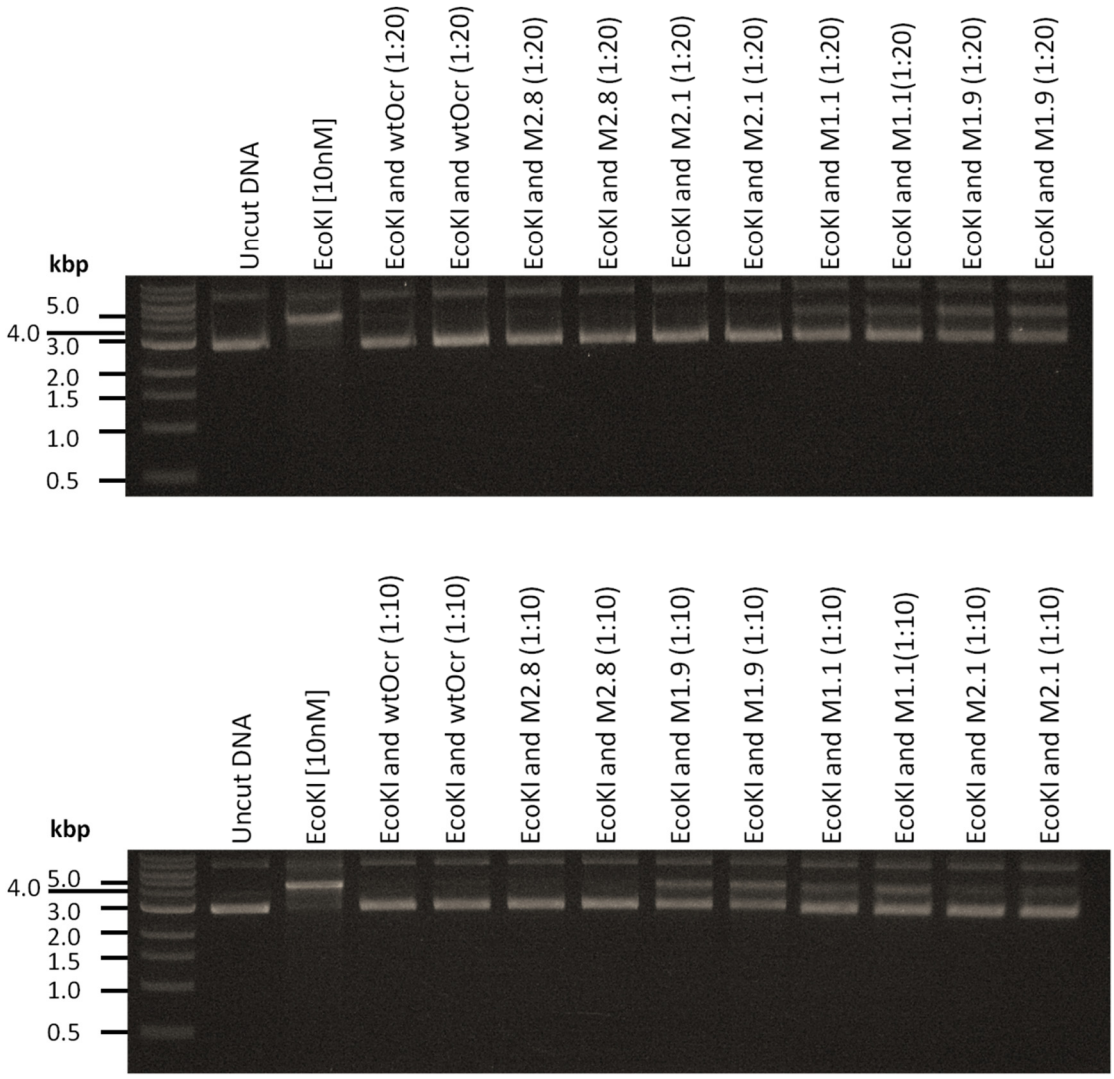


Figure 6.12: *In vitro* nuclease assay, displaying the inhibition activity of the Ocr multimutants and wtOcr protein on EcoKI. The nuclease was incubated with a 10 fold (below) and 20 fold (above) excess amount of each mutant before the reaction was initiated.

The inhibition of the EcoKI enzyme was successful for wtOcr and the mutant M2.8 at a 20-fold and 10-fold excess of the Ocr/mutant protein, with no detectable plasmid digestion. However, some partial digestion was observed when EcoKI was incubated with the mutants; M1.9 and M1.1 at a 20-fold excess. Comparisons of the two gels in Figure 6.12 show that M1.9 and M1.1 exhibited the weakest antirestriction activity whereas, the POcr proteins M2.1 and M2.8 produced the same antirestriction activity as wtOcr.

Further analysis investigated the restriction ability of the nuclease incubated with 5–fold excess of the mutant proteins and wtOcr. In addition, wtOcr and M2.8 were also incubated

with the nuclease using a 2-fold excess in an attempt to assess whether there any difference in antirestriction activity could be distinguished between these two proteins, figure 6.13.

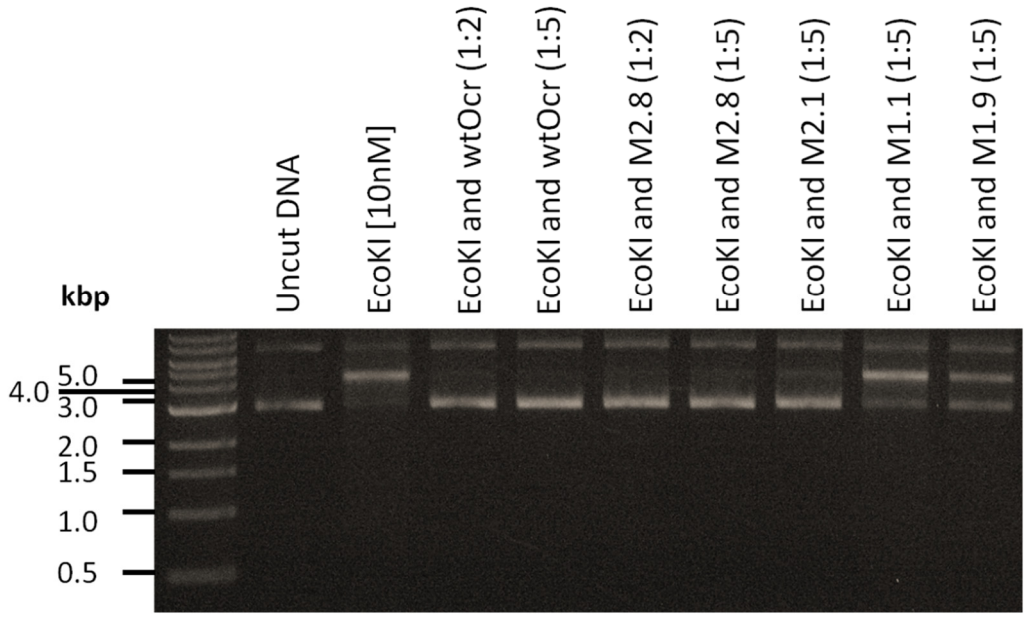


Figure 6.13: The *in vitro* nuclease assay displaying the inhibition activity of the Ocr mutant and wtOcr protein on EcoKI. The nuclease was incubated with a 5 fold excess of each mutant protein and 2 fold excess for the mutants M2.8 and wtOcr before the reaction was initiated.

A similar pattern of inhibition was observed when these results were compared to those shown in Figure 6.12.

In summary although all of the POcr proteins retain some antirestriction activity, this activity appears to vary *in vitro*. M2.8 appears to be the only mutant protein that retains full antirestriction activity comparable to wtOcr. M2.1 retained slightly weaker antirestriction activity and M1.9 and M1.1 displayed the weakest antirestriction activity.

6.9 Discussion

6.9.1 Protein structure

In solution wtOcr exists as a stable homodimer in order to function effectively. Cross linking studies carried out using the reagent glutaraldehyde showed that a stable homodimer was also present in solution for each of the mutant proteins. Furthermore, CD analysis revealed that the removal of 60% of the acidic residues had no effect on the structural fold of the

proteins secondary structure compared to wtOcr. Remarkably the extensive reduction in the number of acidic residues does not affect the secondary structure or dimerisation of the POcr proteins when compared to wtOcr.

6.9.2 Protein stability

The stability of the protein was assessed using two methods, thermal stability by a thermal denaturation assay and chemical stability using the denaturant, Guanidine Hydrochloride (GdmCl).

The results from the chemical stability study exhibited changes in stability related to the hydrophobic properties of each POcr proteins, rather than the electrostatic interactions which are masked by direct interaction with Guanidine ions.[124] The free energy of unfolding calculated from these studies showed that the removal of acidic residues reduced exposed hydrophobic domains which has resulted in an increase in the stability of the POcr proteins. Furthermore, the differences between the numbers of acidic residues in close proximity to the tryptophan residue could affect the resulting free energy value. On the other hand thermal denaturation determines the overall stability of a protein and takes into account all properties that may affect stability. With this in mind the thermal studies showed that the proteins M2.1 and M2.8 were less stable than M1.1 and M1.9. In fact the same trend is noticeable from the calculated $D_{50\%}$ midpoint value from the chemical denaturation study which also suggests that M2.1 and M2.8 are less stable than the proteins M1.1 and M1.9. This suggests that any increases in stability gained by the accessibility of hydrophobic residues in the chemical denaturation of M2.1 and M2.8 are compensated by other properties of the proteins, according to the thermal stability results.

6.9.3 Protein function

The ITC results showed that each of the mutants exhibited tight binding towards the M.EcoKI. The interaction between each POcr protein and M.EcoKI was comparable to wtOcr, all exhibiting subnanomolar binding affinities. As such, ITC is not sensitive enough to differentiate any subtle differences between the affinities of each POcr protein to MEcoKI.

Furthermore, the ΔH of the interaction is more favorable for wtOcr, M1.1 and M2.8 and considerably less favorable for POcr protein M1.9 and M2.8. The decrease in the exothermic value of the enthalpy cannot be correlated to a decrease in the number of acidic residues but the decrease in enthalpy is compensated by an increase in entropy, suggesting the production of a more ordered protein: enzyme complex. Anomalous contributions from the ligand or the

buffer were discounted by carrying out a control experiment which injected the ligands; wtOcr and the POcr proteins into buffer, (see appendix, Figure A.2).

From ITC, the interaction between M.EcoKI and the POcr proteins is indistinguishable from wtOcr with M.EcoKI. In contrast the results from the nuclease assay indicate that the POcr proteins elicit variable antirestriction activities. The POcr protein M2.8 retains antirestriction activity as effective as wtOcr, whereas antirestriction activity exhibited by M2.1 is moderately weaker and for the mutants M1.1 and M1.9 partial antirestriction activity is observed *in vitro*. Each of the POcr proteins possess a similar number of acidic residues therefore the difference in antirestriction activity does not reflect a decrease in the number of acidic residues. However the reduction in antirestriction activity for the mutants M1.9 and M1.1 correlates with an increase in the protein stability of the POcr protein M1.9 and M1.1. This may indicate that increased flexibility is more favourable for binding to the M.EcoKI; this could possibly be due to slight changes in conformation of the Ocr protein, upon binding.

Previous investigations carried out by A. Stephanou et al, 2009 revealed that the removal of 16 acidic residues from each Ocr monomer reduced the binding affinity of Ocr with the M.EcoKI complex from a K_d of 44pM to ~20nM, which incidentally matches the interaction between the M.EcoKI enzyme and dsDNA.[120] In contrast, the results from this chapter indicate that the removal of 21 acidic residues had no distinguishable effect on the binding interaction of the protein to M.EcoKI, which remains subnanomolar for each of the active POcr mutants. The structure of each mutant protein was comparable to wtOcr and revealed no apparent differences. Stability studies showed that the reduction in acidic residues generally increased the stability of the POcr mutants but not enough to affect the interaction of these proteins with M.EcoKI. On the other hand, the antirestriction activity of the POcr proteins was reduced, this implies that the POcr proteins differentially bind to the M.EcoKI and R.EcoKI enzyme. Sequences do show the absence of acidic residues present on Helix D which are thought to interact with the R subunits.[61]

In conclusion the need for acidic residues on the Ocr protein beyond the 13-15 acidic residues essential for the active POcr proteins, selected from the library LibDg, is still unknown. The effect and advantage of these acidic residues which represent ~60% of the total number on the Ocr protein is unknown in regards to the antimethylation activity. However, the reduction in acidic residues does appear to affect the efficiency of the

antirestriction activity of the protein, furthermore this decrease in antirestriction activity *in vitro* does not translate to a reduction *in vivo*.

Chapter 7. Partially active POcr mutants

7.1 Introduction

It is known that the production of the synthetic positive Ocr gene, pOcr, in which all of the acidic residues were substituted by their neutral analogue, induced a loss of activity but also a loss of protein solubility and even protein expression. Expression trials failed to produce the POcr protein in *E. coli* cells (unpublished work carried out by Dr GA Roberts/Dr JH White). This shows that the acidic residues of the protein are clearly important in terms of protein folding/stability. However, two issues need to be addressed: (i) At which point do the acidic residues become such a vital component of the protein?; (ii) how many (and which) of the acidic residues reintroduced into POcr are required for expression, solubility and finally biological function?

7.2 Aims

The aim of this chapter was to investigate the role of acidic residues in every aspect of the activity, structural integrity and expression of the wtOcr protein, by investigating POcr mutants with varying pI values, variable functionality and induce different phenotypes compared to wtOcr. These POcr mutants were assessed by relating their activity to the following important factors;

(i) The number of acidic residues that are present on each POcr mutant protein.

(ii) Positional context of the acidic residues in the primary amino acid structure of POcr.

As a result a large screen was carried out on POcr mutants isolated from each of the nine libraries that were generated in chapter 4. Approximately 50 POcr mutants that displayed various activities and possessed unique pI values were investigated more thoroughly. Furthermore, the sequences from the library LibDg (which includes the active POcr mutants isolated from LibDg in chapter 5) were assessed thoroughly to produce a map of Ocr which demonstrates the relative importance of each acidic residue towards the structural integrity and activity of Ocr.

The aim of the work was to investigate how negative charge can affect the activity of the POcr mutants *in vivo*.

7.3 Initial screening

Table 7.1 identifies the total number of POcr mutants that were screened and which library they were isolate from. In total 102 POcr mutants were isolated and screened using the 2AP assay developed in chapter 3.

POcr Library	Nº of POcr mutants screened
Lib1	7
Lib2	7
Lib3	7
Lib4	7
Shuffle Lib	10
Lib 1D	14
Lib 2D	9
1D+2D	11
Lib Dg	30
Total	**102**

Table 7.1: To carry out a large scale screen 102 POcr mutants were selected at random from each of the POcr libraries generated in chapter 4. The table shows the distribution the 102 POcr sequences screened and from which of the POcr libraries they belong to of the nine POcr libraries generated in chapter 4.

In chapter 3, the 2AP assay managed to elucidate three different phenotypes; inactive, active and partially active mutants. The level of $2AP^R$ induced by the expression of Ocr multimutants in *E. coli* cells was successfully correlated with their antirestriction and antimodification activity. In the same manner the POcr mutants highlighted in table 7.1 were screened and categorised as inactive, active or partially active depending on whether the mutants exhibited 2AP resistance at, 20μg/ml and 80μg/ml or whether induction of the POcr mutant reduced the cell viability of *E. coli* NM1041 cells.

7.4 Initial 2AP studies

Initial studies assessed $2AP^R$ by transforming *E. coli* NM1041 cells with the plasmid encoding the appropriate POcr mutant. Single colonies of each mutant were then streaked onto four types of plate; (i) a control plate supplemented with antibiotics only, (ii) an induction plate supplemented with antibiotics and IPTG, (iii) a plate supplemented with antibiotics and 80µg/ml of 2AP ($2AP_{80}$) and (iv) a plate supplemented with antibiotics and 20µg/ml of 2AP ($2AP_{20}$). The resulting viable cell count from these two concentrations of 2AP and the addition of a plate supplemented with IPTG was used to determine whether the mutants exhibited no 2AP resistance (green), partial resistance (blue) or affected the growth of cells (red).

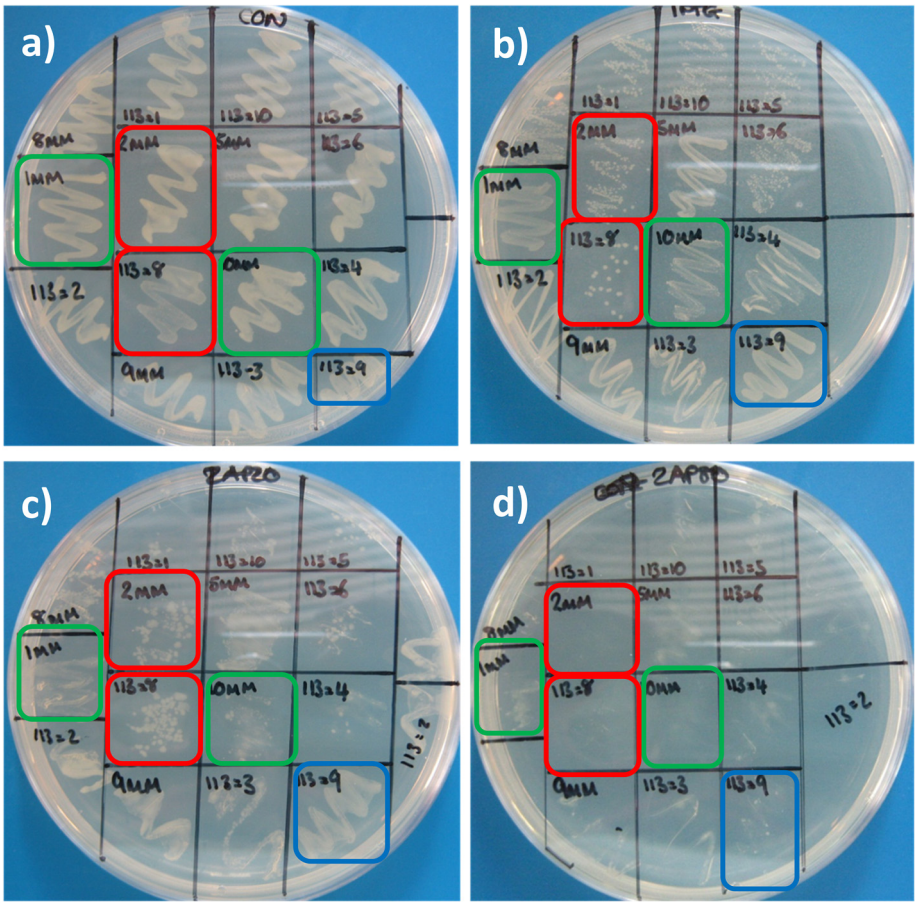


Figure 7.1: The growth of a single colony that has been streaked onto four different LB-agar plates supplemented with a) antibiotics only, b) antibiotics and IPTG, c) antibiotics, IPTG and 2AP (20µg/ml) and d) antibiotics, IPTG and 2AP (80µg/ml). The plates were incubated overnight at 37°C. Highlighted in green are colonies which appear to have no resistance to 2AP, in blue those with some resistance, and in red those mutants that affect the growth of cells.

An example of the mutants that were investigated is shown in Figure 7.1. Plate a) contains antibiotics only and each streak produces a homogenous growth pattern and represents a mutant that has been successfully transformed. Changes in cell viability upon induction of protein expression by the addition of IPTG are displayed in plate b for each of the POcr mutants. Furthermore, plates c and d contain increasing concentrations of the DNA mutagen 2AP. POcr mutants that affect the viability of *E. coli* cells were highlighted; in red were POcr mutants that appeared to severely affect cell viability whereas in green are POcr mutants that induce the same effect but to a lesser extent. Lastly, in blue are POcr mutants that are inactive and are $2AP^R$ to 20µg/ml but are not resistant to 80µg/ml.

The preliminary results provide an insight into the activity profile of each mutant. However, the experiment was repeated by comparing the viable cell count, on each type of plate, using serial dilutions of *E. coli* NM1041 cells transformed with each of the mutant constructs.

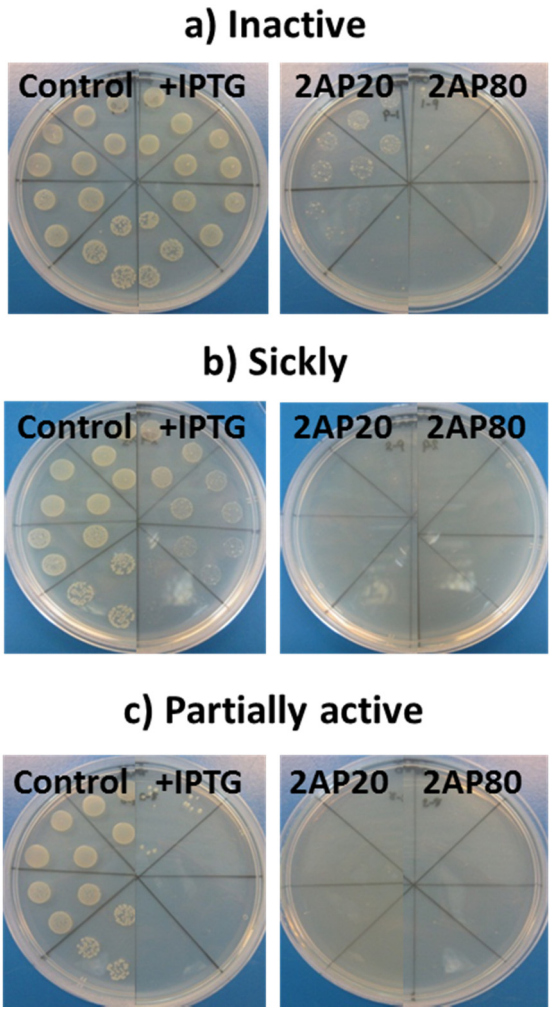


Figure 7.2: Three different phenotypes were induced by the 102 POcr mutants screened. The three phenotypes had been witnessed previously in chapter 3 upon assessment of the Ocr multimutants where the phenotypes were classified as expressing mutants that exhibited no activity (a), partial activity by inducing poor growth (b) or no growth(c). Viable counts were plated as 10-fold serial dilutions in an anticlockwise direction for the control and 2AP20 plates and a clockwise direction for +IPTG and 2AP80 plates.

The cell survival rate of *E. coli* NM1041 cells was measured in the presence of a POcr mutant protein and 2AP. The results in Figure 7.2 demonstrate the typical phenotypes that were displayed as a consequence. Figure 7.2 demonstrates that expression of the mutants in *E. coli* NM1041 cells, in the presence of 2AP produced one of three phenotypes; a 2AP resistance of 20µg/ml and two phenotypes that displayed a reduction (Figure 7.2b) or inhibition of cell growth (figure 7.2c). Results from chapter 3 show that the latter two phenotypes are common among mutants possessing intermediate binding affinities to RM enzymes. Consequently, these will be both termed as partially active.

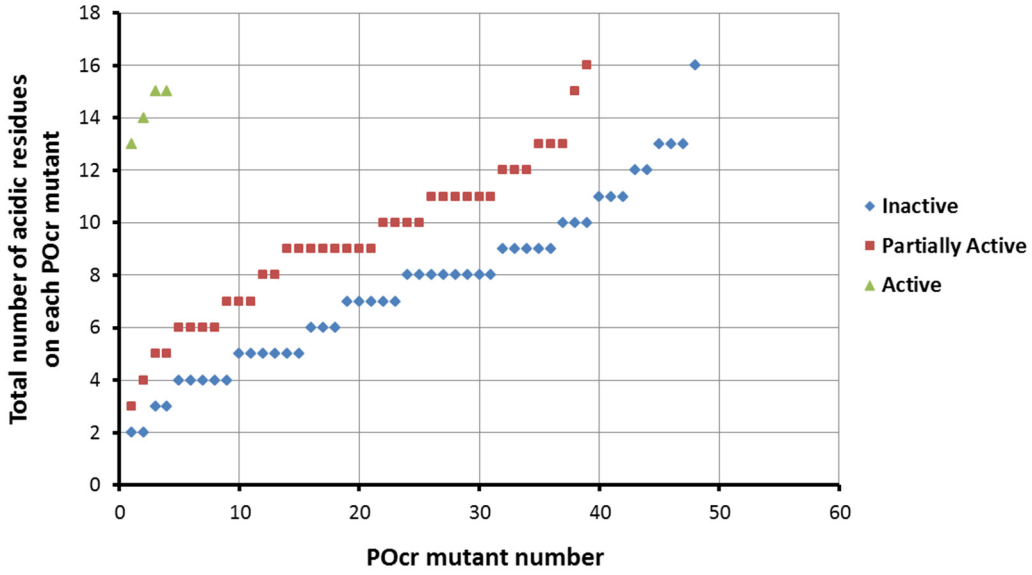


Figure 7.3: A scatter graph showing the relationship between the total numbers of mutations found on each mutant protein and the corresponding phenotype in the 2AP assay (from a total of 102 sequences). The sick phenotype and partially active phenotype demonstrated in figure 7.2 have been combined in this graph. For comparison, the active mutants from chapter 5 and 6 have been included.

The relationship between the number of acidic residues in POcr and the induced phenotype is shown in Figure 7.3 above. A larger proportion of the partially active mutants contain a greater number of acidic residues; on average 9 acidic residues were present on partially active mutants compared to an average of 7 acidic residues on inactive mutant mutants. However, this does not explain why mutants that contain the same number of acidic residues as the active mutants, isolated in chapter 5, remain inactive. This observation suggests that either (i) the position of an acidic residue/the sequence context of acidic residues is important or (ii) the mutant mutants categorised as 'partially active' in fact display different activities that result in the same phenotype.

7.4.1 Positional context of acidic residues

To assess whether the positional context of acidic residues is important and to pinpoint exactly which positions are of greatest importance, further analysis was carried out on POcr mutants that were isolated from the library LibDg only. The reason why only LibDg POcr mutants were analysed was so that direct comparisons could be made with the active POcr mutants isolated from LibDg in chapter 5. Therefore the 2AP screen above showed that of the 26 POcr mutants from LibDg 16 were partially active and 10 were inactive. These were

compared with the additional four active POcr mutants that were characterised in chapters 5 and 6.

The importance of each acidic residue towards the activity of the mutant was assessed by calculating whether antirestriction activity increases if a particular position contains an acidic residue. In order to do this the likelihood of a residue being acidic at each position of interest was calculated for the whole population of 30 mutants as a percentage, and this was compared to the percentage likelihood that a position is acidic from POcr mutants inducing a particular phenotype. From this bar charts were produced for each phenotype; Active, Partially Active and Inactive. These charts present a positive and negative percentage for each position of interest on the primary protein structure of POcr; these percentages represent the probability that a residue is acidic (positive percentage) or neutral (negative percentage) among the active (A)/ partially active (PA)/ inactive (IN) mutants compared to the whole population. Therefore a positive percentage represents an increased probability that an acidic residue at the supposed position is important to the phenotype of interest whereas a negative percentage suggests that an acidic residue is not essential at this position to induce the phenotype of interest.

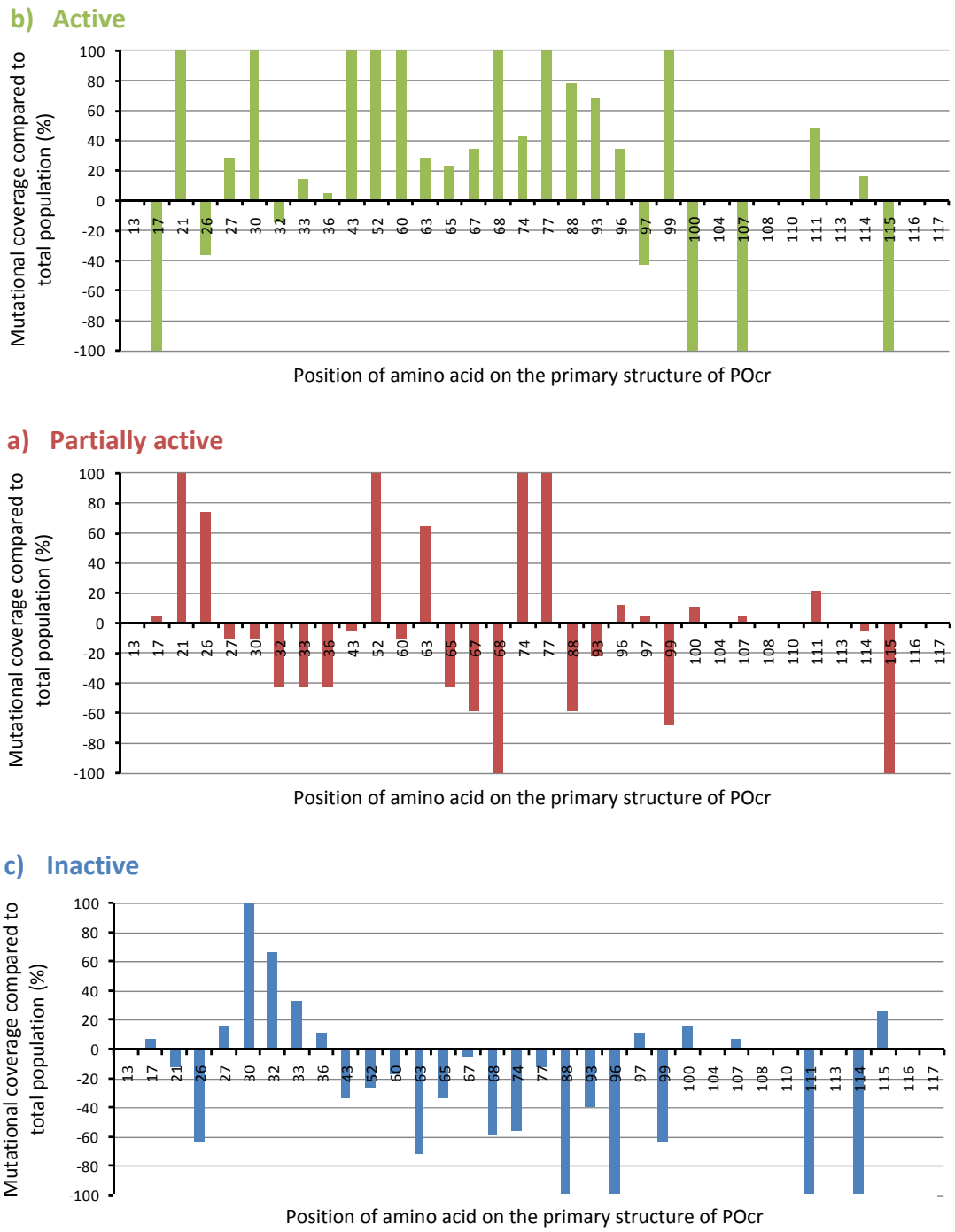


Figure 7.4: Three graphs that compare the % likelihood that a target site is acidic for POcr mutants that induce phenotypes; a) active (green), b) partially active (red) and c) inactive (blue,) as characterised by the 2AP screening procedure. The probability is compared to the % likelihood that the same target site is acidic when selected at random from the total population of POcr mutants characterised from LibDg.

The graphs show the importance of acidic residues at particular positions for POcr mutants that induce either an; active, partially active or inactive phenotype for antirestriction. General

trends show that an increase in activity correlates with an increase in the importance of acidic residues, with many acidic residues becoming more prominent as activity increases.

It was suspected that no acidic residue would be present in abundance among inactive POcr mutants, indeed this was the case for all the acidic residues with the exception of positions 30, 32 and 33 (likely a result of the small sample size). This suggests that the presence of acidic residues on POcr mutants is not important for inactive POcr mutants.

However, POcr mutants that induce partially active and fully active antirestriction properties have a number of acidic residues that appear to be important in producing functional protein. A large number of acidic residues are present at high levels on the active POcr mutants with several conserved acidic residues (100% acidic). These were previously highlighted in chapter 6 and include acidic residues E21, D52, D74 and D77 which are also conserved on the partially active POcr mutants. This suggests that the acidic residues at these positions are important for partially active and fully active POcr mutants. Not only this but that these acidic residues are important towards the structural integrity of the POcr protein, rather than its biological activity.

A fourth graph was produced in figure 7.5 that investigates the importance of acidic residues present on active POcr mutants, compared to inactive and partially active POcr mutants. This would highlight all of the acidic residues that appear to be important in antirestriction activity as opposed to those residues that are structurally important and so are present in high abundance among partially active POcr mutants.

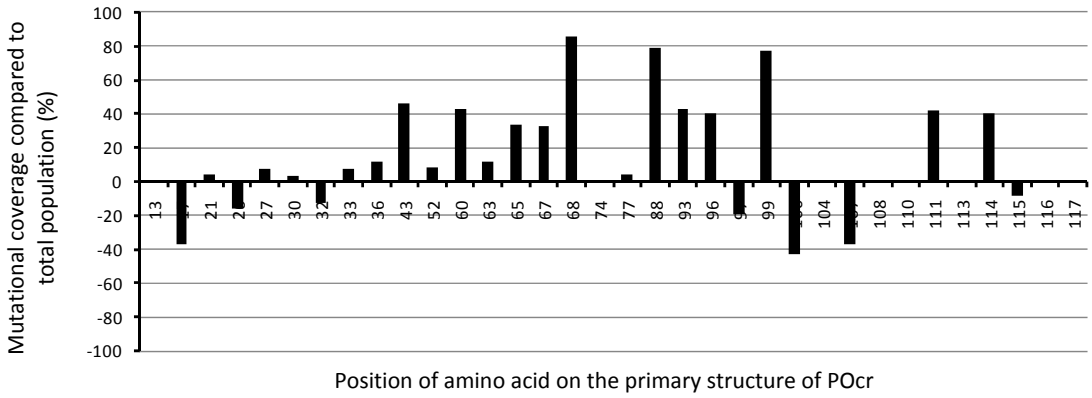


Figure 7.5: A graph that compares the % likelihood that a target site is acidic for POcr mutants inducing an active phenotype, as characterised by the 2AP screening procedure, compared to the % probability that the target site is acidic if selected from the POcr mutants characterised from LibDg that induce an inactive or partially active phenotype.

Figure 7.5 shows that the acidic residues 68, 88 and 99 are in high abundance on POcr mutants that possess an active phenotype. This suggests that these residues are essential components for the activity of the POcr mutants, given that there is a ~80% likelihood that these positions are acidic.

Other residues that appear to have some importance include the midloop region (residues 60-68) and the Helix D region (88-96). The likelihood that these residues are acidic among active POcr mutants is ~40%. Notably this is in agreement with results from chapter 3 which indicate that the loop 2 region and a region on Helix D (Ocr multimutant M4) were important for antimethylation and antirestriction activity respectively.

Interestingly, residues from the 1oop 1 region (26-36), which were also highlighted as important in chapter 3, were not as highly represented in figure 7.5 above. This suggests that these residues could play a minor role in the activity of the POcr mutants. Alignment of the active POcr mutants in chapter 5 shows that specific residues in the loop1 region were not conserved and that a non-specific representation of acidic residues in the two loop regions was sufficient for activity rather than acidic residues at specific positions.

In summary the analysis of the POcr mutants from LibDg highlighted the importance of an acidic residues position on the performance of POcr as an antirestriction protein. The results show that the following acidic residues E21, D43, D52, D74 and D77 could possibly be structurally important. Whereas the acidic residues at D60, D68, E88 and D99 in addition to

a number of non-specific acidic residues in both loop regions on POcr can produce a fully functional antirestriction protein.

The analysis so far implies that the presence of at least five acidic residues is essential in producing a partially active antirestriction protein. However, Figure 7.3 shows that inactive POcr mutants can contain as many as 16 acidic residues and partially active mutants can possess as few as 3 acidic residues. These results suggest that the partially active POcr phenotype, which observed a reduced cell growth in *E. coli* NM1041 cells, may in fact have very different properties.

7.4.2 Investigating the partially active phenotype

A number of intermediate and inactive POcr mutants were selected for further analysis. In total forty two sequences were selected from the initial study. These mutants had a range of 3-15 acidic residues per mutant. Twenty of these POcr mutants were inactive, with a $2AP^R$ of 20µg/ml and possessed 3 to 13 acidic residues. Twenty two mutants were partially active and contained between 5-15 acidic residues. These have been highlighted in Figure 7.6.

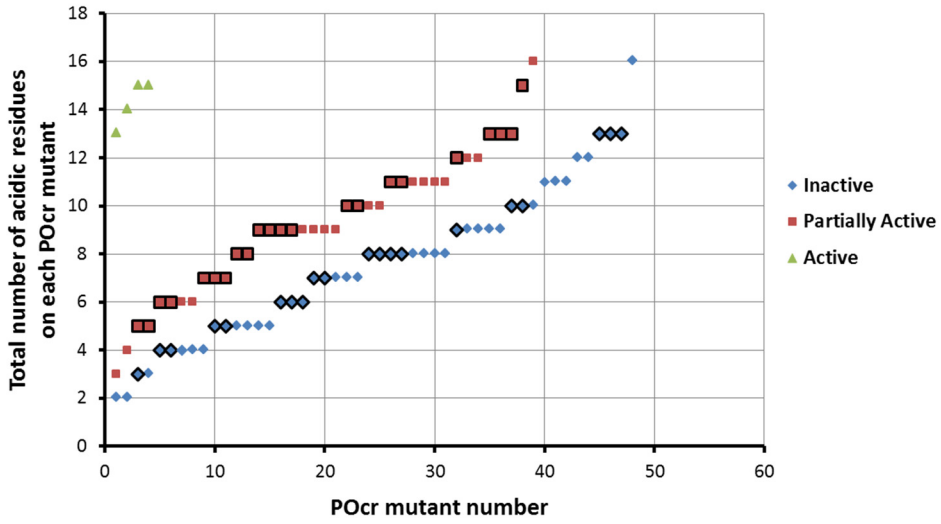


Figure 7.6: The selection of mutants that were chosen to be carried forward for further analysis. The graph describes whether the POcr mutants screened were inactive or partially active and the number of acidic residues each mutant possessed. POcr mutants that were selected for further analysis are highlighted by a black border.

Previous results in chapter 3 have shown that a number of the Ocr multimutants were partially active. Upon overexpression in *E. coli*, these mutants were detrimental to cell growth inducing a phenotype of $m^{+/-}r^{+/-}$. Upon further investigations these Ocr multimutants

were shown to be active from *in vivo* phage infection assays. If like, the Ocr multimutants, the POcr mutants exhibit the $m^{+/-}r^{+/-}$ phenotype then they will also exhibit antirestriction and antimethylation activity when carrying out phage infection assays.

The eop results of the phage titre were used to calculate an arbitrary unit of antirestriction and antimodification for each of the mutants, by determining the ratio of the eop value for each mutant to the eop of the negative control, the empty plasmid (pTrc99A) vector.

wtOcr	7500.00	625.00
pTrc	1.00	1.00

Mutant	Anti restriction	Anti modification
ED10 (3)	0.10	1.23
C8(4)	0.24	0.88
D7(4)	0.12	0.83
E2(5)	0.24	0.6
ED9(5)	0.33	1.20
A11(6)	19.50	1.00
D5(6)	0.40	0.66
ED4(6)	0.15	1.19
A10(7)	0.09	0.95
D10(7)	0.28	0.60
E5(8)	0.17	1.99
ED3(8)	0.50	0.24
Dg10(8)	2.00	1.90
ES5(8)	0.17	1.58
ES4(9)	0.40	1.17
B12(10)	1.17	n/a
113-3(10)	0.67	0.50
ES8(13)	0.38	0.39
Dg12(13)	0.15	3.79
Dg2(13)	0.20	1.11

Mutant	Anti restriction	Anti modification
G2=8 (5)	0.75	2.76
2.5 (5)	0.15	0.15
1+2-8 (6)	0.90	1.17
ED6 (6)	0.38	2.00
1+2-5 (7)	0.86	0.95
3.29 (7)	0.30	1.56
2mm (7)	0.25	2.70
2.1 (8)	0.50	1.17
3.2 (8)	0.21	0.88
2+1-9 (9)	3.90	1.33
2+1-1 (9)	0.35	1.38
G2=7 (9)	0.90	1.50
1+2-7 (9)	3.25	n/a
2+1-6 (10)	0.21	n/a
B7(10)	0.21	0.90
2+1-10 (11)	0.05	0.58
B4(11)	0.38	4.38
Dg4 (12)	0.28	1.31
2+1-8 (10)	0.60	0.88
2.4 (13)	0.33	1.53
Dg3 (13)	0.33	1.07
Bx3.11 (15)	2785.71	5250.00

Table 7.2: The results from the phage infection assay. The unmodified lambda phage was used to infect restriction proficient and restriction deficient cells that were transformed with the mutant of interest. The resulting phage titre and eop value were calculated and the antirestriction value determined. Phage plaques from the restriction proficient cells were subsequently used to make a phage stock which was further used to infect *E. coli* cells in the presence and absence of the HsdR subunit. The resulting phage titre and eop value calculated the antimodification value. The results were carried out with a negative and positive control; the empty vector pTrc99A and wtOcr protein, respectively. Therefore an antirestriction and antimodification value greater than 1 suggests some antirestriction and antimodification activity is present. Errors were calculated from deviation of repeats from a mean average and were approximately +/- 20%.

Table 7.2 shows that each of the POcr mutants demonstrates no antirestriction or antimodification activity from the phage assay. The one exception was the POcr protein Bx3.11, which not only exhibits antirestriction and antimodification activity but this activity is similar to the activity of the wtOcr protein, (incidentally this mutant contains the largest

number of acidic residues of all the POcr mutants that were assessed). The remaining 21 partially active POcr mutants possessed no apparent activity towards EcoKI, suggesting that these mutants may be toxic to the cell.

7.4.3 Cell toxicity and the restriction pathway

If the toxicity induced by the partially active POcr mutants is not dependent on its interaction with the Type I RM enzyme EcoKI, then the POcr mutants could reduce the growth of all *E. coli* cells. Such toxicity can be a common problem in recombinant protein expression in *E. coli* cells.[103]

To assess the toxicity of each POcr mutant in table 7.2, bacterial growth curves were measured for *E. coli* NM1041 and NM1261 cells strains ($HsdR^{+/-}$) with the expression construct of each POcr mutant. The growth curves for each cell strain were carried out with and without induction of heterologous gene expression. OD_{600} measurements were carried out every 5mins, for a total of 500mins at 37°C, with shaking at 180 rpm between readings.

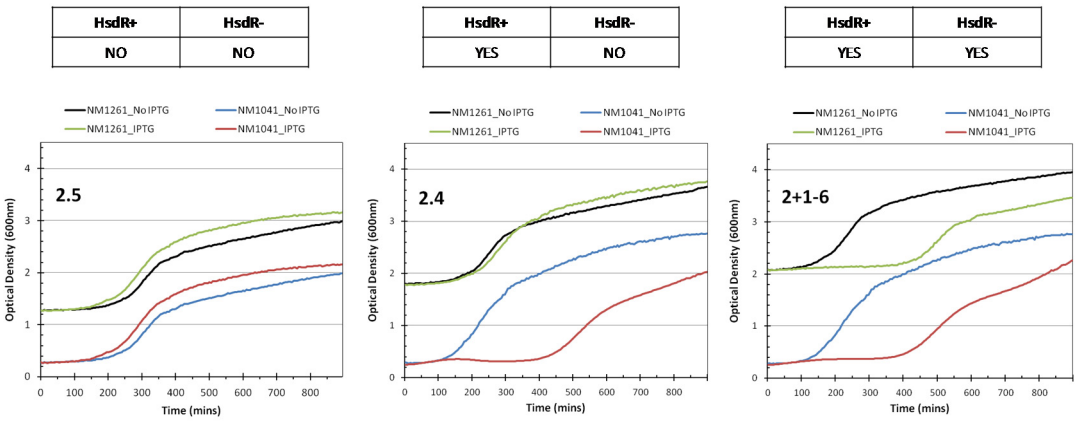


Figure 7.7: Three graphs which demonstrate the three main phenotypes that were recognised from carrying out bacterial growth curves of each of the mutants in the presence and absence of the HsdR subunit from left to right. The growth curves for the *E. coli* NM1261 cells were offset from t=0, OD600=0 for comparison.

The bacterial growth curves in Figure 7.7 demonstrate the difference between *E. coli* cells expressing each of the POcr mutants in the presence and absence of the HsdR subunit. The two sets of graphs refer to the growth curves obtained for *E. coli* NM1041 cells ($HsdR^{+}$) or *E. coli* NM1261 cells ($HsdR^{-}$) in the presence/absence of IPTG. The effect induced by each of the partially active POcr mutants on the growth rate of *E. coli* NM1261/NM1041 cells, could be defined by three major characteristics as; N/N, Y/N and Y/Y (Figure 7.7).

1) N/N- *E. coli* NM1041 and NM1261 cells exhibited no changes in growth upon expression of the mutant 2.5.

2) Y/N- the mutant 2.4 reduces cell growth when expressed in *E. coli* NM1041 cells which contain the RM system EcoKI. However, upon expression of the mutant in the absence of EcoKI (*E. coli* NM1261) cell growth is unaffected. This result suggests that the reduction in cell growth is a direct result of the restriction pathway and the presence of the mutant; implying that the enzyme and the mutant are interacting.

3) Lastly the mutant 2+1-6 demonstrates a Y/Y growth effect. Therefore, the expression of the mutant reduces the cell viability of *E. coli* cells in the presence and absence of the HsdR subunit. Such growth curves show an extended lag period with a growth delay of ~400mins. This suggests that expression of these mutants will be toxic in all *E. coli* cells.

Upon closer inspection of the partially active POcr mutants in figure 7.2, those mutants that induce a Y/N growth pattern generally possess ≥8 acidic residues with the exception of only one Y/N POcr protein which contains 6 acidic residues. However, those POcr mutants that induce a toxic effect (Y/Y) appear to possess a variable number of acidic residues that ranges from 3-15 acidic residues.

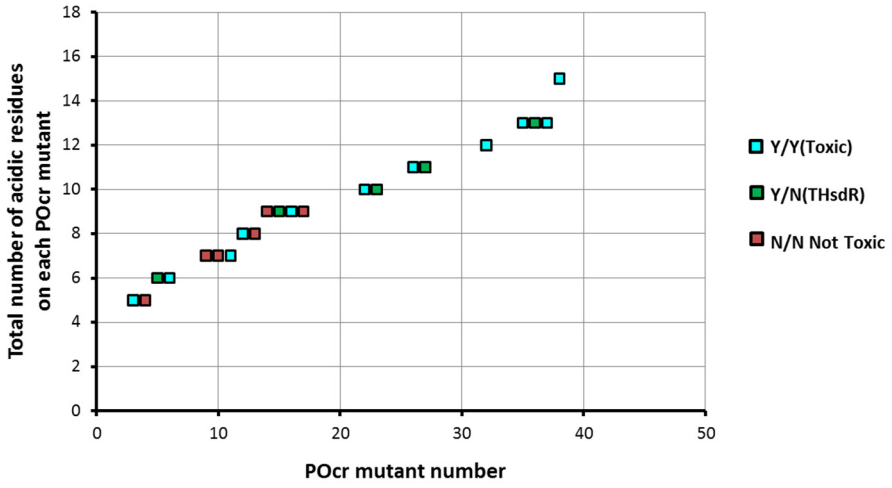


Figure 7.8: The graph above depicts each of the partially active mutants examined in relation to the number of acidic residues each of the POcr mutants contain. The colours denote the growth profile that each of the mutants exhibited from the examples shown in figure 7.7. The graph show that is no obvious trend between the growth profile induced by the partially active mutants and the number of acidic residues they possess.

The results have shown that the POcr mutants that were defined as partially active from the initial 2AP screen in figure 7.1, in fact comprised of POcr mutants that possessed different growth profiles, figure 7.8.Three new phenotypes were discovered;

i) POcr mutants effected the growth of *E. coli* cells, in the presence of an EcoKI enzyme but displayed no antirestriction/antimethylation activity upon phage infection.

ii) POcr mutants reduced/inhibited the growth of *E. coli* cells in the presence and absence of the EcoKI enzyme and so were toxic in all *E. coli* cells.

iii) The POcr mutant Bx3.11, was toxic to all cells but shared the same phenotype as the partially active Ocr multimutants (demonstrating full antirestriction and antimethylation activity upon phage infection).

7.4.4 Evidence of cell filamentation

Previous results in chapter 3 showed that the partially active Ocr multimutants that reduce the growth rate of *E. coli* NM1041 (+HsdR) caused cells to filament as a result of inducing a $m^{+/-}r^{+/-}$ phenotype. Although the POcr mutants do not display any antirestriction or antimethylation activity upon phage infection, a number of them do induce the same growth profile (Y/N) as the Ocr multimutants. Therefore evidence of cell filamentation confirms that

these POcr mutants do possess the same phenotype ($m^{+/-}r^{+/-}$) as the Ocr multimutants and hence possess a weak binding affinity towards EcoKI. Furthermore, cell morphology in the presence of POcr mutants that are toxic (Y/Y) were also assessed to see if the toxicity effect induced the bacterial SOS response/triggered cell filamentation.

The expression of each POcr mutant was induced once an OD_{600} of 0.3 was achieved, after which a further 2 hours incubation was followed by cell fixation using methanol. The cells were subsequently examined under a light and fluorescent microscope, to observe the DAPI stained nucleoid.

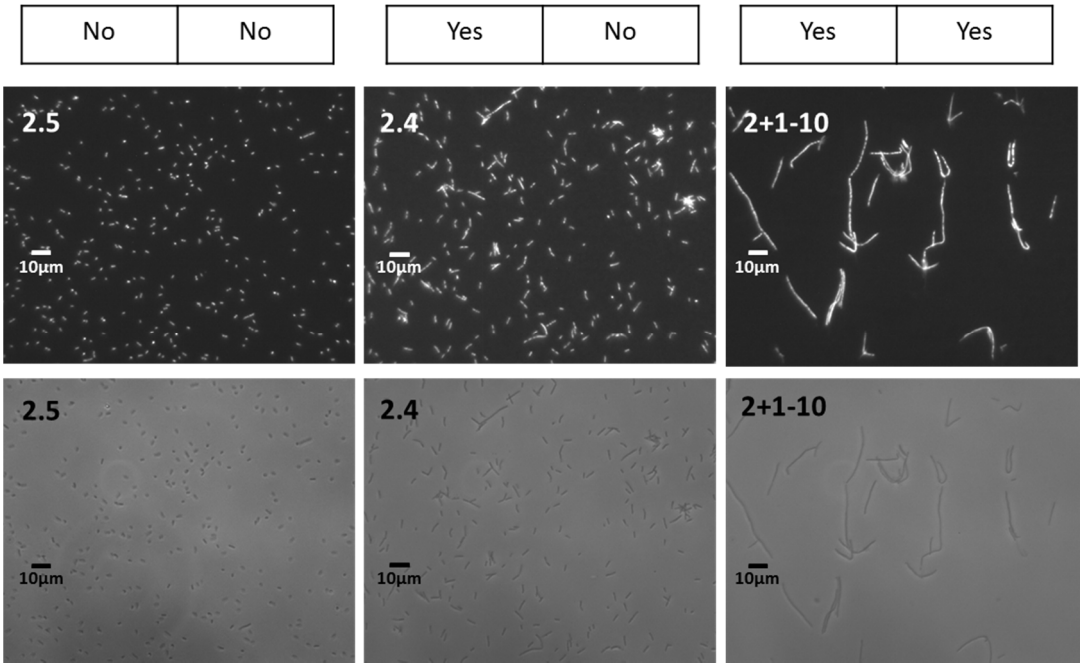


Figure 7.9: Microscopy images of *E. coli* NM1041 cells which have been induced to express mutants that exhibit the three different phenotypes that were expressed from the bacterial growth curves that are shown in Table 7.2. The mutant 2.5 induces no change (N/N), whereas mutant 2.4 induces reduced viability in the presence of the HsdR subunit (Y/N) and the mutant 2+1-10 which induces a reduction in the growth of *E. coli* cells in the presence and absence of an RM system(Y/Y).

The microscopy images in Figure 7.9 are examples of the morphology of *E. coli* NM1041 cells, where POcr mutants have been expressed produce three different growth profiles that define the partially active POcr mutants; N/N, Y/N and Y/Y. These represent the three different growth behaviors that were outlined in Figure 7.7. The figure 7.9 shows that the expression of the inactive (N/N) POcr mutant M2.5 has no effect on the morphology of *E. coli* NM1041 cells this was the case for all of the POcr mutants that were inactive.

However M2.4 represents a POcr mutant that reduces the growth of *E. coli* NM1041 cells only (Y/N, T^{HsdR} phenotype). Here the cells are longer and there are signs of cell filamentation, this is reflected for each of the POcr mutants that effect growth in the presence of EcoKI. However, the extent and size of the filaments produced differed for each POcr mutant analysed, Table 7.3.

Partially active Mutant	Filaments	Small?/Large?	Toxic?/T^{HsdR}?
1+2-8	YES	S	T^{HsdR}
Dg10	YES	L	T^{HsdR}
ED6	YES	S	T^{HsdR}
2.4	YES	S	T^{HsdR}
2+1-8	YES	S	T^{HsdR}
Dg4	YES	L	T
2+1-10	YES	L	T
2+1-1	YES	S	T
2.1	YES	S	T
G2=7	YES	S	T
G2=8	YES	L	T
2+1-6	YES	S	T
2mm	YES	L	T
Bx3/11	YES	S	T

Table 7.3: The table describes the size of filaments for *E. coli* cells that filament when the gene of a partially active POcr mutants is expressed. The size of the filaments was designated as small (S) or large (L) relative to the filaments exhibited for 2.4 and 2+1-10 respectively (figure 7.9). This was associated with the toxicity exhibited for each of the POcr mutants, where T^{HsdR} signifies cell toxicity in the presence of EcoKI whereas T was attributed to POcr mutants that appear to be toxic to all *E. coli* cells.

Lastly, the POcr mutant 2+1-10 which induced a Y/Y growth profile and is toxic to all *E. coli* cells can also cause *E. coli* NM1041 cells to filament. Furthermore, cell filamentation was observed in *E. coli* NM1057 ($HsdR^-$) cells. This shows that the toxic mutants induce cell filamentation and therefore appear to be initiating the bacterial SOS response. The production of filaments occurs as replication and growth of cells proceeds but the cells do not divide producing elongated filaments. Such behavior is typically caused by the induction of the SOS response.[93,97,125]

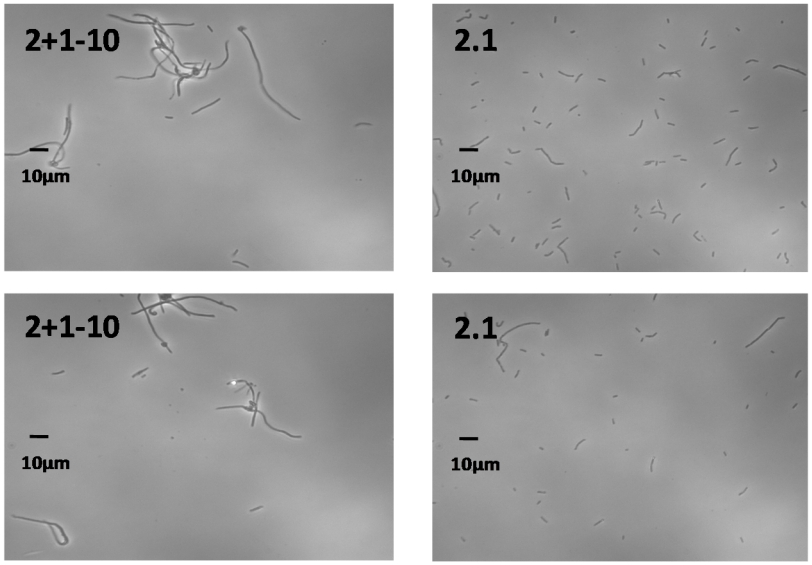


Figure 7.10: Microscope images that were captured for mutants 2+1-10 and 2.1 that were expressed in *E. coli* NM1261 cells.

The extent of filamentation as well as the size and shape of the filaments differs with the expression of each POcr mutant. No obvious trend can be identified which correlates cell filamentation to growth behavior. However, different types of filaments were demonstrated and suggest that the POcr mutants may be behaving differently or that the subtle differences in surface negative charge of the POcr mutants, has an effect on the extent of filamentation. In addition some of the cells possess circular areas of light phase at either end of each cell. These circular patches are indicative of inclusion bodies. Inclusion bodies occur when over expressed protein is unable to fold appropriately causing the protein to aggregate; insoluble protein is then encapsulated into inclusion bodies that are generally observed at the poles of the bacterial cell. [126,103] This suggests that a further distinction could be made on the solubility of the different POcr mutants and the effect this might have on the activity of the mutant and its phenotype.

7.5 Discussion

The POcr expression construct fails to produce protein in *E. coli* cells however, the introduction of 13 acidic residues causes the protein to express, be soluble and function as an antirestriction protein. This suggests that acidic residues play a role in each of these properties. The plasmid libraries that were produced in chapter 4 contained a variety of POcr mutants which covered a range of isoelectric points. Analysis of these POcr mutants could

highlight mutants that straddle each of these properties, by associating the introduction of acidic residues towards the ability of POcr to express, be soluble and function as an antirestriction protein. This should increase our understanding into the effect of the acidic residues towards the properties of the protein wtOcr.

7.5.1 Global analysis of activity

The POcr mutants of LibDg contained the largest number of mutations per POcr mutant ranging from 8-17 acidic residues. From this library four unique POcr mutants were selected containing full antirestriction and antimethylation properties (Chapters 5 and 6). However POcr mutants from this library also induced inactive and partially active mutants, presumably because the acidic residues of importance are absent from their sequence. Therefore the analysis of sequences from the library allowed direct comparisons between these POcr mutants, to highlight which acidic residues are important. Thereby producing a global map across the mutants surface, rating the importance and necessity of each acidic residue.

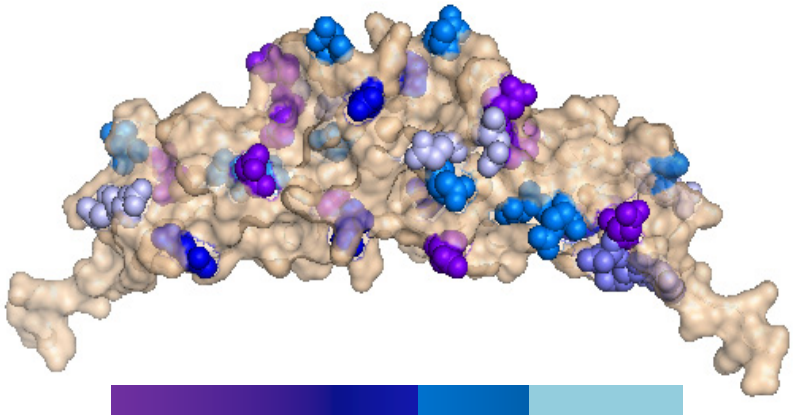
Analysis of partially active POcr mutants indicated that the acidic residues E21, D52, D74, D77 and D63, were commonly present. However, when considering what acidic residues are important for elevating partially active to fully active mutants, the residues of importance are located in the loop 2 region and an area located on Helix D of Ocr. Both of these regions were highlighted previously in chapter 3 as important for binding towards the MTase and nuclease respectively. However these residues are not conserved which suggests that these acidic residues can be interchangeable within this region. Instead conserved residues present on active and partially active POcr mutants refer to acidic residues that surround either the loop1 or loop2 region, suggesting that these residues could stabilise the loop regions and are structurally important. The importance of each acidic residue was rated by;

1) The conservation of acidic residues in active and partially active mutants, indicating their structural and functional importance

2) The acidic residues that are conserved in active POcr mutants, presumably important for biological function

3) Groups of acidic residues that are consistently present in high abundance on active and partially active POcr mutants

4) Acidic residues present on active POcr mutants in high abundance



D43, E60, D68, E88, E99
E21, D52, D77
D30, D63, D74, D93, E96
D26, D27, E65, E67
Residues which are not important:
D13, E17, D32, D33, D36, D93, E99, D100, E104, E107-8, E110-1, E113-4. D115. E116-7

Figure 7.11: Mapping the importance of acidic residues on the structure of wtOcr. Each of the acidic residues was pooled into groups depending on the importance of the residue to the activity of Ocr, which was based on the global analysis results from figures 7.4 and 7.5.

7.5.2 Partially active mutants

2AP was used to assess a selection of POcr mutants from each of the libraries that were produced in chapter 4. POcr mutants (102) were screened and subsequently categorised as either inducing an inactive or partially active phenotype. The POcr mutants contained 2-16 acidic residues producing a pI range of 10.0-3.9. Although the average number of acidic residues was 7 for inactive mutants and 9 for partially active, this marginal difference in the number of acidic residues highlights that positional context of the acidic residues is important. However, the global study carried out above does not consider why partially active mutants might contain as few as 5 acidic residues and inactive mutants as many as 13 acidic residues. As a result 42 of POcr proteins that possessed a range of pI values and demonstrated inactive and partially active phenotypes were investigated further.

An *in vivo* phage assay was performed the results showed that none of the mutants appeared to inhibit phage infection, no antirestriction or antimethylation behavior was detected. The

Ocr multimutants in chapter 3, possessed a weak binding affinity with EcoKI but provide adequate antirestriction activity in the *in vivo* phage assay. The POcr mutants do not share this activity which suggests that the partially active phenotype could encompass mutants with different activities and behaviors.

Indeed each of the partially active POcr mutants observed one of two growth patterns Y/N and Y/Y.

Y/N- The POcr mutants reduced the growth rate of *E. coli* cells in the presence of the HsdR subunit only. No change in cell growth is observed in the absence of the HsdR subunit

Y/Y- The presence of the POcr protein reduces the growth rate of *E. coli* cells in the presence or absence of the HsdR subunit

This indicates that these POcr mutants are interacting with the restriction pathway and suggests that the POcr mutants are only partially inhibiting the activity of the M.EcoKI and EcoKI enzymes, causing the cell's chromosomal DNA to become susceptible to the nuclease. This self-destructive phenomenon was confirmed by the appearance of DNA filaments which was a result of chromosomal DNA becoming susceptible to EcoKI. This phenomenon was also shown in chapter 3 for a number of the Ocr multimutants, although these mutants displayed antirestriction and antimethylation activity from phage *in vivo* experiments. This difference could be due to the instability of the POcr mutants due to the loss of ~60-86% of acidic residues.

On the other hand, several partially active POcr mutants reduced the growth of cells in the presence and absence of the EcoKI enzyme. The toxicity affect is common in the overproduction of recombinant mutants but due to the nature of the protein considered, as a DNA mimic, if mutation has not disrupted the shape and fold of the protein, it could mean that the POcr mutants are interacting with other components of the cell. This is supported by evidence of cell filamentation observed on *E. coli* cells expressing the toxic POcr mutants. The filamentation suggests the initiation of the SOS response as a result of DNA damage, which could be a result of POcr mutants interacting with other DNA–binding proteins. Although less probable, two factors increase its likelihood; the known binding of the wtOcr protein with *E. coli* RNA polymerase and the fact that one of the toxic POcr mutants Bx3.11, also possesses antirestriction and antimethylation activity, suggesting dual functionality.[127]

7.5.3 The enigma that is Bx3.11

The most curious phenotype induced by a POcr mutant was from the mutant: Bx3.11. POcr mutant Bx3.11 demonstrates full antirestriction and antimethylation activity akin to wtOcr from the *in vivo* phage assay, (incidentally the only POcr mutant to do so, with the exception of the active POcr mutants described in chapter 5 and 6). However, POcr protein Bx3,11 also induces a toxic effect in the presence and absence of EcoKI. Microscopy images of *E. coli* NM1041 cells expressing the Bx3.11 construct have shown that a number of filaments were present, although the total number and size of the filaments were smaller than those induced by other toxic mutants. Therefore although this mutant appears to demonstrate full activity *in vivo* towards phage DNA, the expression of the construct is toxic towards the bacterial cell. This could suggest that either the specificity of the mutant has decreased by interacting with other DNA binding proteins in the cell including EcoKI or that expression of the construct is unstable producing insoluble protein which is detrimental to the cell. The POcr mutant Bx3.11 contains the largest number of acidic residues of the POcr mutants analysed in this chapter, with a total number of 15 acidic residues. Therefore sequence alignment studies were carried between the active POcr mutants and the mutant Bx3.11 to ascertain why protein Bx3.11 is toxic.

```
M1.1  MAMSNMTYNNVFNHAYQMLKENIRYDNIRDTNDLHDAIHMAADNAVPHYYADIFSVMASE
M2.8  MAMSNMTYNNVFNHAYQMLKENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIFSVMASE
M1.9  MAMSNMTYNNVFNHAYQMLKENIRYNDIRDTDDLHNAIHMAADNAVPHYYADIFSVMASE
M2.1  MAMSNMTYNNVFNHAYQMLKENIRYNDIRDTNNLHNAIHMAADNAVPHYYADIFSVMASE
3.11  MAMSNMTYNNVFNHAYQMLKENIRYDNIRDTNDLHDAIHMAADNAVPHYYADIFSVMASE
      ************************::****::**:**********************

M1.1  GIDLQFEDSGLMPNTKDVIRILQARIYEQLTIDLWQDAQNLLNQYLQQVQEYQQNQQ
M2.8  GIDLQFQDSGLMPDTKDVIRILQARIYEQLTINLWENAQNLLNQYLQQVQEYQQNQQ
M1.9  GINLQFQDSGLMPDTKDVIRILQARIYEQLTIDLWQNAQNLLNQYLQQVQQYQENQQ
M2.1  GIDLEFEDSGLMPDTKDVIRILQARIYEQLTIDLWENAQNLLNQYLQQVQEYQQNQQ
3.11  GIDLQFEDSGLMPNTKDVIRILQARIYQQLTIDLWQDAQNLLNQYLQQVQEYQQNQQ
      **:*:*:******:************:****:**::************:**:***
```

Figure 7.12: Alignment studies of each of the active mutants with the mutant Bx3-11. Highlighted in red are acidic residues that are present in the protein sequence and yellow signifies where the residues has not been mutated back to an acidic residues and therefore remains a neutral residues from the original POcr template.

The alignment of the protein, Bx3.11, with each of the active POcr mutants shows no striking differences between the protein sequences. Closer inspection of the sequences shows that the active POcr protein M1.1 is nearly identical to the toxic POcr mutant Bx3.11. The exception is the absence of the acidic residue at E88 on the Bx3.11 mutant, incidentally this residue is conserved on all four active POcr mutants. This suggests that this acidic residue is

crucial for the activity of the POcr protein M1.1 and its absence causes the protein to become toxic to the cell, although seemingly does not prevent the activity of the protein in the phage assay. A closer inspection of the E88 acidic residue shows that it forms part of a hydrogen bond network with amine groups of the POcr polypeptide chain located on the loop 2 region. As such it is thought that this hydrogen bond network may stabilise the loop 2 region of the protein. The loop 2 region has been highlighted as an important interaction interface in previous chapters and this result consolidates this theory. Not only this, the EM model of the EcoKI:Ocr interaction also displays this region at the interface of the two proteins and therefore this result also adds merit to this structural model.

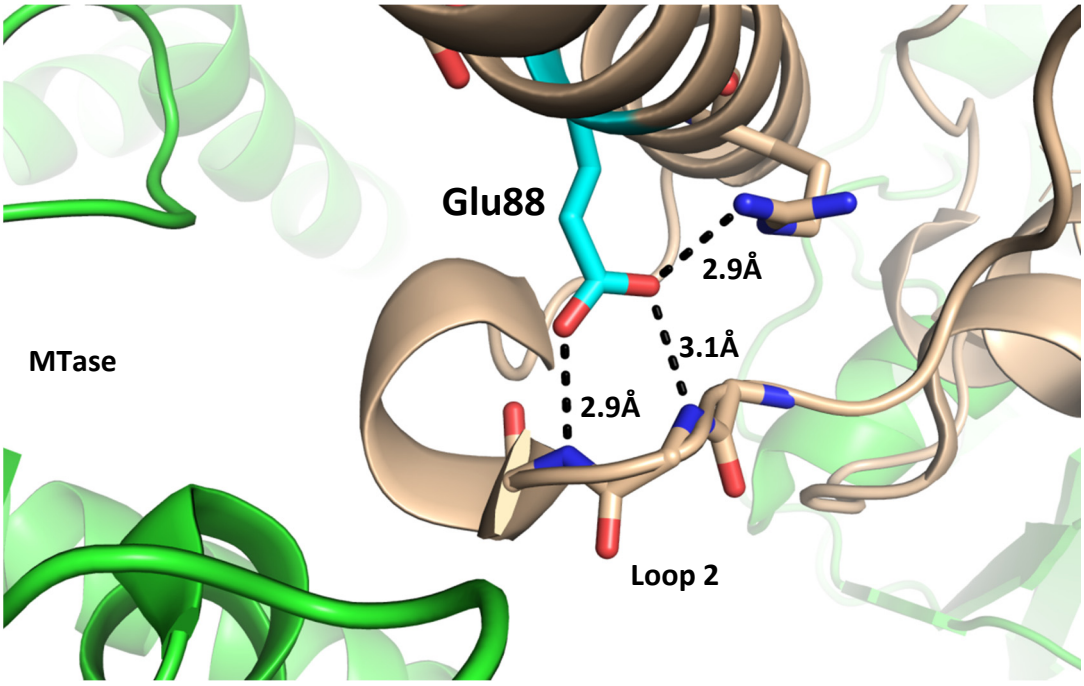


Figure 7.13: The EM structure of the Mtase:Ocr interaction. The Ocr protein is highlighted in wheat and the MTase in green. The E88 residue located on Helix D has been highlighted in cyan and the bond lengths of its side chain with the amine groups of the main peptide chain of the loop 2 region have been displayed using sticks and black dotted lines. The bond lengths are all ~3Å in length therefore could be possible hydrogen bonds.

It is thought that the neutralisation of the E88 residue may destabilise the loop 2 region therefore producing a much more flexible interaction site which can therefore target other DNA-binding proteins as well as EcoKI. The specificity of the interaction between wtOcr and EcoKI is thought to be governed by the mimicry of charge and the prebent shape of the

protein. The changes introduced on the Bx3.11 mutant could change the shape of the interface, decreasing the specificity of the POcr protein to interact as a DNA mimic towards other DNA/protein interactions, of which many distort/bend DNA upon binding.

To conclude, the analysis of many POcr mutants with differing pI values indicate the importance of acidic residues to the antirestriction activity of POcr. The acidic residues were ranked producing a map of POcr that highlighted which acidic residues were most important and possibly what role they might possess. In addition, POcr mutants originally characterised as partially active were revealed to induce several types of behavior; they either exhibited some antirestriction activity or were toxic in *E. coli* cells. Analysis could not relate particular acidic residues with either phenotype but the formation of filaments indicated that the toxic mutants may provide promiscuous binding activity that subsequently induces the SOS response. Finally the mutant Bx3.11 possess its own unique behavior, although it binds to the EcoKI enzyme according to the phage infection assays (table 7.2), it is also toxic to *E. coli* cells. Furthermore its protein sequence shows that it is identical to the active POcr protein M1.1, with the exception of E88Q which is thought to reduce the specificity of Bx3.11 which results in the toxic effect.

Chapter 8. Mapping charge-function relationships

In this study so far, each chapter has investigated the charge to function relationships of the Ocr homodimer using several different strategies.

The results from each of the chapters has built on understanding the character of the protein Ocr; which is essentially a robust protein with a large number of non-essential acidic residues. However certain sequences that appear to follow no apparent trend in pI or hotspot coverage can initiate deleterious effects in the host bacterial cell. Therefore this chapter shall focus on aggregating all of the results to compile a profile of Ocr going from an active protein to an inactive protein.

In order to do this several aspects associated with the charge and activities of the protein were analysed:

1. The relationship between the total number of acidic residues and the isoelectric point for all the POcr and Ocr mutants analysed.

2. The importance of acidic residues residing in the loop regions compared to non loop regions for the activity of the Ocr/POcr mutants

3. The relationship between activity and acidic residues that reside in the loop region, the pI and acidic residues important for the structural integrity of the protein.

Finally this chapter will present overall thoughts about the study, further work that is required and overall conclusions.

8.1 Charge, residues and activity

The number of acidic residues present on a POcr mutant should relate to its pI value. The activity of the Ocr protein is dependent on its ability to mimic DNA; therefore a decrease in the number of acidic residues should lead to a loss of activity. Although this has been demonstrated previously in chapter 7, by analysing all of the mutants from chapters 3-7 the relationships between the active, partially active and inactive mutants can be mapped out in accordance with their pI value.

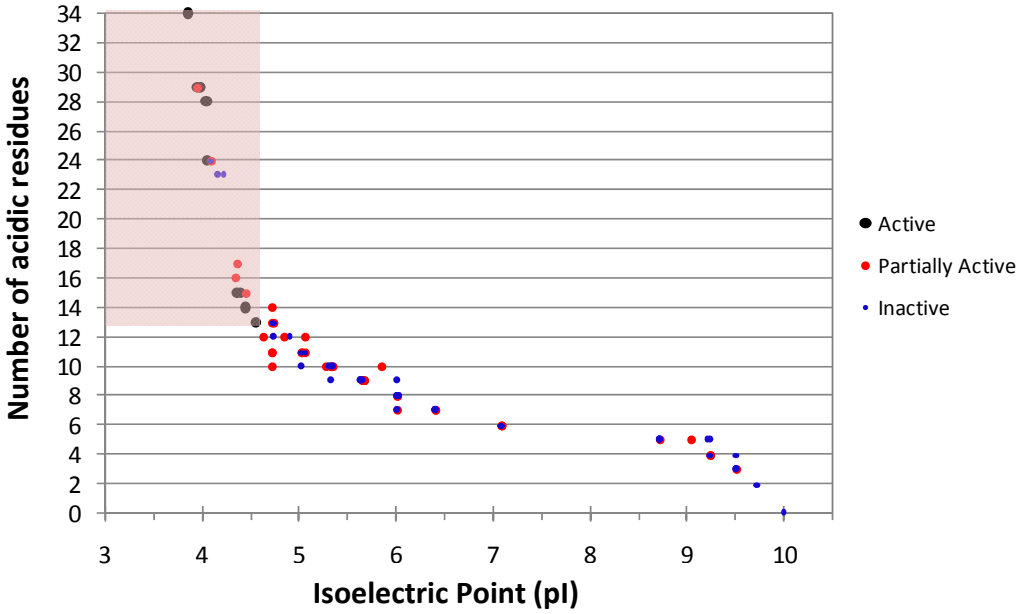


Figure 8.1: Mutants of POcr and Ocr that were studied in each of the chapter 3-7 have been analysed according to their theoretical pI number (calculated by Expasy pI calculator) and the number of acidic residues that the mutant contains. 102 mutants were analysed and have been plotted according to their activity as a result of the 2AP RA assay, where; black = mutants that demonstrated full 2AP resistance (80µg/ml) and named active; Red = mutants that demonstrate deleterious effects in the assay were named partially active; Blue = mutants that had a 2AP resistance of 20µg/ml and so were inactive. The transparent red box highlights all of the mutants that have a smaller pI value and a higher number of acidic residues then most positively charged active POcr mutant (M1.9) that functions like wtOcr.

The Figure 8.1 above shows that the activity of the POcr/Ocr proteins increases with a reduction in pI and an increase in the number of acidic residues. Active mutants appear to have the largest number of acidic residues and the lowest isoelectric point; these have been highlighted in the red oblong in figure 8.1. However there are a number of inactive and partially active mutants which are also in this range, these mutants also possess a large number of acidic residues and low pI values but are not fully active antirestriction proteins. This highlights that the amount of charge on the surface of the protein is not the only factor that governs the activity of Ocr.

The inactive mutants are distributed widely with a pI range of 4.0-10.0 with 0-13 acidic residues. This again demonstrates that pI and the sheer number of acidic residues present on an Ocr/POcr mutant does not allow it to function as an antirestriction protein.

The distribution of the partially active mutants shows a pI range of 3.9-9.5 and a maximum of 29 acidic residues to a minimum of 3 residues. Although a large distribution is apparent

the majority of partially active mutants are concentrated at a pI range of 4.0-6.0. Although the distribution of partially active mutants seems to span the pI range of POcr to Ocr, results in chapter 7 revealed that the partially active phenotype harbored several different phenotypes, figure 8.2. These included; POcr mutants that were toxic to all *E. coli* cells (yellow in figure 8.2) and those specifically toxic to cells with a Type I RM system (green in figure 8.2). As a result the partially active mutants were re-categorised taking into account these two new phenotypes that were observed in chapter 7, Y/Y and Y/N respectively.

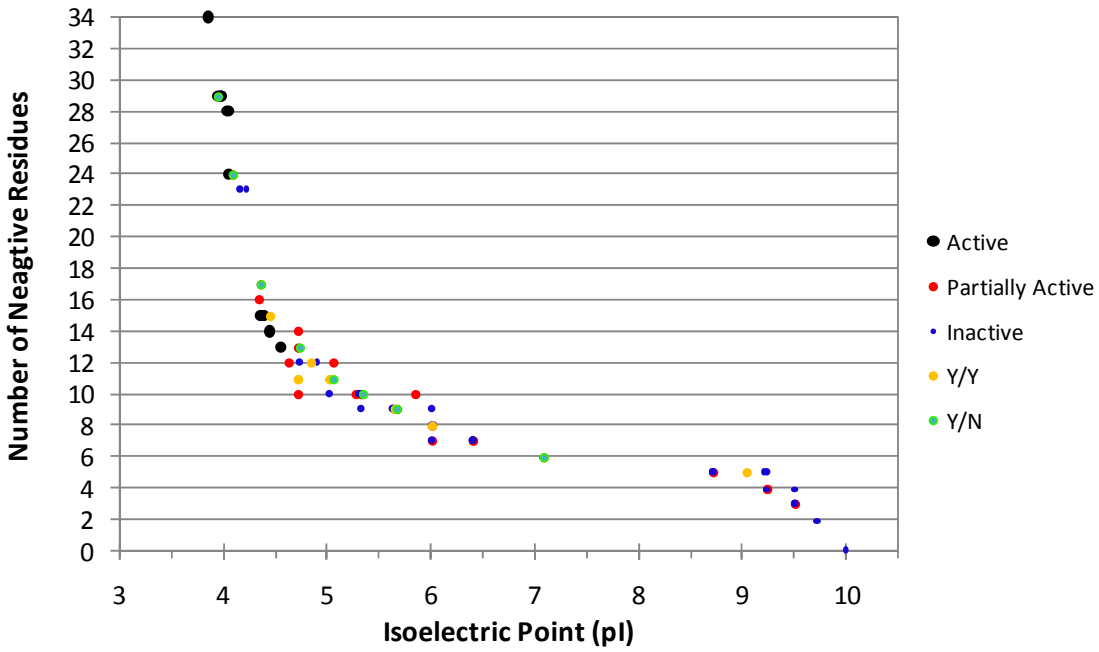


Figure 8.2: Mutants of POcr and Ocr that were studied in each of the chapter 3-7 have been analysed according to their theoretical pI number (calculated by Expasy pI calculator) and the number of acidic residues that the mutant contains. 102 mutants were analysed and have been plotted according to their activity as a result of the 2AP RA assay, where; black = mutants that demonstrated full 2AP resistance (80µg/ml) and named active; red = mutants that demonstrate deleterious effects in the assay were named partially active; blue = mutants that had a 2AP resistance of 20µg/ml and so were inactive. In green and yellow are mutants that were examined in chapter 3 and 7 and revealed a number of partially active mutants; that were toxic in *E. coli* cells in the presence of the HsdR subunits (green) and other mutants that were toxic to all cells (yellow).

The results showed no clear relationship between the pI value and the number of acidic residues on the degree of toxicity induced by the mutants. However, no toxic mutants possessed over 15 acidic residues and a pI value less than 4.5. Instead the partially active mutants that are above this threshold were toxic in the presence of the HsdR subunit only. From chapter 3 and 7 it was shown that these mutants had a weak binding affinity towards

Type I RM systems producing an $M^{+/-}R^{+/-}$ phenotype. The only toxic mutant that appeared to be in the same range as the active mutants was the POcr mutant Bx3.11. Analysis in chapter 7 showed that this mutant is identical to the active POcr mutant M1.1 with the exception of the acidic residue E88, which has presumably lead to instability and/or promiscuous activity with other DNA binding proteins.

Analysis of the active mutants shows that a minimum of 38% of the acidic residues and a pI of 4.5 is all that is required to maintain full antirestriction and antimethylation activity. However there are exceptions, a number of inactive mutants and partially active mutants have the minimum requirements for activity but they do not retain or only partially retain the activity of the Ocr protein. Sequence alignment of these mutants shows that acidic residues are reduced in the loop1 or loop 2 regions. This confirms that the number of acidic residues and the pI are important however they are not the only parameters that govern the activity of the protein and the position of the acidic resides are also important.

8.2 Loop region versus non-loop region

The analysis above and the results in chapter 3 showed that the loop regions are important for the activity of the protein. Therefore the relationship between the activity of the proteins was analysed with the % of acidic residues that were present in Loop regions and non-Loop regions which were defined as; Loop Regions= Loop1 aa25-33, Loop2 aa58-77 and Non-Loop Regions = Helix A aa7-24, Helix B aa34-44, Helix C aa49-57 and Helix D aa78-106.

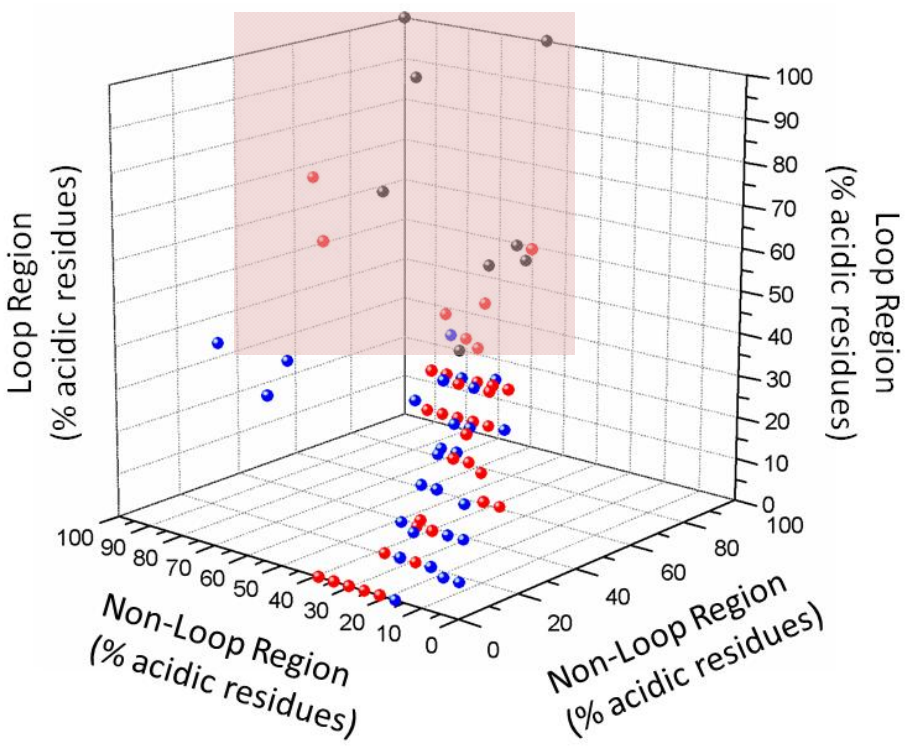


Figure 8.3: Mutants of POcr and Ocr that were studied in each of the chapter 3-7 have been analysed according to the percentage number of residues that were found in the loop regions (aa25-33 and aa58-77) and non loop regions (aa7-24, 34-44, 49-57 and 78-106). The same mutants that were assorted in figure 8.1 were used within the analysis and plotted according to their activity as a result of the 2AP RA assay, where; black = mutants that demonstrated full 2AP resistance (80µg/ml) and named active; Red = mutants that demonstrate deleterious effects in the assay were named partially active; Blue = mutants that had a 2AP resistance of 20µg/ml and so were inactive. The transparent red box highlights all of the mutants that have a greater ratio of loop: non loop acidic residues then most positively charged active POcr mutant (M1.9) that functions like wtOcr.

It is apparent that the proportion of acidic residues in the loop region compared to the non-loop region is important to the activity of the mutants. The results do not take account of acidic residues that reside on the tail region of POcr or Ocr mutants, as it was found that these residues are not important towards the antirestriction activity of the mutants. The results in Figure 8.3 above showed that the minimum requirement of acidic residues for an active mutant contained a ratio of 40%:35%, Loop: Non-loop; this can be reduced to 30% of the acidic residues in the non-loop region however this is counterbalanced by the enrichment of the loop region to ~60%.

Although full activity was observed for POcr mutant M1.9, at a ratio of 40%:35% Loop: Non-Loop, a number of Ocr/POcr mutants, that have been highlighted by the red box in

Figure 8.3, possesses the minimum requirements for activity but do not retain full activity. These mutants will be looked at in more detail and compared to the active POcr mutant, M1.9.

Mutant	Activity	Acidic Residues	Isoelectric point	% Loop Region	% Non Loop	N° of Conserved Residues
s1.9	**A**	**13**	**4.56**	**42**	**35**	**7**
C2	IN	13	4.75	44	39	4
M12	PA (T^{HsdR})	23	3.96	67	100	5
M16	PA (T^{HsdR})	24	4.1	56	87	4
Dg7	PA (T^{HsdR})	12	4.63	44	35	5
Dg11	PA (T^{HsdR})	16	4.35	48	43	3
2+1-8	PA (T)	11	4.72	51	35	2

Table 8.1: POcr and Ocr multimutants that are either inactive or partially active and represented in the red box in figure 8.2 as retaining properties that exceeded the active mutant M1.9. The properties that were analysed in figure 8.1 and 8.2 have been compared the properties of the active mutant that contains the minimum number of protein that are required for full activity where A, IN and PA identify active, inactive and partially active phenotypes according to the 2AP assay and T represents mutants that induce a toxic effect with THsdR signifying mutants that induce a toxic effect only in the presence of the HsdR subunit.

The table shows that each of the POcr mutants have similar properties and contain acidic residues that exceed the numbers found in the active mutant M1.9. However expression of these mutants induces a different phenotype in the 2AP assay. Phage infection showed that the mutants M1.9, M12 and M16 all exhibited full activity but, *in vitro* studies in chapter 3 showed that the Ocr multimutants M12 and M16 possessed a weak binding affinity to M.EcoKI, whereas M1.9 was a tight binding enzyme. The mutant C2 was inactive from all the assays carried out and the mutant 2+1-8 was toxic to all *E. coli* cells. This leaves the mutants Dg7 and Dg11 which were deleterious in the presence of the HsdR subunit and inactive in the phage assay, it is thought that this phenotype is also a result of a weak binding affinity to EcoKI as with the Ocr multimutants M12 and M16.

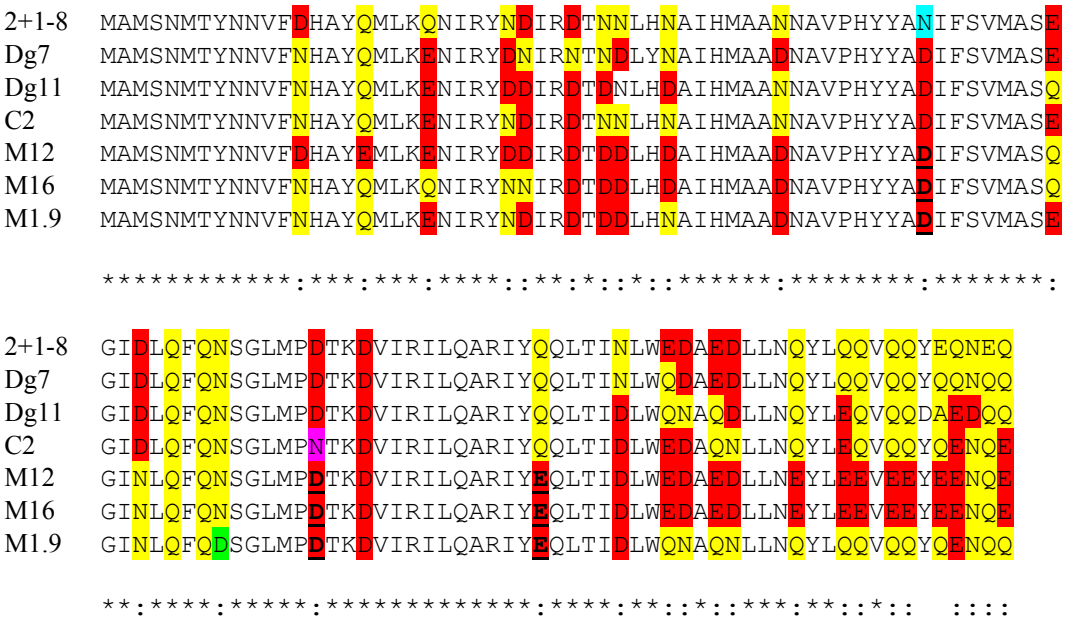


Figure 8.4: Sequence alignment (carried out by Clustal Omega software) of the mutants in table 8.1. Many of these mutants exhibit similar properties (i.e. pI, acidic residues in the loop region) as the active POcr mutant M1.9, however they do not induce full antirestriction activity. In blue is the neutral residue N52 that is conserved only on mutant 2+1-8 which is toxic. In green the acidic residue D68 that is conserved on the active mutant M1.9 only and in pink the neutral residue N74, conserved only on the inactive mutant C2. The acidic residues that are underlined in bold; D52, D74 and E88 show a triplet of residues that are only conserved on the mutants M12, M16 and M1.9.

The alignment of the sequences does not show a clear pattern as to why each of the mutants are similar and yet display different antirestriction and antimodification activities. However a closer inspection of the sequences does reveal some subtle differences between each of the mutants. The only clear difference between the active mutant M1.9 and the other sequences is the absence of the acidic residue at D68 (highlighted in green). This suggests that this particular residue might be important to the activity of Ocr. In fact, this residue is conserved among each of the active POcr mutants. Each of the other mutants appear to be missing at least three of the conserved residues that were discovered in chapter 5 (E21, D43, D52, E60, D68, D77 and E88). Notably a number of these conserved residues are not present in either of the loop regions.

However, comparisons between the inactive mutant C2 and each of the other mutants showed that C2 was the only mutant that did not possess an acidic residue at position 74. N74, highlighted in magenta, appears to be conserved only in the inactive C2 mutant.

Sequence alignment comparisons did not reveal any other clear differences between the inactive mutant C2 and the other mutants although the loop2 region contains only one acidic residue. This suggests that the absence of the acidic residue D74 and the absence of acidic residues in the loop 2 region are detrimental to the activity of the protein.

The properties outlined in Table 8.1 for the inactive mutant C2 shows that it is similar to the mutant 2+1-8 which is toxic to all cells. These mutants possess the highest pI values of the mutants analysed and both contain a similar number of acidic residues in the Non-Loop region. Alignment of the toxic mutant 2+1-8 showed that the neutral residue N52 is exclusively conserved on 2+1-8, which implies that the absence of an acidic residue at this position seems to destabilise the protein. As a result this residue appears to be structurally important as its absence induces a toxic effect. Furthermore, the mutant 2+1-8 only contains 2 of the seven acidic residues that were found to be conserved among each of the active mutants in chapter 6.

The remaining mutants include Dg7 and Dg11, which exhibit partial activity in the 2AP plate assay but are inactive in the phage infection experiments, and the mutants M12 and M16, which are also partially active in the 2AP assay but active in the phage infection assay. It was more difficult to be able to define the residues responsible for each of the phenotypes exhibited as they closely resembled each other. However by looking at groups of residues, in particular at residues that were already highlighted as absent from the other phenotypes (i.e., D52 and D74), the acidic residues D52, D74 and E88 were seen to be conserved in the mutants M12 and M16. The mutants Dg7 and Dg11 also contain the acidic residues D52 and D74 but not the residue E88; this seems to have reduced the activity of these mutants in comparison to the mutants M12 and M16.

Analysis of each of the sequences in Figure 8.4 can reveal differences which enables differentiation between the different phenotypes induced by the mutants. However the differences are hard to define without looking at each sequence carefully. This could be due to the small subset of potential POcr/Ocr multimutants that have been analysed. Perhaps with a larger number of POcr/Ocr proteins, the trends and relationships between activity and charge may be more obvious.

8.3 Structural residues versus loop region residues versus pI

Overall the results above suggest that a number of specific acidic residues may contribute to more than just the electrostatic interactions of the Ocr protein towards Type I RM systems. The absence of acidic residues appears to affect the solubility of the mutants, for example, in chapter 7 the induction of some mutants led to the formation of inclusion bodies. [105] By aligning the sequences of Ocr/POcr mutants that are known to be soluble, the conservation of acidic residues should reveal residues important to the structure and stability of the Ocr protein.

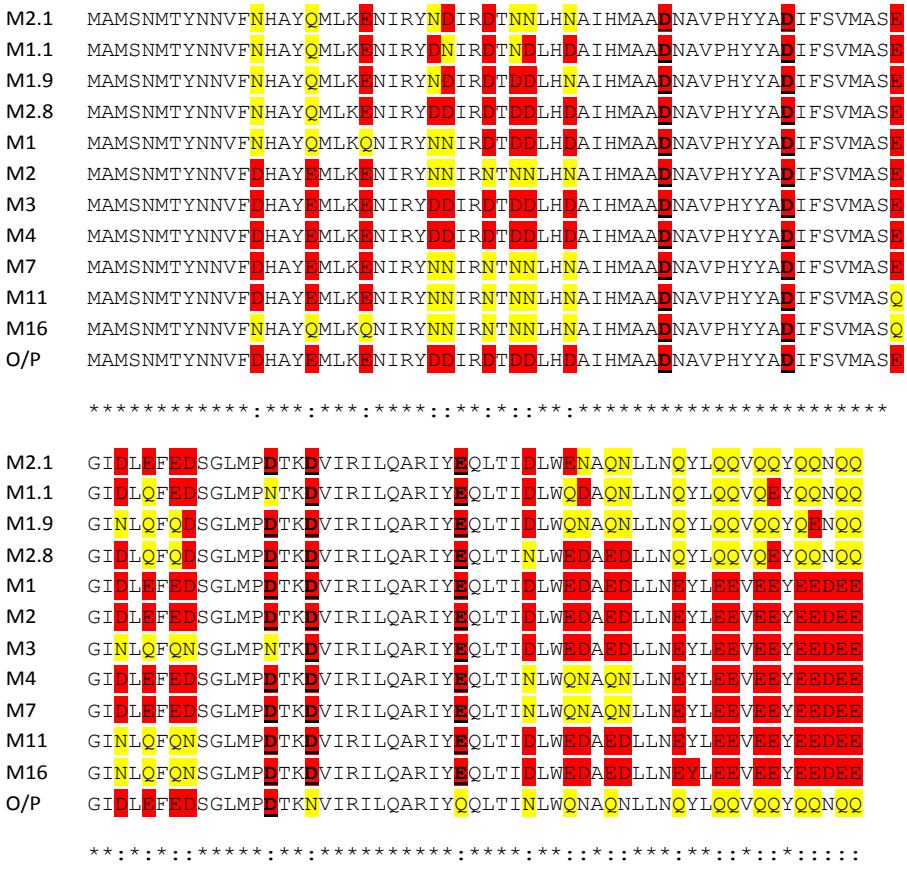


Figure 8.5: Sequence alignment (carried out by Clustal Omega software) of the soluble mutants from chapter 3 and chapter 6. Highlighted in red are all the acidic residues that have been reintroduced onto the mutants and in yellow are the remaining target sites where acidic residues are present in wtOcr from T7 bacteriophage. Underlined are residues that are conserved or almost conserved in each of the mutants, it is possible that these mutants may be structurally important. O/P signifies the Ocr multimutant Ocr/POcr studied in chapter 3.

The acidic residues that are almost completely conserved in the alignment study in Figure 8.5 were; D43, D52, D74, D77 and E88. None of these residues are involved in the loop

regions. As such these residues would not have been considered in the previous analysis in figure 8.3, which showed the relationship between loop and non-loop regions. However the fact that these acidic residues are conserved or almost conserved among soluble Ocr/POcr mutants implies their relevance to the structurally integrity of the Ocr protein. Furthermore, alignment of the mutants in figure 8.5 with Ocr proteins from other phage entities, shows that the residues D77 and E88 were conserved and the remaining target sites were either almost conserved (D43) or had conserved acidic residues, therefore either D or E residues at the following target sites; D/E52 and D/E74.

Considering all of the findings so far, the relationship between negative charge and the function of the protein appears to follow three stages of development. Firstly an adequate number of acidic residues are required to produce a low pI value, secondly, the concentration of negative charge in the loop regions needs to be adequate and thirdly, the presence of acidic residues which appear to stabilise and allow the protein to fold adequately. The following three factors were analysed; the number of acidic residues, the % number of residues that reside in the loop region and the % of residues present which are thought to be structurally important to the protein.

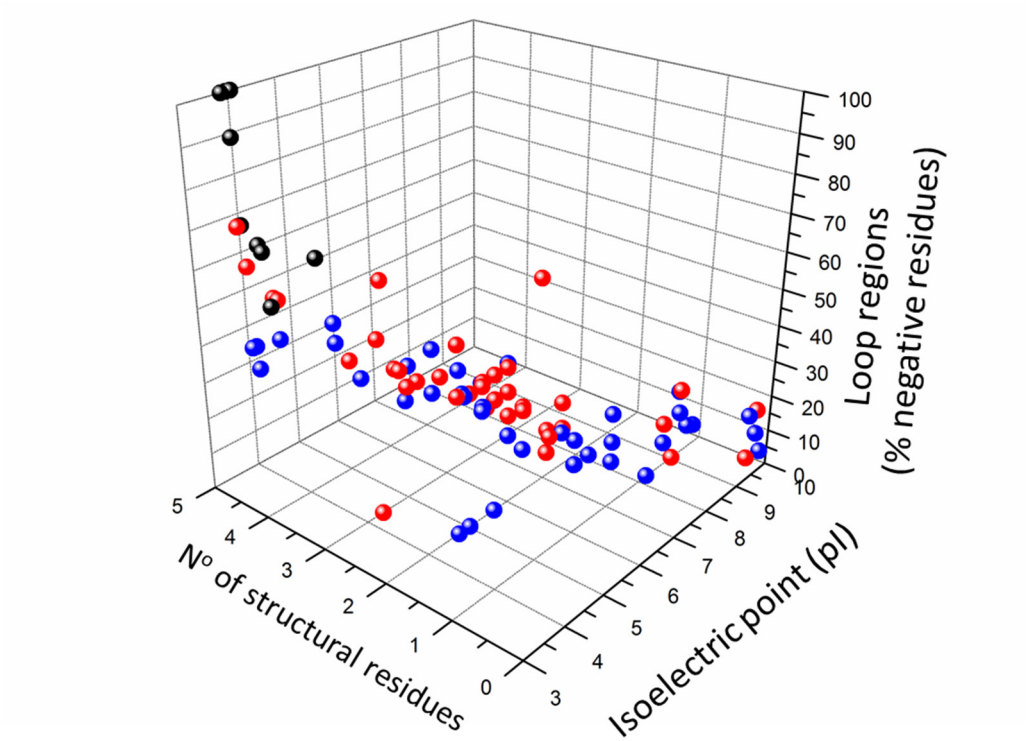


Figure 8.6: Mutants of POcr and Ocr that were studied in chapter 3-7 have been analysed according to the percentage number of residues that were found in the loop regions (aa25-33 and aa58-77) and the theoretical isoelectric point (calculated on Expasy protein calculator) and structurally relevant residues (D43, D52, D74, D77 and E88) The same mutants that were assorted in figure 8.1 and figure 8.2 were used within the analysis and plotted according to their activity as a result of the 2AP RA assay, where; black = mutants that demonstrated full 2AP resistance (80µg/ml) and named active; Red = mutants that demonstrate deleterious effects in the assay were named partially active; Blue = mutants that had a 2AP resistance of 20µg/ml and so were inactive.

From figure 8.6 it is clear that each of the active mutants possesses a low pI number, large % of acidic residues in the loop regions and most if not all of the structurally important residues are present. The graph does illustrate the importance of these three parameters for the active mutants and that the minimum requirements for an active mutant include; a pI value of 4.5, at least 40% of the acidic residues in the loop regions, and 4 out of 5 structurally-relevant acidic residues. Most importantly the properties of the inactive mutants appear to be consistently lower than the fully active mutants. In addition there were only a number of partially active mutants that had the same properties as the active mutants. These were the Ocr mutants, M12 andM16. However these mutants lacked a key feature, namely, 100% of the acidic residues in the midloop region were absent from both mutants.

The partially active and inactive mutants are distributed across a narrow plane of the potential sequence space provided by each of the parameters investigated. The partially active mutants appear to be concentrated between the two extremes of the graph; possessing a pI of 5-6, an average of 3 structural residues and between 20-30% of the acidic residues in the loop regions.

Nevertheless, a number of exceptions do not appear to follow this general pattern, this could be because these mutants have not been fully characterised. In light of the results from chapter 7 a number of unclear phenotypes emerged that suggested that the partially active phenotype encompassed a multitude of activities which included, figure 8.7; toxic mutants (yellow), mutants that appear to be partially active but fail to prevent the restriction of phage DNA (magenta), partially active mutants which are toxic (brown) and partially active mutants (green), this would make the relationship between phenotype and charge more difficult to discern.

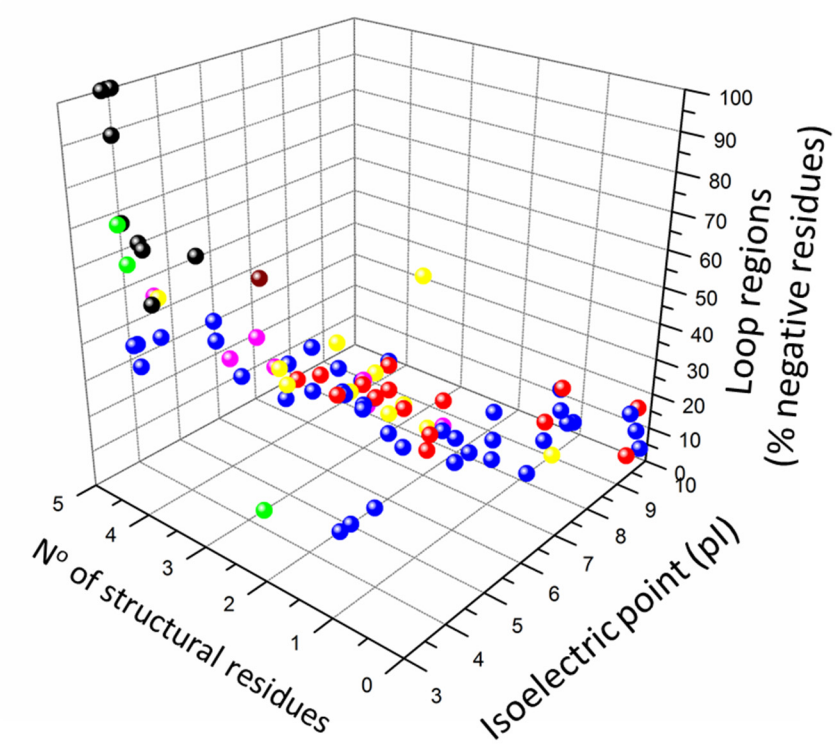


Figure 8.7: Mutants of POcr and Ocr that were studied in chapter 3-7 have been analysed according to the percentage number of residues that were found in the loop regions (aa25-33 and aa58-77) and the theoretical isoelectric point (calculated on Expasy protein calculator) and structurally relevant residues (D43, D52, D74, D77 and E88) The figure is the same as figure 8.6 but the partially active mutants in red have been subdivided into; Yellow, mutants that are toxic to all *E. coli* cells; Green, mutants that exhibit the $M^{+/-}R^{+/-}$ phenotype; Magenta, mutants that are partially active from the 2AP assay but are inactive from the phage assay and dark red, mutants that are partially active and toxic. Highlighted in red were the remaining partially active mutants of the 2AP assay that were not further characterised the other mutants exhibited in black and blue were as previously described; black = mutants that demonstrated full 2AP resistance (80µg/ml) and named active and Blue = mutants that had a 2AP resistance of 20µg/ml and so were inactive.

The graph above tries to differentiate the partially active mutants further. No clear relationship can be resolved between the different phenotypes and the three properties that have been analysed in figure 8.6. However the results do show that those mutants that are partially active (green) possess a low pI value and a large number of structurally relevant acidic residues however, the number of acidic residues in either or both loop regions is diminished. These mutants express protein and even provide antirestriction activity towards phage, however, upon over expression these mutants, in the presence of EcoKI, produce toxic effects in *E. coli*. The mutant Bx3.11 (dark red) induces the same behaviour, however, a reduction in the total number of acidic residues and the absence of the structurally relevant

E88 residue means that this protein is active for phage infection but toxic to *E. coli*. It is more difficult to distinguish between the toxic, inactive POcr mutants in yellow and the partially active mutants in magenta. Unlike the Y/N Ocr multimutants (green partially active mutants) these POcr mutants fail to inhibit the restriction of phage DNA.

8.4 General discussion

To summarise, the analysis carried out in this study has started to piece together the *in vitro* evolution of the Ocr protein. The results have shown that the location as well as the number of acidic residues present on the wtOcr protein is important for antirestriction activity. The acidic residues that reside on the loop regions of the protein appear to be particularly important in the electrostatic interaction of the protein with its binding partner, EcoKI (Type I RM systems). The results show that within these two loop regions, only a few acidic residues need to be present to elicit a fully functional response. Acidic residues within the loop regions that are conserved among the active POcr mutants only include E60 and D68 which reside in loop2 and no residues were conserved in the loop 1 region.

Mutations in Loop 1 in conjunction with either the Loop 2 region or a second region (M4) located on Helix D can completely abolish the activity of Ocr. The M4 region, aa93-100, has a similar effect to the Loop 1 region; results from chapter 3 suggest that the absence of acidic residues in this region itself had no effect on the antirestriction activity of Ocr, however, in combination with the neutralisation of either the Loop 1 or Loop2 region, activity was completely lost. Furthermore *in vitro* characterisation of M7 (neutralisation of both the Loop1 and M4 regions) showed that antirestriction activity was reduced however antimethylation activity was still fully functional. This was confirmed by the reduced antirestriction activity exhibited by the active POcr proteins, analysis showed that a maximum of only 20% of the acidic residues in the M4 region were present on each of the active POcr proteins. However, this had no noticeable effect on the activity of the POcr mutants *in vivo* and does not affect the binding interaction of the protein to the M.EcoKI enzyme. The differentiation *in vitro* between the antirestriction and antimethylation activity displayed for each of the active POcr proteins implies that there might be a secondary binding site located on the Ocr protein (at aa93-100), which binds to the HsdR subunit producing a more stable interaction.

The selection of the minimum number of acidic residues that retained the full activity of the Ocr protein in chapter 5 demonstrated that a loss of over 60% of the acidic residues on the Ocr protein, had no effect on the activity of the protein. In addition truncations of the Ocr

protein showed that the loss of the negatively enriched C-terminal region (10 acidic residues) had no effect on activity. The C-terminal region contains over 40% of acidic residues on the Ocr protein.

Sequence alignments studies of POcr mutants in chapter 7 revealed the importance of acidic residues towards the structural integrity of the protein. The most important finding in chapter 7 was the observation of toxic POcr mutants. The most peculiar of these was Bx3.11 which is; fully active from phage infection assays, induces partial activity from the 2AP assay but was toxic to all *E. coli* cells. This demonstrated a new phenotype where the toxicity effect did not affect its antirestriction activity towards phage DNA. Sequence analysis reveals that the POcr mutant Bx3.11 has a protein sequence identical to the active POcr mutant M1.1, with the exception of the acidic residue E88 which is Q88 on Bx3-11. This shows that reducing the charge of one acidic residue has destabilised the protein, the absence of E88 appears to disrupt a hydrogen bond network which stablises the loop 2 region. Evidence of filamentation suggests that this disruption has initiated the SOS response which could inadvertently be due to promiscuous activity of the mutant Bx3.11 with other DNA binding proteins. [103,102,128,129] Notably, in 1974 D.Ratner et al[127], discovered that the Ocr protein had a strong interaction with *E. coli* RNA polymerase (RNAP). As a RNA binding protein it has many mechanistic similarities to EcoKI, both contain helicase domains and Mg^{2+} binding sites. However, studies need to be carried out to ascertain if the POcr proteins do indeed provide any promiscuous activity.

Overall the results from chapter 7 highlighted that a number of acidic residues that were not located in the loop region were important for the structural integrity of the protein. This has been further recognised by the sequence alignment in figure 8.4 which indicated that the mutants D43, D52, D74, D77 and E88 were likely to be these structural residues.

8.5 Overall conclusions and looking ahead

To conclude, this study has shown that the number of acidic residues in the loop region, a low pI value and the number of structurally relevant proteins are important in producing an effective antirestriction protein. The investigations have shown that many of the acidic residues on the surface of Ocr appear to be redundant for the function of the protein. The conservation of over 60% of the acidic residues on Ocr proteins from different bacteriophage appear to have no obvious selective advantage, with the revelation that 13 out of a possible 34 residues is the minimum number required to retain full activity of the protein Ocr.

Yet the Ocr proteins have evolved with many of these supposedly useless acidic residues conserved. Therefore it is likely that the Ocr protein has not evolved from a neutral protein scaffold and gradually selected for acidic residues but has evolved from a protein already enriched with a large number of acidic residues. The *in vitro* evolution of negative charge on the protein Ocr shows that the loss of charge in the loop regions induces a partially active mutant that is detrimental to *E. coli* in the presence of a Type I RM system. Furthermore, the acidic residues in the M4 region possibly form a secondary interaction hotspot specific to the restriction subunit.

However further investigations need to be performed to assess the effect of charge on the solubility of the protein by carrying out expression trials and possibly by tagging the protein with a fluorescent tag and identifying the production of inclusion bodies. In addition the investigations carried out in this study only managed to elucidate four unique POcr mutants that were active and on the borderline of activity. Further screening of the library LibDg could hopefully reveal more active mutants; this would provide more information and a clearer link between activity and charge.

In addition, further investigations on whether the residues within the loop regions are interchangeable with one another could be carried out using the active POcr mutants such as mutant M2.1.

More long term studies would include identifying if the POcr mutants possess promiscuous interactions with other DNA binding proteins. This could be done by carrying out pull down assays from cell extracts or by attempting to purify the proteins. Due to their toxicity/solubility these mutants could be tagged to help purify the protein. Genetic knockout assays could be designed to pinpoint other interactions.

Being able to understand how the charge of the protein can be used to change the selectivity of a protein like Ocr could lead to the production of tailor-made inhibitors for DNA processes. This could be a useful tool/probe as well as potentially be used diagnostically and therapeutically.

REFERENCES

1. Murray NE. Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol,* **64**, 412–34 (2000).
2. Tock MR and Dryden DTF. The biology of restriction and anti-restriction. *Curr Opin Microbiol,* **8**, 466–72 (2005).
3. Murray NE and Blakely GW. DNA Restriction and Modification. *Encyclopedia of Microbiology* Elsevier, 3, 538-549 (2009).
4. Bheemanaik S, Reddy YVR and Rao DN Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochem J,* **399**, 177–90 (2006).
5. Murray NE. Immigration control of DNA in bacteria : self versus non-self. *Microbiol,* **148**, 3–20 (2002).
6. Meisel A, Mackeldanz P, Bickle TA, Krüger DH and Schroeder C. Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO* **14**, 2958–66 (1995).
7. Bourniquel AA and Bickle TA. Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* **84**, 1047–59 (2002).
8. Stewart FJ, Panne D, Bickle TA and Raleigh EA. Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J Mol Biol,* **298**, 611–22 (2000).
9. Halford SE and Marko JF. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res,* **32**, 3040–52 (2004).
10. Loenen WAM. Tracking EcoKI and DNA fifty years on: a golden story full of surprises. *Nucleic Acids Res*, **31**, 7059-69 (2003).
11. Jindrova E, Schmid-Nuoffer S, Hamburger F, Janscak P and Bickle TA. On the DNA cleavage mechanism of Type I restriction enzymes. *Nucleic Acids Res,* **33**, 1760–6 (2005).
12. Atanasiu C, Byron O, McMiken H, Sturrock SS. and Dryden DT F. Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7. *Nucleic Acids Res,* **29**, 3059–3068 (2001).
13. Bickle TA and Krüger DH. Biology of DNA restriction. *Microbiol Rev* **57**, 434–50 (1993).
14. Taylor IA, Davis KG, Watts D and Kneale GG. DNA-binding induces a major structural transition in a type I methyltransferase. *EMBO,* **13**, 5772–8 (1994).

15. Wilson GG and Murray NE, Restriction and modification systems. *Annu Rev Gene,* **25,** 585–627 (1991).
16. Willcock DF, Dryden DTF and Murray NE. A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO,* **13**, 3902–8 (1994).
17. Cooper LP and Dryden DTF. The domains of a type I DNA methyltransferase. Interactions and role in recognition of DNA methylation. *J Mol Biol,* **236**, 1011–21 (1994).
18. Calisto BM, Pich OQ, Pinol J, Fit F, Querol E and Caperna X, Crystal structure of a putative type I restriction-modification S subunit from *Mycoplasma genitalium*. *J Mol Biol,* **351**, 749–62 (2005).
19. Gao P, Tang Q, An X, Yan X and Liang D. Structure of HsdS subunit from *Thermoanaerobacter tengcongensis* sheds lights on mechanism of dynamic opening and closing of type I methyltransferase. *PloS one,* **6**, 173-46 (2011).
20. Kim YI, Levchenko I, Fraczkowska K, Woodruff RV, Sauer RT and Baker TA. Molecular determinants of complex formation between Clp / Hsp100 ATPases and the ClpP peptidase. *Nat Struct Biol*, **8**, 268–271 (2001).
21. Kim JS, DeGiovanni A, Jancarik J, Adams PD, Yokota H, Kim R and Kim SH. Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc Natl Acad Sci USA,* **102**, 3248–3253 (2005).
22. Sturrock SS and Dryden DTF. A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res,* **25**, 3408–14 (1997).
23. O'Neill M, Dryden DTF and Murray NE. Localization of a protein-DNA interface by random mutagenesis. *EMBO,* **17**, 7118–27 (1998).
24. O'Neill M, Powell LM and Murray NE. Target recognition by EcoKI: the recognition domain is robust and restriction-deficiency commonly results from the proteolytic control of enzyme activity. *J Mol Biol,* **307**, 951–63 (2001).
25. Kennaway CK, Taylor JE, Song CF, Potrzebowski W, Nicholson W, White JH, Swiderska A, Obarska-Kosinska A, Callow P, Cooper LP, Roberts GA, Artero JB, Bujnicki JM, Trinick J, Kneale GG and Dryden DTF. Structure and operation of the DNA-translocating type I DNA restriction enzymes. *Gene Dev,* **26**, 92–104 (2012).
26. Murphy M, Nuoffer SS and Bickle TA. Lack of Regulation of the Modification-Dependent Restriction Enzyme McrBC in Escherichia coli. *J Bacteriol,* **184**, 1794–1795 (2002).

27. Su TJ, Tock MR, Egelhaaf SU, Poon WCK and Dryden DTF. DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Res,* **33**, 3235–44 (2005).
28. Meselson M, Yuan R and Heywood J. Restriction and modification of DNA. *Annu Rev Biochem,* **41**, 447–66 (1972).
29. Lapkouski M, Panjikar S, Janscak P, Smatanova IK, Carey J, Ettrich R and Csefalvay W. Structure of the motor subunit of type I restriction-modification complex EcoR124I. *Nat Struct Mol Biol.* **16**, 94–5 (2009).
30. Loenen WA, Daniel AS, Braymer HD and Murray NE. Organization and sequence of the *hsd* genes of Escherichia coli K-12. *J Mol Biol,* **198**, 159–70 (1987).
31. Szczelkun MD. How to proteins move along DNA? Lessons from type-I and type-III restriction endonucleases. *Essays Biochem,* **35**, 131–43 (2000).
32. Janscak P, MacWilliams MP, Sandmeier U, Nagaraja V and Bickle TA. DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO,* **18**, 2638–47 (1999).
33. Rosamond J, Endlich B and Linn S. Electron microscopic studies of the mechanism of action of the restriction endonuclease of Escherichia coli B. *J Mol Biol,* **129**, 619–635 (1979).
34. Yuan R, Hamilton D L and Burckhardt J. DNA translocation by the restriction enzyme from *E. coli* K. *Cell,* **20**, 237–44 (1980).
35. Studier FW and Bandyopadhyay P K. Model for how type I restriction enzymes select cleavage sites in DNA. *Proc Natl Acad Sci USA,* **85**, 4677–81 (1988).
36. Halford SE and Marko JF. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res,* **32**, 3040–52 (2004).
37. Berge T, Ellis DJ, Dryden DTF, Edwardson JM and Henderson RM. Translocation-independent dimerization of the EcoKI endonuclease visualized by atomic force microscopy. *Biophys J* , **79**, 479–84 (2000).
38. Neaves KJ, Cooper LP, White JH, Carnally SM, Dryden DTF, Edwardson JM and Henderson RM. Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerization and DNA looping. *Nucleic Acids Res,* **37**, 2053–63 (2009).
39. Janscak P and Bickle TA. DNA supercoiling during ATP-dependent DNA translocation by the type I restriction enzyme EcoAI. *J Mol Biol*, **4**, 1089-1099 (2000)

40. Holubová I, Vejsadová S,Weiserová M and Firman K. Localization of the type I restriction-modification enzyme EcoKI in the bacterial cell. *Biochem Biophys Res Co,* **270**, 46–51 (2000).
41. Firman K and Szczelkun MD. Measuring motion on DNA by the type I restriction endonuclease EcoR124I using triplex displacement. *EMBO,* **19**, 2094–102 (2000).
42. Garcia LR and Molineux I J. Translocation and specific cleavage of bacteriophage T7 DNA *in vivo* by EcoKI. *Proc Natl Acad Sci USA,* **96**, 12430–12435 (1999).
43. Seidel R, van Noort J, van der Scheer C, Bloom JG, Dekker NH, Dutta CF, Blundall A, Robinson, Firman K and Dekker C. Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nat Struct Mol Biol,* **11**, 838–43 (2004).
44. Meselson M and Yuan R. DNA restriction enzyme from *E. coli*. *Nature,* **217**, 1110–4 (1968).
45. Kennaway CK, Obarska-Kosinska A, White JW, Tuszynska I, Cooper LP, Bujnicki, Trinick J and Dryden DTF. The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res,* **37**, 762–70 (2009)
46. Walkinshaw MD, Taylor P, Sturrock SS, Atanasiu C, Berge T, Henderson RM, Edwardson JM and Dryden DTF. Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol Cell,* **9**, 187–94 (2002).
47. Iida S, Streiff MB, Bickle TA and Arber W. Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. *Virology,* **157**, 156–66 (1987).
48. Loenen WA and Murray NE. Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J Mol Biol,* **190**, 11–22 (1986).
49. Simmon VF and Lederberg S. Degradation of bacteriophage lambda deoxyribonucleic acid after restriction by Escherichia coli K-12. *J Bacteriol,* **112**, 161–9 (1972).
50. Studier FW. Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host. *J Mol Biol,* **94**, 283–95 (1975).
51. Krüger DH, Gola G, Weisshuhn I and Hansen S. The ocr gene function of bacterial viruses T3 and T7 prevents host-controlled modification. *J Gen Virol,* **41**, 189–92 (1978).
52. Krüger DH, Chernin LS, Hansen S, Rosenthal HA and Goldfarb DM. Protection of foreign DNA against host-controlled restriction in bacterial cells. *Mol Gen Genet,* **159**, 107–110 (1978).

53. Krüger DH, Schroeder C, Hansen S and Rosenthal HA. Active protection by bacteriophages T3 and T7 against *E. coli* B-and K-specific restriction of their DNA. *Mol Gen Genet,* **153**, 99–106 (1977).
54. Studier FW and Movva NR. SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J Virol,* **19**, 136–45 (1976).
55. Spoerel N, Herrlich P and Bickle TA. A novel bacteriophage defence mechanism: the anti-restriction protein. *Nature,* **278**, 30–4 (1979).
56. Dunn JJ, Elzinga M, Mark K and Studier FW. Amino Acid Sequence of the Gene 0.3 Protein of Bacteriophage. *J Biol Chem,* **256**, 2579–2585 (1981).
57. Stephanou AS, Roberts GA, Cooper LP, Clarke DJ, Thomson AR, MacKay CL, Nutley M. Cooper A and Dryden DTF. Dissection of the DNA Mimicry of the Bacteriophage T7 Ocr Protein using Chemical Modification. *J Mol Biol,* **391**, 565–576 (2009).
58. Putnam CD and Tainer JA. Protein mimicry of DNA and pathway regulation. *DNA repair,* **4**, 1410–20 (2005).
59. Atanasiu C, Su TJ, Sturrock SS and Dryden DTF. Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res,* **30**, 3936–3944 (2002).
60. Atanasiu C, Byron O, McMiken H, Sturrock SS and Dryden DTF. Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7. *Nucleic Acids Res,* **29**, 3059–68 (2001).
61. Stephanou AS, Roberts GA, Tock MR, Pritchard EH, Turkington R, Nutley M, Cooper A Dryden DTF. A mutational analysis of DNA mimicry by ocr, the gene 0.3 antirestriction protein of bacteriophage T7. *Biochem Biophys Res Co,* **378**, 129–32 (2009).
62. Salyers AA, Shoemaker NB, Stevens AM and Li LY. Conjugative transposons: an unusual and diverse set of integrated gene transfer elements. *Microbiol Rev,* **59**, 579–90 (1995).
63. Wilkins BM. Plasmid promiscuity: meeting the challenge of DNA immigration control. *Environ Microbiol,* **4**, 495–500 (2002).
64. Zavil'gel'skiĭ, GV, Kotova VI and Rastorguev SM. Antimodification activity of the ArdA and Ocr proteins. *Genetika,* **47**, 159–167 (2011).
65. Zavilgelsky, GB, Kotova VY and Rastorguev SM. Antirestriction and antimodification activities of T7 Ocr: Effects of amino acid substitutions in the interface. *Mol Biol,* **43**, 93–100 (2009).

66. Nekrasov SV, Agafonova OV, Belogurova NG, Delver EP and Belogurov AA. Plasmid-encoded antirestriction protein ArdA can discriminate between type I methyltransferase and complete restriction-modification system. *J Mol Biol,* **365**, 284–97 (2007).
67. McMahon SA, Roberts GA, Johnson KA, Cooper LP, Liu H, White JH, Carter LG, Sanghvi B Oke M Walkinshaw MD, Blakely GW, Naismith JH and Dryden DTF. Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance. *Nucleic Acids Res,* **37**, 4887–97 (2009).
68. Zavilgelsky GB and Rastorguev SM. Antirestriction proteins ArdA and Ocr as efficient inhibitors of type I restriction-modification enzymes. *Mol Biol,* **43**, 241–248 (2009).
69. O'Neill M, Chen A. and Murray NE. The restriction-modification genes of Escherichia coli K-12 may not be selfish: they do not resist loss and are readily replaced by alleles conferring different specificities. *Proc Natl Acad Sci USA,* **94**, 14596–601 (1997).
70. Boyer H. Genetic control of Restriction and Modification in Escherichia coli.. *J Bacteriol,* **88**, 1652–60 (1964).
71. Sain B and Murray NE. The hsd (host specificity) genes of *E. coli* K 12. *Mol Gen Genet,* **180**, 35–46 (1980).
72. Arber W and Linn S. DNA modification and restriction. *Annu Rev Biochem,* **38**, 467–500 (1969).
73. Baharoglu Z, Bikard D and Mazel D. Conjugative DNA transfer induces the bacterial SOS response and promotes antibiotic resistance development through integron activation. *PLoS Genet*, **6**, 10011-65 (2010).
74. Kelleher JE and Raleigh EA. Response to UV damage by four Escherichia coli K-12 restriction systems. *J. Bacteriol,* **176**, 5888–5896 (1994).
75. Fuller-Pace FV, Cowan GM and Murray NE. EcoA and EcoE: alternatives to the EcoK family of type I restriction and modification systems of Escherichia coli. *J Mol Biol,* **186**, 65–75 (1985).
76. Blakely GW and Murray NE. Control of the endonuclease activity of type I restriction-modification systems is required to maintain chromosome integrity following homologous recombination. *Mol Microbiol,* **60**, 883–93 (2006).
77. Doronina VA and Murray NE. The proteolytic control of restriction activity in Escherichia coli K-12. *Mol Microbiol,* **39**, 416–28 (2001).

78. Belogurov AA, Efimova EP, Del'ver EP and Zavil'gel'skiĭ GB. Weakening of type I restriction in *E. coli*: the effect of 2-aminopurine and 5-bromouracil. *Mol Gen Mikrobiol Virusol,* **12**, 34–40 (1987).
79. Efimova EP, Delver EP and Belogurov AA. 2-Aminopurine and 5-bromouracil induce alleviation of type I restriction in Escherichia coli: mismatches function as inducing signals? *Mol Gen Genet,* **214**, 317–320 (1988).
80. Cromie GA and Leach DR. Recombinational repair of chromosomal DNA double-strand breaks generated by a restriction endonuclease. *Mol Microbiol,* **41**, 873–83 (2001).
81. Matic I, Ekiert D, Radman M and Kohiyama M. Generation of DNA-Free Escherichia coli Cells by 2-Aminopurine Requires Mismatch Repair and Nonmethylated DNA. *J Bacteriol,* **188**, 339–342 (2006).
82. Makovets S, Titheradge AJ and Murray NE. ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and IB restriction systems. *Mol Microbiol,* **28**, 25–35 (1998).
83. Makovets S, Doronina VA and Murray NE. Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes. *Proc Natl Acad Sci USA,* **96**, 9757–9762 (1999).
84. Makovets S, Doronina VA and Murray NE. Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes. *Proc Natl Acad Sci USA,* **96**, 9757–62 (1999).
85. Maurizi MR, Thompson MW, Singh SK and Kim SH. Endopeptidase Clp: ATP-dependent Clp protease from Escherichia coli. *Method Enzymol,* **244**, 314–31 (1994).
86. Maurizi MR, Clark WP, Kim SH and Gottesman S. Clp P represents a unique family of serine proteases. *J Biol Chem,* **265**, 12546–52 (1990).
87. Wang J, Hartling JA and Flanagan JM. The structure of ClpP at 2.3 A resolution suggests a model for ATP-dependent proteolysis. *Cell,* **91**, 447–56 (1997).
88. Kessel M, Maurizi MR, Kim B, Kocsis E, Trus BL, Singh SK, Steven AC. Homology in structural organization between *E. coli* ClpAP protease and the eukaryotic 26 S proteasome. *J Mol Biol,* **250**, 587–94 (1995).
89. Hoskins JR, Pak M, Maurizi MR and Wickner S. The role of the ClpA chaperone in proteolysis by ClpAP. *Proc Natl Acad Sci USA,* **95**, 12135–40 (1998).
90. Janion C. Some aspects of the SOS response system — A critical survey. *Acta Biochim Pol,* **48**, 599–607 (2001).

91. Aertsen A, Houdt R Van, Vanoirbeek K and Michiels CW. An SOS Response Induced by High Pressure in Escherichia coli An SOS Response Induced by High Pressure in Escherichia coli. *J Bacteriol*, **186**, 6133-41 (2004).
92. Janion, C. Inducible SOS Response System of DNA Repair and Mutagenesis in Escherichia coli. *Int J Bio Sci,* **4**, 338–344 (2008).
93. Gruenig MC, Renzette N, Long E, Chitteni-Pattu S, Inman RB, Cox MM and Sandler SJ. RecA-mediated SOS induction requires an extended filament conformation but no ATP hydrolysis. *Mol Microbiol,* **69**, 1165–79 (2008).
94. Michel, B. After 30 years of study, the bacterial SOS response still surprises us. *PLoS biology,* **3**, e255 (2005).
95. Ivancic-Bace I, Vlasic I, Salaj-Smic E and Brcic-Kostic K. Genetic evidence for the requirement of RecA loading activity in SOS induction after UV irradiation in Escherichia coli. *J Bacteriol,* **188**, 5024–32 (2006).
96. Krishna S, Maslov S and Sneppen K. UV-induced mutagenesis in Escherichia coli SOS response: a quantitative model. *PLoS Comput Biol,* **3**, e41 (2007).
97. Voloshin ON, Ramirez BE, Bax A and Camerini-Otero RD. A model for the abrogation of the SOS response by an SOS protein: a negatively charged helix in DinI mimics DNA in its interaction with RecA. *Gene Dev,* **15**, 415–27 (2001).
98. Boberek JM, Stach J and Good L. Genetic evidence for inhibition of bacterial division protein FtsZ by berberine. *PloS one,* **5**, e13745 (2010).
99. Huismant O, Arit RD and Gottesmant S. Cell-division control in Escherichia coli: Specific induction of the SOS function SfiA protein is sufficient to block septation. Proc Natl Acad Sci USA. **81**, 4490–4494 (1984).
100. Ari RD, Huisman O, Ari RD and Huisman O. Novel mechanism of cell division inhibition associated with the SOS response in Escherichia coli. *J Bacteriol,* **156**, 243-250 (1983)
101. Mo AH and Burkholder WF. YneA, an SOS-induced inhibitor of cell division in Bacillus subtilis, is regulated posttranslationally and requires the transmembrane region for activity. *J Bacteriol,* **192**, 3159–73 (2010).
102. Jeong KJ and Lee SY. Enhanced Production of Recombinant Proteins in Escherichia coli by Filamentation Suppression Enhanced Production of Recombinant Proteins in Escherichia coli by Filamentation Suppression. *Proc Natl Acad Sci USA*, **14**, 4490-4 (2003).
103. Dahiyat BI, Sarisky CA and Mayo SL. *De novo* protein design: towards fully automated sequence selection. *J Mol Biol,* **273**, 789–96 (1997).

104. Matsuura T, Ernst A, Zechel DL and Plückthun A. Combinatorial approaches to novel proteins. *Chembiochem,* **5**, 177–82 (2004).
105. Guo H H, Choe J and Loeb L.A. Protein tolerance to random amino acid change. *Proc Natl Acad Sci USA,* **101**, 9205–10 (2004).
106. Goldsmith M and Tawfik DS. Enzyme engineering by targeted libraries. *Method Enzymol,* **523**, 257–83 (2013).
107. Gupta RD and Tawfik DS. Directed enzyme evolution via small and effective neutral drift libraries. *Nat Method,* **5**, 939–942 (2008).
108. Moffet DA and Hecht MH. *De novo* proteins from combinatorial libraries. *Chem Rev,* **101**, 3191–203 (2001).
109. Hellinga, H. W. Rational protein design: combining theory and experiment. *Proc Natl Acad Sci USA* **94**, 10015–7 (1997).
110. Gaucher E.A, Gu X, Miyamoto MM and Benner SA. Predicting functional divergence in protein evolution by site-specific rate shifts. *Trends Biochem Sci,* **27**, 315–21 (2002).
111. Pál C, Papp B and Lercher MJ. An integrated view of protein evolution. *Nat Rev Genet,* **7**, 337–48 (2006).
112. Kato S, Han SY, Liu W, Otsuka K, Shibata H, Kanamaru R and Ishioka C. Understanding the function – structure and function – mutation relationships of p53 tumor suppressor protein by high-resolution missense mutation analysis. *Proc Natl Acad Sci USA*, **100,** 8425-8429 (2003).
113. West MW, Wang W, Patterson J, Mancias JD, Beasley JR and Hecht MH. *De novo* amyloid proteins from designed combinatorial libraries. *Proc Natl Acad Sci USA,* **96**, 11211–6 (1999).
114. Fisher MA, McKinley KL, Bradley LH, Viola SR and Hecht MH. *De novo* designed proteins from a library of artificial sequences function in Escherichia coli and enable cell growth. *PloS one,* **6**, e15364 (2011).
115. Go A, Kim S, Baum J and Hecht MH. Structure and dynamics of *de novo* proteins from a designed superfamily of 4-helix bundles. *Protein Sci,* **17**, 821–832 (2008).
116. Beasley JR and Hecht MH. Protein design: the choice of *de novo* sequences. *J Biol Chem,* **272**, 2031–4 (1997).
117. Hecht MH, Das A, Go A, Bradley LH and Wei Y. *De novo* proteins from designed combinatorial libraries. *Protein Sci,* **13**, 1711–1723(2004).

118. Herman A and Tawfik DS. Incorporating Synthetic Oligonucleotides via Gene Reassembly (ISOR): a versatile tool for generating targeted libraries. *Protein Eng,* **20**, 219–26 (2007).
119. Makovets S, Powell LM, Titheradge AJB, Blakely GW and Murray NE. Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? *Mol Microbiol,* **51**, 135–147 (2003).
120. Serfiotis-Mitsa, D, Herbert AP, Roberts GA, Soares DC, White JW, Blakely GW, Uhrin D and Dryden DTF. The structure of the KlcA and ArdB proteins reveals a novel fold and antirestriction activity against Type I DNA restriction systems *in vivo* but not *in vitro*. *Nucleic Acids Res,* **38**, 1723–37 (2010).
121. Stephanou, A. S. Biophysical study of the DNA charge mimicry displayed by the T7 Ocr protein. Thesis, Edinburgh. (2010).
122. Monera OD, Kay CM and Hodges RS Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci,* **3**, 1984–91 (1994).
123. Dryden DTF, Cooper LP, Thorpe PH and Byron O. The *in vitro* assembly of the EcoKI type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochem,* **36**, 1065–76 (1997).
124. Higham, R. G. A Biophysical Analysis of the Ocr Protein Gel. Thesis, Edinburgh.(2007)
125. Kowalczykowski SC, Dixon DA, Eggleston AK, Lauder SD and Rehrauer WM. Biochemistry of homologous recombination in Escherichia coli. *Microbiol Rev,* **58**, 401–65 (1994).
126. Betton JM and Hofnung M. Folding of a mutant maltose-binding protein of Escherichia coli which forms inclusion bodies. *J Biol Chem,* **271**, 8046–52 (1996).
127. Ratner, D. The interaction of bacterial and phage proteins with immobilized Escherichia coli RNA polymerase. *J Mol Biol,* **88**, 373–383 (1974).
128. Shimoni Y, Altuvia S, Margalit H and Biham O. Stochastic analysis of the SOS response in Escherichia coli. *PloS one,* **4**, e5363 (2009).
129. Dillingham MS and Kowalczykowski SC. RecBCD enzyme and the repair of double-stranded DNA breaks. *Microbiol Mol Biol,* **72**, 642–71, Table of Contents (2008).
130. Dryden DTF and Tock MR. DNA mimicry by proteins. *Biochem Soc T,* **34**, 317–9 (2006).

131. Roberts GA, Stephanou AS, Kanwar N, Dawson A, Cooper LP, Chen K, Nutley M, Cooper A, Blakely GW and Dryden DTF. Exploring the DNA mimicry of the Ocr protein of phage T7. *Nucleic Acids Res,* **40**, 8129–43 (2012).

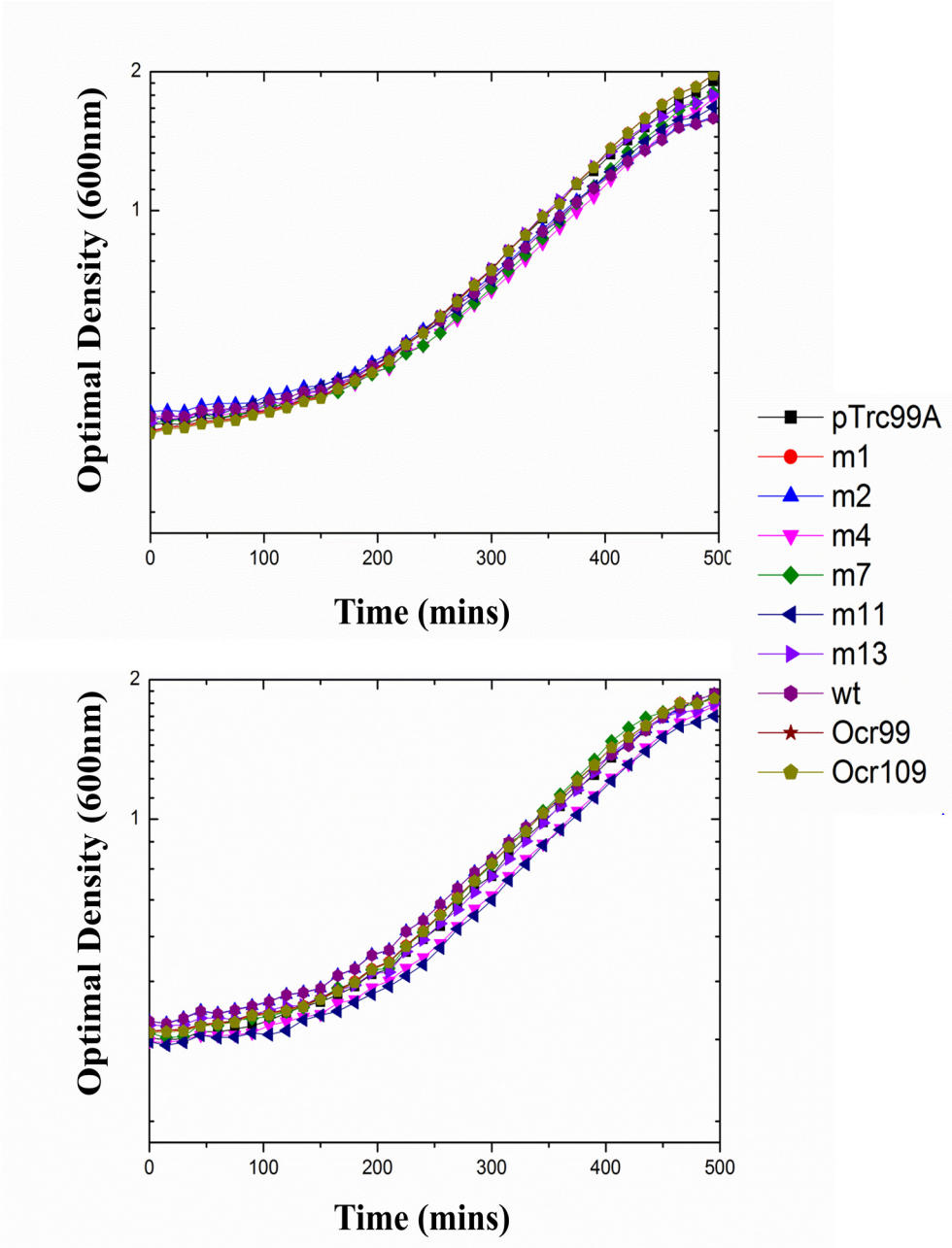


Figure A.8: Bacterial Growth Curves with the Ocr mutants expressed in *E. coli*, NM1041 (*clpXP*-, r+m+) cells in the absence of 2AP. pTrc99A is the vector alone, all other data are for cells expressing Ocr wild-type (wt) or Ocr mutants. OD_{600} measurements were recorded every 15 min for 500 min. Errors are ±7% for triplicate measurements (error bars not shown for clarity).Each of the mutants maintain share the same growth pattern as wtOcr.

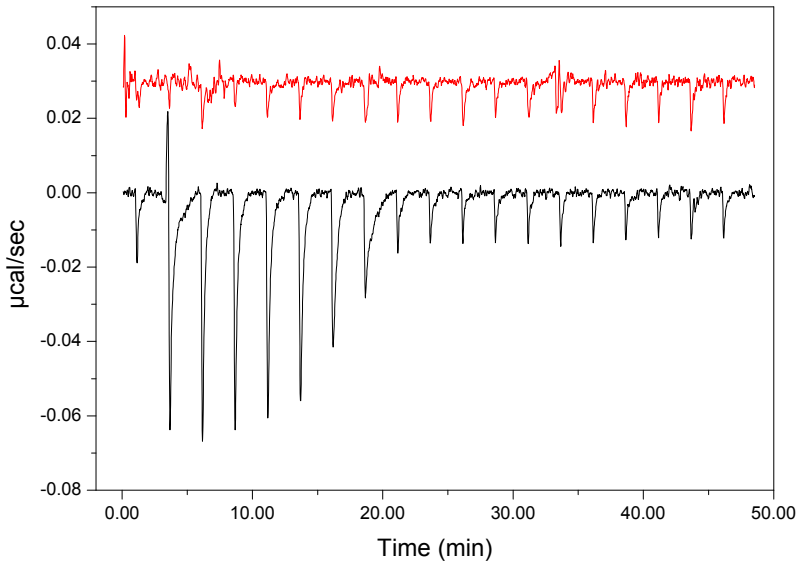


Figure A.9: The ITC traces which follow the reaction between M.EcoKI with wtOcr (black) and wtOcr with buffer (20 mM HEPES, pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol). The "Sample Cell" contained the M.EcoKI, while the "Syringe" contained wtOcr protein. The traces show that the buffer does not contribute to the heat change exhibited in the interaction.

ArdA multimutants

Introduction

ArdA is an antirestriction protein and, like wtOcr, it functions as a competitive inhibitor of EcoKI by mimicking the shape and charge of B-form DNA. However, there are also many fundamental differences between the two antirestriction proteins; the ArdA protein is more elongated mimicking 42bp of dsDNA compared to Ocr which mimics 24bp dsDNA, the secondary structure fold of both proteins is also different, indicating that they belong to two separate protein families. Furthermore, the antirestriction proteins are located on different types of mobile genetic elements. Genes that code for the Ocr protein reside in lytic bacteriophage whereas the ArdA gene locus is highly dispersed among conjugative plasmids and transposons.

Each of these mobile elements behaves differently upon infection; lytic bacteriophage dock onto the cell membrane of the bacterial cell and inject the phage genome into the cytoplasm. Upon replication of the virion, the phage initiates cell lysis and destroys the bacterial cell. Once destroyed the phage progeny are released and able to subsequently infect and replicate. However, conjugative plasmids are symbiotic in nature, consisting of circular dsDNA. These mobile genetic elements are transferred into bacterial cells by direct conjugational transfer. As a result the plasmid can replicate and produce proteins and genes that can be

advantageous to the bacterial cell. The key difference between conjugative transposons and plasmids is that transposons tend to integrate into the chromosomal DNA, whereas plasmids remain separate from the host genome.

Previous studies of ArdA proteins from different transposons and plasmids have shown that, unlike Ocr, the ArdA proteins can exhibit differential activity towards antirestriction and antimodification activity depending on the particular ArdA protein and the particular Type I RM system being targeted. Furthermore comparisons of the binding affinity of wtOcr from T7 bacteriophage and Orf18 ArdA from the transposon Tn916 shows the ArdA has a weaker interaction to EcoKI compared to wtOcr. In addition, ArdA proteins exist as a monomer and homodimer in solution, in fact the monomer is dominant at lower concentrations whereas the homodimer is prevalent at higher concentrations.

During the development of the 2AP *in vivo* assay, a study to investigate the loss of negative surface charge upon the three domains that exist on the Orf18 ArdA protein was underway in the lab (by Dr GA Roberts). In a similar manner to the Ocr mutant investigation, several mutants were designed which targeted acidic residues residing on different regions (in this case the three different domains) of the Orf18 ArdA protein structure. These acidic residues were neutralised by substitutions of aspartic and glutamic acid to asparagine and glutamine respectively. In addition, a mutant which targeted the hydrophobic interactions of the dimer interface, with the aim of producing an ArdA monomer protein, was created. Although *in vivo* characterisation of the ArdA mutants was carried out by assessing the success of phage infection on restriction proficient and deficient cells, further *in vitro* characterisation could not be carried out on a number of the ArdA mutants. Attempts to purify the protein showed that there was variable expression between the mutants. In fact insufficient protein was expressed to be detected on a DEAE column making purification of these mutants difficult. Therefore the mutants were investigated with the 2AP RA assay. This provided the opportunity to test the assay on a set of mutants from a different antirestriction protein, assessing the flexibility and versatility of the assay. Also the study should hopefully allow us to gain a more detailed profile for the activity of the ArdA protein.

ArdA multimutants

The initial design and creation of the ArdA mutants including the subsequent *in vitro* characterisation of the mutants M5, M6, L127E were carried out by Dr G.A. Roberts and were a kind gift for the purposes of testing the RA assay.

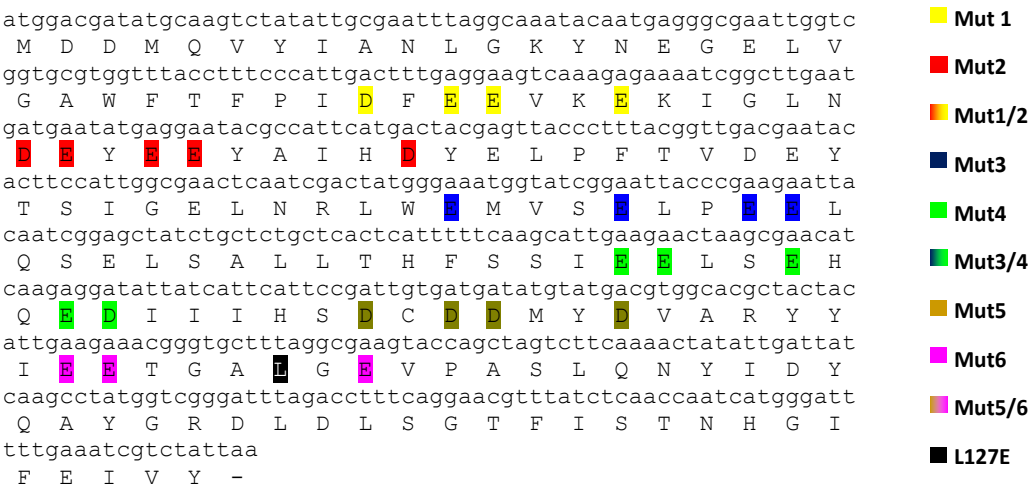


Figure A.10: The primary amino acid and DNA sequence of the Orf 18 ArdA wild type (wt) protein. Highlighted are the targets of mutagenesis, the acidic residues aspartic and glutamic acid were substituted with asparagines and glutamine respectively. Each color represents a different ArdA multimutant.

In vivo restriction alleviation assay

From Figure 3.2 in chapter 3, it was shown that like wtOcr, the presence of an active ArdA protein can rescue the cell viability of *E. coli* NM1041 exposed to 2AP by inhibiting the EcoKI enzyme.

E. coli NM1041 cells were transformed with each of the ArdA mutants. The effect of 2AP on cell survival, in the presence of the ArdA mutants is shown in Figure 10.4. The RA assay was carried out with leaky expression (-IPTG) and over expression (+IPTG) to discern any differences in activity. RM enzymes are present in low concentrations in the cellular environment; this small amount of expression can be represented in the leaky expression of the mutants as well as reflecting the conditions used in the phage infection experiment.

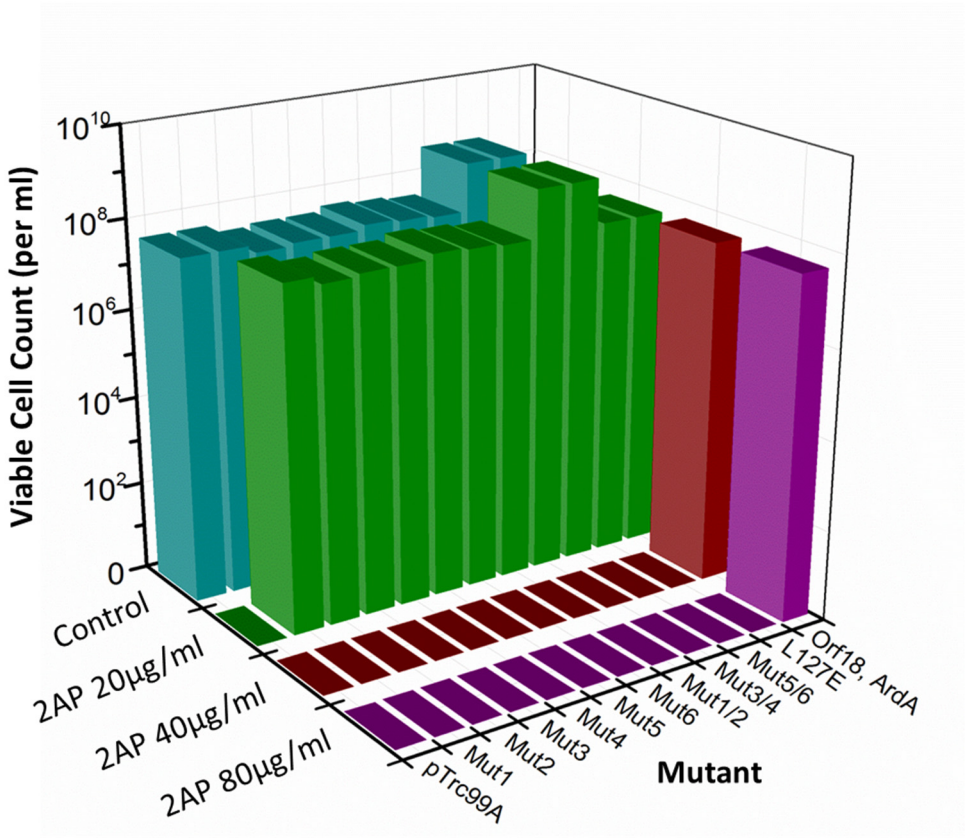


Figure A.11: The RA assay measuring cell survival of *E. coli* NM1041 (clpXP−, r^+m^+) when transformed with various plasmids expressing Orf 18 ArdA or its mutant forms. The graph shows growth on LB-agar plates supplemented with antibiotics (cyan) only, and in the presence of antibiotics plus increasing concentrations of 2AP (20, 40 and 80 µg/ml are green, red, and magenta, respectively).

In the absence of 2AP, the *E. coli* strain grew well for all of the ArdA mutants at 10^7 cells per ml. A loss in cell viability was observed for cells grown with the empty plasmid vector pTrc99A, at a concentration of 20µg/ml. This is consistent with observations from the Ocr experiments and represents the resistance conferred by an inactive antirestriction protein.

Cells grown with each of the ArdA mutants possessed the same resistance to 2AP, each of the mutants lost cell viability at a 2AP concentration of 40µg/ml. However, wtArdA maintains 2AP resistance at a concentration of 80ug/ml. This suggests that all of the mutants provide some antirestriction activity, although this is reduced compared to wtArdA protein (Figure 4.2).

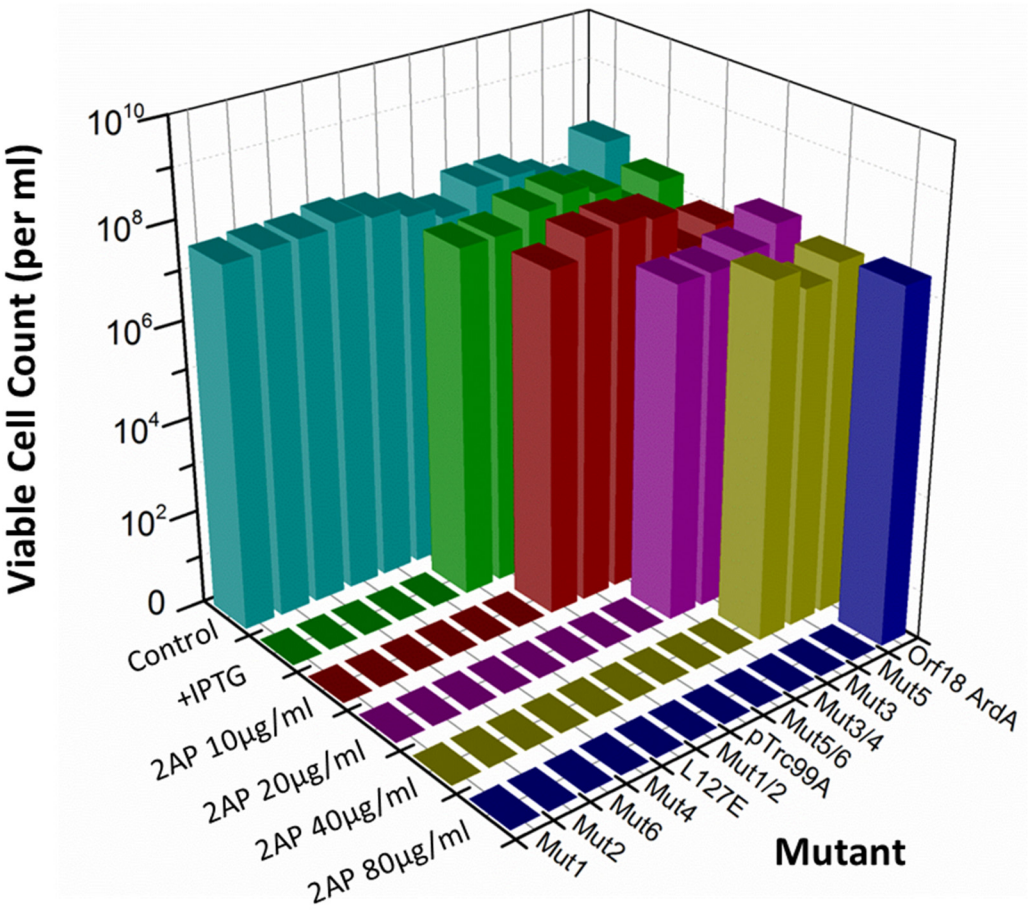


Figure A.12: The RA assay measuring cell survival of *E. coli* NM1041 (clpXP$^-$, r$^+$m$^+$) when transformed with various plasmids expressing Orf18 or its mutant forms. The graph shows growth on LB-agar plates supplemented with antibiotics (cyan) only, antibiotics and IPTG (green) and, in the presence of antibiotics, IPTG plus increasing concentrations of 2AP (10, 20, 40 and 80µg/ml are red, magenta and olive, respectively).

The 2AP resistance exhibited by *E. coli* cells in the presence of pTrc99A vector or wtOrf18 ArdA remained unchanged from the uninduced assay and *E. coli* cells overexpressing each mutant displayed a resistance that was no greater than 40µg/ml, confirming that none of the mutants functioned as a fully active antirestriction protein. However, *E. coli* NM1041 cells upon overexpression of the mutant Mut5/6 reduced 2AP resistance from a concentration of 20µg/ml to 10µg/ml; this is the same resistance as the pTrc99A vector indicating that the mutant Mut5/6 is inactive.

Upon induction by IPTG, cells grown with the mutants, Mut1, Mut2, Mut4, Mut6 and L127E showed toxic effects and loss of cell viability was observed for Mut1/2 at 10µg/ml, these mutants share the same phenotype witnessed by the Ocr mutants M12, M16, Ocr/POcr and M3. However the remaining ArdA mutants have shown new phenotypes.

The 2AP resistance experienced by the remaining mutants ranges from 20µg/ml to 80µg/ml. Cells in the presence of Mut3/4 lost viability at a concentration of 40µg/ml, whereas both Mut3 and Mut5 lost viability at 80µg/ml. The increased sensitivity of the RA assay and over expression of the mutants has allowed us to differentiate between the mutants by increasing the level of 2AP.

Bacterial growth curves

The RA assay shows an effective phenotype for each of the mutants, however, we do not know how the cell cultures are affected as a function of time. Bacterial growth curves of the ArdA mutants identified subtle differences in the growth rate of cells. The growth curves compared the growth rate of cells with and without induction of the protein. Bacterial growth curves of the *E. coli* NM1041 were also measured with increasing levels of 2AP by measuring the OD_{600} of the bacterial cultures every 5mins for a total of 900mins.

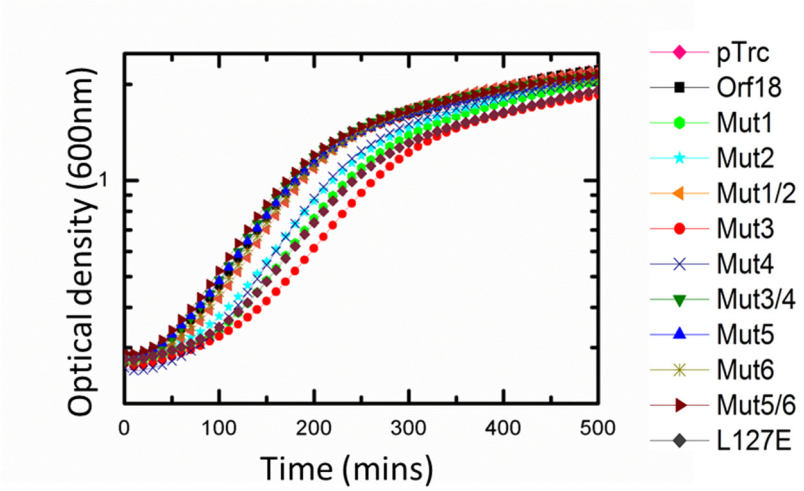


Figure A.13: Cell growth with the ArdA, Orf18 mutants expressed in *E. coli*, NM1041 (clpXP$^-$, r$^+$m$^+$) in the presence of antibiotics only (Km and Carb). It is clear that cells expressing different ArdA mutants showed the same growth rate. Growth behavior of the mutants was compared with wtOrf18 and the empty vector pTrc99A which represent an active and inactive anti-restriction protein respectively.

Bacterial growth curves were recorded for *E. coli* NM1041 cells grown with each of the ArdA mutants. Under selective antibiotic markers only, each of the mutants induced conventional growth patterns, similar to *E. coli* NM1041 cells in the presence of pTrc99A or wtArdA (Figure A6). The bacterial growth curves of the ArdA mutants demonstrated a doubling time of ~101minutes, this doubling time is slightly faster than the doubling rate of *E. coli* NM1041 cells in the presence of wtOcr which was ~130mins. This suggests that unlike the Ocr mutants none of the ArdA mutants caused adverse growth patterns as a result

of leaky expression. This is verified by the uninduced RA assay carried out in Figure A.4, which produces no evidence of deleterious effects.

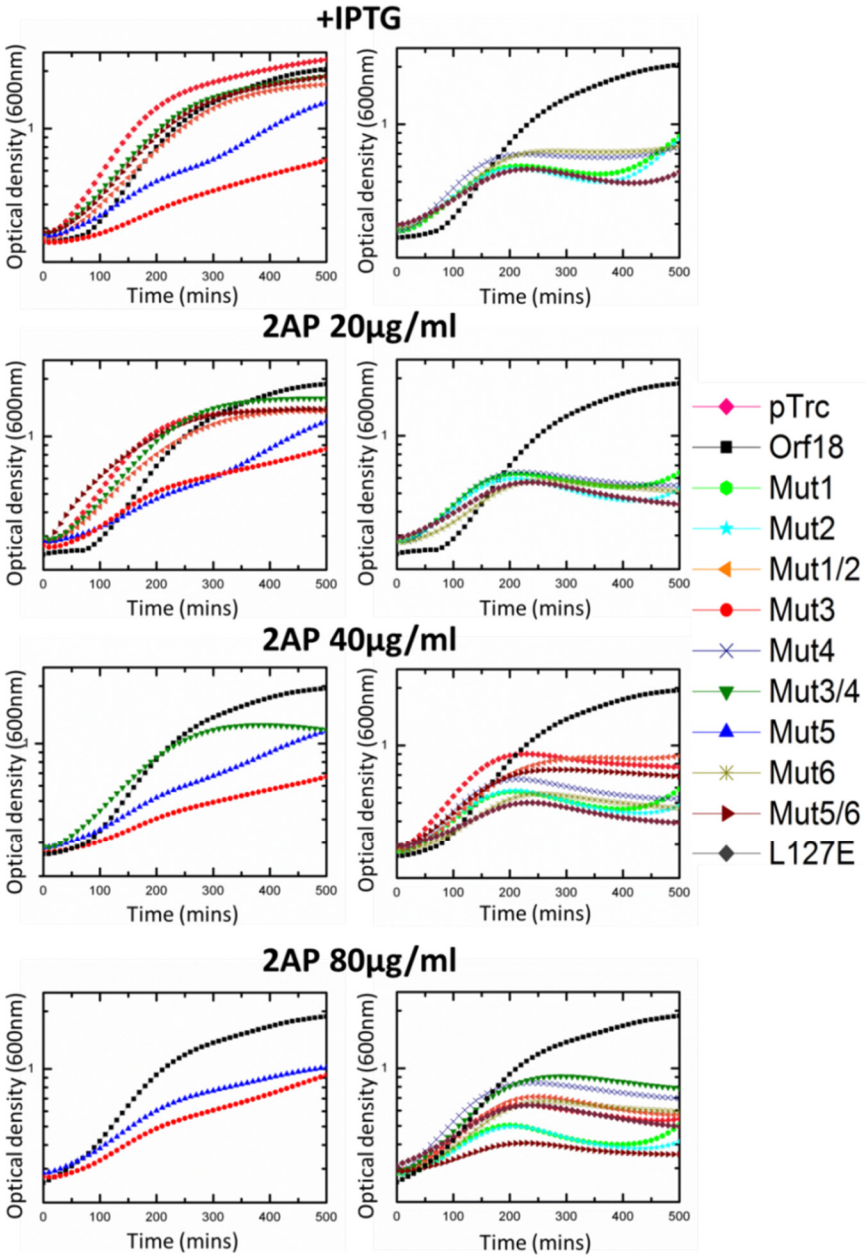


Figure A.14: Cell growth with the ArdA Orf18 mutants expressed in *E. coli*, NM1041 (clpXP^{-}, r^{+}m^{+}) in the presence and absence of 2AP. It is clear that cells expressing different ArdA mutants showed varying growth rates, some cultures were affected by; the expression of the mutant (+IPTG) or when exposed to increasing concentrations of 2AP. Growth behavior of the mutants were compared with wtOrf18 and the empty vector pTrc99A which represent an active and inactive anti-restriction protein respectively. Graphs on the left show mutants which have retained some activity when exposed to increasingly toxic conditions whereas on the right are the mutants which have a sigmoidal curve which represents a loss of activity both types of graph include the growth curve of wtArdA for comparative analysis. OD600 measurements were recorded every 5 min for 800 min. Errors are ±7% for triplicate measurements (error bars not shown for clarity).

Overall the growth of *E. coli* NM1041 in the presence of the ArdA mutants, displays growth curves which reflect the phenotypes produced from the RA plate assay. The exception is the growth profile of cells grown with mutants Mut3 and Mut5. Although the level of 2AP resistance is retained at a concentration of 80μg/ml from the results of the 2AP plate assay and bacterial growth curves, upon over expression with IPTG both mutants induce a slow growth rate but retain antirestriction activity.

Microscopy

The deleterious effect induced by particular ArdA mutants on cell growth has been witnessed previously by the Ocr mutants, M12, M16 and Ocr/pOcr in chapter 3. The severe reduction in cell growth was previously attributed to the production of DNA filaments. On the other hand, a number of the ArdA mutants confer different levels of 2AP resistance/antirestriction activity that have not been identified before.

Therefore *E. coli* NM1041 cells expressing ArdA or its mutants were induced with IPTG, the cells were fixed with methanol and examined under a light microscope. Comparisons that showed that the majority of cells observed were a similar size regardless of the ArdA mutant present. However, cells grown with some of the ArdA mutants did produce a number of filaments although these were not as obvious as those induced by the Ocr mutants. To be able to discern whether the number of filaments is significant and can be attributed to the activity of the ArdA mutant, the cell length was measured for at least 400 cells that were selected at random. Each of these cells were categorised as either normal or as filaments. Conventionally cell length follows a normal distribution with the majority of cells inducing pTrc99A measuring ~2 microns, if the length of a cell was greater than 4 microns then it would be classified as a filament. The percentage frequency of filaments to normal cells is outlined for each mutant in Figure A.9.

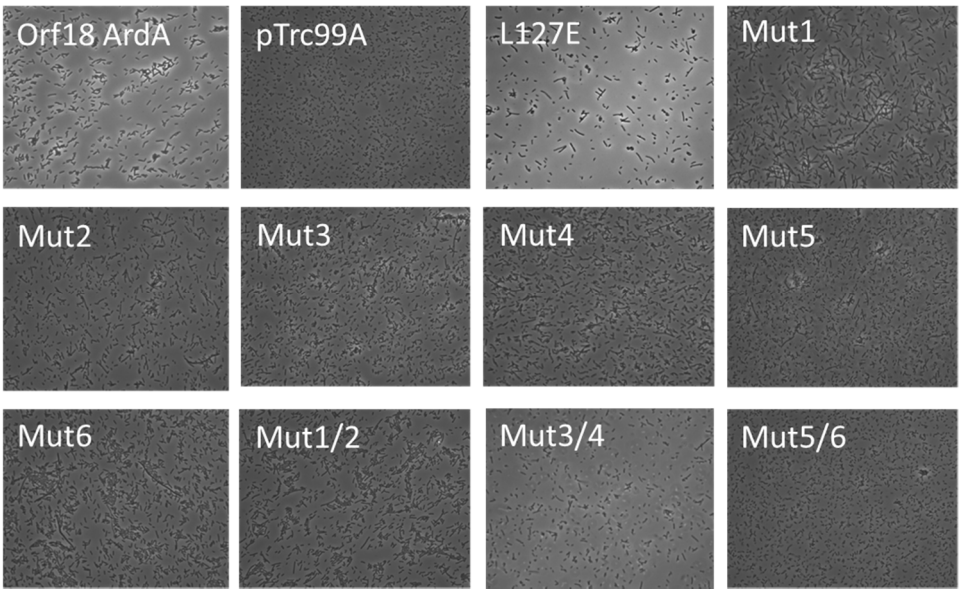


Figure A.15: Light Microscopy images of cell morphology of bacterial cultures of *E. coli* NM1041 (ClpXP^{-}, M^{+}R^{+}) transformed with plasmids expressing wild-type Orf18, ArdA or its variants. Control cells were transformed with the vector pTrc99A.

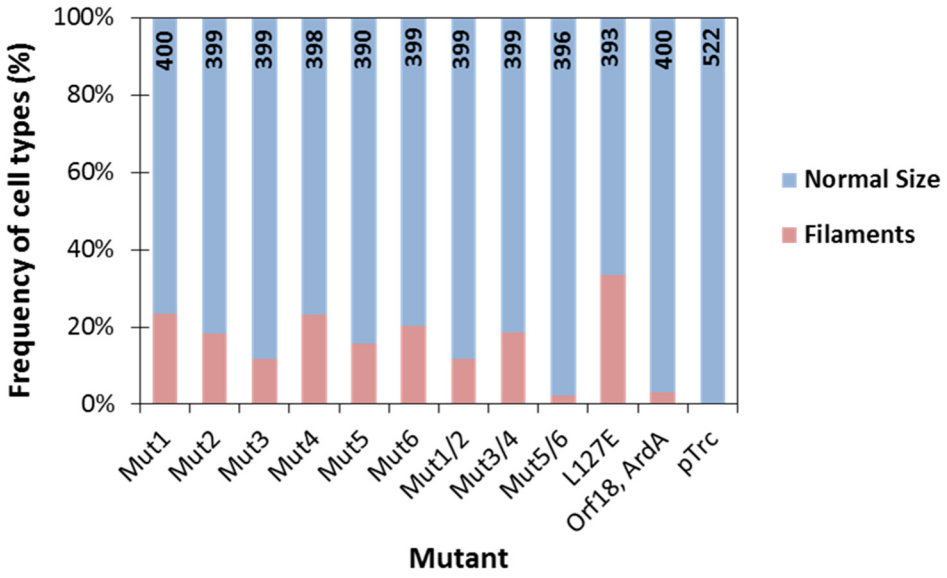


Figure A.16: Morphology types of *E. coli* NM1041 (m^{+}r^{+}, ClpXP^{-}) cells, transformed and induced with wild-type Orf18 ArdA or the mutant derivatives. Filaments were assigned to cells which were twice the unit size of the pTrc99A cell without the introduction of ArdA Orf18 or mutants thereof. The number at the top of each bar is equal to the total number of cells examined.

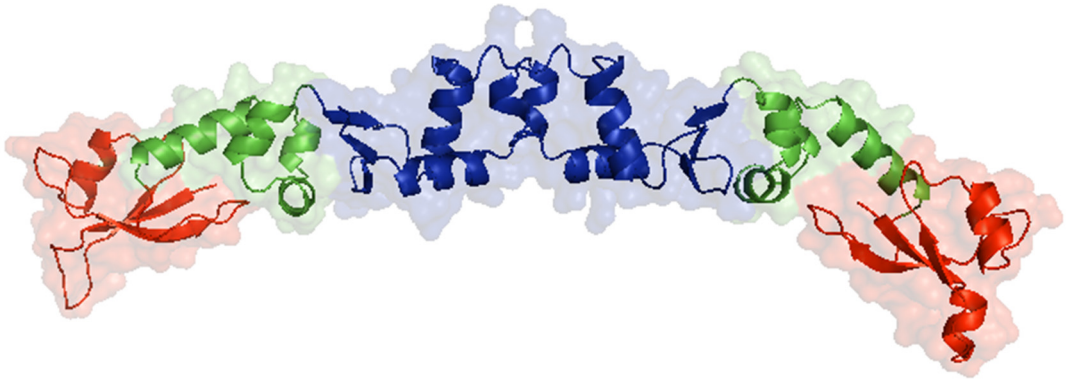
Filaments are observed for several mutants, however the total percentage of filaments does not exceed 23%, except for the culture expressing the mutant L127E. Cell cultures

expressing pTrc99A, wtArdA and Mut5/6 possess minute numbers of filaments, < 2% of the total population. The mutants Mut1/2, Mut3, and Mut5 acquired filaments which contributed to ~10-20% of the population whereas, the mutants Mut1, Mut2, Mut4, Mut6 and L127E all contained the highest percentage of filaments ~20% relative to the population size. The increase in filamentation appears to correspond to a decrease in 2AP resistance, with the mutants Mut1, Mut2, Mut4, Mut6 and L127E producing highly deleterious effects, as shown by the results from the RA assay and the bacterial growth curves.

Discussion

Structure to function relationships

The Orf18, ArdA monomer protein structure is made up of 3 domains. Each of the mutants targeted two negatively enriched clusters on each domain, producing Muts1-6. In addition combinations; Mut1/2, Mut3/4 and Mut5/6 which combine the two negatively enriched clusters were produced to examine the effect of charge from whole of domain I, II or III respectively.



Domain	Amino Acids
I	1-61
II	62-102
III	103-165

Figure A.17: The ArdA homodimer crystal structure, representing the three domains that produce the ArdA monomer structure, the table details which residues that belong to each of the domains.

The ArdA mutants Mut1 and Mut2 are located in domain I which dominates the N- terminal region of the protein. Both Mut1 and Mut2 appear to exhibit deleterious effects in the 2AP

assay. The toxic effect is a result of the restriction activity of the EcoKI complex being active, whereas modification has been inhibited, a so called M^-R^+ phenotype. The combination of both of the mutants to form the mutant Mut1/2, produces a 2AP resistance of up to 10μg/ml. This resistance is just slightly less than the 2AP resistance exhibited by the pTrc99A plasmid, producing an almost inactive ArdA mutant protein. This suggests that the ArdA protein might have additional binding sites with the HsdR subunits which do not include the core interaction of ArdA protein with the M.EcoKI enzyme.

The core interaction of the ArdA protein with M.EcoKI includes the interaction of the HsdS subunit with the central region of the ArdA protein. In chapter 3 a similar, central region was highlighted as an important interaction between the HsdS subunit with the Ocr protein. The central region is dominated by the C-terminal domain III of the ArdA protein and the mutants Mut5 and Mut6 target this particular region of the ArdA structure. The mutant Mut6 produced a deleterious effect similar to the Ocr mutant M12. *In vitro* studies showed that this deleterious effect correlates with mutants that possess weak antirestriction properties with a phenotype of $M^{-/+}R^{-/+}$.

However the Mut5 protein is resistant to high concentrations of 2AP, its resistance is similar to wtArdA although the growth rate of the *E. coli* cells is reduced, which could suggest some deleterious effects, this was confirmed by the formation of filaments (10-20% of the total cell population). This suggests that its antirestriction activity is not as efficient as wtArdA. The combination of the two mutants to form the Mut5/6 protein produces a mutant that is completely inactive. The mutant Mut5/6 shares the same phenotype as the pTrc99A plasmid in the 2AP assay. However upon leaky expression the Mut5/6 protein demonstrates a greater 2AP resistance compared to pTrc99A, as such the protein may possess some slight activity that remains undetectable once the mutant protein is over expressed.

The last domain is domain 2 which lies between the C-terminal (domain 1) and N- terminal (domain III). Mut3 and Mut4 represent the mutations which target this region and the two mutants appear to display some 2AP resistance. The Mut3 protein displays the same change in growth rate that was demonstrated for the Mut5 protein, therefore, Mut3 is active but with reduced antirestriction activity whereas the mutant Mut4 appears to be deleterious. This implies that the mutations in Mut4 are more detrimental to the activity of the ArdA protein then those on Mut3. The inability to inhibit both the antirestriction and antimodification activity fully causes *E. coli* cells with the Mut4 protein to become susceptible to restriction by the EcoKI enzyme. The combination of the mutants in Mut3/4 reduces $2AP^R$ to 20μg/ml,

the resistance of *E. coli* cells in the absence of antirestriction proteins is 10μg/ml, and as such Mut3/4 exhibits limited antirestriction and antimodification activity.

In summary, the mutants that target each domain can affect either the antirestriction or/and antimodification properties of wtArdA. A reduction in the number of acidic residues presented in domain I appear to reduce antirestriction activity to a greater extent. This suggests that the ArdA protein has multiple binding sites which are specific to the M.EcoKI and EcoKI complexes. This is reflected by the elongated nature of the ArdA protein. As a result, upon binding to the M.EcoKI core complex, domain I of the wtArdA homodimer flank either side of the complex whereas the wtOcr protein is completely engulfed by the enzyme.

On the other hand residues that reside in domain III appear to be important to the interaction between the mutant and the MTase core and as a result mutants that target this domain demonstrate weak binding affecting both the antirestriction and antimodification activity of these mutants. Domain II has the same effect as domain III but to a lesser extent. Interestingly the mutants Mut3 and Mut5 which span domain II and domain III appear to have the least effect on the activity of the protein. Both of these mutants retain almost the same activity as the wtArdA protein.

The mutant L127E which results in the ArdA monomer protein is deleterious similar to Mut1, Mut2, Mut4 and Mut6 and this suggests that, although the monomer protein has a weak binding affinity towards EcoKI enzyme, this is not sufficient and the wtArdA homodimer is a better inhibitor.

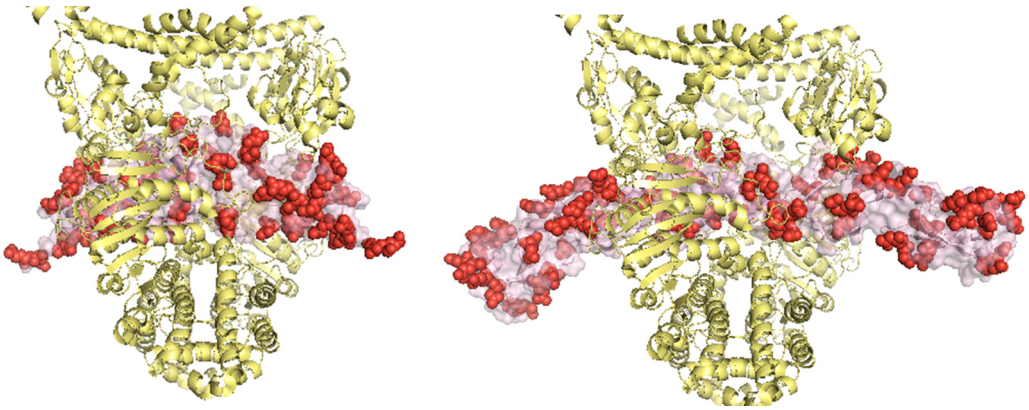


Figure A.18: Aligning both wtOcr and wtOrf18ArdA homodimer structures to dsDNA in a model of the interaction of wtOcr: M.EcoKI complex. The acidic residues of wtOcr and wtOrf18ArdA have been highlighted in red. Comparisons highlight the elongated form of the ArdA dimer which proceeds out of the M.EcoKI complex presumably revealing binding sites towards the HsdR subunits. However the wtOcr protein has been engulfed by the methylase core of the enzyme revealing very few sites of further interaction with the Endonuclease complex. Furthermore the wtOcr protein appears to have a more tight fit within the M.EcoKI compared to the wtArdA protein possibly justifying the smaller Kd value of it interaction.

Comparing results

Although the purification of the ArdA mutants was difficult the mutants Mut5, Mut6 and L127E were all successfully purified and characterised *in vitro,* by Dr GA Roberts and colleagues. In addition previous investigations measured the effect of phage infection on restriction proficient and restriction deficient cells in the presence of each of the mutants. Table A.1 compares the results obtained from the 2AP investigation presented in this chapter and the previous phage infection experiments.

Plasmid	Antirestriction	Antimodification	2AP (-IPTG)	2AP (+IPTG)
pTrc99A	n/a	0.9	20	20
wtOrf18	3.70E+03	122	80	<80
Mut1	2.90E+03	1	40	T
Mut2	2.80E+03	1.9	40	T
Mut1/2	2.40E+03	1	40	10
Mut3	2.90E+03	3.2	40	80
Mut4	2.80E+03	1.5	40	T
Mut3/4	4.00E+02	2	40	40
Mut5	2.00E+02	1	40	80
Mut6	2.80E+03	2.9	40	T
Mut5/6	1.10E+01	1	40	20
L127E	3.00E+03	1	40	T

Table A.2: The table compares the results from the 2AP RA assay for each of the ArdA mutants with results from a previous *in vivo* phage infection experiment (kindly donated by Dr GA Roberts). The 2AP results present the 2AP resistance exhibited by each of the mutants with and without IPTG inducing agent. The values are the concentrations reached before cell death for cultures in the presence of each of the mutants. 'T' represents those mutants which were toxic to the cells, mutant which caused cell death in *E. coli* NM1041 cells upon over expression.

One defining difference between the ArdA proteins and wtOcr, is that ArdA proteins can display differential antirestriction and antimodification activity, these differences are dependent on the type of RM enzyme and the origin of the ArdA protein.[70] The same differences in antirestriction and antimodification activity have been demonstrated by mutations to the Orf18 ArdA protein.

Mut5 and Mut3/4 possessed high levels of $2AP^R$ for Mut5 this was similar to wtArdA and for Mut3/4 a $2AP^R$ of 40µg/ml was recorded. However bacterial growth curves showed a reduction in the total cell population for Mut5 which indicates that the antirestriction and antimodification activity of these mutants is reduced. The phage assays suggest that the antimodification activity of the protein has suffered more compared to antirestriction.

Several ArdA mutants exhibit toxic effects towards *E. coli* NM1041 from the 2AP RA assay and this, together with the formation of cellular filaments, implies a phenotype that prevents methylation but not restriction M^-R^+. On the other hand, the results from the phage assay suggest that restriction is inhibited but methylation is prevalent, hence a m^+r^- phenotype.

ArdA monomers

The low solubility of the ArdA mutants meant that only the mutants Mut5, Mut6 and L127E were successfully purified. Analysis of the mutants L127E and Mut5 revealed that these mutants existed as ArdA monomers in solution rather than homodimer proteins (personal communication by Dr GA Roberts). The Mut5 protein targets acidic residues near the dimer interface, it seems that this has inadvertently caused a change in the local folding pattern causing the Mut5 protein to exist as a monomer. However studies have shown that at high concentrations the equilibrium between the monomer and dimer unit shifts towards increased dimer (personal communication by Dr GA Roberts). This could affect the activity of these mutants at high and low concentrations. Therefore the phage results represent the activity of Mut5 and L127E at low concentrations as monomer units. The phage results show that the monomers are unable to inhibit methylation but do prevent restriction of the EcoKI RM enzyme. However upon over expression in the 2AP RA assay the mutants Mut5 and L127E could exist as a mixture of monomer and homodimer, as such the phenotype that these mutants express in *E. coli* NM1041 cells would differ compared to the phage assay. The mutant Mut5, at high concentrations, retains a $2AP^R$ that is almost the same as wtArdA but the slight decrease in activity seems to affect the rate of cell growth. In contrast the mutant L127E is deleterious, suggesting that methylation is inhibited but restriction is prevalent, and this phenotype is atypical of weak binding ArdA mutants to the M.EcoKI/EcoKI enzyme.

Comparisons between the *in vivo* phage infection assay and the 2AP RA assays in Table A.1 above showed that a multitude of phenotypes have been exhibited by the ArdA mutants. The diversity of the activity demonstrated is much greater than that exhibited by mutants of wtOcr, this shows that even though both of these antirestriction proteins essentially mimic dsDNA the subtleties in the difference of their design and fold means that changes to the ArdA protein are highly sensitive to the activity exhibited however, the wtOcr protein is very robust to changes in its surface charge.

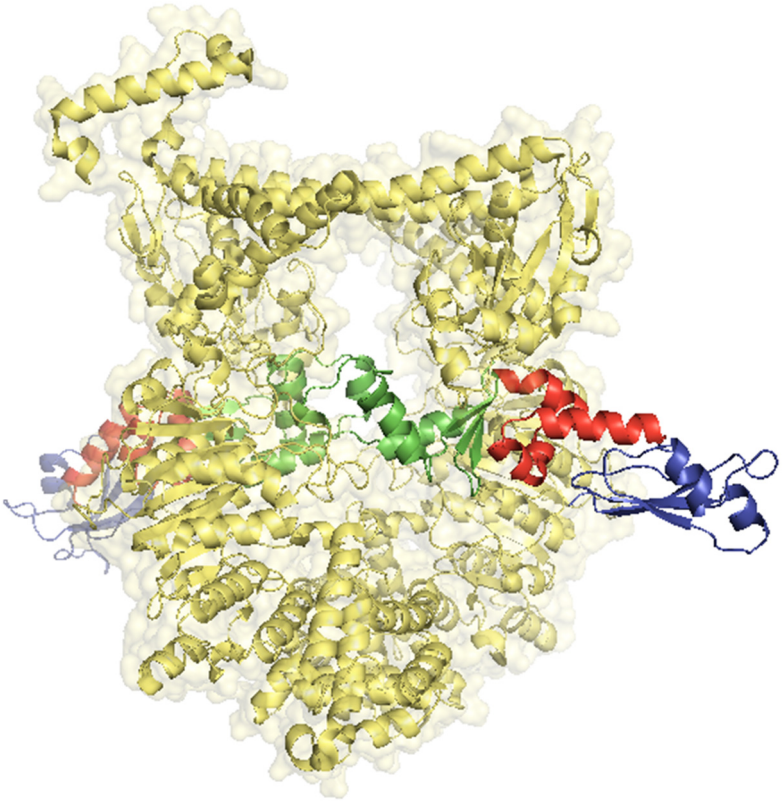


Figure A.19: The ArdA homodimer has been aligned to the dsDNA molecule within a model of the MEcoKI:DNA structure (pdb: 2Y7H). Each of the domains has been highlighted on the ArdA homodimer to highlight the possible location of each of these domains upon ArdA interaction with the M.EcoKI enzyme. Domain 3 is clearly seen to not be involved in the interaction of the ArdA protein to the methylase core of the enzyme.

An overall summary suggests that the acidic residues in domain 1 appear to be important for antirestriction behavior of the protein, the central region domain 3 interacts with the M.EcoKI core inhibiting modification and domain 2 spans both regions. The acidic residues located on Mut1 and Mut2 both contribute to antirestriction activity of the ArdA protein. The residues from mutant Mut4 appear to have a greater effect on M.EcoKI and the residues on Mut3 appear to affect both antirestriction and antimodification activity equally. Finally, the acidic residues of the mutant Mut6 seem important in the binding of the protein to the M.EcoKI core, however, the acidic residues found on Mut5 appear to produce an ArdA monomer which either produces a dimer at high concentrations or the monomer is able to bind with a 2:1 or 1:1 stoichiometry with the EcoKI complex.