A Biophysical Study of the DNA Charge Mimicry displayed by the T7 Ocr Protein

Augoustinos S. Stephanou

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh
2010

Declaration

I declare that this thesis was composed by me, and the research presented is my own except where otherwise stated.

Augoustinos S. Stephanou

2010

To my parents

To my father Stephis A. Stephanou from Famagusta (1943-2007). A man of extremes; of romantic disposition; a true Greek. His fervent and courageous work for the love of God and country, led to his capture and torture by the Turks. Before giving his last breath in a small cell, he let it be known that he wanted his grandchildren to be proud of him. For as he said: "Their grandfather did not yield. He stood up to the Turks". Now he stands amongst those, who stood against many.

To my mother Dr. Celia Bennett who being enormously selfless and loving by nature, has devoted her life to helping and supporting others.

Acknowledgments

I am greatly indebted to my supervisor Dr David Dryden for his support, wise guidance and for his understanding and patience when progress was slow.

I am grateful to Dr. John White for advice on molecular biology, Laurie Cooper for helping me on numerous occasions in preparing and purifying proteins; Prof. Alan Cooper and Margaret Nutley for ITC and also Dr. Anita Jones for making sure progress was afoot.

Special thanks however must go to Dr. Gareth Roberts without whom the lab would be a ship adrift. His depth and breadth of scientific knowledge combined with his tireless temperament of assisting all in the lab, has had a huge and positive impact on this thesis.

Finally, I thank my parents and Fr. John Maitland-Moir.

I would also like to mention that it has been an extra privilege and honour for me to study at Edinburgh University because my grandfather Dr. John Bennett (1915-2008) and his father too, also studied and received their doctorates at this institution.

This work was supported by a grant from the Biotechnology and Biological Research Council.

Abstract

The homodimeric Ocr protein of bacteriophage T7 is a molecular mimic of a bent double-stranded DNA molecule ~24 bp in length. As such, Ocr is a highly effective competitive inhibitor of the bacterial Type I restriction modification (R/M) system. Thus, Ocr facilitates phage infection of the bacterial cell to proceed unhindered by the action of the R/M defense system. The main aim of this work was to understand the basis of the DNA mimicry displayed by Ocr. The surface of the protein is replete with acidic residues, most or all of which mimic the phosphate backbone of DNA. Aspartate and glutamate residues on the surface of Ocr were either mutated or chemically modified in order to investigate their contribution to the tight binding between Ocr and the EcoKI Type I R/M enzyme. Single or double mutations of Ocr had no discernable effect on binding to EcoKI or its methyltransferase component (M.EcoKI). Chemical modification was then used to specifically modify the carboxyl moieties of Ocr, thereby neutralizing the negative charges on the protein surface. Ocr samples modified to varying degrees were analysed to establish the extent of derivatisation prior to extensive biophysical characterisation to assess the impact of these changes in terms of binding to the EcoKI R/M system. The results of this analysis revealed that the electrostatic mimicry of Ocr increases the binding affinity for its target enzyme by at least ~800-fold. In addition, based on the known 3-D structure of the protein, a set of multiple mutations were introduced into Ocr aimed at eliminating patches of negative charge from the protein surface. Specifically, between 5 and 17 acidic residues were targeted for mutation (Asp and Glu to Asn and Gln, respectively). Analysis of the *in vivo* activity of the mutant Ocr along with biophysical characterisation of the purified proteins was then performed. Results from these studies identified regions of the Ocr protein that were critical in forming a tight association with the EcoKI R/M system. Furthermore by comparing the relative contribution of different groups of acidic residues to the free energy of binding, the actual mechanism by which Ocr mimics the charge distribution of DNA has been delineated.

Table of Contents

Chapter 1. Introduction

1.1. Restriction endonucleases

The existence of Restriction-Modification (R/M) systems was first identified with the discovery that phages produced in one strain of a bacterial species would readily infect the same strain but only rarely infect a different strain successfully. Successful infections and propagation of viruses was thought to occur when the phage managed to preserve an “imprint” of the first bacterial strain, providing the phage DNA with protection (Arber, 1965).

This "imprint" was found to be host-controlled methylation or modification, whereby a specific sequence of DNA is recognised by a methyltransferase (MTase) enzyme that can modify the sequence by specific methylation (Smith *et al.,* 1972). The same sequence is recognised by the endonuclease (REase), which proceeds to cleave the phage DNA in the absence of the specific modification. Therefore foreign DNA entering the bacterial cell is recognised by the absence of specific methylation and is degraded by endonucleases, whilst the host DNA is not subject to degradation (Murray, 2000; Dryden *et al.,* 2001). This is why R/M systems are often regarded as bacterial immune systems.

Therefore the two enzymatic activities of R/M systems are methylation and restriction of DNA. R/M systems can exist as combined multi-subunit complexes that can perform both restriction and modification, or as separate enzymes that perform one or the other activity.

R/M enzymes can cleave double-stranded (ds) DNA at a fixed position from the recognition sequence or at a random position far away from the recognition sequence, by hydrolysing one phosphodiester bond in the backbone of each DNA strand. Cleavage occurs on the 5'-side of the phosphate group, leaving DNA fragments with 5'- phosphoryl and 3'-hydroxy termini, which can be rejoined by DNA ligase (Wilson and Murray, 1991) (Figure 1-1).

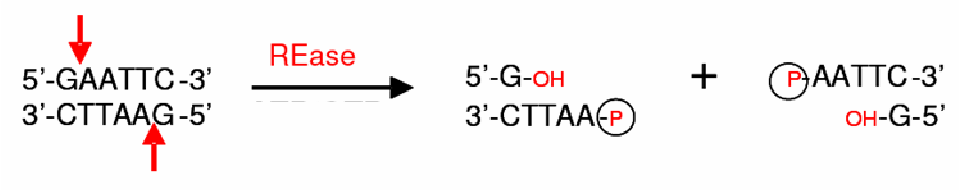


Figure 1-1. Hydrolysis of phosphodiester bonds and cleavage of ds DNA by a REase, in the presence of an energy source.

Methylation by the MTase enzyme of R/M systems usually occurs at adenines or cytosines and requires S-adenosyl-methionine (SAM) as methyl donor (Tock & Dryden, 2005). Methylation of adenines occurs at the N^6 position (Figure 1-2) while cytosines are methylated at the N^4 or C^5 position (Wilson and Murray, 1991).

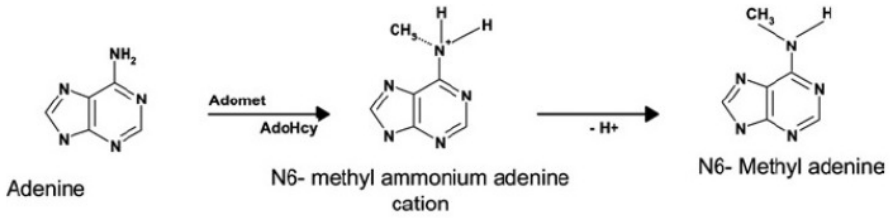


Figure 1-2. Reaction mechanism for the methylation of adenine by a MTase enzyme.
Adomet; S-adenosyl-methionine (SAM). AdoHcy; S-adenosylhomocysteine (SAH) (reprinted from Bheemanaik *et al.*, 2006).

Classification of R/M systems is based upon their enzyme composition, cofactor requirements, the nature of their target sequence, and the position of the site of DNA cleavage with respect to the target recognition sequence. Four different types of R/M systems have been characterised thus far and have been classified as type I, II, III and IV (Tock & Dryden, 2005). Some examples are shown in Table 1-1.

Table 1-1. Classification of R/M systems based on a range of different characteristics .

	Type I	Type II	Type III	Type IV
Example R-M system	EcoKI	EcoRI	EcoP1I	EcoMcrBC
Genes	*hsdR, hsdM, hsdS*	*ecorIR, ecorIM*	*mod, res*	*mcrB, mcrC*
Subunits	Three different subunits (R, M and S) combine to form $R_2M_2S_1$ and M_2S_1	Two different subunits (R and M) combine to form R_2 or M_1	Two different subunits (mod and res) combine to form mod_2res_2	Two different subunits are present, McrB and McrC
Enzyme activities	REase, MTase and ATPase	REase or MTase	REase, MTase and ATPase	REase and GTPase
Co-factors required for DNA cleavage	ATP, SAM, Mg^{2+}	Mg^{2+}	ATP, Mg^{2+} (SAM)	GTP, Mg^{2+}
Co-factors required for methylation	SAM	SAM	SAM	No methylation
Recognition sequence	Asymmetric and bipartite, e.g. EcoKI, 5'AAC(N_6)GTGC	Mostly symmetric, e.g. EcoRI, 5'GAATTC	Asymmetric, e.g. EcoP1I, 5'AGACC	Bipartite and methylated, e.g. EcoMcrBC, 5'RmC($N_{30-4000}$)RmC
Cleavage site	Variable locations 1000 bp from recognition site	Fixed location at or near the recognition site	Fixed location 25–27 bp from recognition site	Between methylated bases at multiple sites
DNA translocation	Yes	No	Yes	Yes

Adapted from Tock and Dryden, 2005.

1.1.1. Type II R/M systems

Most of the Type II R/M systems contain separate MTase and REase enzymes that recognise the same sequence (usually a 4-8 base pair (bp) palindrome). The cofactors required by MTase and REase are SAM and Mg^{2+}, respectively. The MTases generally act as monomers to modify specific bases of their target sequence on both DNA strands. Methylation of adenines occurs at the N^6 position while cytosines are methylated at the N^4 or C^5 position (Sisitia *et al*., 2004). Type II REases are active as dimers (each contributing a single strand break), while other Type II REases act as monomers or tetramers. There is evidence that many Type II

REase enzymes must bind multiple target DNA sites before restriction activity is achieved (Kruger *et al*., 1988; Halford *et al*., 2004). Type II REases have revolutionised molecular biology because they cut DNA at or within their target sequence, creating defined restriction products.

1.1.2. Type III R/M systems

Type III R/M enzymes are hetero-oligomers that consist of a "mod" subunit, which recognises and modifies the DNA target sequence, and a "res" subunit (restriction activity) that is only active when associated in a res_2mod_2 complex (Janscak *et al*., 2001). The mod subunit can act independently, requiring only SAM as cofactor, and conferring protection from restriction by only single strand methylation of the target recognition sequence; methylation occurs at the N^6 position of adenine. Restriction requires Mg^{2+} and ATP as cofactors and occurs when a Type III enzyme interacts with two inversely oriented copies of its 5-6 bp asymmetric recognition sequence (Meisel *et al*., 1992). DNA restriction is preceded by DNA translocation while the complex remains bound to its recognition sequence (Dryden *et al*., 2001). Only after stalled DNA translocation and/or collision of two separate res_2mod_2 complexes does restriction occur at 25-27 bp from the recognition site. The best studied examples are EcoP1I and EcoP15I (Bickle *et al*., 1993). With respect to EcoP15I, it has recently been shown that restriction occurs only after two enzymes have interacted with each other while still bound to their recognition sites which maybe separated by up to 3,500 bp. Fast-scan atomic force microscopy has shown that communication of the two enzymes occurs through diffusive DNA looping and ATPase driven translocation of the DNA between the enzymes (Crampton, 2007).

1.1.3. Type IV R/M systems

Type IV R/M systems are REases with no associated methyltransferase and recognise and cleave DNA that has been modified (methylated, hydroxymethylated and glucosyl-hydroxymethylated). Thus foreign DNA is attacked because it does

contain methylated bases in a particular sequence, rather than because it does not (Stewart *et al*, 2000). The best studied Type IV system is McrBC from *E. coli* K12. McrBC requires Mg^{2+} as cofactor and is the only nuclease known to use GTP for translocation and restriction of DNA (Pieper *et al*., 1997). The enzyme recognizes two methylated or hemimethylated sites, which are separated by between 40 to 3000 base pairs, and cleaves the DNA close to one of the two sites (Sutherland *et al*., 1992).

1.1.4. Type I restriction enzymes

1.1.4.1. Introduction

Type I R/M systems are hetero-oligomeric, multifunctional enzymes, capable of catalysing both restriction and modification. These enzymes recognise specific asymmetric bipartite nucleotide sequences, comprising two 3-5 bp specific sequences, separated by a nonspecific spacer of 6-8 bp. Depending on the methylation state of the target sequence, the enzyme will act as either an endonuclease or a methyltransferase. If the sequence is methylated, no activity takes place. If the sequence is hemimethylated, as is the case with newly replicated DNA, the enzyme methylates the unmethylated strand (Vovis *et al.,* 1974). If the sequence is unmethylated however, the enzyme acts as an endonuclease. Specifically, EcoKI will pull the DNA towards itself from both directions whilst remaining bound to the target site (Bickle *et al.,* 1978; Davies *et al.,* 1999; Yuan *et al.,* 1980). Cleavage of the DNA then occurs when two translocating enzymes collide or translocation is generally impeded, which occurs anywhere between at least 40 bp (Dreier *et al.,* 1996b) and several thousand base pairs from the target site (Studier and Bandyopadhyay, 1988). Endonuclease activity requires ATP, Mg^{2+} and SAM as cofactors, whereas methyltransferase activity requires SAM as the cofactor and methyl donor (Murray, 2000).

Type I enzymes have been subdivided into five families based on genetic complementation, DNA hybridisation and antibody cross-reactivity (Cajthamlova *et al.*, 2007).

1.1.4.2. Subunit composition

Type I enzymes are made up from three subunits; HsdM (~50 kDa), HsdS (50-60 kDa) and HsdR (~140 kDa) (Dryden *et al.,* 2001; Bickle and Kruger, 1993). The acronym Hsd stands for "host specificity of DNA". HsdS is the specificity subunit, and includes two target recognition domains (TRD), which confer the enzyme with DNA target sequence specificity. HsdM is the methylation subunit, which harbours the binding site for SAM as well as the active site for methylation. HsdR is the restriction subunit that is responsible for endonuclease activity, DNA translocation as well as containing the active site for ATP hydrolysis.

The stoichiometry of a fully active type I restriction enzyme is $R_2M_2S_1$ (~440 kDa) (Dryden *et al.,* 1997). However, the methyltransferase enzyme lacking endonuclease activity is also present *in vivo*, with a functional stoichiometry of M_2S_1 (Dryden *et al.,* 1993; Suri and Bickle, 1985; Taylor *et al.,* 1992). The activities of type I R/M systems are summarised in Figure 1-3.

Interestingly, the *hsdR* gene is expressed from a separate promoter, whereas the *hsdM* and *hsdS* genes are expressed from a common promoter, hence the expression of these subunits is independent (Wilson and Murray, 1991). This independent expression is to ensure the *hsdR* gene is not expressed at times when the host DNA may be lacking the required methylation modifications that would protect the host from indiscriminate DNA restriction in the presence of the fully active nuclease enzyme (Wilson and Murray, 1991; Dryden *et al*., 1997).

The host's chromosomal DNA is normally immune to restriction because the unmethylated strand produced by DNA replication is methylated before the next round of replication. In some instances however, specific methylation may be lacking, as is the case when (i) the host DNA is damaged by agents such as UV light, nalidixic acid and 2-aminopurine (Thoms and Wackernagel, 1984); or (ii) when the *hsd* genes encoding EcoKI (for example) are transferred to a new recipient cell that lacks the protective modification of the DNA target sequences recognised by the EcoKI system. In both cases the host DNA would become susceptible to degradation

if it were not for the phenomenon of "restriction alleviation". This phenomenon has been attributed to the ClpXP protease (Makovets *et al*., 1999).

ClpXP is a two component system consisting of ClpP (the proteolytic core) and ClpX (the ATPase domain). ClpX is composed of a hexamer, while ClpP contains two rings of seven subunits (the rings placed on top of each other). The binding of ClpX to ClpP is mediated by a hydrophobic loop referred to as the IGF loop (Kim *et al*., 2001).

The ClpXP protease competes with MTase for the interaction with HsdR, delaying the production of the complete EcoKI nuclease complex (Makovets *et al*., 1998). It has been shown that ClpXP degrades the R subunit from EcoKI so that the latter cannot cleave the host's chromosomal DNA. ClpP digests the protein that is unfolded and translocated by ClpX into uniform peptide fragments of six to eight residues (Choi and Licht, 2005). The HsdR domain is only recognised by the protease after the EcoKI complex has embarked on its restriction pathway, but before any damage has been inflicted on the unmodified chromosomal DNA (Figure 1-4).

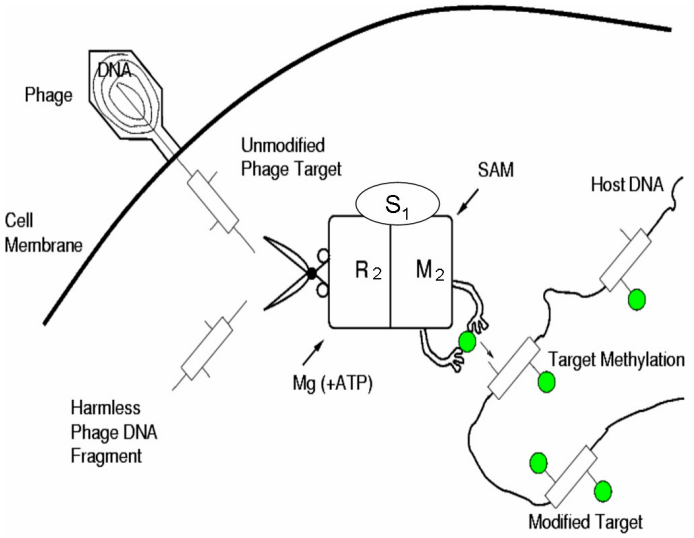


Figure 1-3. Type I restriction/modification systems.

The type I restriction enzymes recognise the methylation state of their specific target sequence. When the enzyme comes across fully methylated DNA, it recognises this DNA as being part of the host bacterial genome and no action is taken. If the DNA is hemimethylated, it is recognised as newly replicated bacterial host DNA, and the enzyme methylates the unmethylated strand. Invading bacteriophage DNA however, that lacks this specific methylation, is recognised as foreign and is therefore cleaved into harmless DNA fragments.

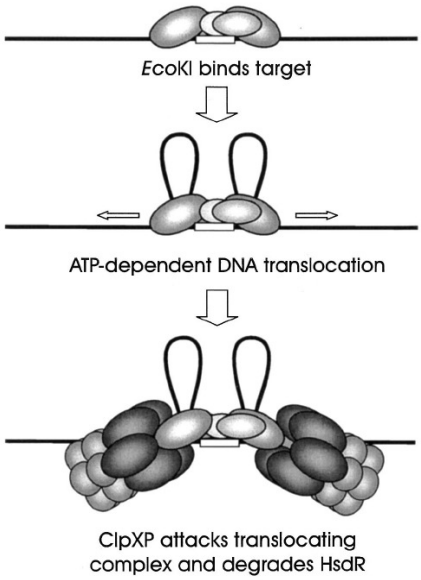


Figure 1-4. Model for the mechanism of ClpXP-dependent restriction alleviation.

When a type IA or IB R/M system binds to an unmodified target sequence, translocation is initiated and ClpXP recognizes the HsdR subunit of the complex. ClpXP degrades HsdR, preventing further translocation and endonuclease activity. Reprinted from Murray, 2000.

1.1.4.3. DNA recognition and specificity

The HsdS subunit of EcoKI confers sequence specificity to the enzyme by recognising the nucleotide sequence AAC(N_6)GTGC, where N can be any base (Kan *et al.,* 1979). HsdS contains two target recognition domains (TRD) each of ~160 amino acids. The N-terminal TRD recognises the 5′ (AAC) section of the bipartite sequence while and C-terminal TRD recognises the 3′ (GTGC) part (Fuller-Pace *et al.,* 1984; O'Neill *et al.,* 1998).

Domain swapping experiments, in which TRDs from members of the same family were combined, resulted in novel specificities, indicating that each TRD functions independently of the other (Fuller-Pace and Murray, 1986; Nagaraja *et al*., 1985; Cowan *et al*., 1989).

In the absence of a crystal structure of any type I enzyme in complex with DNA, little is known about the amino acids of TRDs involved in binding the target DNA. However, sequence alignments and secondary structure predictions of about 51 type TRDs, suggested that type I TRDs have a common tertiary structure with that found in the solved structure of the TRD of the type II methyltransferase HhaI (Sturrock and Dryden, 1997). This has led to the prediction of a DNA interaction region comprising a loop-β-strand-loop within the TRD that could make specific contacts with bases within the major groove (Chen *et al.,* 1995). This is supported by random and site-directed mutagenesis experiments (O'Neill *et al.,* 1998).

1.1.4.4. Methylation

Type I enzymes are able to catalyse the transfer of the methyl group from SAM to the N^6 position of specific adenine residues in their respective target sequences. In the case of EcoKI for example, with a target recognition sequence of A<u>A</u>C(N_6)G<u>T</u>GC), the underlined A is a methylation site, whereas the underlined T represents the site of methylation for adenine on the complementary DNA strand.

Type IA and IC families have a strong preference for methylating only hemimethylated target sites, and cleaving only unmethylated DNA (Dryden *et al.,* 1993; Janscak *et al.,* 1996; Suri and Bickle, 1985). EcoKI methyltransferase

(M.EcoKI) has a much greater binding affinity for its specific target sequence compared to nonspecific DNA. It was also shown that the methylation state of the target sequence had no effect on this binding affinity. This suggests that the preference for methylating hemimethylated over unmethylated DNA is due to recognition of the methylation state after binding has occurred (Kelleher *et al.,* 1991; Powell *et al.,* 1993, 1998a).

DNaseI and exonuclease III footprinting experiments have shown that the M_2S_1 complexes of both EcoKI and EcoR124I cover 25-30 bp of DNA (Mernagh and Kneale, 1996; Powell *et al.,* 1993, 1998a). It has also been found that the partially assembled form (M_1S_1) protects the same length of DNA as the active form, indicating that both of the HsdM subunits must be positioned on either side of the HsdS subunit. Evidence that the HsdM subunits enwrap the DNA by a large-scale movement has been found by small angle X-ray scattering of EcoR124I methyltransferase (Taylor *et al.,* 1994). Recent experiments using electron microscopy also suggest that the M.EcoKI completely envelops its DNA target (Kennaway *et al*., 2009).

Methylation interference experiments have shown that M.EcoKI interacts with its DNA recognition sequence *via* contacts in the major grove (Powell and Murray, 1995). Methylation occurs by a base-flipping mechanism whereby the target base is rotated 180° out of the DNA helix, into the catalytic site of the enzyme (Roberts and Cheng, 1998).

The type I enzyme HsdM subunits are essential for DNA sequence recognition and binding, even though they lack specificity determinants and can be exchanged between closely related R/M systems. This is supported by the finding that the presence of SAM increases the affinity of the methyltransferase for the specific DNA target site by 4-20 fold but has little affect on its affinity for nonspecific DNA. Furthermore the HsdS subunit is not soluble enough and not capable of forming a stable complex with DNA in the absence of the HsdM subunits (Powell *et al.,* 1998a).

1.1.4.5. DNA translocation and restriction

DNA cleavage by Type I R/M systems occurs at variable positions remote from the specific target sites of the enzymes (Horiuchi and Zinder, 1972). Studier and Bandyopadhyay (1988), have suggested that whilst the enzyme remains bound to the target recognition sequence, DNA is translocated towards the complex from both directions using energy from ATP hydrolysis. Cleavage occurs when translocation of DNA causes two translocating complexes to meet. According to this model, the trigger for restriction is the hindrance of translocation, which agrees with the finding that cleavage occurs when two restriction enzymes from different families collide or when the enzyme encounters an unpassable Holliday junction (Janscak *et al.,* 1999a). After cleavage, the enzymes remain bound to the target sequences and do not turn over in the endonuclease reaction (Yuan *et al*., 1980). Bianco *et al.,* 2009 however, claim that Type I enzymes do turn over (*i.e*., the complexes dissociate intact from their DNA targets after cleavage, in order to bind and cleave other possible DNA targets) and hence should be characterised as true enzymes.

Type I restriction enzymes also exhibit the peculiar characteristic of continuing to hydrolyse ATP long after they have carried out their restriction reaction (Bourniquel and Bickle, 2002; Dreier and Bickle, 1996; Eskin and Linn, 1972). Continued ATP hydrolysis has been suggested as a potential method of cell self-destruction, due to massive ATP consumption. This would remove the infected cell from the population, thereby further reducing the chance of phage propagation (Dreier and Bickle, 1996a).

Atomic force microscopy experiments have shown two EcoKI molecules dimerising on DNA prior to addition of ATP and translocation (Ellis *et al*., 1999; Berge *et al.,* 2000; Neaves *et al.*, 2009). According to this model (Neaves *et al.*, 2009) (Figure 1-5; panel b), which builds on the previous model by Studier and Bandyopadhyay (1988), an EcoKI enzyme binds DNA at a random sequence and then by linear diffusion, hopping and jumping locates its target sequence. A second EcoKI molecule then binds to the first one forming a dimer. This EcoKI dimer then binds another flexible DNA region *via* random collision, forming a loop. Once the loop is formed the second EcoKI molecule scans the DNA for the second target

sequence. Only after both target sequences are identified is translocation initiated by both enzymes. As a result, the loop between the monomers becomes fully contracted and cleavage ensues due to DNA stalled translocation.

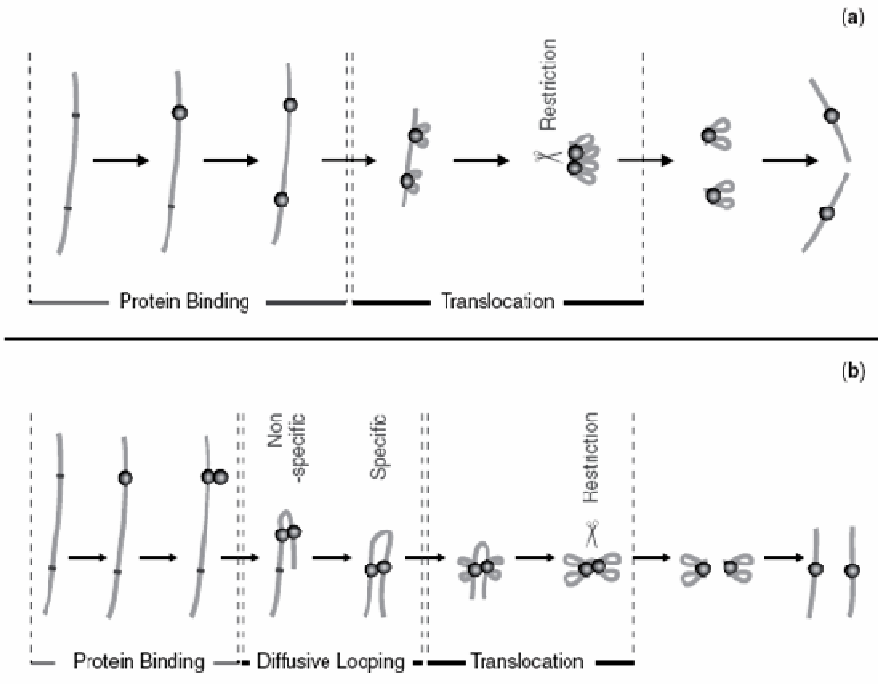


Figure 1-5. Diagrammatic representation of EcoKI binding, translocation and restriction on DNA.
Panel (a) shows the previous model (Studier and Bandyopadhyay, 1988). EcoKI monomers bind individually to each site and translocation occurs independantly from both occupied sites. When the monomers meet restriction occurs due to stalled translocation. Panel (b) shows a new model which builds on the previous one (Yuan *et al.*, 1980; Berge *et al.,* 2000; Neaves *et al.*, 2009). EcoKI binds to one site and then a second "monomer" binds to the same site to form a dimerised complex. The EcoKI dimer then forms diffusive loops with non-specific regions of DNA until it is stabilized 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 (reprinted from Neaves *et al.*, 2009).

DNA-dependent dimerisation however is not an absolute requirement for translocation. It has been shown that only one site is required for translocation (Garcia and Molineux, 1999; Janscak *et al.,* 2000). In addition translocation by individual EcoR124I enzymes has been studied by single molecule techniques (Seidel *et al.,* 2004). However, dimerisation could enhance restriction by facilitating cooperation between two complexes, allowing the enzymes to remain stationary and

translocate DNA, without having to pull another R/M complex through a crowded solution of macromolecules.

The translocation activity of EcoKI has been studied by measuring the rate at which T7 phage DNA is pulled into an *E. coli* cell, where EcoKI is the only method of translocation (*i.e.*, in the absence of the host RNA polymerase which was inhibited; that would otherwise bind and translocate the phage genome within the bacterial cell). In this study, a single EcoKI target site was engineered at the end of the T7 phage genome, and hence enters first into the bacterial host (Garcia and Molineux, 1999). It was found that the entire ~39 kb genome could be translocated at a rate of 100-200 bp/s. This is similar to the *in vitro* results that found the rate to be 200-400 bp/s for EcoKI and EcoR124I (Studier and Bandyopadhyay, 1988; Firman and Szczelkun, 2000).

Single molecule imaging using magnetic tweezers has been used to study the translocation of the type I restriction enzyme, EcoR124I (Seidel *et al.,* 2004). In EcoR124I, the second R subunit is more weakly bound than the first. This allows the formation of an enzyme with either one or two motor subunits, although both subunits are required for cleavage. It was found that EcoR124I with a single R subunit could translocate around 1300 bp DNA at ~550 bp/s. This is in agreement with the translocation rate measured by Firman and Szczelkun, (2000), of ~ 400 bp/s. When two R subunits were present, EcoR124I translocated over three times more DNA at about twice the translocation rate (Seidel *et al.,* 2004). The results indicate that the two motors work independently and without coordination. Furthermore, switches in translocation direction were also seen for the single motor complex, generating positive supercoils ahead of the translocating complex. The fact that around 11 bp were translocated per supercoil introduced, suggests that the R subunit follows the DNA helical pitch during translocation.

Type I HsdR subunits have conserved sequences that share homology with ATP-dependent helicases and putative helicases. These motifs are referred to as DEAD-box motifs (Gorbalenya and Koonin, 1991). It is therefore possible that these conserved motifs are responsible for translocating the DNA by a mechanism similar to helicases. Mutagenesis studies have shown that each of these seven DEAD-box

motifs is essential for restriction and DNA translocation *in vivo* and both ATP hydrolysis and DNA cleavage *in vitro* (Webb *et al*., 1996; Davies *et al*., 1998, 1999a). These finding indicate that the DEAD-box motifs may be coupling ATP hydrolysis to DNA translocation. Mutations in these DEAD-box motifs did not abolish DNA nicking activity nor did they affect the conformational changes that confer tight binding to the target sequences (Davies *et al.,* 1998). Although helicases also possess strand separating ability, type I restriction enzymes are thought to translocate double-stranded DNA without opening the duplex (Janscak and Bickle, 2000). This conclusion is based on the finding that when the DNA strands were cross-linked together at a low level by psoralen, enzyme translocation was not impeded (reviewed in Singleton and Wigley, 2002).

Another conserved region of HsdR, called motif X, shares sequence homology with motifs associated with endonuclease activities in type II, type III restriction enzymes (Pingoud and Jeltsch, 1997; Davies *et al.,* 1999b) and the RecB family of nucleases (Aravind *et al*., 1999). The sequence homology implies that these systems may be utilising a common mechanism for the hydrolysis of phosphodiester bonds. Mutations within this region did not affect either ATP hydrolysis or translocase activities of EcoKI. However, these substitutions did block the nicking and cutting activities of the enzymes, consistent with their proposed function in cleavage (Davies *et al.,* 1999a).

If the enzyme follows the helical path of the DNA during translocation, as occurs with helicases, the three points of contact made by the enzyme to the DNA (two contacts by the HsdR subunits and one by the MTase core) impart topological stresses. As a result, DNA ahead of the translocating complex would become overwound and the DNA behind would become underwound (Davies *et al.,* 1999b). As the built up topological stress does not prevent translocation, it was thought that this would need to be relieved by nicking. However, the nicking ability of motif X was shown not to affect translocation *in vivo* and *in vitro* and no other topoisomerase activity has so far been observed (Davies *et al.,* 1999b; Janscak *et al.,* 1999b).

Footprinting studies have found that the footprint of EcoKI restriction endonuclease (42-46 bp), in the presence of SAM, shortens upon the addition of ATP

to the same size as the methyltransferase (30 bp), even though HsdR remains bound to the complex. It has been suggested that in the presence of cofactors, the tight binding of the target site is weakened in HsdR, allowing the conformational changes required for translocation (Powell *et al.,* 1998b).

A model of the domain structure of a type I restriction enzyme is represented in Figure 1-6.

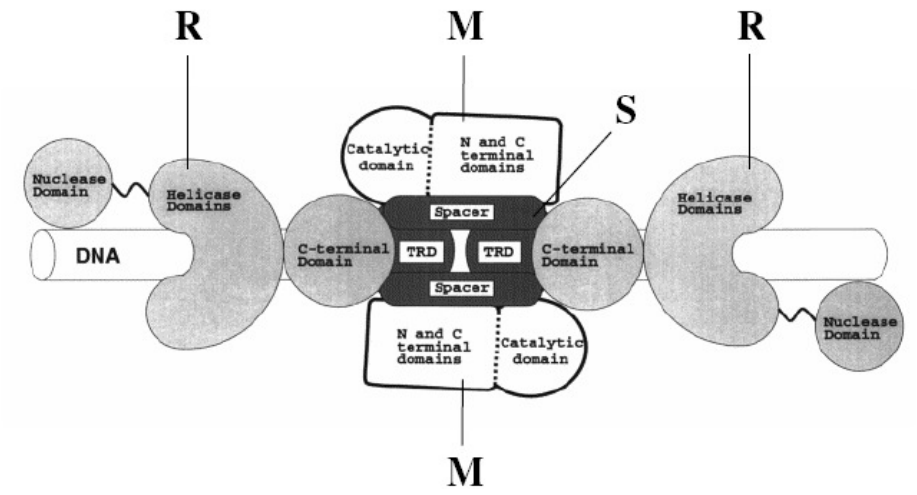


Figure 1-6. Model of the domain structure of a type I restriction enzyme.
The S subunit (black) comprises two target recognition domains (TRDs) which recognise and bind to specific target sites on DNA (white cylinder). The TRDs are joined by two spacer regions. The M subunits (bold outline, white) are made up of a catalytic domain and N- and C-terminal domains. The N-terminal domain is involved in recognition (of the DNA methylation state) and methylation, whereas the C-terminal domain is involved in the assembly of the S subunit. The R subunits (light grey) comprise an endonuclease domain, responsible for cleavage, a helicase domain, and a C-terminal domain involved in the association with the M and S subunits. Image adapted from Davies *et al.,* 1999b.

1.1.4.6. Structures and models of Type I R/M systems

The overall structures of M.EcoKI or the complete EcoKI nuclease enzymes are still unknown although several models have been proposed based on the known X-ray crystal structures of individual M and S subunits.

The crystal structures of HsdS from *Methanococcus jannaschi* and *Mycoplasma genitalium* (Calisto *et al*., 2005; Kim *et al*., 2005) have recently been

published. The S subunit from *M. jannaschi* is a monomer composed of two TRDs separated by a conserved region (CR) (Figure 1-7, panel A). The structure is characterised by four continuous structural motifs: an N-terminal globular TRD (TRD1), a long-helical CCR (central CR), the second globular TRD (TRD2) and the C-terminal long-helical DCR (distal CR). TRD1 and TRD2 have a very similar fold, each interacting extensively with either of the two CR helices; TRD1 interacts with the CCR helix and TRD2 with the DCR helix. CRs form two 10-turn helices running antiparallel to each other with a large number of hydrophobic contacts along the two helical coils. The crystal structure of HsdS from *Mycoplasma genitalium* shows a very similar fold to the HsdS structure from *Methanococcus jannaschi* even though the sequence identity of the two subunits is below 20% (Calisto *et al.*, 2005). A model structure of a Type I R/M endonuclease describing the proposed assembly of the subunits is also shown (Figure 1-7, panel B) (Kim *et al*., 2005).

A proposed structural model of the Type IC M.EcoR124I DNA methyltransferase by Obarska *et al*., (2006) is shown in Figure 1-8. The model is comprised of two HsdM subunits, the HsdS subunit and a substrate DNA molecule. The crystal structures of two HsdM subunits from EcoKI and StySJI are PBD ID 2AR0 and PBD ID 2OKC, respectively.

Recently, Kennaway *et al*., (2009) have produced an ~18Å resolution structure of M.EcoKI by negative-stain electron microscopy, together with an approximate atomic model of it in complex with Ocr (an antirestriction protein) and a DNA substrate (Figure 1-9). The model emphasises the complete engulfment of the bound substrate and the "clamp" method of binding. More specifically, the model suggests that the N-terminal domains of each HsdM subunit come in contact forming a bridge that must open in order to allow DNA to enter and be recognised by HsdS.

Taking both models (Obarska *et al*., 2006; Kennaway *et al*., 2009) into consideration, it is probable that the structure of Type I MTases fluctuates between an open form (Figure 1-8) and a closed form (Figure 1-9). This opening and closing motion is facilitated by twisting and bending of the coiled-coil CR region of the HsdS subunit (Kennaway *et al*., 2009).

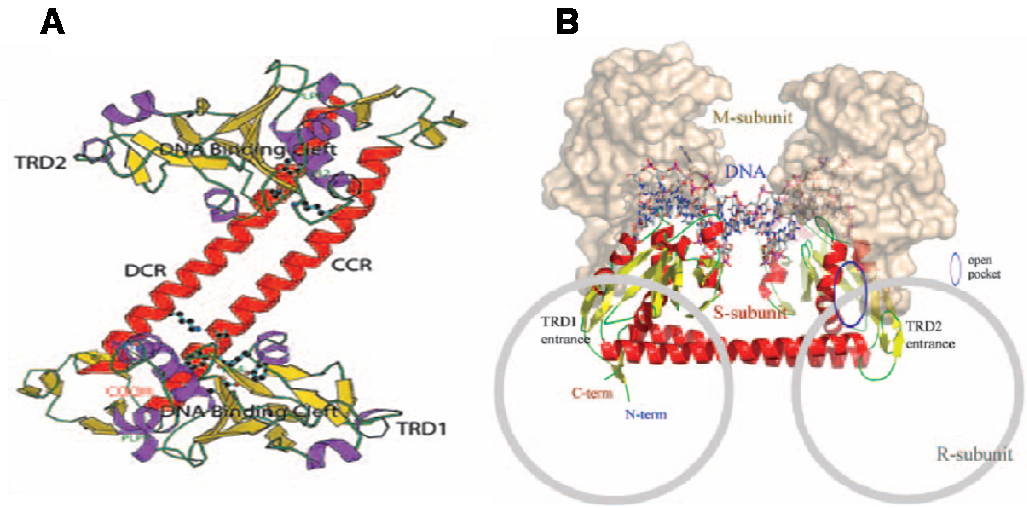


Figure 1-7. Crystal structure of the Type I S subunit from *Methanococcus jannaschi* and a proposed model for the EcoKI subunit assembly.
Panel A shows the S subunit crystal structure that consists of four successive structural domains: the globular TRD1, a long helical CCR domain, the globular TRD2, and a C-terminal DCR helix. Panel B (Model for subunit assembly); The S subunit is drawn as a ribbon diagram and the DNA is shown in sticks. The M subunit of the *Taq*I-MTase is drawn as a surface model and is docked around the TRDs of the S subunit. The possible R subunit regions are drawn as grey circles (reprinted from Kim *et al*., 2005).

Figure 1-8. Structural model showing the M.EcoR124I DNA methyltransferase and DNA substrate.
The HsdS subunit is shown in green. The two HsdM subunits are shown in yellow and orange and the substrate DNA in light blue. The adenines of the recognition sequence, which are to be methylated are shown 'flipped-out' coloured red (reprinted from Obarska *et al*., 2006).

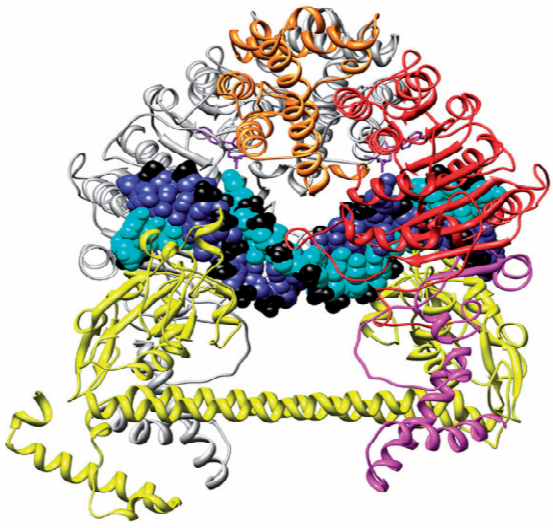


Figure 1-9. Model of M.EcoKI in complex with a DNA substrate and SAM.
Colour coding: Yellow, HsdS; Grey, HsdM; Blue and cyan, DNA; Black, phosphate groups of DNA; Purple, SAM; Orange, N-terminal domain of HsdM; Red, Catalytic domain of HsdM; Magenta, C-terminal domain of HsdM (adapted from Kennaway *et al.*, 2009).

The characterisation of the HsdR subunit of EcoR124I by dynamic light scattering, analytical ultracentrifugation and small angle neutron scattering had indicated that the subunit is monomeric, globular and moderately compact (Obarska *et al.*, 2008). Indeed, the recent crystal structure of the HsdR subunit of EcoR124I (Lapkouski, *et al.*, 2009) revealed four globular domains in a square-planar arrangement. The four domains are: an endonuclease domain, two RecA-like helicase domains (1 and 2) and a helical domain with no apparent structural relatives (Figure 1-10).

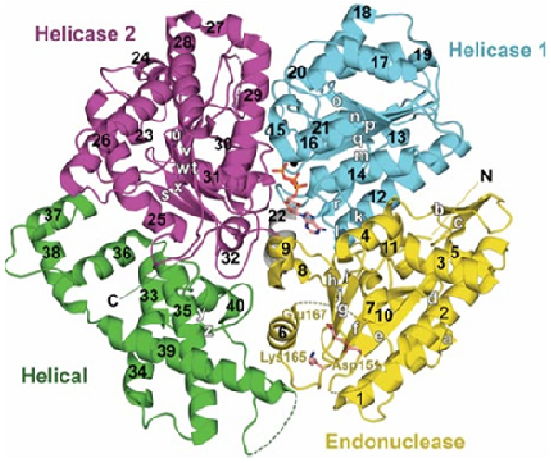


Figure 1-10. Crystal structure of the HsdR subunit of EcoR124I.
Four domains can be seen: Endonuclease (yellow), helicase 1 (blue), helicase 2 (purple) and helical domain (green) (reprinted from Lapkouski *et al.*, 2009).

1.2. Anti-restriction strategies

Phages and transmissible plasmids have evolved several mechanisms in order to avoid restriction by R/M systems (Bickle and Kruger, 1993; Tock and Dryden, 2005).

1.2.1. DNA sequence alteration

Restriction occurs when REases bind and assess the modification state of one or more specific target DNA sites. Phages and plasmids can therefore evade restriction by removing or changing these potential sites on their genome. In T3 and T7 phages for example, a reduction of recognition sites is sufficient to confer protection against cleavage by EcoRII. The EcoRII enzyme must bind two sites for restriction to occur but is unable to do so if the distance between such sites is very large (Bickle and Kruger, 1993). Type III R/M enzymes require sites that have a head to head orientation. However all EcoP1I sites in phage T7 are in the same orientation (Meisel, *et al*., 1992). Some phages make their potential target sites unrecognizable to R/M enzymes by incorporation of unusual bases within their DNA. For example, phages that infect *B. subtilis* replace thymine with 5-hydroxymethyluracil (Warren, 1980).

1.2.2. Transient occlusion of restriction sites

Another method of avoiding restriction is to cover restriction sites with DNA binding proteins. Phage P1, for example, encodes the proteins DarA and DarB, which are co-injected with the phage DNA and occlude Type I restriction sites by binding the phage DNA (Iida *et al*., 1987).

1.2.3. Subversion of restriction-modification activities

Bacterial R/M activities can by subverted by either stimulating the host MTases to modify phage DNA or by destroying REase cofactors. The phage λ protein Ral, for example, stimulates Type IA MTases (Zabeau *et al*., 1980) whereas phage T3 encodes a SAM hydrolase that destroys intracellular SAM soon after infection (Studier and Movva, 1976).

1.2.4. Inhibition of Type I R/M enzymes

Type I R/M enzymes can be prevented from destroying phage and plasmid DNA by specialised proteins that directly bind and occupy the catalytic sites of these enzymes. The gene 0.3 protein of phage T7 for example, also known as the overcome classical restriction (Ocr) protein, is the first protein to be expressed by T7 phage as it enters a bacterium (Studier, 1975). By mimicking the size, shape and charge of 24 bp of DNA, the Ocr protein is able to bind, nearly irreversibly, to Type I enzymes thereby inactivating them (Kruger *et al.*, 1989; Brandyopadhyay *et al*., 1985). Ocr, however, is inactive against Type II R/M enzymes.

Conjugative plasmids and transposons, such as IncN, IncP plasmids and the Tn916 transposon, encode antirestriction proteins that help them evade host restriction. Such an example is the ArdA (alleviation of restriction of DNA) and ArdB family of proteins. These proteins, which like Ocr are small and highly acidic, are amongst the first to be expressed by an immigrant plasmid as it enters a recipient cell during conjugation. There is, however, no significant sequence similarity between the ArdA protein and Ocr (Wilkins, 2002). Although the exact mechanism is not known, there is evidence to suggest that ArdA proteins inhibit both MTase and REase of Type I R/M enzymes (Zavilgelsky and Rastorguev, 2004; Serfiotis-Mitsa *et al.*, 2008; McMahon *et al*., 2009).

1.2.5. DNA mimicry by proteins

Before looking at examples of proteins mimicking DNA, the structure of DNA is briefly introduced.

1.2.5.1. DNA structure

DNA (deoxyribonucleic acid) is found in all cellular life forms and contains all the biological information to replicate and sustain that particular organism. It is a linear, unbranched polymer comprised of monomeric subunits called nucleotides. The nucleotide copy number in any single DNA molecule can range from hundreds to even millions of units in length. As can be seen in Figure 1-11, there are four chemically distinct nucleotides, made up of three different components: A 2-deoxyribose pentose sugar; a phosphate group; and a nitrogenous base, one of cytosine (C), thymine (T), adenine (A) or guanine (G) (Figure 1-11, panel B).

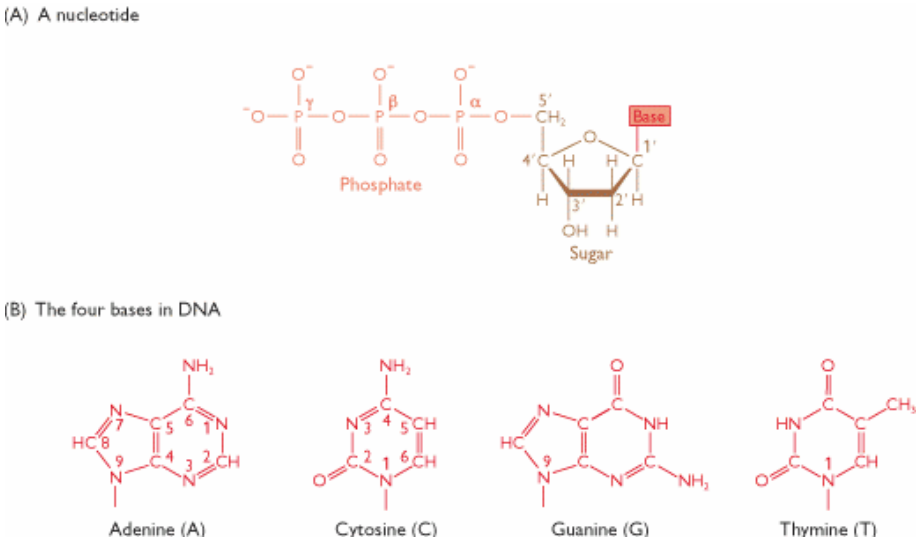


Figure 1-11. DNA components
Panel A: The general structure of a deoxyribonucleotide, found in DNA. Panel B: The four bases that occur in deoxyribonucleotides (Brown, 2002).

The nitrogenous base is attached to the 1'-carbon of the sugar by a β-N-glycosidic bond whereas the phosphate group is attached to the 5'-carbon. Polymerisation occurs with the linkage of individual nucleotides by phosphodiester bonds between their 5'- and 3'-carbons (Figure 1-12). The polymerisation reaction entails the release of a pyrophosphate from one of the nucleotides and a hydroxyl group from the 3'-carbon of the other nucleotide. This gives rise to a polymer that has chemically different termini and therefore a directionality, abbreviated as 3'→ 5' or 5'→ 3.

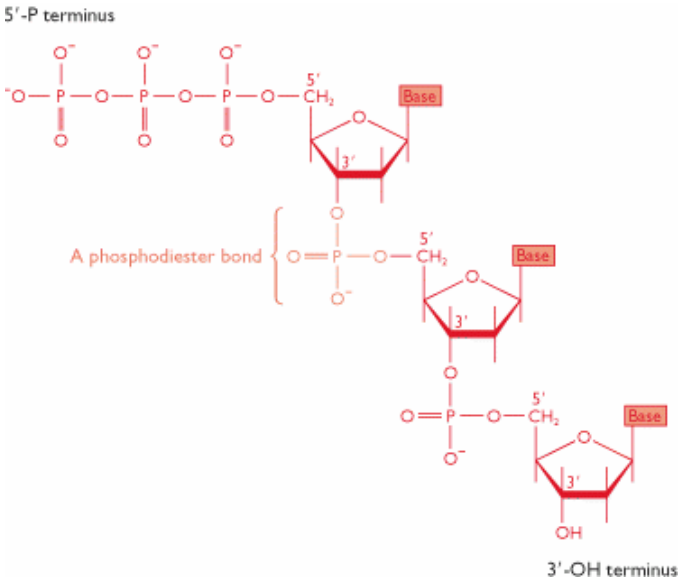


Figure 1-12. A short DNA polynucleotide showing the structure of the phosphodiester bond (Brown, 2002).

1.2.5.2. The double helix

In 1953 James Watson and Francis Crick published the structure of DNA (Figure 1-13, panel A) as found in living cells. It is described as a double helix, composed of two strands of polynucleotides running in opposite direction, wrapping around each other forming a spiral staircase-like structure. The negatively charged sugar-phosphate backbone of the helix is placed on the outside of the structure

whereas the hydrophobic bases (at neutral pH) are restricted in the core. The helix is further stabilised by chemical interactions known as base-pairing and base-stacking. Base-pairing involves the formation of hydrogen bonds between the nitrogenous bases on opposite strands such that adenine on one strand interacts with thymine on the other strand and similarly cytosine interacts with guanine. These two base-pair combinations (A-T and C-G) are the ones that are usually accommodated by the DNA structure (Figure 1-13, panel B). After the two strands have come together with the formation of base-pairing, the structural stability is increased by base-stacking which are π-π electron interactions between the adjacent base pairs.

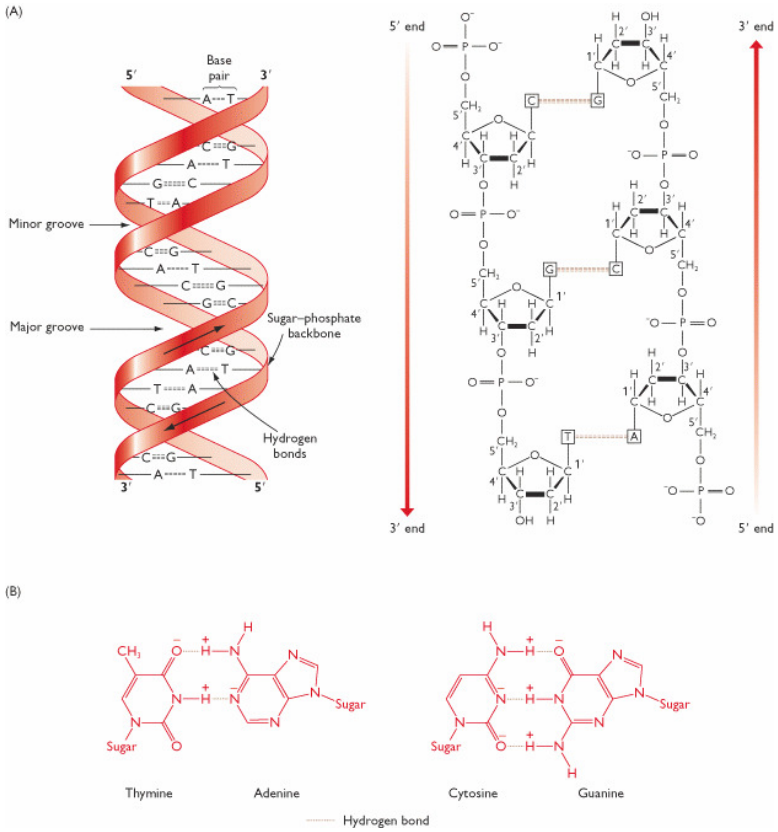


Figure 1-13. The double helix structure of DNA.
Panel A: Two representations of the double helix. On the left the structure is shown with the sugar-phosphate 'backbones' of each polynucleotide drawn as a red ribbon with the base pairs in black. On the right the chemical structure for three base pairs is given. Panel B: Adenine base-pairs with Thymine, and Guanine base-pairs with Cytosine.The bases are drawn in outline, with the hydrogen bonding indicated by dotted lines (Brown, 2002).

The features and characteristics of the double helix described by Watson and Crick correspond to the B-form of DNA. There are many other variations but the most distinct are the B-, A- and Z-DNA forms (Figure 1-14). The size and shape of the major and minor grooves also vary between different DNA forms; the implications of which are important when considering DNA-protein interactions and the accessibility by proteins of the nitrogenous bases within these grooves.

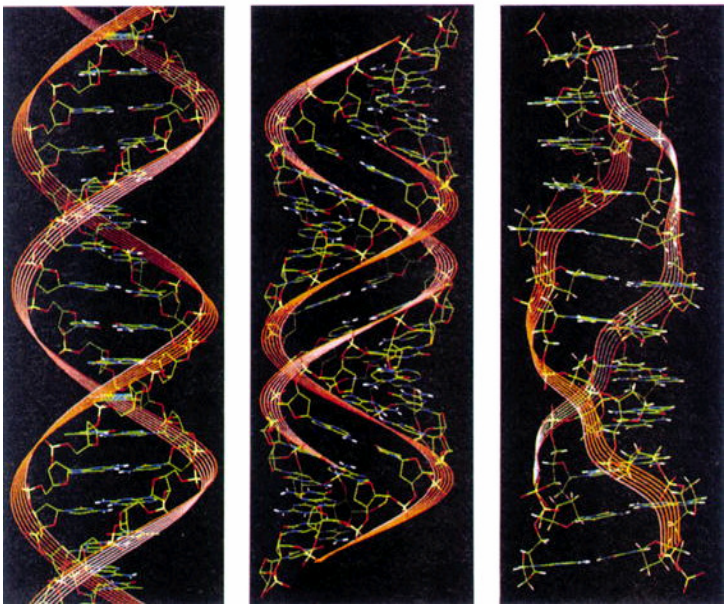


Figure 1-14. Image of B-DNA (left), A-DNA (center) and Z-DNA (right).
(Reprinted from Brown, 2002).

In living cells many different structural forms or intermediates can be found coexisting in a single DNA molecule. This is primarily due to specific nucleotide sequences (and their base-stacking interactions) in the vicinity and the natural flexibility of DNA. DNA flexibility arises from the ability of nucleotides to slightly change their molecular shapes and enables DNA to form circles and supercoils. Changes within individual nucleotides are mainly brought about by adjustments of orientation of the nitrogenous base relative to the pentose sugar, involving rotation around the β-N-glycosidic bond, and also by rotation around the 3'- and 4'- carbon of the sugar. In addition, specific series of nucleotide sequences confer the DNA

structure with increased flexibility and bending capability. For example two or more groups of repeated adenines, each group comprising 3-5 As, with individual groups separated by 10 or 11 nucleotides, will bend at the 3' end of the adenine-rich region (Young and Beveridge, 1998; Brown, 2002).

1.2.5.3. Examples of proteins mimicking DNA

Proteins that mimic DNA are structurally diverse, but share some common themes. Firstly, the mimicry usually resembles distorted rather than normal B-DNA, possibly to prevent the inhibition of critical DNA metabolizing proteins. Secondly, the distribution of side chain carboxylates of Asp and Glu residues resembles the charge pattern of the DNA sugar phosphate backbone. Finally, DNA mimics, being small in size (enforced by the dimensions of DNA) and having a high surface charge density, must possess a highly hydrophobic core, which helps stabilize the protein fold (Putnam and Tainer, 2005).

The crystal structure of the ArdA protein Orf18 has recently been published (McMahon *et al*., 2009). In a similar fashion to Ocr, the Orf18 protein forms an elongated dimer that mimics the overall shape and charge distribution of a DNA molecule (Figure 1-15). Another example is the white spot syndrome virus (a double stranded eukaryotic virus) protein ICP11 (Figure 1-16), which prevents the host DNA from binding and interacting with histone proteins (Wang *et al*., 2008). Other proteins suggested as being DNA mimics are; the Arn protein from phage T4, the gp2 and gp5 protein from phage T7 (Putnam and Tainer, 2005) and a range of other antirepressor proteins from various phages (Dryden, 2006). Some other examples of protein DNA mimics and their target enzymes are shown in Table 1-2.

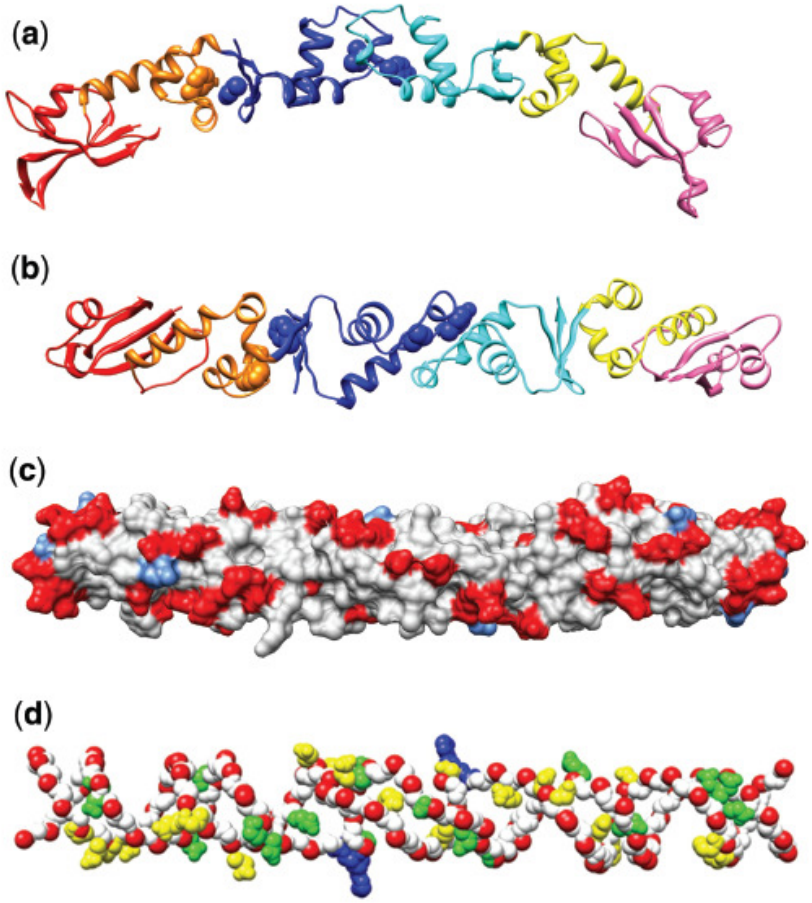


Figure 1-15. The dimeric structure of ArdA and its distribution of negative charge clearly show mimicry of the DNA double helix.
(a) The ArdA dimer with domain 1 coloured red and salmon, domain 2 coloured orange and yellow and domain 3 coloured blue and cyan. (b) The ArdA dimer rotated 90º from (a) looking at the convex face. (c) The surface of ArdA in the same orientation as in (b) with acidic residues coloured red and basic residues coloured light blue. The negatively charged residues form a helical pattern across the surface. (d) An overlay of a selection of the acidic residues in the dimer onto the DNA duplex is shown in green and yellow. Only the phosphate backbone of DNA is shown in red and white. Nucleotides flipped out by the MTase for methylation are shown in blue. Reprinted from McMahon *et al*., 2009.

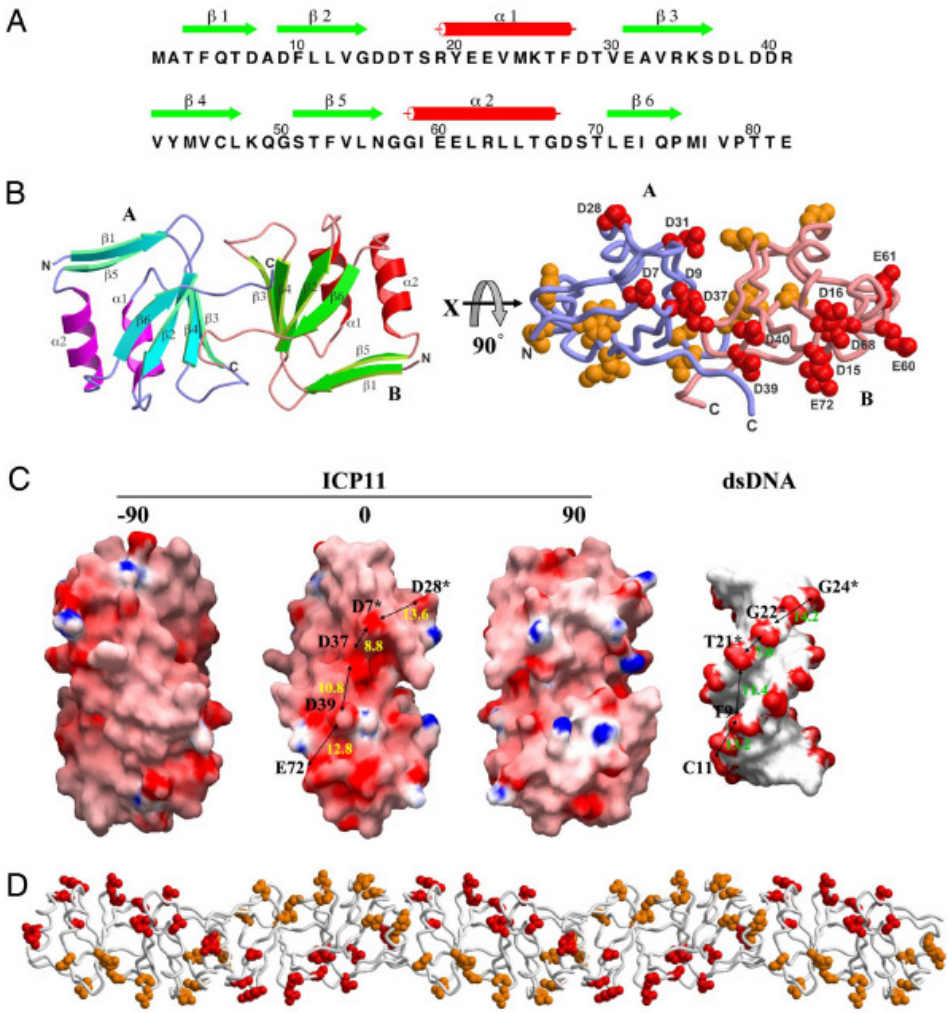


Figure 1-16. The negative charge distribution of ICP11 resembles the B-form DNA helix.

(*A*) The secondary structural elements are shown above the amino-acid sequence, with green arrows and red cylinders representing β-strands and α-helices. (*B*) A ribbon diagram of the ICP11 dimer (*Left*). The strands and helices of monomer A on the right side are colored as in *A*; those of monomer B on the left side are colored in cyan and magenta. (*Right*) A diagram of the ICP11 dimer, in which each negatively-charged amino acid is shown as a cluster of spheres (except for E21 and E22, which are omitted). (*C*) The molecular surfaces of the ICP11 dimer and dsDNA. Color-coding; acidic residues coloured red and basic residues coloured light blue; red to blue representing the electrostatic potential from -15 k_BT to +15 k_BT. The ICP11 dimer is shown in 3 orthogonal views normal to the dyad axis. Labels indicate the ICP11 acidic residues that correspond to the negatively charged spots on the dsDNA. Asterisks indicate the ICP11 monomer B or the complementary strand of the dsDNA. The approximate distances between spots are also shown. (*D*) Distribution of negatively charged amino acids in the ICP11 helical filament. The diagram is coloured as in *B*. Reprinted from Wang *et al.*, 2008.

Table 1-2. Proteins most likely to function by DNA mimicry and their biological role.

DNA mimic or putative mimic	Targeted enzyme or protein	Targeted DNA function
Uracil glycosylase inhibitor	Uracil glycosylase	Repair
Ocr	Type I restriction enzymes	Restriction
ArdA, ArdB	Type I restriction enzymes and some Type II restriction enzymes	
HI1450	Histone like proteins	DNA packaging
MfpA	DNA gyrase	Topology
$dTAF_{II}230$	RNA polymerase	Transcription
Ocr	RNA polymerase	
DinI	RecA	Recombination
p53	Replication protein A	Single-strand binding

Table adapted from Dryden, 2006.

1.3. T7 Ocr protein

The first protein to be produced during infection of *E. coli* by bacteriophage T7 is overcome classical restriction (Ocr), the product of gene 0.3 of bacteriophage T7 (Studier, 1975). The Ocr protein is the best characterised example of an antirestriction protein and of a structural mimic of DNA (Tock and Dryden, 2005; Dryden, 2006; Putnam and Tainer, 2005). The structure of Ocr is an elongated, curved dimer (Figure 1-17). The approximate length of the dimer is around 85 Å with a maximum diameter of 30 Å and an apparent bend of ~34°. Ocr is a highly negatively charged protein (pI 4.02) with a similar shape and charge distribution to that of a bent, double-stranded DNA molecule approximately 24 base pairs in length (Mark and Studier, 1981; Dunn *et al*., 1981; Bandyopadhyay *et al*., 1985; Atanasiu *et al*., 2001, 2002; Blackstock *et al*., 2001; Walkinshaw *et al*., 2002; Stephanou *et al*., 2009; Zavilgelsky *et al*., 2009).

This molecular mimicry accounts for the ability of Ocr to inhibit, virtually irreversibly, all Type I restriction and modification (R/M) enzymes found in most eubacteria and archaea (Murray, 2000; Tock and Dryden, 2005; Walkinshaw *et al*., 2002). Ocr binds to and completely occupies the DNA-binding site on the enzyme thereby inhibiting the restriction endonuclease activity and protecting the phage

genome as it enters the bacterium. Thus, Ocr greatly assists the spread of phage infection in the bacterial population.

Atanasiu *et al.,* (2002) has reported that one Ocr dimer binds to the core methyltransferase and two to the complete EcoKI nuclease enzyme. Furthermore, the strong Ocr:M.EcoKI interaction (Kd ~ 100 pM) renders this association nearly irreversible. It was proposed that virtually the entire Ocr surface is involved in the interaction, with the Type I enzyme completely wrapping around the Ocr molecule (Atanasiu *et al*., 2002). The engulfment of the Ocr dimer by its target enzyme was reinforced by Kennaway *et al*., (2009), where the ~18 Å resolution structure of the Ocr:M.EcoKI complex was achieved by negative-stain electron microscopy.

The fact that Ocr mimics the general shape and charge of DNA, rather than any specific base pair sequence, means that the protein can act against a vast array of Type I R/M enzymes, each of which recognises a different defined base pair sequence (Murray, 2000; Tock and Dryden, 2005). Each 116 amino acid monomer is decorated with 34 surface exposed Asp and Glu residues but only 2 Lys and 4 Arg residues (Dunn *et al*., 1981; Walkinshaw *et al*., 2002). Asp and Glu residues are the only amino acids with side chains possessing multiple hydrogen bond acceptors that are geometrically similar to oxygen atoms in a phosphate group. Moreover, many of the negatively charged residues of Ocr can be superimposed upon the equivalent phosphate groups on the DNA molecule.

There are two different models however, that describe the mimicry of the charge distribution of the DNA molecule by Ocr (Walkinshaw *et al*., 2002; Putnam and Tainer, 2005). The model by Walkinshaw *et al*., (2002) for example, postulates that about eleven carboxyl carbon atoms can be superimposed on the phosphorus atoms of the DNA backbone, as shown in Figure 1-17, panel A. By contrast, the model by Putnam and Tainer (2005) describes two different sets of twelve carboxylates for each Ocr monomer (24 carboxylates per dimer) that mimic the sugar phosphate backbone of DNA (described in Figure 1-18). Table 1-3 lists the carboxylates that mimic the charge distribution of DNA for both of the aforementioned models.

In addition to mimicking the charge distribution, Ocr mimics the bend of approximately 46° in the DNA helical axis induced in DNA when it binds to the R/M enzyme (Walkinshaw *et al*., 2002) (Figure 1-19). The introduction of the bend in the DNA by the R/M enzyme is energetically costly. This cost is "saved" when the enzyme binds to Ocr because the molecule is already "pre-bent" (Su, T-J. *et al*., 2005). By combining the mimicry of DNA in terms of both electrostatics and the bent shape of the molecule, the R/M enzyme:Ocr interaction is energetically more favourable than the equivalent interaction with DNA. Thus, the overall binding affinity of the R/M enzyme for Ocr is 50-fold greater than for DNA (Atanasiu *et al*., 2001, 2002; Bandyopadhyay *et al*., 1985).

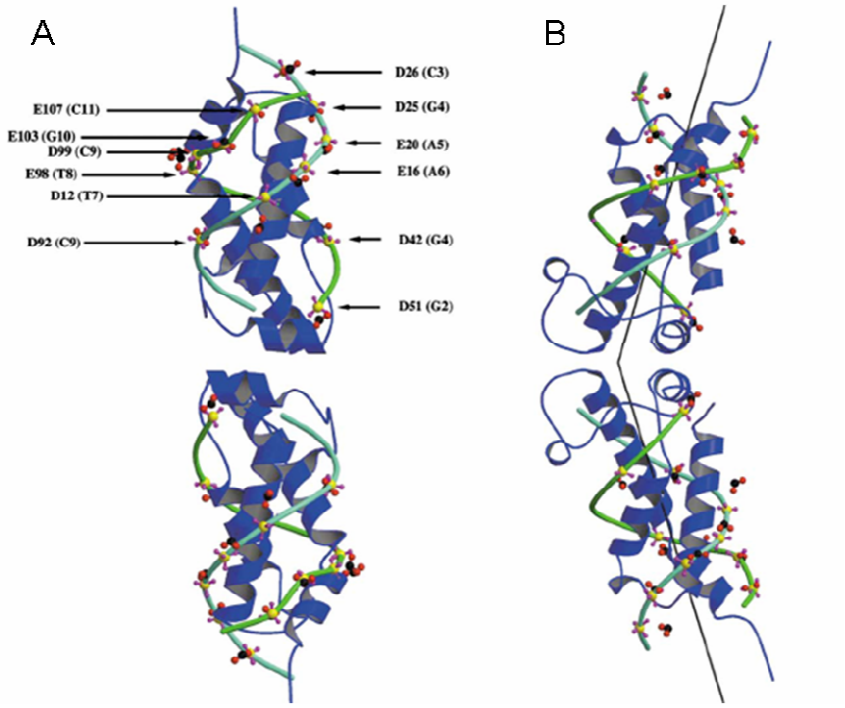


Figure 1-17. Crystal structure of the Ocr protein (pdb code 1BNA) superimposed on that of the sugar-phosphate backbone of two 12bp B-form DNA.
Ocr is a homodimer comprising two 13.7 kDa subunits. The overall structure of Ocr resembles the shape of 24 bp of DNA and an array of negative charges on the surface of the protein matches 24 phosphates from the DNA backbone. Panel A; The arrows highlight the superimposed carboxyl carbon atoms (of the protein) and the phosphorous atoms (of DNA) respectively. Colour code: P (yellow); $O_{phosphate}$ (purple); $O_{carboxy}$ (red); $C_{carboxy}$ (black); Sugar-phosphate backbone of DNA (two shades of green). Panel B; A 90º rotation of panel A, using the same colour code. The 2-fold axis lies in the plane of the paper. The vectors describing the direction of the fitted DNA on both halves of the dimer are drawn as black lines. Their intersection gives a bend angle of 33.6º. (Adapted from Walkinshaw *et al.,* 2002).

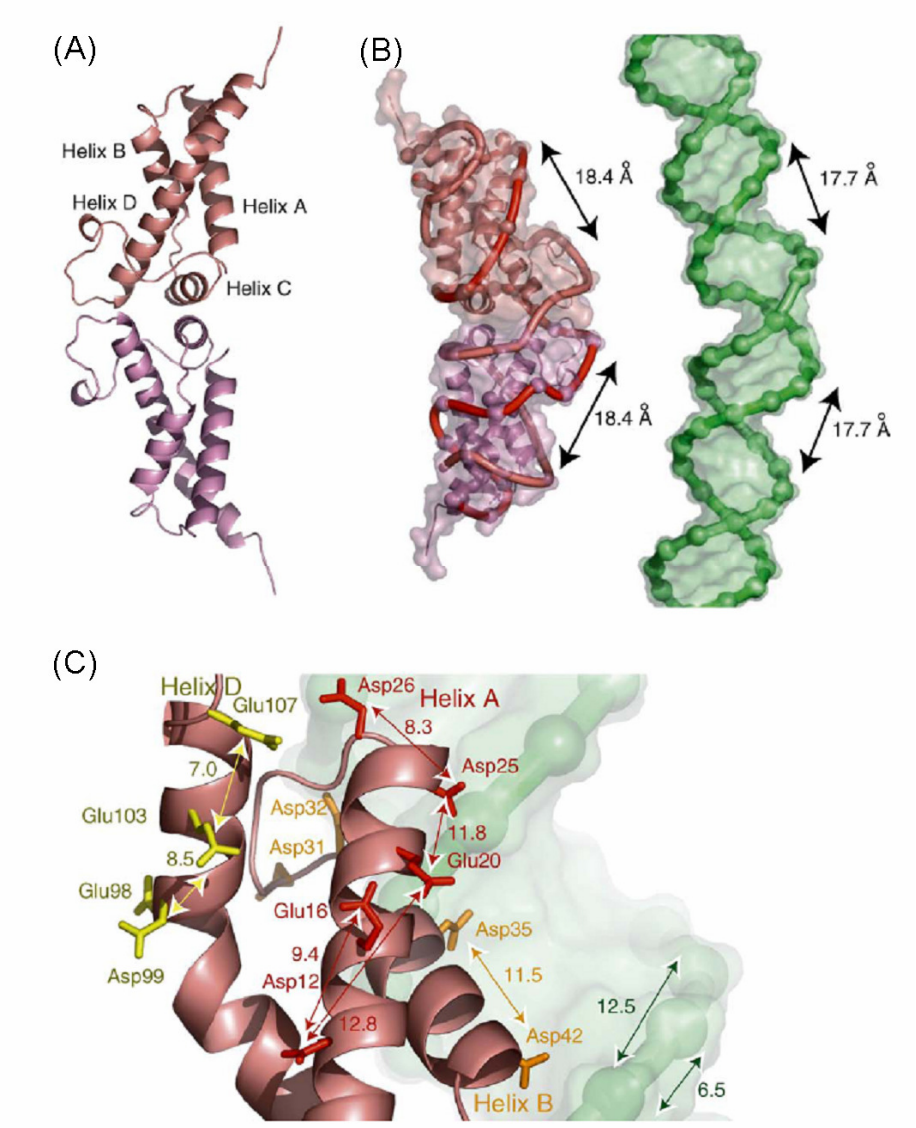


Figure 1-18. Structure of the Ocr dimer and comparison to bent dsDNA.
(A), Protein fold shown as a ribbon diagram for the Ocr dimer with subunits in light red and light magenta (PDB 1s7z). (B), Comparison of Ocr (red transparent surface) to dsDNA bent by 30 (green transparent surface). Red tubes are splines through carboxylate carbon positions (red spheres) in the dimer. Each tube mimics a single strand of DNA and they pass through the residues Glu106, Asp26, Asp25, Glu20, Glu16, Asp12, Glu95/Asp92, Glu66/Asp67, Asp73, Asp76 in one monomer and then Glu64, Asp62, Glu59, Asp51, Asp42, Asp35, Asp32, Asp31, Glu98, Asp99, Glu103, and Glu107 in the second monomer. The bent DNA was generated by tracing the DNA along a short arc over 30°. Note that the two strand mimics in Ocr have spacings reminiscent of the distance between the backbones for both the DNA major an minor grooves. (C), Detail of the phosphate-mimicking helices A (red), B (orange) and D (yellow) showing that the two to four amino acid gaps between negatively charged residues place them on the same face of the helix. Labeled distances are between the carboxylate carbon of the side chains. Residues separated by one turn of the helix can adopt a spacing of 6.5 Å consistent with the spacing between adjacent (*i* and *i* + 1) phosphates on one strand of DNA (shown in green). Residues separated by two turns of the α-helix can adopt spacings of 12.5 Å, reminiscent of the spacings between *i* and *i* + 2 phosphates on a single strand of B-DNA (Adapted from Putnam and Tainer, 2005).

Table 1-3. List of Ocr carboxylates that mimic the charge distribution of DNA according to two different models.

Walkinshaw et al. 2002		Putnam & Tainer, 2005	
Both monomers		1st monomer	2nd monomer
D12		E106	E64
E16		D26	D62
E20		D25	E59
D25		E20	D51
D26		E16	D42
D42		D12	D35
D51		E95	D32
D92		D92	D31
D98		E66	E98
D99		D67	D99
E103		D73	E103
E107		D76	E107

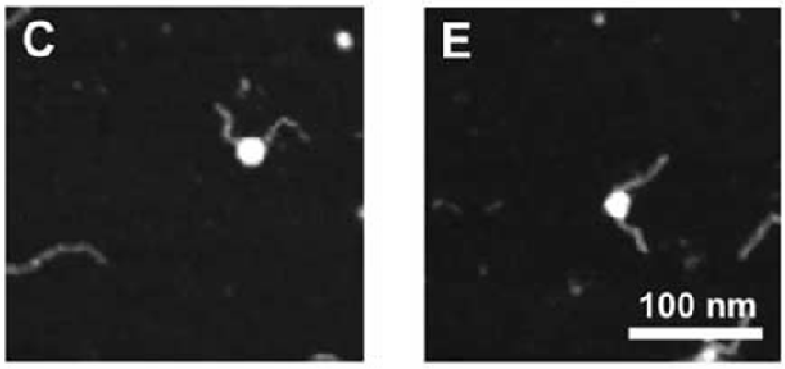


Figure 1-19. DNA bending induced by specific binding of EcoKI, as visualized using atomic force microscopy (AFM).

The EcoKI enzyme is the bright spot in the middle of the bent DNA fragments that were 292 bp long. The bending induced on the DNA molecules was 46°± 2°. Adapted from Walkinshaw *et al.*, 2002.

1.4. Protein-protein interactions

Because this study is mostly concerned with the Ocr: M.EcoKI interaction (a protein- protein interaction), some basic characteristics of protein-protein interactions are briefly introduced.

The majority of proteins operate in concert with other proteins, in complexes and networks, to create and regulate a myriad of biological processes. These processes include cell cycle control, differentiation, protein folding, signalling, transcription, translation, post-translational modification and transport. The study of protein interactions has been vital to the understanding of how proteins function within the cell.

Protein interactions can be characterized as stable or transient. Stable interactions are those associated with multi-subunit complexes. The subunits of the complex can be identical or different. Hemoglobin and EcoKI nuclease are two examples of stable multi-subunit complex interactions. Transient interactions are on/off or temporary in nature and typically require a set of conditions that promote the interaction. Transient interactions can be strong or weak, fast or slow (association rates up to ~5 x 10^9 $M^{-1}s^{-1}$ have been measured and off rates as low as about 1 x 10^{-5} s^{-1}) (Phizicky and Fields, 1995). Binding affinities of different protein-protein interactions range from the millimolar (respiratory electron transfer proteins) to femtomolar (nuclease-inhibitor complexes) (Figure 1-20).

To date, there are three classes of protein-protein interfaces: homodimers, protein complexes (permanent and transient) and crystal-packing interfaces. The type of a protein interaction depends on the physical and chemical properties of the interacting interface (*i.e.*, interface size, shape complementarity and amino acid composition (hydrophobicity and polarity).

The definition of what is considered a protein-protein interface differs slightly from study to study. It usually relies on proximity or whether solvent accessibility on one protein has been blocked by the other protein. More recently an algorithm has been developed that helps define an interface (Ban *et al.,* 2006).

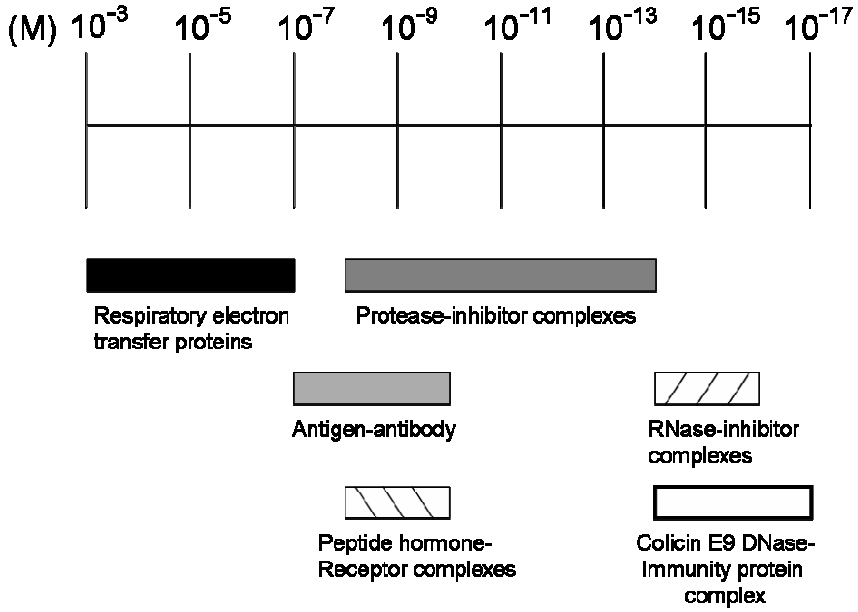


Figure 1-20. Comparison of the dissociation constants for different protein-protein complexes.
Adapted from Wallis *et al.*, (1995).

In general, interfaces tend to be planar with an area that is often proportional to the total protein size (Jones and Thornton, 1996). There is no preference for any type of secondary structure or structural motif in interfaces; while shape complementarity and packing is quite good (Stites, 1997). Crystal structures of dimeric proteins have shown the interface size per subunit, ranging from 670 to 4890 $Å^2$ (Janin *et al.*, 1988). For oligomeric proteins with more subunits, the interface areas are at the higher end of (or above) the range for dimers (Janin *et al.*, 1988). Crystal structures of antibody-protein antigen complexes and protease-inhibitor complexes show 600-1000 $Å^2$ buried per individual protein, which corresponds to 17 ± 4 amino acid residues per monomer (Janin and Cothia, 1990).

Various analyses have shown that the hydropathy and electrostatic potential of one interacting surface is complementary to that of the other interacting surface (Korn and Burnet, 1991; McCoy *et al.*, 1997). Generally the amino acid composition of interfaces is intermediate between the compositions of the protein interior and exterior. However, homodimers are enriched in nonpolar and aromatic side chains while depleted in polar and charged side chains (the exception being Arg). Other

more transient protein complexes rely on salt bridges and hydrogen bonds (Jones and Thornton, 1996). Figure 1-21 shows the average abundance of amino acids in different types of protein-protein interactions and their hydrogen bonding contribution. The extent of hydrogen bonding in interfaces of both homodimers and protein complexes is very similar, with an average of one hydrogen bond per ~100 $Å^2$. On average 20% of hydrogen bonds involve main-chain atoms only, while 40% involve main-chain to side-chain interactions (Lo Conte *et al*., 1999).

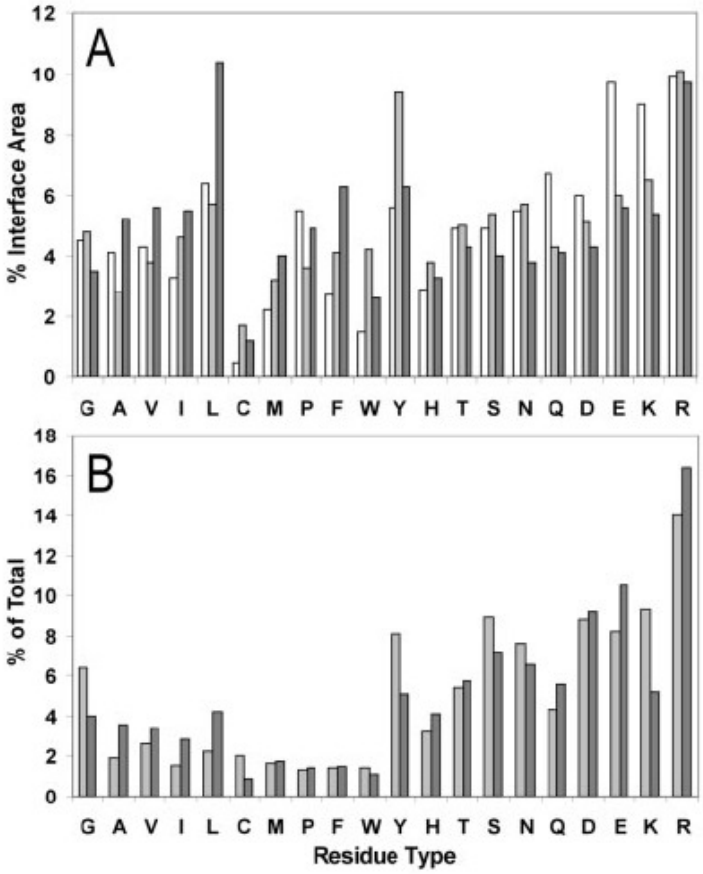


Figure 1-21. Comparison of the average abundance of amino acids in different types of protein-protein interactions and their hydrogen bonding contribution.
Panel A: The average abundance of amino acids in protein–protein interfaces of (white) large crystal-packing interfaces, (light grey) protein complexes and (dark grey) homodimers. Panel B: The average percentage contribution of amino acids to hydrogen bonding in the interfaces found in (light grey) protein complexes (63 interfaces, 588 hydrogen bonds) and in (dark grey) homodimers (115 interfaces, 2,026 hydrogen bonds). Reprinted from Crowley and Golovin, 2005.

Salt bridges occur considerably less frequently with an average number of about two per 1,000 $Å^2$ of interface (Tsai and Nussinov, 1997). Long range electrostatic interactions however, hugely accelerate rates of association (by up to 10^5 times greater) and also steer the interacting partners (pre-orientation prior to binding), even if these charge-charge or charge-dipole interactions are not part of the interacting interface (Janin, 1997; Schreiber and Fersht, 1996).

Though interfaces can be quite large, it was shown that only a small fraction of the residues contribute to the majority of the binding energy (Figure 1-22) (Clackson and Wells, 1995). These hot spots of binding energy are enriched in Trp, Tyr and Arg residues and are surrounded by energetically less important residues that most likely serve to occlude bulk solvent from the hot spot. One reason for why Trp, Arg and Tyr are the three most common amino acids in hot spots is that these residues are capable of making multiple types of favorable interactions. Trp, for example, can contribute aromatic π-interactions, a hydrogen bonding donor, and a large hydrophobic surface. Likewise, Tyr offers a hydrophobic surface, and both aromatic π-interactions and the hydrogen bonding ability of its 4-hydroxyl group (which may explain why it is more than three times more likely to be in hot spots than phenylalanine). Finally, Arg has the ability to form a hydrogen bonding networks with up to five H-bonds and a salt-bridge *via* its positively charged guanidinium motif. In addition, the electron delocalization of the guanidinium π-system has a pseudo-aromatic character and the three methylene carbon atoms of Arg are all hydrophobic in character (Bogan and Thorn, 1998). It is also interesting to note that in hot spots, Asp appears more than twice as often as Glu. Similarly, Asn is more prevalent in hot spots than Gln. These differences are presumably due to differences in side-chain conformational entropy (Lee *et al*., 1994; Bogan and Thorn, 1998).

It has been suggested that cation-π interactions are a primary driving force in heterodimerisation (Meyer *et al.,* 2003). A study reported that approximately half of the protein complexes and one third of the homodimers investigated, contained one or more cation-π interactions (Crowley and Golovin, 2005).

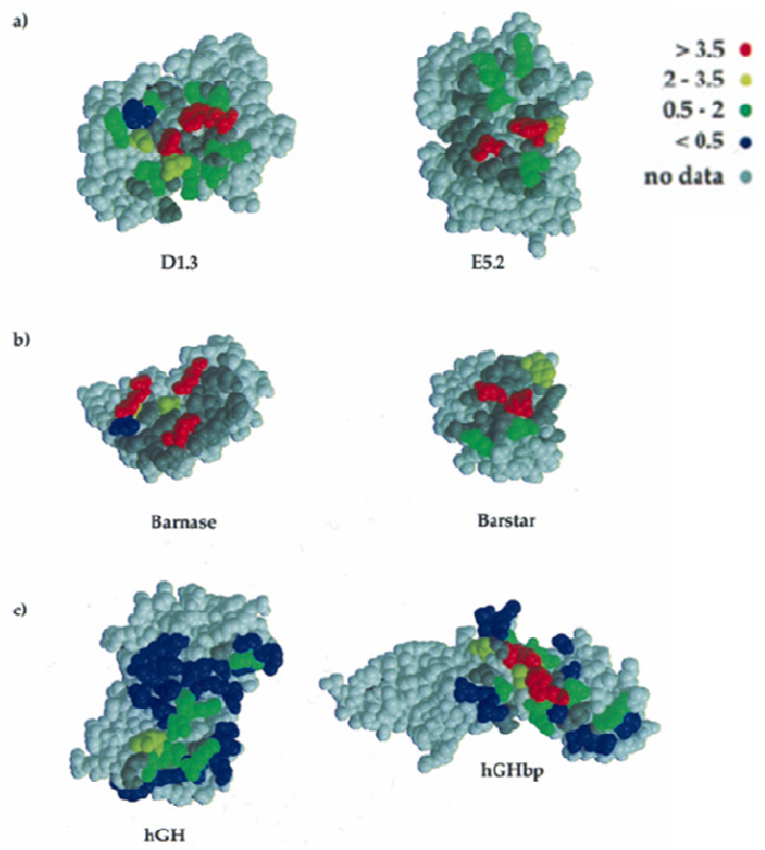


Figure 1-22. Mapping of ΔΔG of individual residues onto their location in the complexes.

The complexes have been opened by 180º rotation around the *x*-axis and water molecules have been omitted. Residues are colored by their energetic contribution to the interface: <0.5 kcal/mol, blue; 0.5 to 2.0 kcal/mol, green; 2.0 to 3.5 kcal/mol, yellow; >3.5 kcal/mol, red; no, data dark gray; not in the interface, white. The energetic contribution of specific residues for (a) the D1.3-E5.2 antibody complex (PBD 1dvf), (b) the barnase-barstar complex (PDB 1brs), and (c) the human growth hormone-receptor site 1 complex (PDB 3hhr). These complexes were selected on the basis of thorough alanine-scanning sampling of the binding interface. The hot spots of binding energy are near the center of the interfaces and are sealed from bulk solvent by surrounding residues of less energetic contribution. While a few hot spot residues appear to be at the edge of the interface, they are largely occluded from solvent by residues on the complementary face of their partner. There is no purely geometric reason why the hot spots should be in the center of the interfaces and not at the edges. Reprinted from Bogan and Thorn, 1998.

These interactions involve a positively charged amine group (usually Arg) interacting with the delocalised π-electrons of the aromatic side chains (Gallivan and Dougherty, 1999) (most frequently Tyr and Trp in heterocomplexes, and Phe in homodimers). Cation-π interactions, which are electrostatic in nature, are found to be more stable and thermodynamically favourable than salt-bridges. These interactions

contribute about 3.0 ± 1.5 kcal/mol and are usually found at the edges of interfaces due to the ability of Arg to form both hydrophobic and hydrophilic interactions. A positively charged amine may interact with an aromatic ring and form a salt-bridge simultaneously. Cation-π interactions have also been assigned an important role in the specificity of protein associations (Crowley and Golovin, 2005).

Occlusion of solvent is found to be a necessary condition for highly energetic interactions (> 3.5 kcal/mol) because water strongly attenuates the strength of polar or charged interactions. It has been suggested that the hydrophobic effect of hot-spot residues is almost double that of other residues in an antibody/antigen interface (Li *et al.*, 2005). Desolvation is therefore a major thermodynamic barrier affecting protein–protein interactions. The other condition for highly energetic interactions is that interacting residues are within van der Waals distances (Bogan and Thorn, 1998). The vast majority of protein heterodimer interfaces are larger than 600 $Å^2$ (Jones and Thornton, 1996). It has been suggested that this is the minimum area required to form a water-tight interface, which favours highly energetic interactions (Bogan and Thorn, 1998). Water molecules however, buried at the protein-protein interface can also mediate many intermolecular hydrogen bonds and increase the shape complementarity of the interface. Crystal structures of homodimers and complexes contain about ten water molecules per 1000 $Å^2$ of interface none of which are present within the "dry", desolvated hotspots (Figure 1-23). Water molecules at interfaces form hydrogen bonds with protein groups, with a preference for the main-chain carbonyl and the charged side-chains of Glu, Asp and Arg. Water-mediated polar interactions are as abundant at the interfaces as direct protein–protein hydrogen bonds, and they may contribute to the stability of the assembly (Rodier *et al*., 2005). Perona *et al.,* (1993) have shown that water-mediated interactions have the same effect on binding affinity as direct hydrogen-bonding contacts.

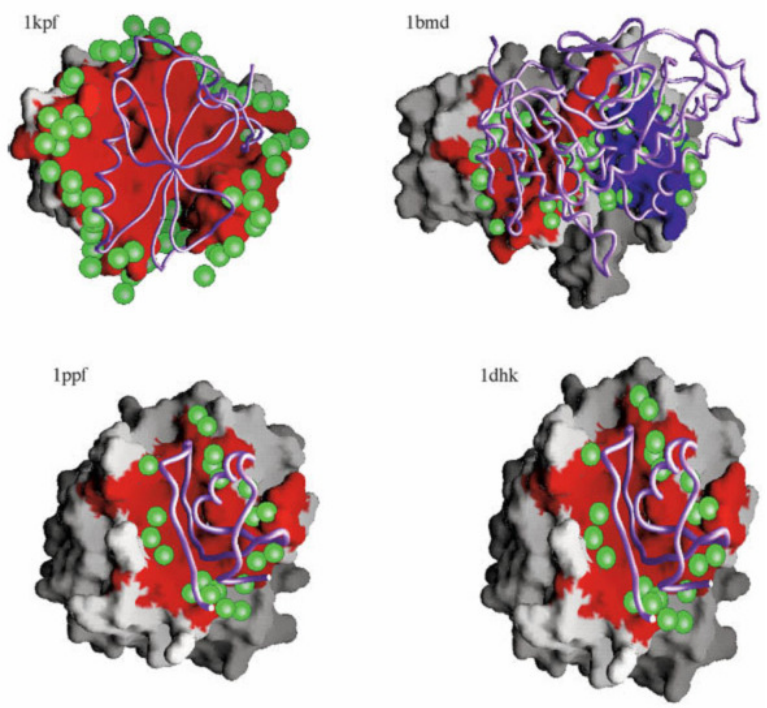


Figure 1-23. Patterns of interface hydration.

In each panel, one protein subunit is shown as a backbone worm, the other as a molecular surface with the interface in color. Green spheres are water molecules. 1kpf: The subunit interface of the dimeric PKC interacting protein is typically dry; its interface area is B = 3700 Å2; Nw = 56 (Nw; number of water molecules). 1bmd: The subunit interface of malate dehydrogenase is in 2 patches colored red and blue; interface waters line both patches; B = 3100 Å2; Nw = 47. 1ppf: Elastase–ovomucoid inhibitor complex; B = 1320 Å2; Nw = 22 1dhk: α-amylase–bean inhibitor complex; B = 3010 Å2; Nw = 40. Adapted from Rodier *et al.*, 2005.

The overall aim and direction of this thesis was to understand the basis of the DNA mimicry displayed by Ocr. The surface of the protein is replete with acidic residues, most or all of which mimic the phosphate backbone of DNA. Therefore, aspartate and glutamate residues on the surface of Ocr were either mutated or chemically modified in order to investigate their contribution to the tight binding between Ocr and the EcoKI Type I R/M enzyme.

Chapter 2. Materials and Methods

2.1. Materials

2.1.1. Bacterial strains, Maintenance and Growth

E. coli JM109 was purchased from Promega (Madison, WI). *E. coli* NM1261 (r^-m^-) and *E. coli* NM1049 (EcoKI Type IA R/M system) were a kind gift of Professor Noreen Murray (School of Biology, University of Edinburgh, UK). The NM strains were converted to DE lysogens as described by McMahon *et al.*, (2009). *E. coli* XL1-Blue supercompetent cells were purchased from Stratagene (La Jolla, CA). The expression strain *E. coli* BL21(DE3)pLysS and *E. coli* DH5α cells were purchased from Invitrogen (Groningen, The Netherlands).

For long term storage of bacterial strains or competent cells, a 0.5 ml aliquot of the LB culture in exponential phase was transferred to a sterile 1.5 ml eppendorf tube to which 15% by volume of glycerol was added. The cultures were then put on dry ice for ~15 min and then stored at -80°C. Viable cells were recovered by scratching the surface of the frozen stock with a sterile platinum loop.

All cell growth was conducted at 37ºC, in either LB (Luria-Bertani) liquid medium (10 g/l NaCl, 10 g/l Difco bacto-tryptone, 5 g/l Difco yeast extract, pH adjusted to 7.2 with NaOH), NZY liquid medium (5 g/l NaCl, 2 g/l $MgSO_4.7H_2O$, 5 g/l Yeast extract, 10 g/l NZ amine (casein hydrolysate), pH adjusted to 7.5 with NaOH) or LB agar plates (LB broth, 1.5 g/l Difco agar, pH adjusted to 7.2 with NaOH), supplemented with appropriate antibiotics to facilitate plasmid maintenance.

2.1.2. Plasmids and Phage

Plasmid pAR2993 harbours the gene encoding wild-type Ocr located just downstream of an isopropyl β-D-1-thiogalactopyranoside (IPTG) inducible promoter (Atanasiu *et al.*, 2001). Plasmids pAR3786 and pAR3790 encoding for C-terminal deletions of the last 7 or 17 amino acids of wild-type Ocr were a kind gift from Dr. Alan Rosenberg and Professor William Studier (Brookhaven National Laboratory,

USA). Plasmid pET24 was purchased from Novagen (Madison, WI). pBRsk1 is an engineered version of pBR322 (4361 bp) in which one of the two EcoKI sites (4024-4036) has been removed by site-directed mutagenesis (Davies *et al*., 1999a). The unmethylated form of pBRsk1 used in the nuclease assays was prepared from *E. coli* NM1261(r^-m^-).

Virulent unmodified bacteriophage lambda λv.o, or modified lambda λv.k were provided by Professor Noreen Murray. The letter "v", stands for virulent, meaning that the phage is unable to form lysogens.

2.1.3. Chemicals

Dimethylamine, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and hydroxybenzotriazole (HOBt) were obtained from Pierce (Rockford, IL). Guanidine hydrochloride (Ultrol grade) was purchased from Calbiochem Inc. (San Diego, CA). All other reagents were purchased from Sigma-Aldrich (St Louis, MO). Broad range pre-stained molecular mass markers for SDS-PAGE were purchased from BioRad (Precision Plus protein standards; Hercules, CA). All solutions were made up in distilled, deioinised water.

2.2. DNA methods

2.2.1. Preparation of Plasmid DNA

A QIAprep Spin Miniprep Kit and a microcentrifuge were used. Cells from 5-10 ml overnight culture of *E. coli* DH5α (that had been previously transformed with the construct of interest) in LB medium, was pelleted by centrifugation in a 15 ml falcon tube at 4000 rpm for 10 min. Thereafter the procedure described in the Qiagen manual (Qiagen Corp., Valencia, CA) was precisely followed. This kit uses an alkaline lysis of the bacterial cells followed by adsorption of the DNA onto a silica membrane in high salt conditions. The DNA is then washed and eluted in 50 µl buffer EB (10 mM Tris·Cl, pH 8.5). The resultant ~50 µl of plasmid solution (at ~50 ng/µl) was stored at either 4°C or -20°C.

2.2.2. Site Directed Mutagenesis

Single, double or multiple codon mutagenesis was achieved using the QuikChange II Site-Directed Mutagenesis kit from Stratagene following the manufacture's guidelines. This procedure is based on a Polymerase Chain Reaction (PCR). A typical reaction mixture contained:

5 µl of 10× reaction buffer

~0.4 µl (at ~50 ng/µl) of dsDNA template

1 µl (at ~125 ng/µl) of oligonucleotide primer #1

1 µl (at ~125 ng/µl) of oligonucleotide primer #2

1 µl of dNTP mix (10mM each dNTP)

ddH_2O to a final volume of 50 µl

Then added

1 µl of *PfuUltra* HF DNA polymerase (2.5 U/µl)

The PCR cycles were carried out using an "Eppendorf Mastercycler personal, with a heated lid" from Eppendorf North America.

Removal of parental plasmid DNA was achieved by treating the reaction mixture with DpnI (~0.4 µl per reaction) and incubating for 1 hour at 37 °C. The DpnI endonuclease (target sequence: 5´-Gm^6ATC-3´) is specific for methylated and hemimethylated DNA and is used to digest the parental DNA template and to select for mutation-containing synthesized DNA. (DNA isolated from almost all *E. coli* strains is *dam* methylated and therefore susceptible to DpnI digestion). The nicked vector DNA containing the desired mutations is then transformed into XL1-Blue supercompetent cells by a heat shock method as described in the Mutagenesis kit manual.

2.2.3. Cloning Techniques

2.2.3.1. Polymerase Chain Reactions

A typical reaction mixture for PCR contained:

5 µl of 10× reaction buffer

~0.4 µl (at ~50 ng/µl) of dsDNA template

1 µl (at ~125 ng/µl) of oligonucleotide primer #1

1 µl (at ~125 ng/µl) of oligonucleotide primer #2

1 µl of dNTP mix (10mM each dNTP)

ddH_2O to a final volume of 50 µl

Then added

1 µl of *PfuUltra* HF DNA polymerase (2.5 U/µl)

The PCR cycles were carried out using an "Eppendorf Mastercycler personal, with a heated lid" from Eppendorf North America. The PCR products were subsequently purified using the QIAquick PCR purification kit (Qiagen) and a Microcentrifuge following the protocol provided by the manufacturer. In the final step of the protocol the DNA was eluted from the QIAquick column in 30 µl of buffer EB.

2.2.3.2. Restriction Endonuclease Digests

Restriction digests of plasmid DNA or of the PCR purified products were performed according to the manufacturer's guidelines using commercial buffer systems (New England Biolabs, Ipswich, MA). Typically digestions contained 1-5 µg of DNA and 5-10 units of restriction enzyme and the digestion time was ~1 hour. Following digestion the samples were loaded directly on gels for electrophoresis.

2.2.3.3. Ligation of DNA

Generally an aliquot of vector and insert DNA was visualised by agarose gel electrophoresis prior to ligation in order to estimate the relative concentrations. The

ligation reaction typically contained 100 ng of vector DNA and 300ng of insert DNA. The reaction volume was 20 µl and included 1.0 mM ATP, 10 mM DTT and 10 units of T4 DNA Ligase (New England Biolabs). The ligation mixtures were incubated at room temperature for ~16 hours. The mixture was then directly used for transformation of XL1-Blue supercompetent cells.

2.2.4. Agarose Gel Electrophoresis of DNA

Agarose gel electrophoresis was used as a common analytical tool to identify and visualise DNA. Agarose gels (0.8-1.5% (w/v) were cast (as horizontal slab gels) and run in 1x TAE buffer (40 mM Tris·Cl, 2 mM EDTA, 24 mM acetic acid, pH 7.7) as described by Sambrook *et al.*, (2001). Ethidium bromide was include in the molten agarose at a final concentration of 0.5 µg/ml. Samples were prepared in 1X DNA loading buffer (25% v/v ficoll 400, 0.25% v/v xylene cyanol, 0.25% v/v bromophenol blue) and loaded into wells using a Gilson pipette. Typically samples were run alongside molecular weight markers at 100 V/h. DNA was visualized using a UV transilluminator (UVP, TFM-30) and images were acquired using a digital camera (Fujifilm FinePix S602Zoom; Fuji Corp., Tokyo, Japan).

2.2.5. Recovery of DNA from Agarose Gels

DNA was extracted from agarose gels by using a QIAquick Gel Extraction Kit, a microcentrifuge and following the protocol provided by the manufacturer. The DNA was eluted from the QIAquick purification columns in either 30 or 50 µl of buffer EB.

2.2.6. Competent cells

E. coli DH5α cells (Invitrogen) were made competent using the $CaCl_2$ method (Sambrook and Russell, 2001). 5 ml of LB was inoculated with a single colony of *E. coli* DH5α previously grown on an LB plate. The culture was grown overnight at 37°C, with vigorous shaking. 1 ml of the overnight culture was inoculated into 50 ml of fresh LB media, and incubated at 37°C with shaking until the optical density of the culture at 600 nm was ~0.5. The cells were then harvested at 13,000 g for 15 minutes. The pellet was resuspended in 20 ml of ice-cold 100 mM

$MgCl_2$, and centrifuged at 13,000 *g* for 10-15 minutes. The supernatant was discarded and the pellet was resuspended again in 10 ml of ice-cold 100 mM $MgCl_2$, and centrifuged once more. The supernatant was discarded, and finally the pellet was resuspended in 2 ml of ice-cold 100 mM $CaCl_2$. 200 µl of the competent cells were transferred to pre-chilled Eppendorf tubes and stored on ice. Because the efficiency of transformation increases four- to sixfold during the first 12-24 hours of storage at 4°C (Sambrook and Russell, 2001) the transformation step, described below, was carried out the following day.

2.2.7. Transformation

1–2 µl of plasmid DNA (10 ng) was used to transform 50 µl of competent cells. The mixture was incubated on ice for 30 minutes. The cells were treated to 30-45 seconds of heat shock at 42°C, and then chilled on ice for 2 minutes. 500 µl of LB or NZY medium was added to the mixture, and the cells were incubated for 1 hour at 37°C with shaking. 10-100 µl was then transferred and distributed evenly over the surface of an LB agar plate containing the appropriate antibiotic, using a sterile polystyrene "Lazy-L spreader" purchased from Sigma-Aldrich. The LB agar plates were incubated at 37°C overnight to select for transformed cells.

2.3. Protein Methods

2.3.1. Gel electrophoresis

Samples were routinely analysed by SDS-PAGE using NuPAGE 4-12% Bis-Tris gels (Invitrogen). The effect of chemical modification on the overall charge of the Ocr molecule was assessed by running gels (15% acrylamide) under non-denaturing conditions (*i.e.*, in the absence of SDS). The unmodified Ocr was also analysed using the NativePAGE Novex® 4-16% Bis-Tris gel system (Invitrogen), which uses Coomassie G250 as the charge-shift molecule. In relation to the Ocr protein about ~10 µl at ~30 µM would be usually loaded (including the appropriate loading buffer) to either of the aforementioned polyacrylamide gels. Following electrophoresis, gels were fixed and stained in a solution of methanol:acetic acid:water (5:1:5 by volume) containing 0.1% (w/v) Coomasie Brilliant Blue R250

for at least 30 min. The gels were destained in 30% (v/v) methanol, 10% (v/v) acetic acid.

2.3.2. Protein Expression and Purification

2.3.2.1. Purification of wild-type and Ocr mutants

E. coli BL21(DE3)pLysS was transformed (sections 2.2.6 and 2.2.7) with the appropriate construct. Cells were grown at 37°C in 2 L conical flasks containing 1 L LB broth supplemented with 34 µg/ml chloramphenicol and 50 µg/ml of carbenicillin and shaking at ~230 rpm. Once the optical density at 600 nm reached ~0.5, heterologous gene expression was induced by the addition of IPTG (final concentration of 1 mM) and growth continued for a further 2.5 hours before harvesting the cells by centrifugation (8000 *g* for 10 min at 4°C). Cell pellets were stored at -20°C until required. The cells were resuspended in ice-cold buffer A (20 mM Tris, 300 mM NH_4Cl, pH 8.0) in the presence of a protease inhibitor cocktail (Roche, Basel, Switzerland). Cells were then broken on ice by sonication using a Soniprep 150 Sonicator (Sanyo, Tokyo, Japan) fitted with a 9 mm probe (1 min/g of cell paste). The cell debris was then removed by centrifugation (15,000 *g* for 1 hr at 4°C). The supernatant was loaded onto a 20 x 1.6 cm 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 48 ml/h. The column was extensively washed with buffer A and then a 500 ml gradient from 0.3-1.0 M NH_4Cl in buffer A was run at 24 ml/h. Fractions containing Ocr, identified by SDS-PAGE, were subsequently pooled. UV-spectroscopy indicated that the sample was contaminated with nucleic acid. The nucleic acid was removed in the following manner. 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 ~10 min. The precipitate was collected by centrifugation at 15,000 *g* for 10 min at 4°C and the pellet was resuspended in 95% ethanol with gentle mixing for ~10 min. After centrifugation (15,000 *g* 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 20 mM Tris HCl pH 8.0, instead of 95% ethanol. The sample was then dialysed against ~4 L 20 mM Tris pH 8.0 for 16 hr at 4°C and concentrated by centrifugation using a Vivaspin concentrator (10,000 MWCO; VivaScience AG, Hannover, Germany). Finally, an equal volume of glycerol was added to the sample, which was then stored at -20°C until required.

2.3.2.2. Purification of M.EcoKI

M.EcoKI was prepared from 10 litres of BL21(DE3) cells transformed with pJFMS, grown in LB broth containing 50 µg/ml carbenicillin. Each of the two 15 g cell pastes was thawed and resuspended in 80 ml of 20 mM Tris, 20 mM MES, 10 mM $MgCl_2$, 7 mM 2-mercaptoethanol, 0.1 mM EDTA pH 6.5 buffer plus protease inhibitors (20 µM benzamidine and 10 µM phenylmethylsulfonyl fluoride). Cells were homogenised and passed twice through a French Press at ~16 p.s.i. The lysate was centrifuged at 20,000 *g* for 3 hours after which the supernatant was loaded on a 40 x 2.5 cm diameter CM-Sepharose fast flow cation exchange column (GE Healthcare, Piscataway, NJ) which had been pre-equilibrated in buffer, at a flow rate of 48 ml/h. The CM column was washed with at least one column volume of buffer and then ~5 ml fractions were collected using a salt gradient 0 to 0.5 M NaCl (using a flow rate of 24 ml/h). The fractions containing the M.EcoKI protein, as determined by running a 10% SDS-PAGE gel, were pooled dialysed against buffer to remove NaCl and loaded on a 15 x 1.6 cm diameter Heparin column (Sigma-Aldrich) at a flow rate of 48 ml/h. After washing the column with at least one column volume of buffer, ~5 ml fractions were collected with a salt gradient of 0 to 1 M NaCl using a flow rate of 24 ml/h. The chosen fractions, determined by running a 10% SDS-PAGE gel were pooled together and proteins were precipitated by 70% saturation with ammonium sulfate. The precipitants were resuspended in a minimal volume of buffer containing 20 mM Tris, 20 mM MES, 10 mM $MgCl_2$, 7 mM 2-mercaptoethanol, 0.1 mM EDTA 200 mM NaCl, pH 8.0. This was then loaded as aliquots on a gel-filtration column (Superdex 200 HiLoad 16/60, Amersham-Pharmacia) at a flow rate of 48 ml/h. Fractions were then analysed by 10% SDS-PAGE gel. The best fractions were then pooled together and stored at -20˚C in 50% glycerol. Protein

concentrations were measured by UV spectroscopy at 280 nm, using the extinction coefficient for M.EcoKI (M_2S_1) 143140 $M^{-1}cm^{-1}$ (Dryden *et al*., 1997).

2.3.3. Chemical Modification Procedure

Surface carboxyl groups of Ocr were chemically modified using 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC) and either ammonium hydroxide or dimethylamine as a nucleophile to give the so called N- and D-series of chemically modified protein, respectively. Chemical modification of Ocr (3 µM) was carried out at 25°C in an aqueous solution (dH_2O) of 750 mM ammonium hydroxide or dimethylamine HCl, 200 mM NaCl, 60 mM EDC, 60 mM HOBt, pH 6.5. The pH of the reaction mixture was adjusted by using either conc. HCl or conc. NaOH for the N- and D-series reactions respectively. The reaction was initiated when a concentrated sample of Ocr (~100 µM) prepared in 10 mM Mes, 150 mM NaCl, pH 6.5 was injected in the reaction mixture to a final Ocr concentration of 3 µM. Aliquots were withdrawn at specific time points (1 to 180 min) and the reaction was quenched by adding a six fold-excess of sodium acetate (using a 3 M stock of sodium acetate pH 6.5). After incubation for 10-20 min the mixture was first dialysed against 100 mM of nucleophile (either ammonium hydroxide or dimethylamine, pH 8.5) for 4 h at 25°C and then against 50 mM ammonium acetate at 4°C for a minimum of 4 h. Hydroxylamine was then added to the samples at a final concentration of 400 mM (using a 3 M stock of hydroxylamine pH 7.0) and the mixture was incubated at 25°C for 4 h. Finally, the solution was dialysed against 20 mM ammonium acetate and then the protein concentration was adjusted to 20-30 µM using a Vivaspin concentrator (10,000 MWKO; VivaScience AG). Samples were stored in 50% v/v glycerol at -20°C.

2.4. Analytical Methods

2.4.1. 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 10 mM Tris-HCl pH 8.0, 50 mM NaF. Far-UV CD spectra (190–260 nm)

were performed at a protein concentration of 30 μM using a 0.2 mm pathlength cell. A bandwidth of 1 nm was used with a scan speed of 10 nm/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.

2.4.2. Ion exchange chromatography

Protein samples were analysed by anion exchange chromatography using a 1 ml Mono Q column (GE Healthcare). Each protein (11.2 μg) was individually loaded onto the column that was pre-equilibrated in 20 mM Tris HCl pH 8.0 at a flow rate of 1 ml/min. After washing the column a linear gradient of 0 to 1 M NaCl in 20 mM Tris HCl pH 8.0 was set up over 30 min using the same flow rate. Protein elution from the column was monitored by measuring the fluorescence (excitation, 295 nm, emission, 350 nm) of the eluate. The elution time of each sample was determined by integrating the peak and calculating the point corresponding to 50% of the peak area (*i.e.*, time at which 50% of the protein has eluted from the column).

2.4.3. MALDI-TOF MS analysis

Mass spectrometry of protein samples was performed by MALDI-TOF using a Voyager DE STR instrument (Applied Biosystems, Foster City, CA). The samples were diluted in 0.1% trifluoroacetic acid to ~0.05 mg/ml and mixed with an equal volume of matrix (saturated solution of sinapinic acid in 50% acetonitrile, 0.1% trifluoroacetic acid) on a stainless steel surface and were then air dried at room temperature to crystallize. The machine was operated in positive ion mode and calibrated with conalbumin and bovine serum albumin.

2.4.4. High Resolution LC-MS analysis

Protein samples were extensively desalted by dialysis prior to mass spectrometry. For LC-MS, an Ultimate 3000 HPLC system (Dionex Corporation, Sunnyvale, CA), equipped with a monolithic PS-DVB (500 μM x 5 mm) analytical column (Dionex Corporation), was used. Samples containing ~1 μg of protein were centrifuged (16,100 *g* for 2 min) immediately prior to injection onto the column.

Solutions B and C were prepared comprising of 2:97.95 and 80:19.95 actonitrile:water with 0.05% formic acid respectively. Samples were injected onto the analytical column, washed with solution B for 5 min, followed by a 20 min linear gradient elution (20 µl/min) into solution C. MS data was acquired on a Bruker 12 Tesla Apex Qe FT-ICR (Bruker Daltonics, Billerica, MA) equipped with an electrospray ionization source. Desolvated ions were transmitted to a 6 cm Infinity cell® penning trap. Trapped ions were excited (frequency chirp 48-500 kHZ at 100 steps of 25 µs) and detected between *m/z* 600 and 2000 for 0.5 s to yield a broadband 512 Kword time-domain data. Fast Fourier Transforms and subsequent analyses were performed using DataAnalysis (Bruker Daltonics) software. Multiple charge states could be observed in this way for each of the major species.

2.4.5. Unfolding Studies

Equilibrium unfolding as a function of guanidinium hydrochloride (GdmCl) concentration was monitored by fluorescence spectroscopy (excitation 280 nm; emission scan from 330-450 nm; bandwidth 5 nm). A stock solution of GdmCl was freshly prepared from ultrapure reagents and the precise concentration determined using a refractometer. The protein samples in 20 mM 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.

2.4.6. Fluorescence Anisotropy

In order to determine the K_d of the Ocr:M.EcoKI interaction, a competition binding assay using fluorescence anisotropy was used, as described below. All measurements were carried out using an Edinburgh Instruments FS 900 CDT T-geometry fluorimeter (Edinburgh Instruments Ltd, Livingston, UK). The anisotropy data were fitted to an equilibrium binding model using DynaFit (BioKin Ltd, Watertown, MA, USA) as described further down.

Assays were performed in 20 mM Tris-HCl pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol (buffer A), including 2 nM of HEX-labelled 21 bp dsDNA and

100 μM SAM (SAM would be added immediately prior to the actual titration experiment). The Ocr protein, if present, was at a concentration of 50 nM (as a dimer). This sample solution was placed in a quartz cuvette whilst mixing with a small magnetic flea. The concentration of M.EcoKI varied between 0.25 and 500 nM during the titration while at the same time continuous readings were recorded. In order to accomplish a gradual logarithmic increase of M.EcoKI concentration in the sample solution without significantly altering the initial Ocr and DNA concentrations, the M.EcoKI was titrated by using three different stock solutions, at about 0.2 μM, 2 μM and 20 μM in buffer A. The final M.EcoKI concentration in the cuvette, for each of the ~ 20 titration points, was then calculated. The DNA labelled duplex had been annealed in ddH_2O, by heating at 95ºC for 5 min and then left to cool gradually over night in a ~500 ml volume of the heated bath water. The unlabelled strand was at a 20% greater concentration than the labelled strand to ensure that the observed anisotropy signal was due to dsDNA.

The actual titration experiment was carried out in the following manner:

Conditions

A 1 cm pathlength quartz cuvette (1 x 0.4 cm; Starna) containing ~800 μl of sample solution (as described above) was placed in the fluorimeter and left for a few minutes for temperature equilibration. The magnetic stirrer was set at ~400 rpm, to induce rigorous but gentle mixing of the sample solution. To prevent bubble formation, buffer A was either degassed or an aliquot was left to equilibrate at room temperature. To remove fine particulates and dust, the buffer was filtered. The fluorimeter water bath (controlling the temperature cuvette holder) was set at 25ºC. Excitation slits were set at 2-3 mm, and emission slits at ~10 mm using a T-format geometry.

Step 1

The G factor was adjusted to ~1 by exciting (λex= 530 nm) with horizontally polarised light (90º) and adjusting the light intensities of emission (λem=549-551 nm; step= 0.1 nm) for both arms to approximately the same value.

Step 2

The following emission scans were recorded to assess the HEX fluorophore.

I_{VV}, I_{VH}, I_{HH}, I_{HV}; λex= 530 nm, λem=550-610 nm, step= 1 nm, dwell time= 0.1 sec, accumulation of 2 scans.

Step3

The G factor was recorded by a Kinetic scan; I_{HH}, I_{HV} for 60 sec; λex= 530 nm, λem= 560 nm.

Step 4

The readings taken while M.EcoKI was titrated in the sample solution using a Hamilton syringe: Kinetic scan for ~45 min (long enough to complete titration); I_{VV}, I_{VH}; λex= 530 nm, λem= 560 nm. (At each titration point the fluoremeter lid is lifted and the an M.EcoKI aliquot is added to the cuvette solution. Each time the lid is lifted the emission slits are automatically closed giving zero intensity readings. This event is important in the analysis procedure for "r" as is described further down).

Step 5

The G factor was calculated by repeating *Step 3*.

Data Analysis

In order to get a Kd of binding for the Ocr:M.EcoKI interaction, a plot of anisotropy *vs* [M.EcoKI] must be fitted to an appropriate binding curve discussed below. The data for this plot was acquired as described in the next few paragraphs.

The final concentration of M.EcoKI in the sample mix after each titration point, was calculated manually.

Manual calculation of anisotropy "r" (where $r = (I_{VV} - GI_{VH})/ (I_{VV} + 2GI_{VH})$) for each titration point, however, is extremely laborious and time consuming. For this reason, by collaborating with Dr Joseph Spadavecchia (University of Edinburgh, Database Research Group, School of Informatics) a script was designed to automate this process (the script is given in Appendix A). Three ASCII text files (Data sets in columns; Tab as delimiter) are required from *Steps 3, 4* and *5* (described above).The script is executed through Command Prompt while in the same directory as the three text files. To execute script type: **mtase Step3filename.txt Step5filename.txt Step4filename.txt**

(Notes: (i) To execute the script, the order in which the txt files are typed is important. **Step3filename.txt** (*i.e.*, G factor file at the START of experiment); **Step5filename.txt** (*i.e.*, G factor file at the END of experiment); **Step4filename.txt** (the titration file). If only one G factor file was recorded then repeat the G factor file twice in the execution command of the script; the order in which the two G factor files are typed in the command is not important since the script calculates the average G factor from the two files. The third txt file typed in the command must always be the titration file. (ii) Emission Arm1 of the fluorimeter should be reading I_{VV} while Emission Arm2 should be reading I_{VH}, because this will affect the order in which the data columns will appear within the txt files. An example of how these txt files should appear is shown in Table 2-1. Refer to Table 2-2 for required parameters in relation to executing the script).

Table 2-1. The format of the txt files required for using a script for the automated calculation of anisotropy for each titration point of a binding assay using T-format geometry fluorescence anisotropy.

The execution of the script file is described in the main text above. Numbers highlighted in red correspond to a titration point (*i.e.*, lifting of the fluorimeter lid accompanied by automatic shutting of emission slits resulting in zero intensity readings). The txt file of the titration (left panel) shows only two titration points out of approximately 20 for each titration experiment.

ASCII txt file of titration

Labels	Decay1 Em1	Decay1 Em2
Type	Kinetic Scan	Kinetic Scan
Comment		
start(nm)	0.00	0.00
stop(nm)	1.8E12	1.8E12
step(nm)	1E9	1E9
fixed/offset(nm)		
Xaxis	Time	Time
Yaxis	Counts	Counts
0.00	6.082E+3	4.127E+3
1E9	6.425E+3	4.025E+3
2E9	6.457E+3	4.247E+3
3E9	6.237E+3	4.082E+3
4E9	6.395E+3	4.200E+3
5E9	6.112E+3	4.285E+3
6E9	6.370E+3	4.060E+3
7E9	6.387E+3	4.187E+3
8E9	6.402E+3	4.337E+3
9E9	6.517E+3	4.145E+3
1E10	6.575E+3	4.085E+3
1.1E10	6.312E+3	4.017E+3
1.2E10	6.250E+3	4.210E+3
1.3E10	6.355E+3	4.267E+3
1.4E10	6.252E+3	4.305E+3
1.5E10	6.480E+3	4.085E+3
1.6E10	6.372E+3	4.295E+3
1.7E10	6.340E+3	4.200E+3
1.8E10	6.210E+3	4.175E+3
1.9E10	6.310E+3	4.240E+3
2E10	6.537E+3	4.205E+3
2.1E10	6.082E+3	4.085E+3
2.2E10	6.222E+3	4.115E+3
2.3E10	6.275E+3	4.265E+3
2.4E10	6.090E+3	4.367E+3
2.5E10	6.370E+3	4.055E+3
2.6E10	6.385E+3	4.245E+3
2.7E10	6.427E+3	4.062E+3
2.8E10	6.325E+3	4.152E+3
2.9E10	6.490E+3	4.157E+3
3E10	6.237E+3	4.175E+3
3.1E10	6.320E+3	4.180E+3
3.2E10	6.265E+3	4.182E+3
3.3E10	6.225E+3	4.250E+3
3.4E10	6.220E+2	5.000E+0
3.5E10	2.000E+0	2.000E+0
3.6E10	7.000E+0	0.000E+0
3.7E10	2.000E+0	0.000E+0
3.8E10	5.000E+0	1.200E+1
3.9E10	2.000E+0	2.670E+2
4E10	6.172E+3	4.185E+3
4.1E10	6.390E+3	4.145E+3
4.2E10	6.165E+3	3.990E+3
4.3E10	6.097E+3	4.290E+3
4.4E10	6.027E+3	4.047E+3
4.5E10	6.445E+3	4.097E+3
4.6E10	6.225E+3	4.045E+3
4.7E10	6.302E+3	4.157E+3
4.8E10	6.182E+3	4.826E+3
4.9E10	6.082E+3	4.367E+3
5E10	6.100E+3	4.040E+3
5.1E10	6.147E+3	4.207E+3
5.2E10	6.335E+3	4.190E+3
5.3E10	6.217E+3	4.177E+3
5.4E10	6.387E+3	4.130E+3
5.5E10	6.197E+3	4.147E+3
5.6E10	5.905E+3	4.070E+3
5.7E10	5.992E+3	4.115E+3
5.8E10	6.282E+3	4.220E+3
5.9E10	6.102E+3	4.097E+3
6E10	8.070E+2	2.000E+0
6.1E10	2.000E+0	5.000E+0
6.2E10	2.000E+0	1.000E+1
6.3E10	2.000E+0	0.000E+0
6.4E10	6.000E+0	7.000E+0
6.5E10	0.000E+0	2.997E+3
6.6E10	6.217E+3	4.187E+3
6.7E10	6.385E+3	4.325E+3
6.8E10	6.310E+3	4.317E+3
6.9E10	6.260E+3	4.127E+3
7E10	6.270E+3	3.985E+3
7.1E10	6.132E+3	4.125E+3
7.2E10	6.080E+3	4.102E+3
7.3E10	6.085E+3	4.205E+3
7.4E10	6.335E+3	4.075E+3
7.5E10	6.302E+3	4.192E+3
7.6E10	6.172E+3	4.077E+3
7.7E10	6.450E+3	4.097E+3
7.8E10	6.422E+3	4.002E+3
7.9E10	6.492E+3	4.240E+3
8E10	6.215E+3	4.110E+3
8.1E10	6.395E+3	3.882E+3
8.2E10	6.250E+3	4.022E+3
8.3E10	6.282E+3	4.055E+3
8.4E10	6.310E+3	4.085E+3
8.5E10	6.300E+3	4.195E+3
8.6E10	6.232E+3	3.942E+3
8.7E10	6.290E+3	3.997E+3
8.8E10	6.372E+3	4.102E+3
8.9E10	6.150E+3	4.272E+3
9E10	6.370E+3	4.085E+3
9.1E10	6.167E+3	4.025E+3
9.2E10	5.880E+3	4.150E+3
9.3E10	6.450E+3	4.127E+3

ASCII txt file of G factor

Labels	Decay1 Em1	Decay1 Em2
Type	Kinetic Scan	Kinetic Scan
Comment		
start(nm)	0.00	0.00
stop(nm)	8E10	8E10
step(nm)	1E9	1E9
fixed/offset(nm)		
Xaxis	Time	Time
Yaxis	Counts	Counts
0.00	1.447E+4	1.287E+4
1E9	1.486E+4	1.311E+4
2E9	1.484E+4	1.315E+4
3E9	1.481E+4	1.289E+4
4E9	1.514E+4	1.313E+4
5E9	1.477E+4	1.285E+4
6E9	1.547E+4	1.297E+4
7E9	1.500E+4	1.296E+4
8E9	1.503E+4	1.299E+4
9E9	1.491E+4	1.336E+4
1E10	1.473E+4	1.332E+4
1.1E10	1.472E+4	1.311E+4
1.2E10	1.503E+4	1.283E+4
1.3E10	1.488E+4	1.315E+4
1.4E10	1.514E+4	1.306E+4
1.5E10	1.471E+4	1.309E+4
1.6E10	1.486E+4	1.333E+4
1.7E10	1.490E+4	1.293E+4
1.8E10	1.481E+4	1.326E+4
1.9E10	1.507E+4	1.290E+4
2E10	1.480E+4	1.299E+4
2.1E10	1.488E+4	1.332E+4
2.2E10	1.502E+4	1.291E+4
2.3E10	1.472E+4	1.335E+4
2.4E10	1.477E+4	1.315E+4
2.6E10	1.484E+4	1.320E+4
2.6E10	1.455E+4	1.295E+4
2.7E10	1.489E+4	1.319E+4
2.8E10	1.495E+4	1.291E+4
2.9E10	1.496E+4	1.305E+4
3E10	1.490E+4	1.386E+4
3.1E10	1.481E+4	1.290E+4
3.2E10	1.475E+4	1.292E+4
3.3E10	1.491E+4	1.257E+4
3.4E10	1.474E+4	1.273E+4
3.5E10	1.482E+4	1.318E+4
3.6E10	1.491E+4	1.326E+4
3.7E10	1.508E+4	1.277E+4
3.8E10	1.517E+4	1.288E+4
3.9E10	1.511E+4	1.292E+4
4E10	1.475E+4	1.335E+4
4.1E10	1.488E+4	1.322E+4
4.2E10	1.493E+4	1.317E+4
4.3E10	1.453E+4	1.297E+4
4.4E10	1.460E+4	1.322E+4
4.5E10	1.439E+4	1.328E+4
4.6E10	1.497E+4	1.329E+4
4.7E10	1.466E+4	1.313E+4
4.8E10	1.473E+4	1.320E+4
4.9E10	1.497E+4	1.318E+4
5E10	1.448E+4	1.310E+4
5.1E10	1.478E+4	1.339E+4
5.2E10	1.474E+4	1.303E+4
5.3E10	1.489E+4	1.304E+4
5.4E10	1.488E+4	1.298E+4
5.5E10	1.472E+4	1.281E+4
5.6E10	1.520E+4	1.323E+4
5.7E10	1.474E+4	1.317E+4
5.8E10	1.470E+4	1.295E+4
5.9E10	1.493E+4	1.305E+4
6E10	1.451E+4	1.305E+4

Table 2-2. Required parameters for executing script in relation to calculating anisotropy.

1. *G-begin* - file containing G-factor measurements of the following format:

#	I_{HV}	I_{HH}
0.00	1.327E+4	8.872E+3
1E9	1.320E+4	8.812E+3
2E9	1.366E+4	8.715E+3
3E9	1.306E+4	8.910E+3
...		

The first line is a comment. In general, lines beginning with '#' or not starting with a digit are treated as comments and, thus, ignored. The first column is ignored, the second two columns are processed.

2. *G-end* - file containing G-factor measurements of the same format

3. *data* - file containing data measurements of the following format:

#	I_{VV}	I_{VH}
0.00	5.175E+3	2.515E+3
1E9	5.480E+3	2.347E+3
2E9	5.397E+3	2.467E+3
3E9	5.347E+3	2.425E+3
...		

Again, lines beginning with '#' or not starting with a digit are treated as comments and, thus, ignored. The first column is ignored, while the second two columns are processed.

In this file, groups of measurements are delimited by lines where either (or both) I_{VV} and I_{VH} are 0, or mapped to zero by the *z-value* (see below).

The first two and the last two lines of each group of measurements are always ignored. Thus, groups of measurements must consist of more than 4 lines to be considered.

The supplied data must have the format above, with I_{VV} in the left column and I_{VH} in the right column.

Optional parameters:

1. *z-value* - integer used in parsing *data* file, maps I_{VV} and I_{VH} values in [0,z] => 0

If this value is not specified then the default *z-value* of 20 will be applied.

After plotting anisotropy (r) *vs* [M.EcoKI], the data was fitted to a binding curve using DynaFit (BioKin Ltd).

In this experiment there are three equilibria occurring:

(M.EcoKI).DNA <===> (M.EcoKI) + DNA : K1 dissoc. (K_d = 2.11 ± 0.35 nM)

(M.EcoKI).(M.EcoKI).DNA <===> (M.EcoKI) + (M.EcoKI).DNA : K2 dissoc. (K_d = 349 ± 73 nM)

(M.EcoKI).Ocr <===> (M.EcoKI) + Ocr : K3 dissoc.

By keeping K1d and K2d fixed in the analysis procedure using DynaFit, it was possible to calculate the dissociation constant of M.EcoKI for Ocr (or any Ocr mutant).

2.4.7. Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) was carried out using a VP-ITC instrument (Microcal, Northampton, MA). The stocks of Ocr and M.EcoKI were buffer exchanged into 20 mM Tris-HCl 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 either by dilution into the same buffer or by concentration using a Vivaspin concentrator (10,000 MWCO for Ocr, 30,000 MWCO for M.EcoKI; VivaScience AG). SAM (New England Biolabs) was then added to a final concentration of 100 µM. All solutions were thoroughly degassed prior to use. Typically, Ocr at a concentration of 30-40 µM was titrated into a M.EcoKI solution at a concentration of 3-4 µM in the VP-ITC cell (1.4 ml active volume). All titrations were carried out at 25°C. The heat of dilution was obtained by injecting Ocr into buffer or buffer into buffer and these values were subtracted from the ITC titration data. 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 using the Microcal LLC Origin software package.

2.4.8. Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) was carried out using a VP-DSC calorimeter (Microcal). The measurements were performed in 20 mM Tris-HCl pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol with a protein concentration of 50 µM. The sample volume was 0.52 mL and scans were carried out from 20 to 100°C at a heating rate of 1°C/min. Samples were thoroughly degassed immediately prior to loading into the calorimeter and experiments were performed at a pressure of 1.72 bar to avoid bubble formation. After the initial scan the samples were cooled down and then rescanned to determine whether the unfolding process was reversible. Data were fitted using a non-2-state unfolding model using Microcal LLC Origin Software.

2.4.9. Nuclease assay

The *in vitro* assay monitored the cleavage of unmethylated pBRsk1 using purified EcoKI in the absence or presence of Ocr protein. Reactions were performed in a 50 µl volume at 37°C and contained 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. Where required, Ocr (~100 nM for example) was briefly preincubated with EcoKI in the reaction mix prior to initiating the reaction. Reactions were initiated by the addition of DNA substrate and performed for 10 min before quenching by incubation at 68°C for 10 min. To act as a convenient size marker pBRsk1 was linearised with EcoRI (New England Biolabs) using the commercial buffer. Samples were mixed with gel loading buffer and loaded onto a 0.9% 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.

2.4.10. *In vivo* spot tests

Spot tests were performed using the prepared phage (unmodified and modified virulent lambda, λv.o and λv.k respectively). These tests gave an indication of the antirestriction activity of the Ocr proteins, measured by the survival of λv.o in the R/M proficient *E. coli* strain NM1049 (EcoKI Type IA R/M system); in relation

to the appropriate controls: Survival of λv.o and λv.k phage in *E. coli* NM1261 (r^-m^-) and NM1049 strains in the presence (pAR2993) and absence (pET14b) of an Ocr protein (mutant or wild-type). Colonies obtained from a fresh transformation of the *E. coli* strains with the Ocr constructs were set up for an overnight growth. 5 ml of LB containing the appropriate antibiotic were inoculated with a single colony, and the cultures were incubated at 37°C overnight. 200 µl of the overnight cultures + 3 ml of molten BBL top agar (maintained at 42°C) was poured onto the BBL bottom media plates and allowed to settle. Then 5 µl of the serial dilutions (10^{-1} – 10^{-9}) from the original phage stocks of λv.o and λv.k were spotted on the bacterial lawn as shown schematically below (Figure 2-1). The plates were incubated at 37°C overnight.

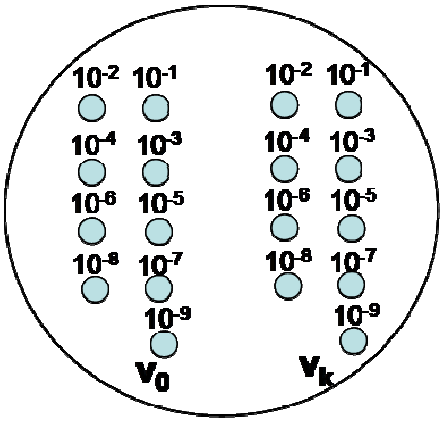


Figure 2-1. Schematic representation of a BBL plate on which spot tests were performed.

Chapter 3. Investigating the DNA mimicry of the T7 Ocr protein by a single and double mutational analysis

3.1. Introduction

According to Walkinshaw *et al.,* (2002) the Ocr protein of bacteriophage T7 is a structural and electrostatic mimic of approximately 24 base pairs of double-stranded B-form DNA (Figure 3-1). As such, it inhibits all Type I Restriction and Modification (R/M) enzymes by blocking their DNA binding grooves thereby inactivating the enzyme. This allows the infection of the bacterial cell by T7 to proceed unhindered by the action of the R/M defence system.

It was anticipated that certain single and double mutations of Asp and Glu residues on the Ocr surface may significantly perturb the tight binding between the EcoKI Type I R/M enzyme and Ocr. The selection of Asp and Glu residues was guided by an overlay of the structure of Ocr on ds DNA (Figure 3-1). According to this model the residues chosen mimic the phosphate backbone of one strand of the DNA duplex. The selection of residues, agrees well with a different model proposed be Putnam and Tainer, 2005 (Figure 3-2). The constructs encoding the Ocr mutants D12N; E16A; E16Q; E20A; D25A; D26N; D12A,D26N; D12N,D26N; D12N,E87D; E16A,D25A; were engineered and made by Dr. David T.F. Dryden, Dr. Mark R. Tock and Rachel Turkington. The Ocr Cys mutant constructs D25C; E59C; D62C; S68C, which were engineered for a different project (Atanasiu *et al.,* 2002), were also used in this study.

3.2. Aims

The aim of the work described in this chapter was to assess the *in vivo* and *in vitro* activity of the aforementioned Ocr mutant proteins. My contribution was to express and purify the Ocr proteins and assess their *in vitro* binding to M.EcoKI by fluorescence anisotropy.

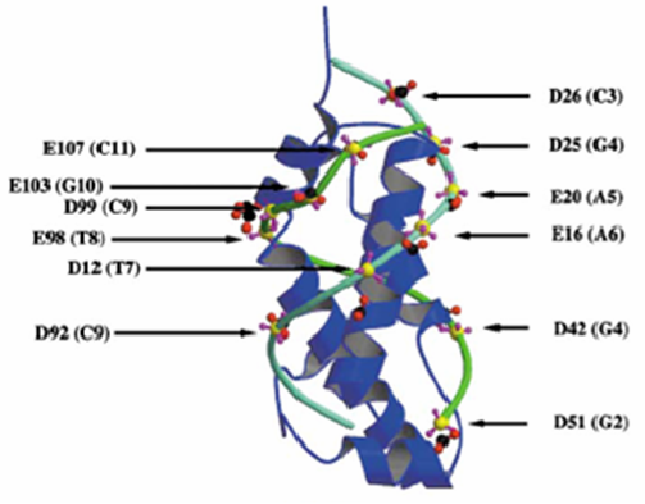


Figure 3-1. Crystal structure of the Ocr monomer from pdb code 1BNA, superimposed on that of the sugar-phosphate backbone of 12bp B-form DNA.
Ocr is a homodimer comprising two 13.7 kDa subunits. The overall structure of the Ocr dimer resembles the shape of 24 bp of DNA and an array of negative charges on the surface of the protein matches 24 phosphates from the DNA backbone. The arrows highlight the superimposed carboxyl carbon atoms (of the protein) and the phosphorous atoms (of DNA) respectively. Colour code: P (yellow); Ophosphate (purple); Ocarboxy (red); Ccarboxy (black); Sugar-phosphate backbone of DNA (two shades of green) (Walkinshaw *et al.*, 2002).

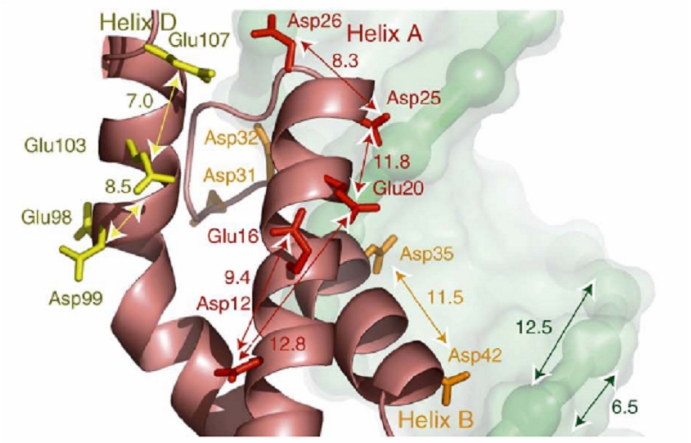


Figure 3-2. Model describing acidic residues of Ocr that mimic DNA phosphate groups.
The carboxylates highlighted in red agree well with the DNA mimicking model proposed by Walkinshaw *et al.*, 2002 shown in Figure 3-1. Figure Adapted from Putnam and Tainer, (2005).

3.3. Overall strategy

The various Ocr mutants were expressed and purified, using the protocol described in section 2.3.2.1. Binding of the Ocr proteins to M.EcoKI was assessed by fluorescence anisotropy (section 2.4.6) and by isothermal titration calorimetry (ITC) (section 2.4.7). *In vivo* phage restriction assays were also carried out (Stephanou AS *et al*., 2009).

3.3.1. Background for ITC and Fluorescence Anisotropy

3.3.1.1. Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a quantitative technique that can measure the thermodynamic properties of protein-protein (or other macromolecular-ligand) interactions (O'Brien and Haq, 2004). ITC measures the binding equilibrium of the complex directly by determining the heat released on association of a ligand with the protein. This technique can potentially determine parameters such as the stoichiometry of binding (*n*), the association constant (Ka) and the enthalpy of binding (ΔHb); all in one single experiment. Using the association constant (Ka), the free energy and the entropy of binding can also be determined. If the titration of the interaction is performed at various temperatures, the heat capacity ΔCp (at constant pressure) of the binding can be determined. The temperature-dependant differences in the binding enthalpy are used to calculate the changes in heat capacity, which is measured as the slope of a plot of ΔHb *vs.* temperature (Plum and Breslauer, 1995). This is defined by the equation:

$$\Delta Cp = \delta(\Delta Hb)/\delta T = T.[\delta(\Delta S)/\delta T]$$

A typical ITC instrument (Figure 3-3) consists of two identical cells made out of a highly efficient thermal conducting material such as gold, surrounded by an adiabatic jacket. The jacket is connected to a circulating water bath in order to keep it cool. Sensitive thermocouple circuits detect the temperature difference between the two cells as well as between the cells and the jacket, while the heaters located on both cells and the jacket are responsible for maintaining identical temperatures

between all the components. The macromolecule solution is placed in the sample cell, and the reference cell contains the respective buffer.

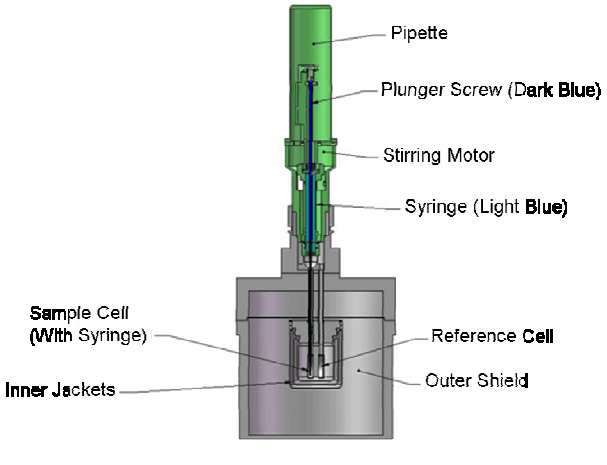


Figure 3-3. Schematic representation of an ITC instrument.
The signal observed in an ITC titration experiment is a direct measure of the heat released or absorbed upon the binding of a ligand to a macromolecule.

The injection of the titrant (ligand) into the sample cell from an injection syringe results in heat absorbed (endothermic reaction) or released (exothermic reaction) upon association of the protein and the ligand. If the reaction is exothermic, the temperature in the sample cell will increase, and the feedback power will be deactivated to maintain equal temperatures between the two cells (negative signal of the differential power). For an endothermic reaction, the feedback circuit will increase power (positive signal of the differential power) to the sample cell to maintain the temperature.

The enthalpy change measured by ITC is a global property of the whole system in the sample cell, and denotes the total heat released or absorbed upon each injection of the ligand (Jelesarov and Bosshard, 1999). Despite this, there is a direct relationship between the fraction of the ligand bound and to the heat absorbed or released during a calorimetric titration. During the initial injections, all or most of the ligand is bound to the macromolecule, which results in large exothermic or endothermic signals. However, as the ligand concentration increases, the macromolecule becomes saturated, leading to less heat being released or absorbed.

A typical set of raw data is shown in Figure 3-4. The figure shows a plot of the power (µcal/sec) against time for the interaction of M.EcoKI and Ocr.

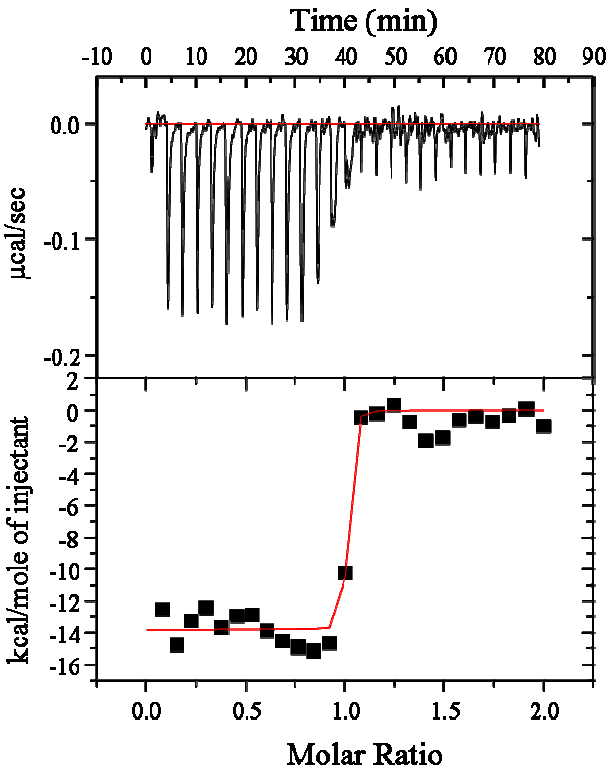


Figure 3-4. Typical data derived from an ITC experiment.
The upper panel shows the raw data output for the interaction of M.EcoKI and Ocr. The peaks are plotted as power against time. The lower panel shows the integrated raw data output plotted as molar change in enthalpy against mole ratio.

3.3.1.2. Fluorescence Anisotropy

Steady state fluorescence anisotropy measurements can describe the dynamics of a biomolecule in solution. Analysis of derived data can provide information on macromolecule size, shape and possible interactions with other macromolecules. Fluorescence anisotropy is sufficiently sensitive and moreover it is a solution-based technique (thus offering a true equilibrium measure of an interaction) making it ideal for the study of protein–nucleic acid or protein–protein interactions. Data acquisition is fairly simple, rapid and can be used in thermodynamic analyses of complex systems.

An isotropic solution is one in where all the fluorophores are oriented randomly. When such a solution is irradiated with polarised light only certain fluorophores will get excited. These selectively excited fluorophores are those whose absorption transition dipoles are aligned partially parallel along the electric vector of the polarised excitation which is oriented parallel to the vertical or *z*-axis (Figure 3-5). The probability that excitation with polarised light will actually result in a population of excited fluorophores that are partially oriented along the *z*-axis is known as photoselection. The principle of anisotropy can be defined as the photoselective excitation of fluorophores by polarised light (Lakowicz, 1999).

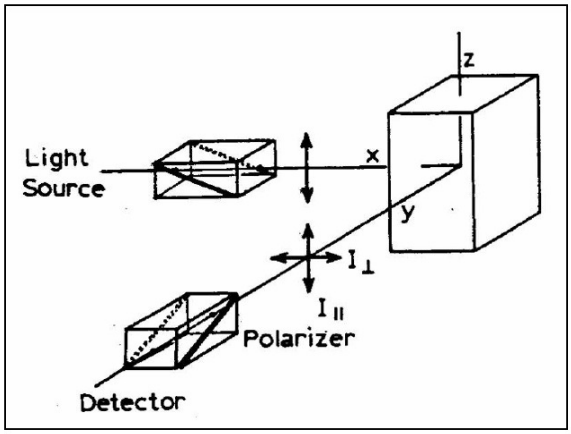


Figure 3-5. Schematic diagram for fluorescence anisotropy measurements.
The sample is excited with vertically polarized light (*i.e.*, the electric vector is oriented parallel to the vertical or *z*-axis). The intensity of the emission is measured through a polarizer which can be oriented parallel or perpendicular to the excitation beam(reprinted from Lakowicz, 1999).

Anisotropy is a unit-less quantity that can also be described as the extent to which a fluorophore rotates during the excited-state lifetime. This phenomenon can be used to calculate the apparent volume of biomolecules or possible interactions, since larger biomolecules will rotate more slowly due to greater mass and volume (Lundblad *et al.,* 1996). Therefore, if a biomolecule binds to another biomolecule, the rotational rate ("tumbling") will decrease, resulting in an increase in anisotropy (Lakowicz, 1999).

Diffraction grating monochromators in fluorescence spectrometers can disperse polychromatic light into its various wavelengths. The transmission

efficiency of a monochromator *i.e.*, the observed fluorescence intensities, is dependent on the polarisation displayed by the gratings of the monochromator. The polarisation properties of the gratings can display different efficiencies of transmission at different wavelengths. This can have important consequences in the measurements of fluorescence anisotropy when the emission of the vertically and horizontally polarised signal is measured simultaneously. These measurements can be corrected (for the varying efficiencies of these monochromators) by calculating the G factor.

The two commonly used methods in anisotropy measurements are the L-format and the T-format. In the L-format, a single emission channel is used; whereas in the T-format, the parallel (VV) and the perpendicular (VH) components of the emission spectra are measured simultaneously through separate channels. In the T-format measurements of fluorescence anisotropy, the intensities of the parallel and perpendicular components of the emission spectrum are measured simultaneously using two separate detection systems (Figure 3-6). The G factor is measured by keeping the excitation beam horizontally polarised (excitation angle = 90°), and measuring the emission spectra of the vertically polarised emission (I_{HV}, emission angle = 0°) as well as the horizontally polarised emission (I_{HH}, emission angle = 90°) simultaneously. The ratio of the two intensities gives the G factor.

$$G = I_{HV} / I_{HH}$$

where

I_{HV} = intensity of vertically polarised emission spectra measured (0°) when excited with a horizontally polarised (90°) beam.

I_{HH} = intensity of horizontally polarised emission spectra measured (90°) when excited with a horizontally polarised (90°) beam.

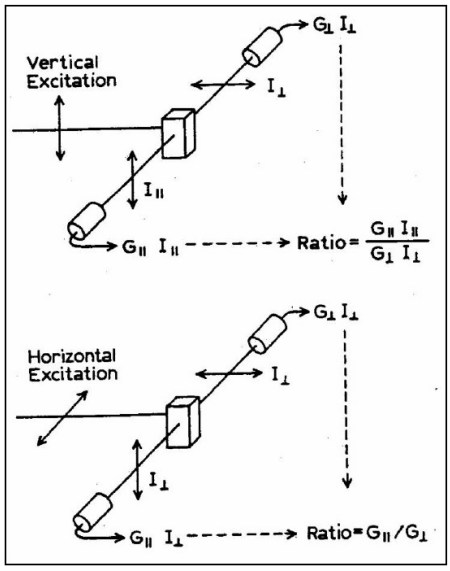


Figure 3-6. Schematic diagram for T-format measurements of fluorescence anisotropy. Adapted from Lakowicz, (1999).

The G factor of the system is recorded at the beginning (and the end) of every experiment and is adjusted to a value of approximately 1.0 as described in section 2.4.6 *step 1*. When the G factor is known, the anisotropy (r) of the sample can be calculated by:

$$r = (I_{VV} - GI_{VH})/ (I_{VV} + 2GI_{VH})$$

where r = anisotropy

$G = I_{HV} / I_{HH}$

The denominator also describes total intensity of emission

I_{VV} = fluorescence intensity of the vertically polarised emission (V = 0°), when the sample is excited with vertically polarised light (V = 0°)

I_{VH} = fluorescence intensity of the horizontally polarised emission (H = 90°), when the sample is excited with vertically polarised light (V = 0°)

Fluorescence anisotropy was used, to study the binding of M.EcoKI to a fluorescently-labelled 21 bp DNA duplex containing the EcoKI recognition site, as well as to indirectly measure the dissociation constant of the Ocr:M.EcoKI interaction. As the intrinsic fluorescence of DNA was too weak, an externally attached fluorescent probe was used. DNA molecules are easier to label with an extrinsic fluorophore than protein molecules, since they can be labelled in a manner

where the binding of the DNA to other biomolecules is not hindered. The rod-like geometry of DNA also means that the binding of protein to an oligonucleotide will have a greater effect on the rotational motion of the molecule, again leading to larger changes in fluorescence anisotropy (Lundblad *et al.,* 1996).

A 21 bp DNA duplex was tagged with a hexachlorofluorescein (HEX) probe (Figure 3-7) at the 5' end of the top strand (ATDBio Ltd). The DNA sequence for the top strand of the oligonucleotide was 5'-HEX-GCCT**AAC**CACGTG**GTGC**GTAC-3', with an EcoKI-specific recognition site (highlighted in bold). HEX is a common extrinsic fluorophore and the fluorescence signal of this fluorophore is monitored to study the "tumbling" of the DNA molecule (Kumke *et al.,* 1995; Sevenich *et al.,* 1998). The HEX probe was attached to the 21 bp DNA duplex *via* a six-carbon linker. The issue was then raised whether the motion of the DNA can be directly correlated with the motion of the linked fluorophore. According to Powell *et al.,* (1998a) the probe is neither too flexible nor too rigid in respect to the DNA molecule. A slight degree of freedom was detected although not large enough to preclude the use of the labelled duplex for the study of protein–DNA interactions. Moreover, HEX can be detected at low concentrations, which is an important feature since tight binding protein-DNA interactions are analysed at very low DNA and protein concentrations (Lundblad *et al.,* 1996).

The overall concept of studying protein-protein or DNA-protein interactions using fluorescence anisotropy is described in Figure 3-8.

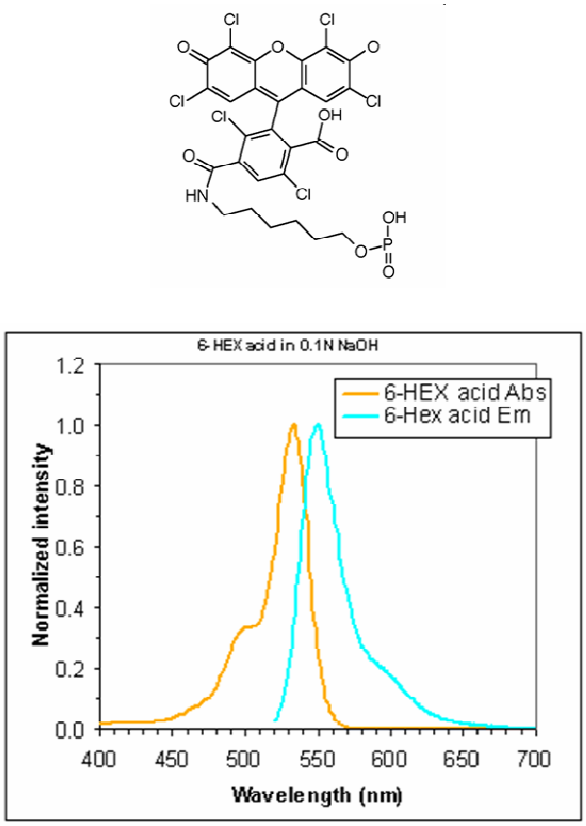


Figure 3-7. The fluorescence probe, hexachlorofluorescein (HEX).
The upper panel shows the chemical structure of HEX, while the lower panel shows the characteristic absorption and emission spectra.

The rotational diffusion of the free oligonucleotide is rapid, and the anisotropy of the fluorescent probe covalently bound to the oligonucleotide is quite low. The binding of a protein to the DNA significantly decreases the rotational motion of the fluoroscein-tagged oligonucleotide, resulting in an increase in the fluorescence emission anisotropy of the tagged DNA (Figure 3-8).

As an example, the data showing the effect on anisotropy of the fluorescently-labelled DNA as the amount of M.EcoKI increases during a titration, is shown in Figure 3-9.

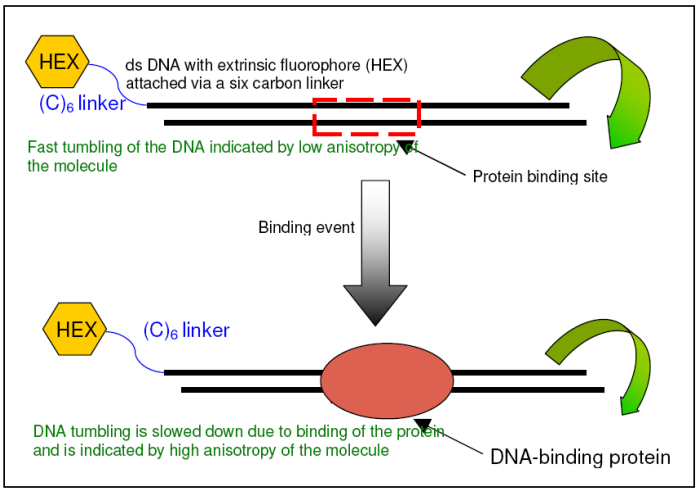


Figure 3-8. Monitoring binding by fluorescence anisotropy.
DNA "tumbles" in solution at a faster rate when unbound, exhibiting lower anisotropy of the HEX extrinsic fluorophore. The binding of a protein molecule to the DNA decreases the mobility of the DNA, and hence the "tumbling" is reduced, causing an increase in the anisotropy of the labelled HEX. Adapted from Lundblad *et al.,* 1996.

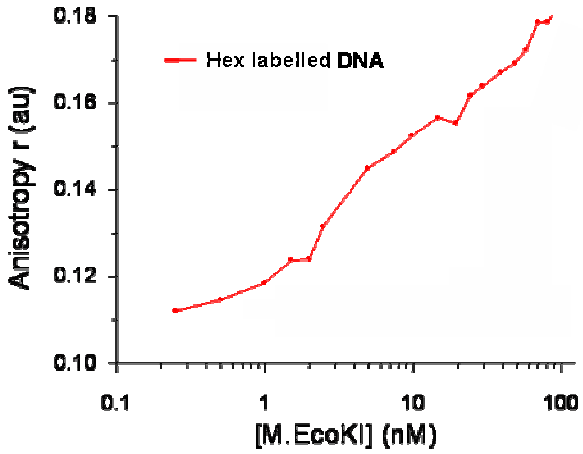


Figure 3-9. Example of a fluorescence anisotropy titration experiment.
M.EcoKI (a DNA-binding protein) was titrated into a solution containing 2 nM of a 21 bp DNA duplex fluorescently labelled at the 5' end. The binding of M.EcoKI to the DNA duplex could be monitored by anisotropy (r); a measurement of the rate of tumbling of the fluorophore. Upon binding of M.EcoKI to DNA the DNA-protein complex tumbles more slowly and r increases. The graphs show raw data that has not been fitted to a model but rather points have been joined by a line.

3.4. Results

3.4.1. Purification of Ocr proteins (wild-type and mutants)

Below is a brief outline of the Ocr protein purification steps. For a detailed description on Ocr purification refer to section 2.3.2.1. The protocol consists of two steps: (i) a DEAE (diethylaminoethyl)–Sepharose fast flow anion exchange column and (ii) a TCA precipitation treatment of the protein solution for removing contaminating nucleic acid.

E. coli expression cells were transformed with the pAR2993 plasmid, containing either one of the wild-type or mutant Ocr proteins. Prior to "large scale" (typically a 4 litre culture) protein expression, a small induction/expression test was usually carried out on a 5-10 ml culture. An example of such an induction test is shown in Figure 3-10.

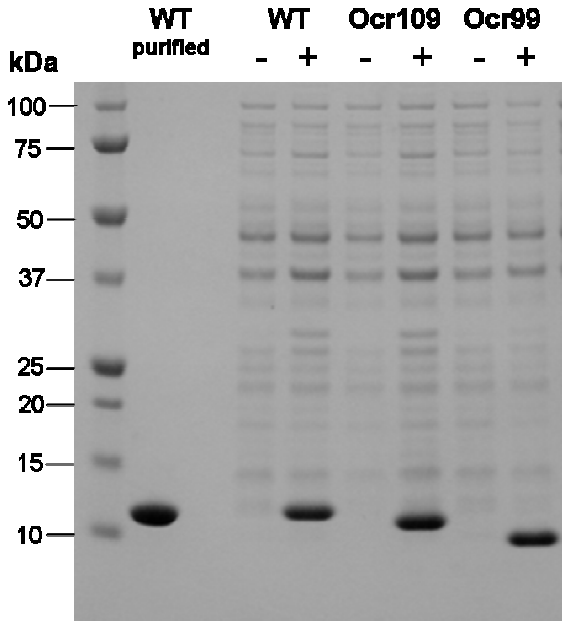


Figure 3-10. Protein expression tests.
Small scale (5-10 ml) induction tests of wild-type Ocr (WT) and two truncated forms of Ocr (Ocr99, Ocr109; the numbers correspond to the first 99 or 109 amino acids that were not removed from the protein). Numbers of Ocr99 and Ocr109 refer to position at which proteins have been truncated. Plus or minus (+/-) indicate addition or not of 1 mM IPTG when the culture reached an optimum optical density (section 2.3.2.1).

Once positive expression was verified, the experiment was repeated at a larger scale (~4 L), giving 10-15 g of cell paste. After cell breakage the clarified cell-free extract was loaded on a DEAE column. After extensive washing, 5 ml fractions were collected with a salt gradient. By referring to the UV A_{280nm} elution profile (Figure 3-11) selected fractions were further analysed by SDS-PAGE gel electrophoresis (Figure 3-12). Selected fractions (*i.e.*, fractions 15-30) containing the Ocr protein were then pooled.

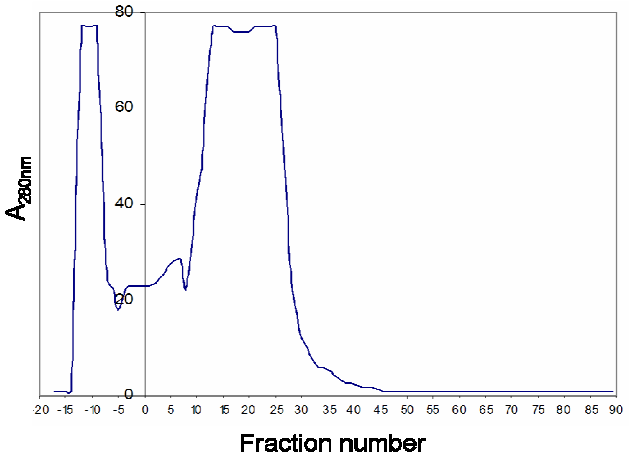


Figure 3-11. Elution profile from a DEAE column with a salt gradient for the purification of Ocr.
The 5 ml fraction numbers are indicated on the *x*-axis. The *y*-axis shows the intensity of absorption at 280 nm by the eluting material.

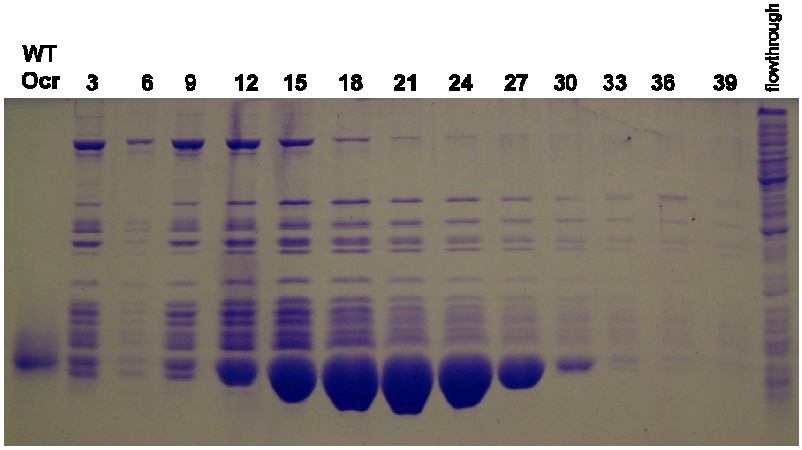


Figure 3-12. An SDS-PAGE electrophoresis gel (12% acrylamide) of eluted fractions from a DEAE column for the purification of Ocr.
WT Ocr, refers to previously purified protein used as a size marker; flowthrough, is the material that did not bind to the column. The 5 ml fraction numbers are indicated above each lane.

The absorption spectrum of the protein solution indicated a high DNA content. Therefore the contaminating DNA was removed by treating the sample with trichloroacetic acid (TCA) (Figure 3-13).

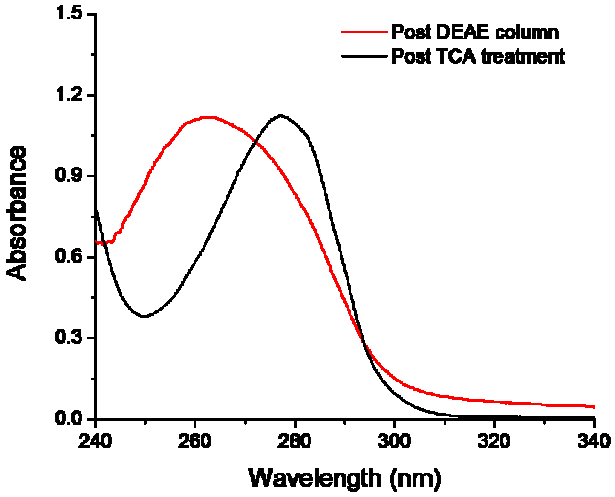


Figure 3-13. Removal of contaminating nucleic acid from the Ocr protein sample by a TCA precipitation method.
The TCA precipitation method is described in section 2.3.2.1.

Apart from removing nucleic acid, the TCA treatment also serves to purify the Ocr sample of other contaminating proteins (as indicated by the comparison of the collected fractions 15-30 of Figure 3-12 and the lane from Figure 3-10 labelled as "WT purified"). The protein solutions composed of pure Ocr protein were then dialysed (in order to remove traces of TCA), concentrated and stored at -20°C in the presence of 50% (by volume) glycerol or at -80°C immediately after flash freezing in liquid nitrogen. The Ocr proteins that were expressed and purified are shown in Table 3-1 and in Figure 3-14.

Table 3-1. Ocr proteins expressed and purified from ~4 L cultures, in order to study their binding to M.EcoKI *in vitro*.

Ocr protein	Amount (mg)	Ocr protein	Amount (mg)
WT Ocr	80	S68C	39
D12N	50	E59C	24
E16A	45	D62C	19
E16Q	5	Ocr109	51
E20A	32	Ocr99	47
D25A	12	D12A, D26N	17
D25C	36	D12N, D26N	20
D26N	16	D12N, E87D	42
		E16A, D25A	12

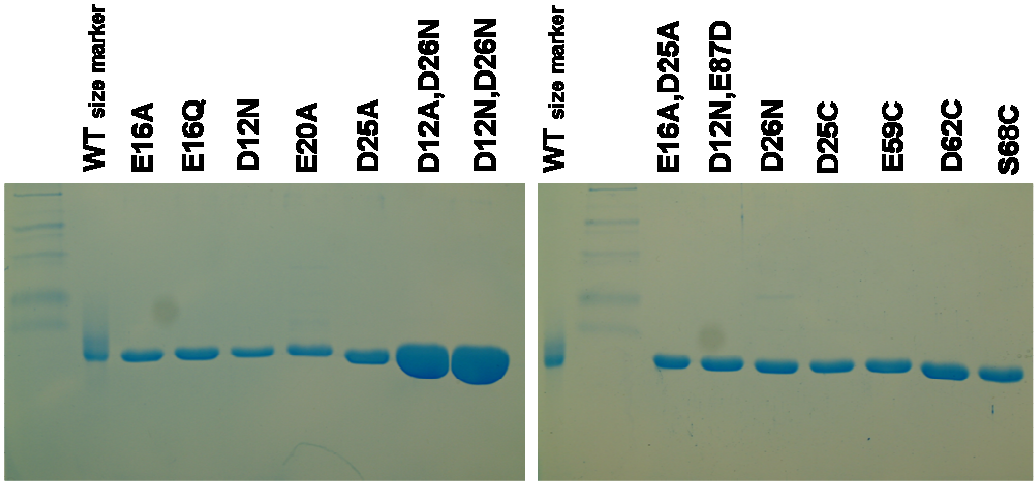


Figure 3-14. SDS-PAGE gel (15% acrylamide) of purified Ocr mutants.
WT refers to wild-type Ocr used as a size marker.

3.4.2. Purification of M.EcoKI

Below is a brief outline of the M.EcoKI purification steps. For a detailed description on M.EcoKI purification refer to section 2.3.2.2. The protocol consists of 3 purification columns; two anion exchange columns (CM-Sepharose and Heparin) and a gel-filtration column.

M.EcoKI was prepared from 10 litres of BL21(DE3) cells transformed with pJFMS. In order not to overload the columns the M.EcoKI purification procedure was repeated separately for each of the 15gr cell pastes. After cell breakage the clarified cell-free extract was loaded on a CM column. The CM column was washed with at least one column volume of buffer and then ~5 ml fractions were collected using a salt gradient. By reference to the UV A_{280nm} elution profile (Figure 3-15) a series of fractions were analysed by SDS-PAGE gel electrophoresis. The fractions containing the protein of interest (*i.e.*, 36-51) as determined by the SDS-PAGE gel (Figure 3-16), were pooled and then dialysed against buffer to remove NaCl.

The protein sample was then loaded on a Heparin column. After washing the column with at least one column volume of buffer, fractions were collected with a salt gradient. By reference to the UV A_{280nm} elution profile (Figure 3-17) selected fractions were further analysed by SDS-PAGE. The fractions containing M.EcoKI (*i.e.*, fractions 23-37), as observed by the SDS-PAGE gel in Figure 3-18, were pooled and the protein mixture was then precipitated by 70% saturation with ammonium sulfate.

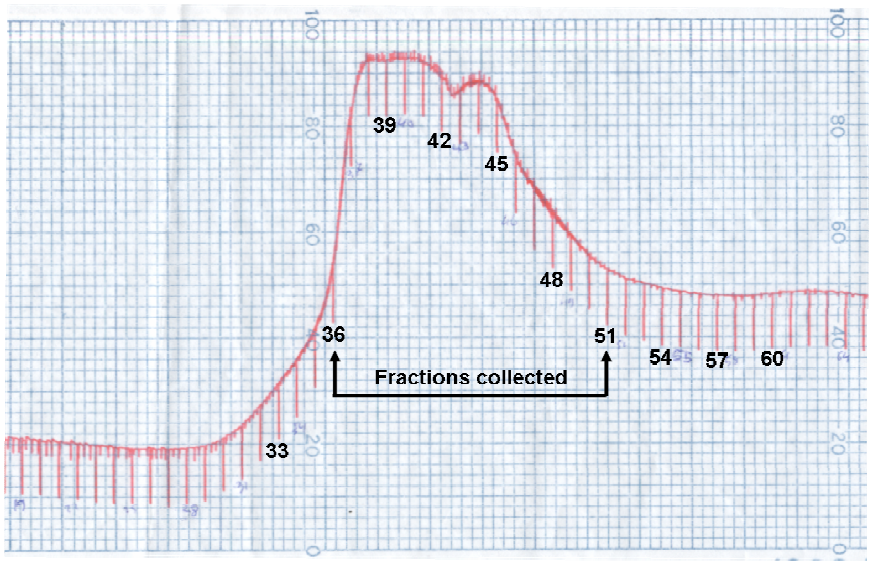


Figure 3-15. Elution profile from a CM column with a salt gradient for the purification of M.EcoKI.
Numbers in bold correspond to the 5 ml fraction numbers. *Y*-axis corresponds to the intensity of absorption at 280 nm by the eluting material.

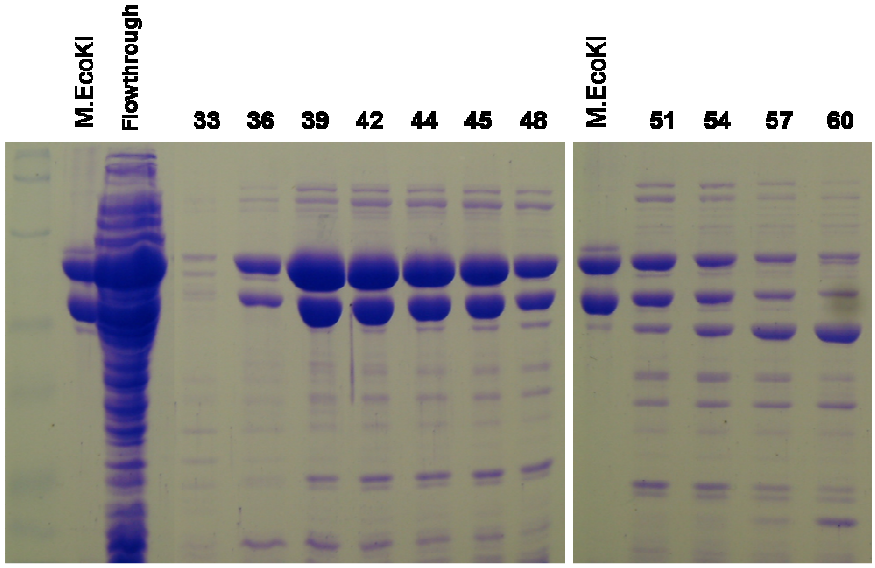


Figure 3-16 An SDS-PAGE electrophoresis gel (10% acrylamide) of eluted fractions from a CM column for the purification of M.EcoKI.
M.EcoKI, refers to previously purified protein, which was used as a size marker; flowthrough, is the material that did not bind to the column. The 5 ml fraction numbers are indicated above each lane.

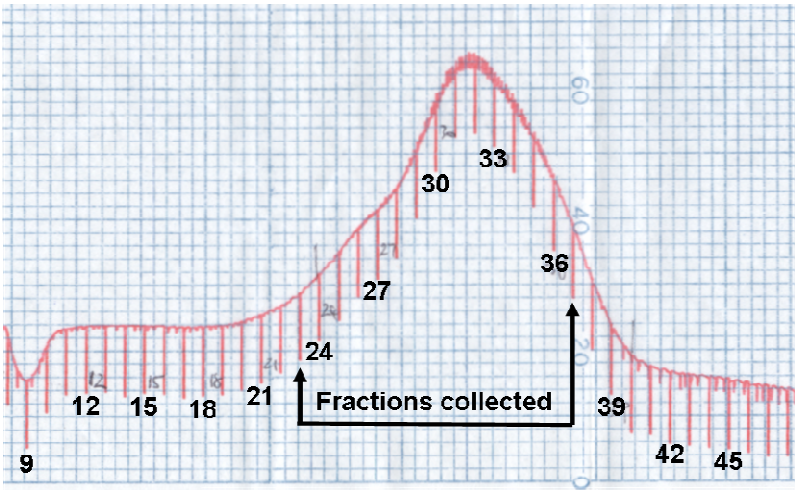


Figure 3-17. Elution profile from a Heparin column with a salt gradient for the purification of M.EcoKI.
Numbers in bold correspond to the 5 ml fraction numbers. *Y*-axis corresponds to the intensity of absorption at 280 nm by the eluting material.

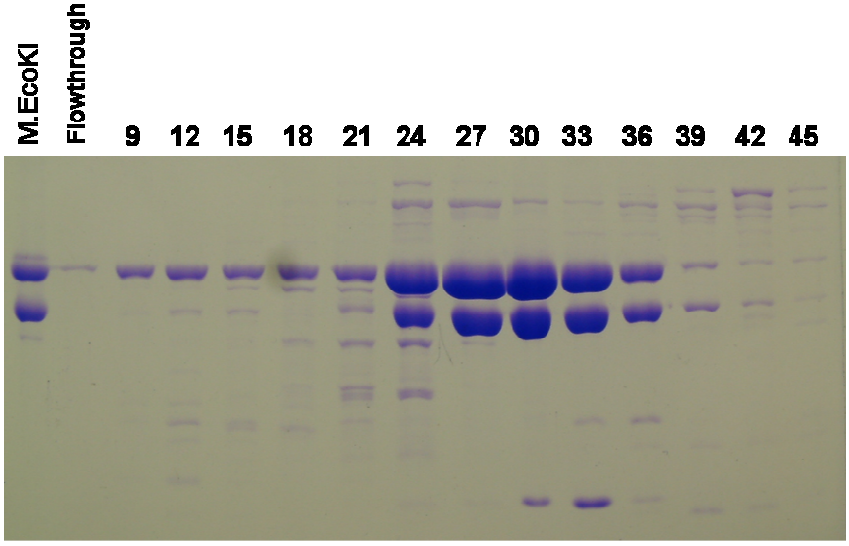


Figure 3-18. An SDS-PAGE electrophoresis gel (10% acrylamide) of eluted fractions from a Heparin column for the purification of M.EcoKI.
M.EcoKI, refers to previously purified protein, which was used as a size marker; flowthrough, is the material that did not bind to the column. The 5 ml fraction numbers are indicated above each lane.

The precipitants were resuspended in a minimal volume of a buffer also containing 0.2 M NaCl. This was then loaded on a gel-filtration column. Fractions were then analysed by SDS-PAGE gel electrophoresis. The fractions believed to contain the M_2S_1 form of the protein (*i.e.*, fractions 33-39) as observed by the

SDS-PAGE gel in Figure 3-19, panel A, were then pooled and stored at -20°C in the presence of 50% glycerol (by volume). Fractions 40-45 were composed of the M_1S_1 form of the enzyme and were therefore stored separately. The overall purification of M.EcoKI from ~30 g of cell paste, yielded about 124 mg of protein (Figure 3-19, panel B).

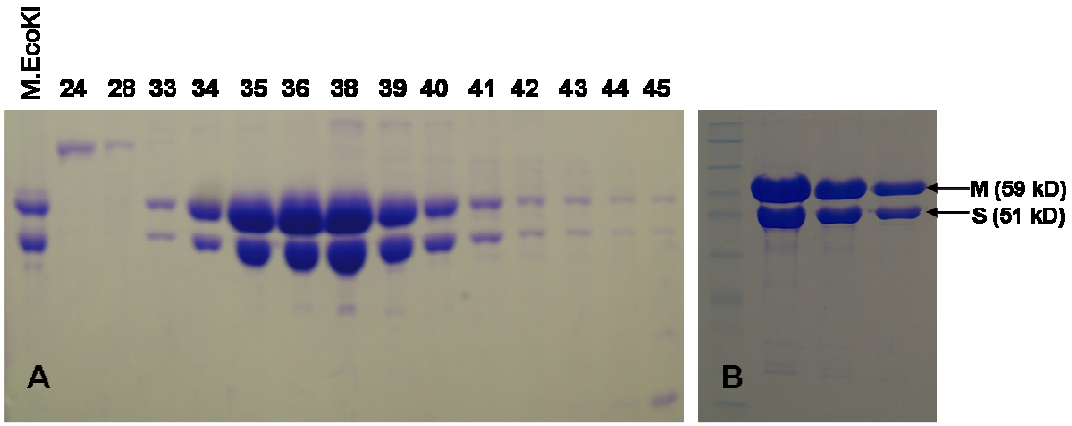


Figure 3-19. SDS-PAGE gels (10% acrylamide) of eluted fractions from a gel filtration column for the purification of M.EcoKI and the final purified product. Panel A showing the eluted fractions. Fraction numbers are indicated above each lane. M.EcoKI, refers to previously purified protein, which was used as a size marker. Panel B shows the purified M.EcoKI, loaded at three different concentrations.

3.4.3. Assessing *in vivo and in vitro* activity of Ocr proteins

The results of the *in vivo* phage restriction assays, ITC measurements and fluorescence anisotropy assays are given in Table 3-2 (page 81).

3.4.3.1. Fluorescence Anisotropy

In order to assess possible changes in the free energy of binding for the Ocr proteins to M.EcoKI, a fluorescence anisotropy competition assay was used as is described in section 2.4.6. This assay is more appropriate than ITC for interactions in the pM K_d range. A control titration of a labelled DNA oligonucleotide duplex with M.EcoKI showed a typical binding behaviour with a K_d of 2.11 ± 0.35 nM similar to previous measurements (Powell *et al.*, 1998a). A weaker non-specific binding event was observed at high concentrations of M.EcoKI with a K_d of 349 ± 73 nM. The addition of 50 nM Ocr dimer to the solution containing the duplex reduces the

proportion of M.EcoKI available to bind to the duplex so that more M.EcoKI has to be added to observe DNA binding. For all of the Ocr proteins tested, the tight binding between Ocr and M.EcoKI resulted in little DNA binding being observed until the M.EcoKI had bound all of the available Ocr. The way in which the data was fitted and hence binding constants calculated is described in section 2.4.6. Examples of raw anisotropy data and how these were averaged and fitted to a binding model are shown in Appendix B. The mutant forms of Ocr have K_d values ranging from roughly equal to 4.7 times larger than wild-type Ocr binding to M.EcoKI. The mutant Ocr proteins with substitutions of amino acids E16, D25 or D26 appear to have the weakest interactions with M.EcoKI although the magnitude of the effect is not large (Figure 3-21).

3.4.3.2. ***In vivo*** **phage restriction assays (work done by Emily H. Pritchard)**

Once the mutations had been verified by sequencing, the activity of the mutated Ocr proteins was tested *in vivo*. These experiments were conducted at low levels of heterologous gene expression in the absence of induction by IPTG. *E. coli* cells with (NM1049) or without (NM1261) the chromosomal EcoKI R/M system were transformed with either the expression vector (a control) or the vector containing the wild-type Ocr gene or the mutated Ocr genes. Each of these transformed strains was then challenged with phage lambda $\lambda_{v.o}$ (unmodified lambda DNA) or $\lambda_{v.k}$ (modified lambda DNA). As expected, cells transformed with the vector alone showed a strong reduction (six orders of magnitude) in the efficiency of plating of $\lambda_{v.o}$ compared to $\lambda_{v.k}$. Cells transformed with the plasmid expressing the wild-type Ocr showed essentially identical numbers of plaques with both $\lambda_{v.o}$ and $\lambda_{v.k}$ showing that Ocr had inhibited the R/M system. Assaying the mutated forms of Ocr showed that all of the mutants were fully active and indistinguishable within the limits of the errors associated with this sort of assay. There was no significant difference between single amino acid substitutions, double substitutions or the C-terminal deletions of 7 or 17 amino acids. The data calculated is shown in Table 3-2.

The purified proteins were then tested for binding to the methyltransferase core of EcoKI. It has previously been shown that Ocr binds extremely tightly to M.EcoKI (Kd ~ 100 pM) (Atanasiu *et al*., 2002). The interaction between the wild-type and mutated Ocr proteins with M.EcoKI was measured using ITC as a direct method, and the fluorescence anisotropy competition assay as an indirect method.

3.4.3.3. Isothermal Titration Calorimetry (work done by Dr. Gareth A. Roberts)

The ITC data showed a highly exothermic interaction between Ocr and M.EcoKI with a stoichiometry consistent with one Ocr dimer per M.EcoKI. The sharp transition (step-like function) of the ITC isotherm (Figure 3-4) does not allow accurate determination of Ka (association constant) and hence free energy of binding; although ΔH values derived from such data are highly reliable. The enthalpy change of the M.EcoKI:Ocr mutant interactions varied by ±40% from the -20.5 kcal/mol observed for the interaction of the wild-type Ocr with M.EcoKI (Figure 3-20). The substitutions D12N, D25A, D62C and S68C gave rise to the largest differences from the wild-type protein.

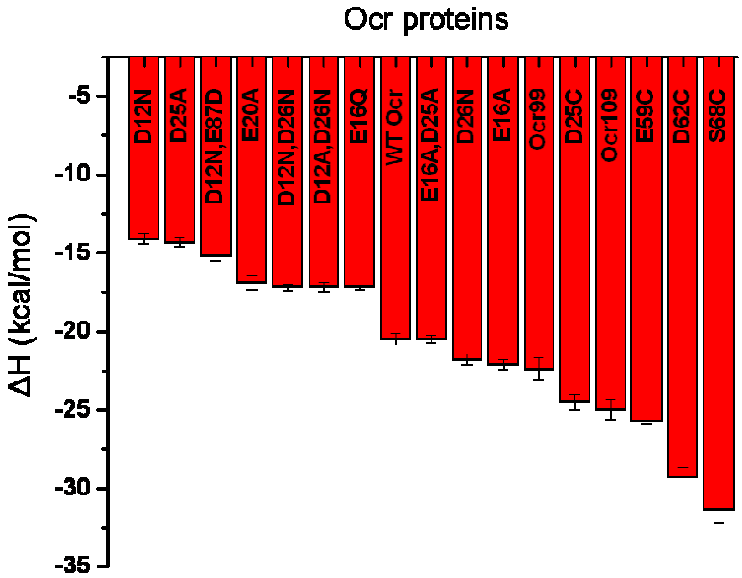


Figure 3-20. A bar chart describing the enthalpy of binding of the different Ocr proteins to M.EcoKI, as measured by ITC.
Experimental details are described in section 2.4.7 and the data is tabulated in Table 3-2.

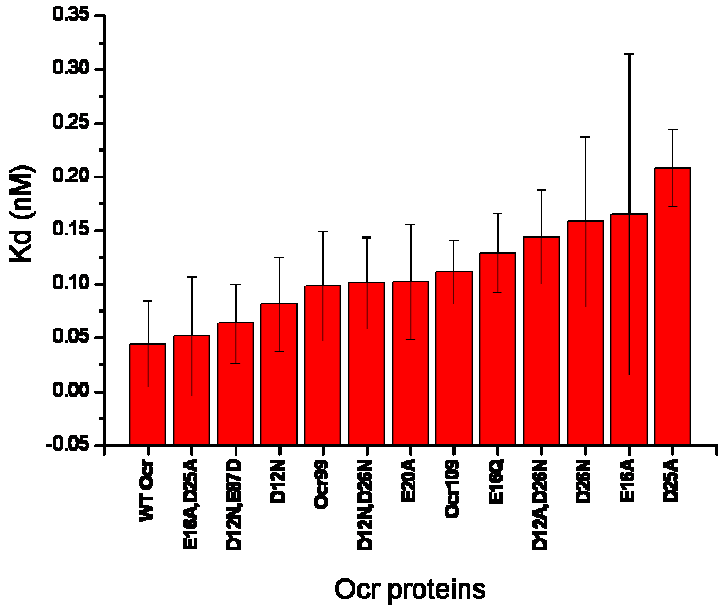


Figure 3-21. Bar chart showing the dissociation constants (Kd) in nM for the various M.EcoKI:Ocr mutant interactions as measured by fluorescence anisotropy.
Experimental conditions and data analysis are described in section 2.4.6. Data is shown in Table 3-2. The Kd errors represent the goodness of the fit of the averaged data of n=3. Refer to Appendix B for some examples of raw anisotropy data.

Table 3-2. A list of the Ocr mutant proteins and their *in vivo* and *in vitro* behaviour.

Protein	Eop with unmodified phage λ_{vir}	Eop with modified phage λ_{vir}	Enthalpy of binding of Ocr to M.EcoKI (kcal/mol)	Stoichiometry of binding, Ocr:M.EcoKI	Dissociation constant (nM) for Ocr binding to M.EcoKI determined by anisotropy
Plasmid vector alone	1.703×10^{-6}	0.973			
Wild-type Ocr	0.792	1.493	-20.5 ± 0.3	0.87 ± 0.01	0.044 ± 0.040
D12N	0.912	1.039	-14.1 ± 0.3	0.94 ± 0.01	0.081 ± 0.044
E16A	2.504	2.985	-22.1 + 0.3	0.82 + 0.01	0.165 + 0.150
E16Q			-17.2 ± 0.1	0.70 ± 0.00	0.129 ± 0.037
E20A	0.681	0.996	-16.9 ± 0.4	0.85 ± 0.01	0.102 ± 0.054
D25A			-14.3 ± 0.3	0.96 ± 0.01	0.208 ± 0.036
D25C	0.541	0.502	-24.5 ± 0.5	0.78 ±0.01	
D26N	0.821	0.761	-21.8 ± 0.3	0.76 ± 0.01	0.158 ± 0.080
E59C	1.135	0.634	-25.7 ± 0.2	0.86 ± 0.00	
D62C	1.056	1.018	-29.3 ± 0.6	0.78 ± 0.01	
S68C	1.066	0.893	-31.4 ± 0.8	0.75 ± 0.01	
D12A, D26N	0.78	0.986	-17.2 ± 0.3	1.00 ± 0.01	0.144 ± 0.044
D12N, D26N	1.506	1.317	-17.2 ± 0.2	0.89 ±0.01	0.101 ± 0.043
D12N, E87D	0.982	1.008	-15.2 ± 0.3	0.98 ± 0.01	0.063 ± 0.037
E16A, D25A	0.728	0.833	-20.5 + 0.2	0.84 + 0.01	0.051 + 0.056
Ocr deletion, 109 amino acids long	0.603	1.395	-25.0 ± 0.6	0.74 ± 0.01	0.111 ± 0.030
Ocr deletion, 99 amino acids long	1.355	1.022	-22.4 ± 0.7	0.71 ± 0.01	0.098 ± 0.051

Eop: Efficiency of plating from phage assays showing the ratio of plaques in NM1049 and NM1261, with each of the two phages. The Kd errors of anisotropy represent the goodness of the fit of the averaged data of n=3.

3.5. Discussion

Here, attention had been focussed on a stretch of amino acids of Ocr that mimic a single strand of DNA. Mutants were generated that contained one or two amino acid changes per monomer in which residues bearing a negatively-charged side chain were mutated to a neutral side chain. It had been anticipated that at least some of the mutations would significantly weaken the interaction of the protein with M.EcoKI or show a ΔΔG of binding greater than 2-3 kcal/mol. The results clearly show that this postulate was wrong as all of the mutants retained full activity *in vivo* and bound strongly to M.EcoKI *in vitro*.

In relation to the Kd values obtained by fluorescence anisotropy, changes at E16, D25 or D26 appeared to show the weakest level of binding. However the effect may not be very reliable because double mutants containing any of the aforementioned three mutations show no significant difference in relation to the binding affinity exhibited by wild-type Ocr (Figure 3-21). The dissociation constant varies between 0.044 nM and 0.208 nM, a four-fold difference. This variation in K_d only amounts to free energy changes between -14.1 kcal/mol (wild-type Ocr) and -13.2 kcal/mol (D25A); a rather small variation (ΔΔG = 0.9 kcal/mol), but still significant in terms of showing a probable role in the association process. Intriguingly, a study by Bogan and Thorn, 1998 indicates that the contribution of charged residues that do not form part of interacting interfaces contribute ~1.0 kcal/mol. Nevertheless, Janin (1997) has postulated that such residues may contribute to the binding mechanism by electrostatic steering of the two interacting partners. Indeed, the C-terminal region of Ocr may play a role in the initial "recruitment" and binding event with Type I enzymes. This hypothesis can be resolved with stop-flow or surface plasmon resonance experiments, which would measure the rate of association of wild-type and mutant Ocr proteins (in particular the truncated versions of the protein) with its target enzyme. Indeed similar experiments with the barnase-barstar association have shown that long range electrostatic interactions increase the rate of binding by a factor of 10^5 (Schreiber and Fersht, 1996; Janin, 1997).

The finding that Ocr99 (which has 10 acidic residues removed *i.e.*, 29% of total) still retains very tight biding to M.EcoKI with no discernable difference compared to wild-type Ocr, may be due to the fact that these residues are not involved in the interaction interface with M.EcoKI. A model of the Ocr:M.EcoKI interaction (Kennaway *et al.*, 2009) positions the C-terminal part of Ocr just outside the region engulfed by the M.EcoKI complex (Figure 3-22, panel B). It is anticipated that the C-terminal tail of Ocr may be interacting with the R subunit of the EcoKI nuclease. A possible effect of these C-terminal deletions on the binding affinity of Ocr to the complete EcoKI nuclease is therefore likely.

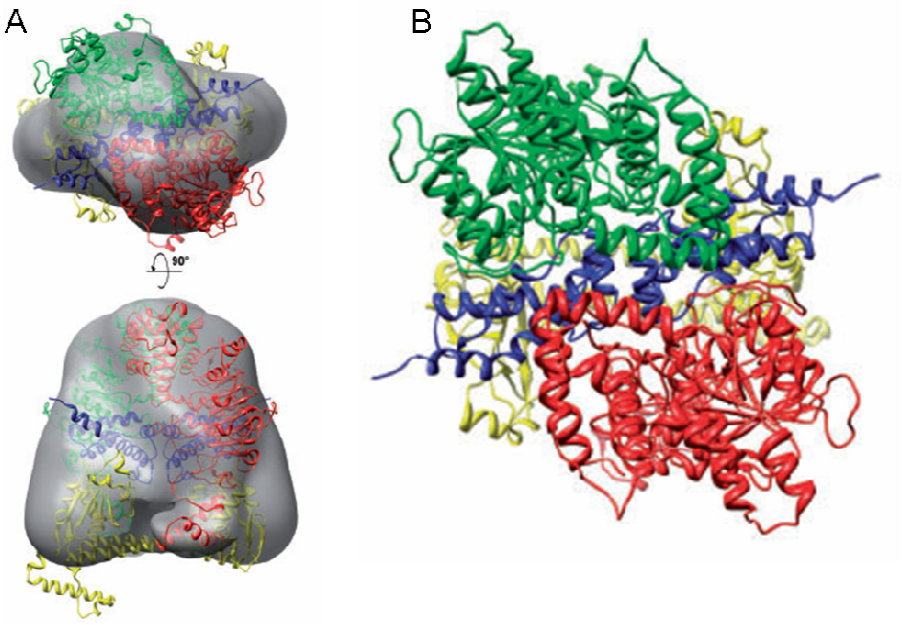


Figure 3-22. A model of M.EcoKI bound to Ocr at ~18Å resolution as determined by negative-stain electron microscopy (EM).
A; Two orthogonal semi-transparent surface representations of the EM 3D reconstruction each with a view of the modelled coordinates fitted as a rigid body (green and red—HsdM; blue—Ocr dimer; yellow— HsdS). B; A view of the M.EcoKI–Ocr atomic model. Adapted from Kennaway *et al.*, 2009.

None of the residues chosen for the single/double mutational analysis of Ocr affect the binding affinity of the Ocr:M.EcoKI association to any great extent. There are three possible reasons for this observation; (i) the mutations are not juxtaposed with regions of highly energetic protein-protein interacting regions (*i.e.*, the residues do not form part of an interaction hot spot) (ii) the loss of a potential electrostatic

interaction at the interface of the Ocr:M.EcoKI complex are compensated by other negatively charged residues in the proximity of the mutation (iii) the very tight protein-protein association, render subtle changes in binding affinity practically undetectable, within the constraints of the methodology used and experimental error.

The enthalpy of binding changed in an unpredictable manner with some mutants showing more exothermic binding and some less exothermic binding to M.EcoKI compared with the wild-type protein. The enthalpy change varied between -31.4 kcal/mol to -14.1 kcal/mol, as shown in Figure 3-20. In relation to the single mutants D12N, D25A, D62C and S68C, which gave rise to the largest enthalpic effects, and assuming an insignificant entropic component, then a possible contribution of these residues to the binding affinity of the interaction could be qualitatively inferred, since $\ln Kd = \Delta H/RT - \Delta S/R$.

Ocr mutants containing a Cys residue (wild-type Ocr does not contain any Cys residues) gave highly exothermic associations with M.EcoKI. This may be due to enthalpically favourable sulfur-aromatic interactions (s-π interactions) between the sulfur atom of the Ocr Cys residue and aromatic residues (Trp, Tyr or Phe) of M.EcoKI at the interacting interface (Meyer *et al*., 2003). Residues of particular interest (D25, E59, D62 and S68) are highlighted in Figure 3-23. In the absence of a crystal structure of the Ocr:M.EcoKI interaction, the reasons for the variable enthalpy changes are difficult to define. Deviation however will arise from the different characteristics of each of the possible s-π interactions (*e.g*., distance, angle). Yet, it is highly likely that residues S68 and D62 (each of which showed a $\Delta\Delta H$ of about -10 kcal/mol) are in close proximity with M.EcoKI, given that s-π interactions have an optimum distance of 4.0±0.5 Å (Ringer *et al*., 2007).

Even though certain residues (shown in Figure 3-23) appear to be part of the Ocr:M.EcoKI interacting interface, their role and contribution to the binding affinity is not yet clear. Therefore mutagenesis studies on a much larger scale than attempted in this chapter are required to fully understand the DNA mimicry exhibited by Ocr. Two strategies can be adopted to target multiple acidic residues on the Ocr protein. One methodology is to chemically modify carboxylic acid groups on the protein surface. These experiments will be discussed in Chapter 4. An alternative approach is to use mutagenesis to target multiple acidic residues. Based on the three-dimensional

structure of Ocr, multiple amino acid residues that correspond to patches of negatively charged regions on the surface of the protein are selected for mutagenesis. In this way, it should be feasible to deduce regions of the protein that are particularly important for interaction with M.EcoKI. This work is described in Chapter 5.

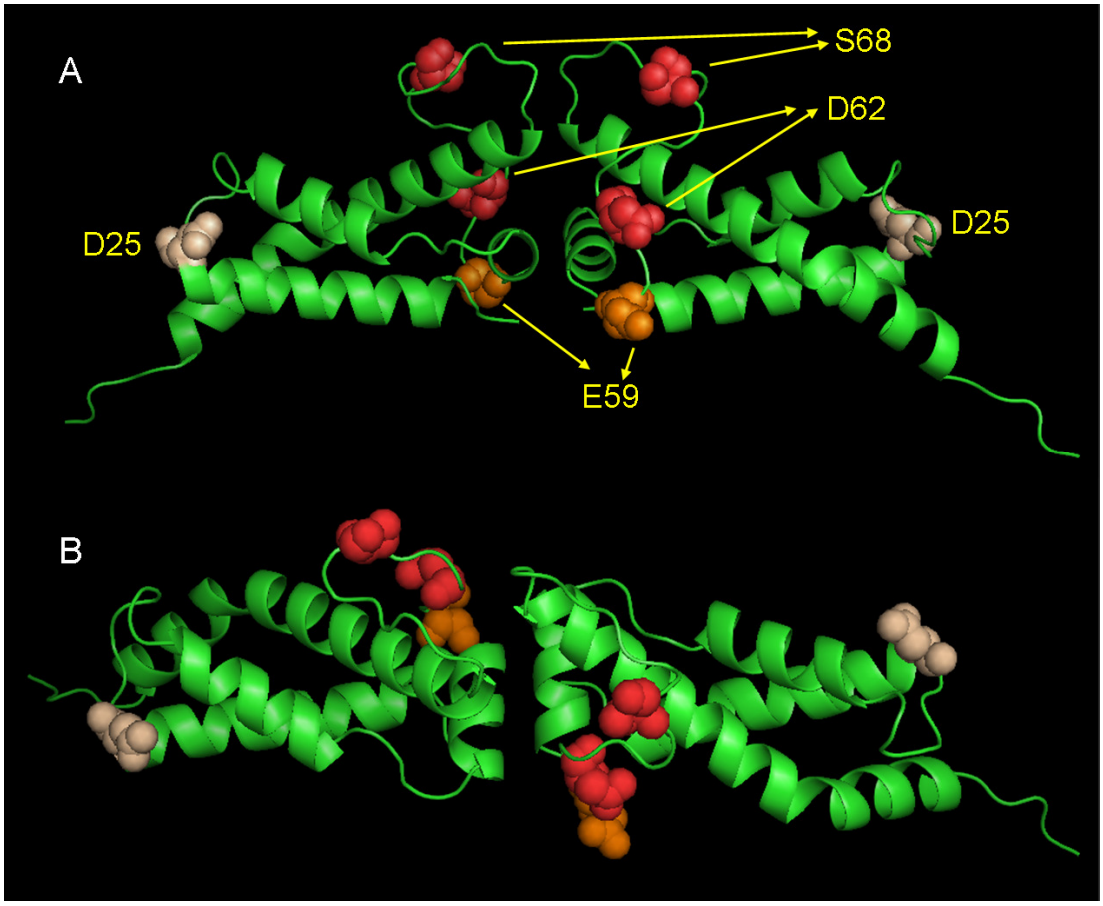


Figure 3-23. Residues of interest on the Ocr molecule.
Panel A; A view of the Ocr dimer from the side with residues S68, D62, E59 and D25 shown as spheres. Panel B; a 90º angle view of the Ocr molecule from panel A. The residues highlighted are thought to be in close proximity with M.EcoKI upon complex formation, because Cys residues at these positions gave large and favourable enthalpy changes, which are associated with s-π interactions. Refer to Figure 3-20.

Chapter 4. Chemical Modification of Surface Exposed Carboxylates of the Bacteriophage T7 Ocr Protein

4.1. Introduction

As has already been mentioned, the surface of Ocr is replete with acidic residues (34 per monomer) many of which mimic the distribution of phosphates on DNA. It is therefore believed that the carboxylates of Ocr play a key role in the anti-restriction activity of the protein. The single or double mutational analysis, described in Chapter 3, did not identify the critical residues involved in the interaction with M.EcoKI. Thus, the non-specific and random removal of surface acidic residues by chemical modification was undertaken as an alternative strategy towards understanding the evolutionary significance of the extremely low pI of Ocr.

4.2. Aims

This chapter explores the hypothesis that chemical modification of the acidic residues on the Ocr protein surface will significantly impair its interaction with a model Type I enzyme. Therefore the specific aims for this chapter were to:

- Achieve the gradational removal of acidic residues from the Ocr surface by chemically derivatising the carboxyl moieties.

- Assess the interaction of the chemically modified Ocr with a Type I R/M system and hence deduce how many negative residues are critically important in the interaction.

- Assess the relative importance of the electrostatic mimicry of Ocr *versus* the Ocr mimicry of the bent shape of DNA.

4.3. Overall strategy

The Ocr protein was produced and purified as described in section 2.3.2.1. Acidic residues of Ocr were chemically modified as described in section 2.3.3. The extent and nature of the derivatisation was initially analysed by: polyacrylamide gel electrophoresis under non-denaturing conditions and denaturing SDS-PAGE gels (section 2.3.1); ion exchange chromatography (section 2.4.2), MALDI-TOF and FT-ICR mass spectrometry (sections 2.4.3 and 2.4.4 respectively). Protein folding and stability were assessed by circular dichroism (CD), differential scanning calorimetry (DSC) and guanidinium hydrochloride (GdmCl)-induced denaturation (sections 2.4.1, 2.4.8 and 2.4.5 respectively). The binding affinities for the interaction of the Ocr samples with M.EcoKI (the methyltransferase core) and EcoKI (the entire nuclease) were investigated by: isothermal titration calorimetry (ITC) of Ocr with M.EcoKI (section 2.4.7); competition between Ocr and DNA for binding to M.EcoKI (section 2.4.6) and inhibition of the nuclease activity of EcoKI by Ocr (section 2.4.9).

4.4. Results

4.4.1. Method development

The water-soluble carbodiimide EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) was used to specifically modify the carboxyl groups of Asp and Glu residues and the C-terminus of Ocr. It is the most frequently used carbodiimide with a wide range of applications in biological and chemical systems such as: peptide synthesis (Anderson *et al*., 1964), protein cross-linking and complex formation (Wu-Chou *et al*., 1984) conjugation of hapten-carrier immunogens (Hermanson, 2008), cellular modification (Thelen *et al*., 1988), enzyme activity (Perfetti *et al*., 1976; Pedemonte and Kaplan, 1986; Nyvall *et al*., 2000) and biomaterials (Sannino *et al*., 2005; Lee *et al*., 1996). The physical and chemical properties of EDC have been extensively studied (Hoare & Koshland, 1967; Carraway & Koshland, 1972; Williams & Ibrahim, 1981; Nakajima & Ikada, 1995; Chan & Cox, 2007). In aqueous solution under acidic conditions EDC is anticipated to react with Asp, Glu, Cys and Tyr residues. The absence of a Cys residue in Ocr rules out any unwanted side reactions with sulfhydryl groups. Regeneration of unsubstituted tyrosyl residues, the phenolic hydroxyl group of which reacts with EDC to form a relatively stable *O*-arylisourea, is achieved by treating the protein with hydroxylamine (Smyth, 1967; Carraway & Koshland, 1968). EDC reacts with surface exposed carboxyl groups to form highly reactive *O*-acylisourea intermediates that are susceptible to nucleophilic attack. In this study, the reactions were performed in the presence of a large molar excess of either dimethylamine or ammonium hydroxide as nucleophile (*i.e*., D- and N-series, respectively). The end result is the formation of a stable amide bond with concomitant loss of one negative charge for each residue modified (Figure 4-1).

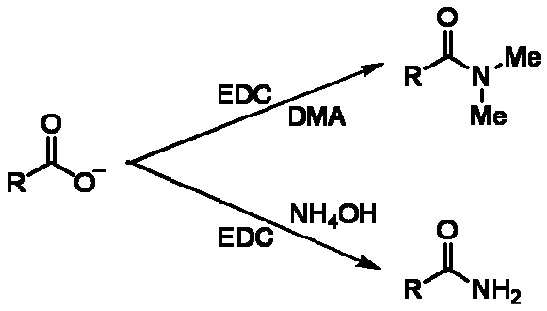


Figure 4-1. Derivatisation of carboxylate side chains of Ocr.

Asp or Glu residues of Ocr were modified by using EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride) in the presence of either DMA (dimethylamine) (D-series) or ammonium hydroxide (N-series) as nucleophile. The end result is the loss of one negative charge per carboxylate modified, under physiological pH conditions.

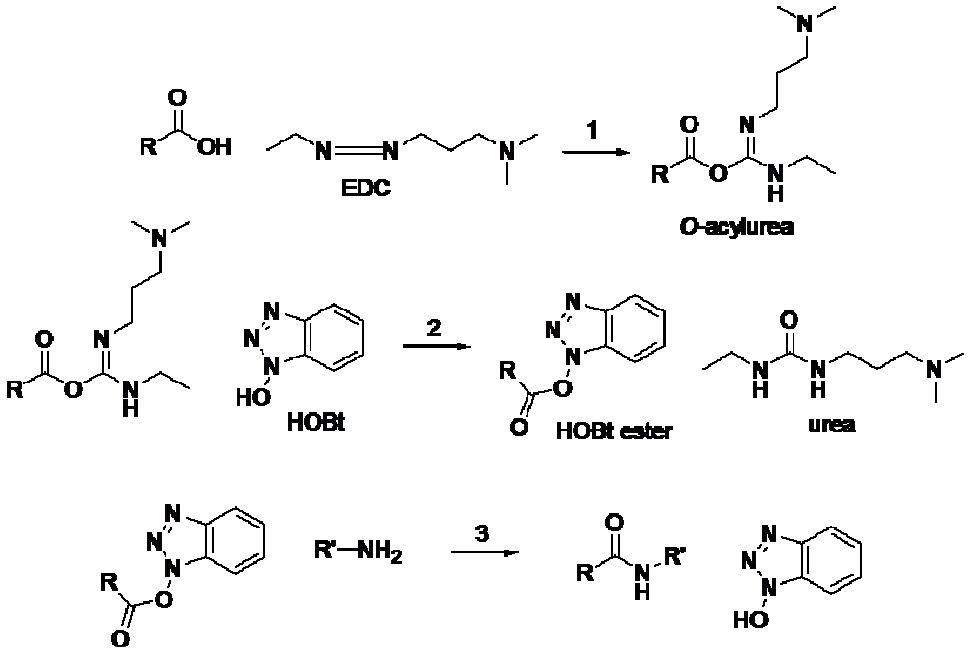


Figure 4-2. Reactions for the coupling of the carboxylate groups of the protein with an amine in the presence of EDC and HOBt (N-hydroxybenzotriazole).

Reaction 1, EDC reacts with a carboxyl group to form the activated ester O-acylisourea. Reaction 2 shows the O-acylisourea reacting with HoBt to form the HoBt ester. Reaction 3 shows the final coupling step of a primary amine and the HoBt ester to form a stable amide bond.

The *O*-acylisourea intermediate (Figure 4-2, reaction 1) however is unstable and can either react with water, regenerating the initial carboxylate and hydrolysing EDC to its urea derivative, or rearrange to an *N*-acylurea, thus forming a stable adduct on the protein. This undesirable *N*-acylurea rearrangement is largely avoided by carrying out the reaction in the presence of (i) a high concentration of nucleophile and (ii) *N*-hydroxybenzotriazole (HOBt). HOBt reacts with the *O*-acylisourea to form a more stable activated ester (Figure 4-2, reaction 2), greatly increasing the overall coupling efficiency (Figure 4-2, reaction 3) (Chan & Cox, 2007) and preferentially mediating the reaction with an amine. This, however, can include the ε- and α-amino groups of Lys residues and the protein N-terminus respectively, leading to intramolecular crosslinking and possibly protein polymerisation. Nevertheless, HOBt minimizes the adventitious formation of ester bond crosslinks because the HOBt ester is less susceptible than the *O*-acylisourea to nucleophilic attack by OH groups (*i.e.*, Tyr, Ser and Thr residues). An additional mechanism has been reported by Nakajima and Ikada (1995), whereby the *O*-acylisourea intermediate may react with a neighbouring free carboxylate to form an acid anhydride. This labile anhydride, which is highly susceptible to hydrolysis, can either react with an amine or a hydroxyl to form an amide or an ester bond, respectively (Gounaris & Perlmann, 1967). It is probable that the modification reaction of surface exposed carboxyl groups on Ocr occurs through a combination of the aforementioned species (*O*-acylisourea, HOBt ester and acid anhydride). Considering the close proximity of side chains within a protein and the variability of their pKa values, depending on their specific microenvironment, it is clear that a limited number of side reactions are unavoidable.

The Ocr surface is replete with carboxylates, which when activated with EDC can react with other nucleophilic amino acids both intra- and intermolecularly. It was therefore apparent that the modification procedure should be performed in dilute protein solution (*i.e.*, as dilute as practically possible) and in the presence of a very large molar excess of nucleophile. A protein concentration of ~3 μM and a nucleophile concentration of ~750 mM were chosen. Due to the high molarity of the nucleophile a buffer did not seem practical and was therefore omitted, since the nucleophile itself acted as a buffering system. In order to prevent aggregation of

proteins in solution and improve protein stability and solubility, about 200 mM NaCl was included in the reaction mixture. According to Chan and Cox (2007) the rate-limiting step for carboxylic acid modification is determined by formation of the *O*-acylisourea intermediate, arising from the reaction of a carboxylate anion with the doubly protonated EDC. Thus the two parameters that greatly impact on the overall rate of the reaction are pH and EDC concentration. Hence, preliminary experiments were carried out in order to optimise these reaction conditions (pH and [EDC]) for this study. Throughout this study a 1:1 molar ratio of EDC:HoBt was employed because HoBt has no effect on the reaction rate of amidation (Chan & Cox 2007). Furthermore, HoBt concentration remains constant with time (due to regeneration after each modification cycle) and is present in a huge molar excess in relation to the Ocr protein.

Methods used to assess the various protocols were: (i) Monitoring the appearance of the reaction mixture (*i.e.*, formation of aggregates); using UV spectral (240-340 nm) analysis for soluble protein determination and observing any increase in Rayleigh scattering, (ii) Analysing the migration patterns of the chemically modified Ocr samples on a non-denaturing polyacrylamide gel, where migration (in relation to unmodified Ocr) depends on the overall charge of the protein, (iii) Far UV circular dichroism (CD) spectra for observing possible changes to the protein secondary structure.

4.4.1.1. Determination of optimum pH

It is widely accepted that when working with peptides and proteins, EDC-mediated amide bond formation occurs most effectively between pH 4.5 and 7.5. EDC, which is highly susceptible to hydrolysis under acidic conditions, is more stable at or above pH 6.5 (Nakajima & Ikada, 1995). With this in mind, two reactions were carried out at pH ~5.3 and ~6.5 in the presence of 50 mM EDC, 50 mM HoBt, 750 mM ammonium hydroxide and 200 mM NaCl. The reactions were otherwise carried out as described in section 2.3.3. Even in the absence of a buffer the pH of the reaction remained constant. The protein samples were then analysed by polyacrylamide gel electrophoresis under non-denaturing conditions (Figure 4-3). As was expected the reaction proceeds at a faster rate at pH 5.3 rather than pH 6.5,

exemplified by the change in the rate of migration of the modified protein samples. The more gradual loss of negative charge from the Ocr surface (Figure 4-3, panel A) was preferred. Thus, all subsequent reactions were performed at pH 6.5.

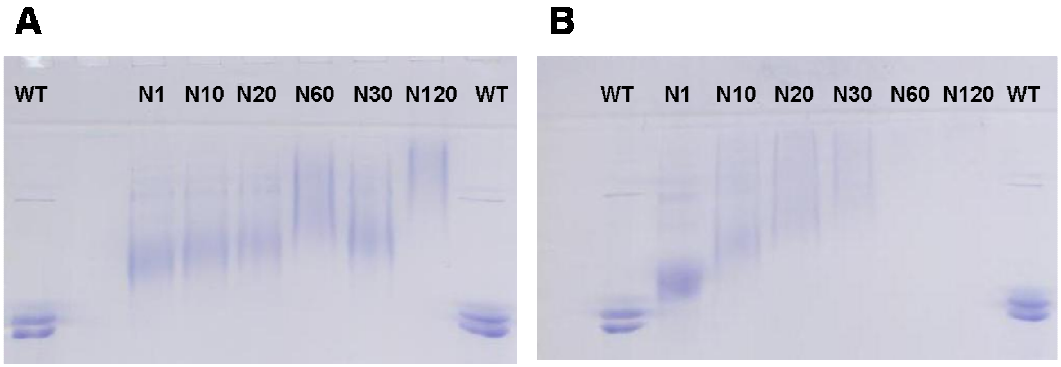


Figure 4-3. Analysis of chemically modified Ocr (varying pH reaction conditions) by 15% polyacrylamide gel electrophoresis under non-denaturing conditions.

Panels A and B show the migratory pattern of proteins modified under identical reaction conditions apart from pH (pH 6.5 and 5.3 for A and B, respectively). Reaction details are described in the main text. WT refers to wild-type unmodified Ocr, while N- refers to ammonium hydroxide used as nucleophile during the modification procedure. Numbers correspond to reaction time in minutes.

4.4.1.2. Determination of the optimum EDC concentration

In order to select the most suitable EDC concentration that would result in a gradual and stepwise loss of negative charge from the Ocr surface, a range of different reactions with varying concentrations of EDC were carried out. The reaction mixtures were composed of an aqueous solution (dH_2O) of 3 μM Ocr protein, 750 mM nucleophile (ammonium hydroxide or dimethylamine), 200 mM NaCl, pH 6.5 and EDC at either 3, 30, 60, 120 or 150 mM with an equivalent amount of HoBt (*i.e.*, 1:1 molar ratio of EDC:HoBt). The reactions were carried out as described in section 2.3.3. The samples were then analysed by polyacrylamide gel electrophoresis under non-denaturing conditions as described in section 2.3.1, (Figure 4-4) and by circular dichroism (Figure 4-5) to assess possible perturbations in protein secondary structure (section 2.4.1).

Far UV CD (Figure 4-5) indicated that sample D150/30 (the first number corresponds to the 150 mM of EDC used in the reaction; the second number refers to

the reaction time in minutes) represented the maximal extent of modification that could be achieved without causing a significant alteration to the secondary structure of the protein. Therefore, the migration of the D150/30 samples was chosen as a standard against which to compare all other modified samples. Thus, reaction conditions were chosen so as not to modify Ocr to a greater extent than sample D150/30 (refer to Figure 4-4, panel F, lane D30). Reaction conditions that gave rise to the migratory patterns on the non-denaturing gels annotated as 60 mM EDC (Figure 4-4, panels C and D) were chosen as ideal for further investigation.

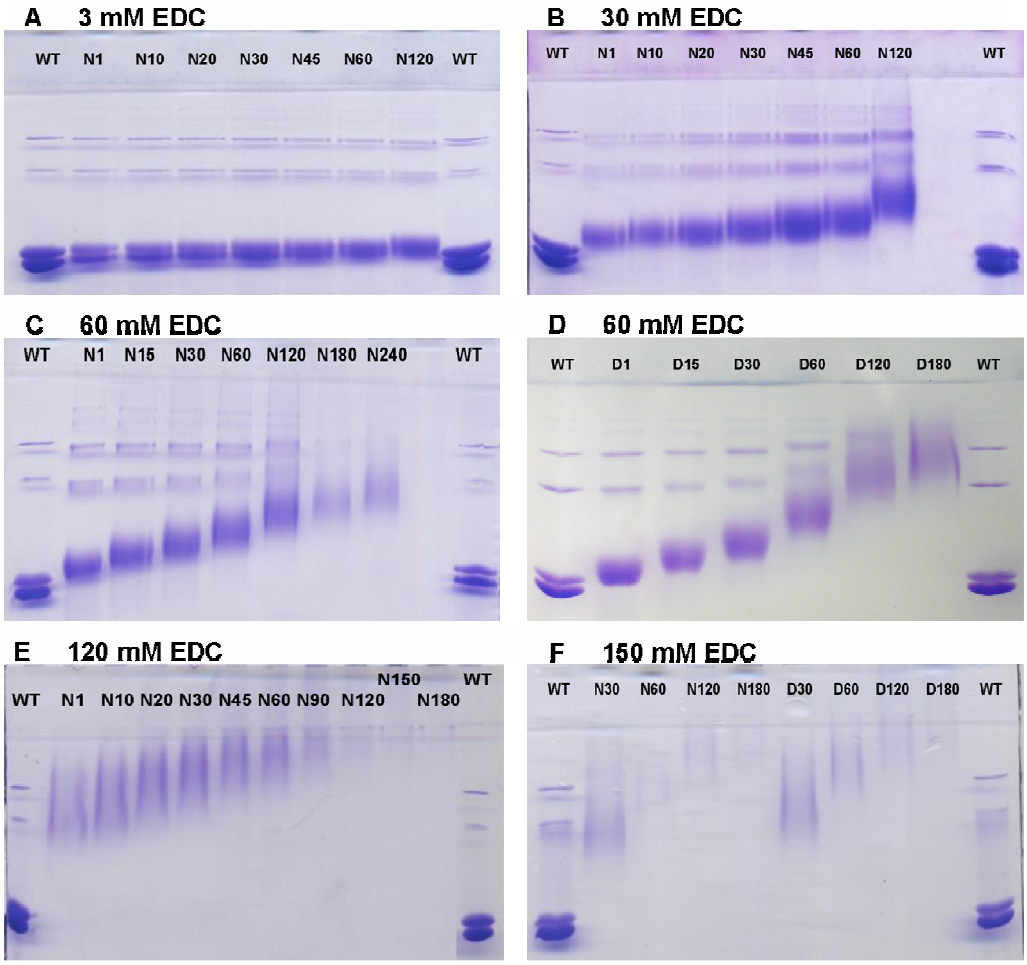


Figure 4-4. Analysis of chemically modified Ocr (varying EDC concentrations) by 15% polyacrylamide gel electrophoresis under non-denaturing conditions.

Reactions were conducted using various concentrations of EDC under otherwise identical conditions. The concentration of EDC is annotated above each gel. Reaction details are described in the main text. WT refers to wild-type unmodified Ocr while N- and D- refer to either ammonium hydroxide or dimethylamine used as nucleophile during the modification procedure. The numbers for each sample correspond to the reaction time in minutes. The gels are shown as a function of increasing EDC concentration and hence as a function of reaction rate and loss of negative charge from Ocr protein surfaces.

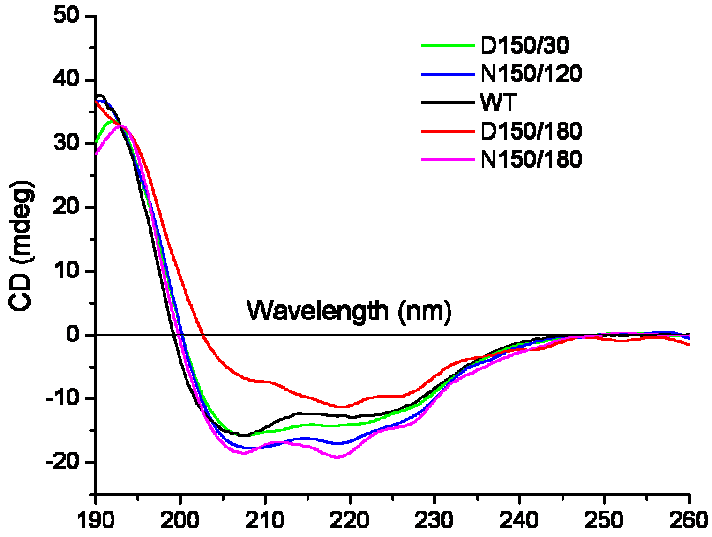


Figure 4-5. Far UV CD spectra of N- and D-series samples chemically modified in the presence of 150 mM EDC.

WT; wild-type unmodified Ocr. N and D correspond to ammonium hydroxide and dimethylamine used as nucleophile in the reaction, respectively. The first number of each sample corresponds to the 150 mM of EDC used in the reaction; the second number refers to the reaction time (min). Reaction details are described in the main text.

4.4.2. Biophysical characterisation of the chemically modified Ocr samples

4.4.2.1. Extent of modification

Panels C and D of Figure 4-4 are reprinted in Figure 4-6 for a more in depth discussion on the extent to which these samples have been modified. As has already been mentioned, the longer the modification reaction progressed, the slower the rate of migration through the gel towards the positive anode, indicating a gradational loss of negative charge. However, the modified samples gave protein bands that were more diffuse than the unmodified Ocr, which reflects the heterogeneous nature of the reaction products (Figure 4-6, panels A and B).

Gel electrophoresis of the unmodified Ocr, under non-denaturing conditions revealed four sharp bands (Figure 4-6, panels A and B; bands *a-d*), even though the same sample analysed by denaturing SDS-PAGE migrates as a single species (Figure 4-6, panel C). The unmodified Ocr was also analysed on a NativePAGE Bis-Tris gel

system (Invitrogen) based on the blue native polyacrylamide gel electrophoresis (BN-PAGE) technique originally developed by Schägger and von Jagow (1991). BN-PAGE confirmed that bands *c* and *d* correspond to Ocr dimer and tetramer, respectively (Figure 4-6, panel D). It is not known precisely why the Ocr monomer (bands *a* and *b*) runs as two distinct bands, but this may be the result of partial denaturation during the running of the gel, thereby causing the protein to adopt a different conformation.

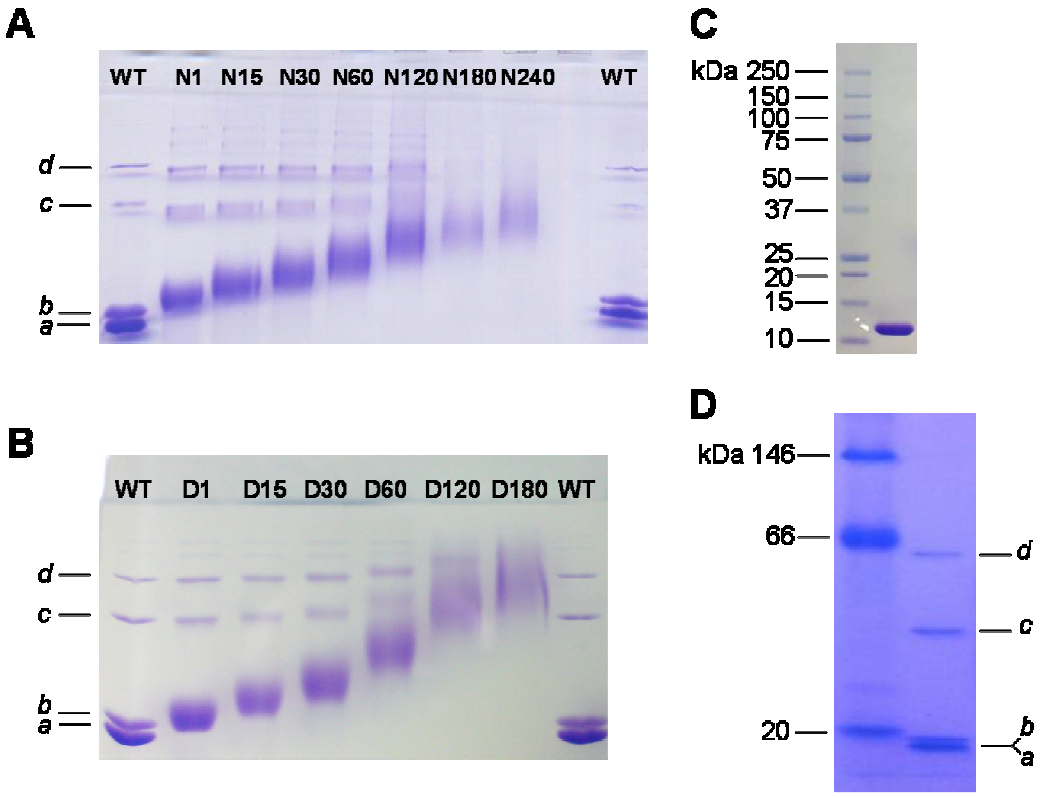


Figure 4-6. Gel electrophoretic analysis of chemically modified Ocr using either ammonium hydroxide (N-series) or dimethylamine (D-series) as nucleophile.

Panels A and B show 15% polyacrylamide gel electrophoresis under non-denaturing conditions of the N- and D-series of chemically modified Ocr samples, respectively. WT, refers to the wild-type unmodified Ocr. The name given to each modified protein sample reflects the reaction time in minutes. Panel C, SDS-PAGE, using a NuPAGE 4-12% Bis-Tris gel, of the unmodified Ocr sample (same sample as shown in lanes WT in panels A and B). Panel D, unmodified Ocr analysed by native-blue gel electrophoresis using the NativePAGE Novex® 4-16% Bis-Tris gel system. The protein markers were supplied by Invitrogen. The Ocr sample is resolved into four discrete bands labelled a-d.

The chemically modified proteins were also analysed by SDS-PAGE (section 2.3.1), which confirmed the presence of crosslinked dimers (Figure 4-7) most probably involving covalent bond formation between Lys75 and a carboxylate on either side of the dimer interface. The formation of multimers by crosslinking of independent dimers was minimised by conducting the modification reaction at low protein concentration (3 μM). For both the N- and D-series of chemical modifications the percentage crosslinked species increased with increasing reaction time. The main upper band, corresponding to dimer, is noticeably broader and more diffusive than the lower band, corresponding to monomer. Indeed, for the highly modified samples the upper band is resolved into two distinct species. This is probably caused by the incomplete unfolding of the more highly crosslinked dimers.

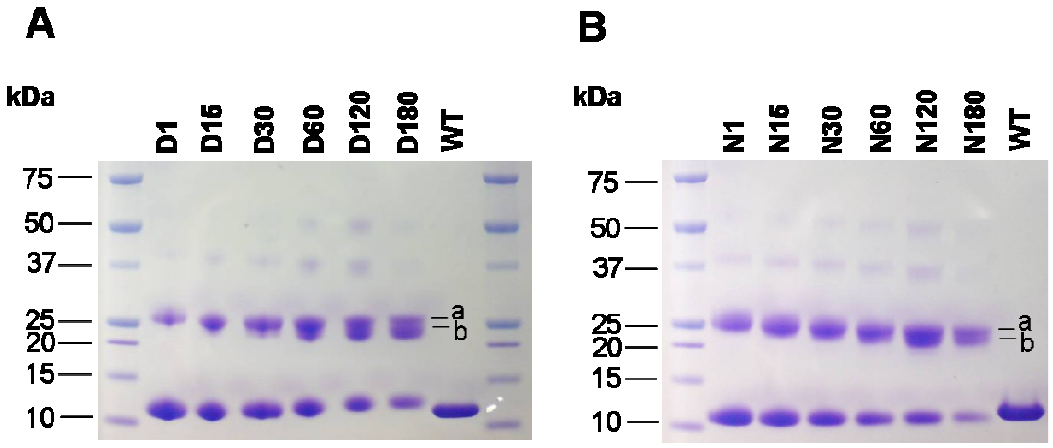


Figure 4-7. SDS-PAGE analysis of chemically modified Ocr using a 15% acrylamide gel.

Panel A, analysis of the Ocr samples using dimethylamine as nucleophile (D-series). The numbers for each sample correspond to the reaction time in minutes. Panel B, analysis of the Ocr samples using ammonium hydroxide as nucleophile (N-series). The numbers for each sample correspond to the reaction time in minutes. In both panels A and B the modified Ocr runs as two main bands: upper and lower band corresponding to Ocr dimer and monomer, respectively. Note that the upper band for the highly modified protein samples from the D- and N-series appear to resolve into two separate species (labelled *a* and *b*).

Additional confirmation of chemical modification and loss of charge on the Ocr dimer was obtained by anion exchange chromatography using a Mono Q column (section 2.4.2). Native Ocr binds very strongly to such media and requires high concentrations of NaCl to be eluted. In this experiment, Ocr eluted as a narrow peak

late in the salt gradient (Figure 4-8). The chemically modified samples eluted at progressively lower salt concentrations as the modification reaction time was increased. Moreover, the eluted material came off in a broad asymmetric peak indicating considerable heterogeneity in each sample due to different numbers of modified residues. As anticipated, the D-series of protein samples shows the most drastic change in elution time because dimethylamine, being a better nucleophile, generates a greater level of modification over ammonium hydroxide after an equivalent reaction time. Noteworthy is the asymmetric shape of the highly derivatised sample D180, the leading edge of which elutes from the column more abruptly than the other samples. The initial portion of the elution profile of D180 most likely represents an endpoint in the chemical modification reaction in which almost all available Glu and Asp residues have been amidated.

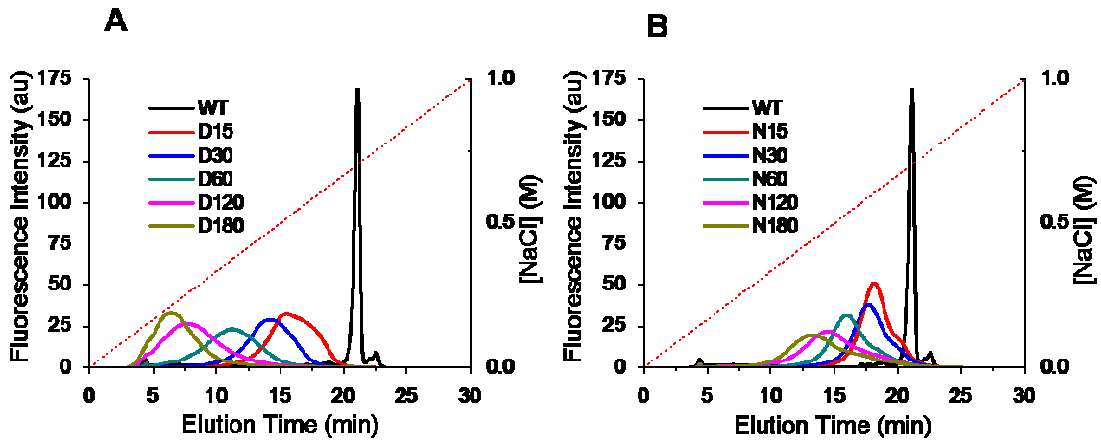


Figure 4-8. Anion exchange chromatography profiles of Ocr and its chemically modified derivatives.

Panels A and B show the analysis of the D- and N-series of chemically modified Ocr on a 1 mL MonoQ column, respectively. In each case, 11.2 µg of protein was loaded onto the column that had been equilibrated in 20 mM Tris HCl pH 8.0 at a flow rate of 1 ml/min. The bound protein was then eluted using a 0-1 M NaCl gradient over 30 min. WT, refers to unmodified Ocr.

Although the gel electrophoresis and ion exchange experiments confirmed that the protein was being derivatised, the number of modifications could not be quantified using these techniques. MS of the D-series of modified samples was performed to obtain this information, as each modification will increase the mass by

27 Da (Δmass +27 Da). Unfortunately, MS is unsuitable for analysis of the N-series because there is only a net reduction in mass of 1 Da after each successive modification. Given the chemical complexity of the samples, data resolution would be inadequate for meaningful analysis.

MALDI-TOF mass spectrometry of the D-series of modified samples was performed (section 2.4.3). Each spectrum showed a broad peak slightly skewed towards greater mass. Nevertheless, an estimate of the approximate minimum and maximum number of modifications for each sample could be determined as follows: D15, 9-18; D60, 16-25; D180, 22-33 (Figure 4-9). A smaller peak corresponding to Ocr dimer was also observed.

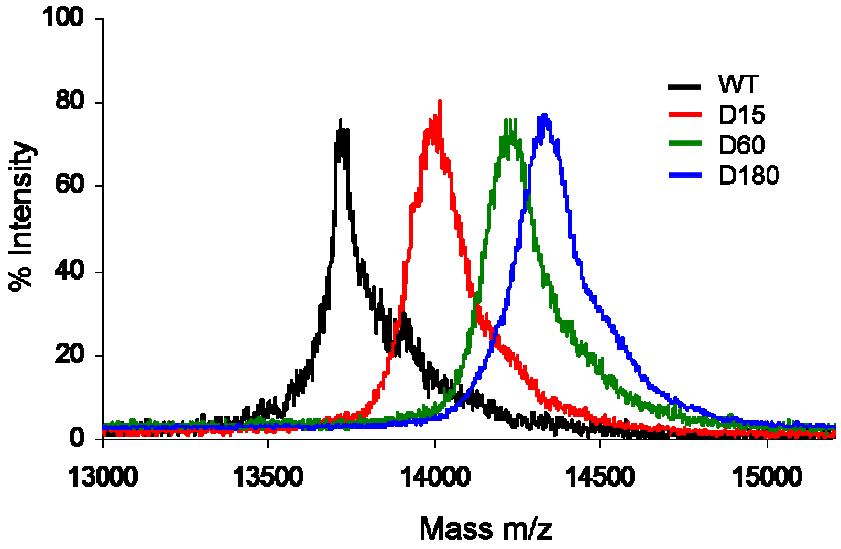


Figure 4-9. Mass specta by MALDI-TOF, of wild-type Ocr and chemically modified Ocr using dimethylamine as nucleophile.

Protein samples (~0.05 mg/ml) were diluted in 0.1% trifluoroacetic acid and mixed with an equal volume of matrix composed of a saturated solution of sinapinic acid in 50% acetonitrile, 0.1% trifluoroacetic acid on a stainless steel surface. The samples were air dried at room temperature prior to analysis. The machine was operated in positive ion mode. Peaks corresponding to the modified and unmodified Ocr samples are labeled.

Samples were also analysed on a FT-ICR instrument (by Dr. David Clarke) to obtain higher resolution (section 2.4.4). The proteins were desalted *via* reverse-phase HPLC and electrospray MS data were collected for the peak eluting from the HPLC column. A nested set of MS species were observed with a mass difference of 27 Da,

corresponding to the anticipated increase in mass after each modification with dimethylamine (Figure 4-10). From this data it was possible to ascertain the number of amidations per Ocr monomer as being: D15, 10-21; D60, 15-25; D180 22-31. Surprisingly, the nested set of peaks, corresponding to each successive modification, display a bimodal distribution that may arise from perturbations in the carboxylate pKa values (due to reduction of negatively charged residues on the protein surface), which will influence reaction rates with the doubly protonated EDC molecule as the modification reaction proceeds.

Whatever the underlying mechanism, for clarity and simplicity in data analysis, each sample is defined as being composed of a single number of modifications distributed randomly and uniformly on the protein surface. Despite the bimodal distribution of the nested set of peaks each sample is represented by a single average number of modifications. In a true statistical sense, taking the average of a bimodal distribution is not recommended. However, in this case I believe it to be an advantage in data analysis and interpretation. This choice in data analysis is further reinforced in the discussion section, when addressing the interactions of the Ocr samples with M.EcoKI.

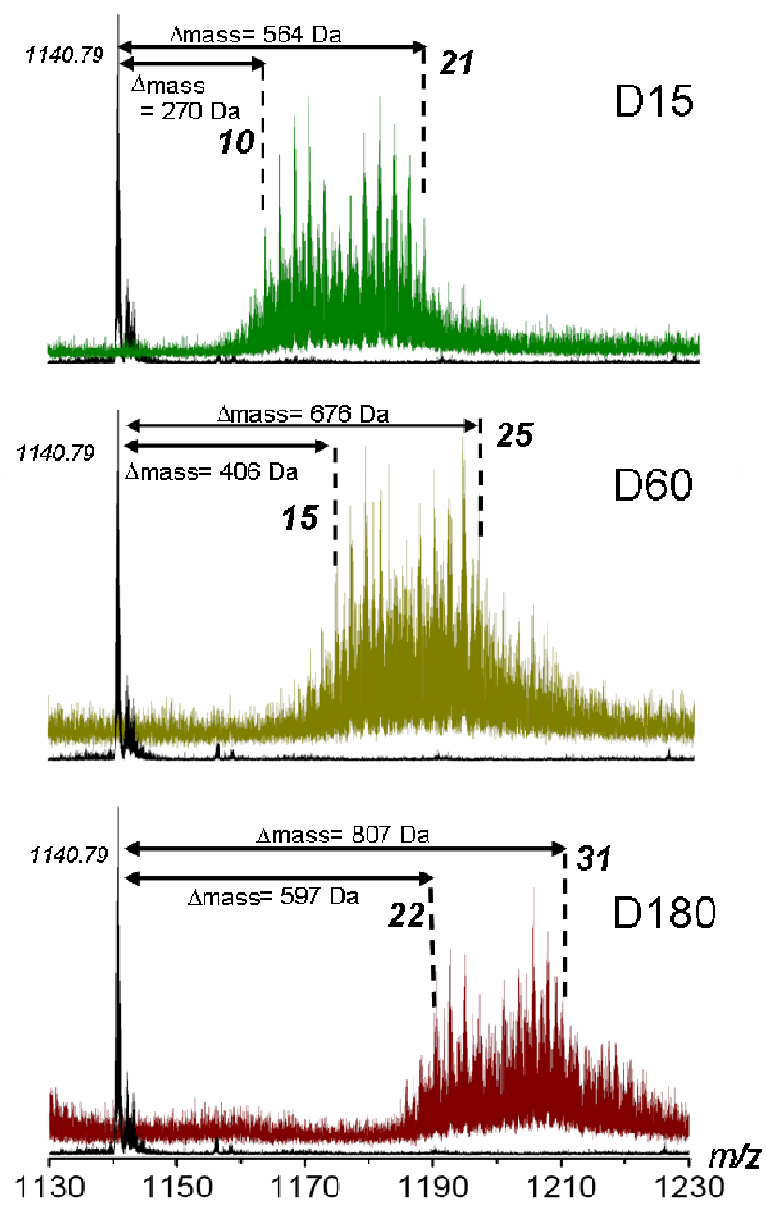


Figure 4-10. FT-ICR MS analysis of the D-series of chemically modified Ocr.

Upper spectra, Ocr before (black trace) and after (green trace) chemical modification with dimethylamine for 15 min as described in *Materials and Methods*. Middle spectra, Ocr before (black trace) and after (light green trace) chemical modification with dimethylamine for 60 min. Lower spectra, Ocr before (black trace) and after (red trace) chemical modification with dimethylamine for 180 min as described in *Materials and Methods*. In each case the $[M+12H]^{12+}$ charge state was analysed. The number in italics is the average m/z value for each species. The number in bold signifies the number of Δ+27 Da modifications for each species.

4.4.2.2. Protein folding and stability

Extensive modification of the 35 carboxyl groups (34 acidic residues plus the C-terminus) on each Ocr monomer could induce an alteration in the folding of the protein, potentially leading to a loss of activity through a gross structural defect rather than a more subtle change in the charge distribution of the protein. Chemical modification of Ocr did result in some loss of protein due to precipitation during sample preparation. However, upon removal of the aggregates by centrifugation, the final protein samples displayed no increased propensity to denature and precipitate, indicating that the structural integrity of the chemically modified Ocr was not compromised.

Chemically modified samples were subjected to circular dichroism (CD) in the far UV region, which detects the presence of protein secondary structure (section 2.4.1). The structure of Ocr is dominated by α-helices as reflected from the CD spectra, which displays two distinct minima at 208 nm (part of the π to π* transition) and 222 nm (*n* to π* transition). CD spectra of the D-series of modified Ocr samples showed no major alteration in secondary structure content, even though they were derivatised to a greater extent and to an unnatural amino acid (as opposed to the N-series samples) (Figure 4-11). CD analysis was also performed on certain N-series samples, but no major alteration in secondary structure was detected (Figure 4-11). Spectra were analysed using the Dichroweb online secondary structure deconvolution programme (Whitmore & Wallace, 2004). Analysis using the CDSSTR method gave an α-helical content of 60% for the unmodified Ocr, which is in good agreement with the crystal structure. Similar values were obtained for the chemically modified samples. It was therefore concluded that the chemically modified proteins retained the overall fold of native Ocr.

The stability of the chemically modified Ocr in comparison with that of the native protein was also investigated. Initially, the GdmCl-induced unfolding of the protein (section 2.4.5) was analysed. Addition of denaturant caused the protein to unfold resulting in a quenching of the fluorescent signal from the single tryptophan, W94. The fluorescent intensity was measured as a ratio of the fluorescent signal at

350 nm *versus* 380 nm to remove any variation in intensity due to slight differences in protein concentration between samples (Dryden *et al*., 1993). For unmodified Ocr the data indicated a two-state transition with no intermediate steps with a ΔG of unfolding of 20 kcal mol^{-1} and a midpoint of 3.85 M GdmCl, which is in good agreement with previously published data (Atanasiu *et al*., 2001). However, the unfolding of the modified Ocr samples could not be fitted to a two-state equation. Rather, the midpoint of unfolding was calculated by fitting the data to a sigmoidal curve (Boltzmann Equation: y = A2 + (A1-A2)/(1 + exp((x-x0)/dx)), Origin Software *version 6.1*) (Figure 4-12). The unfolding midpoint was defined as the transition midpoint of the fitted line, which was calculated by trigonometric means. Surprisingly, the data suggested that unfolding occurred at significantly higher concentrations of GdmCl for the N- and D-series samples (D15, 4.25 M GdmCl; D60, 4.75 M GdmCl; N60, 4.88 M GdmCl; c.f. unmodified Ocr, 3.85 M GdmCl) (Figure 4-13; panel B). It became apparent that Ocr stability was increasing as a function of reaction time of the carboxylate modification procedure.

The thermal stability of the N60 and D15 samples were analysed by DSC (section 2.4.8). Refer to Appendix B for the raw DSC data. The data shows a clear energy uptake during the transition consistent with a cooperative endothermic unfolding event. The T_m of D15 (T_m = 68.2°C) was similar to that of the unmodified Ocr (T_m = 69.0°C). However, the T_m of N60 (T_m = 73.7°C) is slightly elevated relative to the native protein. Noteworthy, however, was the observation that the unfolding of the modified Ocr samples (N60 and D15) was irreversible, which contrasts with unmodified Ocr where the unfolding transition is completely reversible. This irreversible unfolding is presumably due to the formation of covalent intra- and intermolecular crosslinks. Furthermore, a reduction in the net charge/electrostatic repulsions in the native fold could enhance the tendency for aggregation/irreversibility in the unfolded state. After normalisation for concentration and baseline correction, the data were fitted to a simple non-2-state model that afforded two independent estimates of the unfolding enthalpy. For the unmodified Ocr the estimate for the calorimetric enthalpy (ΔH_{cal} = 199 ± 3 kcal mol^{-1}) was somewhat greater than that for the van't Hoff enthalpy (ΔH_{VH} = 123 ± 3 kcal mol^{-1}) indicating that the unfolding transition may involve several intermediates

where the dimer dissociates to folded monomers prior to denaturation of the monomers (Atanasiu *et al*., 2001). The thermodynamic data for the unfolding transition of the modified Ocr samples (N60 and D15), which resulted in a relatively broad transition, could not be fitted to a non-2-state reversible model. This is presumably due to the chemical crosslinking and heterogeneity of the modified samples.

In relation to wild-type Ocr, the discrepancy in the models used to fit the chemical and thermal denaturation data (two-state *versus* a non-2-state reversible model, respectively), indicates that the single Trp residue on the Ocr molecule is not sensitive to changes and dissociation equilibria at the dimer interface (Atanasiu *et al*., 2001).

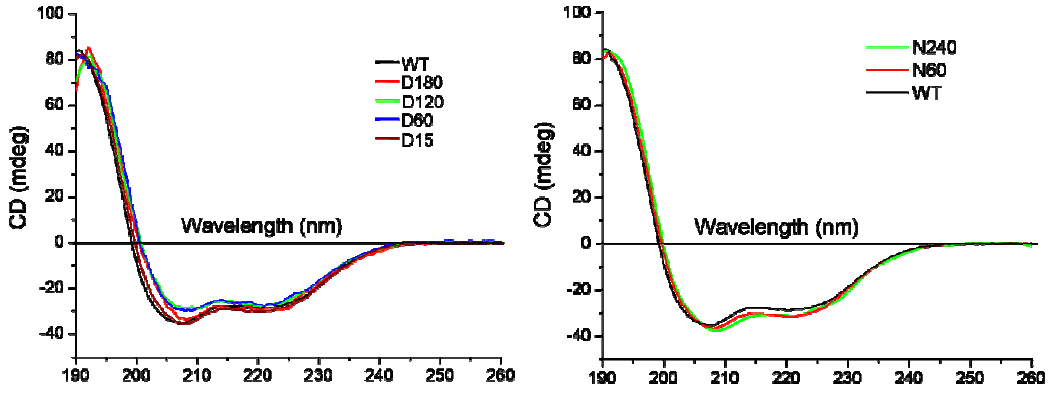


Figure 4-11. Far UV circular dichroism (CD) of chemically modified N- and D-series Ocr samples.

Protein samples were prepared in 10 mM Tris-HCl pH 8.0, 50 mM NaF. Spectra (190–260 nm) were obtained at a protein concentration of 30 μM using a 0.2 mm pathlength cell. Each spectrum was an accumulation of four individual scans.

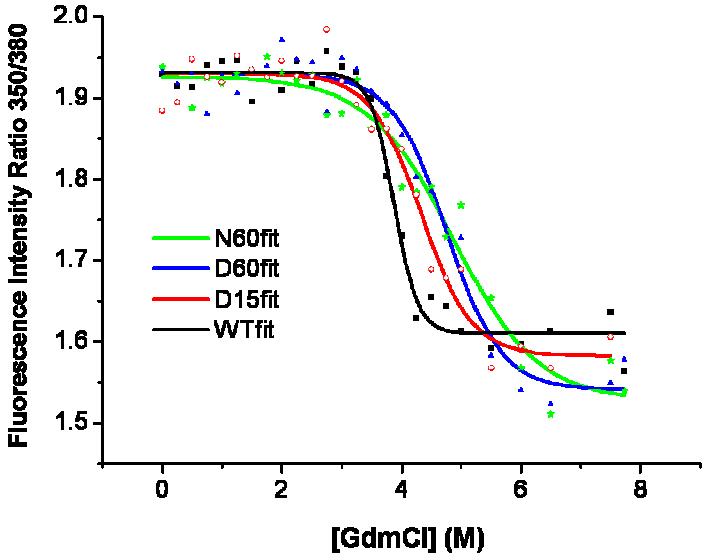


Figure 4-12. Guanidinium hydrochloride (GdmCl)-induced denaturation of Ocr protein samples.

Unmodified Ocr (black boxes), Ocr samples D15 (red circles), D60 (blue triangles) and N60 (green stars) as a function of GdmCl concentration, observed as the fluorescence intensity ratio of 350/380 nm; excited at 280 nm. All data sets were fitted to the Boltzmann Equation: y = A2 + (A1-A2)/(1 + exp((x-x0)/dx)) from which the mid-point of denaturation was determined.

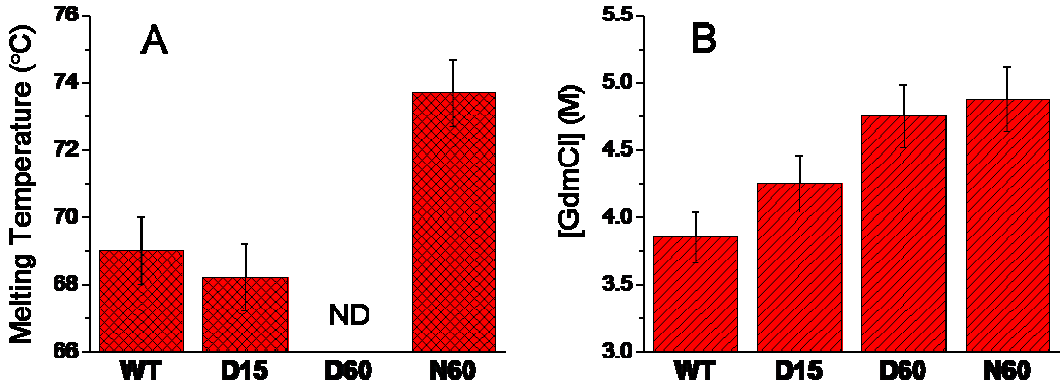


Figure 4-13. Bar charts describing thermal and chemical stability of selected D- and N-series samples in relation to wild-type unmodified Ocr.

Panel A; melting temperatures as determined by DSC. Panel B; midpoints of unfolding by GdmCl. ND; Not Determined. WT; wild-type Ocr.

4.4.3. Interactions of modified Ocr with M.EcoKI and EcoKI

Having determined that the modified Ocr samples were correctly folded and the extent to which they had been modified, their ability to function as DNA mimics was assessed by measuring their interaction with M.EcoKI and EcoKI. Specifically, the following experiments were performed: (i) Isothermal titration calorimetry (ITC) of Ocr with M.EcoKI (section 2.4.7); (ii) competition between Ocr and DNA for binding to M.EcoKI (section 2.4.6) and (iii) inhibition of the nuclease activity of EcoKI by Ocr (section 2.4.9).

4.4.3.1. Isothermal titration calorimetry of Ocr with M.EcoKI

The ITC experiment was initially performed using unmodified Ocr in 20 mM Tris-HCl pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol, 100 µM SAM with or without 500 mM NaCl at a range of different temperatures from 10-30°C (work done by Dr. Gareth Roberts). The experiment at 25°C was also repeated using 20 mM HEPES buffer (heat of ionization = 4.9 kcal mol^{-1}) in place of Tris-HCl (heat of ionization = 11.34 kcal mol^{-1}) in low salt conditions, which gave a very similar enthalpy of interaction (-21.5 kcal mol^{-1} in HEPES compared to -20.6 kcal mol^{-1} in Tris). These experiments indicated that there were no major contributions to ΔH due to effects of buffer ionization arising from protonation changes during binding. The enthalpy change (ΔH) upon interaction was strongly exothermic in the low ionic strength buffer (*i.e.*, no NaCl). Furthermore, upon formation of the M.EcoKI:Ocr complex, the ΔH showed significant temperature dependence, which is characteristic of a heat capacity change ΔCp, This was quantified from the slope of the plot of enthalpy change *versus* temperature, assumed linear, using the standard thermodynamic relationship:

$$\Delta Cp = d\Delta H/dT$$

The ΔCp of unmodified Ocr in low salt buffer was determined to be -1009 ±89 cal mol^{-1} K^{-1} (Figure 4-14). In 500 mM NaCl the transition was less sharp (*i.e.*, sigmoidal) and endothermic at all temperatures and the slope of the linear regression yielded a binding heat capacity of -495 ±89 cal mol^{-1} K^{-1}.

ITC was also carried out on a representative sample from the N- and D-series of chemically modified Ocr in 20 mM Tris-HCl pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol, 100 µM SAM at a range of different temperatures from 10-30°C. Specifically, the N60 and D15 samples were chosen because they elute from the Mono Q column at a similar point in the salt gradient, suggesting the Ocr is modified to a comparable extent. The calorimetric data shows that the interaction between chemically modified Ocr and M.EcoKI was essentially stoichiometric for both N60 and D15, as is the case for the unmodified Ocr (*i.e.*, 1 Ocr dimer per molecule of M.EcoKI). Intriguingly, our results demonstrate that the ΔH values fall between the behaviour of the native Ocr in low ionic strength buffer and high ionic strength buffer (Figure 4-14). Furthermore, the ΔCp values for N60 (-331 ±129 cal mol^{-1} K^{-1}) and D15 (-349 ±92 cal mol^{-1} K^{-1}) are close to the value determined for unmodified Ocr under high salt conditions.

The thermodynamic parameters (as determined by ITC) for the M.EcoKI:Ocr dimer interaction under various conditions are summarised in Table 4-1.

Table 4-1. Thermodynamic parameters of the M.EcoKI:Ocr dimer interaction as determined by ITC.

All experiments were conducted as described in section 2.4.7. The units of ΔG, ΔH, and TΔS are kcal mol^{-1}. TΔS was calculated as ΔH-ΔG. N, stoichiometry of the interaction (Ocr dimer/M.EcoKI). WT, wild-type (unmodified) Ocr. ND, not determined values due to either very sharp or non existent (ΔH= 0) transitions of the raw ITC data.

	WT (no salt)				WT (0.5 M NaCl)				N60 (no salt)				D15 (no salt)			
Temp (K)	ΔG	ΔH	TΔS	N	ΔG	ΔH	TΔS	N	ΔG	ΔH	TΔS	N	ΔG	ΔH	TΔS	N
283	-9.3	-5.9	+3.4	0.67	-9.2	+15.1	+24.3	1.09	ND	0	ND	ND	-9.0	+6.7	+15.7	1.24
288	-9.7	-12.5	-2.8	0.80	-10.3	+10.3	+20.6	1.11	ND	0	ND	ND	-10.5	+5.0	+15.5	1.07
293	ND	-19.1	ND	0.89	-10.1	+8.7	+18.8	1.09	-9.2	-4.7	+4.5	1.05	-12.1	+1.1	+13.2	0.99
298	ND	-20.6	-6.5	0.90	-12.0	+5.2	+17.2	1.06	-9.3	-6.5	+2.8	1.21	ND	0	+11.7	ND
303	ND	-26.6	ND	0.88	-9.9	+5.3	+15.2	0.99	-10.2	-8.5	+1.7	1.11	ND	0	ND	ND

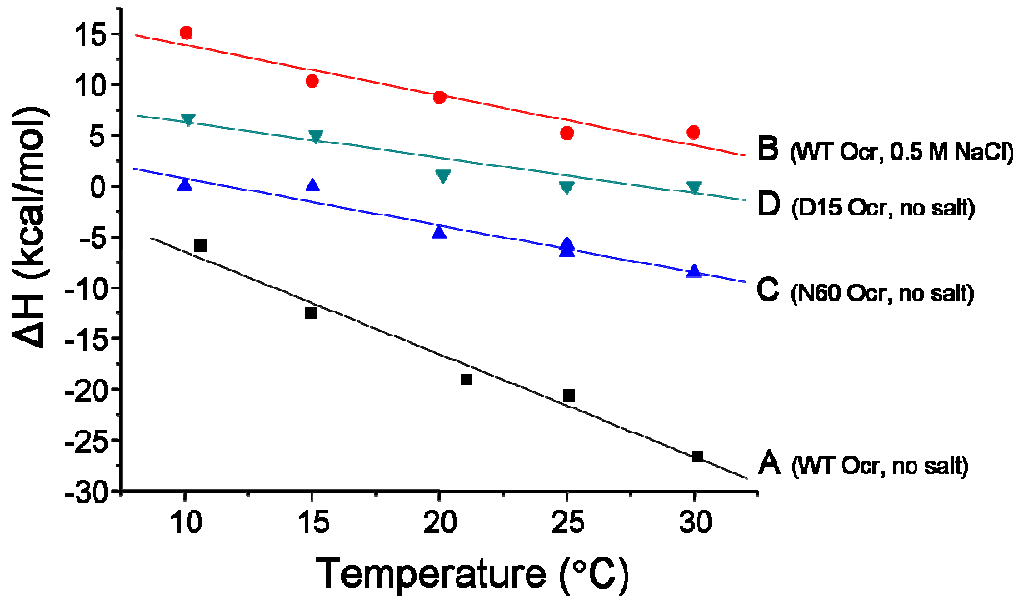


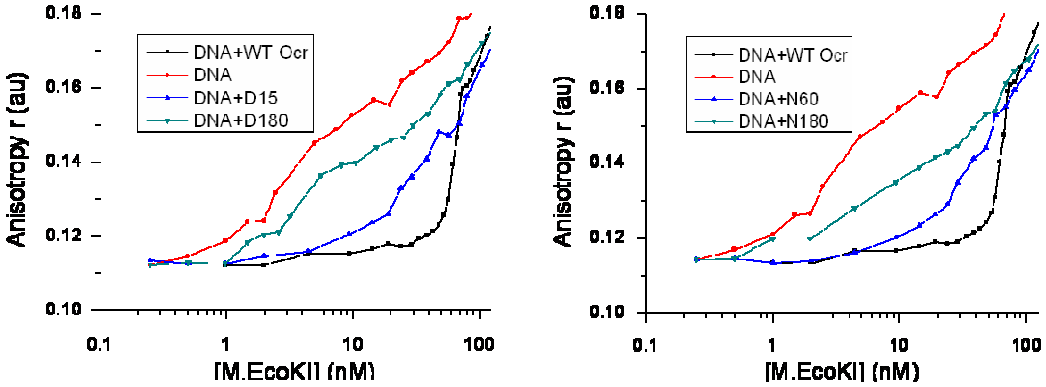
Figure 4-14. Plot of ΔH (kcal/mol) *versus* temperature (°C) for the interaction of either unmodified Ocr or chemically modified Ocr with M.EcoKI.

The ΔH (kcal/mol) of interaction was determined from ITC experiments that were performed at five different temperatures ranging from 10 to 30°C. All experiments were conducted as described in section 2.4.7. Plot A and B, unmodified Ocr in the absence or presence of 500 mM NaCl, respectively. Plot C and D, show the analysis of the interaction between M.EcoKI and N60 or D15 chemically modified samples, respectively.

4.4.3.2. Competition binding assay using fluorescence anisotropy

The interaction between modified Ocr and M.EcoKI was also studied using a sensitive fluorescence anisotropy assay as described in section 2.4.6. Some examples of the raw data showing the effect on anisotropy as the amount of M.EcoKI increases during the titration is shown in Figure 4-15, panel A. The results tabulated in Figure 1-15, panel B, show that the modified Ocr samples were less able to interfere with DNA binding by M.EcoKI than the native Ocr protein. Furthermore, the amount of modification correlated with a loss in binding. Specifically, the most extensively modified Ocr samples showed relatively weaker binding to M.EcoKI.

A



B

D series	K_d (nM)
D15	2.64 ± 0.52
D30	5.8 ± 1.4
D60	19.7 ± 7.3
D120	23.1 ± 8.1
D180	26.6 ± 9.2

N series	K_d (nM)
N15	0.56 ± 0.13
N30	1.82 ± 0.37
N60	1.9 ± 0.2
N120	3.77 ± 0.82
N180	12.6 ± 3.0

Figure 4-15. Competition binding assay for the M.EcoKI:Ocr dimer interaction using fluorescence anisotropy.

Panel A. M.EcoKI was titrated into a solution containing 50 nM Ocr dimer and 2 nM of a 21 bp DNA duplex fluorescently labelled at the 5' end. The binding of M.EcoKI to the DNA duplex could be monitored by anisotropy (r); a measurement of the rate of tumbling of the fluorophore. Upon binding of M.EcoKI to DNA the DNA-protein complex tumbles more slowly and r increases. The graphs show representative raw data that has not been fitted to a model but rather points have been joined by a line. Panel B. K_d values for N- and D-series samples binding M.EcoKI as determined by fluorescence anisotropy (c.f. K_d for wild-type Ocr ~44 pM; K_d for DNA 2.1±0.35 nM). Experimental details and data analysis are described in section 2.4.6. The Kd errors represent the goodness of the fit of the averaged data of n=2 or n=3. An example of data fitting to a binding model is shown in Appendix B.

4.4.3.3. ***In vitro* endonuclease inhibition assay**

The activity of the modified Ocr samples were tested in an endonuclease assay using purified EcoKI. Linearisation of a circular unmethylated plasmid (pBRsk1) containing a unique EcoKI target recognition site was monitored in the absence and presence of Ocr (unmodified or N- or D-modified samples). In each case the reaction mixture minus DNA was prepared and the digestion initiated by addition

of pBRsk1. The reaction was stopped after 10 min and the mixtures were then analysed by agarose gel electrophoresis (Figure 4-16). Incubation of pBRsk1 in the presence of a six-fold excess of EcoKI resulted in complete digestion of the plasmid within 10 min (compare lanes 1 and 3). The experiment was also performed in the presence of a ten-fold excess of Ocr dimer over EcoKI (lanes 4-10). As anticipated, unmodified Ocr completely abolished linearisation of the plasmid DNA (lane 4). However, the modified Ocr samples became progressively poorer endonuclease inhibitors as the extent of modification increased. For example, D15 and D60 display partial inhibitory activity but D120 shows no evident nuclease inhibition (lanes 5, 6 and 7, respectively). A similar trend is also found for the N-series of modified Ocr (N15, N60 and N120), although here the inhibitory activity is greater than for the corresponding samples from the D-series (*e.g.*, compare lanes 5 and 8). This finding is consistent with our earlier biophysical data, which suggests that Ocr is more highly modified by dimethylamine than by ammonium hydroxide after an equivalent incubation period.

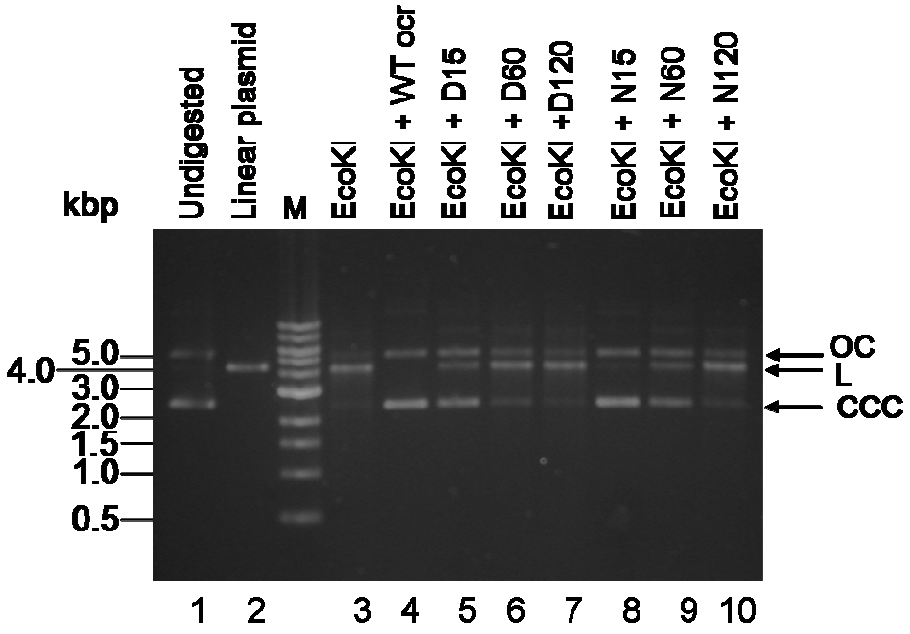


Figure 4-16. *In vitro* inhibition of EcoKI using chemically modified Ocr.

An unmethylated plasmid (pBRsk1) with a single EcoKI recognition site was used as substrate. Samples were analysed using a 0.9% agarose gel. Numbers at the bottom of gel denote lane numbers: Lane 1, undigested pBRsk1; lane 2, pBRsk1 linearised with EcoRI to act as size marker; lane 3, pBRsk1 linearised with EcoKI; lane 4, reaction performed using EcoKI that was pre-incubated with a 10-fold molar excess of unmodified Ocr dimer; lanes 5-10, as for lane 4 except using D15, D60, D120, N15, N60 or N120 in place of the unmodified Ocr, respectively. M, 1 kbp ladder from New England Biolabs.

4.5. Discussion

The DNA mimicry displayed by Ocr comprises two main features; its mimicry of the bent DNA substrate, preferred by EcoKI, and its mimicry of the electrostatics of the phosphate backbone. Other features, such as H-bonding and van der Waals interactions, will also play a role in the binding of Ocr to its target enzyme. However, in the absence of a detailed structure for an Ocr:M.EcoKI complex, these intermolecular forces are not easily defined. Chapter 4 focuses on the electrostatic component of the Ocr:M.EcoKI interaction, which is an important feature of the DNA mimicry.

Chemical modification offers a convenient method of reducing the number of negatively charged groups (*i.e.*, Asp and Glu side chains) on the Ocr surface. The results clearly show the stepwise reduction of negative charge with reaction time. MS FTIR of the D-series of chemically modified samples showed that the protein was subject to an average of 15.5, 20 and 26.5 modifications per Ocr monomer for the D15, D60 and D180 samples, respectively. CD analysis of the various modified Ocr samples confirmed no major changes in the secondary structure of the protein after modification. Such extensive modification without destruction of the protein fold is noteworthy.

The stability of the modified Ocr samples was also tested and compared to that of the native protein. It was initially surprising to find that chemical modification increased the stability of the protein to either chemically or thermally induced denaturation and that the enhancement in the stability of the protein unfolding increases with the extent of chemical modification. This is attributable to two factors. Firstly, the adventitious formation of inter- (*i.e.*, between monomers of an Ocr dimer) and intramolecular crosslinks (*i.e.*, within each Ocr monomer) possibly increasing with reaction time and therefore stabilising the fold (although it will also make it recalcitrant to refolding upon denaturation). Secondly, the closeness of the carboxylates in the unmodified Ocr (less than the Bjerrum length, Bloomfield *et al.*, 2000) leads to electrostatic repulsion forces. Chemical modification increases the

distance between the unmodified carboxylates, reducing electrostatic repulsion and enhancing stability.

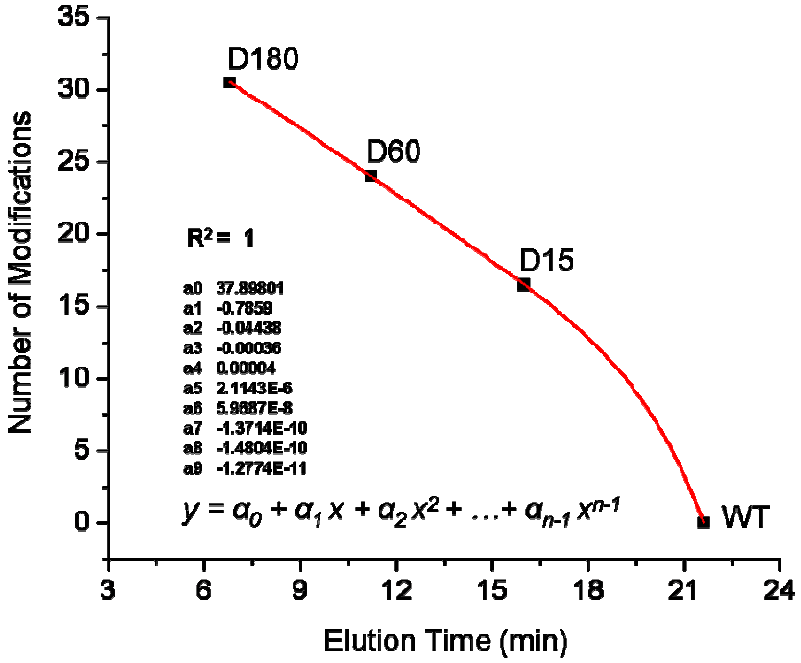
Nevertheless, despite confirming the existence of inter- and intramolecular crosslinks the choice of reaction conditions limits the number of unwanted covalent modifications. This is exemplified by analysis of the D180 sample where the reaction appears to be approaching completion as shown by the anion exchange chromatography, which displays a pronounced asymmetric elution profile, and MS, which gives a maximum pronounced peak showing ~31 modifications (implying predominantly ~4 acidic residues per monomer that may be involved in crosslinking). However, some minor peaks of species with up to ~34 modifications, can also be seen. Indeed, examination of the X-ray crystal structure of Ocr (pdb 1s7z) indicates that eight amino acids (α-amino group of the N-terminus, Lys19, Lys75, Ser54, Ser58, Ser68, Thr30, Thr74) may act as nucleophiles and could therefore be involved in crosslinking with neighbouring activated carboxyl groups. The estimated average number of four intramolecular crosslinks is therefore a good representation of an otherwise chemically complex and heterogeneous sample. It is possible to envisage that this number of ~4 intra- and/or intermolecular crosslinks may also hold for samples modified over shorter reaction periods. This has implications for the mass spectrometric analysis of the D60 sample where the true extent of the modification is masked. It could, therefore, be conjectured that the average number of modifications for D60 is ~24, rather than ~20 as indicated from the nested set of species seen in the mass spectrum. Likewise, the average modification of D180 is likely to be ~30.5, rather than ~26.5. However, it is more difficult to estimate the average number of crosslinks for the D15 sample. The SDS-PAGE gel of Figure 4-7, panel A, suggests that a proportion of the population (less than 50%) has at least one cross-link at the dimer interface. Furthermore, the DSC data demonstrates that the D15 sample was unable to refold after complete denaturation. These data reaffirm that most of the Ocr proteins constituting the heterogeneous D15 sample have at least one crosslink. Whilst bearing this in mind a comparison of the melting temperatures of D15 and N60 (Figure 4-13, panel A), where N60 has been defined as containing an average ~4 intra-crosslinks, suggests that the latter sample contains more crosslinks than the D15 sample. Thus, for

simplicity we define the D15 sample as containing an average number of ~1 crosslinks and hence an average number of ~16.5 modifications rather than ~15.5.

If we assume a direct correlation between elution time (from anion exchange chromatography) and the number of negative charges on the protein surface, it is then possible to determine the average extent of modifications for any of the N- and D-series samples. A graph was plotted of elution time (from the MonoQ column) *versus* the average number of modifications (derived from MS data for D15, D60, D180 plus unmodified Ocr). This data was then fitted to a polynomial equation (Figure 4-17, panel A). The elution time for each sample (defined as the time at which 50% of the protein has eluted from the column) was calculated as described in section 2.4.2. From these values (of elution time) it was then possible to extrapolate the average number of modifications of each sample by using the calibration curve to directly establish the corresponding *y*-axis value (*i.e.*, number of modifications). The data derived from this analysis is tabulated in Figure 4-17, panel B.

An extensive programme of mutational analysis that targeted one or two acidic residues per Ocr monomer showed that such changes do not significantly perturb the interaction with M.EcoKI *in vitro* or EcoKI nuclease activity *in vivo* (Stephanou *et al*., 2009; Chapter 3). Furthermore, removal of the C-terminal 17 amino acids of Ocr, which included 10 acidic residues, did not significantly affect binding to M.EcoKI (Dunn *et al*., 1981). Thus, it appears that not all Asp or Glu residues of Ocr contribute equally (if at all) to the interaction with M.EcoKI and that certain regions of the protein are relatively unimportant in terms of binding affinity. Hence, positional context as well as the number of negatively charged residues is critical for the interaction with M.EcoKI.

A



B

Sample	Number of Modifications	Standard Deviation
D180 *	30.5	2.9
D120	28.8	3
D60 *	24	3
D30	19	3
D15 *	16.5	3.45
N180	19.6	3.45
N120	17.9	3.3
N60	15.7	3
N30	12.5	3
N15	11.6	3
WT *	0	0

Figure 4-17. Estimating the average number of modifications per Ocr monomer, based on elution time from anion exchange chromatography.

Panel A; Calibration curve for estimating the average number of modifications. Elution times were calculated as described in section 2.4.2. The equation represents the most appropriate fitting of four samples with a defined number of modifications. The polynomial equation chosen reflects the non-linearity of the initial modification rate from wild-type (WT) to sample D15, as observed from the elution profiles of the anion exchange column (Figure 4-8). Panel B; Tabulated data as extrapolated from panel A, above. Standard deviations were calculated from the MS data (Figure 4-10) or estimated as appropriate by comparing the broadness of elution peaks from the anion exchange column (Figure 4-8). Samples annotated with an asterisk signify that the data shown was experimentally determined (*i.e.*, MS data or as known for WT Ocr).

Whilst bearing this caveat in mind, we can consider the relationship between the number of modifications and the binding affinity for M.EcoKI. First however, the following four simplifying assumptions are made that will assist in the data analysis of the M.EcoKI:Ocr interactions.

1. *The Ocr carboxyl groups are distributed evenly and uniformly on the entire protein surface.*

By observing the electrostatic representation of Ocr (Figure 4-18), it can be observed that the acidic residues are very nearly distributed equally and uniformly on the protein surface.

2. *The modification procedure is a random process, with all carboxylates being equally prone to modification*

All 70 carboxylates of the Ocr dimer may act as potential targets for modification since they are surface exposed and easily accessible by the small EDC and nucleophilic (amine) molecules. It is likely that the rates of modification of each carboxylate vary slightly due the complexity of the chemistry and the different microenvironments on the protein surface, which will also change as the reaction proceeds. Nevertheless, on the time scales (minutes to hours) and hence resolution of this experiment, all carboxylates are deemed as being equally prone to modification in a completely random fashion.

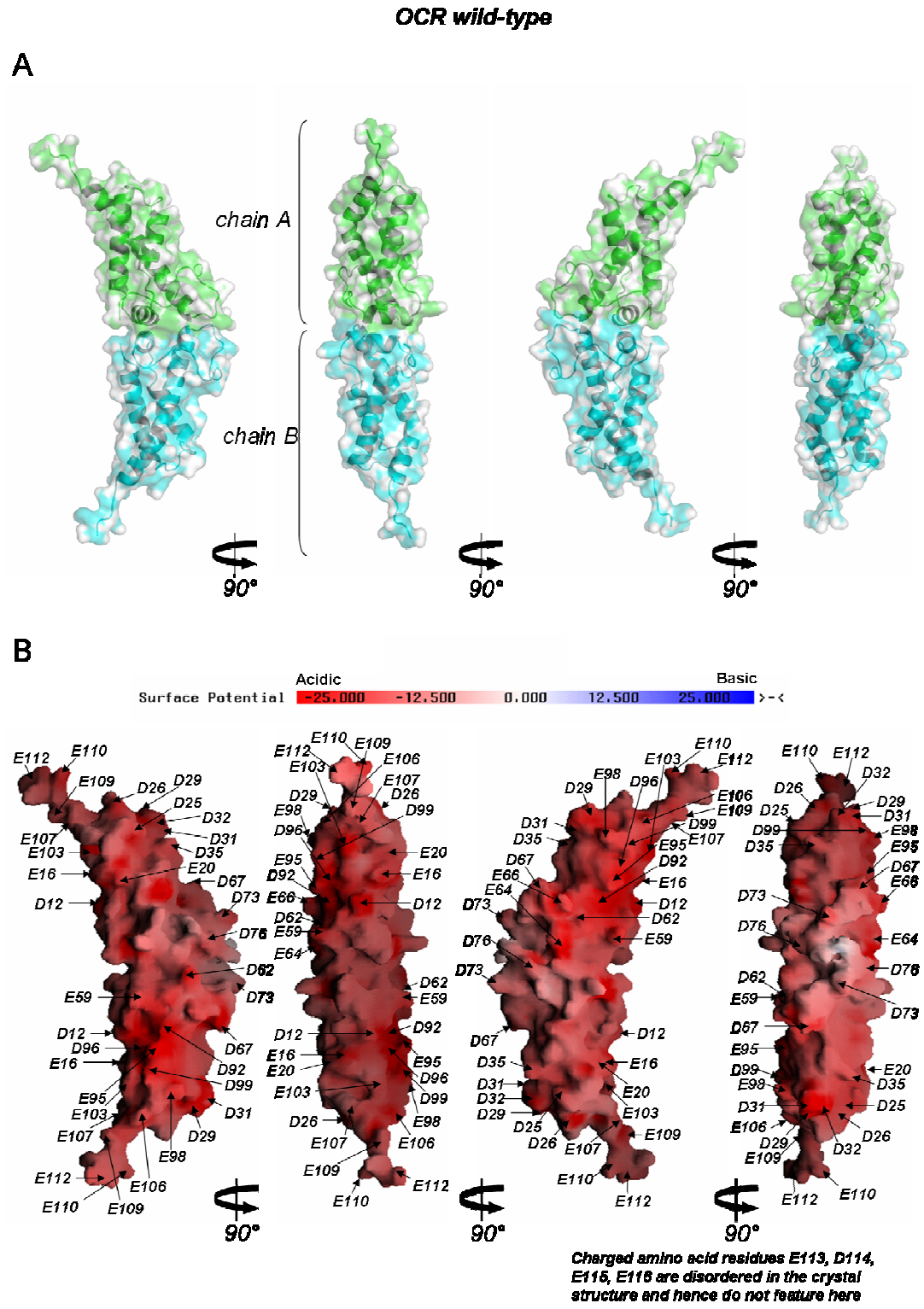


Figure 4-18. The electrostatic representation for wild-type Ocr.

Panel A: A cartoon representation of the Ocr molecule showing 4 views rotated by 90 degrees about the *y*-axis along with a semi-transparent surface for comparison purposes with panel B. Panel B: Four views of the protein, rotated by 90 degrees about the *y*-axis. All Asp and Glu residues in each view are highlighted.

3. *The Ocr modified samples are a homogeneous mixture composed of molecules modified to the same degree and equal to the extrapolated average number* (Figure 4-17, panel B).

It has been shown that each N- and D-series sample is a heterogeneous mixture containing a range of different modifications. For example the D60 sample has been shown to contain 15-25 modifications. Each modification, however, is not equally represented within the sample population; some modifications being present at much higher numbers than others. This is exemplified by the elution profiles of Figure 4-8, which qualitatively describe the distribution (*i.e.*, the relative amounts or population density) of each modification within a sample. Because the elution profiles from the anion exchange chromatography (Figure 4-8) display either a normal or only a very slight skewed distribution, the peaks of these profiles also correspond to the calculated average number of modifications. It is assumed that the Kd measured by ITC or fluorescence anisotropy reflects that of the most prevalent molecular species within the heterogeneous population of modified Ocr. Consequently, whenever relating the modified samples to their binding affinities with M.EcoKI, we can consider these samples as being homogeneous and containing a defined number of modifications equal to the calculated average number (Figure 4-17, panel B).

4. *The modifications on the Ocr protein are distributed uniformly and the probability that any specific carboxyl group has been modified is equal to the ratio of the average number of modifications versus the total number of carboxylates.*

Due to the randomness of the chemical modification process, any protein within a specific sample will contain a different combination of carboxylates modified, resulting in a heterogeneous sample with a whole spectrum of modification combinations. For example in the D60 sample different proteins will contain different combinations of 24 carboxyl groups modified out of the possible 35 targets. This spectrum of combinations can be envisaged as having two extremes; from the very clustered to those evenly and uniformly spaced out on the Ocr surface. Assumptions 2 and 3 require that this spectrum of modification topology for each N- and D-series sample is described by a normal distribution. For simplicity however

these samples are defined as being composed of all possible modification combinations belonging to the "evenly and uniformly distributed" extreme of the spectrum. In this sense the samples can be considered as heterogeneous. Accordingly, in any given sample, the collection of all modifications from all protein surfaces will show that all carboxylates are targeted for modification. It follows that on any one protein surface the probability of a specific carboxyl group being modified is equal to the average number of modifications over the total number of carboxylates on the protein surface.

The combined use of assumptions 1 and 4, allows for correlations between number of modifications, binding affinities and Ocr protein surface areas.

A plot of K_d (for the M.EcoKI: Ocr dimer interaction) *versus* average number of modifications (per Ocr monomer) is shown in Figure 4-19, panel A. When approximately 16.5 acidic residues in each Ocr subunit (*i.e.*, equivalent to sample D15) have been removed the protein displays the same binding affinity as DNA for M.EcoKI (*i.e.*, ~50-fold loss in binding affinity compared to unmodified Ocr). It is noteworthy that an additional loss of eight acidic residues (*e.g.*, D60) leads to a sudden decrease in binding affinity (*i.e.*, ~500-fold loss in binding affinity relative to unmodified Ocr). Any further modification appears to have no significant effect on binding. The evident sigmoidal nature of this plot suggests that the overall stability of the M.EcoKI:Ocr complex is not equal to the sum of the individual contributions of each interacting residue. This may imply that certain residues, directly involved in the binding mechanism with M.EcoKI, are clustered together to form part of an interacting region(s) (or hot spots) on the Ocr surface (Keskin *et al.*, 2005). On the other hand, it may be envisaged that the mid point of the sigmoidal transition (of the Kd plot) represents a critical number of carboxylates involved in the binding of the S subunit TRDs (this suggestion follows from the assumption that the S subunit TRDs contribute most of the interacting interface with Ocr).

The free energy of the interaction was calculated from the K_d values. A plot of the free energy of interaction *versus* number of modifications appears to show a linear relationship until one reaches ~23 modified residues per monomer (Figure 4-19, panel B). Any further modification has very little effect on the free energy of

interaction with M.EcoKI. This transition point is very similar to the mid-point of the sigmoidal transition in panel A.

These numbers, however, should be treated with caution since they reflect the relative amount of "random modification cycles" required to target the most important residues involved in the interaction (keeping in mind the relatively unimportant contribution to binding affinity by the 10 acidic residues at the C-termini). Therefore, these results are an overestimate of the total number of negatively charged residues that are critically involved in the interaction between Ocr and M.EcoKI. A more targeted approach, using site-directed mutagenesis, is likely to require fewer modifications in order to generate an equivalent loss in binding affinity.

The "critical" number of 23 residues represents a loss of about 66% of carboxylates per monomer. This number would also include 66% of the 10 acidic residues on the C-terminus that are known not to contribute to the binding affinity. By subtracting this percentage of residues, results in an upper limit of about 16 negative residues per monomer that seem to be critically important for the binding affinity of the Ocr:M.EcoKI interaction.

By taking into account assumptions 1 - 4 described previously, it follows that the upper limit of 16 acidic residues, corresponding to ~ 46% of the Ocr surface, is directly involved in the interacting interface with M.EcoKI. Thus, up to 5170 $Å^2$ of the surface area per Ocr dimer and 2585 $Å^2$ per Ocr monomer may be directly involved in the interaction with M.EcoKI.

The Ocr:M.EcoKI interaction was also investigated by ITC. The thermodynamics of protein-protein interactions are typically made up of numerous small changes in free energy with both enthalpic and entropic components (Cooper *et al.*, 2001). One might anticipate a major entropic component to the interaction given that the interface between the two proteins is substantial with the M.EcoKI almost completely engulfing the Ocr protein (Atanasiu *et al.*, 2002), which has a Connolly surface of 11238 $Å^2$ for the dimer (Walkinshaw *et al.*, 2002).

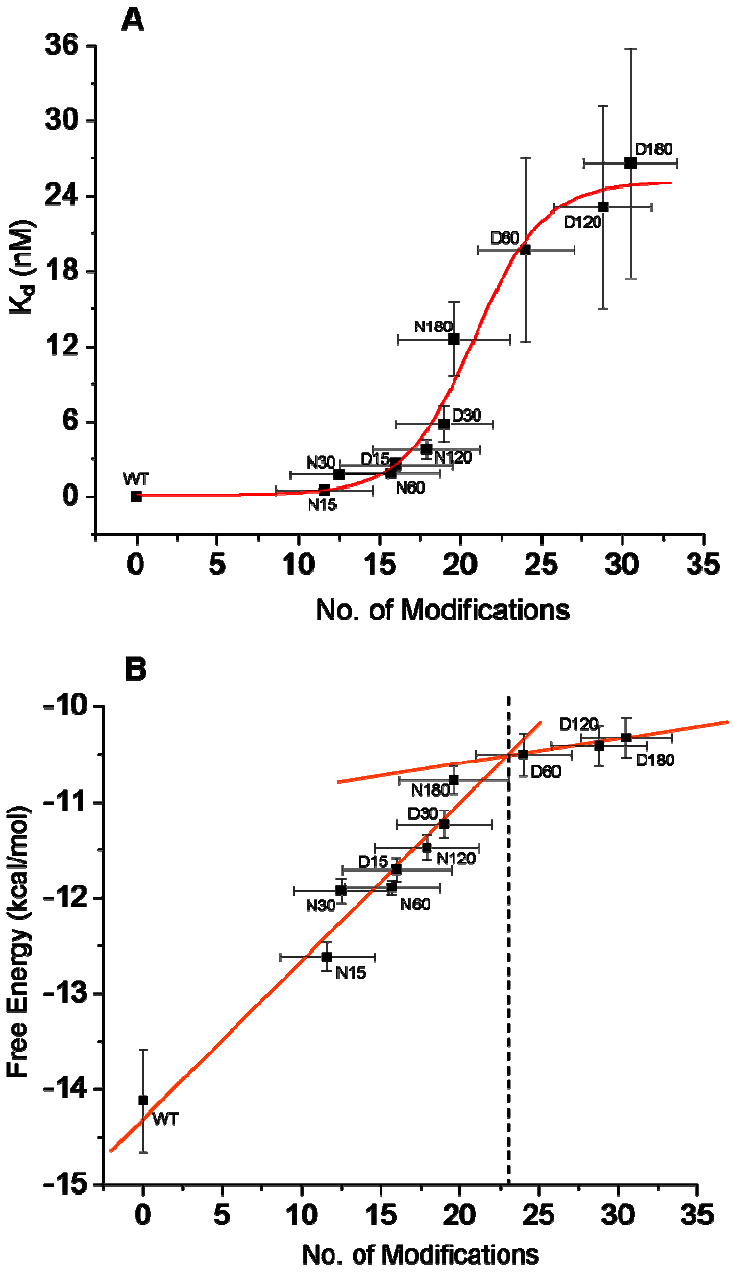


Figure 4-19. Analysis of the interaction affinity between M.EcoKI and chemically modified Ocr samples.

Panel A; plot of K_d (nM) *versus* number of acidic residues modified from the Ocr surface. The data was fitted to a sigmoidal plot using Origin software (*version 6.1*). Panel B; plot of free energy of interaction calculated from the Kd values (Figure 4-15, panel B) against number of modifications. The dotted line highlights the point at which any further modification to the Ocr molecule has no significant effect on binding. For both plots, the number of modifications was determined as described in the main text. The *y*-axis error bars were determined experimentally; *x*-axis error bars represent the standard deviation.

Indeed, a recent model for the Ocr:M.EcoKI interaction has been proposed that is consistent with this hypothesis (Kennaway *et al*., 2009). However, the precise surface area of interaction is still not known. Nevertheless, a large number of water molecules with all of their H-bonding potential are presumably displaced during the formation of the complex and the dynamics of residues on the interface must become rather constrained. This is particularly important under high salt conditions where the counterions decorating the surface of Ocr must be displaced in order to allow interaction with M.EcoKI. Indeed, the transition in the presence of salt is more endothermic than in the absence of salt at any given temperature (Figure 4-14).

Initially, the interaction of unmodified Ocr with M.EcoKI was studied in the absence and presence of monovalent salt. Bearing in mind the average ΔCp for protein-protein interactions is reported to be -333±202 cal/molK (Stites, 1997), the results show a surprisingly large ΔCp effect for the M.EcoKI:Ocr interaction that is strongly dependent on ionic strength. Similarly, the ΔCp of the D15 and N60 protein samples under low salt conditions, were found to be similar to the value determined for unmodified Ocr under high salt conditions.

Conventionally, large ΔCp effects are associated with hydrophobic interactions. However, theoretical considerations and empirical observations show that long range electrostatic interactions and other effects can also make a significant contribution to ΔCp (Gallagher and Sharp, 1998; Cooper, 2005). The results presented here clearly reinforce this notion. Furthermore, several algorithms have been devised that attempt to correlate the magnitude of ΔCp with changes in solvent accessible surface area of interaction (Gomez and Freire, 1995). Assuming that the structure of the Ocr:M.EcoKI interface is essentially unchanged by high ionic strength, or by chemical modification, then conventional theories would predict the same ΔCp. The results shown in Figure 4-14 clearly contradict this prediction.

An *in vitro* nuclease assay was also used to investigate the interaction of the N- and D-series of modified Ocr with EcoKI (Figure 4-16). The assay involved monitoring the linearisation of unmethylated plasmid DNA (pBRsk1) using purified EcoKI. As anticipated, pre-incubation of EcoKI with unmodified Ocr (ratio of EcoKI to Ocr of 1:10) resulted in complete inhibition of nuclease activity. By contrast,

incubation of EcoKI with chemically modified Ocr caused only partial inhibition. Moreover, the Ocr was less effective in reducing nuclease activity as the degree of chemical modification increased. Interestingly, D15 and N60, which are modified to a similar level, display comparable anti-restriction activities.

In order to address the relative contribution of the electrostatic mimicry *versus* the mimicry of the overall shape and bend of the DNA molecule upon binding of M.EcoKI, it is useful to consider the highly modified sample D180 in which almost all (~87%) the negatively charged residues have been removed. Hence, we can observe the relative contribution of the overall dimensions and shape of Ocr to the binding process in the absence of electrostatic mimicry. It is noteworthy that the sample still displays significant binding (nM range) presumably due to the favourable shape complementarity of the Ocr:M.EcoKI association involving a multiplicity of weak intermolecular interactions. Therefore, the major contributor to the biological activity of Ocr is its shape mimicry of a bent DNA molecule.

The results show that the electrostatic mimicry of Ocr, although not the major contributor to its inhibitory activity, increases its binding affinity for an archetypal Type I R/M enzyme by up to 800-fold (when comparing the Kd of wild-type Ocr ~ 0.044 nM and that of D180 ~ 35.8 nM which is the upper limit value). Indeed, the DNA mimicry displayed by Ocr is extremely robust and it is remarkable that Ocr can still act as an effective DNA mimic even after removal of ~47% of the negative charges (sample D15) on the surface of the protein.

In conclusion and with respect to the aims set out at the beginning of this chapter, the gradational removal of negative charge from the Ocr surface was achieved; the relative contribution of the electrostatic mimicry *versus* the mimicry of the overall dimensions of a bent DNA molecule was addressed; and an upper limit of about 16 acidic residues were identified as playing an important part in the binding affinity of Ocr to M.EcoKI.

Chapter 5. Investigating the DNA mimicry of the T7 Ocr protein by a multi-mutational approach

5.1. Introduction

The single or double mutational analysis (described in Chapter 3) failed to clarify the contribution of the carboxylate groups of Ocr to the tight binding observed for the Ocr:M.EcoKI complex. Furthermore, this approach did not provide an insight into how the Ocr molecule mimics the charge distribution of DNA. The results from Chapter 4 suggest that up to ~16 acidic residues per monomer play a crucial role in the binding with M.EcoKI. An alternative strategy is to target multiple amino acid residues corresponding to patches of negative charge on the surface of the protein. In this fashion, it was reasoned that the relative contribution and importance of specific regions of negative charge to the anti-restriction activity of Ocr might be deduced. Thus, the aim of the work described in this chapter was to understand the mechanism by which Ocr mimics the charge distribution of DNA.

5.2. Aims

The specific aims for this chapter were to:

- Create a series of Ocr mutants, in which the negative charges from certain areas or patches on the protein surface, are neutralised (*i.e.*, mutating aspartate to asparagine and glutamate to glutamine).

- Assess the *in vivo* and *in vitro* activity of these mutants against the model Type I R/M system, EcoKI.

- Delineate the relative contribution of groups of acidic residues to the biological activity of Ocr and hence deduce the mechanism by which Ocr mimics the charge distribution of DNA.

5.3. Overall strategy

A range of different Ocr mutants were designed and created as described in section 5.4.1. Ocr mutant proteins were expressed and purified using the protocol described in section 2.3.2.1. Secondary structure determination of proteins was carried out by circular dichroism (CD) (section 2.4.1). Binding of the Ocr proteins to M.EcoKI was assessed by fluorescence anisotropy (section 2.4.6) and by isothermal titration calorimetry (ITC) (section 2.4.7). The anti-restriction activity of the Ocr mutants was assessed by the *in vitro* nuclease assay (section 2.4.9). *In vivo* phage restriction assays (spot tests) were also carried out (section 2.4.10).

5.4. Results

5.4.1. Design and creation of Ocr mutants

The residues targeted for mutagenesis are colour-coded in Figure 5-1 and are also highlighted on the Ocr structure (Figures 5-2 to 5-8).

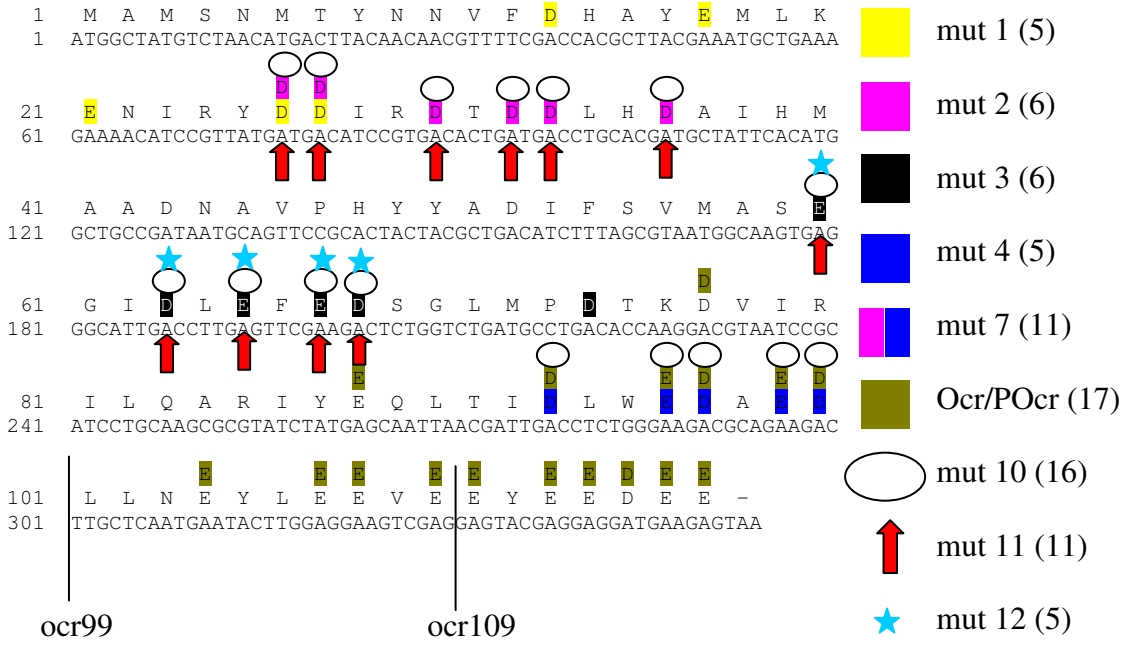


Figure 5-1. Colour coding of the amino acids targeted for mutagenesis for each Ocr mutant.
The protein and corresponding DNA sequence of wild-type Ocr (pdb 1S7Z) is shown. The panel on the right shows the colour coding used to highlight the amino acids targeted for mutagenesis; the numbers in parentheses correspond to the total number of amino acid substitutions for each mutant.

In all, nine multiple mutants of Ocr were created as shown in Figure 5-1. The approach was to mutate a contiguous series of acidic residues in the primary amino acid sequence of Ocr, which corresponded to a specific patch of negative charge on the protein surface. On occasion, it was necessary to perform two or three rounds of mutagenesis in order to introduce the desired number of amino acid substitutions. Sometimes however, existing mutated forms of Ocr that had been prepared for other projects were employed as template for the mutagenic PCR. In such instances (*i.e.,* mut2, mut3, mut7 mut10 and mut11), one of the acidic residues was substituted to Cys.

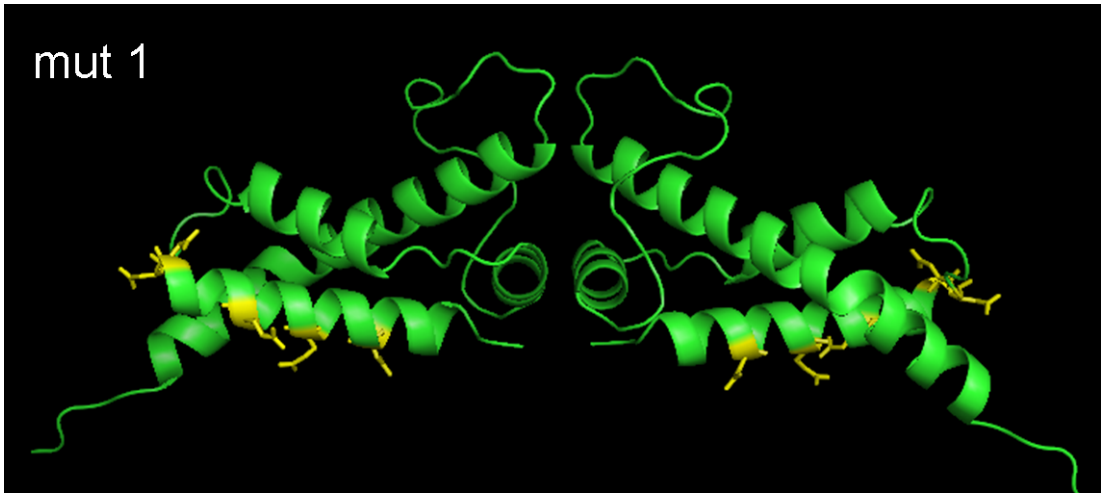


Figure 5-2. The Ocr mutant mut 1.
The mut 1 residues targeted for mutagenesis (D12, E16, E20, D25, D26) are highlighted as yellow sticks on the structure of the Ocr dimer.

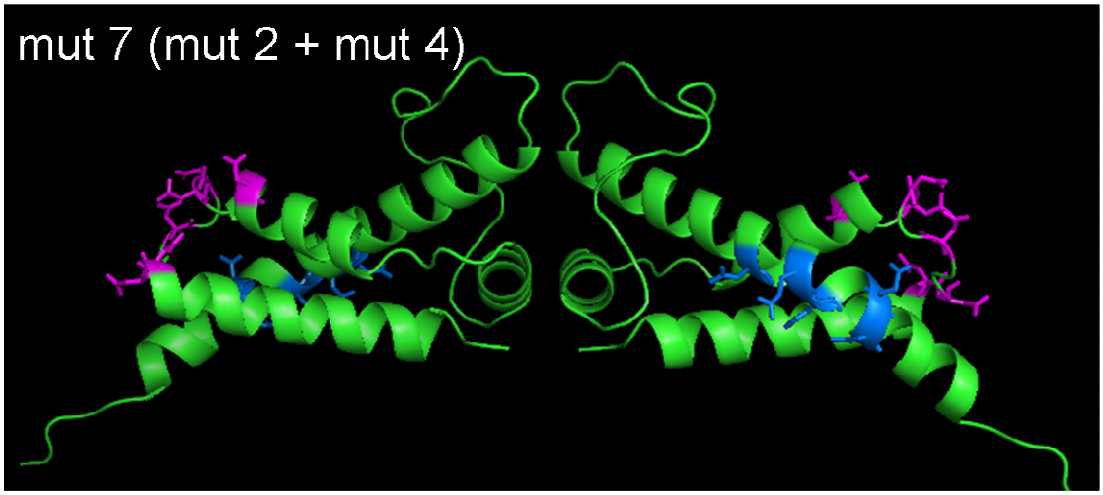


Figure 5-3. The Ocr mutants mut 2, mut 4 and mut 7.
The residues targeted for mutagenesis for mut 2 (magenta- D25, D26, D29, D31, D32, D35) and mut 4 (blue- D92, E95, D96, E98, D99) are highlighted on the structure of the Ocr dimer. Mut 7 incorporates mutations from both mut 2 and mut 4. For clarity, all three mutant proteins are shown in the same figure.

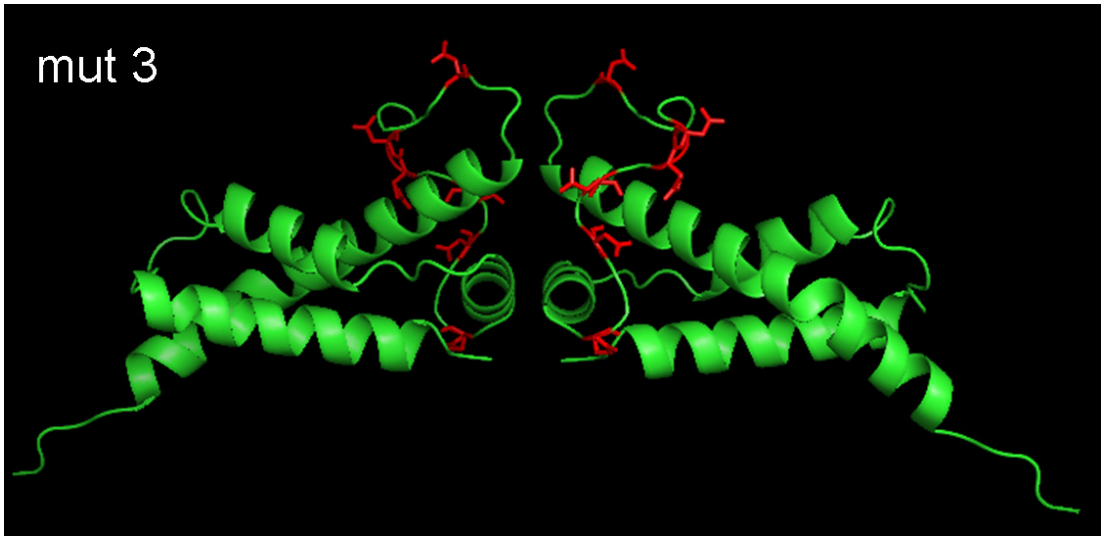


Figure 5-4. The Ocr mutant mut 3.
The mut 3 residues targeted for mutagenesis (E59, D62, E64, E66, D67, and D73) are shown as red sticks, which span the length of the loop at the interface of the Ocr dimer.

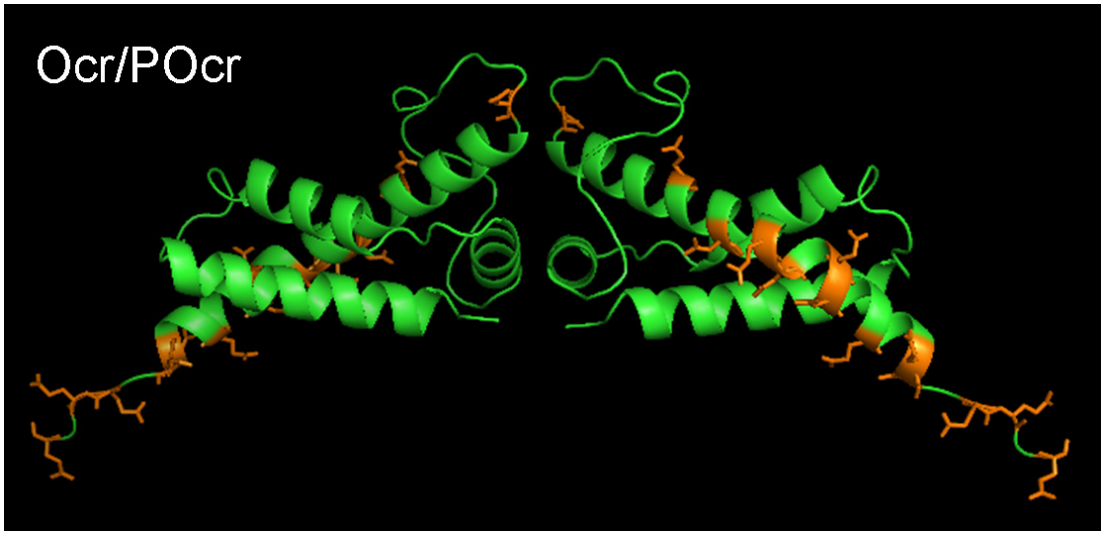


Figure 5-5. The Ocr mutant Ocr/POcr.
The orange sticks correspond to the negatively charged residues on Helix-D of Ocr that were targeted for mutagenesis (D76, E87, D92, E95, D96, E98, D99, E103, E106, E107, E109, E110, E112, E113, D114, E115, E116).

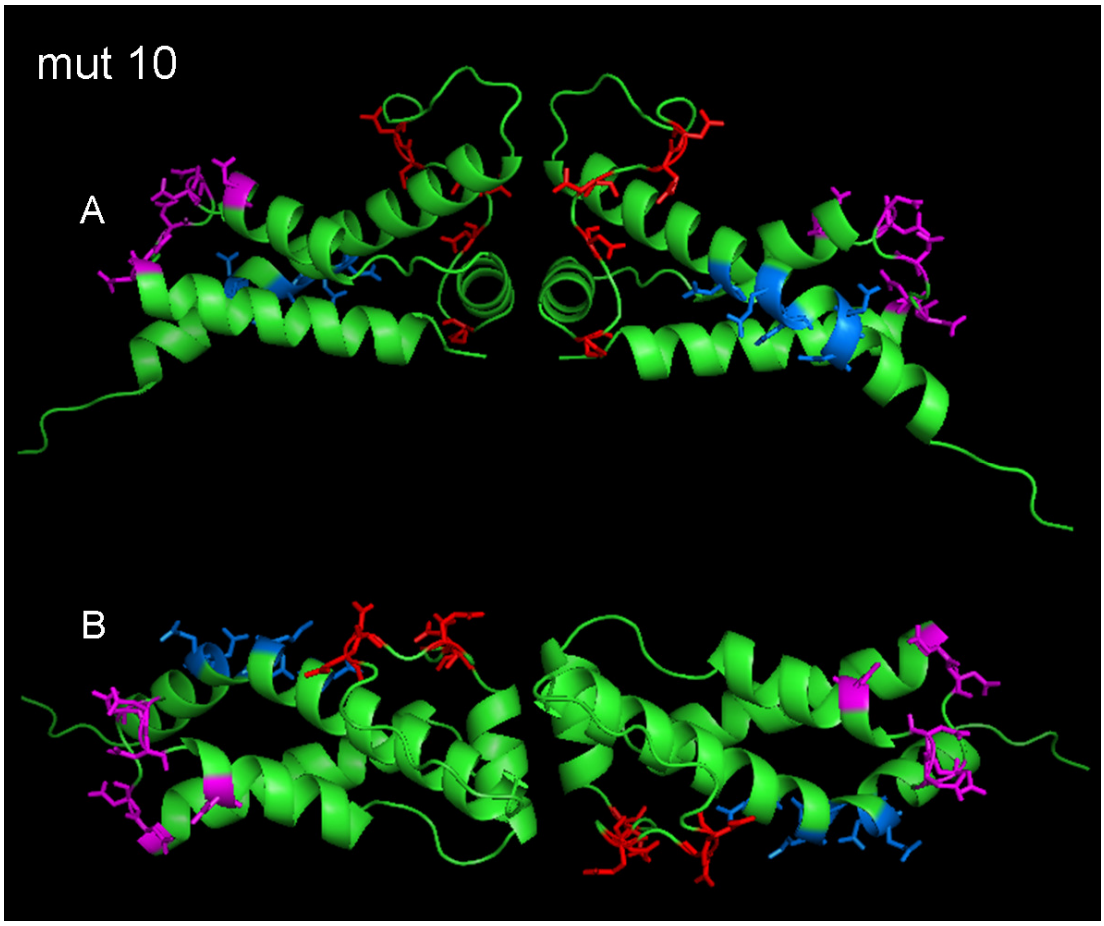


Figure 5-6. The Ocr mutant mut 10.
The residues targeted for mutagenesis for mut 10 (D25, D26, D29, D31, D32, D35N, D92, E95, D96, E98, D99, E59, D62, E64, E66, D67) are highlighted as coloured sticks. Note that mut 10 combines all the mutations from mut 2 (magenta), mut 3 (red - excluding D73) and mut 4 (blue). Panel B is a view of the molecule in panel A, rotated by 90°.

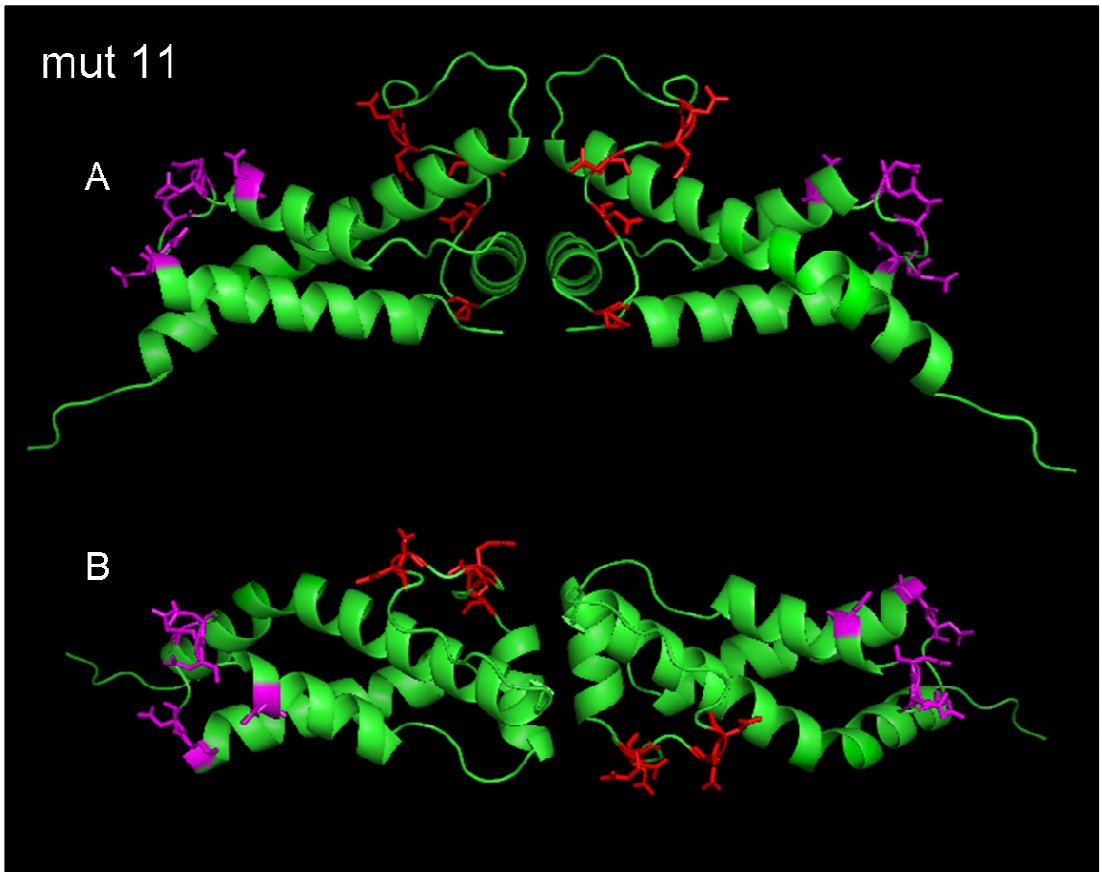


Figure 5-7. The Ocr mutant mut 11.
The residues targeted for mutagenesis for mut 11 (D25, D26, D29, D31, D32, D35, E59, D62, E64, E66, D67) are highlighted as coloured sticks. The mutations found in mut 11 combine those for mut 2 (magenta), mut 3 (red - excluding D73). This mutant was created in retrospect of the biophysical analysis of all the other Ocr mutants. Panel B is a view of the molecule in panel A, rotated by 90°.

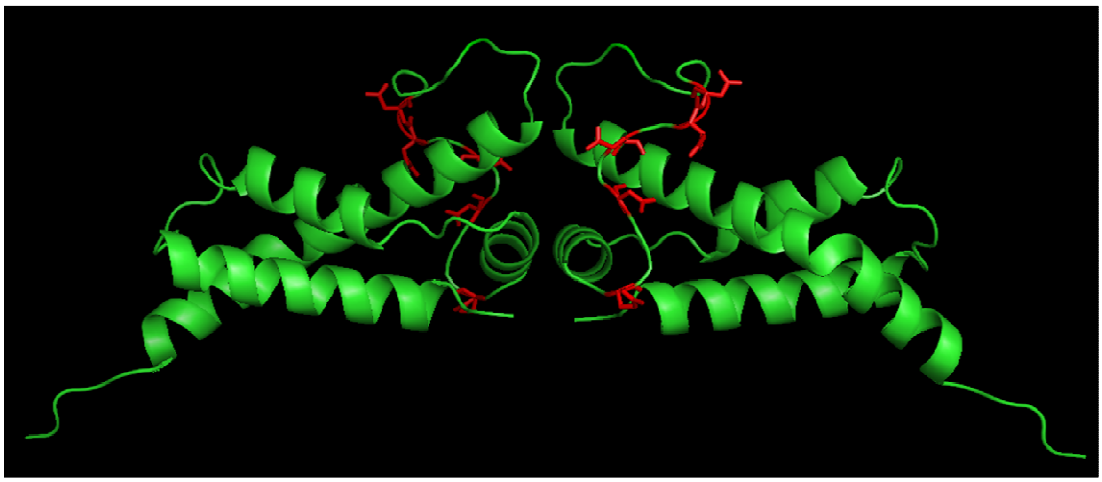


Figure 5-8 The Ocr mutant mut 12.
The mut 12 residues targeted for mutagenesis (E59, D62, E64, E66 and D67) are shown as red sticks. This mutant was created in retrospect of the *in silico* mutagenesis of mut 3 (Figure 5-16).

The Ocr mutants; mut 1, mut 2, mut 3, mut 4, mut 7, mut 10, mut 11 and mut 12 were created as described in section 2.2.2. The PCR cycle parameters are shown in Table 5-1.

Table 5-1 PCR cycle parameters for site directed mutagenesis

Segment	Cycles	Temperature	Time
1	1	95°C	30 seconds
2	12	95°C	30 seconds
		55°C	1 minute
		68°C	6:30 minutes

The mutagenic primers and template used for each Ocr mutant are shown below (letters in blue correspond to mutations already present on the constructs used as template. Letters in red correspond to mutations introduced during PCR):

mut 1 (giving rise to pOcrmut1)
(Using pOcrD12ND26N as template):
Forward primer: 5'-CCACGCTTACCAAATGCTGAAACAAAACATCCGTTATAATAACATCCG-3'
Reverse primer: 5'-CGGATGTTATTATAACGGATGTTTTGTTTCAGCATTTGGTAAGCGTGG-3'

mut 2 (giving rise to pOcrmut2)
(Using pOcrD25C DNA as template):
Forward primer: 5'-CCGTTATTGTAACATCCGTAACACTAATAACCTGCACAATGCTATTCA-3'
Reverse primer: 5'-TGAATAGCATTGTGCAGGTTATTAGTGTTACGGATGTTACAATAACGG-3'

mut 3 (giving rise to pOcrmut3)
(Using pOcrD73C DNA as template):
Forward primer: 5'-AATGGCAAGTCAGGGCATTAACCTTCAGTTCCAAAACTCTGGTCT-3'
Reverse primer: 5'-AGACCAGAGTTTTGGAACTGAAGGTTAATGCCCTGACTTGCCATT-3'

mut 4 (giving rise to pOcrmut4)
(Using pAR2993 (wild-type Ocr) as template):
Forward primer: 5'-GCAATTAACGATTAACCTCTGGCAAAACGCACAAAACTTGCTCAATGA-3'
Reverse primer: 5'-TCATTGAGCAAGTTTTGTGCGTTTTGCCAGAGGTTAATCGTTAATTGC-3'

mut 7 (giving rise to pOcrmut7)
(Using primers for mut 4 and pOcrmut2 as template)

mut 10 (giving rise to pOcrmut10)
(Using primers for mut 3 and pOcrmut7 as template)

mut 11 (giving rise to pOcrmut11)
(Using primers for mut 3 and pOcrmut2 as template)

mut 12 (giving rise to pOcrmut12)
(Using primers for mut 3 and pAR2993 as template)

The strategy adopted for the creation of pOcrpocr is outlined below. The aim was to remove all negative charges from Helix-D of the Ocr structure (*i.e.*, mutate aspartates and glutamates to asparagines and glutamines, respectively) (Figure 5-5). The overall strategy (Figure 5-9) was to use PCR to amplify and fuse together two different fragments of DNA encoding Ocr (using pAR2993) and POcr (using pPOcr). The pPOcr construct encodes Ocr in which all aspartate and glutamate residues have been mutated to asparagine and glutamine, respectively (referred to as positive Ocr; POcr).

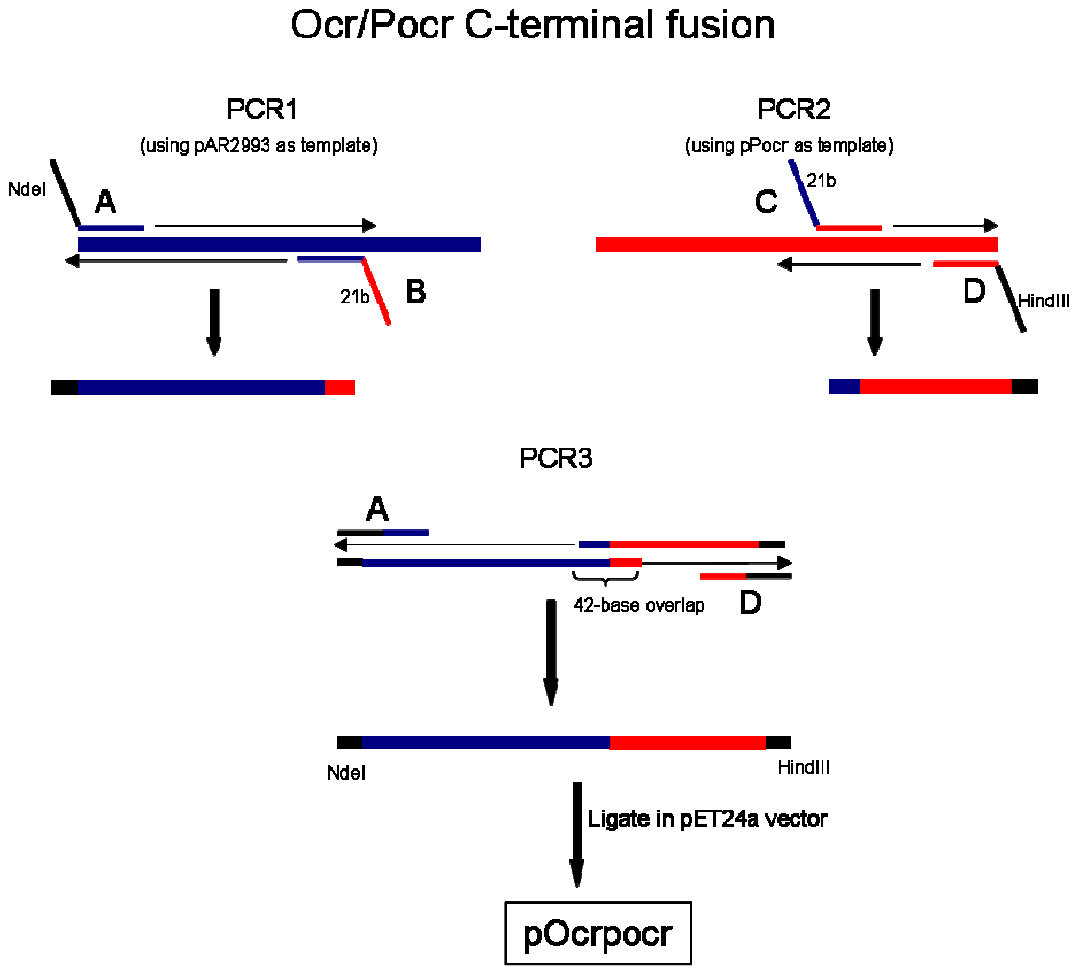


Figure 5-9. The overall strategy for the creation of the Ocr/POcr C-terminal fusion.
The methodology involved three different PCRs (PCR1-3), four different primers (A-D) and the DNA templates for Ocr (blue line) and POcr (red line) (constructs pAR2993 and pPOcr, respectively). The product of PCR3 was then ligated in a pET24a vector as is described in the main text.

The first two PCRs (PCR1 and PCR2) were conducted separately using the primer pairs A, B and C, D to amplify specific sections from pOcr and pPocr, respectively. The PCR cycle parameters are shown in Table 5-2 and the reaction conditions are described in section 2.2.3.1. Both PCRs generated amplified products of the anticipated size (Figure 5-10). The two products were gel purified (sections 2.2.4 and 2.2.5) and then used as template for PCR3 (Table 5-2) using primers A and D as described in section 2.2.3.1. Note that primers B and C (of PCR1 and PCR2) were designed to generate a partial overlap, which facilitated the generation of full length template during the first extension cycle of PCR3. This full length product then allowed amplification of Ocr/POcr using primers A and D.

Table 5-2. PCR cycle parameters for DNA cloning

Segment	Cycles	Temperature	Time
1	1	95°C	30 seconds
2	25	95°C	30 seconds
		55°C	1 minute
		72°C	1 minute

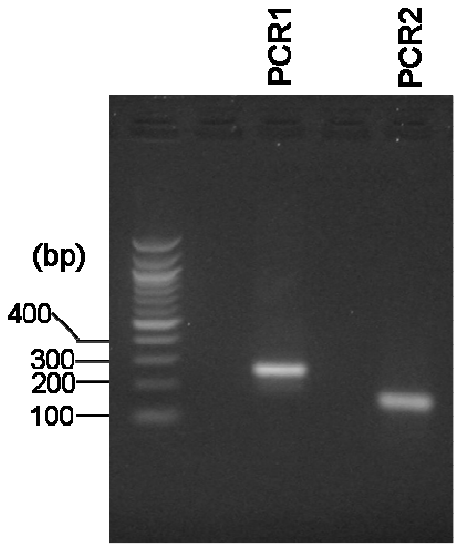


Figure 5-10. A 1.6% agarose gel of the DNA fragments amplified by PCR1 and PCR2.
The product of PCR1 encodes the N-terminal region of Ocr (~240 bp in length). The PCR2 product encodes the C-terminus region of Pocr (~150 bp in length).

Primers A and D incorporated NdeI and HindIII restriction recognition sites, respectively, which facilitated directional cloning of the amplified product from PCR3 into the corresponding sites of pET24a (section 2.2.3.2 and 2.2.3.3). The ligation mix was used to transform XL1-Blue supercompetent cells (section 2.2.7) and recombinant clones were then selected by growth on LB plates supplemented with 50 µg/ml kanamycin. Primers A, B, C and D are listed below (5' to 3'):

Primer A

AGCTCATATGGCTATGTCTAACATGACTTACAACAACGTTTTCGAC

Primer B

CTGCAGAATACGGATTACGTTCTTGGTGTCAGGCATCAGACC

Primer C

GGTCTGATGCCTGACACCAAGAACGTAATCCGTATTCTGCAG

Primer D

GAATTCAAGCTTTTACTGTTGGTTCTGCTGATATTGCTG

The Ocr mutant constructs were verified by sequencing on both strands of the DNA insert. An example of an electrophoretogram from one such set of sequencing data (mut 3) is shown alongside the wild-type Ocr sequence in Figure 5-11. The mutated bases are highlighted.

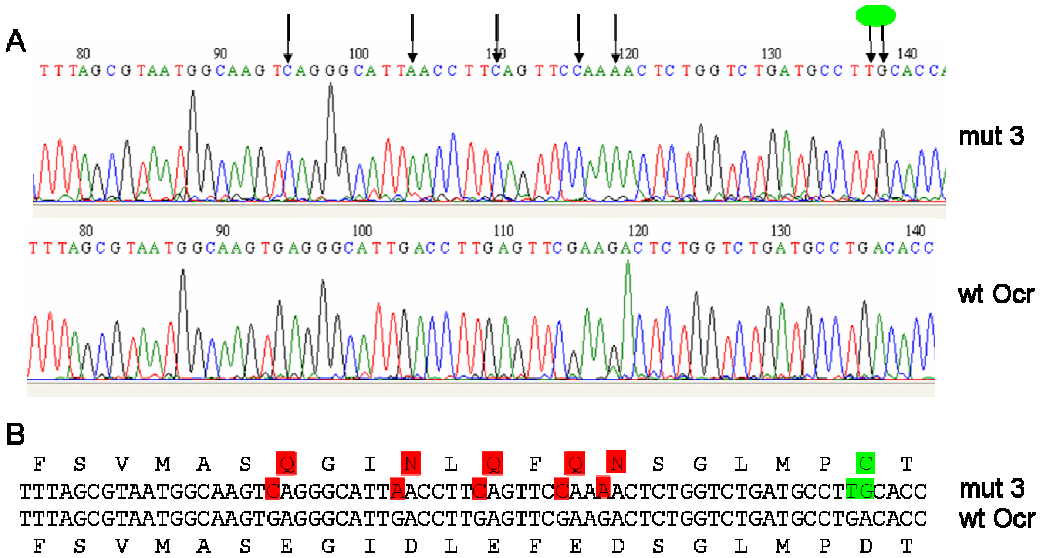


Figure 5-11. Example of sequencing data for mut 3.
(A) Raw data from an electrophoretogram of the region containing the mut 3 mutations (highlighted by arrows). The two arrows labeled with a green circle denote mutations already present in the DNA template (pOcrD73C). (B) Amino acid and nucleic acid alignments of both wild-type (wt) and mut 3. Mutations created by the PCR are colored red; mutations present in the template used are shown in green.

All the Ocr mutants created are shown in the construct list below (all constructs are pET1-based apart from pOcrpocr which is pET24-based. Numberings of residues do not take into account the N-terminal methionine of the 1S7Z sequence file shown in Figure 5-1):

pOcrmut1 D12N, E16Q, E20Q, D25N, D26N
pOcrmut2 D25C, D26N, D29N, D31N, D32N, D35N
pOcrmut3 E59Q, D62N, E64Q, E66Q, D67N, D73C
pOcrmut4 D92N, E95Q, D96N, E98Q, D99N
pOcrmut7 (mutations of mut 2 and mut 4 combined)
D25C, D26N, D29N, D31N, D32N, D35N, D92N, E95Q, D96N, E98Q, D99N
pOcrmut10 (mutations of mut 2, mut 4, mut 3 combined but excluding D73C)
D25C, D26N, D29N, D31N, D32N, D35N, D92N, E95Q, D96N, E98Q, D99N, E59Q, D62N, E64Q, E66Q, D67N
pOcrmut11 (mutations of mut 2, mut 3 combined but excluding D73C)
D25C, D26N, D29N, D31N, D32N, D35N, E59Q, D62N, E64Q, E66Q, D67N
pOcrmut12 (mutations of mut 3 but excluding D73C)
E59Q, D62N, E64Q, E66Q, D67N
pOcrpocr (all aspartates and glutamates were converted to asparagines and glutamines respectively from and including D76 onwards) D76N, E87Q, D92N, E95Q, D96N, E98Q, D99N, E103Q, E106Q, E107Q, E109Q, E110Q, E112Q, E113Q, D114N, E115G, E116Q

5.4.2. *In silico* mutagenesis and electrostatic representations

In silico electrostatic representations for wild-type Ocr and all mutant proteins, apart from mut 12, were created (work done by Dinesh Soares) (Figures 5-13 to 5-21). These involve 4 views for each protein structure/model rotated by 90 degrees about the *y*-axis including a PyMol rendition on Figure 5-12 to show the cartoon representation of the Ocr molecule along with a semi-transparent surface for orientation guidance. All figures were first superimposed and are in equivalent orientation for ease of comparison. The influence of each set of mutations on the electrostatic properties of the protein can clearly be seen. All relevant side-chains and mutations in view have been carefully labelled.

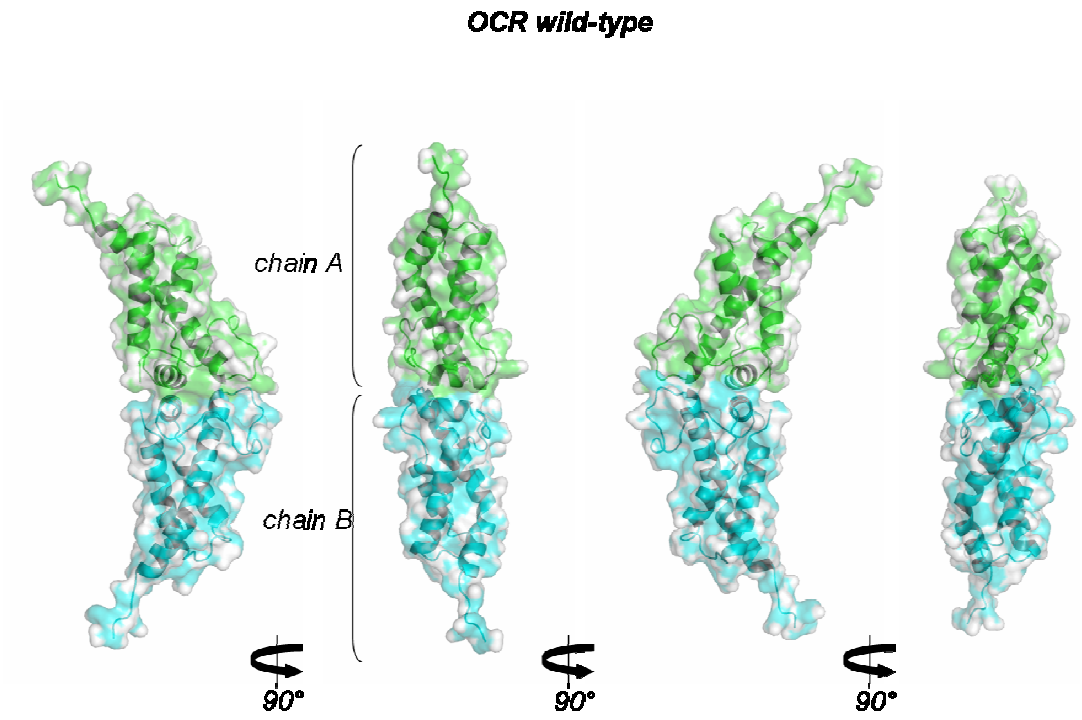


Figure 5-12. A cartoon representation of the Ocr molecule along with a semi-transparent surface.
Four views, rotated by 90 degrees about the *y*-axis, are displayed.

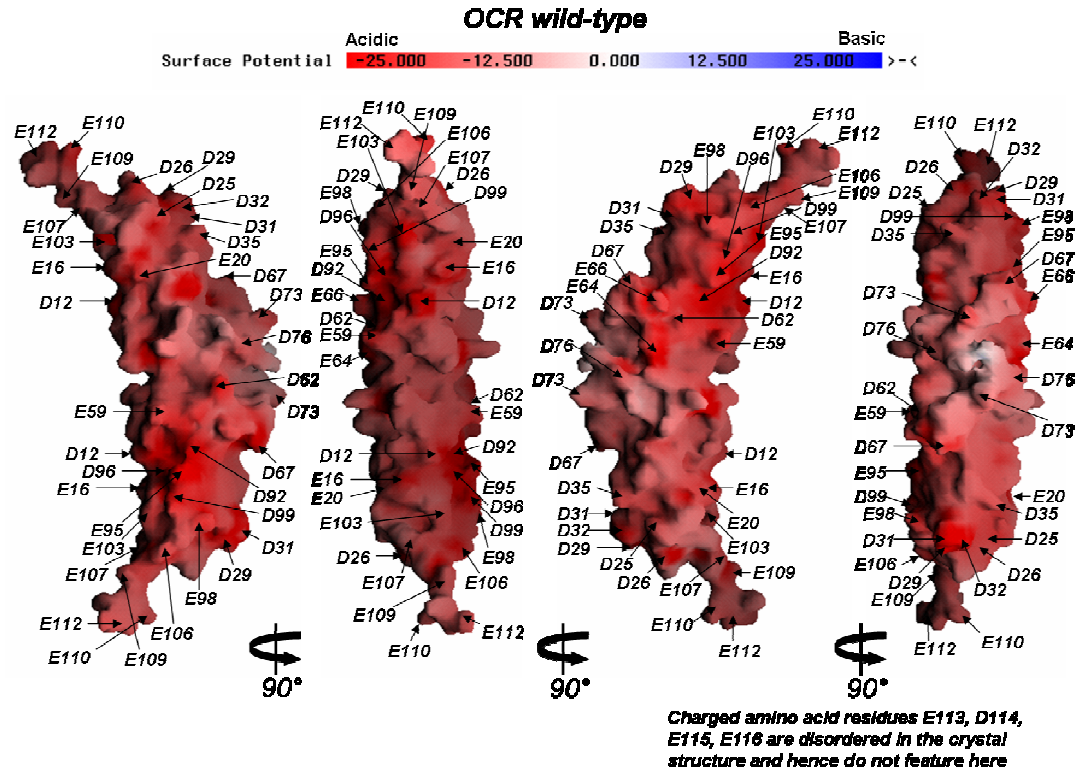


Figure 5-13. The electrostatic representations for wild-type Ocr.
All Asp and Glu residues in view are highlighted.

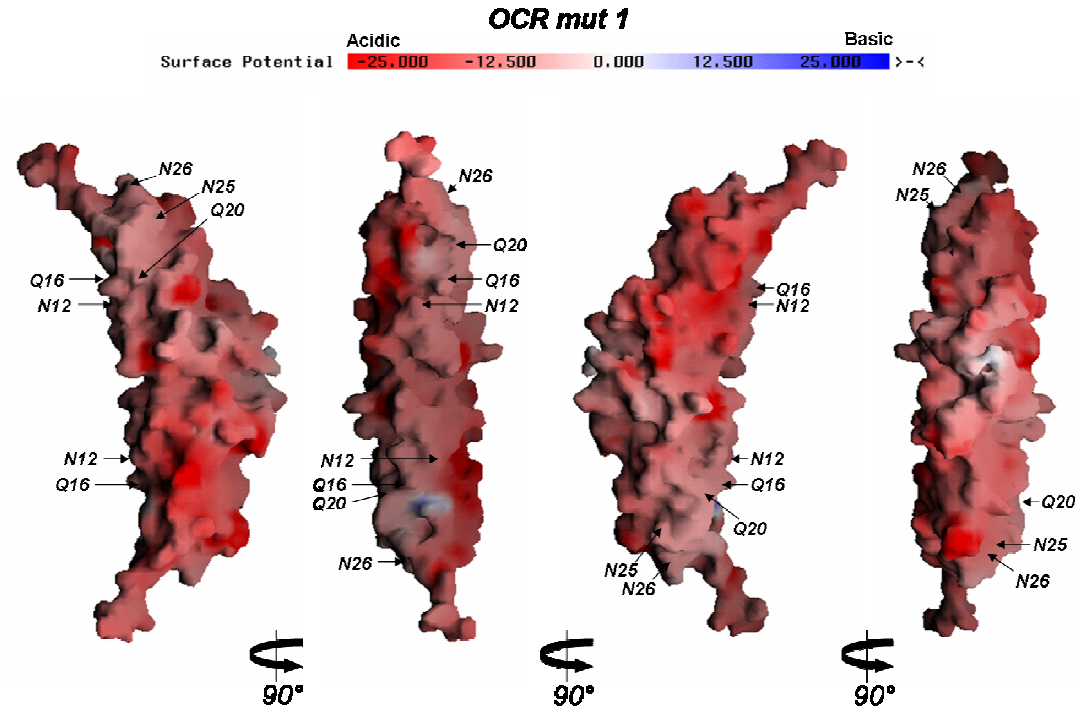


Figure 5-14. The electrostatic representation of mut 1.
The mutations in each view are labeled. Ocr mut 1 is composed of the following mutations: D12N, E16Q, E20Q, D25N, D26N.

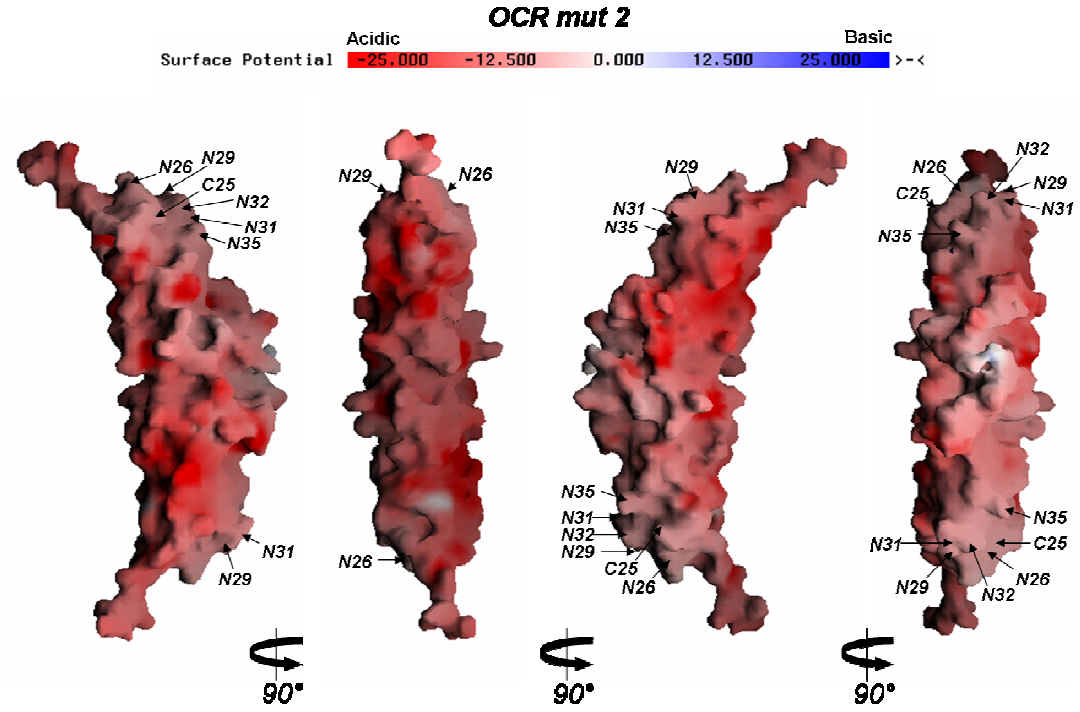


Figure 5-15. The electrostatic representation of mut 2.
Ocr mut 2 is composed of the following mutations: D25C, D26N, D29N, D31N, D32N, D35N.

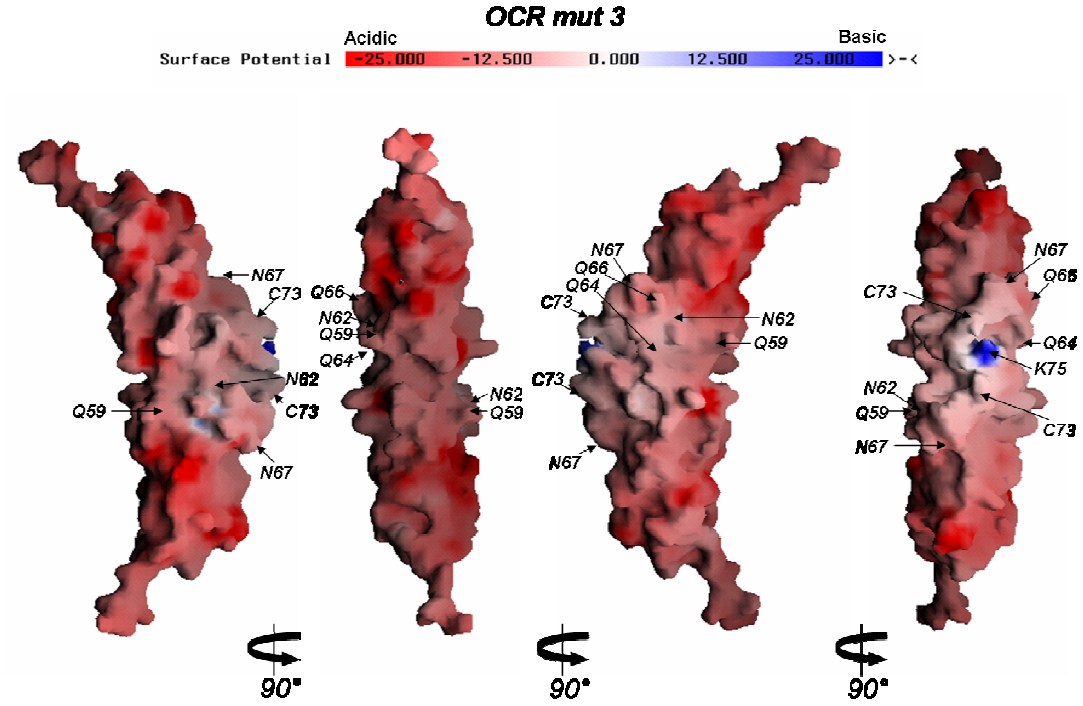


Figure 5-16. The electrostatic representation of mut 3.
Ocr mut 3 is composed of the following mutations: E59Q, D62N, E64Q, E66Q, D67N, D73C.

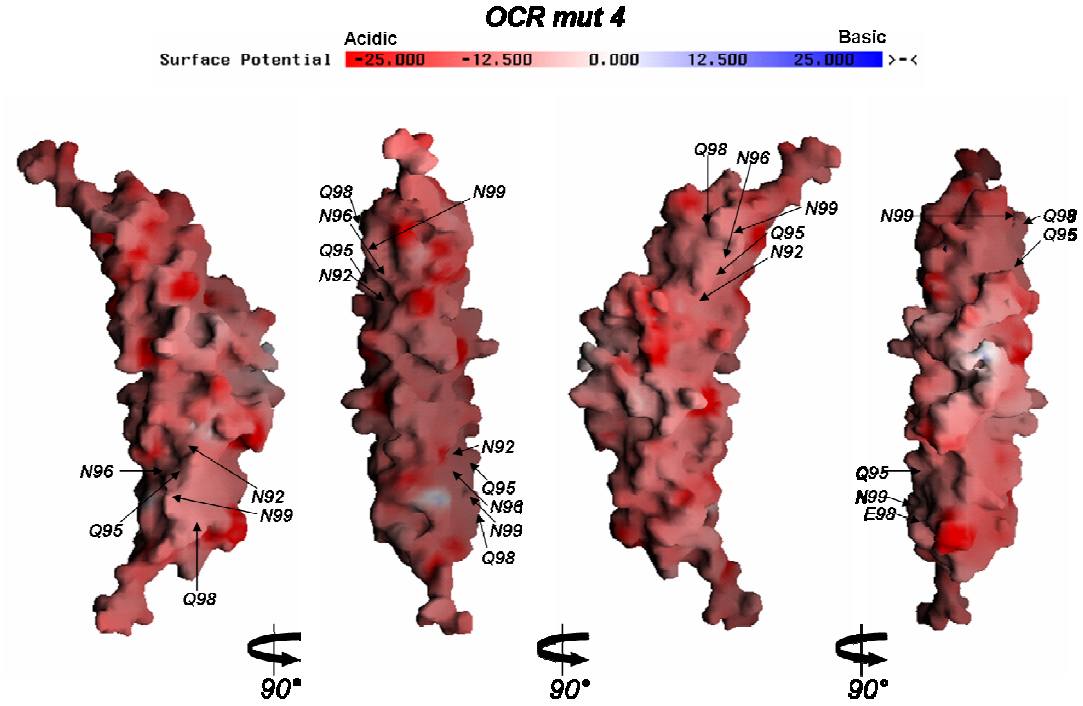


Figure 5-17. The electrostatic representation of mut 4.
Ocr mut 4 is composed of the following mutations: D92N, E95Q, D96N, E98Q, D99N.

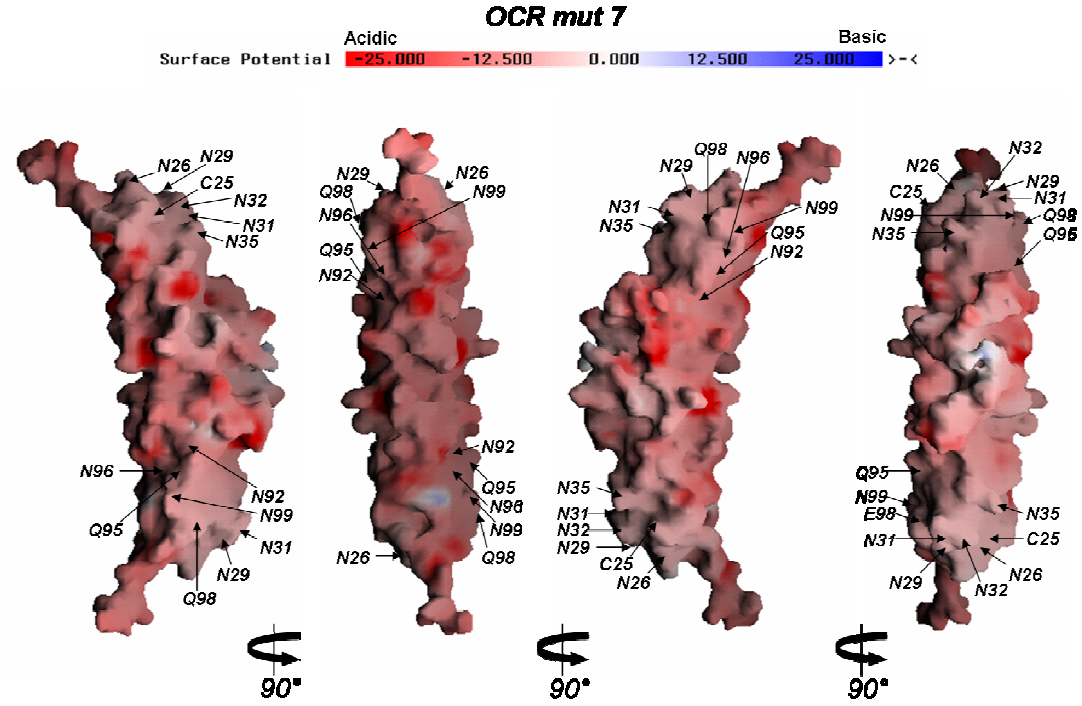


Figure 5-18. The electrostatic representation of mut 7.
Ocr mut 7 is composed of the following mutations: D25C, D26N, D29N, D31N, D32N, D35N, D92N, E95Q, D96N, E98Q, D99N.

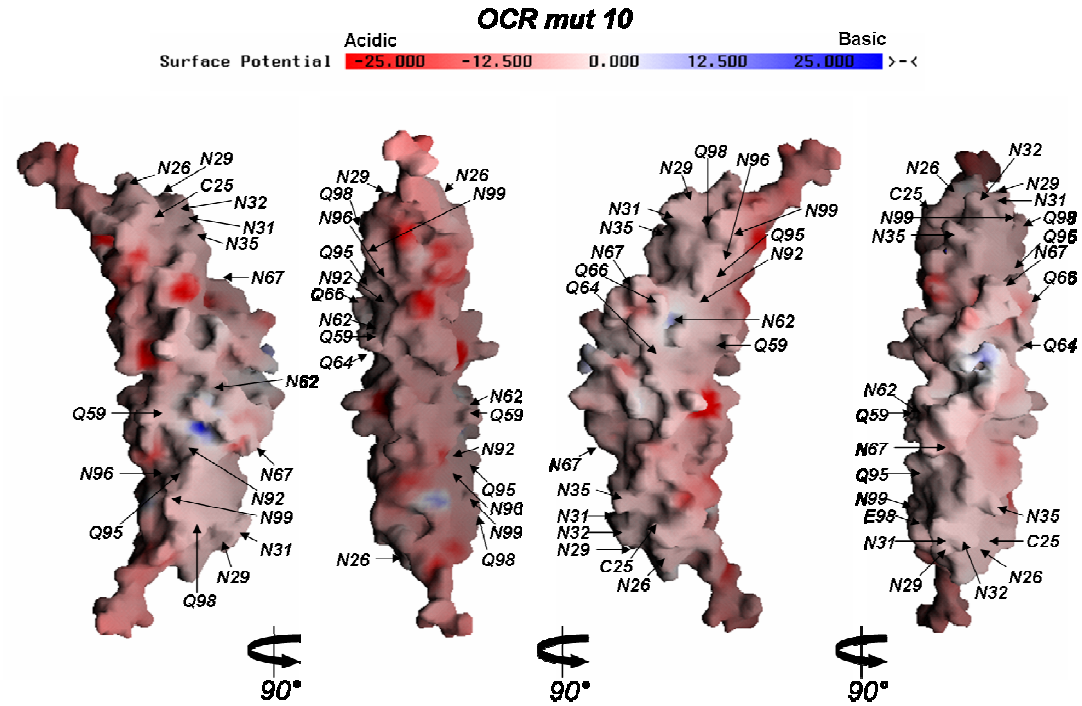


Figure 5-19. The electrostatic representation of mut 10.
Ocr mut 10 is composed of the following mutations: D25C, D26N, D29N, D31N, D32N, D35N, D92N, E95Q, D96N, E98Q, D99N, E59Q, D62N, E64Q, E66Q, D67N.

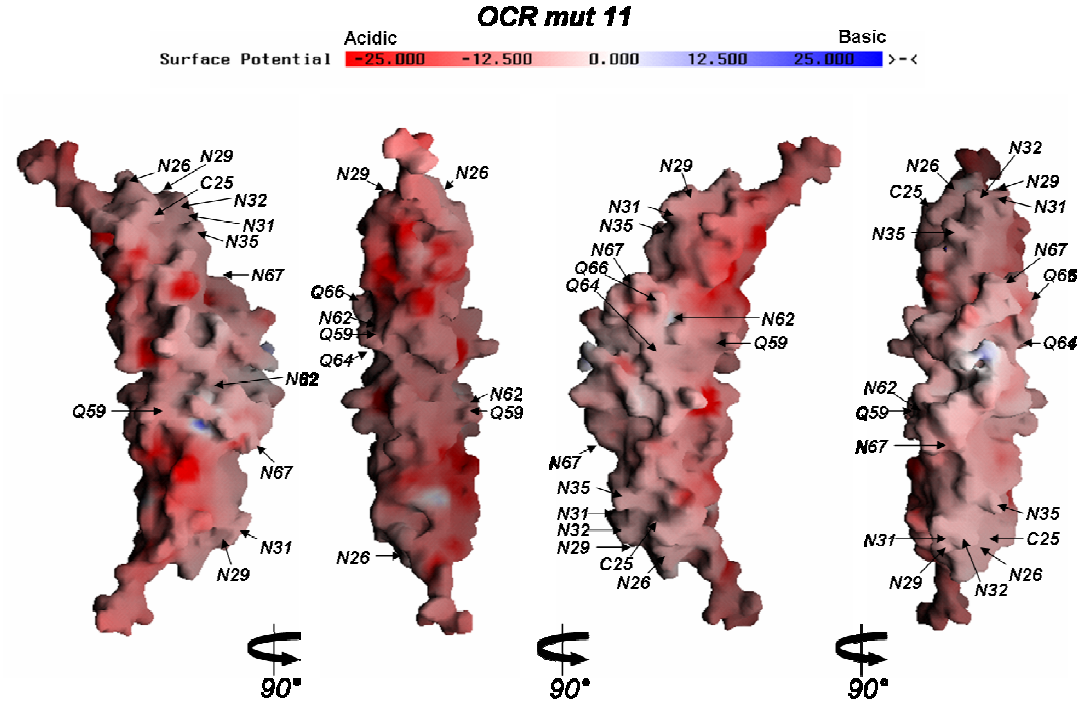


Figure 5-20. The electrostatic representation of mut 11.
Ocr mut 11 is composed of the following mutations: D25C, D26N, D29N, D31N, D32N, D35N, E59Q, D62N, E64Q, E66Q, D67N.

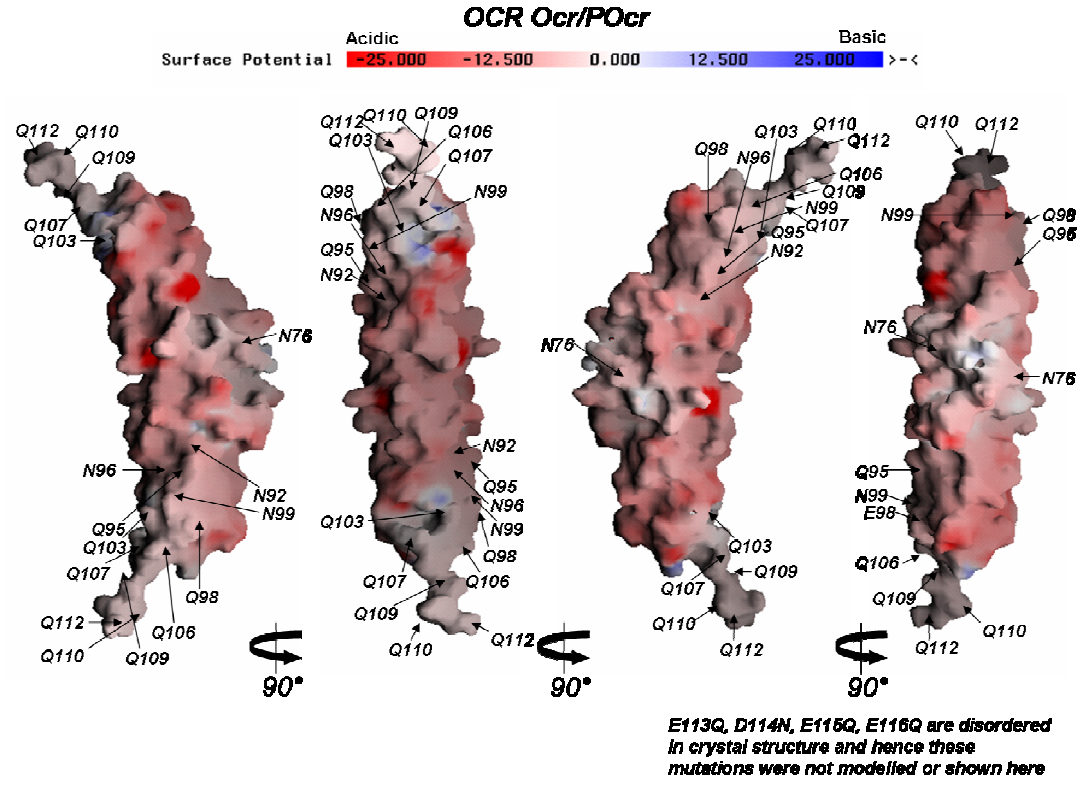


Figure 5-21. The electrostatic representation of Ocr/POcr.
Ocr mutant Ocr/POcr is composed of the following mutations: D76N, E87Q, D92N, E95Q, D96N, E98Q, D99N, E103Q, E106Q, E107Q, E109Q, E110Q, E112Q, E113Q, D114N, E115G, E116Q.

5.4.3. Protein expression and purification

The Ocr mutant proteins were expressed and purified as described in section 2.3.2.1 During the purification procedure, all mutant proteins behaved essentially the same as wild-type Ocr (section 3.4.1) and showed greater than 95% purity.

5.4.4. Circular dichroism (CD)

The Ocr mutant proteins were subjected to circular dichroism (CD) in the far UV region, which detects the presence of protein secondary structure (section 2.4.1). The structure of Ocr is dominated by α-helices as reflected from the CD spectra, which displays two distinct minima at 208 nm (part of the π to π* transition) and 222 nm (*n* to π* transition). CD spectra of all Ocr mutant proteins that were analysed showed no major alteration in secondary structure content (Figure 5-22). Note that mut 2 and mut 4 were not subjected to CD analysis because these mutations were incorporated in mut 7, which does not display any structural changes by comparison to wild-type Ocr. Spectra were analysed using the Dichroweb online secondary structure deconvolution programme (Whitmore & Wallace, 2004). Analysis using the CDSSTR method gave an α-helical content of 60% for wild-type Ocr, which is in good agreement with the crystal structure. Similar values (± 2%) were obtained for the mutant proteins. It was therefore concluded that the mutant proteins retained the overall fold of wild-type Ocr.

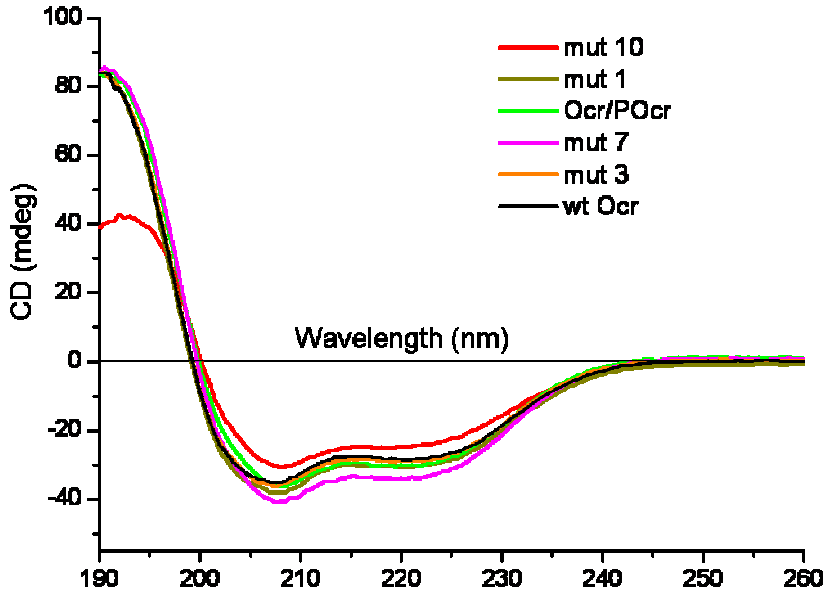


Figure 5-22. Secondary structure determination of Ocr mutants by Far UV circular dichroism.

All Ocr proteins were analysed at ~30μM.

5.4.5. *In vivo* phage restriction assays (spot tests)

Once the mutations had been confirmed by sequencing, and their secondary structure verified, the biological activity of the mutated Ocr proteins was then tested *in vivo* (section 2.4.10). These experiments were conducted at low levels of heterologous gene expression in the absence of induction by IPTG (*i.e.*, leaky expression). *E. coli* cells with (NM1049DE3) or without (NM1261DE3) the chromosomal EcoKI R/M system were transformed with either an expression vector (a control lacking the Ocr DNA insert) or the vector containing the wild-type Ocr gene or the mutated Ocr genes. Each of these transformed strains was then challenged with phage lambda $\lambda_{v.o}$ (unmodified lambda DNA) or $\lambda_{v.k}$ (modified lambda DNA).

As expected, cells transformed with the vector alone showed a strong reduction (four orders of magnitude) in the efficiency of plating of $\lambda_{v.o}$ compared to $\lambda_{v.k}$. Cells transformed with the plasmid expressing the wild-type Ocr showed

essentially identical numbers of plaques with both $\lambda_{v.o}$ and $\lambda_{v.k}$, demonstrating that Ocr had inhibited the R/M system. Assaying the mutated forms of Ocr showed that all of the mutants were fully active apart from Ocr mutants mut 10, mut 11 and mut POcr which displayed no anti-restriction activity (Figure 5-23). However, no apparent expression of the mut POcr protein could be detected in the corresponding recombinant strain even after IPTG induction. This observation suggests that mut POcr is inherently unstable and subject to degradation within the cell.

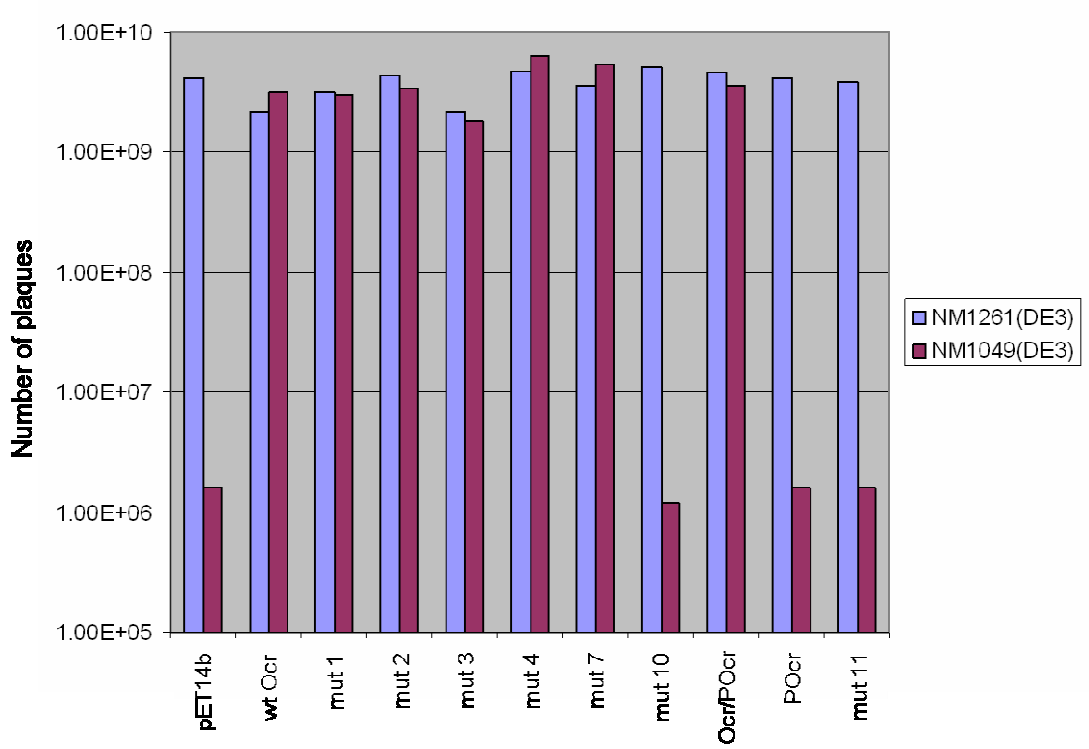


Figure 5-23. Data for the *in vivo* phage restriction assays (spot tests).

All mutants retain anti-restriction activity *in vivo* apart from mut 10, mut 11 and mut Pocr, which are completely inactive.

5.4.6. Isothermal Titration Calorimetry

The ITC data showed a highly exothermic interaction between Ocr and M.EcoKI with a stoichiometry consistent with one Ocr dimer per M.EcoKI (Figures 5-24 and 5-25). The dissociation constants derived from these experiments are tabulated in Table 5-3 and plotted in Figure 5-26. ITC showed no apparent binding for mut 10 or mut 11 ($\Delta H = 0$). The apparent Kd for the wild-type Ocr:M.EcoKI interaction as determined by ITC was ~1.8 nM instead of ~0.044 nM (as calculated by fluorescence anisotropy). The discrepancy between these two values is due to the limit of sensitivity for ITC, which is known to be about 10 nM. The same holds true for mut 1, mut 2, mut 4, mut 7 and Ocr/Pcr, which show sharp binding transitions. Nevertheless, the ITC data clearly indicate that these mutants display binding constants below 10 nM. By contrast, the binding transitions for mut 3 and mut 12 were sigmoidal (Figure 5-24 and 5-25) and give Kd values of 79.6 nM and 29.3 nM respectively (Table 5-3).

Table 5-3. Dissociation constants and free energy of binding for the various M.EcoKI:Ocr mutant interactions as measured by ITC.

ND; not detectable due to no apparent transition *i.e.*, $\Delta H = 0$.

Ocr protein	Kd (nM)	ΔG (kcal/mol)
mut 1	0.15 ±0.4	13.39 ±1.59
mut 2	1.87 ±0.77	-11.9 ±0.25
mut 3	79.55 ±21.25	-9.68 ±0.16
mut 4	0.2 ±0.18	-13.22 ±0.54
mut 7	3.57±2.02	-11.52 ±0.34
Ocr/POcr	2.19 ±0.87	-11.81 ±0.24
mut 10	ND	ND
mut 11	ND	ND
mut 12	29.29 ±8.94	-10.27 ±0.18
wt Ocr	1.82 ±0.83	-11.92 ±0.27

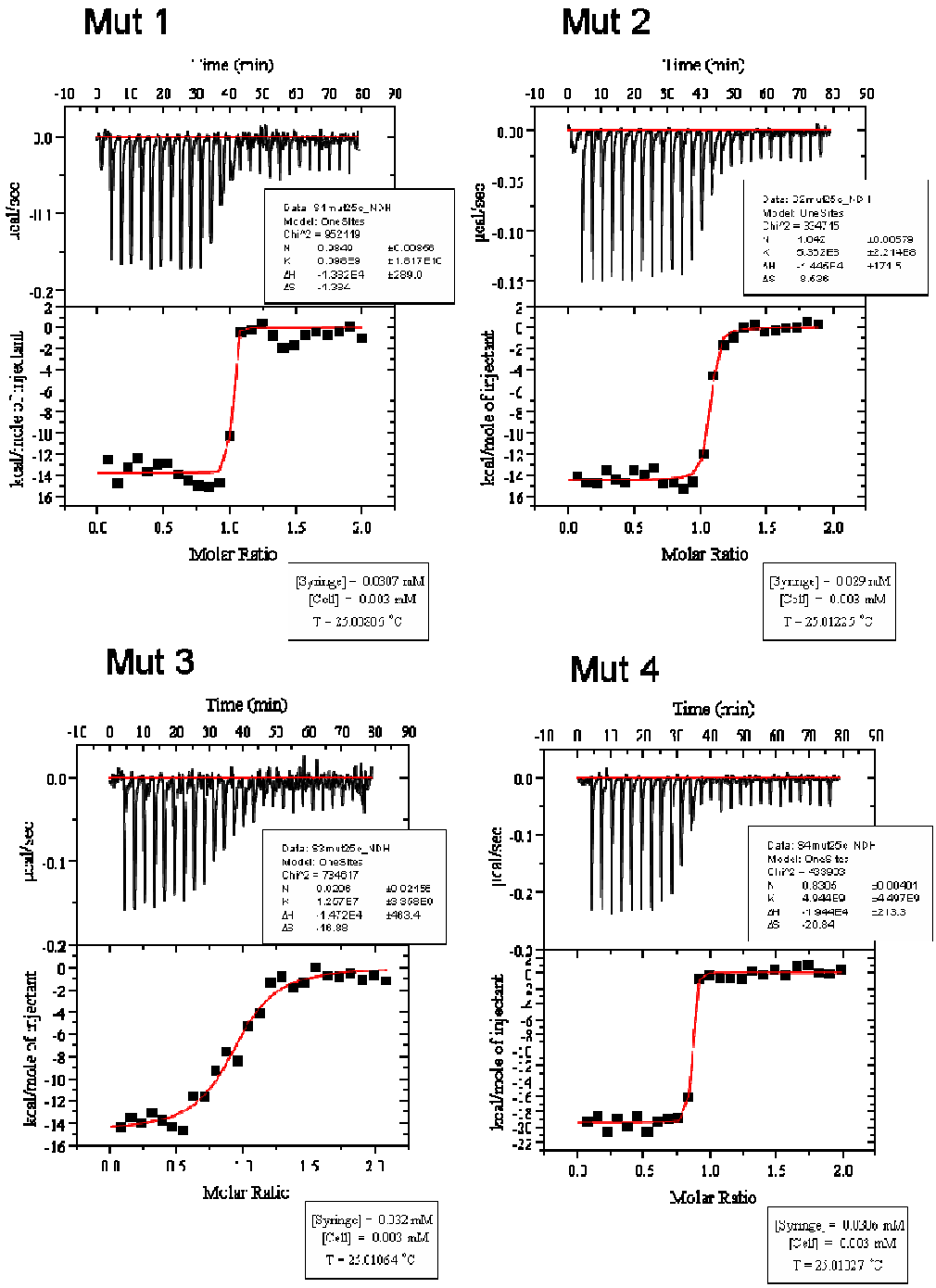


Figure 5-24. ITC traces for the various M.EcoKI:Ocr mutant interactions

The "Cell" contained the M.EcoKI, while the "Syringe" contained corresponding Ocr protein. Protein concentrations are shown for each titration.

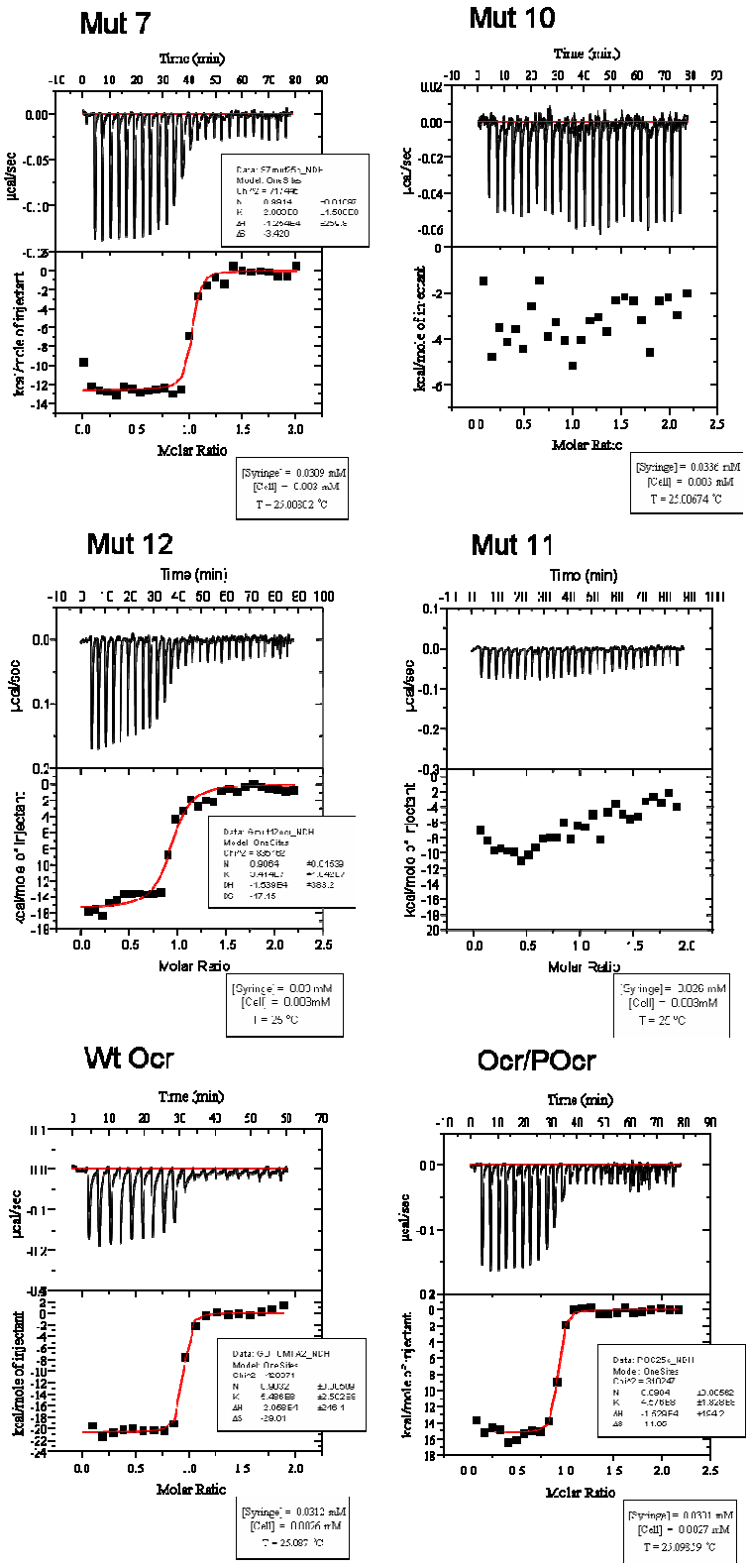


Figure 5-25. ITC traces for the various M.EcoKI:Ocr mutant interactions

The "Cell" contained the M.EcoKI, while the "Syringe" contained corresponding Ocr protein. Protein concentrations are shown for each titration.

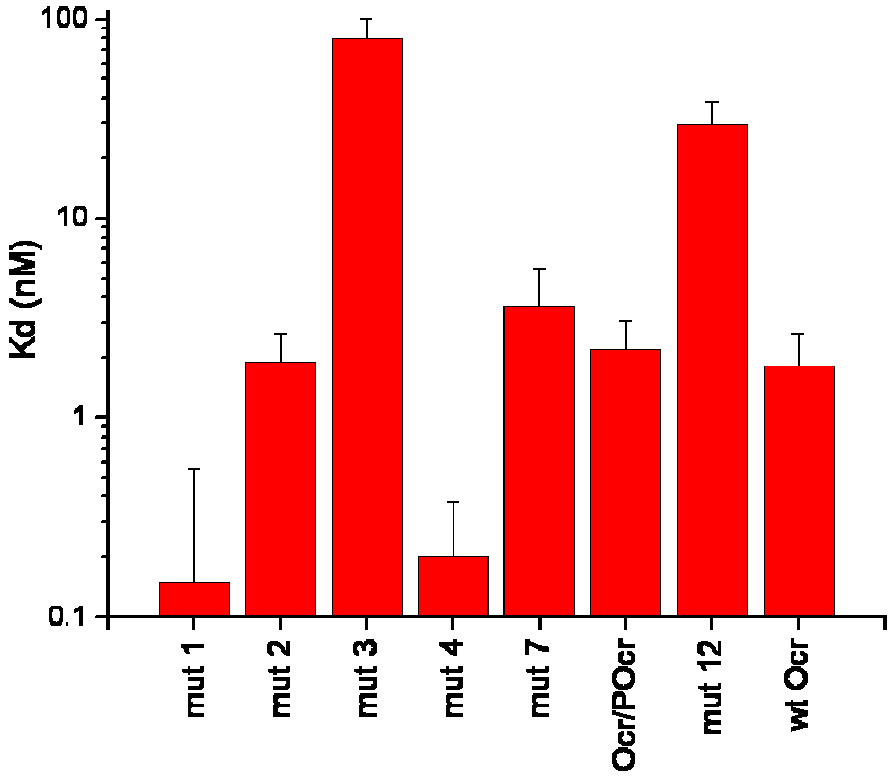


Figure 5-26. Bar chart of the dissociation constants derived from ITC for the various M.EcoKI:Ocr mutant interactions.

Mut 11 and mut 10 showed no apparent transitions (i.e, ΔH=0) and therefore no binding affinity could be calculated.

5.4.7. Fluorescence anisotropy

5.4.7.1. Fluorescence anisotropy competition assay

In order to assess possible changes in the free energy of binding for the Ocr proteins to M.EcoKI, a fluorescence anisotropy competition assay was used as described in section 2.4.6. Given a K_d of 2.11 nM for the M.EcoKI-DNA interaction, we find that the mutant forms of Ocr have apparent K_d values ranging up to ~88 times greater (mut 10) than wild-type Ocr. The Kd values calculated from the fluorescence anisotropy assay are tabulated in Table 5-4 and are also displayed as a bar chart in Figure 5-27. The Kd errors represent the goodness of the fit of the averaged raw data of n=2 (refer to Appendix B for examples of raw data).

Table 5-4. Dissociation constants and free energy of binding for the various M.EcoKI:Ocr mutant interactions as measured by the fluorescence anisotropy competition assay.

Ocr protein	Kd (nM)	ΔG (kcal/mol)
mut 1	0.56 ± 0.21	-12.62 ±0.23
mut 2	0.55 ± 0.08	-12.63 ±0.10
mut 3	1.65 ± 0.25	-11.97 ±0.10
mut 4	0.33 ± 0.11	-12.93 ± 0.20
mut 7	0.60 ± 0.14	-12.57 ±0.14
mut 10	3.86 ± 0.39	-11.47 ± 0.07
Ocr/POcr	0.91 ± 0.19	-12.33 ±0.13
wt Ocr	0.04 ± 0.04	-14.12 ± 0.54

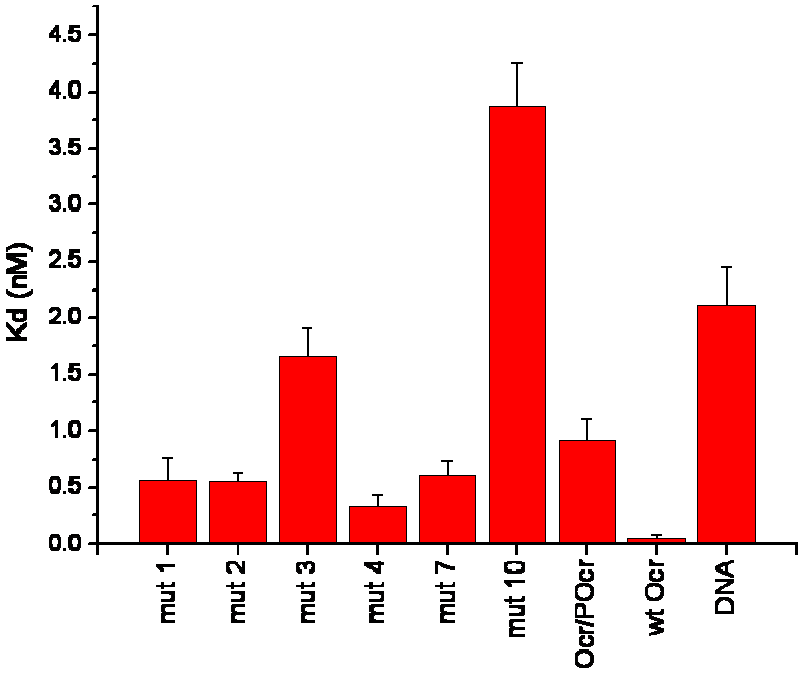


Figure 5-27. Bar chart showing the dissociation constants (Kd) in nM for the various M.EcoKI:Ocr mutant interactions as measured by the fluorescence anisotropy competition assay.

Also shown is the Kd for the M.EcoKI:DNA interaction (K_d of 2.11 nM). In each case a 21 bp ds DNA molecule containing the target recognition sequence for M.EcoKI was used. The Kd errors represent the goodness of the fit of the averaged raw data of n=2 (refer to Appendix B for examples of raw data averaging and fitting).

5.4.7.2. Fluorescence anisotropy of M.EcoKI-GFP

There was a clear discrepancy between the Kd values for mut 3 and mut 10 measured either by ITC or using the fluorescence anisotropy competition assay. Thus, a more direct approach was employed to calculate the Kd for the mut 10:M.EcoKI interaction. For this assay a C-terminal GFP (green fluorescence protein) fusion of the S subunit of M.EcoKI (generated by Dr. John White, School of Chemistry, University of Edinburgh) was used. The overall strategy involved titrating the Ocr protein into a constant amount of M.EcokI-GFP and observing the change in anisotropy of GFP.

Given that the M.EcoKI:Ocr complex displays only a small change in the Stokes radius in relation to unbound M.EcoKI, the change in anisotropy (r) was expected to be small. Indeed, the Δr (change of anisotropy of bound *versus* unbound M.EcoKI-GFP) was ~0.02. Nevertheless, this small change in r proved to be more than sufficient for calculating a Kd.

The Kd of ~97 nM for the subunits of M.EcoKI (Dryden *et al*., 1997), precluded the use of this assay at concentrations below 100 nM. Thus, Kd values below 100 nM could not be reliably determined using this methodology. As the dissociation constant for the M.EcoKI:mut 10 interaction was expected to be quite large (μM range), the initial concentration of M.EcoKI-GFP in the cuvette was set at 300 nM. The buffer used was, 20 mM Tris-HCl pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol including 100 μM SAM (SAM would be added to the cuvette immediately prior to the actual titration experiment). The experimental conditions were kept as described in section 2.4.6 apart from: λex= 395 nm, λem= 510 nm; excitation slits were set at ~ 0.5 mm and emission slits at ~10 mm, using a T-format geometry. Ocr mut 10 was then titrated into the cuvette (using a stock at ~19 μM) while taking continuous measurements. The cuvette concentrations of mut 10 ranged from ~ 20 nM to a final concentration of ~2.5 μM. A graph of the concentration of mut 10 *versus* the change in anisotropy observed for GFP (of the M.EcoKI-GFP fusion) is plotted in Figure 5-28. The Kd for the M.EcoKI: mut 10 interaction was thus calculated as 1453 ±145 nM.

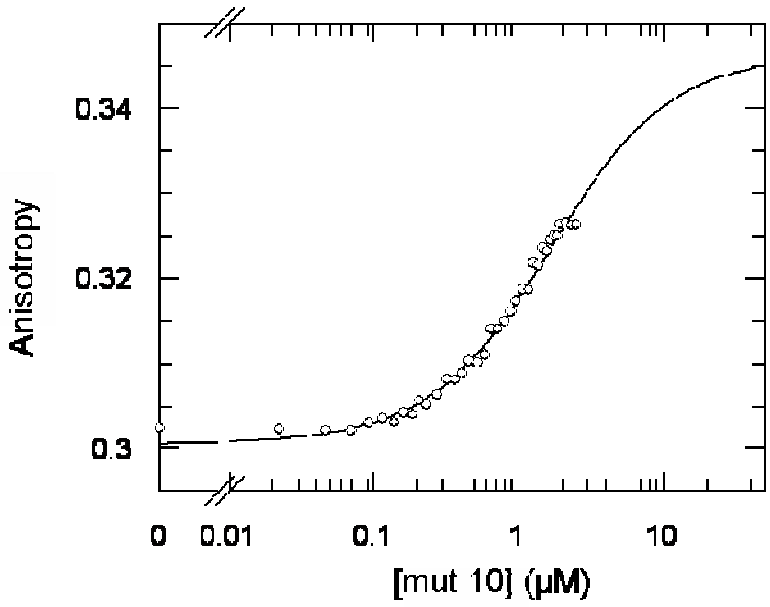


Figure 5-28. The binding assay for the mut 10:M.EcoKI-GFP interaction using fluorescence anisotropy.

Ocr mut 10 was titrated into a solution containing 300 nM M.EcoKI-GFP. The binding of mut 10 to M.EcoKI could be monitored by measuring the change in anisotropy of GFP. Upon binding of Ocr to M.EcoKI-GFP the complex tumbles more slowly and r increases. The data was fitted to a one binding site model using Dynafit (Biokin Ltd).

5.4.8. *In vitro* nuclease assay

The activity of the mutant Ocr proteins was tested in an endonuclease assay using purified EcoKI. Linearisation of a circular unmethylated plasmid (pBRsk1) containing a unique EcoKI target recognition site was monitored in the absence and presence of Ocr. In each case the reaction mixture minus DNA was prepared and the digestion initiated by addition of pBRsk1. The reaction was stopped after 10 min and the mixtures were then analysed by agarose gel electrophoresis (Figure 5-29). Incubation of pBRsk1 in the presence of a five-fold excess of EcoKI resulted in complete digestion of the plasmid within 10 min (compare the first two lanes;"uncut" and "EcoKI digest"). The experiment was also performed in the presence of a ten-fold and twenty-fold excess of Ocr dimer over EcoKI (Figure 5-29, panels A and B respectively). As anticipated, wild-type Ocr completely abolished linearisation of the plasmid DNA, as did the Ocr mutant proteins; mut 1, mut 2, mut 4 and the truncated versions of Ocr (ocr199 and ocr99). However, the mutant Ocr proteins mut 3, mut 10

and Ocr/Pocr showed no anti-restriction activity at the conditions used. Of particular interest is Ocr mut 7 which displayed partial inhibitory activity at an EcoKI:Ocr dimer ratio of 1:10 (panel A); changing to near complete inhibition of DNA linearization at the 1:20, EcoKI:Ocr dimer ratio (panel B).

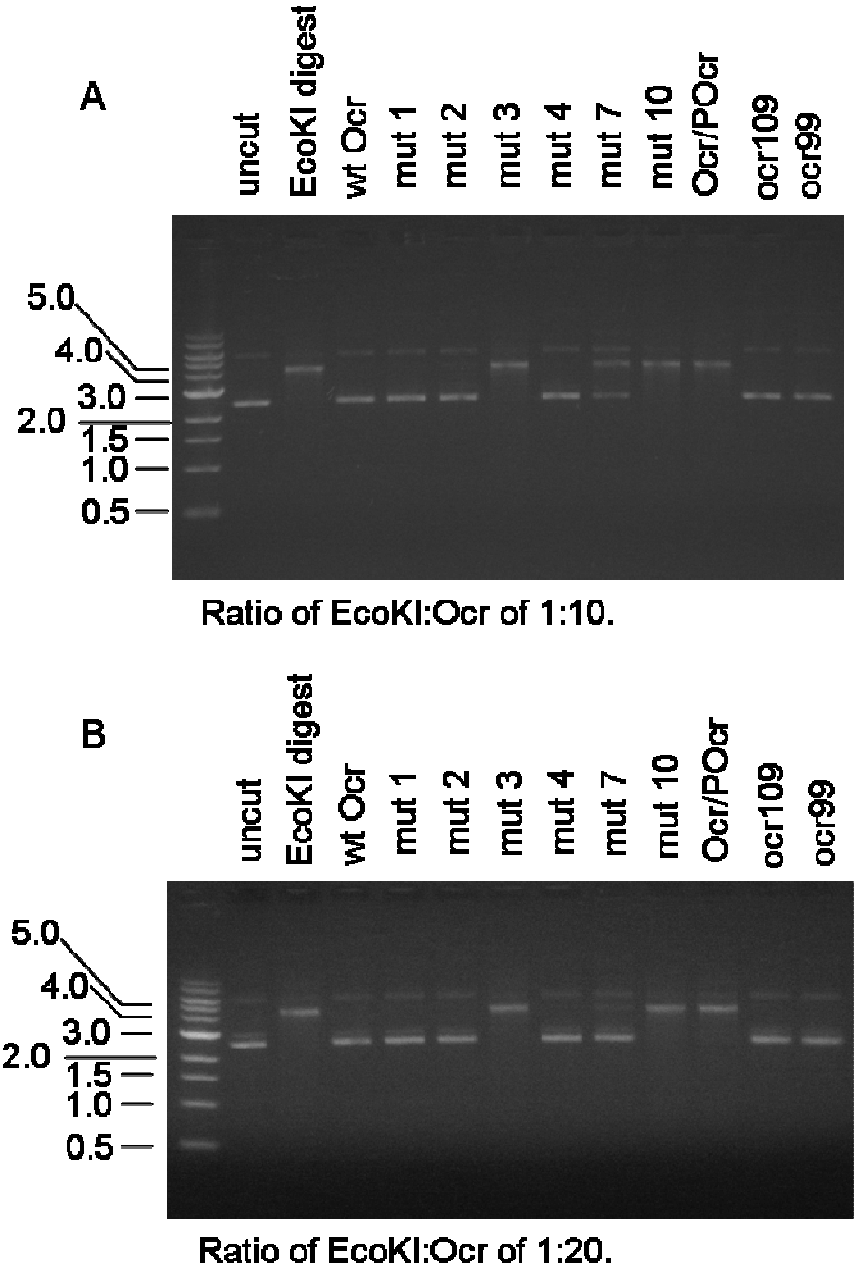


Figure 5-29. *In vitro* inhibition of EcoKI using Ocr mutants.

An unmethylated plasmid (pBRsk1) with a single EcoKI recognition site was used as substrate. Digests were performed in the presence of 10 nM EcoKI. Samples were analysed using a 0.9% agarose gel. All lanes apart from the first two ("uncut" and "EcoKI digest") denote EcoKI digests, performed in the presence of a specific Ocr protein as is labeled above each lane. M, 1 kbp ladder from New England Biolabs. Further details in section 2.4.9.

5.5. Discussion

The aim of the work described in this chapter was twofold; (i) to understand the importance and contribution of different groups of acidic residues to the activity of the Ocr protein; (ii) to establish how the Ocr molecule mimics the charge distribution of DNA.

For the purposes of this project, the simplest and quickest way to mutate multiple residues from the Ocr surface was to use the QuikChange II Site-Directed Mutagenesis kit from Stratagene. A drawback of this technique however, was that the primers used could not exceed ~45 bases in length. This set a limitation on the number of acidic residues that could be mutated in any single round of PCR. Therefore, two different strategies were adopted to mutate regions of the Ocr molecule that were distant from each other in the primary amino acid sequence; (i) more than one round of PCR was carried out to mutate several regions on the protein surface, (ii) Ocr constructs with pre-existing mutations (such as single Cys residues or constructs created for the single or double mutational analysis of Ocr for Chapter 1) were used as PCR templates. This is the reason why some mutants were generated in which acidic residues are mutated to Cys residues, rather than Glu to Gln or Asp to Asn. For example, Ocr mut 3 contains D73C whereas Ocr mutants mut 10 does not. It was simpler to use the existing template of pOcrmut7 that had already been created, whilst missing a single mutation (D73C) from mut 3 (the PCR of which used pOcrD73C as template). This strategy was deemed to be an acceptable compromise.

In retrospect, this approach proved to be adventitious because subsequent *in silico* analysis of the mutants revealed that removal of D73 creates a positive charge loci on the Ocr surface by exposure of K75 (Figure 5-16). It is very likely that the introduction of the surface positive charge may be exacerbating the reduction in binding affinity of mut 3. Indeed, this affect is probably masking the true contribution of the negative charges removed from mut 3. A modified version of mut 3 that retains D73, referred to as mut 12, was therefore prepared. Mut 12 (Figure 5-1 and 5-8) has the same loop region mutations as mut 10 and mut 11 and will facilitate a more accurate analysis of the contribution made by the negatively charged residues making up this region of the protein surface.

Another mutant created as a result of the biophysical analysis presented in this chapter is mut 11. Because mut 10 is biologically inactive, it is logical to proceed in the reverse direction by replacing Asp and Glu residues in order to establish which residues contribute the most to the binding affinity. The data so far collated for mut 11 (*i.e.*, the *in vivo* phage restriction assay and ITC for the mut 11:M.EcoKI interaction) have shown that the removal of the 11 amino acids targeted (Figure 5-1 and 5-7) results in loss of activity for the Ocr molecule.

Having purified and established the correct fold of the Ocr mutant proteins by Far UV CD analysis, their *in vivo* and *in vitro* activities were then assessed.

With the exception of mut 3, mut 12, mut 10 and mut 11, ITC data for the multiple mutants show too sharp a transition to obtain reliable dissociation constants (*i.e.*, Kd <10 nM). Because mut 10 and mut 11 displayed no apparent transition ($\Delta H=0$), only ITC for mut 3 and mut 12 could be used to obtain quantitative data. The obvious large discrepancy between the values obtained from ITC and the fluorescence anisotropy competition assay for mut 3 and mut 10, led to the analysis of the M.EcoKI-GFP:mut 10 interaction by monitoring the anisotropy of GFP. This is a simpler and more direct technique. However, due to problems associated with the dissociation of M.EcoKI at low concentrations (*i.e.*, below ~100 nM; Kd for M.EcoKI ~97 nM), the assay could only be used to assess dissociation constants above ~200-300 nM. Using this technique, a more reliable Kd for the M.EcoKI-GFP:mut 10 interaction was determined (~1.5 μM). These experiments led to the realization that the fluorescence anisotropy competition assay (as described in section 2.4.6) was unsuitable for analysing Kd values much greater than 2 nM (the Kd for the Tagged-DNA:M.EcoKI interaction). The definition of this assay as a "competition" assay requires that the Ocr:M.EcoKI dissociation constant should be in the same region as that for the Tagged-DNA:M.EcoKI interaction. Furthermore, due to the complexity of this assay with three different equilibria (see section 2.4.6), data fitting may be problematic.

Nevertheless, the competition assay is appropriate for analysing interactions with Kd values at or below ~2 nM. Indeed this conclusion is reaffirmed by the observation that the ITC data for mut 1, mut 2, mut 4, mut 7 and Ocr/POcr gives Kd

values of the same order of magnitude (~ 0.1-1.0 nM). Secondly, the competition assay agrees well with the *in vitro* nuclease assay, which also confirms how the binding affinities of these mutants relate to each other. Table 5-5 and Figure 5-30 show the most reliable data determined using the most appropriate of the three different techniques in each case. Note that the data presented in Figure 5-30 is entirely consistent with the qualitative analysis of the various Ocr mutants obtained using the *in vitro* nuclease assay (see Figure 5-29).

Table 5-5. The most reliable binding affinity data for the various M.EcoKI:Ocr mutant interactions.

Ocr protein	Kd (nM)	ΔG (kcal/mol)	ΔΔG (kcal/mol)
mut 1 (competition anisotropy)	0.56 ±0.21	-12.62 ±0.23	1.50
mut 2 (competition anisotropy)	0.55 ±0.08	-12.63 ±0.10	1.50
mut 3 (ITC)	79.55 ±21.25	-9.68 ±0.16	4.44
mut 4 (competition anisotropy)	0.33 ±0.11	-12.93 ±0.20	1.19
mut 7 (competition anisotropy)	0.60 ±0.14	-12.57 ±0.14	1.55
mut 10 (GFP anisotropy)	1453 ±145	-7.96 ±0.07	6.16
mut 12 (ITC)	29.29 ±8.94	-10.27 ±0.18	3.85
Ocr/POcr (competition anisotropy)	0.91 ±0.19	-12.33 ±0.13	1.80
wt Ocr (competition anisotropy)	0.04 ±0.04	-14.12 ±0.54	0

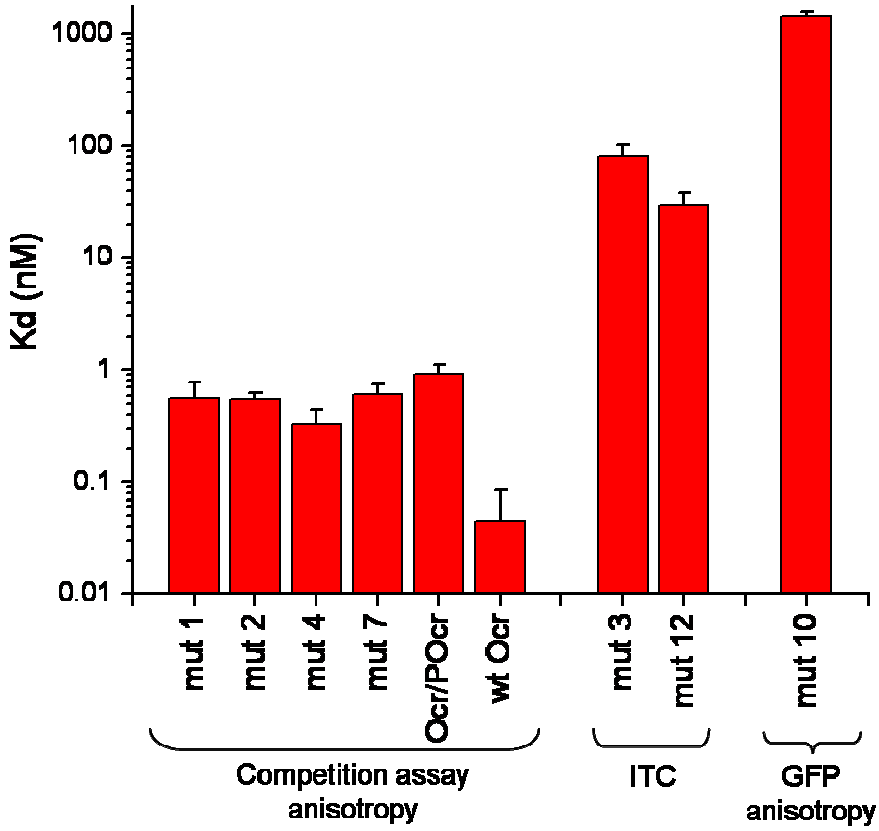


Figure 5-30. A column bar chart representing the most accurate estimates for the dissociation constants of the various Ocr mutant:M.EcoKI interactions.

By studying the data in Figure 5-30 and the ITC data in Figure 5-26, an attempt is made to describe certain regions of negative charge on the Ocr protein, in descending order of importance and contribution to the binding affinity of the M.EcoKI:Ocr interaction (Figure 5-31).

Mut 3 (6) Figure 5-4 > Mut 2 (6) Figure 5-3 > Mut 1 (5) Figure 5-2 > Mut 4 (5) Figure 5-3

Figure 5-31. Descending order of importance of certain regions of negative charge on the Ocr surface.

Numbers in parentheses correspond to the number of acidic residues mutated.

Intriguingly, there is a non-additive effect on the loss of binding affinity when mut 3 and mut 7 are combined to form the Ocr mutant mut 10. While the Kd values for mut 3 and mut 7 are ~80 and 0.6 nM, respectively, the mut 10 Kd was determined to be ~1.5 μM (Figure 5-30). This non-additive effect is even more pronounced for mut 11, since replacement of the five negative residues in the relatively unimportant mut 4 region does not restore binding activity to the mut 10 protein. Such analysis highlights the importance of the mut 3 and mut 2 regions in terms of the Ocr:M.EcoKI interaction. The non-additive effect of mut 2 and mut 3 may be explained by looking at the proposed model structure of the M.EcoKI:Ocr complex published by Kennaway *et al*., 2009 (Figure 5-32). According to this model both the aforementioned mutant regions of Ocr are in direct contact with M.EcoKI, forming the majority of the interacting interface with the highly basic TRD regions of the S subunit.

In conclusion, the data presented in this Chapter have helped to highlight regions on the Ocr surface that are likely to participate in the interaction interface with M.EcoKI. Together with the results from the previous two Chapters, a picture of the way in which Ocr mimics the charge distribution of DNA is emerging that can explain its specific interaction with the Type I R/M system. The implications of these findings are discussed in the next Chapter.

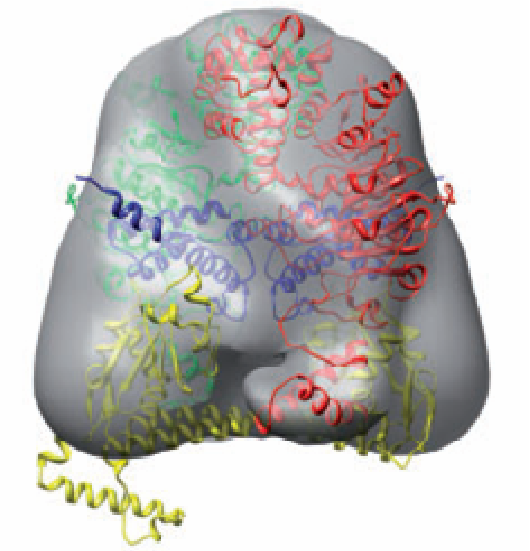


Figure 5-32. A model of M.EcoKI bound to Ocr at ~18 Å resolution as determined by negative-stain electron microscopy (EM).

A semi-transparent surface representations of the EM 3D reconstruction each with a view of the modelled coordinates fitted as a rigid body (green and red—HsdM; blue—Ocr dimer; yellow— HsdS). Adapted from Kennaway *et al*., 2009.

Chapter 6. General remarks and future work

The multi mutational work presented in Chapter 5 highlights the primary importance of the loop region (E59-D76) of Ocr in terms of its interaction with M.EcoKI. This conclusion is consistent with the suggestion (Chapter 3, section 3.5) that residues S68, D62 and probably E59 are in close proximity with M.EcoKI (4.0 ± 0.5 Å) and hence constitute part of the interacting interface. It is well known that loops are frequently exposed on the protein surface and often form part of an active site or protein–protein interface (Caffrey D. *et al*., 2004). For example the deep catalytic cleft of serine proteases are bound by the highly conserved 'canonical' loop conformation of their cognate protein inhibitors (Jackson and Russell, 2000). Furthermore, loops play a prominent part in protein-protein interfaces in antibody-antigen complexes. The complementarity determining regions (CDRs) of antibodies, which determine specificity and constitute the antigen binding interface, show a highly conserved protein framework formed by a number of loops (Burgoyne and Jackson, 2006). Indeed, as can be seen from Figure 6-1, both the mut 2 and mut 3 regions are the two most prominent loop regions of the Ocr molecule containing very highly conserved acidic residues. It is also interesting to note that the large loop region of Ocr (mut 3 region) also has the characteristics of a typical loop, containing both Gly and Pro residues. These two residues are thought to play a more structural role given that they are secondary structure breakers. For this reason, Gly and Pro are often located in loops (Crasto and Feng, 2001) and helix caps (Fetrow *et al*., 1997). Given the importance of these loops on the Ocr surface in terms of the interaction with M.EcoKI, it is not surprising that mutation of the highly conserved acidic residues on the mut 2 and mut 3 loop regions renders the molecule biologically inactive.

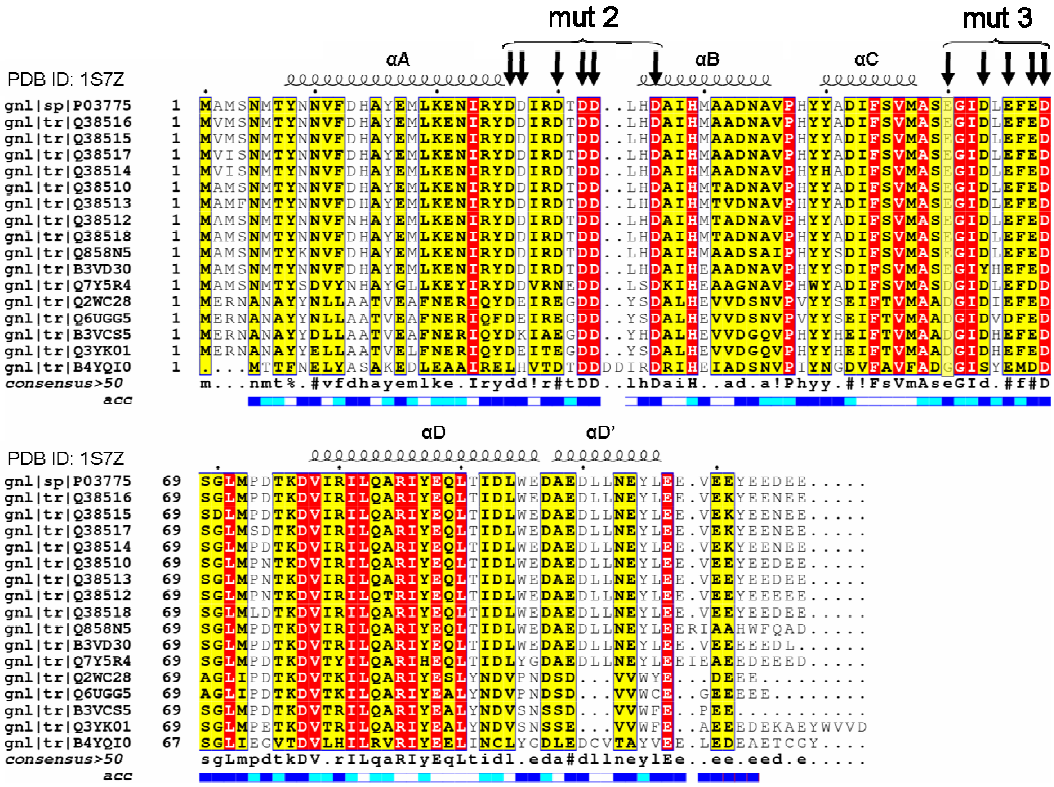


Figure 6-1 Multiple sequence alignment of phage Ocr-related sequences.
A sub-set of related sequences from a BLAST search against the UniProt database (EValue: 10-6). Fragmented sequences were removed from the list and the remaining sequences aligned using ClustalX, and depicted using ESPript v2 (red: strictly conserved; yellow: highly conserved/conservatively substituted; white: non-conserved). The secondary structure of the ocr bacteriophage T7 structure (PDB ID: 1S7Z) is mapped on the top of the alignment and extent of alpha-helices shown according to Walkinshaw *et al.*, 2002. The majority sequence consensus is given at the bottom of the alignment and corresponding residue solvent-accessibility from the crystal structure is also indicated (blue: accessible; cyan: intermediate; white: buried). The black arrows highlight the mut 2 and mut 3 mutations, the combination of which forms mut 11 (a none active form of Ocr).

By looking at the positions of mut 2 and mut 3 on the Ocr molecule one can visualise the mechanism by which Ocr mimics the charge distribution of DNA. As discussed in section 1.3, there are two different models that describe the way in which Ocr mimics the DNA duplex (Walkinshaw *et al*., 2002; Putnam and Tainer, 2005). In light of the findings presented in Chapter 5, the model by Putnam and Tainer, 2005 (Figure 1-18 and 6-2) looks more attractive because it takes into account the two very important Ocr loop regions (E59-D76 and D25-D35) (Figure 6-2). In addition, the model concludes that most of the acidic residues (24 out of 34 in total) play a role in the charge mimicry. Indeed, this may account for the fact the Ocr protein retains a plethora of negative charges throughout the evolutionary process. The mimicry suggested by the Putnam and Tainer model is that of a distorted rather than normal B-DNA, possibly to prevent the inhibition of critical DNA metabolising proteins. Moreover the Putnam model corroborates the importance and probable interaction of the mut 2 and mut 3 regions with the S subunit TRDs as discussed below.

The Kennaway atomic model in Figure 5-32 was based on an 18 Å resolution structure, which is unable to define the key regions of Ocr critical for the Ocr:M.EcoKI interaction. Unequivocal evidence for the spatial arrangement of Ocr bound to EcoKI can be obtained by either solving the X-ray crystal structure of the Ocr:M.EcoKI complex or by a series of cross-linking experiments. The latter experiments (which are already underway) involve using a range of different single Cys mutant Ocr proteins. The sulfhydryl group of the Cys residues are then derivatised with a short heterobifunctional cross-linker. The cross-linker is composed of a maleimide group for specific derivatisation of the Cys residues on the Ocr protein and a photo-activatable group, which is activated by UV irradiation only after the Ocr:M.EcoKI (or Ocr:EcoKI nuclease) complex is formed. Photoactivation leads to the generation of a highly reactive triplet-state intermediate, which can rapidly insert or add to organic components within van der Waals distance. In this fashion, covalent cross-linked species are generated that can be analysed to determine the relative orientation of the bound Ocr molecule with respect to M.EcoKI.

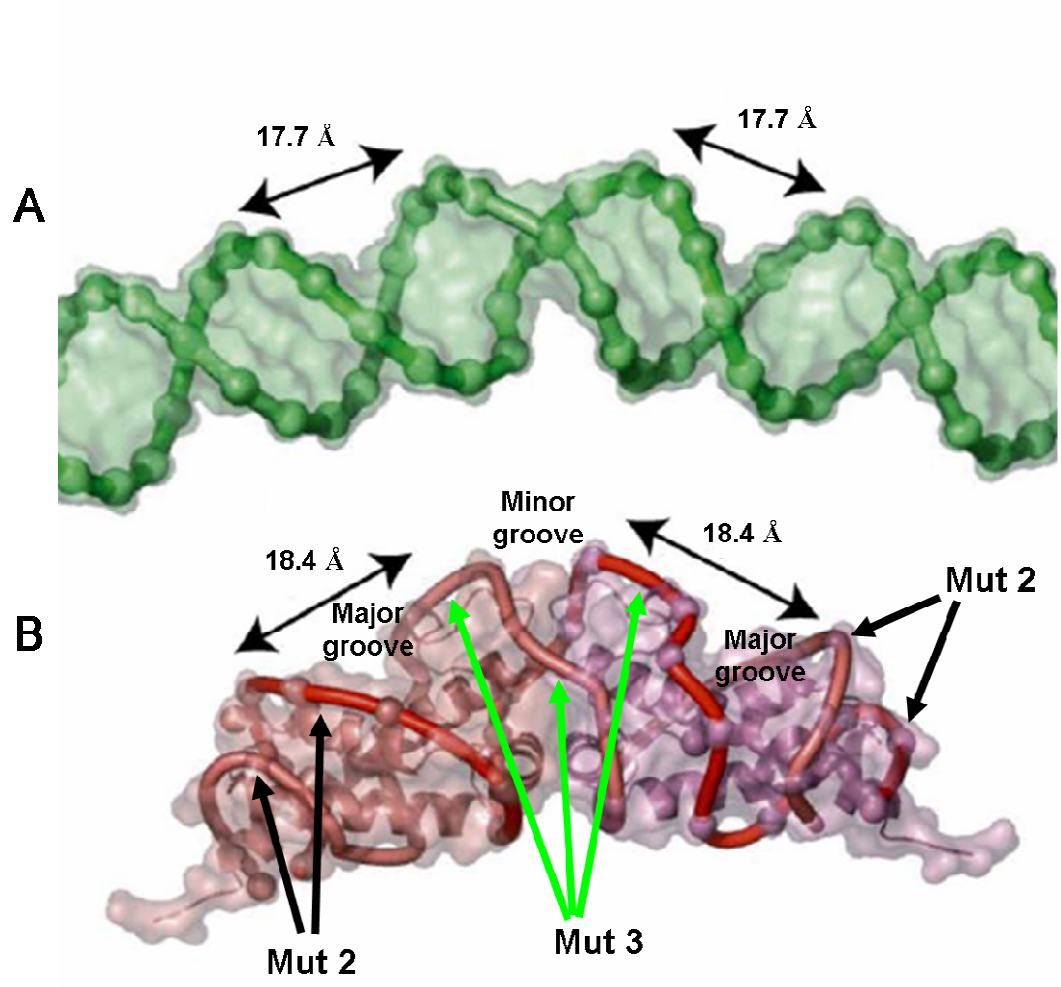


Figure 6-2. A model of the DNA charge mimicry of Ocr, proposed by Putnam and Tainer, 2005.
Comparison of Ocr (red transparent surface- Panel B) to dsDNA bent by 30 (green transparent surface Panel A). Red tubes are splines through carboxylate carbon positions (red spheres) in the dimer. Each tube mimics a single strand of DNA and they pass through the residues Glu106, Asp26, Asp25, Glu20, Glu16, Asp12, Glu95/Asp92, Glu66/Asp67, Asp73, Asp76 in one monomer and then Glu64, Asp62, Glu59, Asp51, Asp42, Asp35, Asp32, Asp31, Glu98, Asp99, Glu103, and Glu107 in the second monomer. The bent DNA was generated by tracing the DNA along a short arc over 30°. Note that the two strand mimics in Ocr have spacings reminiscent of the distance between the backbones for both the DNA major an minor grooves. The black and green arrows highlight the very important mut 2 and mut 3 regions to the anti-restriction activity of Ocr, which also confirm and prove the proposed model of the DNA charge mimicry by Ocr. Figure adapted and modified from Putnam and Tainer, 2005.

Despite poor resolution, the Kennaway model is useful with regard to the orientation of both Ocr and the S subunit of M.EcoKI. The model suggests the Ocr molecule adopts an orientation that mimics the bend of the DNA molecule shown in Figure 6-3 (a model representation of the proposed DNA:M.EcoKI interaction). The two models proposed by Kennaway describing the interaction of M.EcoKI with either Ocr or DNA are both reminiscent of the X-ray crystal structure of the interaction of a transcriptional activator BmrR with DNA. The structure of the BmrR:DNA complex has been solved to a resolution of 2.4 Å (Figure 6-4, panel A) (Zheleznova and Brennan, 2001). The Kennaway model also helps explain the non-additive effect on the loss in binding affinity when mut 2 and mut 3 are combined to form mut 11 (discussed in section 5.5). Finally the importance of the mut 2 and mut 3 regions and the way in which they are anticipated to generate the interacting interface with the S subunit of M.EcoKI is shown in Figure 6-4. The grey transparent circles highlight the critical points of interaction for both the Ocr (panel B) and the DNA molecule (panel A), based on the BmrR:DNA interaction. It is also anticipated that the conserved Trp and Tyr residues of the Ocr molecule may be involved in forming highly stable and favourable cation-π interactions with Arg or Lys residues of M.EcoKI.

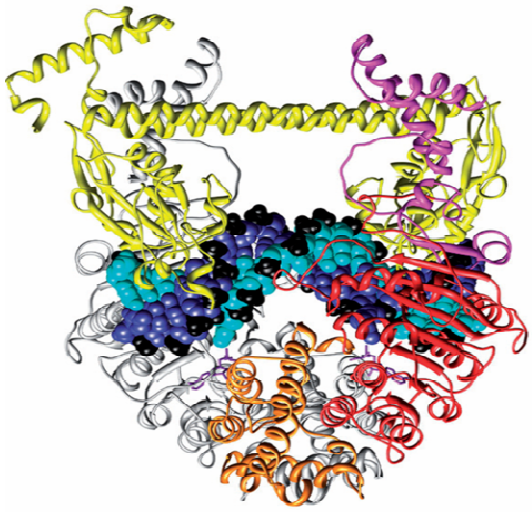


Figure 6-3. Model of M.EcoKI in complex with a DNA substrate and SAM.
Colour coding: Yellow, HsdS; Grey, HsdM; Blue and cyan, DNA; Black, phosphate groups of DNA; Orange, N-terminal domain of HsdM; Red, Catalytic domain of HsdM; Magenta, C-terminal domain of HsdM (adapted from Kennaway *et al.*, 2009).

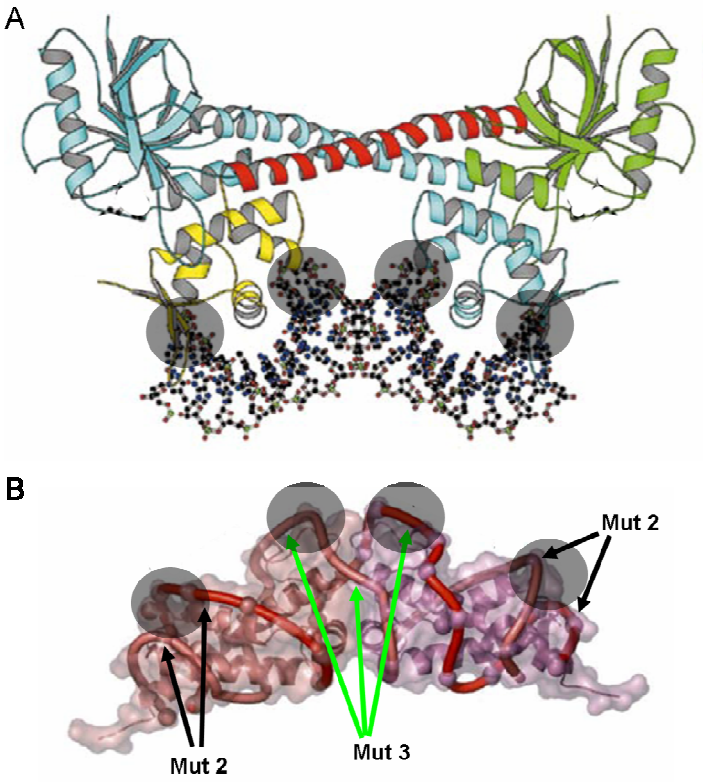


Figure 6-4. Proposed model of the Ocr : S subunit interaction.
Panel A: The crystal structure at 2.4 Å resolution of the transcription activator BmrR bound to DNA (Zheleznova and Brennan, 2001); DNA represented as balls and sticks (carbon,black; nitrogen, blue; oxygen, red; and phosphorus, green). The transparent grey circles highlight the points of contact with the DNA molecule. Panel B: The Ocr molecule represented as a DNA mimic. The grey transparent circles highlight the equivalent points of contact (as in panel A) of the very important mut 2 and mut 3 regions of the Ocr molecule with the S subunit of M.EcoKI.

Out of a total of 34 Asp/Glu residues per Ocr monomer, it has been shown that up to 11 of these residues play a critical and direct role in the interaction with M.EcoKI. This agrees well with the initial estimate of an upper limit of 16 acidic residues as being critically important (Chapter 4). The possible involvement and function of the clustered C-terminal acidic residues of Ocr in its interaction with EcoKI has not yet been studied. Apart from a likely role in the interaction with the R subunits of EcoKI, these residues may have evolved to influence the kinetics of Ocr binding. As has been mentioned previously, charged residues not directly involved in a protein interface may play a crucial role in the initial stage of protein-protein

association (*i.e.*, protein steering and pre-orientation prior to binding). Therefore, in addition to solving the X-ray crystal structure of the Ocr:M.EcoKI (or EcoKI) complex, a complete picture of the interaction can only be achieved *via* a thorough kinetic analysis. For example, stop-flow or surface plasmon resonance experiments need to be conducted to analyse the interaction between the Type I enzyme and wild-type or mutated forms of Ocr (including truncated versions of the protein). It may be that the majority of acidic residues of Ocr have evolved in order to maximise the rate of formation of enzymatically inactive complexes. Indeed, the ability to rapidly and efficiently neutralise the Type I R/M system is crucial to the bacteriophage lifecycle.

References

Anderson GW, Zimmerman JE and Callahan FM (1964) Esters of N-hydroxysuccinimide in peptide synthesis. *J. Am. Chem. Soc.* **86**: 1839-1842

Aravind L, Walker DR and Koonin EV (1999) Conserved domains in DNA repair proteins and evolution of repair systems. *Nucl. Acids Res*. **27**: 1223-1242

Arber W (1965) Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.* **19**: 365-378

Atanasiu C, Byron O, McMiken H, Sturrock SS, Dryden DTF (2001) Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7. *Nucl. Acids Res.* **29**: 3059-3068

Atanasiu C, Su T-J, Sturrock SS, Dryden DTF (2002) Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucl. Acids Res.* **30**: 3936-3944

Ban Y-EA, Edelsbrunner H, Rudolph J (2006) Interface surfaces for protein-protein complexes. *JACM* **53**: 361-378

Berge T, Ellis DJ, Dryden DT, Edwardson JM, and Henderson RM (2000) Translocation-independent dimerization of the EcoKI endonuclease visualized by atomic force microscopy. *Biophys. J.* **79**: 479-484

Bianco PR, Cuiling Xu , and Min Chi (2009) Type I restriction endonucleases are true catalytic enzymes. *Nucl. Acids Res*. **37**: 3377-3390

Bickle TA, Brack C and Yuan R (1978) ATP-induced conformational changes in the restriction endonuclease from Escherichia coli K-12. *Proc. Natl. Acad. Sci. USA*. **75**: 3099-3103

Bickle TA and Kruger DH (1993) Biology of DNA restriction. *Microbiol. Rev*. **57**: 434-450

Blackstock JJ, Egelhaaf SU, Atanasiu C, Dryden DTF, Poon WCK (2001) Shape of Ocr, the gene 0.3 protein of bacteriophage T7: modeling based on light scattering experiments *Biochemistry* **40**: 9944-9949

Bloomfield VA, Crothers DM, Tinoco JrI (2000) In *Nucleic Acids: Structures, Properties and Functions.* pp 335-443 University Science Books, California, USA

References

Bourniquel AA and Bickle TA (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie*. **84**: 1047-1059

Bandyopadhyay PK, Studier FW, Hamilton DL and Yuan R (1985) Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.* **182**: 567-578

Bogan AA and Thorn KS (1998) Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* **280**: 1–9

Brown TA (2002) *Genomes* (Second Edition), Wiley-Liss, New York, USA

Burgoyne NJ and Jackson RM (2006) Predicting protein interaction sites: binding hot-spots in protein-protein and protein-ligand interfaces. *Bioinformatics* **22**: 1335-1342

Caffrey D, Somaroo S, Hughes J, Mintseris J, Huang ES (2004) Are protein protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Sci*. **13**: 190-202

Cajthamlova K, Sisakova E, Weiser J and Weiserova M (2007) Phosphorylation of Type IA restriction-modification complex enzyme EcoKI on the HsdR subunit. *FEMS Microbiol. Lett.* **270**: 171-177

Calisto RM, Pich OQ, Pinol J, Fita I, Querol E and Carpena X (2005) Crystal structure of a putative type I restriction–modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.* **351**: 749–762

Carraway KL, Koshland Jr DE (1968) Reaction of tyrosine residues in proteins with carbodiimide reagents. *Biochim. Biophys. Acta* **160**: 272-274

Carraway KL, Koshland Jr DE (1972) Carbodiimide modification of proteins. *Methods Enzymol.* **25**: 616-623

Chan CL, Cox GB (2007) Kinetics of amide formation through carbodiimide/Nhydroxybenzotriazole (HOBt) couplings. *J. Org. Chem.* **72**: 8863-8869

Chen A, Powell LM, Dryden DTF, Murray NE, Brown T (1995) Tyrosine 27 of the specificity polypeptide of EcoKI can be UV crosslinked to a bromodeoxyuridine-substituted DNA target sequence. *Nucl. Acids Res.* **23**: 1177-1183

References

Choi KH and Licht S (2005) Control of peptide product sizes by the energy-dependent protease ClpAP. *Biochemistry*. **44**: 13921–13931

Clackson T and Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. *Science*. **267**: 383–386

Cooper A (2005) Heat capacity effects in protein folding and ligand binding: a reevaluation of the role of water in biomolecular thermodynamics. *Biophys. Chem.* **115**: 89-97

Cooper A, Johnson CM, Lakey JH, Nollmann M (2001) Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.* **93**: 215-230

Cowan GM, Gann AA and Murray NE (1989) Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell* **56**: 103-109

Crampton N, Yokokawa M, Dryden DTF, Edwardson MJ, Rao DN, Takeyasu K, Yoshimura SH and Henderson MR (2007) Fast-scan atomic force microscopy reveals that the type III restriction enzyme EcoP15I is capable of DNA translocation and looping. *Proc. Natl. Acad. Sci. U. S. A.* **104**: 12755-12760

Crasto CJ and Feng J (2001) Sequence codes for extended conformation: A neighbour dependent sequence analysis of loops in proteins. *Proteins* **42**: 399–413

Crowley PB and Golovin A (2005) Cation-p interactions in protein-protein interfaces. *PROTEINS: Structure, Function and Bioinformatics* **59**: 231-239

Davies GP, Kemp P, Molineux IJ, Murray NE (1999a) The DNA translocation and ATPase activities of restriction-deficient mutants of EcoKI. *J. Mol. Biol.* **292**: 787–796

Davies GP, Martin I, Sturrock SS, Cronshaw A, Murray NE and Dryden DTF (1999b) On the structure and function of type I DNA restriction enzymes. *J. Mol. Biol.* **290**: 565-579

Davies GP, Powell LM, Webb JL, Cooper LP and Murray NE (1998) EcoKI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities. *Nucl. Acids Res.* **26**: 4828-4836

References

Dreier J and Bickle TA (1996a) ATPase activity of the type C restriction-modification system EcoR124II. *J. Mol. Biol.* **257**: 960-969

Dreier J, MacWilliams MP and Bickle TA (1996b) DNA cleavage by the type IC restriction-modification enzyme EcoR124II. *J. Mol. Biol.* **264**: 722-733

Dryden DTF (2006) DNA mimicry by proteins and the control of enzymatic activity on DNA. *Trends Biotechnol.* **24**: 378-382

Dryden DTF, Cooper LP, Murray NE (1993) Purification and characterization of the methyltransferase from the Type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.* **268**: 13228-13236

Dryden DTF, Cooper LP, Thorpe PH, Byron O (1997) The *in vitro* assembly of the EcoKI Type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry* **5**: 1065-1076

Dryden DTF, Murray NE and Rao DN (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucl. Acids Res*. **29**: 3728-3741

Dunn JJ, Elzinga M, Mark KK, Studier FW (1981) Amino acid sequence of the gene 0.3 protein of bacteriophage T7 and nucleotide sequence of its mRNA. *J. Biol. Chem.* **256**: 2579-2585

Ellis DJ, Dryden DTF, Berge T, Edwardson JM and Henderson RM (1999) Direct observation of DNA translocation and cleavage by the EcoKI endonuclease using atomic force microscopy. *Nat. Struct. Biol*. **6**: 15-17

Eskin B and Linn S (1972) The deoxyribonucleic acid modification and restriction enzymes of Escherichia coli. *J. Biol. Chem.* **247**: 6192-6196

Fetrow JS, Palumbo MJ, and Berg G (1997) Patterns, structures, and amino acid frequencies in structural building blocks, a protein secondary structure classification scheme. *Proteins* **27**: 249–271

Firman K and Szczelkun MD (2000) Measuring motion on DNA by the type I restriction endonuclease EcoR124I usingtriplex displacement. *EMBO J.* **19**: 2094-2102

Frisch C, Schreiber G, Johnson CM, Fersht AR (1997) Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation. *J. Mol. Biol.* **267**: 696-706

References

Fuller-Pace FV, Bullas LR, Delius H and Murray NE (1984) Genetic recombination can generate altered restriction specificity. *Proc. Natl. Acad. Sci. USA.* **81**: 6095-6099

Fuller-Pace FV and Murray NE (1986) Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. *Proc. Natl. Acad. Sci. USA.* **83**: 9368-9372

Gallagher K, Sharp K (1998) Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* **75**: 769-776

Gallivan JP and Dougherty DA (1999) Cation-pi interactions in structural biology. *Proc. Natl. Acad. Sci. USA.* **96**: 9459–9464

Garcia LR and Molineux IJ (1999) Translocation and specific cleavage of bacteriophage T7 DNA in vivo by EcoKI. *Proc. Natl. Acad. Sci. USA.* **96**: 12430-12435

Gomez J and Freire E (1995) Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. *J. Mol. Biol.* **252**: 337-350

Gorbalenya AE and Koonin EV (1991) Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett.* **291**: 277-281

Gounaris AD, Perlmann GE (1967) Succinylation of pepsinogen. *J. Biol. Chem.* **242**: 2739-2745

Halford SE and Marko JF (2004) How do site-specific DNA-binding proteins find their targets? *Nucl. Acids Res.* **32**: 3040-3052

Hermanson GT., (2008) *Bioconjugate Techniques* (Second Edition) pp 755-766, Elsevier, Oxford, UK

Hoare DG, Koshland Jr DE (1967) A method for the quantitative modification and estimation of carboxylic acid groups in proteins. *J. Biol. Chem.* **242**: 2447- 2453

Horiuchi K and Zinder ND (1972) Cleavage of bacteriophage f1 DNA by the restriction enzyme of Escherichia coli B. *Proc. Natl. Acad. Sci. USA* **69**: 3220-3224

Iida S, Streiff MB, Bickle TA and Arber W (1987) Two DNA antirestriction systems of bacteriophage P1, darA, and darB: characterization of darA- phages. *Virology* **157**: 156-66

Jackson RM and Russell RB (2000) The serine protease inhibitor canonical loop conformation: examples found in extracellular hydrolases, toxins, cytokines and viral proteins. *J. Mol. Biol.* **296**: 325–334

Janin J (1997) The kinetics of protein-protein recognition. *Proteins Struct. Funct. Genet.* **28**: 153–161

Janin J and Chothia C (1990) The structure of protein-protein recognition sites. *J. Biol. Chem.* **265**: 16027-16030

Janin J, Miller S and Chothia C (1988) Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**: 155-164

Janscak P, Abadjieva A and Firman K (1996) The type I restriction endonuclease R.EcoR124I: over-production and biochemical properties. *J. Mol. Biol.* **257**: 977-991

Janscak P and Bickle TA (2000) DNA supercoiling during ATP-dependent DNA translocation by the type I restriction enzyme EcoAI. *J. Mol. Biol.* **295**: 1089-1099

Janscak P, MacWilliams MP, Sandmeier U, Nagaraja V and Bickle TA (1999a) DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J.* **18**: 2638-2647

Janscak P, Sandmeier U and Bickle TA (1999b) Single amino acid substitutions in the HsdR subunit of the type IB restriction enzyme EcoAI uncouple the DNA translocation and DNA cleavage activities of the enzyme. *Nucl. Acids Res*. **27**: 2638-2643

Janscak P, Sandmeier U, Szczelkun MD, Bickle TA (2001) Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *J. Mol. Biol.* **306**: 417-431

Jelesarov I and Bosshard HR (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J. Mol. Recognit.* **12**: 3-18

Jones S and Thornton J M (1996) Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**: 13–20

Kan NC, Lautenberger JA, Edgell MH and Hutchison CA (1979) The nucleotide sequence recognized by the Escherichia coli K12 restriction and modification enzymes. *J. Mol. Biol.* **130**: 191-209

References

Kelleher JE, Daniel AS and Murray NE (1991).Mutations that confer de novo activity upon a maintenance methyltransferase. *J. Mol. Biol.* **221**: 431-440

Kennaway CK, Obarska-Kosinska A, White JH, Tuszynska I, Cooper LP, Bujnicki JM, Trinick J, Dryden DTF (2009) The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucl. Acids Res.* **37:** 762-770

Kim YI, Levchenko I, Fraczkowska K, Woodruff RV, Sauer RT and Baker TA (2001) Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat. Struct. Biol.* **8**: 230-3

Kim JS, DeGiovanni A, Jancarik J, Adams PD, Yokota H, Kim R and Kim SH (2005) 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

Korn AP and Burnett RM (1991) Distribution and complementarity of hydropathy in multisubunit proteins. *Proteins Struct. Funct. Genet*. **9**: 37-55

Kruger DH, Barcak GJ, Reuter M and Smith HO (1988) EcoRII can be activated to cleave refractory DNA recognition sites. *Nucl. Acids Res*. **16**: 3997-4008

Kruger DH, Schroeder C, Santibanez-Koref M, Reuter M (1989) Avoidance of DNA methylation. A virus-encoded methylase inhibitor and evidence for counter selection of methylase recognition sites in viral genomes. *Cell Biophys.* **15**: 87-95

Kumke, MU, Löhmannsröben HG and Roch T (1995) 'Fluorescence spectroscopy of Polynuclear aromatic compounds in environmental monitoring'. *J. Fluorescence* **5**: 139–153

Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy* (Second Edition), Plenum Press, New York

Lapkouski M, Panjikar S, Janscak P, Smatanova KI, Carey J, Ettrich R and Csefalway E (2009) *Nat. Struct. & Mol. Biol.* **16**: 94-95

Lee JM, Edwards HHL, Pereira CA, Samii SI. (1996) Crosslinking of tissue-derived biomaterials in 1-ethyl-3-(3-dimethylaminoprpyl)-carbodiimide (EDC). *J. of Mat. Sci.: Materials in Medicine* **7**: 531-541

References

Lee K H, Xie D, Freire E and Amzel LM (1994) Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. *Proteins: Struct. Funct. Genet*. **20**: 68-84

Lo Conte L, Chothia C and Janin J (1999) The Atomic Structure of Protein-Protein Recognition Sites. *J. Mol. Biol*. **285**: 2177-2198

Lundblad J, Laurance M and Goodman R (1996) Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol. Endocrinol.* **10**: 607 612

McCoy A J, Epa VC and Colman PM (1997) Electrostatic complementarity at protein-protein interfaces. *J. Mol. Biol.* **268**: 570–584

McMahon SA, Roberts GA, Johnson KA, Cooper LP, Liu H, White JH, Carter LG, Sanghvi B, Oke M, Walknishaw MD, Blakely GW, Naismith JH and Dryden DTF (2009) Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance *Nucl. Acids Res.* **37**: 4887-4897

Makovets S, Doronina VA and Murray NE (1999) 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

Mark KK, Studier FW (1981) Purification of the gene 0.3 protein of bacteriophage T7, an inhibitor of the DNA restriction system of Escherichia coli. *J. Biol. Chem.* **256**: 2573-2578

Meisel A, Bickle TA, Kruger DH and Schroeder C (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature* **355**: 467-469

Meisel A, Mackeldanz P, Bickle TA, Kruger DH, Schroeder C (1995) Type III restriction endonucleases translocate DNA in a reaction driven by recognition site-specific ATP hydrolysis. *EMBO J.* **14**: 2958-2966

Mernagh DR and Kneale GG (1996) High resolution footprinting of a type I methyltransferase reveals a large structural distortion within the DNA recognition site. *Nucl. Acids Res*. **24**: 4853-4858

Meyer EA, Castellano RK and Diederich F (2003) Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem. Int. Edit.* **42**: 1210–1250

Murray NE (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* **64**: 412-434

References

Nagaraja V, Shepherd JC and Bickle TA (1985) A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature* **316**: 371-372

Nakajima N, Ikada Y (1995) Mechanism of amide formation by carbodiimide for bioconjugation in aqueous media. *Bioconjugate Chem.* **6**: 123-130

Neaves KJ, Cooper LP, White JH, Carnally SM, Dryden DTF, Edwardson JM and Henderson RM (2009) Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerisation and DNA looping. *Nucl. Acid Res.* **37**: 2053-2063

Nyvall P, Pedersén M, Kenne L, Gacesa P (2000) Enzyme kinetics and chemical modification of α-1,4-glucan lyase from Gracilariopsis sp. *Phytochem.* **54**: 139–145

Obarska, A, Blundell A, Feder M, Vejsadová S, Sisáková E, Weiserová M, Bujnicki JM and Firman, K (2006) Structural model for the multisubunit Type IC restriction-modification DNA methyltransferase M.EcoR124I in complex with DNA *Nucl. Acid Res.* **7**: 1992-2005

O'Brien R and Haq I (2004) In *Biocalorimetry 2; Applications of Calorimetry in the Biological Sciences,* pp 3-31, John Wiley and Sons Ltd, West Sussex, UK

O'Neill M, Dryden DTF and Murray NE (1998) Localization of a protein-DNA interface by random mutagenesis. *EMBO J.* **17**: 7118-7127

Pedemonte CH and Kaplan JH (1986) Carbodiimide inactivation of Na,K-ATPase. *J. Biol. Chem.* **261**: 3632-3639

Perfetti RB, Anderson CD and Hall L (1976) The chemical modification of Papain with EDC. *Biochemistry* **15**: 1735-1742

Perona JJ, Craik CS and Fletterick RJ (1993) Locating the catalytic water molecule in serine proteases. *Science* **261**: 620 - 622

Phizicky EM and Fields S (1995) Protein-protein interactions: methods for detection and analysis. *Microbiol. Rev.* **59**: 94–123

Pieper U, Brinkman T, Kruger T, Noyer-Weidner and Pingoud A (1997) Characterisation of the interaction between the restriction endonuclease McrBC from E. coli and its cofactor GTP. *J. Mol. Biol.* **272**:190-199

References

Pingoud A and Jeltsch A (1997) Recognition and cleavage of DNA by type-II restriction endonucleases. *Eur. J. Biochem.* **246**: 1-22

Plum, GE and Breslauer KJ (1995) Calorimetry of proteins and nucleic acids. *Curr. Opin. Struct. Biol.* **5**: 682-690

Powell LM, Connolly BA, Dryden DTF (1998a) The DNA binding characteristics of the trimeric EcoKI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting. *J. Mol. Biol.* **283**: 947-961

Powell LM, Dryden DTF and Murray NE (1998b) Sequence-specific DNA binding by EcoKI, a type IA DNA restriction enzyme. *J. Mol. Biol.* **283**: 963-976

Powell LM, Dryden DTF, Willcock DF, Pain RH and Murray NE (1993) DNA recognition by the EcoK methylatransferase: the influence of DNA methylationand the cofactor S-adenosyl-L-methionine. *J. Mol. Biol.* **234**: 60-71

Powell LM and Murray NE (1995) S-adenosyl methionine alters the DNA contacts of the EcoKI methyltransferase. *Nucl. Acids Res.* **23**: 967-974

Putnam DC, Tainer AJ (2005) Protein mimicry of DNA and pathway regulation. *DNA Repair* **4**: 1410-1420

Ringer AL, Senenko A and Sherrill CD (2007) Models of S/π interactions in protein structures: Comparison of the H_2S–benzene complex with PDB data. *Protein Sci.* **16**: 2216–2223

Roberts RJ and Cheng X (1998) Base flipping. *Annu. Rev. Biochem.* **67**: 181-198

Rodier F, Bahadur RP, Chakrabarti P and Janin J (2005) Hydration of protein-protein interfaces. *Proteins* **60**: 36-45

Sambrook and Russell, (2001) *Molecular Cloning, A Laboratory Manual (Third edition)*, pp 1.116-1.118, Cold Spring Harbor Laboratory Press, New York

Sannino A, Pappada S, Madaghiele M, Maffezzoli A, Ambrosio L and Nicolais L (2005) Crosslinking of cellulose derivatives and hyaluronic acid with water-soluble carbodiimide. *Polymer* **46**: 11206–11212

Schägger H, von Jagow G (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199**: 223 231

Schreiber G and Fersht AR (1996) Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.* **3**: 427–431

Seidel R, van Noort J, van der Scheer C, Bloom JGP, Dekker NH, Dutta CF, Blundell A, Robinson T, Firman K and Dekker C (2004) Real-time observation of DNA translocation by the type I restriction modification enzyme EcoR124I. *Nat. Struct. Mol. Biol*. **11**: 838-843

Serfiotis-Mitsa D, Roberts GA, Cooper LP, White JH, Nutley M, Cooper A, Blakely GW, Dryden DTF (2008) The Orf18 gene product from conjugative transposon Tn916 is an ArdA antirestriction protein that inhibits type I DNA restriction-modification systems. *J. Mol. Biol.* **383**: 970-981

Sevenich FW, Langowski JR, Weiss V and Rippe K (1998) DNA binding and oligomerization of NtrC studied by fluorescence anisotropy and fluorescence correlation spectroscopy. *Nucl. Acids Res.* **26**:1373-81

Singleton MR and Wigley DB (2002) Modularity and specialization in superfamily 1 and 2 helicases. *J. Bacteriol.* **184**:1819-1826

Sistla S and Rao DN (2004) S-adenosyl-l-methionine-dependent restriction enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**: 1-19

Smith JD, Arber W and Kuhnlein U (1972) Host specificity of DNA produced by Escherichia coli XIV. The role of nucleotide methylation in *in vivo* B-specific modification. *J. Mol. Biol.* **63**: 1-8

Smyth GD (1967) Acetylation of amino and tyrosine hydroxyl groups. *J. Biol. Chem.* **242**: 1592-1598

Stephanou AS, Roberts GA, Tock MR, Pritchard EH, Turkington R, Nutley M, Cooper A, Dryden DTF (2009a) A mutational analysis of DNA mimicry by ocr, the gene 0.3 antirestriction protein of bacteriophage T7. *Biochem. Biophys. Res. Comm.* **378**: 129-132

Stephanou AS, Roberts GA, Cooper LP, Clarke DJ, Thomson AR, MacKay CL, Nutley M, Cooper A and Dryden DTF (2009b) Dissection of the DNA mimicry of the bacteriophage T7 Ocr protein using chemical modification. *J. Mol. Biol.* **39**: 565-76

Stewart FJ, Panne D, Bickle TA and Raleigh EA (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.* **298**: 611-622

Stites WE (1997) Protein-protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev.* **97**: 1233-1250

Studier FW (1975) Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host. *J. Mol. Biol.* **94**: 283-295

Studier FW and Bandyopadhyay PK (1988) Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl. Acad. Sci. USA* **85**: 4677-4681

Studier FW and Movva NR (1976) SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J. Virol.* **19**: 136-145

Sturrock SS and Dryden DTF (1997) A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res.* **25**: 3408-3414

Suri B and Bickle TA (1985) EcoA: the first member of a new family of type I restriction modification systems. Gene organization and enzymatic activities. *J. Mol. Biol.* **186**: 77-85

Sutherland E, Coe L and Raleigh EA (1992) McrBC: a multi-subunit GTP-dependent restriction endonuclease. *J. Mol. Biol.* **225**: 327-348

Su T-J, Tock MR, Egelhaaf SU, Poon WCK, Dryden DTF (2005) DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucl. Acids Res.* **33**: 3235-3244

Taylor I, Patel J, Firman K and Kneale G (1992) Purification and biochemical characterisation of the EcoR124 type I modification methylase. *Nucl. Acids Res.* **20**: 179-186

Taylor I, Watts D and Kneale GG (1994) DNA binding induces a major structural transition in a type I methyltransferase. *EMBO J.* **13**: 5772-5778

Thelen P and Deuticke B (1988) Chemo-mechanical leak formation in human erythrocytes upon exposure to a water-soluble carbodiimide followed by very mild shear stress. II. Chemical modifications involved. *Biochim. Biophys. Acta.* **944**: 297-307

Thom, B and Wackernagel W (1984) Genetic control of damage-inducible restriction alleviation in Escherichia coli K12: an SOS function not repressed by lexA. *Mol. Gen. Genet.* **197**: 297-303

References

Tock MR, Dryden DTF (2005) The biology of restriction and anti-restriction. *Curr. Op. Microbiol.* **8:** 466-472

Tsai CJ and Nussinov R (1997) Hydrophobic folding units at protein-protein interfaces: implications to protein folding and protein-protein association. *Protein Sci.* **6**: 1426 –1437

Vovis GF, Horiuchi K and Zinder ND (1974) Kinetics of methylation of DNA by a restriction endonuclease from Escherichia coli B. *Proc. Natl. Acad. Sci. USA* **71**: 3810-3813

Walkinshaw MD, Taylor P, Sturrock SS, Atanasiu C, Berge T, Henderson RM, Edwardson JM, Dryden DTF (2002) Structure of ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell* **9**: 187-194

Wang HC, Wang HC, Ko TP, Lee YM, Leu JH, Ho CH, Huang WP, Lo CF and Wang AH (2008) White spot syndrome virus protein ICP11: A histone-binding DNA mimic that disrupts nucleosome assembly. *Proc. Natl. Acad. Sci. USA* **105**: 20758-20763

Warren RA (1980) Modified bases in bacteriophage DNAs. *Annu. Rev. Microbiol.* **34**: 137-158

Webb JL, King G, Ternent D, Titheradge AJ and Murray NE (1996) Restriction by *Eco*KI is enhanced by cooperative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J.* **15**: 2003–2009

Whitmore L, Wallace BA (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucl. Acids Res.* **32**: W668-W673

Wilkins BM (2002) Plasmid promiscuity: meeting the challenge of DNA immigration control. *Env. Microbiol.* **4**: 495-500

Williams A, Ibrahim IT (1981) Carbodiimide chemistry: recent advances. *Chem. Rev.* **81**: 589-636

Wilson GG and Murray NE (1991) Restriction and modification systems. *Annu. Rev. Genet.* **25**: 585-627

Wu-Chou S, Robinson AE, Hrabeta E and Packer L (1984) Cross-linking of bacteriorhodopsin using specific carboxyl modifications and proteolytic cleavage. *Biochem. Biophys. Res. Comm.* **124**: 565-571

Young MA and Beveridge DL (1998) Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.* **281**: 675-687

Yuan R, Hamilton DL and Burckhardt J (1980) DNA translocation by the restriction enzyme from E. coli K. *Cell* **20**: 237-244

Zabeau M, Friedman S, Van Montagu M and Schell J (1980) The *ral* gene of phage lambda. I. Identification of a non-essential gene that modulates restriction and modification in *E. coli. Mol. Gen. Genet.* **179**: 63-73

Zavilgelsky GB, Rastorguev SM (2004) Antirestriction activity of ArdA encoded by the IncI1 transmissive plasmid R64. *Mol. Biol. (Mosc)* **38**: 901-906

Zavilgelsky GB, Kotova VY, Rastorguev SM (2008) Comparative analysis of anti-restriction activities of ArdA (ColIb-P9) and Ocr (T7) proteins. *Biochemistry (Mosc)* **73:** 906-911

Zavilgelsky GB, Kotova VY, Rastorguev SM (2009) Antirestriction and antimodification activities of T7 Ocr: Effects of amino acid substitutions in the interface. *Mol. Cell. Biol. (Mosc)* **43**: 93-100

Zheleznova EEH and Brennan RG (2001) Crystal structure of the transcription activator BmrR bound to DNA and a drug. *Nature* **409**: 378-382

Appendix A – Script for data analysis of the anisotropy competition assay

```
#!/usr/bin/perl

# By Augustinos Stephanou & Joseph Spadavecchia
# April 2007, Edinburgh University
#
# This program is free software; you can redistribute it and/or
modify
# it under the terms of the GNU General Public License as published
by
# the Free Software Foundation; either version 2 of the License, or
# (at your option) any later version.
#
# This program is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE.  See the
# GNU General Public License for more details.
#
# You should have received a copy of the GNU General Public License
along
# with this program; if not, write to the Free Software Foundation,
Inc.,
#  51 Franklin Street, Fifth Floor, Boston, MA 02110-1301 USA.


sub average {
  my $array = $_[0];
  my $count = $_[1];

  my $sum = 0;

  for($i = 0; $i < $count; $i++) {
    $sum += @{$array}[$i];
  }

  return $sum/$count;
}


if ($#ARGV + 1 < 3 || $#ARGV + 1 > 4 ||
    ($#ARGV + 1 == 4 && $ARGV[3] !~ /^[0-9]+$/)) {
  print "mtase-titrations v1.0\n";
  print "\n";
  print "usage: ./mtase-titrations.pl <g-begin> <g-end> <data> [z-
value]\n";
  print "  Arguments:\n";
  print "    <g-begin>  -  required gfactor begin file\n";
  print "    <g-end>    -  required gfactor end file\n";
  print "    <data>     -  required data file\n";
  print "    [z-value]  -  optional integer value for mapping values
[0,z] => 0 (default z=20)\n";
```

```
  exit;
}

$begin_file = $ARGV[0];
$end_file   = $ARGV[1];
$data_file  = $ARGV[2];

$z = $ARGV[3];

if($z eq '') {
  $z = 20;
}

if(!open(BEGIN_FILE, "<" . $begin_file)) {
  print "unabled to open g-begin file: $begin_file\n";
  exit;
}

if(!open(END_FILE,   "<" . $end_file)) {
  print "unabled to open g-end file: $end_file\n";
  exit;
}

if(!open(DATA_FILE,  "<" . $data_file)) {
  print "unabled to open data file: $data_file\n";
  exit;
}

# calculate G_begin

my $sumHV = 0;
my $sumHH = 0;

while(<BEGIN_FILE>) {
  $line = $_;

  if($line !~ /^[0-9]/ ||
     $line =~ /^#/) {
    next;
  }

  $line =~ s/E\+/e/g;

  my ($junk, $hv, $hh) = split(/\s+/, $line);

  $sumHV += $hv;
  $sumHH += $hh;
}

my $G_begin = $sumHV/$sumHH;

print "G_begin = $G_begin\n";


# calculate G_end

$sumHV = 0;
$sumHH = 0;
```

```
while(<END_FILE>) {
  $line = $_;

  if($line !~ /^[0-9]/ ||
     $line =~ /^#/) {
    next;
  }

  $line =~ s/E\+/e/g;

  my ($junk, $hv, $hh) = split(/\s+/, $line);

  $sumHV += $hv;
  $sumHH += $hh;
}

my $G_end = $sumHV/$sumHH;

print "G_end   = $G_end\n";

# calculate G

my $G = ($G_begin + $G_end)/2.0;

print "G       = $G\n";
print "\n";

my $i = 0;
my $values = 0;
my $VV_block = ();
my $VH_block = ();
my $VV_min = -1;
my $VH_min = -1;

while(<DATA_FILE>) {

  $line = $_;

  if($line !~ /^[0-9]/ ||
     $line =~ /^#/) {
    next;
  }

  $line =~ s/E\+/e/g;

  my ($junk, $vv, $vh) = split(/\s+/, $line);

  if($vv <= $z) {
    $vv = 0;
  }

  if($vh <= $z) {
    $vh = 0;
  }

  if($vv != 0.0 && $vh != 0.0) {
    $values++;
```

```
    if($values > 2) {
      @VV_block[$i] = $vv;
      @VH_block[$i] = $vh;

      if($VV_min == -1 || $VV_min > @VV_block[$i]) {
        $VV_min = @VV_block[$i];
      }

      if($VH_min == -1 || $VH_min > @VH_block[$i]) {
        $VH_min = @VH_block[$i];
      }

      $i++;
    }
  } else {
    if($i > 2) {
      my $VV_avg = average(\@VV_block, $i - 2);
      my $VH_avg = average(\@VH_block, $i - 2);

      my $r = ($VV_avg - $G*$VH_avg)/($VV_avg + 2.0*$G*$VH_avg);

      print "$r\n";

      $i = 0;
      $values = 0;
      @VV_block = ();
      @VH_block = ();
    }
  }



}

print "\n";

print "VV_min = $VV_min\n";
print "VH_min = $VH_min\n";
```

Appendix B – Supplementary data

Fluorescence anisotropy competition assay data

Representative raw data from the fluorescence anisotropy competition assay, used to describe the different Ocr:M.EcoKI interactions are shown below. M.EcoKI was titrated into a solution containing 2 nM of a 21 bp DNA duplex fluorescently labelled at the 5' end in the presence (or absence) of 50 nM of an Ocr protein. The binding of M.EcoKI to the DNA duplex could be monitored by anisotropy; a measurement of the rate of tumbling of the fluorophore. Upon binding of M.EcoKI to DNA the DNA-protein complex tumbles more slowly and r increases. By fitting the data to a specific binding model (as described in section 2.4.6) it is possible to calculate the Ocr:M.EcoKI binding affinity. The graphs B1 to B3 show raw data that has not been fitted to a model but rather points have been joined by a line. An example of data fitting to a binding model using Dynafit, is shown in Figure B4, on a linear scale.

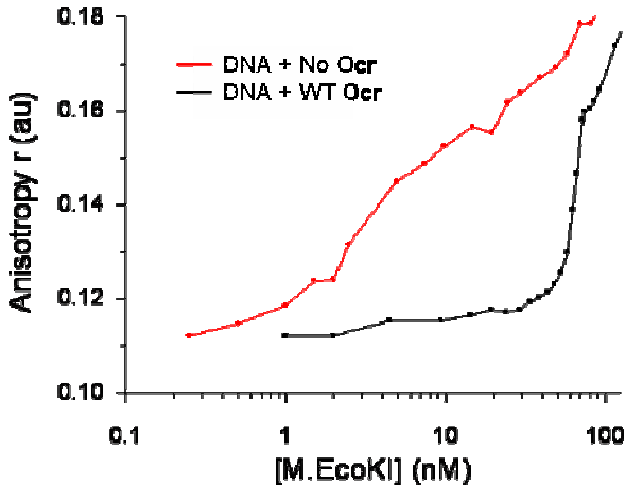


Figure B1. Fluorescence anisotropy measurements of a hex-labelled DNA molecule during a titration of a competition binding assay in the presence and absence of 50 nM wild-type Ocr.

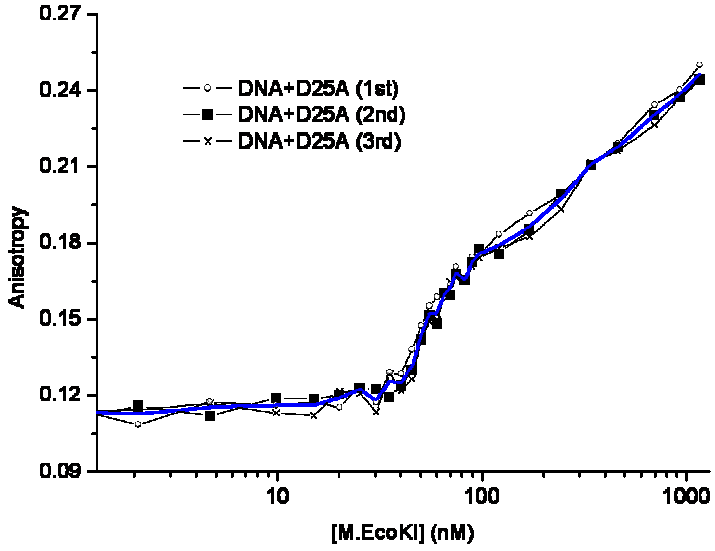


Figure B2. Fluorescence anisotropy measurements of a hex-labelled DNA molecule during titrations of a competition binding assay in the presence of 50 nM of the D25A Ocr protein.

The graph shows three repeats of the titration. The blue line represents the averaged set of data using Origin7.5.

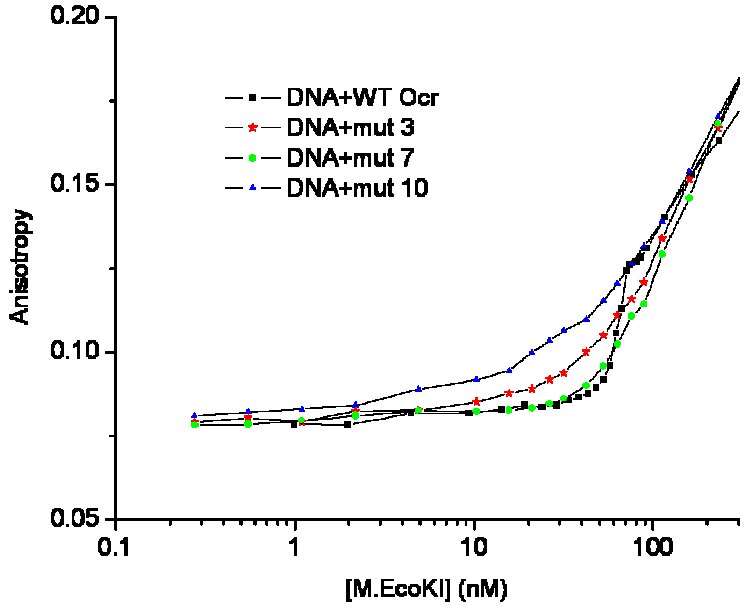


Figure B3. Fluorescence anisotropy measurements of a hex-labelled DNA molecule during different titrations of a competition binding assay in the presence of 50 nM of Ocr (specified in the figure).

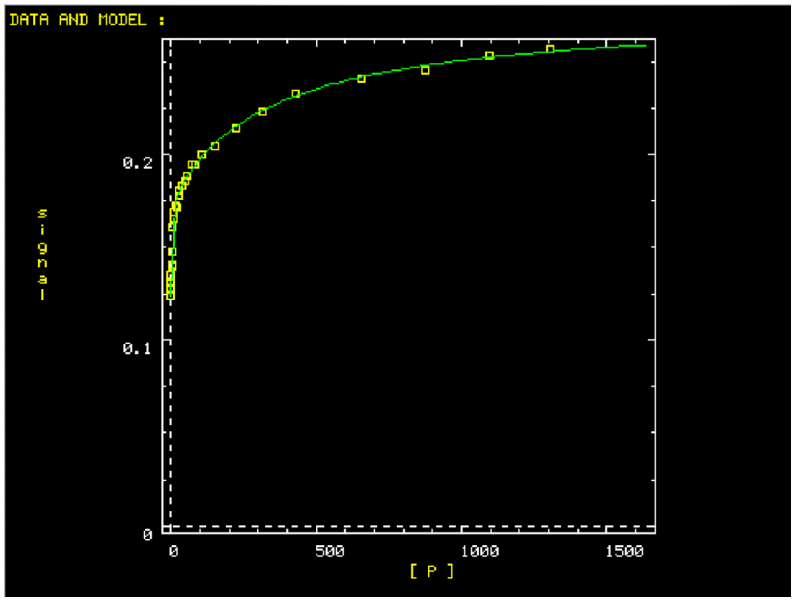


Figure B4. An example of data fitting to a binding model using Dynafit, of a representative set of fluorescence anisotropy data (section 2.4.6), is shown in linear scale.

Differential scanning calorimetry (DSC) data

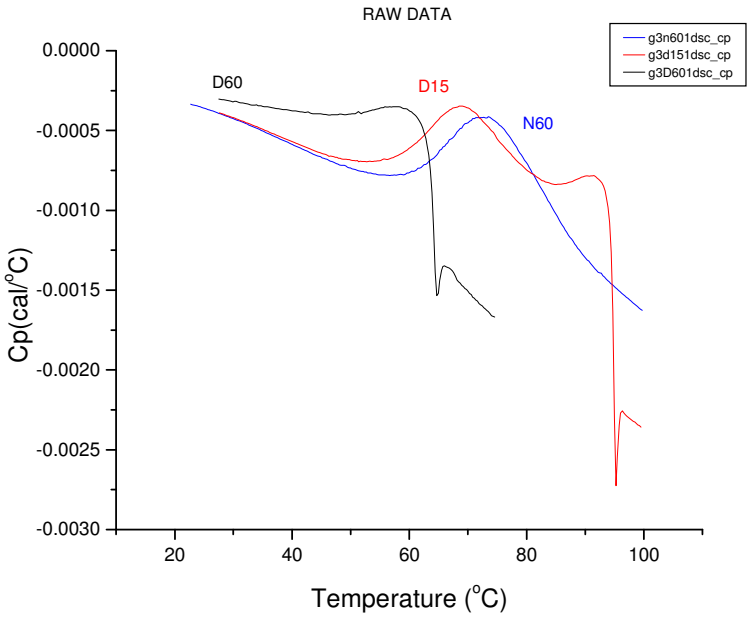


Figure B5. DSC raw data for the chemically modified Ocr proteins (Chapter 4).

Appendix C – Published papers

Biochemical and Biophysical Research Communications 378 (2009) 129–132

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

A mutational analysis of DNA mimicry by ocr, the gene 0.3 antirestriction protein of bacteriophage T7

Augoustinos S. Stephanou, Gareth A. Roberts, Mark R. Tock [1], Emily H. Pritchard [2], Rachel Turkington, Margaret Nutley [3], Alan Cooper [3], David T.F. Dryden *

EaStChem School of Chemistry, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3JJ, UK

ARTICLE INFO

Article history:
Received 4 November 2008
Available online 21 November 2008

Keywords:
DNA mimicry
Restriction–modification systems
EcoKI
Gene 0.3 protein

ABSTRACT

The ocr protein of bacteriophage T7 is a structural and electrostatic mimic of approximately 24 base pairs of double-stranded B-form DNA. As such, it inhibits all Type I restriction and modification (R/M) enzymes by blocking their DNA binding grooves and inactivates them. This allows the infection of the bacterial cell by T7 to proceed unhindered by the action of the R/M defence system. We have mutated aspartate and glutamate residues on the surface of ocr to investigate their contribution to the tight binding between the EcoKI Type I R/M enzyme and ocr. Contrary to expectations, all of the single and double site mutations of ocr constructed were active as anti-R/M proteins *in vivo* and *in vitro* indicating that the mimicry of DNA by ocr is very resistant to change.

The ocr protein encoded by gene 0.3 of bacteriophage T7 [1,2] is the best characterised example of an antirestriction protein and of a structural mimic of DNA [3,4]. Phage- or plasmid-encoded antirestriction proteins alleviate the effects of host restriction endonucleases during invasion of the host by the phage or plasmid DNA [5,6]. Ocr has the same shape and charge distribution as a bent, double-stranded DNA molecule approximately 24 base pairs in length [7,8]. This molecular mimicry accounts for the ability of ocr to inhibit virtually irreversibly, all known Type I restriction and modification (R/M) enzymes found in most eubacteria and archaea [8–12]. Ocr binds to and completely occupies the DNA-binding site on the enzyme and prevents the enzyme from acting as a restriction endonuclease on the phage genome as it enters the bacterium. Ocr thereby greatly assists the spread of the phage infection in the bacterial population. The mimicry by ocr of the general shape and charge of DNA rather than any specific base pair sequence means that ocr acts against any of the vast variety of Type I R/M enzymes, each of which recognises a different base pair sequence [13].

The structure of ocr is an elongated, banana-shaped dimer decorated on its surface with a surfeit of aspartate and glutamate residues which can be superimposed upon the equivalent phosphate groups on the DNA molecule [8]. In addition to mimicking the charge distribution, ocr mimics the bend of approximately 50 in the DNA helical axis which is induced upon DNA binding to the R/M enzyme [8]. The introduction of the bend in DNA by the R/M enzyme is energetically costly and this cost is "saved" when the R/M binds to ocr as ocr is already "pre-bent" [14]. This would account for the ~50-fold tighter binding of the enzyme to ocr than to DNA.

We have explored the effect of replacing aspartate and glutamate residues in the ocr molecule on the interaction between this DNA mimic and Type I R/M systems *in vivo* and *in vitro* in an attempt to assess the importance of electrostatic mimicry of the DNA. Our selection of Asp and Glu residues was guided by the overlay of the structure of ocr on duplex DNA, Fig. 1. The residues chosen mimic the phosphate backbone of one strand of the duplex DNA recognised by the R/M enzyme. We found that the activity of ocr and its binding affinity for the archetypal purified Type I R/M enzyme, EcoKI, were extremely robust to mutations removing one or two acidic side chains from each monomer of ocr.

Materials and methods

Escherichia coli JM109 was purchased from Promega (Madison, WI). *E. coli* NM1261 (r^-m^-, no R/M system) and *E. coli* NM1049 (r^+m^+, EcoKI Type IA R/M system) were a kind gift of Professor Noreen Murray (School of Biology, University of Edinburgh, UK). The expression strain *E. coli* BL21(DE3)pLysS was purchased from Invitrogen (Groningen, The Netherlands). All cell growth was conducted

* Corresponding author. Fax: +44 (0)131 650 6453.
E-mail address: david.dryden@ed.ac.uk (D.T.F. Dryden).
[1] Present address: Liverpool Innovation Park, Baird House, Edge Lane, Liverpool L7 9NJ, UK.
[2] Present address: MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK.
[3] Present address: WestChem Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK.

0006-291X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2008.11.014

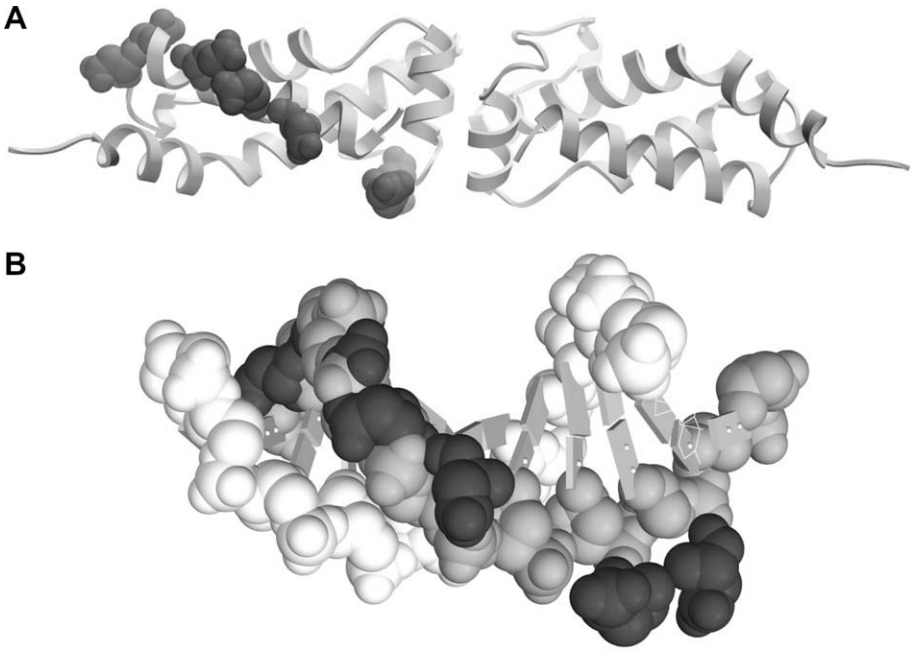


Fig. 1. (A) A ribbon diagram of the ocr dimer showing amino acids 5–110 in each subunit. The C-termini are at the extreme left and right and the N-termini are located at the interface between the two subunits. The amino acids changed by site-directed mutagenesis are shown as space-filling spheres in the left hand subunit. These charged residues run around the surface of the ocr protein. The ocr monomer has pdb code 1s7z. (B) A DNA duplex is shown with the base pairs as slabs and the phosphate backbone as space-filling spheres in light grey and grey. The atoms of the amino acids in ocr targeted for mutagenesis are shown as very dark grey space-filling spheres. It can be seen that they overlap well with the phosphate backbone of the grey strand of DNA [8].

at 37 °C unless stated otherwise. *In vivo* restriction assays using virulent unmodified bacteriophage lambda, $\lambda_{v.o}$, or modified lambda $\lambda_{v.k}$ were performed as described previously [15]. Plasmids pAR3786 and pAR3790 encoding for C-terminal deletions of the last 7 or 17 amino acids of wild-type ocr were a kind gift from Dr. Alan Rosenberg and Professor William Studier (Brookhaven National Laboratory, USA).

Single or double codon mutagenesis was achieved using the QuikChange II Site-Directed Mutagenesis kit from Stratagene (La Jolla, CA) following the manufacturers guidelines. All constructs were sequenced to ensure no mistakes had been introduced during amplification. Edman sequencing of purified ocr [16] indicated that the N-terminal methionine was subject to post-translational cleavage. Therefore for clarity, mutated amino acids in the present work are numbered by taking this into account (i.e. the ocr sequence starts AMSNM).

To prepare the mutant versions of ocr, we adapted the procedure used to purify wild-type ocr [17]. *E. coli* BL21(DE3)pLysS was transformed with the appropriate construct. Cells were grown at 37 °C in 2 L conical flasks containing 1 L LB broth supplemented with 34 μg/ml chloramphenicol and 50 μg/ml of carbenicillin and shaking at ~230 rpm. Once the optical density at 600 nm reached ~0.5, heterologous gene expression was induced by the addition of IPTG (final concentration of 1 mM) and growth continued for a further 2.5 h before harvesting the cells by centrifugation at 8000*g* for 10 min at 4 °C. Cell pellets were stored at −20 °C until required. The cells were resuspended in ice-cold buffer A (20 mM Tris–HCl, 300 mM NH_4Cl, pH 8.0) in the presence of a protease inhibitor cocktail (Roche, Basel, Switzerland). Cells were then broken on ice by sonication using a Soniprep 150 Sonicator (Sanyo, Tokyo, Japan) fitted with a 9-mm probe (1 min/g of cell paste). The cell debris was then removed by centrifugation (20,000*g* for 1 h at 4 °C). The supernatant was loaded onto a 20 × 1.6 cm diameter DEAE-Sepharose fast flow ion-exchange column (GE Healthcare, Piscataway, NJ), which had been pre-equilibrated in buffer A, at a flow rate of 48 ml/h. The column was extensively washed with buffer A and then a 500 ml gradient from 0.3 to 1.0 M NH_4Cl in buffer A was run at 24 ml/h. Fractions containing ocr, identified by SDS–PAGE, were subsequently pooled. UV-spectroscopy indicated that the sample was contaminated with nucleic acid. The nucleic acid was removed in the following manner. The crude ocr preparation was precipitated by the addition of 1.2 volumes of 10% trichloroacetic acid (TCA) and incubated on ice for ~10 min. The precipitate was collected by centrifugation at 20,000*g* for 20 min at 4 °C and the pellet was resuspended in 95% ethanol with gentle mixing for ~10 min. After centrifugation (20,000*g* for 20 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 20 mM Tris–HCl, pH 8.0, instead of 95% ethanol. The sample was then dialysed against ~4 L 20 mM Tris–HCl, pH 8.0, for 16 h at 4 °C and concentrated by centrifugation using a Vivaspin concentrator (10,000 MWCO; VivaScience AG, Hannover, Germany). Finally, an equal volume of glycerol was added to the sample, which was then stored at −20 °C until required. The M.EcoKI methyltransferase was purified as previously described [18,19].

Isothermal titration calorimetry (ITC) was carried out using a VP-ITC instrument (Microcal, Northampton, MA). The stocks of ocr (wild-type or mutated versions) and M.EcoKI were buffer exchanged into 20 mM Tris–HCl 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 adjusted either by dilution into the same buffer or by concentration using a Vivaspin concentrator (10,000 MWCO for ocr, 30,000 MWCO for M.EcoKI; VivaScience AG). *S*-Adenosyl-L-methionine (New England Biolabs, Ipswich, MA, USA) was then added to a final concentration of 100 μM. All solutions were thoroughly degassed prior to use. Typically, ocr at a concentration of 30 μM was titrated into a M.EcoKI solution at a concentration of 3 μM in the VP-ITC cell (1.4 ml active volume). All titrations were carried out at 25 °C. The heat of dilution was obtained by injecting ocr into buffer or

buffer into buffer and these values were subtracted from the ITC titration data. The calorimetric data were converted into differential binding curves by integration of the resultant peaks. The data were fitted with a single-site binding model using the Microcal LLC Origin software package.

Competition between DNA and the wild-type or mutated ocr for binding to M.EcoKI was determined using the fluorescence anisotropy assay described previously [9]. Assays were performed in 20 mM Tris–HCl pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol, 100 μM *S*-adenosyl-L-methionine with the concentration of M.EcoKI varying between 0.25 and 500 nM. The 21 base pair DNA duplex containing the M.EcoKI target sequence and end-labelled with hexachlorofluorescein was present at 2 nM [20]. The ocr dimer, if present, was at a concentration of 50 nM. The anisotropy data were fitted to an equilibrium binding model using Dynafit (Biokin Ltd., Watertown, MA, USA) as described previously [9].

Results

The results of all our in phage infection assays, ITC measurements and fluorescence anisotropy assays are given in Table 1.

Once the mutations had been verified by sequencing of the entire gene on both strands, we tested whether the mutated ocr protein was active *in vivo*. These experiments were conducted at low levels of heterologous gene expression in the absence of induction by IPTG. *E. coli* cells with (NM1049) or without (NM1261) the chromosomal EcoKI R/M system were transformed with either the expression vector (a control) or the vector containing the wild-type ocr gene or the mutated ocr genes. Each of these transformed strains was then challenged with phage lambda $\lambda_{v.o}$ or $\lambda_{v.k}$. As expected, cells transformed with the vector alone showed a strong reduction (six orders of magnitude) in the efficiency of plating of $\lambda_{v.o}$ compared to $\lambda_{v.k}$. Cells transformed with the plasmid expressing the wild-type ocr showed essentially identical numbers of plaques with both $\lambda_{v.o}$ and $\lambda_{v.k}$ showing that ocr had knocked out the R/M system. Assaying the mutated forms of ocr showed that all of the mutants were fully active and indistinguishable within the limits of the errors associated with this sort of assay. There was no significant difference between single amino acid substitutions, double substitutions or the C-terminal deletions of 7 or 17 amino acids.

The mutated proteins all overexpressed and purified in the same way as the wild-type ocr indicating that their structures were all likely to be similar to that of the wild-type protein as already suggested by the *in vivo* assay.

The purified proteins were then tested for binding to the methyltransferase core of EcoKI. The core protein, M.EcoKI, comprises one DNA specificity subunit and two DNA methylation subunits with a total molecular mass of 169 kDa. It has previously been shown that ocr binds extremely tightly to M.EcoKI (K_d ~50 pM) and can easily displace a DNA duplex from the enzyme [9,10,14]. We measured the interaction between the mutated ocr proteins and M.EcoKI using ITC as a direct method, and the fluorescence anisotropy competition assay as an indirect method.

The ITC data showed a highly exothermic interaction between ocr and M.EcoKI with a stoichiometry consistent with one ocr dimer per M.EcoKI. The enthalpy change on binding the mutated ocr varied by ±40% from the −85.8 kJ/mol observed for the interaction of the wild-type ocr with M.EcoKI. The substitutions D25C, E59C and D62C gave rise to the largest differences from the wild-type with D25A being less exothermic and the others more exothermic. It is interesting that removal of a charged residue can give both exothermic and endothermic effects. The ITC data indicate tight binding (K_d <10 nM), as evidenced from the sharp end point of the titration (data not shown), and it was not possible to obtain a reliable measure of the free energy change for the interaction from ITC.

To attempt to estimate whether the free energy change for binding was altered by the amino acid substitutions or the C-terminal deletions, we used a fluorescence anisotropy competition assay as previously described [9]. This assay is more appropriate than ITC in this K_d range. A control titration of a labelled DNA oligonucleotide duplex with M.EcoKI showed a typical binding behaviour with a K_d of 2.106 ± 0.350 nM similar to previous measurements [14,20]. A weaker non-specific binding event was observed at high concentrations of M.EcoKI with a K_d of 349 ± 73 nM. The addition of 50 nM ocr dimer to the solution containing the duplex reduces the proportion of M.EcoKI available to bind to the duplex so that more M.EcoKI has to be added to observe DNA binding. The binding of wild-type ocr to M.EcoKI is much stronger than the binding of DNA to M.EcoKI so no change in anisotropy is observed until al-

Table 1
A list of the mutant forms of the ocr protein created by mutagenesis and their behaviour *in vivo* and *in vitro*.

Protein	Eop[1] with unmodified phage $\lambda_{v.0}$	Eop with modified phage $\lambda_{v.k}$	Enthalpy[2] of binding of ocr to M.EcoKI (kJ/mol)	Stoichiometry[2] of binding, ocr:M.EcoKI	Dissociation constant (nM) for ocr binding to M.EcoKI determined by anisotropy
Plasmid vector alone	1.703×10^{-6}	0.973			
Wild-type ocr	0.792	1.493	−85.8 ± 1.3	0.87 ± 0.01	0.044 ± 0.040
D12N	0.912	1.039	−59.0 ± 1.3	0.94 ± 0.01	0.081 ± 0.044
E16A	2.504	2.985	−92.5 ± 1.3	0.82 ± 0.01	0.165 ± 0.150
E16Q			−72.0 ± 0.4	0.70 ± 0.00	0.129 ± 0.037
E20A	0.681	0.996	−70.7 ± 1.7	0.85 ± 0.01	0.102 ± 0.054
D25A			−59.9 ± 1.3	0.96 ± 0.01	0.208 ± 0.036
D25C	0.541	0.502	−102.6 ± 2.1	0.78 ± 0.01	
D26N	0.821	0.761	−91.2 ± 1.3	0.76 ± 0.01	0.158 ± 0.080
E59C	1.135	0.634	−107.6 ± 0.8	0.86 ± 0.00	
D62C	1.056	1.018	−122.6 ± 2.5	0.78 ± 0.01	
D12A, D26N	0.780	0.986	−72.0 ± 1.3	1.00 ± 0.01	0.144 ± 0.044
D12N, D26N	1.506	1.317	−72.0 ± 0.8	0.89 ± 0.01	0.101 ± 0.043
D12N, E87D	0.982	1.008	−63.6 ± 1.3	0.98 ± 0.01	0.063 ± 0.037
E16A, D25A	0.728	0.833	−85.8 ± 0.8	0.84 ± 0.01	0.051 ± 0.056
Ocr deletion, 109 amino acids long	0.603	1.395	−104.6 ± 2.5	0.74 ± 0.01	0.111 ± 0.030
Ocr deletion, 99 amino acids long	1.355	1.022	−93.8 ± 2.9	0.71 ± 0.01	0.098 ± 0.051

[1] Eop, efficiency of plating from phage assays showing the ratio of plaques in NM1049–NM1261, with each of the two phages.
[2] The errors quoted are from the ITC analysis programme but we estimate that the real experimental errors are up to ±20% taking into account uncertainties in the determination of protein and DNA concentrations using UV absorption of around 5%.

most all of the ocr has been bound and excess M.EcoKI is added. The experiment involves three equilibria; two for binding of M.EcoKI to DNA and one for M.EcoKI binding to ocr. By keeping the binding affinities fixed in the analysis procedure for the DNA–EcoKI interactions, one can calculate the binding affinity of M.EcoKI for ocr. The accuracy of this method is not high because the affinity for ocr is so much greater than for DNA. Nevertheless, this methodology sets an upper limit for the dissociation constant, K_d. For all of the mutant proteins tested, tight binding between ocr and M.EcoKI was observed with little DNA binding being observed until the M.EcoKI had bound most of the available ocr. Given a K_d of 2.106 nM for the M.EcoKI–DNA interaction, we find that the mutant forms of ocr have K_d values ranging from roughly equal to the wild-type ocr to 4.7 times larger than wild-type ocr for binding to M.EcoKI. The mutant ocr proteins with substitutions of amino acids E16, D25 or D26 appear to have the weakest interactions with M.EcoKI although the magnitude of the effect is not large.

Discussion

Here, we have focussed our attention on a stretch of amino acids of ocr that mimic a single strand of the DNA recognised by the EcoKI R/M enzyme. Mutants were generated containing one or two amino acid changes of negatively-charged side chains to a neutral side chain per ocr monomer. These changes apparently do not perturb the ocr structure but we had postulated that they would seriously weaken the interaction with M.EcoKI due to electrostatic and structural mimicry of DNA by the ocr molecule. However, our results clearly show that this postulate is wrong as all of mutants were still active *in vivo* and bound strongly to M.EcoKI *in vitro*.

The binding to M.EcoKI is virtually unaffected by any of the amino acid substitutions, at least within the sensitivity limits of the ITC and fluorescence experiments. Only changes at E16, D25 or D26 appeared to slightly weaken binding but the effect was only marginally greater than the error in the measurements. The enthalpy of binding changed in an unpredictable manner with some mutants showing more exothermic binding and some less exothermic binding to M.EcoKI than the wild-type ocr. The enthalpy change varies by about a factor of two between −112.6 kJ/mol and −59.0 kJ/mol and the dissociation constant varies between 0.044 nM and 0.208 nM, a fourfold variation. This variation in K_d only amounts to free energy changes between −59.1 kJ/mol and −55.2 kJ/mol, a rather trivial variation. Given the small changes in the free energy of binding between all of our variants of ocr and M.EcoKI, it is apparent that changes in the enthalpy of binding are being largely compensated by changes in entropy.

There are three possible reasons for our observations: (i) the mutations are not in the ocr:M.EcoKI interface, (ii) the loss of a potential electrostatic interaction at the ocr:M.EcoKI interface is compensated by the proximity of other electrostatic interactions, or (iii) the ocr:M.EcoKI complex is based upon so many electrostatic interactions that it is largely unaffected by the removal of only a few interactions. Given that M.EcoKI is known to wrap around ocr with a resultant large interfacial area [10], the first reason is improbable hence we favour a combination of local (reason ii) and global effects (reason iii) as the explanation for our results. Thus, mutagenesis on a much larger scale than attempted in this paper is required to fully understand the DNA mimicry exhibited by ocr. Based on the three-dimensional structure of ocr, it should be possible to mutate multiple amino acid residues that correspond to entire regions of negative charge on the surface of the protein. In this way, we may be able to deduce regions of the protein that are particularly important for the interaction with M.EcoKI. This work is currently underway. The mimicry of DNA seems to be very robust from a structural viewpoint. However, it is also robust from an evolutionary perspective. Once a mimic has been evolved, it would appear to be resistant to incremental changes in its ability to function.

Acknowledgments

This work was supported by grants from the Wellcome Trust (GR080463MA) and the Biotechnology and Biological Research Council (15/17860).

References

[1] F.W. Studier, Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host, J. Mol. Biol. 94 (1975) 283–295.
[2] D.H. Kruger, C. Schroeder, Bacteriophage T3 and bacteriophage T7 virus-host cell interactions, Microbiol. Rev. 45 (1981) 9–51.
[3] D.T.F. Dryden, DNA mimicry by proteins and the control of enzymatic activity on DNA, Trends Biotechnol. 24 (2006) 378–382.
[4] D.T.F. Dryden, M.R. Tock, DNA mimicry by proteins, Biochem. Soc. Trans. 34 (2006) 317–319.
[5] T.A. Bickle, D.H. Kruger, Biology of DNA restriction, Microbiol. Rev. 57 (1993) 434–450.
[6] M.R. Tock, D.T.F. Dryden, The biology of restriction and anti-restriction, Curr. Opin. Microbiol. 8 (2005) 466–472.
[7] J.J. Blackstock, S.U. Egelhaaf, C. Atanasiu, D.T.F. Dryden, W.C.K. Poon, Shape of ocr, the gene 0.3 protein of bacteriophage T7: modelling base on light scattering experiments, Biochemistry 40 (2001) 9944–9949.
[8] M.D. Walkinshaw, P. Taylor, S.S. Sturrock, C. Atanasiu, T. Berge, R.M. Henderson, J.M. Edwardson, D.T.F. Dryden Structure of Ocr from bacteriophage T7. A protein that mimics B-form DNA, Mol. Cell 9 (2002) 187–194.
[9] C. Atanasiu, O. Byron, H. McMiken, S.S. Sturrock, D.T.F. Dryden, Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7, Nucleic Acids Res. 29 (2001) 3059–3068.
[10] C. Atanasiu, T.-J. Su, S.S. Sturrock, D.T.F. Dryden, Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme, Nucleic Acids Res. 30 (2002) 3936–3944.
[11] P.K. Bandyopadhyay, F.W. Studier, D.L. Hamilton, R. Yuan, Inhibition of the Type I restriction–modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7, J. Mol. Biol. 182 (1985) 567–578.
[12] G.B. Zavilgelsky, V.Y. Kotova, S.M. Rastorguev, Comparative analysis of anti-restriction activities of ArdA (Collb-P9) and Ocr (T7) proteins, Biochemistry (Mosc) 73 (2008) 906–911.
[13] N.E. Murray, Type I restriction systems: sophisticated molecular machines: a legacy of Bertani and Weigle, Microbiol. Mol. Biol. Rev. 64 (2000) 412–434.
[14] T.-J. Su, M.R. Tock, S.U. Egelhaaf, W.C.K. Poon, D.T.F. Dryden, DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping, Nucleic Acids Res. 33 (2005) 3235–3244.
[15] D. Serfiotis-Mitsa, G.A. Roberts, L.P. Cooper, J.H. White, M. Nutley, A. Cooper, G.W. Blakely, D.T.F. Dryden, The Orf18 gene product from conjugative transposon Tn916 is an ArdA antirestriction protein that inhibits type I DNA restriction–modification systems, J. Mol. Biol. 383 (2008) 970–981.
[16] J.J. Dunn, M. Elzinga, K.-K. Mark, F.W. Studier, Amino acid sequence of the gene 0.3 protein of bacteriophage T7 and nucleotide sequence of it mRNA, J. Biol. Chem. 256 (1981) 2579–2585.
[17] K.-K. Mark, F.W. Studier, Purification of the gene 0.3 protein of bacteriophage T7, an inhibitor of the DNA restriction system of Escherichia coli, J. Biol. Chem. 256 (1981) 2573–2578.
[18] D.T.F. Dryden, L.P. Cooper, N.E. Murray, Purification and characterization of the methyltransferase from the Type I restriction and modification system of *Escherichia coli* K12, J. Biol. Chem. 268 (1993) 13228–13236.
[19] D.T.F. Dryden, L.P. Cooper, P.H. Thorpe, O. Byron, The *in vitro* assembly of the EcoKI Type I DNA restriction/modification enzyme and its *in vivo* implications, Biochemistry 5 (1997) 1065–1076.
[20] L.M. Powell, B.A. Connolly, D.T.F. Dryden, The DNA binding characteristics of the trimeric EcoKI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting, J. Mol. Biol. 283 (1998) 947–961.

doi:10.1016/j.jmb.2009.06.020 J. Mol. Biol. (2009) 391, 565–576

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Dissection of the DNA Mimicry of the Bacteriophage T7 Ocr Protein using Chemical Modification

Augoustinos S. Stephanou[1], Gareth A. Roberts[1], Laurie P. Cooper[1], David J. Clarke[1], Andrew R. Thomson[1], C. Logan MacKay[1], Margaret Nutley[2], Alan Cooper[2] and David T. F. Dryden[1]*

[1]*EastChem School of Chemistry, University of Edinburgh, Edinburgh EH9 3JJ, UK*

[2]*West Chem Department of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK*

Received 13 May 2009; received in revised form 4 June 2009; accepted 5 June 2009
Available online 10 June 2009

The homodimeric Ocr (*o*vercome *c*lassical *r*estriction) protein of bacteriophage T7 is a molecular mimic of double-stranded DNA and a highly effective competitive inhibitor of the bacterial type I restriction/modification system. The surface of Ocr is replete with acidic residues that mimic the phosphate backbone of DNA. In addition, Ocr also mimics the overall dimensions of a bent 24-bp DNA molecule. In this study, we attempted to delineate these two mechanisms of DNA mimicry by chemically modifying the negative charges on the Ocr surface. Our analysis reveals that removal of about 46% of the carboxylate groups per Ocr monomer results in an ~50-fold reduction in binding affinity for a methyltransferase from a model type I restriction/modification system. The reduced affinity between Ocr with this degree of modification and the methyltransferase is comparable with the affinity of DNA for the methyltransferase. Additional modification to remove ~86% of the carboxylate groups further reduces its binding affinity, although the modified Ocr still binds to the methyltransferase *via* a mechanism attributable to the shape mimicry of a bent DNA molecule. Our results show that the electrostatic mimicry of Ocr increases the binding affinity for its target enzyme by up to ~800-fold.

Keywords: DNA mimic; chemical modification; restriction/modification system

Edited by B. Connolly

Introduction

The first protein to be produced during infection of *Escherichia coli* by bacteriophage T7 is *o*vercome *c*lassical *r*estriction (Ocr), the product of gene 0.3.[1] The Ocr protein is the best characterised example of an antirestriction protein and of a structural mimic of DNA.[2–4] Ocr is a highly negatively charged protein (p*I* of 4.02) with a shape similar to that of a bent double-stranded DNA molecule approximately 24 bp in length (Fig. 1).[5–13] This molecular mimicry accounts for the ability of Ocr to inhibit, virtually irreversibly, all type I restriction/modification (R/M) enzymes found in the majority of eubacteria and archaea.[3,5,14] Ocr binds to and completely occupies the DNA-binding site on the enzyme, thereby inhibiting the restriction endonuclease activity and protecting the phage genome as it enters the bacterium. Thus, Ocr greatly assists in the spread of phage infection in the bacterial population. The fact that Ocr mimics the general shape and charge of DNA, rather than any specific base pair sequence, means that the protein can act against all type I R/M enzymes, each of which recognises a different defined base pair sequence.[3,14] The type I R/M enzymes are complex oligomeric multifunc-

**Corresponding author.* E-mail address: david.dryden@ed.ac.uk.

Current address: A. R. Thomson, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.

Abbreviations used: Ocr, overcome classical restriction; R/M, restriction/modification; EDC, 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride; HOBt, hydroxybenzotriazole; MS, mass spectrometry; MALDI-TOF, matrix-assisted laser desorption/ionization time of flight; FT-ICR, Fourier transform ion cyclotron resonance; GdmCl, guanidinium hydrochloride; SAM, S-adenosyl-L-methionine; ITC, isothermal titration calorimetry; WT, wild type.

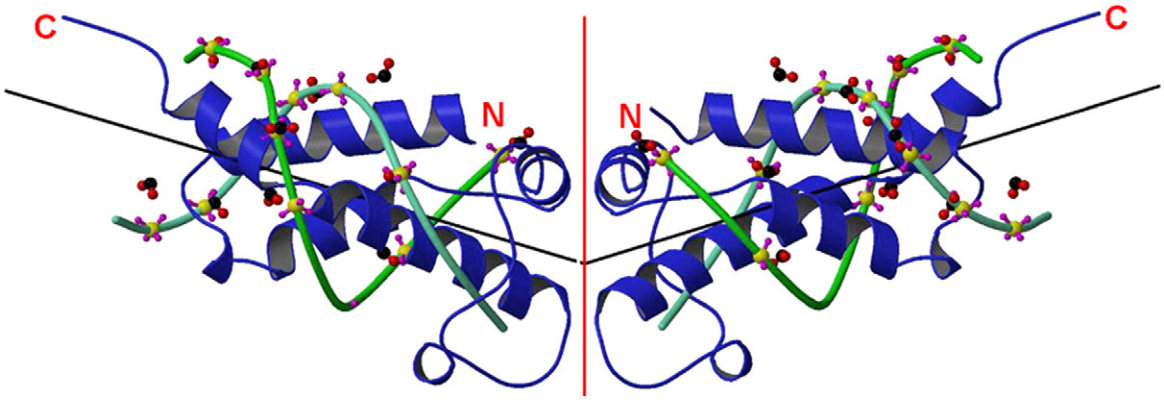


Fig. 1. Superposition of two 12-bp B-DNA molecules on the ocr dimer.[5] Ocr is shown in blue ribbon form, with the N- and C-termini indicated and the dimer interface shown as a red line. A fit of phosphate groups of a B-DNA complex onto 12 carboxyl groups of ocr gave an rms fit of 1.9 Å. Phosphate groups are shown in yellow (phosphorus) and purple (oxygen). The carboxyl groups are shown in red (oxygen) and black (carbon). The sugar backbones of the DNA chains are shown in two shades of green with the base pairs omitted for clarity. Vectors for the DNA helical axes are drawn as black lines.

tional enzymes comprising an S subunit for DNA sequence recognition, two M subunits for DNA methyltransferase activity and two R subunits for ATP-hydrolysis-dependent DNA translocation and restriction endonuclease activities. The M_2S_1 complex can function as a sequence-specific methyltransferase, and the $R_2M_2S_1$ complex can switch between methyltransferase and restriction endonuclease activities. The restriction endonuclease only operates on DNA containing unmethylated recognition sequences typically found on invading foreign phage DNA.

The structure of Ocr is an elongated, curved dimer. Each 116-amino-acid monomer is decorated with 34 surface-exposed Asp and Glu residues but only 2 Lys and 4 Arg residues.[5,6] Asp and Glu residues are the only amino acids with side chains possessing multiple hydrogen-bond acceptors that are geometrically similar to oxygen atoms in a phosphate group. Many of the negatively charged residues of Ocr can be superimposed upon the equivalent phosphate groups on the DNA molecule recognised by the R/M enzyme. In addition to mimicking the charge distribution, Ocr mimics the bend of approximately 46° in the DNA helical axis induced in DNA when it binds to the R/M enzyme.[5] The introduction of the bend in the DNA by the R/M enzyme is energetically costly. This cost is "saved" when the R/M enzyme binds to Ocr because Ocr is already "pre-bent."[15] By combining electrostatic mimicry of DNA and mimicry of the bent shape, binding of the R/M enzyme to Ocr is energetically more favourable than binding to DNA. Thus, the overall binding affinity of the R/M enzyme for Ocr is 50-fold greater than that for DNA with a K_A of $\sim 2\times10^{10}$ M^{-1}.[8,9,11,13,16]

In this study, we explored the effect of chemically modifying the acidic residues in the Ocr molecule on its interaction with the archetypal EcoKI type I R/M system from *E. coli* K12 and assessed the relative importance of the electrostatic mimicry *versus* the mimicry of the bent shape of the DNA.

Results

Extent of modification

We used 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (EDC), a water-soluble carbodiimide, to specifically modify the carboxyl groups of Asp and Glu residues and the C-terminus of Ocr (Fig. 2). The physical and chemical properties of EDC have been extensively studied.[17–20] In aqueous solution under acidic conditions, EDC is anticipated to react with Asp, Glu, Cys and Tyr residues. The absence of a Cys residue in Ocr rules out any unwanted side reactions with sulfhydryl groups. Regeneration of unsubstituted tyrosyl residues, the phenolic hydroxyl group of which reacts with EDC to form a relatively stable *O*-arylisourea, is achieved by treating the protein with hydroxylamine.[21,22] EDC reacts with surface-exposed carboxyl groups to form highly reactive *O*-acylisourea intermediates (Fig. 3, reaction 1) that are susceptible to nucleophilic attack. The reactions were done in the presence of a large molar excess of dimethylamine (i.e., D-series) or ammonium hydroxide (i.e., N-series) as a nucleophile. The end result is the formation of a stable amide bond with concomitant loss of one negative charge for each residue modified (Fig. 3). The *O*-acylisourea intermediate is

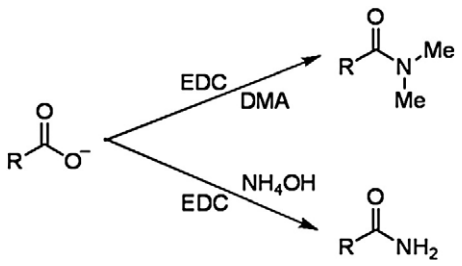


Fig. 2. Derivatisation of carboxylate side chains of Ocr (Asp or Glu residues) using either dimethylamine (D-series) or ammonium hydroxide (N-series) as a nucleophile.

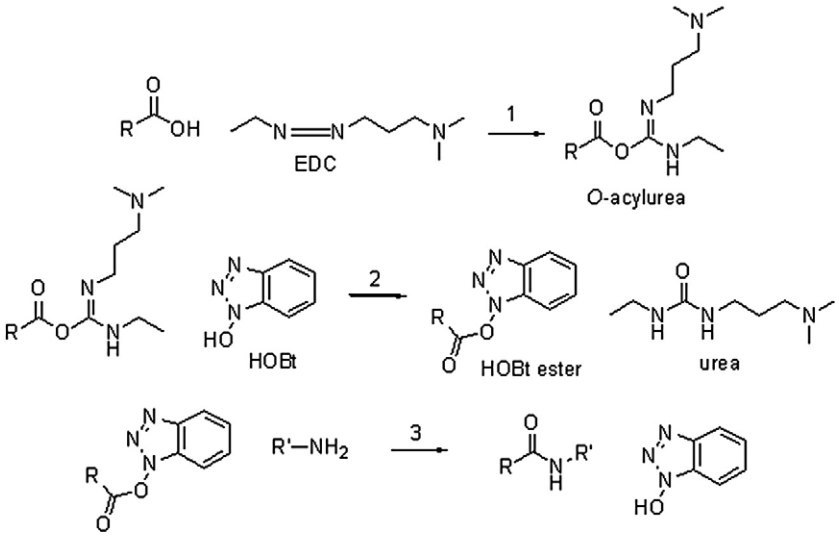


Fig. 3. Reaction for the coupling of the carboxylate groups of the protein with an amine in the presence of EDC and HOBt.

however unstable and can either react with water, regenerating the initial carboxylate and hydrolysing EDC to its urea derivative, or rearrange to an *N*-acylurea, thus forming a stable adduct on the protein. This undesirable *N*-acylurea rearrangement is largely avoided by carrying out the reaction in the presence of (i) a high concentration of nucleophile and (ii) *N*-hydroxybenzotriazole (HOBt). HOBt reacts with the *O*-acylisourea to form a more stable activated ester (Fig. 3, reaction 2), greatly increasing the overall coupling efficiency (Fig. 3, reaction 3) and preferentially mediating the reaction with an amine.[23] This can however include the ε- and α-amino groups of Lys residues and the protein N-terminus, respectively, leading to intramolecular cross-linking and possibly protein polymerisation. Nevertheless, HOBt minimises the adventitious formation of ester-bond cross-links because the HOBt ester is less susceptible than the *O*-acylisourea to nucleophilic attack by OH groups (i.e., Tyr, Ser and Thr residues). An additional mechanism has been reported by Nakajima and Ikada, whereby the *O*-acylisourea intermediate may react with a neighbouring free carboxylate to form an acid anhydride.[20] This labile anhydride, which is highly susceptible to hydrolysis, can react with an amine and a hydroxyl to form an amide bond and an ester bond, respectively.[24] It is probable that the modification reaction of surface-exposed carboxyl groups on Ocr occurs through a combination of the aforementioned species (*O*-acylisourea, HOBt ester and acid anhydride). Considering the close proximity of side chains within a protein and the variability of their pK_a values, depending on their specific microenvironment, it is clear that a limited number of side reactions are unavoidable.

In order to ascertain the degree of modification, we initially analysed the various protein samples by polyacrylamide gel electrophoresis (PAGE) under non-denaturing conditions (Fig. 4). The longer the

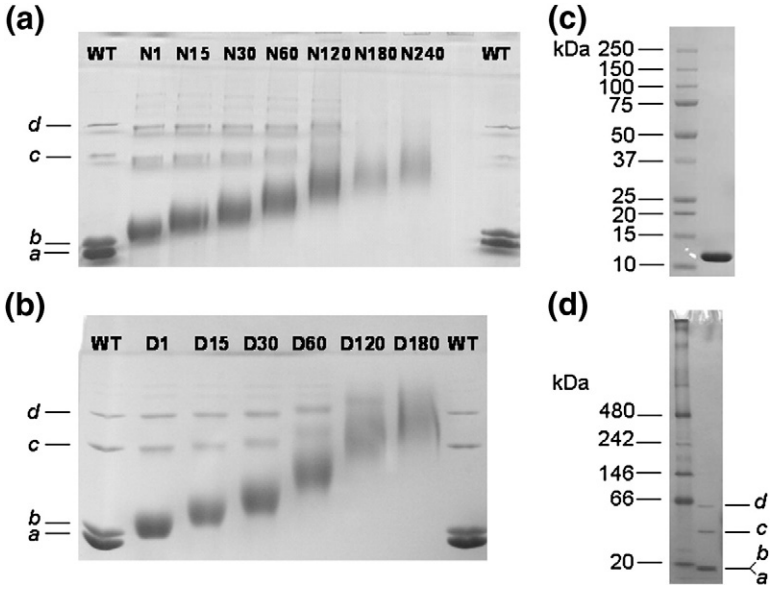


Fig. 4. Gel electrophoretic analysis of chemically modified Ocr using either ammonium hydroxide (N-series) or dimethylamine (D-series) as a nucleophile. Panels (a) and (b) show native 15% PAGE of the N- and D-series of chemically modified Ocr samples, respectively. Gels were run under non-denaturing reducing conditions. "WT" refers to the WT unmodified Ocr. The name given to each modified protein sample reflects the reaction time in minutes. (c) SDS-PAGE, using NuPAGE 4%–12% Bis–Tris gel, of the unmodified Ocr sample [same sample as shown in lanes WT in panels (a) and (b)]. (d) Unmodified Ocr analysed by blue native gel electrophoresis using the NativePAGE Novex® 4%–16% Bis–Tris gel system. The protein markers were supplied by Invitrogen. The Ocr sample was resolved into four discrete bands labelled *a*–*d*.

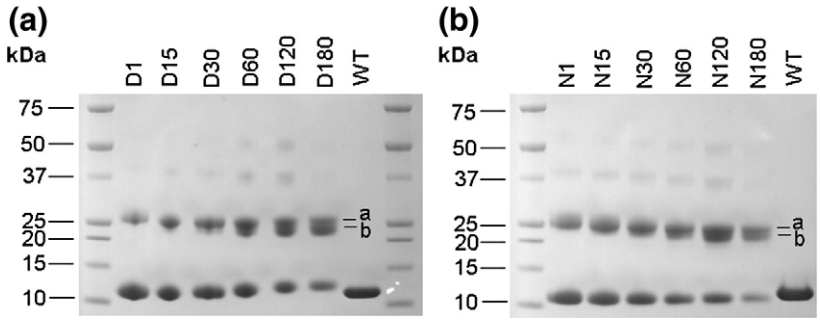


Fig. 5. SDS-PAGE analysis of chemically modified Ocr using 15% acrylamide gel. (a) Analysis of the Ocr samples using dimethylamine as a nucleophile (D-series). The numbers for each sample correspond to the reaction time in minutes. (b) Analysis of the Ocr samples using ammonium hydroxide as a nucleophile (N-series). The numbers for each sample correspond to the reaction time in minutes. In both panels (a) and (b), the modified Ocr runs as two main bands: upper and lower bands corresponding to Ocr dimer and monomer, respectively. Note that the upper bands for the highly modified protein samples from the D- and N-series appear to resolve into two separate species (labelled *a* and *b*).

modification reaction proceeded, the slower was the rate of migration through the gel, indicating progressively greater loss of negative charge. The modified samples gave protein bands that were more diffuse than the unmodified Ocr, which reflects the heterogeneous nature of the reaction products (Fig. 4a and b). As anticipated, the D-series of protein samples showed a greater change in migration rate since dimethylamine, being a better nucleophile, generated a greater level of modification than ammonium hydroxide after an equivalent reaction time. This is more apparent in the chromatographic analyses presented below.

Native gel electrophoresis of the unmodified Ocr revealed four sharp bands (Fig. 4a and b, bands *a*–*d*), even though the same sample analysed by denaturing SDS-PAGE migrated as a single species (Fig. 4c). We also analysed the unmodified Ocr on a NativePAGE Bis–Tris Gel System (Invitrogen) based on the blue native PAGE technique originally developed by Schägger and von Jagow.[25] Blue native PAGE confirmed that bands *c* and *d* correspond to Ocr dimer and tetramer, respectively (Fig. 4d). As Ocr is stable in solution, we do not know precisely why the Ocr monomer (bands *a* and *b*) runs as two distinct bands, but this may be the result of partial denaturation during the running of the gel, thereby causing the protein to adopt a different conformation.

The chemically modified proteins were also analysed by SDS-PAGE, which confirmed the presence of cross-linked dimers (Fig. 5) most probably involving covalent-bond formation between Lys75 and a carboxylate on either side of the dimer interface. The formation of multimers by cross-linking of independent dimers was minimised by conducting the modification reaction at a low protein concentration (3 μM). For both the N- and D-series of chemical modifications, the percentage of cross-linked species increased with increasing reaction time. The upper band, corresponding to dimer, was noticeably broader and more diffuse than the lower monomer band. Indeed, for the highly modified samples, the upper band is resolved into two distinct species. This is probably caused by the incomplete unfolding of the more highly cross-linked dimers.

Additional confirmation of chemical modification and loss of charge on the Ocr dimer was obtained by anion-exchange chromatography using a Mono Q anion-exchange column. Native Ocr binds very strongly to such media and requires high concentrations of NaCl to be eluted. In this experiment, Ocr eluted as a narrow peak late in the NaCl gradient (Fig. 6). The chemically modified samples eluted at progressively lower NaCl concentrations as the modification reaction time was increased. Moreover, the eluted material came off in a broad asymmetric peak, indicating considerable heterogeneity in each

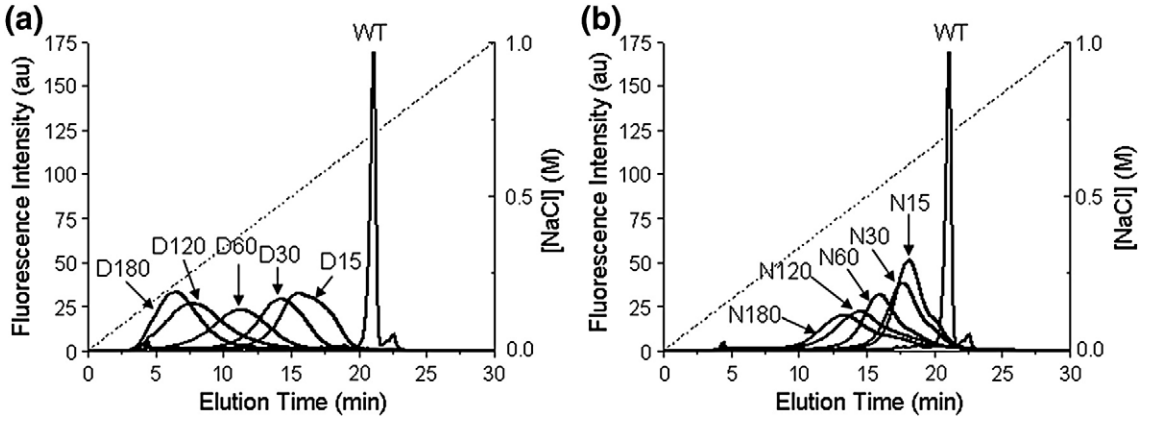


Fig. 6. Anion-exchange chromatography profiles of Ocr and its chemically modified derivatives. Panels (a) and (b) show the analysis of the D- and N-series of chemically modified Ocr on a 1-ml Mono Q column, respectively.

sample due to different numbers and locations of modified residues. Noteworthy is the asymmetric shape of the highly derivatised sample D180, the leading edge of which eluted from the column more abruptly than the other samples. We believe that the initial portion of the elution profile of D180 represents an endpoint in the chemical modification reaction in which almost all available Glu and Asp residues have been amidated.

Although the gel electrophoresis and ion-exchange experiments indicated that the protein was being derivatised, the number of modifications could not be quantified. Mass spectrometry (MS) of the D-series of modified samples was performed to obtain this information as each modification will increase the mass by 27 Da (Δmass +27 Da). Unfortunately, MS is unsuitable for analysis of the N-series because there is only a net reduction in mass of 1 Da after each successive modification.

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS of the D-series of modified samples was performed. Each spectrum showed a broad, slightly asymmetric peak skewed toward greater mass. Nevertheless, we were able to estimate the number of modifications for each sample as follows: D15, 9–18 modifications; D60, 16–25 modifications; D180, 22–33 modifications (data not shown). A smaller peak corresponding to Ocr dimer was also observed.

D-series samples were also analysed on a Fourier transform ion cyclotron resonance (FT-ICR) instrument to obtain higher resolution. The proteins were desalted *via* reversed-phase HPLC, and electrospray MS data were collected for the peak eluting from the HPLC column. A nested set of MS species was observed with a mass difference of 27 Da, corresponding to the anticipated increase in mass after each modification with dimethylamine (Fig. 7). From these data, we were able to ascertain the number of amidations per Ocr monomer as being 10–21 modifications for D15, 15–25 modifications for D60 and 22–31 modifications for D180. Surprisingly, the nested set of peaks, corresponding to each successive modification, displayed a bimodal distribution that may arise from perturbations in the pK_a values of the carboxylates. This would influence reaction rates with the doubly protonated EDC molecule.

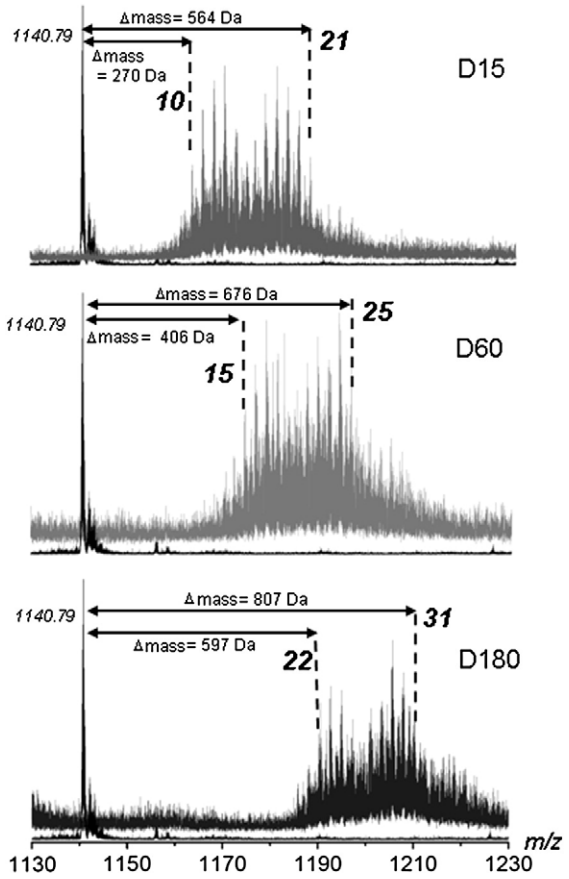


Fig. 7. FT-ICR MS analysis of the D-series of chemically modified Ocr. Upper panel, Ocr before (left-hand peak) and after (nested set of peaks) chemical modification with dimethylamine for 15 min as described in Methods. Middle panel, Ocr before (left-hand peak) and after (nested set of peaks) chemical modification with dimethylamine for 60 min. Lower panel, Ocr before (left-hand peak) and after (nested set of peaks) chemical modification with dimethylamine for 180 min. In each case, the $[M+12H]^{12+}$ charge state was analysed. The number in italics is the average m/z value for unmodified Ocr. The number in bold signifies the number of Δ+27 Da modifications for each species.

Protein folding and stability

Modification of a large fraction of the 35 carboxyl groups (34 acidic residues plus the C-terminus) in each Ocr monomer could induce an alteration in the folding of the protein, potentially leading to a loss of activity through a gross structural defect rather than a more subtle change in the charge distribution of the protein. Chemical modification of Ocr did result in some loss of protein due to precipitation during sample preparation. However, upon removal of the aggregates, the final protein samples displayed no increased propensity to denature and precipitate, indicating that the structural integrity of the chemically modified Ocr was not compromised.

Chemically modified samples were subjected to CD spectroscopy in the far-UV region to detect the presence of protein secondary structure. The structure of Ocr is dominated by α-helices,[5] and this is reflected in the CD spectra that display two distinct minima at 208 nm (part of the π-to-π* transition) and 222 nm (*n*-to-π* transition). CD spectra of the modified Ocr samples showed no major alteration in secondary structure content for the D-series of modifications (Fig. 8), nor for the N-series of modifications (data not shown). Spectra were analysed using the DichroWeb online secondary structure deconvolution programme.[26] Analysis using the CDSSTR method gave an α-helical content of 60% for the unmodified Ocr, which is in good agreement with the crystal structure. Similar values were ob-

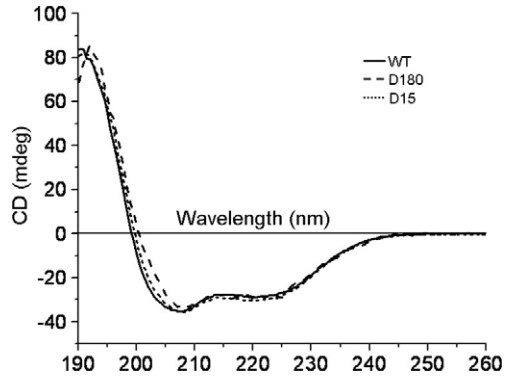


Fig. 8. Far-UV CD spectroscopy of Ocr and chemically modified Ocr (D-series). There is no apparent change in the secondary structure of Ocr upon modification. CD analysis for the N-series of chemically modified Ocr samples gave similar results.

tained for the chemically modified samples. We therefore concluded that the chemically modified proteins retained the overall fold of native Ocr.

We also investigated the stability of the chemically modified Ocr in comparison with that of the native protein. Initially, we analysed the guanidinium hydrochloride (GdmCl)-induced unfolding of the protein. Addition of denaturant caused the protein to unfold, resulting in a quenching of the fluorescent signal from the single tryptophan W94 (data not shown). For unmodified Ocr, the data indicated a two-state transition with no intermediate step with a ΔG of unfolding of 20 kcal mol^{-1} and a midpoint of 3.85 M GdmCl, in good agreement with previously published data.[9] However, the unfolding of the modified Ocr samples could not be described by a two-state transition. Instead, the midpoint of unfolding was calculated by fitting a sigmoidal curve (Origin Software version 6.1) to the data and found to occur at significantly higher concentrations of GdmCl for the N- and D-series of modified Ocr samples (D15, 4.25 M GdmCl; D60, 4.75 M GdmCl; N60, 4.88 M GdmCl; cf. unmodified Ocr, 3.85 M GdmCl). Thus, our data indicate that Ocr is actually stabilised by chemical modification.

The thermal stability of the N60 and D15 samples was also analysed by differential scanning calorimetry (data not shown). The data showed a clear energy uptake during the transition consistent with a cooperative endothermic unfolding event. The T_m of D15 (T_m = 68.2 °C) was similar to that of the unmodified Ocr (T_m = 69.0 °C). However, the T_m of N60 (T_m = 73.7 °C) was slightly elevated relative to that of the native protein. Noteworthy, however, was the observation that the unfolding of the modified Ocr samples (N60 and D15) was irreversible and showed a broader unfolding transition in contrast to unmodified Ocr, where the unfolding transition was completely reversible. This irreversible unfolding is presumably due to the formation of covalent cross-links within the Ocr dimer. Furthermore, a reduction in the net charge/electrostatic repulsions in the native fold could enhance the tendency for aggregation/irreversibility in the unfolded state.

Interactions of modified Ocr with M.EcoKI and EcoKI

Having determined that the modified Ocr samples were still folded and the extent to which they had been modified, we assessed their ability to function as DNA mimics by measuring their interaction with the methyltransferase core, M.EcoKI, and the entire nuclease, EcoKI. Specifically, we performed (i) isothermal titration calorimetry (ITC) of Ocr with M.EcoKI to determine the enthalpy and stoichiometry of binding (the binding was too strong to give accurate values of the binding affinity), (ii) competition between Ocr and a fluorescently labelled 21-bp DNA duplex for binding to M.EcoKI to obtain the binding affinity and (iii) inhibition of the nuclease activity of EcoKI on plasmid DNA by Ocr.

The ITC experiment was initially performed using unmodified Ocr in 20 mM Tris–HCl, pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol and 100 μM *S*-adenosyl-L-methionine (SAM) with or without 500 mM NaCl at a range of temperatures from 10 to 30 °C. The experiment at 25 °C was repeated using 20 mM Hepes buffer (heat of ionization = 20.5 kJ/mol) in place of Tris–HCl (heat of ionization = 47.4 kJ/mol) in the absence of NaCl, which gave a very similar enthalpy of interaction (ΔH of −90.0 kJ/mol in Hepes compared with −86.2 kJ/mol in Tris). These experiments indicated that there was no major contribution to ΔH due to effects of buffer ionization arising from protonation changes during binding. The enthalpy change upon interaction was strongly exothermic in the absence of NaCl but showed significant temperature dependence characteristic of a heat capacity change, ΔCp, upon formation of the M.EcoKI–Ocr complex. This was quantified from the slope of the plot of enthalpy change *versus* temperature (assumed to be linear) using the standard thermodynamic relationship:

$$\Delta Cp = d\Delta H/dT$$

We determined the ΔCp of unmodified Ocr in low salt buffer to be −4221 ± 372 J/mol K (Fig. 9). In 500 mM NaCl, the transition was less sharp (i.e., sigmoidal and indicating a slightly weaker interaction) and endothermic at all temperatures, and the slope of the linear regression yielded a binding heat capacity of −2071 ± 372 J/mol K.

We also carried out ITC on a representative sample from the N- and D-series of chemically modified Ocr in the same zero NaCl buffer described above. Specifically, we chose the N60 and D15 samples because they eluted from the Mono Q column at a similar point in the NaCl gradient, suggesting that the Ocr was modified to a comparable extent. The calorimetric data showed that the interaction between chemically modified Ocr and M.EcoKI was stoichiometric for both N60- and D15-modified Ocr samples, as found for the unmodified

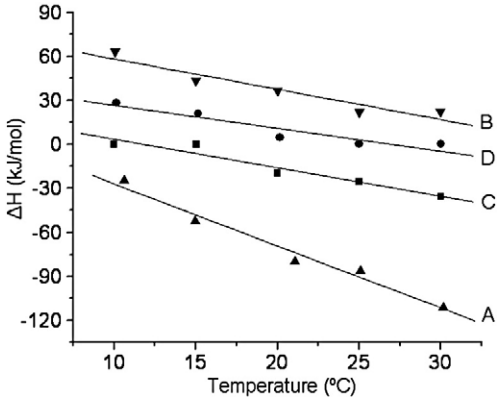


Fig. 9. Plot of enthalpy, ΔH (kJ/mol), *versus* temperature (°C) for the interaction of either unmodified Ocr or chemically modified Ocr with M.EcoKI. The ΔH of interaction was determined from ITC experiments at temperatures ranging from 10 to 30 °C. "A" and "B" show unmodified Ocr in the absence and in the presence of 500 mM NaCl, respectively. "C" and "D" show the analysis of the interaction between M.EcoKI and the N60 and the D15 chemically modified Ocr samples in the absence of NaCl, respectively.

Ocr (i.e., one Ocr dimer per molecule of M.EcoKI). In addition, the ΔH values fell between the behaviour of the unmodified Ocr in zero NaCl buffer and that in 500 mM NaCl buffer (Fig. 9). Furthermore, the ΔCp values for N60 Ocr (-1385 ± 540 J/mol K) and D15 Ocr (-1460 ± 385 J/mol K), although determined in the absence of NaCl, were close to the value determined for unmodified Ocr in 500 mM NaCl.

We also studied the interaction between modified Ocr and M.EcoKI using a sensitive fluorescence anisotropy assay to determine the binding affinity.[12,27] The values of the dissociation constant, K_d, given in Table 1 show that the modified Ocr samples were less able to interfere with DNA binding by M.EcoKI than the native Ocr protein. Furthermore, the amount of modification correlated with a loss in binding affinity. Specifically, the most extensively modified Ocr samples bound more weakly than DNA to M.EcoKI.

Table 1. K_d values for N- and D-series samples binding M. EcoKI as determined by fluorescence anisotropy (cf. K_d for unmodified Ocr is ~44 pM, and K_d for DNA is ~2.1 ±0.35 nM)

	K_d (nM)
D-series	
D15	2.64±0.52
D30	5.8±1.4
D60	19.7±7.3
D120	23.1±8.1
D180	26.6±9.2
N-series	
N15	0.56±0.13
N30	1.82±0.37
N60	1.9±0.2
N120	3.77±0.82
N180	12.6±3.0

M.EcoKI was titrated into a solution containing 50 nM Ocr and 2 nM concentration of a 21-bp DNA duplex fluorescently labelled at the 5′ end. The binding of M.EcoKI to the DNA duplex was monitored by anisotropy. Upon binding of M.EcoKI to DNA, the DNA–protein complex tumbles more slowly, resulting in an increase in fluorescence anisotropy from the fluorophore on the DNA.

The activity of the modified Ocr samples was tested in an endonuclease assay using purified EcoKI. Linearisation of a circular unmethylated plasmid (pBRsk1) containing a unique EcoKI target recognition site was monitored in the absence and in the presence of Ocr (unmodified and N- or D-modified samples). In each case, the reaction mixture minus DNA was prepared and the digestion was initiated by addition of pBRsk1. The reaction was stopped after 10 min, and the mixtures were then analysed by agarose gel electrophoresis (Fig. 10). Incubation of pBRsk1 in the presence of a 6-fold excess of EcoKI resulted in complete cutting of the plasmid to a linear form within 10 min (compare lanes 1 and 3). The experiment was also performed in the presence of a 10-fold excess of Ocr dimer over EcoKI (lanes 4–10). As anticipated, unmodified Ocr completely abolished linearisation of the plasmid DNA (lane 4). However, the modified Ocr samples became progressively poorer endonuclease inhibitors as the extent of modification increased. For example, D15 and D60 displayed partial inhibitory activity and D120 showed almost no nuclease inhibition (lanes 5, 6 and 7, respectively). A similar trend was also found for the N-series of modified Ocr (N15, N60 and N120), although here the inhibitory activity was greater than that for the corresponding samples from the D-series (e.g., compare lanes 5 and 8).

Discussion

The DNA mimicry displayed by Ocr comprises two main features: its mimicry of the bent DNA

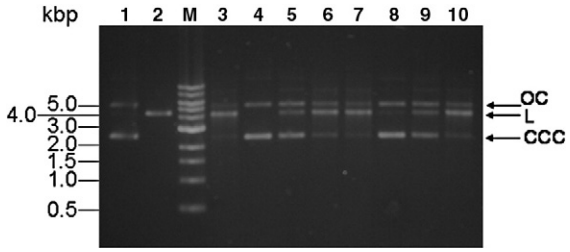


Fig. 10. *In vitro* inhibition of EcoKI using chemically modified Ocr. An unmethylated plasmid (pBRsk1) with a single EcoKI recognition site was used as substrate. Samples were analysed using 0.9% agarose gel. Lane 1, undigested pBRsk1; lane 2, pBRsk1 linearised with EcoRI endonuclease to act as a size marker; lane 3, pBRsk1 linearised with EcoKI; lane 4, reaction performed using EcoKI pre-incubated with a 10-fold molar excess of unmodified Ocr dimer; lanes 5–10, as for lane 4 except the use of D15, D60, D120, N15, N60 and N120 in place of the unmodified Ocr, respectively. M, 1-kbp ladder from New England Biolabs. CCC is the covalently closed circular form of the plasmid, OC indicates the open circular form containing one or more single-strand nicks and L indicates the linear product of cleavage.

substrate, preferred by EcoKI, and its mimicry of the electrostatics of the phosphate backbone. Other features, such as H-bonding and van der Waals interactions, will also play a role in the binding of Ocr to its target enzyme. However, in the absence of a detailed structure for an Ocr–M.EcoKI complex, these intermolecular forces are not easily defined. Recently, Kennaway *et al*. produced a 16-Å resolution structure of M.EcoKI and an approximate atomic model of it in complex with Ocr, which shows it completely enveloping the Ocr molecule.[28] Therefore, any of the residues of Ocr could potentially be involved in the interaction. Here, we focus on the electrostatic component of the Ocr–M.EcoKI interaction, which appears to be an important feature of the mimicry.

Chemical modification offers a convenient method of reducing the number of negatively charged groups (i.e., Asp and Glu side chains) on the Ocr surface. Our results clearly show the stepwise reduction of negative charge with reaction time. MS of the D-series of chemically modified samples showed that the protein was subject to averages of 16 (12/13/17/18/19 modifications most common), 20 (16/17/22/23/24 most common) and 27 (23/24/29/30 most common) modifications per Ocr monomer for the D15, D60 and D180 samples, respectively. The presence of cross-links was also demonstrated despite our effort to minimise this side reaction, so these numbers of modifications may be a slight underestimate by perhaps 1 or 2 modifications per Ocr monomer. Our MS measurements of the number of modified residues and the additional uncertainty due to cross-linking reflect the relative number of "random modification cycles" required to target the most important residues involved in the interaction. Therefore, our results are most probably an overestimate of the total number of negatively charged residues that are critically involved in the interaction between Ocr and M.EcoKI. This is supported by mutational studies showing that positional context and the number of negatively charged residues are critical for the interaction with M.EcoKI.[6,12,13,16]

The degree of chemical modification did not appear to have any deleterious effect on protein folding, as shown by CD, or stability. Such extensive modification without destruction of the protein fold is noteworthy. Indeed the stability to denaturation even increased. This is attributable to two factors. First, the adventitious formation of intermolecular (i.e., between monomers of an Ocr dimer) and intramolecular (i.e., within each Ocr monomer) cross-links during the chemical modification will stabilise the fold (although it will also make it recalcitrant to refolding upon denaturation). Second, the closeness of the carboxylates in the unmodified Ocr (approximately equal to the Bjerrum length) leads to electrostatic repulsion energies of the order k_BT.[29] Chemical modification increases the distance between the unmodified carboxylates, reducing electrostatic repulsion and enhancing stability.

Concomitant with the degree of modification, we observed a clear drop in the binding affinity for an anion-exchange column, for binding to M.EcoKI and for inhibiting the EcoKI nuclease. The first two features are more easily quantifiable and should be correlated given the importance of electrostatics in protein–DNA (and protein–DNA mimic) interactions. Additionally, the binding also gave rise to well-defined enthalpy and heat capacity changes, which can be discussed in terms of protein–protein interfaces.

Using our data of binding affinity and elution time (or [NaCl]) from the anion-exchange column, we can plot RTlog(K_d) (i.e., the free energy) against [NaCl] or elution time (Fig. 11). Our results indicate that approximately 16 acidic residues in each Ocr subunit, equivalent to sample D15 if one ignores cross-linking modifications, must be lost for the protein to have the same binding affinity as DNA for M.EcoKI (i.e., 2.1 nM or −49.5 kJ/mol). This is an ~50-fold loss in binding affinity compared with unmodified Ocr. Further loss of acidic residues leads to further decrease in binding affinity. For example, sample D60, which has lost an average of 20 acidic residues, not including cross-linking modifications, displays an ~500-fold loss in binding affinity relative to unmodified Ocr and binds more poorly than DNA. If one were to assume that each acidic residue of Ocr contributed equally to the interaction with M.EcoKI, one might anticipate a linear free energy relationship between log(K_d) and the number of residues modified. As all degrees of modification from the wild type (WT; zero modifications) up to and including D30 and N180 fall on a straight line, this would appear to be the case starting from unmodified Ocr until one reaches ~16 to ~20 modified residues per monomer (~32 to ~40 for the dimer).

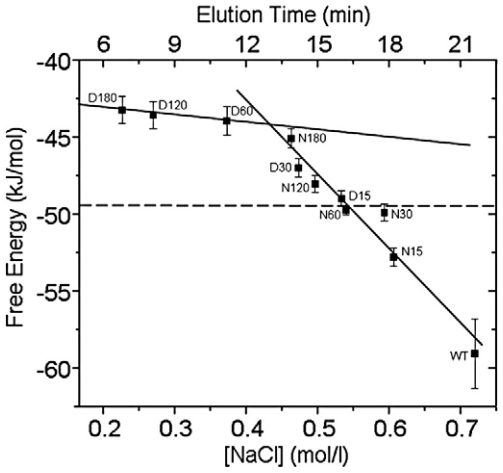


Fig. 11. Correlation of the interaction between M.EcoKI and chemically modified Ocr samples with the NaCl concentration (mol/l) at which the Ocr sample elutes from an ion-exchange chromatography column (equivalent to the elution time from the column). The free energy of interaction calculated from the K_d values (Table 1) is plotted against [NaCl] (or elution time). Two regions showing a linear correlation of free energy of binding with [NaCl] are apparent. The dotted line represents the free energy of interaction between M.EcoKI and DNA.

Thus, each of these individual negative charges contributes only ~0.4 to ~0.5 kJ/mol to the free energy of binding. Since this energy is less than the thermal energy, k_BT, then it is clear that a very large number of such weak interactions need to be summed to account for the effectiveness of Ocr as an inhibitor. This value per charge is also small when compared with the electrostatic effects observed with the conceptually similar barnase–barstar protein–protein interaction system.[30] Additional modification further reduces binding affinity although the effect of individual single modifications is greatly reduced as the free energies of binding of samples D60, D120 and D180 are very similar. We can see that these highly modified bent Ocr molecules have a binding affinity of the order 20 nM (−44 kJ/mol) for M.EcoKI. Thus, it appears that once one exceeds 16 to 20 modifications per ocr monomer, one observes only the contribution of the general shape of Ocr to the binding and the electrostatic mimicry is lost.

Hence, we have separated the relative contributions of the mimicry by Ocr, of the shape of the bent DNA molecule bound by EcoKI and of the charge distribution on the DNA molecule. It is noteworthy that the highly modified Ocr still displays significant binding (in the nanomolar range), presumably due to the favourable shape complementarity of the Ocr–M.EcoKI association involving a multiplicity of weak intermolecular interactions. Our results show that the addition of electrostatic mimicry to the DNA-shape mimicry of the Ocr molecule further increases the binding affinity for M.EcoKI by up to ~800-fold.

We also investigated the Ocr–M.EcoKI interaction by ITC, and this gives us some further insight into the nature of the protein–protein interface. The thermodynamics of protein–protein interactions are typically made up of numerous small changes in free energy with both enthalpic and entropic components.[31] One might anticipate a major entropic component to the interaction given that the interface between the two proteins is substantial, with the M.EcoKI almost completely engulfing the Ocr protein, which has a Connolly surface of 11,238 ^{2} for the dimer.[5,11] A recent model for the Ocr–M.EcoKI interaction that is consistent with this hypothesis has been proposed.[28] A large number of water molecules with all of their H-bonding potential are presumably displaced during the formation of the complex, and the dynamics of residues on the interface must become rather constrained. This is particularly important under high salt conditions where the counterions decorating the surface of Ocr must be displaced in order to allow interaction with M.EcoKI. Indeed, the transition in the presence of salt is more endothermic than that in the absence of salt at any given temperature (Fig. 9).

Initially, we studied the interaction of unmodified Ocr with M.EcoKI in the absence and in the presence of monovalent salt. Bearing in mind the average ΔCp for protein–protein interactions is reported to be −1393±845 J/mol K,[32] our results show a surprisingly large ΔCp effect for the M.EcoKI–Ocr interaction that is strongly dependent on ionic strength. Similarly, we determined the ΔCp values of the D15 and N60 protein samples under low salt conditions, which were found to be similar to the value determined for unmodified Ocr under high salt conditions. Large ΔCp values similar to those obtained here have been found for the interaction between the highly charged inhibitor protein barstar and its target barnase enzyme.[30]

Conventionally, large ΔCp effects are associated with hydrophobic interactions. However, theoretical considerations and empirical observations show that long-range electrostatic interactions and other effects can also make a significant contribution to ΔCp.[30,33,34] Our results clearly reinforce this notion. Furthermore, several algorithms that attempt to correlate the magnitude of ΔCp with changes in the solvent-accessible surface area of interaction have been devised.[35] Assuming that the structure of the Ocr–M.EcoKI interface is essentially unchanged by high ionic strength, or by chemical modification, then conventional theories would predict the same ΔCp. Our results clearly contradict this prediction. Similar discrepancies with this surface area theory have been observed for the strong electrostatic interaction between the enzyme barnase and its protein inhibitor barstar.[30]

In conclusion, our results show that the DNA mimicry displayed by Ocr is extremely robust and can be separated into mimicry of the shape and charge of DNA. After extensive modification, potentially removing ~86% of the negative charge from the carboxyl side chains and C-terminus, the modified Ocr protein still binds to M.EcoKI with an affinity of 27 nM (−43.9 kJ/mol). This is only marginally weaker than the M.EcoKI–DNA interaction (−49.4 kJ/mol) and is presumably the contribution to binding of M.EcoKI by Ocr's mimicry of the shape of the bent DNA molecule preferred by M.EcoKI.[5,15] The introduction of negative charge onto this simple "shape mimic" further increases binding affinity for M.EcoKI until Ocr far surpasses the affinity of DNA for M.EcoKI. The introduction of the mimicry of the charge pattern on the DNA increases affinity to 44 pM (a further −14.6 kJ/mol in addition to the shape mimicry). Thus, although the charge mimicry makes a lesser contribution than shape mimicry to the effectiveness of Ocr, it is a crucial addition to push the affinity of Ocr for M.EcoKI past the affinity of DNA for M.EcoKI, thereby transforming Ocr into an extremely effective inhibitor of type I R/M enzymes.

Methods

Enzymes, chemicals and plasmids

EcoKI ($R_2M_2S_1$) and the DNA methyltransferase component M.EcoKI (M_2S_1) were purified as described previously.[36,37] The Ocr protein was purified as described previously.[12] Plasmid pAR2993 (a kind gift from Alan

Rosenberg and William Studier, Brookhaven National Laboratory) harbours the gene encoding WT Ocr located just downstream of an isopropyl-β,D-thiogalactopyranoside-inducible promoter.[9] pBRsk1 is an engineered version of pBR322 (4361 bp) in which one of the two EcoKI sites (4024–4036) has been removed by site-directed mutagenesis.[38] The unmethylated form of pBRsk1 used in the nuclease assays was prepared from *E. coli* NM1261 (r^-m^-) (a kind gift of Prof. Noreen E. Murray, University of Edinburgh). Dimethylamine, EDC and HOBt were obtained from Pierce (Rockford, IL). Guanidine hydrochloride (ULTROL grade) was purchased from Calbiochem (San Diego, CA). SAM was from New England Biolabs (Ipswich, MA). All other reagents were purchased from Sigma-Aldrich (St. Louis, MO). Broad-range pre-stained molecular mass markers for SDS-PAGE were purchased from BioRad (Precision Plus Protein Standards; Hercules, CA). All solutions were made up in distilled, deionized water.

Chemical modification procedure

Surface carboxyl groups of Ocr were chemically modified using EDC and either ammonium hydroxide (to give the so-called N-series of chemically modified proteins) or dimethylamine (to give the so-called D-series) as a nucleophile. Chemical modification of Ocr (3 μM) was carried out at 25 °C in 750 mM ammonium hydroxide or dimethylamine HCl, 150 mM NaCl, 60 mM EDC and 60 mM HOBt, pH 6.5. Aliquots were withdrawn at specific time points (1 to 180 min), and the reaction was quenched by adding a 6-fold excess of sodium acetate (from a 3 M stock of sodium acetate, pH 6.5). After incubation for 10–20 min, the mixture was first dialysed against 100 mM nucleophile (either ammonium hydroxide or dimethylamine, pH 8.5) for 4 h at 25 °C and then against 50 mM ammonium acetate at 4 °C for a minimum of 4 h. Hydroxylamine was added to the solution to a final concentration of 400 mM (from a 3 M stock of hydroxylamine, pH 7.0), and the mixture was incubated at 25 °C for 4 h. Finally, the solution was dialysed against 20 mM ammonium acetate and then the protein concentration was adjusted to 20–30 μM using a Vivaspin concentrator (10,000 MWCO; VivaScience AG). Samples were stored in 50% v/v glycerol at −20 °C.

Ion-exchange chromatography

WT and chemically modified Ocr samples were analysed by anion-exchange chromatography using a 1-ml Mono Q column (GE Healthcare). Each protein (11.2 μg) was individually loaded onto the column pre-equilibrated in 20 mM Tris–HCl, pH 8.0, at a flow rate of 1 ml/min. After washing the column, a linear gradient of 0–1 M NaCl in 20 mM Tris–HCl, pH 8.0, was run over 30 min using the same flow rate. Protein elution from the column was monitored by measuring the tryptophan fluorescence (excitation, 295 nm; emission, 350 nm) of the eluate. The elution time of each sample was determined by integrating the peak and calculating the point corresponding to 50% of the peak area (i.e., time at which 50% of the protein has eluted from the column).

MALDI-TOF MS analysis

MS of the D-series of chemically modified Ocr was performed by MALDI-TOF using a Voyager DE STR instrument (Applied Biosystems, Foster City, CA). Protein samples were diluted in 0.1% trifluoroacetic acid to 0.05 mg/ml and mixed with an equal volume of matrix (saturated solution of sinapinic acid in 50% acetonitrile and 0.1% trifluoroacetic acid) on a stainless steel surface. The samples were air dried at room temperature to crystallize. The machine was operated in positive ion mode and calibrated with conalbumin and bovine serum albumin.

High-resolution LC-MS analysis

Protein samples were extensively desalted by dialysis into 20 mM ammonium acetate prior to MS. For LC-MS, an Ultimate 3000 HPLC system (Dionex Corporation, Sunnyvale, CA) equipped with a monolithic PS-DVB (500 μm×5 mm) analytical column (Dionex Corporation) was used. Solutions B and C were prepared comprising 2:97.95:0.05 and 80:19.95:0.05 of acetonitrile/water/formic acid, respectively. Samples in solution B containing ~1 μg of chemically modified Ocr were centrifuged (16,100*g* for 2 min) immediately prior to injection onto the column. After injection, the column was washed with solution B for 5 min, followed by a 20-min linear gradient elution (20 μl/min) into solution C. The eluate was passed into the mass spectrometer. MS data were acquired on a Bruker 12-Tesla Apex Qe FT-ICR (Bruker Daltonics, Billerica, MA) equipped with an electrospray ionization source. Desolvated ions were transmitted to a 6-cm Infinity Cell® Penning trap. Trapped ions were excited (frequency chirp of 48–500 kHz at 100 steps of 25 μs) and detected between m/z values of 600 and 2000 for 0.5 s to yield broadband 512-kWord time-domain data. Fast FTs and subsequent analyses were performed using DataAnalysis (Bruker Daltonics) software. Multiple charge states could be observed in this way for each of the major species.

Circular dichroism (CD) analysis

CD measurements were performed as described previously.[39] Protein samples were prepared in 10 mM Tris–HCl, pH 8.0, and 50 mM NaF. Spectra (190–260 nm) were obtained at a protein concentration of 30 μM using a 0.2-mm path-length cell. Each spectrum was an accumulation of four individual scans. The spectra were corrected for buffer contribution.

Unfolding studies

Equilibrium unfolding of Ocr as a function of GdmCl concentration was monitored by tryptophan fluorescence spectroscopy as described previously.[9] The fluorescent intensity was measured as a ratio of the fluorescent signal at 350 *versus* 380 nm to remove any variation in intensity due to slight differences in protein concentration between samples.[36]

Calorimetry

Differential scanning calorimetry and ITC were carried out as described previously.[12] All ITC experiments were conducted using 20 mM Tris–HCl, pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol and 100 μM SAM. Typically, Ocr at a concentration of 40 μM was titrated into an M.EcoKI solution at a concentration of 4 μM.

Fluorescence anisotropy

Competition for binding WT or chemically modified Ocr to M.EcoKI was determined using the fluorescence anisotropy assay as described previously.[12]

Nuclease assay

The *in vitro* assay monitored the cleavage of unmethylated circular pBRsk1 using purified EcoKI in the absence or in the presence of WT or chemically modified Ocr essentially as described elsewhere.[39]

Acknowledgements

We gratefully acknowledge financial support from the Wellcome Trust (GR080463MA to D.T.F.D.) and the Biotechnology and Biological Sciences Research Council (BB/D001870/1 to D.T.F.D. and B20089 to A.C. for microcalorimetry facilities). A.R.T. was supported by funding from the Engineering and Physical Sciences Research Council (EP/D056756/1 to D.T.F.D.) and a RASOR grant from the Biotechnology and Biological Sciences Research Council (BB/C511599/1). We thank Prof. Noreen Murray (Institute of Cell and Molecular Biology, University of Edinburgh) for kindly providing the NM1261 strain, Alan Rosenberg and William Studier (Brookhaven National Laboratory) for providing pAR1992 and Dr. Joseph Spadavecchia (Database Research Group, School of Informatics, University of Edinburgh) for assistance with data analysis. We also thank Dr. Patrick Langridge-Smith for access to the SIRCAMS FT-ICR MS facility and Dr. Andy Cronshaw for assistance with the MALDI-TOF experiment.

References

1. Studier, F. W. (1975). Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host. *J. Mol. Biol.* **94**, 283–295.
2. Putnam, D. C. & Tainer, A. J. (2005). Protein mimicry of DNA and pathway regulation. *DNA Repair*, **4**, 1410–1420.
3. Tock, M. R. & Dryden, D. T. F. (2005). The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472.
4. Dryden, D. T. F. (2006). DNA mimicry by proteins and the control of enzymatic activity on DNA. *Trends Biotechnol.* **24**, 378–382.
5. Walkinshaw, M. D., Taylor, P., Sturrock, S. S., Atanasiu, C., Berge, T., Henderson, R. M. *et al.* (2002). Structure of ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell*, **9**, 187–194.
6. Dunn, J. J., Elzinga, M., Mark, K. K. & Studier, F. W. (1981). Amino acid sequence of the gene 0.3 protein of bacteriophage T7 and nucleotide sequence of its mRNA. *J. Biol. Chem.* **256**, 2579–2585.
7. Mark, K. K. & Studier, F. W. (1981). Purification of the gene 0.3 protein of bacteriophage T7, an inhibitor of the DNA restriction system of *Escherichia coli*. *J. Biol. Chem.* **256**, 2573–2578.
8. Bandyopadhyay, P. K., Studier, F. W., Hamilton, D. L. & Yuan, R. (1985). Inhibition of the type I restriction–modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.* **182**, 567–578.
9. Atanasiu, C., Byron, O., McMiken, H., Sturrock, S. S. & Dryden, D. T. F. (2001). Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7. *Nucleic Acids Res.* **29**, 3059–3068.
10. Blackstock, J. J., Egelhaaf, S. U., Atanasiu, C., Dryden, D. T. F. & Poon, W. C. K. (2001). Shape of Ocr, the gene 0.3 protein of bacteriophage T7: modeling based on light scattering experiments. *Biochemistry*, **40**, 9944–9949.
11. Atanasiu, C., Su, T.-J., Sturrock, S. S. & Dryden, D. T. F. (2002). Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res.* **30**, 3936–3944.
12. Stephanou, A. S., Roberts, G. A., Tock, M. R., Pritchard, E. H., Turkington, R., Nutley, M. *et al.* (2009). A mutational analysis of DNA mimicry by ocr, the gene 0.3 antirestriction protein of bacteriophage T7. *Biochem. Biophys. Res. Commun.* **378**, 129–132.
13. Zavilgelsky, G. B., Kotova, V. Y. & Rastorguev, S. M. (2009). Antirestriction and antimodification activities of T7 Ocr: effects of amino acid substitutions in the interface. *Mol. Cell. Biol. (Mosc.)*, **43**, 93–100.
14. Murray, N. E. (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* **64**, 412–434.
15. Su, T.-J., Tock, M. R., Egelhaaf, S. U., Poon, W. C. K. & Dryden, D. T. F. (2005). DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Res.* **33**, 3235–3244.
16. Zavilgelsky, G. B., Kotova, V. Y. & Rastorguev, S. M. (2008). Comparative analysis of anti-restriction activities of ArdA (ColIb-P9) and Ocr (T7) proteins. *Biochemistry (Mosc.)*, **73**, 906–911.
17. Hoare, D. G. & Koshland, D. E., Jr (1967). A method for the quantitative modification and estimation of carboxylic acid groups in proteins. *J. Biol. Chem.* **242**, 2447–2453.
18. Carraway, K. L. & Koshland, D. E., Jr (1972). Carbodiimide modification of proteins. *Methods Enzymol.* **25**, 616–623.
19. Williams, A. & Ibrahim, I. T. (1981). Carbodiimide chemistry: recent advances. *Chem. Rev.* **81**, 589–636.
20. Nakajima, N. & Ikada, Y. (1995). Mechanism of amide formation by carbodiimide for bioconjugation in aqueous media. *Bioconjugate Chem.* **6**, 123–130.
21. Smyth, G. D. (1967). Acetylation of amino and tyrosine hydroxyl groups. *J. Biol. Chem.* **242**, 1592–1598.
22. Carraway, K. L. & Koshland, D. E., Jr (1968). Reaction of tyrosine residues in proteins with carbodiimide reagents. *Biochim. Biophys. Acta*, **160**, 272–274.
23. Chan, C. L. & Cox, G. B. (2007). Kinetics of amide formation through carbodiimide/*N*-hydroxybenzotriazole (HOBt) couplings. *J. Org. Chem.* **72**, 8863–8869.
24. Gounaris, A. D. & Perlmann, G. E. (1967). Succinylation of pepsinogen. *J. Biol. Chem.* **242**, 2739–2745.
25. Schägger, H. & von Jagow, G. (1991). Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* **199**, 223–231.
26. Whitmore, L. & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668–W673.

27. Powell, L. M., Connolly, B. A. & Dryden, D. T. F. (1998). The DNA binding characteristics of the trimeric EcoKI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting. *J. Mol. Biol.* **283**, 947–961.
28. Kennaway, C. K., Obarska-Kosinska, A., White, J. H., Tuszynska, I., Cooper, L. P., Bujnicki, J. M. *et al.* (2009). The structure of M.EcoKI type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res.* **37**, 762–770.
29. Bloomfield, V. A., Crothers, D. M. & Tinoco, I. (2000). In *Nucleic Acids: Structures, Properties and Functions*, (Connolly, B., ed.), pp. 335–443, University Science Books, Sausalito, CA.
30. Frisch, C., Schreiber, G., Johnson, C. M. & Fersht, A. R. (1997). Thermodynamics of the interaction of barnase and barstar: changes in free energy *versus* changes in enthalpy on mutation. *J. Mol. Biol.* **267**, 696–706.
31. Cooper, A., Johnson, C. M., Lakey, J. H. & Nollmann, M. (2001). Heat does not come in different colours: entropy–enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions. *Biophys. Chem.* **93**, 215–230.
32. Stites, W. E. (1997). Protein–protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev.* **97**, 1233–1250.
33. Gallagher, K. & Sharp, K. (1998). Electrostatic contributions to heat capacity changes of DNA-ligand binding. *Biophys. J.* **75**, 769–776.
34. Cooper, A. (2005). Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics. *Biophys. Chem.* **115**, 89–97.
35. Gomez, J. & Freire, E. (1995). Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. *J. Mol. Biol.* **252**, 337–350.
36. Dryden, D. T. F., Cooper, L. P. & Murray, N. E. (1993). Purification and characterization of the methyltransferase from the type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.* **268**, 13228–13236.
37. Dryden, D. T. F., Cooper, L. P., Thorpe, P. H. & Byron, O. (1997). The *in vitro* assembly of the EcoKI type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry*, **5**, 1065–1076.
38. Davies, G. P., Kemp, P., Molineux, I. J. & Murray, N. E. (1999). The DNA translocation and ATPase activities of restriction-deficient mutants of EcoKI. *J. Mol. Biol.* **292**, 787–796.
39. Serfiotis-Mitsa, D., Roberts, G. A., Cooper, L. P., White, J. H., Nutley, M., Cooper, A. *et al.* (2008). The Orf18 gene product from conjugative transposon Tn916 is an ArdA antirestriction protein that inhibits type I DNA restriction–modification systems. *J. Mol. Biol.* **383**, 970–981.