The use of green fluorescent protein for the analysis of protein-protein and protein-DNA interactions

Kai Chen

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh
2011

Declaration

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

Kai Chen

2011

Acknowledgements

Lots of people helped me during the last four years. Firstly, I would like to thank my supervisor Dr David Dryden for his advice, help, support and his continuous encouragement during my PhD. I also want to thank Dr Anita Jones for her useful advice and discussion on the analysis of the fluorescence lifetime data.

I am very grateful to all present and past members of David's research group, Laurie Copper, Gareth Roberts, John White, Bansi Sanghvi, Marcel Reuter, Dimitra Serfiotis-Mitsa, Augoustinos Stephanou who provided me with an excellent environment for my research. I am grateful to Dr John White for advice on molecular biology, Laurie Cooper for helping me on preparing and purifying proteins, Dr Gareth Roberts for advice, discussion and help. I also thank Prof. Alan Cooper and Margaret Nutley at the University of Glasgow for ITC experiments. I also wish to thank Andy Gary, David Paden and Stuart Mains for their assistance in building the fluorescence instrument. Special thanks to David, Gareth and Laurie for their correction proof reading of the thesis.

Many thanks to my wife, Pinghui, for her massive support and understanding. Thanks also go to my family who have supported and encouraged me.

This work was supported by EaStCHEM, the joint Chemistry Research School of Edinburgh and St. Andrews funded by the Scottish Funding Council for Further and Higher Education.

Abstract

Restriction modification (RM) systems play a crucial role in preventing the entry of foreign DNA into the bacterial cell. The best studied Type I RM system is EcoKI from *Escherichia coli* K12. Both bacteriophage and conjugative plasmids have developed a variety of strategies to circumvent the host RM system. One such strategy involves the production of antirestriction proteins that mimic a short segment of DNA and efficiently inhibit the RM system. The main aim of this project was to analyse the interaction of EcoKI and its cognate methylase (MTase) with the T7 antirestriction protein, known as overcome classical restriction (Ocr), and various ArdA antirestriction proteins. Currently, there is a paucity of structural data on the complex formed between the Type I system and the antirestriction proteins. The aim of this work was twofold; (i) compare the interaction of MTase with DNA and Ocr and (ii) quantify the strength of interaction between MTase and various ArdA proteins.

The MTase was fused to the Green Fluorescent Protein (GFP) to facilitate determination of the orientation of interaction with DNA and Ocr. Time resolved fluorescence measurements were carried out using the GFP-MTase fusion to determine the fluorescence lifetime and anisotropy decay. These experiments were conducted using a time resolved fluorescence instrument fabricated in-house. The values determined in these experiments were then used to perform fluorescence resonance energy transfer (FRET) measurements with fluorescently labelled DNA or Ocr. These measurements gave information concerning the relative orientation of the MTase with either DNA or Ocr.

The GFP-MTase fusion was also used to quantify the strength of interaction with various ArdA proteins. Previous attempts to determine the strength of interaction between MTase and ArdA proteins by employing conventional techniques have been unsuccessful. Therefore, a novel method was developed that exploits the interaction of MTase with a cation exchange medium, which can subsequently be displaced upon binding to ArdA. This method facilitated the determination, for the first time, of a set of binding affinities for the MTase and ArdA interaction.

Table of contents

Chapter 1 Introduction

1.1 Restriction-modification systems

The phenomenon of restriction and modification (R-M) was first identified as a prokaryotic immune system in the early 1950s. Certain strains of bacteria were found to "restrict" (inhibit) the propagation of foreign DNA that entered the cell [Wilson *et al*., 1991, Arber, 1965]. R-M systems comprise two enzymatic activities with opposing intracellular effects: a restriction endonuclease (REase) and a methyltransferase. By recognising a specific DNA sequence, the REase is able to distinguish between "self" and "non-self" DNA. "Non-self" DNA is cleaved to small harmless fragments. The methyltransferase protects "Self" DNA at the recognition site by adding a methyl group to either a cytosine or an adenine nucleotide in the recognition sequence. The absence of the methyl group at the recognition site triggers REase activity resulting in a double strand cut of the foreign DNA [Wilson *et al*., 1991, Dryden *et al*., 2001].

R-M enzymes can cleave double-stranded (ds) DNA at the recognition sequence or at a random position far away from the recognition sequence, and this is accomplished by hydrolysis of one phosphodiester bond in the backbone of each DNA strand. Restriction REases commonly require Mg^{2+} or a similar divalent cation; some also require, or are stimulated by, ATP or *S*-adenosylmethionine (SAM). Cleavage generally occurs on the 5' -side of the phosphate group, leaving DNA fragments with 5'-phosphoryl and 3 ' -hydroxyl termini. Unlike fragments with 3'-phosphoryl and 5'-hydroxyl ends, 5'-phosphoryl and 3'-hydroxy termini can be rejoined again by a DNA ligase [Wilson *et al*., 1991].

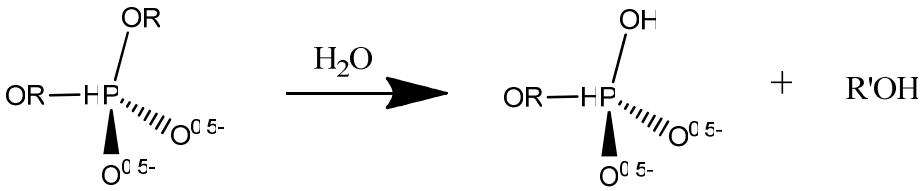


Figure 1-1: Cleavage of a phosphodiester bond in double-stranded DNA by a REase. The hydrolysis products have 5'-phosphoryl and 3'-hydroxy termini.

Modification methyltransferase require SAM as a methyl group donor to supply a methyl group to one nucleotide in each strand of the recognition sequence. Thus, SAM is an essential cofactor for methylation. Usually, the same DNA is methylated on both strands. Adenine and cytosine are the only bases known to be methylated by R-M systems.

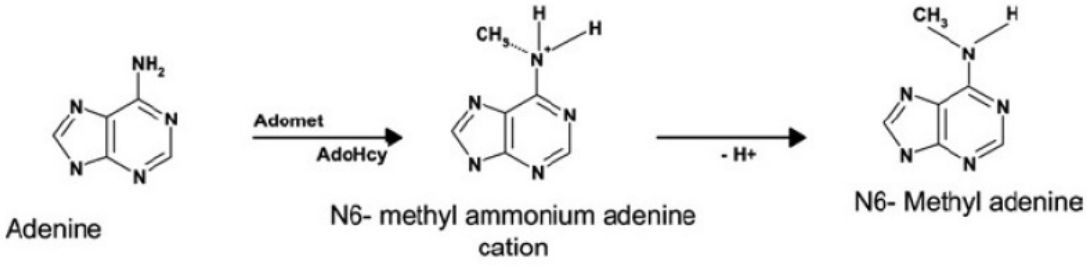


Figure 1-2: Proposed reaction mechanism of N6-adenine methylation (reprinted from Bheemanaik *et al*., 2006). Adomet is another name given to SAM; AdoHcy – also known as SAH (S-adenosylhomocysteine).

R-M systems are classified according to their cofactor requirements, enzyme composition, nature or symmetry of their recognition sequences and the position of the site of DNA cleavage with respect to the target sequence. Four different types of R-M systems have been characterised thus far and have been classified as types I – IV [Wilson *et al*., 1991]. The characteristic features of these systems are listed in Table 1-1. The R-M systems have not been found in eukaryotes. All of the R-M systems operate on DNA by recognising a specific sequence of bases.

Table 1

Characteristics and organization of the genetic determinants of different classes of R-M systems[a].

	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

[a] A full description of R-M classifications is given in [7••].

Table 1-1: Classification of R-M systems based on their enzyme composition, cofactor requirements, symmetry of the recognition sequences and the position of DNA cleavage (adapted from Murray, 2000; Dryden *et al.*, 2001 and Tock *et al.*, 2005).

1.1.1 Type I systems

Type I systems were originally found in *E. coli* and its relatives, *Citrobacter* and *Salmonella* sp. These enzymes recognise specific asymmetric bipartite nucleotide sequences, comprising two 3-5 bp specific sequences, separated by a nonspecific spacer of 6-8 bp. Type I enzymes are comprised of three polypeptides, HsdR (restriction ~140 kDa), HsdM (modification ~50-60 kDa) and HsdS (specificity ~50 kDa). The S subunit includes two target recognition domains (TRDs) that recognise target DNA sequence specificity for both the restriction and modification activities of the complex. The M subunit includes the binding site of the methyl donor SAM and determines the methylation status of the target DNA sequence. The R subunit includes the active site for ATP hydrolysis and other sequences essential for DNA translocation and REase activity and is involved in the cleavage of the unmethylated DNA. The resulting complex is both a REase and a methyltransferase.

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

al., 2007]: Type IA (e.g.EcoKI), Type IB (e.g. EcoAI), Type IC (e.g. EcoR124I), Type ID (e.g. StySBLI) and Type IE (e.g. KpnBI).

The archetypal type I R-M enzyme is EcoKI ($R_2M_2S_1$ ~440 kDa) from *E. coli* [Dryden *et al.,* 1997]. EcoKI was the first DNA R-M enzyme to be discovered [Bertani 1953] and purified [Meselson, 1968]. The EcoKI nuclease recognises the ds-DNA specific sequence 5'-A<u>A</u>C(N_6)G<u>T</u>GC-3', where N can be any base [Kan *et al*., 1979], and methylates the N^6 position of adenine at and complementary to the underlined nucleotides within both parts of the target recognition sequence. MTase (M_2S_1) is a maintenance EcoKI methyltransferase, which is very efficient at methylating newly replicated, hemimethylated host DNA in common with many eukaryotic methyltransferase. Unmethylated targets are typically found in the DNA of invading phage and plasmid and trigger the restriction enzyme R.EcoKI to cleave the foreign DNA into harmless fragments (Figure 1-3).

Atomic force microscopy (AFM) was used to investigate EcoKI translocation during the restriction reaction. The results show two EcoKI molecules dimerising on the DNA prior to addition of ATP and translocation [Berge *et al*., 2000, Ellis *et al*., 1999, Neaves *et al*., 2009]. After analysing recent AFM results [Neaves *et al*., 2009] and the previous model of Studier and Bandyopadhyay [Studier *et al*., 1988, Neaves *et al*., 2009] built a new model (Figure 1-4). First, an EcoKI monomer binds to DNA at a non-specific site where it may either dissociate or conduct a limited degree of linear diffusion up and down the DNA contour [Halford, 2004]. During this diffusion process, the enzyme may find its specific target sequence and a second EcoKI monomer could bind to the EcoKI that is already bound on the DNA to form a dimer. Due to the inherent flexibility of the DNA, the EcoKI dimer that is bound at the target site could collide with another region of the DNA to form a loop. This second binding region could be either a second specific sequence site or a non-specific DNA region. Once the loop has formed, the second EcoKI molecule can search the DNA

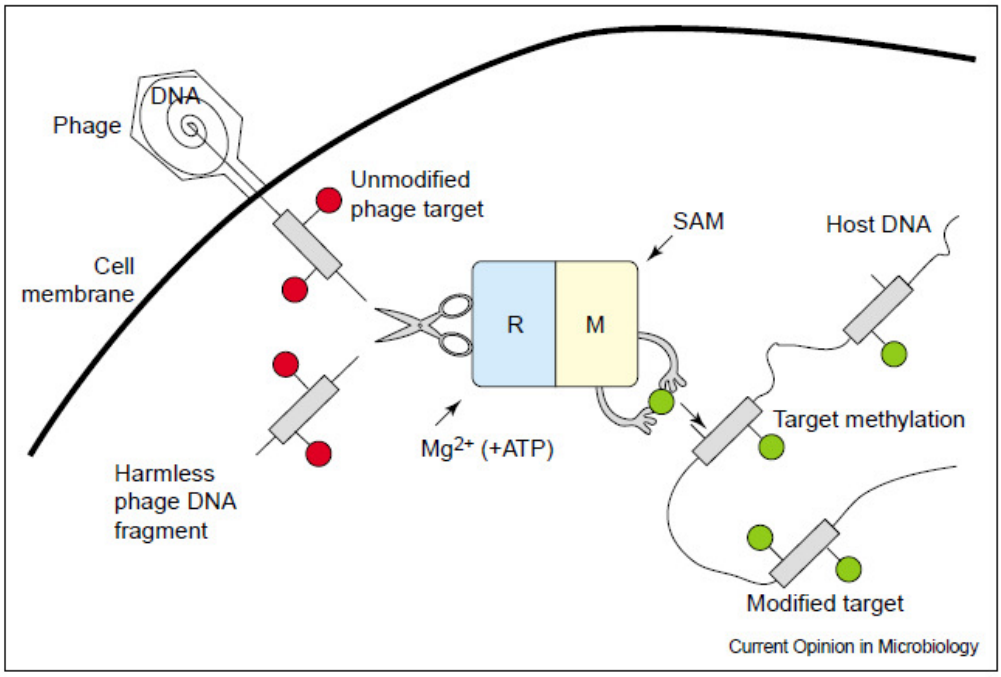


Figure 1-3: The function of R-M systems, as illustrated by a Type I R-M enzyme. These enzymes recognise the methylation state of their specific target sequence. Fully methylated DNA (shown as two green circles on the target sequence on the host DNA) is recognised to be part of the bacterial genome. Hemimethylated DNA (a single green circle on the host DNA target sequence) is recognised as newly replicated bacterial DNA, and the methyltransferase (M) modifies the other strand by methylation using the cofactor SAM. However, invading DNA, for example a phage genome, generally lacks specific modification (red circles on the target sequence of phage DNA) and is recognised to be foreign by the REase (R) and cleaved into harmless fragments (including legend repriented from Tock *et al*., 2005).

to find another target sequence. During the locating process, the loop may disassociate and then a new loop or even multiple loops may form. Once both EcoKI molecules of the dimer find target sequences, ATPase-driven translocation and DNA cleavage can be initiated. This translocation and restriction model is shown in Figure 1-4.

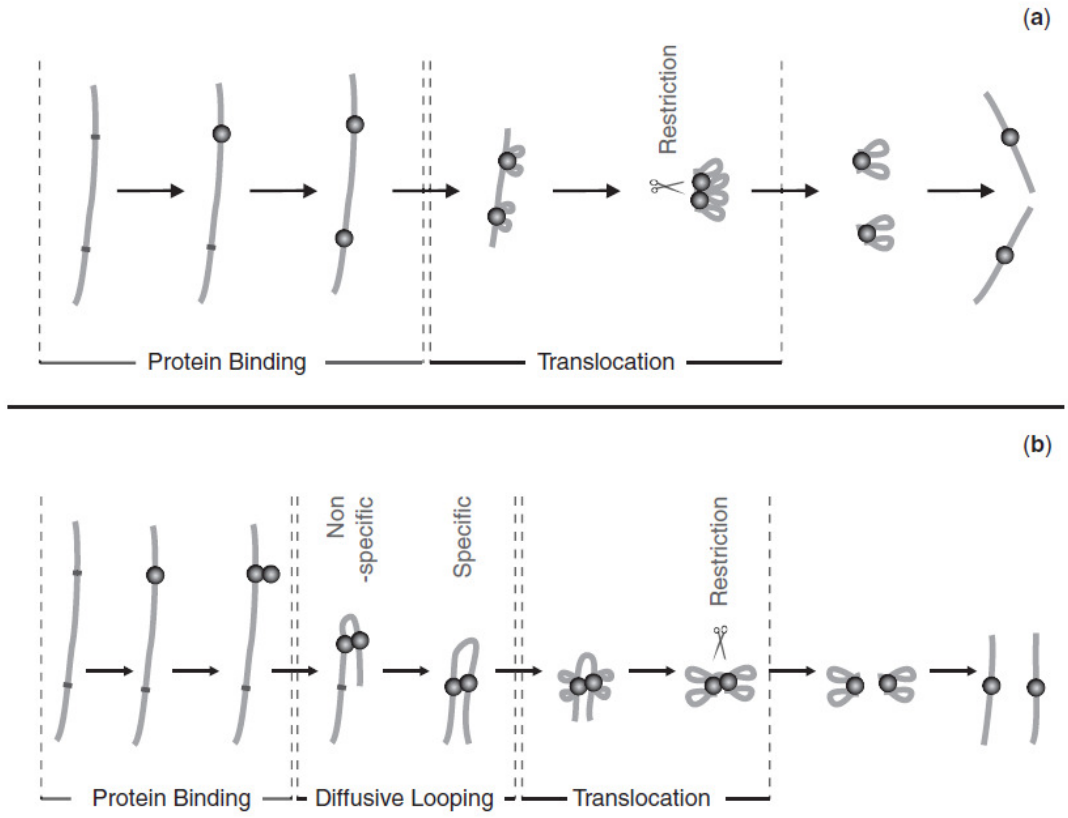


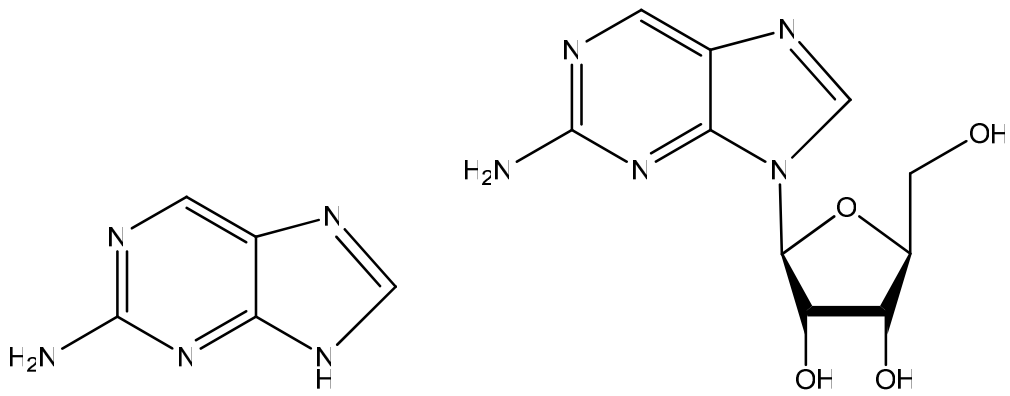
Figure 1-4: Diagrammatic representations of EcoKI binding, translocation and restriction of DNA. Part (a) shows the model of Studier & Bandy [Studier *et al*., 1988]. In this model EcoKI monomers individually bind to each site and translocation occurs independently from both occupied sites. When the monomers meet, translocation is stalled and restriction occurs. Part (b) shows a new model of Neaves *et al*., [Neaves *et al*., 2009] in which a monomer of EcoKI binds to one site and then a second monomer binds to the same site to form a dimerised complex at that site. This dimerised complex then forms diffusive loops with non-specific regions of DNA until it is stabilized by contact with the secondary EcoKI site. Translocation then occurs from both sides of both monomers and, in agreement with the previous model, restriction occurs when the translocation process is stalled (this time because the diffusive loop between the monomers becomes fully contracted). It remains uncertain whether translocation is triggered by occupation of a secondary site or whether both processes occur concurrently. In both models 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 (including legend reprinted from Neaves *et al*., 2009).

Depending upon the methylation state of the DNA, the Type I R-M enzyme can function as either a REase or a methyltransferase. If the target is fully methylated DNA, it recognises this DNA as being part of self-DNA and no action taken. If the target sequence is hemi-methylated, the DNA is targeted for further methylation. If the target sequence is unmethylated DNA then it is recognised as foreign-DNA and is targeted for restriction. Unmethylated DNA will be cleaved into harmless DNA fragments (Figure 1-3). The restriction reaction requires SAM, ATP and Mg^{2+}. Cleavage occurs anywhere between at least 40 bp [Dreier *et al*., 1996] and several thousand base pairs from the target site [Studier *et al*., 1988].

1.1.1.1 Base flipping

Before the methyltransferase catalyses the transfer of a methyl group from the cofactor SAM to the target base within the recognition sequence, the enzyme needs to bend the bases of the DNA around the recognition sequence, thereby leading to base flipping. Unfortunately, a crystal structure of a type I methyltransferase the in complex with DNA is not available. Indirect methods, such as spectroscopy, have to be used to investigate the mechanism of the base flipping. The base analogue 2-aminopurine (2AP) is a fluorescent molecule with a structure very similar to that of adenine (Figure 1-5). 2AP can form two hydrogen bonds with thymine in DNA [Nordlund *et al*., 1989]. The fluorescence quantum yield of 2AP is about 0.6 in aqueous solution [Neely *et al*., 2004] compared with that for adenine of 10^{-3}. Figure 1-6 is the fluorescence excitation and emission spectrum of 2AP in aqueous solution. When 2AP is incorporated into DNA, the fluorescence of 2AP is quenched [Allan *et al*., 1996]. When methyltransferase binds to the 2AP DNA there is a significant increase in both fluorescence intensity [Allan *et al*., 1996] and fluorescence lifetime [Neely, 2004 and 2005]. Figure 1-7 is the model of 2AP base flipping.



(a) (b)

Figure 1-5: The structures of (a) 2AP, (b) 2AP riboside.

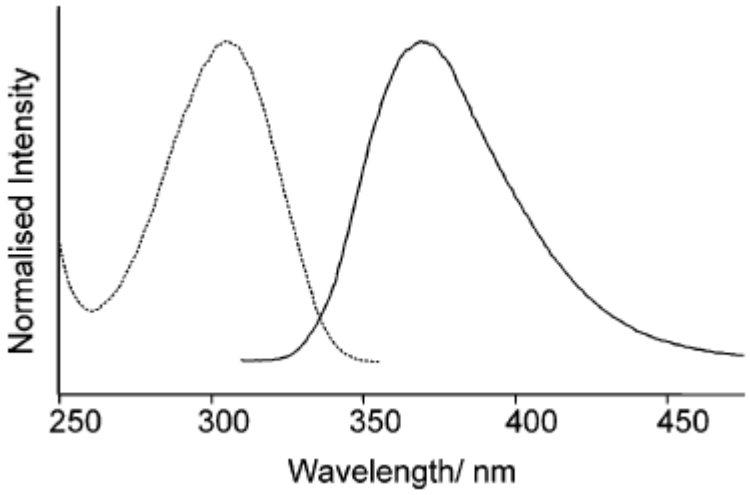


Figure 1-6: Fluorescence excitation (dotted line) and emission (solid line) spectra of 2AP in aqueous solution (reprinted from Neely *et al*., 2004).

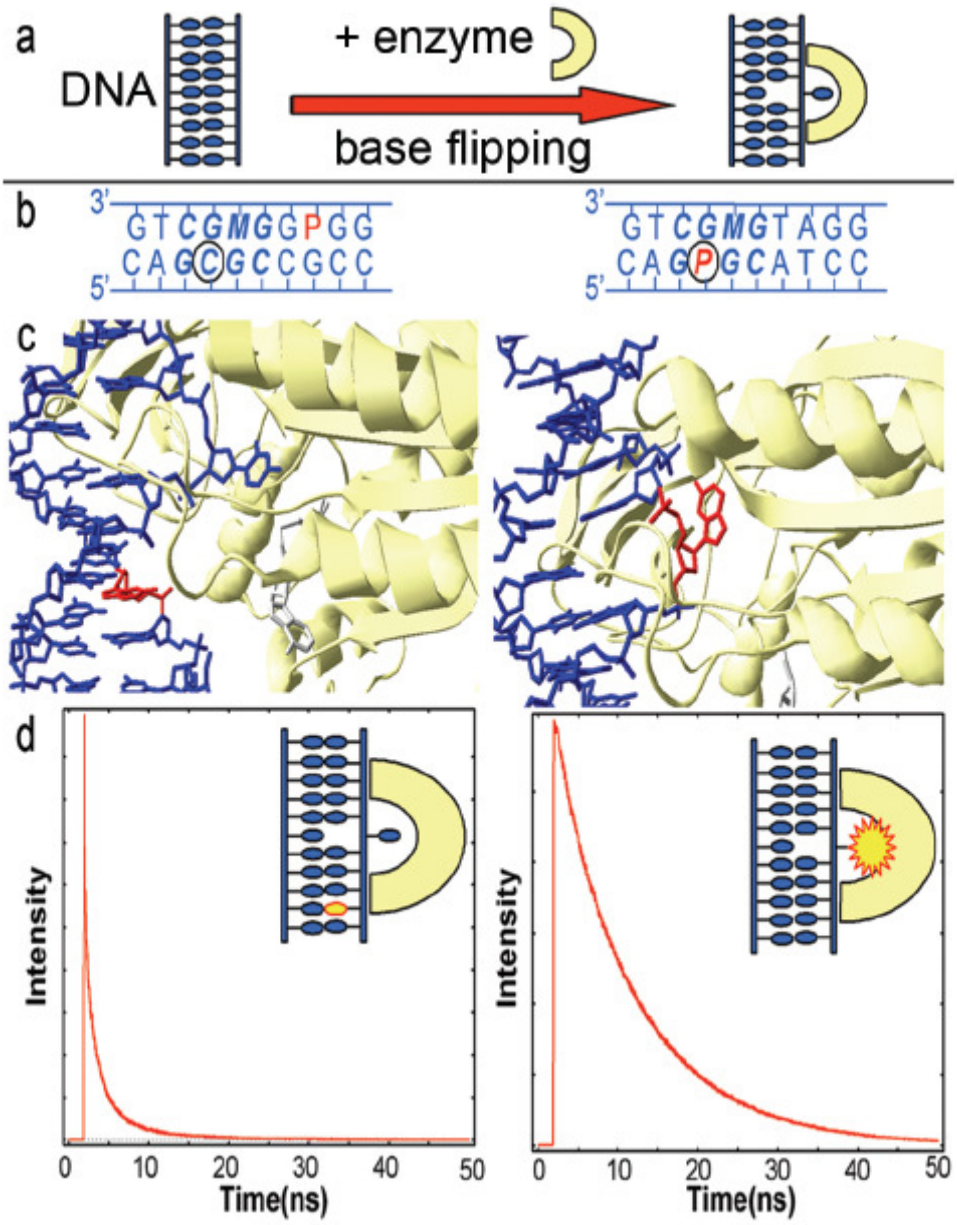


Figure 1-7: The 2AP fluorescence decay response with base flipping. (a) A schematic representation of base flipping. (b-d) show after M.HhaI bound the DNA with AP which is outside (c, lefthand panel) or within (c, righthand panel) of recognition sequence, the fluorescence decay of AP is different (reprinted from Neely *et al*., 2005).

1.1.1.2 MTase structure

To date, no X-ray crystal structure of a Type I methyltransferase, including MTase, has been solved. However, based on the known X-ray crystal structures of individual M and S subunits, several models of methyltransferase have been proposed [Obarska *et al*., 2006, Kennaway *et al*., 2009].

The crystal structures of HsdS from *Methanococcus jannaschi* and *Mycoplasma genitalium* [Kim *et al*., 2005, Calisto *et al*., 2005] have recently been published. The S subunit from *M. jannaschi* is a monomer composed of two target recognition domains (TRDs) and two conserved regions (CRs) (Figure 1-8). A monomer consists of four successive structural domains; (i) globular TRD1, (ii) long helical CCR domain, (iii) globular TRD2, and (iv) C-terminal DCR helix. 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].

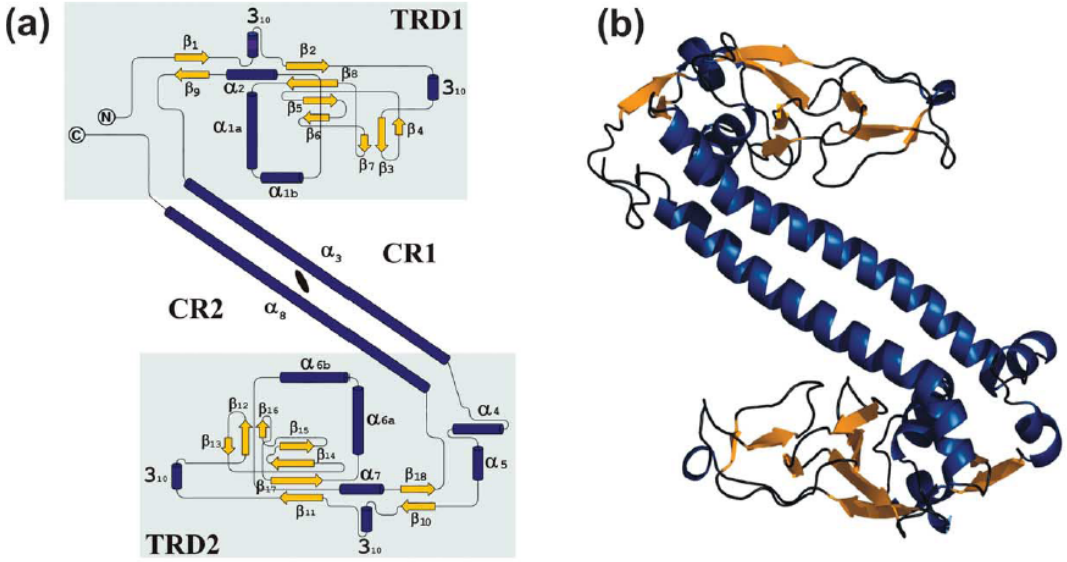


Figure 1-8: (a) Scheme of the HsdS MG438 topology. The helices are depicted as blue cylinders and β-strands as yellow arrows. (b) Ribbon structure of the HsdS MG438 (reprinted from Calisto *et al*., 2005).

A model structure of a Type I R-M REase subunits assembly is also shown (Figure 1-9) [Dryden *et al*., 1995].

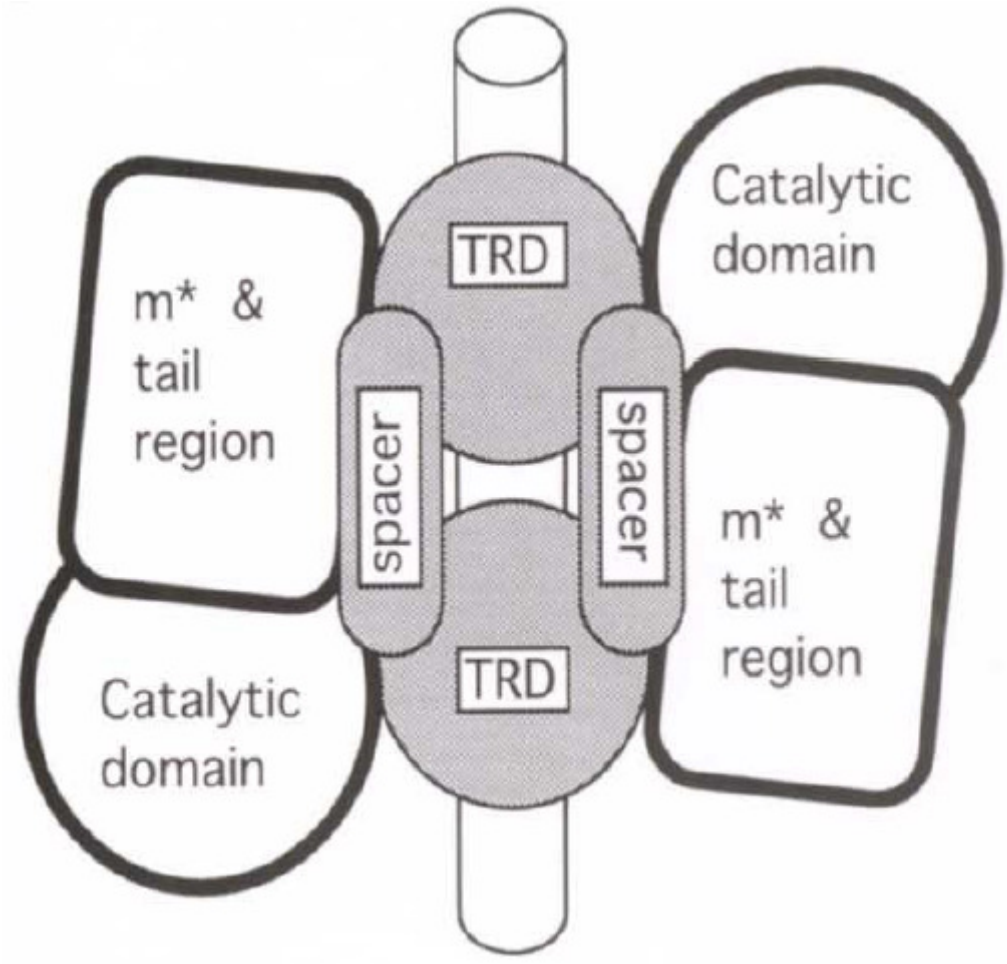


Figure 1-9: Model of MTase bound to DNA (reprinted from Dryden *et al*., 1995).

Obarska *et al*., [Obarska *et al*., 2006] have produced a structural model of the M.EcoR124I MTase comprising the HsdM (EcoR124I) and HsdS (EcoR124I) subunits, based on the crystal structures of HsdS (MjaXIP) [Kim *et al*., 2005], HsdS (MgeORF438P) [Calisto *et al*., 2005] and M.TaqI–DNA complex [Goedecke *et al*., 2001] (Figure 1-10).

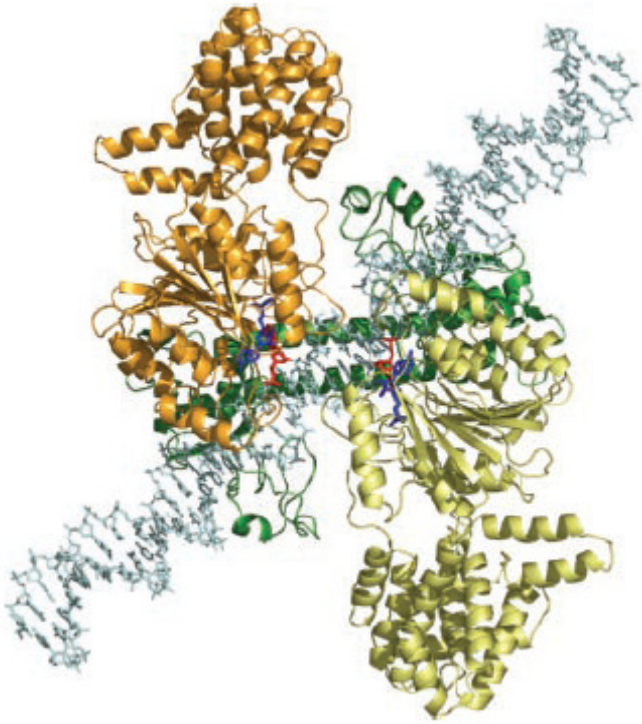


Figure 1-10: Predicted structure of the M.EcoR124I DNA MTase and substrate (reprinted from Obarska *et al*., 2006).

Recently, Kennaway *et al*., [Kennaway *et al*., 2009] have produced a ~18Å resolution structure of MTase bound to Ocr, an antirestriction protein, by negative-stain electron microscopy (EM). Combining the EM model with known crystallographic structures of individual subunits allowed the generation of an atomic model of the MTase enzyme (Figure 1-11). This model suggests a dynamic opening and closing of the protein during binding of MTase with either DNA or an antirestriction protein, such as Ocr or ArdA. This opening and closing of the protein is driven by a flexing and twisting of the conserved coiled-coil region within HsdS, which is required to expose the HsdM–HsdM interface, in a clamp-like fashion.

Figure 1-11: Model of the MTase in complex with substrate DNA and AdoMet cofactors. HsdS is shown in yellow, the HsdM at the rear is shown in grey, DNA strands are shown in blue and cyan with phosphates in black, the two AdoMet molecules in purple (reprinted from Kennaway *et al*., 2009).

Characterisation of the HsdR subunit of EcoR124I indicate that the subunit is monomeric and globular by using analytical ultracentrifugation and small angle neutron scattering data [Obarska *et al*., 2008]. Recently, the crystal structure of the HsdR subunit of EcoR124I was found. The crystal structure revealed that there are four globular domains with a square-planar arrangement in the HsdR subunit; endonuclease domain (residues 13–260), two RecA-like helicase domains (261–461 and 470–731) and a helical domain (732–892). Unresolved residues 893 to 1038 are not shown (Figure 1-12) [Lapkouski *et al*., 2009].

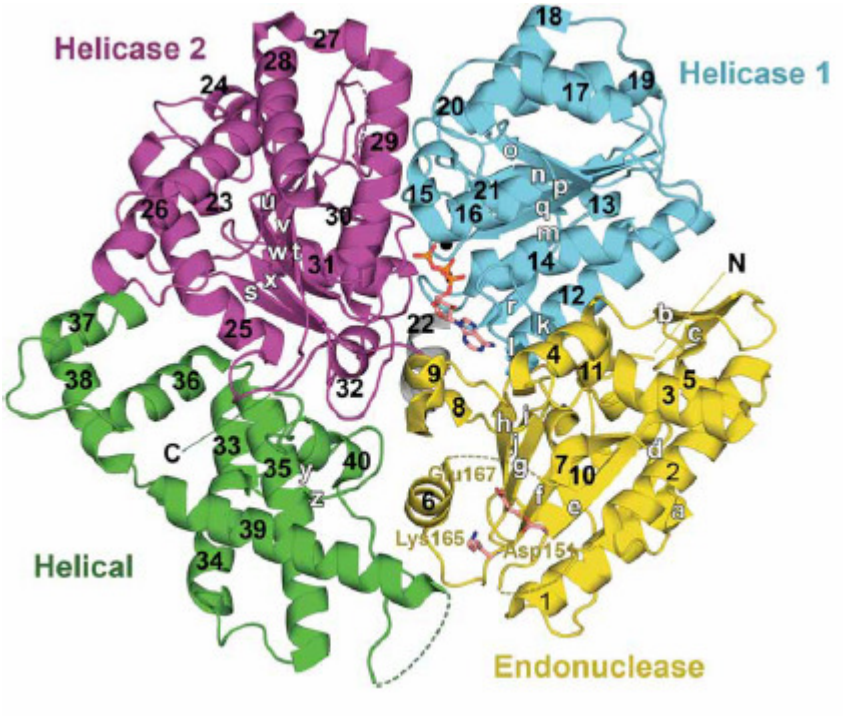


Figure 1-12: Crystal structure of the HsdR subunit of EcoR124I. Four domains are in different colour: Endonuclease (residues 13-260), helicase 1 (261-461), helicase 2 (470-731) and helical domain (732-892) (reprinted from Lapkouski *et al*., 2009).

1.1.2 Type II systems

The Type II R-M systems are the simplest and the most abundant of the R-M systems. In Type II R-M systems, REases and methyltransferases are usually independent enzymes and recognise the same sequence (usually a 4-8 base pair (bp) palindrome). Type II R-M systems require fewer cofactors than Type I systems. The restriction REase only requires Mg^{2+}, and the methyltransferase requires SAM as a cofactor. Some Type II REases act as homodimers, each monomer cuts one strand of DNA, one homodimer generate double-strand breaks. Other Type II REases are monomers or tetramers. Before cleaving DNA, many of them must bind to more than one recognition site. Cleavage by Type II REases occurs symmetrically within or very close to the recognition sequence, creating defined restriction products. Thus Type II R-M systems are very useful tools in molecular biology.

1.1.3 Type III systems

Type III R-M enzymes are similar to Type I R-M enzymes and consist of two kinds of subunits-Mod and Res. Mod is about 75 kDa and responsible for DNA recognition and modification, Res is about 106 kDa and response for DNA cleavage. When they associate to res_2mod_2, it acts as a functional restriction enzyme. Type III R-M enzymes recongize asymmetric and interrupted 5-6 bp sequency. The Mod subunit can act as a functional methyltransferase on its own, methylating on only one strand to produce a hemi-methylated DNA using SAM as a cofactor. ATP is used to recognise, translocate and restrict DNA. ATP and Mg^{2+} are the cofactors for cleavage [Meisel *et al*., 1995]. Type III REases require the presence of two oppositely oriented DNA recognition sites [Meisel *et al*., 1992] and only cleave the DNA that is unmethylated on both sides. The restriction reaction is stimulated by SAM, cleaving 25-27 bases away from the recognition site. EcoP1I and EcoP15I are best studied examples [Bickle *et al*., 1993]

1.1.4 Type IV systems

The biggest difference compared to other systems is that Type IV Systems only cleavage modified DNA (methylated, hydroxymethylated and glucosyl-hydroxymethylated). Thus Type IV Systems have no associated modification system and are called methylation-independent restriction systems. McrBC from *E. coli* K12 is the best studied Type IV system. For McrBC, Mg^{2+} is its cofactor and GTP hydrolysis is used to cleave and translocate DNA [Pieper *et al*., 1997]. This enzyme recognises two sites that are seperated by between 40 and 3000 nucleotides, and the cleavage point is 30 bp away from one of the sites.

1.2 Anti R-M systems

R-M systems are not perfect defence systems to protect bacteria from invading foreign DNA. Antirestriction strategies have been employed by phages and plasmids in order to avoid restriction and survive in the bacterial host cells. Strategies such as

modification of phage genome, transient occlusion of restriction sites, subversion of host R-M activities, and direct inhibition of restriction enzymes are used [Tock *et al*., 2005].

1.2.1 DNA sequence alteration

REases only restrict foreign DNA with one or more copies of the specific sequence. Phages and plasmids could remove or change these potential sites on their genome to evade restriction. In phage T3 and phage T7, the distance of recognised sites is too long to be recognised by EcoRII, an enzyme which requires two copies of target sequence. Thus these phages are resistant to cleavage [Kruger *et al*., 1988, Halford *et al*., 2004]. Some phages incorporate unusual bases into the DNA sites. T-even phage genomes contain the unusual base hydroxymethylcytosine [Kruger *et al*., 1983]. Some phages even encode their own methyltransferase that methylates and protects their DNA within the bacterial host. This phenomenon was found in SPβ phage of *B. Subtilis* [Warren, 1980].

1.2.2 Transient occlusion of restriction sites

Phage and plasmid eject some DNA-binding proteins which transiently occlude restriction sites. Thus their DNA molecules gain protection from host R-M. For example, phage P1 encodes proteins DarA and DarB, which are coinjected with the phage DNA into bacteria. Type I R-M enzymes cannot restrict the phage DNA where DarA and DarB are bound [Iida *et al*., 1987].

1.2.3 Subversion of R-M activities

There are two ways to subvert bacterial R-M activities: destruction of SAM cofactor and stimulation of host methyltransferase to modify phage DNA.

As Type I and Type III R-M enzymes require SAM for activity, phage T3 reduces the concentration of SAM in the host cell. The function of the REase is alleviated in

this way. Phage T3 encodes a SAM hydrolase that destroys SAM within the host cell after infection [Studier *et al*., 1976]. It does not inhibit R-M enzymes that are already bound to SAM or do not require it.

1.2.4 Inhibition of Type I R-M enzymes

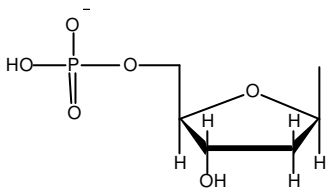
Some specific proteins inhibit the function Type I R-M enzymes by directly binding to the specific DNA binding sites of Type I R-M enzymes. The most well studied anti-restriction proteins are the gene 0.3 protein of phage T7 (also known as the overcome classical restriction (Ocr) protein) and the ard family (ArdA, ArdB, ArdC). Ocr is the first product to be expressed by phage T7 after it infects a bacterium [Kruger *et al*., 1989, Studier *et al*., 1976]. Before injection of the remaining phage DNA, the phage first injects the Ocr gene for transcription. By mimicing the size, shape and electrical charge of 24 bp of bent B-form DNA, it blocks the DNA binding site of Type I R-M enzymes which then cannot cleave the phage DNA. Thus the phage propagate in the infected bacterium [Bandyopadhyay *et al*., 1985]. However, Ocr cannot deactive Type II R-M enzymes.

Before discussing how these proteins mimic DNA in detail, we briefly introduce the structure of DNA.

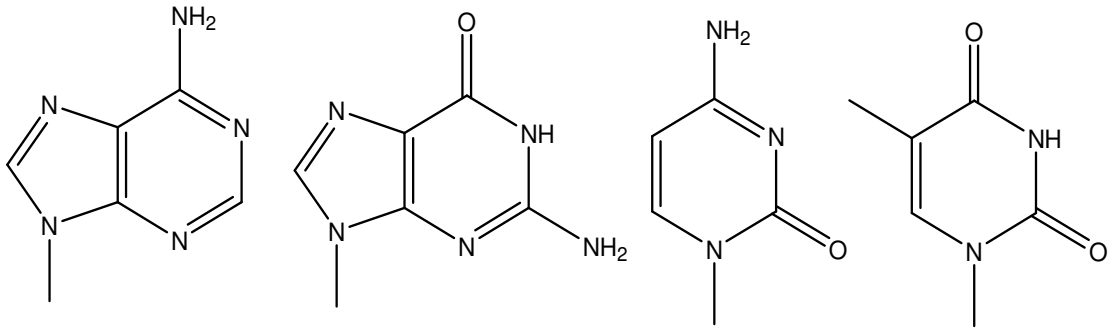
1.2.4.1 DNA structure

DNA (deoxyribonucleic acid) is found in all life forms and contains all the genetic information to replicate and sustain that particular organism. DNA is a long linear, unbranched and two-stranded polymer made from repeating units called nucleotides linked by phosphodiester bonds (Figure 1-12). 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).

(A) A nucleotide phosphate



B The four bases in DNA



Adenine (A) **Guanine (G)** **Cytosine (C)** **Thymine (T)**

Figure 1-12: Panel A: The general structure of a deoxyribonucleotide, found in DNA. Panel B: The four bases that occur in deoxyribonucleotides.

Although each nucleotide is very small, DNA polymers can be enormous molecules containing millions of nucleotides. Each nucleotide that has been incorporated into the polynucleotide is called a nucleotide residue. The sequence starts with a 5' phosphate and ends with a 3' hydroxyl.

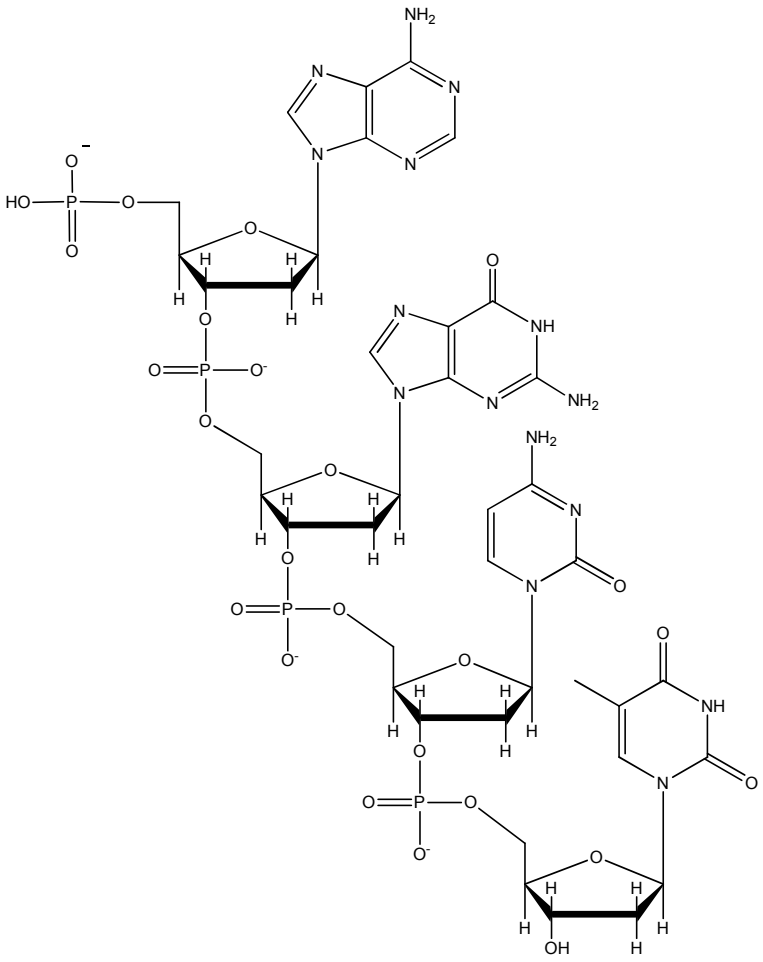


Figure 1-13: A short DNA polynucleotide showing the structure of the phosphodiester bond and the base sequence AGCT.

In the Watson-Crick model of DNA, the two polynucleotide chains of DNA wind around a common axis to form a double helix (Figure 1-14). Each polynucleotide chain is right-handed, but they are antiparallel. 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 in the core. Each base is linked by hydrogen bonds to a base in the opposite strand to form a planar base pair. Only two types of base pairs could be formed: adenine must pair with thymine and guanine residue must pair with cytosine. This phenomenon is knows as complementary base pairing. In this way the polynucleotides associate to form a double helix. The surface of the double helix contains two unequal width grooves: the major and minor grooves.

The features and characteristics of the double helix described by Watson and Crick correspond to the B-form of DNA. Double-helical DNA can assume several distinct structures depending on the solvent composition and base sequence. The major structural variants of DNA are A-DNA, B-DNA and Z-DNA (Table 1-2).

Table 1-2: The structural features of ideal A-, B-, and Z-DNA are summarised (reprinted from Donald *et al*., 1999).

	A	B	Z
Helical sense	Right handed	Right handed	Left handed
Diameter	~26 Å	~20Å	~18Å
Base pairs per helical turn	11	10	12 (6dimers)
Helical twist per base pair	33°	36°	60° (per dimer)
Helix pitch (rise per turn)	28 Å	34 Å	45 Å
Helix rise per base pair	2.6 Å	3.4 Å	3.7 Å
Base tilt normal to the helix axis	20°	6°	7°
Major groove	Narrow and deep	Wide and deep	Flat
Minor groove	Wide and shallow	Narrow and deep	Narrow and deep
Sugar pucker	C3'-endo	C2'-endo	C2'-endo for pyrimidines; C3'-endo for purines
Glycosidic bond	Anti	Anti	Anti for pyrimidines syn for purines

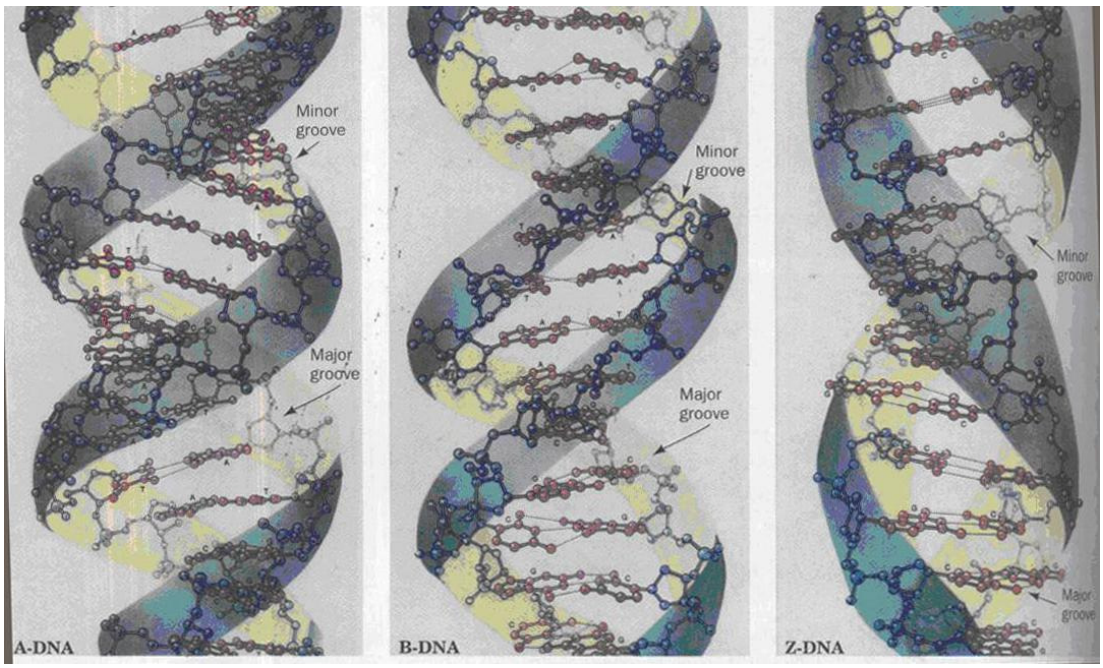


Figure 1-14: Structure of A-, B- and Z-DNA. View perpendicular to the helix axis. The sugar-phosphate backbones are outlined by blue-green ribbon and the bases are red (reprinted from Donald *et al*., 1999).

1.2.4.2 Ocr structure

The surface of the 116 amino acid wild-type Ocr, shows many negatively charged aspartate and glutamate side chain and imitates the distribution of negatively charged phosphate groups along the DNA double helix [Walkinshaw *et al.,* 2002]. Therefore, Ocr is a highly negatively charged protein (pI 4.02). The major structure of the Ocr monomer is constituted of four α-helices: A (residues 7–24), B (residues 34– 44), C (residues 49–57) and a long and bent one, D (residues 73– 106) (Figure 1-15). Ocr exists in solution as a dimer. The helix C is a part of the interface determining the contact of monomers and stable dimer formation. The approximate length of the Ocr dimer is around 85 Å with a maximum diameter of 30 Å and an apparent bend of ~34°. It is unusual that no hydrogen bonds are formed at Ocr dimer interface. There are 15 nonbonded contacts around the hydrophobic interface.

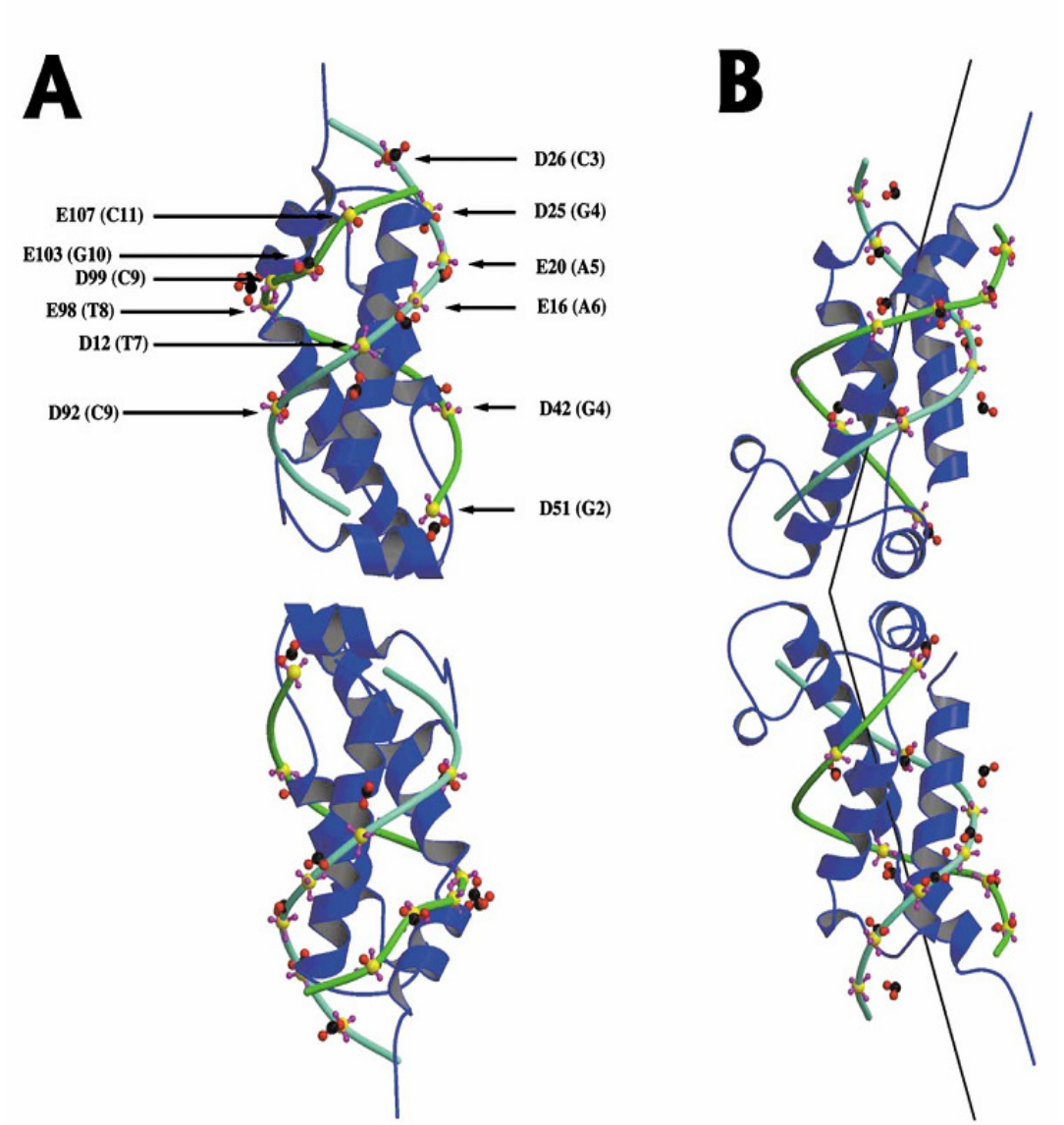


Figure 1-15: Superposition of two 12 bp B-DNA Molecules on the Ocr Dimer (A) A least squares fit of phosphate groups of the B-DNA complex (pdb code 1BNA) was made onto 12 carboxyl groups of ocr (see text) giving an rms fit of 1.9 Å (pairs of amino acid side-chains and phosphate used in the fit are labelled). The phosphorus atoms are coloured yellow, and oxygen atoms of the phosphate groups are coloured purple. The 12 carboxyl groups are coloured red (oxygen) and black (carbon). The sugar backbones of the DNA dimer chains are coloured in two shades of green. For clarity the base pairs are omitted. (B) The view is identical to that in Figure 1B and coloured as in (A). The 2-fold axis lies

in the plane of the paper. The vectors describing the direction of the fitted DNA (17) on both halves of the dimer are drawn as black lines. Their intersection gives a bend angle of 33.6^0 (including legend reprinted from Walkinshaw *et al*., 2002).

Figure 1-15 indicates that Ocr monomer shape and charge distribution mimic B-DNA. The structure of the Ocr dimer, both in solution and in crystal form is similar in length and charge distribution to 24 bp of bent B-DNA. The bend angle between the fitted DNA duplexes is 33.60 [Walkinshaw *et al*., 2002]. The stoichiometry of binding of WT-Ocr dimer to Mtase is 1:1 [Atanasiu *et al*., 2002]. The binding affinity of the Type I R/M enzyme MTase to DNA target sequence (nanomolar) [Powell *et al*., 1993] was far weaker than the affinity for binding to Ocr (picomolar) [Atanasiu *et al*., 2002].

1.2.4.3 ArdA protein structure

Ard (alleviation of restriction of DNA) proteins are used by many plasmids to overcome the host restriction system. ArdA, like Ocr, is a small, very acidic protein comprised of 140-170 amino acid residues and full of negative charge on its surface [Zavilgelsky *et al*., 2009]. It is a very efficient inhibitor against type I RM systems, allowing the DNA of the plasmid to evade R-M. Based on the amino acid sequence homology, Ard proteins are divided into three groups, ArdA, ArdB and ArdC. Every Ard has a small similar antirestriction motif made by 14 amino acid residues [Belogurov *et al*., 1995].

The first plasmid antirestriction protein was discovered on the enterobacterial plasmid pKM101 [Belogurov *et al*., 1985]. Two nonhomologous antirestriction proteins, ArdA and ArdB, are encoded by pKM101. The ArdA from pKM101 inhibits the Type I R-M enzyme, but the ArdB is ineffective against the modification system [Belogurov *et al*., 1993].

Orf18 from the Tn 916 transposon is the best investigated ArdA. The crystal structure of Orf18 has recently been reported [McMahon *et al*., 2009]. Each monomer can be decomposed into three domains: the N-terminal domain 1 (residues 3-61), the central domain 2 (62-103) and the C-terminal domain 3 (residues 104-165) (Figure 1-16). The three domains of the ArdA monomer are arranged in an approximately linear manner giving a very elongated molecular shape (7.0 nm×2.0 nm) with a definite curve. The two monomers of protein formed a tail-to-tail dimer. The dimer, like the monomer, is highly elongated and curved and has a length of 140 Å with a large negative charge on its surface. Its structure is a very good mimic of ~42 bp of bent B form DNA Figure 1-16.

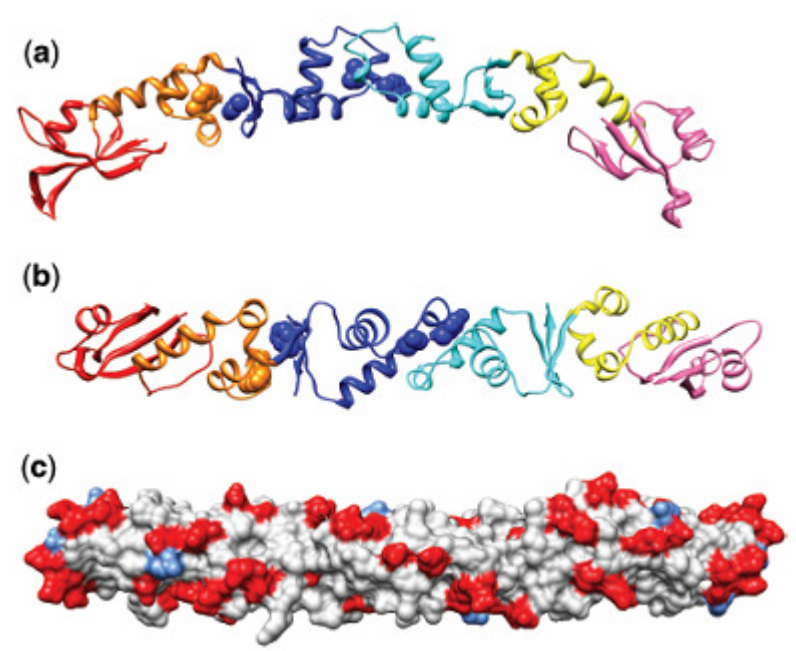


Figure 1-16: The dimeric structure of ArdA and its distribution of negative charges clearly show mimicry of the DNA double helix. In panel a and b the ArdA dimer with domain 1 coloured red and salmon, domain 2 coloured orange and yellow and domain 3 coloured blue and cyan for the left hand and right hand monomer, respectively. In panel c, acidic residues coloured red and basicresidues coloured light blue (reprinted from McMahon *et al*., 2009).

1.3 Fluorescence Spectroscopy

Fluorescence spectroscopy is a particularly valuable, extremely sensitive and simple technique in biological science. After a molecule is excited by absorption of one or two photons, one electron in the molecule jumps to an excited orbital. Because the excited state is not as stable as the ground state, the electron has to go back to the ground state and a large amount of energy must be lost. The excited electrons go back to the lowest virbrational level of the first excited state by internal conversion and vibrational relaxation. During this period heat will be released. Then three things may happen:

- the election may drop to the ground state by emitting a photon (this phenomena is fluorescence)

- it may convert to a triplet state via intersystem crossing, the molecule may either drop to the ground state and emit a photon (this phenomena is phosphoresce) or undergo a nonradiative transition.

- may go to the ground state without emitting a photon via a nonradiative transition

All these processes are illustrated in Figure 1-17.

Time-resolved fluorescence use a pulsed light source. Steady state fluorescence is obtained by using a continuous light source to excite sample and continually collecting the fluorescence.

In proteins three aromatic amino acids, tryptophan, tyrosine and phenylalanine, are fluorescent and act as intrinsic fluorescent probes. The UV absorption maxima of these amino acids in water are 280 nm for tryptophan, 275 nm for tyrosine and 258 nm for phenylalanine. In solution, phenylalanine exhibits an emission spectrum with a maximum around 282 nm, tyrosine has a maximum around 303 nm and tryptophan

has a maximum around 350 nm. Usually tryptophan is excited at 295 nm, which avoids excitation of tyrosine. Tryptophan is the most valuable fluorescent probe in proteins, since it is highly sensitive to its local environment and polarity.

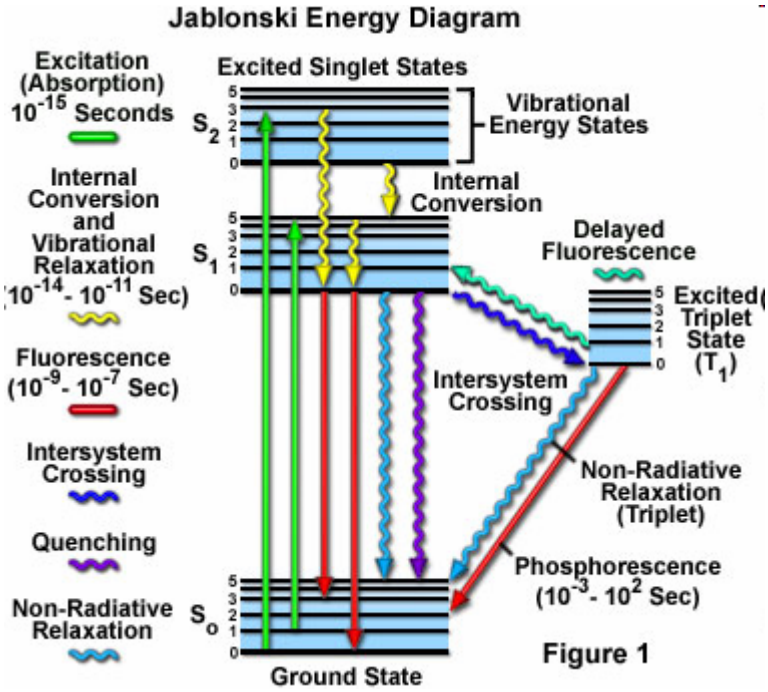


Figure 1-17: One form of a Jablonski diagram (reprinted from http://micro.magnet.fsu.edu/primer/techniques/fluorescence/fluorescenceintro.html) showing the lifetimes of absorption, fluorescence and phosphorescence. The singlet ground, first and second electronic states are depicted by S_0, S_1 and S_2, respectively. At each of these electronic energy levels, the fluorophores can exist in a number of vibrational energy levels, depicted by 0, 1, 2, etc.

When intrinsic protein fluorescence is not enough, fluorescence can be obtained by labelling with fluorescent probes which usually have longer excitation and emission wavelengths than aromatic amino acids. The protein-labeling probes can be covalently or noncovalently attached to the protein.

1.3.1 Time-resolved fluorescence

Time-resolved fluorescence is widely used to investigate biological macromolecules. Time-resolved measurements contain more information than steady-state data. For example, consider a protein containing two tryptophan residues. It is impossible to distinguish the emission from the two tryptophan residues from steady state fluorescence data because their individual absorption and emission spectrum overlap. However, time-resolved data may reveal each one has a distinct lifetime.

The fluorescence lifetime refers to the average time the electron stays in its excited state before returning to the ground state. Time-domain and frequency-domain methods are two methods of measuring time-resolved fluorescence in widespread use. In the time-domain, the sample is excited with a pulse of light duration shorter than the decay time of the sample. After the excitation pulse, the time dependent intensity is measured, and the decay time τ is calculated from the slope of a plot of log I(t) versus t Figure 1-18 and Equation 1.1. The fluorescence intensity decays are often obtained by exciting through a polarizer oriented at 54.7 degree from the vertical z-axis. In this way, the effects of rotational diffusion and/or anisotropy on intensity decay are avoided.

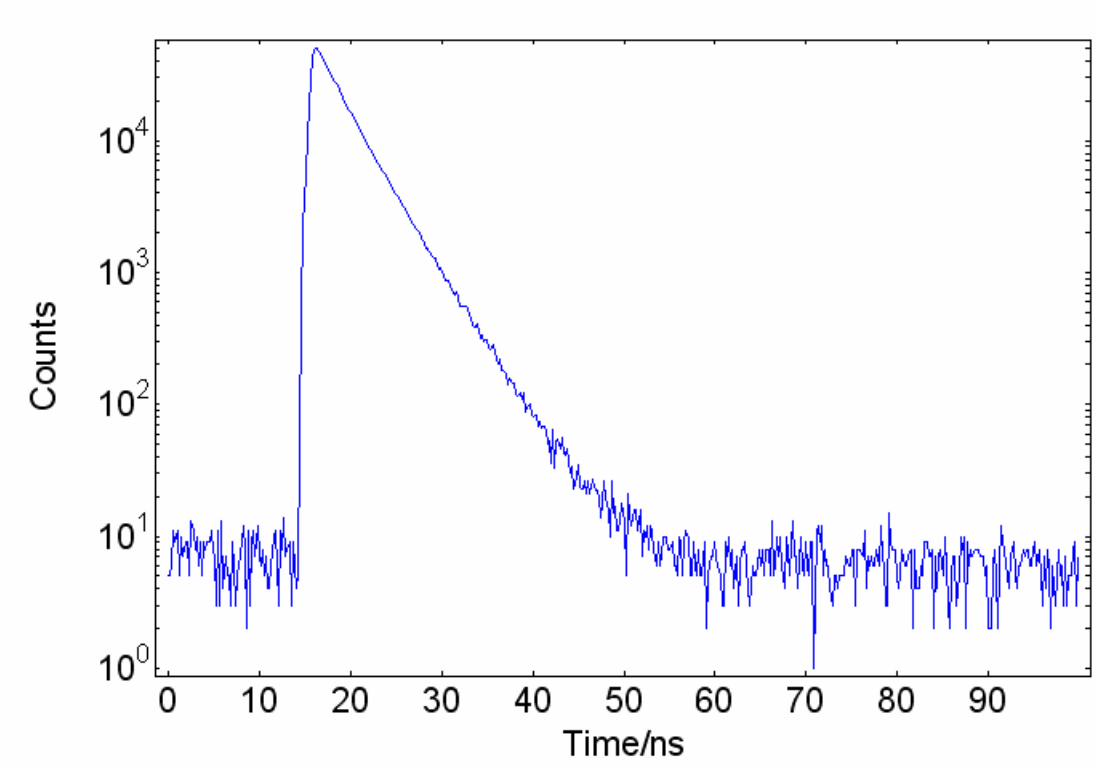


Figure 1-18: This is a typical fluorescence decay. Y-axis is cumulative photon counts.

Equation 1-1: $I(t) = I_0 \exp(-t/\tau)$

Where I_0 is the intensity at time 0, I(t) is the intensity at time t.

1.3.2 Time-correlated single-photon counting

Time-correlated single-photon counting (TCSPC) is the most popular method in time-domain measurements. TCSPC instruments use picosecond (ps) or femtosecond (fs) laser light sources, including pulsed lasers and light-emitting diodes (LEDs), and high-speed microchannel plate photomultiplier tubes. The sample is excited with a pulse of light, (Figure 1-19 A). For TCSPC, the conditions are adjusted so that less than one photon is detected per laser pulse. The excitation pulse that excites the sample also sends a signal to the electronics to start timing. After the detector receives one photon, timing stops for this pulse excitation (Figure 1-19 B). The time difference is measured between the excitation pulse and the observed photon and

stored in a histogram. The x- axis is time difference and the y-axis the number of photons detected in this time difference (Figure 1-19 C). A histogram of the decay is measured by repeating the pulse light excitation process numerous times.

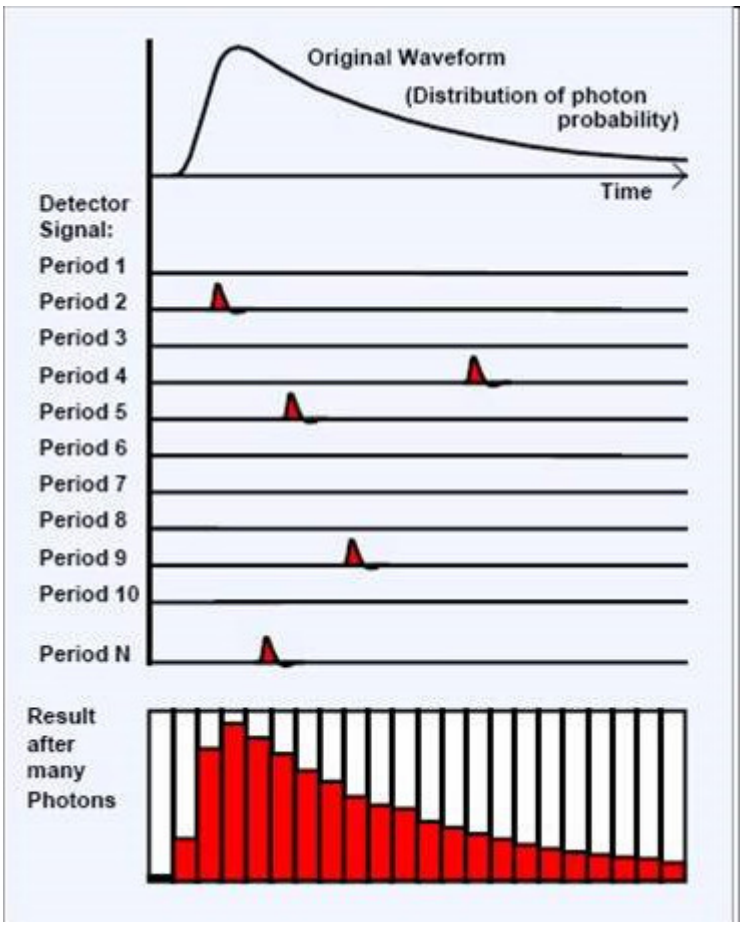


Figure 1-19: Principle of TCSPC. Panel a is the original waveform; panel b is the detected photons per laser pulse; c sum of all of detected photons.

http://microscopy.tw/molweb/experiment/MOL_FLIM_FRET/FLIMofPrinciple.html

1.3.3 Fluorescence anisotropy

Fluorescence anisotropy measurements are commonly used in biological fluorescence. Anisotropy data provide information on molecular environments, size and shape of proteins, and also the association of biomacromolecules. In a homogeneous solution, the ground state fluorophores are all randomly oriented. When samples are excited by polarised light, fluorophores preferentially absorb photons whose electric vectors are aligned parallel to the transition moment of the fluorophore. The extent of polarization of the emission is described in terms of the anisotropy (r).

The measurement of fluorescence anisotropy is illustrated in Figure 1-20. Usually the sample is excited with vertically polarised light. When the emission polariser is oriented parallel ($\parallel$)to the direction of polarised excitation the observed intensity is called $I_{\parallel}$ and when oriented perpendicular it is called $I_{\perp}$. These intensity values are used to calculate the anisotropy.

Equation 1-2:

$$r = \frac{I_{\parallel}/I_{\perp} - 1}{I_{\parallel}/I_{\perp} + 2}$$

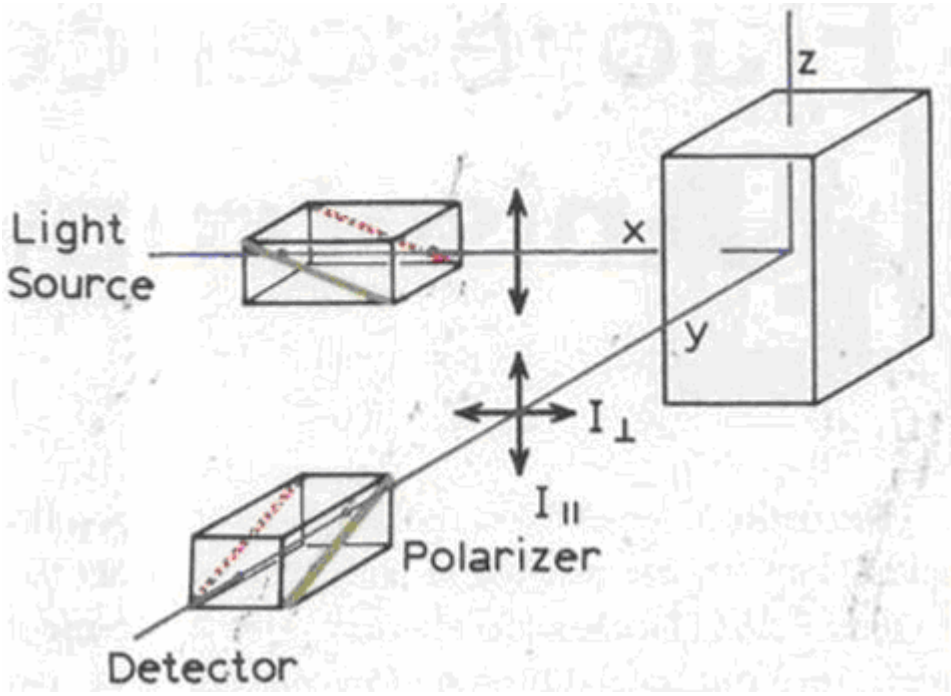


Figure 1-20: Schematic diagram for measurement of fluorescence anisotropy (reprinted from Lakowicz 1996).

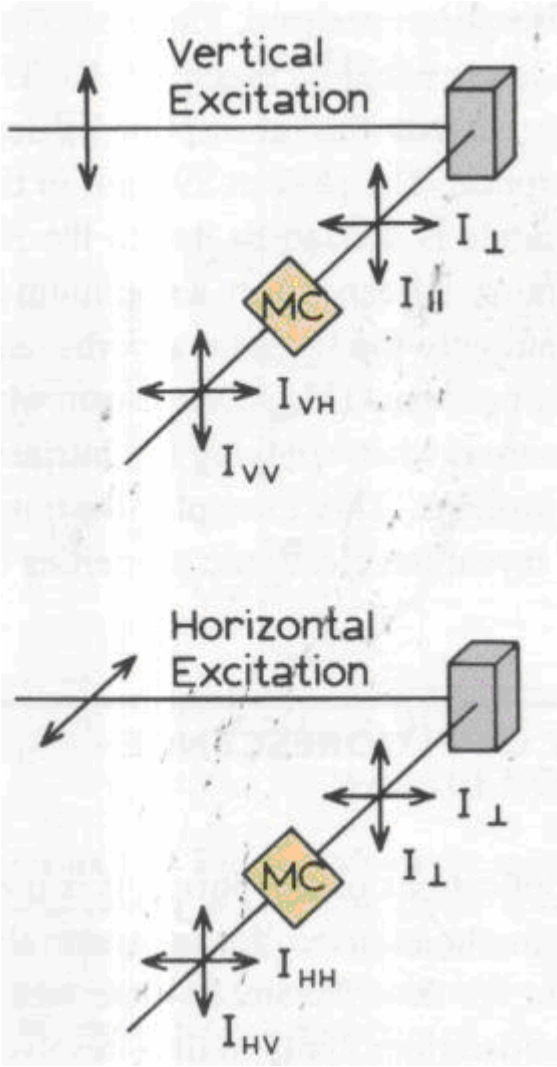


Figure 1-21: Schematic diagram for L-format measurements of fluorescence anisotropy. MC represents the monochromator. To calulate the anistropy the ratio of I_{VV} and I_{VH} is measured with the G factor. The G factor accounts for instrumental error by measuring the I_{HH} and I_{HV} (reprinted from Lakowicz 1996).

Figure 1-21 is a typical L format fluorometer. Assume the sample is excited by polarised light, and emission through the monochromator is observed by the detector. Unfortunately, the monochromator has a different transmission efficiency for vertical and horizontal polarised light. So the measured parallel and perpendicular intensities are changed by monochromator. The objective is to measure these actual $I_{\parallel}$ and $I_{\perp}$

intensites, rather than the detected I_{VV} and I_{VH}. Thus we need a factor G to get the true anisotropy value. To obtain the G factor, we need to excite the sample with horizontal polarised light, and measure I_{HH} and I_{HV}. The G factor will be obtained by Equation 1-3.

Equation 1-3: $$G = \frac{I_{HV}}{I_{HH}}$$

The anisotropy is given by

Equation 1-4: $$r = \frac{I_{VV} - GI_{VH}}{I_{VV} + 2GI_{VH}}$$

Factors that effect rotational diffusion such as the size and shape of the rotating molecule and the viscosity of solvent will affect anisotropy. For small fluorophores in low viscosity solutions, the rate of rotational diffusion is very fast and the anisotropy close to zero. Thus, the anisotropy measured through steady-state fluorescence can be used to follow molecular size changes that result from binding. When a biomacromolecule binds to a ligand, the rotational correlation time of the ligand decreases and the anisotropy of ligand fluorescence increases. The rotational rate is related to the rotational correlation time (θ) according to the Equation 1-5:

Equation 1-5: $$\theta = \frac{\eta V}{RT}$$

Where η is the viscosity of the solution, V the molecular volume, R the gas constant and T the temperature in Kelvin. For a single-exponential fluorescence intensity decay, the relationship between the anisotropy and rotational correlation time θ is given by the Perrin equation.

Equation 1-6: $$r = \frac{r_0}{1 + (\tau/\theta)}$$

Where r_0 is the intrinsic anisotropy from the angular difference between absorption and emission dipole.

1.3.4 Fluorescence resonance energy transfer

Because fluorescence resonance energy transfer (FRET) can be used to obtain the distance between energy transfer pairs, FRET is widely used in DNA/protein analysis and optical imaging. FRET occurs between a donor (D) molecule in the excited state and an acceptor (A) molecule in the ground state. The emission of the donor overlaps with the absorption spectrum of the acceptor. Energy transfer occurs via long range dipole-dipole interactions between donor and acceptor without emission of a photon from the donor. There are four factors affecting the efficiency of FRET:

1 The extent of spectral overlap of the emission spectrum of the donor with the absorption spectrum of the acceptor,
2 The relative orientation of dipoles of the donor and acceptor transition,
3 The quantum yield of the donor,
4 The distance between the donor and acceptor molecules

The most common application of FRET is to measure the actual distance between donor and acceptor on two sites of one macromolecule or between a macromolecule binding to a ligand or to DNA.

To calculate the FRET distance between donor and acceptor, we need to first calculate the Förster distance for 50% transfer efficiency (R_0) for donor and acceptor on the basis of Equation 1-7.

Equation 1-7: $R_0^6=8.78*10^{-5}\kappa^2\Phi Jn^{-4}$

Where n is the refractive index of the medium, κ^2 is the orientation factor of the donor and acceptor, n is the quantum yield of donor in the absence of acceptor, and J is the spectral overlap integral. The spectral overlap integral J is in units of $M^{-1}cm^{-1}(nm)^4$. The Förster distance in angstroms.

The spectral overlap integral J, between the donor emission spectrum and the acceptor absorbance spectrum is determined by using Equation 1-8,

Equation 1-8: $J(\lambda)=\int F_d(\lambda)\varepsilon_a(\lambda)\ \lambda 4 d\lambda / \int F_d(\lambda) d\lambda$

Where $F_d(\lambda)$ and $\varepsilon_a(\lambda)$ represent the fluorescence intensity of the donor and the molar extinction coefficient of the acceptor, respectively, at wavelength λ.

The efficiency of the energy transfer is usually calculated based on the decrease in the donor fluorescence intensity, Equation 1-9.

Equation 1-9: $$E = 1 - \frac{F_{ad}}{F_d}$$

Where F_{ad} and F_d represent the donor fluorescence intensity measured in the presence (F_{ad}) and absence of acceptor (F_d), respectively.

The efficiency of the energy transfer can also be calculated from the decrease in the fluorescence lifetime of the donor fluorescence, Equation 1-10.

Equation 1-10: $$E = 1 - \frac{\tau_{ad}}{\tau_d}$$

Where τ_{ad} and τ_d are the fluorescence lifetime in the presence (τ_{ad}) and absence of acceptor (τ_d), respectively.

The FRET distance (r) between donor and acceptor is calculated from Equation 1-11.

Equation 1-11: $$r^6 = \frac{R_0^6(1-E)}{E}$$

1.4 Green fluorescent protein

1.4.1 Introduction

In the early 1960s, green fluorescent protein (GFP) was identified in the jellyfish *Aequorea victorea* [Shimomura *et al*., 1962]. The work was originally carried out in order to understand the photophysics involved in *Aequorea* bioluminescence. Wild-type (WT) GFP is comprised of a single-polypeptide chain of 238 amino acid residues which folds to form a soluble globular protein (molecular mass of 27 kDa). In 1996, the crystal structure of GFP was resolved [Ormo *et al*., 1996, Yang *et al*., 1996]. GFP was found to adopt an 11 β-stranded fold that forms an almost perfect cylinder of dimensions 4.2 nm by 2.4 nm (Figure 1-22). GFP spontaneously forms a chromophore which is buried in the centre of the cylinder by the autocatalytic cyclization of the polypeptide backbone. The β-barrel isolates the chromophore from the bulk solvent. This encapsulation is beneficial for the fluorescence quantum yield (the ratio of absorbed to emitted photons) of GFP. An advantageous feature of GFP as a biological tool is that it can emit a visible green fluorescence without the need for any substrate. WT GFP absorbs 395 nm blue light with a smaller peak at 475 nm. When excitating WT GFP at 395 nm, it emits green light at 508 nm, or it can be excited at 475 nm and then the emission is at 503 nm. WT GFP has a high quantum yield about 0.72–0.85 [Morise *et al*., 1974, Ward *et al*., 1980]. Some factors, such as temperature, pH and ionic strength, effect the ratio between the two absorption peaks [Ward *et al*., 1982]. These results indicate that the chromophore of GFP has two different forms. The larger 395 nm peak is caused by a neutral chromophore and the minor peak at 475 nm is due to an ionised chromophore [Heim *et al*., 1994]. When GFP and GFP-like proteins are fused to another biomacromolecule, they retain their fluorescence and tend not to affect the function of the fusion partner. Thus GFP and GFP-like proteins have been widely used to track the movement of

biomacromolecules within cellular environments. They are a revolution in the study of cell biology and biomedical sciences. For this reason, three researchers, Osamu Shimomura, Martin Chalfie and Roger Tsien, won the 2009 Nobel Prize in chemistry, for their contributions to the discovery and development of the *Aequorea victoria* green fluorescent protein [Sample *et al*.,2009].

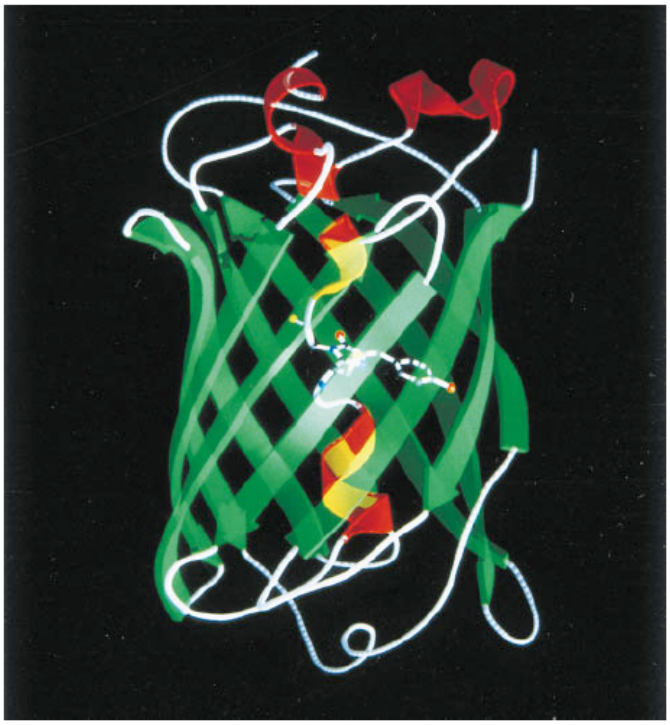


Figure 1-22: Ribbon diagram of WT GFP structure. The α-helices are shown in red, the β-strands are shown in green (reprinted from Kraulis, 1991).

GFP is a commonly used molecular imaging tool in biology, chemistry, genetics and medicine [Zimmer, 2009]. In 2006 more than 10,000 papers were published that used GFP or GFP-like proteins [Chalfie *et al.,* 2006].

1.4.2 The GFP chromophore

Synthesis of the GFP fluorophore occurs spontaneously after protein folding without cofactors or accessory proteins. The GFP chromophore, 4-(*p*-hydroxybenzylidene)-

5-imidazolinone (p-HBI), is formed in the central helix located in the hydrophobic core of the protein. The fluorescent p-HBI species is composed of a conjugated ring structure derived from a triplet of amino acids, Ser65-Tyr66- Gly67. Wild-type GFP can absorb light at 395 nm and 475 nm, due to the neutral and anionic chromophores, respectively [Heim, 1994]. WT-GFP emits green fluorescence after excitation at either of these wavelengths [Zimmer, 2009].

During chromophore maturation, the first chemical step involves the G67 amide nitrogen atom attacking the S65 carbonyl carbon to create a five-membered imidazolone ring, and dehydration of the S65 carbonyl oxygen. During the rate-determining step, molecular oxygen catalys dehydrogenation along the Cα–Cβ bond of Tyr66 to form the chromophore (Figure 1-23) [Zimmer, 2009].

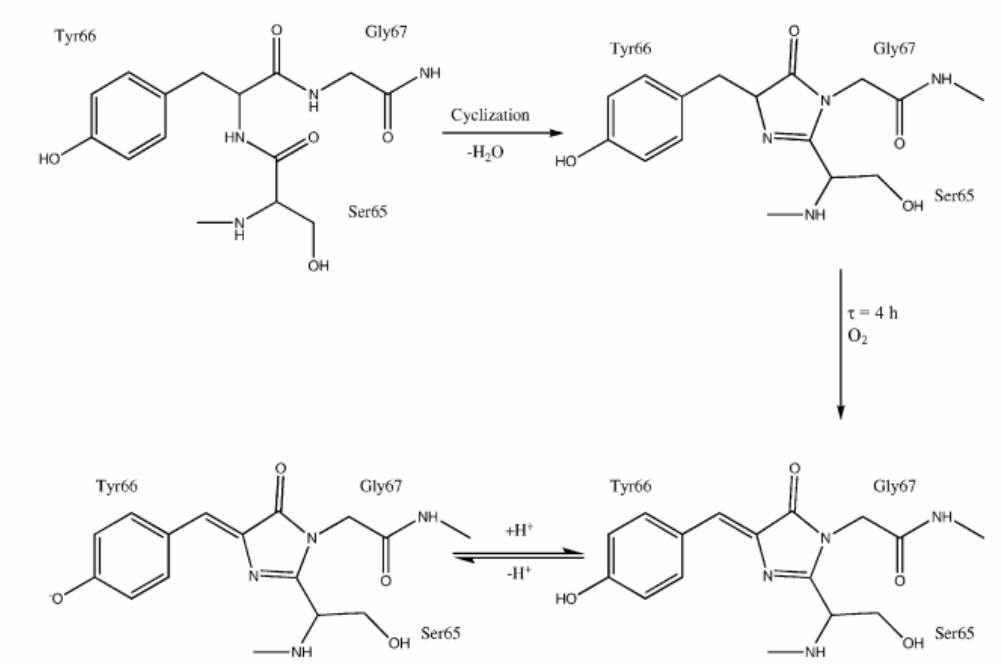


Figure 1-23: Proposed scheme for the formation of the GFP chromophore. The neutral form is excited at 395 nm while the anionic form is excited at 475 nm (reprint from Zimmer, 2009, Heim *et al*., 1994, Brejc *et al*., 1997).

Tsien found that the S65T GFP mutant has a six-fold increased brightness compared to the wild-type protein and only one excitation peak. The S65T GFP mutant is the most commonly used GFP, and is called enhanced green fluorescent protein (EGFP) [Heim *et al*., 1995, Cubitt *et al*., 1995].

1.4.3 Other fluorescent proteins

A variety of changes have been introduced to further improve the characteristics of GFP. These factors include enhancing the brightness, improving the folding efficiency, changing the absorption and emission properties and decreasing sensitivity to pH. Currently the fluorescence protein (FP) colour palette contains a broad range of genetic variants whose emission profiles span nearly the entire visible spectrum. FPs are generally divided into seven spectral classes based upon their emission maxima, including those emitting in the blue (BFPs; 440–470 nm), cyan (CFPs; 471–500 nm), green (GFPs; 501–520 nm), yellow (YFPs; 521–550 nm), orange (OFPs; 551–575 nm), red (RFPs: 576–610 nm) and far-red (FRFPs; 611–660 nm) spectral regions (Figure 1-24). To date, over 150 fluorescent or colourful GFP-like proteins have been reported [Zimmer, 2009].

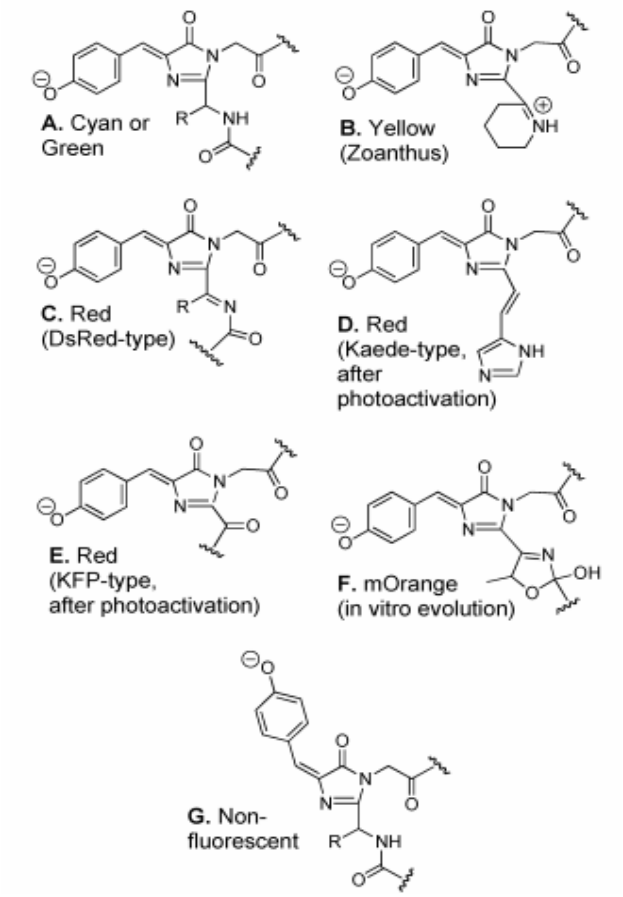


Figure 1-24: Chemical diversity of chromophores generated in GFP-like proteins [reprinted from Zimmer, 2009].

1.4.4 Some applications of GFP and GFP-like proteins

1.4.4.1 Cellular imaging

Compared with a chemical fluorescent marker, there are a few advantages for using GFP as biological marker,

Firstly, one can easily fuse GFP to peptides or proteins,

Secondly, GFP and GFP-like proteins are water soluble,

Thirdly, GFP and GFP-like proteins are not toxic for biological samples.

Therefore, GFP and GFP-like proteins are widely used as biological markers in biology. For example, GFP fusions have been used to investigate the location of two proteins, FtsZ and FtsA, in *Escherichia coli*, during bacterial division. Both FtsZ-GFP and FtsA-GFP are located in cells to form rings at the division site (midpoints of *E. coli*). It supports the idea of multiprotein septum complex [Ma *et al*., 1996]. GFP and GFP-like proteins are not only used for *in vivo* cell imaging but also for tissue imaging of live animals [Day *et al*., 2009]. For example, GFP was expressed in tumour cells in mice to investigate real-time studies of tumour progression, metastasis, and drug–response evaluations.

1.4.4.2 The bimolecular fluorescence complementation assay

GFP and GFP-like proteins are widely used in the bimolecular fluorescence complementation (BiFC) assay, which provides a way to visualize protein interactions and modifications in living cells. In this assay, two complementary fragments of a fluorescent protein are fused to interaction partners, respectively. Once the two interaction partners form a complex, the two complementary fragments of a fluorescent protein associate together and fluorescence can be observed. Thus, this assay can be used to visualise macromolecular interactions and modifications, such as proteins and nucleic acids in cellular functions. Figure 1-26 is a typical description of the BiFC assay. BiFC analysis has been used to visualise a wide range of cellular processes in several plants and animals. Recently Kerppola published two reviews describing the BiFC assay [Kerppola, 2008 and 2009].

1.5 Aim of the project

The aim of this project was to study the interaction between MTase, the cognate methylase of EcoKI, and various antirestriction proteins, including Ocr and four different ArdA proteins. In order to facilitate the analysis of complex formation between MTase and the various antirestriction proteins, a fused version of the methylase with GFP (GFP-MTase) was used. In this way, the fluorescent properties of the GFP molecule could be exploited to study the relative binding affinities of the ArdA proteins to MTase. GFP-MTase was also used in a series of FRET experiments

to determine the relative orientation of the molecule with respect to a bound DNA duplex or antirestriction protein. Here, GFP and a fluorescent tag on the DNA or antirestriction protein acted as the donor and acceptor pair for FRET, respectively. These experiments allowed the determination of the distances between the donor acceptor pairs. Such calculations will be useful for refining the model of the MTase molecule in complex with either DNA or Ocr.

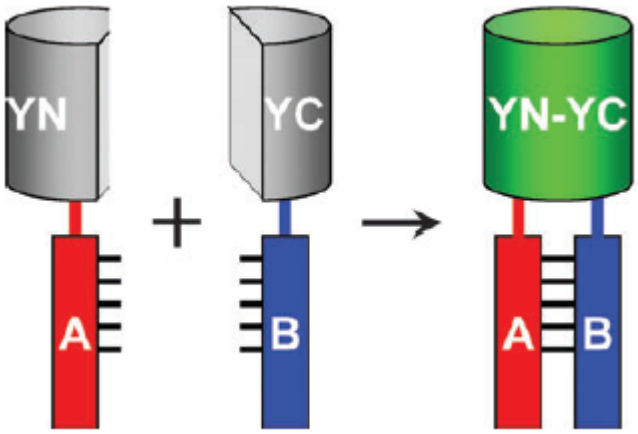


Figure 1-26: The principle of BiFC analysis (reprinted from Rajaram *et al.*, 2004). A and B represent interaction partners. YN and YC are two complementary fragments of a fluorescent protein. They are fused to A and B, respectively. When A interacts with B, the two fragments, YN and YC associate to form a fluorescent complex.

Chapter 2 Materials and methods

2.1 Chemicals and reagents

2.1.1 Chemicals

Dylight sulfhydryl-reactive fluors was purchased from Thermo scientific (Waltham, USA). Guanidine hydrochloride (Ultrol grade) was purchased from Calbiochem Inc. (San Diego, USA). Broad range pre-stained molecular mass markers for SDS-PAGE were purchased from BioRad (Precision Plus protein standards; Hercules, USA). All other reagents were purchased from Sigma-Aldrich (MO, USA). All solutions were made up in distilled, deioinised water, except HPLC/gel filtration buffer which used HPLC grade water.

2.1.2 Reagents

DNA loading buffer (10x): 20 % Ficoll 400, 0.1 M Na_2EDTA (pH 8.0), 1 % SDS, 0.25 % (w/v) bromophenol blue.

SDS PAGE destaining solution: 10 % acetic acid; 30 % methanol; 60 % H_2O.

SDS PAGE staining solution: 20 % acetic acid, 50 % methanol, 29.75 % H_2O, 0.25 % (w/v) Coomassie brilliant blue.

SDS Laemmli sample buffer (2x): bought from Sigma.

TBE (10x): 0.89 M Tris base; 0.89 M boric acid; 20 mM EDTA.

TE buffer: 10 mM Tris-HCl, 1 mM EDTA, pH 8.0.

TAE buffer: 40 mM Tris-HCl, 2 mM EDTA, 24 mM acetic acid, pH 7.7.

Buffer A: 20 mM Tris-HCl, 20 mM MES, 200 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 7 mM 2-mercaptoethanol (2ME), pH 6.5.

Buffer B: 20 mM Tris-HCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 7 mM 2ME, pH 8.0.

Buffer C: 20 mM Tris-HCl, 20 mM MES, 0.1 mM EDTA, 10 mM $MgCl_2$, 7 mM 2ME, pH 6.5.

Buffer D: 20 mM Tris-HCl, 10 mM $MgCl_2$, 7 mM 2ME, pH 8.0.

2.1.3 Proteins

All of proteins in my thesis were expressed and purifed by Mr. L. Cooper.

2.2 DNA methods

2.2.1 DNA agarose gel electrophoresis

Agarose gel electrophoresis was used to identify and visualise DNA. 0.9% agarose gels (w/v) were prepared in 1X TAE buffer. The agarose in 1X TAE buffer was melted in a microwave and ethidium bromide solution added just prior to pouring the gel. The final concentration of ethidium bromide was 0.5 μg/ml. Samples were prepared in 10X DNA loading buffer and 10 μl of the sample volume was loaded by using a Gilson pipette. Electrophoresis was carried out at a constant voltage of 100 Volts at room temperature.

DNA was visualised using a UV transilluminator (UVP, TFM-30) and images were acquired using a digital camera (Fujifilm FinePix S602Zoom).

2.2.2 Oligonucleotides

All oligonucleotides were purchased from ATDBio Ltd (Southampton UK).
21B 5'-GTA CGC ACC ACG TGG TTA GGC-3'

21T 5'-GCC TAA CCA CGT GGT GCG TAC-3'

21TH or 21BH has same base sequence as 21T or 21B, but the hexachlorofluorescein (HEX) label is on the 5' end of the top "21TH" strand or on the 5' end of the bottom "21BH" strand.

21B21TH was annealed using 21B and 21TH; 21BH21T was annealed using 21BH and 21T.

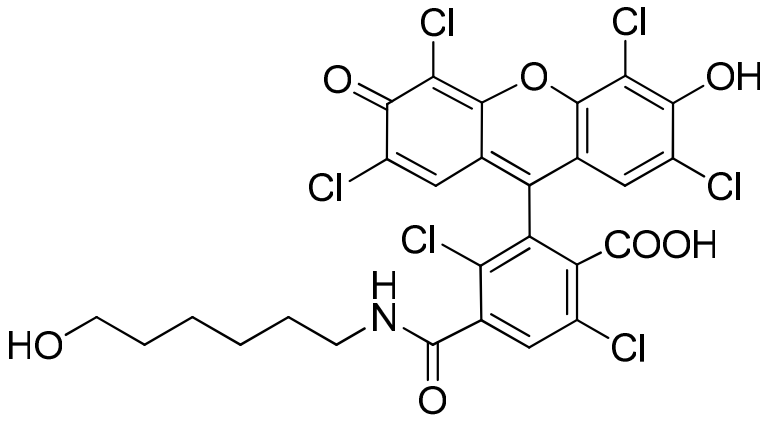


Figure 2-1: The structure of the DNA labeled dye hexachlorofluorescein (HEX) (reprinted from ATDBio Ltd).

The DNA labelled duplex was annealed in H_2O by incubating at 95 °C for 5 min and then cooling gradually overnight in a water bath in the dark. The unlabelled strand was

at a 20% higher concentration than the labelled strand to ensure that the observed fluorescence signal of HEX was only from double stranded DNA (DS-DNA). Aliquots of DS-DNA were stored at -20°C until required.

2.3 Protein methods

2.3.1 Determining the molar extinction coefficient of a protein

The molar extinction coefficient (ε) is the parameter defining how strongly a substance absorbs light at a given wavelength per molar concentration. The units for the molar extinction coefficient can be written as $M^{-1}cm^{-1}$. So the molar extinction coefficient can be thought of as the absorption of a one molar sample in 1 cm distance. In a specific buffer, the molar extinction coefficient of a protein is constant at a specific wavelength. In biochemistry, the extinction coefficient of a protein at 280 nm is usually used to determine the concentration of the protein. Different amino acids have different absorption at 280 nm. Tryptophan (W) has a large absorption. The aromatic rings of tyrosine (Y) and the disulphide bridges of cysteine (C) have a small absorption at this wavelength. Thus, the theoretical extinction coefficient of a protein can be calculated from the number of tryptophan, tyrosine and cysteine amino acids residues [Gill *et al*., 1989 and Pace *et al*., 1995].

$$\varepsilon = (n\mathrm{W} \times 5500) + (n\mathrm{Y} \times 1490) + (n\mathrm{C} \times 125)$$

Where ε is molar extinction coefficient (unit $M^{-1}cm^{-1}$).

2.3.2 ArdA details

Table 2-1 summarises some information concerning the ArdA proteins. Their amino acid sequences are listed at Appendix A.

Table 2-1: The source of the ArdA genes investigated (adapted from McMahon 2009) pI, molecular weight and molar extinction coefficient calculated by (http://www.scripps.edu/~cdputnam/protcalc.html).

Reference	Organism	Gene ID	Amino acid length	Predicted pI	Molecular weight	Molar extinction coefficient ($M^{-1}cm^{-1}$)
Flannagan *et al*., 1994	*Enterococcus faecalis*	Tn 916 (Orf 18)	165	4.02	19125	28020
Cerdeno *et al*., 2005	*Enterococcus faecalis* V583	EF2335	166	4.00	19400	23610
Amann *et al*., 1988	*Staphylococcus aureus* Mu50	SAV0405	166	3.98	19390	28020
Terwilliger *et al*., 1999	*Bacteroides fragilis* NCTC 9343	BF1222	177	4.13	21143	39400

*Molar extinction coefficient estimated by the method of Gill *et al*., 1989 does not count the absorption of disulfides.

*pI estimate assumes all residues have pKa values that are equivalent to the isolated residues. For a folded protein this is not valid. However, this rough value can be useful for planning protein purfications. pKa values for the individual amino acids from Stryer [Stryer, 1988].

2.3.3 SDS Polyacrylamide gel electrophoresis

SDS Polyacrylamide gel electrophoresis (SDS-PAGE) separates denatured polypeptides/proteins based on their molecular size. The protein samples were prepared by adding an equal volume of 2x sample buffer. NuPAGE 4-12 % Bio-Tris gels (Invitrogen) we were used. The running buffer used was NuPAGE MES SDS running buffer (Invitrogen). The gels ran for 45min at 150V. The SDS-PAGE gels were removed from their casts and gently shaken in staining solution for 1hr. Gels were then transferred to a destain solution and gently shaken for 3-5hr, replenishing the solution every hour. Images of the gels were captured using a digital camera (Fujifilm FinePix S602Zoom) and optimised using Fuji Finepix software.

2.3.4 Purification of GFP-MTase

Approximately 15 g of cell pellet was defrosted on ice for 15 mins and resuspended in 50 mls of buffer A (20 mM sodium phosphate, 500 mM NaCl, 20mM imidazole, pH 7.4). A protease inhibitor tablet was added to the buffer (*Complete* EDTA-free protease inhibitor – Roche). The cells were disrupted by sonication on ice using a Soniprep 150 sonicator (Sanyo, Tokyo, Japan) fitted with a 9mm diameter probe for approximately 15 min with intermittent cooling between bursts. The extract was centrifuged at 20,000 g for 1 hour at 4 °C, and the supernatant filtered through a filter unit (0.45 µm; Sartorius AG, Goettingen, Germany). The clarified extract was then applied to a HisTrap FF 5 ml column (GE Healthcare, Uppsala, Sweden) equilibrated in buffer A at 3 ml/min. Once the sample was loaded, the column was washed with 100 ml of buffer A. Elution of the bound protein was done with buffer B (20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.4). Protein elution was monitored by observing the green colour and was completed in 5 ml total volume. Analysis on SDS-PAGE showed there were still impurites present and further purification was necessary. The sample was loaded onto a HiLoad 16/60 Superdex 200 pg gel filtration column (GE Healthcare, Uppsala, Sweden) pre-equilibrated in buffer C (20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM

$MgCl_2$, 7 mM 2ME) and the protein eluted at 12 ml/hour. The fractions were analysed by SDS-PAGE and those containing the protein were pooled. The protein still contained impurities and was further purified using anion exchange chromatography. The sample was dialysed against 2 litres of buffer D (20 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 7 mM 2ME) for 3 hours and then loaded onto a DEAE anion exchange column (30cm×1.6 cm diameter), equilibrated in buffer D, at a flow rate of 50 ml/h. The column was washed with one column volume of buffer D and the bound protein eluted using a 500 ml gradient of 0-0.4 M NaCl in buffer D at a flow rate of 25 ml/h. The green eluted fractions were analysed by SDS-PAGE. The protein had separated out into GFP-M_2S_1 and GFP-M_1S_1 components (as does the WT protein. Dryden *et al*., 1993). The purest fractions for each were pooled separately and concentrated using spin concentrators with a 30 kDa cutoff membrane (VivaScience AG, Hanover, Germany). The purified proteins were stored at -20 °C containing 50% (v/v) glycerol. The final yield was approximately 40 mgs of pure GFP-M_2S_1 and 12 mgs pure GFP-M_1S_1. Extinction coefficients were obtained by inputting sequence data into the Protein Calculator program the extinction coefficient at 280 nm was taken to be 167,560 $M^{-1}cm^{-1}$ for MTase (M_2S_1, protein calculator:
http://www.scripps.edu/~cdputnam/protcalc.html).

2.3.5 Chemical modification procedure

The Thermo Scientific DyLight sulfhydryl-reactive fluorophore 549 (DyLight 549) was used to label Ocr mutants. The sulfhydryl-reactive fluorophores contain maleimide groups that react predominantly with free –SH groups forming a stable thioether bond in water solution (Figure 2-2).

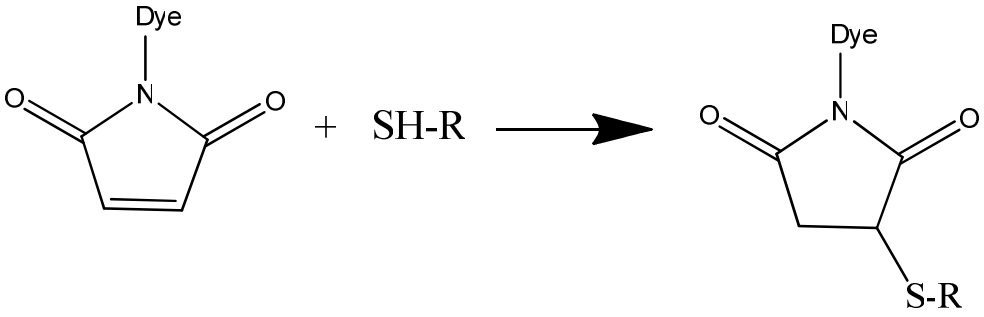


Figure 2-2: The mechanism of sulfhydryl-reactive fluorophore 549 binding to protein with a sulphydryl group. R could be protein, peptide or another ligand.

Chemical modification steps:

1. Equilibrate the frozen vial of DyLight 549 to room temperature before opening to avoid moisture condensation onto the product;

2. Dissolve the DyLight 549 by adding 100 µl of Dimethylformamide (DMF) into 1 mg DyLight and mix well;

3. Prepare 1 ml stock of the Ocr mutants (E20C, S68C and E117C are prepared and purified as in Stephanou *et al.*, 2008. Dr. Stephanou provided the mutants) in PBS buffer (0.1 M sodium phosphate, 0.15 M sodium chloride, 10 mM EDTA, pH 7.2) at 20 µM, respectively;

4. Add 20 µl of the dye to the tube containing the Ocr mutant solution and mix well;

5. Allow the reaction to proceed overnight at room temperature in the dark;

6. Remove nonreacted dye from the proteins by dialysis against 2 L 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME, pH 8.0, 3 times, each time last 2 h;

7. Using liquid N_2, flash freeze the proteins and store at -80 °C.

To calculate the degree of labeling

1. Dilute a small amount of labeled, purified protein in PBS.

2. Using a 1 cm path length cuvette, measure the Uv-vis absorbance spectrum (240-650 nm) of the specific sample.

3. Calculate protein concentration as follows:

$$\text{Protein concentration (M)} = \frac{[A_{280} - (A_{max} \times CF)]}{\varepsilon_{protein}} \times \text{dilution factor}$$

Equation 2-1:

$\varepsilon_{protein}$ is the protein molar extinction coefficient (31860 $M^{-1}cm^{-1}$ for Ocr mutants)

CF is correction factor $\frac{A_{280} \text{ of the fluor}}{A_{max} \text{ of the fluor}}$

A_{max} is the absorption at 562 nm

4. Calculate the degree of labeling:

$$\text{Moles dye per mole protein} = \frac{A_{max} \text{ of the labeled protein} \times \text{ dilution factor}}{\varepsilon_{fluor} \times \text{protein concentration (M)}}$$

Equation 2-2:

Where ε_{fluor} is 150,000 $M^{-1}cm^{-1}$ at 562 nm.

2.4 Analytical methods

2.4.1 Size-Exclusion HPLC

Size-exclusion HPLC, also known as gel filtration, separates molecules based on their size, shape and hydrodynamic volume. The size exclusion HPLC column was prepacked

with gel particles which form a separation matrix through which a buffer solution is passed. The small molecules which diffuse into the gel beads are delayed in their passage through the column compared with the large molecules which cannot diffuse into the gel and move continuously through the column in the flowing eluent. Thus the large molecules pass through the column first followed by the smaller molecules.

Size-exclusion HPLC was used to determine the molecular mass of ArdA and the complex of ArdA and MTase. The experiments were conducted using a -SEC-s 3000 gel filtration column (Phenomenex, Torrance, USA) with 0.5 ml/min flow rate. The column was equilibrated in 20 mM Tris-HCl, 20 mM MES, 200 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 7 mM 2ME, pH 6.5 buffer (Buffer A). The sample volume was 10 µl. Tryptophan fluorescence was used to detect proteins in the eluate. The column eluate was monitored by UV-vis absorption at 280 nm and was excited at 295 nm and continuously monitored at 350 nm using a flow-through fluorimeter. Before using the column, it was calibrated using standard globular proteins (bovine serum albumin, alcohol dehydrogenase, carbonic anhydrase and cytochrome c), which were run individually. The equation derived from plotting the elution volume vs. log molecular mass of standards was used to calculate the molecular mass of ArdA proteins. The temperature used was constant at 20 °C.

2.4.2 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) was carried out using a VP-ITC instrument (Microcal, Northampton, MA). The stocks of ArdA, MTase and GFP-MTase were buffer exchanged into 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME, pH 8.0, by using a PD-10 gel filtration column (GE Healthcare, USA) before ITC measurement. SAM (New England Biolabs, Ipswich, USA) was then added to a final concentration of 100 µM. The concentration of the protein solution was then adjusted appropriately: typically, ArdA at a concentration of 30-40 µM; MTase and GFP-MTase at a concentration of 3.0 µM. All solutions were thoroughly degassed, before ArdA was titrated into a MTase

solution 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 ArdA solution 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.

2.4.3 Fluorescence measurements

A 3 mm pathlength quartz cuvette (3 x 5 mm, Starna) containing ~100 µl of sample solution was placed in the fluorimeter and left for a few minutes to temperature equilibrate before measurements were taken for both steady state and time-resolved fluorescence measurements.

2.4.3.1 Static fluorescence and fluorescence anisotropy

Static fluorescence and fluorescence anisotropy were used to investigate protein-protein and protein-DNA interactions. All measurements were carried out using an Edinburgh Instruments FS 900 CDT T-geometry fluorimeter (Edinburgh Instruments Ltd, Livingston, UK) with a 5 nm bandwidth.

2.4.3.2 Time-resolved fluorescence experiments

Time correlated single photon counting was performed with a home made time-resolved fluorimeter equipped with an Edinburgh Instruments (Edinburgh, UK) TCC900 single photon counting card, 456 nm or 500 nm pulsed LED driven by a PDL 800-B pulsed diode laser driver (PicoQuant Gmbh, Berlin, German) and a PMH-100-3 single photon counting photomultiplier tube (Becker & Hickl Gmbh, Berlin, German). A 405 nm pulsed laser (Edinburgh Instruments) was also sometimes used. Emission wavelengths were selected with a monochromator. Polarisation was applied using quartz Glan-

Thompson polarisers. Excitation pathlengths were 3 mm and emission pathlengths were 3 mm in the cuvette. The emission bandpass was 20 nm.

For time-resolved fluorescence measurement, first of all, we need to measure the instrument response function (IRF) by exciting buffer solution at zero degree (Figure 1-21 vertical excitation) and collecting the emission at the magic angle at the same wavelength as excitation wavelength.

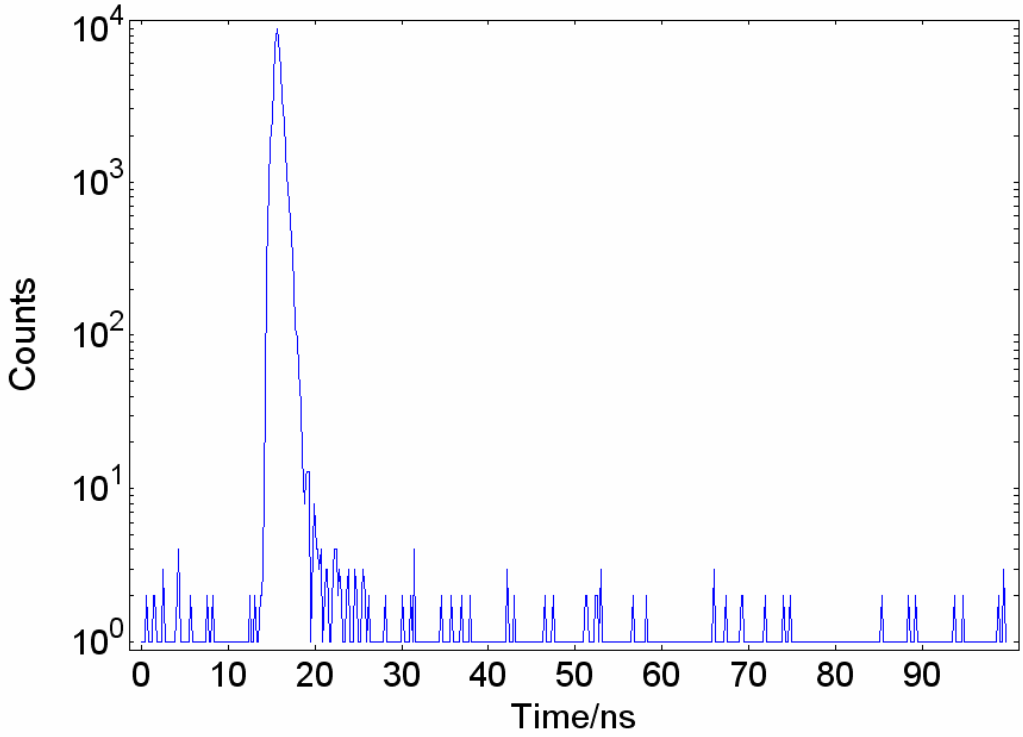


Figure 2-3: The instrument response function of LED. The sample was excited with zero degree polarised light and collected the magic angle emitted light until collected 10^4 counts in the peak.

Figure 2-4 is the typical fluorescence decay data. Non-linear least squares curve fitting was use to analyze each individual set of data with the Edinburgh Instruments F900 software.

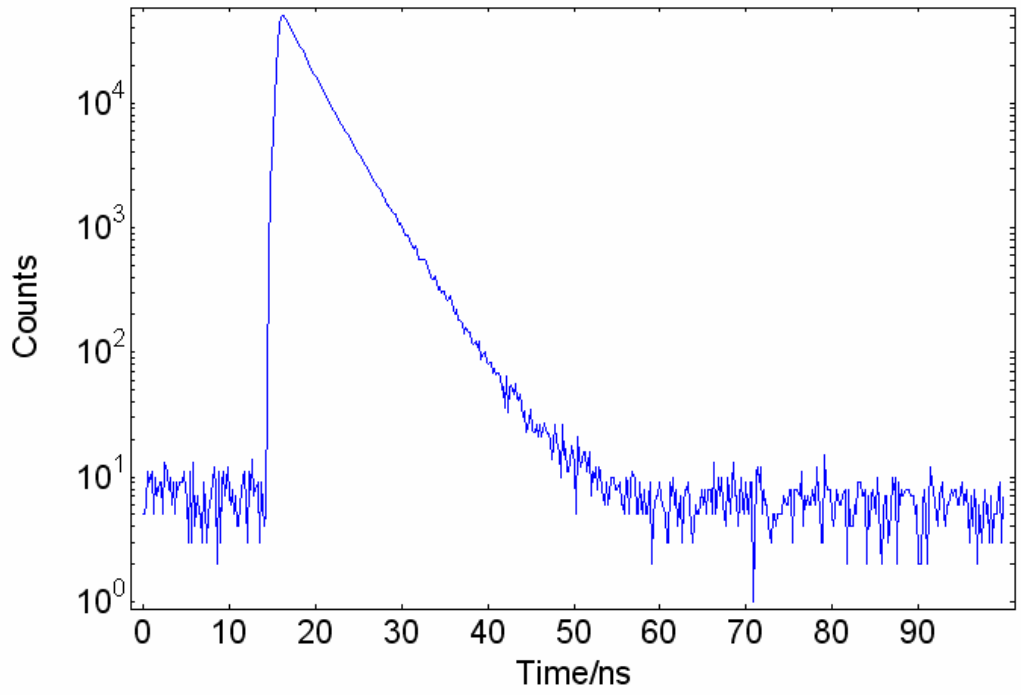


Figure 2-4: Fluorescence decay data for 500 nM 21B21TH, collected with the time-correlated single photon counting set up (blue trace). The buffer was 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME, pH 8.0.

The observed decay, I(t), is a convolution of the IRF and the experimental decay of the GFP, G(t):

Equation 2-3: I(t)=IRF*G(t)

G(t) is obtained by iterative reconvolution and used a previously recorded IRF.

It is assumed that G(t) follows an exponential decay and thus is fitted to Equation 2-4.

Equation 2-4: $G(t) = B + \sum_i A_i \exp\left(\frac{-t}{\tau_i}\right)$

Where B is an integer corresponding to the background level, t is time, τ_i is the lifetime of the i^{th} component and A_i is the fractional amplitude of the i^{th} component.

The F900 software creates a test function G(t)' with A_i and τ_i as variable parameters and B a constant. G(t)' is convoluted with the IRF (Equation 2-3) to generate a calculated fitting function Y(t). Data points for the calculated function, $Y(t_i)$, are compared with the experimentally observed data points, $y(t_i)$, through the *chi squared* parameter, χ^2:

Equation 2-5: $$\chi^2 = \sum_i W_i [y(t_i) - Y(t_i)]^2$$

where W_i is the weighting factor, the square of the deviations of corresponding points from the experimental test function. Equation 2-6 expresses χ^2 with the commonly used weighting factor:

Equation 2-6: $$\chi^2 = \sum_i \left\{ \frac{[I_0(t_i) - Y(t_i)]^2}{I(t_i)} \right\}$$

where $I_0(t)$ is the background-corrected intensity decay Equation 2-7 and I(t) is the uncorrected function:

Equation 2-7: $$I_0(t_i) = I(t_i) - \mathrm{B}$$

During the fitting procedure, χ^2 is minimised and is further evaluated through the *reduced chi squared* parameter χ_v^2:

Equation 2-8: $$\chi_v^2 = \frac{\chi^2}{n_2 - n_1 + 1 + p}$$

Where n_1 and n_2 are the first and last data point of the data fitting range, respectively, and p is the number of variable parameters. For a good fit, χ_v^2 is close to unity.

The suitability of the calculated function is further assessed by visual inspection and by the inspection of a plot of the weighted residual $r(t_i)$ for each point.

Equation 2-9: $r(t_i) = \sqrt{W_i[I_0(t_i) - Y(t_i)]} = \frac{I_0(t_i) - Y(t_i)}{\sqrt{I(t_i)}}$

Plotted residuals from successful fits should be randomly distributed about zero.

Decay curves are fitted using the beginning of the decay curve, (a few channels from the peak) and the end of the decay curve by Equation 2-4 using as a small a number of exponential terms as possible.

For example, Figure 2-5 presents the decay from Figure 2-4 fitted with one exponential term in Equation 2-4. Figure 2-6 shows the same decay fitted with two exponentials. The result of adding a second exponential term is a significantly lower χ_v^2, a more uniform variation of the residuals and, visually, a better fit of the calculated function to the experimental data. Thus the decay in Figure 2-4 is best described by a two-exponential function.

The contribution of the n^{th} emitting species to the steady state intensity (for a multiexponential decay):

Equation 2-10: $\frac{A_n \tau_n}{\sum_i A_i \tau_i}$

and the average lifetime of the sample is:

Equation 2-11: $\frac{\sum_i A_i \tau_i}{\sum_i A_i}$

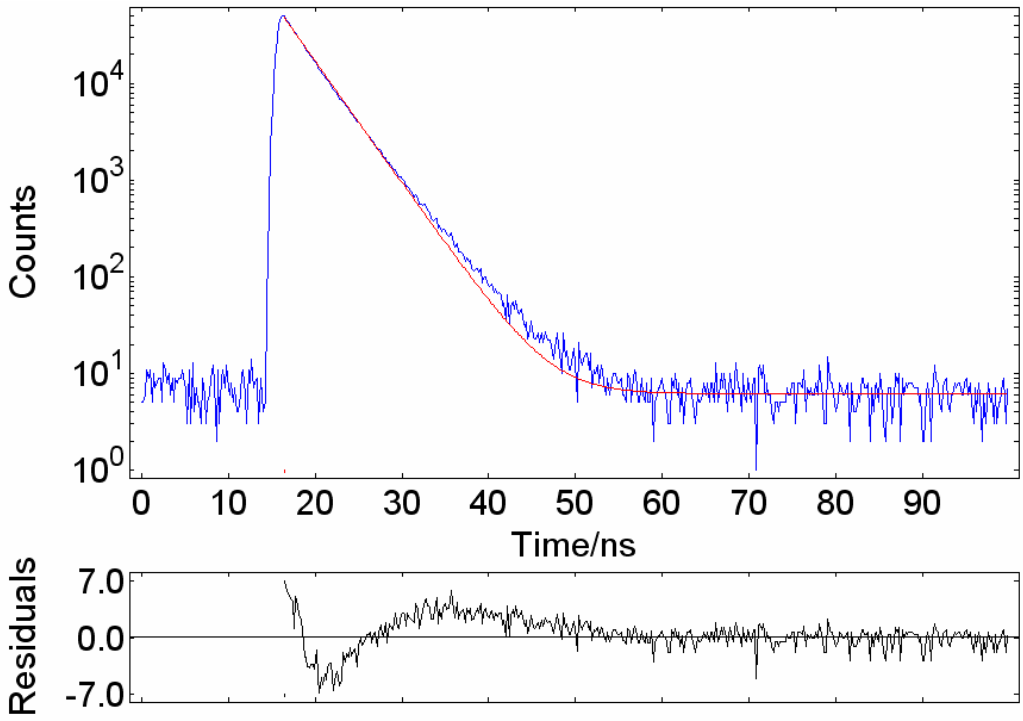


Figure 2-5: Fluorescence decay data for 500 nM 21B21TH (blue), the calculated function generated by F900 (red) and a plot of the residuals between the decay data and the calculated function (black). The decay was fit with one exponential, between the 16.2 and 100.0 ns. The resulting lifetime calculated for the red function is 3.46±0.08 ns and the reduced chi squared = 4.538. The buffer was 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME, pH 8.0.

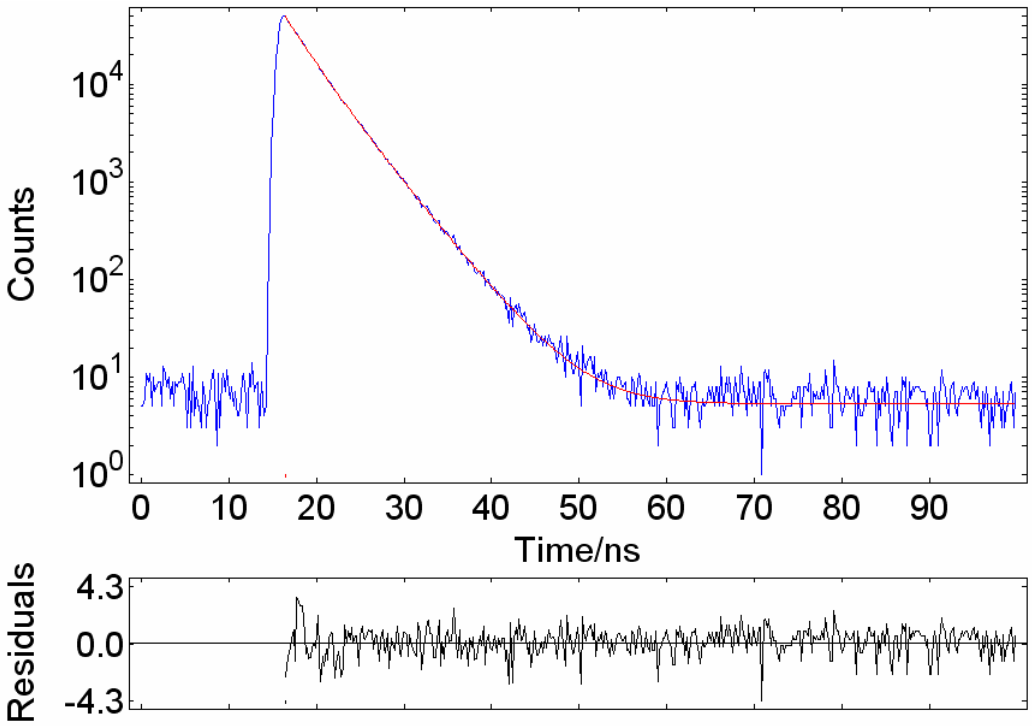


Figure 2-6: Fluorescence decay data for 500 nM 21B21TH (blue), the calculated function generated by F900 (red) and a plot of the residuals between the decay data and the calculated function (black). The decay was fitted with two exponentials, by iterative reconvolution, between the 16.2 and 100.0 ns. The resulting lifetimes calculated for the red function are 2.62±0.06 and 4.15±0.06 ns, and the reduced chi squared = 1.207. The buffer was 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME, pH 8.0.

In fluorescence anisotropy decays measurements, the sample is excited with ninety degree polarised light (Figure 1-21 horizontal excitation), the emission light is collected at ninety and zero degrees for 5 min, to give I_{HH} and I_{HV}, respectively (Figure 2-7). Using these results the F900 software package can calculate the G factor using the following Equation 1-3. From Figure 2-7, we can see, between 16.2 and 60.0 ns was the real fluorescence decay. Other part was just noise.

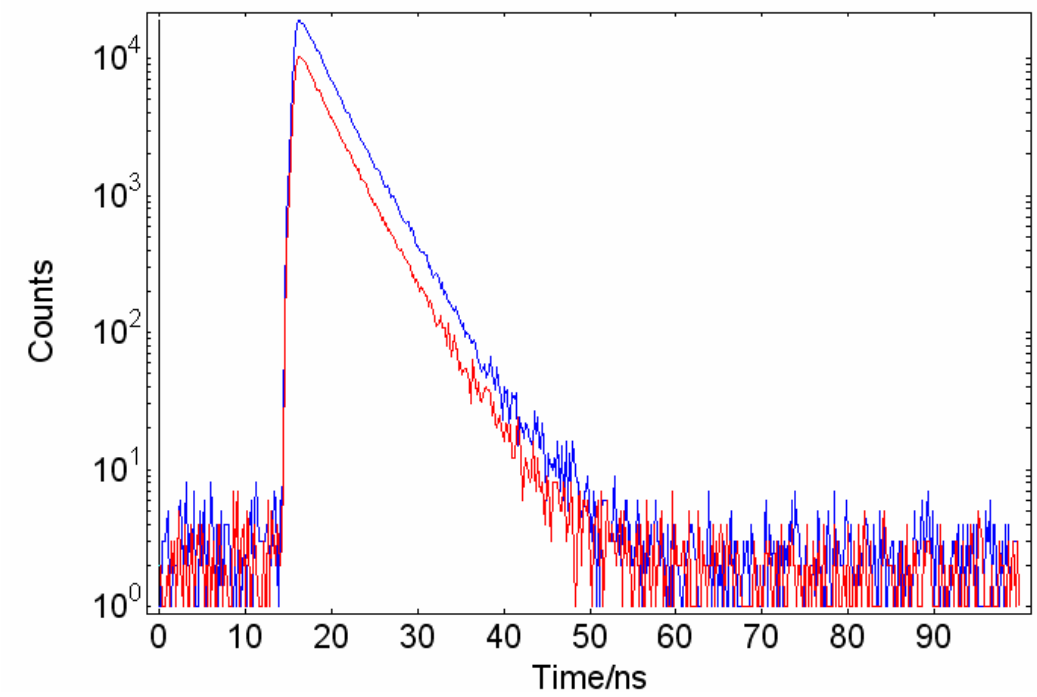


Figure 2-7: Fluorescence decay data for 500 nM 21B21TH, blue curve is I_{HH} and red curve is I_{HV}. The experiments last 5 min, respectively. The G factor is calculated by using these data. Excitation was 500 nm and emission at 565 nm. The buffer was 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME, pH 8.0.

After collection of I_{HH} and I_{HV}, the excitation light is changed to zero degree (Figure 1-21 vertical excitation), and the emission collected at ninety and zero degrees for 30 min to obtain I_{VH} and I_{VV}, respectively.

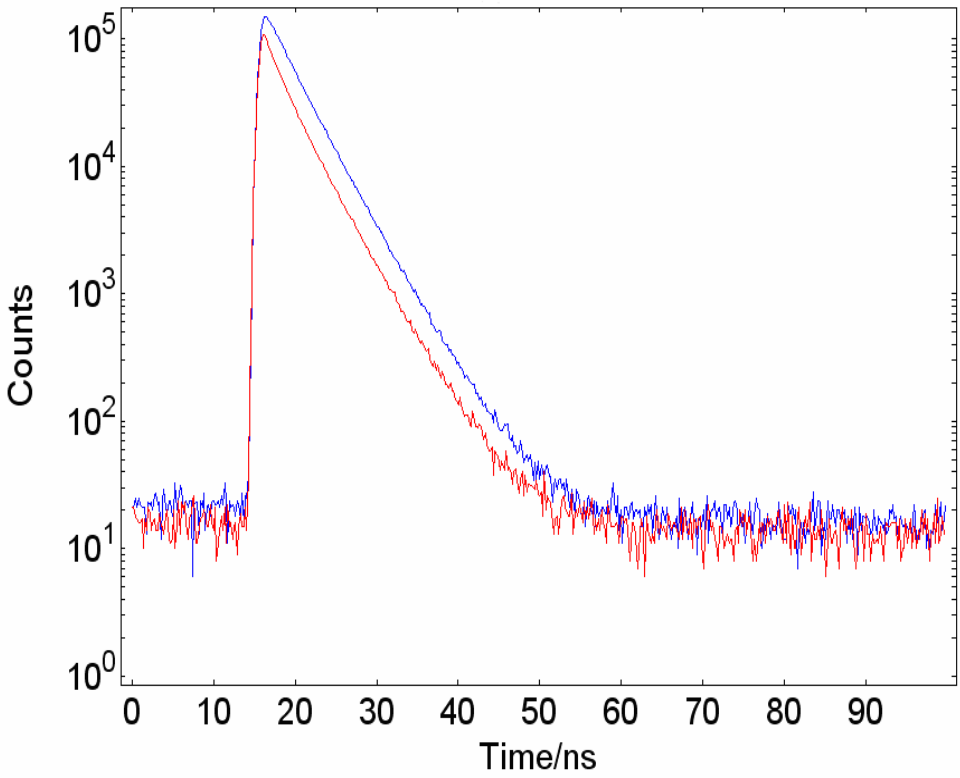


Figure 2-8: Fluorescence decay data for 21B21TH, blue curve is I_{VH} and red curve is I_{VV}. The experiments last 30 min, respectively. Excitation was 500 nm and emission at 565 nm. The buffer was 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME, pH 8.0.

Then by using the G factor, I_{VH} and I_{VV}, the fluorescence anisotropy decay of 21B21TH (Figure 2-8) can be calculated by F900 by using the **Equation 2-12:**

$$r(t) = \frac{I_{VV}(t) - GI_{VH}(t)}{I_{VV}(t) + 2GI_{VH}(t)}$$

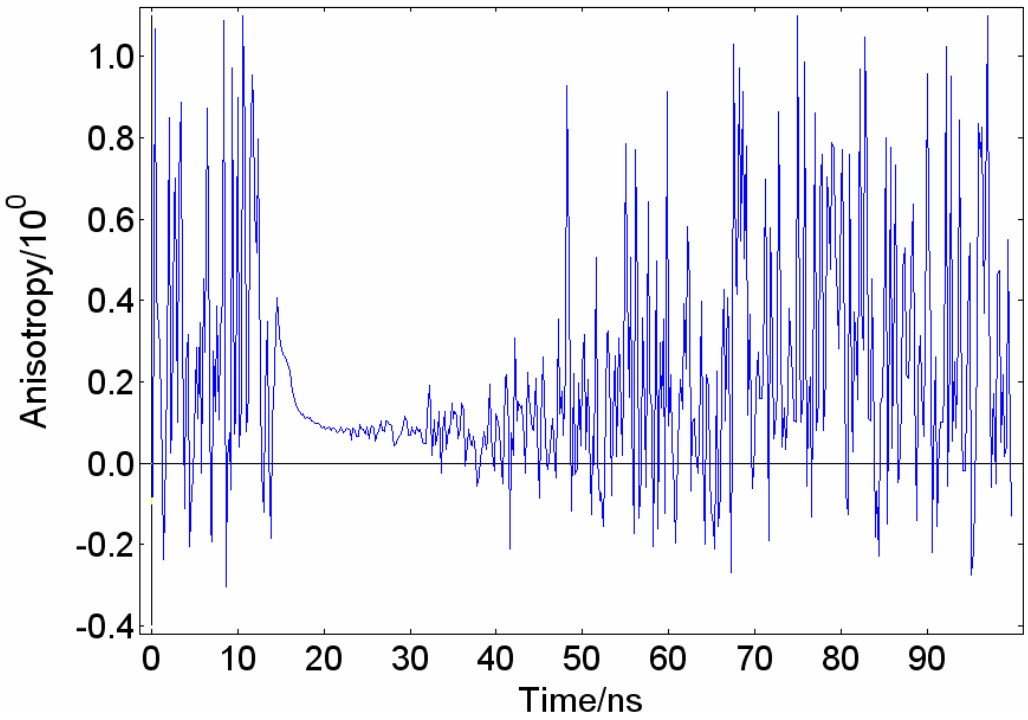


Figure 2-9: Fluorescence anisotropy decay of 500 nM 21B21TH. Calculated by using I_{HH}, I_{HV}, I_{VH} and I_{VV} in the previous Figures 2-5 and 2-6. The buffer was 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME, pH 8.0.

In Figure 2-7, before 16.2 ns and after 60.0 ns, the data are useless and can be ignored. Using I_{VH} and I_{VV} measurements (Figure 2-8), the data beyond 16.2 ns and 60.0 ns were not fluorescence decays, but simply background. So we fitted the results between 16.2 ns and 60.0 ns to get the anisotropy decay (1.73 ns and the reduced chi squared = 1.061) using the same fitting method which was used to fit the normal fluorescence lifetime.

Chapter 3 GFP-MTase as a new probe of Type I DNA restriction and modification enzymes

3.1 Introduction

Since their introduction into genetic engineering, the green fluorescent protein (GFP) and its many spectral variants have proved to be extraordinarily useful probes of protein structure and function both *in vitro* and *in vivo* [Tsien, 2009]. In particular, Förster resonance energy transfer to measure distances between two fluorophores, a donor and an acceptor, has been the subject of many uses of GFP despite its complex photophysics and its relatively large size compared to more traditional small molecule fluorophores such as fluorescein [Piston *et al*., 2007].

Sequence-specific DNA-binding enzymes such as methyltransferases and endonucleases comprising bacterial restriction-modification (R/M) systems would seem to present excellent targets for analysis via fusion to GFP given that many of them introduce complex rearrangements of DNA structure including for example DNA looping to bring distant sites on a single DNA molecule into close proximity. However, as yet few investigations of R/M systems have utilised these versatile fluorescent probes [Senejani *et al*., 2007].

Wild-type Ocr contains no Cys residues. Therefore, mutating Ocr to introduce a single Cys residue into each monomer at positions E20, S68 or E117 (Figure 3-1) allows the protein to be modified with Dylight 549 Maleimide to generate a FRET acceptor for GFP-MTase.

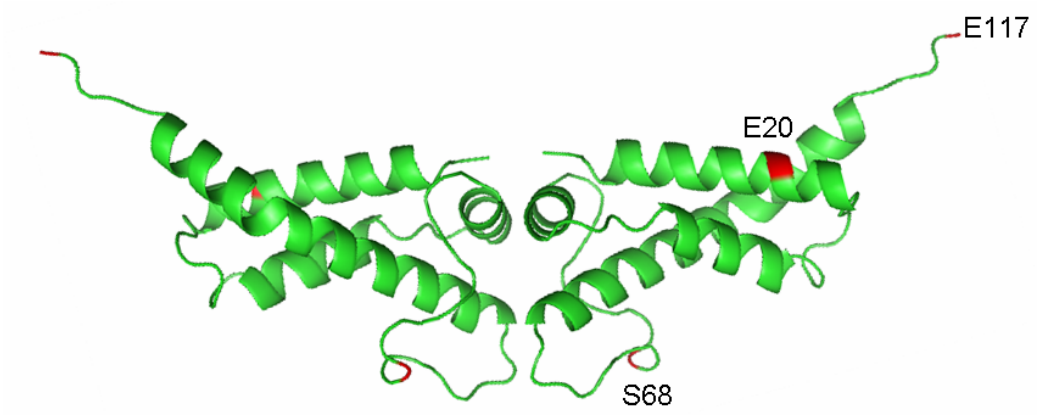


Figure 3-1: The structure of Ocr dimer. The red amino acid is replaced by Cys.

Because no crystal structure of MTase is currently available, we have to find an indirect experimental method to compare and improve the structural model [Kennaway *et al*., 2009]. In this chapter we demonstrate the preparation of an active MTase fused to GFP and measure, via FRET, the distance from the GFP to a HEX label on a duplex bound to the MTase and to a fluorescently labelled Ocr protein bound to the MTase. These distances are then compared to predictions from the structural model [Kennaway *et al*., 2009].

3.2 Results

3.2.1 UV-vis and fluorescence spectra results

The purified GFP-MTase showed the absorption and fluorescence emission properties, Figure 3-2, expected with aromatic absorption and fluorescence in the UV and absorption due to the GFP chromophore in the visible region. The absorption of GFP-MTase in 280 nm is the absorption of aromatic amino acids, and the absorption of GFP-MTase around 400 and 500 nm is the absorption of the chromophore of GFP. GFP-MTase was excitated at 395 nm to show GFP emission around 510 nm.

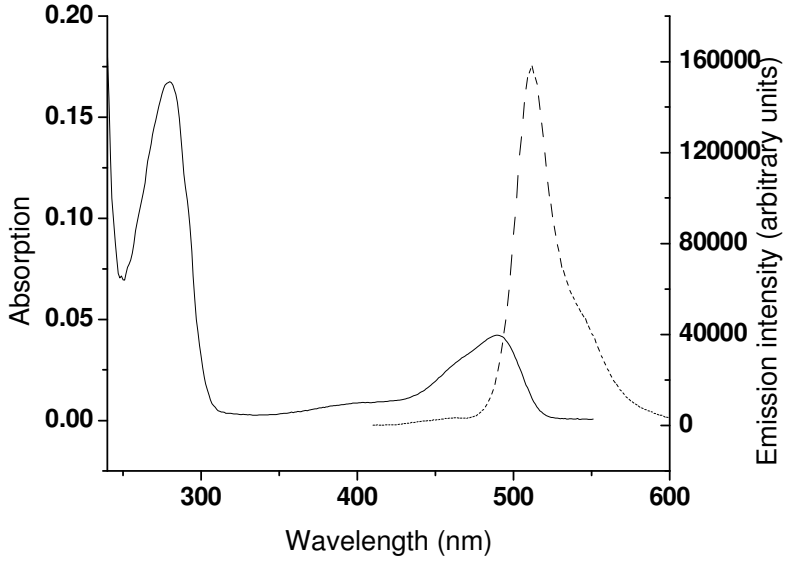


Figure 3-2: Spectrophotometric analysis of MTase and the fusion with GFP. Uv-vis (solid) and fluorescence emission spectrum (dash) of GFP-MTase fusion protein. Protein concentration was 1.0 µM in 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME (pH 8.0). Excitation wavelength was 395 nm.

Figure 3-3 is the absorption and fluorescence of 21B21TH with the absorption in the visible region due to the Hex chromophore. As 21BH21T has same chromophore and environment as 21B21TH, the absorption and fluorescence of 21BH21T is same as 21B21TH.

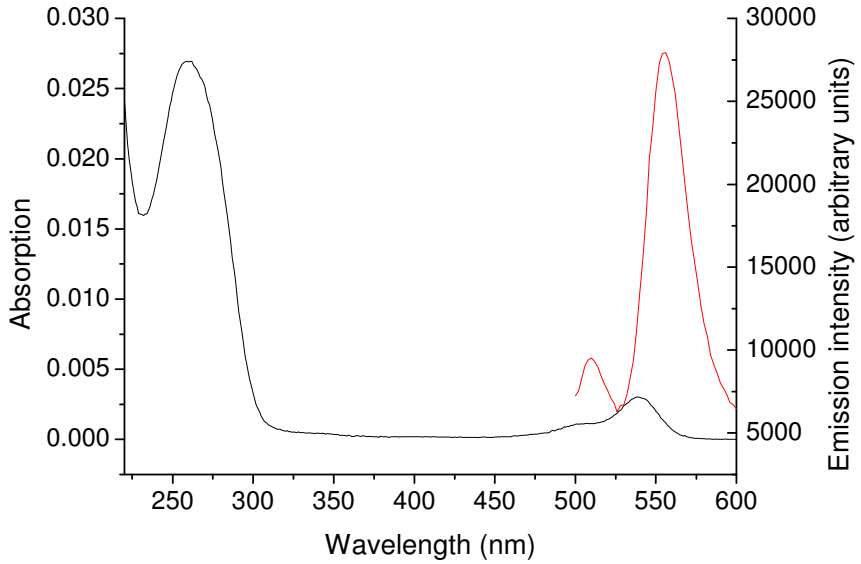


Figure 3-3: Spectrophotometric analysis of 21B21TH. Uv-vis (black) and fluorescence emission spectrum (red) of 21B21TH. 21B21TH DNA concentration was 5 µM in 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME (pH 8.0).

The Uv-vis and fluorescence spectrum of the labelled Ocr mutants are shown below, Figure 3-4.

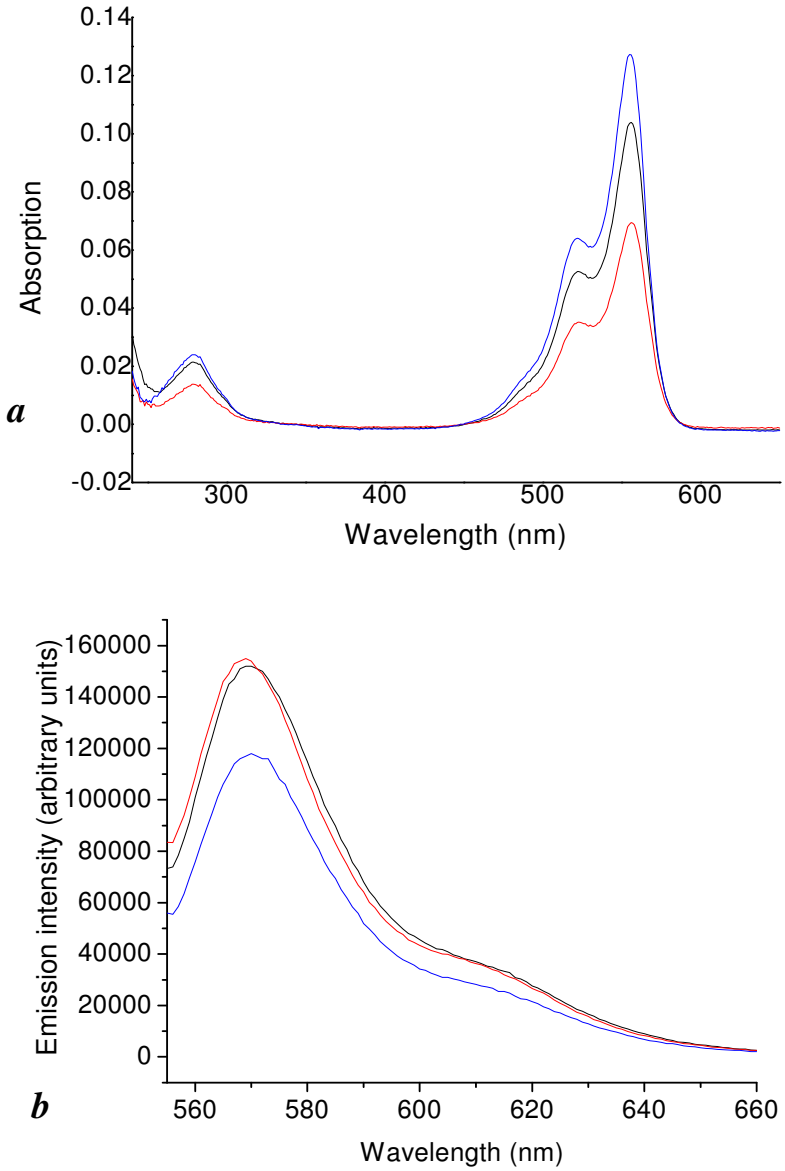


Figure 3-4: Spectrophotometric analysis of Ocr mutant absorption and emission. Panel *a* Spectrophotometric analysis of Ocr mutant absorption. In each case the protein concentration was 5 µM. Panel *b* Spectrophotometric analysis of Ocr mutants. Black curve is E20C, red one is S68C and blue one is E117C. The protein concentration was 0.5 µM. The buffer was 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME (pH 8.0).

From above excitation and emission spectrum, we can clearly see that the emission of GFP-MTase overlaps with HEX-DNA and the labelled Ocr mutant very well. Therefore, once the molecules are close enough, FRET will occur between them. GFP-MTase will be the donor and the HEX-DNA or labelled Ocr mutant will be the acceptor. The results from Figure 3-8 and Table 3-2 indicate that the rotation of GFP is unaffected by linkage to the GFP-MTase. Thus, 2/3 can be used as the value of orientation factor. The overlap of the emission of the GFP with the HEX and Dylight549 labels allowed R_o distances of 6.14±0.37 nm and 6.53±0.46 nm, respectively to be calculated.

3.2.2 Dissociation constant from continuous variation titration

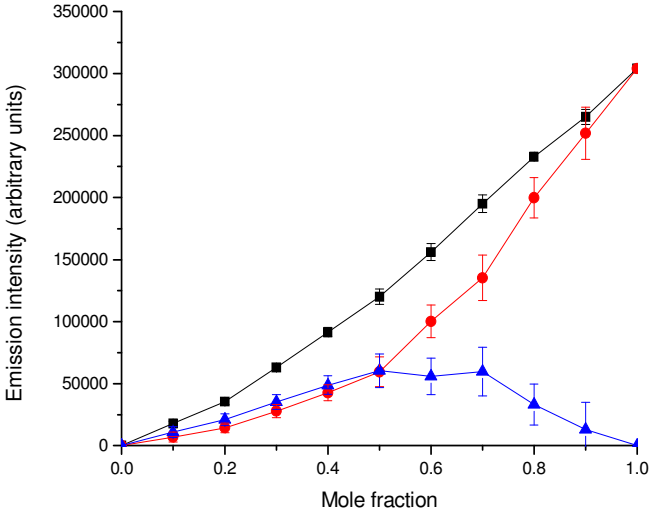
The binding affinity of GFP-MTase and DNA was investigated by measuring the fluorescence intensity of mixtures in which the combined concentration of the two components was constant (200 nM). Initially, solutions of GFP-MTase (0, 20, 40, 60, 80, 100, 120, 140, 160, 180 and 200 nM) were prepared and their fluorescence measured (black data in Figure 3.5). Then, solutions containing different proportions of GFP-MTase and 21B21TH, where the overall total concentration of both components was always 200 nM, were prepared. Energy transfer between GFP-MTase and 21B21TH causes the fluorescence intensity of the complex to be lower than that of GFP-MTase alone (red data in Figure 3-5). The difference in intensity for any given concentration of protein was then plotted (blue data in Figure 3-5). The data shown in Figure 3-5 was normalised and plotted again in Figure 3-6. The difference in fluorescence intensity (blue line in Figure 3-6) was fitted using **Equation 3-1:**

$$(C(1-\chi)-[AB_n])(C\chi-n[AB_n])^n=Kd[AB_n]$$

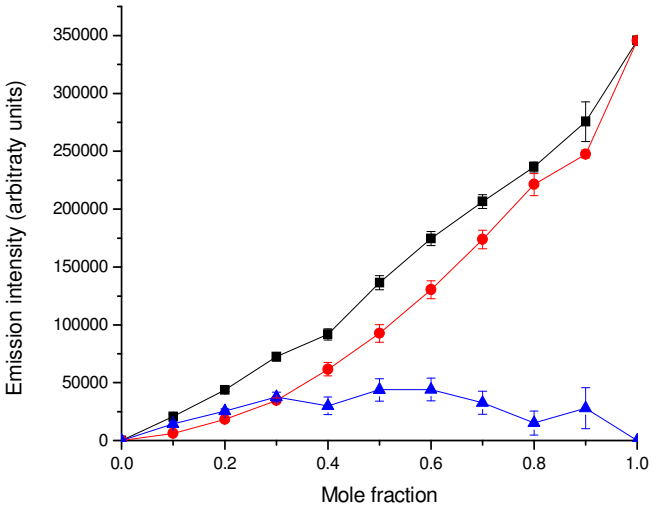
Where C is the total concentration of protein and DNA, A is the concentration of 21B21TH, B is the concentration of GFP-MTase and χ is the mole fraction of GFP-MTase in the sample. We assumed the stoichiometry (n) is 1 to fit the data.

Equation 3-1 is used to determine the K_d of interaction between GFP-MTase and 21B21TH (Figure 3.7). The experiment was conducted at different concentrations of NaCl (0, 25, 50 or 100 mM). The binding affinity was determined at various salt levels (refer to Table 3-1). In 100 mM NaCl, the fluorescence difference was extremely small, indicating that the binding at high salt concentration is very weak. These results are similar to previously reported data (Su *et al.,* 2005).

a



b



c

d

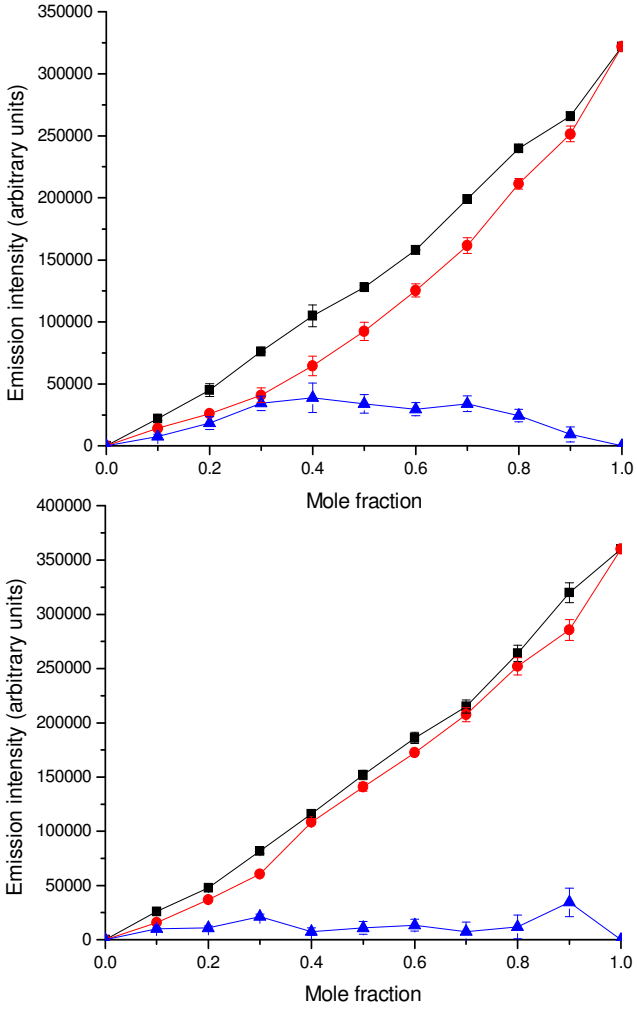
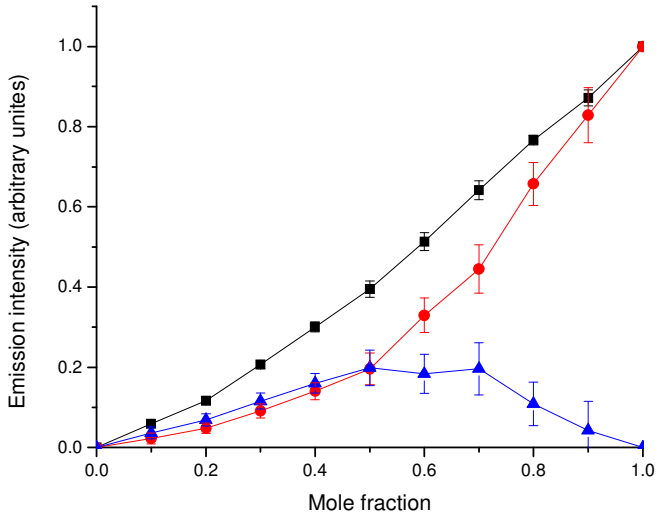


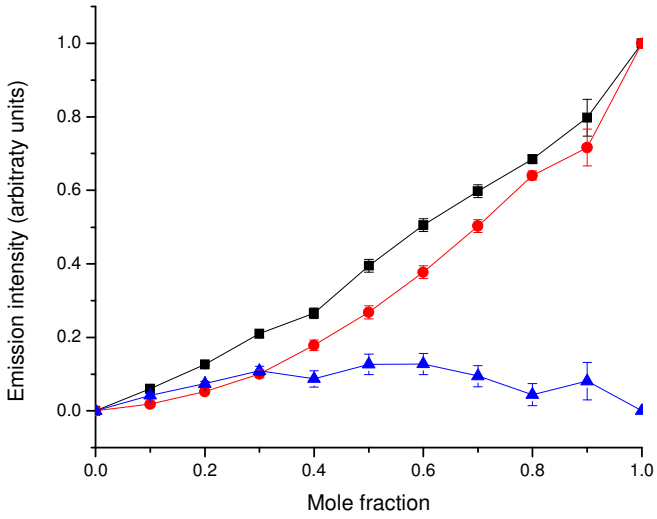
Figure 3-5: Experiment to assess binding of GFP-MTase with a fluorescently labelled 21 bp DNA duplex containing the recognition sequence for MTase (21B21TH). A range of solutions containing GFP-MTase from 0 to 200 nM in 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME (pH 8.0) were prepared (black). In addition, solutions containing different proportions of GFP-MTase and 21B21TH, where the overall total concentration of both components was always 200 nM, were prepared (red). The emission spectrum of each solution was then analysed (excitation 395 nm). The intensity of the emission peak (508 nm) was recorded and plotted against the mole fraction of GFP-MTase. The difference between the intensity for GFP-MTase and GFP-MTase plus 21B21TH at each point on the graph is also plotted (blue). Experiments were conducted in the

absence of salt (panel *a*) or in the presence of 25 mM NaCl (panel *b*), 50 mM NaCl (panel *c*) or 100 mM NaCl (panel *d*).

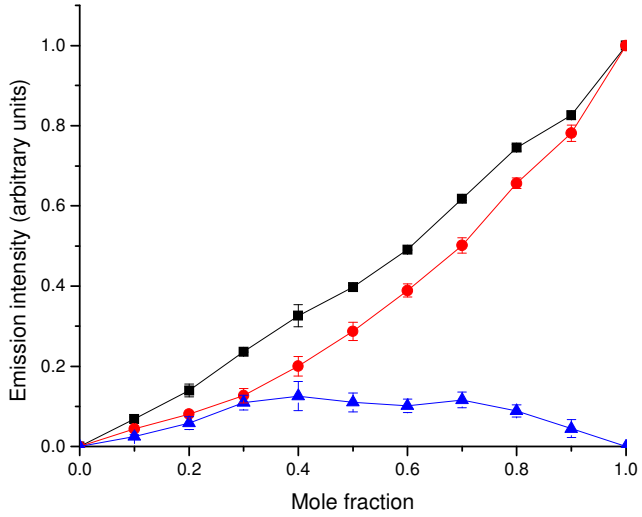
a



b



c



d

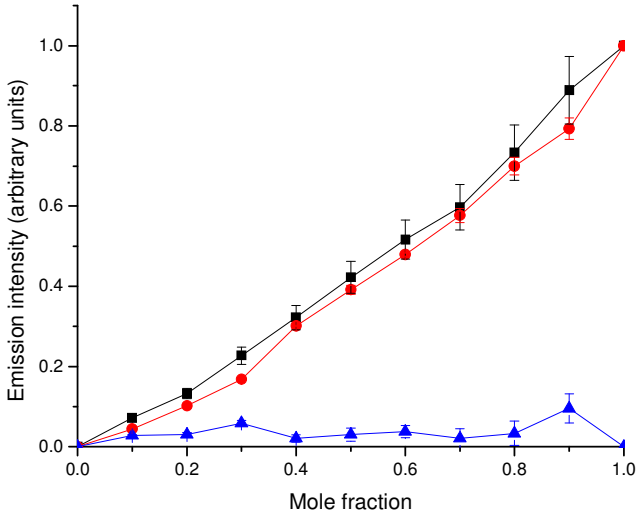
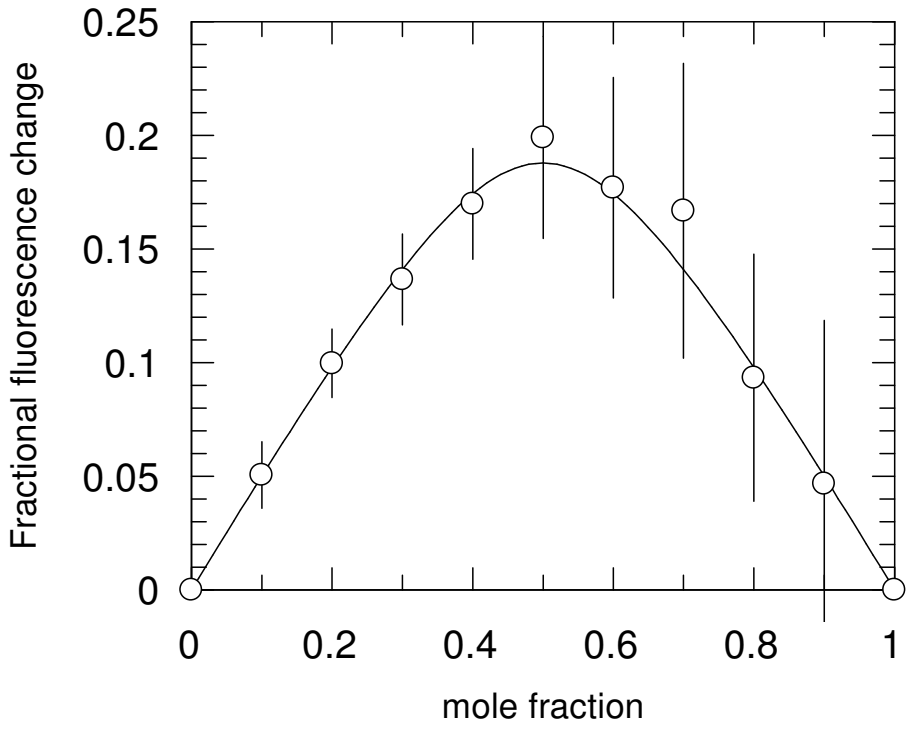
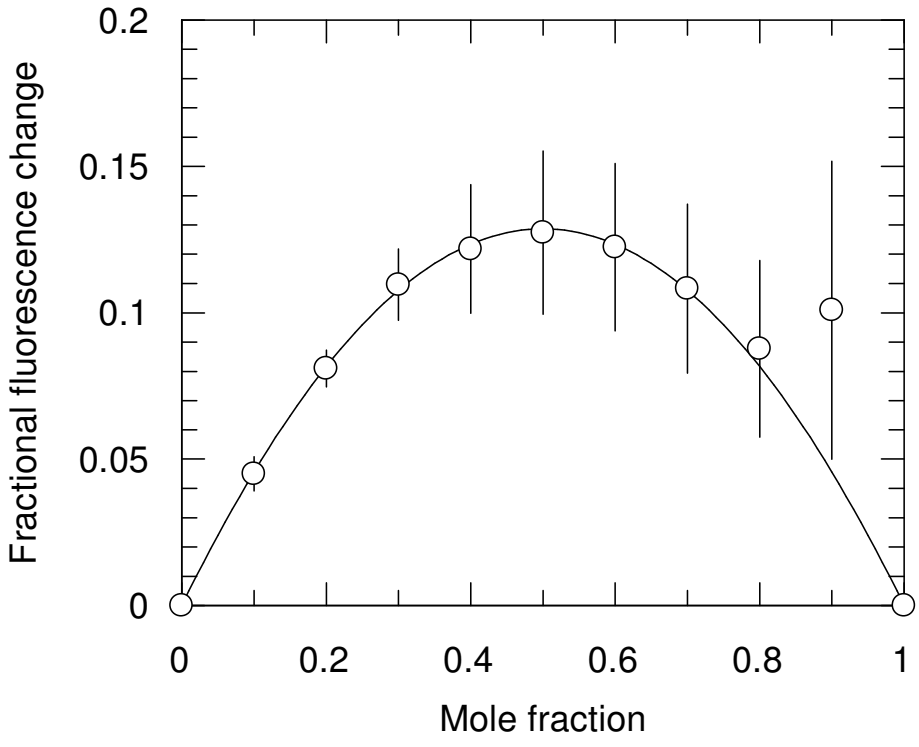


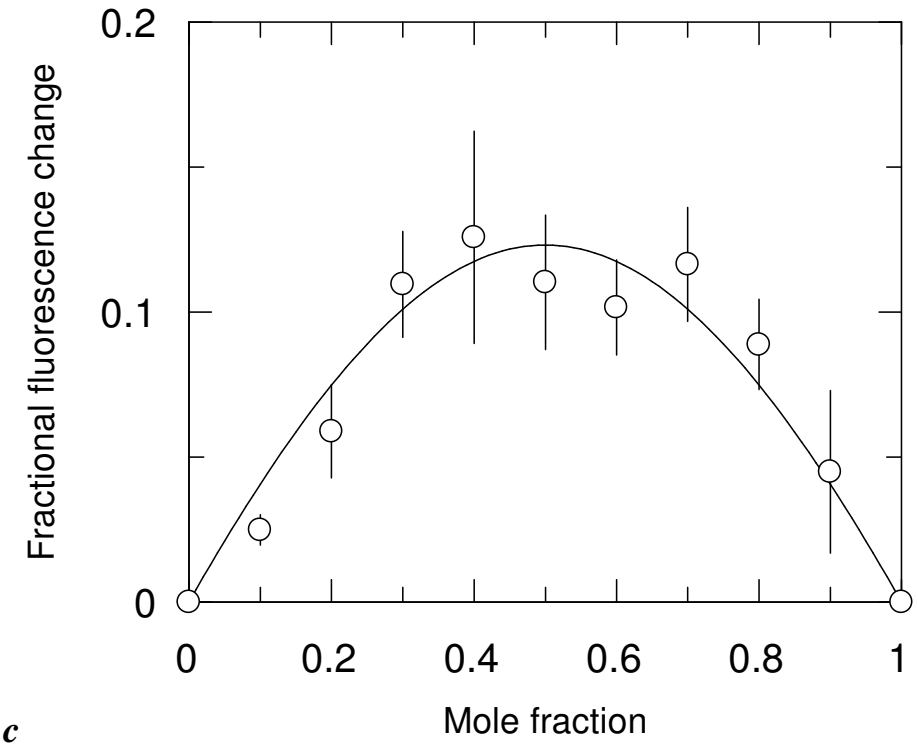
Figure 3-6: Results after normalisation of Figure 3-5. *a*, no salt; *b*, 25 mM NaCl; *c*, 50 mM NaCl; *d*, 100 mM NaCl.


Fractional fluorescence change
0.25
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
mole fraction

a


Fractional fluorescence change
0.2
0.15
0.1
0.05
0
0
0.2
0.4
0.6
0.8
1
Mole fraction

b



c

Figure 3-7: The fitting of the data from Figure 3-6 at a range of salt conditions. The difference in the fluorescence intensity in the presence and absence of 21B21TH (i.e. blue line in Figure 3-6) was fitted to Equation 3-1. *a*, no salt; *b*, 25 mM NaCl; *c*, 50 mM NaCl. The data could not be fitted for the experiment carried out in the presence of 100 mM NaCl (Figure 3-6*d*).

Table 3-1: Dissociation constants (nM) for binding of GFP-MTase to 21B21TH determined in different concentrations of NaCl.

Sodium chloride concentration (mM)			
0	25	50	100
13.9±6.0	38.3±42.2	21.3±20.6	cannot be fitted

The titration results (Figure 3-6 and 3-7) show that the fluorescence change of GFP is much smaller in the presence of high concentrations of salt than in no salt buffer. Indeed, almost no fluorescence change was observed in 100 mM NaCl. These results

indicate that salt decreases the binding affinity between GFP-MTase and labelled DNA. This phenomenon was observed previously [Su *et al*., 2005, Powell *et al*., 1998]. Furthermore, the data (shown in Table 3-1 and Figure 3-7) is entirely consistent with the 1:1 binding model.

3.2.3 Anisotropy decay of the fluorescent labels

To use energy transfer quantitatively, one ideally should determine whether the donor and acceptor chromophores are free to rotate or are sterically hindered on the nanosecond timescale as this indicates that the κ^2 orientation parameter can be reasonably set at 2/3 as assumed in Equation 1-7. The time-resolved data, Figure 3-8 and Table 3-2, indicated that GFP alone gave a rotational decay time of 14.9±0.2 ns as expected for a globular protein of 27 k Da [Volkmer *et al*., 2000]. The anisotropy decay time of the emission from the GFP-MTase fusion was 14.1±0.1 ns. The GFP was rotating on the nanosecond timescale despite its attachment to the MTase but the degree of rotational freedom on GFP when fused to the MTase was slightly less than that of the free GFP as indicated by the higher value of the anisotropy at infinite time. The rotational correlation time of the HEX label on the DNA duplex was unaffected by GFP-MTase binding as was the degree of rotation of the Dylight549 label when attached to the E20C and E117C mutant Ocr proteins, Table 3-2 and Figure 3-9. The label attached to the S68C mutant protein showed an unusual anisotropy decay shape in the absence of GFP-MTase and a long anisotropy decay time in the presence of the GFP-MTase, Figure 3-9. These data indicate that the label attached to the S68C position is not free to rotate and hence that the κ^2 orientation parameter is not 2/3 in the FRET experiments.

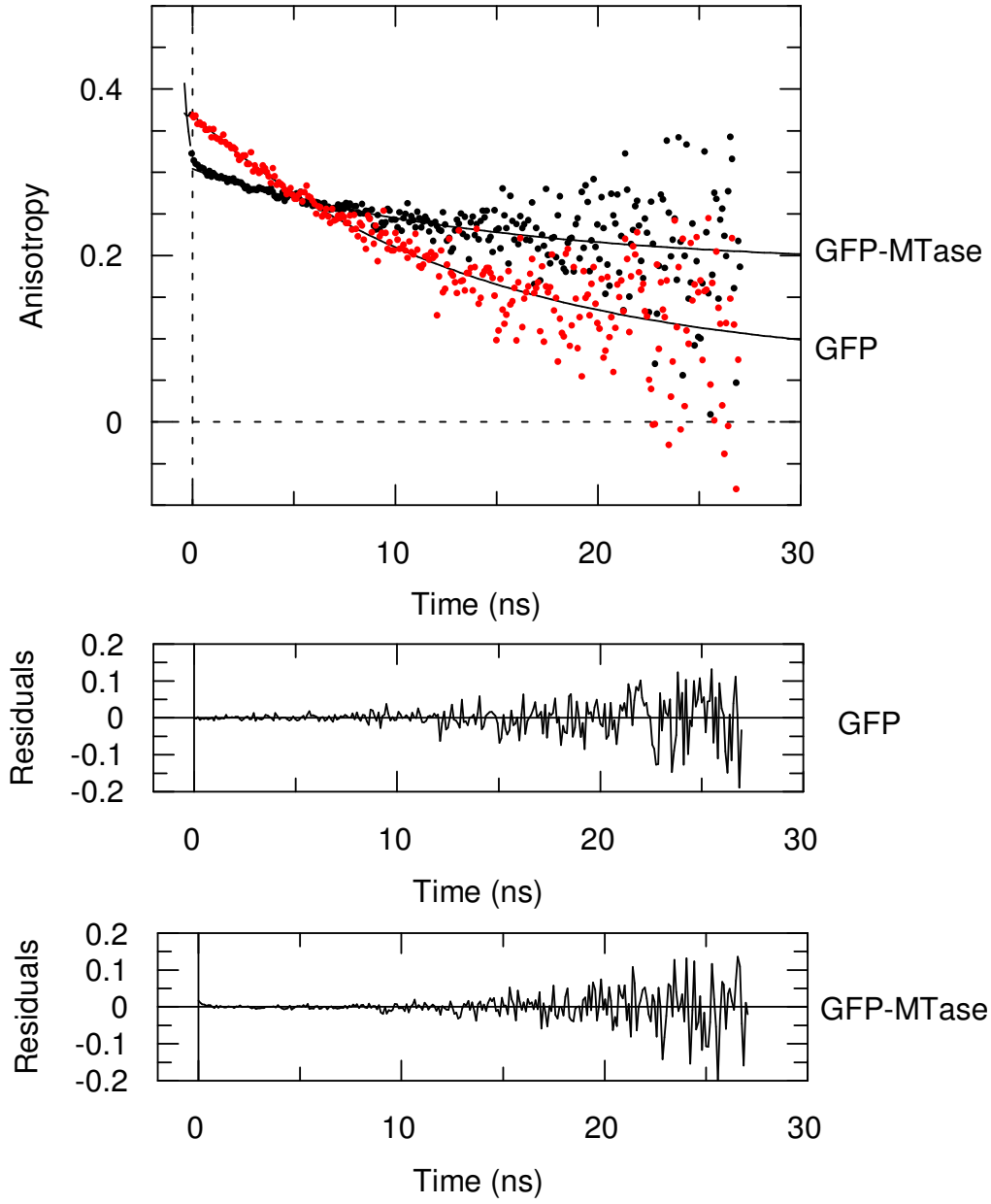
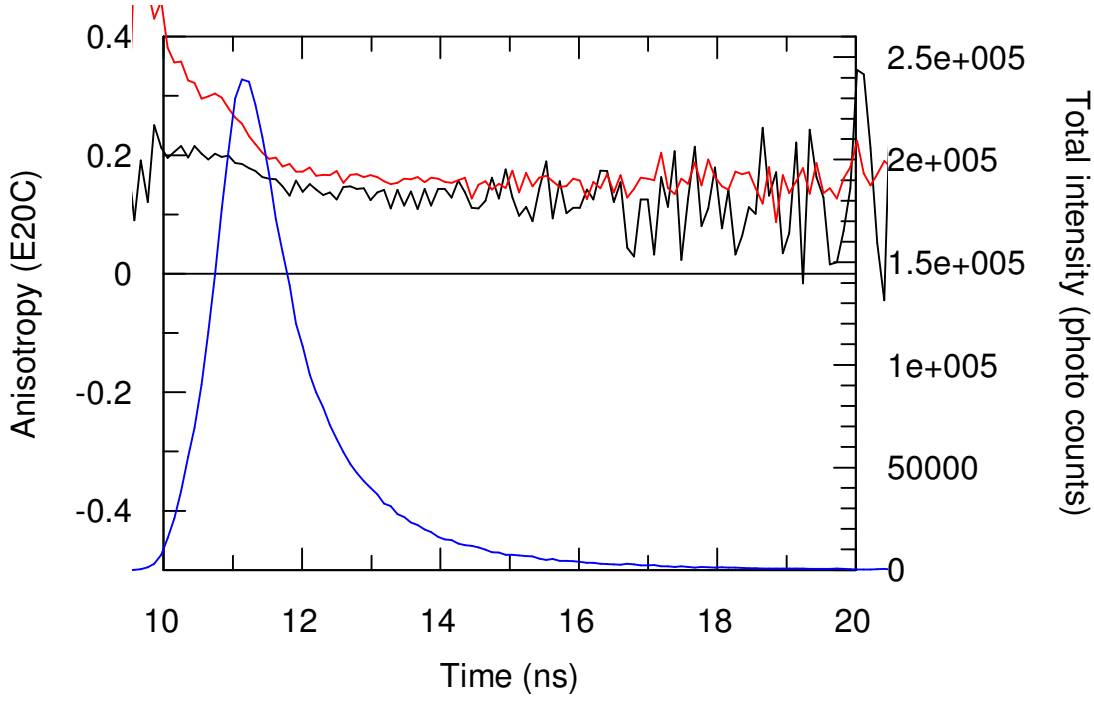
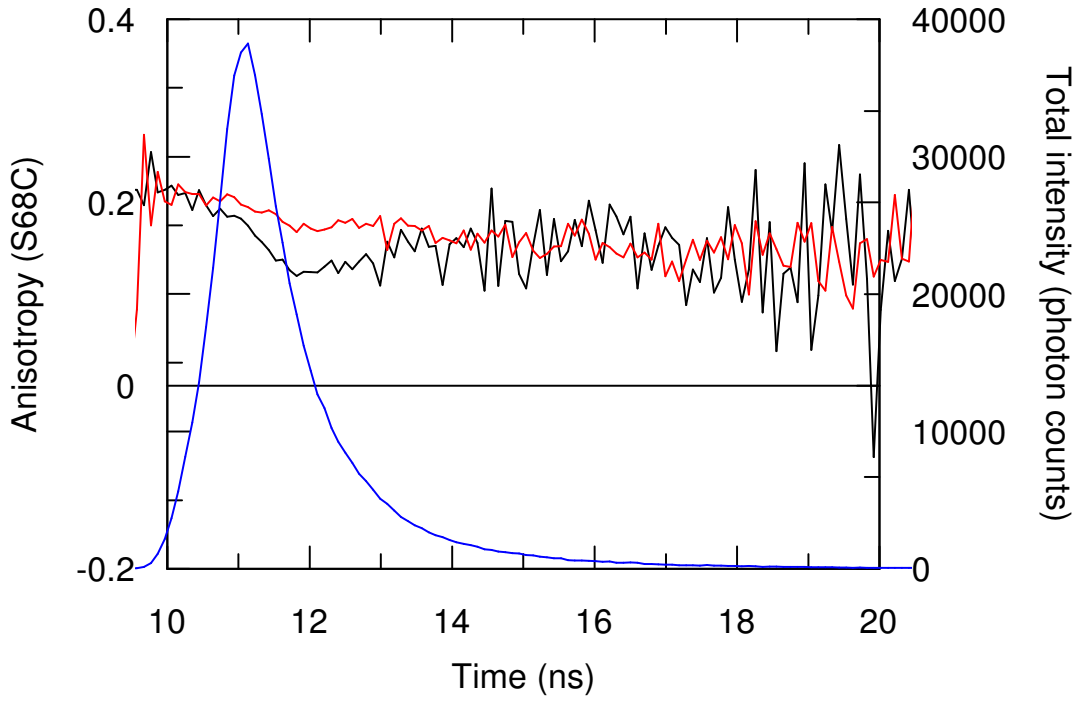


Figure 3-8: Analysis of fluorescence anisotropy decay of GFP (red) and GFP-MTase (black) using one lifetime component. Experiments were conducted in 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME (pH 8.0). Excitation was carried out at 405 nm and decay in anisotropy was monitored at 510 nm. The relaxation time of GFP was determined to be 14.9±0.2 ns (χ^2 =1.146). The initial anisotropy was 0.36 and the final anisotropy was 0.13. The relaxation time of GFP-MTase was 14.1±0.1 ns (χ^2=1.219). The initial anisotropy was 0.31 and the final anisotropy was 0.21.


0.4
0.2
0
-0.2
-0.4
Anisotropy (E20C)
2.5e+005
2e+005
1.5e+005
1e+005
50000
0
Total intensity (photo counts)
10
12
14
16
18
20
Time (ns)


0.4
0.2
0
-0.2
Anisotropy (S68C)
40000
30000
20000
10000
0
Total intensity (photon counts)
10
12
14
16
18
20
Time (ns)

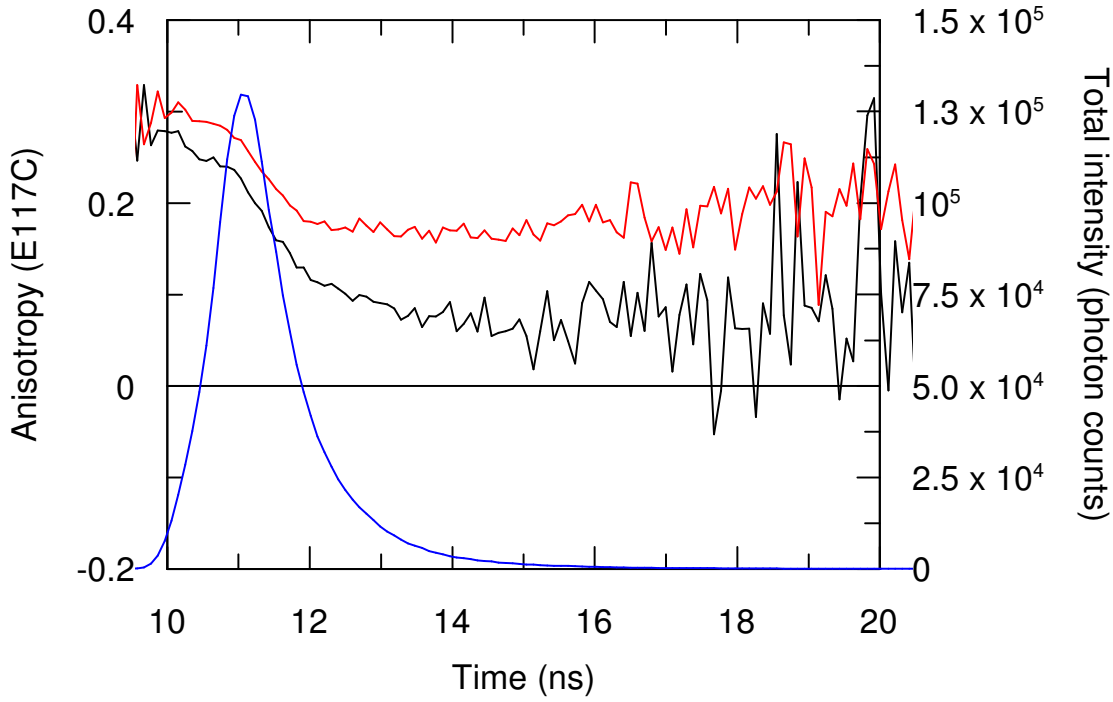


Figure 3-9: Anisotropy decays of labelled mutant Ocr proteins in the absence (black) or presence of GFP-MTase (red). Excitation was at 500 nm and emission at 570 nm. The total fluorescence intensity decay curve for the Ocr alone is shown in blue. The decays for labelled E20C and E117C Ocr decay exponentially in a few ns. The anisotropy decays for the S68C Ocr are either unusual showing a rise and decay (S68C Ocr by itself) or very long (with GFP-MTase).

Table 3-2: Time-resolved fluorescence decay analysis of samples showing FRET between GFP and HEX or Dylight549. Excitation at 405 nm, emission at 510 nm. The pre-exponential factor for each lifetime is given in the brackets.

Sample	τ1(ns)	τ2(ns)	τ3(ns)	χ^2	<τ> (ns)
GFP-MTase		2.20±0.01 (0.36)	3.01±0.02 (0.64)	1.082	2.72
GFP-MTase + 21B21TH DNA	0.29±0.01 (0.38)	1.42±0.01 (0.29)	2.84±0.01 (0.33)	1.076	1.45
GFP-MTase + 21BH21T DNA		1.61±0.01 (0.31)	2.84±0.02 (0.69)	1.061	2.46
GFP-MTase + E20C Ocr	0.28±0.01 (0.27)	1.42±0.01 (0.40)	2.72±0.01 (0.33)	1.052	1.55
GFP-MTase + S68C Ocr		1.37±0.01 (0.44)	2.67±0.02 (0.56)	1.173	2.09
GFP-MTase + E117C Ocr	0.33±0.01 (0.38)	1.28±0.01 (0.32)	2.84±0.01 (0.29)	1.012	1.38

3.2.4 Steady state FRET

In Figure 3-10, GFP-MTase alone shows a typical GFP emission. In the presence of 21B21TH or 21B21TH, the emission of GFP around 510 nm decreased. The emission of HEX around 555 nm appeared. The energy from GFP transferred to HEX Note that significantly more decrease of GFP emission occurs in the presence of 21B21TH rather than 21BH21T. The efficiency of FRET between GFP-MTase and 21B21TH (50.1 % transfer) was much higher than GFP-MTase and 21BH21T (8.5 % transfer). Therefore, the distance between the donor and acceptor is different in these samples. Using the calculated Förster distance, the GFP is separated from the HEX label of 21B21TH by 6.10±0.24 nm and from the HEX label on 21BH21T by 9.12±0.27 nm.

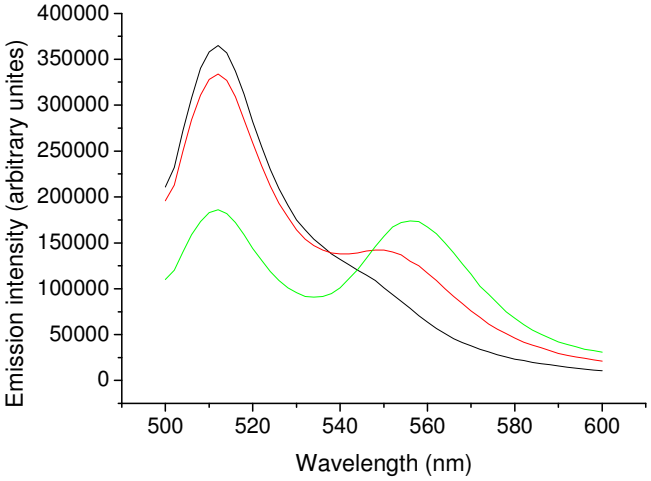


Figure 3-10: Fluorescence emission scan of either GFP-MTase in the absence (black) or presence of a fluorescently labelled 21 bp DNA duplex containing the bipartite recognition sequence for MTase (21BH21T, red; 21B21TH, green). The fluorescent HEX label was located at the 5' end of either the top (21B21TH) or bottom (21BH21T) strand of the 21 bp DNA duplex. All measurements were performed in 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME (pH 8.0) using 200 nM GFP-MTase and if required 200 nM DNA.

Figure 3-11 shows the induction of FRET when the GFP-MTase bound to the mutant Ocr proteins labelled with Dylight549. The amount of energy transfer depended on the mutant used. However, the introduction of a single cysteine into each Ocr subunit means that there are two FRET acceptors and, given the elongated shape of Ocr, these acceptors are highly likely to be located at different distances from the GFP donor. Hence the observed FRET was a complex average of the two distances given the $1/r^6$ dependence of FRET on distance. In the absence of further information, we simply calculated this "average" distance to be 7.62±0.23 nm, 10.21±0.31 nm and 6.60±0.20 nm for the E20C, S68C and E117C mutant Ocr proteins assuming κ^2 is 2/3.

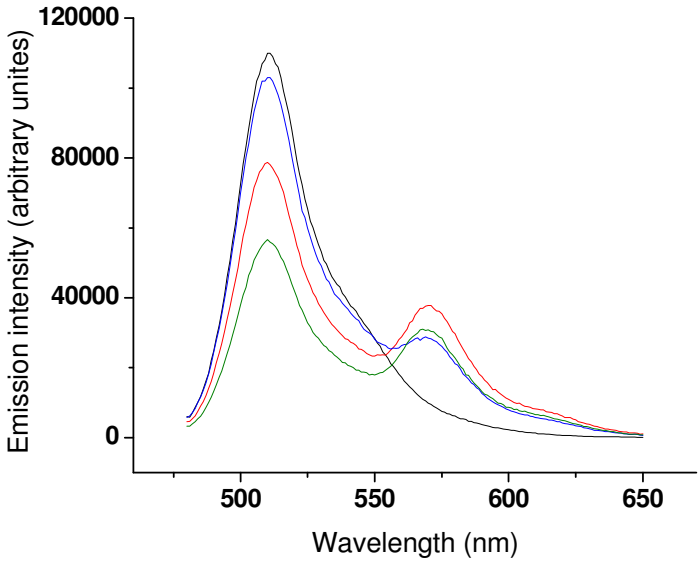


Figure 3-11: Fluorescence emission scan of either 500 nM GFP-MTase in the absence (black) or presence of 1.0 µM labelled Ocr mutations (E20C red, S68C blue and E117C green). Bottom (21BH21T) strand of the 21 bp DNA duplex. All measurements were performed in 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM 2ME (pH 8.0).

3.2.5 FRET measurements of GFP to HEX using time-resolved fluorescence

The fluorescence decay of the GFP for 1:1 mixtures of DNA and GFP-MTase and of labelled Ocr with GFP-MTase were determined. The emission was collected at the magic angle to remove undesirable anisotropy effects on the fluorescence decay and the fitted lifetimes are shown in Table 3-2.

It was apparent that the presence of the HEX label on DNA or the Dylight549 label on Ocr reduced the average fluorescence lifetime, $<\tau>$, of the GFP donor. This was indicative of energy transfer and an average distance between the donor and acceptor could be calculated (Table 3-3). This distance was in all cases except those using the S68C Ocr mutant protein, very similar to the distance calculated from the fluorescence intensity data.

It was also clear that the expected bi-exponential decay of GFP became a three exponential decay in some complexes so changes in individual lifetimes due to FRET could also be calculated. Considering first the bi-exponential decays, we assumed that since the pre-exponential factors remained roughly constant in the presence or absence of acceptor, that FRET shortens the 2.20 ns lifetime to 1.61 or 1.37 ns and the 3.01 ns lifetime to 2.84 or 2.67 ns for the 21B21TH and S68C samples, respectively allowing FRET efficiencies and distances to be calculated (Table 3-3). In the three-exponential decays, we assumed that the 3.01 ns lifetime split in to two components; the 2.7-2.8 ns component and the ~0.3 ns component as the sum of the two pre-exponential factors approximately equalled the initial 0.64 pre-exponential factor. The 2.20 ns lifetime, which once again hardly changed its pre-exponential factor, was assumed to decrease to the 1.3-1.4 ns lifetime. These assumptions allowed distances to be calculated. These interpretations imply multiple locations for the GFP with the electronic transition responsible for the 3.01 ns lifetime being particularly sensitive to an interaction with the acceptors. However, the photophysics of GFP and its derivatives is so complex in FRET experiments [Millington *et al.,* 2007] that it may be wise not to over interpret the distances calculated from the individual lifetimes, particularly since there are two acceptors on the Ocr mutant proteins, but rather to use the distance from the average lifetime when examining the location of GFP on the MTase structural model. This is particularly the case for the FRET between GFP and the label in the S68C Ocr mutant protein where the acceptor was not free to rotate.

Table 3-3: FRET distances (nm) calculated using fluorescence decay times of GFP-MTase in the absence or presence of the fluorescence acceptor compared to distances calculated from fluorescence intensity measurements. All distances are in nm.

Sample	Distance from τ3 to τ1	Distance from τ2 to τ2	Distance from τ3 to τ3	Distance from <τ>	Distance from intensity
GFP-MTase + 21B21TH DNA	4.23±0.13	6.79±0.20	9.80±0.30	6.27±0.19	6.10±0.24
GFP-MTase + 21BH21T DNA		7.26±0.22	9.90±0.30	8.97±0.27	9.38±0.27
GFP-MTase + E20C Ocr	4.47±0.14	7.22±0.22	9.48±0.31	6.84±0.20	7.62±0.25
GFP-MTase + S68C Ocr		7.10±0.22	9.21±0.30	7.98±0.23	10.21±0.36
GFP-MTase + E117C Ocr	4.61±0.15	6.90±0.25	10.4±0.38	6.56±0.30	6.60±0.26

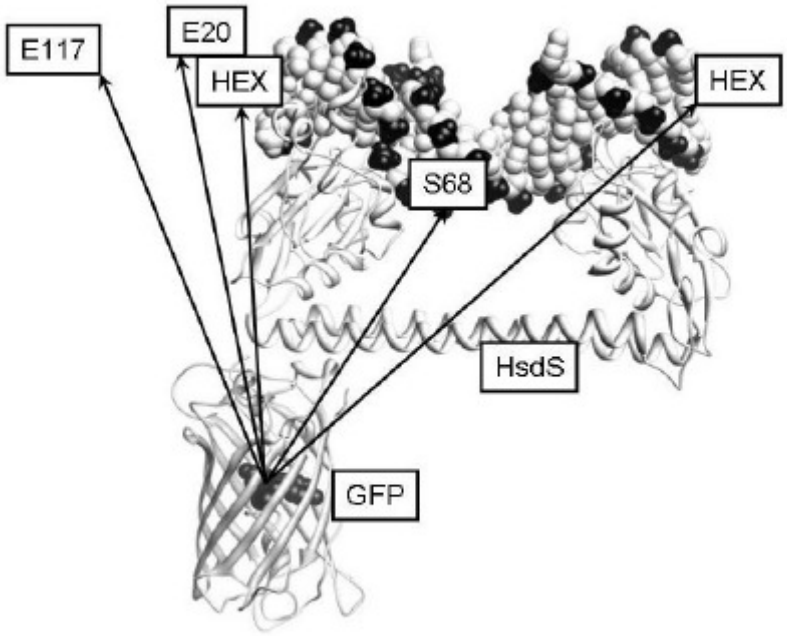


Figure 3-12: The HsdS subunit bound to a DNA duplex as proposed from electron microscopy data [Kennaway *et al.,* 2009] is shown above a GFP model with the chromophore shown in the centre of the GFP β-barrel. The locations of the HEX labels (21B21TH is on the left and 21BH21T is on the right) and of the locations of the Ocr residues labelled with Dylight549 are indicated (Ocr is not shown but superimposes on and extends further out than the DNA duplex shown). The arrows show the FRET distances determined from <τ> given in Table 3.2 except for the distance to S68 on Ocr where the distance in the actual model is shown (the FRET distance is longer for this pair but is incorrect due to rotational constraints on the acceptor, reprinted from Chen *et al*., 2010).

3.3 Discussion

Our results show that it is possible to fuse GFP to the C-terminus of HsdS in a Type I MTase without any deleterious effect on *in vivo* methylation or *in vitro* binding to either DNA or to a DNA mimic. The assembly of the trimeric MTase is also not

affected because the GFP appears to be able to adopt a range of conformations with respect to the MTase and freely move between them.

Recently Kennaway *et al*. [Kennaway *et. al.,* 2009] have published an atomic model of the MTase bound to a DNA duplex and to Ocr. Figure 3-12 shows the HsdS subunit bound to DNA with the GFP chromophore placed roughly at the distances determined by FRET using the average fluorescence lifetimes. It can be seen that the results all converge on approximately the same location for the GFP apart from the distance to the S68C location on the Ocr protein. The GFP is best located directly below one end of the HsdS subunit to satisfy the FRET distances. This location is what would be expected from the model of MTase.

We now use the GFP-MTase to investigate binding to the ArdA proteins in the next chapter.

Chapter 4 Investigation of the interaction between ArdA and MTase/GFP-MTase by different methods

ArdA proteins comprise a family of antirestriction proteins that inhibit Type I R-M systems and are often encoded on conjugative plasmids and transposons. Like Ocr, the ArdA family of proteins are highly acidic [Belogurov *et al*., 1995 Wilkins, 2002]. Despite the wide distribution of ArdA in many important pathogens, there is little biochemical data on their mode of action and behaviour in solution. It has been shown that ArdA from pathogens is very effective against the archetypal Type I RM systems of *Escherichia coli* by directly binding to the enzymes. Dryden and coworkers [McMahon *et al*., 2009] reported the first structure of an ArdA protein (Orf18) and its implications for the structure of Type I RM enzymes. Orf18 ArdA forms an extremely elongated molecule with a highly charged surface. Its structure is reminiscent of 42 bp duplex of B-form DNA making it the largest DNA mimic yet characterised. The structure of ArdA allows us to envisage how this interaction occurs. However, no binding affinity of MTase and ArdA has been reported. Here I used a range of techniques to investigate four ArdA proteins referred to as (Orf 18, Mu50, V583 and NCTC) and their mode of binding to MTase or GFP-MTase.

4.1 Structure of ArdA proteins

4.1.1 Structure of Orf18

The crystal structure of Orf18 was already reported [McMahon *et al*., 2009]. The two monomers of the protein form one independent dimer. Structural analysis shows that each monomer can be split into three domains: an N-terminal domain 1 (residues 3-61), the central domain 2 (62-103) and the C-terminal domain 3 (residues 104-165) (Figure 1-16). The three domains of the ArdA monomer are arranged in an approximately linear manner giving a very elongated molecular shape (7.0 nm×2.0

nm) with a definite curve. The two monomers are arranged in a head-to-head orientation. The dimer, like the monomer, forms a highly elongated and curved structure and has a length of 14 nm.

4.1.2 Comparison of Orf18 with other ArdA sequences

Unfortunately, no crystal structures of NCTC, Mu50 and V583 are available. Figure 4-1 shows an alignment of the amino acid sequences of the ArdA investigated in this study and by others. As can be seen, there is considerable variability in sequence in this set but the pattern of charged residues and the two motifs identified previously are well conserved, indicating that all ArdA are very likely to have the same structure and to operate in the same manner. Biochemical data on the ArdA listed in Table 4-1 [McMahon *et al*., 2009] when compared to that from Tn916 ArdA (Orf18), support this assertion.

Table 4-1: The source of the ArdA genes investigated (reprinted from McMahon *et al*., 2009).

Reference	Organism	Gene ID	Amino acid length	Predicted pI	Optimal protein expression temperature;
1	*Enterococcus faecalis*	Tn 916 (Orf18)	165	3.91	37
2	*Enterococcus faecalis* V583	EF2335	166	3.91	30
2	*Staphylococcus aureus* Mu50	SAV0405	166	3.92	30
3	*Bacteroides fragilis* NCTC 9343	BF1222	177	3.98	37

```
Tn916          ----MDDMQVYIANLGKYNEGELVGAWFTFP-----IDFEEVKEKIGLNDEYEEYAIHD 50
Mu50           -----MEMKVYVANLGRYNEGELVGAWFTPP-----IDEEEMAERIGLNEDYEEYAIHD 49
V583           ----MEQMRVYIANLGKYNEGELVGAWFTPP-----VDFDEVKERIGLNDDYEEYAIHD 50
NCTC          MEAVTLSEARVYVGTYNKYNNGSLFGKWLDLSDYSDKDEFMEACRELHKDDQDPEFMFQD 60
                     : :**:.. .:**:*.*.* *:  .      :  *  ..:  :::  *: ::*

Tn916        YELP--FTVDEYTSIGELNRLWEMVSELP-EELQSELSALL-THFSSIEELSEHQEDIII 106
Mu50         FELP--FDVDEYTPISEINRLCEAIQEIEGTPIYNELKEIQGMWFSSLEELLENKEDIHC 107
V583         YELP--FEIDEYTSIEEINRLCGLAEELEGTPIGEVASEIQHAFFNSFEEMVEHVDDIVY 108
NCTC         YENIPEALISESWLSEKFFELRDAIEKLSETQQEAFFVWCDHHNSDISEEDADDLISSFE 120
              :*      :.*     :: .*    .::                 .  **  :.  .
```

```
Tn916        HSDCDDMYDVARYYIEETGALGEVPASLQNYIDYQAYGRDLDLSGTFISTNHGIFEIVY 165
Mu50         YSDCDSMEDVARYYVEETGQLGEVPSNLQNYIDYQSLGRDMEIEGNYLVTSHGVFEYCQ 166
V583         YPDCNDMEDLA-YQMVNEGYLGDAPENFVRYFNYSSFARDLEIEGNYLVTNRGIFEYPI 166
NCTC         DEYQGEYKDEEDYAYEIVEQCYDLPEFAKTYFDYSAFARDLFIT-DYWMDNGFVFRCA- 177
                 ..  *   *         : *     *::*.: .**: :   :   .  :*.
```

Figure 4-1: CLUSTALW multiple sequence alignment of the four ArdA proteins. Positions of absolute conservation are highlighted underneath with an asterisk. Regions displaying good to moderate conservation are highlighted with a double or single dot, respectively. The amino acids in the sequence are colour coded as follows: blue, acidic residues; magenta, basic residues; green, hydroxyl, amine, basic (Q) residues; red, small plus hydrophobic residues (including aromatic Y).

4.2 Isothermal titration calorimetry results

Isothermal titration calorimetry (ITC) was used to study the binding of ArdA to MTase and GFP-MTase. Titrations of ArdA into MTase or GFP-MTase solution were carried out at 25°C using a VP-ITC calorimeter (MicroCal, Northampton, UK). Both the ArdA and MTase solutions were buffer exchanged into 20 mM Tris-HCl, 6 mM $MgCl_2$, 7 mM ME, pH 8.0. Experiments were performed in duplicate and each titration consisted of 30 injections of 10 µl of ArdA into 3.0 µM MTase/GFP-MTase.

Typical ITC results of the interaction between MTase and ArdA are shown in Figure 4-2. The interaction of MTase with ArdA is exothermic with a clear end point corresponding to the dimer concentration of ArdA at which MTase becomes saturated. After integration of each injection peak and subtraction of heats of dilution, the values of the heat of interaction of each injection were normalised for the ArdA dimer concentration and plotted *versus* the molar ratio of ArdA to MTase. The data were fitted to a theoretical titration curve (single-site model) to determine the stoichiometry (n) of interaction and enthalpy (ΔH) by MicroCal LLC Origin software. ITC analysis indicates significant differences between the four ArdAs in their ability to bind MTase /GFP- MTase.

4.2.1 ArdA binds MTase

Figure 4-2 shows a typical titration of ArdA into MTase. These results show the binding between the four ArdA proteins and MTase is quite strong (Kd is about 1 nM). However, the very sharp transition precluded reliable determination of the binding affinity using ITC. The calculated ΔH values are listed on Table 4-2.

The results show the binding stoichiometry between the four ArdA proteins and MTase were not the same (Table 4-3). Roughly, one monomer of Orf18 and V583 binds to one molecule of MTase. By contrast, one dimer of NCTC and Mu50 binds one molecule of MTase. This result was surprising because we had assumed all the ArdA proteins possess a similar structure. Thus, the mode of binding to MTase should be very similar i.e., same stoichiometry of interaction. However, as shown in Table 4-3, the n values were totally different. It must, however, be remembered that the n value is the average number of binding sites per mole of protein in solution and that this is based on a number of assumptions. Firstly, all the binding sites are assumed to be identical and independent. Secondly, the two interacting proteins (MTase and ArdA) are assumed to be homogeneous and their concentration needs to be accurately known. SDS-PAGE analysis clearly demonstrates that the proteins used in this study were highly purified and essentially homogeneous. Furthermore, the homogeneous nature of the ArdA proteins was further verified by electrospray mass spectrometry (D. Clarke & G. A. Roberts, personal communication). However, there is typically an error of ±5% associated with the extinction coefficient based on amino acid sequence data [Gill *et al*., 1989 and Pace *et al*., 1995]. Thirdly, it is assumed that all the protein in the sample is correctly folded (i.e., protein concentration corresponds to active protein concentration). However, there is no evidence of aggregate formation within the sample based upon examination of the baseline of the UV-visible absorption scan. Specifically, the baseline between 300 to 400 nm was flat in each case.

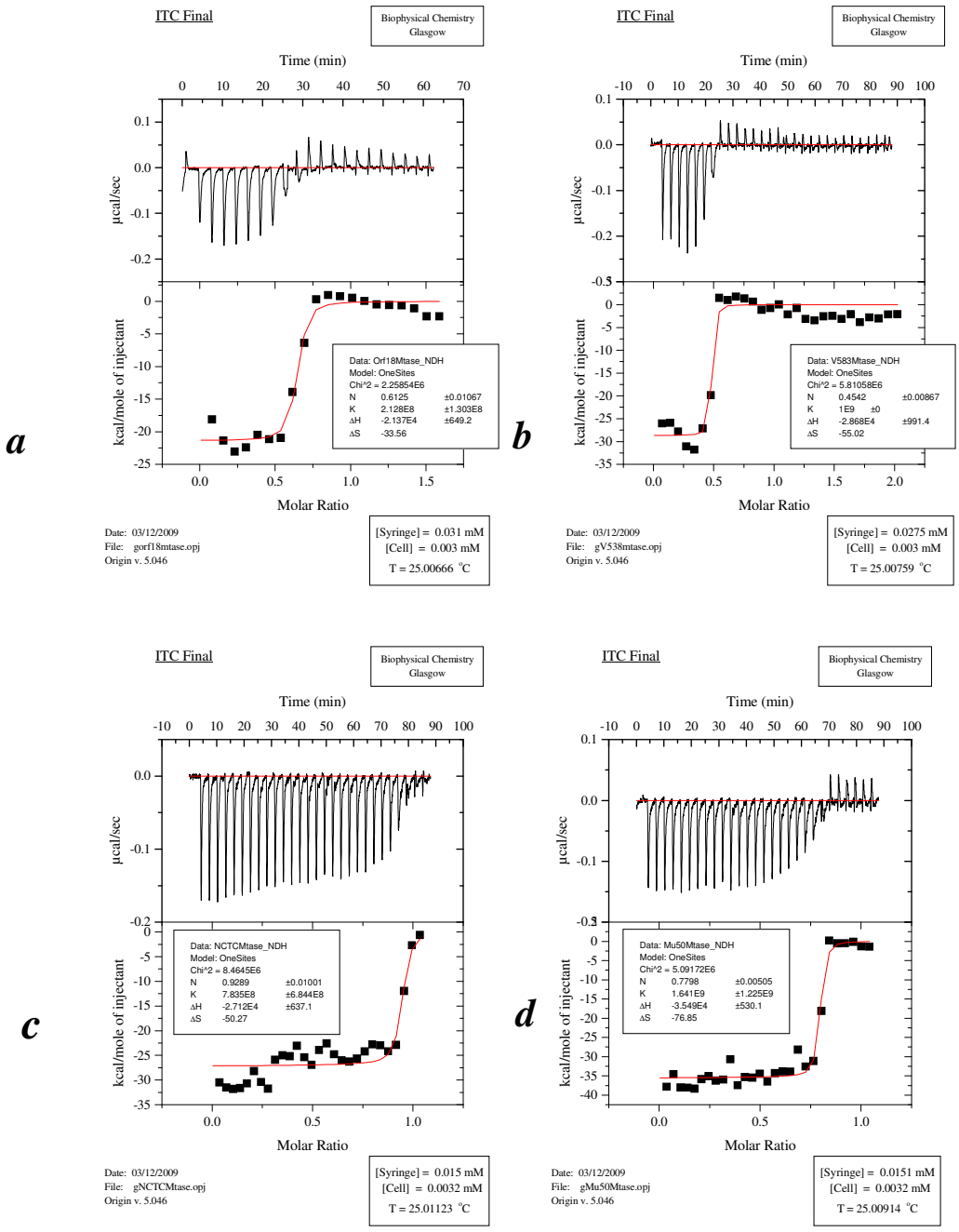


Figure 4-2: Calorimetric titration of ArdA into MTase. The ITC titration curve is shown in the upper panel. The lower panel shows the calorimetric binding isotherm for the ArdA-MTase interaction. The x axis shows the molar ratio of ArdA (dimer concentration) to MTase. All experiments were carried at 25 °C. Panel a is Orf18 titrating MTase, panel b is V583 titrating MTase, panel c is NCTC titrating MTase, panel d is Mu50 titrating MTase. The results are presented in the box next to the respective titration. N is the stoichiometry of

dimer ArdA:MTase; K is the association constant (Ka) of the interaction (M^{-1}); ΔH is the enthalpy change (cal/mol); ΔS is the entropy change (cal/K mol).

4.2.2 ArdA binds GFP-MTase

Figure 4-3 shows the results of ArdA being titrated into GFP-MTase. These results show the binding between the four ArdA proteins and GFP-MTase is, like ArdA and MTase, very strong (Kd is about 1 nM). However, the ITC results cannot be used to determine the binding affinity because the transition is too sharp. The ΔH values, stoichiometry and Kd/Ka of interaction between ArdA and GFP-MTase are listed in Table 4-2, 4-3 and 4-4, respectively.

Compared with the interaction of ArdA and MTase, binding of ArdA and GFP-MTase is less exothermic. This means the interaction of ArdA and GFP-MTase releases less heat than the interaction of ArdA and MTase. Generally, the stoichiometry of interaction of ArdA and GFP-MTase is similar to the interaction of ArdA and MTase. Only the interaction of V583 binding to the two MTases show a significant difference.

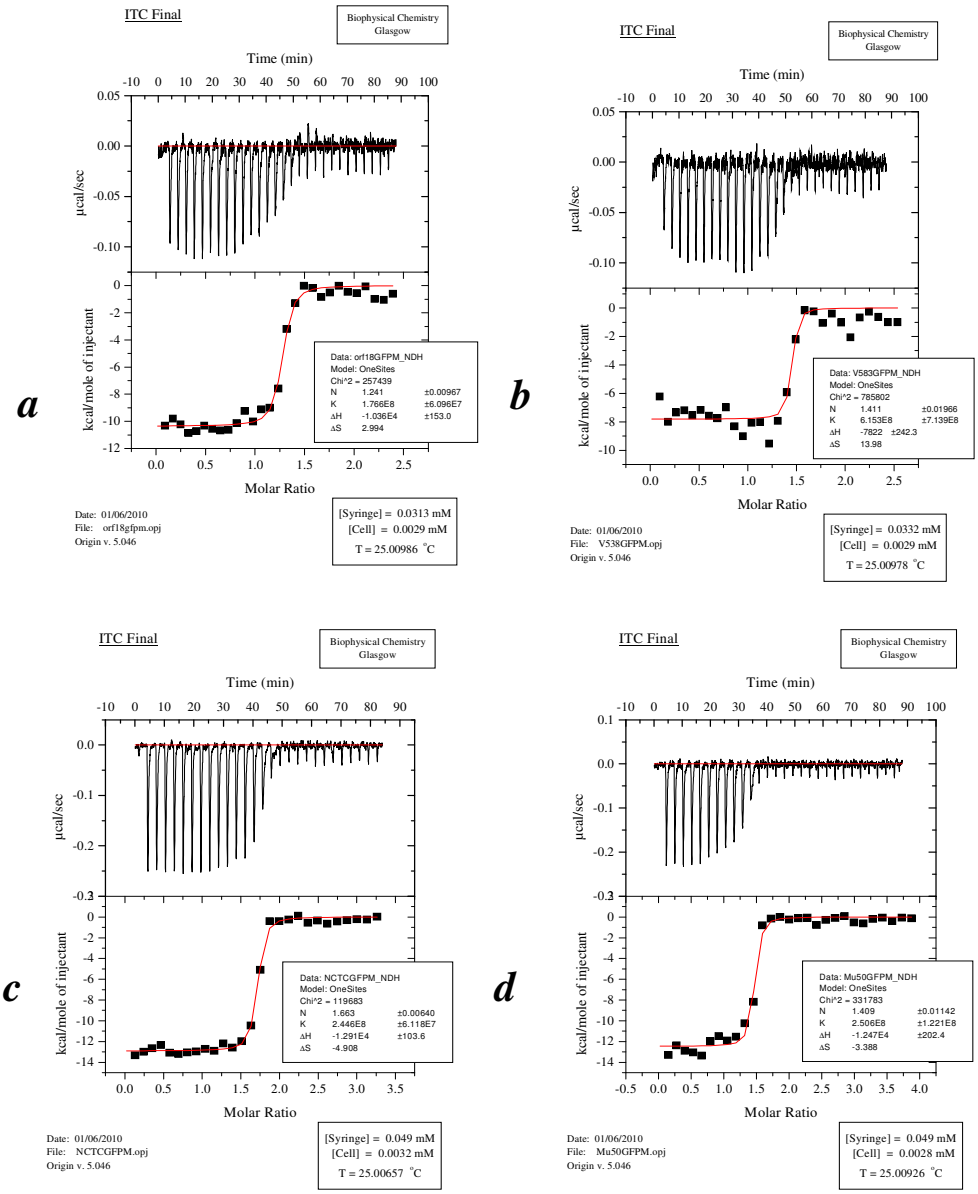


Figure 4-3: Calorimetric titration of ArdA into GFP-MTase. The ITC titration curve is shown in the upper panel. The lower panel shows the calorimetric binding isotherm for the ArdA-MTase interaction. The x axis shows the molar ratio of ArdA (monomer concentration) to MTase. All experiments were carried at 25 °C. Panel a is Orf18 titrating MTase, panel b is V583 titrating MTase, panel c is NCTC titrating MTase, panel d is Mu50 titrating MTase. The results are summarised in the box next to the respective titration. N is the stoichiometry of dimer ArdA:GFP-MTase; K is the association constant of the interaction (M^{-}

[1]); ΔH is the change of enthalpy (cal/mol); ΔS is the change of entropy (cal/K mol).

Table 4-2: ΔH values of interaction between MTase or GFP-MTase with dimer ArdA (kcal/mol).

	MTase	GFP-MTase
Orf18	-21.4±0.6	-20.8±0.4
Mu50	-35.5±0.5	-25.0±0.4
V583	-28.7±1.0	-15.6±0.4
NCTC	-27.1±0.6	-25.8±0.2

Table 4-3: The stoichiometry of interaction between MTase or GFP-MTase with ArdA (i.e., monomer ArdA: MTase or GFP-MTase).

	MTase	GFP-MTase
Orf18	1.22±0.07	1.24±0.07
Mu50	1.56±0.09	1.41±0.08
V583	0.91±0.06	1.41±0.08
NCTC	1.86±0.10	1.66±0.09

Table 4-4: The Ka (M^{-1}) and Kd (nM) transfer. The value of Ka and Kd is not accurate due to tight binding.

	MTase		GFP-MTase	
	Ka	Kd	Ka	Kd
Orf18	2.1E8±1.3E8	4.8±3.0	1.8E8±0.6E8	5.6±0.19
Mu50	1.6E9±1.2E9	0.63±0.47	2.5E8±1.2E8	4.0±1.9
V583	1E9±0*	1±0*	6.2E8±7.1E8	1.6±1.8
NCTC	7.8E8±6.8E8	1.3±1.1	2.4E8±0.6E7	0.42±0.10

* Very tight interaction that does not allow correct fitting. In this case fitting was fixed at a Kd of 1 nM.

4.2.3 Summary of ITC results

Based on the ITC data, the stoichiometry of interaction between ArdA proteins and MTase/GFP-MTase are different. Previous results had shown that one Ocr dimer binds to one molecule of MTase (i.e., binding ratio of 2 Ocr monomers per MTase molecule; refer to Atanasiu *et al*., 2002). Therefore, it was perhaps expected that the ArdA proteins would display a similar stoichiometry to that of Ocr in terms of their interaction with MTase. However, in each case the stoichiometry is not one molecule MTase/GFP-MTase binding one dimer ArdA protein. The endpoint between 1 and 2 indicates monomer ArdA protein could be bound by MTase/GFP-MTase. Due to the interactions of ArdA proteins and MTase/GFP-MTase being very tight, we cannot determine an accurate binding affinity from the ITC results and another approach must be employed. Thus, size-exclusion HPLC experiments were performed in order to determine the binding affinity for the various ArdA proteins and MTase.

4.3 Size-exclusion HPLC

Size-exclusion HPLC experiments were used to determine the binding affinity of ArdA and MTase or GFP-MTase and to establish the quaternary structure of ArdA and the complex of ArdA and MTase over a range of protein concentrations. In size-exclusion chromatography, molecules are separated according to their molecular size. The size exclusion HPLC column was prepacked with gel particles that form a separation matrix through which a buffer solution is passed. The small molecules, which freely diffuse into the gel beads, are delayed in their passage through the column compared with the large molecules, which either cannot enter the gel (i.e. are totally excluded) or can only partially enter the gel. Thus, large molecules elute from the column ahead of the smaller molecules in accordance to their respective size.

Tryptophan fluorescence was used to detect protein in the eluate. The column was equilibrated in 20 mM Tris-HCl, 20 mM MES, 200 mM NaCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 7 mM 2ME, pH 6.5 buffer (Buffer A). The flow rate was set to 0.5 ml/min and the sample volume was 10 µl (i.e. 0.2 % of column volume). The column eluate was excited at 295 nm and continuously monitored at a wavelength of 350 nm using a flow-through spectrophotometer. All size-exclusion experiments were

conducted at 20°C. All ArdA protein concentration in section 4.3 are dimer concentrations. All samples were prepared at high concentration then dilution to the low concentration.

4.3.1 Calibration of column

A Biosep-SEC-s 3000 gel filtration column (Phenomenex, Torrance, USA) was first calibrated with standard proteins and the resulting calibration curve is shown in figure 4-4. Apoferritin (443 kDa), β-amylase (200 kDa), alcohol dehydrogenase (150 kDa), bovine serum albumen (66 kDa) and carbonic anhydrase (29 kDa) were individually employed as molecular mass markers for calibration of the column. All of these standard proteins are known to be globular. Plotting the elution volume vs. log molecular mass of standards was used to calculate the molecular mass of other proteins.

The Orf 18 ArdA has a highly elongated structure [McMahon *et al.*, 2009]. Therefore, the molecular weight of ArdA proteins determined by size-exclusion chromatography will not be accurate. This is because an elongated protein runs anomalously quickly on a gel filtration column due to a much larger than anticipated Stokes radius (or hydrodynamic radius) as the molecule tumbles in solution.

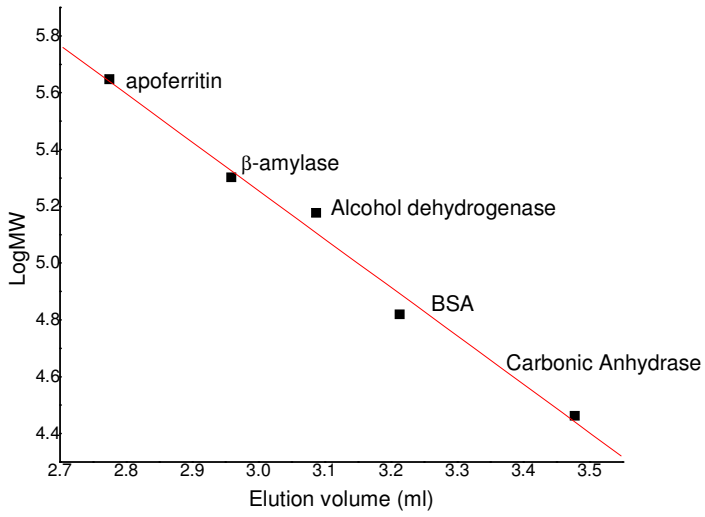
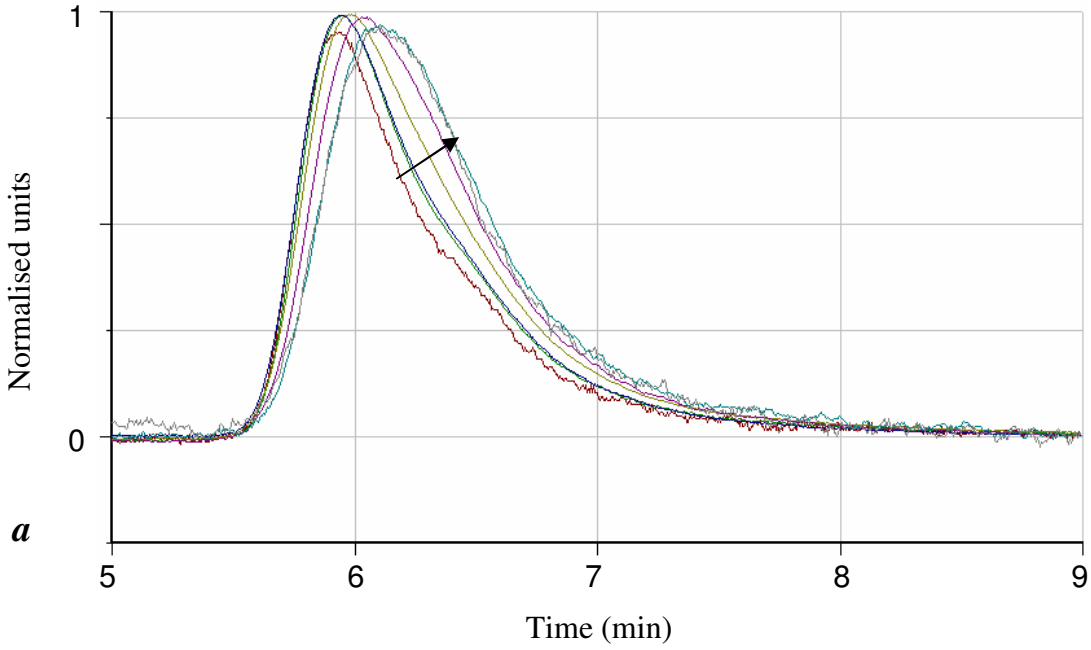


Figure 4-4: Calibration graph of standard proteins. A Biosep-SEC-s 3000 gel filtration column was calibrated with standard proteins. The Equation derived from plotting the elution volume vs. log molecular mass of standards was used to calculate the molecular mass of ArdA. The column was equilibrated with buffer A. Flow rate was 0.5 ml/min and sample volume was 10 µl. The column eluate was monitored using a fluorescence detector (excitation at 295 nm, emission 350 nm). The elution time of each sample was multiplied by the flow rate (0.5 ml/min) to obtain the elution volume. All size-exclusion experiments were done at 20°C. The calibration equation is: LogMW=10.37-1.705×(elution volume).

4.3.2 MTase behaviour

Figure 4-5 shows the results of gel filtration chromatography using different concentrations of MTase. As the concentration of MTase decreased, the elution volume increased. This finding indicates that MTase disassociates at low protein concentration. The Kd was 148±41 nM and this result is similar to the previously published experiment. In Dryden paper [Dryden *et al*., 1997] at that time Kd was estimated to be in 10-100 nM range. At very low protein concentrations (e.g., 20 nM) no clear signal corresponding to protein can be detected in the eluate. Figure 4-5 b shows the fitting curve of a 1:1 binding ratio reaction/interaction. At very high concentration, the molecular weight of MTase is about 208,000. This value is very close to 169,000, which corresponds is the anticipated molecular weight of M_2S_1 based on the known amino acid sequence. The apparent molecular weight of MTase decreased dramatically when the protein concentration was reduced. At very low concentrations, the experimentally determined molecular weight was about 105,000. This species presumably corresponds to M_1S_1 whose predicted molecular weight is 110,000. Thus, because MTase corresponds to a globular protein it was possible to determine a reasonably accurate molecular weight using the calibration graph shown in Figure 4-4. These findings are similar to the previously published results.



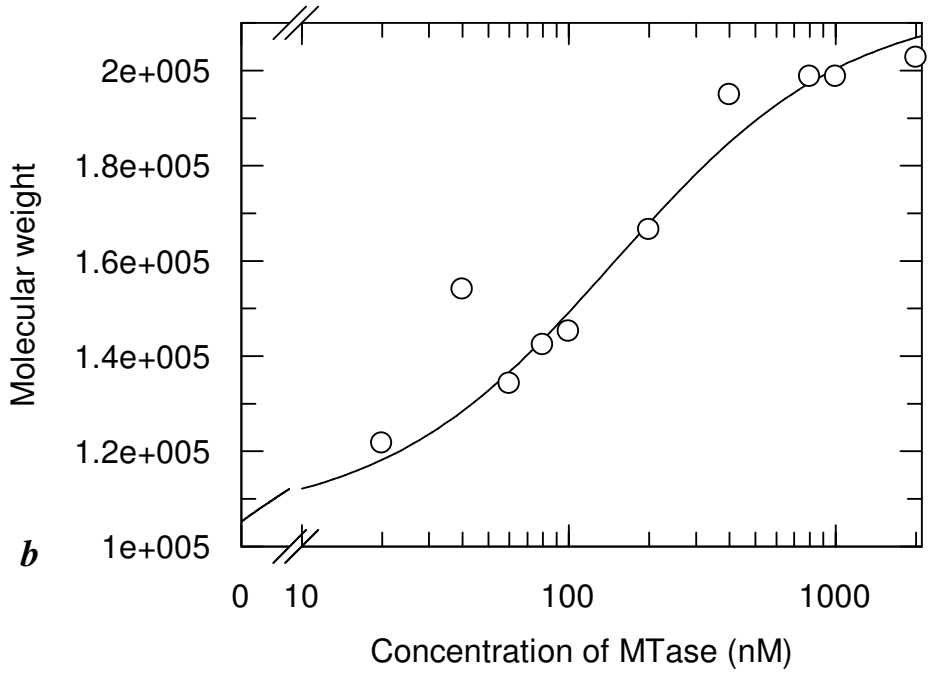


Figure 4-5: Size-exclusion chromatography analysis of MTase to investigate the solution molecular mass at different MTase concentrations. The column was equilibrated in buffer A. Experimental conditions were the same as described

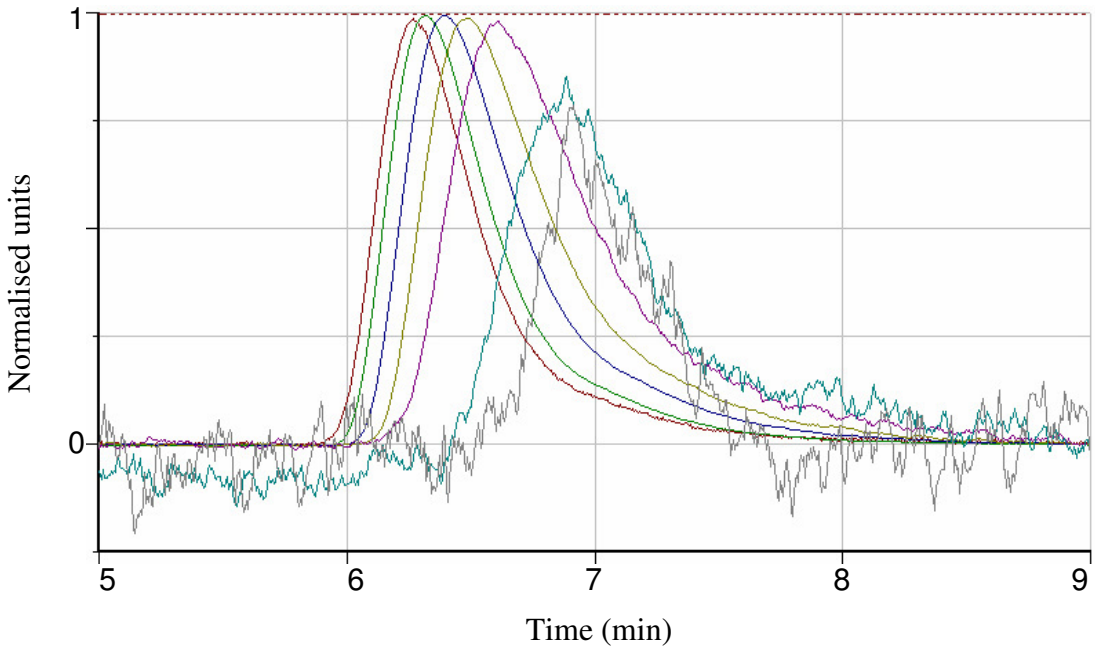
previously (see Section 4.3.1). Panel *a* is the raw data showing the fluorescence signal of the eluate against elution time (min). For clarity, only a selected number of traces are shown. The arrow indicates the effect of decreasing the concentration of MTase. The profiles that intersect the arrow from left to right correspond to an MTase concentration of 2, 1, 0.8, 0.4, 0.2, 0.1, 0.08 µM, respectively. Panel *b* apparent molecular weight of MTase at different concentrations as determined from the elution time on the calibrated gel filtration column. The black curve is a 1:1 binding fitting curve for $M1S1 + M \rightleftharpoons M2S1$.

4.3.3 NCTC interaction with MTase

4.3.3.1 NCTC behaviour

NCTC was tested at the following dimer concentrations: 20 nM, 40 nM, 100 nM, 200 nM, 400 nM, 1 µM and 2 µM. Tryptophan fluorescence was used to detect the protein. Figure 4-6 a shows some typical gel filtration results. Experimental conditions were the same as described previously (see Section 4.3.1). Decreasing concentration of NCTC caused the elution time of NCTC to increase. This showed that at higher concentrations of NCTC the molecular weight was greater than at lower concentrations. The results shown in Figure 4-6 b were used to calculate the molecular weight of NCTC at different concentrations of protein. The elution time of each sample was multiplied by the flow rate (0.5 ml/min) to obtain the elution volume. This elution volume was used as the "x" parameter in the Equation of the calibration line and the molecular mass of the NCTC was obtained from y. All of the standard proteins were globular. Thus, the molecular mass of the ArdA protein obtained from the calibrated column is expected to give an estimate of the molecular weight for a globular protein. However, it is anticipated that the ArdA molecule will be highly elongated and therefore the molecular weight from size-exclusion HPLC will be an overestimate. Nonetheless, we can use the data to assess the relative size of the protein.

Figure 4-6 b shows at very low concentrations of NCTC (i.e., 10 nM monomer) the apparent molecular weight was 28,000 (i.e., monomer). However, at higher protein concentrations (400 nM and 800 nM monomer) NCTC has an apparent molecular weight of 70,000 corresponding to a dimer. At concentrations greater than 2 µM NCTC the molecular weight was about 110,000 (i.e., over three-fold greater molecular weight than at low protein concentration). This result indicates that NCTC probably aggregates at very high concentration, possibly forming tetramers. The data were fitted using a 1:1 binding Equation so this fit is only to guide the eye.



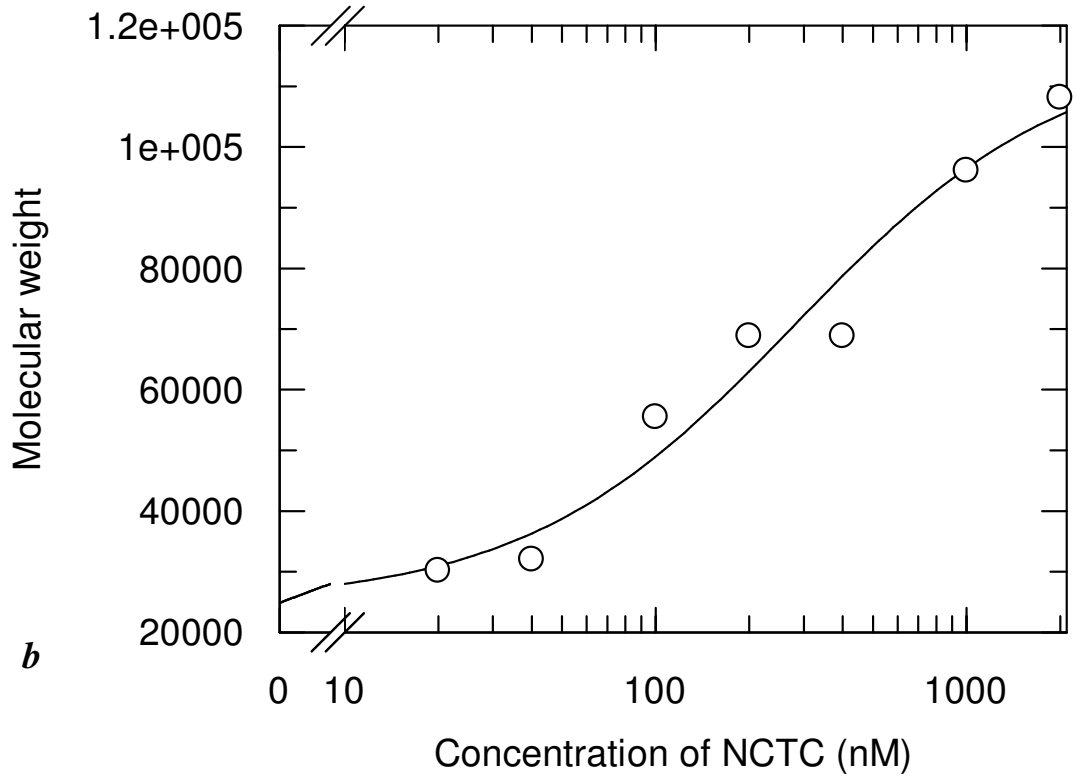
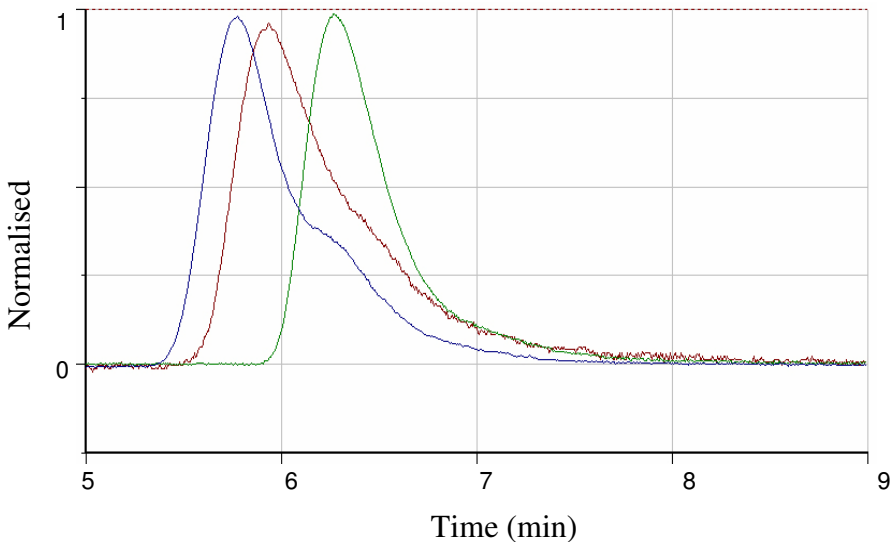


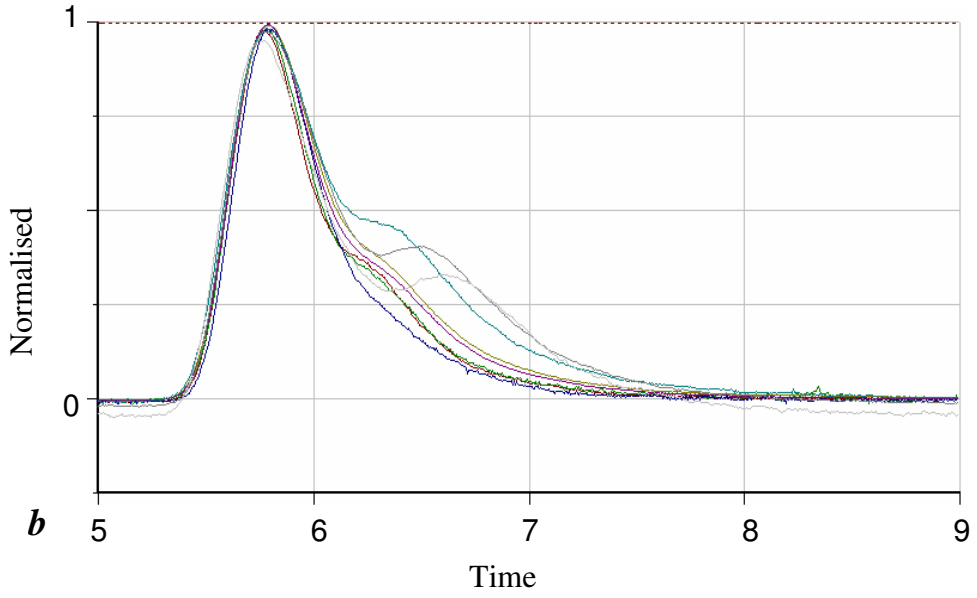
Figure 4-6: Size-exclusion chromatography analysis of NCTC to investigate the solution molecular mass at different NCTC dimer concentrations. The column was equilibrated in buffer A. Experimental conditions were the same as described previously (see Section 4.3.1). Panel *a* is the raw data showing the fluorescence signal of the eluate against elution time (min). The arrow indicates the effect of decreasing the concentration of NCTC. The profiles that intersect the arrow from left to right correspond to a NCTC concentration of 2, 1, 0.4, 0.2, 0.1, 0.04, 0.02 µM, respectively. Panel *b* apparent molecular weight of NCTC at different concentrations as determined from the elution time on the calibrated gel filtration column. The black curve is a 1:1 binding curve.

4.3.3.2 The complex of MTase and NCTC

The behaviour of a 1:1 mixture of MTase and NCTC dimer during size-exclusion HPLC is shown in Figure 4-7. From Figure 4-7 a we can see that the peak of the complex of NCTC and MTase is earlier than the other two peaks (corresponding to free NCTC and MTase). This means NCTC and MTase forms a stable complex,

which elutes from the column first. In addition, some dissociated or unbound NCTC was also detected. This observation infers some of the complex dissociated during the gel filtration experiment or that not all of the NCTC protein formed a complex initially. Figure 4-7 b and c shows the complex did not dissociate during the gel filtration process at any detectable protein concentration. Thus NCTC stabilises the trimeric structure of MTase. The molecular mass was around 280,000. This result further indicates that once NCTC was bound, it remains bound to MTase for periods significantly longer than 6 min (i.e., approximate duration of the gel filtration experiment). Therefore, the association between NCTC and MTase must be very tight with a very slow dissociation rate for the complex. Unfortunately, it was not possible to calculate the binding constant using these results alone.





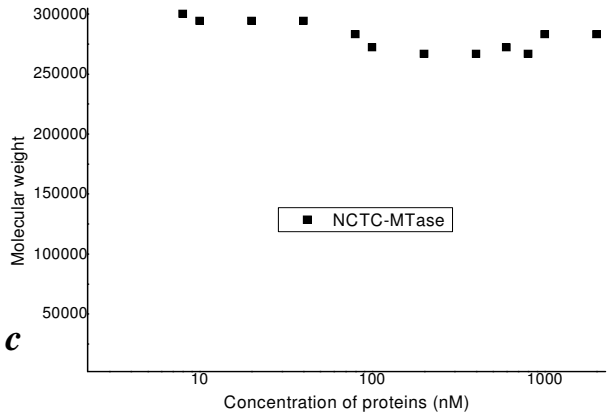


Figure 4-7: Size-exclusion chromatography analysis of binding between MTase and NCTC to investigate the solution molecular mass at different protein concentrations. The column was equilibrated in buffer A. Experimental conditions were the same as described previously (see Section 4.3.1). Panel *a* is the data for the high protein concentrations showing the fluorescence signal of the eluate against elution time (min). Blue is the mixture of 2 µM MTase and 2 µM dimer NCTC; red is 2 µM MTase; green is 2 µM dimer NCTC. Panel *b* is data for a range of protein concentrations (2, 1, 0.8, 0.6, 0.4, 0.2, 0.1, 0.08 µM) complex of MTase and NCTC showing the fluorescence signal of the eluate against elution time (min). Panel *c* apparent molecular weight of the complex at

different concentrations as determined from the elution time on the calibrated gel filtration column.

4.3.3.3 Summary

Size-exclusion HPLC was performed in a buffer containing 200 mM NaCl to prevent unwanted interaction between the protein and gel matrix. We already know that salt decreases the binding affinity of many DNA binding proteins and this can be attributed to the polyelectrolyte nature of DNA [Powell *et al*., 1998]. Specifically, the protein is competing for binding sites on the DNA with a high concentration of cations and this reduces the apparent binding affinity of the protein. Thus, the interaction of ArdA and MTase/GFP-MTase maybe considerably weakened by the presence of 200 mM NaCl. In the ITC experiments, no NaCl was included in the buffer. Hence, the binding observed in the ITC experiment will be much stronger than in the size-exclusion HPLC conditions.

However, Figure 4-8, it is obvious that NCTC stabilises the trimeric structure of MTase, even in the presence of 200 mM NaCl. The stability of the NCTC:MTase complex may be slightly less than that for Ocr and MTase, where no free Ocr was observed. Nonetheless, the binding of NCTC and MTase is still too tight to calculate a binding constant from these results.

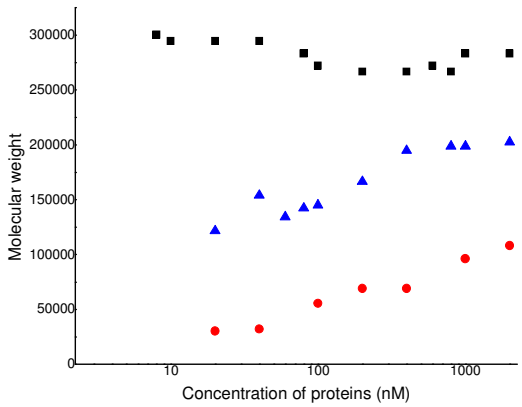


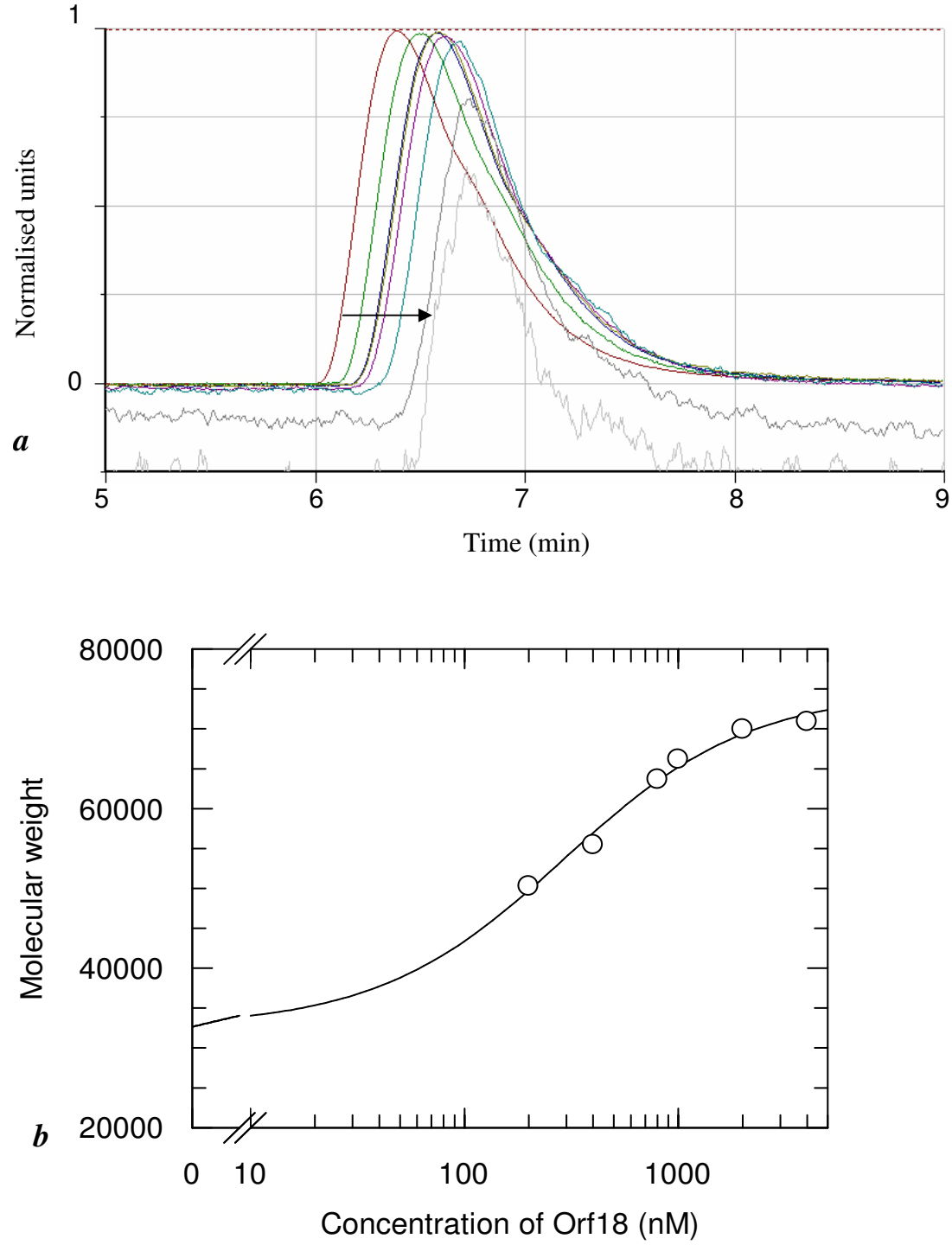
Figure 4-8: Size-exclusion chromatography analysis of NCTC and MTase to investigate the solution molecular mass at different protein concentrations. Red

circles: molecular mass values of NCTC at different dimer concentrations; Blue triangles: molecular mass values of MTase at different concentrations; Black squares: molecular mass values of the complex of NCTC and MTase at different concentrations.

4.3.4 Other ArdA proteins: Orf18, V583 and Mu50

4.3.4.1 ArdA proteins in absence of MTase

Figure 4-9 ***a*** shows that decreasing the concentration of Orf18 caused the elution time of Orf18 to increase. This indicates that at higher concentrations of Orf18, the molecular weight was greater than at lower protein concentrations. Figure 4-9 ***b*** shows that at a protein concentration of 2 μM or above the Orf18 molecular weight did not change too much. However, there was a significant decrease in the apparent solution molecular weight of Orf18 as the protein concentration was reduced from 1000 nM to 200 nM. At concentrations lower than 200 nM the protein peak could not be detected (due to very bad signal to noise ratio). However, the curve fitting clearly shows at very low concentration the molecular weight of Orf18 is about 33,000. The predicted monomer molecular weight of Orf18, based on its amino acid sequence, is 19,125. However, the shape of Orf18 is highly elongated. Therefore, at low concentrations Orf18 most probably exists as a monomer in solution. At high concentration (≥2 μM) the molecular weight of Orf18 is about 70,000, which is slightly more than twice the molecular weight at low concentration. This result indicates that most of the Orf18 is present as a dimer at concentrations of ≥2 μM. At intermediate concentrations, Orf18 exists as a mixture of monomer and dimer. This result is accordance with the previous published paper (Serfiotis-Mitsa *et al*., 2008).

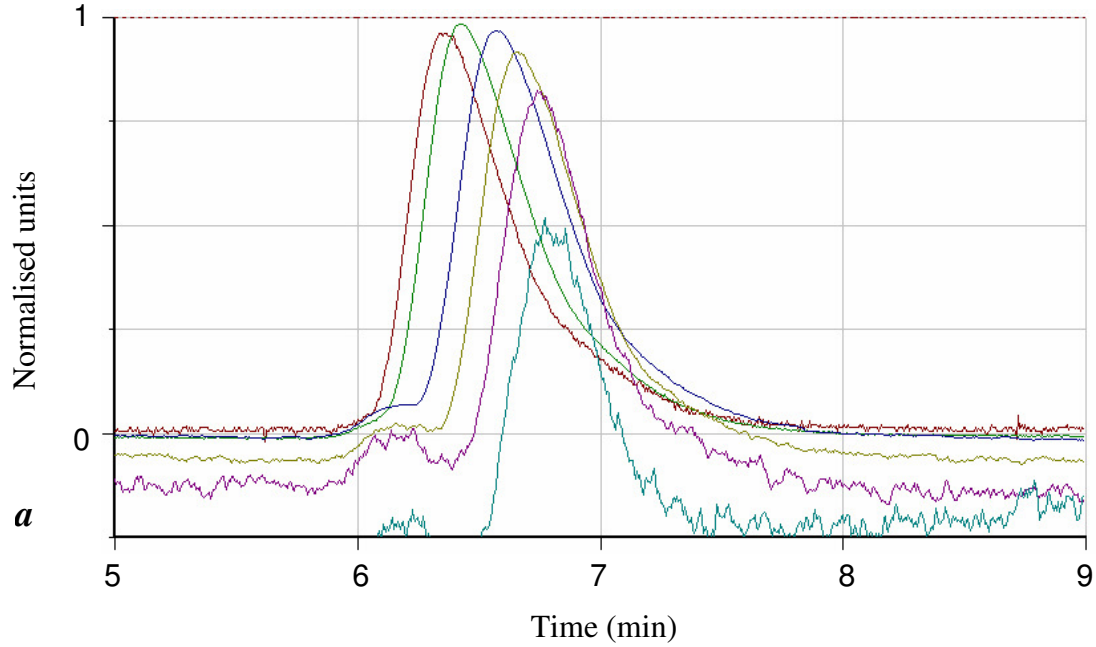


.

Figure 4-9: Size-exclusion chromatography analysis of Orf18 to investigate the solution molecular mass at different Orf18 dimer concentrations. The column was equilibrated in buffer A. Experimental conditions were the same as described previously (see Section 4.3.1). Panel *a* is the raw data showing the fluorescence signal of the eluate against elution time (min). The arrow indicates

the effect of decreasing the concentration of Orf18. The profiles that intersect the arrow from left to right correspond to an Orf18 concentration of 4, 2, 1, 0.8, 0.6, 0.4, 0.2, 0.1 µM, respectively. Panel *b* apparent molecular weight of Orf18 at different concentrations as determined from the elution time on the calibrated gel filtration column. The black curve is a 1:1 binding curve.

As anticipated from the results with the other ArdA proteins, decreasing the concentration of V583 caused the elution time to increase (Figure 4-10 a). From the fitting curve, it can be seen that at very low concentrations of protein the solution molecular weight of V583 is about 31,000. The predicted molecular weight of V583 based on its amino acid sequence is 19,125. However, by analogy to the crystal structure of Orf18, we expect the V583 molecule to be highly elongated. Thus, at low concentration V583 exists as a monomer in solution (with an apparent molecular weight of 31,000). At very high concentration (i.e. 2 µM) the apparent solution molecular weight of V583 is probably about 90,000, which is almost three times its molecular weight at very low concentration. This result indicates that most of the V583 is present as a dimer at a concentration of ~2 µM. At intermediate concentrations, V583 is present as a mixture of monomer and dimer.



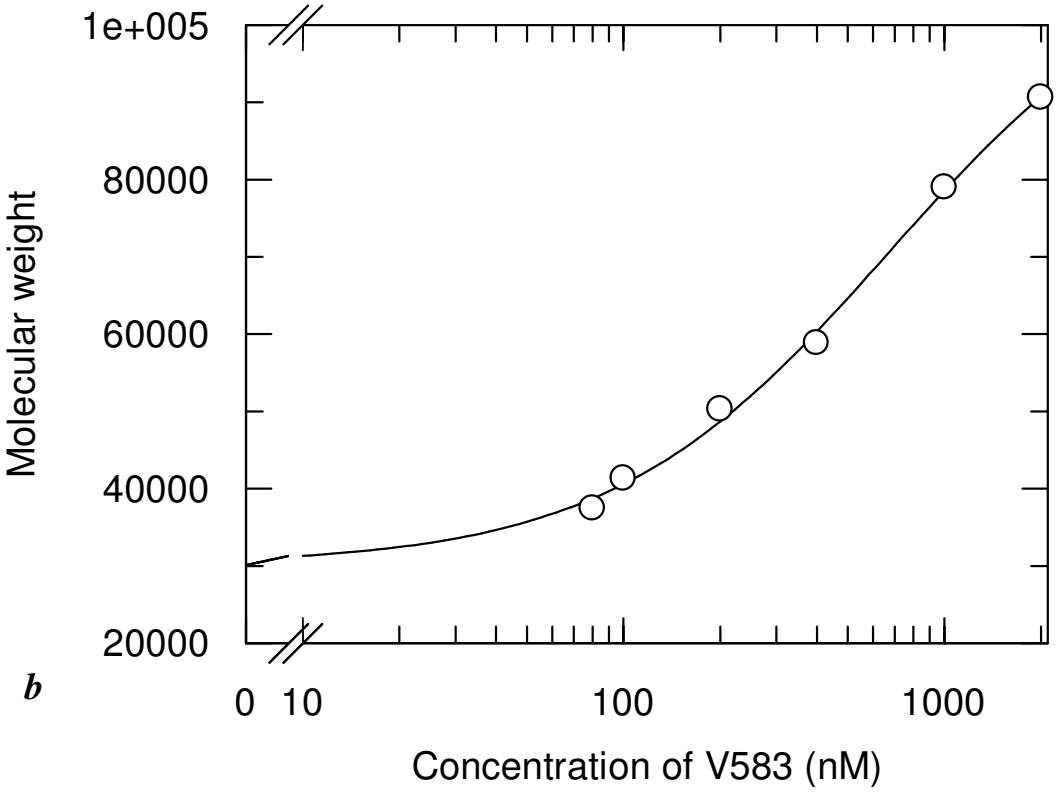


Figure 4-10: Size-exclusion chromatography analysis of V583 to investigate the solution molecular mass at different V583 dimer concentrations. The column was equilibrated in buffer A. Experimental conditions were the same as described previously (see Section 4.3.1). Panel *a* is the raw data showing the fluorescence signal of the eluate against elution time (min). The arrow indicates the effect of decreasing the concentration of V583. The profiles that intersect the arrow from left to right correspond to a V683 concentration of 2, 1, 0.4, 0.2, 0.1, 0.08 µM, respectively. Panel *b* apparent molecular weight of V583 at different concentrations as determined from the elution time on the calibrated gel filtration column. The black curve is a 1:1 binding curve.

As anticipated from the results of the other ArdA proteins, decreasing the concentration of Mu50 caused the molecular weight to decrease (Figure 4-11). From the fitting curve, it can be seen that at very low concentrations of protein the solution molecular weight of Mu50 is about 45,000 which is much bigger than other ArdA proteins. The predicted molecular weight of Mu50 based on its amino acid sequence is 19,390. Even, by analogy to the crystal structure of Orf18, we expect the Mu50

molecule to be highly elongated. However, the molecular weight of Mu50 is still much bigger than other ArdA proteins'. Thus, at low concentration Mu50 exists as a mixture of monomer/dimer in solution (with an apparent molecular weight of 45,000). At very high concentration (i.e. 10 μM) the apparent solution molecular weight of Mu50 is over 120,000, which is more than three times its molecular weight at very low concentration. This result indicates that most of the Mu50 is present as a dimer or larger aggregate at a concentration of ~10 μM. At 1.0 μM the apparent solution molecular weight of Mu50 is about 90,000.

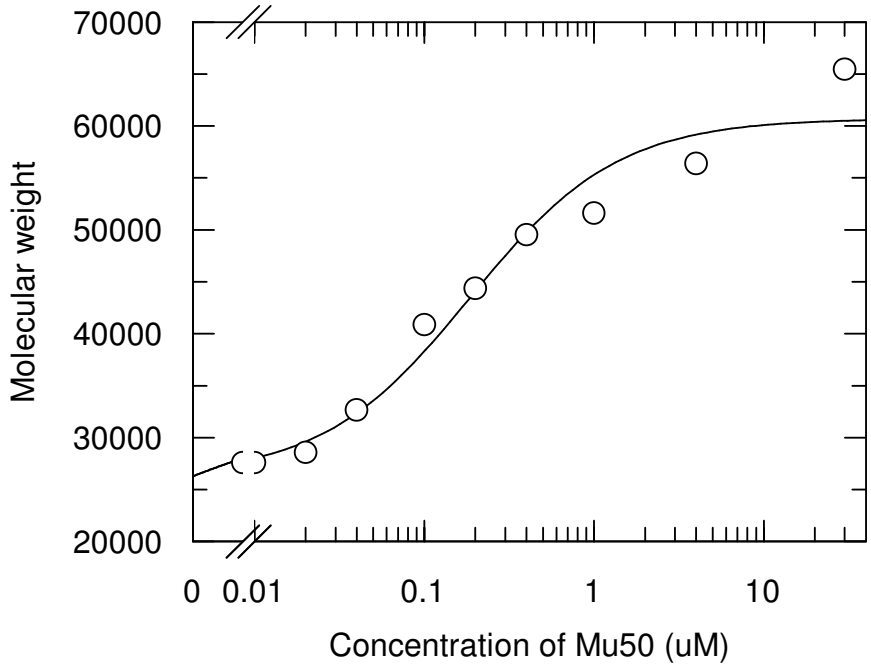
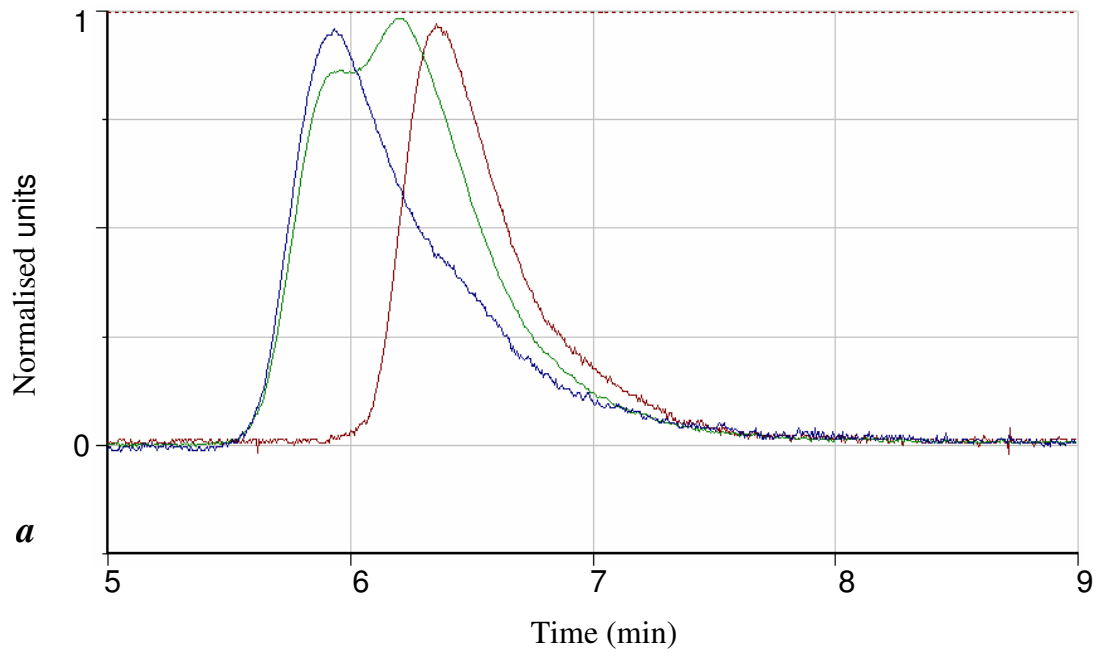


Figure 4-11: Size-exclusion chromatography analysis of Mu50 to investigate the solution molecular mass at different Mu50 dimer concentrations. The molecular weight of Mu50 at different concentrations as determined from the elution time on the calibrated gel filtration column. The black curve is a 1:1 binding curve.

4.3.4.2 ArdA and MTase complexes

Figure 4-12 shows the size-exclusion HPLC results of Orf18, V583 and Mu50 bound to MTase. Complexes of 2 μM MTase and 2 μM ArdA dimer were prepared and applied to the column. Figure 4-12 ***a*** shows two peaks in the V583 and MTase sample. The elution time of the earlier peak (i.e., shoulder of the green trace in figure

4-12 ***a***) is similar to that of MTase alone. However, the elution time of the main peak is less than that of V583 alone. This result indicates that V583 and MTase form an unstable complex, which then subsequently dissociates on the column during chromatography. The main peak corresponds to V583 that has dissociated from the complex during the chromatography which explains why the elution time is less than that of V583 when run alone (red trace in Figure 4-12 ***a***). Similarly, one might anticipate that the peak corresponding to the MTase which has dissociated from the complex would elute fractionally earlier than that of the MTase alone. However, the difference in the elution time between the dissociated MTase peak and the MTase alone peak is not obvious. This observation is probably due to the fact that the molecular species are not completely resolved. Consequently, the elution profile of the dissociated MTase gives a shoulder, rather than a discrete peak, which makes it difficult to define an accurate elution time.



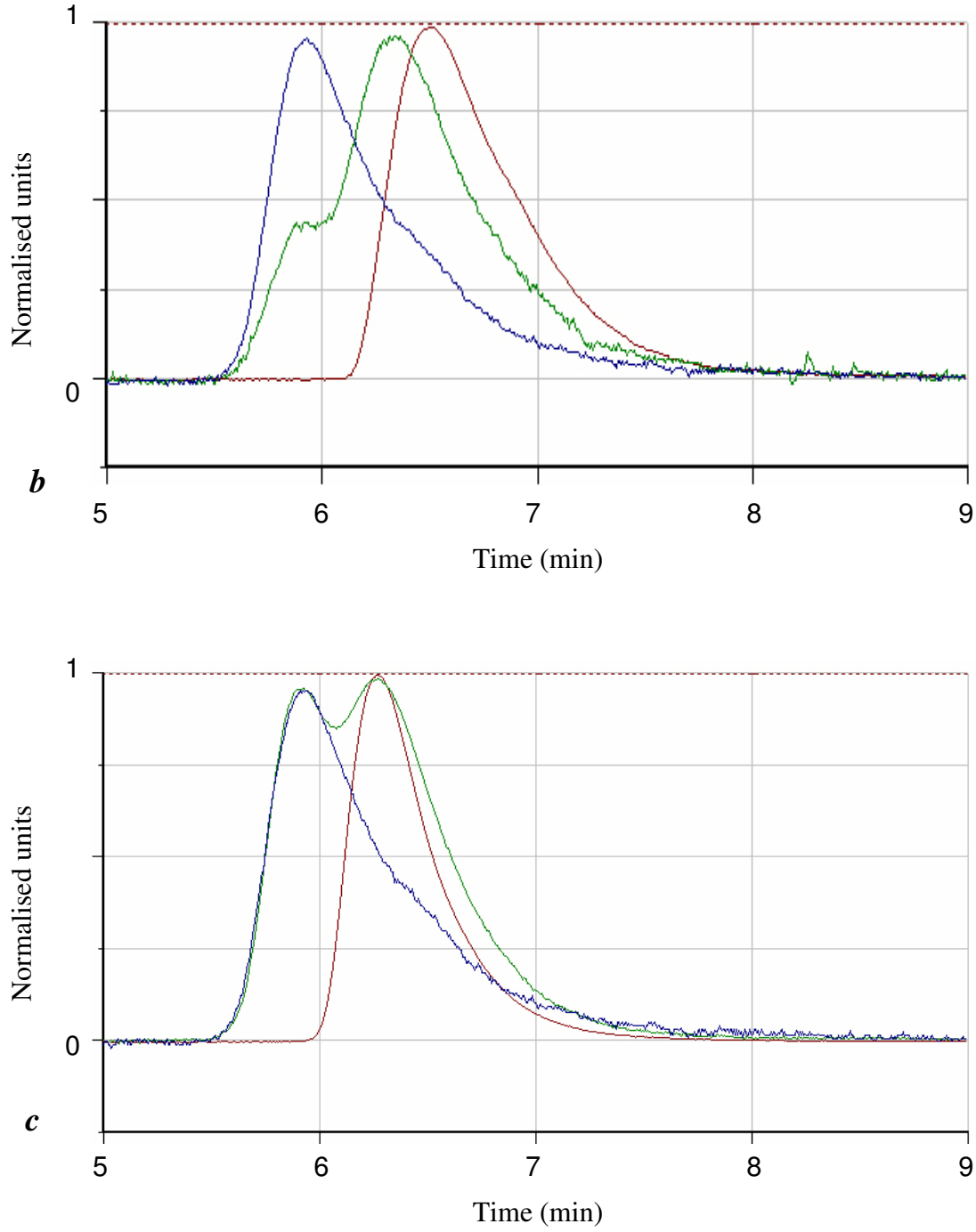


Figure 4-12: Size-exclusion chromatography analysis of the interactions between ArdA proteins (Mu50, Orf18 and V583) and MTase to investigate the solution molecular mass at different protein concentrations. No complex between ArdA protein and MTase was observed. Green trace is mixture of 2 µM dimer ArdA protein and 2 µM MTase, blue trace is 2 µM MTase, red trace is 2 µM dimer ArdA protein. Panel *a* is V583 and MTase, panel *b* is Orf18 and

MTase, Panel *c* is Mu50 and MTase. Salt (200 mM NaCl) was present in the buffer used for size-exclusion chromatography, which may decrease the binding affinity of the proteins.

Figure 4-12 ***b*** shows the gel filtration experiment of Orf18 in complex with MTase. As with V583 and MTase, no complex could be detected. The elution time of the earlier peak (i.e., shoulder of the green trace in Figure 4-12 ***b***) is similar to that of MTase alone and the elution time of the main peak is less than that of Orf18 alone. This result indicates that Orf18 and MTase form an unstable complex dissociation during chromatography. Thus, as with the V583:MTase complex, the complex of Orf18 and MTase appears to be unstable under these experimental conditions.

No complex of Mu50 and MTase was observed in Figure 4-12 ***c***. The two peaks are very similar to MTase or Mu50 alone. These indicate that the interaction between Mu50 and MTase is very weak under these experimental conditions.

4.3.5 Summary of ArdA size-exclusion HPLC results

Gel filtration analysis shows that ArdA proteins can exist as a monomer or dimer and maybe a tetramer in solution depending on the concentration [Serfiotis-Mitsa *et al*., 2008, McMahon *et al*., 2009]. At concentrations of ArdA proteins less than 100 nM, most of the protein is present as a monomer. At about 2 μM concentration the ArdA protein exists as a dimer or tetramer. At intermediate concentrations the protein is present as a mixture of monomer and dimer. All four ArdA proteins showed this behavior.

No complex of MTase and Orf18, V583 and Mu50 was observed in contrast to the mixture of NCTC with MTase. This shows that the interactions of Orf18, V583 and Mu50 and MTase were weaker than the complex of NCTC with MTase. The complex of NCTC with MTase was stable despite the presence of 200 mM NaCl during the chromatographic procedure.

4.4 Displacement assay to determine relative binding affinities

ITC and size-exclusion HPLC cannot be used to obtain the binding affinity of ArdA proteins and MTase/GFP-MTase. Analysis of the ITC results showed the stoichiometry of interaction between ArdA proteins and MTase/GFP-MTase varied. Therefore, an alternative methodology, termed a displacement assay, was used to establish the relative binding affinity of the ArdA proteins with MTase.

The displacement assay uses CM Sepharose beads (Pharmacia, Buckinghamshire, UK) to initially bind GFP-MTase by electrostatic forces in a Tris-HCl buffered solution at pH 6.5 or 8.0. After removing the beads by centrifugation through a filter tube (Pall Portsmouth UK) no fluorescence was observed in the remaining solution as all the GFP-MTase was stuck to the beads. ArdA was then added to the GFP-MTase beads. ArdA competes with the charged surface of the beads for binding to GFP-MTase. Centrifugation of the suspension through a filter tube is performed for a second time. The amount of GFP-MTase:ArdA complex in solution is then determined by quantifying the fluorescence signal in solution. The more ArdA protein is added to the GFP-MTase bound beads, the more GFP-MTase is displaced into solution. In order to analyse the data it is assumed the stoichiometry of interaction between the each ArdA dimer protein and GFP-MTase is one GFP-MTase molecule to one ArdA dimer. The data were then fitted to Equation 4-1 (derived in Appendix B).

Equation 4-1:

$$F = (F_f - F_0)\frac{[P2]_t + [P1]_t + K_d \pm \sqrt{([P2]_t + [P1]_t + K_d)^2 - 4[P1 \cdot P2]_t[P2]_t}}{2[P1 \cdot P2]_t}$$

Where F is the observed fluorescence, F_0 is the fluorescence background (before adding any ArdA), F_f is the fluorescence at the end of titration, P1 is ArdA, P2 is GFP-MTase, P1·P2 is the complex of ArdA and GFP-MTase, $[P1]_t$ is the concentration of total ArdA, $[P2]_t$ is the concentration of total GFP-MTase, $[P1{\cdot}P2]_t$ is the concentration of total complex, Kd is dissociation constant.

Using this method we can determine the relative binding affinity of the interaction between the ArdA proteins and MTase. However, this is ignoring dissociation of the ArdA dimer.

4.4.1 Experimental details

4.4.1.1 Preconditioning of the beads

The CM sepharose beads are supplied in 20% ethanol to act as a preservative. Therefore, it was necessary to thoroughly wash the beads in the desired buffer solution before starting the experiment. The thick slurry of beads was gently inverted for several minutes and 800 µl of suspension was transferred into a 1.5 ml tube. The beads were then collected by centrifugation at 1000 *g* for 1 min and the upper solution was gently removed and discarded. A 800 µl aliquot of buffer B (20 mM Tris-HCl, 10 mM $MgCl_2$, 0.1 mM EDTA, 7 mM 2 ME, pH 8.0) or buffer C (20 mM Tris-HCl, 20 mM MES, 0.1 mM EDTA, 10 mM $MgCl_2$, 7 mM 2 ME, pH 6.5) was then used to resuspend the pellet of beads. After mixing, the suspension was centrifuged as described previously. In total, the washing procedure was repeated a further five times. Finally, about 500 µl of washed bead suspension was obtained.

4.4.1.2 Performing the dissociation assay

A 200 µl aliquot of buffer B or C was added to 500 µl of wet beads. Then 150 µl of 100 nM GFP-MTase and 50 µl of the bead suspension was added to a filter tube to give 75 µM MTase in total. The resulting suspension was gently agitated on a rotary mixer at 20 rpm, 4 °C for 30 min. The sample was then centrifuged in the filter tube at 4000 *g*, 4 °C for 2 min to pellet the beads. The buffer solution that passed through the filter device by centrifugation was discarded. A 150 µl aliquot of different ArdA solutions of known concentration was added to the tubes containing the beads bound to GFP-MTase. The resulting suspension was then gently agitated on a rotary mixer at 20 rpm, 4 °C for 30 min. After pelleting the beads by centrifugation (4000 *g*, 4 °C,

2 mins) the solution was sampled and the fluorescence measured (excitation 395 nm, emission 510 nm).

Some of the control results are shown in Figure 4-11. Firstly, there was no difference in fluorescence signal between buffer and the same batch of buffer that had been passed through the filter tube. Secondly, approximately 20 % of GFP-MTase was found to adhere to the filter of the tube device. Thirdly, after 30 mins mixing of 100 nM GFP-MTase and beads at 4 °C, only 5% of unbound GFP-MTase was found to pass through the tube. These results show the filter tubes can used to separate the GFP-MTase bound with beads (solid) and unbound GFP-MTase (in solution).

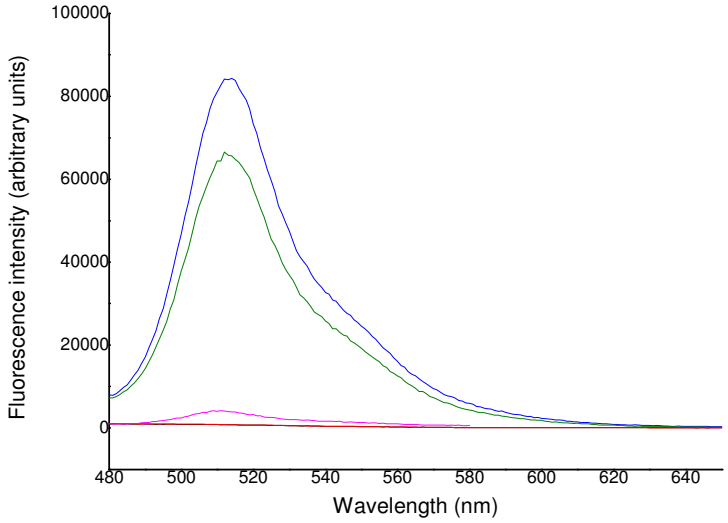


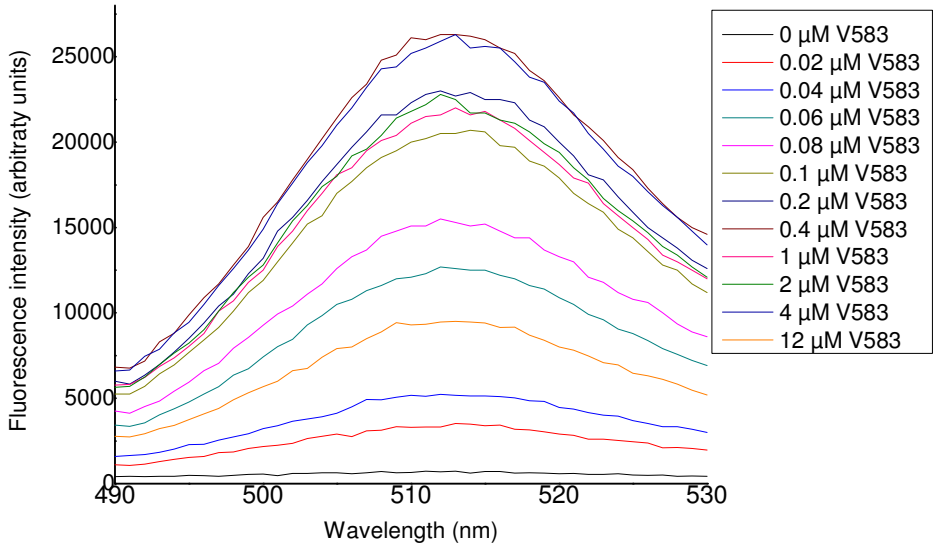
Figure 4-11: Spectrophotometric analysis of the displacement assay. Black line shows the fluorescence spectrum of buffer B; red line is the fluorescence spectrum of buffer B after single passage through the filter tube; blue line is the fluorescence spectrum of 100 nM GFP-MTase; green line is 100 nM GFP-MTase after single passage through the filter tube; pink line is the buffer after 100 nM GFP-MTase was bound to the beads. Excitation wavelength was 395 nm. All spectra were acquired at 20 °C.

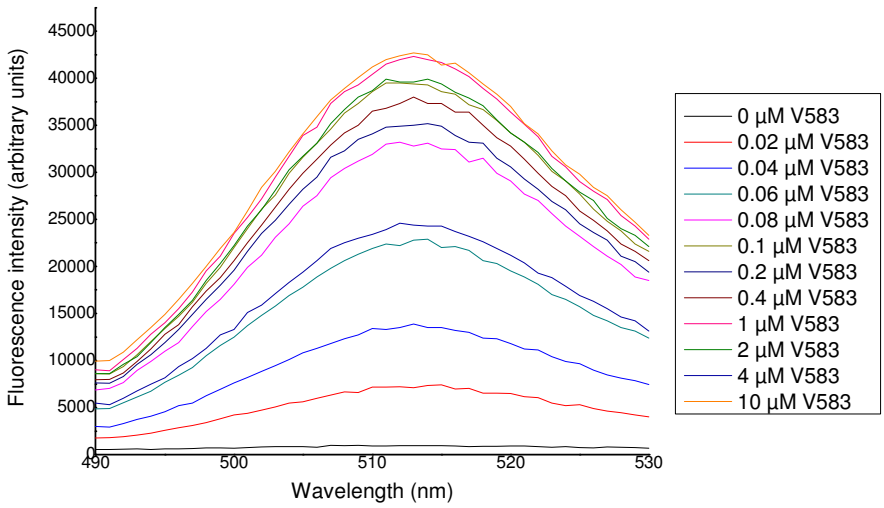
4.4.2 Displacement experiment to determine Kd

4.4.2.1 V583 binding to GFP-MTase

A 150 µl aliquot of GFP-MTase solution (0.1 µM) and 50 µl of bead suspension were added to each tube. The suspension in each tube was mixed as described earlier and then the tubes were centrifuged at 4000 *g*, 4 °C for 2 min to recover the beads. A series of different concentrations of V583 were used to resuspend the beads. After gently mixing the suspension for 20 min as described earlier, samples were centrifuged at 4000 *g*, 4 °C for 2 min. The fluorescence signal in the resulting solution is shown in Figure 4-12. Figure 4-12 shows the effect of using different concentrations of V583. Increasing concentrations of V583 resulted in a greater fluorescence signal in solution after the centrifugation step to remove the beads in both buffer B and C. Each experiment was repeated three times.

a





b

Figure 4-12: A series of different concentrations of V583 were added to GFP-MTase bound beads. Note all V583 concentrations refer to that of dimer rather than monomer. Panel *a*, experiment was conducted in buffer C; panel *b*, experiment was conducted in buffer B. Excitation wavelength was 395 nm.

The fluorescence intensity (ranging from 490-530nm) for each concentration of V583 shown in Figure 4-12 was integrated. The values obtained at each concentration were averaged, scaled to a maximum change and then plotted against V583 dimer concentration (Figure 4-13). It is clear that very similar results were obtained at the two different pH values. The results were fitted to Equation 4-1 twice. First, the results of proteins in buffer B (pH 8.0) were fitted to Equation 4-1 with F_0 and F_f variable. The Kd was 1.2±2.2 nM, F_f was 0.91±0.04 and F_0 was -0.039±0.060. When F_0 and F_f were fixed at 0 and 1, respectively, the Kd was 8.3±4.5 nM. The fitting with more variable parameters gives better fitting. Since this fitting result is not too much different, the simpler one is fine. Second, the results of proteins in buffer C (pH 6.5) were fitted to Equation 4-1 with F_0 and F_f variable. The Kd was 4.8±2.2 nM, F_f was 0.99±0.02 and F_0 was -0.026±0.061. When F_0 and F_f were fixed at 0 and 1, respectively, the Kd was 6.0±1.7 nM. Thus, V583 binds GFP-MTase slightly better at the lower pH value but the difference is not significant compared to the error.

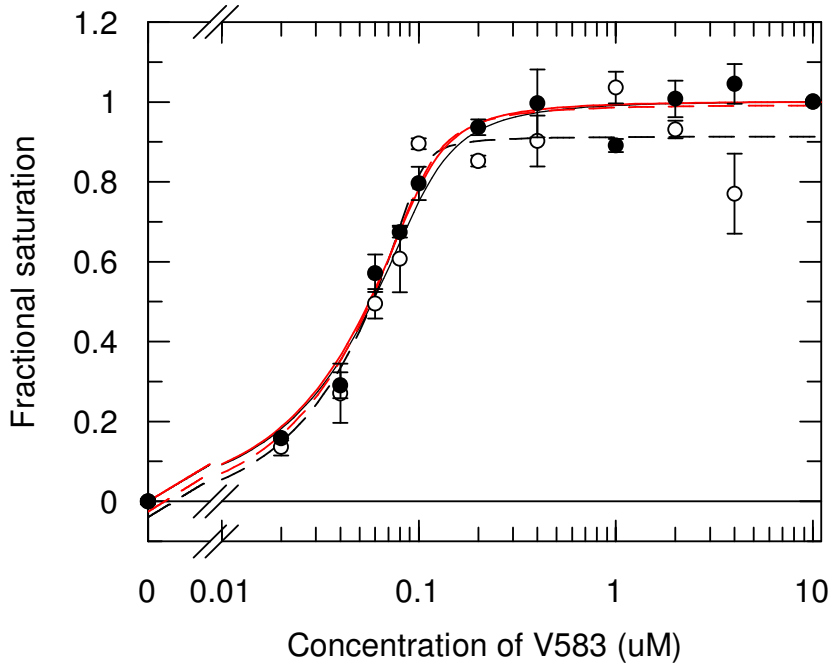


Figure 4-13: V583 titration of beads bound with GFP-MTase. Black circles: proteins were prepared in buffer C (pH 6.5). White circles: proteins were prepared in buffer B (pH 8.0). The data was fitted using Equation 4-1. Black curves are the fitting curves of proteins in buffer B; red curves are the fitting curves of proteins in buffer C. The dash curves were fitted when F_0 and F_f were variable; the continuous curves were fitted when F_0 and F_f were fixed at 0 and 1, respectively.

4.4.2.2 Orf18 binding to GFP-MTase

Figure 4.14 shows the result of the Orf18 titration of GFP-MTase beads in buffer B and C. The experimental details were the same as those described in Section 4.4.1.2.

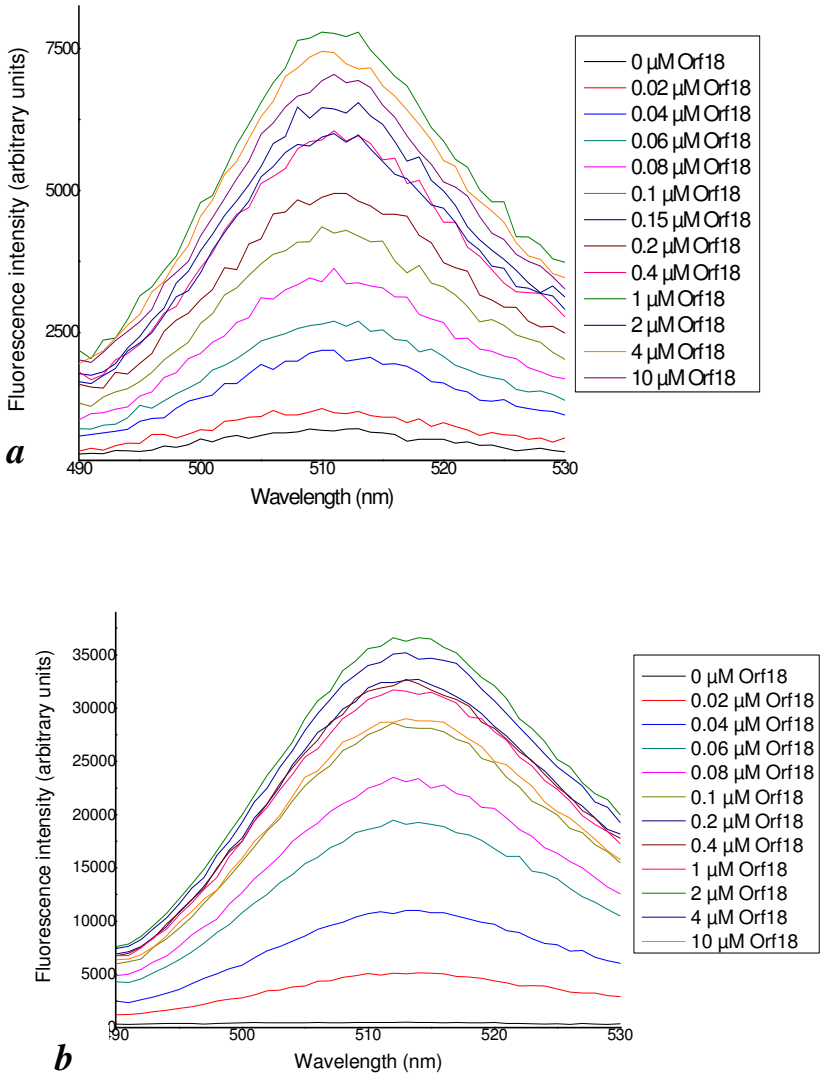


Figure 4-14: A series of different dimer concentrations of Orf18 bound to GFP-MTase. Panel *a*; experiment was conducted in buffer C. Panel *b*; experiment was conducted in buffer B. Excitation wavelength was 395 nm.

The fluorescence intensity results (ranging from 490-530nm) for each concentration of Orf18 displayed in Figure 4-14 were integrated. The resulting values obtained at each concentration were averaged and scaled. The results were then plotted against Orf18 dimer concentration (Figure 4-15). The graph in figure 4-15 demonstrates the effect of binding at two different pH values. The complex in buffer B (pH 8.0)

produces a sharper transition compared to the same complex in buffer C (pH 6.5). An increase in the gradient of the slope indicates greater binding affinity of the two proteins. Therefore, this result suggests that the interaction in buffer B is stronger than in buffer C. The results were fitted to Equation 4-1. First, the results of proteins in buffer B (pH 8.0) were fitted Equation 4-1 with F_0 and F_f variable. The Kd was 1.4±1.3 nM, F_f was 1.01±0.02 and F_0 was -0.091±0.038. When F_0 and F_f were fixed at 0 and 1, respectively, the Kd was 1.2±1.0 nM. Second, the results of proteins in buffer C (pH 6.5) were fitted Equation 4-1 with F_0 and F_f variable. The Kd was 79±20 nM, F_f was 1.00±0.03 and F_0 was 0.031±0.038. When F_0 and F_f were fixed at 0 and 1, respectively, the Kd was 79±10 nM. The simpler model gives same Kd as more complex one, so it is preferable. Therefore, Orf18 binds more tightly to GFP-MTase at the higher pH value.

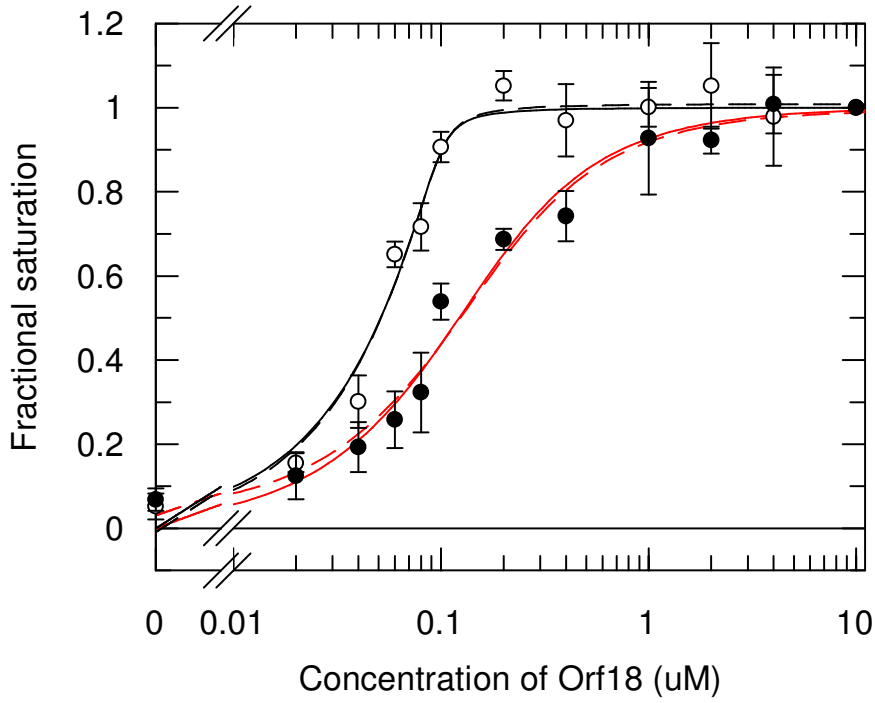


Figure 4-15: Orf18 titration of beads bound with GFP-MTase. Black circles: proteins were prepared in buffer C (pH 6.5). White circles: proteins were prepared in buffer B (pH 8.0). The data was fitted using Equation 4-1. Black curves are the fitting curves of proteins in buffer B; red curves are the fitting curves of proteins in buffer C. The dash curves were fitted when F_0 and F_f were

variable; the continuous curves were fitted when F_0 and F_f were fixed at 0 and 1, respectively.

4.4.2.3 Mu50 binding with GFP-MTase

Figure 4-16 shows the result of the Mu50 titration of GFP-MTase beads in buffer B and C. The experimental details were the same as those described in Section 4.4.1.2.

The fluorescence intensity results (ranging from 490-530nm) for each concentration of Mu50 displayed in figure 4-16 were integrated. The resulting values obtained at each concentration were averaged and scaled. The results were then plotted against Mu50 dimer concentration (Figure 4-17). The graph in figure 4-17 demonstrates the effect of binding at two different pH values. The complex in buffer B (pH 8.0) produces a sharper transition compared to the same complex in buffer C (pH 6.5). An increase in the gradient of the slope indicates greater binding affinity of the two proteins. Therefore, this result suggests that the interaction in buffer B is stronger than in buffer C. Indeed, the results were fitted to Equation 4-1 twice. First, the results of proteins in buffer B (pH 8.0) were fitted to Equation 4-1 with F_0 and F_f variable. The Kd was 11±4 nM, F_f was 1.01±0.03 and F_0 was -0.070±0.035. When F_0 and F_f were fixed at 0 and 1, respectively, the Kd was 16±5 nM. Second, the results of proteins in buffer C (pH 6.5) were fitted to Equation 4-1 with F_0 and F_f variable. The Kd was 2.1±0.9 nM, F_f was 1.03±0.01 and F_0 was -0.082±0.023. When F_0 and F_f were fixed at 0 and 1, respectively, the Kd was 1.2±1.0 nM. Therefore, the binding affinity of Mu50 binding to GFP-MTase is similar at these pH values and the simpler model is preferred.

Figure 4-18 shows the result of the NCTC titration of GFP-MTase beads in buffer B and C. The experimental details were the same as those described in Section 4.4.1.2.

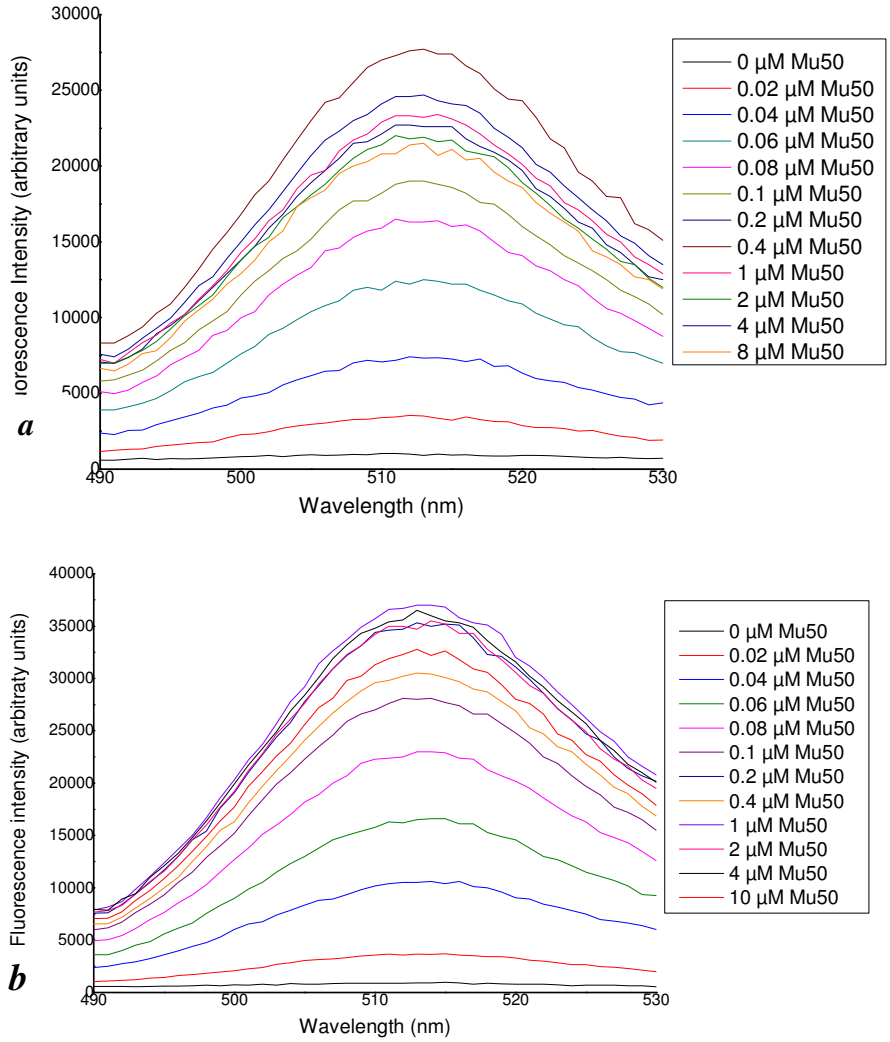


Figure 4-16: A series of different concentrations of Mu50 were mixed with GFP-MTase bound to the beads. Panel *a*, experiment conducted in buffer C. Panel *b*, experiment conducted in buffer B. Excitation wavelength was 395 nm.

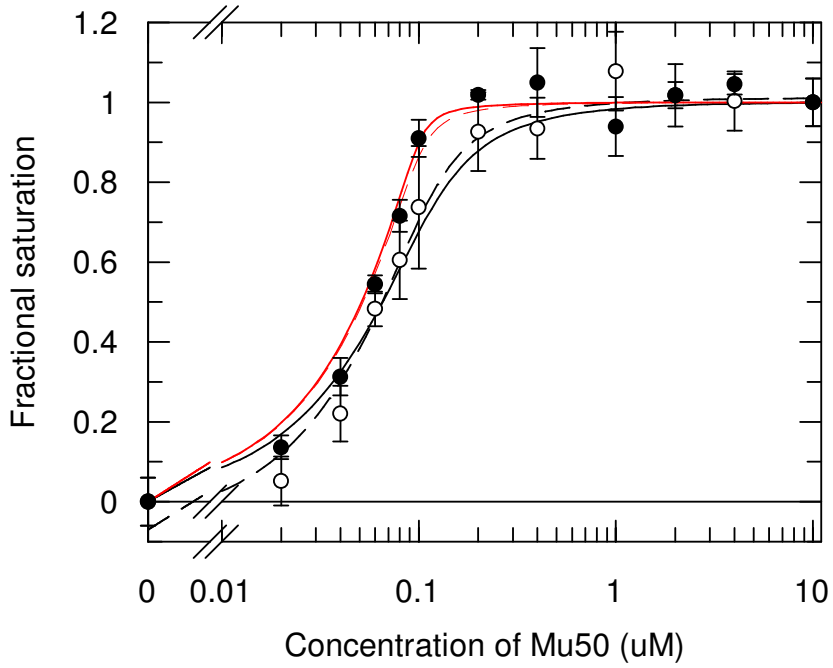
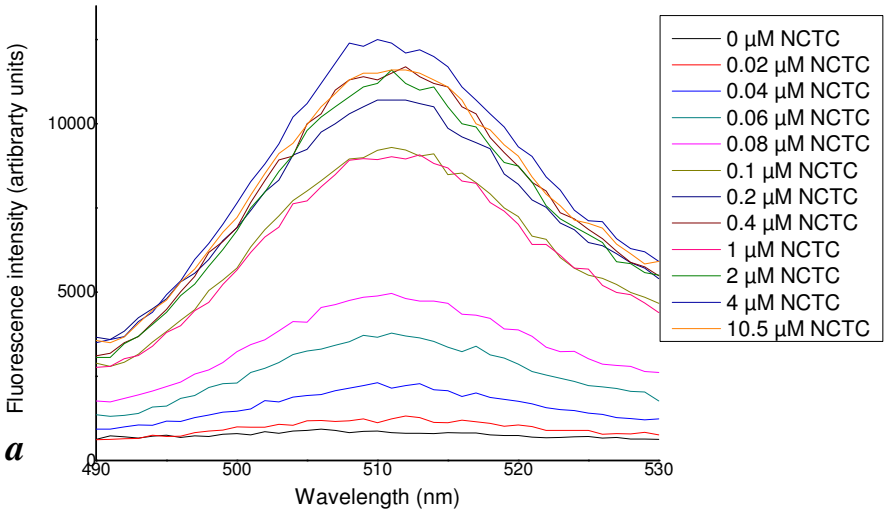


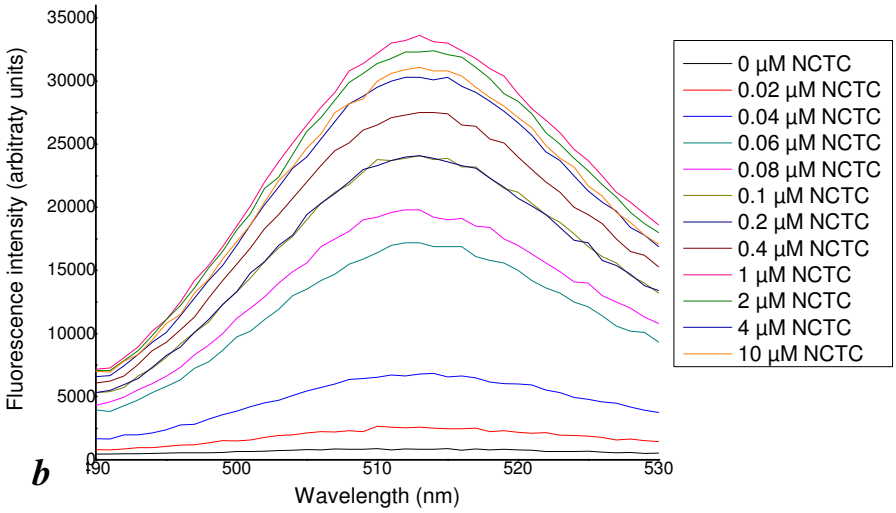
Figure 4-17: Mu50 titration of beads bound with GFP-MTase. Black circles: proteins were prepared in buffer C (pH 6.5). White circles: proteins were prepared in buffer B (pH 8.0). The data was fitted using Equation 4-1. Black curves are the fitting curves of proteins in buffer B; red curves are the fitting curves of proteins in buffer C. The dash curves were fitted when F_0 and F_f were variable; the continuous curves were fitted when F_0 and F_f were fixed at 0 and 1, respectively.

The fluorescence intensity results (ranging from 490-530nm) for each concentration of NCTC displayed in Figure 4-18 were integrated. The resulting values obtained at each concentration were averaged and scaled. The results were then plotted against NCTC dimer concentration (Figure 4-19). The graph in Figure 4-19 demonstrates the effect of binding at two different pH values. The complex in buffer B (pH 8.0) produces a sharper transition compared to the same complex in buffer C (pH 6.5). An increase in the gradient of the slope indicates greater binding affinity of the two proteins. Therefore, this result suggests that the interaction in buffer B is stronger

than in buffer C. The results were fitted to Equation 4-1. First, the results of proteins in buffer B (pH 8.0) were fitted to Equation 4-1 with F_0 and F_f variable. The Kd was 3.7±3.5 nM, F_f was 0.94±0.04 and F_0 was -0.037±0.055. When F_0 and F_f were fixed at 0 and 1, respectively, the Kd was 9.7±4.1 nM. Second, the results of proteins in buffer C (pH 6.5) were fitted to Equation 4-1 with F_0 and F_f variable. The Kd was 6.7±2.8 nM, F_f was 0.98±0.03 and F_0 was 0.060±0.050. When F_0 and F_f were fixed at 0 and 1, respectively, the Kd was 7.3±4.2 nM. The binding affinity of NCTC binding to GFP-MTase is very similar at these pH values.



a



b

Figure 4-18: A series of different concentrations of NCTC were mixed with GFP-MTase bound to the beads. Panel *a*, experiment conducted in buffer C. Panel *b*, experiment conducted in buffer B. Excitation wavelength was 395 nm.

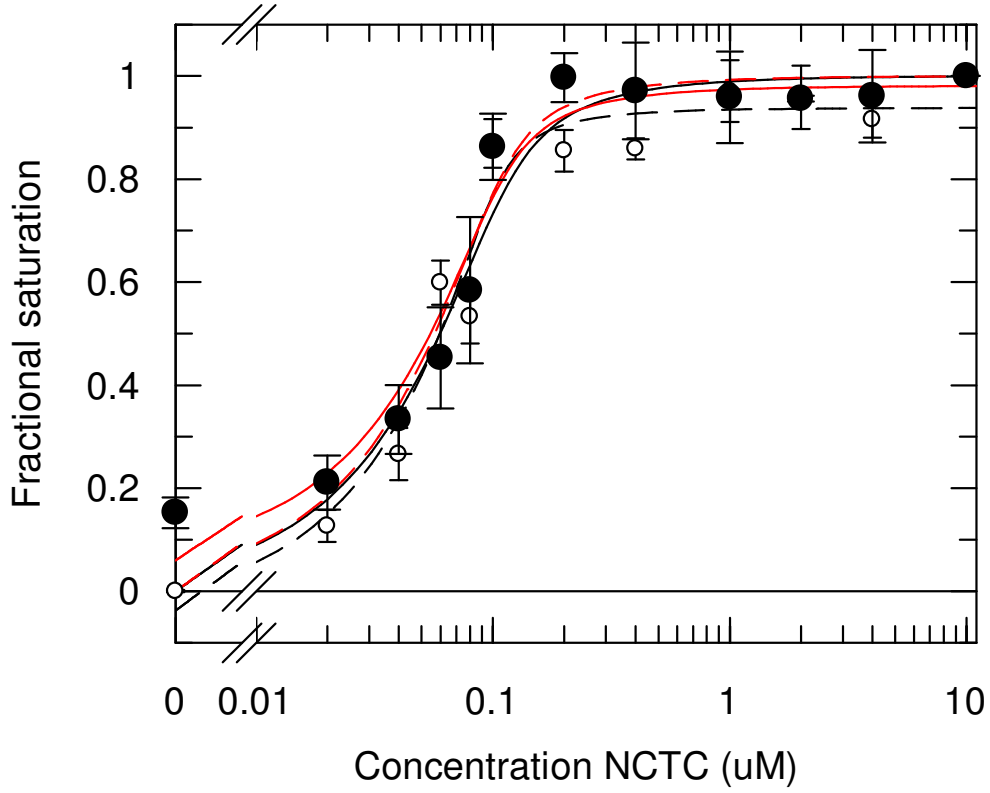


Figure 4-19: NCTC titration of beads bound with GFP-MTase. Black circles: proteins were prepared in buffer C (pH 6.5). White circles: proteins were prepared in buffer B (pH 8.0). The data was fitted using Equation 4-1. Black curves are the fitting curves of proteins in buffer B; red curves are the fitting curves of proteins in buffer C. The dash curves were fitted when F_0 and F_f were variable; the continuous curves were fitted when F_0 and F_f were fixed at 0 and 1, respectively.

4.4.3 Summary of ArdA proteins binding to GFP-MTase

From Table 4-5 and 4-6 we can see the binding between Mu50, V583 and NCTC and GFP-MTase no change in Kd with pH. Orf18 shows a strong change with pH and the

binds best at high pH. The sample model gives roughly same Kd as more complex model. The F_0 and F_f values are pretty close to 0 and 1 so the sample model is fine.

Table 4-5: Kd (nM) of interaction between ArdA proteins and GFP-MTase the F_0 and F_f were fixed at pH 6.5 and 8.0.

	V583	Mu50	Orf18	NCTC
Buffer B (pH 8.0)	5.6±1.7	2.1±0.9	1.2±1.0	9.7±4.1
Buffer C (pH 6.5)	8.3±4.5	1.2±1.0	79±10	7.3±4.2

Table 4-6: Kd (nM) of interaction between ArdA proteins and GFP-MTase the F_0 and F_f were not fixed at pH 6.5 and 8.0.

	V583	Mu50	Orf18	NCTC
Buffer B (pH 8.0)	1.2±2.2	11±4	1.4±1.3	3.7±3.5
Buffer C (pH 6.5)	4.8±2.2	2.1±0.9	79±20	6.7±2.8

Table 4-7: F_0 and F_f values of table 4-6 at pH 6.5 and 8.0.

	V583		Mu50		Orf18		NCTC	
	F_0	F_f	F_0	F_f	F_0	F_f	F_0	F_f
Buffer B (pH 8.0)	-0.039 ±0.060	0.91 ±0.04	-0.070 ±0.035	1.01 ±0.03	-0.091 ±0.038	1.01 ±0.02	-0.037 ±0.055	0.94 ±0.04
Buffer C (pH 6.5)	-0.026 ±0.061	0.99 ±0.02	-0.082 ±0.023	1.03 ±0.01	0.031 ±0.038	1.00 ±0.03	0.060 ±0.050	0.98 ±0.03

4.5 Discussion

Analysis of the interaction between the ArdA proteins and MTase was studied by size-exclusion chromatography. The running buffer used for this experiment included a high concentration of salt (i.e. 200 mM NaCl) at pH 6.5. The presence of salt in the buffer is required to prevent non-specific interaction between the protein and the inert column matrix.

The results from size-exclusion chromatography showed that, in most cases, the ArdA:MTase complex was too unstable to detect in the eluate. The complex between NCTC and MTase is the only one that was stable enough to pass through the column intact. Indeed, NCTC stabilises the trimeric structure of MTase so that the complex did not dissociate during the column run even at very low protein concentrations.
However, it was clear from the results of the size-exclusion chromatography that it is not possible to determine the affinity of binding between the ArdA proteins and MTase using this approach. Therefore an alternative method was developed that relied on the displacement of MTase bound to a negatively charged medium (CM sepharose) upon complexation with ArdA. It was reasoned that these experiments could be conducted in the absence of salt, which should increase the stability of the protein complex. However, in order to accurately determine the amount of displaced MTase it was necessary to use MTase fused to the fluorescent protein GFP (i.e., GFP-MTase). According to this strategy, very small amounts of protein in solution can be readily quantified from the fluorescence spectrum. Using this so called displacement experiment the interaction between the various ArdA proteins and MTase was studied.

Displacement experiments were conducted in buffer B (pH 6.5, no salt), this resulted in similar Kd values expressed for the interaction of ArdA proteins; NCTC, V583 and Mu50 to GFP-MTase. What is surprising is that in size-exclusion HPLC, only NCTC gave a stable complex with MTase as the displacement bead assay suggests that V583 and Mu50 should also give a stable complex. It must, however, be remembered that the two experiments were conducted under different buffer conditions. The two different conditions between those experiments were:

(i) The buffer used in the size-exclusion HPLC experiments included 200 mM NaCl, which prevents unwanted interaction between the protein and gel matrix. However, the buffer used for in the displacement experiment does not include any salt.

(ii) The binding partner in the two experiments is different. In the size-exclusion HPLC experiments WT MTase was used to bind with ArdA. By contrast, the

displacement experiment used GFP-MTase to facilitate detection of the displaced protein complex.

Hence, there are two possible reasons for the apparent discrepancy between the data obtained from size-exclusion HPLC and the displacement experiment. From the structure of GFP-MTase shown in Figure 3-12 it is clear that the GFP fused to the S subunit of MTase is distant from the proposed ArdA binding site. Thus, it appears unlikely that the presence of fused GFP will have a significant impact on the binding affinity of MTase with the ArdA proteins. However, it is more likely that the presence of salt significantly decreases the binding affinity of MTase for ArdA. Thus, the structure of NCTC around its binding site with MTase presumably differs from that of the other ArdA proteins. These differences may mean a high concentration of salt has a different effect on the interaction of NCTC with MTase by comparison with the other ArdA proteins.

Based on the ITC results (Table 4-3), we know the stoichiometry of interaction between MTase or GFP-MTase with ArdA (i.e., monomer ArdA: MTase or GFP-MTase) is lower than 2. Previous binding studies of MTase and Ocr show that the stoichiometry of interaction is one molecule of MTase per Ocr dimer (Atanasiu *et al*., 2002) and the dimer Orf18 structure (McMahon *et al*., 2009). Therefore, one might anticipate that the MTase:ArdA complex would form with the same ratio. The size-exclusion experiments of the four different ArdA proteins show that the molecules can exist as both dimer and monomer in solution. Specifically, the proteins are present in solution as a monomer at low concentration and dimer at high concentration. Some of the ArdA proteins are even present as larger forms, perhaps tetramer at high concentration. I believe that the oligomerisation of ArdA proteins in solution has an effect on the stoichiometry of interaction between ArdA and MTase.

The size-exclusion chromatography experiments show that at relatively low concentration the ArdA proteins exist as both dimer and monomer in solution. Hence, this means there is a reversible equilibrium of dimer and monomer for the four ArdA proteins. This can be expressed in terms of Equation 4-2.

Equation 4-2: $$A + A \rightleftharpoons A2$$

where A is ArdA monomer and A2 is ArdA dimer. Of course at high concentration of ArdA, there is more dimer ArdA protein than in dilute solution.

Moreover, we know MTase can bind to the dimer ArdA (A2).

Equation 4-3: $$A2 + MTase \rightleftharpoons A2 - MTase$$

However, we do not know if MTase can bind monomer ArdA.

Firstly, consider the situation where the MTase cannot bind the monomer of ArdA. Based on Equation 4-3, if one unit of MTase binds A2 to form A2-MTase, we need one unit of A2. However, we know in solution that some A2 dissociates to A. Therefore, in order to generate one unit of A2-MTase complex, we need more than one unit of A2 in total (NB precisely how much more A2 is required will depend on the equilibrium shown in Equation 4-2). So if MTase cannot bind the monomer of ArdA, the stoichiometry of ArdA:MTase or GFP-MTase should be more than 2 (assuming the ArdA concentration is expressed as less monomer). However, the experimental data from ITC gives n values of greater than 2 for all of the ArdA proteins. These findings suggest this situation is incorrect.
Secondly, assume monomer ArdA can interact with MTase and analyse how this affects the interaction.

Equation 4-4: $$A + MTase \rightleftharpoons A - MTase$$

In Equation 4-4, one unit of MTase forms a complex A-MTase, which only requires one unit of monomer A. Thus, when both ArdA monomer and dimer (Equation 4-3) can form a complex with MTase, the stoichiometry of ArdA:MTase or GFP-MTase should be more than 1 but less than 2. This corresponds with the ITC results are shown earlier (Table 4-3) where n varies between 0.91 and 1.86.

Another possibility is that the A-MTase complex could bind A to form A2-MTase.

Equation 4-5: $$A + A - MTase \rightleftharpoons A2 - MTase$$

This seems unlikely because the bound ArdA will presumably have to shift position within the MTase to allow the second monomer to bind.

Initially, I surmised that the stability of the ArdA dimer might be a key factor in determining the stoichiometry of interaction with MTase. If this is the case, ArdA proteins that form more stable dimers are anticipated to give a greater *n* value. The ITC results (Table 4-3) indicate that NCTC has greatest stoichiometry followed by Mu50, then Orf18 and finally V583 for interaction with MTase. Thus, if this hypothesis is correct NCTC should most readily form dimers followed by Mu50, then V583 and finally Orf18. Based on size-exclusion HPLC results (Figure 4-20), the solution molecular weight of all four ArdA proteins increases with concentration. However, NCTC most readily forms a dimer followed by V583, then Orf18 and finally Mu50. Hence, the size-exclusion HPLC results do not follow the trend predicted from the stoichiometry of interaction with MTase as determined by ITC. The stability of the ArdA dimer is only one factor in determining its interaction. Another important factor is the binding affinity of interaction. When the binding affinity of interaction of dimer and MTase is bigger than that of monomer and MTase, the *n* value is increased.

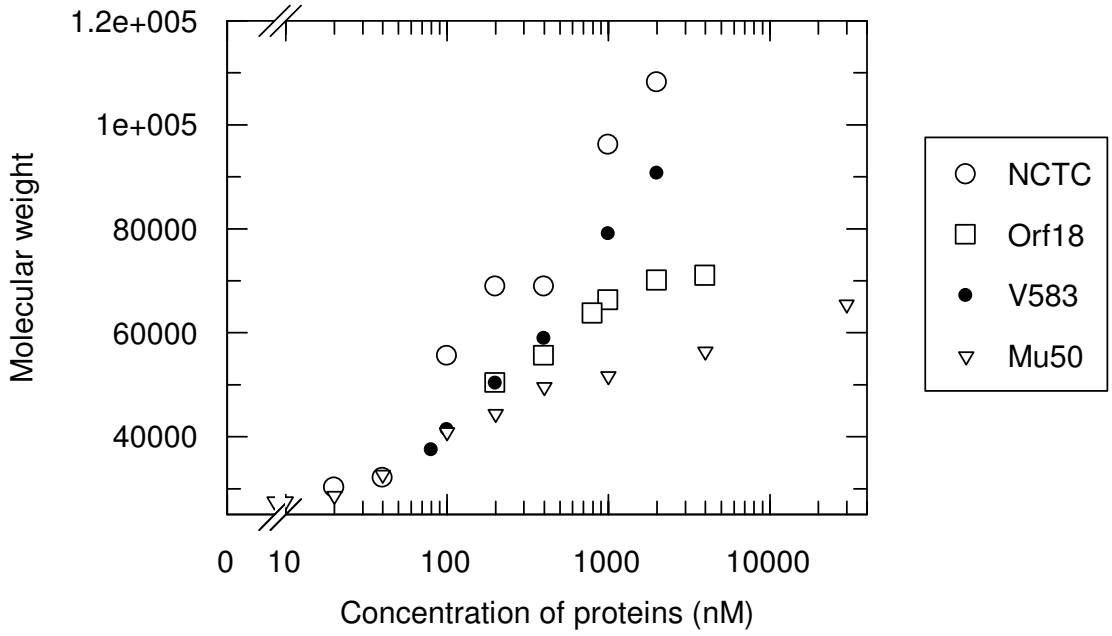


Figure 4-20: Determination of the solution molecular weight of all four ArdA proteins as determined by size-exclusion HPLC.

In this chapter though I tried to use several different methods (ITC, size-exclusion HPLC and displacement assay) to determine the binding affinity of ArdA proteins and MTase/GFP-MTase, unfortunately no accurate Kd was obtained by these methods. However, by these results we can get an idea of how tight the interaction between MTase and ArdA.

One disadvantage is we have to work on different experimental conditions with these methods. For ITC the buffer was at pH 8.0 and no salt. For size-exclusion HPLC experiments, to avoid unwanted interactions between proteins and gel matrix, the buffer must contain 200 mM NaCl. The column also had to use pH 6.5 buffer. In the displacement assay, the experiments in pH 8.0 and pH 6.5 buffer were performed but without NaCl to make the GFP-MTase bind. However, in displacement assay we also have to fuse GFP to MTase to determine. Thus, we had to work on different conditions, this is not perfect situation.

Due to the interaction of ArdA proteins and MTase/GFP-MTase being very tight, I was unable to determine an accurate binding affinity from the ITC data. Specifically, the Kd was in the 0.1-10 nM range for each ArdA. Furthermore, it was not possible to obtain Kd values from the size-exclusion HPLC results. In particular, the association between NCTC and MTase is very tight with a very slow dissociation rate for the complex. Unfortunately, it was not possible to calculate the binding constant using these results alone. For the other ArdA proteins (Mu50, Orf18 and V583), no complex with MTase was observed. Interestingly, however, the size-exclusion results also show that ArdA proteins are present as monomer at low concentrations. As the concentration of ArdA increases the protein is present as a mixture of monomer, dimer and even tetramer. The displacement assay results indicate that the Kd between ArdA proteins and GFP-MTase was in the 1-10 nM range except Orf18 and GFP-MTase.

Chapter 5 Building of a fluorescence detector

Fluorescence spectroscopy is a very valuable technique for studying biochemistry and biophysics. It is extremely sensitive and simple to use. To broaden the results presented in chapters 3 and 4, I would like to use a highly sensitive fluorescence detector to investigate the fluorescence of small sample volumes, as well as studying kinetic reactions/interactions and time-resolved fluorescence. Here, I will introduce the building of these detectors.

5.1 Capillary fluorescence detector

As fluorescence spectroscopy is very sensitive, it makes even single molecule fluorescence imaging possible. This is useful as my biological samples are quite expensive and time consuming to produce. If I could build a detector which could detect the fluorescence of tiny bio-samples, it would be very useful in my research. The first question is where to put the tiny amount of sample. The normal cuvette is too big for this job, so a capillary can be used instead. There are a few advantages in using a capillary:

Firstly, capillaries with varying accurate diameters are easy to obtain enabling the accurate control of sample volume. Secondly, fused silica capillaries covered with polyimide are very durable and thirdly, once the polyimide coating is removed, light can pass through the silica capillary and be detected.

This capillary fluorescence detector could be used with capillary electrophoresis (CE), HPLC and time-resolved fluorescence to investigate the interaction of biomolecules. As the capillary uses such small amounts of sample, the method could be quick and low cost.

5.1.1 Initial design

Figure 5-1 is a typical capillary fluorescence detector design. The laser light (excitation) is focused on the window of the capillary by an excitation lens. The emission light in the vertical direction was focused on the detector by an emission lens. A filter was used to cut off the excitation light between the detector and the lens. As the emission of fluorescent light radiates out from a very small volume, it is better for the emission lens to be close to the window of the capillary to collect more light. Since the emission filter must be placed after the capillary, this makes focussing the light the most difficult aspect.

Initially I built a capillary fluorescence detector with no emission filter. 5.0 μM 21B21TH in 12.5 mM $Na_2B_4O_7$ (pH=9.20) was injected into a capillary with an inside diameter (I.D.) of 50 μm and an outside diameter (O.D.) of 375 μm. Figure 5-2 shows the results from this. As there was no emission filter included, there was a large narrow peak present where scattered excitation light was recorded. However, the emission of 21B21TH was very clear.

The volume of sample can be calculated, assuming the laser light hits the capillary over a length of 0.5 mm, and the I.D. of the capillary is 50 μm

$\pi r^2 l = 3.14*(50*10^{-6}/2)^2*0.5*10^{-3}\ m^3 = 981*10^{-12}\ L$

The moles of sample which were excited was:
$5\ \mu M*981*10^{-12}\ L = 4.9*10^{-15}$ mole

The number of molecules of 21B21TH which was excited was:
$6.02*10^{23}*4.9*10^{-15} = 3.0*10^9$

These results were encouraged us a lot. So one design incorporating an emission filter and one allowing fluid mixing and time-resolved detection were constructed.

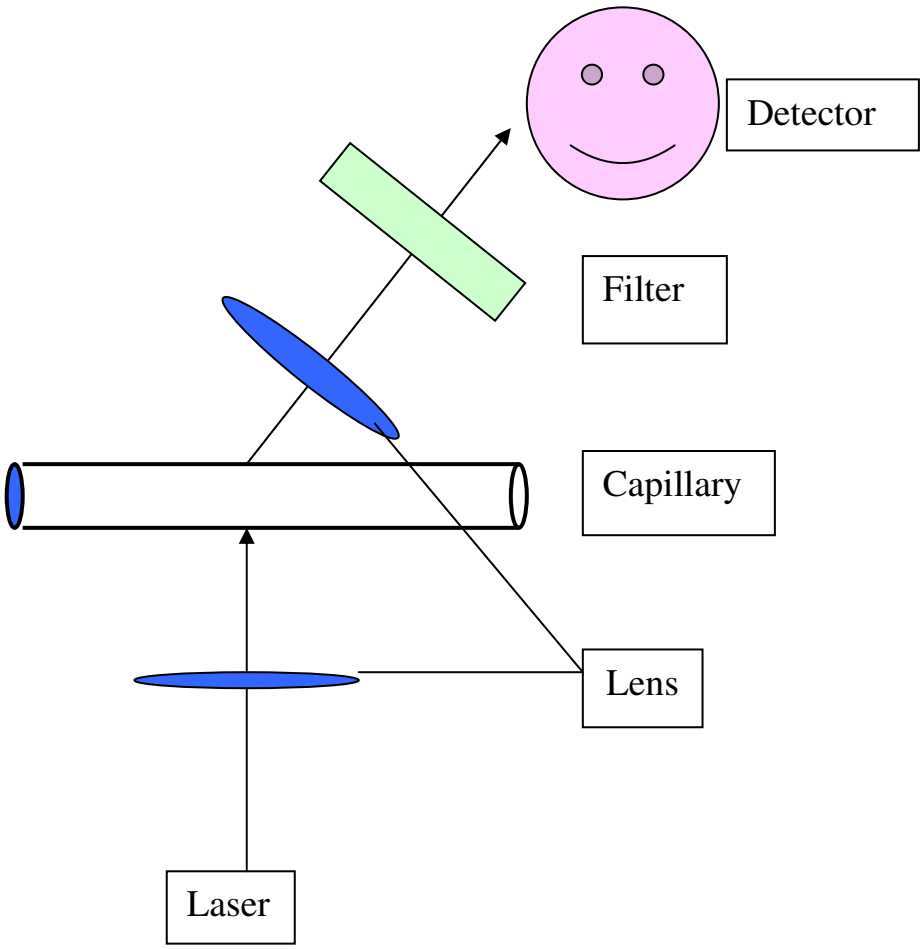


Figure 5-1: Design of the capillary fluorescence detector. Laser light is focused on the window of the capillary by the first lens. Some of the emission light is focused on the filter using a second lens, which cuts out the scattered light. The emission light then enters the optical fiber to the detector.

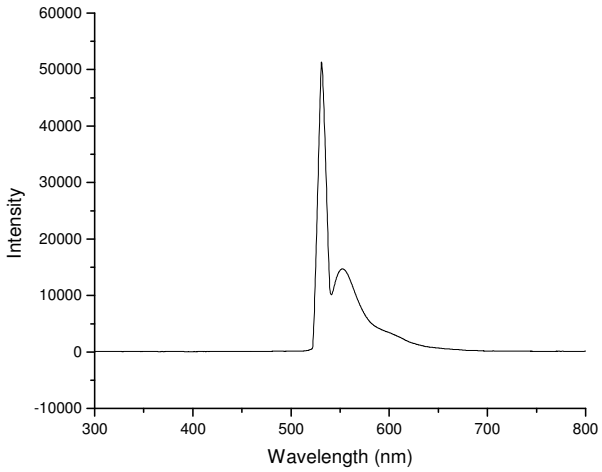


Figure 5-2: 5 µM 21B21TH in 12.5 mM $Na_2B_4O_7$ (pH 9.20) in a capillary with an I.D. of 50 µm and an O.D. of 375 µm. A green laser pen (excitation wavelength ~530 nm) was used to excite the sample. A USB4000-FL spectrometer (Ocean Optics, Dunedin, USA) was used to collect the data.

5.1.2 Incorporation of emission filter

Thus, we changed the original design and used a dichroic mirror and filter set in order to separate fluorescent and excitation photons (Figure 5-3). Light from a laser source was directed at the dichroic mirror (Di01-R532-25*36, Semrock, NY, USA) and focused by a planoconvex (PCX) lens (LA 1951, Thorlabs, Newton, USA) onto the capillary window. Fluorescent light was captured by the same focusing lens and collimated through the dichroic mirror and an emission filter (BrightLine 543/22 Semrock, NY, USA), prior to collection through an objective lens (Din 10 0.25, Edmund Optics, Barrington, USA). The focused light was directed into an optical fibre which terminated in a USB-driven diode array spectrometer (USB4000, Ocean Optics, Dunedin, USA) allowing the peak emission wavelength and whole spectrum if required to be monitored over a period of time. Components were positioned securely inside a custom built aluminium housing that allowed optimisation of the output fibre position through X-Y translation, a fixed capillary on the aluminium housing and vertical adjustment of the laser source. This is shown in Figure 5-4.

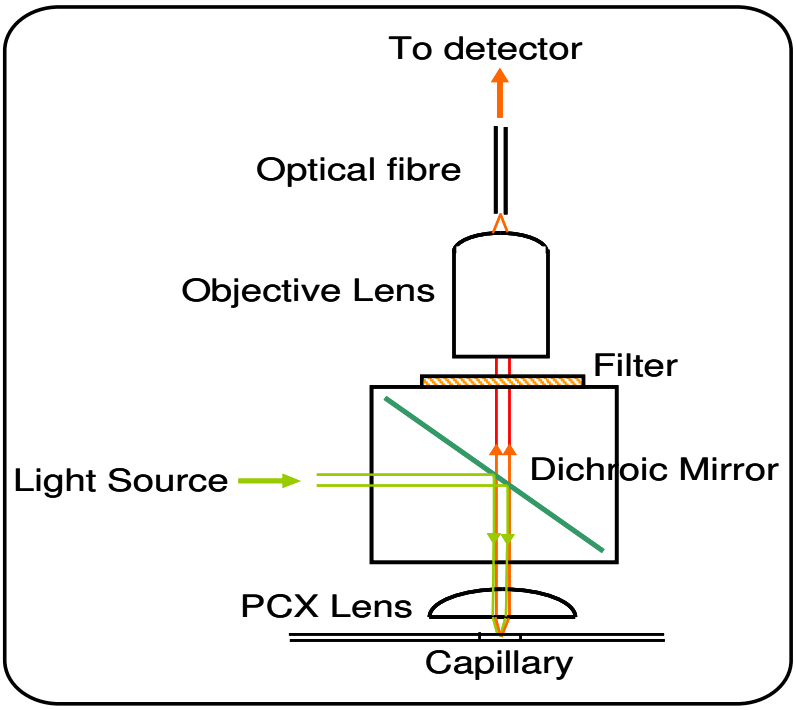


Figure 5-3: Fluorescence detector setup. Excitation light from a laser source was directed at the dichroic mirror and focused by a PCX lens onto the capillary window. Fluorescent light was captured by the same focusing lens and collimated through the dichroic mirror and an emission filter prior to collection through an objective lens. The focused light was directed into an optical fibre, which terminated in a USB-driven diode array spectrometer.

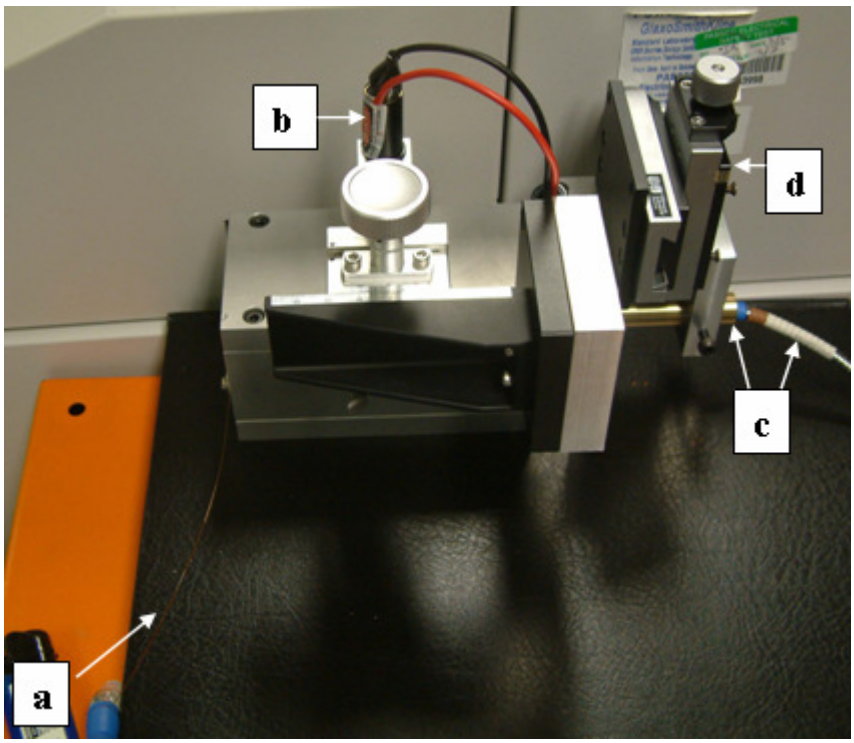


Figure 5-4: Custom built detector. a, capillary; b, laser pen; c, optical fibre; d, x-y flexure stage.

The detector was used to analyse protein-DNA interactions of samples eluting from an HPLC column. A Biosep-SEC-s-3000 (Phenomenex, Torrance, CA, USA) column was used to separate the complex from unbound DNA. Using 0.2M phosphate buffered saline (pH 6.4) 0.5 μM 21TH with varying concentrations of single-strand binding protein (SSB) which was given from Mr. Cooper, very good results were achieved. Addition of SSB showed elution of a protein-DNA complex prior to elution of unbound DNA (Figure 5-5). . Flow speed was 0.5 ml/min. The excitation light was 532 nm and the fluorescence of the eluent was monitored at 575 nm. The complex of SSB and 21TH elutes first followed by SSB alone.

From Figure 5-2 and 5-5 we can see that capillaries are suitable for detection of steady state fluorescence from low concentrations of sample.

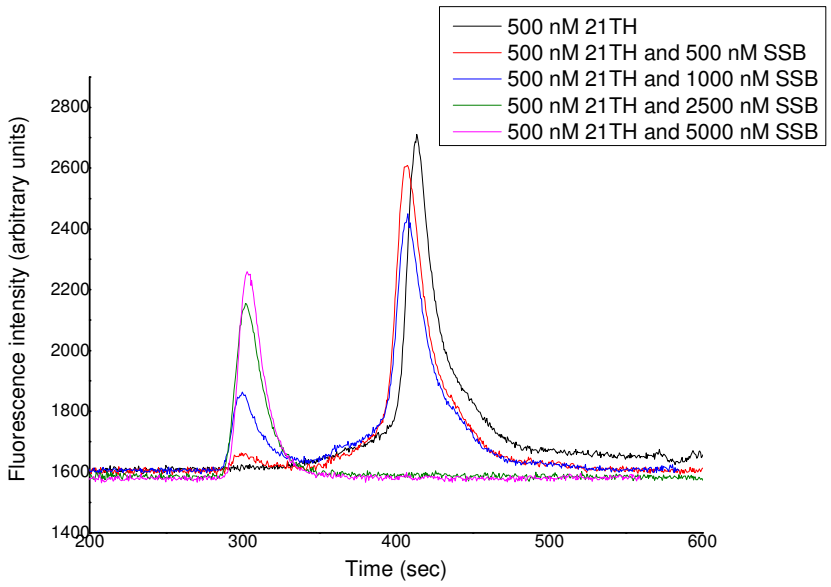


Figure 5-5: 0.5 µM 21TH bound with varying concentrations of SSB. The solution volume is 20 µl.

5.2 Time-resolved fluorescence instruments

Pulsed LEDs (PicoQuant Gmbh, Berlin, German) and a 405 nm pulsed laser were obtained and driven by a PDL 800-B pulsed diode laser driver (PicoQuant Gmbh, Berlin, German). A PMH-100-3 single photon counting photomultiplier tube (Becker & Hickl Gmbh) and a time-resolved fluorimeter equipped with an Edinburgh Instruments TCC900 single photon counting card was built. Emission wavelengths were selected with a monochromator and polarisation was applied using quartz Glan-Thompson polarisers on both excitation and emission. This instrument has 301, 376, 405, 456 and 502 nm light sources to excite the sample and any emission wavelength can be selected by the monochromator. This home made time-resolved instrument not only could detect the normal time-resolved fluorescence decay at different wavelengths, but could also detect time-resolved anisotropy decay of a sample at different wavelengths. All of the time-resolved fluorescence and anisotropy decay experiments in my thesis were performed using this instrument using a cuvette

sample. Now we wished to place a capillary in the LED beam and see if time-resolved data could be collected.

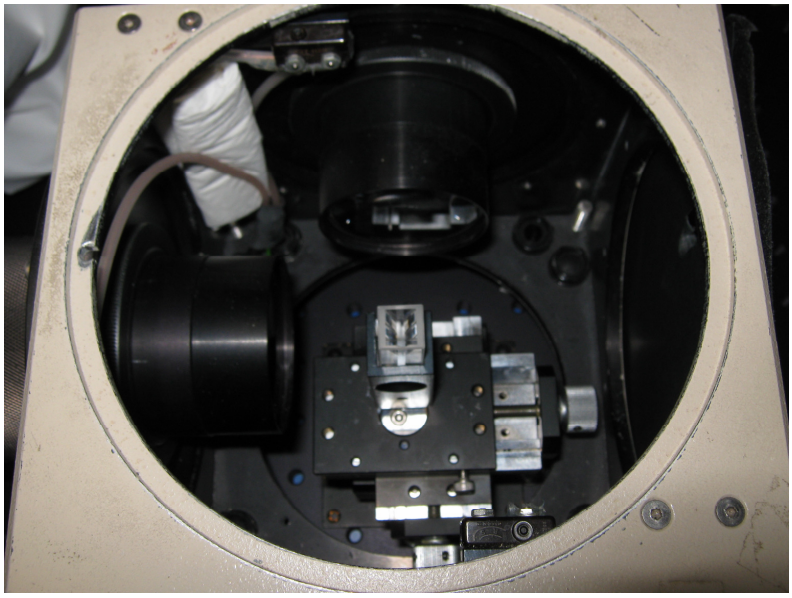


Figure 5-6: Picture of the time-resolved fluorescence instrument. The excitation light passes through the upper lens and focuses on the cuvette. Emission light is captured by the left lens and then goes to the detector after passing through a monochromator.

5.2.1 Time-resolved fluorescence capillary detector

The design in Figure 5-1 was used on the time-resolved instrument to measure 2.5 µM kiton red in water in a capillary with an I.D. of 50 µm and an O.D. of 375 µm (Figure 5-7 a). The excitation light used was 500 nm polarise LED (PLS-8-2-524) and a monochromator was used to collect the light at 580 nm at magic angle with the emission bandpass at 20 nm.

Initially it appeared as though the monochromator was not functioning property as scattering was observed at the time of the pulse. Adjusting the setup so the emission filter was before the monochromator made no difference to the light scattering effect. This suggested the monochromator was working correctly and indicated the problem may be the light source. On investigating the LED literature, it appeared that when the LED head is used at high power its light range is very broad (further than 580

nm). This meant whatever monochromator and emission filter settings were used, the light scattering remained a problem. To solve this excitation filter to remove light longer than 540 nm from the light source was introduced, and the results in Figure 5-7 were obtained. Using this setup the detector could measure the time-resolved fluorescence of samples down to micromolar concentrations in the capillary, in reasonable data collection times.

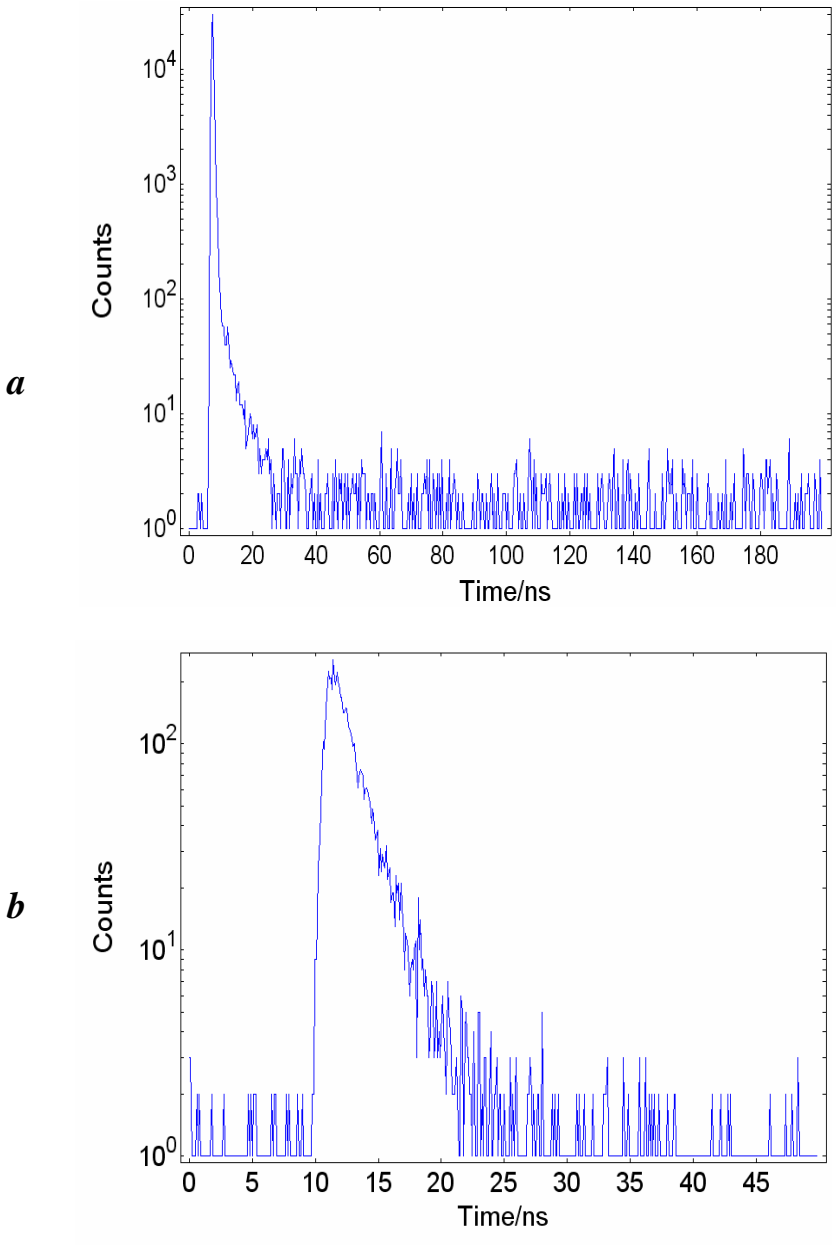


Figure 5-7: The time-resolved result of kiton red in water solution using a capillary with an I.D. of 50 µm and an O.D. of 375 µm. Panel *a* 2.5 µM kiton red

was detected by the detector without the excitation filter inside scatter is visable contributing to the size and sharpness of the initial decay, panel *b* 1.0 µM kiton red was detected by the detector with the excitation filter inside the scattering peak has been eliminated. Excitation was at 500 nm and emission at 580 nm. Both panel *a* and *b* experiments lasted about 30 sec.

So our time-resolved fluorescence capillary instrument can be used to measure fluorescence lifetime of low concentration, small volume samples by TCSPC methods.

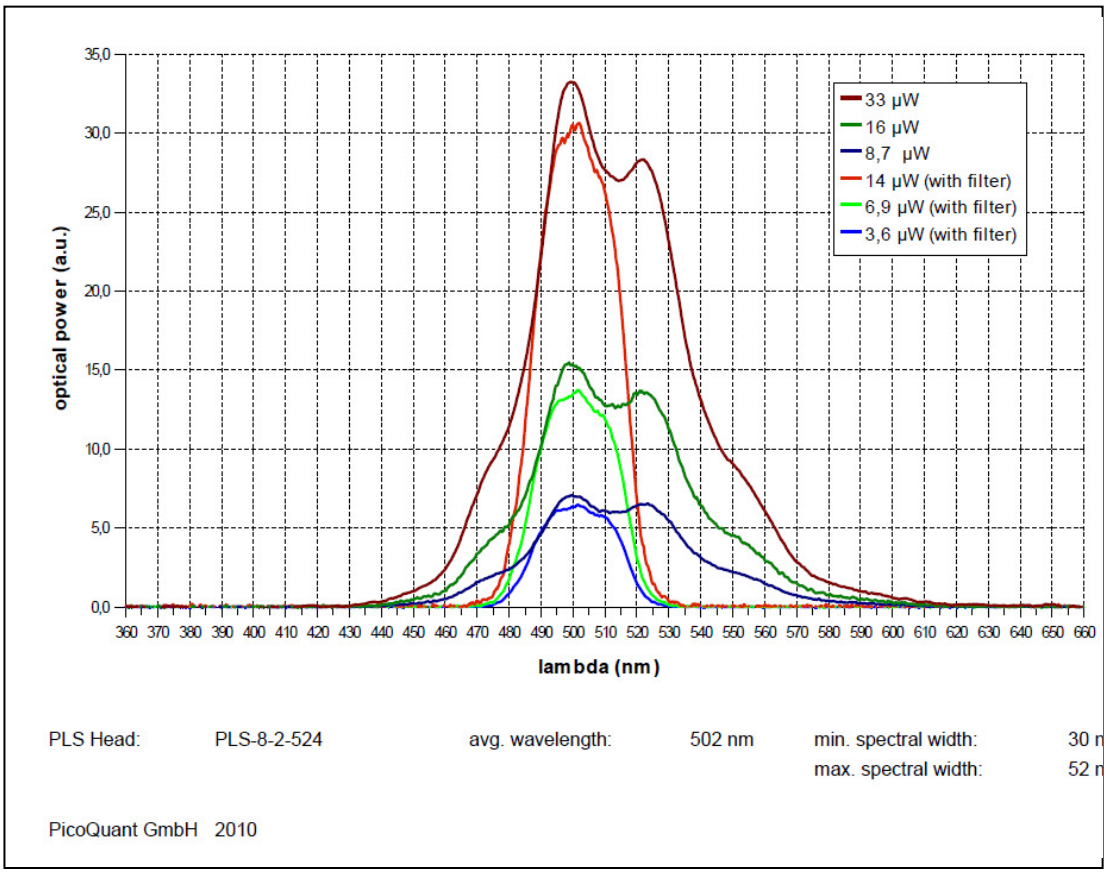


Figure 5-8: The spectral output of LED PLS-8-2-524 which was used to excite kiton red in Figure 5-7 (Reprinted from PicoQuant GmbH).

5.3 Kinetic experiment

The final question to ask was whether TCSPC could be collected during a reaction in a capillary? The rate of reaction/interaction is one of the most important parameters of a reaction/interaction. Continuous and stopped flows are very useful techniques to measure the kinetics of a reaction.

5.3.1 Stopped flow technique

Figure 5-9 shows a typical stopped flow technique. In this method, the reactants are injected into two syringes which are driven by a pump. The solutions in the syringes are mixed in the mixing chamber and proceed into a stopping syringe. When the stopping syringe fills and drives its plunger back against a stop, then the flow stops and triggers the activation of data acquisition. The reaction continues in the mixed solutions. Usually spectroscopic techniques such as UV-vis absorption, circular dichroism and fluorescence emission are used to detect the sample.

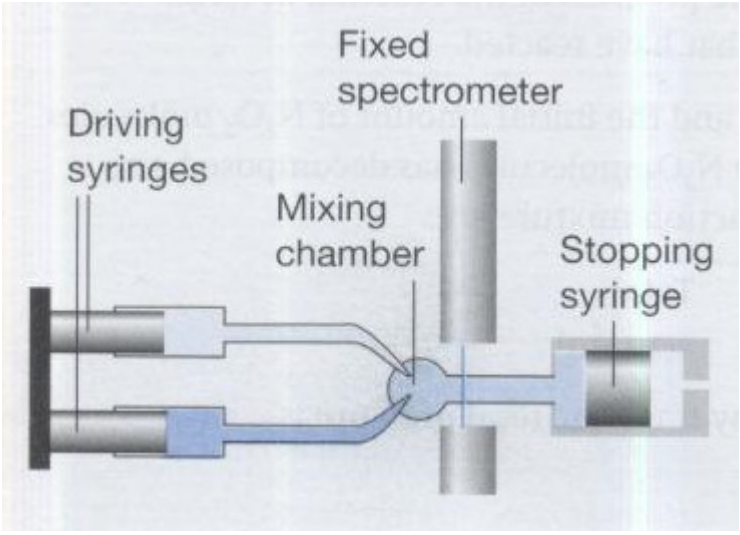


Figure 5-9: Diagrammatic representation of a typical stopped-flow instrument (reprinted from Atkins *et al*., 2002). The samples are first loaded into the two driving syringes. The experiment is initiated by simultaneously introducing the

two solutions into the mixing chamber. Once the stopping syringe is filled a trigger is activated that stops the driving syringes and switches on the spectrometer to record the properties of the mixture.

In the stopped flow technique, the reaction/interaction begins in the mixing chamber. However, the reaction mix needs to travel to the spectrometer where the product(s) can be detected. The time needed for the mixture (reactants/products) to travel to the detector is the so-called dead time of the instrument ($t_{dead\ time}$). Once the stopping syringe is full, the flow is stopped the detector starts recording. The sample is then monitored until the reaction/interaction is complete ($t_{complete}$). Therefore the time during which measurements of the sample are recorded is from $t_{dead\ time}$ to $t_{complete}$. So actually the detector is recording the data of the same sample in the spectrometer for the whole reaction/interaction. Several experiments can give you a rate constant, k_{on} and k_{off}, and binding constant for an interaction or rate constant for chemical reactions.

5.3.2 Continuous flow technique

Figure 5-10 shows the typical continuous flow technique. Compared with stopped flow, the biggest difference is that in continuous flow there is no stopping syringe, so the flow is constant and the spectrometer records the results. The flow is continuous, the speed of the syringes is constant and the reactants/products travel through the spectrometer at the same time. One experiment records the data at a specific time.
Then by changing the position of the spectrometer or the speed of the driving syringes, data for a different time can be recorded. The limitation of using continuous flow is that usually it needs lots of solution due to the flow being continuous. However, if the capillary fluorescence detector is used on continuous flow, this may not be a problem anymore, as the sample volume used will be very small. In my detector, the I.D. of capillary is just 50 μm.

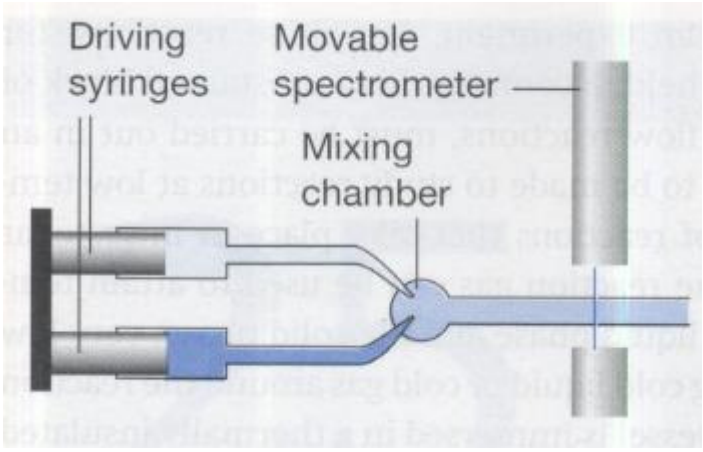


Figure 5-10: Schematic representation of a typical continuous flow instrument (reprinted from Atkins *et al*., 2002). The samples are first loaded into the two driving syringes. The experiment is initiated by simultaneously introducing the two solutions into the mixing chamber. Unlike the stopped flow instrument (see Figure 5-9), the spectroscopic properties of the mixture are then monitored continuously until all the required data has been accumulated.

Compared with the stopped flow technique, the continuous flow technique could be better at investigating the details of intermediate times. For example, fluorescence lifetime TCSPC measurement usually needs a few mins to collect enough decay data. In stopped flow experiments it is impossible to measure the lifetime of the sample. However, in continuous flow the sample age of the sample in the capillary observation window is always the same and one can collect data for as long as necessary.

5.4 Future work

We intend to combine the capillary detector, TCSPC and continuous flow in one instrument.

From Figures 5-5 and 5-7 b we can see that the detector can detect the both static fluorescence and fluorescence lifetime of a tiny amount of sample in the capillary. So it can be used with continuous flow techniques as a detector.

The continuous flow device is described below. A nano-mixer (Figure 5-11 Upchurch, Washington, USA) was obtained to mix the reactants. There are two flow paths into the mixing cell and one flow path out. It can mix about 100 nL samples in tens of milliseconds. The nano-mixer can be linked to the time-resolved fluorescence instrument or static fluorescence detector. Then we can measure the static and time-resolved fluorescence results of reactants. So we can use it to do both stopped and continuous flow experiments by controlling the syringe pump. During pumping, we can measure both static and time-resolved fluorescence at any intermediate time of the reaction/interaction. Then we will know the details of the mixture at any given time. When the pump is stopped, the sample will stop flowing instantly, and data can be collected to obtain the stopped flow result. So both stopped flow and continuous flow experiments can be done on one instrument. This complete instrument will be tested in the future.

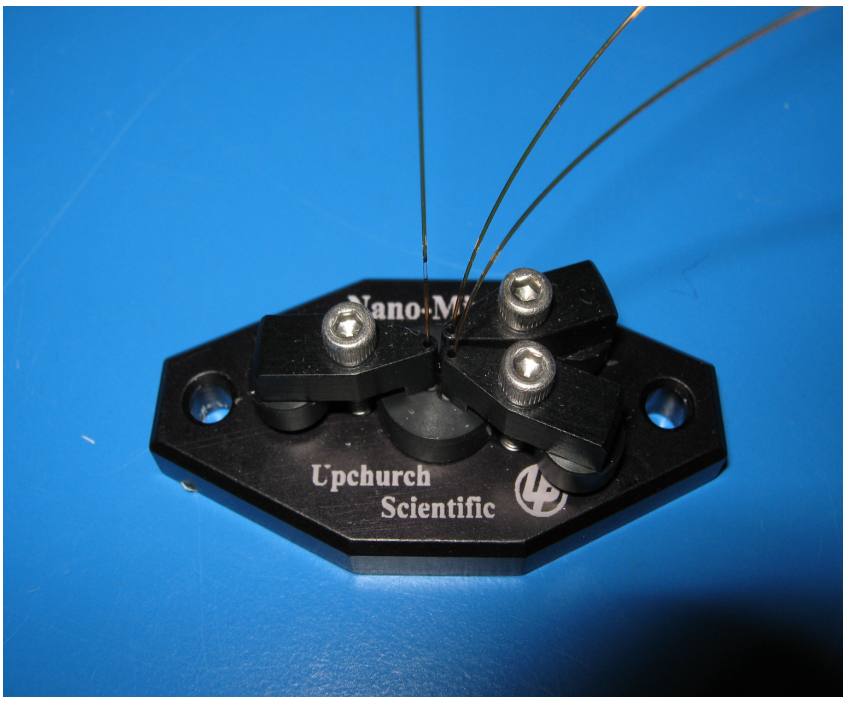


Figure 5-11: The nano-mixer. The two reactants are injected from right two capillary separately into the nano-mixer and after mixture in milliseconds, the sample will go out from left capillary and fluorescence detected from the window as in Figure 5-2 and 5-7.

Chapter 6 Conclusions and future work

EcoKI is the archetypal Type I restriction endonuclease. However, so far no crystal structure has been solved. Recently, Kennaway [Kennaway, *et. al*., 2009] have published an atomic model of the M.EcoKI MTase bound to a DNA duplex and to Ocr by using EM and computer modelling. However, this result still needs more experimental data to fully validate the model. In chapter 3, GFP was fused to the C-terminus of HsdS in a Type I MTase without any deleterious effect on its *in vitro* binding to either DNA or to a DNA mimic. The assembly of the trimeric MTase is also unaffected because the GFP appears to form an independent domain that has little interaction with the remainder of the molecule. Indeed, the anisotropy of the GFP is almost unaffected by fusion to the S subunit of MTase. The in-house time-resolved fluorescence instrument (chapter 5) has been successfully used to measure fluorescence lifetime and anisotropy decay in chapter 3. FRET distances between the chromophore of GFP and labelled DNA/Ocr were obtained and analysed to test the model of MTase. The model agrees with these FRET data.

In chapter 4, I used different methods (ITC, size-exclusion HPLC and a displacement assay) to investigate the binding affinity of the interaction between ArdA proteins (Orf18, V583, Mu50 and NCTC) and MTase/GFP-MTase. Unfortunately, no accurate Kd was obtained by these methods. Using ITC, all of the interactions between ArdA proteins and MTase/GFP-MTase were too tight to calculate the accurate binding affinity. However, the Kd was in the 0.1-10 nM range by ITC results. Size-exclusion HPLC showed the complex between NCTC and MTase was the only one that was stable enough to pass through the column intact. Indeed, NCTC stabilises the trimeric structure of MTase so that the complex did not dissociate during the column run even at very low protein concentrations. However, it is impossible to determine the binding affinity of the interactions by using the size-exclusion HPLC results. The size-exclusion results indicate that at any given

concentration there is a mixture of dimer and monomer of the ArdA proteins present in solution, which complicates analysis. The displacement assay results indicate that the Kd between ArdA proteins and GFP-MTase was in the 1-10 nM range. However, the stoichiometry of interactions between MTase or GFP-MTase with ArdA (i.e., monomer ArdA: MTase or GFP-MTase) is lower than 2. This result surprised us. Previous binding studies of MTase and Ocr show that the stoichiometry of interaction is one molecule of MTase per Ocr dimer (Atanasiu *et al*., 2002) and the dimer Orf18 structure (McMahon *et al*., 2009). Thus, one might anticipate that the MTase:ArdA complex would form with the same ratio. Therefore, for the first time it was realised that probably both monomers and dimers of ArdA proteins can form a complex with the MTase. In the future, by generating an ArdA mutant in which we change one or two amino acids at the dimer interface to introduce a negative charge, we may learn more about the interaction. Using this method of introducing a negative charge at the interface, we may find it stops the formation of the ArdA dimer. We can use circular dichroism (CD) to test if the mutant forms secondary structure and size-exclusion HPLC to test if the mutant forms a dimer in solution. Using a mutant ArdA protein which is always present as a monomer in solution, we could repeat the ITC experiments. Hopefully, we may also get the Kd of the interaction between the mutant ArdA protein and the MTase.

The in-house time-resolved fluorescence instrument is sensitive enough to detect low concentrations of fluorescent sample in a capillary. The device contains a nano-mixer that offers two flow path options selected by simply changing the orientation of the enclosed glass/silicon channel and can mix sample volumes as small as 100 nL (Figure 5-11). A pump is used to drive two syringes to mix the two interaction/reaction samples together through the nano-mixer, and then measure the static or time-resolved fluorescence of the sample. If we measure the fluorescence change of the sample after we stop the pump we obtain stopped-flow data. Compared to the stopped flow technique, the continuous flow technique is better at investigating the details of intermediate times. For example, fluorescence lifetime measurement usually needs a few minutes to collect sufficient decay data. In stopped flow experiments it is impossible to measure the lifetime of the sample, whereas in

continuous flow the sample age in the capillary is always the same and one can collect data for as long as necessary. However, because the flow is continuous, usually this technique needs lots of sample. Since our detector only needs tiny amounts of sample to detect the fluorescence, we can do the continuous flow experiments using only a few hundred micro litres. In the future this device will be used to further the analyses presented in chapter 3 and 4.

References

Allan, B. W., Reich, N. O. (1996) Targeted base stacking disruption by the EcoRI DNA methyltransferase. *Biochemistry* **35** 14757–14762.

Amann, E., Ochs, B., Abel, K. J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in Escherichia coli. *Gene* **69** 301–315.

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

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. *Nucl. Acids Res.* **29** 3059-3068.

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. *Nucl. Acids Res.* **30** 3936-3944.

Atkins, P., Paula, J. D. (2002) Physical chemistry. 7th Ed. *Oxford University Press* Oxford 864.

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.

Barcus, V. A., Titheradge, A. J., Murray, N. E. (1995) The diversity of alleles at the hsd locus in natural populations of Escherichia coli. *Genetics* **140** 1187-1197.

Barrett, W. A., Norbert, O. R. (1996) Targeted Base Stacking Disruption by the *Eco*RI DNA Methyltransferase. *Biochemistry* **35** 14757-14762.

Belogurov, A. A., Delver, E. P. (1995) A motif conserved among the type I restriction-modification enzymes and antirestriction proteins: a possible basis for mechanism of action of plasmid-encoded antirestriction functions. *Nucl. Acids Res.* **23** 785-787.

Belogurov, A. A., Delver, E. P., Rodzevich, O. V. (1993) Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *J. Bacteriol.* **175** 4843-4850.

Belogurov, A. A., Yussifov, T. N., Kotova, V. U., Zavilgelsky, G. B. (1985) The novel gene(s) ARD of plasmid pKM101: alleviation of EcoK restriction. *Mol. Gen. Genet.* **198** 509-513.

Berge, T., Ellis, D. J., Dryden, D. T. F., Edwardson, J. M., Henderson, R. M. (2000) Translocation-independent dimerization of the EcoKI endonuclease visualized by atomic force microscopy. *Biophys. J.* **79** 479–484.

Bertani, G., Weigle, J. J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.* **65** 113-121.

Bheemanaik, S., Reddy, Y. V. R., Rao, D. N. (2006) Structure, function and mechanism of exocyclic DNA methyltransferases. *Biochem. J.* **399** 177-190.

Bickle, T. A., Kruger, D. H. (1993) Biology of DNA restriction. *Microbiol. Rev*. **57** 434-450.

Blackstock, J. J., Egelhaaf, S. U., Atanasiu, C., Dryden, 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.

Brejc, K., Sixma, T. K., Kitts, P. A., Kain, S. R., Tsien, R. Y., Ormo, M., Remington, S. J. (1997) Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl. Acad. Sci. USA*. **94** 2306-2311.

Calisto, B. M., Pich, O. Q., Piñol, J., Fita, I., Querol, E., Carpena, X. (2005) Crystal Structure of a Putative Type I Restriction–Modification S Subunit from Mycoplasma genitalium. *J. Mol. Biol.* **351** 749-762.

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

Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C., *et al*. (2005) Extensive DNA inversions in the B. fragilis genome control variable gene expression. *Science* **307** 1463-1465.

Chalfie, M., Kain, S. R. (2006) Green Fluorescent Protein: Properties, Applications and Protocols. 7th Ed. *John Wiley & Sons, Inc.* New York.

Chen, K., Roberts, G. A., Stephanou, A. S., Cooper, L. P., White, J. H., Dryden, D. T. F. (2010) Fusion of GFP to the M.EcoKI DNA methyltransferase produces a new probe of Type I DNA restriction and modification enzymes. *Biochem. Biophys. Res. Commun.* **398** 254-259.

Cubitt, A. B., Heim, R., Adams, S. R., Boyd, A. E., Gross, A., Tsien, R. Y. (1995) Understanding, improving and using green fluorescent proteins. *Trends Biochem. Sci.* **20** 448-455.

Day, R. N., Davidson, M. W. (2009) The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev.* **38** 2887-2921.

Donald, V., Judith, G. V., Charlotte, W. P. Fundamentals of biochemistry. *John Wiley & Sons, Inc.* New York, 728.

Dreier, D., MacWilliams, M. P., Bickle, T. A., (1996) DNA cleavage by the type IC restriction-modification enzyme EcoR124II. *J. Mol. Biol.* **264** 722-733.

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.

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.

Dryden, D. T. F., Sturrock, S. S., Winter, W. (1995) Structural modelling of a type I DNA methyltransferase. *Nat. Struct. Biol.* **7** 632-635.

Dybvig, K., Sitaraman, R., French, C. T. (1998) A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements. *Proc. Natl. Acad. Sci. USA*. **95** 13923-13928.

Ellis, D. J., Dryden, D. T. F., Berge, T., Edwardson, J. M., Henderson, R. M. (1999) Direct observation of DNA translocation and cleavage by the EcoKI endonuclease using atomic force microscopy. *Nat. Struct. Biol.* **6** 15–17.

Flannagan, S. E., Zitzow, L. A., Su, Y. A., Clewell, D. B. (1994) Nucleotide sequence of the 18-kb conjugative transposon Tn916 from *Enterococcus faecalis*. *Plasmid* **32** 350-354.

Gill, S. C., von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* **182** 319-326.

Goedecke, K., Pignot, M., Goody, R. S., Scheidig, A. J., Weinhold, E. (2001) Structure of the N6-adenine DNA methyltransferase M-TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.* **8** 121-125.

Guest, C. R., Hoschstrasser, R. A., Sowers, L. C., Millar, D. P. (1991) Dynamics of mismatched base pairs in DNA. *Biochemistry* **30** 3271-3279.

Halford, S. E., Marko, J. F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucl. Acids Res.* **32** 3040-3052.

Heim, R., Cubitt, A., Tsien, R. Y. (1995) Improved green fluorescence. *Nature* **373** 663-664.

Heim, R., Prasher, D. C., Tsien, R. Y. (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl. Acad. Sci. USA*. **91** 12501-12504.

Iida, S., Streiff, M. B., Bickle, T. A., Arber, W. (1987) Two DNA anti-restriction systems of bacteriophage P1, darA, and darB: characterization of darA-phages. *Virology* **157** 156-166.

Hoffman, R. M. (2002) Green fluorescent protein imaging of tumour growth, metastasis, and angiogenesis in mouse models. *The Lancet Oncol.* **3** 546-556.

Kan, N. C., Lautenberger, J. A., Edgell, M. H., Hutchison, C. A. (1979) The nucleotide sequence recognized by the *Escherichia coli* K12 restriction and modification enzymes. *J. Mol. Biol.* **130** 191-209.

Kennaway, C. K., Obarska-Kosinska, A., White, J. H., Tuszynska, I., Cooper, L. P., Bujnicki, J. M., Trinick, J., Dryden, D. T. F. (2009) The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucl. Acids Res.* **3** 762-770.

Kerppola, T. K. (2008) Bimolecular fluorescence complementation (BiFC) analysis as a probe of protein interactions in living cells. *Annu. Rev. of Biophys.* **37** 465-487.

Kerppola, T. K. (2009) Visualization of molecular interactions using bimolecular fluorescence complementation analysis: Characteristics of protein fragment complementation. *Chem. Soc. Rev.* **38** 2876-2886.

Kim, J. S., DeGiovanni, A., Jancarik, J., Adams, P. D., Yokota, H., Kim, R., Kim, S. H. (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.

Kraulis, P. (1991) a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24** 946-950.

Kruger, D. H., Barcak, G. J., Reuter, M., Smith, H. O. (1988) EcoRII can be activated to cleave refractory DNA recognition sites. *Nucl. Acids Res.* **16** 3997-4008.

Kruger, D. H., Bickle, T. A. (1983) Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.* **47** 345-360.

Kruger, D. H., Schroeder, C., Santibanez, K. 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.

Lakowicz, J. R. (1996) Principles of Fluorescence Spectroscopy. 3rd Ed. *Springer* New York 354-362.

Lapkouski, M., Panjikar, S., Janscak, S. K. I., Carey, J., Ettrich, R., Csefalway, E. (2009) Structure of the motor subunit of type I restriction-modification complex EcoR124I. *Nat. Struct. Mol. Biol.* **16** 94-95.

Ma, X., Ehrhardt, D. W., Margolin, W. (1996) Colocalization of cell division proteins FtsZ and FtsA to cytoskeletal structures in living *Escherichia coli* cells by using green fluorescent protein. *Proc. Natl. Acad. Sci. USA.* **93** 12998-13003.

McMahon, S. A., Roberts, G. A., Johnson, K. A., Cooper, L. P., Liu, H., White, J. H., Carter, L. G., Sanghvi, B., Oke, M., Walkinshaw, M. D., Blakely, G. W., Naismith, J. H., Dryden, D. T. F. (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.

Meisel, A., Bickle, T. A., Kruger, D. H., 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, T. A., Kruger, D. H., Schroeder, C. (1995) Type III restriction endonucleases translocate DNA in a reaction driven by recognition sitespecific ATP hydrolysis. *EMBO J.* **14** 2958-2966.

Meselson, M., Yuan, R. (1968) DNA restriction enzyme from *E. coli. Nature* **217**, 1110-1114.

Millington, M., Grindlay, G. J., Altenbach, K., Neely, R. K., Kolch, W., Bencina, M., Read, N. D., Jones, A. C., Dryden, D. T. F., Magennis, S. W. (2007) High-precision FLIM-FRET in fixed and living cells reveals heterogeneity in a simple CFP-YFP fusion protein. *Biophys. Chem.* **127** 155-164.

Morise, H., Shimomura, O., Johnson, F. H., Winant, J. (1974) Intermolecular energy transfer in the bioluminescent system of *Aequorea*. *Biochemistry* **13** 2656-2662.

Neaves, K. J., Cooper, L. P., White, J. H., Carnally, S. M., Dryden, D. T. F., Edwardson, J. M., Henderson, R. M. (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.

Neely, R. K., Daujotyte, D., Grazulis, S., Magennis, S, W., Dryden, D. T. F., Klimašauskas, S., Jones, A. C. (2005) Time-resolved fluorescence of 2-aminopurine as a probe of base flipping in M.HhaI-DNA complexes *Nucl. Acids Res.* **33** 6953-6960.

Neely, R. K., Magennis, S. W., Dryden, D. T. F., Jones, A. C. (2004) Evidence of tautomerism in 2-aminopurine from fluorescence lifetime measurements *J. Phys. Chem. B* **108** 17606-17610.

Nekrasov, S. V., Agafonova, O. V., Belogurova, N. G., Belogurov, A. A. (2007) Plasmid-encoded antirestriction protein ArdA can discriminate between type I methyltransferase and complete restriction-modification system. *J. Mol. Biol.* **365** 284-297.

Nienhaus, G. U., Wiedenmann, J. (2009) Structure, dynamics and optical properties of fluorescent proteins: perspectives for marker development. *ChemPhysChem* **10** 1369-1379.

Nordlund, T. M., Andersson, S., Nilsson, L., Rigler, R., Graslund, A., McLaughlin, L. W. (1989) Structure and dynamics of a fluorescent DNA oligomer containing the EcoRI recognition sequence: fluorescence, molecular dynamics, and NMR studies. *Biochemistry* **28** 9095-9103.

Obarska, A., Taylor, J. E., Callow, P., Orlowski, J., Bujnicki, J. M., Kneale, G. (2008) HsdR subunit of the Type I restriction-modification enzyme EcoR124I: biophysical characterisation and structural modelling. *J. Mol. Biol.* **376** 438-452.

Obarska, A., Blundell, A., Feder, M., Vejsadová, S., Sisáková, E., Weiserová, M., Bujnicki, J. M., Firman, K. (2006) Structural model for the multisubunit TypeIC

restriction-modification DNA methyltransferase M.EcoR124I in complex with DNA *Nucl. Acids Res.* **7** 1992-2005.

Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4** 2411-2423.

Pieper, U., Brinkman, T., Kruger, T., Noyer, W., 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.

Powell, L. M., Dryden, D. T. F., Willcock, D. F., Pain, R. H., Murray, N. E. (1993) DNA recognition by the EcoK methyltransferase: The infuence of DNA methylation and the cofactor *S*-adenosyl-L-methionine. *J. Mol. Biol.* **234** 60-71.

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.

Rajaram, N., Kerppola, T. K. (2004) Synergistic transcription activation by maf and sox and their subnuclear localization are disrupted by a mutation in maf that causes cataract. *Mol. Cell. Biol.* **24** 5694-5709.

Roberts, R. J., Belfort, M. T., Bestor, T., *et al*., (2003) A nomenclature for restriction enzymes, DNA MTases, homing endonucleases and their genes. *Nucl. Acids Res.* **31** 1805-1812.

Roberts, R. J., Vincze, T., Posfai, J., Macelis, D. (2007) REBASE-enzymes and genes for DNA restriction and modification. *Nucl. Acids Res.* **35** 269-270.

Sample, V., Newmana, R. H., Zhang, J. (2009) The structure and function of fluorescent protein. *Chem. Soc. Rev.* **38** 2852-2864.

Senejani, A. G., Gogarten, J. P. (2007) Structural stability and endonuclease activity of a PI-SceI GFP-fusion protein. *Int. J. Biol. Sci.* **3** 205-211.

Serfiotis-Mitsa D., Roberts, G. A., Cooper, L. P., White, J. H., Nutley, M., Cooper, A., Blakely, G. W., Dryden, D. T. F. (2008) The Orf18 gene product from conjugative transposon Tn*916* is an ArdA antirestriction protein that inhibits Type I DNA restriction-modification systems. *J. Mol. Biol.* **383** 970-981.

Shimomura, O., Johnson, F. H., Saiga, Y. (1962) Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan *Aequorea*. *J. Cell. Comp. Physiol.* **59** 223-239.

Stephanou, A. S., Roberts, G. A., Tock, M. R., Pritchard, E. H., Turkington, R., Nutley, M., Cooper, A., Dryden, D. T. F. (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.

Stryer, L. (1988) Biochemistry, 3rd Ed. *W. H. Freeman and Company* New York.

Studier, F. W., Bandyopadhyay, P. K. (1988) Model for how type I restriction enzymes select cleavage sites in DNA. *Proc. Natl. Acad. Sci. USA.* **85** 4677-4681.

Studier, F. W. (1973) Analysis of bacteriophage T7 early RNAs and proteins on slab gels. *J. Mol. Biol.* **79** 237-242.

Studier, F.W., Movva, N.R. (1976) SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J. Virol.* **19** 136-145.

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. *Nucl. Acids Res.* **33** 3235-3244.

Terwilliger, T. C., Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Cryst. D* **55** 849-861.

Tock, M. R., Dryden, D. T. F. (2005) The biology of restriction and anti-restriction. *Curr. Op. Microbiol.* **8** 466-472.

Voet, D., Voet, J. G., Pratt, C. W. (1999) Fundamentals of biochemistry. *John Wiley & Sons, Inc.* New York 725-771.

Volkmer, A., Subramaniam, V., Birch, D. J. S., Jovin, T. M. (2000) One- and Two-Photon Excited Fluorescence Lifetimes and Anisotropy Decays of Green Fluorescent Proteins. *Biophys. J.* **78** 1589-1598.

Walkinshaw, M. D., Taylor, P., Sturrock, S. S., Atanasiu, C., Berge, T., Henderson, R. M., Edwardson, J. M., Dryden, D. T. F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell* **9** 187-194.

Ward, D. C., Reich, E., Stryer, L. (1969) Chemistry and metabolism of macromolecules. *J. Biol. Chem.* **244** 1228-1237.

Ward, W. W., Cody, C. W., Hart, R. C., Cormier, M. J. (1980) Spectrophotometric identity of the energy transfer chromophores in *Renilla* and *Aequorea* green-fluorescent proteins. *Photochem. Photobiol.* **31** 611-615.

Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W., Reeves, S. C. (1982) Spectral perturbations of the *Aequorea* green-fluorescent protein. *Photochem. Photobiol.* **35** 803-808.

Warren, R. A. (1980) Modified bases in bacteriophage DNAs. *Annu. Rev. Microbiol.* **34** 137-158.

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

Wilson, G. G., Murray, N. E. (1991) Restriction and modification systems. *Annu. Rev. Genet.* **25** 585-627.

Zavilgelsky, G. B. (2000) Antirestriction. *Mol. Biol. (Translation of Molekulyarnaya Biologiya (Moscow)* **34** 854-862.

Zavilgelsky, G. B., Rastorguev, S. M. (2009) Antirestriction proteins ArdA and Ocr as efficient inhibitors of type I restriction–modification enzymes. *Mol. Biol.* **43** 241-248.

Zimmer, M. (2009) GFP: from jellyfish to the Nobel Prize and beyond. *Chem. Soc. Rev.* **38** 2823-2832.

Appendixes

Appendix A: Amino acid sequence of proteins

ArdA protein sequences

Bacteroides fragilis NCTC 9343

MEAVTLSEAR VYVGTYNKYN NGSLFGKWLD LSDYSDKDEF MEACRELHKD
DQDPEFMFQD YENIPEALIS ESWLSEKFFE LRDAIEKLSE TQQEAFFVWC
DHHNSDISEE DADDLISSFE DEYQGEYKDE EDYAYEIVEQ CYDLPEFAKT
YFDYSAFARD LFITDYWMDN GFVFRCA

Staphylococcus aureus Mu50

MEMKVYVANL GRYNEGELVG AWFTPPIDEE EMAERIGLNE DYEEYAIHDF
ELPFDVDEYT PISEINRLCE AIQEIEGTPI YNELKEIQGM WFSSLEELLE
NKEDIHCYSD CDSMEDVARY YVEETGQLGE VPSNLQNYID YQSLGRDMEI
EGNYLVTSHG VFEYCQ

Enterococcus faecalis Orf18

MDDMQVYIAN LGKYNEGELV GAWFTFPIDF EEVKEKIGLN DEYEEYAIHD
YELPFTVDEY TSIGELNRLW EMVSELPEEL QSELSALLTH FSSIEELSEH
QEDIIIHSDC DDMYDVARYY IEETGALGEV PASLQNYIDY QAYGRDLDLS
GTFISTNHGI FEIVY

Enterococcus faecalis V583

MEQMRVYIAN LGKYNEGELV GAWFTPPVDF DEVKERIGLN DDYEEYAIHD
YELPFEIDEY TSIEEINRLC GLAEELEGTP IGEVASEIQH AFFNSFEEMV
EHVDDIVYYP DCNDMEDLAY QMVNEGYLGD APENFVRYFN YSSFARDLEI
EGNYLVTNRG IFEYPI

Sequence of EGFP, MTase and HsdS-EGFP with His tag

HsdM

MNNNDLVAKL WKLCDNLRDG GVSYQNYVNE LASLLFLKMC KETGQEAEYL
PEGYRWDDLK SRIGQEQLQF YRKMLVHLGE DDKKLVQAVF HNVSTTITEP
KQITALVSNM DSLDWYNGAH GKSRDDFGDM YEGLLQKNAN ETKSGAGQYF
TPRPLIKTII HLLKPQPREV VQDPAAGTAG FLIEADRYVK SQTNDLDDLD
GDTQDFQIHR AFIGLELVPG TRRLALMNCL LHDIEGNLDH GGAIRLGNTL
GSDGENLPKA HIVATNPPFG SAAGTNITRT FVHPTSNKQL CFMQHIIETL
HPGGRAAVVV PDNVLFEGGK GTDIRRDLMD KCHLHTILRL PTGIFYAQGV
KTNVLFFTKG TVANPNQDKN CTDDVWVYDL RTNMPSFGKR TPFTDEHLQP
FERVYGEDPH GLSPRTEGEW SFNAEETEVA DSEENKNTDQ HLATSRWRKF
SREWIRTAKS DSLDISWLKD KDSIDADSLP EPDVLAAEAM GELVQALSEL
DALMRELGAS DEADLQRQLL EEAFGGVKE

HsdS

MSAGKLPEGW VIAPVSTVTT LIRGVTYKKE QAINYLKDDY LPLIRANNIQ
NGKFDTTDLV FVPKNLVKES QKISPEDIVI AMSSGSKSVV GKSAHQHLPF
ECSFGAFCGV LRPEKLIFSG FIAHFTKSSL YRNKISSLSA GANINNIKPA
SFDLINIPIP PLAEQKIIAE KLDTLLAQVD STKARFEQIP QILKRFRQAV
LGGAVNGKLT EKWRNFEPQH SVFKKLNFES ILTELRNGLS SKPNESGVGH
PILRISSVRA GHVDQNDIRF LECSESELNR HKLQDGDLLF TRYNGSLEFV
GVCGLLKKLQ HQNLLYPDKL IRARLTKDAL PEYIEIFFSS PSARNAMMNC
VKTTSGQKGI SGKDIKSQVV LLPPVKEQAE IVRRVEQLFA YADTIEKQVN
NALARVNNLT QSILAKAFRG ELTAQWRAEN PDLISGENSA AALLEKIKAE
RAASGGKKAS RKKS

EGFP

MVSKGEELFT GVVPILVELD GDVNGHKFSV SGEGEGDATY GKLTLKFICT
TGKLPVPWPT LVTTLTYGVQ CFSRYPDHMK QHDFFKSAMP EGYVQERTIF

FKDDGNYKTR AEVKFEGDTL VNRIELKGID FKEDGNILGH KLEYNYNSHN
VYIMADKQKN GIKVNFKIRH NIEDGSVQLA DHYQQNTPIG DGPVLLPDNH
YLSTQSALSK DPNEKRDHMV LLEFVTAAGI TLGMDELYK

HsdS-EGFP with His tag sequence

MSAGKLPEGW VIAPVSTVTT LIRGVTYKKE QAINYLKDDY LPLIRANNIQ
NGKFDTTDLV FVPKNLVKES QKISPEDIVI AMSSGSKSVV GKSAHQHLPF
ECSFGAFCGV LRPEKLIFSG FIAHFTKSSL YRNKISSLSA GANINNIKPA
SFDLINIPIP PLAEQKIIAE KLDTLLAQVD STKARFEQIP QILKRFRQAV
LGGAVNGKLT EKWRNFEPQH SVFKKLNFES ILTELRNGLS SKPNESGVGH
PILRISSVRA GHVDQNDIRF LECSESELNR HKLQDGDLLF TRYNGSLEFV
GVCGLLKKLQ HQNLLYPDKL IRARLTKDAL PEYIEIFFSS PSARNAMMNC
VKTTSGQKGI SGKDIKSQVV LLPPVKEQAE IVRRVEQLFA YADTIEKQVN
NALARVNNLT QSILAKAFRG ELTAQWRAEN PDLISGENSA AALLEKIKAE
RAASGGKKAS RKKSMVSKGE ELFTGVVPIL VELDGDVNGH KFSVSGEGEG
DATYGKLTLK FICTTGKLPV PWPTLVTTLT YGVQCFSRYP DHMKQHDFFK
SAMPEGYVQE RTIFFKDDGN YKTRAEVKFE GDTLVNRIEL KGIDFKEDGN
ILGHKLEYNY NSHNVYIMAD KQKNGIKVNF KIRHNIEDGS VQLADHYQQN
TPIGDGPVLL PDNHYLSTQS ALSKDPNEKR DHMVLLEFVT AAGITLGMDE
LYKHHHHHH

The polyhistidine region (i.e. six consecutive histidine residues known as a His tag) at the C-terminus is highlighted by underline. The His tag simplified purification of the recombinant protein by using metal chelation (Ni^{2+}) chromatography.

Appendix B: Derivation of the binding equation for displacement assay

For reversible reaction/interaction of protein and protein,

$$P1 + P2 \rightleftharpoons P1 \cdot P2$$

Where P1 is ArdA, P2 is GFP-MTase. P1·P2 is the complex of ArdA and GFP-MTase. The dissociation constant (*Kd*) can be calculate by Equation B-1.

Equation B-1:
$$Kd = \frac{[P1][P2]}{[P1 \cdot P2]}$$

Where [P1], [P2] and [P1·P2] are the concentration of P1, P2 and P1·P2, respectively.

Equation B-2: $[P1]=[P1]_t-[P1 \cdot P2]$

Equation B-3: $[P2]=[P2]_t-[P1 \cdot P2]$

where $[P1]_t$ and $[P2]_t$ are the total concentration of P1, P2, respectively.
By substituting equation B-2 and B-3 into B-1 we get Equation B-4.

Equation B-4:
$$Kd = \frac{[P1][P2]}{[P1 \cdot P2]} = \frac{([P1]_t - [P1 \cdot P2])([P2]_t - [P1 \cdot P2])}{[P1 \cdot P2]}$$

Figure B-1 is a typical ArdA titration of GFP-MTase bound to beads. As more ArdA is added to the mixture more GFP-MTase goes into solution.

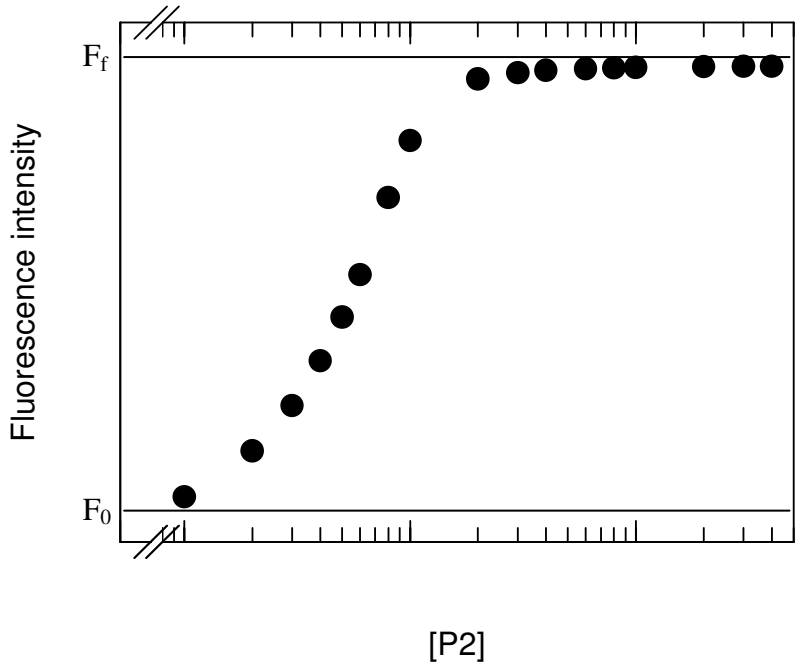


Figure B-1: Typical titration results of P2 into P1 solution. X axis is the concentration of P2. Y axis is the fluorescence intensity of complex.

Equation B-5: $$\frac{F - F_0}{F_f - F_0} = \frac{[P1 \cdot P2]}{[P1 \cdot P2]_t}$$

where F is the observed fluorescence of GFP-MTase, F_0 is the fluorescence background (before addition of any ArdA), F_f is the fluorescence at the end of the titration, $[P1 \cdot P2]_t$ is the concentration of total complex, P1·P2.

From Equation B-5 we get Equation B-6,

Equation B-6: $$\frac{F - F_0}{F_f - F_0} \cdot [P1 \cdot P2]_t = [P1 \cdot P2]$$

And make **Equation B-7:** $$\frac{F - F_0}{F_f - F_0} = a$$

Substitution Equation B-7 into B-6 we get Equation B-8,

Equation B-8:

$$Kd = \frac{([P1]_t - a \cdot [P1 \cdot P2]_t)([P2]_t - a \cdot [P1 \cdot P2]_t)}{a \cdot [P1 \cdot P2]_t}$$

From Equation B-8 we get

Equation B-9:

$$a^2 \cdot [P1 \cdot P2]_t - a \cdot ([P2]_t + [P1]_t + Kd) + [P2]_t = 0$$

So get **Equation B-10:**

$$a = \frac{[P2]_t + [P1]_t + Kd \pm \sqrt{([P2]_t + [P1]_t + Kd)^2 - 4[P1 \cdot P2]_t[P2]_t}}{2[P1 \cdot P2]_t}$$

Equation B-11: $$F = a(F_f - F_0) + F_0$$

Substitution Equation B-10 into Equation B-11 we get Equation B-12,

Equation B-12:

$$F = (F_f - F_0)\frac{[P2]_t + [P1]_t + K_d \pm \sqrt{([P2]_t + [P1]_t + K_d)^2 - 4[P1 \cdot P2]_t[P2]_t}}{2[P1 \cdot P2]_t}$$

where F is the observed fluorescence, F_f, F_0, $[P1 \cdot P2]_t$ and $[P2]_t$ are constants, $[P1]_t$ is known titration ArdA concentration. Thus, we can get the k_d by titration of ArdA into GFP-MTase.

Published online 8 September 2009

Nucleic Acids Research, 2009, Vol. 37, No. 20 ***6859–6870***
doi:10.1093/nar/gkp688

Time-resolved fluorescence studies of nucleotide flipping by restriction enzymes

Robert K. Neely[1,*], Gintautas Tamulaitis[2], Kai Chen[3], Marta Kubala[2], Virginijus Siksnys[2] and Anita C. Jones[3,4,*]

[1]Department of Chemistry, Katholieke Universiteit Leuven, Celestijnenlaan 200F, 3001 Heverlee, Belgium, [2]Institute of Biotechnology, Graiciuno 8, LT–02241, Vilnius, Lithuania, [3]School of Chemistry, University of Edinburgh, West Mains Road, Edinburgh, EH9 3JJ and [4]Collaborative Optical Spectroscopy, Micromanipulation and Imaging Centre (COSMIC), University of Edinburgh, West Mains Road, Edinburgh, EH9 3JZ, UK

Received July 2, 2009; Revised August 4, 2009; Accepted August 5, 2009

ABSTRACT

Restriction enzymes Ecl18kI, PspGI and EcoRII-C, specific for interrupted 5-bp target sequences, flip the central base pair of these sequences into their protein pockets to facilitate sequence recognition and adjust the DNA cleavage pattern. We have used time-resolved fluorescence spectroscopy of 2-aminopurine-labelled DNA in complex with each of these enzymes in solution to explore the nucleotide flipping mechanism and to obtain a detailed picture of the molecular environment of the extrahelical bases. We also report the first study of the 7-bp cutter, PfoI, whose recognition sequence (T/CCNGGA) overlaps with that of the Ecl18kI-type enzymes, and for which the crystal structure is unknown. The time-resolved fluorescence experiments reveal that PfoI also uses base flipping as part of its DNA recognition mechanism and that the extrahelical bases are captured by PfoI in binding pockets whose structures are quite different to those of the structurally characterized enzymes Ecl18kI, PspGI and EcoRII-C. The fluorescence decay parameters of all the enzyme-DNA complexes are interpreted to provide insight into the mechanisms used by these four restriction enzymes to flip and recognize bases and the relationship between nucleotide flipping and DNA cleavage.

INTRODUCTION

The complete rotation of a nucleotide from the DNA duplex, around the phosphate backbone, and into the catalytic pocket of an enzyme is commonly known as nucleotide flipping. It was first observed for the methyltransferase enzyme HhaI (1) and has since been observed in many other DNA methyltransferase systems (2–6) and several DNA-repair enzyme complexes (7,8). Recently, it was demonstrated that the Ecl18kI restriction enzyme uses nucleotide flipping as part of its DNA recognition mechanism (9). The co-crystal structure of Ecl18kI-bound DNA showed that the enzyme binds to the DNA and that both nucleotides that constitute the central base pair in its 5′-CCNGG-3′ (where *N* is any base) recognition sequence are flipped from the duplex, into the enzyme binding pockets. This report was the first example of an enzyme using nucleotide flipping in a situation where the flipped base is not the target of some chemical modification by the enzyme. This seemed a surprising and energetically extravagant method for recognizing a short-DNA sequence, but was soon found to apply to the evolutionarily related enzymes, PspGI (6,10–14) and EcoRII-C (11,15) the catalytic subunit of EcoRII (Table 1).

Recently, using the fluorescent analogue of adenine, 2-aminopurine (2AP), we demonstrated that Ecl18kI, EcoRII-C and PspGI unstack bases at the centre of their recognition sequences in solution (11). The increase in 2AP fluorescence intensity upon complex formation varied significantly in the Ecl18kI, PspGI and EcoRII-C complexes. The 2AP fluorescence increase was largest for PspGI (64-fold), intermediate for EcoRII-C (~12-fold) and smallest for Ecl18kI (6.5-fold).

Although an increase in steady-state intensity is useful as an indicator of destacking of 2AP from the duplex, the magnitude of the intensity change is difficult to interpret, since different distributions of the 2AP population amongst the various conformational states of the duplex-enzyme complex may give similar

*To whom correspondence should be addressed. Tel: +32 16 327399; Fax: +32 16 327990; Email: robert.neely@chem.kuleuven.be
Correspondence may also be addressed to Anita Jones. Tel. +44 131 6506449; Fax: +44 131 6506453; Email: a.c.jones@ed.ac.uk

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

Table 1. Restriction enzymes and their recognition sequences

Enzyme	Recognition sequence
Ecl18kI	↓**CCNGG**
EcoRII	↓**CCWGG**
PspGI	↓**CCWGG**
PfoI	T↓**CCNGGA**

W denotes either A or T and downward arrows show the site targeted for phosphodiester cleavage. The common CCGG interrupted tetranucleotide is indicated in boldface.

steady-state fluorescence intensities. Time-resolved fluorescence spectroscopy gives a more detailed picture of the environment around a fluorophore and the heterogeneity of the environment. A fluorophore in a homogeneous environment exhibits one lifetime; its fluorescence decay is monoexponential. The decay is multiexponential if the fluorophore is partitioned between several environments or conformations that provide distinctly different quenching efficiencies. Thus, 2AP-containing DNA duplexes typically show fluorescence decays described by four lifetime components, reflecting the existence of the duplex in a variety of conformational states. Changes in the 2AP fluorescence decay on enzyme binding can be interpreted to give a detailed picture of the conformational distortion that is induced, including an assessment of the extent of base flipping (16–19).

Here, we have used fluorescence spectroscopy to investigate the nucleotide flipping mechanism of the structurally characterized 5-bp cutters, Ecl18kI, EcoRII-C and PspGI, and the 7-bp cutter, PfoI, for which the crystal structure is unknown. Steady-state fluorescence experiments indicate that the central base pair is unstacked in the PfoI-DNA complex, resulting in ~1000-fold increase of the 2AP fluorescence intensity, compared with the free DNA duplex. Time-resolved fluorescence experiments give new insight into the conformational dynamics of base flipping and the nature of the interactions between the extrahelical 2AP bases and the enzymes. We seek to interpret this insight to address pertinent mechanistic questions regarding these novel base flipping systems, such as is there cooperativity between the enzyme monomers to produce stable base flips; what is the orientation (extent of rotation) of the flipped base relative to the DNA duplex in the solution phase and what is the relationship between base orientation and catalysis?

MATERIALS AND METHODS

Oligonucleotides

2AP-containing oligodeoxynucleotides were obtained from Integrated DNA Technologies (HPLC grade, Coralville, USA) and non-modified oligodeoxynucleotides were purchased from Metabion (HPLC grade, Martinsried, Germany). In order to assemble duplexes, appropriate oligodeoxynucleotides (Table 2) containing 2AP or non-fluorescent control strands were mixed with a 1.05-fold molecular excess of complementary strands in the reaction buffer A [33 mM Tris-acetate (pH 7.9 at 25°C), 66 mM potassium acetate], heated to 85°C and allowed to cool slowly over several hours at room temperature. For the DNA binding and cleavage studies, both strands of the 25-bp duplexes were 5′-end labelled with [γ-^{33}P]ATP (Hartmann Analytic, Braunschweig, Germany) using a DNA labelling kit (Fermentas, Vilnius, Lithuania).

Table 2. Oligoduplexes used in this study[a]

Oligoduplex	Sequence
T/2[a]	5′ CGCACGCCT**TCCTGGA**AGCACACTA 3′ 3′ GCGTGCGGA**AGG2CCT**TCGTGTGAT 5′
2/2[a]	5′ CGCACGCCT**TCC2GGA**AGCACACTA 3′ 3′ GCGTGCGGA**AGG2CCT**TCGTGTGAT 5′
2NS[b]	5′ CGCACGCCTT**CCTGG**AAGCACACTA 3′ 3′ GCGTGCGGA**2GGACC**TTCGTGTGAT 5′
2NS′[b]	5′ CGCACGCCT**TCCTGGA**AGCACACTA 3′ 3′ GCGTGCGG**2AGGACCT**TCGTGTGAT 5′
T/A	5′ CGCACGCCT**TCCTGGA**AGCACACTA 3′ 3′ GCGTGCGGA**AGGACCT**TCGTGTGAT 5′

[a]Overlapping Ecl18kI/EcoRII-C/PspGI/PfoI recognition site is in boldface; 2 = 2AP, is indicated in boldface and underlined.
[b]Duplexes 2NS and 2NS′ contain 2AP introduced immediately adjacent to the target site. The 2NS duplex was used in the Ecl18kI/EcoRII-C/PspGI experiment and the 2NS′ duplex was used in the PfoI experiment.

Strains, plasmids and proteins

Cloning of the PfoI restriction-modification system from *Pseudomonas fluorescens* biovar 126 will be described elsewhere. Recombinant wild-type (wt) *pfoIR* gene was cloned in the pBAD24 [Ap^R] vector. Strain DH10B (ara$^-$) (F$^-$ *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *deo*R *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K16 *gal*E15 l$^-$ *rps*L *nup*G) containing plasmid pHSG415ts_M.Ecl18kI [Cm^R] (20) bearing the *ecl18kIM* gene was used as a host for transformation of [pBAD24_R.PfoI] [Ap^R] containing the *pfoIR* gene. To express the PfoI protein cells were grown in LB medium with appropriate antibiotics at 30°C to an OD_{600} of 0.7 and PfoI expression was induced by addition of arabinose to the final 0.2% concentration. The PfoI protein was purified as described previously (21). Fractions containing PfoI enzyme were pooled and dialysed against the storage buffer [10 mM Tris-HCl (pH 8.0 at 25°C), 300 mM KCl, 1 mM DTT, 1 mM EDTA (ethylenediaminetetraacetic acid), 50% glycerol] and stored at −20°C. The protein preparation was >90% pure according to a Coomassie-blue-stained sodium dodecyl sulphate (SDS)-gel.

The preparations of wt Ecl18kI, the W61Y and W61A variant of Ecl18kI, PspGI and EcoRII-C were carried out as described previously (13). All protein concentrations were determined from the absorption at 280 nm and refer to the dimers. An extinction coefficient 56 300 M^{-1} cm^{-1} was calculated for the PfoI dimer using the Vector NTI^{TM} software.

Gel mobility shift assay

^{33}P-labelled DNA duplexes (Table 2) at 0.2 nM concentration were mixed with increasing amounts of protein in the

presence of 5 mM Ca^{2+}. Apparent K_d values were determined as described previously (22).

DNA cleavage activity

The DNA cleavage activities of PfoI were monitored using a 25-bp DNA duplex containing a ^{33}P-label in both DNA strands (Table 2). Cleavage reactions were conducted at 15°C in the reaction buffer A containing 10 mM magnesium acetate and 0.1 mg ml^{-1} BSA (bovine serum albumin) using 100 nM of duplex and 1000 nM of protein. Aliquots were removed at timed intervals and quenched by mixing with loading dye [95% (v/v) formamide, 0.01% (w/v) bromophenol blue] before denaturing gel electrophoresis. The samples were analysed and quantified as described previously (23).

Steady-state fluorescence spectroscopy

PfoI steady-state fluorescence measurements were acquired in photon counting mode on a Fluoromax–3 (Jobin Yvon, Stanmore, UK) spectrofluorometer. Sample temperatures were maintained at 25°C by a circulating water bath. Emission spectra (340–420 nm) were recorded at an excitation wavelength λ^{ex} = 320 nm with excitation and emission bandwidths of 2 and 8 nm, respectively. At least three scans were averaged for each spectrum. Emission spectra were collected in reaction buffer A in the presence of 10 mM calcium acetate using 250 nM DNA alone or 250 nM DNA mixed with a 5-fold excess of the protein to ensure complete binding of the DNA. Control spectra used for the background subtraction corrections were collected under identical conditions except that duplex T/A was used instead of the fluorescent DNA. For the DNA duplex titration experiment, emission spectra of the 250 nM DNA with protein in the range of 0–2000 nM were collected.

Enzymatic digestion of the DNA duplexes was done by addition of P1 nuclease (0.5 U, Sigma Aldrich, Taufkirchen) to a solution (200 µl) containing thermically denaturated 250 nM duplex T/2 or 2/2 in 10 mM Bis-Tris-HCl (pH 5.8 at 25°C), 100 mM NaCl buffer supplemented with 5 mM zinc acetate and incubating for 1–16 h at 37°C or 56°C.

Time-resolved fluorescence spectroscopy

In order to ensure complete binding of the DNA by restriction enzymes, samples were prepared containing 1 µM DNA duplex, 5 µM enzyme and 10 mM calcium acetate in reaction buffer A. For the time-resolved measurements of the DNA duplexes alone, 5 µM of DNA duplex in the same buffer was used.

Time-resolved fluorescence spectroscopy was performed using the technique of time-correlated single photon counting. The samples were measured in an Edinburgh Instruments spectrometer equipped with TCC900 photon counting electronics. The excitation source was a Ti-Sapphire femtosecond laser system (Coherent, 10 W Verdi and Mira Ti-Sapphire) producing pulses of ~200 fs at 76 MHz repetition rate. The output of the Ti-Sapphire laser was passed through a pulse picker to reduce the repetition rate to 4.75 MHz and then frequency tripled to give an output at 316 nm. The emission from the sample was collected orthogonally to the excitation beam through a polarizer set at the magic angle (54.7°) with respect to the vertically polarized excitation. The fluorescence was passed through a monochromator (bandpass 10 nm) and detected by a Hamamatsu microchannel plate photomultiplier (R3809U-50). The instrument response of the system, measured using a dilute aqueous solution of Ludox (Sigma-Aldrich) was 70 ps full-width at half maximum.

Fluorescence decay curves were recorded on a 50 ns timescale, resolved into 4096 channels, to a total of 10 000 counts in the peak channel. Decay curves were analysed using a standard iterative reconvolution method, assuming a multiexponential decay function,

$$I(t) = \sum_{i=1}^{n} A_i \exp\left(\frac{-t}{\tau_i}\right) \quad \mathbf{1}$$

where A_i is the fractional amplitude and τ_i is the fluorescence lifetime of the i-th decay component.

Fluorescence was excited at 316 nm and decay curves were recorded at three emission wavelengths, 370, 380 and 390 nm. The three decays were analysed globally using Edinburgh Instruments 'FAST' software, i.e. they were fitted simultaneously, with lifetimes, τ_i, as common parameters. The quality of fit was judged on the basis of the reduced chi-square statistic, χ^2, and the randomness of residuals.

RESULTS

Nucleotide flipping by PfoI

Restriction endonuclease PfoI from *P. fluorescens* biovar 126 recognizes the interrupted hexanucleotide palindromic sequence 5′-TCCNGGA-3′ and cuts phosphodiester bonds on the 5′-sides of the outer cytosines (21). DNA cleavage by PfoI produces the same protruding pentanucleotide 5′-ends as Ecl18kI and the EcoRII-C/PspGI enzymes (Table 1). The functional similarity of PfoI to the base flipping enzymes Ecl18kI, PspGI and EcoRII-C, prompted us to investigate whether PfoI flips central nucleotides from its recognition sequences, using the fluorescent adenine analogue, 2AP. Gel shift experiments revealed that the binding affinity of PfoI with duplex DNA is independent of the 2AP substitution at the site for possible base flipping (data not shown). In the buffer supplemented with Mg^{2+} ions, PfoI cleaved duplexes containing or lacking 2AP at identical rates (data not shown). All 2AP fluorescence experiments were performed in the presence of Ca^{2+} ions, which do not support catalysis but are required for stable, specific DNA binding for restriction enzymes (20,22,24).

The steady-state fluorescence intensity of the T/2 and 2/2 DNA duplexes increased dramatically when bound by PfoI, as shown in Figure 1A. The relative increase in intensity was much larger than for other base flipping restriction endonucleases and other flipping enzymes (25–37) and indicates that the 2AP is flipped from the

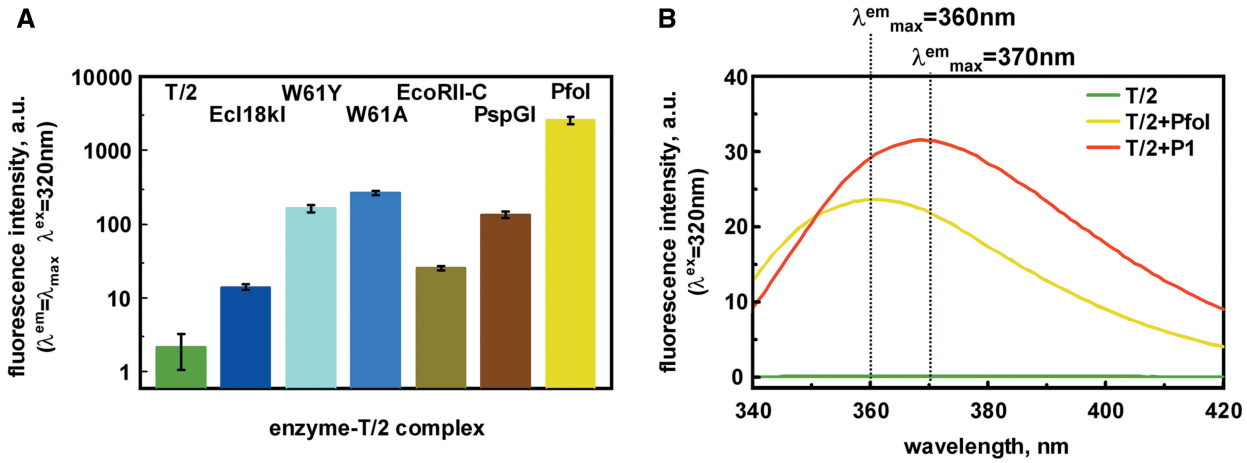


Figure 1. 2AP fluorescence intensity in the ternary complexes with PfoI. (**A**) Bar chart comparing fluorescence intensity values of the corrected fluorescence emission spectra at fluorescence maximum (360 nm for EcoRII-C/PspGI/PfoI, 367 nm for wt Ecl18kI, 365 nm for Ecl18kI W61Y and W61A and 370 nm for T/2 duplex alone) in the restriction enzyme-T/2-Ca^{2+} ternary complexes. Reactions contained 1250 nM enzyme, 250 nM DNA duplex T/2 and 10 mM Ca^{2+} ions. The fluorescence intensity data for Ecl18kI, EcoRII-C and PspGI have been published previously (11,13). (**B**) Fluorescence emission spectra of free 250 nM duplex T/2, 250 nM T/2 duplex in the presence of saturating concentration of PfoI (1250 nM) and of 250 nM T/2 duplex after digestion with nuclease P1.

duplex in the PfoI-DNA complex. When duplex 2/2 was bound by PfoI, the increase in fluorescence intensity was approximately double than that observed for T/2, implying that both target bases are flipped by PfoI, as seen previously for Ecl18kI, EcoRII-C and PspGI (11). Digestion of the T/2 (or 2/2) duplex with nuclease P1 to give the free 2AP deoxyribonucleotide, revealed that the intensity of the flipped 2AP in the PfoI complex was ~80% of that of the free deoxyribonucleotide, as shown in Figure 1B.

Time-resolved fluorescence of restriction enzyme-DNA complexes

To obtain a more detailed picture of the environment of the extrahelical bases, we carried out time-resolved fluorescence measurements on 2AP-labelled DNA in enzyme-DNA–Ca^{2+} ternary complexes. Recorded decays were modelled by the minimum number of lifetime components required to give an acceptable fit. In a simple interpretation, each lifetime (τ_i) can be considered to represent a distinct conformational state and its fractional amplitude (A_i) indicates the fraction of the population occupying this state. This simple model should not be regarded as an exact physical description of the ensemble of duplex conformations, since each lifetime is likely to represent a distribution of conformations in which 2AP experiences similar quenching rates. Nevertheless, it provides valuable insight into the conformational properties of the duplex and, more significantly in the present context, the changes in these properties induced by enzyme binding.

DNA duplex 2NS

Enzymatic binding to the 2NS duplex caused minimal perturbation to the fluorescence decay of the 2AP, as shown in Table 3. The small changes in decay parameters indicate a slight perturbation of base stacking in the duplexes on enzyme binding. As will be shown below, much greater effects are seen on enzyme binding to the duplexes with 2AP in the recognition sequence.

Table 3. Fluorescence lifetimes, τ_i and corresponding fractional amplitudes, A_i, for the DNA duplex 2NS, where the 2AP lies outside the enzymatic recognition sequence, and its complexes with PfoI, PspGI, Ecl18kI and EcoRII-C

Solution	τ_1/ns	τ_2/ns	τ_3/ns	τ_4/ns	A_1	A_2	A_3	A_4
2NS	0.02	0.30	2.1	–	0.98	0.01	0.01	–
PfoI + 2NS	0.05	0.91	3.4	8.6	0.87	0.05	0.04	0.04
PspGI + 2NS	0.05	0.49	2.6	9.4	0.92	0.05	0.02	0.01
Ecl18kI + 2NS	0.04	0.44	2.3	7.9	0.94	0.02	0.03	0.02
EcoRII-C + 2NS	0.05	0.50	2.8	8.5	0.88	0.06	0.04	0.02

DNA duplexes T2 and 2/2

The fluorescence decay parameters for the 2AP-labelled duplexes, T2 and 2/2 are shown in Table 4.

The fluorescence decays are described by four exponential components, as seen typically for 2AP-labelled DNA (16–18,38). The shortest lifetime ($\tau_1 < 100$ ps) is attributed to 2AP that is intrahelical and whose excited state is rapidly deactivated by electron transfer from guanine bases in close proximity (39,40). Approximately 90% of the free duplexes, both T/2 and 2/2, exist in this conformation. The longest decay component ($\tau_4 > 7$ ns) is attributed to 2AP that is extrahelical and experiencing a solvated environment where quenching of the fluorescence by electron transfer from guanine is not favourable. The value of the latter decay time is comparable with that found for free 2AP-riboside in solution, 10.6 ns (41,42). This component accounts for a tiny fraction (~1%) of the population. The intermediate lifetime species

Table 4. Fluorescence lifetimes, τ_i and corresponding fractional amplitudes, A_i, for duplexes T/2 and 2/2

Solution	τ_1/ns	τ_2/ns	τ_3/ns	τ_4/ns	A_1	A_2	A_3	A_4
T/2	0.03	0.41	2.4	9.7	0.94	0.03	0.02	0.01
2/2	0.07	0.46	2.0	7.8	0.89	0.05	0.05	0.01

Table 5. Fluorescence lifetimes, τ_i and the corresponding fractional amplitudes, A_i, for the PfoI-DNA-Ca^{2+} ternary complexes

Solution	τ_1/ns	τ_2/ns	τ_3/ns	τ_4/ns	A_1	A_2	A_3	A_4
PfoI + T/2	–	–	3.7	8.5	–	–	0.14	0.86
PfoI + 2/2	–	–	3.4	8.3	–	–	0.09	0.91

represent imperfectly stacked duplex conformations in which 2AP is intrahelical, but cannot be readily quenched by electron transfer from guanine. The near absence of extrahelical conformations and the overwhelming predominance of highly quenched conformations indicate the 2AP base to be firmly stacked within the duplex, in both the cases. There are some differences in decay parameters between T/2 and 2/2 duplexes, although the intrastrand sequence context of 2AP is identical in both. Thus, the conformational properties are influenced by the different base pair strengths. Notably, τ_1 is significantly shorter in T/2, indicating enhanced stacking of the 2AP with guanine neighbours.

PfoI

The fluorescence decay parameters of the ternary complexes of the duplexes, T/2 and 2/2, with PfoI and calcium differ dramatically from those of the free duplexes, as shown in Table 5.

The fluorescence decays of the PfoI-T/2-Ca^{2+} and PfoI-2/2-Ca^{2+} complexes are essentially identical. Each can be well described by only two components with lifetimes of 3.7 (3.4) ns and 8.5 (8.3) ns. Hence, the 2AP is found in only two distinct environments in these complexes and >80% of the population is free from quenching interactions. This is consistent with the large increase in fluorescence intensity observed on enzyme binding. The complete absence of sub-nanosecond decay components indicate that all of the target bases are ejected from the duplex. The essentially identical decay parameters for both duplexes demonstrate that both target bases in duplex 2/2 are flipped and experience the same extrahelical environment. This is consistent with the observation that the fluorescence intensity for PfoI bound to 2/2 is double that observed when bound to T/2.

PspGI

The fluorescence decays of the PspGI-DNA-Ca^{2+} ternary complexes resemble those of the free duplexes, in being described by four exponential components. However, the values of the decay parameters are substantially different

Table 6. Fluorescence lifetimes, τ_i and the corresponding fractional amplitudes, A_i, for the PspGI-DNA-Ca^{2+} ternary complexes

Solution	τ_1/ns	τ_2/ns	τ_3/ns	τ_4/ns	A_1	A_2	A_3	A_4
PspGI + T/2	0.20	0.74	2.6	8.8	0.55	0.34	0.07	0.04
PspGI + 2/2	0.16	0.63	2.7	8.7	0.64	0.26	0.05	0.05

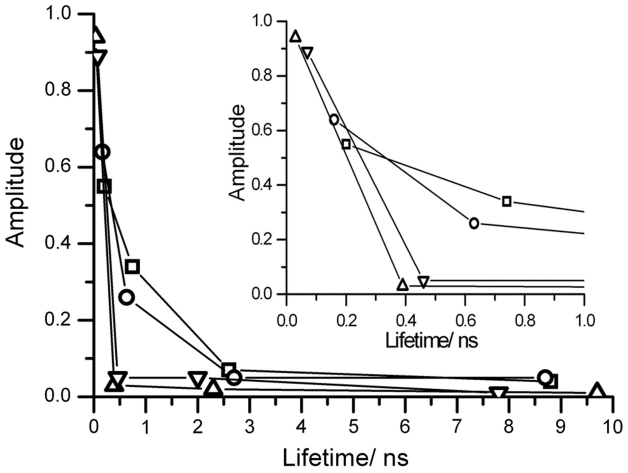


Figure 2. The decay components of the free duplexes, free T/2 (triangles) and 2/2 (inverted triangles) and the ternary complexes, PspGI-T/2-Ca^{2+} (squares) and PspGI-2/2-Ca^{2+} (circles), plotted as fractional amplitude versus fluorescence lifetime. Inset shows the same data with an expanded *x*-axis to highlight changes in the two subnanosecond decay components on enzyme binding.

from those of the free duplexes, as shown in Table 6 and illustrated graphically in Figure 2.

Binding of each duplex by PspGI results in a significant increase in the shortest lifetime, τ_1, from 30 to 200 ps for the T/2 duplex, and from 70 to 160 ps for the 2/2 duplex. The amplitude of this shortest component also decreases significantly in both the cases. These changes are indicative of destacking of 2AP from the duplex. However, the molecular environment seen by 2AP in these complexes differs markedly from that in the PfoI complexes, with 90% of the population in highly quenched conformations, described by the two subnanosecond decay times, τ_1 and τ_2.

Ecl18kI

The decays of the duplexes complexed with wt Ecl18kI are also described by four exponential components, and are rather similar to those of the free duplexes. Nevertheless, as shown in Table 7 and illustrated graphically in Figure 3, there are significant differences between the decay parameters of the bound and free duplexes.

The decay parameters of the two ternary complexes are virtually identical. Around 75% of the ternary complex population shows a very short, 70 ps, lifetime characteristic of highly quenched 2AP. In comparison, in the free duplexes, a significantly greater fraction, 90%, of the 2AP population is highly quenched. The intermediate

Table 7. Fluorescence lifetimes, τ_i and the corresponding fractional amplitudes, A_i, for the ternary complexes of T/2 and 2/2 with wt Ecl18kI and its mutants W61Y and W61A

Solution	τ_1/ns	τ_2/ns	τ_3/ns	τ_4/ns	A_1	A_2	A_3	A_4
Ecl18kI + T/2	0.07	0.77	3.1	9.0	0.74	0.09	0.11	0.06
Ecl18kI + 2/2	0.07	0.70	3.1	8.9	0.80	0.07	0.08	0.05
W61Y + T/2	0.18	0.62	1.7	7.0	0.59	0.29	0.11	0.01
W61Y + 2/2	0.17	0.53	1.5	6.1	0.55	0.29	0.15	0.01
W61A + T/2	0.14	0.91	3.8	9.2	0.61	0.12	0.08	0.19
W61A + 2/2	0.16	1.3	4.6	9.0	0.38	0.12	0.24	0.26

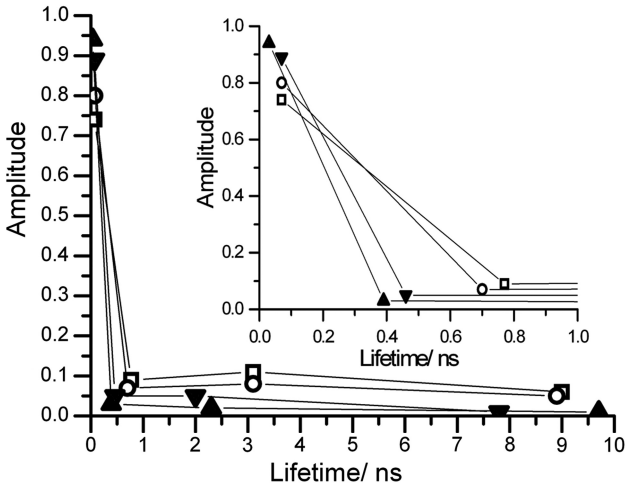


Figure 3. Plot of fractional amplitude versus fluorescence lifetime for free T/2 (filled triangles) and 2/2 (filled inverted triangles) and the ternary complexes, Ecl18kI-T/2-Ca^{2+} (squares) and Ecl18kI-2/2-Ca^{2+} (circles). Inset shows magnified view of the subnanosecond lifetime range.

lifetimes (τ_2 and τ_3) increase significantly in the ternary complexes, compared with the free duplexes, as do their amplitudes. These changes in the decay parameters are clear indicators that the DNA duplex structure is perturbed by the binding of Ecl18kI, but are not typical of base flipping. To further investigate the possibility of base flipping, the tryptophan residue (W61) within the Ecl18kI flipping pocket was mutated to W61Y and W61A variants.

As shown in Table 7, both the amplitudes and lifetimes of the 2AP decays change when the W61 residue is mutated. The W61Y mutation leads to a significant increase in the shortest lifetime, τ_1, and a decrease in its amplitude. There is a corresponding increase in the amplitude of component 2. The W61A mutation results in a general increase in lifetimes (except for the longest component). As for W61Y, the amplitude of the shortest lifetime component decreases appreciably with this mutation. For this mutant alone, the decay parameters of the ternary complexes containing the T/2 and 2/2 duplexes are quite different, particularly the amplitudes of components 1 and 3. This signifies a transfer of population from highly quenched to relatively unquenched conformations in the W61A-2/2-Ca^{2+} complex, compared with W61A-T/2-0Ca^{2+}.

Table 8. Fluorescence lifetimes, τ_i and the corresponding fractional amplitudes, A_i, for the ternary complexes of T/2 and 2/2 with EcoRII-C

Solution	τ_1/ns	τ_2/ns	τ_3/ns	τ_4/ns	A_1	A_2	A_3	A_4
EcoRII-C + T/2	0.08	0.53	2.5	8.5	0.81	0.12	0.04	0.03
EcoRII-C + 2/2	0.08	0.56	2.7	8.4	0.82	0.10	0.05	0.03

EcoRII-C

As shown in Table 8, the decay parameters of the EcoRII-C complexes are very similar to those of the wt Ecl18kI complexes (Table 7). The decays of the bound duplexes are virtually identical and 80% of the 2AP population has a very short, 80 ps, lifetime. As for Ecl18kI, there is clear evidence that duplex conformation in the vicinity of the target base is affected by binding of EcoRII-C, but the decay parameters do not conform to our expectations for base flipping.

DISCUSSION

Comparison of steady-state and time-resolved measurements

We note here a significant discrepancy between the fluorescence enhancement seen in the steady-state data and that which is readily calculated from the time-resolved fluorescence measurements (Supplementary Data). Figure 1 shows a fluorescence enhancement of the order of 1000-fold for base flipping (on the T/2 duplex) by PfoI. However, from the time-resolved data, we would predict an increase in the steady-state intensity of only ~50-fold. The discrepancy is not so large for the 2/2 duplex, which shows a steady-state intensity increase of ~130-fold and a predicted increase of ~30-fold from the time-resolved results. The disagreement between the two measurements can be explained by the occurrence of extremely rapid quenching of some fraction of the excited state 2AP that is stacked within the free DNA duplex, such that it decays on a timescale that is faster than the time resolution of our experiments and has a vanishingly small quantum yield. Hence, there is a population of non-fluorescent 2AP in the free duplexes whose quantum yields increase dramatically upon base flipping. Effectively, the concentration of 'fluorescent' 2AP is increased by base flipping. This has a dramatic effect on the steady-state intensity, but is not apparent in the time-resolved measurements, since the decay parameters are independent of the concentration of emitting molecules. Previous ultrafast spectroscopic measurements of 2AP in DNA demonstrate the existence of quenching on the picoseconds timescale (43,44). The better agreement between steady-state and time-resolved measurements on the 2/2 duplex implies a smaller population of ultra-quenched 2AP which is consistent with the mis-pairing of 2AP in this duplex.

Nucleotide flipping by PfoI

Previous time-resolved fluorescence studies of methyltransferases have shown the base-flipped complexes to

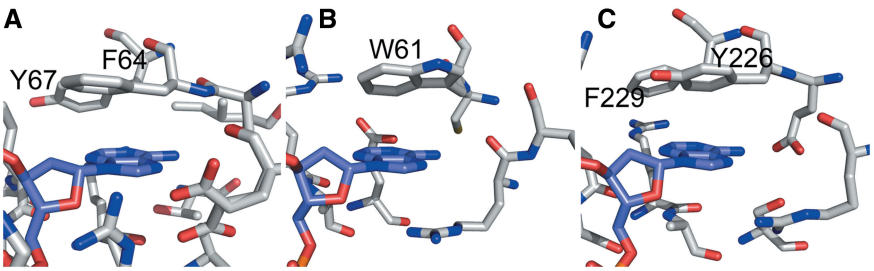


Figure 4. The flipping pockets of PspGI (**A**), Ecl18kI (**B**) and EcoRII-C (**C**) (enzymes shown in grey) with adenine (blue) in the flipping pocket. Amino acids shown contain atoms that lie within 5 Å of the adenine base.

occupy a number of conformational states, requiring at least three exponential components (usually four) to describe the 2AP decay (16–19). In these systems, base flipping is a complex and dynamic process. In comparison, the PfoI complex is extraordinary in its conformational simplicity and in the highly unquenched environment of the flipped 2AP. Base flipping by PfoI results in the complete expulsion of the entire population of target 2AP bases from the DNA duplex. The flipped 2AP base is located in an environment that is almost uniform and where its fluorescence lifetime is largely similar to that of the free 2AP-ribonucleoside. It does not make excursions into the stacked interior of the duplex and does not interact significantly with apolar enzymatic residues, which would shorten its lifetime. This evidence, together with the lack of specificity for the flipped base, suggests that the binding pocket of PfoI is large and filled with water or that the flipped base does not reside within a binding pocket but lies on the exterior of the enzyme. The apparent lack of specific interactions between enzymatic residues and the flipped base, and the completeness of flipping, despite the very low probability (<1%) of finding the target base spontaneously extrahelical, imply base flipping by PfoI to be an active process. The enzyme must actively expel the target base from the DNA duplex rather than opportunistically catching it in a transient, extrahelical state.

Nucleotide flipping by PspGI

The change in decay parameters on binding of PspGI to each duplex are clearly indicative of extraction of 2AP from its highly stacked intrahelical conformation, but the environment of the flipped base differs markedly from that in the PfoI complexes. As illustrated in Figure 4A, the co-crystal structure of PspGI with its natural substrate shows that the flipped adenine lies close to two aromatic residues, F64 and Y67 in the binding pocket (14). Previously, we have found that the fluorescence of 2AP flipped into the active site of methyltransferase TaqI is efficiently quenched by such interactions with tyrosine and phenylalanine residues, resulting in subnanosecond decay times (16). We anticipate that 2AP flipped into the binding pocket of PspGI will be subject to similar quenching and the decay parameters are consistent with this expectation. The two subnanosecond decay components, which account for 90% of the population of ternary complexes, can be ascribed to quenched conformations of 2AP in the binding pocket, together with some contribution from imperfectly stacked intrahelical states. The significant difference between the decay components of the two duplexes when complexed with this enzyme (compared with the essentially identical parameters of the complexes with PfoI) supports the notion of partitioning of the target base between extrahelical and intrahelical states.

Nucleotide flipping by wt Ecl18kI and EcoRII-C

The co-crystal structure of Ecl18kI with the natural substrate (9) shows the flipped base closely stacked with tryptophan (W61) in the binding pocket, as illustrated in Figure 4B. 2AP fluorescence is known to be rapidly quenched by electron transfer from tryptophan (45). This occurs on the picoseconds timescale, at a rate comparable with (or possibly faster than) electron transfer quenching of intrahelical 2AP by guanine (43). Thus, the observation of a 70-ps decay time for 2AP in the wild-type Ecl18kI ternary complexes in the solution is entirely consistent with a base-flipped conformation equivalent to that observed for the natural substrate in the crystalline complex. Moreover, the vast majority of the population exists in this conformation. An alternative interpretation would be that the 2AP remains entirely unflipped in a tightly stacked intrahelical state. However, the X-ray structure, the change in decay parameters on enzyme binding, the similarity of the decay parameters of both duplexes when enzyme-bound (although they are different when free) and the effect of mutation on the decay parameters (*vide infra*) constitute convincing evidence for base flipping. A contribution from intrahelical conformations cannot be ruled out, but the close similarity between the decay parameters of the two duplexes and the lack of heterogeneity suggest that this is not significant. Hence, in the wt Ecl18kI, base flipping of 2AP is an efficient process and the binding pocket of Ecl8kI provides an environment that is similar to the DNA duplex in order to stabilize the extrahelical base.

The decay parameters of the EcoRII-C complexes are very similar to those of the wt Ecl18kI complexes, with the vast majority of the emitting population showing a very short decay time (80 ps). The crystal structure of EcoRII-C with DNA (15) shows a tyrosine residue

(Y226) that coincides with the tryptophan residue (W61) in the binding pocket of Ecl18kI (Figure 4C). The oxidation potential of tyrosine (0.93 V) is slightly less than that of tryptophan (1.0 V) (46). Therefore, it can be inferred that, like Ecl18kI, EcoRII-C flips the target 2AP base into the binding pocket where it is tightly stacked with tyrosine and subject to efficient electron transfer quenching. We have not observed such rapid quenching of 2AP by tyrosine in other enzymes, such as PspGI and TaqI methyltransferase, implying that in these systems, the geometrical relationship between 2AP and tyrosine is less conducive to fast electron transfer.

Ecl18kI: the effect of mutations in the binding pocket

Changes in the structure of the binding pocket of Ecl18kI, as a result of mutation of W61, confer significant changes on the 2AP fluorescence decay. In the W61Y mutant, the replacement of the tryptophan residue in the binding pocket by tyrosine results in a substantial reduction in the efficiency of quenching of the flipped 2AP and an increase in its conformational mobility. The decay parameters are very similar to those observed for PspGI and can be interpreted in a similar fashion.

In the W16A mutant, the elimination of an interacting aromatic residue greatly reduces the quenching of the flipped 2AP, with ~20% of the population occupying an unquenched state (with 9 ns lifetime). With this lifetime, 2AP could be in an aqueous, extrahelical state, between the DNA duplex and the binding pocket, or in a hydrated environment within the flipping pocket. The W61A mutation likely creates a void in the flipping pocket that could readily fill with water molecules. The significant increase in the value of τ_2, relative to W61Y, supports the assignment of this component primarily to conformations with 2AP extrahelical, while the negligible change in τ_1 suggests a substantial intrahelical contribution to this component.

There is a high degree of conformational heterogeneity in the W61A complexes, particularly, that with duplex 2/2, in which the 2AP population is rather uniformly distributed amongst the four photophysically distinct states. It is evident that the W61A mutation weakens the interaction between the enzyme and the flipped base and there is no strong energetic preference for a particular orientation of the 2AP in the binding pocket. There is a clear difference in conformational distribution between the complexes of T/2 and 2/2 with W61A, with the former showing a greater population of intrahelical states. The final distribution of the 2AP population in these complexes is thus sensitive to the strength of the base pair that is targeted for flipping. The effects of destabilization of the interactions between the binding pocket and the flipped nucleotide, in the W61A mutant, reveal that base flipping by Ecl18kI is controlled by two factors: the stability of the base pair that is the target for flipping and the interaction of the flipped base with the residues in the enzymatic flipping pocket. Hence, consistent with our previous kinetic studies (13), we find that Ecl18kI uses a two-step mechanism to achieve nucleotide flipping; first, breaking the DNA duplex and second, capturing the extruded base.

Ecl18kI mutant W61A: relationship between base flipping and cleavage activity

The greater occupancy of intrahelical states by the target base in the W61A-T/2 complex correlates with a much reduced binding affinity for T/2 and a loss of cleavage activity (13). Complexes of both T/2 and 2/2 with W61A show much greater conformational flexibility than any of the other complexes studied here and a very low cleavage rate. Thus, there appears to be a relationship between the tightness of binding of the flipped base in the enzyme pocket and the cleavage rate. The cleavage activity of the W61A mutant of Ecl18kI on a DNA duplex containing abasic sites at the target sites for flipping is near to that of the wild-type enzyme (13). Hence, base capture is not critical for catalysis to proceed. Yet, the dramatic loss of activity that results from introducing the W61A mutation seems contradictory to this argument. Thus, we propose a model for flipping where the initial disruption of the base pair is a transient, reversible event, temporarily expelling the target base(s) from the DNA duplex. Further to our previous predictions (13), the time-resolved results now show definitively that the base capture in the W61A mutant of Ecl18kI is an inefficient process. In the absence of a binding pocket that is capable of base capture, the transiently flipped base readily returns to the DNA duplex. DNA cleavage is slow relative to the rapid movement of the transiently flipped base to and from the duplex and only takes place once a stable nucleotide flip, in both strands of the duplex, and subsequent collapse of the DNA duplex is achieved.

Ecl18kI mutant W61Y: altered specificity results from unfavourable interactions with thymine in the binding pocket

The similar fluorescence decay parameters of the ternary W61Y-DNA-Ca^{2+} complexes are surprising given the markedly different cutting activities of the W61Y variants on the duplexes examined (~100-fold less on the T/2 duplex as compared with the 2/2 duplex) (13). In fact, Tamulaitis *et al.* (13) have previously noted that this mutant shows altered specificity compared with wt Ecl18kI, preferring not to cut on duplexes containing a thymine at the flipping site, despite strong binding of these substrates. The time-resolved fluorescence results confirm that the disparity in cutting activity of the W61Y mutant on the T/2 and 2/2 duplexes is not due to different interactions of the enzyme with the 2AP bases and, therefore, is due to its treatment of the flipped thymine base. Further, the flipped 2AP is in an identical, environment for both of the complexed duplexes. Since cleavage is slow in the T/2 duplex, we infer that, despite stable 2AP flipping, the thymine base is not stably flipped by the W61Y mutant enzyme. Hence, in this complex, with the T/2 duplex, the two Ecl18kI monomeric units interact with their target bases in quite different ways. This clearly demonstrates that the two units of the

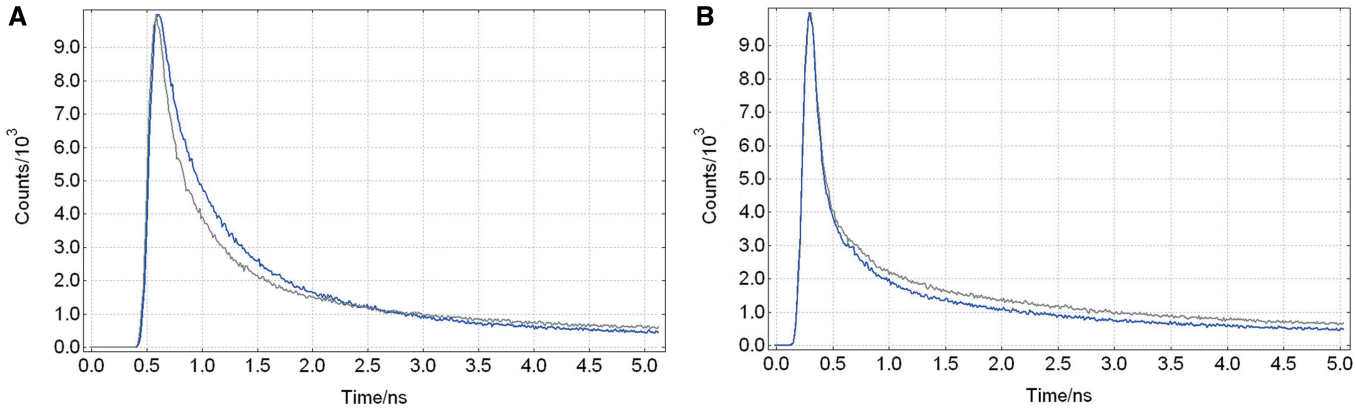


Figure 5. The first 5 ns of the fluorescence lifetime decays of PspGI (**A**) and EcoRII-C (**B**) bound in ternary complexes to the T/2 (blue) and 2/2 (grey) duplexes.

dimeric Ecl18kI complex are capable of independently producing stable flips of their target nucleotides. Regarding the question of cooperativity between the monomers in the base flipping process, we cannot say whether the stable flipping of 2AP influences the probability of a thymine flip, but it is evident that the unstable flip of thymine certainly does not decrease the likelihood of a stable 2AP flip.

PspGI and EcoRII-C specificity at the flipping site

Both PspGI and EcoRII-C target the same recognition sequence, i.e. 5′-CCWGG-3′ (where W stands for A or T). The crystal structures of PspGI and EcoRII-C complexed with their cognate DNA show that both enzymes flip the central bases of this recognition sequence (14,15), as shown in Figure 4A and C.

Unlike the Ecl18kI binding pocket, in the PspGI and EcoRII-C pockets, there are several amino acids that form 'walls' to the pocket. These amino acids may be necessary for the recognition of the A/T bases over G/C bases. Tamulaitis *et al.* (13) previously proposed that a 'double-check' mechanism is used by PspGI and EcoRII-C during the cleavage of their substrates. In this model, the strength of the target base pair for flipping is critical for stable binding of the enzymes to the DNA. Prior to cleavage, the enzyme makes a second check of the flipped base identity once it is located within its flipping pocket. The time-resolved measurements focus solely on the second step of this check, the recognition tied to flipping, since we employ conditions that ensure complete binding of the DNA. One key question with regard to the second check, is whether this is achieved through differential affinities for bases within the binding pocket, leading to more or less flipping of the target base, or whether all bases are strongly bound in the binding pocket but that binding of non-target G/C bases occurs in such a way as to prevent catalysis from occurring.

The fluorescence decays of the PspGI-T/2-Ca^{2+} and PspGI-2/2-Ca^{2+} complexes are dissimilar on the short timescale (<1 ns), especially when compared with those for the analogous EcoRII-C complexes, as shown in Tables 6 and 8 and Figure 5. There is a greater proportion of the shortest lifetime in the PspGI-2/2-Ca^{2+} decay, as compared with the PspGI-T/2-Ca^{2+} complex, which indicates a greater proportion of intrahelical 2AP in the 2/2 complex. PspGI cleaves the 2/2 duplex ~20-fold slower than it does the T/2 duplex (13). This is consistent with the relationship between base flipping and cleavage activity that was discussed in the context of Ecl18kI (see above). This seems counter-intuitive when considering the relatively weak 2AP-2AP base pair in the 2/2 duplex, compared with the Watson–Crick-type T-2AP base pair of the T/2 duplex. It appears that the conformational distribution is determined, not by the strength of the base pair of the target bases but, rather, through the interaction of the flipped base with the PspGI. This interaction overrides the effect of the base pair strength on the overall dynamics of the system and favours more intrahelical 2AP in the 2/2 duplex than in the T/2 duplex. Since PspGI cleaves duplexes containing only A/T at the target site for flipping it is probable that discrimination against flipping of 2AP results from the presence of the 2-amino group that 2AP has in common with guanine.

The time-resolved fluorescence results provide evidence of two important features of the second step of the double-check mechanism used by PspGI for recognition of its target base for flipping. First, 2AP is less firmly held within the PspGI binding pocket than in Ecl18kI or EcoRII-C; the target 2AP bases are flipped but only transiently, as a result of base recognition by PspGI. Second, the differential treatment of the flipped 2AP in the two complexes investigated illustrates that there is apparent communication between the two subunits of the dimeric PspGI complex. That is to say, that the stable flipping of thymine, which is a natural target base for flipping by PspGI, results in a different distribution of the 2AP population in the T/2-containing complex as compared with the transient flipping of the two 2AP bases of the 2/2 duplex. Whether this effect is the result of DNA duplex collapse, whose likelihood is increased by the stable flipping of thymine, or of some direct enzyme-enzyme contact is uncertain.

In contrast to PspGI, the catalytic subunit of EcoRII-C appears to have no mechanism by which to distinguish 2AP

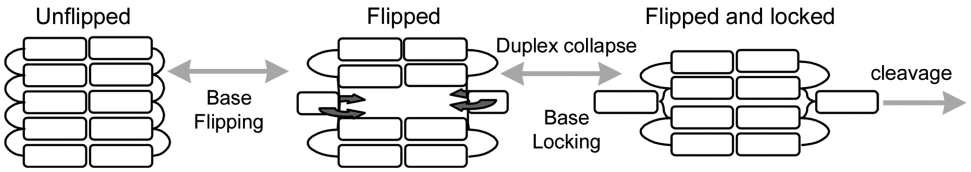


Figure 6. Schematic representation of the DNA bases (white rectangles) during the flipping and locking steps that precede DNA cleavage. The target base can be unflipped (as largely seen in the Ecl18kI W61A complex), in a highly dynamic flipped (but not locked) state (predominant in the PspGI complex) or locked in the enzyme's flipping pocket (as in wt Ecl18kI and EcoRII-C complexes).

from A or T. The decay parameters of the EcoRII-C-T/2-Ca^{2+} and EcoRII-C-2/2-Ca^{2+} complexes are remarkably similar and in both the cases flipping and binding of the flipped 2AP are efficient. Furthermore, binding affinities and cleavage rates of the 2AP-containing duplexes are similar to those on duplexes containing the natural EcoRII-C recognition sequence (13).

SUMMARY

Base flipping and DNA cleavage by all of the enzymes studied go hand-in-hand. The general mechanism employed for DNA recognition and cleavage is shown schematically in Figure 6.

For each of the enzymes studied, their target bases must be flipped from the duplex before DNA duplex collapse occurs, which is necessary for catalysis to proceed. However, we have shown that there are significant differences in the mechanisms used by this group of enzymes to produce a base-flip that is sufficiently stable to allow the collapse of the DNA duplex. Table 9 compares the key mechanistic attributes of the enzymes.

We have shown that PfoI flips the base into an environment where enzymatic contacts to the base are minimal. In contrast, the other enzymes we investigated flip the base into an orientation that affords interaction with aromatic residues in the binding pocket. In Ecl18kI and EcoRII-C, the flipped 2AP is stacked particularly tightly with tryptophan or tyrosine, respectively, and sees an environment that closely resembles that within the DNA duplex and in which its fluorescence is highly quenched. Regardless of the close functional relationship between these enzymes, the flipping pocket of PfoI is quite distinct from that of the other Ecl18kI-type enzymes (i.e. Ecl18kI, EcoRII-C and PspGI).

Studies of W61 mutants of Ecl18kI have shown that base flipping is dependent on both the strength of the base pair that is the target for flipping and the ability of the enzyme to capture the flipped base. These studies have also demonstrated that base flipping is part of a concerted catalytic cycle and must be complete before DNA cleavage can proceed. Both the bases of the DNA duplex must be flipped and stably bound within the flipping pocket before cutting takes place. The linking of these processes may provide an opportunity for rational design of the base selectivity at the flipping position in the future.

PspGI discriminates between A/T and G/C bases once the bases are flipped from the duplex. Flipping of target 2AP bases is a dynamic process, the 2AP being shuttled between the DNA duplex and the enzyme's flipping pocket. The presence of the 2-amino group in 2AP, as in guanine, is a likely cause of the reduced affinity of the binding pocket for the flipped base.

Table 9. A comparison of the mechanistic behaviour of the enzymes investigated

Enzyme	Does 2AP occupy the given state in the ternary complex?			Active flipping from duplex?	Flipping pocket capable of base capture?	Do enzymes show cooperativity to capture flipped bases?
	UF[a]	F	FL			
PfoI	No	Yes		Yes	Possibly	Not applicable
PspGI	No	Yes	Yes	Unknown	Yes	Yes
wt-Ecl18kI	No	No	Yes	Yes	Yes	Unlikely
W61Y	No	No	Yes	Yes	Yes	Unlikely
W61A	Yes	Yes	No	Yes	No	Not applicable
EcoRII-C	No	No	Yes	Unknown	Yes	Unknown

[a]UF, Unflipped; F, flipped; FL, flipped and locked as illustrated in Figure 6.

EcoRII-C behaves in a similar way to wt Ecl18kI, in that base flipping of 2AP appears to be an efficient process that results in similar 2AP populations in complexes containing either the T/2 or 2/2 duplexes. Hence, EcoRII-C does not distinguish 2AP from A or T, implying that it is not sensitive to the presence of the 2-amino group.

Time-resolved fluorescence measurements of 2AP provide a means by which to examine the heterogeneity of the highly dynamic base flipping process, an insight that is not achievable using either crystallographic or ensemble kinetic measurements. However, the time-resolved measurements are used to greatest effect in combination with crystallographic, mutational and kinetic analyses and using this approach we have begun to build a remarkably detailed understanding of the structural and dynamic behaviour of the known base-flipping restriction enzymes.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Dr David Dryden, Dr Patricia Richardson and Dr Mindaugas Zaremba for

helpful discussions and comments on the manuscript. We thank New England Biolabs for the PspGI clone, Dr Monika Reuter (Institute of Virology, Berlin, Germany) for the EcoRII clone and Fermentas UAB for making the PfoI clone available. We are also grateful to Dr Sonata Jurenaite-Urbanaviciene for PfoI cloning.

FUNDING

Engineering and Physical Sciences Research Council (EP/C53543X/1 postdoctoral fellowship at the life-sciences interface to R.K.N.); the FP6 Marie Curie Research Training Networks grant 'DNA Enzymes'; the Lithuanian State Science and Studies Foundation (T-14/07 and T-27/08); the European Commission (Marie Curie Intra-European Fellowship to R.K.N.]; EastChem [studentship to K.C.]. Funding for open access charge: Katholieke Universiteit Leuven.

Conflict of interest statement. None declared.

REFERENCES

1. Klimasauskas,S., Kumar,S., Roberts,R.J. and Cheng,X.D. (1994) Hhal methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–369.
2. Reinisch,K.M., Chen,L., Verdine,G.L. and Lipscomb,W.N. (1995) The crystal-structure of HaeIII methyltransferase covalently complexed to DNA – an extrahelical cytosine and rearranged base-pairing. *Cell*, **82**, 143–153.
3. Goedecke,K., Pignot,M., Goody,R.S., Scheidig,A.J. and Weinhold,E. (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.*, **8**, 121–125.
4. Horton,J.R., Liebert,K., Hattman,S., Jeltsch,A. and Cheng,X.D. (2005) Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of Dam methyltransferase. *Cell*, **121**, 349–361.
5. Yang,Z., Horton,J.R., Zhou,L., Zhang,X.J., Dong,A.P., Zhang,X., Schlagman,S.L., Kossykh,V., Hattman,S. and Cheng,X.D. (2003) Structure of the bacteriophage T4 DNA adenine methyltransferase. *Nat. Struct. Biol.*, **10**, 849–855.
6. Carpenter,M.A. and Bhagwat,A.S. (2008) DNA base flipping by both members of the PspGI restriction-modification system. *Nucleic Acids Res.*, **36**, 5417–5425.
7. Hosfield,D.J., Guan,Y., Haas,B.J., Cunningham,R.P. and Tainer,J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*, **98**, 397–408.
8. Banerjee,A., Yang,W., Karplus,M. and Verdine,G.L. (2005) Structure of a repair enzyme interrogating undamaged DNA elucidates recognition of damaged DNA. *Nature*, **434**, 612–618.
9. Bochtler,M., Szczepanowski,R.H., Tamulaitis,G., Grazulis,S., Czapinska,H., Manakova,E. and Siksnys,V. (2006) Nucleotide flips determine the specificity of the Ecl18kl restriction endonuclease. *EMBO J.*, **25**, 2219–2229.
10. Carpenter,M., Divvela,P., Pingoud,V., Bujnicki,J. and Bhagwat,A.S. (2006) Sequence-dependent enhancement of hydrolytic deamination of cytosines in DNA by the restriction enzyme PspGI. *Nucleic Acids Res.*, **34**, 3762–3770.
11. Tamulaitis,G., Zaremba,M., Szczepanowski,R.H., Bochtler,M. and Siksnys,V. (2007) Nucleotide flipping by restriction enzymes analyzed by 2-aminopurine steady-state fluorescence. *Nucleic Acids Res.*, **35**, 4792–4799.
12. Daujotyte,D., Liutkeviciute,Z., Tamulaitis,G. and Klimasauskas,S. (2008) Chemical mapping of cytosines enzymatically flipped out of the DNA helix. *Nucleic Acids Res.*, **36**, e57.
13. Tamulaitis,G., Zaremba,M., Szczepanowski,R.H., Bochtler,M. and Siksnys,V. (2008) How PspGI, catalytic domain of EcoRII and Ecl18kI acquire specificities for different DNA targets. *Nucleic Acids Res.*, **36**, 6101–6108.
14. Szczepanowski,R.H., Carpenter,M.A., Czapinska,H., Zaremba,M., Tamulaitis,G., Siksnys,V., Bhagwat,A.S. and Bochtler,M. (2008) Central base pair flipping and discrimination by PspGI. *Nucleic Acids Res.*, **36**, 6109–6117.
15. Golovenko,D., Manakova,E., Tamulaitiene,G., Grazulis,S. and Siksnys,V. (2009) Structural mechanisms for the 5′-CCWGG sequence recognition by the N- and C-terminal domains of EcoRII. *Nucleic Acids Res.* doi: 10.1093/nar/gkp699.
16. Lenz,T., Bonnist,E.Y.M., Pljevaljcic,G., Neely,R.K., Dryden,D.T.F., Scheidig,A.J., Jones,A.C. and Weinhold,E. (2007) 2-Aminopurine flipped into the active site of the adenine-specific DNA methyltransferase M.Taql: crystal structures and time-resolved fluorescence. *J. Am. Chem. Soc.*, **129**, 6240–6248.
17. Neely,R.K., Daujotyte,D., Grazulis,S., Magennis,S.W., Dryden,D.T.F., Klimasauskas,S. and Jones,A.C. (2005) Time-resolved fluorescence of 2-aminopurine as a probe of base flipping in M.HhaI-DNA complexes. *Nucleic Acids Res.*, **33**, 6953–6960.
18. Youngblood,B., Bonnist,E., Dryden,D.T.F., Jones,A.C. and Reich,N.O. (2008) Differential stabilization of reaction intermediates: specificity checkpoints for M.EcoRI revealed by transient fluorescence and fluorescence lifetime studies. *Nucleic Acids Res.*, **36**, 2917–2925.
19. Hariharan,C. and Reha-Krantz,L.J. (2005) Using 2-Aminopurine fluorescence to detect bacteriophage T4 DNA polymerase−DNA complexes that are important for primer extension and proofreading reactions. *Biochemistry*, **44**, 15674–15684.
20. Tamulaitis,G., Solonin,A.S. and Siksnys,V. (2002) Alternative arrangements of catalytic residues at the active sites of restriction enzymes. *FEBS Lett.*, **518**, 17–22.
21. Gaigalas,M., Maneliene,Z., Kazlauskiene,R., Petrusyte,M. and Janulaitis,A. (2002) PfoI, a unique type II restriction endonuclease that recognises the sequence 5′-T/CCNGGA-3′. *Nucleic Acids Res.*, **30**, e98.
22. Tamulaitis,G., Mucke,M. and Siksnys,V. (2006) Biochemical and mutational analysis of EcoRII functional domains reveals evolutionary links between restriction enzymes. *FEBS Lett.*, **580**, 1665–1671.
23. Zaremba,M., Sasnauskas,G., Urbanke,C. and Siksnys,V. (2006) Allosteric communication network in the tetrameric restriction endonuclease Bse634I. *J. Mol. Biol.*, **363**, 800–812.
24. Pingoud,A. and Jeltsch,A. (2001) Structure and function of type II restriction endonucleases. *Nucleic Acids Res.*, **29**, 3705–3727.
25. Allan,B.W. and Reich,N.O. (1996) Targeted base stacking disruption by the EcoRI DNA methyltransferase. *Biochemistry*, **35**, 14757–14762.
26. McCullough,A.K., Dodson,M.L., Scharer,O.D. and Lloyd,R.S. (1997) The role of base flipping in damage recognition and catalysis by T4 endonuclease V. *J. Biol. Chem.*, **272**, 27210–27217.
27. Holz,B., Klimasauskas,S., Serva,S. and Weinhold,E. (1998) 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases. *Nucleic Acids Res.*, **26**, 1076–1083.
28. Stivers,J.T., Pankiewicz,K.W. and Watanabe,K.A. (1999) Kinetic mechanism of damage site recognition and uracil flipping by Escherichia coli uracil DNA glycosylase. *Biochemistry*, **38**, 952–963.
29. Gowher,H. and Jeltsch,A. (2000) Molecular enzymology of the Eco RV DNA-(adenine-N6)-methyltransferase: kinetics of DNA binding and bending, kinetic mechanism and linear diffusion of the enzyme on DNA. *J. Mol. Biol.*, **303**, 93–110.
30. Reddy,Y.V.R. and Rao,D.N. (2000) Binding of EcoP15I DNA methyltransferase to DNA reveals a large structural distortion within the recognition sequence. *J. Mol. Biol.*, **298**, 597–610.
31. Su,T.J., Connolly,B.A., Darlington,C., Mallin,R. and Dryden,D.T.F. (2004) Unusual 2-aminopurine fluorescence from a complex of DNA and the EcoKI methyltransferase. *Nucleic Acids Res.*, **32**, 2223–2230.
32. Liebert,K., Hermann,A., Schlickenrieder,M. and Jeltsch,A. (2004) Stopped-flow and mutational analysis of base flipping by the Escherichia coli Dam DNA-(adenine-N6)-methyltransferase. *J. Mol. Biol.*, **341**, 443–454.
33. Szegedi,S.S., Reich,N.O. and Gumport,R.I. (2000) Substrate binding in vitro and kinetics of RsrI [N6-adenine] DNA methyltransferase. *Nucleic Acids Res.*, **28**, 3962–3971.

34. Malygin,E.G., Evdokimov,A.A., Zinoviev,V.V., Ovechkina,L.G., Lindstrom,W.M., Reich,N.O., Schlagman,S.L. and Hattman,S. (2001) A dual role for substrate S-adenosyl-L-methionine in the methylation reaction with bacteriophage T4 Dam DNA-[N6-adenine]-methyltransferase. *Nucleic Acids Res.*, **29**, 2361–2369.
35. Christine,K.S., MacFarlane,A.W. IV, Yang,K. and Stanley,R.J. (2002) Cyclobutylpyrimidine Dimer Base Flipping by DNA Photolyase. *J. Biol. Chem.*, **277**, 38339–38344.
36. Malta,E., Moolenaar,G.F. and Goosen,N. (2006) Base flipping in nucleotide excision repair. *J. Biol. Chem.*, **281**, 2184–2194.
37. Walker,R.K., McCullough,A.K. and Lloyd,R.S. (2006) Uncoupling of nucleotide flipping and DNA bending by the T4 pyrimidine dimer DNA glycosylase. *Biochemistry*, **45**, 14192–14200.
38. Rachofsky,E.L., Seibert,E., Stivers,J.T., Osman,R. and Ross,J.B.A. (2001) Conformation and dynamics of abasic sites in DNA investigated by time-resolved fluorescence of 2-aminopurine. *Biochemistry*, **40**, 957–967.
39. O'Neill,M.A., Dohno,C. and Barton,J.K. (2004) Direct chemical evidence for charge transfer between photoexcited 2-aminopurine and guanine in duplex DNA. *J. Am. Chem. Soc.*, **126**, 1316–1317.
40. Kelley,S.O. and Barton,J.K. (1999) Electron transfer between bases in double helical DNA. *Science*, **283**, 375–381.
41. Rachofsky,E.L., Osman,R. and Ross,J.B.A. (2001) Probing structure and dynamics of DNA with 2-aminopurine: effects of local environment on fluorescence. *Biochemistry*, **40**, 946–956.
42. Neely,R.K., Magennis,S.W., Dryden,D.T.F. and Jones,A.C. (2004) Evidence of tautomerism in 2-aminopurine from fluorescence lifetime measurements. *J. Phys. Chem. B.*, **108**, 17606–17610.
43. Fiebig,T., Wan,C.Z. and Zewail,A.H. (2002) Femtosecond charge transfer dynamics of a modified DNA base: 2- aminopurine in complexes with nucleotides. *ChemPhysChem*, **3**, 781–788.
44. O'Neill,M.A., Becker,H.C., Wan,C.Z., Barton,J.K. and Zewail,A.H. (2003) Ultrafast dynamics in DNA-mediated electron transfer: Base gating and the role of temperature. *Angew. Chem. Int. Ed.*, **42**, 5896–5900.
45. Xia,T.B., Becker,H.C., Wan,C.Z., Frankel,A., Roberts,R.W. and Zewail,A.H. (2003) The RNA-protein complex: direct probing of the interfacial recognition dynamics and its correlation with biological functions. *Proc. Natl Acad. Sci. USA*, **100**, 8119–8123.
46. Harriman,A. (1987) Further comments on the redox potentials of tryptophan and tyrosine. *J. Phys. Chem.*, **91**, 6102–6104.

Biochemical and Biophysical Research Communications 398 (2010) 254–259

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

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

Fusion of GFP to the M.EcoKI DNA methyltransferase produces a new probe of Type I DNA restriction and modification enzymes

Kai Chen, Gareth A. Roberts, Augoustinos S. Stephanou, Laurie P. Cooper, John H. White, David T.F. Dryden *

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

ARTICLE INFO

Article history:
Received 12 June 2010
Available online 19 June 2010

Keywords:
DNA restriction/modification
DNA methyltransferase
Forster resonance energy transfer
Time-resolved fluorescence anisotropy
Time-resolved fluorescence
Green fluorescent protein

ABSTRACT

We describe the fusion of enhanced green fluorescent protein to the C-terminus of the HsdS DNA sequence-specificity subunit of the Type I DNA modification methyltransferase M.EcoKI. The fusion expresses well *in vivo* and assembles with the two HsdM modification subunits. The fusion protein functions as a sequence-specific DNA methyltransferase protecting DNA against digestion by the EcoKI restriction endonuclease. The purified enzyme shows Förster resonance energy transfer to fluorescently-labelled DNA duplexes containing the target sequence and to fluorescently-labelled ocr protein, a DNA mimic that binds to the M.EcoKI enzyme. Distances determined from the energy transfer experiments corroborate the structural model of M.EcoKI.

1. Introduction

Since their introduction into genetic engineering, the green fluorescent protein (GFP) and its many spectral variants have proved to be extraordinarily useful probes of protein structure and function both *in vitro* and *in vivo* [1]. In particular, Förster resonance energy transfer (FRET) to measure distances between two fluorophores, a donor and an acceptor, has been the subject of many uses of GFP despite its complex photophysics and its relatively large size compared to more traditional small molecule fluorophores such as fluorescein [2].

Sequence-specific DNA-binding enzymes such as methyltransferases (MTases) and endonucleases comprising bacterial restriction–modification (R/M) systems would seem to present excellent targets for analysis via fusion to GFP given that many of them introduce complex rearrangements of DNA structure including for example DNA looping to bring distant sites on a single DNA molecule into close proximity. However, as yet few investigations of R/M systems have utilised these versatile fluorescent probes [3].

Bacterial host restriction endonucleases (REase) attack invading foreign DNA lacking the imprinted modification pattern characteristic of the host DNA [4]. R/M systems typically comprise a REase that recognises a specific nucleotide sequence prior to cleavage, and a cognate DNA MTase able, by methylating adenine or cytosine within the same sequence, to confer protection from the REase. The REase cuts unmethylated DNA but not hemimethylated DNA, the substrate for the MTase. R/M systems are classified according to their subunit composition, recognition site, cofactor requirement and DNA cleavage position. The R/M systems display an extraordinary diversity in structure and activity leading to four distinct groupings [5]. The most common R/M systems are the Type II R/M systems, which primarily consist of separate MTase and REase enzymes that recognise 4–8 base pair (bp) palindromic sequences.

In contrast, Type I R/M enzymes [4] such as EcoKI are complex hetero-oligomers of two REase (HsdR) subunits, two MTase (HsdM) subunits and one DNA sequence-specificity (HsdS) subunit. Depending on the methylation status of the DNA substrate, this complex functions as either a REase or an MTase. These enzymes recognise an asymmetric, bipartite sequence (13–15 bp) and require ATP to affect cleavage at a distant site. Over 600 confirmed and putative Type I R/M systems are known and they appear to be as widely spread in bacteria as the Type II R/M systems [6]. The complex of 2 HsdM and 1 HsdS, M_2S_1, forms an active MTase, M.EcoKI, and is the core part of the Type I R/M enzyme. The M.EcoKI MTase recognises the sequence AACNNNNNNGTGC and the methylation status of the adenines at the underline locations. A detailed structural model of M.EcoKI in complex with DNA has recently been proposed based upon electron microscopy of the complex and crystallographic structures of the individual subunits [7].

The genes for R/M systems are found in virtually every sequenced bacterial and archaeal genome and many genomes contain multiple R/M systems [6] often with the capability to switch between different systems and DNA specificities depending upon conditions [8,9]. R/M systems are also extensively represented

* Corresponding author. Fax: +44 (0)131 650 6453.
E-mail address: david.dryden@ed.ac.uk (D.T.F. Dryden).

0006-291X/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.bbrc.2010.06.069

within clinical strain collections such as the *Escherichia coli* ECOR collection [10]. Given that resident R/M systems limit phage propagation in a bacterial population by factors reaching 10^8 (for EcoKI), there is a huge evolutionary pressure on mobile genetic elements such as phage and conjugative plasmids and transposons to evolve 'anti-restriction' counter measures including, for example, the acquisition of proteins which inhibit DNA binding by the R/M enzymes [4]. These inhibitors are structural and electrostatic mimics of double stranded DNA with the gene 0.3 protein, ocr, from phage T7 and the ArdA protein from conjugative Tn*916* mimicking 24 base pairs and 42 base pairs, respectively [11,12]. Their tight binding to M.EcoKI physically fills the DNA binding groove on the enzyme resulting in the inactivation of the R/M system [13–17].

In this paper we demonstrate the preparation of an active M.EcoKI fused to GFP and measure via FRET the distance from the GFP to a HEX label on a duplex bound to the MTase and to a fluorescently-labelled ocr protein bound to the MTase. These distances are then compared to predictions from the structural model [7].

2. Materials and methods

2.1. Plasmid pJFMSEGFP for production of GFP-MTase

The expression construct is derived from pJFMS [18] and pEGFP-N1 (Clonetech) as detailed in supplementary information. This plasmid was named pJFMSEGFP and we call the protein GFP-MTase.

2.2. In vivo activity

pJFMSEGFP and control plasmid pBIO2 were introduced into the r^-m^- mutant, *E. coli* NM1261(DE3). This strain contains a mutation in *hsdS*. pBIO2 is a non-functional derivative of pJFMS lacking the entire *hsdM* and half of *hsdS*. In NM1261(DE3), function of the MTase was dependent upon the plasmid-encoded HsdS forming a complex with HsdM encoded on the chromosome and the plasmid. Bacteriophage lambda virulent containing unmodified EcoKI sites:$\lambda_{v.o}$) was plated on NM1261(DE3) pJFMSEGFP and plaques were picked for assay against the EcoKI tester stains *E. coli* NM1049(DE3) (r^+m^+) and NM1261(DE3). Serial dilutions of plaques resuspended in phage buffer were spotted in 10 μl aliquots on the tester strains plated on BBL top agar supplemented with carbenicillin, 100 μg/ml. Titres were scored after overnight incubation at 37 °C [19]. Note that heterologous gene expression was not induced by addition of IPTG in these experiments but instead relied upon leaky expression from the promoter.

2.3. Purification of GFP-MTase

GFP-MTase was purified to homogeneity after overexpression in *E. coli* BL21(DE3) cells using His-tag affinity, gel filtration and anion exchange chromatography as detailed in supplementary information. The protein occurred in both M_1S_1 and M_2S_1 forms as found for the native protein [18] with the M_2S_1 form being used in further experiments. All subsequent measurements were performed at 20 or 25 °C in 20 mM Tris–HCl pH 8.0, 6 mM $MgCl_2$, 7 mM 2-mercaptoethanol supplemented with NaCl when stated.

2.4. DNA binding activity in vitro

DNA binding was measured using FRET and employed 21 base pair duplexes labelled at their 5′ ends with hexachlorofluorescein (HEX). The interaction of these duplexes with M.EcoKI has been previously analysed using fluorescence anisotropy [15]. Two duplexes were used: 21TH21B has the top "21TH" strand sequence 5′-HEX-GCC TAA CCA CGT GGT GCG TAC-3′ with the complementary unlabelled bottom strand ("21B") and 21T21BH has the same sequence but the HEX label is on the 5′ end of the bottom "21BH" strand.

A range of solutions containing GFP-MTase from 0 to 200 nM and NaCl concentrations of 0, 25, 50 and 100 mM, were prepared. In addition, solutions containing different proportions of GFP-MTase and 21TH21B, where the sum of the concentration of the two components was 200 nM, were prepared. The emission spectrum of each solution was then recorded and the intensity of the emission peak plotted against the mole fraction of GFP-MTase after subtracting the intensity of the GFP-MTase alone. The binding affinity was determined using the continuous variation method [20].

2.5. Preparation of ocr mutant proteins and their interaction with GFP-MTase

Site directed mutagenesis and protein purification was performed as described previously [17] to create the single substitutions, E20C, S68C and E117C in the ocr protein. 1 ml samples of 10 μM of each mutant ocr protein (assuming an ocr dimer) were incubated overnight at 4 °C in the dark with a 20-fold molar excess of Dylight549 Maleimide (Molecular Probes) in 100 mM sodium phosphate buffer, 150 mM NaCl, 1 mM EDTA, pH 7.2. Unreacted probe was removed by extensive dialysis. The concentration of Dylight549 bound to the ocr dimer was calculated from absorption using a molar extinction coefficient of 150 000 $M^{-1}\,cm^{-1}$ at 562 nm. The concentrations of all ocr mutant proteins were calculated using a molar extinction coefficient of 31,860 $M^{-1}\,cm^{-1}$ at 280 nm for the ocr dimer [13]. The concentrations of the labelled proteins were calculated from the absorption spectra at 280 nm after subtracting the Dylight549 absorbance at this wavelength (12,150 $M^{-1}\,cm^{-1}$). A comparison of the concentration of Dylight549 with the concentration of ocr then allowed the degree of labelling to be calculated. Labeling levels of 81.5%, 77.0% and 86.3% were achieved for E20C, S68C and E117C mutant ocr proteins, respectively.

Binding of the labelled mutant ocr proteins to the GFP-MTase was assessed using size exclusion chromatography as previously described [14].

2.6. Fluorescence measurements

Steady state fluorescence intensity measurements were performed on an Edinburgh Instruments FS900 spectrofluorometer (Edinburgh Instruments) with a 5 nm bandwidth. The cuvette path lengths were 3 mm.

Time correlated single photon counting was performed with a home built time-resolved fluorimeter equipped with an Edinburgh Instruments TCC900 single photon counting card, 465 nm or 500 nm pulsed LED driven by a PDL 800-B pulsed diode laser driver (PicoQuant Gmbh) and a PMH-100-3 single photon counting photomultiplier tube (Becker & Hickl Gmbh). A 405 nm pulsed laser (Edinburgh Instruments) was also sometimes used. Emission wavelengths were selected with a monochromator. Polarisation was applied using quartz Glan–Thompson polarisers. Excitation pathlengths were 10 mm and the emission bandpass was 20 nm. Fluorescence decays were fitted using a multiexponential decay equation with the minimum number of decay components required to obtain a χ^2 value close to 1. Anisotropy decays were fitted to Eq. (1).

$$r(t) = r_\infty + r_o \exp(-t/\phi) \quad (1)$$

where $r(t)$ is the anisotropy value at time t, r_o is the initial anisotropy, r_∞ is the anisotropy value at infinite time and ϕ is the rotational correlation time.

2.7. FRET calculations

The Förster distance for 50% transfer efficiency (R_0) for GFP to HEX or Dylight549 was calculated on the basis of Eq. (2). [21]

$$R_0^6 = 8.78 * 10^{-5} \kappa^2 \Phi J n^{-4} \quad (2)$$

where n is the refractive index of the medium (n = 1.33), the orientation factor (κ^2) was considered to be two-thirds on the assumption that the donor and acceptor can adopt random conformations, the quantum yield of GFP was Φ = 0.8. The spectral overlap integral, J, between the donor emission spectrum and the acceptor absorbance spectrum was determined by using Eq. (3),

$$J(\lambda) = \int F_d(\lambda)\epsilon_a(\lambda)\lambda 4 d\lambda / \int F_d(\lambda) d\lambda \quad (3)$$

where $F_d(\lambda)$ and $\varepsilon_a(\lambda)$ represent the fluorescence intensity of the donor and the molar extinction coefficient of the acceptor, respectively, at wavelength λ.

The efficiency of the energy transfer was calculated based on the decrease in the donor (GFP) fluorescence intensity, Eq. (4).

$$E = 1 - (F_a/F_d) \quad (4)$$

where F_a and F_d represent the donor fluorescence intensity measured in the absence and presence of acceptor, respectively.

The efficiency of the energy transfer was also calculated from the decrease in the fluorescence lifetime of the donor (GFP) fluorescence, Eq. (5).

$$E = 1 - (\tau_a/\tau_d) \quad (5)$$

where τ_a and τ_d are, respectively, the fluorescence lifetime in the absence and presence of acceptor.

3. Results

3.1. Protein overexpression

The structural model [9] predicts that the C-terminus of HsdS should be exposed to solvent and would therefore present a suitable site for fusion to the N-terminus of GFP. This fusion gene construct (pJFMSEGFP) was engineered and produced large amounts of GFP-MTase, which could be purified to homogeneity (see supplementary Figure S1).

3.2. In vivo activity

We tested whether the fusion had any effect on the activity of M.EcoKI *in vivo* using phage lambda. Expression of the GFP-MTase was sufficient to modify the five EcoKI target sites on $\lambda_{v.o}$ as shown by the survival of these phage, when passaged through *E. coli* NM1261(DE3) pJFMSEGFP, on an EcoKI restriction proficient strain, NM1149(DE3), or a restriction deficient strain, NM1261(DE3), Table 1. The titre of phage isolated from NM1261(DE3) pJFMSEGFP, was the same on both the restricting and non-restricting strain. Thus the fusion does not interfere with the operation of M.EcoKI and the enzyme is still a functional sequence-specific MTase.

3.3. Absorption and fluorescence spectra

The purified GFP-MTase showed the absorption and fluorescence emission properties expected, Fig. 1A, B. The overlap of the emission of the GFP with the HEX and Dylight549 labels allowed R_o distances of 6.14 nm and 6.53 nm, respectively, to be calculated.

Table 1
Modification of phage $\lambda_{v.o}$ by GFP-MTase protects the phage DNA against the EcoKI R/M system.

Phage recovered from NM1261(DE3) containing the following plasmids	Strain used for plating of recovered phage	Titre of phage on plating strain (pfu/ml)
pBIO2	NM1261(DE3) r^-m^-	3.0×10^8
pBIO2	NM1049(DE3) r^+m^+	1.4×10^4
pJFMS	NM1261(DE3) r^-m^-	0.8×10^8
pJFMS	NM1049(DE3) r^+m^+	1.2×10^8
pJFMSEGFP	NM1261(DE3) r^-m^-	1.4×10^8
pJFMSEGFP	NM1049(DE3) r^+m^+	1.5×10^8

3.4. Fluorescence and anisotropy decay of the fluorescent labels

To use energy transfer quantitatively, one ideally should determine whether the donor and acceptor chromophores are free to rotate or are sterically hindered on the nanosecond timescale as this indicates that the κ^2 orientation parameter can be reasonably set at 2/3 as assumed in Eq. (2). The time-resolved data, supplementary Table S1, indicated that the GFP was rotating on the nanosecond timescale despite its attachment to the MTase but that the degree of rotational freedom on GFP when fused to the MTase was slightly less than that of the free GFP as indicated by the higher value of the anisotropy at infinite time. The rotational correlation time of the HEX label on the DNA duplex was unaffected by GFP-MTase binding as was the degree of rotation of the Dylight549 label when attached to the E20C and E117C mutant ocr proteins, supplementary Table S1 and Figure S3. The label attached to the S68C mutant protein showed an unusual anisotropy decay shape in the absence of GFP-MTase and a long anisotropy decay time in the presence of the GFP-MTase, supplementary Figure S3. These data indicate that the label attached to the S68C position is not free to rotate and hence that the κ^2 orientation parameter is not 2/3 in the FRET experiments.

3.5. DNA binding by GFP-MTase

We additionally checked that the assumption of 1:1 binding to DNA and ocr was correct for the GFP-MTase as previously established for the normal MTase. The interaction between GFP-MTase and both ligands was found to be the same as for the normal MTase [14,15], supplementary information and supplementary Figure S2. The use of concentrations more than 10-fold greater than the dissociation constants for DNA and ocr ensured that there was little unbound donor or acceptor to complicate FRET analysis.

3.6. Steady state fluorescence analysis of FRET between GFP-MTase, DNA and labelled ocr

Fig. 2A shows the induction of FRET when the GFP-MTase was bound to a HEX-labelled 21 bp duplex DNA containing the specificity sequence. Note that significantly more energy transfer occurred to the HEX label in duplex 21TH21B (50.1% transfer) than to HEX in the 21T21BH duplex (8.5% transfer) indicating that one end of the duplex was further from the GFP than the other end. Using the calculated Förster distance, the GFP is separated from the HEX label of 21TH21B by 6.10 nm and from the HEX label on 21T21BH by 9.12 nm.

Fig. 2B shows the induction of FRET when the GFP-MTase bound to the mutant ocr proteins labelled with Dylight549. The amount of energy transfer depended on the mutant used. However, since the

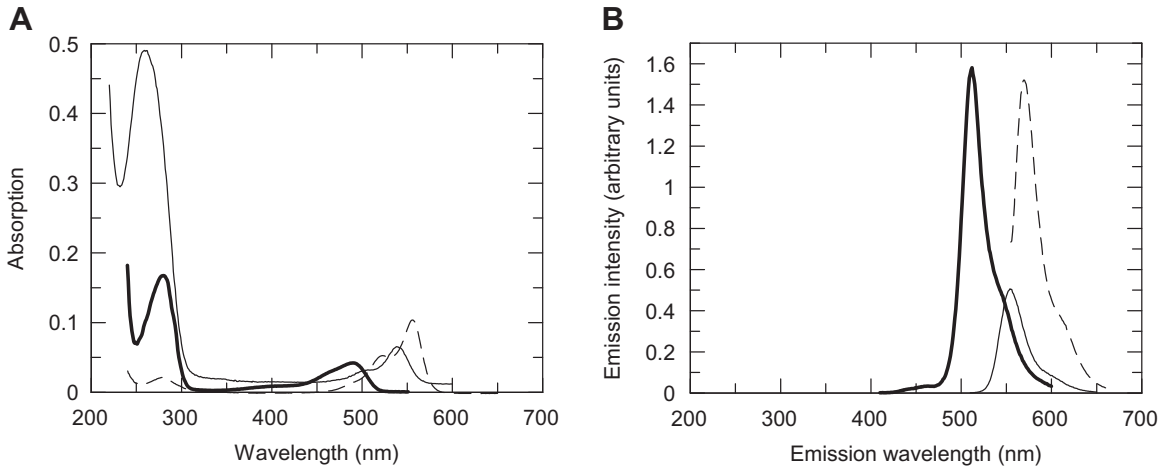


Fig. 1. Spectrophotometric analyses. (A) Absorption spectra of 1 μM GFP-MTase (bold solid line), 1 μM 21TH21B DNA (thin solid line) and 5 μM Dylight549-labeled ocr E20C mutant protein (dashed line). Other labeled proteins had similar spectra. (B) Emission spectra of 1 μM GFP-MTase (bold solid line, excitation at 395 nm), 400 nM 21TH21B DNA (thin solid line, excitation at 530 nm) and 1 μM Dylight549-labeled ocr E20C mutant protein (dashed line, excitation at 550 nm). Other labeled proteins had similar spectra.

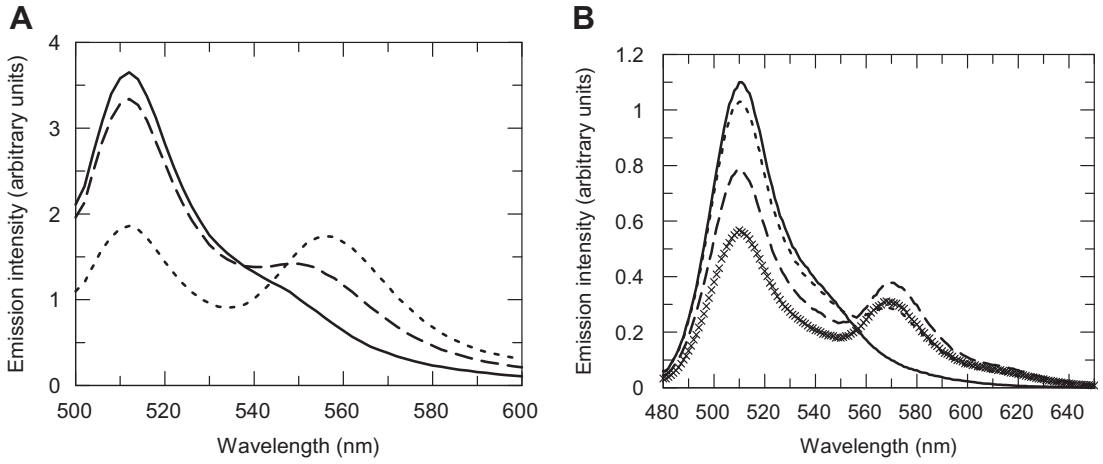


Fig. 2. Fluorescence energy transfer. (A) Fluorescence emission scans of 200 nM GFP-MTase showing the effects of FRET to 200 nM HEX-labelled DNA. GFP-MTase alone (line), GFP-MTase – 21TH21B DNA complex (dashed line), GFP-MTase – 21T21BH DNA (dotted line). Excitation was at 395 nm. (B) Fluorescence emission scans of 500 nM GFP-MTase showing FRET to 500 nM Dylight549-labelled mutant ocr proteins. GFP-MTase alone (line), GFP-MTase – E20C ocr complex (dashed line), GFP-MTase – S68C ocr complex (dotted line), GFP-MTase – E117C ocr complex (small crosses). Excitation was at 395 nm.

introduction of a single cysteine into each ocr subunit means that there are two FRET acceptors and, given the elongated shape of ocr, these acceptors are highly likely to be located at different distances from the GFP donor. Hence the observed FRET was a complex average of the two distances given the $1/r^6$ dependence of FRET on distance. In the absence of further information, we simply calculated this "average" distance to be 7.62, 10.21 and 6.60 nm for the E20C, S68C and E117E mutant ocr proteins assuming κ^2 is 2/3.

3.7. FRET measurements of GFP to HEX using time-resolved fluorescence

The fluorescence decay of the GFP for 1:1 mixtures of DNA and GFP-MTase and of labelled ocr with GFP-MTase were determined. The emission was collected at the magic angle to remove undesirable anisotropy effects on the fluorescence decay and the fitted lifetimes are shown in Table 2.

Table 2
Time-resolved fluorescence decay analysis of samples showing FRET between GFP and HEX or Dylight549. Excitation at 405 nm, emission at 510 nm. The pre-exponential factor for each lifetime is given in the brackets.

Sample	τ1(ns)	τ2(ns)	τ3(ns)	χ^2	<τ>(ns)
GFP-MTase		2.20 (0.36)	3.01 (0.64)	1.082	2.72
GFP-MTase + 21TH21B DNA	0.29 (0.38)	1.42 (0.29)	2.84 (0.33)	1.076	1.45
GFP-MTase + 21T21BH DNA		1.61 (0.31)	2.84 (0.69)	1.061	2.46
GFP-MTase + E20C ocr	0.28 (0.27)	1.42 (0.40)	2.72 (0.33)	1.052	1.55
GFP-MTase + S68C ocr		1.37 (0.44)	2.67 (0.56)	1.173	2.09
GFP-MTase + E117C ocr	0.33 (0.38)	1.28 (0.32)	2.84 (0.29)	1.012	1.38

Table 3
FRET distances (nm) calculated using fluorescence decay times of GFP-MTase in the absence or presence of the fluorescence acceptor compared to distances calculated from fluorescence intensity measurements. All distances are in nm.

Sample	Distance from τ3 to τ1	Distance from τ2 to τ2	Distance from τ3 to τ3	Distance from <τ>	Distance from intensity
GFP-MTase + 21TH21B DNA	4.23	6.79	9.80	6.27	6.10
GFP-MTase + 21T21BH DNA		7.26	9.90	8.97	9.38
GFP-MTase + E20C ocr	4.47	7.22	9.48	6.84	7.62
GFP-MTase + S68C ocr		7.10	9.21	7.98	10.21
GFP-MTase + E117C ocr	4.61	6.90	10.4	6.56	6.60

It was apparent that the presence of the HEX label on DNA or the Dylight549 label on ocr reduced the average fluorescence lifetime, <τ>, of the GFP donor. This was indicative of energy transfer and an average distance between the donor and acceptor could be calculated (Table 3). This distance was in all cases except those using the S68C ocr mutant protein, very similar to the distance calculated from the fluorescence intensity data.

It was also clear that the bi-exponential decay of GFP became a three exponential decay in some complexes so changes in individual lifetimes due to FRET could also be calculated. Considering first the bi-exponential decays, we assumed that since the pre-exponential factors remained roughly constant in the presence or absence of acceptor, that FRET shortens the 2.20 ns lifetime to 1.61 or 1.37 and the 3.01 ns lifetime to 2.84 or 2.67 ns for the 21TH21B and S68C samples, respectively, allowing FRET efficiencies and distances to be calculated (Table 3). In the three exponential decays, we assumed that the 3.01 ns lifetime split into two components; the 2.7–2.8 ns component and the ~0.3 ns component as the sum of the two pre-exponential factors approximately equalled the initial 0.64 pre-exponential factor. The 2.20 ns lifetime, which once again hardly changed its pre-exponential factor, was assumed to decrease to the 1.3–1.4 ns lifetime. These assumptions allowed distances to be calculated. These interpretations imply multiple locations for the GFP with the electronic transition responsible for the 3.01 ns lifetime being particularly sensitive to an interaction with the acceptors. However, the photophysics of GFP and its derivatives is so complex in FRET experiments [22] that it may be wise not to over interpret the distances calculated from the individual lifetimes, particularly since there are two acceptors on the ocr mutant proteins, but rather to use the distance from the average lifetime when examining the location of GFP on the MTase structural model. This is particularly the case for the FRET between GFP and the label in the S68C ocr mutant protein where the acceptor was not free to rotate.

4. Discussion

Our results show that it is possible to fuse GFP to the C-terminus of HsdS in a Type I MTase without any deleterious effect on *in vivo* methylation or *in vitro* binding to either DNA or to a DNA mimic. The assembly of the trimeric MTase is also not affected because the GFP appears to be able to adopt a range of conformations with respect to the MTase and freely move between them.

Recently Kennaway et al. [7] have published an atomic model of the M.EcoKI MTase bound to a DNA duplex and to ocr. Fig. 3 shows the HsdS subunit bound to DNA with the GFP chromophore placed roughly at the distances determined by FRET using the average fluorescence lifetimes (the ocr protein roughly takes the place of the DNA in the atomic model of M.EcoKI and ocr). It can be seen that the results all converge on approximately the same location for the GFP apart from the distance to the S68C location on the ocr protein. The GFP is best located directly below one end of the HsdS subunit to satisfy the FRET distances. This location is what would be expected from the model of M.EcoKI MTase as the location of the C-terminus of HsdS.

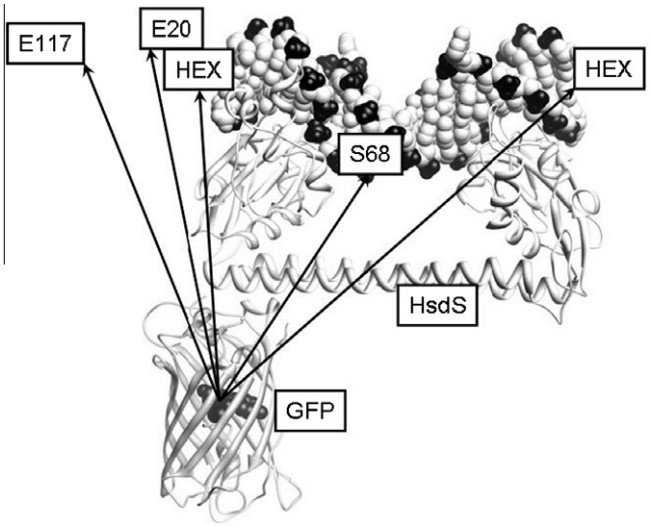


Fig. 3. The HsdS subunit bound to a DNA duplex as proposed from electron microscopy data [7] is shown above a GFP model with the chromophore shown in the centre of the GFP β-barrel. The locations of the HEX labels (21TH21B is on the left and 21T21BH is on the right) and of the locations of the ocr residues labelled with Dylight549 are indicated (ocr is not shown but superimposes on and extends further out than the DNA duplex shown). The arrows show the FRET distances determined from <τ> given in Table 3 except for the distance to S68 on ocr where the distance in the actual model is shown (the FRET distance is longer for this pair but is incorrect due to rotational constraints on the acceptor).

Given the similarity in the structures of the Type I R/M enzymes, this GFP-fusion strategy should work for the other well studied Type I R/M enzymes such as EcoR124I and EcoAI and will facilitate single-molecule experiments both *in vitro* and *in vivo*. It will also allow fluorescence microscopy of the R/M systems in living cells and we note that overexpression of the fusion protein turns the cytoplasm of *E. coli* bright green (unpublished results).

Acknowledgments

We gratefully acknowledge financial support from the Wellcome Trust (GR080463MA), Biotechnology and Biological Sciences Research Council (BB/D001870/1 and BB/C511599/1). Kai Chen was supported by an EaStChem studentship. DTFD gratefully acknowledges The Master and Fellows of Emmanuel College, Cambridge for the award of a Derek Brewer Visiting Fellowship during which time this manuscript was prepared.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2010.06.069.

References

[1] R.Y. Tsien, Constructing and exploiting the fluorescent protein paintbox (Nobel Lecture), Angew. Chem. Int. Ed. Engl. 48 (2009) 5612–5626.

[2] D.W. Piston, G.-J. Kremers, Fluorescent protein FRET: the good, the bad and the ugly, Trends. Biol. Sci. 32 (2007) 407–414.
[3] A.G. Senejani, J.P. Gogarten, Structural stability and endonuclease activity of a PI–SceI GFP-fusion protein, Int. J. Biol. Sci. 3 (2007) 205–211.
[4] M.R. Tock, D.T.F. Dryden, The biology of restriction and anti-restriction, Curr. Opin. Microbiol. 8 (2005) 466–472.
[5] R.J. Roberts, M. Belfort, T. Bestor, et al., A nomenclature for restriction enzymes, DNA MTases, homing endonucleases and their genes, Nucl. Acids Res. 31 (2003) 1805–1812.
[6] R.J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE-enzymes and genes for DNA restriction and modification, Nucl. Acids Res. 35 (2007) D269–D270.
[7] C.K. Kennaway, A. Obarska-Kosinska, J.H. White, I. Tuszynska, L.P. Cooper, J.M. Bujnicki, J. Trinick, D.T.F. Dryden, The structure of M.EcoKI Type I DNA MTase with a DNA mimic antirestriction protein, Nucl. Acids Res. 37 (2009) 762–770.
[8] K. Dybvig, R. Sitaraman, C.T. French, A family of phase-variable restriction enzymes with differing specificities generated by high-frequency gene rearrangements, Proc. Natl. Acad. Sci. USA 95 (1998) 13923–13928.
[9] A.M. Cerdeno-Tarraga, S. Patrick, L.C. Crossman, et al., Extensive DNA inversions in the *B. fragilis* genome control variable gene expression, Science 307 (2005) 1463–1465.
[10] V.A. Barcus, A.J. Titheradge, N.E. Murray, The diversity of alleles at the *hsd* locus in natural populations of *Escherichia coli*, Genetics 140 (1995) 1187–1197.
[11] 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.
[12] S.A. McMahon, G.A. Roberts, K.A. Johnson, L.P. Cooper, H. Liu, J.H. White, L.G. Carter, B. Sanghvi, M. Oke, M.D. Walkinshaw, G.W. Blakely, J.H. Naismith, D.T.F. Dryden, Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance, Nucl. Acids Res. 37 (2009) 4887–4897.
[13] 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, Nucl. Acids Res. 29 (2001) 3059–3068.
[14] 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.
[15] 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.
[16] 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.
[17] A.S. Stephanou, G.A. Roberts, M.R. Tock, E.H. Pritchard, R. Turkington, M. Nutley, A. Cooper, D.T.F. Dryden, A mutational analysis of DNA mimicry by ocr, the gene 0.3 antirestriction protein of bacteriophage T7, Biochem. Biophys. Res. Commun. 378 (2009) 129–132.
[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] Roger W. Hendrix, Jeffrey W. Roberts, Franklin W. Stahl, A. Robert, Lambda II(Cold Spring Harbor Monograph Series 13), Weisberg, 1983.
[20] 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.
[21] J.R. Lakowicz, Principles of Fluorescence Spectroscopy, third ed., Springer, 2006.
[22] M. Millington, G.J. Grindlay, K. Altenbach, R.K. Neely, W. Kolch, M. Bencina, N.D. Read, A.C. Jones, D.T.F. Dryden, S.W. Magennis, High-precision FLIM–FRET in fixed and living cells reveals heterogeneity in a simple CFP–YFP fusion protein, Biophys. Chem. 127 (2007) 155–164.