Characterization of Ocr, the product of gene 0.3 from bacteriophage T7

Constandache Atanasiu

Thesis presented for the degree of Doctor of Philosophy

Institute of Cell and Molecular Biology
University of Edinburgh

2000

To my family

Acknowledgements

I am in debt to a lot of people who helped me during the last four years. Firstly, I would like to thank my supervisor Dr David Dryden for his advice, help and for his continuous encouragement through my PhD. I also want to thank Prof Noreen Murray for her useful advice, particularly concerning *in vivo* antirestriction experiments.

I am very grateful to all present and past members of David's and Noreen's groups, Mary O'Neill, Lynn Powell, Graham Davies, Shane Sturock, Victoria Doronina, Laurie Cooper, Annette Titheradge, Svetlana Makovets and Julie Webb who provide me with an excellent environment for learning new techniques. Mary supervised most of my molecular biology experiments, particularly DNA sequencing and Site-Directed Mutagenesis, Shane and Graham encouraged (and helped) me writing the thesis in LaTeX 2_ε, Victoria and Svetlana helped me with the conjugation assays and Laurie introduced me to the cool technique (in both senses) of protein purification.

Many thanks to my wife, Doina, for her enormous support and understanding. Thanks also go to my family who have supported and encouraged me during my whole life. Last but not least, I would like to thank the Darwin Trust particularly Kenneth Murray for offering me this wonderful opportunity to do a PhD in such a nice town as Edinburgh.

Abstract

The Ocr protein of bacteriophage T7 is a small, dimeric protein which inhibits the cleavage of the phage DNA by type I restriction enzymes in the infected host cells. I have studied the structure and the stability of Ocr and analysed its binding to the *Eco*KI type I restriction-modification enzymes. Ocr has an unique Trp residue in position 94 which is solvent exposed and important for protein stability but not activity. The protein has a melting temperature of 72.2 ± 0.01 °C and a molar extinction coefficient of 32095 M^{-1} cm^{-1} when studied in 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$, pH 8 buffer. Asn4 residue from one monomer is in close proximity to Asn4 in the other monomer. Ser68 residues are also at the monomer-monomer interface. Six surface exposed amino acids were substituted to cysteine and labelled with cysteine-specific fluorophores. The interaction between labelled Ocr(Cys) proteins and *Eco*KI methylase revealed a huge surface area buried at the interface of the two proteins. Ocr binds tightly to the R and S subunits and weakly to the M subunit of the *Eco*KI enzyme. One Ocr dimer binds to *Eco*KI methylase and two dimers to *Eco*KI nuclease. Both Ocr-*Eco*KI methylase and Ocr-*Eco*KI nuclease complexes have a K_d of about 10^{-11} M.

Abbreviations

A	adenine or adenosine; one-letter code for alanine
AdoMet	S-adenosyl methionine
ADP	adenosine diphosphate
Am	ampicillin
AMPS	amonium persulphate
ATP	adenosine triphosphate
BBL	Baltimore Biological Laboratory
BSA	bovine serum albumin
bp	base pair(s)
C	cytosine or cytidine; one-letter code for cysteine
°C	degree Celsium
COOH	carboxyl
ddNTP	dideoxynuccleoside triphosphate
Cm	chloramphenicol
cm	centimetre
Da	Dalton
DSC	Differential Scanning Calorimetry
dsDNA	double strand DNA
DEAE	diethylaminoethyl
dH_2O	deionised water
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
DTT	dithiothreitol
EDTA	diaminoethanetetraacetic acid
e. o. p.	efficiency of plating
EtOH	ethanol
FRET	Fluorescence Resonance Energy Transfer
g	gram
GdnHCl	guanidinium chloride
G	guanine or guanosine; one-letter code for glycine
hsd	host specificity for DNA
h	hour
HsdS	specificity subunit
HsdM	modification subunit
HsdR	restriction subunit

HSV-1	herpes simplex virus 1
1,5 I-AEDANS	N-iodoacetylaminoethyl)-5-naphtylamine-1-sulphonic acid
IPTG	isopropyl-β-D-thiogalactopyranoside
5-IAF	5-iodoacetamidofluorescein
ITC	Isothermal Titration Calorimetry
kb	kilobase
kDa	kilodalton
K_d	dissociation constant
l	liter
M	molar
mA	miliampere
MES	2-(N-morpholino)ethanesulfonic acid
min	minute(s)
NH_2	amino
nm	nanometer
$O.D.^x$	optical density at x nm
PAGE	polyacrylamide gel electrophoresis
PCR	Polymerase Chain Reaction
PM	N-(1-pyrene) maleimide
RNA	ribonucleic acid
RNase	ribonuclease
R-M	restriction-modification
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
ssDNA	single strand DNA
T	thymine or thymidine; one letter code for threonine
TBE	Tris-borate-EDTA
TE	Tris-EDTA
TEMED	N,N,N',N'-tetramethylethylendediamine
T_m	melting temperature
TRD	target recognition domain
Tris	2-amino-2(hydroxymethyl)-1,2,-propandiol
U	uracil or uridine
UGDase	Uracil-DNA-Glycosylase
UGI	Uracil-DNA Glycosylase Inhibitor
UV	ultra violet
V	volt
W/v	weight per volume
ε_M	molar extinction coefficient

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Type I Restriction-Modification enzymes

Restriction and modification (R-M) systems are often regarded as primitive bacterial immune systems (King and Murray, 1994). They protect bacteria against phage infection by restricting (cutting) the viral DNA that enters the cell, a function fulfilled by a R-M endonuclease. R-M systems are able to distinguish between "self" and "non-self" DNA. Their own DNA is protected at the recognition site by methylation, a function fulfilled by a methyltransferase. The absence of the methyl group at the recognition site triggers the endonuclease activity resulting in a double strand cut of the foreign DNA (Meselson and Yuan, 1968).

R-M systems were classified in three classes based on enzyme complexity, cofactor requirement and position of DNA cleavage (Yuan, 1981). Type II R-M systems have played an important role in the development of recombinant DNA technology. They are the simplest of the R-M systems and have separate restriction and modification enzymes (Bickle and Kruger, 1993; Barcus and Murray, 1995). Usually, the endonuclease is a dimer which recognizes and cuts dsDNA within a rotationally symmetric target sequence, and the methylase is monomeric (with a few exceptions) and modifies the same target sequence.

Type III R-M systems contain only four members: *Eco*P1, *Eco*P15, *Hinf*III and *Sty*LT1 and are intermediate in complexity between type I and type II R-M systems (Bickle and Kruger, 1993). Type III enzymes are multi-functional

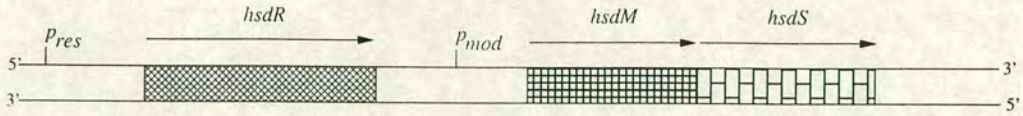


Figure 1.1: The organization of the *Eco*KI *hsdR*, *hsdM* and *hsdS* genes.

proteins able to both cleave and methylate unmodified DNA. They are composed of two different subunits: the *mod* gene product which can function alone as a modification methylase requiring AdoMet and Mg^{2+} and the *res* gene product which has no enzymatic activity and requires ATP and Mg^{2+} for endonuclease activity. Although S-adenosyl-methionine (AdoMet) is required for restriction by type I enzymes, it only stimulates the restriction by type III enzymes (Yuan, 1981).

Type I R-M systems were the first studied in detail and the most complex so far discovered. They are complex, multi-functional proteins which exhibit three enzymatic activities: an endonuclease activity specific for unmodified DNA requiring ATP, AdoMet and Mg^{2+}, (Meselson and Yuan, 1968) a methylation activity that requires AdoMet (Haberman *et al.*, 1972; Hadi *et al.*, 1975) and an ATP-ase activity that is coupled to restriction (Bickle *et al.*, 1978).

Type I R-M systems have been divided into four families (IA-ID) based on genetic complementation tests, antigenic cross-reaction and DNA hybridisation (Murray *et al.*, 1982; Barcus and Murray, 1995; Titheradge *et al.*, 1996). Table 1.1 shows a list of the members of all four families.

The best studied member of the type I systems is *Eco*KI. This enzyme comprises three kinds of polypeptides: R (restriction), M (modification) and S (specificity) (Yuan, 1981), encoded by *hsdR*, *hsdM* and *hsdS* respectively. The *hsd* genes are transcribed in the same direction from two promoters (Fig 1.1), p_{res} situated upstream of *hsdR* gene and p_{mod} between the *hsdR* and *hsdM* genes (Sain and Murray, 1980; Loenen *et al.*, 1987)

The STOP codon of the *hsdM* gene overlaps the initiation codon of the *hsdS*

Table 1.1: Type I R-M systems identified in *Enterobacteriaceae*

Family	Enzyme	Target sequence	Reference
IA	*Eco*KI	AAC(N_6)GTGC	(Kan *et al.*, 1979)
	*Eco*BI	TGA(N_8)TGCT	(Ravetch *et al.*, 1978)
	*Eco*DI	TTA(N_7)GTCY	(Nagaraja *et al.*, 1985c)
	*Sty*LTIII	GAG(N_6)RTAYG	(Nagaraja *et al.*, 1985b)
	*Sty*SPI	AAC(N_6)GTRC	(Nagaraja *et al.*, 1985b)
	*Eco*R5I	not done	(Barcus *et al.*, 1995)
	*Eco*R10I	not done	(Barcus *et al.*, 1995)
	*Eco*R23I	not done	(Barcus *et al.*, 1995)
IB	*Eco*AI	GAG(N_7)GTCA	(Suri *et al.*, 1984)
	*Eco*EI	GAG(N_7)ATGC	(Cowan *et al.*, 1989)
	*Cfr*AI	GCA(N_8)GTGG	(Kannan *et al.*, 1989)
	*Sty*SKI	CGAT(N_7)GTTA	(Thorpe *et al.*, 1997)
	*Sty*STI	not done	(Barcus and Murray, 1995)
	*Eco*R17I	not done	(Barcus *et al.*, 1995)
	*Eco*R42I	not done	(Barcus *et al.*, 1995)
IC	*Eco*R124I	GAA(N_6)RTCG	(Price *et al.*, 1989)
	*Eco*DXXI	TCA(N_7)RTTC	(Gubler *et al.*, 1992)
	*Ecoprr*I	CCA(N_7)RTGC	(Tyndall *et al.*, 1994)
ID	*Sty*SBLI	CGA(N_6)TACC	(Titheradge *et al.*, 1996)
	*Eco*R9I	not done	(Barcus *et al.*, 1995)
	*Kpn*AI	not done	(Lee *et al.*, 1997)

gene (Gough and Murray, 1983), suggesting a translational coupling of the two genes (Oppenheim and Yanofsky, 1980).

M and S subunits are sufficient to form a functional DNA methyltransferase (also called methylase). The ratio between M and S subunits is 2:1 for both *Eco*KI and *Eco*R124I. a member of type IC systems (Dryden *et al.*, 1993; Taylor *et al.*, 1992). During purification of *Eco*KI methylase an M_1S_1 methylase was also identified but this form has been proven to be inactive (Dryden *et al.*, 1993). It appears that formation of the M_1S_1 dimer is the initial step in the assembly of *Eco*KI (Dryden *et al.*, 1997). Two R subunits bind tightly to the methylase forming an active endonuclease with a stoichiometry of $R_2M_2S_1$ (Dryden *et al.*, 1997).

Based on detailed studies *in vitro*, Bickle *et al.* (1978) and Yuan (1981) proposed the following scheme for the mechanism of *Eco*KI enzymes (Fig.1.2). First, *Eco*KI binds rapidly to AdoMet in a non-covalent manner. AdoMet acts as an allosteric cofactor, allowing the enzyme to undergo a slow transition to an activated form, *Eco*KI* (Hadi *et al.*, 1975). *Eco*KI* is able to bind any DNA regardless of the presence or the absence of the specific recognition sites. However, after the initial binding, the activated enzyme seeks and binds to recognition sites (Powell *et al.*, 1998b).

Depending on the methylation state of the recognition site, the oligomeric complex acts as a methyltransferase or endonuclease (Meselson and Yuan, 1968). If the DNA is hemimethylated the activated enzyme switches to the methylation mode and methylates the unmethylated strand.

Unmethylated DNA will be regarded as foreign DNA and will trigger the endonuclease activity. In the presence of ATP, AdoMet dissociates from the complex. ATP acts as an allosteric cofactor inducing conformational changes that lead to a new activated form, *Eco*KI$^+$. The activated enzyme translocates DNA in an ATP-dependent manner, then cleaves the target DNA up to several

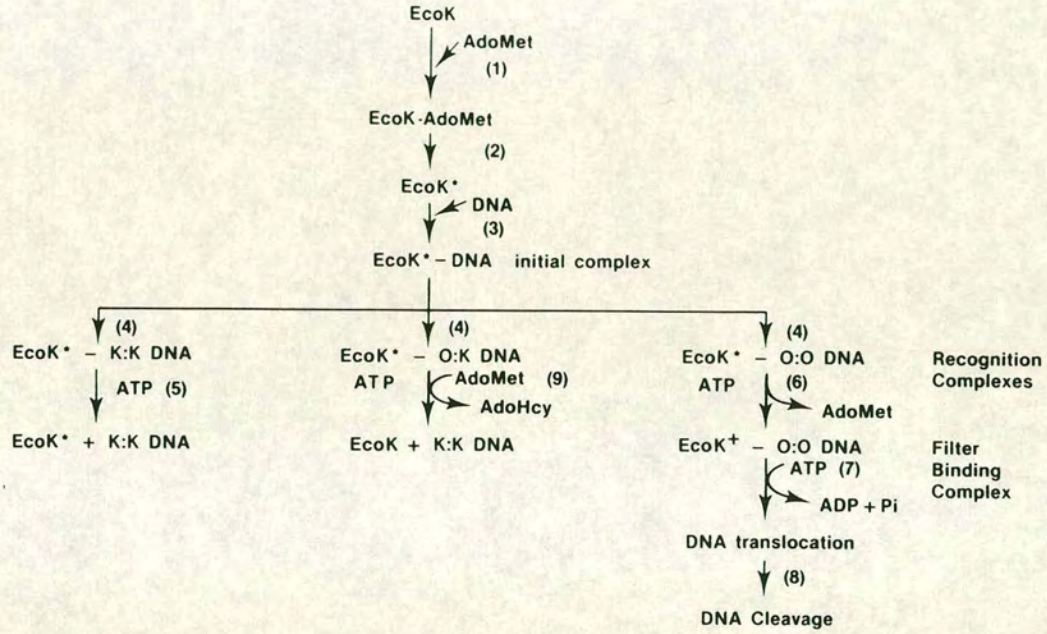


Figure 1.2: The reaction mechanism of the *Eco*KI methylase and nuclease. The first three steps (AdoMet binding, activation to *Eco*KI*, and formation of the initial complex with DNA) are common for both *Eco*KI methylase and nuclease. K:K represents DNA methylated in both strands, 0:K hemimethylated DNA and 0:0 unmethylated DNA. AdoHcy is S-adenosyl-L-homocysteine. From reference Burckhardt *et al.* (1981b)

thousands base-pairs away from the recognition sequence (Horiuchi and Zinder, 1972; Rosamond *et al.*, 1979; Studier and Bandyopadhyay, 1988) . If the DNA is fully methylated at the recognition site (methylated in both strands) the activated enzyme dissociates from DNA without any enzymatic activity.

The S subunits of type I R-M systems dictate DNA sequence specificity. They recognise two asymmetrical bipartite sequences, each of 3 to 5 nucleotides, separated by a spacer of 6-8 non-specific nucleotides. For *Eco*KI, the DNA target sequence is $5' - \mathrm{AAC}(\mathrm{N}_6)\mathrm{GTGC} - 3'$ (Kan *et al.*, 1979).

Sequence alignments of *hsdS* genes identified two long variable target recognition domains (150-180 amino acids each) flanking a central conserved region (Gough and Murray, 1983; Bickle and Kruger, 1993; Barcus and Murray, 1995). The N-terminus target recognition domain (TRD) recognises the AAC trinucleotide and the second TRD recognises the GTGC sequence (Fuller-Pace and Murray, 1986; Gubler *et al.*, 1992).

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

The *hsdS* genes of *Eco*R124I and *Eco*R124II enzymes are identical except for a central conserved sequence of 12 bp which is repeated twice in *Eco*R124I and three times in *Eco*R124II. Their DNA target sequences are identical except for the length of the nonspecific nucleotide sequence which is 6 bp for *Eco*R124I and 7 bp for *Eco*R124II (Price *et al.*, 1989; Gubler and Bickle, 1991). The central conserved domain seems to act as a spacer between the two TRDs and thus controls the length of the nonspecific base sequence (Price *et al.*, 1989). It was also suggested that the conserved region is involved in forming contacts with the other subunits of type I R-M enzymes (Sharp *et al.*, 1992; Cooper and Dryden, 1994; Weiserova and Firman, 1998)

In addition to the regions conserved between members of the same family, there are regions of similarity within each S subunit (Argos, 1985). For type IC systems, a central conserved region was repeated at the carboxyl end (Tyndall *et al.*, 1994). Further investigations revealed that the repeat at the carboxyl terminus is incomplete but the missing part could be identified at the N-terminus (Kneale, 1994). This "split repeat" led Kneale to propose a "circular" organization model of the S domains that allows two M subunits to interact in a symmetrical way (Figure 1.3).

In the absence of any crystal structure of type I R-M enzymes, little is known about the amino acids within TRDs involved in binding the target DNA. Sequence alignment between *Hha*I methylase and 51 type I TRDs suggested that type I TRDs contain a tertiary structure similar to the *Hha*I methylase TRDs (Sturrock and Dryden, 1997). The co-crystal structure of *Hha*I methylase with its target DNA shows that the region involved in DNA binding has a loop-β strand-loop structure (Klimasauskas *et al.*, 1994).

Most of the 101 random mutations introduced in *Eco*KI TRDs did not impair either the restriction or the modification activities (O'Neill *et al.*, 1998). Only seven mutations confered an r^-m^- phenotype. They are not randomly distributed but clustered between residues 80 and 110, a region predicted to be loop-β strand-loop according to Sturrock's model (Sturrock and Dryden, 1997). Based on the tertiary structure of *Hha*I methylase bound to DNA, a model for *Eco*KI methylase bound to its target DNA was proposed (Figure 1.4).

The subunit responsible for methylation of the host DNA is M. The recognition sequence of *Eco*KI enzyme, A**A**C(N)$_6$G**T**GC, is methylated at the adenine shown in bold or the adenine on the complementary strand, opposite the thymine shown in bold (Burckhardt *et al.*, 1981a). The methyl group is transfered from AdoMet which plays a dual role, cofactor and methyl donor. The prefered substrate for methylation is hemimethylated DNA (Vovis *et al.*, 1974; Burckhardt *et al.*, 1981b;

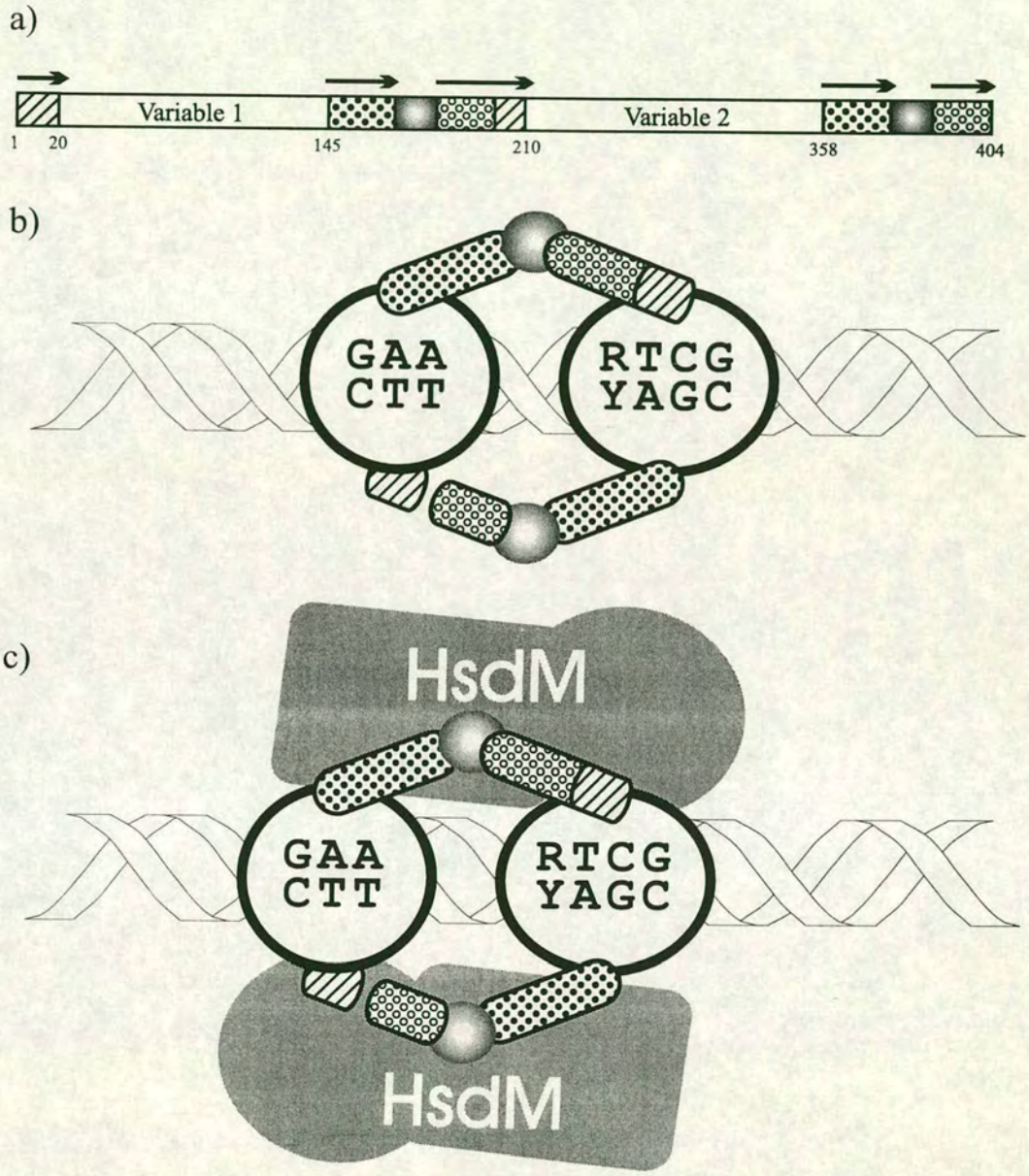


Figure 1.3: HsdS subunit of *Eco*R124I enzyme. (a) Organization indicating the two TRDs and the conserved sequences for type IC family. (b) Model of Kneale, in which the repeated sequences form linkers joining the TRDs in a rotationally symmetrical configuration. (c) Model of the *Eco*R124I methyltransferase, in which the two HsdM subunits bind to the linker region imposing a pseudo-dyad symmetry on the HsdS subunit. From reference Murray (2000).

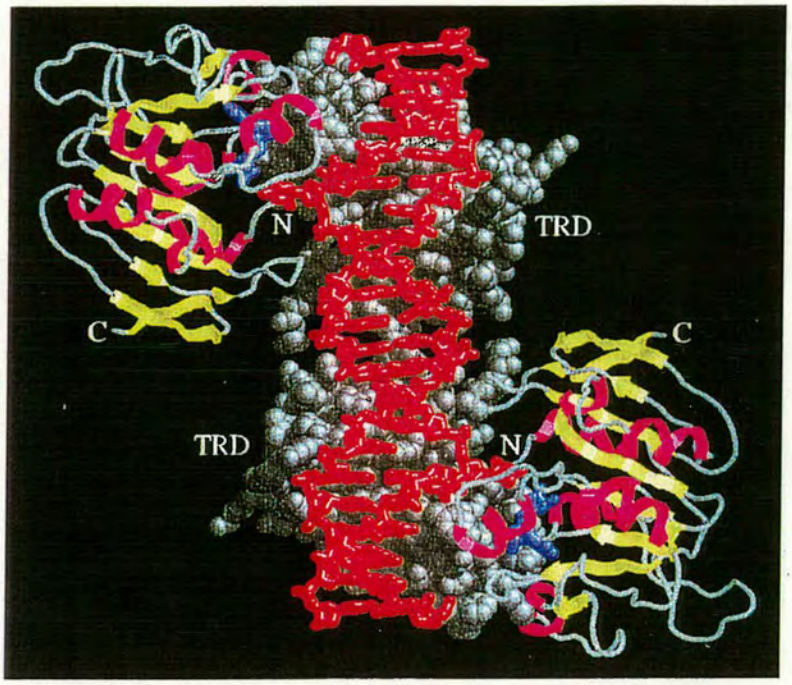


Figure 1.4: A model of *Eco*KI methylase bound to DNA based on the tertiary structure of *Hha*I methylase. α-helices, β-strands and AdoMet are colored magenta, yellow and blue respectively. The target bases are flipped out of the DNA helix and project into the methylation catalytic site. The grey space filled structures are the two TRDs of the S subunit. From reference Dryden *et al.* (1995).

Suri *et al.*, 1984).

The binding affinity for unmethylated and hemimethylated DNA is the same even in the presence of AdoMet, suggesting that the preference for hemimethylated DNA is exerted mainly at the level of catalysis (Powell *et al.*, 1993, 1998a). The affinity of methylase for both specific and non-specific DNA increases in the presence of AdoMet but the effect is more evident for non-specific DNA.

Comparative analysis between sequences of different type I *hsdM* genes (Sharp *et al.*, 1992) revealed a high intraspecific divergence (no more than 32 % identity). This is surprising because within the same family the level of similarity can be as high as 94 %.

All cytosine and adenine methyltransferases have in common a sequence similar to (D/E/S)XFXGXG (also called motif I) responsible for AdoMet binding (Posfai *et al.*, 1989; Kumar *et al.*, 1994; Klimasauskas *et al.*, 1989). In type I systems, the F residue is not so well conserved, the sequence of motif I being DPAX**G**XA for *Eco*KI and DPAXGXG for *Eco*R124I (Sharp *et al.*, 1992). Substitution of the G shown in bold *Eco*KI led to a methylase unable to bind AdoMet (Willcock *et al.*, 1994).

M subunits of N^6-adenine DNA methyltransferases also contain a conserved motif II, (D/N)PP(Y/F) which could be aligned with the PC motif of C^5-cytosine methyltransferases, responsible for methylation activity (Klimasauskas *et al.*, 1989; Posfai *et al.*, 1989). Indeed, mutations in DPPY motif of T4 Dam methylase and *Escherichia coli* Dam methylase abolished the methylation activity (Kossykh *et al.*, 1993; Guyot *et al.*, 1993). For *Eco*KI methylase, a N^6-DNA methyltransferase, the conserved motif II was identified as being NPPY (Loenen *et al.*, 1987). Amino acid substitutions in this motif abolished methylation activity but not the affinity for AdoMet (Willcock *et al.*, 1994).

The circular organization model of S subunit proposed by Kneale, imposes symmetry on the two M subunits positioned on either side of the S subunit.

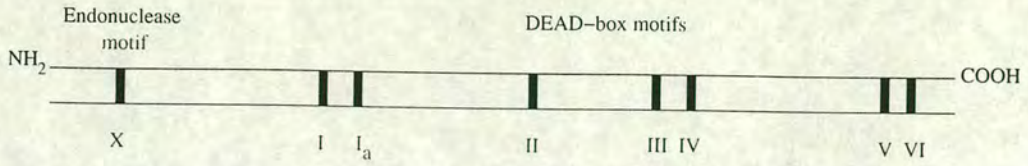


Figure 1.5: DEAD-box motifs in *hsd*R

(Kneale, 1994: Dryden *et al.*, 1995). The M subunit does not bind DNA by itself but asists S to bind (Powell *et al.*, 1998a).

Methylation interference experiments have shown that *Eco*KI methylase interacts with its DNA recognition sequence via contacts in the major groove (Powell and Murray, 1995). Methylation is presumed to occur via a base-flipping mechanism which was proposed to be common to all cytosine and adenine methyltransferases (Klimasauskas *et al.*, 1994; Roberts and Cheng, 1998; Mernagh *et al.*, 1998).

The endonuclease, ATPase and translocase activities are specified by the R subunit. The level of sequence similarity of the R subunit among members of the type I R-M nucleases is quite limited, only 20-30 % identity (Titheradge *et al.*, 1996). However, all of them include motifs similar to those identified in ATP-dependent helicases or putative helicases (Gorbalenya and Konin, 1991). One of these helicase motifs, Asp-Glu-Ala-Asp (or DEAD in a single letter code), gave its name to this family of proteins (Linder *et al.*, 1989). A conserved motif (called X in *Eco*KI), similar to motifs found in the active sites of other endonucleases was also identified at the N terminus (Titheradge *et al.*, 1996). Mutations in motif X abolished restriction and nicking activities but had no effect on DNA binding (Davies *et al.*, 1999b), suggesting that this motif is part of the endonuclease catalytic domain of the R subunit (Fig 1.5).

Two R subunits bind to the *Eco*KI methylase forming a $R_2M_2S_1$ endonuclease, which is a multimeric complex of about 440 kDa (Dryden *et al.*, 1997). Exonuclease III foot printing experiments have shown that *Eco*KI methylase, in

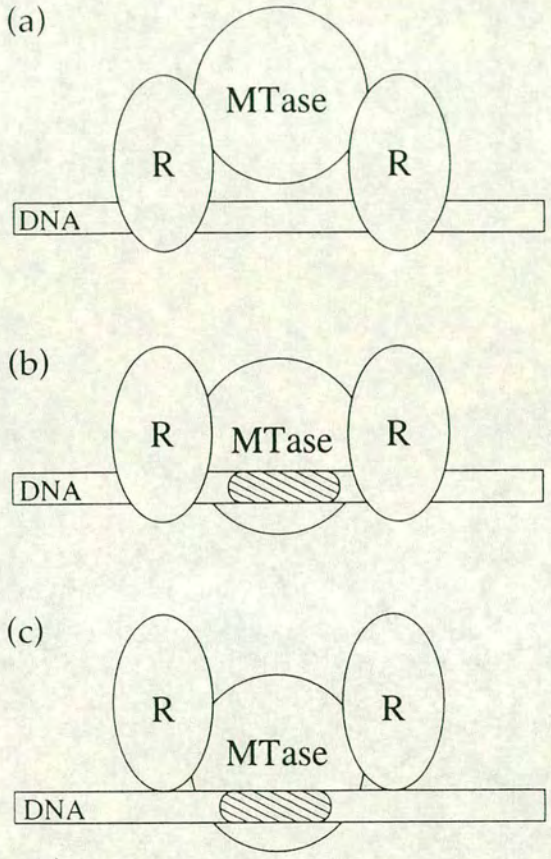


Figure 1.6: An outline of the interaction between *Eco*KI nuclease and DNA showing the effect of the cofactors on binding. (a) Complex formed with non-specific DNA in the absence of ATP. AdoMet does not affect the binding. (b) Complex formed with specific DNA in the absence of AdoMet and ATP. (c) Complex formed with specific DNA in the presence of both cofactors. From reference Powell *et al.* (1998b).

the presence of AdoMet, protects a region of DNA of about 30 bp (Powell *et al.*, 1998a). In the absence of AdoMet and ATP, *Eco*KI nuclease protects a DNA sequence of 42-46 bp, implying that the R subunits project along the DNA helical axis making contacts with it. When both ATP and AdoMet are added, the region protected by the *Eco*KI nuclease is only 30 bp long (Powell *et al.*, 1998b) suggesting that the R subunits have adopted a conformation that allows greater cleavage by exonuclease III. It is possible that this new conformation of the R subunits is a prerequisite for translocation on unmethylated DNA (Fig 1.6).

The mechanism of restriction is a complex process which is not fully understood. One model (Studier and Bandyopadhyay, 1988) proposes that two endonuclease molecules bound to two neighboring recognition sites translocate DNA towards themselves. At the meeting point, usually half distance between two recognition sites, they cleave the DNA. However, atomic force microscopy experiments have shown two endonucleases dimerize prior to translocation (Ellis *et al.*, 1999). Translocation needs ATP hydrolysis, a process that continues long after the cleavage reaction has stopped (Yuan *et al.*, 1972; Eskin and Linn, 1972a,b).

The presence of DEAD-box motifs led to speculation that translocation could occur by a helicase mechanism (Murray *et al.*, 1993). In order to load on the DNA and start unwinding, a helicase needs either free ends or regions of single strand DNA. In addition, since *Eco*KI remains bound to its target site during translocation (Bickle *et al.*, 1978; Rosamond *et al.*, 1979; Yuan *et al.*, 1980; Studier and Bandyopadhyay, 1988; Dreier *et al.*, 1996; Janscak *et al.*, 1996; Szczelkun *et al.*, 1996, 1997), a nicking activity is required in order to relieve the topological constrain that may appear. However, contrary to these suggestions, Davies *et al.* (1999a) have shown that the nicking activity is not necessary either for translocation *in vivo* or for ATPase activity *in vitro*.

The helicase activity has not been proven beyond doubt, but some experiments

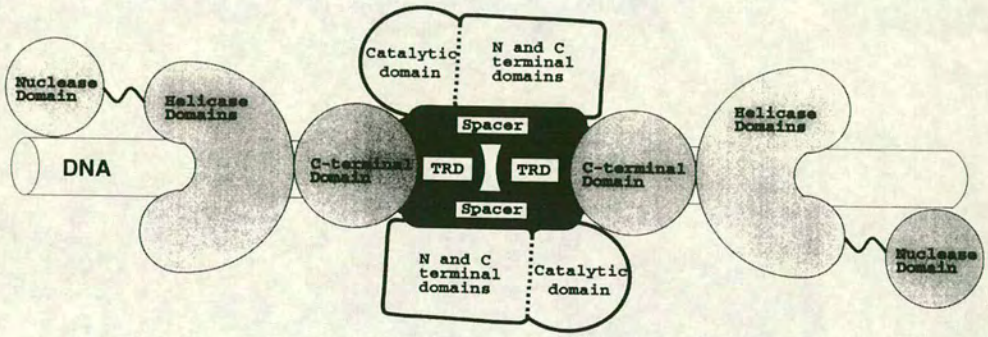


Figure 1.7: The model of the structure of a type I restriction enzyme. The S subunit (heavy shading) comprises two TRDs and two spacer regions. The M subunits (bold outline, no shading) comprise a methylase catalytic domain, an N-terminal domain involved in recognition of DNA methylation and a C-terminal domain required for binding to the S subunit. The R subunits (light shading) comprise an N-terminal endonuclease domain, two DNA helicases domains and a C-terminal domain involved in binding to the M and S subunits. From Davies *et al.* (1999b)

plead in favor of a helicase-like translocation. Mutagenesis in all seven DEAD-box motifs have been shown to impair both restriction and ATPase activities (Webb *et al.*, 1996; Davies *et al.*, 1998). They also impair translocation *in vivo* (Davies *et al.*, 1999a).

Limited proteolysis experiments supplied useful information concerning the structure of R subunit (Davies *et al.*, 1999b). From these assays, it seems that C-terminus region of the R subunit is required for binding to the S subunit.

Based on the accumulated structural information about R, M and S subunits, Davies *et al.* (1999b) proposed the following structural model of the type I restriction enzymes (Fig. 1.7) However, the real breakthrough in understanding the structure of the type I R-M enzymes will be made when X-ray diffraction data become available.

1.2 Phage antirestriction

Many bacteria have developed Restriction-Modification systems as a defensive mechanism against bacteriophage propagation. R-M systems exert a selection pressure on bacteriophage genomes. In response, bacteriophages have evolved strategies to avoid the host R-M systems.

Some bacteriophages encode enzymes able to modify their own DNA. For example bacteriophages SP*3*, SPR, Φ3T and *ρ*11 (Cregg *et al.*, 1980; Noyer-Weidner *et al.*, 1981, 1983) express a DNA methyltransferase which methylates the phage DNA. The methylated sequences are recognized by the restriction endonuclease *Bsu*RI of *Bacillus subtilis* that cuts the sequence GGCC if the central cytosine is not methylated. Due to the methylation at the target sites, all four bacteriophages are protected against restriction.

The resistance of bacteriophage Mu to the restriction exerted by *Eco*KI, *Eco*BI, *Eco*AI and *Eco*P1 is due to a viral gene product called *mom* (standing for *mo*dification of *M*u), able to modify the phage DNA. The modification

does not consist of methylation but acetimidation of the N^6 position of the adenine residues. Approximately 15 to 20 % of the adenine residues are modified in this way. Modification of the phage DNA requires not only an active Mom protein but also a host DNA methylase (Dam) and a phage encoded protein C (Hattman, 1982; Hattman *et al.*, 1985; Sun and Hattman, 1998). Dam methylates the p_{mom} promoter region blocking the binding of another host protein, OxyR, which repress *mom* transcription in dam^- strains (Sun and Hattman, 1996).

Phage λ is severely restricted immediately after its entry in the bacterial cells. However, if the phage manages to survive, a viral gene called *ral* (*r*estriction *al*leviation) is expressed (Zabeau *et al.*, 1980; Debrouwere *et al.*, 1980; Loenen and Murray, 1986). Ral protein is efficient against type IA but not type IB systems (Loenen and Murray, 1986). Unlike other antirestriction proteins, Ral does not inhibit type IA R-M enzymes. The phage is still restricted but a fraction of it escapes. This fraction is then protected by Ral protein which enhances the methylation activity of type IA methylase. Consequently, the unmodified λ DNA is methylated at a high rate, a process normally very inefficient for type IA systems (Suri *et al.*, 1984). Once methylated, λ phage is protected against any further restriction.

A different antirestriction strategy is used by phage P1. This phage is relatively insensitive to the type I R-M enzymes. The antirestriction function is confered by two phage head proteins encoded by the non-essential phage genes *darA* and *darB* (standing for *d*efence *a*gainst *r*estriction) that are injected into the host cells along with the phage DNA (Iida *et al.*, 1987).

Some bacteriophages synthesize antirestriction proteins able to bind and to inactivate R-M enzymes. For example Φ1rH and ΦNR2H phages encode a small protein that binds and inactivates *Bam*Nx, a typical type II restriction enzyme produced by *Bacillus amiloliquefaciens* N (Makino *et al.*, 1979, 1980). The direct interaction was demonstrated by preincubation of *Bam*Nx inhibitor protein with

*Bam*Nx enzyme, that resulted in enzyme inactivation. The direct binding was also supported by the fact that the protein inhibitor did not reduce the activity of *Ava* II restriction enzyme which recognises and cleaves the same target sequence as *Bam*Nx. T3 and T7 are examples of bacteriophages that are protected against type I systems due to a viral protein that impairs both restriction and modification activities (Kruger and Bickle, 1983).

The nucleotide sequences of many phages have been searched for the frequencies of the target sequences recognised by certain restriction enzymes (Schroeder *et al.*, 1986; Bickle and Kruger, 1993; Wilkins *et al.*, 1996). The number of target sites for most of the type II systems is lower than expected. This underrepresentation of the target sites is not so obvious for type I systems. For bacteriophage ϕ1, statistical calculations predict that the sequence recognised by *Eco*RII enzyme should occur 223 times. However, no site for *Eco*RII was found. This shows that selection was directed against these sequences which were considered disadvantageous for phage survival and consequently eliminated during evolution. This is not a unique example. Phage SPO1 have five *Bsu*RI sites although it would be expected to have $\approx$ 300 sites (Kruger and Bickle, 1983).

For phage T7, 56 recognition sites for *E. coli* Dcm and 114 for Dam methylases (Schroeder *et al.*, 1986) are expected. However, only two and six respectively were identified (Dunn and Studier, 1983). The same mechanism of counterselection has drastically reduced the number of recognition sites.

Strand bias is another mechanism by which bacteriophages fight against host restriction enzymes. *Eco*P15 type III R-M enzyme recognises a short asymmetrical recognition site which can be methylated only in one strand whereas the complementary strand remains unmethylated (Hadi *et al.*, 1979; Schroeder *et al.*, 1986; Meisel *et al.*, 1991). During replication, the recognition site in one daughter DNA will be methylated whereas the same site in the other daughter molecule will be unmethylated and ought to be subject to restriction (Fig 1.8). In spite of this

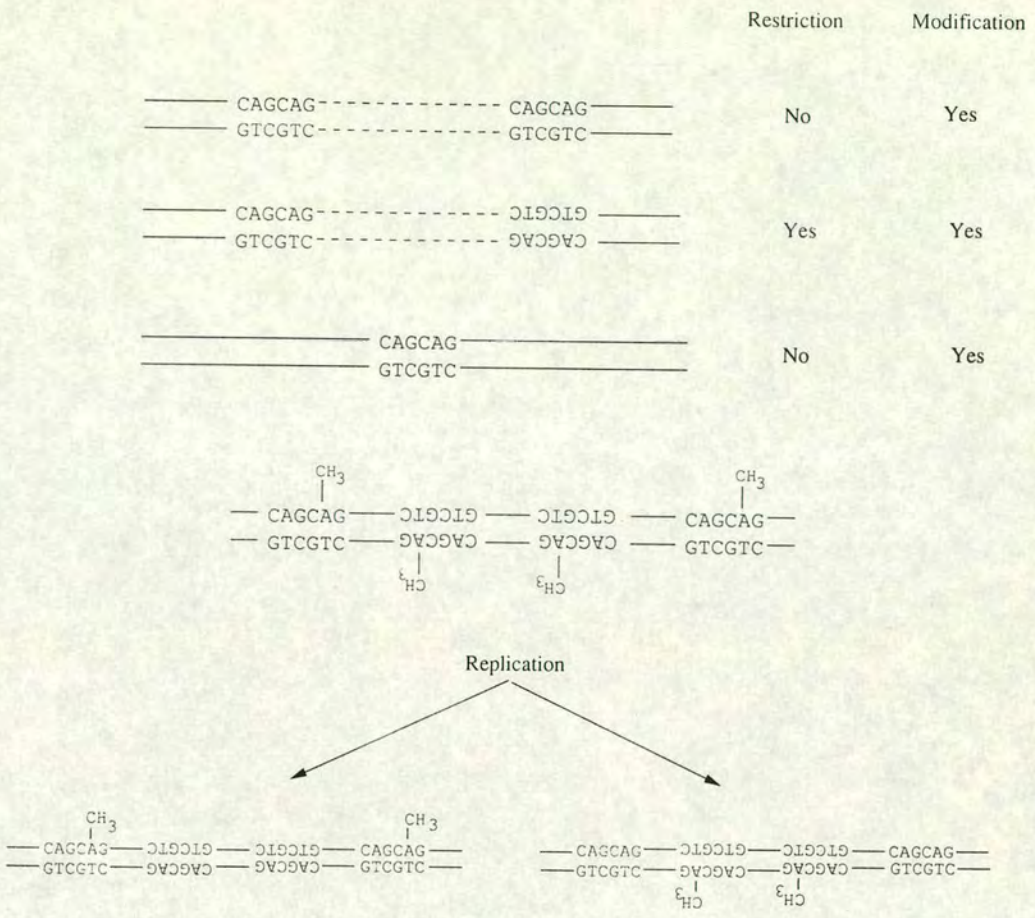


Figure 1.8: Substrate requirement for *Eco*P15 enzyme

fact, the unmodified strand is not restricted because the only substrate for restriction is represented by two unmethylated recognition sequences inversely oriented (Meisel *et al.*, 1992). Bacteriophage T7 has 36 recognition sites for *Eco*P15 but all of them are in the same orientation. Consequently, T7 DNA is refractory to the restriction activity exerted by *Eco*P15 endonuclease. The recognition site for *Eco*P15, CAGCAG, could be redefined as CAGCAG(N_x)CTGCTG, where x is the number of non-specific base pairs that separates the two halves (Kruger *et al.*, 1995). The length of the spacer is not very important for *Eco*P15 activity.

In other cases, the distance between two recognition sites is essential for enzymes activities. *Eco*RII is an atypical type II enzyme that requires simultaneous interaction with two recognition sites for its activity. The unique *Eco*RII site of

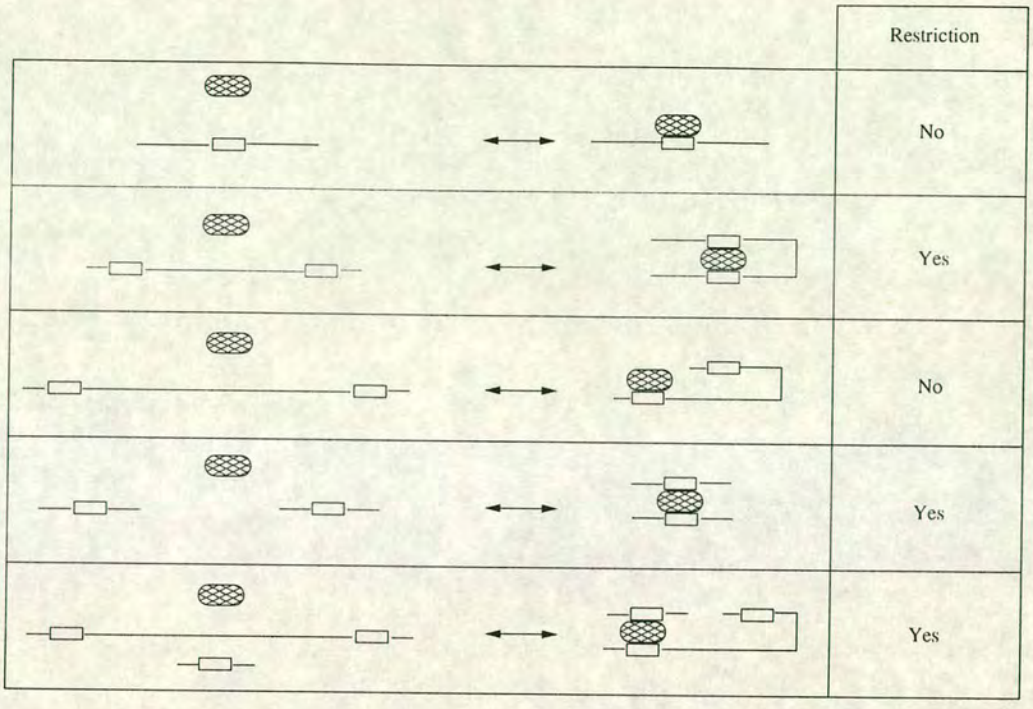


Figure 1.9: Model of productive and refractive *Eco*RII enzyme-DNA interaction. Solid line represent DNA molecule, open boxes represent *Eco*RII recognition sites and hatched ovals represent *Eco*RII enzyme.

T7 DNA as well as the three recognition sites of the T3 DNA are cleaved by *Bst*NI (an isoschizomer of *Eco*RII) but not by *Eco*RII (Kruger *et al.*, 1985). It seems that the distances between *Eco*RII sites of T3 were too great for the enzyme to function. However, when fragments of DNA containing recognition sequences for *Eco*RII were added, the enzyme was able to cut T3 and T7 DNA. Figure 1.9 summarises these observations (Reuter *et al.*, 1998; Kruger *et al.*, 1995).

T-even phage DNAs contain unusual bases which confer resistance to all of the R-M systems that they encounter in their host. These unusual bases, such as 5-hydroxy-methylcytosine (hmC) are further modified by glucosylation (Kruger and Bickle, 1983). In response to the incorporation of hmC bases in the phage DNA, bacteria evolved nucleases specific for unusual bases. These enzymes, called Mcr (standing for *m*ethyl*c*ytosine *r*estricting), are active against non-glucosylated hmC bases. Unusual bases are also incorporated in phages PBS1, PBS2, SPO1,

SP8 and ϕ25 of *Bacillus subtilis*. Thymine is completely replaced by uracil in phages PBS1 and PBS2 (Hemphill and Whiteley, 1975; Kruger and Bickle, 1983) and by 5-hydroxymethyluracil in SPO1, SP8 or ϕ25.

Not only bacteriophages but also self-transmissible plasmids are able to encode antirestriction proteins. Plasmids pKM101, ColIb and pSa of the incompatibility groups N, I and W respectively encode so-called Ard (*a*lleviation of *r*estriction of *D*NA) proteins (Belogurov *et al.*, 1992, 1993, 2000). Ard proteins inhibit both restriction and modification activities of type I systems, have little effect against type II and no effect against type III enzymes.

1.3 Bacteriophage T7

Bacteriophage T7 belongs to the classical T series of bacteriophages. The virion consists of a polyhedral nucleocapsid from which a short, noncontractile tail projects (Serwer *et al.*, 1997). When T7 attacks a bacterium it adheres to the bacterial surface by its tail (Fig 1.10). The tail penetrates the bacterial cell-wall and the DNA is ejected into the cell in a polar fashion, starting with the left end of the T7 genome. (Pao and Speyer, 1973; Saigo, 1975).

T7 genome is a linear dsDNA molecule of about 40000bp and codes for about 30 proteins (Kruger and Schroeder, 1981). Figures 1.11 and 1.12 show an outline of the T7 genome organization and the genetic and physical map of nucleotides 1 to 12,100 of T7 DNA respectively. Positions of different genetic elements are indicated either as bp from the left end or as T7 units (Dunn and Studier, 1981). A T7 unit represents about 400 bp. By convention, genes are numbered according to their relative positions on the genetic map. Studier (1969) has identified 19 genes that were numbered from 1 to 19. The new genes discovered afterwards were given decimal numbers in accordance with their relative positions.

E. coli K-12 is one of the permissive strains which allows a productive T7 growth cycle. After infection, transcription of the T7 genome is done in several

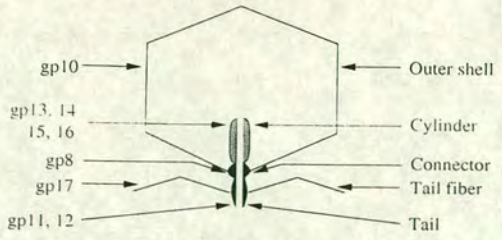


Figure 1.10: A schematic drawing of T7 capsid

stages. The early genes (class I) are transcribed by *E. coli* RNA polymerase that recognises three strong promoters (A_1, A_2, A_3) located at the left end of the T7 genome (Pribnow, 1975; Siebenlist, 1979). Transcription of the early genes results in 5 mRNAs (0.3, 0.7, 1, 1.1 and 1.3) that code for 9 proteins (Studier, 1979). The function of five of them are known. Gene 0.3 protein is an antirestriction protein able to inactivate type I R-M systems (Studier, 1975; Bandyopadhyay *et al.*, 1985). Gp1.2 is a non-essential protein (Studier *et al.*, 1979) responsible for the inactivation of host dGTP triphosphohydrolase (Huber *et al.*, 1988). T7 ligase encoded by gene 1.3 is involved in T7 DNA metabolism, but its function can be substituted by the host DNA ligase (Kruger and Schroeder, 1981; Subramanya *et al.*, 1996; Doherty *et al.*, 1996; Doherty and Wigley, 1999).

Gp0.7 is a protein kinase that inhibits *E. coli* RNA polymerase by the phosphorylation of β and β' subunits (Pai *et al.*, 1975; Zillig *et al.*, 1975; Robertson and Nicholson, 1990; Michalewicz and Nicholson, 1992). Inhibition of *E. coli* RNA polymerase shuts off transcription of the T7 early genes as well as the host genes (Kruger and Schroeder, 1981; Michalewicz and Nicholson, 1992). The host RNA polymerase transcribes only 20 % of the T7 genome. The remaining 80 % (class II and class III genes) is transcribed by the newly synthesized T7 RNA polymerase encoded by gene 1 (Chamberlin *et al.*, 1970; Kruger and Schroeder, 1981).

Besides the A_1, A_2, A_3 promoters mentioned above, three weak promoters for host RNA polymerase have also been identified at the left end of the T7 genome

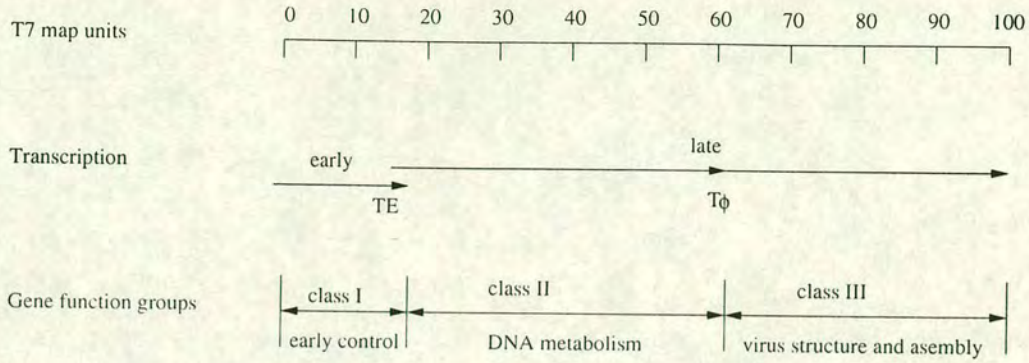


Figure 1.11: Outline of T7 genome organization

(Minkley and Pribnow, 1973; Stahl and Chamberlin, 1977; Studier, 1979). One of them, called A_0, is situated upstream of promoter A_1. The other two, B and C, are located just before gene 0.5 and gene 1 respectively. The exact role of these weak promoters is not very well understood. *In vitro* assays showed that transcription could be initiated at B and C promoters if RNA polymerase is added in such an excess that A promoters are saturated (Stahl and Chamberlin, 1977) or if all A promoters are deleted (Moffatt and Studier, 1988). Another weak promoter for the host RNA polymerase, called E, was identified at the right end of T7 genome, just ahead of gene 19 (Koller *et al.*, 1978). Transcription from all but the A_0 promoter is from left to right (Studier and Maizel, 1969).

The first 160 bp at the left end of T7 are repeated exactly at the right end (Dunn and Studier, 1983). These directly repeated sequences are involved in concatemer formation. Concatemers are tandemly repeated copies of the T7 genome that appear after replication (Chung *et al.*, 1990; Watson and Hayes, 1992).

Transcription of the class I genes terminates either in a rho-independent manner, at the major termination site (TE) at position 18.9 map units or in rho-dependent way at termination sites behind genes 0.3, 0.7 and 1 (Kruger and Schroeder, 1981; Dunn and Studier, 1983).

Transcription of class II genes starts at the ϕ1.1 A and ϕ1.1 B promoters

recognised by T7 RNA polymerase. As easily noticed, there is an overlap region (genes 1.1 to 1.3) between class I and class II genes which is transcribed by both host and T7 RNA polymerases. Transcription of the late genes begins 4 minutes after infection at 37 °C or after 6 minutes at 30 °C (Kruger and Schroeder, 1981).

Most of the class II proteins are involved in DNA metabolism. Recombination requires gp2.5, gp3, gp4, gp5, and gp6 . Besides recombination, gp3 (endonuclease) and gp6 (exonuclease) are also involved in the breakdown of the host DNA (Araki and Ogawa, 1981; Kruger and Schroeder, 1981).

Gene 2 encodes an inhibitor protein that inactivates *E. coli* RNA polymerase by direct interaction, forming an 1:1 complex (Hesselbach and Nakada, 1977; Nechaev and Severinov, 1999). The inhibition is stronger than that exerted by gp0.7. In the absence of functional gp2, T7 infection is abortive, probably because gp2 interferes with the viral DNA packaging and assembly (LeClerc and Richardson, 1979; DeWyngaert and Hinkle, 1980).

At the T7 replication fork three T7 proteins and one host protein are responsible for the basic reactions: gp5 (DNA polymerase), gp4 (helicase/primase), gp2.5 (ssDNA binding protein) and *E. coli* thioredoxin (Richardson, 1983; Kim and Richardson, 1993; Notarnicola *et al.*, 1997). Gp5 and thioredoxin form a 1:1 complex able to catalyze the duplication of the template DNA in a processive manner (Bedford *et al.*, 1997). Gp2.5 interacts with both gp5-thioredoxin and gp4, stimulating their activities(Nakai and Richardson, 1986; Kim *et al.*, 1992). There is also an interaction between the carboxyl terminus of gp4 and gp5 (Notarnicola *et al.*, 1997).

Replication of the T7 DNA is initiated at a site near position 14.75 to 15.0 map units (Saito *et al.*, 1980). This region lies between the ϕ1.1 B promoter and the beginning of gene 1.1 and is rich in A-T base pairs (Fig 1.12). The sequence contains seven TTAA repeats that allow formation of hairpins and also a GGCCC sequence recognised by T7 primase that initiates the synthesis of the

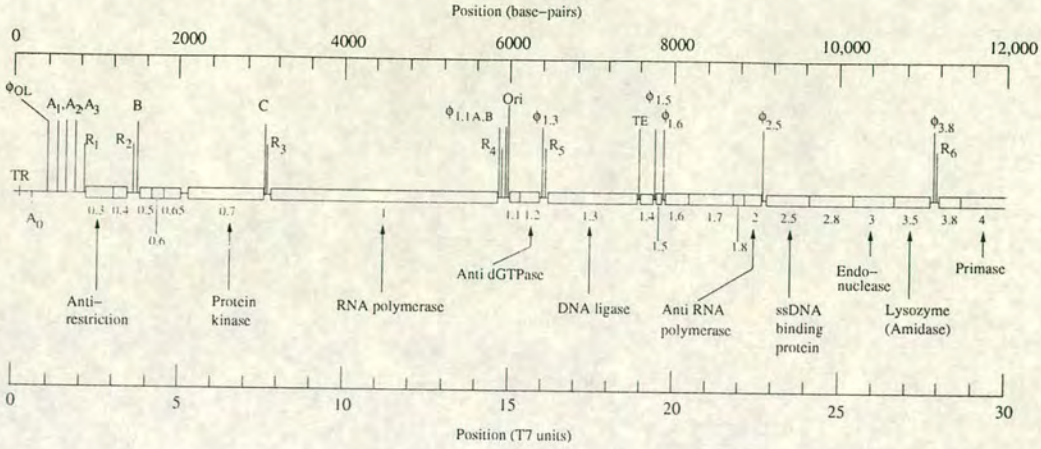


Figure 1.12: Genetic and physical map of nucleotides 1 to 12,100 of T7 DNA. The open reading frames are indicated by boxes and the non-coding sequences are indicated by single lines. R1-R6 are the RNase III cleavage sites. Ori stands for origin of replication and TR for terminal repetition.

leading strand (Kruger and Schroeder, 1981).

A strong promoter for T7 RNA polymerase was found between the A_0 and A_1 promoters (Dunn and Studier, 1981). The role of this promoter was initially not obvious, since no mRNA transcribed from this promoter had been identified at late times after infection (McAllister and McCarron, 1977). Computational searches revealed a very good match with known promoters. It seems that the promoter is functional but the RNA synthesis terminates a short distance after the initiation site which can explain the lack of any detectable mRNA. The sequence that follows this promoter is very rich in A-T base pairs. Comparative sequence analysis between this region and the origin of replication near position 14.75 suggested that the T7 RNA polymerase promoter near position 1.02 plus the following A-T rich sequence form a second origin of replication. This promoter was named ϕOL, standing for phage promoter, origin of replication, left (Dunn and Studier, 1981).

T7 lysozyme is a 17 kDa protein encoded by gene 3.5. It functions both as an amidase, cutting the peptidoglycan layer of the cell wall, and as a transcriptional inhibitor shutting off transcription of the class II genes (Inouye *et al.*, 1973; Mof-

fatt and Studier, 1987; Zhang and Studier, 1995, 1997). The inhibition occurs by direct binding to T7 RNA polymerase at a site remote from the polymerase active site (Jeruzalmi and Steitz, 1998). Transcription of gene 3.5 is done by T7 RNA polymerase. As levels of lysozyme increase, more and more T7 RNA molecules are trapped in inactive lysozyme-RNA polymerase complexes. The remaining active polymerase molecules binds preferentially to class III promoters that are much stronger than those of class II. Thus, transcription of the class II genes is shut off. Transcription of the class III genes continues until lysis (Studier, 1972).

Just beyond gene 6 lie the class III genes that are also transcribed by T7 RNA polymerase. Class III proteins are involved in DNA maturation and packaging (Kruger and Schroeder, 1981).

Gp7 and gp13 (coat proteins) are required for the infectivity of the mature phage but are not required for assembly. Gp8, gp9, gp10, gp14, gp15, gp16, gp17 (head proteins) and gp13 assemble into procapsids. Packaging is initiated in the presence of gp18, gp19 and T7 DNA. Maturation and packaging also require T7 RNA polymerase and is stimulated by T7 lysozyme (Zhang and Studier, 1995). Gp11 and gp12 form the phage tail and gp17 the tail fibers (Kruger and Schroeder, 1981).

T7 RNA polymerase recognises at least two types of termination sites. The first is typified by the signal found in the late region of T7 genome (Tϕ) and involves the formation of a stable stem-loop structure in the newly made RNA, ahead of the transcription point. The second class of terminators was first identified in the the cloned human preproparathyroid hormone (PTH) gene and lies in the concatemer junction region. This termination site does not generate a stem-loop structure in the RNA but contains a conserved sequence ahead of the point of termination (Macdonald *et al.*, 1994; Lyakhov *et al.*, 1997, 1998).

1.4 Ocr protein

Closely related bacteriophages T3 and T7 are not subjected to either restriction or modification by *Eco*KI and *Eco*BI R-M systems *in vivo* (Kruger *et al.*, 1977; Kruger and Bickle, 1983). However, *Eco*BI cuts purified T7 DNA *in vitro*, demonstrating that the ability of bacteriophage T7 to avoid restriction is not due to the lack of recognition sites for this enzyme (Eskin *et al.*, 1973). Indeed T7 DNA contains 6 recognition sites for *Eco*BI and 4 for *Eco*KI, meanwhile T3 DNA has 5-7 recognition sites for *Eco*BI (Kruger *et al.*, 1977; Dunn and Studier, 1983). What prevents the activity of these restriction enzymes *in vivo*?

Studier (1975) has demonstrated that *in vivo* protection of bacteriophage T7 is due to the product of gene 0.3. Gene 0.3 is the leftmost gene on T7 DNA and the first to be expressed after infection. The antirestriction protein was called either gene 0.3 protein according to its position on T7 map or Ocr (standing for *o*vercome *c*lassical *r*estriction) due to the phenotype associated with the gene (Studier, 1975; Kruger *et al.*, 1977).

The early region of T3 DNA is organized in the same way as that of T7 DNA. Five early mRNAs, numbered 0.3, 0.7, 1, 1.1 and 1.3, were identified in both T3 and T7. Proteins translated from these mRNA have similar functions in both bacteriophages. Gene 1 codes for RNA polymerase in T3 as well as in T7, gene 0.7 for protein kinase and gene 1.3 for ligase. Gene 0.3 of T3 codes also for an antirestriction protein that protects T3 against host restriction. However, homologous proteins in T3 and T7 migrate slightly different on SDS-PAGE gel, suggesting that they have different sequences (Studier and Movva, 1976).

Gene 0.3 is located at the left end on T3 DNA and is the first to be expressed after infection. Unlike the corresponding antirestriction protein from T7, gene 0.3 protein of T3 is also a S-adenosyl methionine hydrolase (SAMase) able to cleave AdoMet (Gefter *et al.*, 1966). Since AdoMet is one of the essential cofactors required for *Eco*KI and *Eco*BI restriction enzymes, the antirestriction function

of SAMase could be due to the AdoMet hydrolysis. This explanation is not satisfactory since T3 *sam*$^-$ and T7 (which is *sam*$^-$) are also protected *in vivo* against type I R-M systems. T7 and T3 *sam*$^-$ retain their Ocr function despite the lack of any SAMase activity (Kruger *et al.*, 1977). Further studies confirmed that T3 SAMase and T3 Ocr activities are distinct functions of the same protein and SAMase is not necessary for phage protection (Spoerel *et al.*, 1979).

Despite their common antirestriction function, T3 and T7 Ocr proteins have little if any homology. The unrelatedness is shown by the following observations: antibodies raised against T3 Ocr do not crossreact with T7 Ocr (Spoerel *et al.*, 1979), T3 Ocr is less acidic than T7 Ocr (Mark and Studier, 1981) and their DNA sequences have little if any homology (Davis and Hyman, 1971).

The mechanism by which Ocr manages to inhibit type I R-M enzymes is not well understood. Ponta *et al.* (1976, 1977) suggested that Ocr alters the permeability of the *E. coli* so that the intracellular environment is changed and becomes optimum for the synthesis and function of phage-induced enzymes. Herrlich *et al.* (1974) ascribed a translational repressor function for Ocr protein, able to shut off the synthesis of host proteins after infection (in fact gp0.7 and gp2 shut off the synthesis of the host proteins). None of these proposals were confirmed. Later on, Mark and Studier (1981) have shown that inhibition of *Eco*KI requires stoichiometric rather than catalytic amounts of T7 Ocr, suggesting that inhibition occurs by direct binding to one of the enzyme subunits (Bandyopadhyay *et al.*, 1985). T3 Ocr inhibits restriction enzymes in a similar manner, by direct binding to the *Eco*KI rather than by hydrolysis of AdoMet (Spoerel *et al.*, 1979).

T7 Ocr protein was purified and its amino acid sequence accurately determined (Mark and Studier, 1981; Dunn *et al.*, 1981). The amino acid sequence corresponds to the sequence of the mRNA, except for the first methionine which is missing in the purified protein (Fig 3.2). 0.3 mRNA, the first mRNA that appears after infection, codes for two proteins: Ocr and gp0.4, a small viral protein

of unknown function. The STOP codon of the *ocr* gene overlaps the START codon of gene 0.4, suggesting a translational coupling of the two genes (Dunn *et al.*, 1981).

Ocr protein from T7 is a dimer with a molecular weight of 27,356 Da, each monomer containing 116 amino acids (Mark and Studier, 1981). Ocr is very acidic, 34 of its amino acids are either Asp or Glu and only 6 are Arg or Lys. An Ocr mutant which contains the first 87 amino acids has no antirestriction activity, one with 93 amino acids has a weak activity, whereas one containing 94 amino acids is able to prevent host restriction *in vivo*. The key amino acid for activity seems to be Trp94. If Trp94 is present, the last 22 amino acids at the carboxyl end are not required for antirestriction activity *in vivo* (Dunn *et al.*, 1981).

Purified Ocr binds tightly to one of the small *Eco*KI subunits specified by *hsdM* or *hsdS* genes (Bandyopadhyay *et al.*, 1985). Once formed, the Ocr-*Eco*KI complex is essentially irreversible under physiological conditions and *Eco*KI is no longer able to bind DNA. The large excess of acidic residues raises the possibility that Ocr could act as a polyanion and compete for the DNA-binding site of the restriction enzyme. Since the subunit responsible for binding the DNA target sequence is HsdS, it is very likely that the same subunit interacts with Ocr. Keeping in mind that both nuclease and methylase have only one HsdS subunit, a 1:1 stoichiometry for Ocr-*Eco*KI is very possible. Data obtained so far are quite contradictory. Bandyopadhyay *et al.* (1985) reported that one or two Ocr molecules are enough to inhibit *Eco*KI enzymes, meanwhile Kruger *et al.* (1985) have shown that 4 or more Ocr dimers are required for complete inhibition.

*Eco*KI binds either Ocr or specific DNA *in vitro*, but not both at the same time. If Ocr binds first, the complex is irreversible and the enzyme is completely inhibited. If *Eco*KI-DNA complex is allowed to form first, *Eco*KI is trapped by the Ocr molecules upon dissociation from DNA. The possibility that Ocr binds *Eco*KI-DNA complex and subsequent conformational changes decrease the

affinity of *Eco*KI to DNA is unlikely since no binding of Ocr to *Eco*KI-DNA has been detected (Bandyopadhyay *et al.*, 1985). If the dissociation-trapping model was true, the rate of binding of Ocr to *Eco*KI should depend on the rate of *Eco*KI dissociation from DNA. Indeed, *Eco*KI enzymes bound to hemimethylated DNA were more easily trapped by Ocr than those bound to unmethylated DNA (Bandyopadhyay *et al.*, 1985). It is known that *Eco*KI dissociates less readily from unmethylated DNA than from methylated DNA (Yuan *et al.*, 1975).

The mechanism by which *Eco*KI enzyme cuts foreign DNA is very complex and occurs in many steps (Yuan, 1981). The question is: at which of these stages does Ocr interact with *Eco*KI and inhibit activity? The activation of *Eco*KI by AdoMet seems not to be important since Ocr binds *Eco*KI molecules whether or not they were previously activated by AdoMet (Bandyopadhyay *et al.*, 1985). Low amounts of Ocr added to an *Eco*KI-unmodified DNA complex before adding ATP, did not prevent ATPase but were enough to prevent nuclease activity. When ATP was added before Ocr, even large amounts of Ocr could not prevent ATPase activity. Once ATPase activity has started, it is no longer sensitive to Ocr inhibition. Since low levels of Ocr are able to prevent nuclease but not ATPase activities, it seems that Ocr inhibits *Eco*KI after the onset of ATPase and after DNA translocation but before cleavage (Bandyopadhyay *et al.*, 1985).

Restriction enzymes usually cleave foreign DNA immediately after its entry into the cell. It was not clear initially how Ocr confers such an instantaneous protection of T7 DNA, taking into account that the first Ocr molecule has been detected only 2 to 4 minutes after phage injection (Studier, 1972; Moffatt and Studier, 1988; Garcia and Molineux, 1995). The answer came when the mechanism of T7 DNA entry into *E. coli* was elucidated (Zavriev and Shemyakin, 1982).

The entry of T7 genome is a polar, multistep process, dependent on both *E. coli* and T7 RNA polymerases. The left end of the T7 genome (about 850

bp) is ejected into the cell by a transcription-independent mechanism (Zavriev and Shemyakin, 1982; Garcia and Molineux, 1995, 1996). Within this 850 bp sequence lie three strong promoters for *E. coli* RNA polymerase. The host RNA polymerase attaches to these promoters and starts transcription at a rate of 40 bp per second (Garcia and Molineux, 1995). By transcription, *E. coli* RNA polymerase mechanically pulls T7 genome into the cell (Zavriev and Shemyakin, 1982). About 7 kb of the T7 genome enters via a transcription-dependent process.

Gp0.7 and gp2 inhibit host RNA polymerase (Michalewicz and Nicholson, 1992; Nechaev and Severinov, 1999). Transcription and translocation of the remaining T7 genome is done by the newly synthesized T7 RNA polymerase at a rate of 200 to 300 bp per second.

The fact that translocation of T7 DNA is a transcription-mediated process was confirmed by using rifampicin, an antibiotic that inhibits host polymerase but not the T7 RNA polymerase. In the presence of rifampicin, the entrance of T7 DNA was blocked (Zavriev and Shemyakin, 1982).

The leading 850 bp sequence is not restricted probably because no *Eco*KI recognition sites lie in this region (Zavriev and Shemyakin, 1982). According to Studier's model, linear dsDNA needs at least two *Eco*KI recognition sites for restriction to occur (Studier and Bandyopadhyay, 1988). However, even a single *Eco*KI site on T7 genome leads to restriction *in vivo*, because replicating concatemers contain multiple recognition sites (Garcia and Molineux, 1999). The first *Eco*KI recognition site (located at 15,173 bp from the left end) enters after 7 minutes and the second (located at position 26,603 bp), 45 seconds later (Dunn and Studier, 1983). By this time, all the *Eco*KI molecules are already inhibited by the large amount of Ocr produced.

However, the lack of the restriction sites is not the real reason for the leading sequence resistance to *Eco*KI cleavage. Moffatt and Studier (1988) artificially introduced *Eco*KI recognition sites ahead of gene 0.3 (in position 836 bp), behind

it (1.379 bp) or on the both sides of it but the mutant phage grew normally. Thus, they proved that the ability of T7 to avoid restriction by *Eco*KI is not due to the absence of the recognition sites ahead of gene 0.3, but because the leading sequence is somehow protected against *Eco*KI for a limited period, long enough to allow Ocr expression. Further studies have confirmed the model suggested by Moffat and Studier, in which T7 DNA initially enters a compartment of the cell which is accessible to *E. coli* RNA polymerase but not to the restriction enzymes. Protection against restriction enzymes could be due to bacterial components and/or phage proteins injected into the cell along with T7 DNA (Garcia and Molineux, 1995, 1996). The 836 bp sequence is also temporally inaccessible to Dam methylase and *Eco*P1 restriction enzyme (Moffatt and Studier, 1988; Garcia and Molineux, 1996).

Garcia and Molineux (1996) have isolated mutant virions able to translocate the entire genome by a transcription-independent mechanism, at a constant rate of 75 bp per second. All mutants presented mutations in gene 16. Gp16 is a phage head protein that seems to function as a clamp to prevent transcription-independent DNA translocation. In wild type T7, gp16 does not allow more than 836 bp to be ejected in a transcription-independent manner (Garcia and Molineux, 1996; Struthers-Schlinke *et al.*, 2000).

When T7 DNA was packed in a bacteriophage λ head, then ejected into *E. coli* K-12 cells, the DNA was immediately degraded. Ejection from λ head was fast and *Eco*KI nuclease had immediate access to its recognition sites, before any Ocr was expressed. This experiment proved once more that the entrance of T7 genome into host cells is a delayed, multistep process (Garcia and Molineux, 1999).

The impaired growth of T3 and T7 on P1-lysogenic cells has been known for a long time. The viral DNA is cut by *Eco*P1, a type III R-M enzyme, which is not inhibited by either T3 or T7 Ocr proteins (Moffatt and Studier, 1988). Another type III enzyme, *Eco*P15, cuts T3 but not T7 DNA. The resistance of T7 DNA is

not confered by Ocr protein but by the fact that *Eco*P15 needs for cleavage two recognition sites inversely oriented. T7 DNA has 36 sites for *Eco*P15 but all of them are in the same orientation (Kruger *et al.*, 1995).

Purified T7 Ocr is ineffective against type II restriction enzymes *Hpa*I, *Hpa*II, *Hinc*II and *Hae*III, even at concentrations 100 times higher than needed to inhibit *Eco*BI (Mark and Studier, 1981). T3 and T7 Ocr also have no effect against *Eco*RV, another type II restriction enzyme (Kruger *et al.*, 1983). T3 DNA contains 5 recognition sites for *Eco*RV and is restricted either in the presence or in the absence of T3 Ocr protein. T7 DNA and its *ocr*$^-$ derivatives are not restricted by *Eco*RV because they do not contain any recognition site for this enzyme (Dunn and Studier, 1981; Stahl and Zinn, 1981; Dunn and Studier, 1983).

The finding that Ocr protein overcomes host restriction does not exclude other possible roles for it. Analysis of protein-protein interactions by the yeast two-hybrid system identified 25 interactions among T7 proteins expressed after infection. One of these interactions is between Ocr and gp4.5, a T7 late protein of unknown function. The significance of this interaction has not been yet identified (Bartel *et al.*, 1996).

1.5 Protein-protein interactions

Protein-protein interactions are fundamental to a diversity of biological processes, from replication, transcription, translation to secretion, cell cycle control, signal transducing or intermediary metabolism (Phizicky and Fields, 1995). Binding affinities of different protein-protein complexes range from millimolar for respiratory electron transfer proteins to femtomolar for nuclease-inhibitor complexes (Fig. 1.13).

There is an interesting parallel between the Ocr-*Eco*KI complex and the interaction between Uracil-DNA-Glycosylase (UDGase) and its inhibitor (UGI). PBS-1 and PBS-2 bacteriophages of *Bacillus subtilis* contains uracil instead of

thymine in their genomes. To survive in the host cell, they need to inactivate the host UGDase, which is the first enzyme in the pathway responsible for uracil excision and repair. This inactivation is carried out by UGI, a small, monomeric and very acidic phage protein which forms a tight complex with UGDase, completely inactivating the enzyme (Cone *et al.*, 1980; Bennett and Mosbaugh, 1992).

UGI also inactivates UGDase from other biological sources, including human, herpes simplex virus type-1 (HSV-1) and *E. coli* (Mol *et al.*, 1995; Bennett and Mosbaugh. 1992; Bennett *et al.*, 1993; Savva and Pearl, 1995; Putnam *et al.*, 1999). The co-crystal structure of HSV-1 UDGase-UGI complex reveals an $\alpha-\beta-\alpha$ structure for UGI, which interacts with enzyme residues involved in substrate binding. The electrostatic contacts between $\alpha - \beta - \alpha$ structure of UGI and UGDase appear to mimic the enzyme-DNA contacts previously observed in the UGDase-oligonucleotide complex (Savva *et al.*, 1995). Ocr is a very acidic dimer protein which probably also acts as a DNA mimic. The UGI monomer only has to block one active site, whereas the Ocr dimer has to block two sites, one in each part of the type I recognition sequence.

Colicins are protein toxins produced by bacteria under stress conditions. They are released into the extracellular medium in order to protect the producing bacteria against other bacteria. Colicin-producing bacteria protect themselves against the toxic effect of the colicins by coexpressing a small inhibitor protein called immunity (Im). Each immunity protein protects cells against its cognate colicin. For example Colicin E9. a 61 kDa nonspecific DNase. is inhibited by Im9, a 9.5 kDa acidic protein. The inbition occurs by direct binding between the Im9 and the C-terminus of Colicin 9 (Wallis *et al.*, 1995a,b, 1998).

For tight binding complexes like Colicin9-Im9, a direct estimation of the equilibrium dissociation constant (K_d) is very difficult. In such cases, the K_d is determined from the ratio of the dissociation (k_{off}) and association (k_{on}) rate constants (Lee *et al.*, 1989; Longstaff *et al.*, 1990; Schreiber and Fersht, 1993;

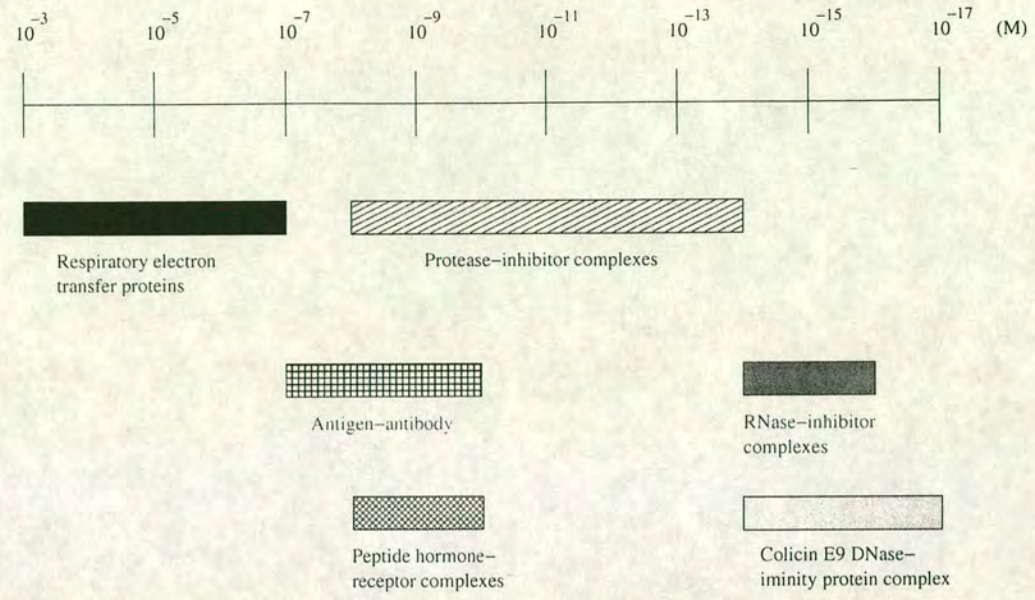


Figure 1.13: Comparison of dissociation constants for different protein-protein complexes. From reference Wallis *et al.* (1995a)

Wallis *et al.*, 1995b). k_{on} was determined by stopped-flow fluorimetry and k_{off} by subunit exchange. Im9 was labeled with [^{3}H] then incubated with the Colicin9 to form a stoichiometric complex. An excess of unlabelled Im9 protein was added to the preformed complex. k_{off} was followed by monitoring the dissappearance of [^{3}H]Im9 from the complex with time. Complex was separated from unbound Im9 by gel filtration. According to the K_d determined for Colicin9-Im9 complex (10^{-17} M) this is one of the tightest complexes ever characterized (Wallis *et al.*, 1995b).

One of the best characterized nuclease-inhibitor complexes is barnase-barstar (Schreiber and Fersht, 1993; Frisch *et al.*, 1997; Vaughan *et al.*, 1999). Barnase is an extracellular RNase of *Bacillus amiloliquefaciens*, which consists of 110 amino acids and has a molecular weight of 12,832 Da. Barnase is inhibited by barstar, a small intracellular inhibitor of 89 amino acids and the two proteins form a very tight complex. The K_d of the complex is 10^{-14} M, which corresponds to an energy of binding of -18.9 kcal/mol (Schreiber and Fersht, 1993). The interaction is mainly electrostatic, with barnase having positive electrostatic potential and

barstar negative (Buckle *et al.*, 1994).

Three mutants of the barnase-barstar complex have been crystallized. In each mutant either a van der Waals, long range charge-charge or hydrogen bond interaction have been deleted by simultaneous mutation of both residues involved in the interaction. Despite each of the double mutants being destabilised compared to the wild type, only local rearrangements were detected. In the crystal structure, the cavities created by the mutations were filled by water molecules (Vaughan *et al.*, 1999). The ability of the water molecules to mimic the deleted side-chains has also been noted in the case of lysozyme-antibody interaction (Chacko *et al.*, 1995).

Water molecules buried at the protein-protein interface mediate many intermolecular hydrogen bonds and increase the shape complementarity. Perona *et al.* (1993) have shown that water-mediated interactions have the same effect on binding affinity as direct hydrogen-bonding electrostatic contacts.

Tight binding protein-protein complexes usually involve a large surface area (more than 600 $\mathring{A}^2$ of each protein) buried at the interface (Janin and Chothia, 1990; Janin, 1995). Chothia and Janin (1975) have suggested that the surface area involved in binding is proportional to the binding affinity of the complex. However, there are many exceptions. For example the binding surface between actin and DNase I is greater than that of barnase-barstar, however the binding affinity is five orders of magnitude higher for the latter complex (Mannherz *et al.*, 1980; Schreiber and Fersht, 1993). It seems likely the surface area involved in binding is not the only determinant of affinity.

Probably a limited number of amino acids (10-30) from each protein contribute substantially to the stability of the complex (Cunningham and Wells, 1991; Hawkins *et al.*, 1993; Kelley and O'Connell, 1993). The neighboring amino acids may also be important in determining the specificity of interaction by electrostatic or steric effects (Clackson and Wells, 1995).

Chapter 2

Materials and Methods

2.1 Media and buffers

1) L-Broth
10g Difco Bacto tryptone
5 g Difco Bacto yeast extract
5g NaCl
dH_2O to 1 liter

pH adjusted to 7.2 with NaOH before autoclaving

2) L-Agar
10 g Difco Bacto tryptone
15 g Difco Bacto yeast extract
15 g Difco agar
5 g NaCl

pH adjusted to 7.2 with NaOH before autoclaving

3) BBL-Agar
10g Baltimore Biological Labs. Trypticase
5 g NaCl
10 g Difco agar
dH_2O to 1 liter

4) BBL-Top Agar
10g Baltimore Biological Labs. Trypticase
5 g NaCl
6.5 g Difco agar
dH_2O to 1 liter

5) Phage buffer
7 g Na_2HPO_4
3 g KH_2PO_4
5 g NaCl
10 ml 0.01 M $MgSO_4.7H_2O$
10 ml 0.01 M $CaCl_2$
1 ml 1 % (w/v) gelatin
dH_2O to 1 liter

6) Antibiotics	Ampicillin at 100 μg/ml Chloramphenicol at 25 μg/ml Tetracycline at 100 μg/ml
7) TE buffer	10 mM Tris-HCl pH 8.0 1 mM EDTA pH 8.0

Adjusted to appropriate pH with HCl

8) 10x TBE buffer	108 g Tris base 55 g boric acid 40 ml 0.5 M EDTA, pH 8.0 dH_2O to 1 liter

2.2 Plasmid construction

The pAR2993, pAR3786 and pAR3790 vectors that contains the T7 *ocr* gene were kindly provided by Alan H. Rosenberg and William Studier (Brookhaven Laboratory). The pET-1 plasmid derives from pBR322 in which the ϕ10 promoter of T7 RNA polymerase was inserted at the *Bam*HI site. The orientation is such that transcription is directed counterclockwise, opposite to that from the *tet* promoter (Studier *et al.*. 1990). The full-length Ocr clone pAR2993 has *Alu* fragment of T7 DNA (bp 837-1379) inserted into the *Bam*HI site (GGATCC) of pET-1 using *Bam*HI linkers (CCGGATCCGG). The truncated clones were made from the full-length clone. The plasmid pAR2993 was cut with *Rsa*I enzyme at T7 bp 1258 or with *Hga*I enzyme at T7 bp 1222. Appropriate palindromic linkers were added to supply a TAA termination codon and a *Bam*HI site. The newly obtained plasmids were termed pAR3786 and pAR3790 respectively. The linker attached to the *Rsa*I site, ACTAAGGATCCTTAGT, defines the right end of the T7 DNA in pAR3786 (bp 837-1260) and the linker attached to the *Hga*I site, TAAGGTCCTTA, defines the right end of the T7 DNA in pAR3790 (bp 837-1227).

2.3 Plasmid purification

A Quiafilter Midi Kit from Qiagen was used for plasmid purification. Plasmids used in this thesis are derivatives of pBR322, a high copy number plasmid. Usually a 5 ml LB medium bottle containing suitable antibiotics was inoculated with a single colony. The culture was grown over night at 37 °C on shaker. Next day 100 μl over night culture was inoculated in 50 ml LB medium (plus antibiotics) and the culture was grown up at 37 °C. About 25 ml over night culture was centrifugated for 15 minutes at 8000 rpm. The supernatant was removed and the bacterial pellet was resuspended in 4 ml resuspension buffer from the Quiafilter kit. The buffer contains RNase A that digest the bacterial RNA released. The following purification steps were done according to the Quiafilter Midi Kit manual. The resuspended bacterial cells were lysed with a buffer containing NaOH (5 minutes at room temperature), then the solution was neutralized with a buffer that contains sodium acetate. A Quiafilter cartridge barrel was used in order to separate the soluble fraction that contains the plasmid DNA from the insoluble fraction. For further purification steps a Quiafilter column was used. The column contains a resin that retains the plasmid DNA and allow the other lysate contaminants to flow through. The plasmid DNA was then eluted with 5 ml Tris buffer, pH 8.0 and precipitated with 3.5 ml isopropanol followed immediately by centrifugation for 30 minutes at 10000 rpm. The DNA pellet was washed with 2 ml 70 % EtOH. After another centrifugation for 2 min at 10000 rpm, the supernatant was removed and the DNA pellet air-dried for 10 minutes at room temperature. Finally the pellet was resuspended in 200 μl TE buffer and stored at 4 °C.

2.4 Plasmid sequencing

The sequencing was done using the "Sequenase Version 2.0 DNA Sequencing Kit" purchased from Amersham.

2.4.1 Alkaline–denaturation

The Sequenase kit gives good results when the template to be sequenced is a ssDNA. However the Sequenase kit allows direct sequencing of the double strand plasmid DNA if the dsDNA is denaturated in advance. The best denaturing method is alkaline denaturation. The denaturing solution contains 10 μl pAR2993 (stock concentration of 86.5 μg/ml), 5 μl of 2mM EDTA, 5 μl of 2M NaOH and 30 μl of sterile water. The mixture was incubated for 30 minutes at 37°C, then neutralized by adding 5 μl of 3M Sodium Acetate (pH 5.0). The DNA was precipitated by adding 100 μl of 96 % EtOH and incubating at -70 °C for one hour. A centrifugation at 13000 rpm for 10 minutes was carried out to collect the precipitate. The supernatant was removed and 150 μl 70 % EtOH was added. After a short centrifugation at 13000 rpm the EtOH was removed and the DNA pellet was placed in a desiccator for 5 minutes. The DNA pellet was resuspended in 7 μl dH_2O.

2.4.2 Annealing reaction

For control DNA, 5 μl M 13 ssDNA, 2 μl Sequenase buffer, 1 μl Primer (-40) and 2 μl of water were mixed together. The sample DNA was obtained by combining 7 μl pAR 2993 plasmid DNA (after alkaline-denaturation), 2 μl of Sequenase buffer and 1 μl primer (50-100 ng). Both sample and control DNA were annealed by heating for 2 minutes at 65 °C then cooling slowly to 35-40 °C over 20-30 minutes. The samples were briefly centrifuged and kept on ice for further use. While cooling, four different ependorff tubes were filled with 2.5 μl of each Termination Mixture (G. A. T and C).

2.4.3 Labelling reaction

To ice-cold annealed DNA Mixture (10 μl) was added 1 μl of 0.1 mM DTT, 2 μl Diluted Labelling Mix (5 times diluted), 0.5 μl [^{35}S] dATP, 2 μl Diluted

Sequenase Polymerase. Samples were briefly centrifuged then incubated at room temperature for 2 minutes.

2.4.4 Termination reactions

First of all, the termination tubes were prewarmed for 1 minute at 37°C. Then 3.5 μl of the labelling reaction were transfered to each termination tube (G, A, T and C), followed by a briefly centrifugation, then incubation at 37 °C for 2 minutes.

2.4.5 Stop reaction

Reactions were stopped by adding 4 μl of Stop Solution. Labelled samples can be stored at -20 °C for a week.

2.4.6 Denaturing gel electrophoresis

1)Preparation and pouring the sequencing gel

A gel-electrophoresis sandwich was made by clamping two glass plates, one washed with EtOH and the other with 5 % Repelcote (octamethyl cyclotetrasiloxane). The two glass plates were separated with two 0.4 mm Plastikard spacers. 50 ml of denaturing acrylamide solution was prepared for each gel by mixing 20 g urea, 5.6 ml 10x TBE buffer, 4.7 ml Long Range Acryl/Bisacryamide solution and 22 ml prewarmed dH_2O. The solution was stirred for 3 minutes. After adding 43 μl TEMED and 120 μl 10 % AMPS, the denaturing acrylamide solution was poured between the glass plates of the gel-electrophoresis sandwich. When solution reached the top of the plates, the gel sandwich was laid down and a 0.4 mm shark-tooth comb was inserted into solution, 2 mm below the top of the plates. The gel was allow to set for at least 2 h.

2)Prerunning

About 800 ml 1x TBE buffer was poured in the top reservoir and 200 ml in the bottom reservoir. The shark-tooth comb was removed and the top of the gel was rinsed with 1x TBE buffer. After cleaning, the shark-tooth comb was reinserted into gel sandwich with points just barely sticking into gel. The gel was preheated at 65 W for about 30 minutes until the temperature became 55 °C.

3)Loading and running

When the denaturing gel was ready for loading, the samples were heated at 75-80 °C (in steam) for 2 minutes, then loaded immediately on the gel. Just before loading, the wells were rinsed with TBE buffer to remove the urea that leached into them. When sequence of interest lay close to the primer, the gel was run at 40 W for about 2 h until the marker dye reached the bottom of the gel. When sequences far away from primer was to be sequenced, the run was continued for about 4 h. In this case the power used was decreased to 35 W because the gel temperature tended to increase beyond 55 °C.

4)Gel drying

The gel was transfer to blotting paper, covered with Saranwrap plastic film and dried on a vacuum drier at 80 °C. The Saranwrap film was removed before overnight exposure to autoradiographic film.

2.4.7 Solutions

1) Control DNA	0.2 μg/μl
2) Primer (-40)	0.5 pmol/μl
3) 5x Labeling Mix	7.5 μM dGTP 7.5 μM dCTP 7.5 μM dTTP
4) ddG Termination mix	80 μl dGTP, 80 μl dATP, 80 μl dCTP, 80 μl dTTP, 8 μl ddGTP, 50 mM NaCl
5) ddA Termination mix	80 μl dGTP, 80 μl dATP, 80 μl dCTP, 80 μl dTTP, 8 μl ddATP, 50 mM NaCl
6) ddT Termination mix	80 μl dGTP, 80 μl dATP, 80 μl dCTP, 80 μl dTTP, 8 μl ddTTP, 50 mM NaCl
7) ddC Termination mix	80 μl dGTP, 80 μl dATP, 80 μl dCTP, 80 μl dTTP, 8 μl ddCTP, 50 mM NaCl
8) 5x Sequenase buffer	200 mM Tris-HCl, pH 7.5 100 mM $MgCl_2$ 250 mM NaCl
9) 2x Stop Solution	95 % Formamide 20 mM EDTA 0.05 % Bromophenol Blue 0.05 % Xylene Cyanol FF
10) Sequenase Version 2.0 T7 DNA Polymerase	13 units/μl in 20 mM KPO_4, pH 7.4 1 mM DTT 0.1 mM EDTA 50 % Glycerol
11) Enzyme Dilution Buffer	10 mM Tris-HCl, pH 7.5 5 mM DTT 0.5 mg/ml BSA

The concentrations of the dNTP and ddNTP solutions are not indicated in the "Sequenase Version 2.0 DNA Sequenching Kit" manual.

2.5 Determination of the extinction coefficient

Wild type Ocr was desalted in 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$, 7 mM β-mercaptoethanol, pH 8.0 buffer. Then 100 μl buffer was placed in a 0.3 cm pathlength cuvette and the baseline was set to zero. The buffer was removed and 100 μl desalted Ocr (83 μM concentration) was added so that the $O.D.^{280}$ was about 0.8. The UV spectrum was stored in the spectrophotometer memory then 267 μl GdnHCl from 8.25 M stock was added in order to get a 6.0 M GdnHCl final concentration. The UV spectrum of the denatured protein was also stored and both the native and denatured spectra were scaled to the same concentration (the spectrum of the native protein was multiplied with $100/367 = 0.2725$).

It was also necessary to subtract the UV spectrum of the GdnHCl. For this purpose 100 μl buffer was added in 0.3 cm pathlength cuvette and the baseline was set to zero. Then 267 μl GdnHCl from 8.25 M stock was added. The spectrum of the GdnHCl was then subtracted from the spectrum of the denatured protein. The corrected spectrum of the denatured protein was subtracted from the spectrum of the native Ocr and the difference at λ 280 nm was calculated.

The protein was desalted using a small G-25 gel filtration column.

Determination of GdnHCl concentration

GdnHCl is quite hygroscopic and is difficult to prepare stock solutions by weight. Consequently, the molarity of the GdnHCl stock solution is based on refractive index measurements and the following equation (Nick and Scholtz, 1997):

$$Molarity = 57.147(\Delta N) + 38.68(\Delta N)^2 - 91.60(\Delta N)^3 \qquad (2.1)$$

where

ΔN is the difference in refractive index between the denaturant solution and buffer at the sodium D line

2.6 Absorbance

2.6.1 General considerations

Almost all biological molecules have a distinct absorption in the ultraviolet or sometimes, in the visible range so quantization can be done easily by measuring at one specific wavelength. Proteins absorb light and emit radiation in the UV range of the spectrum. The aromatic residues Trp and Tyr are responsible for more than 90 % of the absorption in the 270-300 nm range. To some extent the absorbance is also due to the Cys and Phe residues. According to the Beer-Lambert law, the absorbance of a compound has a linear relationship to its absorbance at a defined wavelength.

$$A = \varepsilon cd \tag{2.2}$$

where

A is the absorbance at a defined wavelength
ε is the molar extinction coefficient ($M^{-1}cm^{-1}$)
c is the concentration (mol/liter)
d is the pathlength (1 cm)

2.6.2 Buffers and chemicals

It is better to avoid buffers that absorb in the UV range of interest. The purity of the chemicals used for making buffers is of utmost importance because impurities may lead to UV absorbance in the range of protein absorbance. All buffers used were made from ultrapure chemicals and were filtered using a 0.2 μM Acrodisk filter. Buffers were kept at 4 °C for maximum one week.

2.6.3 Cuvettes and instruments

A 0.15 or 0.3 cm pathlength cuvette was used, depending on the volume of the sample. The cuvettes were cleaned using dH_2O and EtOH and then air-dried. From time to time the cuvettes must be washed with chromic acid in order to

remove the biological sample that could stick onto the cuvettes walls. All spectroscopic measurements were done in a Perkin-Elmer Lambda spectrometer. The spectrometer lamp must be switch on at least 10-15 minutes before the actually measurements start. The following settings were used:

Ordinate mode	Absorption
slitwidth	2.0 nm
Scan speed	240 nm/min
Response	0.5 s
Lamp	332.8 nm
Cycle/Time	1/0.05 min
Y axis (O.D.)	minimum 0
	maximum 1
X axis (wavelength)	minimum 240 nm
	maximum 340 nm

Depending on the absorbance values, the Y axis could be set at different values.

2.6.4 Calculations

For proteins

A suitable buffer was placed in a 0.15 or 0.3 cm pathlegth cuvette and the baseline was set to zero. The cuvette was air-dried and the protein sample was added. The spectrum was recorded and printed out and the distance between the baseline and the spectrum at λ 280 nm was measured.

In the Beer-Lambert equation the pathlength (d) is 1 cm. If different pathlength are used, a correction must be made. The molar concentration would be obtained dividing the corrected absorbance value to the molar extinction coefficient of the protein.

For nucleic acids

The same settings as for proteins were used. The only difference was that the X axis was set at minimum 200 nm and maximum 300 nm. First the buffer baseline was set to zero with TE buffer. Then buffer was removed from the cuvette, DNA sample was added and the spectra recorded. The absorbance was

measured at λ 260 nm making corresponding corrections for using a 0.15 or 0.3 cm pathlength cuvette. The concentration of the nucleic acids was calculated assuming that 1 $O.D.^{260}$ unit correspond to $\approx$ 50 μg/ml (Sambrook *et al.*, 1989). This approximation is accurate enough for the experiments done in this thesis.

2.7 Site Directed Mutagenesis

2.7.1 Protocol

Site-specific mutations in the coding sequence of the T7 Ocr gene were created by PCR mutagenesis using a "QuickChangeTM Site-Directed Mutagenesis Kit" from Stratagene. Pfu DNA polymerase, that has 12-fold higher fidelity in DNA synthesis than Taq DNA polymerase, was used to replicate both strands with high fidelity. For replacing the chosen amino acids with cysteine, it was necessary to replace one or two base pairs. Two oligonucleotide primers (one for each strand) containing the desirable mutation were chosen for each mutation. Following temperature cycling, the reaction mixture will contain both the mutated and parental plasmids. In order to remove the parental plasmid DNA the reaction mixture was treated with *Dpn*I endonuclease, a restriction enzyme specific for methylated and hemimethylated DNA. The parental DNA from almost all *Escherichia coli* strains is *dam* methylated at the recognition site for *Dpn*I (5'-G^{m6}ATC-3'). However, the newly synthesized DNA will not be methylated and will be protected against *Dpn*I enzyme. The mutated plasmid is then transformed into *Epicurian Coli* XL1-Blue supercompetent cells.

The following considerations should be taken into account when designing oligonucleotide primers:

- both primers must contain the desired mutation and anneal to the same sequence on opposite strands of the plasmid.

- the primers should be between 25 and 45 bases in length

Table 2.1: Conditions fullfiled by the chosen primers

Mutation	primer length	GC content (%)	base(s) changed
N4C	44	43.2	A to T, A to G
D25C	41	43.9	G to T, A to G
N43C	44	54.5	A to T, A to G
D62C	39	53.8	G to A, A to G
S68C	39	53.8	C to G
W94C	37	45.9	G to T

- the GC content to be minimum 40 % and should be terminated in one or more G or C bases
- the desired mutation should be in the middle of the primer with about 10-15 correct bases on both sides.

All of these characteristics are summarized in table 2.1

Table 2.2 shows the primer sequences for all six mutation.

2.7.2 Setting up the reactions

1) control reaction
5 μl of 10x reaction buffer
2 μl (10 ng, 0.003 nM) of pWhitescript 5.7 control plasmid
1.25 μl (125 ng, 22 nM) of oligonucleotide control primer #1
1.25 μl (125 ng, 22 nM) of oligonucleotide control primer #2
1 μl of 10 nM dNTP mix (2.5 mM each NTP)
dH_2O to a final volume of 50 μl

Then add 1 μl of native Pfu DNA polymerase (2.5 U/μl)

2) sample reaction
5 μl of 10x reaction buffer
1 μl (86 ng) of dsDNA template
1 μl (125 ng) of oligonucleotide primer #1 (forward strand)
1 μl (125 ng) of oligonucleotide primer #2 (reverse strand)
1 μl of 10 mM dNTP mix (2.5 mM each NTP)
dH_2O to a final volume of 50 μl

Then add 1 μl of native Pfu DNA Polymerase (2.5 U/μl). Overlay each reaction with 30 μl of mineral oil.

The primers were purchased from Oswel.

Table 2.2: The oligonucleotide sequence of the primers used for Site-Directed Mutagenesis

Protein	Strand	Sequence
N4C	Forward	5'-CAAGATGGCTATGTCT*TGC*ATGACTTACAACAACGTTTTCGACC-3'
	Reverse	3'-GTTCTACCGATACAGA*ACG*TACTGAATGTTGTTGCAAAAGCTGG-5'
D25C	Forward	5'-GCTGAAAGAAAACATCCGTTAT*TGT*GACATCCGTGACACTG-3'
	Reverse	3'-CGACTTTCTTTTGTAGGCAATA*ACA*CTGTAGGCACTGTGAC-5'
N43C	Forward	5'-GCTATTCACATGGCTGCCGAT*TGT*GCAGTTCCGCACTACTACGC-3'
	Reverse	3'-CGATAAGTGTACCGACGGCTA*ACA*CGTAACGGCGTGATGATGCG-5'
D62C	Forward	5'-GGCAAGTGAGGGCATT*TGC*CTTGAGTTCGAAGACTCTGG-3'
	Reverse	3'-CCGTTCACTCCCGTAA*ACG*GAACTCAAGCTTCTGAGACC-5'
S68C	Forward	5'-GACCTTGAGTTCGAAGAC*TGT*GGTCTGATGCCTGACACC-3'
	Reverse	3'-CTGGAACTCAAGCTTCTG*ACA*CCAGACTACGGACTGTGG-5'
W94C	Forward	5'-GCAATTAACGATTGACCTC*TGT*GAAGACGCAGAAGAC-3'
	Reverse	3'-CGTTAATTGCTAACTGGAG*ACA*CTTCTGCGTCTTCTG-5'

2.7.3 Cycling the Reactions

The mutagenic conditions recommended by the QuickChange Site-Directed Mutagenesis Method protocol were optimized to the following values:

Segment	Cycles	Temperature	Time
1	1	95 °C	30 seconds
2	12	95 °C	30 seconds
		58 °C	2 minutes
		68 °C	12 minutes

Following temperature cycling, the samples were kept on ice for two minutes to cool the reaction to 37°C.

2.7.4 Product digestion

1 μl *Dpn*I enzyme (10 U/μl) was added. Each reaction sample was gently mixed by pipetting the solution up and down. The reaction mixture was spanned down for 1 min at 10000 rpm then incubated at 37 °C for 1 hour to digest the parental supercoiled DNA.

2.7.5 Transformation

One μl of *Dpn*I treated DNA was added to 50 μl *Epicurial Coli* XL-Blue supercompetent cells. After incubation on ice for 30 minutes, the transformation reaction was heat pulsed for 45 seconds at 42 °C. Then 500 μl of NZY + broth (preheated to 42 °C) was added and the mixture was incubated for 1 h at 37 °C. 250 μl control reaction was plated on LB-ampicillin-methycillin agar plates (the plates also contain 20 μl of 10 % X-Gal and 20 μl of 100 mM IPTG). The entire volume of each sample reaction was plated on the LB-ampicillin plate. All plates were incubated overnight at 37 °C.

2.7.6 Isolating and sequencing of the mutated plasmid DNA

After colony purification, 50 ml LB-ampicilin medium was inoculated with a single colony and incubated at 37 °C over night. From this overnight culture the plasmid DNA was isolated using the "Quiafilter Midi Plasmid Kit" from Quiagen. The newly-obtained plasmid DNA was sequenced using "Sequenase Version 2.0 DNA Sequencing Kit" purchased from Amersham following the same strategy as for pAR2993 plasmid.

2.7.7 Solutions

1) NZY + Broth

10g NZ (casein hydrolysate)
5 g yeast extract
5 g NaCl
Autoclave

Prior to use add:
12.5 ml 1 M $MgCl_2$
12.5 ml 1 M $MgSO_4$
20 ml 20 % (w/v) glucose

2)10x Reaction Buffer	100 mM KCl 100 mM $(NH_4)_2SO_4$ 200 mM Tris-HCl (pH 8.8) 20 mM $MgSO_4$ 1 % Triton X-100 1 mg/ml nuclease-free bovine serum albumin (BSA)

2.8 Labelling of Ocr(Cys) proteins

2.8.1 Labelling with 1,5 I-AEDANS

100 μl of Ocr(Cys) proteins of 50 μM concentration were incubated overnight at 4 °C, in the dark, with 50-fold molar excess of N-(iodoacetylaminoethyl)-5-naphtylamine-1-sulphonic acid (1,5 I-AEDANS) in 20 mM Tris , 20 mM NH_4Cl, 6 mM $MgCl_2$ pH 8 buffer. The incubation was stopped by adding β-mercaptoethanol to 100 mM concentration. The unbound 1,5 I-AEDANS was removed using a PD-10 Sephadex G50 gel filtration column. The PD-10 column was equilibrated in the same buffer plus 7 mM β-mercaptoethanol. The two cysteine residues in Ocr(N4C) protein (one in each monomer) form a disulphide bond (as showed in the Results chapter) so that it was necessary for the labelling buffer to also contain β-mercaptoethanol to keep the cysteine residues in the reduced state. The β-mercaptoethanol also contains -SH moieties that will react with 1,5 I-AEDANS so that it was necessary to add the fluorophore in a huge excess to be sure that the final 1.5 I-AEDANS concentration is higher that the final β-mercaptoethanol concentration. I used a molar ratio of 50:1 excess fluorophore (calculated as the difference between the final 1,5 I-AEDANS concentration and the β-mercaptoethanol concentration) to protein.

Measurements of the fluorescence intensity in the steady state were performed on a Perkin Elmer LS50B spectrofluorometer using an excitation wavelength of 330 nm and emission at 480 nm. The concentration of 1,5 I-AEDANS bound to proteins was calculated from the absorption spectrum, using a molar extinction coefficient (ε_M) of 6100 at 337 nm (Hudson and Webb, 1973).

The concentrations of all Ocr(Cys) except Ocr(W94C) proteins were calculated from the absorption spectrum using the ε_M of the wild type Ocr (32095 M^{-1} cm^{-1} at 280 nm). For Ocr(W94C) protein I used a theoretical ε_M (calculated from the amino acid sequence) of 20600 at 280 nm. The concentration of the labelled protein was calculated from the absorbance at 280 nm after subtracting the absorbance of 1.5 I-AEDANS at 280 nm (the ε_M of 1.5 I-AEDANS at 280 nm is 1060 M^{-1} cm^{-1}.

2.8.2 Labelling with Pyrene Maleimide

N-(1-pyrene)maleimide (PM) was purchased from Molecular Probes. A PM stock solution was made by dissolving PM powder in Dimethyl Sulphoxide (DMSO) due to the high PM insolubility in water. The labelling reaction was done by using a 100 mM phosphate, pH 7.0 buffer instead of Tris buffer. Ocr(Cys) proteins (100 μl of 50 μM) were incubated over night at 4 °C , in the dark, with PM (molar ratio 50:1). The labelling reaction was stopped by adding β-mercaptoethanol to 100 mM final concentration and the unbound fluorophore was removed using a PD-10 gel-filtration column equilibrated in 100 mM phosphate, 7 mM β-mercaptoethanol, pH 7 buffer (Strasburgh *et al.*, 1985).

The labelling of the Ocr(N4C) protein was done in 100 mM phosphate, 7mM β-mercaptoethanol, pH 7.0 buffer using a molar ratio of 50:1 excess PM to protein concentration.The degree of labelling was determinated spectrometrically using a value of $\varepsilon_M = 40000$ M^{-1} cm^{-1} for PM at 345 nm (Wu *et al.*, 1976). The protein concentration was calculated from the absorbance at 280 nm after subtracting the absorbance of PM at this wavelength ($\varepsilon_M = 25000$ M^{-1} cm^{-1} for PM at 280 nm).

2.9 Transformation and overexpresion

One μl pAR2993 plasmid (86.5 μg/μl stock concentration) was added to 100 μl BL21(DE3)pLysS competent cells. The mixture was heat-shocked for 45 seconds at 42 °C, then 1 ml LB was added, followed by an incubation at 37 °C for 1 h. After incubation, 10 ul mixture was platted on a LB-Am-Cm plate and incubated overnight at 37 °C. Next day, 200 ml LB-Am-Cm were inocculated with a single colony and incubated overnight at 37 °C. Eight 2 l flasks, each containing 500 ml fresh LB-Am-Cm inocculated with 25 ml overnight culture, were incubated at 37 °C for about 3 h (on a shaker) until the $O.D.^{600}$ became $\approx$ 0.5, then IPTG was added to a final concentration of 0.5 mM. After addition of IPTG the cells were incubated for another 2 h at 37 °C. The cells were then pelleted at 6000 g for 30 minutes. The obtained cell pellet ($\approx$ 5 g) was kept at -20 °C.

As an expression strain, BL21 has the potential advantage that it should be deficient in the *lon* protease and the *ompT* outer membrane protease that can degrade proteins during purification (Grodberg and Dunn, 1988).

Bacteriophage DE3 is a λ derivative that has the immunity region of phage 21 and carries a DNA fragment containing the *lacI* gene, the *lac*UV5 promoter, the beginning of the *lacZ* gene, and the gene for T7 RNA polymerase. Sometimes the target gene codes for a protein that is toxic for the BL21 cell. When this problem occurs it is necessary to reduce the basal level of T7 RNA polymerase that is responsible for the transcription of the target gene in the uninduced cells.

The T7 RNA polymerase is inhibited by the T7 lysozyme that is expressed by the pLysS plasmid. The plasmid contains the ϕ3.8 promoter for T7 RNA polymerase situated upstream to the lysozyme gene. The lysozyme provided by pLysS plasmid will not impair the expression of the *ocr* gene because the ϕ10 promoter is stronger than ϕ3.8 promoter.

2.10 Protein purification

2.10.1 Cell lysis

The cell pellet obtained after overexpresion was taken out from the freezer and left for 15 minutes to defrost. The defrosted pellet was resuspended in the purification buffer (usually I added about 5 ml buffer per gram of cell pellet). Benzaminidine and PMSF were also added to a final concentration of 10 mM, in order to inhibit the proteases released after the cell lysis. The obtained solution was ultrasonicated for a number of minutes equal to the grams of cell pellet, with intermittent cooling on ice after each minute. Cell debris was removed by centrifugation at 12000 g for 3 h.

2.10.2 Ion-exchange chromatography

The supernatant that contain the protein of interest was applied to a DEAE-Sepharose column. The column was first regenerated with 20 mM Tris, 300 mM NH_4Cl, 4 % glycerol, pH 8.0 buffer for 3 h at 48ml/h flow rate. The protein sample was loaded to column at the same flow rate. After all the protein sample was loaded, the column was eluted with buffer for about 2 h at 48 ml/h in order to remove the proteins that did not stick to the column. Five fractions were collected per hour. Ocr was eluted with 500 ml 0.3-1.0 M NH_4Cl gradient at 24 ml/h in the above buffer. The fractions containing the protein of interest were identified using UV spectroscopy and SDS-PAGE.

2.10.3 Gel-filtration

A Superdex 200 gel-filtration column (3x110 cm) was equilibrated with 240 ml of 20 mM Tris, 300 mM NH_4Cl, 4 % glycerol buffer, pH 8.0 for 10 h at 24 ml/h. Fractions containing the desired protein from DEAE were pooled and concentrated down to $\approx$ 3 ml using 10 Kd cutoff centrifugal concentrators (Flowgen). The protein sample was applied to Superdex 200 column and the protein was

eluted over night at 24 ml/h with about 300 ml of buffer. Four fractions were collected per hour. UV spectroscopy and SDS-PAGE were used to identify the fractions containing the protein of interest.

Note

The above conditions were used for purifying Ocr 2993 (wild type). After two purification steps, the wild type Ocr looks pure on UV spectroscopy and SDS-PAGE gel. For Ocr 3786, Ocr 3790 as well as the Ocr(Cys) proteins a further purification step (hydrophobic interaction chromatography) was necessary, since high amounts of contaminating DNA were still present.

We were not sure if Ocr 3786, Ocr 3790 or Ocr(Cys) proteins would stick on DEAE column as well as the wild type so we used a buffer with a lower salt concentration (50 mM instead of 300 mM NH_4Cl). The same buffer was used in the gel-filtration step. The buffer used for the purification of Ocr(Cys) proteins also contained 7 mM β-mercaptoethanol.

2.10.4 Hydrophobic interaction chromatography

A Phenyl Sepharose column was equilibrated with 20 mM Tris, 2.4 M $(NH_4)_2SO_4$, 4 % glycerol, pH 8 buffer for two hours at 48 ml/h. Fractions from Superdex, identified as containing Ocr protein were pooled and solid $(NH_4)_2SO_4$ was added to a final concentration of 2.4 M. The protein sample was applied on the column then washed with 2.4 M $(NH_4)_2SO_4$ for 1 h. The Ocr was eluted over night with 500 ml 2.4-0 M $(NH_4)_2SO_4$ gradient buffer using a flow rate of 24 ml/h.

The fractions containing the pure protein were pooled and concentrated down to about 1 ml. An equal amount of glycerol was added and the purified protein was kept at -20 $^\circ$ C.

2.11 High Performance Liquid Chromatography

A Rainin Dynamax 4.6 x 250 mm Hydropore-5-SEC column and a guard column were used for HPLC gel filtration assay. First of all wild type Ocr and Ocr(Cys) proteins were desalted using two PD 10 Sephadex G50 columns. One column was equilibrated in 20 mM Tris, 20 mM MES, 200 mM NH_4Cl, 10 mM $MgCl_2$, 0.1 mM EDTA, 5 M GdnHCl, pH 6.5 and the other in the same buffer plus 7 mM β-mercaptoethanol. Proteins samples were centrifuged at 10000 rpm for 15 minutes at 4 °C. 50 μl samples (about 20 μM) were injected onto the column. The flow rate was 0.5 ml/min and the absorbance was recorded at λ 280 nm.

2.12 SDS-PAGE

For all experiments a 15 % acrylamide gel was prepared.

Separating gel

Using two glass plates and two 0.75 mm spacers, a glass-plate sandwich was assembled. 10 ml solution was made by adding 5 ml solution A, 2.5 ml solution B, 0.1 solution C and dH_2O to 10 ml. Take 1 ml from this mixture add 1 μl TEMED. 10 μl 10 % AMPS and seal the edge of the glass-plate sandwich. In the remaining 9 ml mixture, add 9 μl TEMED and 90 μl 10 % AMPS. Pour the solution between glass plates and allow 15 minutes for solidification.

Stacking gel

Make 10 ml solution by mixing 2.0 ml solution A, 2.5 ml solution D, 0.1 ml solution C, 10 μl TEMED, 100 μl 10 % AMPS, H_2O to 10 ml. Pour the solution in the top of the separating gel, insert a teflon comb and allow to set for 1 h.

Sample loading

Sample protein was mixed with sample buffer (v:v) then the mixture was loaded onto the gel. A 25 mA/gel was applied for about 1 h until the bromphenol blue marker reached the bottom of the gel. The gel was stained for 1 h in Coomassie Blue, then destained for 5 h in destaining solution.

Solutions

A) 30 % acrylamide/bis acrylamide

B) 1.5 M Tris-HCl pH 8.8	Tris 18.15 g H_2O to 80 ml adjust pH to 8.8 H_2O to 100 ml
C) 10 % SDS	10 g 99 % pure SDS H_2O to 100 ml
D) 0.5 M Tris pH 6.8	6 g Tris H_2O to 80 ml adjust pH to 6.8 with HCl H_2O to 100 ml
E) Sample buffer	2.5 ml solution D 2.0 ml solution C 2.0 ml glycerol 0.1 ml 0.1 % Bromphenol Blue 200 μl β-mercaptoethanol H_2O to 10 ml
F) Electrophoresis buffer pH 8.8	6 g Tris 28.8 g glycine 10 ml solution C H_2O to 990 ml adjust pH to pH 8.8 H_2O to 1000 ml
G) Staining solution	1 g Coomassie Blue R250 100 ml Acetic Acid 500 ml methanol H_2O to 1000 ml

H) Destaining solution	100 ml Acetic Acid 100 ml methanol H_2O to 1000 ml

2.13 Agarose gel electrophoresis

One gram of agarose was weighted then added to 100 ml 1x TBE buffer. The mixture was boiled, cooled down to 55 °C, poured in a sealed gel casting platform then a gel comb was inserted. EtBr was added to the agarose solution so that was no need for staining and destaining. After the gel has hardened, the comb was removed and the gel was placed into a electrophoresis tank. 1x TBE buffer was added until the gel was covered. The DNA samples were mixed with the loading buffer and 10 μl of the mixture was loaded onto the agarose gel. The electrophoresis was carried out at 10 V/cm of gel, until the bromphenol blue band reached the bottom of the gel.

2.14 Restriction assay

BL21(DE3)pLysS cells were transformed with 9 different plasmids: pAR2993, pAR3786, pAR3790, and six pAR2993(Cys). A conjugation between NK308 as a donor strain and the transformed BL21(DE3)plysS as a recipient strain was carried out. NK308 strain contains the F'101 plasmid which is transfered to BL21 cells by conjugation. F'101 plasmid contains *hsdR*, *hsdM* and *hsdS* genes.

The BL21(DE3)pLysS cells are resistant to chloramphenicol, whereas pAR plasmids confer resistance to ampicillin and F'101 plasmid to tetracycline. Using over night cultures, both donor and acceptor were grown up for about 2 h at 37 °C until the $O.D.^{600}$ became $\approx$ 0.4. The NK308 strains was grown up in LB and the transformed Bl21(DE3)pLysS cells in LB, chloramphenicol and ampicillin. Just before conjugation, the transformed cells were washed with LB in order to remove all the antibiotics. For conjugation, 2 ml of donor was mixed with 0.2 ml

of recipient and the mixture was gently shaken for 1 h at 37 °C. As a control, an untransformed BL21(DE3)pLysS culture was conjugated with NK308 strain. The conjugation mixture was diluted in phage buffer (10^{-2} and 10^{-4}) then plated on LB, Cm, Am, Tc plates. The plates were incubated over night at 37 °C. Next day the colonies were purified to get single colonies. A single colony was grown up in 5 ml LB, Am, Cm, Tc, at 37 °C overnight. One ml overnight culture was mixed with 3 ml warm BBL Top agar then the mixture was plated in the top of the BBl Bottom agar, Cm, Am, Tc, plates. After solidification, bacteriophage λ v.0 was spotted. Different λ phage dilutions were used: $10^{-2}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}$. λ v.k was also spotted as a control. The plates were incubated overnight at 37 °C. Plaques were then counted and the e.o.p. calculated as a ratio between the titer of lambda phage of r^- and r^+ strains.

2.15 Methylase assay

2.15.1 *In vivo* assay

In order to check the anti-methylation activity *in vivo* of the Ocr proteins, the following experiment was performed. One lambda phage plaque from the BBL plates obtained after performing the anti-restriction activity test *in vivo* was resuspended in 1 ml phage buffer. Two serial dilutions (10^{-1} and 10^{-2}) were made in phage buffer. A few drops of chloroform were added and the mixture was centrifuged for few a seconds. The obtained lambda phage was spotted on the top of two plates, one containing 3 ml BBL Top agar plus 1 ml NK301 strain and the other 3 ml BBL Top agar plus 3 ml 5k strain. The plates were incubated over night at 37 °C. Next day, plaques were counted and a relative e.o.p. calculated. NK301 strain has a r^+m^+ phenotype and 5k a r^-m^+ phenotype. Both strains have a C600 background.

Table 2.3: Conditions used for *Hpa*I assay

Reaction	Ocr (μl)	M_2S_1	SAM	BSA	H_2O (μl)	linearized DNA	uncut DNA
Sample	0.4	+	+	+	17.1	+	-
	1.2	+	+	+	16.3	+	-
	4.0	+	+	+	13.5	+	-
Controls	+	+	-	+	17.5	+	-
	-	+	-	+	21.5	-	+
	-	+	+	+	20.0	-	+
	-	+	-	+	21.5	+	-
	-	+	+	+	20.0	+	-
	-	-	+	+	21.0	+	-
	-	-	+	+	21.0	-	+

2.15.2 *In vitro* assay

A DH5α ΔMS strain was transformed with pBlue$Kan^R Hpa^+$ plasmid, then a plasmid DNA stock of 771 nM was made using a Quiagen Midiprep kit. The plasmid has a *Hpa*I recognition site (GTTAAC) and a *Eco*RI recognition site. The AAC sequence is also part of the recognition site of *Eco*KI methylase. If methylation occurs, the second adenine from AAC sequence is methylated and the *Hpa*I nuclease is refractory to *Hpa*I nuclease.

*Eco*RI digestion

First the plasmid DNA must be linearized with *Eco*RI. For one reaction I used 10 μl plasmid DNA (50 nM stock), 5 μl buffer 4 (from New England Biolab), 1 μl *Eco*RI and 34 μl nuclease-free H_2O. The samples were incubated for 1 hour at 37 °C. Reaction was stopped by heat inactivation for 5 minutes at 68 °C.

25 μl digested DNA, 2.5 μl buffer 4, 1.5 μl SAM (32 mM stock) and 2.5 μl BSA (10x 10mg/ml stock) were added in each sample. For each Ocr protein (2993, 3786, 3790, W94C, N43C, S68C, N4C, D25C, D62C), three reactions were made by adding 0.4, 1.2 or 4.0 μl respectively (from 250 μM stock) of Ocr, giving a

Table 2.4: The run parameters for DSC scan

No of scans	1
Post Cycles Thermostat (°C)	20
Starting Temperature (°C)	20
Final Temperature (°C)	100
Scan Rate (°C/h)	60
Prescan Thermostat (min)	20
Filtering Period (sec)	16

molar ratio of 0.4:1, 1.2:1 and 4.0:1 respectively. A few control reactions, where one or more components were missing, were also made. Table 2.3 summarise all the conditions used.

All sample and control reactions were incubated for 6 h at 37 °C. After 6 h the reactions were heat inactivated at 68 °C for 5 minutes. The last two control reactions were kept at room temperature. Half of the amount was digested with *Hpa*I, and the other half was left undigested.

*Hpa*I digestion

For each reaction 20 μl *Eco*RI digestion mixture was mixed with 2 μl buffer 4 and 1.0 μl *Hpa*I (500 u/ml). After 1 h incubation at 37 °C all reactions were heat inactivated for 5 minutes at 68 °C. Four μl of DNA loading buffer was added and 10 μl from this mixture was loaded onto an 0.8 % agarose gel. A 1 kb DNA Ladder was used as marker.

2.16 Differential Scanning Calorimetry(DSC)

All samples were desalted using a PD-10 column equilibrated in 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$ pH 8.0 buffer. The protein concentration was checked by UV spectroscopy (Shimatzu spectometer). Both sample and buffer were degassed before loading into a VP-DSC calorimeter. The calorimeter cell was washed with detergent, H_2O and buffer before each experiment. The reference cell

Table 2.5: The run parameters for ITC assay

Run Parameters	
Total Injections	26
Cell Temperature (°C)	25
Reference Power (μCal/sec)	20
Initial Delay (sec)	20
Syringe Concentration (μM)	64.54
Cell Concentration (μM)	4.8
Stirring Speed	310

was washed only once, then the buffer was left inside for all the other experiments. 750 μl protein was added in the sample cell and 750 μl in the reference cell. The cell volume is about 500 μl but a higher volume was added in order to remove the air bubbles that could appear, then the excess was removed. The cells were closed with a small led that allows the solutions to be kept under pressure in order to avoid evaporation. The run parameters are sumarised in table 2.4.

A DSC scan of the buffer used was also done. For each sample the contribution of the buffer to the DSC scan was substracted using the Origin 5.0 software.

2.17 Isothermal Titration Calorimetry (ITC)

The experiment was done in a VP-ITC calorimeter. Both the Ocr and methylase sample were desalted using a PD-10 column, previously equilibrated in 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$. The experiment consist of titrating Ocr protein (64.54 μM stock solution) into methylase (4.8 μM concentration) placed in the VP-ITC cell.

1700 μl methylase was added into the cell and aliquots of 10 μl Ocr were injected. The first Ocr aliquot contained only 1 μl. The heat of dilution was obtained by injection of Ocr into buffer and also buffer in buffer solution. The run parameters are sumarised in table 2.5.

All the corrections were done using Origin 5.0 software.

2.18 Steady state fluorescence

2.18.1 Stoichiometry

Steady state fluorescence data were collected with a Perkin Elmer LS 50B spectrofluorometer at 25 °C. The stoichiometry assay is based on the fact that the fluorescence intensity of a complex between two proteins is different from the sum of the individual fluorescence intensities of the two proteins. In order to get the stoichiometry of the Ocr-methylase complex, three sets of data were recorded: the fluorescence intensity of the Ocr, the fluorescence intensity of the methylase and the fluorescence intensity of the Ocr-methylase complex.

The *Eco*KI methylase sample was temperature equilibrated for 10 minutes at 25 °C. Then 100 μl of methylase (0.3 μM concentration) was placed in a 0.3 cm pathlength quartz cuvette. The protein sample was excited at 295 nm and the emission spectra recorded between 300 and 500 nm. Using a Hamilton syringe, a concentrated solution of Ocr (22 μM) was added until the molar ratio Ocr:methylase became about 7:1. At each titration point the fluorescence intensity of the complex was recorded. From the emission spectra, the fluorescence intensity at 340 nm was measured.

The fluorescence intensity of the methylase at the begining of the experiments (when no Ocr was added) was measured. Then, this value was multiplied with the dilution coefficient to get the methylase fluorescence intensity at each titration point.

There is a linear dependence between the fluorescence intensity and the concentration of Ocr. 100 μl buffer were placed in a 0.3 cm pathlength cuvette then aliquots of Ocr were added. The fluorescence intensity values were ploted against the Ocr concentration. The Ocr intensities at each titration point were calculated using the values obtained for the slope and the intercept with Y axis.

The fluorescence intensity sum of the Ocr and methylase at each titration point was calculated. The fluorescence intensities of the Ocr-methylase complex for

each titration point were substracted from the fluorescence sum of the individual components. The difference was plotted against the molar ratio Ocr:methylase and the stoichiometry was determined from the intersection of the extrapolated linear portions of the graph.

The fluorescence contribution of the buffer was substracted from the protein sample fluorescence. The absorbances of both methylase and Ocr were less than 0.1 at 295 nm which minimize the inner filter effect. Despite its very low value, the inner filter effect was calculated and the fluorescence intensities were corrected accordingly. The formula used for calculating the inner filter effect was:

$$F_{corr} \cong F_{obs} antilog\left(\frac{O.D.^{ex} + O.D.^{em}}{2}\right) \tag{2.3}$$

where

F_{corr}	corrected fluorescence
F_{obs}	observed fluorescence
$O.D.^{ex}$	optical density of the sample at the excitation wavelength
$O.D.^{em}$	optical density of the sample at the emission wavelength

The slitwidth used for excitation and emission were 5 and 10 nm respectively.

The stoichiometry between Ocr and nuclease was calculated in the same manner, using the same slitwidth for excitation and emission, the same excitation and emission wavelength and the same temperature. However, the protein concentrations used were were different: 0.1 μM for nuclease and 7 μM for Ocr.

2.18.2 Quenching

The fluorescence intensity of the Ocr(Cys)-AEDANS proteins was quenched with acrylamide. 100 μl of Ocr(Cys)-AEDANS proteins (1.0 μM protein concentration) were placed in a 0.3 cm pathlength cuvette. The sample was excited at λ_{ex} 337 nm and the emission intensity was monitored at λ_{em} 480 nm. The labelled proteins

were quenched with acrylamide (4.5 M stock) in 6 steps until the final acrylamide concentration became 0.5 M.

50 μl Ocr(Cys)-AEDANS (2.0 μM stock) and 50 μl methylase (4.0 μM stock) were placed in a clean, air-dried 0.3 cm pathlength cuvette so that the molar ratio methylase:Ocr was 2:1. This complex was then titrated with 4.5 M acrylamide.

The slitwidth for excitation and emission was 5 and 10 nm respectively. The experiment was done at 25 °C.

2.18.3 Fluorescence Resonance Energy Transfer

Labeling

Ocr(Cys) proteins from 50 % glycerol stock were applied onto a PD-10 G-25 column equilibrated in 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$, 7 mM β-mercaptoethanol, 5 M GdnHCl, pH 8.0 buffer. The proteins were eluted as monomers because of the GdnHCl present in the buffer. Half of the monomers were labelled with 1,5 I-AEDANS and half of them with 5-IAF. About 150 μl protein (100 μM concentration) was incubated with 1,5 I-AEDANS or 5-IAF at 4 °C, over night, in the dark (molar ratio excess fluorophore:Ocr(Cys) was about 50:1). The labelling reaction was stopped by adding β-mercaptoethanol to a final concentration of 100 mM. The unbound fluorophore was removed using a PD-10 column equilibrated in the same buffer.

The concentration of the fluorophores bound to the protein was determinated by U.V. spectroscopy using a molar extinction coefficient of $\varepsilon_M^{337} = 3600\ M^{-1}\ cm^{-1}$ for bound 1,5-I-AEDANS and $\varepsilon_M^{492} = 71000\ M^{-1}\ cm^{-1}$ for bound 5-IAF. To calculate the protein concentration, the absorbance contribution of 1,5-I-AEDANS and 5-IAF at λ 280 nm was first subtracted.

Renaturation

About 100 μl Ocr(Cys)-AEDANS and 100 μl Ocr(Cys)-AF (50 μM concentration) were mixed together. Four ml of buffer without GdnHCl was added and the

mixture was spun down to about 100 μl in a Whatman Vectaspin 3 10 kd cutoff concentrator.

The renatured protein mixture should contain 25 % AEDANS-AEDANS homodimers, 25 % AF-AF homodimers and 50 % AEDANS-AF heterodimers. Ocr(Cys)-AEDANS and Ocr(Cys)-AF proteins were renaturated in two separate concentrators to get AEDANS-AEDANS and AF-AF homodimers.

Assay

The sample protein was equilibrated for 10 minutes at 25 °C, then 70 μl renatured protein mixture (2.0 μM protein concentration) was placed in a 0.3 cm pathlength cuvette. The sample was excited at λ 337 nm and an emission spectrum between 400 and 600 nm was recorded.

An emission spectrum of an AEDANS-AEDANS protein homodimer was also recorded between 400 and 600 nm (λ_{ex} 337 nm). This spectra was normalized to the AEDANS concentration from the renatured protein mixture. The normalized spectrum was then subtracted from the emission spectrum of the renatured protein mixture. In the difference spectrum obtained, the fluorescence intensity at 520 nm was measured.

An excitation spectrum was also recorded for the renatured protein mixture using an emission wavelength of 520 nm. The fluorescence intensity at λ 492 nm was measured. A slitwidth of 10 nm was used for both excitation and emission.

2.18.4 Dissociation constant (K_d)

The fluorescence was measured with a FS 900 CDT Steady State T Geometry Fluorimeter. The Ocr(N43C) protein was labelled with 5-IAF over night at 4 °C in the dark. The unbound 5-IAF was removed using a PD-10 G-25 column equilibrated in 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$, 7 mM β-mercaptoethanol, pH 8.0 buffer. The concentration of the AF bound to the protein was calculated from the absorbance at 492 nm using a molar extinction coefficient ε_M = 71000

M^{-1} cm^{-1}. The protein concentration was determinated from the absorbance at λ 280 nm after subtracting the absorbance of 5-IAF at this wavelength. 1000 μl of Ocr(N43C)-AF were placed in a QS 1000 4x10 mm cuvette. Sample protein was excited at 492 nm and the emission spectrum was recorded between 510 and 570 nm. The protein concentration used was 50 pM, 100 pM, 200 pM or 1000 pM depending on experiment.

The assay was performed at 25 °C and the slitwidth used for excitation and emission was 5.4 nm.

Then methylase or nuclease proteins were added. The spectra were integrated between 510 and 570 nm and the values were fitted using Grafit software (Erithacus). Data were fitted to ligand one binding site equation which gave the estimated K_d.

$$F = \frac{F_M + F_{ML} K_d [L]}{1 + K_d [L]} \tag{2.4}$$

where

F_M	relative intensity of free form
F_{ML}	relative intensity of complexed form
[L]	concentration of ligand

2.19 Fluorescence anisotropy

2.19.1 Stoichiometry

The anisotropy was recorded in an L arrangement of a Perkin-Elmer LS 50B spectrofluorimeter. 100 μl of 21 bp ds oligonucleotide DNA (1.0 μM concentration) that contains the recognition site for *Eco*KI were placed in a 0.3 cm pathlength cuvette. The dsDNA was obtained by hybridization between two complementary single strands. An hexachlorofluorescein tag was attached at the end of the top strand.

*Eco*KI methylase or nuclease were titrated into the DNA solution. For each titration step five anisotropy values were recorded then a mean of these values was calculated. The protein was titrated until the anisotropy did not increase any further.

A concentrated solution of Ocr protein was added which displaced the labelled DNA and consequently the anisotropy started to decrease. The titration was stopped when the anisotropy reach a plateau. For each Ocr titration the anisotropy was recorded at 15", 30" 1', 1.30', 2', 3', 4', 5', 6', 7', 8', 9', 10' after titration. A mean of these values was calculated. A slightly time dependent changes in anisotropy were noticed.

The λ_{ex} was 510 nm and the emitted light was passed through an optical filter that allows light at λ 570 nm. The assay was performed at 25 °C and the slitwidth was 15 and 20 nm for excitation and emission respectively.

2.19.2 Ocr-methylase interface

Ocr(Cys)-AEDANS and Ocr(Cys)-PM were prepared as described previously. The anisotropy was recorded with a Perkin-Elmer LB 50S fluorimeter. 100 μl of 1.0 μM Ocr(Cys) labelled protein was titrated with increasing amounts of methylase until the molar ratio methylase:Ocr(Cys)-labelled protein became 2:1. The Ocr(Cys)-AEDANS sample was excited at λ_{ex} 330 nm and the anisotropy was measured at λ_{em} 480 nm. The Ocr(Cys)-PM sample was excited at 342 nm and the anisotropy was measured at 379 nm.

The assays were carried out at 25 °C and the slitwidth used was 15 and 20 nm for excitation and emission.

Chapter 3

Ocr biochemistry

3.1 DNA sequencing

Cloning of the *ocr* gene (wild type and two truncated forms) into the pET-1 vector was done in Studier's laboratory. After cloning, the newly obtained plasmids were not sequenced. Sequencing of the obtained plasmids was done in our laboratory before starting any overexpression.

The sequence obtained (Figure 3.2) for wild type *ocr*, as well as the two truncated forms, corresponded exactly with the sequence previously published (Dunn *et al.*, 1981).

3.2 Ocr purification

Ocr was purified from BL21(DE3)pLysS cells carrying pAR2993 plasmid. After induction with IPTG, cells were pelleted by centrifugation, resuspended in 20 mM Tris, 300 mM NH_4Cl, 4 % glycerol, pH 8 buffer then sonicated. After a further centrifugation, the pellet was discarded and the soluble fraction was applied onto DEAE-Sepharose column. Ocr stuck on DEAE-Sepharose column because of its overall negative charge at pH 8 (the isoelectric point is 3.7). The protein was eluted using a gradient between 0.3 and 1.0 M NH_4Cl. Ocr was eluted in one peak at about 0.8 M NH_4Cl.

Because of the high levels of overproduction, Ocr was followed during purifica-

tion by SDS-polyacrylamide gel electrophoresis. The SDS-gel (figure 3.1) shows a partially purified Ocr protein. In order to separate Ocr from the *E. coli* proteins that also stuck on the DEAE column, it was necessary to use a further step of purification.

Fractions containing Ocr protein were pooled, concentrated down to $\approx$ 3 ml and applied to a Superdex 200 gel-filtration column, equilibrated in the above buffer. The obtained Ocr protein was pure according to the SDS-PAGE gel (Figure 3.1). UV spectra of the fractions containing Ocr had an absorption ratio $A^{280}/A^{260} \approx 2$ indicating that protein sample was not contaminated with free DNA.

We also purified Ocr 3786 and Ocr 3790, two Ocr proteins that have 7 and 17 respectively amino acids deleted from the C-terminal end. The first two steps of purification were almost identical with those used for purification of wild type Ocr, the only difference being the ionic strength of the buffer. For Ocr 3786 and Ocr 3790 we used only 50 mM NH_4Cl because it was possible that the two truncated Ocr proteins have a lower affinity to the DEAE-Sepharose column than the wild type Ocr. Indeed Ocr 3786 came off the DEAE column at 0.54 M NH_4Cl and Ocr 3790 at 0.51 M NH_4Cl.

After two purification steps both proteins were pure according to the SDS-PAGE gel. However, the UV spectra showed a huge contamination with DNA. Consequently, a further step of purification was necessary.

Fractions eluted from gel-filtration column containing Ocr 3786 or Ocr 3790 proteins were pooled together and solid $(NH_4)_2SO_4$ was added to 2.4 M final concentration. The protein sample was then loaded on a Phenyl-Sepharose column equilibrated in 20 mM Tris, 50 mM NH_4Cl, 2.4M $(NH_4)_2SO_4$ pH 8 buffer. The elution was performed with 500 ml 2.4 $\rightarrow$ 0 M $(NH_4)_2SO_4$ gradient. The elution profiles showed two peaks, a broad one containing DNA and a sharp one containing the protein. The absorption ratio A^{280}/A^{260} was found to be ≈ 2.

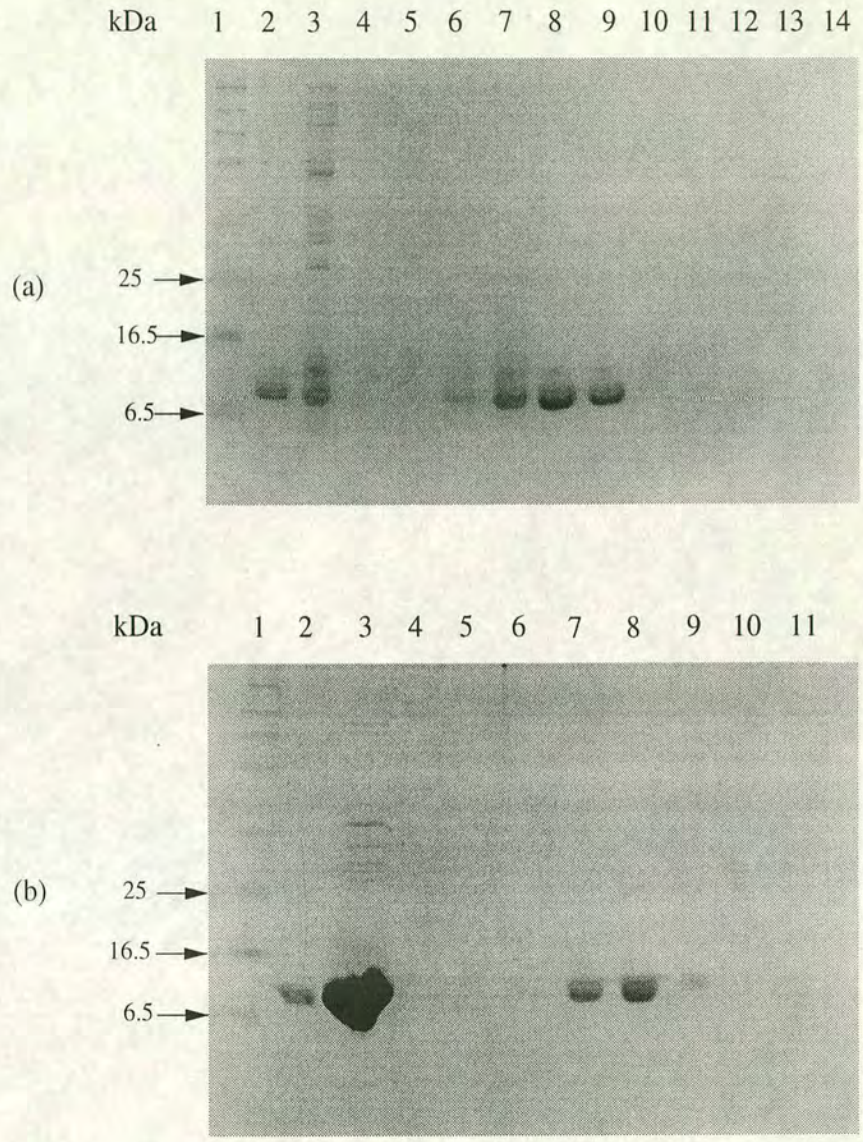


Figure 3.1: Analysis of wild type Ocr by SDS-gel electrophoresis. 15 % acrylamide gels were used and stained with Coomassie Blue. (a) Purification of Ocr on DEAE Sepharose column. Lane 1: molecular marker, lane 2: pure Ocr, lane 3: crude Ocr (before loading onto DEAE Sepharose column), lanes 4-14: Ocr fractions eluted with a gradient of NH_4Cl. (b) Purification of Ocr on Superdex 200 column. Lane 1: molecular marker, lane 2: pure Ocr, lane 3: Ocr collected from DEAE column but before loading onto Superdex 200, lanes 4-11 Ocr fractions eluted from the Superdex column.

3.3 Site-Directed Mutagenesis

3.3.1 Selection of the amino acids for SDM

One of the most important aims of this project was to identify the Ocr amino acids involved in binding of *Eco*KI methylase.

Certain fluorophores have the ability to report any change that could occur in their environment. These fluorophores could be used for labelling certain amino acids that are exposed on the Ocr surface. Then the labelled Ocr could interact with *Eco*KI methylase. If the labelled amino acid is not in the binding site then the environment of the attached fluorophore will not be changed after binding with the methylase. Consequently, the fluorescence intensity and the emission maximum will remain at the same values as for the unbound labelled Ocr. However, if the labelled amino acid is in the binding site, the environment of the fluorophore will change and either an increase or a decrease in the fluorescence intensity will occur and possibly a shift in the maximum emission wavelength.

There are fluorophores that specifically bind to arginine, lysine or cysteine (Lakowicz, 1999). When labelling a particular amino acid (say cysteine) then if more than one cysteine residue is exposed on the protein surface all of them will be labelled. The experiment will be meaningless because it will be difficult to say which amino acid was really in the binding site and which one was not. The problem was easily solved by exploiting the fact that Ocr protein has no cysteine residues, so that by replacing certain amino acid with cysteine and labelling it with a cysteine-specific fluorophore the exact position of the fluorophore will be known.

Six surface exposed Ocr amino acids were replaced with cysteine by Site-Directed Mutagenesis. Choosing these six amino acids was not an easy task. Ocr has a unique sequence in the data base, so that it was really difficult to get a very good secondary structure prediction.

If one of the predicted amino acids is not exposed on the surface it will be

```
                 ...*,....1....,....2....*....3....,....4..*.,....5....,....
        AA       AMSNMTYNNVFDHAYEMLKENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIFSVMASE|
2ndary structure         HHHHHHHHHHHHH    BBB    HHHHHHHHHHHHH    HHHHHHHHHHHH|
Reliability      9876685378999999988245522135514667778678763363266478999998 1|
Accessibility    bbbbbbbbebbebbbebbeebbbbbbbbb eebeebbebbbbebbbeb eebbeebbee|

                 6.*..,..*.7....,....8....,....9...*,....10....,...11...,.
        AA       GIDLEFEDSGLMPDTKDVIRILQARIYEQLTIDLWEDAEDLLNEYLEEVEEYEEDEE|
2ndary structure    BBBB      HHHHHHHHHHHHHHHHHHHHHHH                      |
Reliability      6874465178999956899999999998765766139                     |
Accessibility    bbebbb bbeebeeeeebbbbbebbbbeebbbebbeee                     |
```

Figure 3.2: Secondary structure prediction of the Ocr protein using PHD program. "H", "B" and "L" represents regions which are predicted to be α-helix, β-strand, and loops respectively. "e" indicates the amino acids predicted to be exposed on the protein surface and "b" amino acids predicted to be buried.

impossible to label it because the fluorophores are not able to penetrate the protein matrix. According to a secondary structure prediction program called PHD (Rost and Sander, 1993) Ocr is predicted to have 4 α-helices and 4 loops. As seen in figure 3.2 the PHD program indicates the degree of exposure for each amino acid. It is well known that hydrophilic amino acids are more likely to be solvent exposed than the hydrophobic ones and also that loop regions are usually on the proteins surface. Consequently, the amino acids chosen for Site Directed Mutagenesis were Asn4 from the first loop, Asp25 from the second, Asn43 from the third and Asp62 from the fourth. Ser68 from the fourth loop was also chosen, without any particular reason. The sixth amino acid to be replaced with cysteine was Trp94 from the fourth α-helix. Two main reasons were taken into consideration when choosing Trp94: first, previous experiments clearly indicate that this amino acid is solvent exposed and second, Trp94 seems to be involved in protein stability and/or activity. By replacing Trp94 with cysteine, it will be easy to demonstrate if an Ocr protein devoid of Trp is less stable than the wild type and if it is able to knock out the *Eco*KI methylase or nuclease.

3.3.2 Plasmid sequencing and Ocr(Cys) proteins purification

Site-specific mutations in the coding sequence of the T7 Ocr gene were created by PCR mutagenesis.

Figure 3.3 shows the sequence of the bottom strand for all six mutated plasmids. The whole strand was sequenced, not only in the region of interest. The upper strand was also sequenced, in order to confirm the presence of the desired mutation. All these precautions were necessary because it was very important to know that no other mutation except the desired one was introduced by Site-Directed Mutagenesis.

To emphasize this point it is worth mentioning the case of pAR2993(S68C) plasmid. The plasmid was sequenced in the region of interest and it was con-

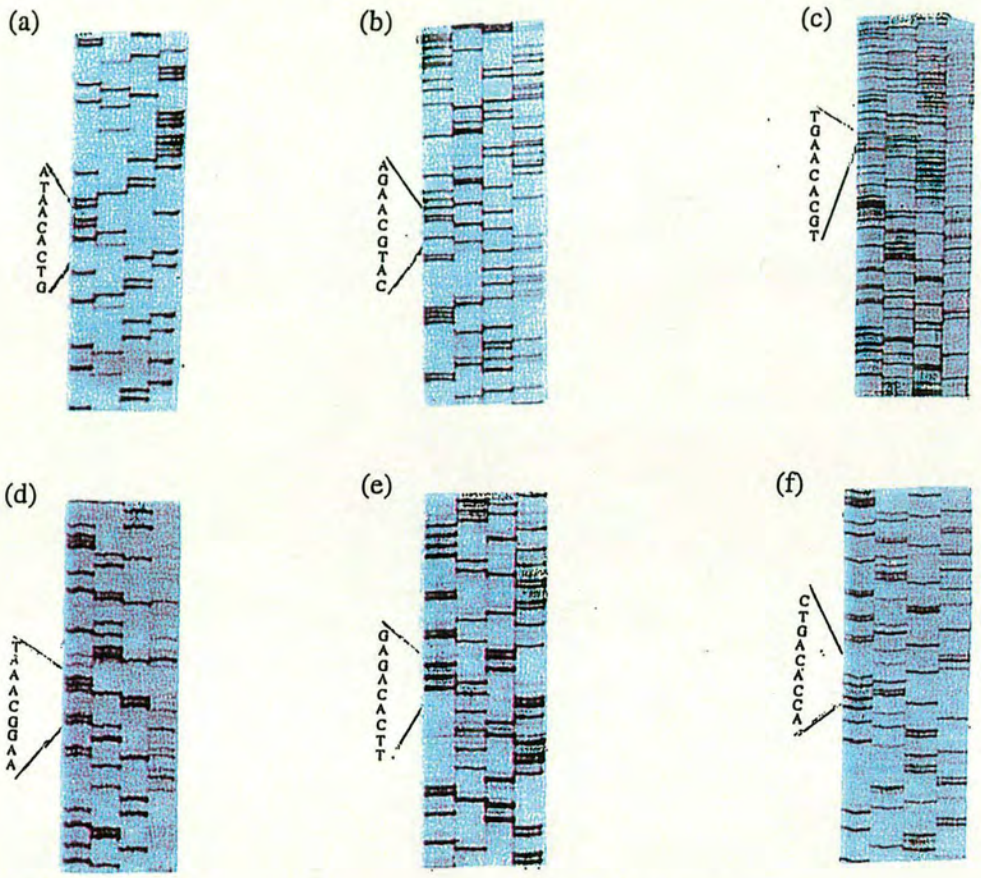


Figure 3.3: The nucleotide sequence of the mutated plasmids (bottom strand). In each of the six sequences the first lane represents A, the second C, the third G and the fourth T. For each sequence the codon for Cys, TGT or TGC (ACA or GCA respectively in the bottom strand, reading from 5' to 3'), is easily identified. (a) D25C, (b) N4C, (c) N43C, (d) D62C, (e) W94C, (f) S68C.

firmed that the Ser codon was replaced by a Cys codon. The attempt to obtain an Ocr(S68C) protein was unsuccessful. Careful sequencing of both upper and bottom strands revealed that a C-G base was deleted. Due to the frameshift, a STOP codon appeared and the obtained protein was only 69 amino acids long. As known from Dunn's work, any Ocr protein with less than 87 amino acids is unstable and is not seen on SDS-PAGE gels (Dunn *et al.*, 1981).

All six Ocr(Cys) proteins were purified in the same way using three columns: DEAE-Sepharose, Superdex 200 and Phenyl-Sepharose. After the last purification step, all six proteins looked pure on SDS-PAGE gel after Coomassie blue staining.

3.4 Ocr stability

Differential Scanning Calorimetry (DSC) is one of the techniques used in biochemistry to determine the stability of a protein (Burova *et al.*, 1999; Richardson 3rd *et al.*, 2000).

Figure 3.5 shows a typical DSC scan for wild type Ocr. The DSC thermogram is a plot of the heat capacity of a solution of Ocr protein *vs.* a range of temperatures chosen for the experiment (20 °C to 100 °C). The same experiment was done using buffer instead of Ocr and the obtained plot was subtracted from the Ocr one. The baseline sloped even after the subtraction of the buffer contribution. In order to calculate the calorimetric enthalpy (ΔH_{cal}), the calorimetric baseline was subtracted and the obtained plot was normalized to the Ocr concentration.

At the end of the experiment, Ocr sample was cooled down to 20 °C and rescanned a few times. The shape of DSC scans were identical with the first scan demonstrating that the thermal denaturation of Ocr is a reversible process.

The corrected data were first fit to a two-state folding curve using Origin 5.0 software. A two-state transition implies that Ocr protein remains dimeric up to the denaturation temperature, then simultaneous denaturation and dissociation

occurred. The following equation characterizes two-state transitions:

$$D_2 \rightleftharpoons 2U \tag{3.1}$$

where

D native state as dimer
U denatured monomeric state

As seen in figure 3.4, the fitting is not very good. Consequently, we tried to fit the data to a dissociation model which assumes that Ocr first dissociates to two monomers then unfolding occurs. The nice fit between the theoretical model and the experimental data and the similarities between the calorimetric enthalpy and the van't Hoff enthalpy demonstrates that Ocr unfolding is indeed best described by a dissociation model (fig 3.5).

The equation that describes this type of transition is:

$$D_2 \rightleftharpoons 2M \rightleftharpoons 2U \tag{3.2}$$

where

D native state as dimer
M native state as monomer
U denatured state

The melting temperature (T_m) is the temperature at which 50 % of the protein is folded and 50 % unfolded. For the two-state model this temperature corresponds to the peak of the DSC scan. For the dissociation model, T_m is a bit higher than the temperature at which the heat capacity has an maximum. For example for wild type Ocr the transition midpoint (i. e. the peak) is at 69.3 °C but the T_m is 72.2 °C.

T_m values were used to compare the stability of the wild type Ocr with different mutants in which certain amino acids were replaced by Cys or the C-terminal

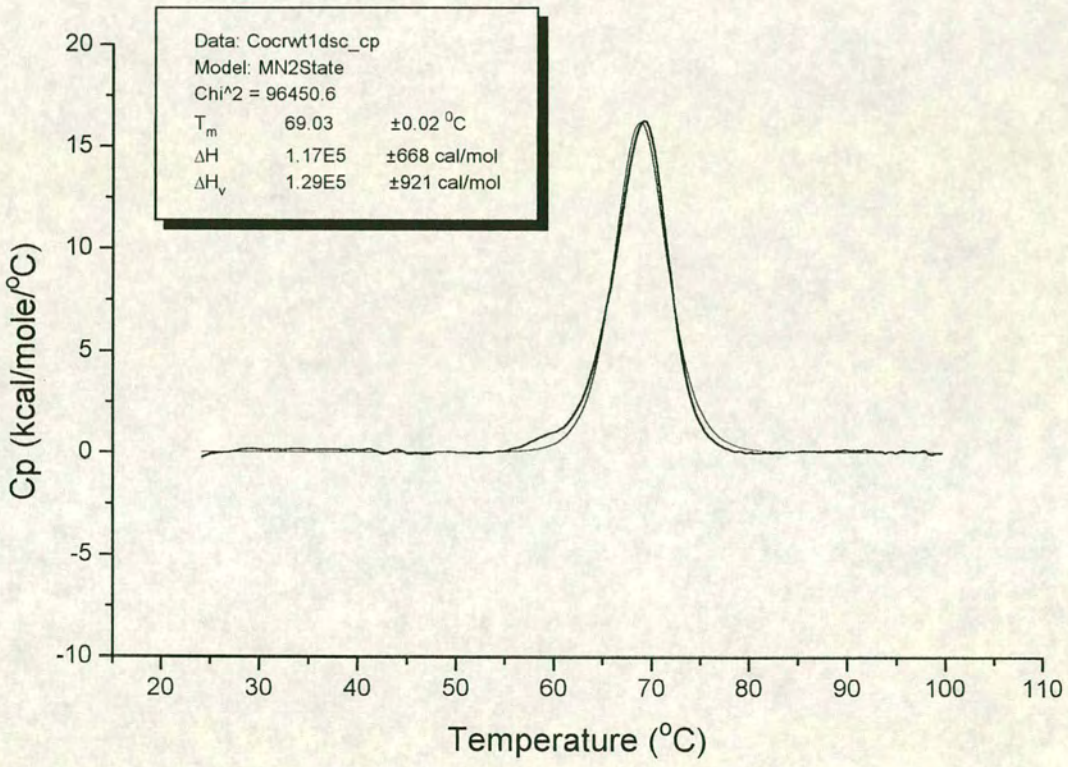


Figure 3.4: Denaturation heat capacity curve of wild type Ocr. The protein sample was degassed before injecting into the cell. The sample was scanned between 20 and 100 °C at a scan rate of 60 °C/h. The buffer was 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$ pH 8. The protein concentration was 31.53 μM. The scan was fitted to two state model. T_m is the melting point, ΔH is the calorimetric enthalpy and ΔH_v the van't Hoff enthalpy.

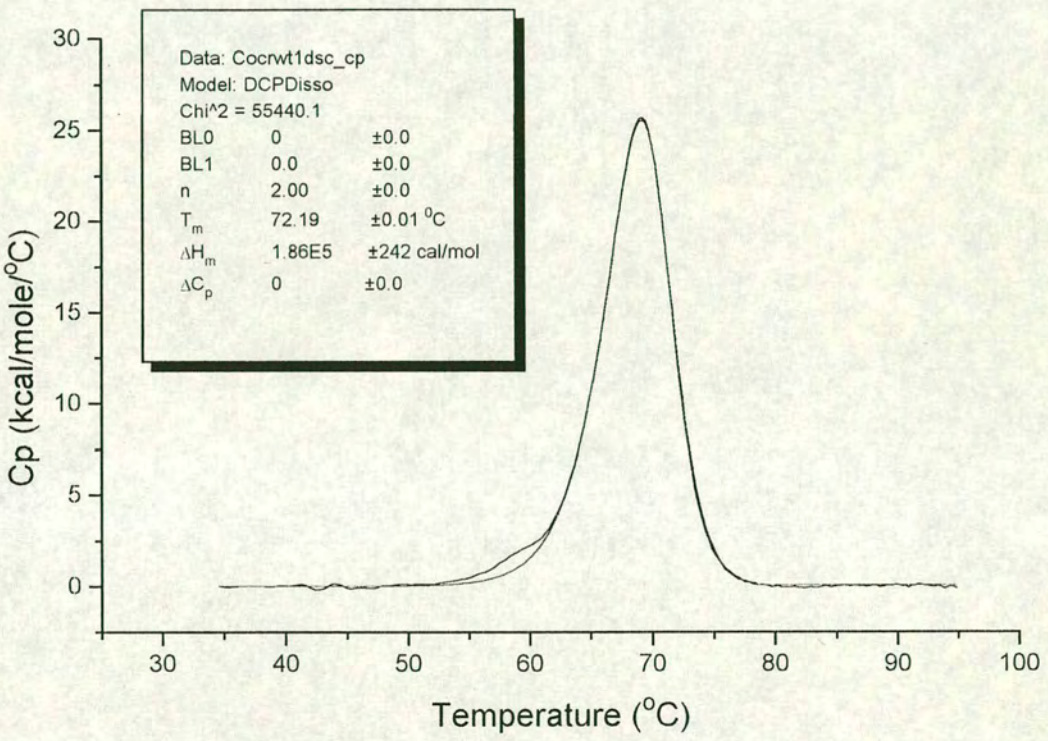


Figure 3.5: DSC of Ocr protein, dissociation model. The scan shown in figure 3.4 was fitted to a dissociation model. The buffer was 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$ pH 8. The protein concentration was 31.53 μM. T_m is the melting point, ΔH_m is the calorimetric enthalpy, BL_0 and BL_1 are the base lines and ΔC_p is the change in heat capacity.

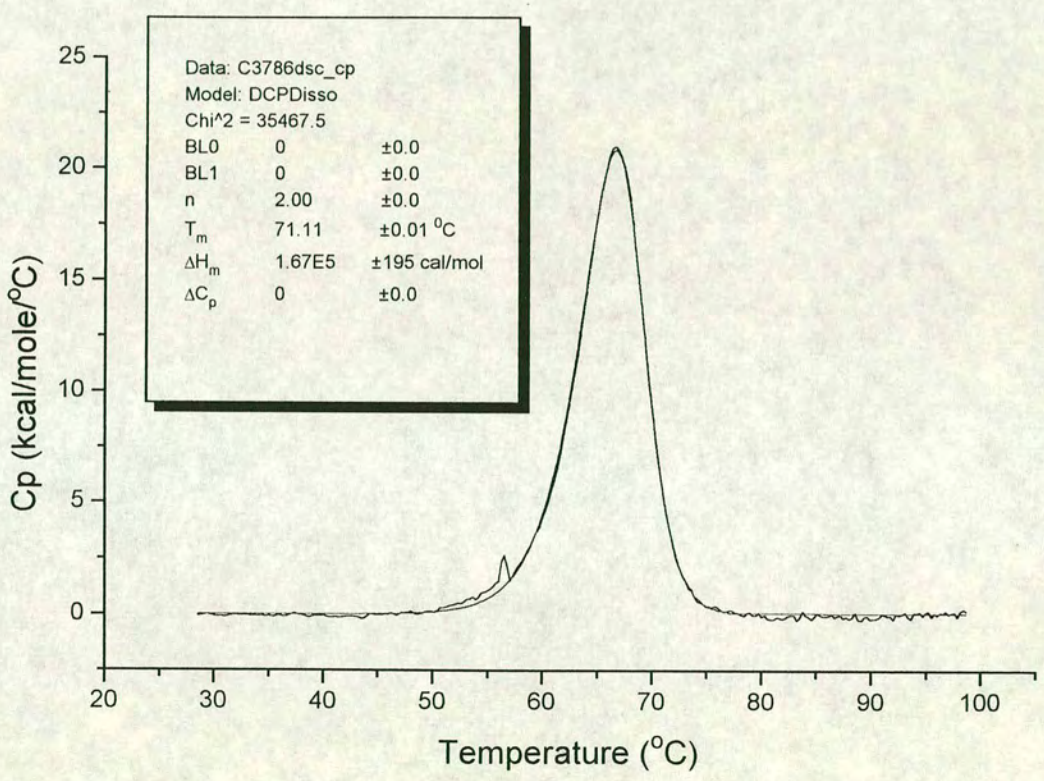


Figure 3.6: DSC scan of Ocr 3786 protein. The experiment was performed in exactly the same conditions as those described in figure 3.5 except the protein concentration which was 13.53 μM.

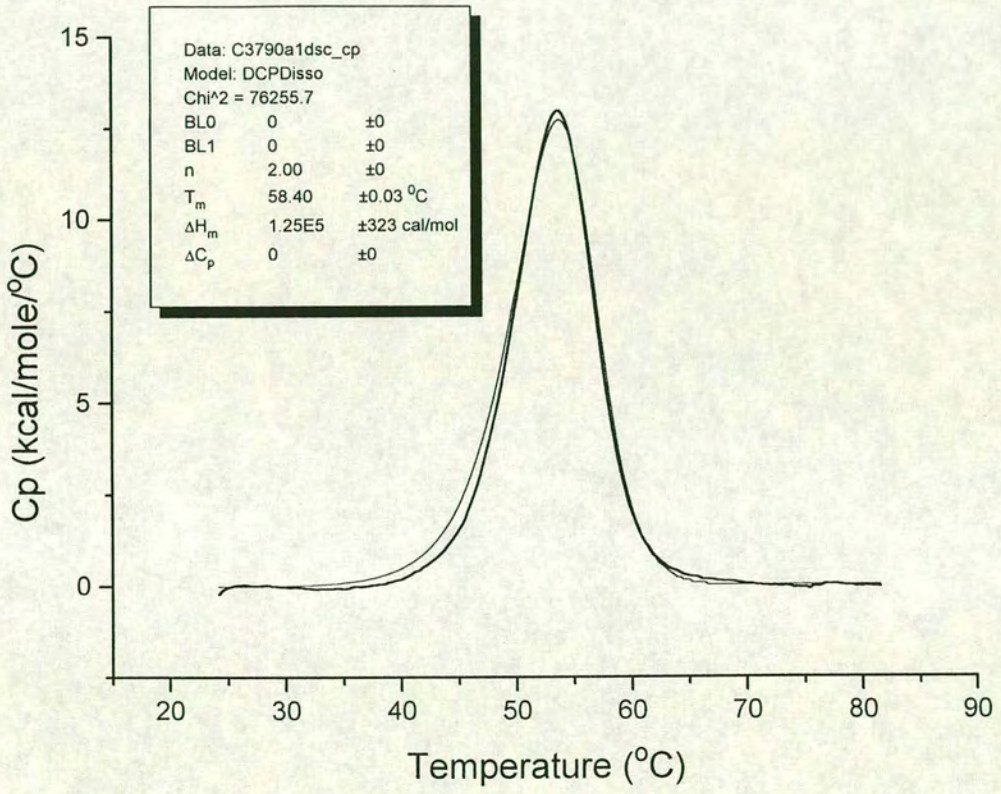


Figure 3.7: DSC scan of Ocr 3790 protein. The experiment was performed in exactly the same conditions as those described in figure 3.5 except the protein concentration which was 22.58 μM.

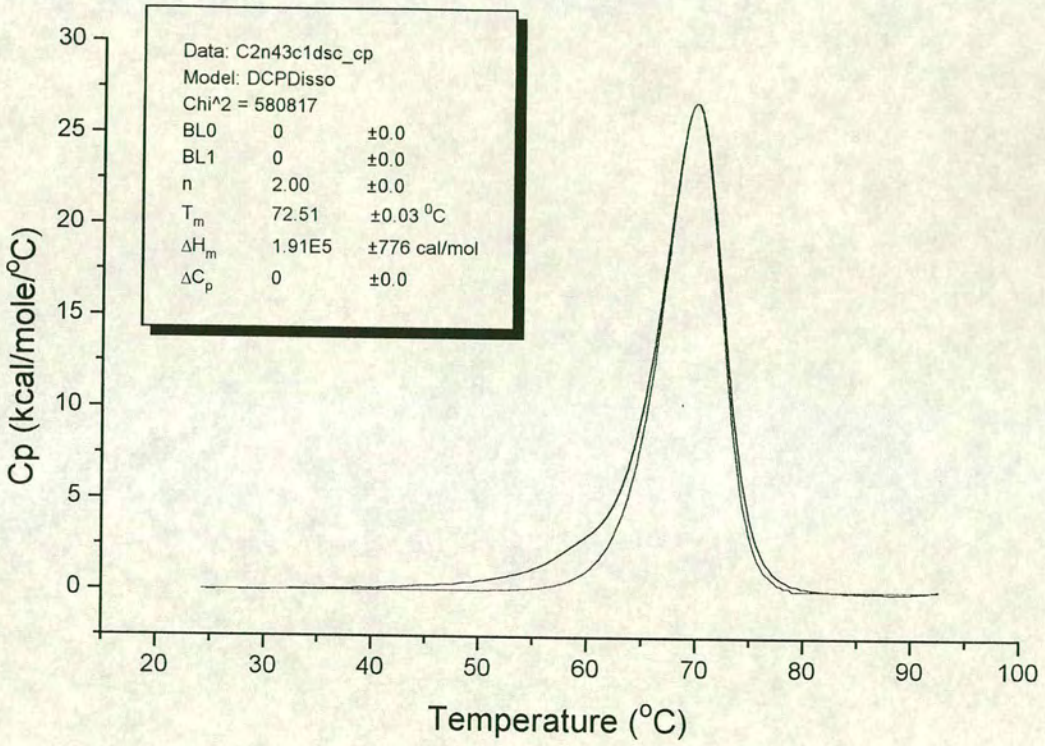


Figure 3.8: DSC scan of Ocr(N43C) protein. The experiment was performed in exactly the same conditions as those described in figure 3.5 except the protein concentration which was 41.29 μM.

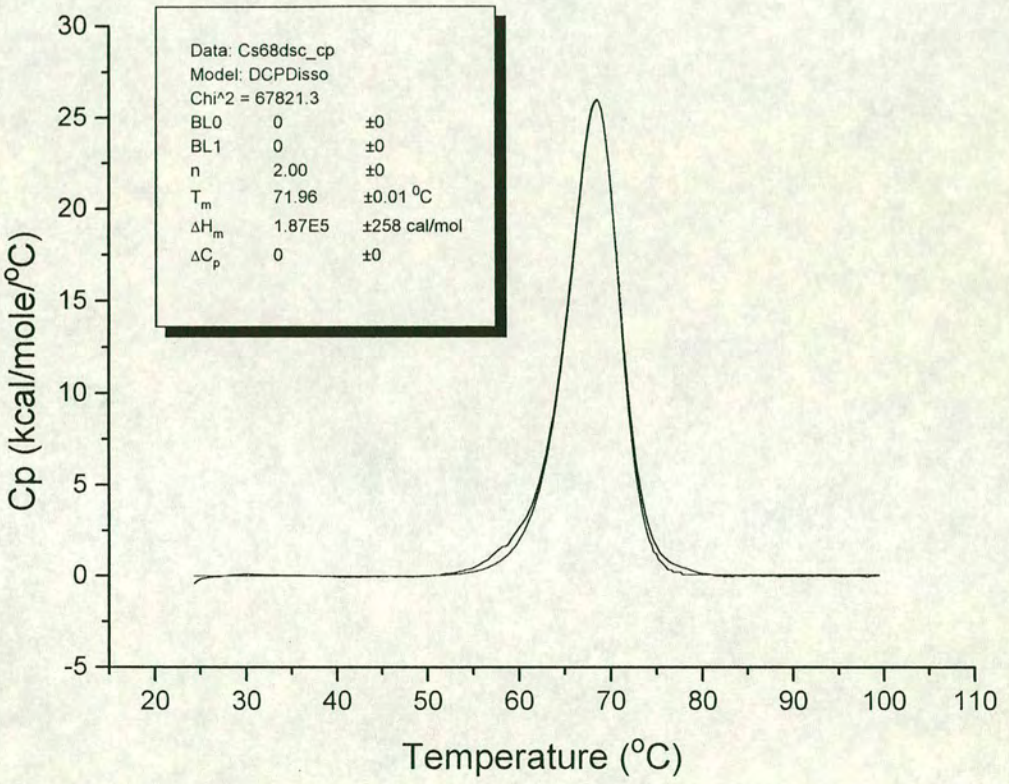


Figure 3.9: DSC scan of Ocr(S68C) protein. The experiment was performed in exactly the same conditions as those described in figure 3.5 except the protein concentration which was 21.76 μM.

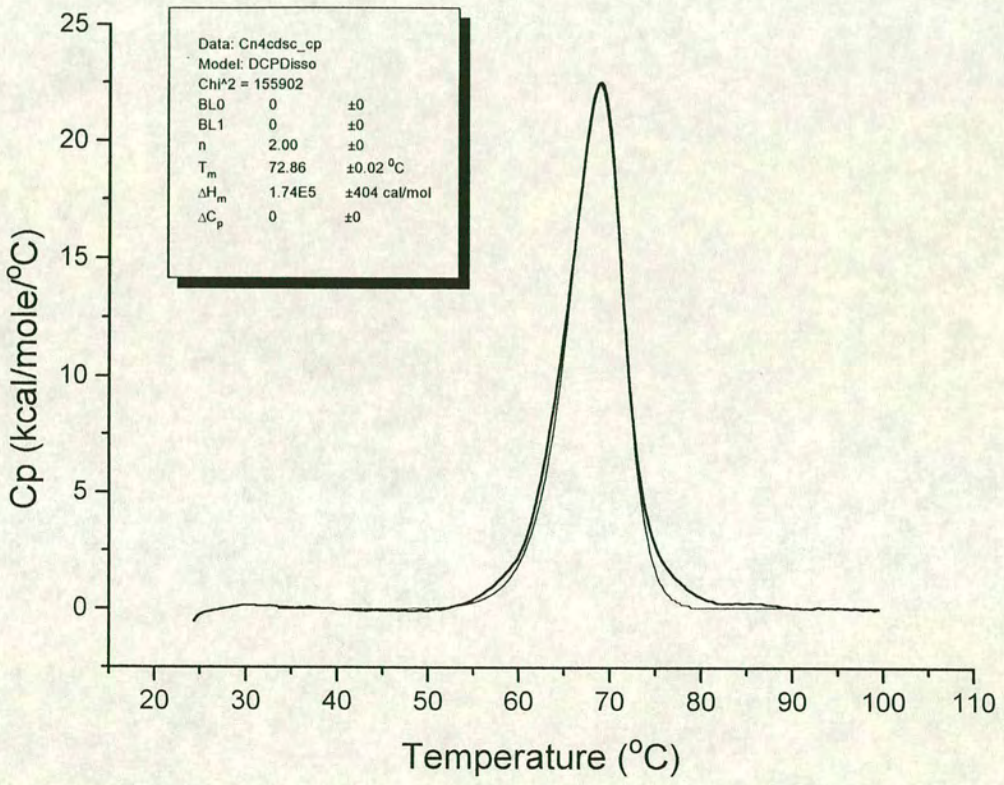


Figure 3.10: DSC scan of Ocr(N4C) protein. The experiment was performed in exactly the same conditions as those described in figure 3.5 except the protein concentration which was 17.01 μM.

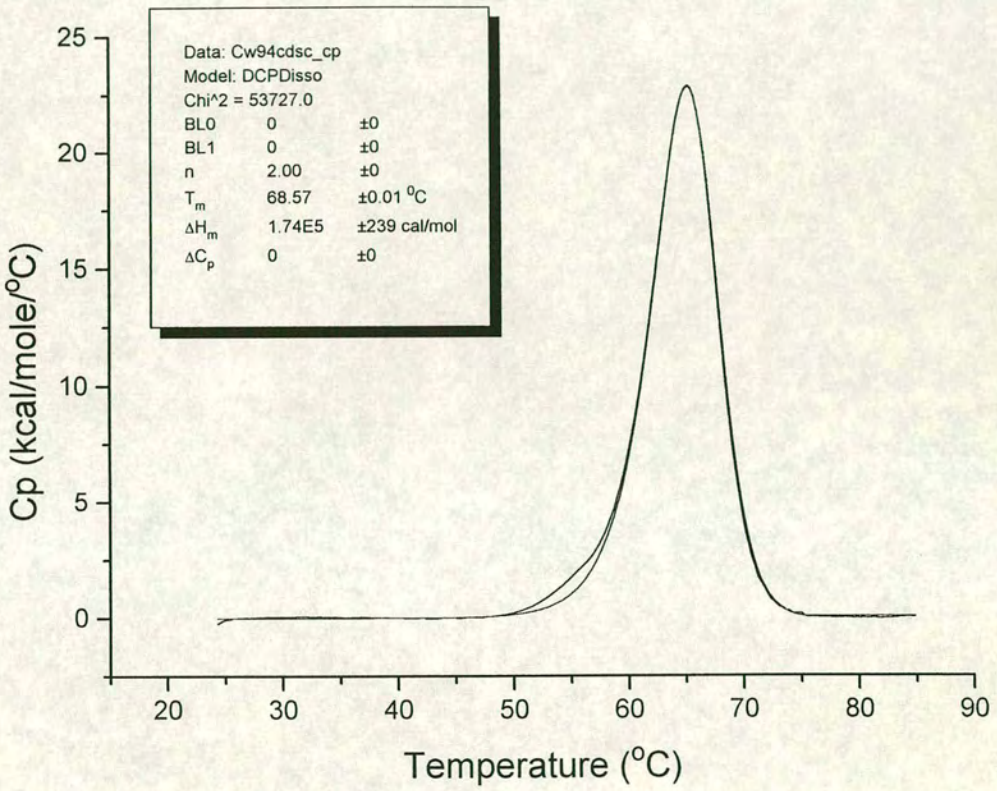


Figure 3.11: DSC scan of Ocr(W94C) protein. The experiment was performed in exactly the same conditions as those described in figure 3.5 except the protein concentration which was 27.87 μM.

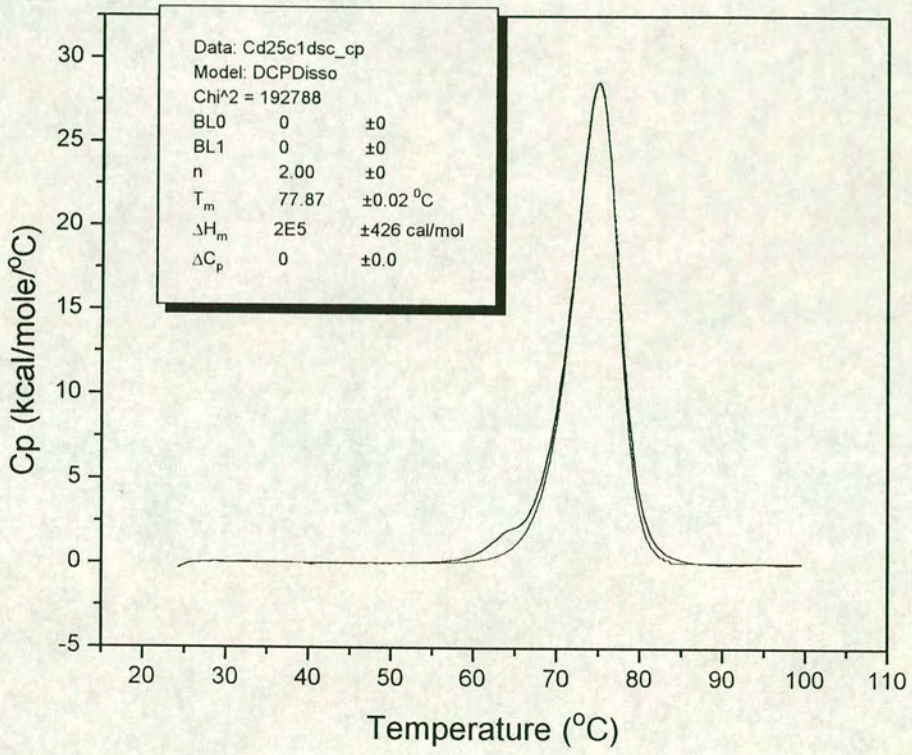


Figure 3.12: DSC scan of Ocr(D25C) protein. The experiment was performed in exactly the same conditions as those described in figure 3.5 except the protein concentration which was 29.10 μM.

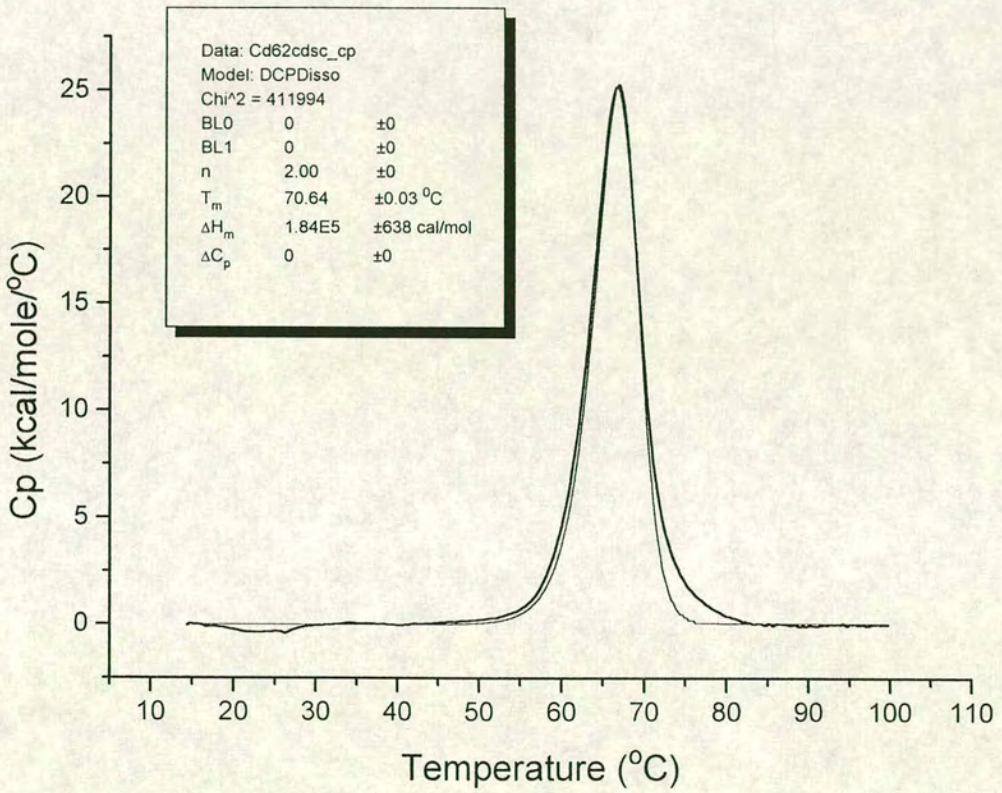


Figure 3.13: DSC scan of Ocr(D62C) protein. The experiment was performed in exactly the same conditions as those described in figure 3.5 except the protein concentration which was 17.97 μM.

Table 3.1: Protein stability

Protein	*Concentration* (μM)	T_m (°C)	ΔT_m (°C)
Ocr	31.53	72.2	0.0
3786	13.53	71.1	-1.1
3790	22.58	58.4	-13.8
S68C	21.76	71.9	-0.3
N43C	41.29	72.5	+0.3
N4C	17.01	72.9	+0.7
D62C	17.97	70.6	-1.6
D25C	29.10	77.9	+5.7
W94C	27.87	68.6	-3.6

end was deleted. As seen above, for wild type Ocr, T_m was found to be 72.2 °C. Table 3.1 shows the T_m values obtained for the proteins under investigation. The accuracy with which the T_m values are measured with this method is ± 0.5 °C. Three out of 6 Cys mutations did not dramatically disturb the stability of Ocr protein. Ocr(S68C) was proved to be almost as stable as the wild type (ΔT_m -0.3 °C) (Figure 3.9). Also substitution of Asn43 residue with Cys does not seem to affect the stability of the protein (ΔT_m +0.3 °C) (Figure 3.8), meanwhile the T_m value obtained for N4C mutant revealed a slightly increase in the stability (ΔT_m +0.7 °C) (Figure 3.10). D62C substitution seems to decrease the protein stability (ΔT_m -1.6 °C). However, the apparent decrease in T_m could be due to the poor fitting to the dissociation model rather than to a real instability introduced by mutation (Figure 3.13). The most dramatic effect was observed with the D25C mutant where the value obtained for T_m was 5.7 °C higher than that obtained for wild type (Figure 3.12). Of a particular importance was the T_m obtained for W94C mutant which is 3.6 °C lower than the value obtained for wild type Ocr (Figure 3.11), confirming an early assumption that the unique Trp residue in position 94 is involved in Ocr stability (Mark and Studier, 1981).

We tried then to see if deletion of the C-terminal end affects Ocr stability.

For Ocr 3786 (Figure 3.6), the instability induced by the deletion of the last 7 amino acids from the C-terminal region was not too drastic (ΔT_m -1.1°C). However, Ocr 3790 where 17 amino acids from C-terminal region were deleted, showed a dramatic decrease in stability (ΔT_m -13.8 °C) (Figure 3.7). The obvious conclusion that arises from the last two assays is that C-terminal end is involved in the stability of Ocr protein.

3.5 Trp94

3.5.1 Steady-state fluorescence. Background

Certain substances have the ability to emit light from electronically excited states, emission called either fluorescence or phosphorescence depending on the nature of the excited states. The processes of fluorescence and phosphorescence are usually explained using a Jablonski diagram (Figure 3.14). In a typical Jablonski diagram, S_0, S_1 and S_2 represent the singlet ground, first and second electronic state respectively, meanwhile T_1 represents the first triplet electronic state. Each of these electronic levels contain different vibrational energy levels denoted 0,1,2, etc (Lakowicz, 1999).

The fluorescence is induced by light not by heat because the energy difference between S_0 and S_1 is too large to be covered by the thermal energy. After absorbing light of a certain wavelength, fluorophore is excited from S_0 to S_1 or S_2. In a very short time (10^{-12} s) the excited molecule relaxes to the lowest level of S_1, a process called internal conversion. Return to the ground state occurs at the expense of the emission of a photon. Emission from S_1 is termed fluorescence and is spin-allowed because the electron in the excited orbital is paired with opposite spin to the second electron in the ground state orbital. The fluorescence lifetimes are typically near 10^{-8} s.

A fraction of the excited molecules return to the ground state by thermal relaxation without emitting light. The fluorescence quantum yield is the ratio of

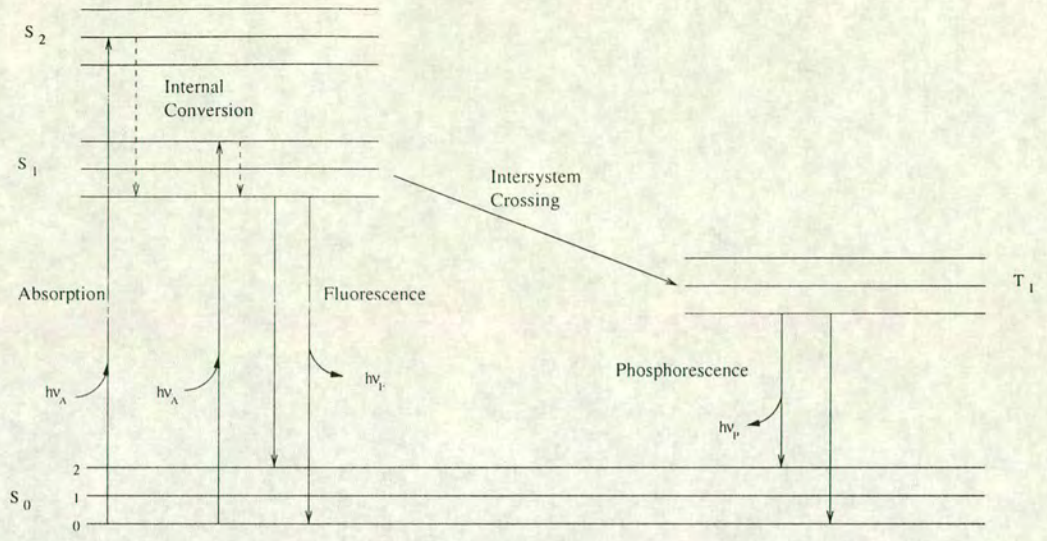


Figure 3.14: Jablonski diagram.

photons emitted to the number absorbed (Lakowicz, 1999). The quantum yield is given by:

$$Q = \frac{\Gamma}{\Gamma + k_{nr}} \tag{3.3}$$

where

Q	quantum yield
Γ	emissive rate of the fluorophore
k_{nr}	rate of the radiationless decay

From S_1, an excited molecule can undergo a spin conversion to the first triplet state (T_1), in a process called intersystem crossing. Emission of light from T_1 is called phosphorescence, and is a slow process ($10^{-3} - 10^{0}$ s). The reason for this slow phosphorescence emission rate is that transitions from T_1 to S_0 are forbidden because the electron in the excited orbital has the same spin orientation as the ground-state electron.

3.5.2 Exposure of Trp94

Fluorescence is a powerful tool for studying the structure and function of proteins. Proteins contain only three aromatic amino acids and all of them are fluorescent.

However, their contribution to the total protein fluorescence is unequal, the fluorescence emission spectrum being dominated by tryptophan.

Several reasons make tryptophan the most valuable amino acid in studying protein fluorescence: it has the strongest absorption, the largest fluorescence quantum yield, and both the fluorescence spectrum and intensity are the most sensitive to the local environment.

Following light absorption a variety of dynamic processes occur resulting in loss of the energy of the excited tryptophan. As a consequence of this loss, the energy of the emission is less than that of absorption and emission occurs at longer wavelength (350 nm) than absorption (295 nm).

One of the factors that influences the tryptophan emission spectrum is solvent polarity. The dipole moment of a polar fluorophore like tryptophan becomes larger in the excited state than in the ground state. The solvent dipoles can rapidly reorient around the excited state dipole moment, in this way lowering the energy of the excited state. This effect becomes larger as the solvent polarity increases, resulting in emission at lower energies or longer wavelength. Consequently, a tryptophan emission spectrum gives information about the polarity of the tryptophan microenvironment (Lakowicz, 1999).

When excited at 295 nm, proteins often display a fluorescence emission spectrum with a maximum around 340 nm. As shown in figure 3.15, the Ocr protein has an emission spectrum shifted to 350 nm. This is a clear indication that the only Trp residue in position 94 is solvent exposed. Emission at 350 nm was also observed in the case of adenocorticotropin hormone, glucagon and for denaturated proteins when buried Trp residues were forced to come in contact with solvent dipoles (Eftink and Giron, 1976; Wu and Yang, 1980; Lakowicz, 1999).

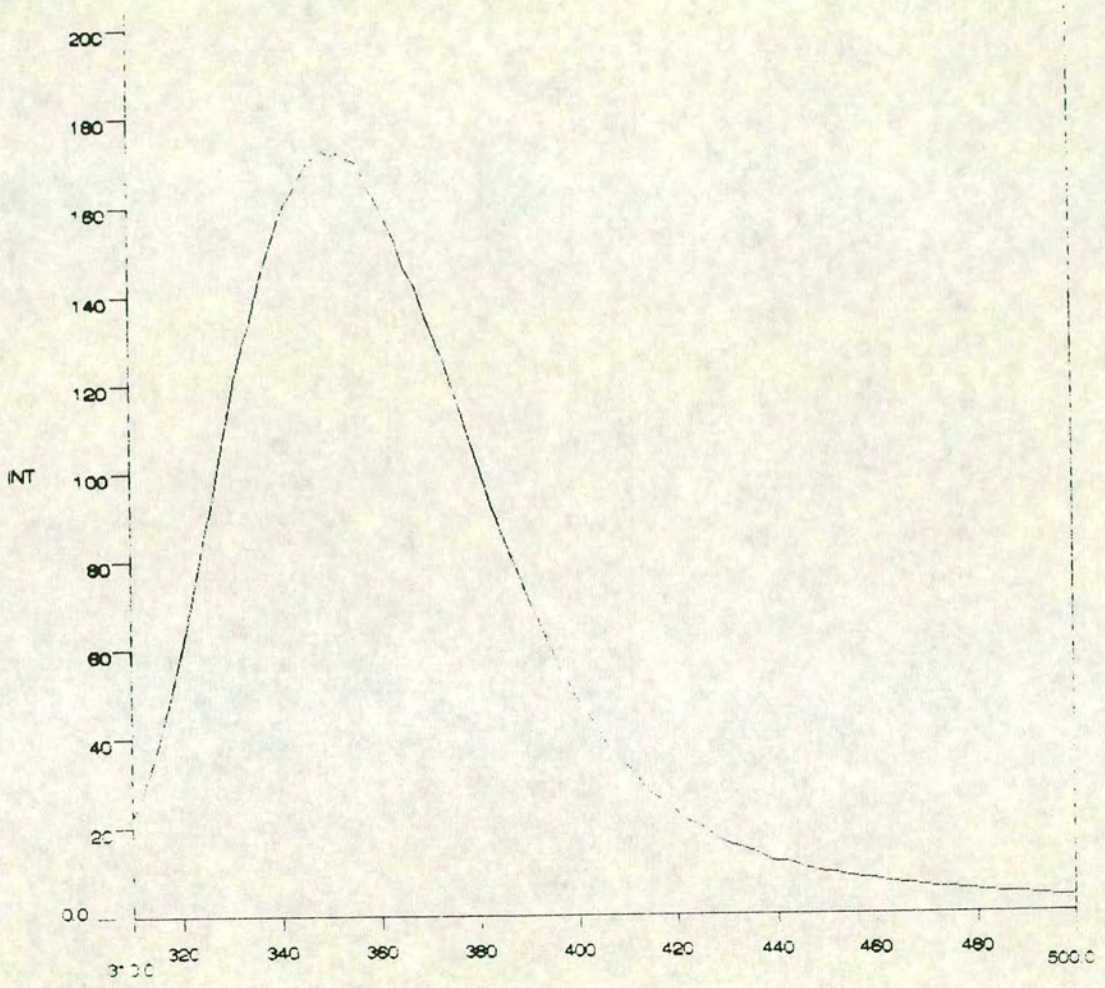


Figure 3.15: The tryptophan emission spectrum of wild type Ocr. Protein sample was excited at 295 nm and the emission was recorded between 300 and 500 nm. The excitation and emission slitwidth were 5 and 10 nm respectively. The buffer was 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$, 7 mM β-mercaptoethanol pH 8 and the temperature 25 °C.

3.6 Determination of the molar extinction coefficient

Accurate determination of protein concentration is essential in quantitative studies of protein-protein interactions. Colorimetric methods commonly used for determination of protein concentration (such as Biuret or Lowry methods) are associated with large degrees of error. The most accurate way to determine the concentration of a protein is to use UV-spectroscopy, if the extinction coefficient of that protein is known (Pace and Schmid, 1997). In this case the concentration is easily calculated according to Beer Lambert law.

Here we calculate the molar extinction coefficient of Ocr protein using a method developed by Gill and von Hippel (1989). The method is based on earlier work done by Edelhoch (1967).

The absorbance of a protein between 270 nm and 300 nm is due to Trp, Tyr residues and to a certain extent, disulphide bonds (Cystine). The absorption spectra of Trp and Tyr depend on their environment and are slightly blue-shifted when transfered from a non-polar to a polar environment. Consequently, Trp or Tyr residues exposed on surface and those buried in the protein matrix will contribute differently to the molar extinction coefficient.

It is not possible to measure directly the molar extinction coefficient. However, it is possible to calculate the molar extinction coefficient of a protein in denaturing conditions when all Trp and Tyr residues are fully exposed. Edelhoch has measured the molar extinction coefficient of Trp, Tyr and Cystine in 6.0 M GdnHCl (Table 3.2).

If the number of Trp, Tyr and Cystine residues is known it is possible to calculate the molar extinction coefficient of the denatured protein from the formula:

$$\varepsilon_{den} = a\varepsilon_{den,Trp} + b\varepsilon_{den,Tyr} + c\varepsilon_{den,Cys} \tag{3.4}$$

Table 3.2: Molar Extinction Coefficients of Model compounds

Model compound	Extinction coefficient at (nm)				
	276	278	279	280	282
N-Acetyl-L-tryptophanamide	5400	5600	5660	5690	5600
Gly-L–tyr-Gly	1450	1400	1345	1280	1200
Cystine	145	127	120	120	100

where

ε_{den}	molar extinction coefficient of protein in 6.0 M GdnHCl
$\varepsilon_{den,Trp}$	molar extinction coefficient of Trp
$\varepsilon_{den,Tyr}$	molar extinction coefficient of Tyr
$\varepsilon_{den,Cys}$	molar extinction coefficient of Cystine
a,b,c	number of Trp, Tyr and Cystine residues respectively per molecule of protein

In the case of Ocr protein the calculated theoretical molar extinction coefficient at λ 280 nm was 15930 M^{-1} cm^{-1}.

$$\varepsilon_{Ocr} = 1x5690 + 8x1280 + 0x120 \tag{3.5}$$

This value corresponds to Ocr monomer. For the dimer, the molar extinction coefficient of denatured Ocr will be 31860 M^{-1} cm^{-1}.

To determine the molar extinction coefficient of the native Ocr, the absorbance spectra of the native Ocr and denatured protein were measured at identical protein concentration. According to the Beer Lambert law:

$$A_{den} = \varepsilon_{den} c_{den} l \tag{3.6}$$

Also we may write:

$$A_{nat} = \varepsilon_{nat} c_{nat} l \tag{3.7}$$

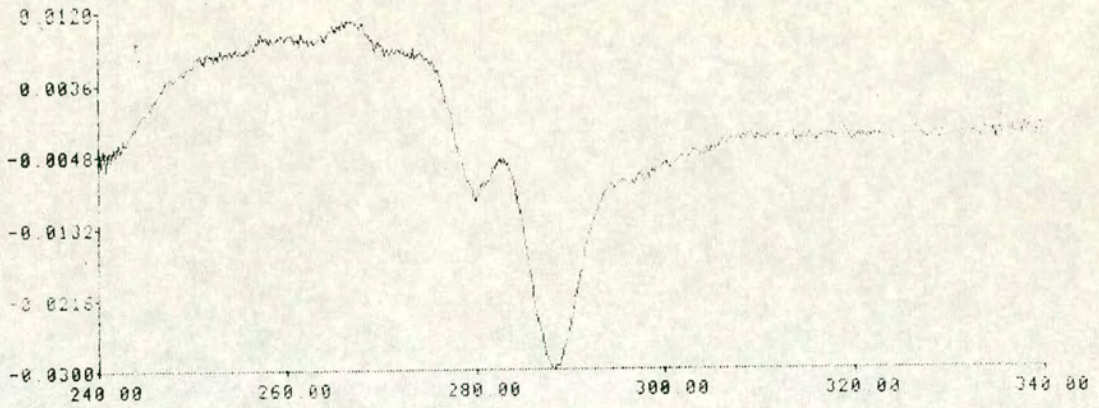


Figure 3.16: Difference spectra between native Ocr and Ocr unfolded in 6M GdnHCl. The absorbance of wild type Ocr was scanned between 240 and 340 nm. The buffer used was 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$, 7 mM β-mercaptoethanol, pH 8.0. GdnHCl was added to a final concentration of 6 M. The spectra of denaturated Ocr was corrected for dilution and subtracted from the spectra of native Ocr.

Since the concentration of the native and denaturated Ocr were set experimentally equal, combining equations (3.6) and (3.7) we obtained:

$$\varepsilon_{nat} = \frac{A_{nat}\varepsilon_{den}}{A_{den}} \tag{3.8}$$

A_{nat} and A_{den} were experimentally measured and ε_{den} was calculated as previously shown. For Ocr dimer we obtained a molar extinction coefficient of 31095 M^{-1} cm^{-1} which is an average of two experiments done in the same conditions. Figure 3.16 shows a typical difference spectra between native and denaturated Ocr.

3.7 Close proximity between N4-N4 and S68-S68 residues

3.7.1 Pyrene Maleimide excimer

Background

Pyrene Maleimide (PM) is a fluorophore that in addition to emitting fluorescence from the excited monomer state can display an excimer fluorescence. The excimer fluorescence is a consequence of an interaction between an excited pyrene monomer with a neighboring ground state pyrene. The first condition for excimer formation is that the two pyrene fluorophores are in close proximity. Secondly, the two pyrenes must be stacked in a precise symmetrical configuration (Lehrer, 1997; Lakowicz, 1999).

In organic solvents, pyrene excimer fluorescence has different properties than in aqueous solutions (Lehrer, 1997). In an organic solvent the pyrenes do not interact in the ground state. If there is no steric hindrance, upon excitation an excited monomer can reorient next to its neighbor ground state pyrene forming an excited state dimer. There is an equilibrium between the monomer and excimer fluorescence. The excimer fluorescence appears a few nanoseconds after the monomer fluorescence. The reason for this delay arises from the fact that the excimer is formed after a monomer is excited.

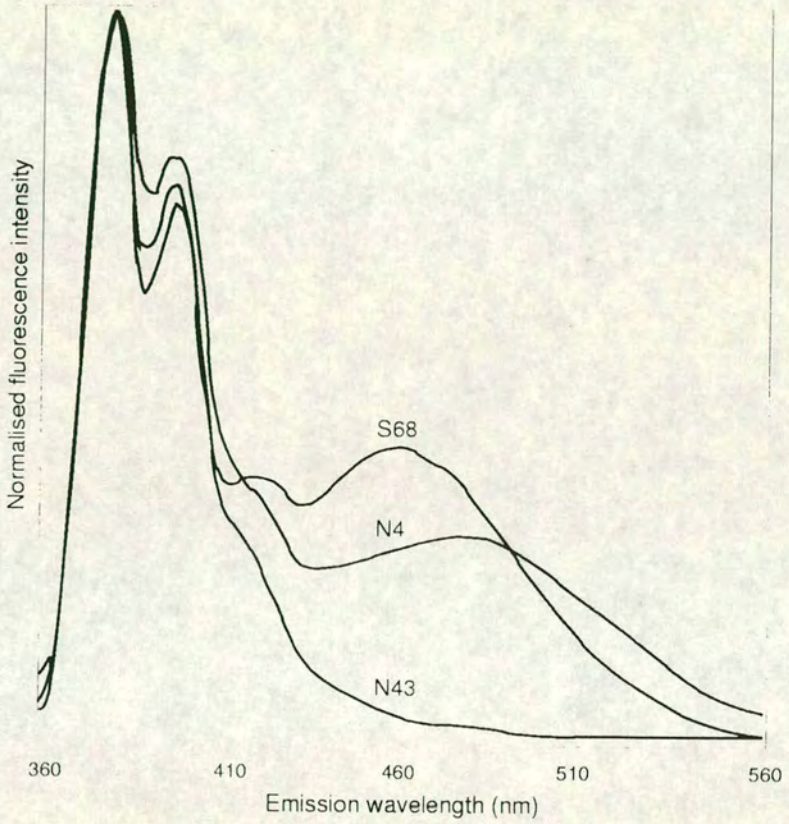


Figure 3.17: Normalized pyrene emission spectra for chemically-modified Ocr containing the N4C and S68C substitutions compared to one of the other mutant proteins, N43C. All pyrene-modified mutant proteins except N4C and S68C gave the same emission spectra. The presence of a broad emission band at long wavelengths for the N4C and S68C proteins indicates the formation of an excimer due to close proximity of the two pyrene moieties across the monomer-monomer interface in the Ocr dimer. Protein samples were excited at 342 nm. The experiment was performed at 25 °C using the excitation and emission slit widths of 5 nm and 10 nm respectively. The buffer was 0.1 M phosphate, pH 7.

In aqueous solutions, pyrenes tend to stack because of their hydrophobicity. There is an equilibrium between stacked and unstacked pyrenes. The shorter time delay between monomer and excimer emission indicate that excimer fluorescence come from the stacked pyrenes that slightly reorient during the lifetime (Graceffa and Lehrer. 1980). It is worth noting that two stacked pyrenes does not necessarily lead to an excimer emission spectrum. The ground-state pyrene could quench the fluorescence of the excited pyrene before the reorientation can take place.

Results

Six solvent exposed Ocr amino acids were individually replaced with cysteine and labelled with cysteine-specific fluorophores. The labelling was done in the absence of β-mercaptoethanol to avoid the competition for fluorophore between SH moieties of β-mercaptoethanol and those of cysteine residues. Five out of six amino acids were labelled in these conditions. The failure in labelling the Ocr(N4C) protein made us believe that N4C residue was not actually solvent exposed. Consequently, we tried to label it in denaturing conditions, in the presence of 5 M GdnHCl, but the labelling failed again. The logical explanation remaining was that the two N4C residues (one in each monomer) were in close proximity and formed a disulphide bond in the absence of reducing reagents. The assumption was proven to be correct because labelling N4C residues in the presence of β-mercaptoethanol succeeded.

An elegant way to prove the close proximity of two cysteine residues is to label them with Pyrene Maleimide and to observe excimer formation (Strasburgh *et al.*, 1985; Ziegler *et al.*, 1994). A typical PM emission spectra is shown in figure 3.17 where Ocr(N43C)-PM sample was excited at 342 nm. When Ocr(N4C)-PM was used as sample, besides the two typical emission peaks at 378 nm and 395 nm another one appeared at 480 nm. The extra peak at 480 nm is due to excimer formation between two adjacent pyrene molecules, confirming the close proximity between N4C residues from the two monomers.

Using the same technique we determined that S68C residues are also in close proximity in spite of the fact that labelling was easily achieved in the absence of the reducing reagents (Figure 3.17). The excimer peak appeared at 460 nm instead of 480 nm as for Ocr(N4C)-PM, which is due probably to a more hydrophobic pocket for S68C than for N4C.

3.7.2 HPLC

The close proximity between N4C-N4C and S68C-S68C residues was an exciting result and we wanted to confirm it by other methods. The question that arises from the close proximity of two cysteine residues is: are they close enough to make a disulphide bond? The present experiment was designed to answer this question.

From previous assays we know that Ocr dimers appear as monomers in the presence of 5 M GdnHCl solution. However, GdnHCl is not able to break up disulphide bonds so that if such a bond exists it is expected that Ocr(N4C) or Ocr(S68C) will appear as dimers even in denaturing conditions.

Ocr(N4C) was desalted using two buffers, one containing 5 M GdnHCl and the other 5 M GdnHCl plus 7 mM β-mercaptoethanol. The samples were then applied to a gel-filtration column. In the presence of β-mercaptoethanol, Ocr(N4C) appeared as one peak corresponding to the monomer position. When β-mercaptoethanol was missing, two protein peaks appeared, one corresponding to the monomer and the other to the dimer position (Figure 3.18).

The same procedure was applied to Ocr(S68C). It appeared as a monomer either in the presence or in the absence of β-mercaptoethanol (Figure 3.18). Putting together the pyrene excimer and HPLC assays, we conclude that N4C residues are in close proximity, and form disulphide bonds. S68C residues are also close to each other but not so close as to form a disulphide bond.

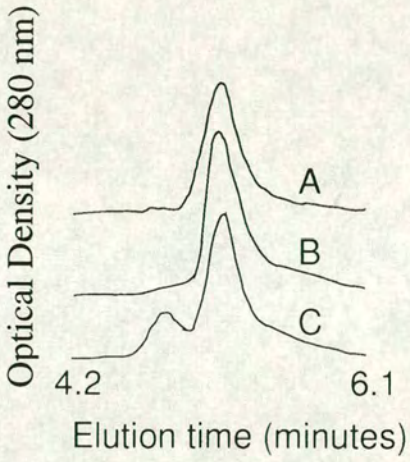


Figure 3.18: HPLC denaturing gel filtration for the N4C and S68C mutant proteins. In the absence (top trace) of β-mercaptoethanol, the S68C protein elutes in a single peak characteristic of the unfolded protein monomer. The protein containing the N4C substitution elutes as two peaks in the absence of β-mercaptoethanol (bottom trace), corresponding to unfolded dimer joined covalently by a disulphide bond, and predominantly as a single peak in the presence of the reducing agent (middle trace). The buffer used was 20 mM Tris, 20 mM MES, 200 mM NH_4Cl, 10 mM $MgCl_2$, 0.1 mM EDTA, 5 M GdnHCl, pH 6.5. The flow rate was 0.5 ml/min and the elution was monitored at 280 nm.

3.8 FRET

Fluorescence Resonance Energy Transfer (FRET) is a distant-dependent interaction between the electronic excited states of two dye molecules in which excitation is transfered from a donor to an acceptor molecule. It is worth mentioning that energy transfer occurs without the appearance of a photon and is the result of dipole-dipole interactions between donor and acceptor (Lakowicz, 1999).

Two conditions must be fulfilled for FRET to occur: first, the absorbance spectra of the acceptor must overlap the fluorescence emission spectra of the donor and second. the distance between donor and acceptor must be in the range of 10 to 100 Å.

The rate of the energy transfer is given by:

$$k_T = \frac{1}{\tau_D}\left(\frac{R_0}{r}\right)^6 \tag{3.9}$$

where

τ_D	decay time of the donor in the absence of acceptor
R_0	Förster distance
r	donor to acceptor distance

The Förster radius (R_0) is the distance at which energy transfer is 50 % efficient. Föster distances are typically in the range of 20 to 90 Å, comparable to the diameter of many proteins.

The efficiency of energy transfer (E) is the fraction of photons absorbed by the donor that are transfered to the acceptor.

$$E = \frac{k_T}{\tau_D^{-1} + k_T} \tag{3.10}$$

Combining the two equations , the efficiency of the energy transfer will be given by:

$$E = \frac{R_0^6}{R_0^6 + r^6} \tag{3.11}$$

E is determined experimentally from:

$$E = 1 - \Phi_{DA}/\Phi_D = 1 - \tau_{DA}/\tau_D \tag{3.12}$$

where

Φ_{DA}	quantum yield of the donor in the absence of the acceptor
Φ_D	quantum yield of the donor in the absence of the acceptor
τ_{DA}	lifetime of the donor in the presence of the acceptor
τ_D	lifetime of the donor in the absence of the acceptor

FRET was used to determine the distances between Cys residues introduced at different positions in Ocr dimer. For all six Ocr(Cys) dimers, if one monomer is labeled with a donor and the other with an acceptor it is possible to calculate 21 distances. In our experiments we used 1,5 I-AEDANS as donor and 5-IAF as acceptor. This donor-acceptor pair has a R_0 of 41-44 Å (Marsh and Lowey, 1980).

First of all Ocr(Cys) proteins were denatured in 5 M GdnHCl (Jonsson *et al.*, 1996). There are two reasons for this denaturation: to obtain 100 % labelling and to get a solution of monomers. Then the monomer solutions were labelled either with 1,5 I-AEDANS or 5-IAF. When labelling was completed, AEDANS labelled proteins were combined with AF labelled proteins in a 1:1 molar ratio. Proteins were then renatured by removing the GdnHCl. The renatured Ocr dimer solution should contain a mixture of AEDANS-AF heterodimers and AEDANS-AEDANS and AF-AF homodimers. Unfortunately, even in denaturing conditions we did not manage to get 100 % labelling. The uncertainty in the percentage of labelling led us to meaningless results.

However, there are three ways by which the experiment can be repeated successfully in the future: to find the conditions for 100 % labelling, to find a way to separate AEDANS-AF heterodimers from the mixture (probably by reverse phase HPLC), or to use Trp94 as donor and 1,5 AEDANS as acceptor. In the latter case the donor will be 100 % labelled but, of course, it will be possible to measure only 6 distances instead of 21.

Chapter 4

Ocr-*Eco*KI interactions

4.1 OCR-methylase stoichiometry

4.1.1 Trp fluorescence

Protein-ligand and protein-protein interactions are usually accompanied by changes in absorbance, light scattering, anisotropy, ligand or protein fluorescence. Consequently, spectrophotometric methods are used with great success to characterize the strength and stoichiometry of the complex formed, to characterize the kinetics of association and dissociation, and the structural changes in the proteins following binding.

Fluorescence is much more sensitive than absorbance and consequently is the method of choice to study protein-ligand or protein-protein interactions. The conformational changes that appear when two proteins interact can be quantified either by monitoring an intrinsic fluorescence from the aromatic amino acids (Trp, Tyr, Phe) or an extrinsic fluorescence due to a fluorophore attached to the ligand or protein (Eftink, 1997).

When studying a protein-ligand interaction it is better that either the change in the fluorescence of the protein or ligand be monitored. Sometimes fluorescence changes appear in both protein and ligand, a situation that leads to complications in interpreting the data (Eftink, 1997).

The emission spectrum of proteins is dominated by tryptophan. Phe and Tyr have a small contribution to the protein emission spectrum. When in water, Phe

displays an emission spectrum with a maximum around 282 nm, and Tyr has an emission spectrum near 303 nm. Usually proteins emit near 340 nm due to Trp residues. However, proteins can emit at lower or longer wavelength depending on the polarity of the environment of the Trp residues. For example in the the case of azurin protein, where Trp is buried in a very hydrophobic environment, the emission is at 308 nm (Finazzi-Agro *et al.*, 1970). For fully exposed Trp as in adenocorticotropin hormone, glucagon or in denatured proteins, emission is near 350 nm (Eftink and Giron, 1976; Wu and Yang, 1980; Lakowicz, 1999). The red-shift of the emission spectra is due to solvent molecules that interact with Trp residues in the excited state (and lower the fluorescence intensity). The phenomenon is known as solvent relaxation (Lakowicz, 1999). The small contributions of the Phe and Tyr residues to the emission spectra of proteins is easily avoided by exciting proteins at λ 295 nm where neither Phe nor Tyr can absorb light.

Fluorescence is a very sensitive technique, suitable to monitor protein-protein interactions. However, the interpretation of the changes which occuerin the Trp fluorescence intensity when two proteins interact is not always straightforward. For example, burial of a solvent exposed Trp upon binding usually leads to an increase in the emission fluorescence and a blue-shift of the emission maximum, since the Trp residue is no longer quenched by the solvent molecules. On the other hand, this increase in the Trp fluorescence intensity could be compensated by the quenching effect exerted by the disulphide bonds, His, Met or Arg residues with which the buried Trp could come in contact in the new environment (Sanyal *et al.*, 1989; Loewenthal *et al.*, 1991; Clark *et al.*, 1996; Hennecke *et al.*, 1997).

In addition to the multitude of factors that could influence the fluorescence of a Trp residue, more complications appear in a multi-Trp protein where every Trp is in a different environment. In these proteins, an energy transfer from a fluorescent Trp to a neighboring non-fluorescent Trp could also appear (Pokalsky

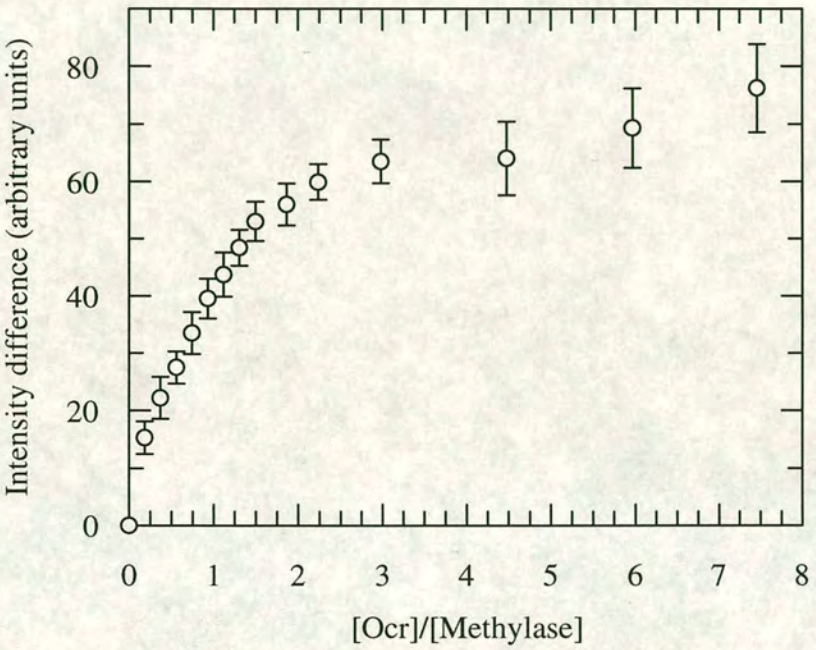


Figure 4.1: Ocr:methylase stoichiometry. Increasing concentrations of Ocr were titrated into a solution containing a constant concentration of 0.3 μM methylase. Sample was excited at 295 nm and the Trp emission fluorescence was observed at 340 nm. The fluorescence intensities at 340 nm for Ocr and methylase at different titration points were measured in two separate assays and the sum of the two values was calculated. From this sum, the Trp fluorescence intensity of the Ocr:methylase mixture was subtracted and the difference plotted *vs.* the Ocr/methylase molar ratio. The stoichiometry of binding was determined from the intersection of the extrapolated linear portions of the graphs. The assay was performed at 25 °C.

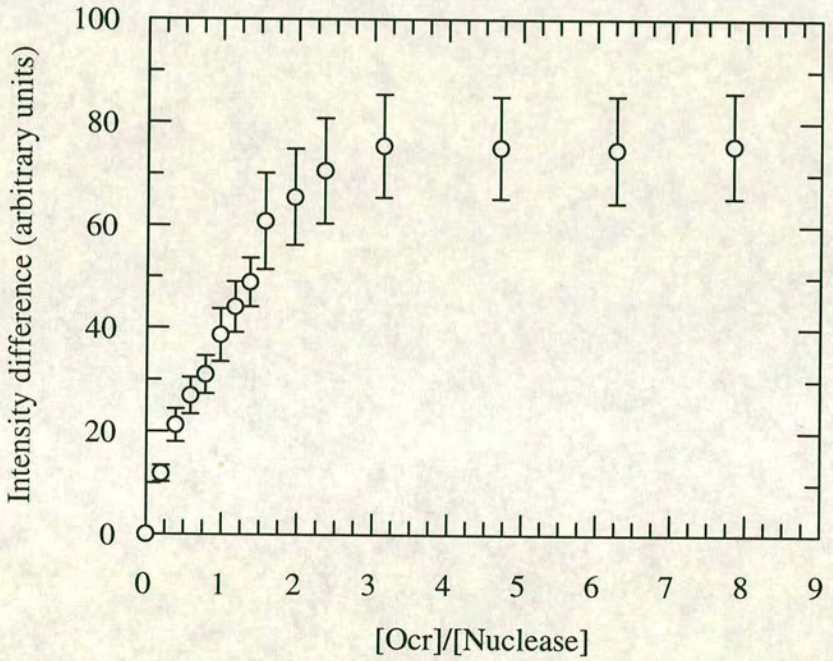


Figure 4.2: Ocr:nuclease stoichiometry. The assay was performed in the same conditions as in figure 4.1 except the concentrations of nuclease that was 0.1 μM.

et al., 1995). Keeping in mind all of these considerations, one can understand that the emission spectra of a protein reflects the average environment of all Trp residues.

In spite of the complications in data analysis, Trp fluorescence was used to study the stoichiometry of Ocr-methylase and Ocr-nuclease complexes. The advantages of the Trp fluorescence consists in the simplicity of the experiment: no extrinsic fluorophore need be attached to either protein, there is no concern about the percentage of labelling or the influence of the bound extrinsic fluorophore on the structure or function of the proteins.

Ocr contains one Trp residue in each monomer, located at position 94. The emission spectral maximum at 350 nm and quenching experiments (Atanasiu *et al.*, unpublished) suggest that Trp94 is exposed on the protein surface. *Eco*KI methylase and nuclease are multi-Trp proteins that display an emission spectrum maximum around 340 nm.

Conformational changes undergone following binding of Ocr to the methylase result in a decrease in the overall fluorescence intensity at λ 340 nm. The stoichiometry of the Ocr-methylase complex was determined from a graph where the difference between the sum of the individual intensities of Ocr and methylase at λ 340 nm and the fluorescence intensity of the Ocr-methylase complex was plotted against the molar ratio [Ocr]/[methylase].

For tight binding systems the stoichiometry is easily determined from the intersection of the extrapolated linear portions of the graphs (Tiedge *et al.*, 1982). Figure 4.1 show that the stoichiometry of the Ocr-methylase complex determined with this method is 1.78 ± 0.02

The same procedure was applied in order to get the stoichiometry of Ocr-nuclease complex which was found to be 2.3 ± 0.07 (Figure 4.2). The values are greater than expected and other techniques were used in order to confirm the obtained stoichiometry values. A 1:1 stoichiometry was expected for both Ocr-

methylase and Ocr-nuclease complexes, since Ocr is likely to interact with the S subunit of *Eco*KI (Bandyopadhyay *et al.*, 1985).

4.1.2 Extrinsic fluorescence

Extrinsic fluorescence is a widely-used method to study protein-protein interactions. The advantage of this method is that the fluorescence signal comes only from one protein or the other. Additionally, the excitation is done at a wavelength where aromatic residues do not absorb so that no protein signal is observed.

Six Ocr amino acids localized on the surface of the Ocr protein were individually changed to Cys by Site-Directed Mutagenesis. The new Ocr(Cys) proteins were labelled with 1,5 I-AEDANS and Pyrene Maleimide, two fluorophores that interact with -SH moieties. Ocr contains no Cys so that the location of the fluorophore was accurately known.

However, certain assumptions must be made before the actual experiment starts. The first question to be answered is how the introduced cysteine amino acids affect the structure and the function of the Ocr protein? DSC experiments confirms that Ocr(Cys) proteins are as stable as the wild type. Only Ocr(W94C) shown a certain instability in the structure. The function of the Ocr(Cys) proteins seem to be as good as the wild type according to the methylation and restriction experiments, both *in vivo* and *in vitro* (see sections 4.3 and 4.4). The percentages of labelling varied from 20 to 55 % for 1,5 I-AEDANS and 6 to 35 % for Pyrene Maleimide. Ocr(Cys)-labelled protein is actually a mixture of labelled and unlabelled protein. It was assumed that methylase binds Ocr(Cys)-labelled protein as well as Ocr(Cys) unlabelled one. A further assumption is related to the fact that Ocr is a dimer. The consequence of this fact is that Ocr(Cys)-labelled fraction is a mixture of labelled-unlabelled and labelled-labelled proteins. It was assumed that both of these species bind equally to the methylase protein. Experiments done on different days when different percentages of labelling were obtained showed no

difference in the values of stoichiometry.

The experiment consists in titrating a certain amount of Ocr(Cys)-labelled protein (usually 1.0 μM protein concentration, no matter the percentage of labelling) with increasing amounts of methylase or nuclease. The fluorescent intensity of bound 1.5 I-AEDANS or PM increased until the end point when adding further methylase or nuclease produced no increase in the fluorescence intensity.

The fluorescence intensities were corrected for dilution and plotted against the molar ratio [methylase]/[Ocr(Cys)-labelled]. Once again the stoichiometry was calculated from the intersection of the two linear parts of the binding curve (Figure 4.3).

From these experiments, a stoichiometry of 1.2 $\pm$ 0.03 to 1 for Ocr(Cys)-labelled to methylase was found. The value is close to 1:1 stoichiometry that was expected for Ocr-methylase complex. The stoichiometry value for Ocr(Cys)-AEDANS and Ocr(Cys)-PM proteins complexed with methylase are shown in table 4.1.2. In practice, the same plots were used for determination of the Ocr-methylase complex stoichiometry and for studying the Ocr-methylase interface (see section 4.2).

The anisotropy experiments done in the same conditions as steady-state fluorescence assay confirm the stoichiometry values obtained (Figure 4.5, 4.6, 4.7 and table 4.1.2). D25C and D62C mutants labelled with I-AEDANS did not give reproducible results when the interaction with methylase was observed as the increase in the fluorescence intensity of the AEDANS moiety (Figures 4.12 and 4.13). However, when the binding was monitored as the increase in the AEDANS anisotropy, the results were reproducible (Figures 4.6 and 4.7). When the two mutants were labelled with PM, both steady state fluorescence and anisotropy gave reproducible results. Ocr(W94C) was not used in this assay due to its less stable structure which led to non-reproducible results.

As previously shown, for Ocr(S68C)-PM and Ocr(N4C)-PM, besides the emis-

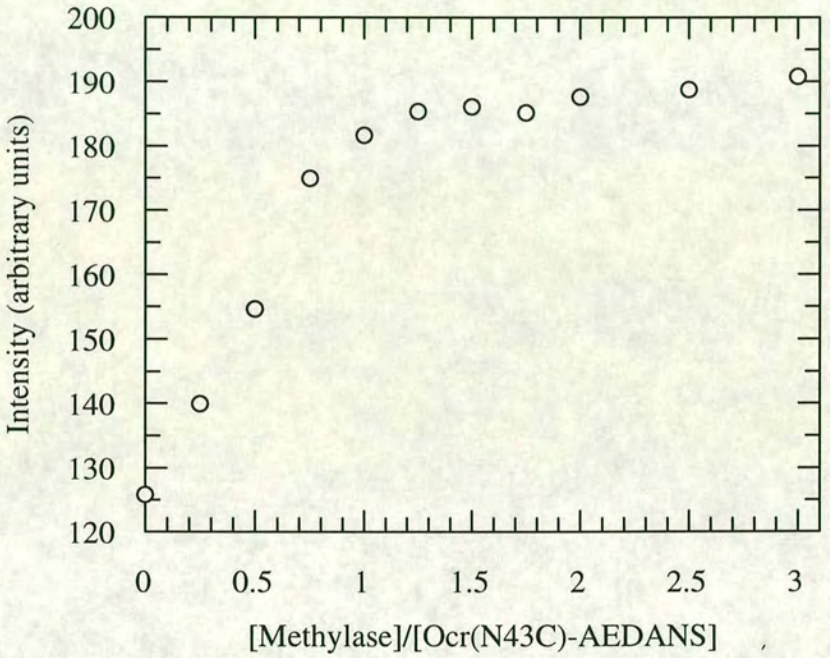


Figure 4.3: Ocr(N43C)-AEDANS:methylase stoichiometry. Increasing concentrations of methylase were titrated into a constant concentration of 1 μM Ocr(N43C)-AEDANS. The sample was excited at 330 nm and the emission maximum was observed at 480 nm. An emission spectrum of the buffer was recorded in exactly the same conditions and subtracted from the spectra of the samples. Spectra were corrected for any change in the volume due to addition of methylase and for the inner filter effect. The stoichiometry of binding was determined from the intersection of the extrapolated linear portions of the graphs. The experiment was performed at 25 °C using the excitation and emission slit widths of 5 nm and 10 nm respectively. The buffer was 20 mM NH_4Cl, 6 mM $MgCl_2$, 7 mM β-mercaptoethanol pH 8.

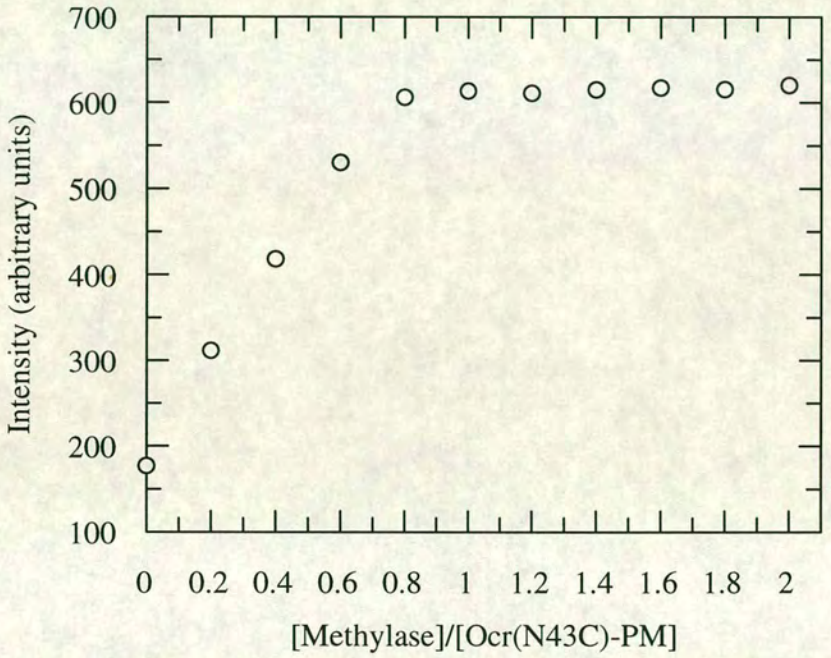


Figure 4.4: Ocr(N43C)-PM:methylase stoichiometry. Increasing concentrations of methylase were titrated into a constant concentration of 1 μM Ocr(N43C)-PM. The sample was excited at 342 nm and the emission was observed at 379 nm. An emission spectrum of the buffer was recorded in exactly the same conditions and subtracted from the spectra of the samples. Spectra were corrected for any change in the volume due to addition of methylase and for the inner filter effect. The stoichiometry of binding was determined from the intersection of the extrapolated linear portions of the graphs. The experiment was performed at 25 °C using the excitation and emission slit widths of 5 nm and 10 nm respectively. The buffer was 100 mM phosphate pH 7.

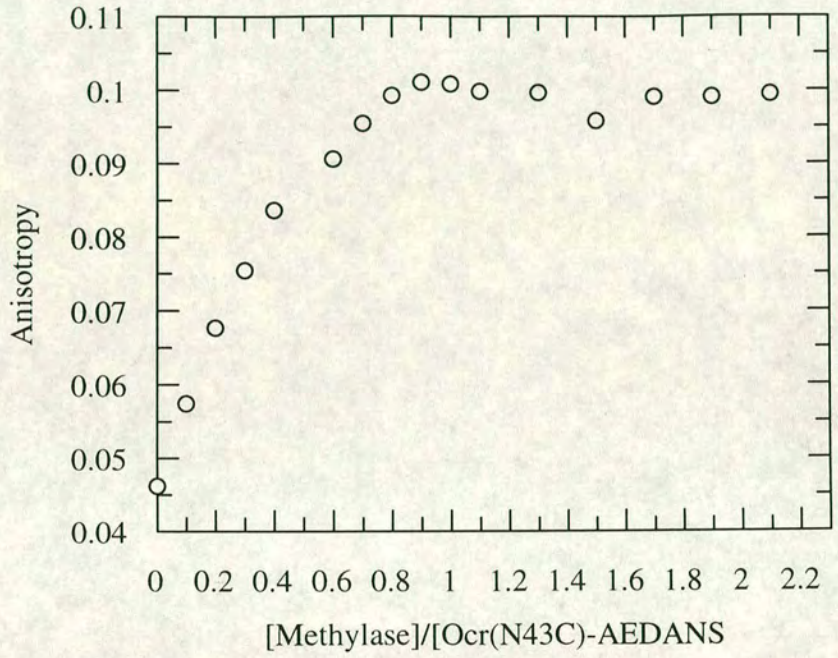


Figure 4.5: The stoichiometry of Ocr(N43C)-AEDANS:methylase complex. The increase in the fluorescence anisotropy of AEDANS-labelled Ocr(N43C) protein when titrated with *Eco*KI methylase. The sample was excited at 330 nm and the emission recorded at 480 nm. Five anisotropy values were measured after each addition of methylase and the calculated mean value was plotted *vs.* the molar ratio methylase:Ocr(N43C)-AEDANS. The experiment was performed at 25 °C and the excitation and emission slit widths were 15 and 20 nm respectively.

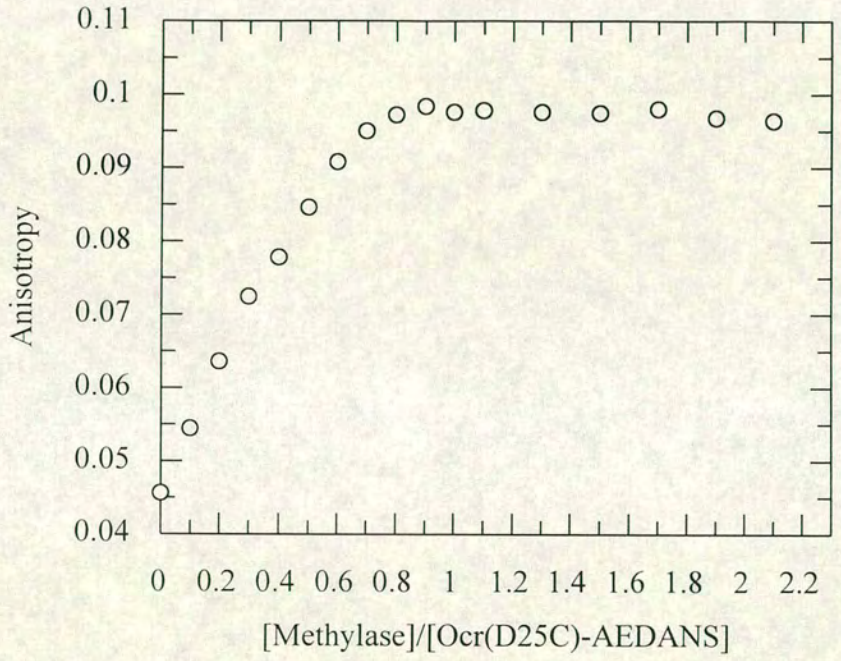


Figure 4.6: The stoichiometry of Ocr(D25C)-AEDANS:methylase complex. Anisotropy assay. The experiment was performed in exactly the same conditions as in figure 4.5.

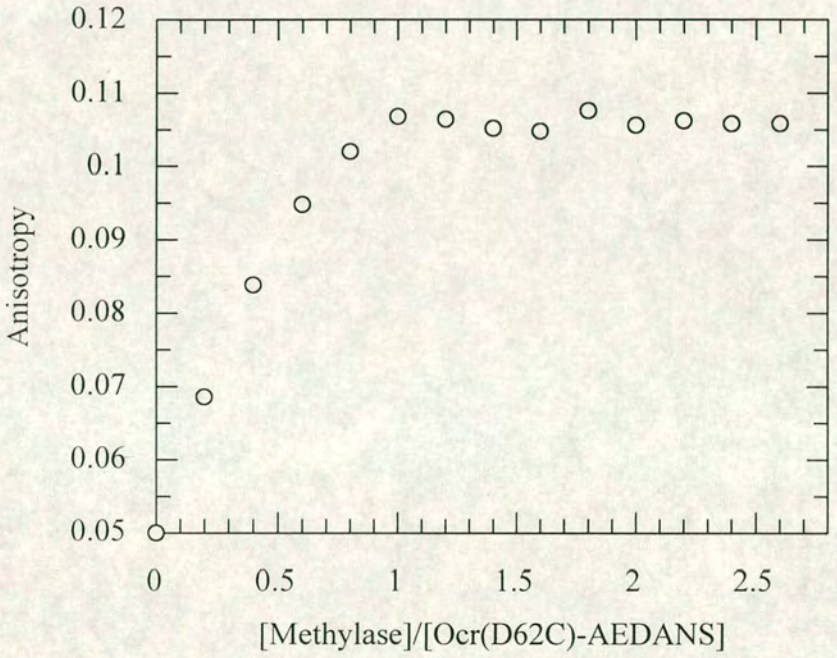


Figure 4.7: The stoichiometry of Ocr(D62C)-AEDANS:methylase complex. Anisotropy assay. The experiment was performed in exactly the same conditions as in figure 4.5.

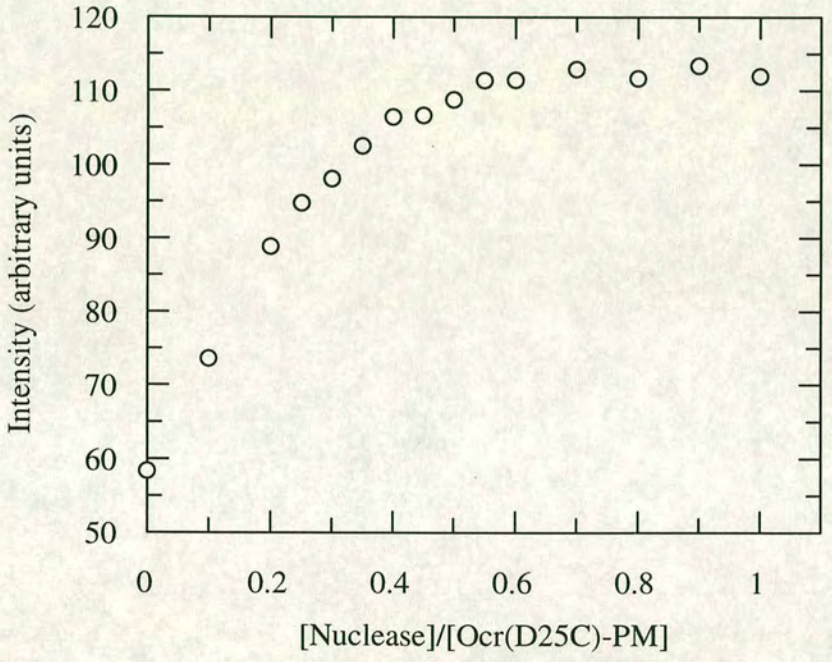


Figure 4.8: The stoichiometry of the Ocr(D25C)-PM:nuclease complex. Increasing concentrations of *Eco*KI nuclease were titrated into a constant concentration of 1 μM Ocr(D25C)-PM. The assay was performed in exactly the same conditions as those described in figure 4.4.

Table 4.1: The stoichiometry of the Ocr-methylase complex. Ocr(Cys) proteins were labelled either with I-AEDANS or pyrene maleimide, and the stoichiometry was determined by steady state fluorescence and anisotropy (nd stands for not determined).

Mutation	I-AEDANS		Pyrene Maleimide	
	Fluorescence	Anisotropy	Fluorescence	Anisotropy
N43C	1 ± 0.1	1.2 ± 0.07	1.2 ± 0.07	1.25
S68C	1.25	1.42	nd	1.25
D62C	nd	1	1.1 ± 0.1	1
D25C	nd	1.25	1.18 ± 0.05	1.4
N4C	1.1	1.33 ± 0.08	nd	nd
W94C	nd	nd	nd	nd

sion peaks at 378 and 395 nm , another one appears at 460 and 480 nm respectively, due to excimer formation between two close spaced PM molecules. When adding methylase, the intensity of the excimer decreased to the molar ratio 0.6:1 then increased again until 1:1 but never reached the initial value. This is probably due to conformational changes appearing after binding when either the distance between the two pyrene molecules changes or the pyrene molecules do not stack in such a precise symmetrical configuration (Figures 4.16 and 4.19).

Concomitant with excimer disappearance at 460 or 480 nm, an increase at 378 and 395 nm occured. Therefore, the increase in the fluorescence intensity at 378 nm is partially due to the solvent protection conferred by methylase and partially to the energy transfer from the excimer. As seen in figures 4.16 and 4.19, the shape of the binding curves for Ocr(N4C)-PM and Ocr(S68C)-PM are not formed by linear parts, one ascending till the 1:1 stoichiometry and the other as a plateau as obtained for the other four Ocr(Cys)-PM proteins. Due to this fact, these two binding curves were not used for determination of the Ocr-methylase stoichiometry.

In the case of Ocr(D25C)-PM protein titrated with nuclease a stoichiometry of 2.1 ± 0.1 to 1 was found (Figure 4.8). The stoichiometry of nuclease with the

other Ocr(Cys) proteins has not been yet determined.

4.1.3 Isothermal Titration Calorimetry

ITC is a very sensitive method that enables determination of the thermodynamic parameters and stoichiometry of a reaction by direct measurement of the released or absorbed heat (Cooper, 1998; Jelesarov and Bosshard, 1999; Burova *et al.*, 1999). Due to the fact that in many cases the heat of the reaction is relatively small, high concentration of proteins are required. This is the main limitation of this method.

Figure 4.9 shows a typical titration calorimetry measurement which consisted of adding aliquots of 64.54 μM Ocr solution to 4.8 μM methylase solution at 25 °C. The reaction was exothermic and each addition of Ocr lowered the amount of heat released until the reaction was saturated. In two separate control experiments, the heat of dilution was obtained by injection of Ocr into buffer and also buffer into buffer solution. After integration of each injection peak and subtraction of the heats of dilution, the values of the heat of interaction normalized for Ocr concentration was plotted *vs.* the molar ratio Ocr:methylase.

The upper panel shows the raw data. The lower panel shows the experimental data fitted to a theoretical titration curve with ΔH^0, K, ΔS^0 and n as floating parameters.

where

ΔH^0	enthalpy of reaction (Kcal/mol)
K	association constant (M^{-1})
ΔS^0	the entropy (cal/mol·K)
n	number of the binding sites

Because of the high concentration of Ocr and methylase required for ITC experiment only ΔH^0 and n can be accurately determined. Indeed the association constant obtained from ITC assay ($4.1 \cdot 10^8$ M^{-1}) was 1000 times lower than the real association constant determinated from a fluorescence titration experiment

(see latter this chapter).

The most desired parameter from the ITC experiment was the stoichiometry of binding (n) whose value was found to be 0.95 ± 0.005. We confirmed in this way the 1:1 stoichiometry obtained in the extrinsic fluorescence experiment.

Besides the stoichiometry value, the ITC experiments give some information concerning the thermodynamics of binding. The real dissociation constant (K_d) obtained from a fluorescence titration experiment, having the value of $4.8 \cdot 10^{-11}$ M, was used in the calculation of the Gibbs free energy (ΔG^0)

$$\Delta G^0 = RTln(K_d) \tag{4.1}$$

where

R gas constant
T temperature in degrees Kelvin

The value obtained for ΔG^0 was -14090 cal/mol. The Gibbs free energy could be also expressed as:

$$\Delta G^0 = \Delta H^0 - T\Delta S^0 \tag{4.2}$$

The entropy of binding was easily calculated from equation 4.2 by using ΔH^0 value of -18.240 ± 197.3 cal/mol obtained directly from the ITC assay. The value obtained for ΔS^0 was -13.93 cal/mol·K. As easily noticed, enthalpy contributes favorably to the Gibbs free energy while the contribution of entropy is unfavorable. It seems that an enthalpy-entropy compensation occurred. However, the data collected are not sufficient to draw this conclusion. It will be necessary to repeat the ITC experiment at different temperatures in order to see if the antagonistic contribution effect of enthalpy and entropy to the total free energy obtained at 25 °C is consistent at all temperatures used. When performing the ITC experiment at different temperatures it will be also possible to determine (Tame *et al.*, 1998) the change in the constant pressure heat capacity (ΔC_p) for the methylase-Ocr interaction from the formula:

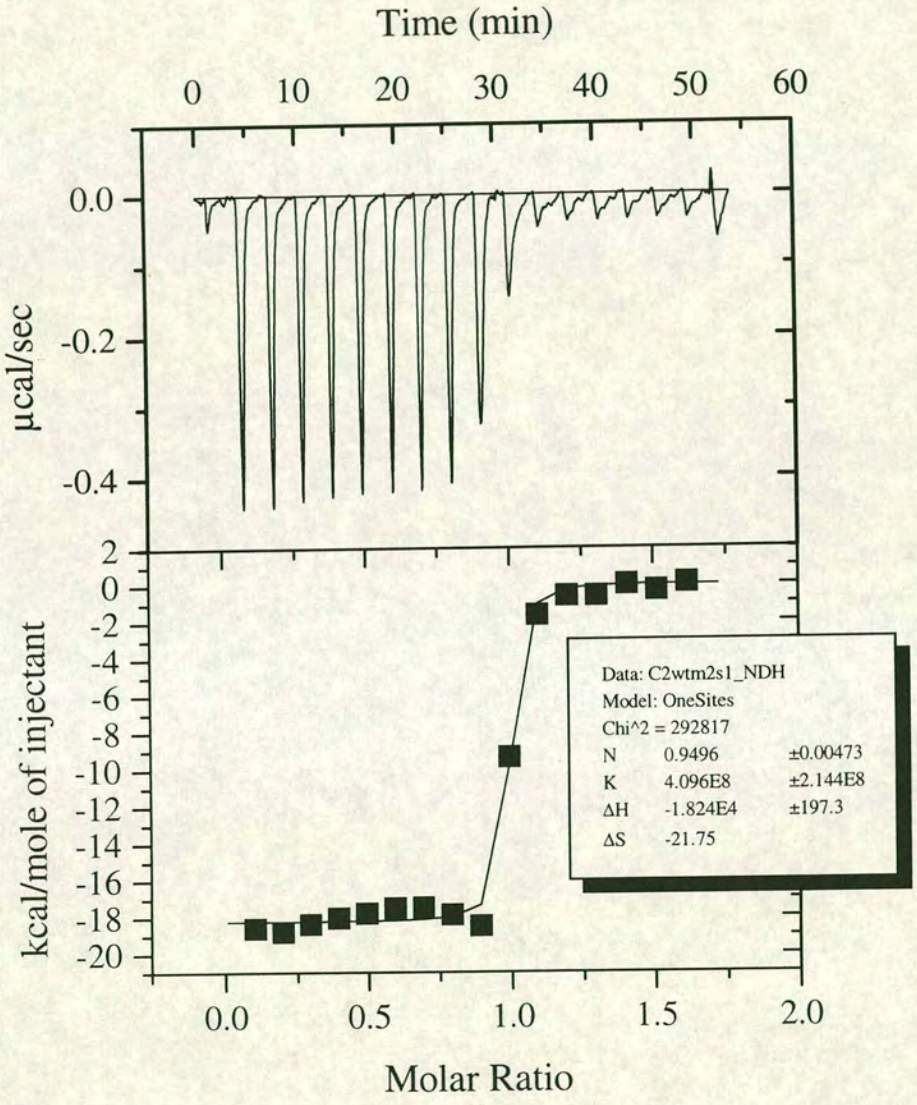


Figure 4.9: Isothermal calorimetric titration curves of *Eco*KI methylase with Ocr. (a) Raw data. Methylase is placed in the cell and Ocr in the syringe. (b) The calorimetric binding isotherm for the Ocr-methylase system. The experimental data were fitted to a theoretical titration curve with ΔH, K, ΔS and n as floating parameters. The buffer was 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$, pH 8 and the temperature 25 °C.

$$\Delta C_p = \frac{\Delta H_{T_2} - \Delta H_{T_1}}{T_2 - T_1} = \frac{\Delta S_{T_2} - \Delta S_{T_1}}{ln(\frac{T_2}{T_1})} \tag{4.3}$$

Depending on the values of ΔC_p obtained it will be possible to predict the surface area involved in binding and also the nature of the interactions involved. In the past, large negative changes in heat capacity have been correlated with large surface areas buried at the interface (Tame *et al.*, 1998). However, we must be cautions when interpreting this kind of data. In the last few years many examples of large ΔC_p which are not correlated with large surfaces involved in binding have been reported (Guinto and Di Cera, 1996).

Due to the low overexpression of the nuclease, it is not possible to perform an ITC experiment to get an accurate stoichiometry value for Ocr-nuclease complex. However, it seems that the extrinsic fluorescent assay is accurate enough to offer a reliable stoichiometry value. Consequently, we consider the stoichiometry of 2:1 for Ocr-nuclease found in extrinsic fluorescence assay to be the true stoichiometry for this complex.

4.2 Ocr-methylase interface

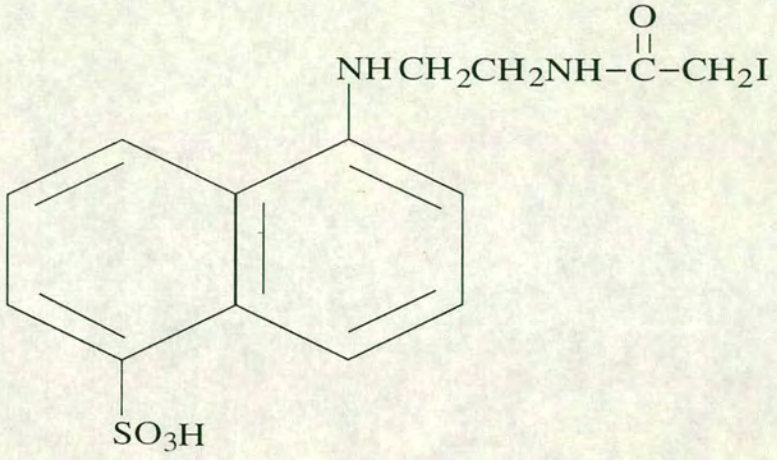
4.2.1 Interaction of Ocr(Cys)-AEDANS proteins with methylase

One of the fluorophores used for labelling Ocr(Cys) proteins was 1,5 I-AEDANS (Figure 4.10). The fluorophore is able to interact with SH moieties of the Cys residues making a covalent bond. A molecule of HI is released during the reaction.

It is well known that 1,5 I-AEDANS is very sensitive to any changes in its environment (Latham *et al.*, 1997). This sensitivity was exploited in order to identify which of the Cys-substituted amino acids are in the binding site of methylase.

All six Ocr(Cys) proteins were labelled which is an indication that all of them are exposed on the surface. The difficulties in labelling N4C residue was due to the disulphide bond formed between N4C residues, one in each monomer. After

(a)



(b)

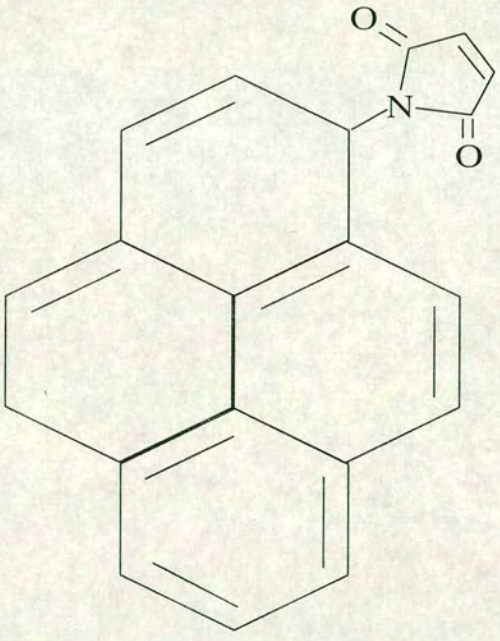


Figure 4.10: Structures of the sulfhydryl-specific reporter groups attached to Ocr(Cys) proteins. (a) 1,5 I-AEDANS. (b) Pyrene Maleimide.

reducing the disulphide bond with β-mercaptoethanol, the N4C residue was easily labelled.

The first amino acid to be labelled was N43C. After labelling, Ocr(Cys)-AEDANS was titrated with increasing amounts of methylase. Keeping in mind that the fluorescence comes only from the excited fluorophore one could expect that no change in the fluorescence intensity would occur when a nonfluorescent protein is added. However, the fluorescence intensity of AEDANS increased linearly at each titration point until the molar ratio methylase: Ocr(Cys)-AEDANS became 1:1 when adding more methylase did not produce further increase of the fluorescence intensity. The only way to explain this increase was either that the N43C residue was located at or near the binding site for methylase or this site was sensitive to a conformational change due to binding of the methylase (Latham *et al.*, 1997). Coming in contact with the methylase the AEDANS moiety was protected from the solvent molecules that are known to lower the fluorescence intensity. A blue shift in the maximum emission spectra was also noticed, another indication that the environment of AEDANS became more hydrophobic in the presence of methylase.

The fluorescence intensity at the emission wavelength of 480 nm was recorded after each titration step, then values were corrected for any change in the volume due to addition of methylase and for the inner filter effect and plotted against the molar ratio methylase: Ocr(Cys)-AEDANS.

Titration assays done with all six labelled Ocr(Cys)proteins led to the same result, an increase in the fluorescence intensity for every amount of methylase added till the molar ratio 1:1 was reached (Figures 4.3, 4.11, 4.12, 4.13, 4.14, 4.15). It seems that all six labelled cysteine are located at or near the binding site. However, another explanation could be possible. For example, one of the labelled amino acids could be away from the binding site but conformational changes appearing in the Ocr after binding to methylase leads to an increase

in the hydrophobicity of the fluorophores environment which causes in turn an increase in the fluorescence intensity. Further experiments tried to rule out the last hypothesis.

4.2.2 Interaction of Ocr(Cys)-PM proteins with methylase

1,5 I-AEDANS specifically binds Cys residues. However, especially in the absence of the Cys residues, a small percentage of His residues could be also labelled. This raises concerns about non-specific labelling.

In order to eliminate this possibility another Cys-specific fluorophore was used. This fluorophore, Pyrene Maleimide (PM), does not label anything but Cys residues. Due to its hydrophobicity, a lower percentage of labelling was obtained. All six Cys-replaced amino acids were labelled confirming that all of them are solvent exposed.

After labelling, all six Ocr(Cys)-PM proteins were titrated with methylase. Samples were excited at 342 nm and the emission was followed at 379 nm. As seen in figures 4.4, 4.16, 4.17, 4.19 and 4.20, the fluorescence intensity of PM increased at each addition of methylase until the molar ratio between methylase and labelled Ocr became 1:1, then remained constant, except for Ocr(D62C)-PM where fluorescence decreased (Figure 4.18). The results were reproducible, confirming that the increase in the fluorescence intensity was due to the labelled amino acids being in the binding site, not because of an anomalous behavior of 1.5 I-AEDANS.

The cause for the decrease in pyrene intensity for Ocr(D62C)-PM mutant is not very clear. Probably the increase in pyrene intensity due to binding to methylase is compensated by a stronger quenching effect exerted by amino acids which belong to methylase. The decrease in pyrene intensity is still an indication that D62C residue is in the binding site otherwise the intensity would be expected to remain constant after Ocr(D62C)-PM binds methylase.

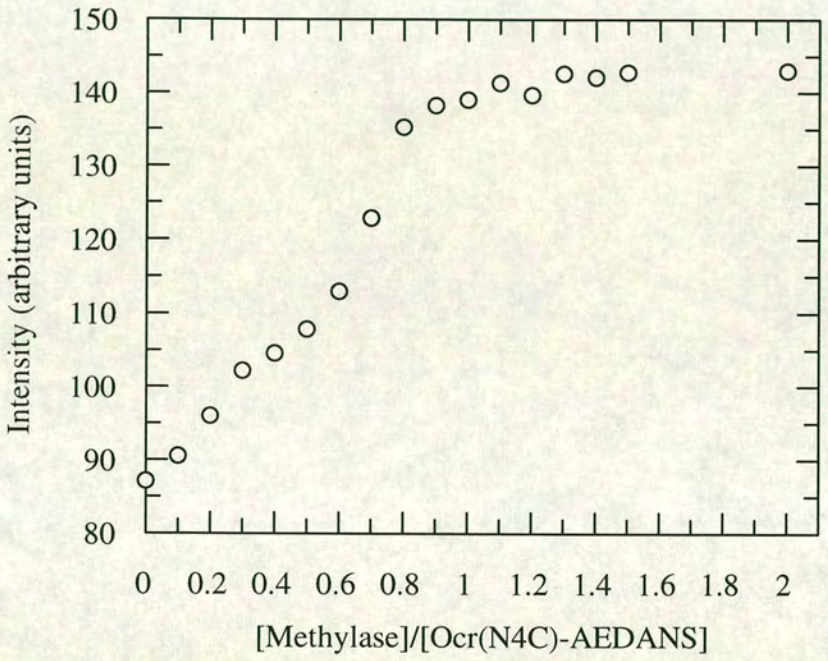


Figure 4.11: Titration of Ocr(N4C)-AEDANS with *Eco*KI methylase. The experiment was performed in exactly the same conditions as those described in figure 4.3 legend

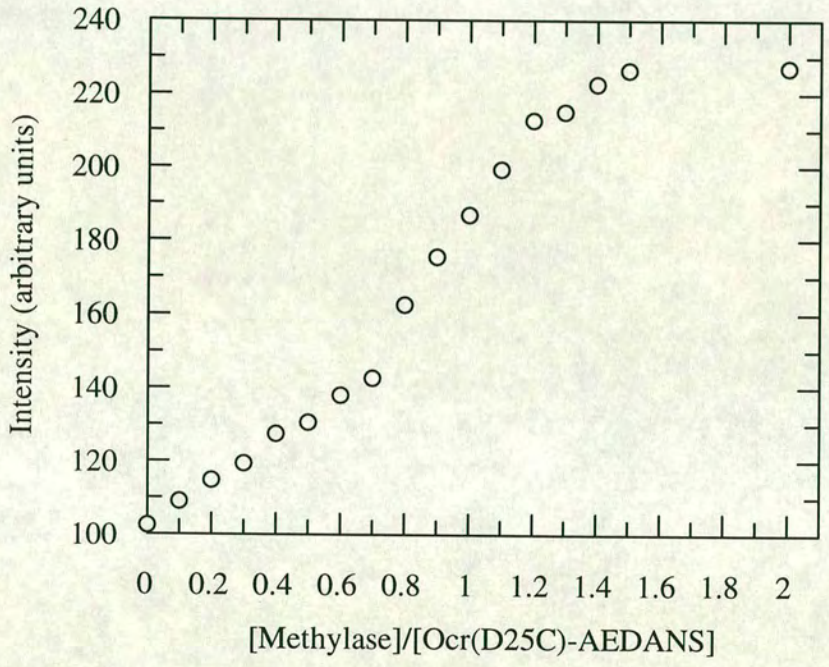


Figure 4.12: Titration of Ocr(D25C)-AEDANS with *Eco*KI methylase. The experiment was performed in exactly the same conditions as those described in figure 4.3 legend

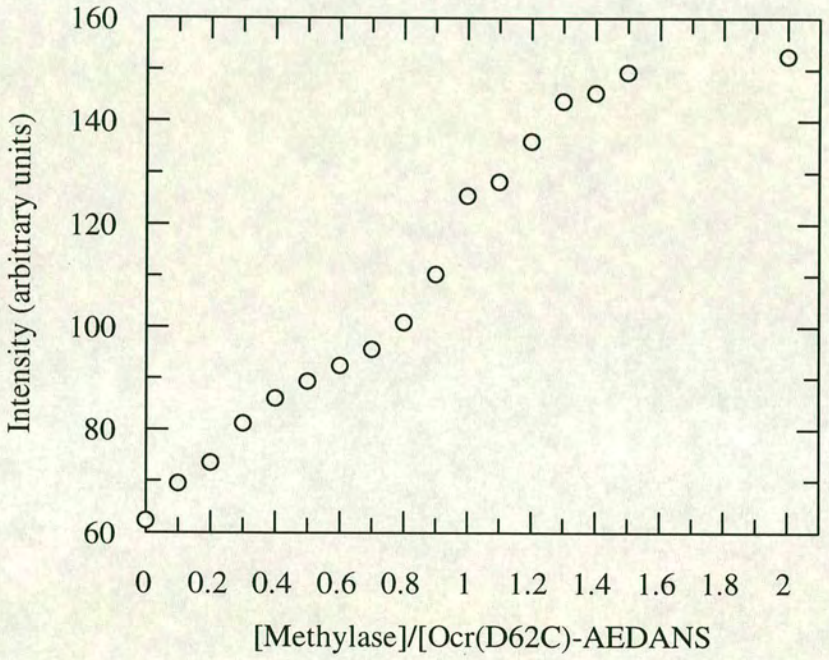


Figure 4.13: Titration of Ocr(D62C)-AEDANS with *Eco*KI methylase. The experiment was performed in exactly the same conditions as those described in figure 4.3 legend

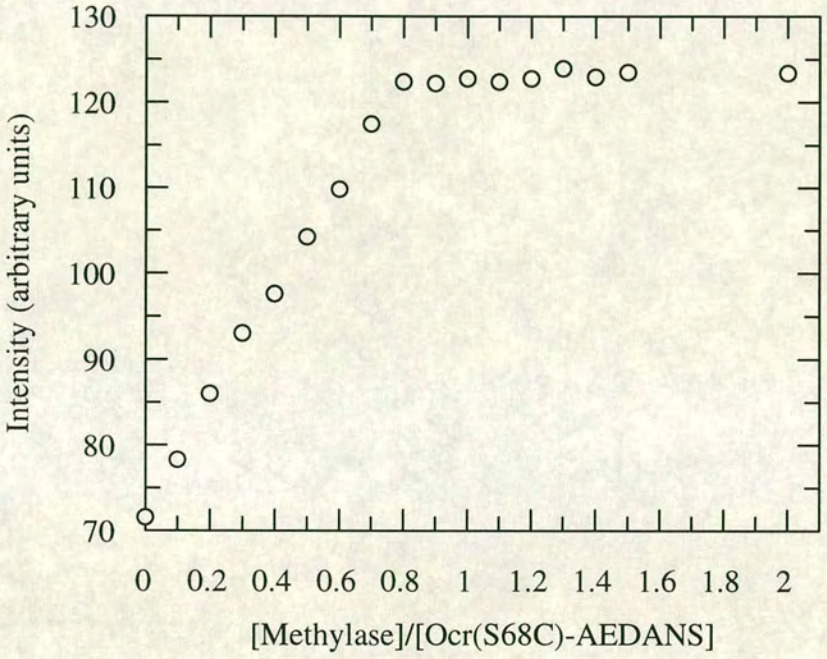


Figure 4.14: Titration of Ocr(S68C)-AEDANS with *Eco*KI methylase. The experiment was performed in exactly the same conditions as those described in figure 4.3 legend

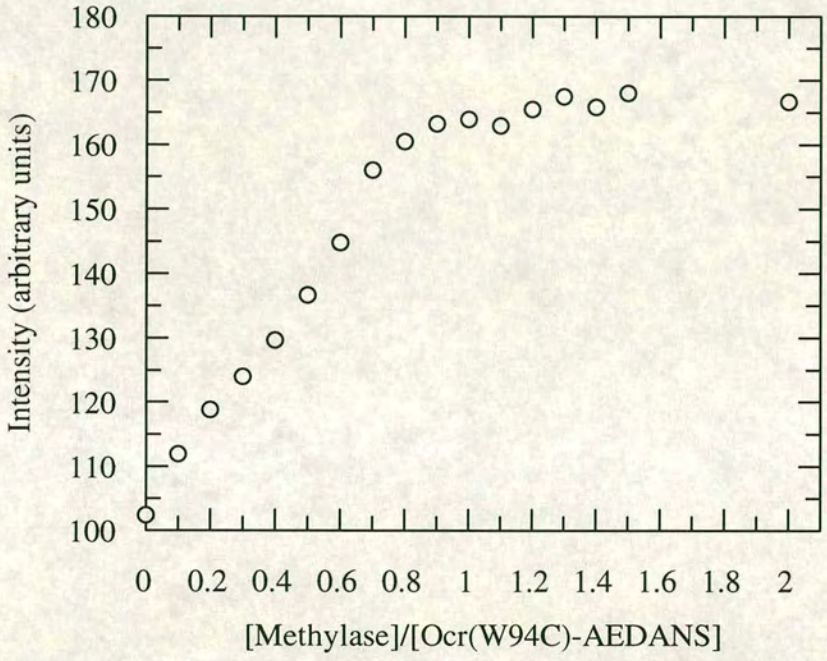


Figure 4.15: Titration of Ocr(W94C)-AEDANS with *Eco*KI methylase. The experiment was performed in exactly the same conditions as those described in figure 4.3 legend

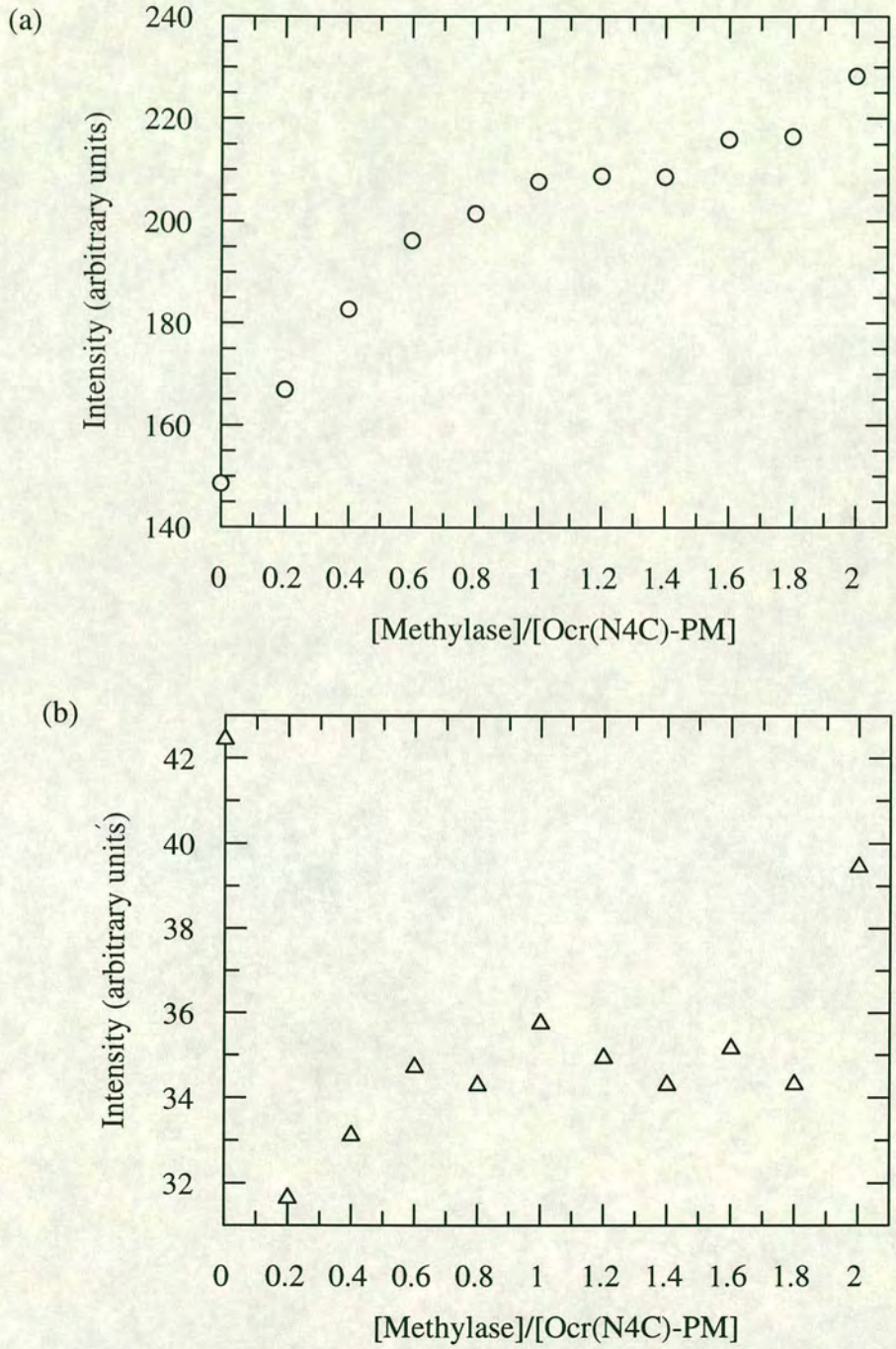


Figure 4.16: Titration of Ocr(N4C)-PM with *Eco*KI methylase. (a) The increase in the fluorescence intensity of PM at 379 nm in the presence of methylase (○). (b) The decrease in the fluorescence intensity at the excimer position (480 nm) in the presence of methylase (△). The experiment was performed in exactly the same conditions as those described in figure 4.4 legend

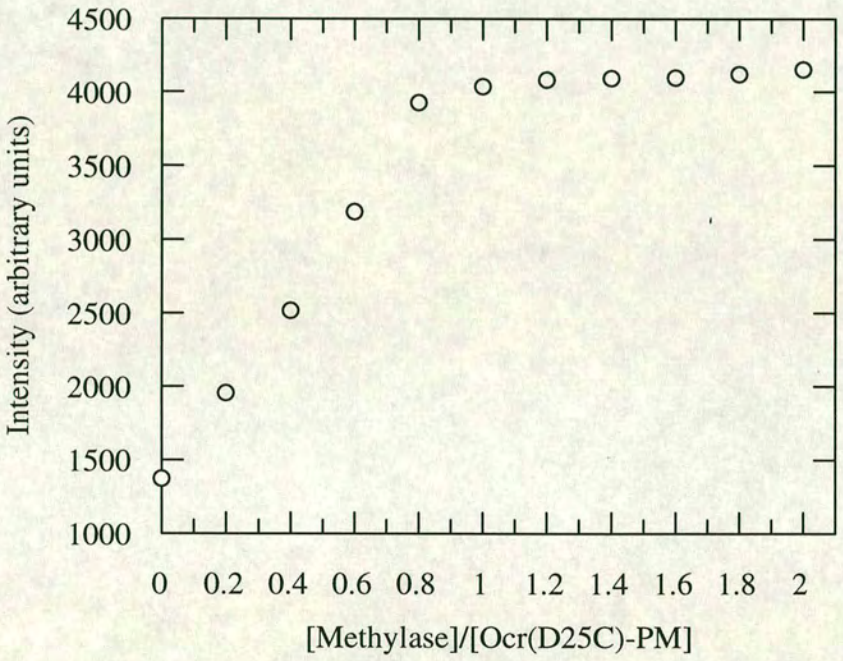


Figure 4.17: Titration of Ocr(D25C)-PM with *Eco*KI methylase. The experiment was performed in exactly the same conditions as those described in figure 4.4 legend

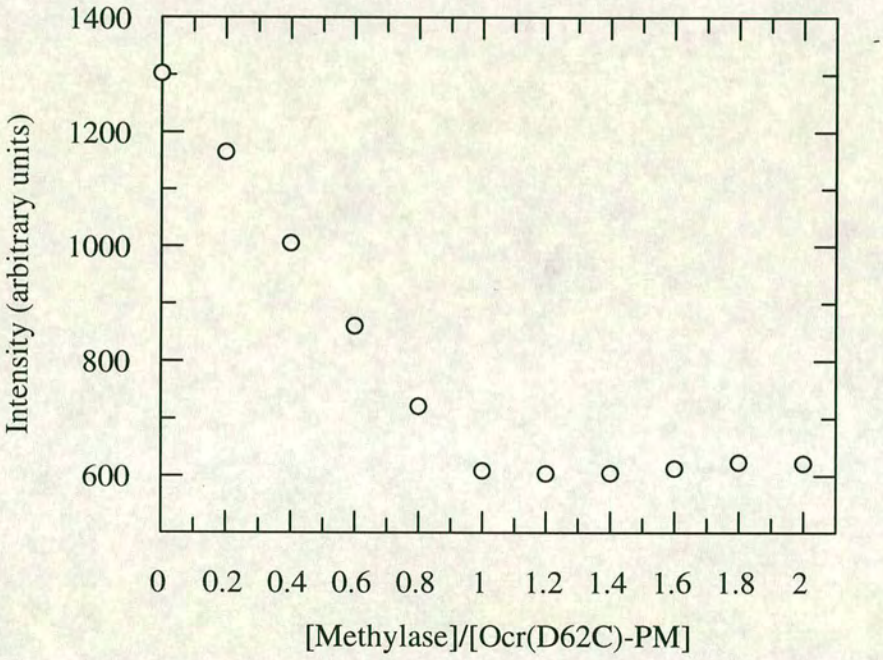


Figure 4.18: Titration of Ocr(D62C)-PM with *Eco*KI methylase. The experiment was performed in exactly the same conditions as those described in figure 4.4 legend

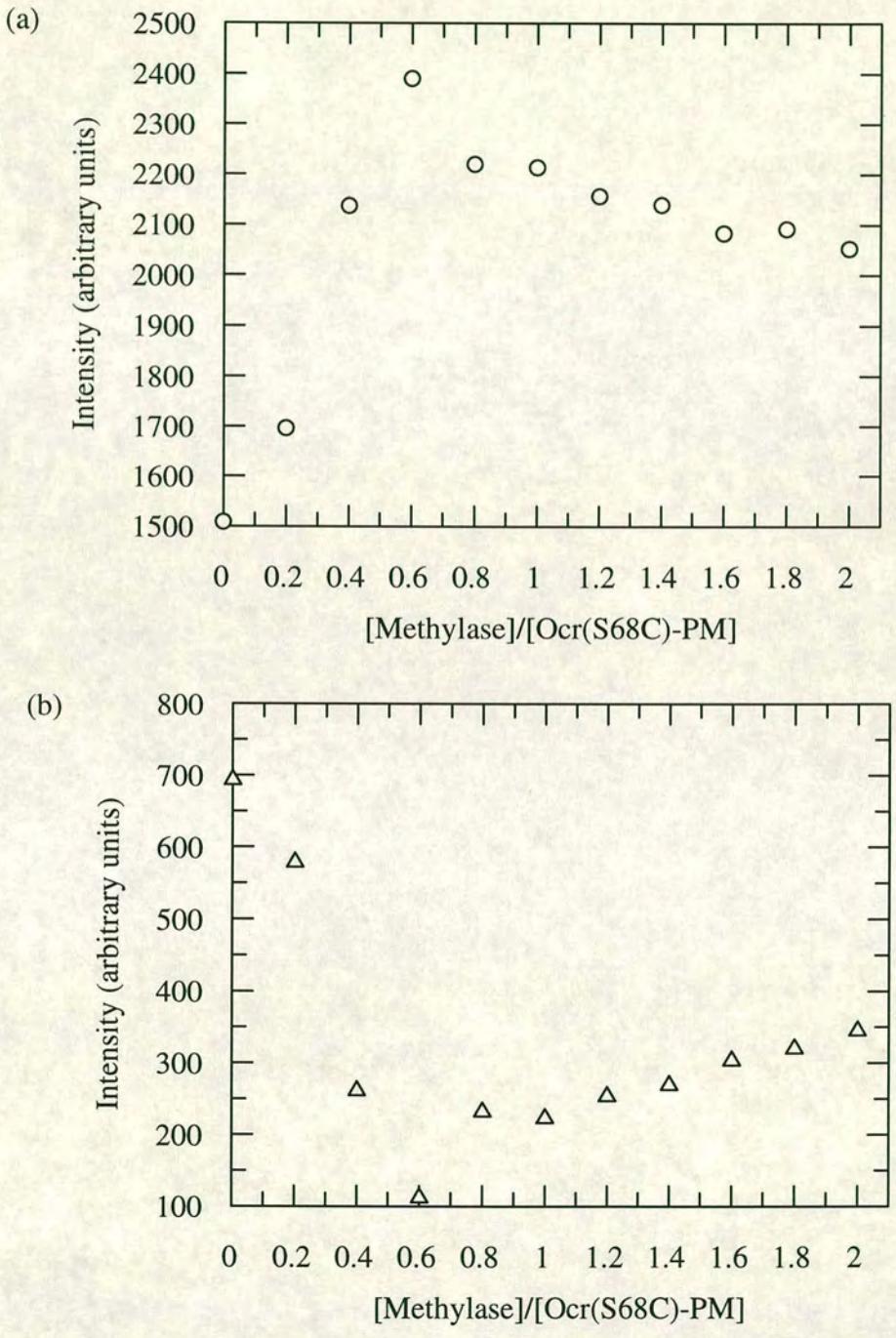


Figure 4.19: Titration of Ocr(S68C)-PM with methylase. (a) The increase in the fluorescence intensity of PM at 379 nm (○) in the presence of methylase. (b) The decrease in the fluorescence intensity at the excimer position (460 nm) in the presence of methylase (△). The experiment was performed in exactly the same conditions as those described in figure 4.4 legend

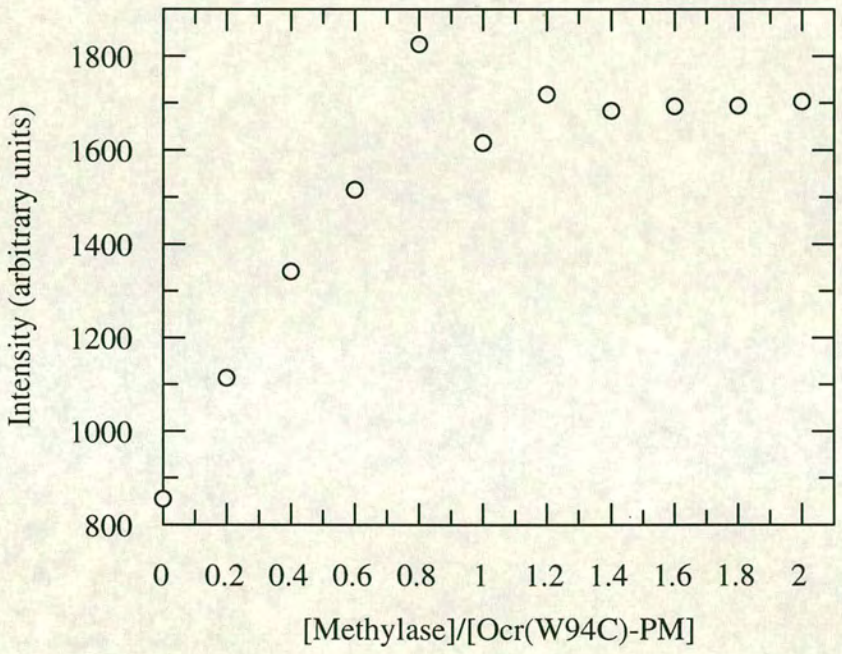


Figure 4.20: Titration of Ocr(W94C)-PM with methylase. The experiment was performed in exactly the same conditions as those described in figure 4.4 legend

4.2.3 Quenching

Background

The intrinsic or extrinsic fluorescence of proteins can be quenched by a variety of processes: collisional quenching, static quenching, energy transfer, excited-state reactions and molecular rearrangements (Lakowicz, 1999).

Collisional quenching requires molecular contact between the fluorophore and the quencher. The quencher interacts with an excited fluorophore causing the fluorophore to return to the ground state without emission of a photon. The quencher must diffuse to the fluorophore during the lifetime of the excited state. Consequently, the efficiency of quenching will be proportional to the lifetime of the excited state (τ) and the quenching diffusion coefficient (D). The average distance, $\sqrt{\Delta x^2}$, a quencher can diffuse during the excited-state lifetime is given by the Einstein equation:

$$\sqrt{\Delta x^2} = \sqrt{2D\tau} \tag{4.4}$$

The oxygen molecule is known as a good collisional quencher. Practically, oxygen quenches almost any fluorophore. Indole and carbamazole are quenched by I^+, Cs^+, cysteine, $NO_3^-, Cu^{2+}, Mn^{2+}, Pb^{2+}, Cd^{2+}$ by a mechanism which probably involve transfer of an electron to the quencher. Indole is also quenched by succinimide, methionine, acrylamide, and Ag^+ (Lakowicz, 1999). Collisional quenching is described by the Stern-Volmer equation:

$$\frac{F_0}{F} = 1 + K_q\tau_0[Q] = 1 + K_D[Q] \tag{4.5}$$

where

F_0	fluorescence intensity in the absence of quencher
F	fluorescence intensity in the presence of quencher
K_q	bimolecular quenching constant
τ_0	lifetime of the fluorophore in the absence of quencher

A plot of F_0/F vs. [Q], which is expected to be linear, will give a slope equal to K_D. Knowing the fluorophore lifetime allows the determination of K_q which characterizes the quencher accessibility to fluorophore.

Results

The results obtained in sections 4.2.1 and and 4.2.2, when either Ocr(Cys)-AEDANS or Ocr(Cys)-PM interacted with methylase were quite surprising. All six labelled amino acids were not expected to be in the binding site. Before drawing any conclusion it was vital to be absolutely sure that no artifact had occured.

As pointed out before, the increase in the fluorescence intensity could also be due to a change in the neighboring environment of the labelled amino acids. We designed a quenching experiment in order to prove or to disprove this assumption. We used acrylamide, a polar noncharged electron deficient molecule that efficiently quenches the 1,5 I-AEDANS fluorophore when in direct contact with it (Wasylewski *et al.*, 1988). If a certain Cys-AEDANS amino acid is really in the binding site, the acrylamide will not be able to quench the AEDANS because the fluorophore will be protected by the methylase. However, if the increase in the fluorescence intensity is due to the changes in in the local environment around the AEDANS moiety, acrylamide will be able to quench the AEDANS fluorescence because there will be nothing to prevent the interaction between acrylamide and AEDANS.

First, a solution of Ocr(Cys)-AEDANS (1 μM as protein concentration) was titrated with a concentrated acrylamide solution. Being exposed on the surface, AEDANS was easily quenched by acrylamide. However, the quenching of the AEDANS moiety bound to the protein was less efficient than the quenching of a 1,5 I-AEDANS solution. Different Ocr(Cys)-AEDANS proteins showed different quenching efficiency.

Ocr(Cys)-AEDANS proteins were allowed to form a complex with methylase

at a molar ratio of 1:1. A two-fold excess of methylase was present ensuring that there is no free Ocr(Cys)-AEDANS in solution. Virtually no quenching appeared when acrylamide was added, suggesting that the acrylamide has no direct access to the fluorophore. This is a clear indication that all six labelled amino acids are really in the binding site.

The fact that Ocr amino acids in positions 4, 25, 43, 62, 68 and 94 are all at the Ocr-methylase interface suggests a huge binding site for methylase. If our first assumption that Ocr mimics the structure of DNA is correct, the above results are not so surprising. In the case of UGI, a DNA mimic protein that interacts with human UDG-ase enzyme, almost one quarter of the total UGI solvent-accessible surface is involved in binding (Mol *et al.*, 1995).

4.2.4 Importance of Trp94

Dunn *et al.* (1981) made the first attempt to demonstrate the importance of the unique Trp in position 94 (one in each monomer). They were looking for the antirestriction and antimethylation activity of different Ocr proteins in which a certain number of amino acids from the C-terminal end were deleted. Ocr proteins that contained 94 or more amino acids behaved like the wild type, but one that had only 93 amino acids was unable to knock out the methylase or the nuclease. Hence, they suggested that Trp94 could be important for Ocr activity or stability.

In order to determine the role of Trp94, this amino acid was replaced with Cys by Site-Directed Mutagenesis. As shown in figure 3.11 the T_m of Ocr(W94C) is 68.57 °C, which is 3.52 °C lower than that of the wild type, indicating that the Cys-substituted protein is less stable than the wild type.

We tried then to see if Ocr(W94C) is still able to bind the methylase protein. For this purpose a short experiment was designed, in which methylase was titrated with increasing amounts of Ocr(W94C). An emission spectrum between 300 and 500 nm, after excitation at 295 nm, was recorded for each titration step.

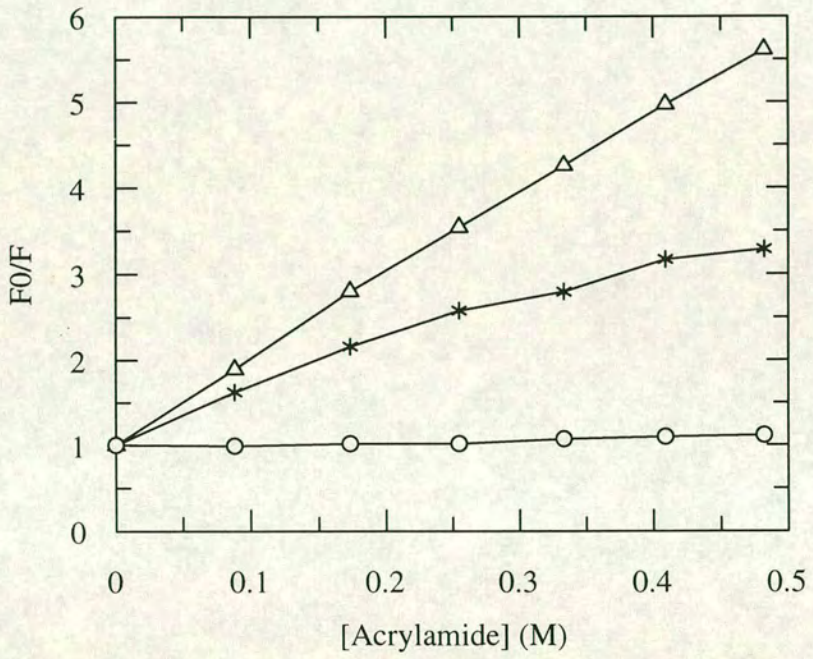


Figure 4.21: Stern-Volmer plots for acrylamide quenching of I-AEDANS ($\triangle$), Ocr(N43C)-AEDANS ($*$) and Ocr(N43C)-AEDANS:methylase ($\circ$). The concentration of methylase and Ocr(N43C)-AEDANS were 2 μM and 1 μM respectively. Samples were excited at 337 nm and the quenching effect was observed at 480 nm. The assay was performed at 25 $^\circ$ using the excitation and emission slitwidths of 5 nm and 10 nm respectively. The buffer used was 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$, pH 8.

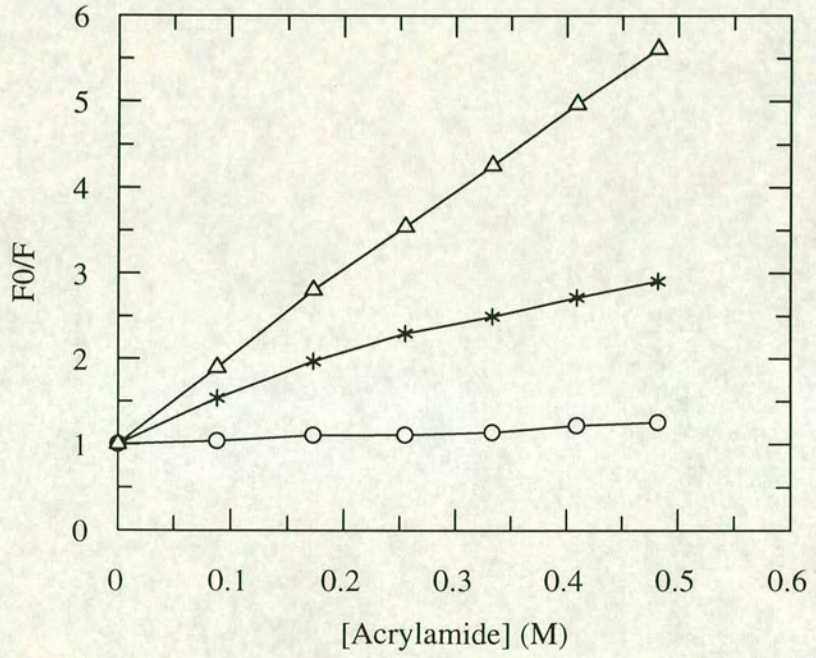


Figure 4.22: Stern-Volmer plots for acrylamide quenching of I-AEDANS ($\triangle$), Ocr(N4C)-AEDANS ($*$) and Ocr(N4C)-AEDANS:methylase ($\circ$). The experiment was performed in exactly the same conditions as in figure 4.21.

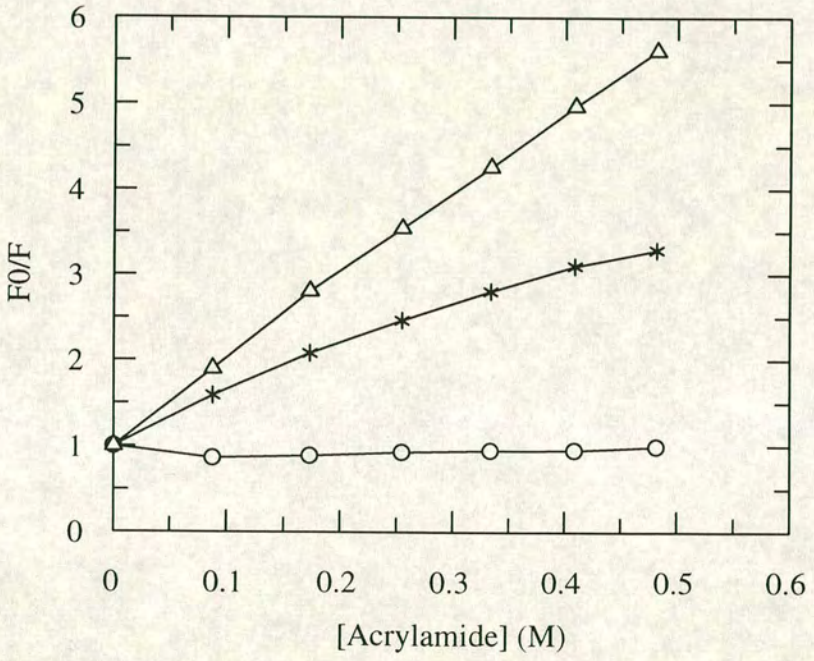


Figure 4.23: Stern-Volmer plots for acrylamide quenching of I-AEDANS ($\triangle$), Ocr(D25C)-AEDANS ($*$) and Ocr(D25C)-AEDANS:methylase ($\circ$). The experiment was performed in exactly the same conditions as in figure 4.21.

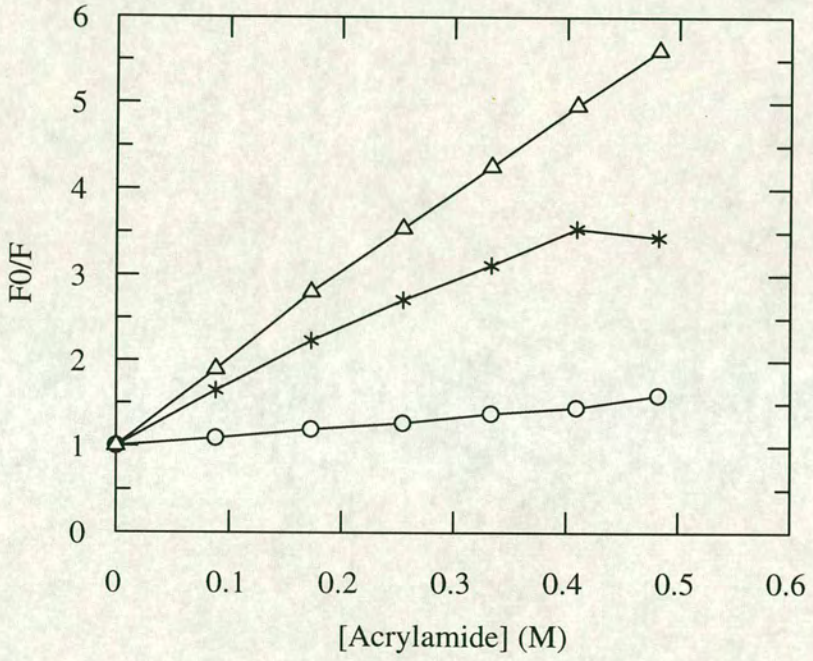


Figure 4.24: Stern-Volmer plots for acrylamide quenching of I-AEDANS ($\triangle$), Ocr(D62C)-AEDANS ($*$) and Ocr(D62C)-AEDANS:methylase ($\circ$). The experiment was performed in exactly the same conditions as in figure 4.21.

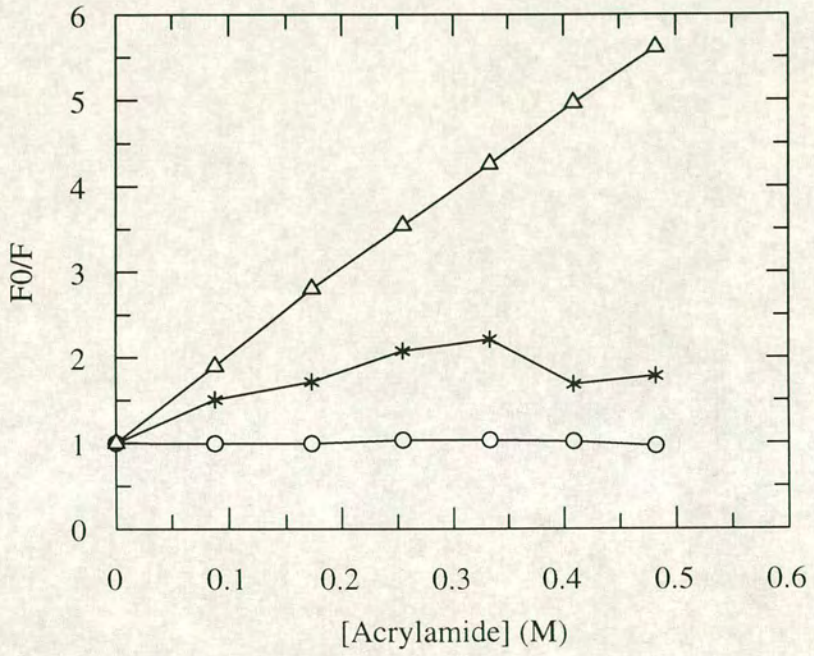


Figure 4.25: Stern-Volmer plots for acrylamide quenching of I-AEDANS (△), Ocr(S68C)-AEDANS (∗) and Ocr(S68C)-AEDANS:methylase (○). The experiment was performed in exactly the same conditions as in figure 4.21.

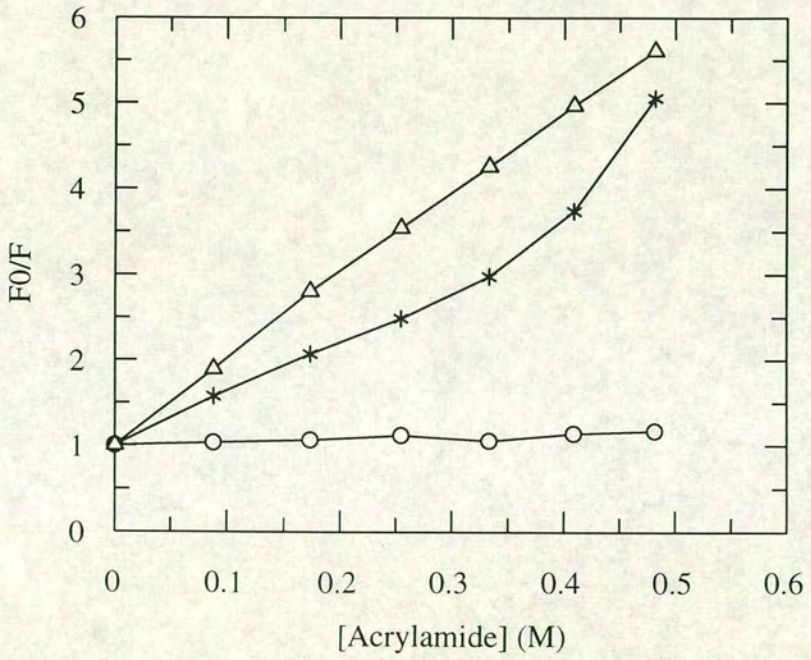


Figure 4.26: Stern-Volmer plots for acrylamide quenching of I-AEDANS ($\triangle$), Ocr(W94C)-AEDANS ($*$) and Ocr(W94C)-AEDANS:methylase ($\circ$). The experiment was performed in exactly the same conditions as in figure 4.21.

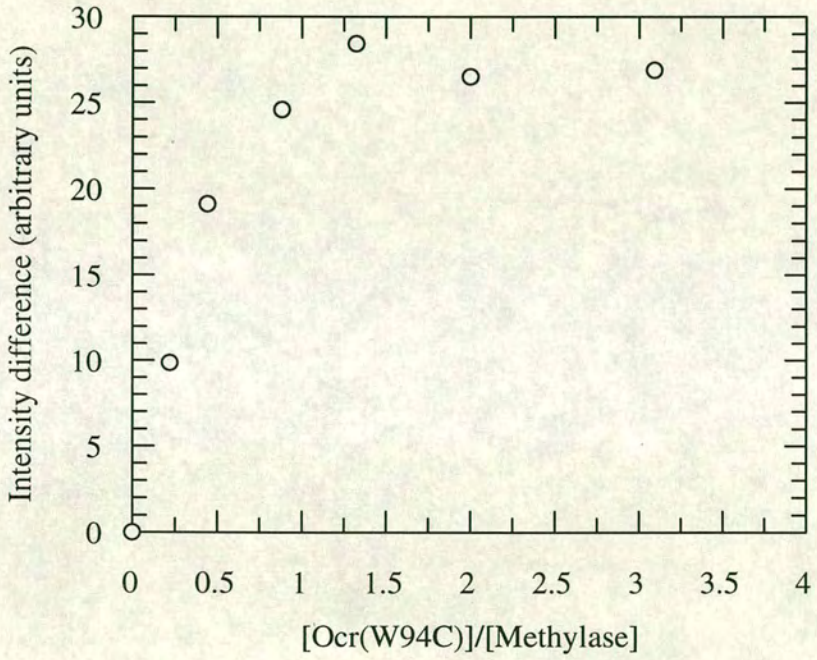


Figure 4.27: Interaction of *Eco*KI methylase with Ocr(W94C). The assay was performed in exactly the same conditions as those described in figure 4.1.

Being devoid of any Trp, Ocr(W94C) will not give any fluorescence signal after excitation at 295 nm. The only fluorescence will come from the Trp residues of the methylase protein. The interpretation of the results is very easy: if Ocr(W94C) is able to bind methylase, the conformational changes that appear following binding will produce a change in the fluorescence intensity of the Trp residues of methylase. If no binding occurs, the fluorescence intensity will remain the same, no matter how much Ocr(W94C) will be added (of course, after the fluorescence intensity is corrected for dilution effect).

Plotting the difference between the fluorescence intensity of the methylase in the absence and in the presence of Ocr(W94C) *vs.* Ocr(W94C) concentration, clearly indicate that Ocr(W94C) is able to bind *Eco*KI methylase. However, from the shape of the binding curve it is easily noticed that Ocr(W94C) does not bind the methylase as strongly as the wild type Ocr (Figure 4.2.4).

Three more experiments were done in order to establish the importance of Trp94. As we will see later (Table 4.2), Ocr(W94C) exhibits antirestriction activity *in vivo* and effectively inhibits *Eco*KI methylase *in vivo* as well as in *in vitro*.

In conclusion, Trp94 seems to have some importance for protein stability, but the antirestriction and the antimethylation activity are almost as good as for the wild type Ocr.

4.3 Ocr antirestriction activity

It is well known that wild type Ocr is able to inhibit *Eco*KI methylase and nuclease both *in vivo* and *in vitro* (Studier, 1975; Bandyopadhyay *et al.*, 1985). The following assay was designed to find out if the Ocr(Cys) and the two Ocr proteins with the C-terminal region deleted are able to knock out the *Eco*KI restriction activity.

BL21(DE3)pLysS cells were transformed with pET1 derivative plasmids con-

taining the *ocr* gene. We chose this strain because the same strain was used for Ocr overexpression and we know from previous assays that even in the absence of IPTG a low amount of Ocr is produced which should be enough for the antirestriction assay.

BL21(DE3)pLysS cells are are deficient in both restriction and methylation ($r^- m^-$). In order to check the antirestriction activity of the Ocr proteins, a conjugation between NK308 and BL21 cells was carried out. NK308 strain contains F'101 plasmid. By conjugation, F'101 plasmid (which contains *hsdR,hsdM* and *hsdS* genes) is transfered to BL21 cells.

Lambda DNA is susceptible to restriction by *Eco*KI nuclease if the recognition sites for *Eco*KI are unmethylated. Plaques appeared when unmodified lambda virus (λ v.0) was added to the transconjugants which suggest that the *Eco*KI nuclease was inhibited by Ocr. Indeed, in a control experiment where λ v.0 was added to a conjugant between NK308 cells and untransformed BL21(DE3)pLysS cells, restriction occurred.

As seen in table 4.2, all Ocr proteins tested were able to knock out the *Eco*KI nuclease. However the e.o.p. was lower than 1, probably because the amount of the expressed Ocr was not high enough to inhibit all the nuclease present. The result is particularly interesting for Ocr(W94C), Ocr 3786 and Ocr 3790 proteins. These results showed that the the unique Trp in position 94 is not essential for the antirestriction activity and also confirmed an earlier result that the C-terminal region does not play any role in the inhibition of *Eco*KI nuclease (Dunn *et al.*, 1981).

4.4 Ocr antimethylation activity

4.4.1 *In vivo* assay

The antirestriction activity of the Ocr proteins *in vivo* demonstrated that the the *Eco*KI nuclease was unable to restrict an unmethylated λ DNA when Ocr

Table 4.2: Antirestriction activity of the Ocr proteins

Donor strain	Recipient strain	e.o.p.
NK308	BL21 + 2993	0.26
NK308	BL21 + 3786	0.32
NK308	BL21 + 3790	0.3
NK308	BL21 + W94C	0.32
NK308	BL21 + N43C	0.32
NK308	BL21 + S68C	0.38
NK308	BL21 + N4C	0.26
NK308	BL21 + D25C	0.3
NK308	BL21 + D62C	0.16
NK308	BL21	$0.8 \cdot 10^{-3}$
NK308	-	$0.6 \cdot 10^{-3}$
-	BL21	1

proteins were present. However, the experiment gave no information about the inhibition of the *Eco*KI methylase. If methylase was still active in the presence of the Ocr proteins the unrestricted λ DNA should be methylated at the recognition site for *Eco*KI. The methylated λ DNA (λ v.k.) would survive on a r^+ strain. However if the methylase was also inhibited, the λ DNA would be unmethylated and susceptible to restriction on a r^+ strain.

λ DNA from the antirestriction experiment was added to NK301 which is r^+m^+ and also on a 5k strain (as a control) which is r^-m^+. As seen in table 4.3, λ DNA was restricted on a r^+m^+ strain confirming that *Eco*KI methylase was also inhibited by Ocr proteins.

4.4.2 *In vitro* assay

There is a very elegant way to prove that methylase is active *in vitro*. Plasmid pBlue$Kan^R Hpa^+$ has a *Hpa*I recognition site GTTAAC. The AAC sequence is also part of the *Eco*KI methylase recognition site. If the methylase is active, the second adenine from the AAC sequence is methylated. When methylation occurs the *Hpa*I site becomes refractory to *Hpa*I nuclease.

	Marker															
Ocr	–	–	–	–	–	–	–	2993	–	–	2993	2993	2993	3786	3786	3786
M_2S_1	–	–	–	–	–	+	+	+	+	+	+	+	+	+	+	+
Ocr:M_2S_1								4:1			0.4:1	1.2:1	4:1	0.4:1	1.2:1	4:1
Uncut DNA	–	+	–	+	–	+	+	–	–	–	–	–	–	–	–	–
Linearized DNA	–	–	+	–	+	–	–	+	+	+	+	+	+	+	+	+
AdoMet	–	–	–	–	–	–	+	–	–	+	+	+	+	+	+	+
*Hpa*I	–	–	–	+	+	+	+	+	+	+	+	+	+	+	+	+

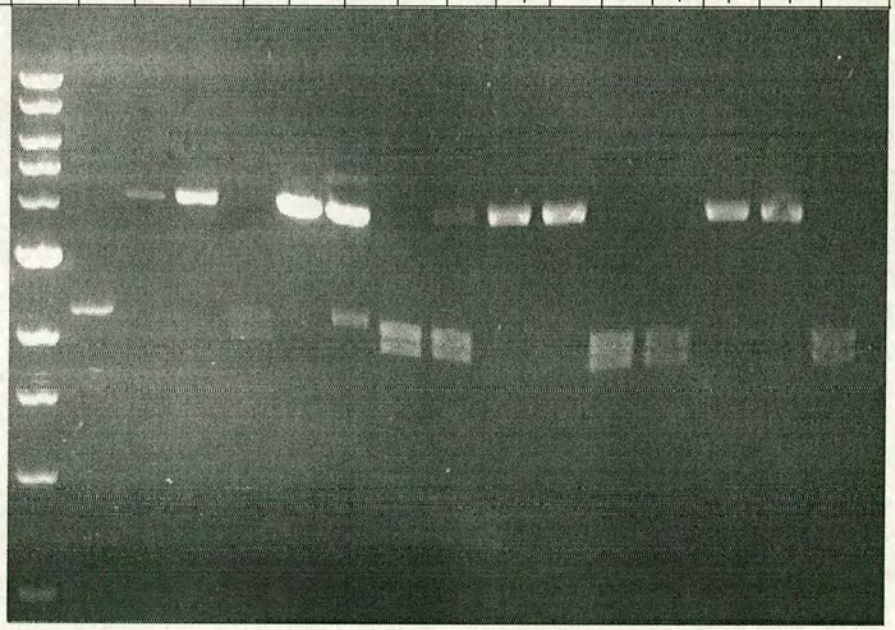


Figure 4.28: Inhibition of *Eco*KI methylase by Ocr proteins. Plasmids were linearized with *Eco*RI. The linearized DNA was incubated with methylase for 6 hours. If the methylase is active, the methylated DNA becomes resistant to *Hpa*I nuclease and only one band is observed. When methylase is inactivated by the Ocr proteins, the *Hpa*I cut the DNA and two adjacent bands are observed.

	Marker														
Ocr		3790	3790	3790	W94C	W94C	W94C	N43C	N43C	N43C	S68C	S68C	S68C	N4C	N4C	N4C
M_2S_1		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Ocr:M_2S_1		0.4:1	1.2:1	4:1	0.4:1	1.2:1	4:1	0.4:1	1.2:1	4:1	0.4:1	1.2:1	4:1	0.4:1	1.2:1	4:1
Linearized DNA		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
AdoMet		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
*Hpa*I		+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

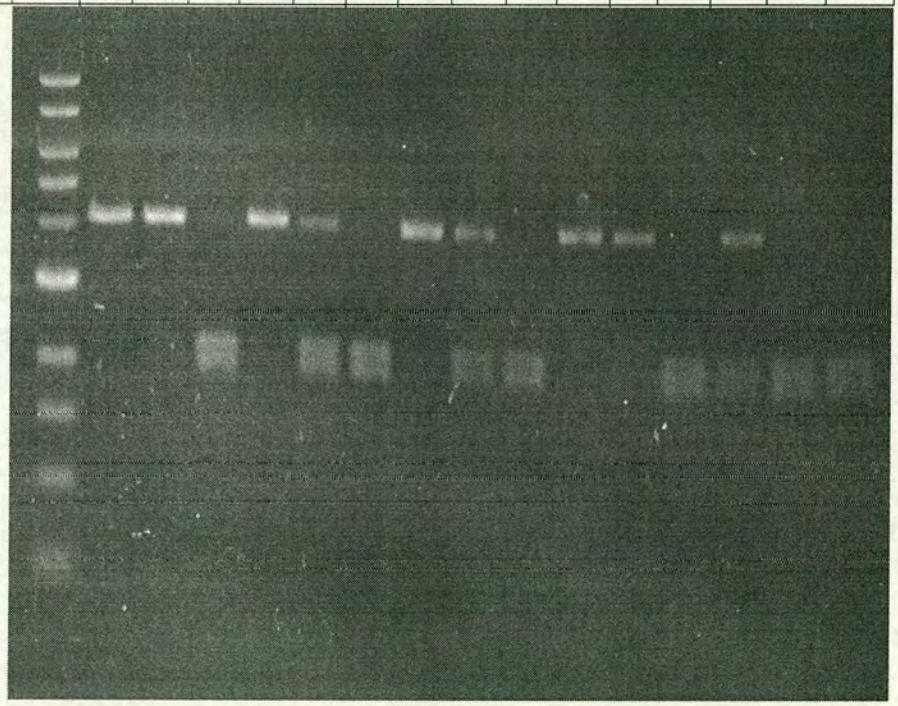


Figure 4.29: Inhibition of *Eco*KI methylase by Ocr(Cys) proteins. The assay was performed in exactly the same conditions as those described in figure 4.28.

	Marker						
Ocr		D25C	D25C	D25C	D62C	D62C	D62C
M_2S_1		+	+	+	+	+	+
linearized DNA		+	+	+	+	+	+
Ocr:M_2S_1		0.4:1	1.2:1	4:1	0.4:1	1.2:1	4:1
AdoMet		+	+	+	+	+	+
*Hpa*I		+	+	+	+	+	+

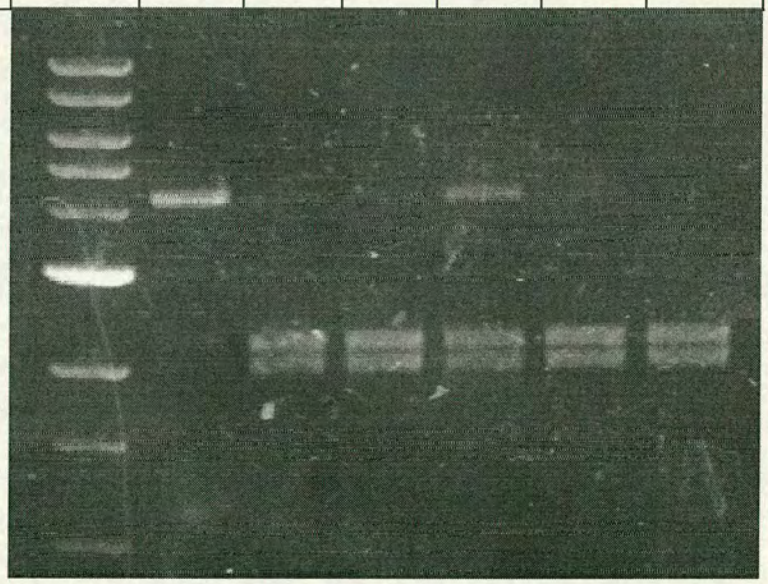


Figure 4.30: Inhibition of *Eco*KI methylase by Ocr(Cys) proteins. The assay was performed in exactly the same conditions as those described in figure 4.28.

Table 4.3: Antimethylation activity of the Ocr proteins

Source of λ	e.o.p.
NK308+BL21+2993	$< 8{\cdot}10^{-3}$
NK308+BL21+3786	$< 10^{-3}$
NK308+BL21+3790	$< 7{\cdot}10^{-3}$
NK308+BL21+W94C	$< 9{\cdot}10^{-3}$
NK308+Bl21+N43C	$< 7{\cdot}10^{-3}$
NK308+BL21+S68C	$< 0.6{\cdot}10^{-3}$
NK308+BL21+N4C	$< 3{\cdot}10^{-3}$
NK308+BL21+D25C	$< 10^{-2}$
NK308+BL21+D62C	$< 3{\cdot}10^{-3}$
BL21	$< 5{\cdot}10^{-4}$
NK308	1.5
NK308+BL21	0.6

The pBlue$Kan^R Hpa^+$ plasmid is first linearized with *Eco*RI enzyme. When methylase is active, *Hpa*I is unable to cut the linearized DNA and consequently a single DNA band can be noticed on an agarose gel. If methylase is inactive, the GTTAAC site remains unmethylated and is sensitive to *Hpa*I nuclease. In this case two DNA bands will appear on an agarose gel.

Methylase in combination with all Ocr proteins at different ratios was incubated with linearized pBlue$Kan^R Hpa^+$ plasmid for 6 hours which is long enough for methylation to be completed. Then *Hpa*I nuclease was added and the effect against the plasmid DNA was monitored on an agarose gel.

As expected, part of the methylase was still active when incubated with Ocr proteins at a molar ratio Ocr:methylase 0.4:1. At a molar ratio 1.2:1 which is closed to the stoichiometry of the complex, most of the methylase was inhibited. At 4.0:1 all the methylase was knocked out (Figures 4.28, 4.29, 4.30).

This result is in agreement with the *in vivo* experiment confirming that all Ocr(Cys) proteins and also the truncated forms Ocr 3786 and Ocr 3790 preserved the antimethylation activity of the wild type Ocr.

4.5 Ocr interacts with *Eco*KI R, M and S subunits

Having established the stoichiometry of binding, the next question to be answered was which *Eco*KI subunit interacts with Ocr?

As a potential nucleotide mimic it was expected that Ocr could compete with target DNA for the binding site of S subunit. The S subunit is insoluble so that it was not possible to perform a direct binding experiment and indirect evidence was sought. A solution containing 1 μM of 21 bp oligonucleotide which has the recognition site for *Eco*KI was titrated with *Eco*KI methylase until no oligonucleotide was free in solution. A hexachlorofluorescein molecule was attached to the top strand of the 21 bp oligonucleotide at the 5' end. The sample was excited at a wavelength specific for the hexachlorofluorescein (Hex) fluorophore (530 nm) and emission was recorded at 570 nm. The binding of methylase to the labelled DNA was monitored as an increase in the fluorophore anisotropy after each addition of methylase. As expected, at a 1:1 molar ratio, the anisotropy did not increase any further. When Ocr was titrated the anisotropy decreased step by step to the initial value which is an indication that Ocr displaced the labelled DNA attached to the S subunit of methylase (Figure 4.31). One Ocr dimer seems to be enough to displace all the labelled DNA.

Ocr was also able to displace the DNA-Hex bound to the recognition site of the *Eco*KI nuclease. As seen in figure 4.36 only half of the DNA-Hex was displaced, the other half remaining bound to the recognition site even in the presence of a huge excess of Ocr. When ATP was added the anisotropy suddenly dropped, indicating that all DNA-Hex was displaced. It is very difficult to explain this result. It is possible that our *Eco*KI nuclease stock is a mixture of two *Eco*KI isoforms, one having the binding site for Ocr exposed and the other hidden. Or maybe nuclease binds two pieces of DNA, one to each Ocr binding site, and only one is displaced by Ocr, the other being displaced only in the presence of ATP.

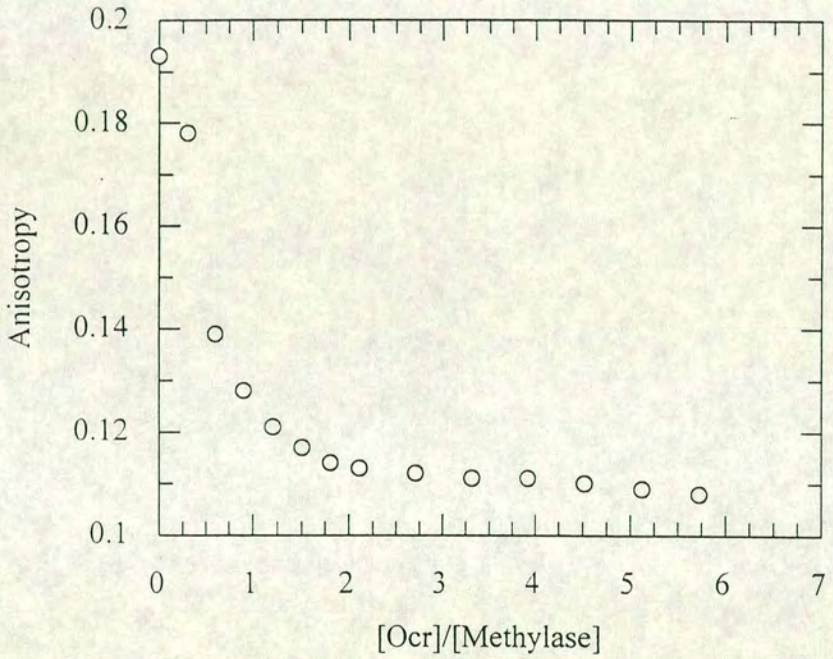


Figure 4.31: Displacement of 1 μM hexachlorofluorescein-labelled 21 bp dsDNA bound to the target site of *Eco*KI methylase with Ocr. The decrease in the fluorescence anisotropy of the labelled fluorophore was observed at 570 nm after an excitation at 530 nm. For each titration point the anisotropy was measured for 10 minutes, a mean of these values was calculated and plotted *vs.* the molar ratio Ocr/methylase. The experiment was performed at 25 °C and the slitwidth were 15 and 20 nm for excitation and emission respectively.

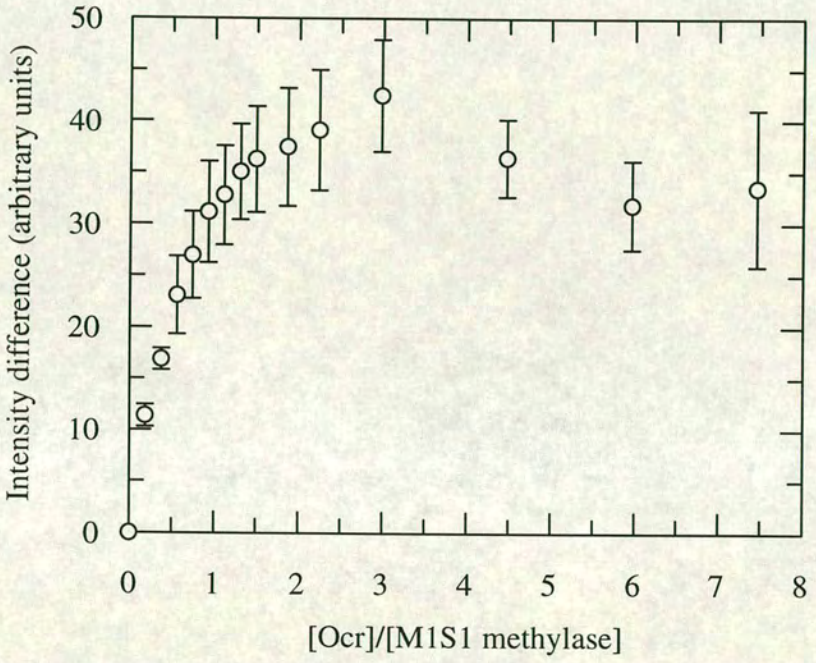


Figure 4.32: Titration of M1S1 methylase with Ocr. The assay was performed in the same conditions as those mentioned in figure 4.1.

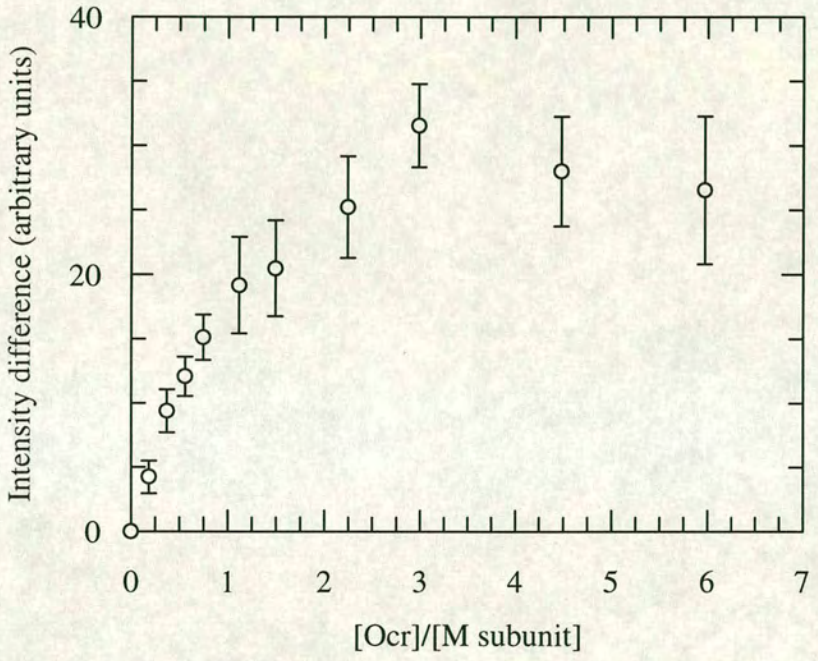


Figure 4.33: Interaction of Ocr with the M subunit. The assay was performed in the same conditions as those mentioned in figure 4.1.

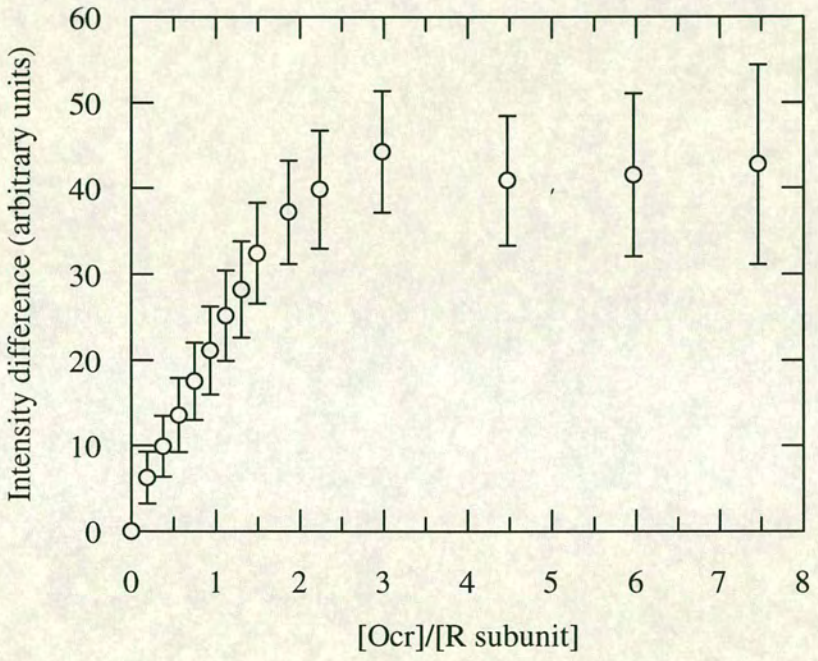


Figure 4.34: Interaction of Ocr with R subunit. The assay was performed in the same conditions as those mentioned in figure 4.1.

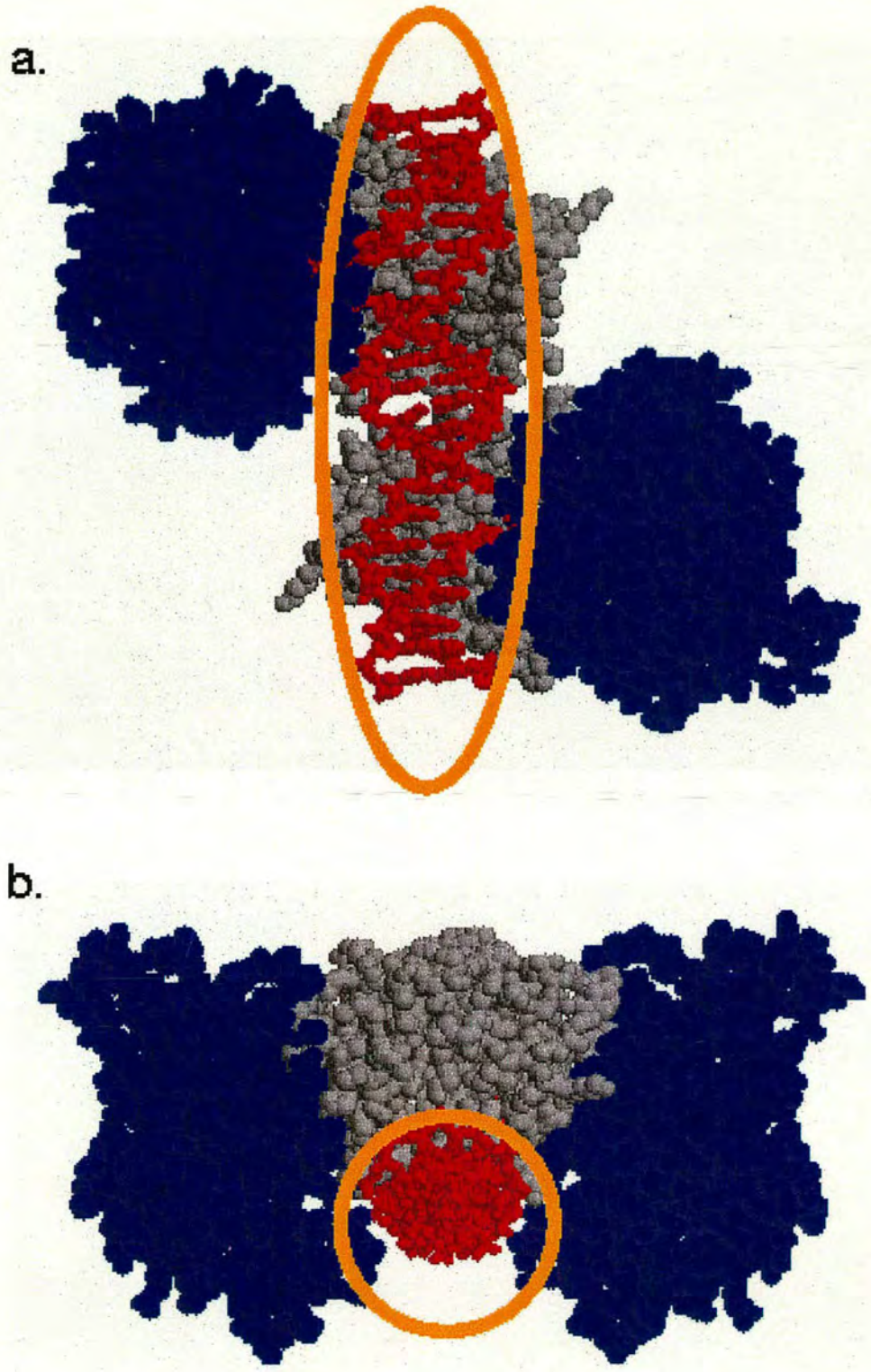


Figure 4.35: Model of interaction between Ocr and *Eco*KI methylase. Ocr (shown as an yellow ellipsoid) is superimposed on top of the structural model of *Eco*KI methylase. This model contains partial structures of two M subunits (blue spacefilling) and one S subunit (grey spacefilling) bound to 24 base pairs of DNA (red wireframe). a) front view, b) top view.

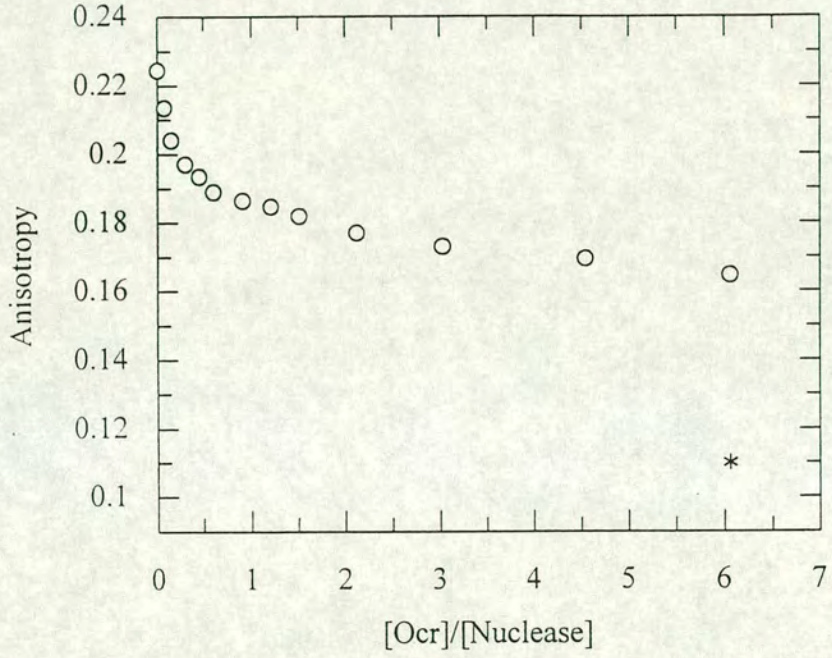


Figure 4.36: Displacement of 1 μM hexachlorofluorescein-labelled 21 bp dsDNA bound to the target site of *Eco*KI nuclease with Ocr. The decrease in the fluorescence anisotropy of the labelled fluorophore was observed at 570 nm after excitation at 530 nm. For each titration point the anisotropy was measured for 10 minutes, a mean of these values was calculated and plotted *vs.* the molar ratio Ocr/nuclease. Only half of the DNA-Hex was displaced by Ocr. When 2 mM ATP was added the other half was suddenly displaced ($*$).

These are only speculations since there is no evidence for the existence of two nuclease isoforms or for the fact that more than one piece of DNA is bound at the nuclease recognition site. It is worth noting that the inability of nuclease to expose binding sites for Ocr was observed only when nuclease was allowed to form complexes with the specific DNA. The free nuclease is able to bind two Ocr dimers, as determined in sections 4.1.1 and 4.1.2.

The binding sites for Ocr on both *Eco*KI methylase and nuclease probably overlap the binding sites for DNA. The possibility that Ocr binds at sites remote from the two TRDs and conformational changes appeared upon binding decrease the affinity for DNA is unlikely since *Eco*KI binds either Ocr or DNA, but not both at the same time (Bandyopadhyay *et al.*, 1985).

Only the active methylase (M_2S_1) is capable of tight binding with Ocr as determined from a Trp fluorescence experiment Figure 4.1. A weak binding was noticed when an inactive methylase (M_1S_1) was titrated with Ocr (Figure 4.32). Ocr has also a weak affinity for the M subunit alone (Figure 4.33). Probably Ocr binds to the S subunit but also has some contacts with M subunit.

In the case of *Eco*KI nuclease, not only the fully active form ($R_2M_2S_1$) but also the R subunit alone binds tightly to Ocr (Figure 4.34). Probably Ocr and the R subunit has a reasonable binding interface in order to get such a strong binding. If each R subunit bind also one Ocr dimer (as shown from the interaction between Ocr and R subunit alone) the stoichiometry of the complex should be 3:1 not 2:1. The only way to fit together all these results is to consider that Ocr binds nuclease in a different way than methylase. S subunit has two target recognition domains (TRDs) each one responsible for the recognition of the specific DNA. In the case of Ocr-methylase complex it is possible that one Ocr monomer binds to one TRD and the other to the second TRD (Figure 4.35). For Ocr-nuclease complex it seem likely that the two TRDs are occupied by two monomers belonging to two different Ocr molecules. The remaining monomer in each Ocr molecule could bind

the two adjacent R subunits.

As noticed in figure 4.36 it seems that one Ocr dimer is enough to displace the labelled DNA. This result does not come in contradiction with our model because even only one TRD is blocked by Ocr the nuclease will not be able to bind the target DNA.

4.6 Ocr-*Eco*KI dissociation constant

The dissociation constant (K_d) of *Eco*KI methylase for binding to 21 bp Hex-DNA (in 50 mM NaCl) is 9.43 ± 14.05 nM (Powell *et al.*, 1998a). For *Eco*KI nuclease-DNA complex the apparent K_d determined by gel retardation was estimated to be between 2 and 5 nm (Powell *et al.*, 1998b). Since Ocr displaces the DNA bound to the target site of the *Eco*KI enzymes, a K_d in the range of picomolar or femtomolar for Ocr-*Eco*KI methylase and nuclease complexes was expected.

Obtaining a reliable K_d value for tight binding systems such as *Eco*KI-Ocr is not an easy task because the interaction must be measured at a very low protein concentration, in the range of the expected K_d. Fluorescence is a very sensitive technique that allows collection of data at very low protein concentrations and consequently was the method of choice.

As determined from previous assays (sections 4.2.1 and 4.2.2), adding methylase or nuclease to Cys-labelled Ocr proteins changes the fluorescence intensity of the labelled fluorophore. The protein chosen for labelling was Ocr(N43C) due to its identical behavior with the wild type Ocr. The fluorophore used for labelling was IAF because of its high molar extinction coefficient (71000) which is higher than those of PM (40000) or I-AEDANS (6100). Unlike PM and I-AEDANS whose fluorescence increase in the presence of *Eco*KI, IAF fluorescence is quenched when *Eco*KI is added.

Table 4.4: Single site binding model

Model and parameters	Reactions and equations
Single site	$M + L \overset{K}{\rightleftharpoons} ML$
Partition function	$Q = 1 + K[L]$
Mole fractions	$X_M = 1/Q; X_{ML} = K[L]/Q$
Conservation of mass	$[M]_0 = [M] + [ML]; [M] = [M]_0 X_M$ $[L]_0 = [L] + [ML]; [L] = [L]_0 - [M]_0 X_{ML}$
Free ligand	$[L]^2 + [L]([M]_0 - [L]_0 + K^{-1}) - [L]_0/K = 0$

where

$[L]_0$	total ligand concentration
[L]	free ligand concentration
$[M]_0$	total macromolecule concentration
[M]	free macromolecule concentration
K	association constant

When studying the dissociation constant of a ligand-protein system, usually the ligand is titrated into a fixed concentration of protein. For a tight binding systems either of the two proteins can be regarded as the ligand. In our case it was convenient to consider *Eco*KI as ligand and titrate it into a fixed concentration of Ocr(N43C)-AF protein and monitor the quenching of fluorescein fluorescence.

Table 4.4 shows the equations that relate the free ligand concentration to the total ligand concentration and total macromolecule concentration in the case of a one site binding model (Eftink, 1997).

The observed fluorescence intensity at each titration point comes from both free Ocr(N43C)-AF and labelled protein complexed with *Eco*KI. Consequently, the observed fluorescence is:

$$F = X_M F_M + X_{ML} F_{ML} \tag{4.6}$$

where

F_M relative intensity of free form
F_{ML} relative intensity of complexed form

The most direct way to relate fluorescence intensity to a binding model is to combine equation 4.6 with a binding model and then to perform a nonlinear least-squares regression analysis (Eftink, 1997). For a single binding site model equation 4.6 becomes:

$$F = \frac{F_M + F_{ML}K[L]}{1 + K[L]} \tag{4.7}$$

Fitting the theoretical model to the data was done by computer, using the Grafit program. For Ocr-methylase complex, a K_d value of $4.8 \cdot 10^{-11}$ M was obtained (Figure 4.37). The K_d of Ocr-nuclease complex had a similar value, $7.7 \cdot 10^{-11}$ M (Figure 4.38).

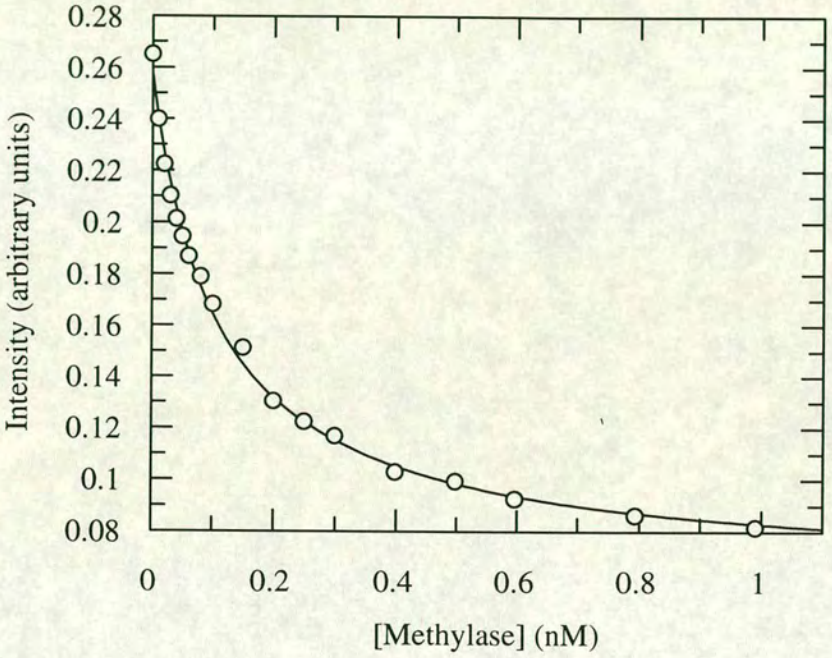


Figure 4.37: The dissociation constant of the Ocr-methylase complex. Increasing concentrations of methylase were titrated into a constant concentration of 100 pM Ocr(N43C)-AF protein. The sample was excited at 492 nm and the emission was recorded between 510 and 570 nm. The area of the spectrum between 510 and 570 nm was integrated. The integrated area of the buffer was subtracted from the spectra of the samples. Spectra were corrected for any change in the volume due to addition of methylase. The corrected data were fitted to a one-binding site theoretical curve using Grafit software and the values plotted *vs.* the concentration of methylase. The assay was performed at 25 °C using a slit width of 5.4 nm for both excitation and emission. The buffer was 20 mM Tris, 20 mM NH_4Cl, 6 mM $MgCl_2$ 7 mM β-mercaptoethanol pH 8.

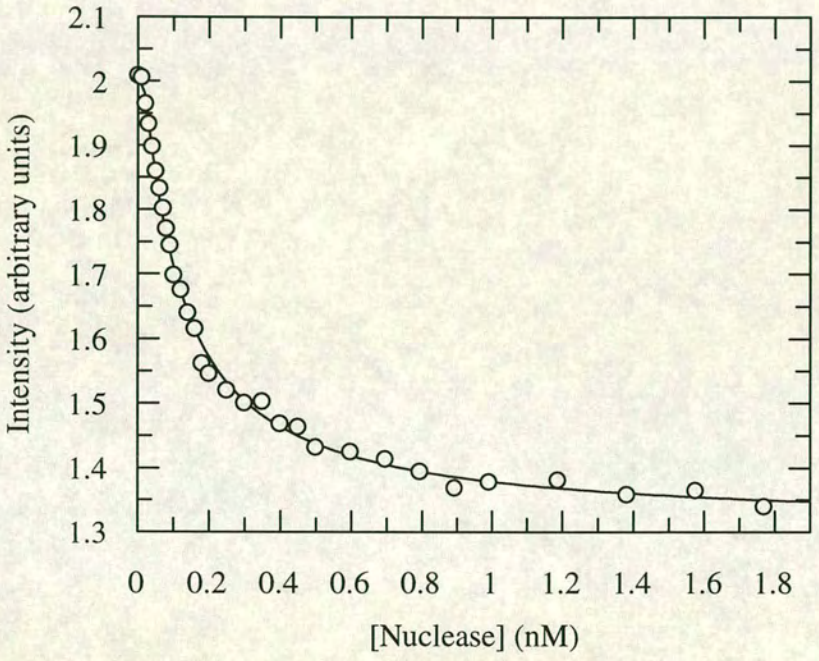


Figure 4.38: The dissociation constant of the Ocr-nuclease complex. The experiment was performed in exactly the same conditions as those described in figure 4.37.

Chapter 5

Discussion

There has been considerable genetic analysis of the interaction between Ocr and *Eco*KI enzymes but very little is known about the biochemistry of Ocr and its interaction with *Eco*KI methylase and nuclease. The aim of this work was to fill this gap, to provide biochemical information about the Ocr protein, its interaction with *Eco*KI enzyme and, in the end, to propose a model of interaction.

One of the most challenging parts of this project was to identify those Ocr amino acids located at the binding interface with *Eco*KI methylase. For this purpose, a technique developed by Latham *et al.* (1997) for studying the interaction between gp45 and gp44/62, two proteins that interact at the DNA replication fork of bacteriophage T4 was used.

Six residues of wild type Ocr, which are predicted to be exposed on the protein surface, were chosen for substitution with cysteine. The newly obtained Ocr(Cys) proteins were labelled with 1,5 IAEDANS or Pyrene Maleimide, two cysteine-specific fluorophores. Labelling was successfully done, indicating that all six amino acids are exposed on the surface. Both fluorophores are very sensitive to any change in their local environment (Latham *et al.*, 1997). Any increase in the hydrophobicity around these fluorophores leads to an increase in their fluorescence intensity and a shift of the maximum of the emission spectrum to lower wavelength (Latham *et al.*, 1997; Lakowicz, 1999).

An increase in the fluorescence intensity and a blue shift in the emission

maximum were observed when any of the six Ocr(Cys)-AEDANS or Ocr(Cys)-PM proteins were titrated with methylase (except Ocr(D62C)-PM which showed a decrease in pyrene intensity when methylase was added). The most probable explanation for this increase in fluorescence intensity is that all six chosen amino acids are located at or near the Ocr binding site for methylase. Upon binding, the fluorophores become protected from the solvent molecules which are known to lower the fluorescence intensity (Lakowicz, 1999). Since the amino acids chosen for labelling were located in the Ocr amino acid sequence in positions 4, 25, 43, 62, 68 and 94, which covers almost the entire protein surface except the C-terminal region, a few of these amino acids would be expected to be at the binding site and the rest not. Surprisingly, all of them are located at or near the binding site which suggests a large Ocr surface area buried at the Ocr-methylase interface.

However, the high sensitivity of the fluorophores to changes in their local environment could lead to misinterpretations. For example, it is possible that one of the labelled amino acids is not actually in the binding site but dramatic conformational changes appearing in Ocr after binding lead to an increase in the hydrophobicity of the fluorophores environment which in turn causes an increase in the fluorescence intensity. Although unlikely, the last possibility was not ruled out.

An acrylamide quenching experiment was designed to discriminate between these two possible explanations for the increase of the fluorescence intensity upon binding. Acrylamide is a polar, non-charged electron deficient molecule that efficiently quenches 1,5 IAEDANS when in direct contact with it (Wasylewski *et al.*, 1988). The fluorescence of the AEDANS moieties bound to all six exposed amino acids was quenched by acrylamide. However, when the Ocr(Cys)-AEDANS proteins were allowed to form a 1:1 complex with methylase, no quenching was observed. This assay clearly proved that acrylamide had no direct access to the AEDANS moieties attached to any of the six amino acids. A slight quenching

was noticed for D62C-AEDANS mutant bound to methylase, probably because the D62 residue is located at the edge of the binding site.

The amino acid substitutions introduced were not intended to disturb the structure or the activity of Ocr. At least for the conservative mutation Ser68 to Cys, the modified protein was expected to behave like the wild type Ocr. The main concern was about the Ocr(W94C) protein since Trp94 seems to be involved in protein activity or stability (Dunn *et al.*, 1981).

Both *in vivo* and *in vitro* assays showed that all Ocr(Cys) proteins efficiently inactivate *Eco*KI methylase. *In vivo* assays also demonstrated that all six proteins are still able to inhibit the nuclease activity so that there is no doubt that all six Ocr(Cys) proteins are still active.

The effect of mutations on the structure of Ocr was analyzed by differential scanning calorimetry. The DSC scan of the wild type Ocr nicely fits to the dissociation model but not to a two-state transition model. This implies that Ocr dimer first dissociates to two monomers then thermal unfolding occurs. Wild type Ocr is a stable protein with a T_m of 72.2 °C.

All six Ocr(Cys) proteins were also fitted to the dissociation model and the T_m was calculated. As expected, the melting temperature of the S68C mutant and also N4C and N43C were very close to the T_m value obtained for the wild type Ocr, suggesting that the newly introduced mutations did not severely affect the structure of Ocr. Ocr(D62C) protein showed a T_m of 70.6 °C which is 1.6 °C lower than that of the wild type Ocr. In theory, any change in T_m bigger than $\pm$ 0.5 °C reflects a change in the protein stability. However, the decrease in T_m value for D62C mutant could be due to the poor fitting to the dissociation model rather than to a real instability introduced by mutation. Substitution of Asp25 residue with Cys increases the stability of the protein since the T_m was 5.7 °C higher than that obtained for the wild type.

As previously shown, substitution of Trp94 for Cys did not impair the ability

to inhibit both methylase and nuclease, suggesting that Trp94 is not important for Ocr activity. However, the T_m was 3.5 °C lower for this mutant than that of wild type, indicating that Trp94 is important for protein stability.

Since Ocr(W94C) was successfully labelled with 1,5 IAEDANS and PM, Trp94 residue must be exposed on the protein surface. The same conclusion can be drawn from the tryptophan emission spectra which has a maximum at 350 nm (Figure 3.15), a position characteristic for proteins having solvent exposed tryptophan.

According to the work done by Dunn *et al.* (1981) the last 22 amino acids at the C-terminus are not important for Ocr activity *in vivo*. Two truncated proteins, Ocr 3786 and Ocr 3790 were purified. They have deleted the last 7 or 17 amino acids respectively but still retain the ability to prevent the restriction and modification activities *in vivo* as well as *in vitro*. Deletion of the last 7 amino acids did not decrease too much the protein stability; Ocr 3786 has a T_m of 71.1 °C compared to 72.2 °C for wild type Ocr. However, deletion of the last 17 amino acids at the C-terminal end severely decreased the stability; the melting temperature for Ocr 3790 is only 58.4 °C. These results suggest that the C-terminal end is important for Ocr stability but not activity.

Since Ocr is a dimer, each Ocr(Cys) protein contains the same mutation in each monomer. The ability of Ocr(N4C) to be labelled in the presence but not in the absence of the thiol reducing reagents raised the possibility that the two N4C residues are at the monomer-monomer interface and form a disulphide bond.

Pyrene Maleimide is often used to probe the close proximity of labelled amino acids (Graceffa and Lehrer, 1980; Lehrer, 1997). When two pyrene moieties are close to each other they form an excimer characterized by a fluorescence emission maximum at about 470 nm. The emission spectra of Ocr(N4C)-PM displays such an excimer with a maximum at 480 nm, indicating that the two N4C residues are in close proximity. Excimer formation was also used to probe the close proximity

of S68C residues. In this case, the fluorescence emission spectra was shifted to 460 nm, probably because S68 residues lie in a more hydrophobic environment than those of N4C.

Additional information came from a gel filtration assay done in denaturing conditions (5 M GdnHCl), both in the presence and in the absence of thiol reducing reagents. Ocr(N4C) appeared only as monomer when β-mercaptoethanol was present. However, two protein peaks appeared in the absence of β-mercaptoethanol, one corresponding to Ocr(N4C) dimer and one to monomeric form, clearly indicating that the two N4C residues form a disulphide bond. The S68C mutant appeared as monomer both in the presence and in the absence of β-mercaptoethanol. No disulphide bond is formed between the S68C residues despite their close proximity.

Accurate determination of protein concentration is essential in the quantitative study of protein-protein interactions, such as determination of Ocr-*Eco*KI stoichiometry. The most accurate method to determine protein concentration is UV-spectroscopy, which requires knowledge of the extinction coefficient. The molar extinction coefficient of *Eco*KI methylase and nuclease is known (Dryden '97), but not that of Ocr. The molar extinction coefficient of wild type Ocr was calculated using a method developed by Gill and von Hippel (1989). The value obtained (32,095 M^{-1} cm^{-1}) is likely to be accurate since it is only 1 % higher than the theoretical value calculated from the amino acid sequence (31,860 M^{-1} cm^{-1}). In practice, a difference higher than 5 % between these two values is an indication that the measured molar extinction coefficient is wrong (Gill and von Hippel, 1989).

According to the ITC experiment, the stoichiometry of Ocr-methylase complex is 0.95 $\pm$ 0.005. The stoichiometry obtained from the interaction between Ocr(Cys)-labelled proteins and methylase indicated a value of 1.2 $\pm$ 0.03. For this experiment we relied mostly on the values obtained for the S68C, N4C and

N43C mutants, since they are structurally and functionally very close to the wild type Ocr. The most reproducible results were obtained when using N43C mutant labelled with either IAEDANS or PM. That was the reason for using this mutant to determine the K_d of the Ocr-*Eco*KI complex. The D25C mutant showed the same stoichiometry of binding, despite its increased stability induced by the mutation. Ocr(W94C) was not considered at all due to its less stable structure which led to non-reproducible results. In addition, it was difficult to get an accurate molar extinction coefficient for this mutant since Gill's method is suitable only for proteins that have at least one Trp residue in their structure (Pace and Schmid, 1997). Ocr(N4C)-PM and Ocr(S68C)-PM were also omitted due to complications introduced by the presence of the excimer.

Ocr(D25C)-PM protein complexed with nuclease showed a stoichiometry of 2.1 ± 0.1. No ITC experiment was done for Ocr-nuclease complex. However, the extrinsic fluorescence experiment seems to be accurate enough to provide a good estimation of the stoichiometry.

Less sensitive methods like intrinsic Trp fluorescence, seems to overestimate the stoichiometry of binding. This method indicates a stoichiometry of 1.78 ± 0.02 for Ocr-methylase and 2.3 ± 0.07 for Ocr-nuclease complex.

The large excess of acidic residues raises the possibility that Ocr could act as a polyanion and compete for the DNA-binding site of *Eco*KI enzymes (Dunn *et al.*, 1981). The idea is sustained by the finding that *Eco*KI enzymes bind either DNA or Ocr but not both at the same time (Bandyopadhyay *et al.*, 1985), suggesting that Ocr and DNA compete for the same binding site represented by the two TRDs of the S subunit. Unfortunately the S subunit is insoluble so that a direct interaction experiment between this subunit and Ocr cannot be designed. Indirect evidence came from an anisotropy experiment in which methylase was allowed to form a 1:1 complex with a 21 bp unmethylated, specific oligo-DNA, labelled with hexachlorofluorescein. The decrease in the fluorophore anisotropy

suggest that the added Ocr has displaced the labelled DNA from its binding site.

More difficult to interpret was the displacement of the labelled DNA from a nuclease-DNA complex. As seen in figure 4.36, only half of the DNA is displaced by Ocr. The other half remained bound to the nuclease even in the presence of a large excess of Ocr. However, when ATP was added the anisotropy immediately decreased, indicating that all DNA was displaced by Ocr.

In a similar experiment, Bandyopadhyay *et al.* (1985) have shown that interaction between *Eco*KI nuclease and unmethylated, specific DNA results in a complex in which the binding site for Ocr is hidden. When ATP was added, the binding site for Ocr became available again. If this is true how can my result where a fraction of the DNA is displaced in the absence of ATP be explained? Maybe there are two nuclease isoforms, one having the binding site for Ocr exposed and one not when in complex with DNA. Or maybe *Eco*KI binds two pieces of DNA, one to each Ocr binding site and only one is directly displaced by Ocr and the second is displaced only in the presence of ATP. It was also noticed that one Ocr dimer was enough to displace the DNA despite the fact that nuclease has two binding sites for Ocr. However, there is no evidence so far that *Eco*KI binds more than one piece of DNA. Definitely more work should be done in the future to clarify this matter.

The above data together with the observation that Ocr strongly interacts with the R subunit but weakly with M subunit leads to the following model of binding (Figure 4.35). One Ocr dimer binds to *Eco*KI methylase, each monomer occupying one TRD of the S subunit. The inability of Ocr to bind to a nuclease which already has a DNA bound to its recognition site suggests that the only binding site for Ocr is located within the S subunit. Since nuclease has two binding sites for Ocr, they must be adjacent. Unlike methylase, the two TRDs are occupied by monomers that belong to two adjacent Ocr molecules. The remaining monomer from each Ocr molecule binds tightly to the R subunit. Some additional

contacts are made with the M subunit. It is difficult to understand what is the biological significance of the second binding site for Ocr, since one Ocr dimer seems to be enough to hinder the binding of the DNA.

For tight binding complexes the K_d is usually determined from the ratio of the dissociation and association rate constants (Lee *et al.*, 1989; Longstaff *et al.*, 1990; Schreiber and Fersht, 1993; Wallis *et al.*, 1995b). However, for Ocr-*Eco*KI complex the K_d was directly estimated from a binding curve were the fluorescence intensity of the 5-iodoacetamido-fluorescein bound to Ocr(N43C) mutant was quenched by addition of methylase or nuclease. The direct estimation was possible because the experiment was performed at very low protein concentrations, in the range of the expected K_d. The K_d for Ocr-methylase is $4.8 \cdot 10^{-11}$ M, which corresponds to a free energy of binding of -14090 cal/mol at 25 °C. The K_d for Ocr-nuclease complex is in the same range, $7.7 \cdot 10^{-11}$ M. Ocr binds tightly to both methylase and nuclease, however not as tightly as barstar binds barnase or Immunity protein9 binds Colicin E9 (Figure 1.13), two well-known examples of nuclease-inhibitor complexes (Schreiber and Fersht, 1993; Wallis *et al.*, 1995b).

The enthalpy of binding (ΔH^0) for Ocr-methylase complex is -18,240 cal/mol according to the ITC assay, meanwhile the calculated value for the change in entropy (ΔS^0) is -13.93 cal/mol·K. The negative sign of ΔS^0 indicates that the decrease in entropy due to binding is not compensated by the removal of the highly ordered water molecules surrounding nonpolar or charged groups from the Ocr-methylase interface, a favorable entropic factor. Binding of Ocr to methylase is an enthalpically driven process which may suggest dominant contributions of van der Waals contacts and/or polar interactions (Fisher and Singh, 1995; Makhatadze and Privalov, 1995).

However, it is difficult to say only from the above data if electrostatic interactions are involved in binding, as suggested by Dunn *et al.* (1981). Future work where the K_d of Ocr-methylase interaction will be analysed at different salt

concentrations will show if the electrostatic interactions play any role in binding.

The Ocr surface area involved in binding seems to be huge, much higher than the 25 % of the accessible area that is buried by UGI when it interacts with human UDGase. It seems likely that methylase wraps around the Ocr dimer after binding since almost all the surface exposed Ocr amino acids are likely to be involved in binding. The crystal structure of Ocr which is currently under investigation will show if this is the case or if the amino acids chosen for mutagenesis are clustered, by chance, in the same area. Once the Ocr structure is obtained it will be possible to fit this structure to the previously published model of *Eco*KI methylase (Dryden *et al.*, 1995).

Bibliography

Araki H. and Ogawa H. (1981). The participation of T7 DNA-binding protein in T7 genetic recombination. *Virology*, **111**, 509–15.

Argos P. (1985). Evidence for a repeating domain in type I restriction enzymes. *EMBO J*, **4**, 1351–55.

Bandyopadhyay P.K., Studier F.W., Hamilton D.L. and Yuan R. (1985). Inhibition of the type I restriction-modification enzymes *Eco*B and *Eco*K by the gene 0.3 protein of bacteriophage T7. *J Mol Biol*, **182**, 567–78.

Barcus V.A. and Murray N.E. (1995). Barriers to recombination:restriction. *In Population Genetics of Bacteria, ed. S. Baumberg, J.P.W. Young, S.R. Saunders and E.M.H. Wellington, Society for General Microbiology, Symposium 52, Cambridge Univ. Press*, 31–58.

Barcus V.A., Titheradge A.J. and Murray N.E. (1995). The diversity of alleles at the hsd locus in natural populations of *Escherichia coli*. *Genetics*, **140**, 1187–97.

Bartel P.L., Roecklein J.A., SenGupta D. and Fields S. (1996). A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat Genet*, **12**, 72–7.

Bedford E., Tabor S. and Richardson C.C. (1997). The thioredoxin binding domain of bacteriophage T7 DNA polymerase confers processivity on *Escherichia coli* DNA polymerase I. *Proc Natl Acad Sci U S A*, **94**, 479–84.

Belogurov A.A., Delver E.P., Agafonova O.V., Belogurova N.G., Lee L.Y. and Kado C.I. (2000). Antirestriction protein Ard (Type C) encoded by IncW plasmid pSa has a high similarity to the "protein transport" domain of TraC1 primase of promiscuous plasmid RP4. *J Mol Biol*, **296**, 969–77.

Belogurov A.A., Delver E.P. and Rodzevich O.V. (1992). IncN plasmid pKM101 and IncI1 plasmid ColIb-P9 encode homologous antirestriction proteins in their leading regions. *J Bacteriol*, **174**, 5079–85.

Belogurov A.A., Delver E.P. and 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**, 4843–50.

Bennett S.E. and Mosbaugh D.W. (1992). Characterization of the Escherichia coli uracil-DNA glycosylase.inhibitor protein complex. *J Biol Chem*, **267**, 22512–21.

Bennett S.E., Schimerlik M.I. and Mosbaugh D.W. (1993). Kinetics of the uracil-DNA glycosylase/inhibitor protein association. Ung interaction with Ugi, nucleic acids, and uracil compounds. *J Biol Chem*, **268**, 26879–85.

Bickle T.A., Brack C. and Yuan R. (1978). ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli* K-12. *Proc Natl Acad Sci U S A*, **75**, 3099–103.

Bickle T.A. and Kruger D.H. (1993). Biology of DNA restriction. *Microbiol Rev*, **57**, 434–50.

Buckle A.M., Schreiber G. and Fersht A.R. (1994). Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-A resolution. *Biochemistry*, **33**, 8878–89.

Burckhardt J., Weisemann J., Hamilton D.L. and Yuan R. (1981a). Complexes formed between the restriction endonuclease *Eco*K and heteroduplex DNA. *J Mol Biol*, **153**, 425–40.

Burckhardt J., Weisemann J. and Yuan R. (1981b). Characterization of the DNA methylase activity of the restriction enzyme from *Escherichia coli* K. *J Biol Chem*, **256**, 4024–32.

Burova T.V., Choiset Y., Jankowski C.K. and Haertle T. (1999). Conformational stability and binding properties of porcine odorant binding protein. *Biochemistry*, **38**, 15043–51.

Chacko S., Silverton E., Kam-Morgan L., Smith-Gill S., Cohen G. and Davies D. (1995). Structure of an antibody-lysozyme complex unexpected effect of conservative mutation. *J Mol Biol*, **245**, 261–74.

Chamberlin M., McGrath J. and Waskell L. (1970). New RNA polymerase from *Escherichia coli* infected with bacteriophage T7. *Nature*, **228**, 227–31.

Chothia C. and Janin J. (1975). Principles of protein-protein recognition. *Nature*, **256**, 705–8.

Chung Y.B., Nardone C. and Hinkle D.C. (1990). Bacteriophage T7 DNA packing III. A "hairpin" end formed on T7 concatemers may be an intermediate in the processing reaction. *J Mol Biol*, **216**, 939–48.

Clackson T. and Wells J.A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science*, **267**, 383–6.

Clark P.L., Liu Z.P., Zhang L. and Gierasch L.M. (1996). Intrinsic tryptophans of CRABPI as probes of structure and folding. *Protein Sci*, **5**, 1108–17.

Cone R., Bonura T. and Friedberg E.C. (1980). Inhibitor of uracil-DNA glycosylase induced by bacteriophage PBS2. Purification and preliminary characterization. *J Biol Chem*, **255**, 10354–8.

Cooper A. (1998). Microcalorimetry of protein-protein interactions. *In Biocalorimetry. Applications of calorimetry in the biological sciences, John Wiley and sons Ltd.*, 103–11.

Cooper L.P. and Dryden D.T. (1994). The domains of a type I DNA methyltransferase. Interactions and role in recognition of DNA methylation. *J Mol Biol*, **236**, 1011–21.

Cowan G.M., Gann A.A. and Murray N.E. (1989). Conservation of complex DNA recognition domains between families of restriction enzymes. *Cell*, **56**, 103–9.

Cregg J.M., Nguyen A.H. and Ito J. (1980). DNA modification induced during infection of *Bacillus subtilis* by phage ϕ3T. *Gene*, **12**, 17–24.

Cunningham B.C. and Wells J.A. (1991). Rational design of receptor-specific variants of human growth hormone. *Proc Natl Acad Sci U S A*, **88**, 3407–11.

Davies G.P., Kemp P., Molineux I.J. and Murray N.E. (1999a). The DNA translocation and ATPase activities of restriction-deficient mutants of *Eco*KI. *J Mol Biol*, **292**, 787–96.

Davies G.P., Martin I., Sturrock S.S., Cronshaw A., Murray N.E. and Dryden D.T. (1999b). On the structure and operation of type I DNA restriction enzymes. *J Mol Biol*, **290**, 565–79.

Davies G.P., Powell L.M., Webb J.L., Cooper L.P. and Murray N.E. (1998). *Eco*KI with an amino acid substitution in any one of seven DEAD-box motifs has impaired ATPase and endonuclease activities. *Nucleic Acids Res*, **26**, 4828–36.

Davis R.W. and Hyman R.W. (1971). A study in evolution: the DNA base sequence homology between coliphages T7 and T3. *J Mol Biol*, **62**, 287–01.

Debrouwere L., Zabeau M., Van Montagu M. and Schell J. (1980). The *ral* gene of phage lambda. II. Isolation and characterization of *ral* deficient mutants. *Mol Gen Genet*, **179**, 75–80.

DeWyngaert M.A. and Hinkle D.C. (1980). Characterization of the defects in bacteriophage T7 DNA synthesis during growth in the Escherichia coli mutant tsnB. *J Virol*, **33**, 780–8.

Doherty A.J., Ashford S.R. and Wigley D.B. (1996). Characterization of proteolytic fragments of bacteriophage T7 DNA ligase. *Nucleic Acids Res*, **24**, 2281–7.

Doherty A.J. and Wigley D.B. (1999). Functional domains of an ATP-dependent DNA ligase. *J Mol Biol*, **285**, 63–71.

Dreier J., MacWilliams P. and Bickle T.A. (1996). DNA cleavage by the type IC restriction-modification enzyme *Eco*R124II. *J Mol Biol*, **264**, 722–33.

Dryden D.T., Cooper L.P. and Murray N.E. (1993). Purification and characterization of the methyltransferase from the type 1 restriction and modification system of *Escherichia coli* K12. *J Biol Chem*, **268**, 13228–36.

Dryden D.T., Cooper L.P., Thorpe P.H. and Byron O. (1997). The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry*, **36**, 1065–76.

Dryden D.T., Sturrock S.S. and Winter M. (1995). Structural modelling of a type I DNA methyltransferase [letter]. *Nat Struct Biol*, **2**, 632–5.

Dunn J.J., Elzinga M., Mark K.K. and Studier F.W. (1981). Amino acid sequence of the gene 0.3 protein of bacteriophage T7 and nucleotide sequence of its mRNA. *J Biol Chem*, **256**, 2579–85.

Dunn J.J. and Studier F.W. (1981). Nucleotide sequence from the genetic left end of bacteriophage T7 DNA to the beginning of gene 4. *J Mol Biol*, **148**, 303–30.

Dunn J.J. and Studier F.W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J Mol Biol*, **166**, 477–535.

Edelhoch H. (1967). Spectroscopic determination of tryptophan and tyrosine in proteins. *Biochem*, **7**, 1948–1954.

Eftink M.R. (1997). Fluorescence methods for studying equilibrium macromolecule-ligand interactions. *Methods Enzymol*, **278**, 221–257.

Eftink M.R. and Giron C.A. (1976). Exposure of tryptophanyl residues in proteins. Quantitative determination by fluorescence quenching studies. *Biochemistry*, **15**, 672–680.

Ellis D.J., Dryden D.T., Berge T., Edwardson J.M. and Henderson R.M. (1999). Direct observation of DNA translocation and cleavage by the *Eco*KI endonuclease using atomic force microscopy. *Nat Struct Biol*, **6**, 15–7.

Eskin B., Lautenberger J.A. and Linn S. (1973). Host-controlled modification and restriction of bacteriophage T7 by *Escherichia coli* B. *J Virol*, **11**, 1020–3.

Eskin B. and Linn S. (1972a). The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. *J Biol Chem*, **247**, 6192–6.

Eskin B. and Linn S. (1972b). The deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B. II. Purification, subunit structure, and catalytic properties of the restriction endonuclease. *J Biol Chem*, **247**, 6183–91.

Finazzi-Agro A., Rotilio G., Avigliano L., Guerrieri P., Boffi V. and Mondovi B. (1970). Environment of copper in *Pseudomonas fluorescens* azurin: fluorometric approach. *Biochemistry*, **9**, 2009–14.

Fisher H.F. and Singh N. (1995). Calorimetric methods for interpreting protein-ligand interactions. *Methods Enzymol*, **259**, 194–221.

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

Fuller-Pace F.V. and Murray N.E. (1986). Two DNA recognition domains of the specificity polypeptides of a family of type I restriction enzymes. *Proc Natl Acad Sci U S A.* **83**. 9368–72.

Garcia L.R. and Molineux I.J. (1995). Rate of translocation of bacteriophage T7 DNA across the membranes of *Escherichia coli. J Bacteriol*, **177**, 4066–76.

Garcia L.R. and Molineux I.J. (1996). Transcription-independent DNA translocation of bacteriophage T7 DNA into *Escherichia coli. J Bacteriol*, **178**, 6921–9.

Garcia L.R. and Molineux I.J. (1999). Translocation and specific cleavage of bacteriophage T7 DNA in vivo by *Eco*KI. *Proc Natl Acad Sci U S A*, **96**, 12430–5.

Gefter M., Hausmann R., Gold M. and Hurwitz J. (1966). The enzymatic methylation of ribonucleic acid and deoxyribonucleic acid. X. Bacteriophage T3-induced S-adenosylmethionine cleavage. *J Biol Chem*, **241**, 1995–2006.

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

Gorbalenya A.E. and Konin E.V. (1991). Endonuclease (R) subunits of type-I and type-III restriction modification enzymes contain a helicase-like domain. *FEBS Letters*, **291**. 277–81.

Gough J.A. and Murray N.E. (1983). Sequence diversity among related genes for recognition of specific targets in DNA molecules. *J Mol Biol*, **166**, 1–19.

Graceffa P. and Lehrer S.S. (1980). The excimer fluorescence of pyrene-labeled tropomyosin. A probe of conformational dynamics. *J Biol Chem*, **255**, 11296–300.

Grodberg J. and Dunn J.J. (1988). *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. *J Bacteriol*, **170**, 1245–53.

Gubler M. and Bickle T.A. (1991). Increased protein flexibility leads to promiscuous protein–DNA interactions in type IC restriction-modification systems. *EMBO J*, **10**, 951–7.

Gubler M., Braguglia D., Meyer J., Piekarowicz A. and Bickle T.A. (1992). Recombination of constant and variable modules alters DNA sequence recognition by type IC restriction-modification enzymes. *EMBO J*, **11**, 233–40.

Guinto E.R. and Di Cera E. (1996). Large heat capacity change in a protein-monovalent cation interaction. *Biochemistry*, **35**, 8800–4.

Guyot J.B., Grassi J., Hahn U. and Guschlbauer W. (1993). The role of the preserved sequences of Dam methylase. *Nucleic Acids Res*, **21**, 3183–90.

Haberman A., Heywood J. and Meselson M. (1972). DNA modification methylase activity of *Escherichia coli* restriction endonucleases K and P. *Proc. Natl. Acad. Sci.*, **69**, 3138–3141.

Hadi S.M., Bachi B., Shepherd J.C., Yuan R., Ineichen K. and Bickle T.A. (1979). DNA recognition and cleavage by the *Eco*P15 restriction endonuclease. *J Mol Biol*, **134**, 655–66.

Hadi S.M., Bickle T.A. and Yuan R. (1975). The role of S-adenosylmethionine in the cleavage of deoxyribonucleic acid by the restriction endonuclease from *Escherichia coli* K. *J Biol Chem*, **250**, 4159–64.

Hattman S. (1982). DNA methyltransferase-dependent transcription of the phage Mu *mom* gene. *Proc Natl Acad Sci U S A*, **79**, 5518–21.

Hattman S., Ives J., Margolin W. and Howe M.M. (1985). Regulation and expression of the bacteriophage Mu *mom* gene: mapping of the transactivation (dad) function to the C region. *Gene*, **39**, 71–6.

Hawkins R.E., Russell S.J., Baier M. and Winter G. (1993). The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen. The interaction of mutant D1.3 antibodies with lysozyme. *J Mol Biol*, **234**, 958–64.

Hemphill H.E. and Whiteley H.R. (1975). Bacteriophages of *Bacillus subtilis*. *Bacteriol Rev.*, **39**, 257–315.

Hennecke J., Sillen A., Huber-Wunderlich M., Engelborghs Y. and Glockshuber R. (1997). Quenching of tryptophan fluorescence by the active-site disulphide bridge in the DsbA protein from *Escherichia coli*. *Biochemistry*, **36**, 6391–400.

Herrlich P., Rahmsdorf H.J., Pai S.H. and Schweigher M. (1974). Translational control induced by bacteriophage T7. *Proc Natl Acad Sci U S A.*, **71**, 1088–92.

Hesselbach B.A. and Nakada D. (1977). I protein: bacteriophage T7-coded inhibitor of *Escherichia coli* RNA polymerase. *J Virol*, **24**, 746–60.

Horiuchi K. and Zinder N.D. (1972). Cleavage of bacteriophage fl DNA by the restriction enzyme of *Escherichia coli* B. *Proc Natl Acad Sci U S A*, **69**, 3220–4.

Huber H.E., Beauchamp B.B. and Richardson C.C. (1988). *Escherichia coli* dGTP triphosphohydrolase is inhibited by gene 1.2 protein of bacteriophage T7. *J Biol Chem*, **263**, 13549–56.

Hudson E.N. and Webb G. (1973). Synthesis and characterization of two fluorescenct sulfhydryl reagents. *Biochemistry*, **12**, 4154–61.

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

Inouye M., Arnheim N. and Sternglanz R. (1973). Bacteriophage T7 lysozyme is an N-acetylmuramyl-L-alanine amidase. *J Biol Chem*, **248**, 7247–52.

Janin J. (1995). Elusive affinities. *Proteins*, **21**, 30–9.

Janin J. and Chothia C. (1990). The structure of protein-protein recognition sites. *J Biol Chem*, **265**, 16027–30.

Janscak P., Abadjieva A. and Firman K. (1996). The type I restriction endonuclease R.*Eco*R124I: over-production and biochemical properties. *J Mol Biol*, **257**, 977–91.

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

Jeruzalmi D. and Steitz T.A. (1998). Structure of T7 RNA polymerase complexed to the transcriptional inhibitor T7 lysozyme. *EMBO J*, **17**, 4101–13.

Jonsson T., Waldburger C.D. and Sauer R.T. (1996). Nonlinear free energy relationships in Arc represor unfolding imply the existence of unstable, native-like folding intermediates. *Biochemistry*, **35**, 4795–802.

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

Kannan P., Cowan G.M., Daniel A.S., Gann A.A. and Murray N.E. (1989). Conservation of organization in the specificity polypeptides of two families of type I restriction enzymes. *J Mol Biol*, **209**, 335–44.

Kelley R.F. and O'Connell M.P. (1993). Thermodynamic analysis of an antibody functional epitope. *Biochemistry.*, **32**, 6828–35.

Kim Y.T. and Richardson C.C. (1993). Bacteriophage T7 gene 2.5 protein: an essential protein for DNA replication. *Proc Natl Acad Sci U S A*, **90**, 10173–7.

Kim Y.T., Tabor S., Churchich J.E. and Richardson C.C. (1992). Interactions of gene 2.5 protein and DNA polymerase of bacteriophage T7. *J Biol Chem*, **267**, 15032–40.

King G. and Murray N.E. (1994). Restriction enzymes in cells, not eppendorfs. *Trends Microbiol.* **2**, 465–9.

Klimasauskas S.S., Roberts R.J. and Cheng X. (1994). *Hha*I methyltransferase flips its target base out of the DNA helix. *Cell*, **76**, 357–69.

Klimasauskas S.S., Timinskas A., Menkevicius S., Butkiene D., Butkus V. and Janulaitis A. (1989). Sequence motifs characteristic of DNA[cytosine-N4] methyltransferases: similarity to adenine and cytosine-C5 DNA-methylases. *Nucleic Acids Res*, **17**, 9823–31.

Kneale G.G. (1994). A symmetrical model for the domain structure of type I DNA methyltransferases. *J Mol Biol*, **243**, 1–5.

Koller T., Kubler O., Portmann R. and Sogo J.M. (1978). High resolution physical mapping of specific binding sites of *Escherichia coli* RNA polymerase on the DNA of bacteriophage T7. *J Mol Biol*, **120**, 121–31.

Kossykh V.G., Schlagman S.L. and Hattman S. (1993). Conserved sequence motif DPPY in region IV of the phage T4 Dam DNA-[N6-adenine]-methyltransferase is important for S-adenosyl-L-methionine binding . *Nucleic Acids Res*, **21**, 3563–66.

Kruger D.H. and Bickle T.A. (1983). Bacteriophage survival: multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol Rev*, **47**, 345–60.

Kruger D.H., Kupper D., Meisel A., Reuter M. and Schroeder C. (1995). The significance of distance and orientation of restriction endonuclease recognition sites in viral DNA genomes. *FEMS Microbiol Rev*, **17**, 177–84.

Kruger D.H., Reuter M., Schroeder C., Glatman L.I. and Chernin L.S. (1983). Restriction of bacteriophage T3 and T7 *ocr*$^+$ strains by the type II restriction endonuclease *Eco*RV. *Mol Gen Genet*, **190**, 349–51.

Kruger D.H. and Schroeder C. (1981). Bacteriophage T3 and bacteriophage T7 virus-host cell interactions. *Microbiol Rev*, **45**, 9–51.

Kruger D.H., Schroeder C., Hansen S. and Rosenthal H.A. (1977). Active protection by bacteriophages T3 and T7 against *E. coli* B- and K-specific restriction of their DNA. *Mol Gen Genet*, **153**, 99–106.

Kruger D.H., Schroeder C., Reuter M., Bogdarina I.G., Buryanov Y.I. and Bickle T.A. (1985). DNA methylation of bacterial viruses T3 and T7 by different DNA methylases in *Escherichia coli* K12 cells. *Eur J Biochem*, **150**, 323–30.

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

Lakowicz J.R. (1999). Introduction to fluorescence. *In Principles of fluorescence spectroscopy*, Kluwer Academic/Plenum Publishers, New York.

Latham G.J., Bacheller D.J., Pietroni P. and von Hippel P.H. (1997). Structural analyses of gp45 sliding clamp interactions during assembly of the bacteriophage T4 DNA polymerase holoenzyme. *J Biol Chem*, **272**, 31677–84.

LeClerc J.E. and Richardson C.C. (1979). Gene 2 protein of bacteriophage T7: purification and requirement for packaging of T7 DNA *in vitro*. *Proc Natl Acad Sci U S A*, **76**, 4852–6.

Lee F.S., Shapiro R. and Vallee B.L. (1989). Tight-binding inhibition of angiogenin and ribonuclease A by placental ribonuclease inhibitor. *Biochemistry*, **28**, 225–30.

Lee N.S., Rutebuka O., Arakawa T., Bickle T.A. and Ryu J. (1997). *Kpn*AI, a new type I restriction-modification system in *Klebsiella pneumoniae*. *J Mol Biol*, **271**, 342–8.

Lehrer S.S. (1997). Intramolecular pyrene excimer fluorescence: a probe of proximity and protein conformational change. *Methods Enzymol*, **278**, 286–295.

Linder P., Laski P.F., Ashburner M., Leeroy P.L., Nielson P., Nishi K., Shnier J. and Slonimski P.P. (1989). Birth of the d-e-a-d box. *Nature*, **337**, 121–2.

Loenen W.A., Daniel A.S., Braymer H.D. and Murray N.E. (1987). Organization and sequence of the *hsd* genes of *Escherichia coli* K-12. *J Mol Biol*, **198**, 159–70.

Loenen W.A. and Murray N.E. (1986). Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *J Mol Biol*, **190**, 11–22.

Loewenthal R., Sancho J. and Fersht A.R. (1991). Fluorescence spectrum of barnase: contributions of three tryptophan residues and a histidine-related pH dependence. *Biochemistry*, **30**, 6775–9.

Longstaff C., Campbell A.F. and Fersht A.R. (1990). Recombinant chymotrypsin inhibitor 2: expression, kinetic analysis of inhibition with alpha-chymotrypsin and wild-type and mutant subtilisin BPN', and protein engineering to investigate inhibitory specificity and mechanism. *Biochemistry*, **29**, 7339–47.

Lyakhov D.L., He B., Zhang X., Studier F.W., Dunn J.J. and McAllister W.T. (1997). Mutant bacteriophage T7 RNA polymerases with altered termination properties. *J Mol Biol*, **269**, 28–40.

Lyakhov D.L., He B., Zhang X., Studier F.W., Dunn J.J. and McAllister W.T. (1998). Pausing and termination by bacteriophage T7 RNA polymerase. *J Mol Biol*, **280**, 201–13.

Macdonald L.E., Durbin R.K., Dunn J.J. and McAllister W.T. (1994). Characterization of two types of termination signal for bacteriophage T7 RNA polymerase. *J Mol Biol*, **238**, 145–58.

Makhatadze G.I. and Privalov P.L. (1995). Energetics of protein structure. *Adv Protein Chem*, **47**, 307–425.

Makino O., Kawamura J., Saito H. and Ikeda Y. (1979). Inactivation of restriction endonuclease *Bam*Nx after infection with phage ϕNR2. *Nature*, **277**, 64–6.

Makino O., Saito H. and Ando T. (1980). *Bacillus subtilis*-phage ϕ1 overcomes host-controlled restriction by producing *Bam*Nx inhibitor protein. *Mol Gen Genet*, **179**, 463–8.

Mannherz H.G., Goody R.S., Konrad M. and Nowak E. (1980). The interaction of bovine pancreatic deoxyribonuclease i and skeletal muscle actin. *Eur J Biochem*, **104**, 367–79.

Mark K.K. and Studier F.W. (1981). Purification of the gene 0.3 protein of bacteriophage T7, an inhibitor of the DNA restriction system of *Escherichia coli*. *J Biol Chem*, **256**, 2573–8.

Marsh D.J. and Lowey S. (1980). Fluorescence energy transfer in myosin subfragment-1. *Biochemistry*, **19**, 774–84.

McAllister W.T. and McCarron R.J. (1977). Hybridization of the *in vitro* products of bacteriophage T7 RNA polymerase to restriction fragments of T7 DNA. *Virology.*, **82**, 288–98.

Meisel A., Bickle T.A., Kruger D.H. and Schroeder C. (1992). Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*, **355**, 467–9.

Meisel A., Kruger D.H. and Bickle T.A. (1991). M.*Eco*P15 methylates the second adenine in its recognition sequence. *Nucleic Acids Res*, **19**, 3997.

Mernagh D.R., Taylor I.A. and Kneale G.G. (1998). Interaction of the type I methyltransferase M.*Eco*R124I with modified DNA substrates: sequence discrimination and base flipping. *Biochem J*, **336**, 719–25.

Meselson M. and Yuan R. (1968). DNA restriction enzyme from *E. coli*. *Nature*. **217**, 1110–4.

Michalewicz J. and Nicholson A.W. (1992). Molecular cloning and expression of the bacteriophage T7 0.7 (protein kinase) gene. *Virology*, **186**, 452–62.

Minkley E.G. and Pribnow D. (1973). Transcription of the early region of bacteriophage T7: selective initiation with dinucleotides. *J Mol Biol*, **77**, 255–77.

Moffatt B.A. and Studier F.W. (1987). T7 lysozyme inhibits transcription by T7 RNA polymerase. *Cell*, **49**, 221–7.

Moffatt B.A. and Studier F.W. (1988). Entry of bacteriophage T7 DNA into the cell and escape from host restriction. *J Bacteriol*, **170**, 2095–105.

Mol C.D., Arvai A.S., Sanderson R.J., Slupphaug G., Kavli B., Krokan H.E., Mosbaugh D.W. and Tainer J.A. (1995). Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell*, **82**, 701–8.

Murray N.E. (2000). Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol Mol Biol Rev*, **64**, 412–34.

Murray N.E., Daniel A.S., Cowan G.M. and Sharp P.M. (1993). Conservation of motifs within the unusually variable polypeptide sequences of type I restriction and modification enzymes. *Mol Microbiol*, **9**, 133–43.

Murray N.E., Gough J.A., Suri B. and Bickle T.A. (1982). Structural homologies among type I restriction-modification systems. *EMBO J*, **1**, 535–9.

Nagaraja V., Shepherd J.C. and Bickle T.A. (1985a). A hybrid recognition sequence in a recombinant restriction enzyme and the evolution of DNA sequence specificity. *Nature*, **316**, 371–2.

Nagaraja V., Shepherd J.C., Pripfl T. and Bickle T.A. (1985b). Two type I restriction enzymes from *Salmonella* species. Purification and DNA recognition sequences. *J Mol Biol*, **182**, 579–87.

Nagaraja V., Stieger M., Nager C., Hadi S.M. and Bickle T.A. (1985c). The nucleotide sequence recognised by the *Escherichia coli* D type I restriction and modification enzyme. *Nucleic Acids Res*, **13**, 389–99.

Nakai H. and Richardson C.C. (1986). Interactions of the DNA polymerase and gene 4 protein of bacteriophage T7. Protein-protein and protein-DNA interactions involved in RNA-primed DNA synthesis. *J Biol Chem*, **261**, 15208–16.

Nechaev S. and Severinov K. (1999). Inhibition of *Escherichia coli* RNA polymerase by bacteriophage T7 gene 2 protein. *J Mol Biol*, **289**, 815–26.

Nick C. and Scholtz J.M. (1997). Measuring the conformational stability of a protein. *In Protein structure, a practical approach, Oxford University Press*, 299–321.

Notarnicola S.M., Mulcahy H.L., Lee J. and Richardson C.C. (1997). The acidic carboxyl terminus of the bacteriophage T7 gene 4 helicase/primase interacts with T7 DNA polymerase. *J Biol Chem*, **272**, 18425–33.

Noyer-Weidner M., Jentsch S., Pawlek B., Gunthert U. and Trautner T.A. (1983). Restriction and modification in *Bacillus subtilis*: DNA methylation potential of the related bacteriophages Z, SPR, SP beta, ϕ3T, and ρ11. *J Virol*, **46**, 446–53.

Noyer-Weidner M., Pawlek B., Jentsch S., Gunthert U. and Trautner T.A. (1981). Restriction and modification in *Bacillus subtilis*: gene coding for a *Bsu*R-specific modification methyltransferase in the temperate bacteriophage ϕ3T. *J Virol*, **38**, 1077–80.

O'Neill M., Dryden D.T. and Murray N.E. (1998). Localization of a protein-DNA interface by random mutagenesis. *EMBO J*, **17**, 7118–27.

Oppenheim D.S. and Yanofsky C. (1980). Translational coupling during expression of the tryptophan operon of *Escherichia coli*. *Genetics*, **95**, 785–95.

Pace C.N. and Schmid F.X. (1997). How to determine the molar absorbance coefficient of a protein. *In Protein structure. A practical approach*, 253–59.

Pai S.H., Rahmsdorf H.J., Ponta H., Hirsch-Kauffmann M., Herrlich P. and Schweiger M. (1975). Protein kinase of bacteriophage T7. 2. Properties, enzyme synthesis *in vitro* and regulation of enzyme synthesis and activity in vivo. *Eur J Biochem*, **55**, 305–14.

Pao C.C. and Speyer J.F. (1973). Order of injection of T7 bacteriophage DNA. *J Virol*, **11**, 1024–6.

Perona J.J., Tsu C.A., McGrath M.E., Craik C.S. and Fletterick R.J. (1993). Relocating a negative charge in the binding pocket of trypsin. *J Mol Biol*, **230**, 934–49.

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

Pokalsky C., Wick P., Harms E., Lytle F.E. and van Etten R.L. (1995). Fluorescence resolution of the intrinsic tryptophan residues of bovine protein tyrosyl phosphatase. *J Biol Chem*, **270**, 3809–15.

Ponta H., Altendorf K.H., Schweiger M., Hirsch-Kaufmann M., Pfennig-Yeh M.L. and Herrlich P. (1976). *E. coli* membranes become permeable to ions following T7-virus-infection. *Mol Gen Genet*, **149**, 145–50.

Ponta H., Oratzel M., Pfennig-Yeh M., Hirsch-Kauffmann M. and Schweiger M. (1977). Membrane alteration induced by T7 virus infection. *FEBS Lett*, **73**, 207–9.

Posfai J., Bhagwat A.S., Posfai G. and Roberts R.J. (1989). Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res*, **17**, 2421–35.

Powell L.M., Connolly B.A. and Dryden D.T. (1998a). The DNA binding characteristics of the trimeric *Eco*KI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting. *J Mol Biol*. **283**, 947–61.

Powell L.M., Dryden D.T. and Murray N.E. (1998b). Sequence-specific DNA binding by *Eco*KI. a type IA DNA restriction enzyme. *J Mol Biol*, **283**, 963–76.

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

Powell L.M. and Murray N.E. (1995). S-adenosyl methionine alters the DNA contacts of the *Eco*KI methyltransferase. *Nucleic Acids Res*, **23**, 967–74.

Pribnow D. (1975). Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. *J Mol Biol*, **99**, 419–43.

Price C., Lingner J., Bickle T.A., Firman K. and Glover S.W. (1989). Basis for changes in DNA recognition by the *Eco*R124 and *Eco*R124/3 type I DNA restriction and modification enzymes. *J Mol Biol*, **205**, 115–25.

Putnam C.D., Shroyer M.J., Lundquist A.J., Mol C.D., Arvai A.S., Mosbaugh D.W. and Tainer J.A. (1999). Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with *Escherichia coli* uracil-DNA glycosylase. *J Mol Biol*, **287**, 331–46.

Ravetch J.V., Horiuchi K. and Zinder N.D. (1978). Nucleotide sequence of the recognition site for the restriction-modification enzyme of *Escherichia coli* B. *Proc Natl Acad Sci U S A*, **75**, 2266–70.

Reuter M., Kupper D., Meisel A., Schroeder C. and Kruger D.H. (1998). Cooperative binding properties of restriction endonuclease *Eco*RII with DNA recognition sites. *J Biol Chem*, **273**, 8294–300.

Richardson C.C. (1983). Bacteriophage T7: minimal requirements for the replication of a duplex DNA molecule. *Cell*, **33**, 315–7.

Richardson 3rd J.M., Lemaire S.D., Jacquot J.P. and Makhatadze G.I. (2000). Difference in the mechanisms of the cold and heat induced unfolding of thioredoxin h from chlamydomonas reinhardtii: spectroscopic and calorimetric studies. *Biochemistry*, **39**, 11154–62.

Roberts R.J. and Cheng X. (1998). Base flipping. *Annu Rev Biochem*, **67**, 181–198.

Robertson E.S. and Nicholson A.W. (1990). Protein kinase of bacteriophage T7 induces the phosphorylation of only a small number of proteins in the infected cell. *Virology*, **175**, 525–34.

Rosamond J., Endlich B. and Linn S. (1979). Electron microscopic studies of the mechanism of action of the restriction endonuclease of *Escherichia coli* B. *J Mol Biol*, **129**, 619–35.

Rost B. and Sander C. (1993). Prediction of protein secondary structure at better than 70 % accuracy. *J Mol Biol*, **232**, 584–99.

Saigo K. (1975). Polar DNA ejection in bacteriophage T7. *Virology*, **65**, 120–7.

Sain B. and Murray N.E. (1980). The *hsd* (host specificity) genes of *E. coli* K 12. *Mol Gen Genet*, **180**, 35–46.

Saito H., Tabor S., Tamanoi F. and Richardson C.C. (1980). Nucleotide sequence of the primary origin of bacteriophage T7 DNA replication: relationship to adjacent genes and regulatory elements. *Proc Natl Acad Sci U S A*, **77**, 3917–21.

Sambrook J., Fritsch E.F. and Maniatis T. (1989). In Molecular cloning. *Cold Spring Harbour Laboratory*.

Sanyal G., Kim E., Thompson F.M. and Brady E.K. (1989). Static quenching of tryptophan fluorescence by oxidized dithiotreithol. *Biochem Biophys Res Commun*, **165**, 772–81.

Savva R., McAuley-Hecht K., Brown T. and Pearl L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature*, **373**, 487–93.

Savva R. and Pearl L.H. (1995). Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. *Nat Struct Biol*, **2**, 752–7.

Schreiber G. and Fersht A.R. (1993). Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry*, **32**, 5145–50.

Schroeder C., Jurkschat H., Meisel A., Reich J.G. and Kruger D. (1986). Unusual occurrence of *Eco*P1 and *Eco*P15 recognition sites and counterselection of type II methylation and restriction sequences in bacteriophage T7 DNA. *Gene*. **45**, 77–86.

Serwer P., Khan S.A., Hayes S.J., Watson R.H. and Griess G.A. (1997). The conformation of packaged bacteriophage T7 DNA: informative images of negatively stained T7. *J struct Biol*, **120**, 32–43.

Sharp P.M., Kelleher J.E., Daniel A.S., Cowan G.M. and Murray N.E. (1992). Roles of selection and recombination in the evolution of type I restriction-modification systems in enterobacteria. *Proc Natl Acad Sci U S A*, **89**, 9836–40.

Siebenlist U. (1979). Nucleotide sequence of the three major early promoters of bacteriophage T7. *Nucleic Acids Res*, **6**, 1895–907.

Spoerel N., Herrlich P. and Bickle T.A. (1979). A novel bacteriophage defence mechanism: the anti-restriction protein. *Nature*, **278**, 30–4.

Stahl S.J. and Chamberlin M.J. (1977). An expanded transcriptional map of T7 bacteriophage. Reading of minor T7 promoter sites in vitro by *Escherichia coli* RNA polymerase. *J Mol Biol*, **112**, 577–601.

Stahl S.J. and Zinn K. (1981). Nucleotide sequence of the cloned gene for bacteriophage T7 RNA polymerase. *J Mol Biol*, **148**, 481–5.

Strasburgh G.M., Leavis P.C. and Gergely J. (1985). Troponin-C-mediated calcium-sensitive changes in the conformation of troponin I detected by pyrene excimer fluorescence. *J Biol Chem*, **260**, 366–70.

Struthers-Schlinke J.S., Robins W.P., Kemp P. and Molineux I.J. (2000). The internal head protein Gp16 controls DNA ejection from the bacteriophage T7 virion. *J Mol Biol*, **301**, 35–45.

Studier F.W. (1969). The genetics and physiology of bacteriophage T7. *Virology*, **39**, 562–74.

Studier F.W. (1972). Bacteriophage T7. *Science*, **176**, 367–76.

Studier F.W. (1975). Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host. *J Mol Biol*, **94**, 283–95.

Studier F.W. (1979). Relationships among different strains of T7 and among T7-related bacteriophages. *Virology*, **95**, 70–84.

Studier F.W. and Bandyopadhyay P.K. (1988). Model for how type I restriction enzymes select cleavage sites in DNA. *Proc Natl Acad Sci U S A*, **85**, 4677–81.

Studier F.W. and Maizel J.V. (1969). T7-directed protein synthesis. *Virology*, **39**, 575–86.

Studier F.W. and Movva N.R. (1976). SAMase gene of bacteriophage T3 is responsible for overcoming host restriction. *J Virol*, **19**, 136–45.

Studier F.W., Rosenberg A.H., Dunn J.J. and Dubendorff J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol*, **185**, 60–89.

Studier F.W., Rosenberg A.H., Simon M.N. and Dunn J.J. (1979). Genetic and physical mapping in the early region of bacteriophage T7 DNA. *J Mol Biol*, **135**, 917–37.

Sturrock S.S. and Dryden D.T.F. (1997). A prediction of the amino acids and structures involved in dna recognition by type i dna restriction and modification enzymes. *Nucleic Acids Res*, **25**, 3408–14.

Subramanya H.S., Doherty A.J., Ashford S.R. and Wigley D.B. (1996). Crystal structure of an ATP-dependent DNA ligase from bacteriophage T7. *Cell*, **85**, 607–15.

Sun W. and Hattman S. (1996). *Escherichia coli* OxyR protein represses the unmethylated bacteriophage Mu *mom* operon without blocking binding of the transcriptional activator C. *Nucleic Acids Res*, **24**, 4042–9.

Sun W. and Hattman S. (1998). Bidirectional transcription in the mom promoter region of bacteriophage Mu. *J Mol Biol*, **284**, 885–92.

Suri B., Shepherd J.C. and Bickle T.A. (1984). The *Eco*A restriction and modification system of *Escherichia coli* 15T-: enzyme structure and DNA recognition sequence. *EMBO J*, **3**, 575–9.

Szczelkun M.D., Dillingham M.S., Janscak P., Firman K. and Halford S.E. (1996). Repercussions of DNA tracking by the type IC restriction endonuclease *Eco*R124I on linear, circular and catenated substrates. *EMBO J*, **15**, 6335–47.

Szczelkun M.D., Janscak P., Firman K. and Halford S.E. (1997). Selection of non-specific DNA cleavage sites by the type IC restriction endonuclease *Eco*R124I. *J Mol Biol*, **271**, 112–23.

Tame J., O'Brien R. and Ladbury J. (1998). Isothermal titration calorimetry of biomolecules. *in Biocalorimetry. Applications of calorimetry in the biological science, John Wiley and sons Ltd.*, 27–38.

Taylor I., Patel J., Firman K. and Kneale G. (1992). Purification and biochemical characterisation of the *Eco*R124 type I modification methylase. *Nucl Acids Res*, **20**, 179–86.

Thorpe P.H., Ternent D. and Murray N.E. (1997). The specificity of *Sty*SKI, a type I restriction enzyme, implies a structure with rotational symmetry. *Nucleic Acids Res*, **25**, 1694–700.

Tiedge H., Lucken U., Weber J. and Shafer G. (1982). Hihg-affinity binging of ADP analogues to mithocondrial F_1-ATP-ase. *Eur J Biochem*, **127**, 291–294.

Titheradge A.J., Ternent D. and Murray N.E. (1996). A third family of allelic hsd genes in *Salmonella enterica*: sequence comparisons with related proteins identify conserved regions implicated in restriction of DNA. *Mol Microbiol*, **22**, 437–47.

Tyndall C., Meister J. and Bickle T.A. (1994). The *Escherichia coli prr* region encodes a functional type IC DNA restriction system closely integrated with an anticodon nuclease gene. *J Mol Biol*, **237**, 266–74.

Vaughan C.K., Buckle A.M. and Fersht A.R. (1999). Structural response to mutation at a protein-protein interface. *J Mol Biol*, **286**, 1487–506.

Vovis G.F., Horiuchi K. and Zinder N.D. (1974). Kinetics of methylation of DNA by a restriction endonuclease from *Escherichia coli* B. *Proc Natl Acad Sci U S A*, **71**, 3810–3.

Wallis R., Leung K.Y., Osborne M.J., James R., Moore G.R. and Kleanthous C. (1998). Specificity in protein-protein recognition: conserved Im9 residues are the major determinants of stability in the colicin E9 DNase-Im9 complex. *Biochemistry*, **37**, 476–85.

Wallis R., Leung K.Y., Pommer A.J., Videler H., Moore G.R., James R. and Kleanthous C. (1995a). Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 2. Cognate and noncognate interactions that span the millimolar to femtomolar affinity range. *Biochemistry*, **34**, 13751–9.

Wallis R., Moore G.R., James R. and Kleanthous C. (1995b). Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 1. Diffusion-controlled association and femtomolar binding for the cognate complex. *Biochemistry*, **34**, 13743–50.

Wasylewski Z., Kaszycki P., Guz A. and Stryjewski W. (1988). Fluorescence-quenching-resolved spectra of fluorophores in mixtures and micellar solutions. *Eur J Biochem*, **178**, 471–476.

Watson P.S.R.H. and Hayes S.J. (1992). Formation of the right before the left mature DNA end during packing-cleavage of bacteriophage T7 DNA concatemers. *J Mol Biol*. **226**, 311–17.

Webb J.L., King G., Ternent D., Titheradge A.J. and Murray N.E. (1996). Restriction by *Eco*KI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J*, 15(8), 2003–9.

Weiserova M. and Firman K. (1998). Isolation of a non-classical mutant of the DNA recognition subunit of the type I restriction endonuclease R.*Eco*R124I. *Biol Chem*, **379**, 585–9.

Wilkins B.M., Chilley P.M., Thomas A.T. and Pocklington M.J. (1996). Distribution of restriction enzyme recognition sequences on broad host range plasmid rp4: molecular and evolutionary implications. *J Mol Biol*, **258**, 447–56.

Willcock D.F., Dryden D.T. and Murray N.E. (1994). A mutational analysis of the two motifs common to adenine methyltransferases. *EMBO J*, **13**, 3902–8.

Wu C.S. and Yang J.T. (1980). Helical conformation of glucagon in surfactant solutions. *Biochemistry*, **19**, 2117–22.

Wu C.W., Yarbrough L.R. and Wu F.Y.H. (1976). N-(1-Pyrene)maleimide: a fluorescent cross-linking reagent. *Biochemistry*, **15**, 2863–68.

Yuan R. (1981). Structure and mechanism of multifunctional restriction endonucleases. *Annu Rev Biochem*, **50**, 285–319.

Yuan R., Bickle T.A., Ebbers W. and Brack C. (1975). Multiple steps in DNA recognition by restriction endonuclease from *E. coli* K. *Nature*, **256**, 556–60.

Yuan R., Hamilton D.L. and Burckhardt J. (1980). DNA translocation by the restriction enzyme from *E. coli* K. *Cell*, **20**, 237–44.

Yuan R., Heywood J. and Meselson M. (1972). ATP hydrolysis by restriction endonuclease from *E. coli* K. *Nat New Biol*, **240**, 42–3.

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

Zavriev S. and Shemyakin M. (1982). RNApolymerase-dependent mechanism for stepwise T7 phage DNA transport from the virion into *E. coli*. *Nucl Acids Res*, **5**, 1635–1651.

Zhang X. and Studier F.W. (1995). Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. *J Mol Biol.*, **269**, 10–27.

Zhang X. and Studier F.W. (1997). Mechanism of inhibition of bacteriophage T7 RNA polymerase by T7 lysozyme. *J Mol Biol*, **269**, 10–27.

Ziegler M., Xiao R. and Penefsky H.S. (1994). Close proximity of Cys^{64} and Cys^{140} in the δ subunit of *Escherichia coli* F_1-ATPase. *J Biol Chem.*, **269**, 4233–39.

Zillig W., Fujiki H., Blum W., Janekovic D., Schweiger M., Rahmsdorf H., Ponta H. and Hirsch-Kauffmann M. (1975). *In vivo* and *in vitro* phosphorylation of DNA-dependent RNA polymerase of *Escherichia coli* by bacteriophage-T7-induced protein kinase. *Proc Natl Acad Sci U S A*, **72**, 2506–10.

Appendix A.
Crystal structure of Ocr

During the writing of this thesis, the crystal structure of Ocr protein was solved (Sturrock *et al.* , unpublished). The protein has an elongated form and displays many acidic residues on the surface. Asn4 residues are in close proximity, at the interface between the two monomers. Ser68 residues are also close to each other but a bit further away than Asn4 residues. The six amino acids are not clustered in a particularly area but spread out through the entire Ocr surface.

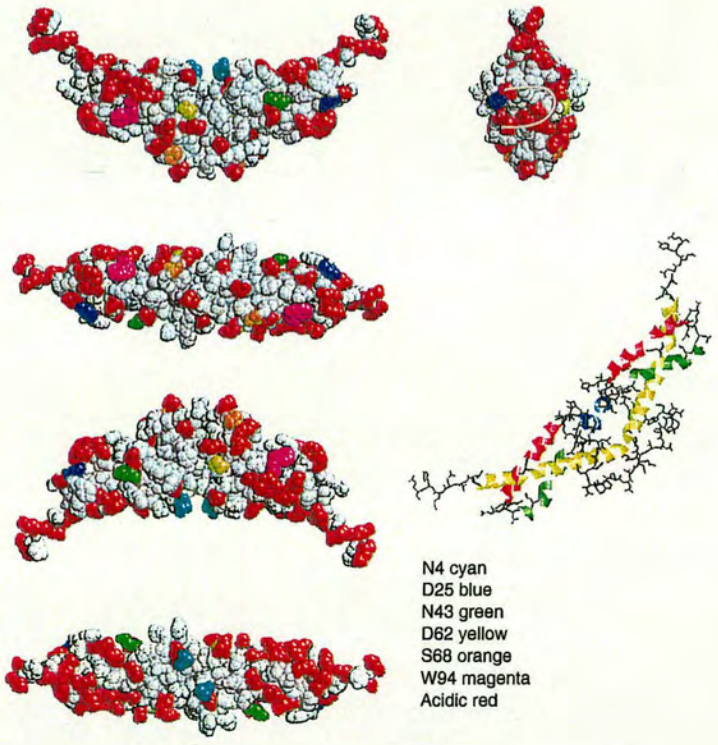


Figure 5.1: The crystal structure of Ocr protein. N4 residues are colored in cyan, D25 in blue, N43 in green, D62 in yellow, S68 in orange, W94 in magenta and the acidic residues in red.

Appendix B.
Paper submitted for publication

DNA structural mimicry by the gene 0.3 protein of bacteriophage T7, an inhibitor of type I DNA restriction enzymes.

C. Atanasiu[1], O. Byron[2], J. Blackstock[3], S.S. Sturrock[1], H. McMiken, S. Egelhaaf[3] & D.T.F. Dryden*

Department of Chemistry
The King's Buildings
University of Edinburgh
Edinburgh
EH9 3JJ
United Kingdom
Telephone 0131-650-4735, Fax 0131-650-6453, email David.Dryden@ed.ac.uk

* author for correspondence
1 Institute of Cell & Molecular Biology, The King's Buildings, University of Edinburgh, Edinburgh, EH9 3JR, United Kingdom
2 Institute of Biomedical & Life Sciences, Division of Infection & Immunity, Joseph Black Building, University of Glasgow, G12 8QQ, United Kingdom
3 Department of Physics & Astronomy, The King's Buildings, University of Edinburgh, Edinburgh, EH9 3JZ, United Kingdom

Summary

Abbreviations used: aa, amino acid(s); AdoMet, S-adenosyl-methionine; DLS, dynamic light scattering; DSC, differential scanning calorimetry; EDTA, ethylene-diamine-tetra-acetic acid; esms, electrospray mass spectrometry: GuCl, guanidinium hydrochloride; HPLC, high performance liquid chromatography; ILT, inverse Laplace transform; 1,5-I-AEDANS, N-iodoacetyl-N-(5-sulpho-1-napthyl)ethylenediamine; M, modification subunit; ocr, overcome classical restriction; R, restriction subunit; R/M restriction and modification; S, specificity subunit; TRD, target recognition domain; UGI, uracil glycoslyase inhibitor protein; UNG, uracil glycosylase.

Introduction

Upon injection of bacteriophage DNA from the viral particle into a bacterial cell, the viral DNA is often vulnerable to attack by bacterial restriction/modification (R/M) systems. In general, these R/M systems comprise a restriction endonuclease activity which cleaves unmodified DNA and a modification methyltransferase activity to modify specific target sequences on the bacterial DNA. They are grouped into three main type, I, II and III (reviewed by Wilson & Murray, 1991; Bickle & Kruger, 1993; Murray, 2000). The modification of the host DNA protects it from degradation by the restriction endonuclease but the viral DNA, being unmethylated, is vulnerable to the endonuclease. Bacteriophage have developed numerous mechanisms to defend themselves from R/M systems including the synthesis of antirestriction proteins which block the action of the restriction endonuclease (Kruger & Bickle, 1983; Bickle & Kruger, 1993). One such protein is the product of gene 0.3 of bacteriophage T7. This gene, often called *ocr* (overcome classical restriction) is the first to be transcribed and translated upon injection of T7 phage DNA into the *E.coli* host. The DNA sequence preceding the gene is inaccessible to restriction endonucleases, probably due to masking of the DNA by other protein-DNA complexes, and this allows time for gene 0.3 to be transcribed (Moffat & Studier, 1976; Garcia & Molineux,1995). Once translated, the ocr protein targets the cell's type I R/M systems and inactivates them by binding very tightly to the type I R/M enzyme (Mark & Studier, 1981; Bandyopadhyay *et. al.*, 1985). The type I R/M enzymes contain both restriction endonuclease and modification methyltransferase activities within one large, oligomeric enzyme (Wilson & Murray, 1991; Bickle & Kruger, 1993; Murray, 2000). The blockage of the DNA binding site by the ocr protein prevents the R/M enzyme from binding to unmethylated phage DNA and allows the phage to propagate. Ocr operates against type I R/M enzymes which possess different DNA target specificities suggesting that ocr can operate in a non-specific manner against these R/M systems (Kruger *et. al.*, 1977; Mark & Studier, 1981; Kruger *et. al.*, 1983; Bandyopadhyay *et. al.*, 1985; Kruger *et. al.*, 1985), figure 1.

In order to inhibit all type I R/M enzymes, ocr must recognise some common feature. Type I R/M enzymes comprise three types of subunits combining DNA sequence specificity, methyltransferase activity (modification) and endonuclease activity (restriction) in one large oligomeric protein (see Murray, 2000 for review). R (restriction) subunits cleave the DNA at a site thousands of base pairs distant from the unmethylated DNA target recognition sequence in a complex reaction requiring DNA translocation powered by ATP hydrolysis. The M (modification) subunit contains the S-adenosyl-methionine-dependent methyltransferase activity and methylates hemimethylated DNA target recognition sequences. The S (specificity) subunit forms the core of the enzyme binding both M and R subunits in addition to recognising the DNA target sequence. The S subunits contain two target recognition domains (TRDs), each of which recognises one part of the bipartite DNA target recognition sequence characteristic of these enzymes, held in appropriate orientation by conserved polypeptide linkers. The subunit stoichiometry of active type I R/M enzymes is $R_2M_2S_1$. Apart from conserved catalytic motifs associated with the multiple enzyme activities displayed by type I R/M enzymes and conserved sequences linking the TRDs, there is little sequence similarity between type I R/M enzymes from a diverse range of eubacteria and archaea. It was found that ocr would displace the DNA recognition sequence from a type I R/M enzyme but given the diversity of type I R/M enzymes and their DNA target sequences, the ocr protein must be acting in a manner not dependent upon the DNA sequence recognised by the type I R/M enzyme (Bandyopadhyay *et. al.*, 1985). It was suggested that ocr may be acting as a mimic of DNA structure rather than of DNA sequence and that this would allow ocr to trick type I R/M enzymes of different DNA sequence specificity into binding to the inhibitor rather than to phage DNA. However, to avoid binding to other proteins which interact with DNA, ocr must mimic a distorted DNA conformation specific for type I R/M enzymes. Mimicry of a distorted DNA structure has been observed for the uracil glycosylase inhibitor protein, UGI, which binds to uracil glycosylase, UNG (Mol *et. al.*, 1995; Savva & Pearl, 1995; Putnam *et. al.*, 1999).

Ocr has been found to be a dimer comprising two 13.5kDa subunits (Mark & Studier, 1981). Of the 116 amino acids in each subunit, only the first 94 are essential for activity (Dunn *et. al.*, 1981), figure 2. Overall the protein is highly acidic containing 34 aspartic or glutamic acid residues and only 6 arginine or lysine residues (Dunn *et. al.*, 1981). Most of the acidic residues, 14 out of 34, are located in the non-essential C-terminus while none of the basic residues are in this region (Dunn *et. al.*, 1981). Proteins of similar mass to ocr are widespread in T7-like phage, which infect both enteric and non-enteric bacteria (Hausmann, 1988). However, their gene coding sequences and function are not known A weak sequence similarity between T7 ocr, T3 ocr and the S subunits of type I R/M enzymes has been postulated (Belogurov & Delver, 1995).

In this paper we present a characterisation of the structure and shape of the ocr protein using a range of hydrodynamic, biochemical and spectroscopic techniques to provide a framework for analysing the interaction of ocr with type I R/M enzymes and to test the hypothesis that ocr is a structural mimic of DNA.

Results

Size and shape of ocr.

To determine the molecular mass of ocr and its shape we employed sedimentation equilibrium, sedimentation velocity and dynamic light scattering.

Sedimentation equilibrium measurements using ultracentrifugation were initially performed with the truncated form of ocr containing only the N-terminal 99 amino acids. From the amino acid sequence, the monomer mass is known to be 11,512 Da. Thus a dimer would have a mass of 23,024 Da. The extrapolations in figure 3 yield masses of 24.7 ± 0.2 kDa and 23.9 ± 0.2 kDa for the 25,000 and 32,000 rpm data respectively. These data indicate that truncated ocr is a dimer in solution at the concentrations studied. Good fits to the raw data were obtained with the simplest form of equation 2 for a thermodynamically ideal system comprising one molecular species, although a concentration

dependence in the variance was observed: at higher loading concentrations the residuals of the fit ceased to be random and instead displayed a pattern indicative of thermodynamic non-ideality. This trend was more apparent for the data obtained at 32,000 rpm than for the 25,000 rpm data. This behaviour is characterised by B, the second virial coefficient, equation 3. It was possible to estimate the excess charge (Z) on the monomer at pH 8 as being -18. Thus the second term in equation 3 has a value of 47 ml/g for the dimer. The gradients of the least-squares fits in figure 3a yield B: for the 25,000 rpm data BM = 101.1 ± 17.5 ml/g and for the 32,000 data BM = 49.3 ± 14.8 ml/g if the dimer mass is used. The much higher value of BM for the 25,000 rpm data most probably stems from the presence of a small amount of higher molecular weight contaminant that appears to have been largely removed from the bulk of the solution column at the higher rotor speed of 32,000 rpm. Taking the contribution to the excluded volume of protein due to charge (47 ml/g) from the total determined experimentally leaves a measure of the excluded volume due to shape, equation 4. For the 25,000 rpm data this value is 54.1 ± 17.5 ml/g, whilst for the 32,000 rpm data this reduces to 2.3 ± 14.8 ml/g. For a sphere this shape term is 4v = 2.89 ml/g thus the 32,000 rpm data imply that the truncated protein is globular, although the upper error limit extends this shape to a cylindrical rod of axial ratio 24. The 25,000 rpm data on the other hand imply that the protein is highly asymmetrical, with an axial ratio of 75 ± 24 if it were a cylindrical rod.

Sedimentation velocity experiments were performed to obtain a better estimate of the shape of the truncated ocr. Analysis of the sedimentation velocity data using the program SVEDBERG yielded, after correction to standard conditions, the weight-average sedimentation coefficients ($s_{20,w}$) plotted in figure 3b. The increase in $s_{20,w}$ with concentration is hard to reconcile with the non-ideality indicated by sedimentation equilibrium studies: this should decrease at higher concentrations. However, linear extrapolation of the data to infinite dilution gives a sedimentation coefficient of $s^0_{20,w}$= 2.2 ± 0.1 S. From equations 5, 7 and 8, the frictional ratio is f/f_0= 1.33 ± 0.11 and the frictional ratio due only to molecular asymmetry $(f/f_0)_{sh}$ calculated at various values of protein hydration are given in table 1. Modelling the protein as a prolate ellipsoid of revolution gives axial ratios between 3.50 ± 1.31 and 4.64 ± 1.00 using equation 9. Perhaps unsurprisingly then, the sedimentation velocity data more closely agree with the 32,000 rpm sedimentation equilibrium than those acquired at 25,000 rpm and their combined strong implication is that the truncated form of ocr containing only 99 amino acids is a slightly elongated, tightly associated dimer.

From the amino acid sequence of full length ocr the monomer mass is known to be 13,678 Da. Thus a dimer would have a mass of 27,356 Da. The extrapolation of sedimentation equilibrium data in figure 3 yields a mass close to this value: 29.4 ± 0.2 kDa indicating that ocr is also a dimer in solution at the concentrations studied in agreement with earlier experiments (Mark & Studier, 1981). The fits to the raw data obtained with the simplest form of equation 2 were good, with little or no concentration dependence in the variance observed.

The excess charge on the ocr monomer at pH 8 was estimated as being -28. Thus the second term in equation 3 has a value of 95.5 ml/g for the dimer. The gradients of the least-squares fit in figure 3a yield BM = 49.9 ± 15.6 ml/g if the dimer mass is used. How can it be possible that the experimental excluded volume due to charge effects is less than that predicted? This could appear to be the case if non-ideality effects were being masked by further self-association. Perhaps the elevated mass of 29.4 kDa instead of 27.4 kDa obtained by extrapolation of the data in figure 3a is evidence of an oligomerisation event with a high dissociation constant.

Analysis of the sedimentation velocity data for ocr was performed in the same manner as for the truncated protein. The reduced data are plotted in figure 3b. Extrapolation of these data to infinite dilution yields $s^0_{20,w}$= 2.5 ± 0.1 S and f/f_0= 1.35 ± 0.06. The frictional ratio due to shape at different protein hydrations ranges between 1.165 ± 0.05 and 1.244 ± 0.05 and the axial ratio varies between 3.75 ± 0.72 and 4.91 ± 0.71 as shown in table 2. An important difference between the sedimentation velocity data for ocr and the truncated protein is that the latter shows a trend with increasing concentration indicative of non-ideality.

Analysis of the DLS data using the ILT method produced a translational diffusion constant of 79.73 ± 0.56 μm^2 / sec. The quoted errors are the standard error of the diffusional constants obtained over a number of angles and runs. Figure 4 shows the homodyne autocorrelation function {g2(Q,t)} of several representative data sets at a scattering angle of 60°. A frictional coefficient and frictional ratio were calculated from equations 5, 7 and 8, and the frictional ratio due to asymmetry at various values of protein hydration are given in table 2. The axial ratio, depending upon the level of hydration of the protein used in equation 9, varies between 3.822 ± 0.110 and 4.990 ± 0.115 in excellent agreement with the shape calculated from sedimentation velocity measurements.

We conclude from the sedimentation velocity and DLS measurements that the full length ocr protein is more elongated than the truncated 99 amino acid form. The sedimentation equilibrium measurements show that both proteins are dimeric but do not define a shape particularly accurately.

Detailed environments of selected amino acids.

1. Tryptophan 94

The fluorescence emission, when excited at 295nm to produce only excitation of the tryptophan residues, showed an emission maximum of 350nm (data not shown). This is close to that found for pure tryptophan in water indicating that the tryptophan at position 94 in ocr is highly exposed to the aqueous solvent. The fluorescence decay of the tryptophan emission could be fitted to a sum of two exponentials with lifetimes of 1.74±0.13ns (33.5%) and 4.30±0.04ns (66.5%) with a χ^2 value of 1.2 (data not shown) giving an average fluorescence lifetime of 3.44 ns. The Perrin plot showing the variation in anisotropy of the tryptophan emission as a function of temperature and viscosity was multivalued indicating that the fluorophores were free to move independently of the rest of the protein, figure 5. The intercept of the Perrin plot for data obtained at fixed glycerol concentrations gives the intrinsic anisotropy, r_a, of the fluorophore as 0.166, rather lower than expected for tryptophan excited at 295nm (Lakowicz, 1999). From the

intercepts of data collected at fixed temperatures, r_{int}, one can calculate that the tryptophan residue is effectively rotating in a cone with a semi-angle varying between 7.5° to 18° depending upon temperature, equation 11. From the slopes at fixed temperatures and equation 10, one can also calculate that at 25°C in 0% glycerol, ocr has a rotational correlation time of 65ns. This time is much longer than that calculated from equation 12 for a hydrated sphere of the same mass as ocr, table 2, indicating that ocr is not spherical. The error in this determination of overall rotational correlation time is large due to the shortness of the fluorescence decay relative to the tumbling time of the protein and does not support a calculation of the axial ratio for ocr from the Perrin plot. Time-resolved fluorescence anisotropy decay measurements would resolve this ambiguity.

The ocr protein contains a large excess of acidic amino acid side chains over basic side chains. This observation prompted the suggestion that the excess negative charge could be responsible for the tight binding of ocr to the DNA binding site of *Eco*KI (Bandyopadhay *et. al.*, 1985). To determine whether W94 was located in a highly charged region of the surface, fluorescence quenching using the two charged quenchers Cs^+ and I^- was performed. The ionic strength of the solution was kept constant by the addition of NH_4Cl which has no effect upon the fluorescence emission. It is known that Cs^+ is a poor quencher of tryptophan with only about 20% quenching efficiency whilst I^- is 100% efficient as a dynamic quencher and that their charge prevents them from penetrating into the protein matrix to quench buried fluorophores (Eftink & Ghiron, 1981). Therefore, it was expected that I^- would quench the fluorescence better than Cs^+. In fact the opposite was observed with I^- being virtually unable to cause any quenching and Cs^+ being an exceptionally good quencher, figure 6. Cs^+ quenching produced a curved quenching profile requiring a two component fit, equation 13. The quenching constants for these data were $2.26 \pm 0.06\ M^{-1}$ (80%) and $99.60 \pm 18.63\ M^{-1}$ (20%). The smaller quenching constant found for Cs^+ is identical to that found for free tryptophan, $2.25 \pm 0.02\ M^{-1}$ (data not shown), indicating that the ion has complete access to a fraction of the conformational space available to the fluorophore. The very high quenching constant is far greater than physically possible by a normal collisional quenching mechanism. It is likely that this fraction of the quenching is due to the preferential attraction of Cs^+ to a negatively charged location near to the tryptophan residue and that it may even be bound at this location. The two fluorescence lifetimes of the tryptophan were both quenched by 0.2M Cs^+ to values of 1.10 ± 0.13 ns (42.3%) and 2.62 ± 0.04 ns (57.7%) indicating that the environments giving rise to the different lifetimes are equally accessible to the quencher. Quenching by I^- could be fitted with a one component Stern-Volmer equation with a quenching constant of 0.70 ± 0.03. This constant is far lower than the value of $12.23 \pm 0.09\ M^{-1}$ found for free tryptophan (data not shown). These results imply that W94 is located within a highly negatively charged patch upon the surface of ocr. Equivalent quenching effects were also observed for the two truncated forms of ocr containing 109 and 99 amino acids (data not shown) ruling out the possibility that the negatively charged patch around Trp94 is actually part of the highly acidic tail of ocr. This is consistent with the observation that the tail of ocr is not required for activity *in vivo* (Dunn *et. al.*, 1981).

2. Site-directed mutations

Six amino acids, N4, D25, N43, D62, S68 and W94, where chosen as targets for replacement by cysteine, figure 2. Cysteine was chosen as the replacement amino acid as native ocr contains no cysteine. The single cysteine introduced into each subunit would therefore be a suitable target for chemical labelling. Each protein was purified in the same manner as the native ocr and each formed a stable dimeric protein as shown by their purification being the same as the wild type protein. It is possible that the single amino acid substitutions made within ocr may perturb the structure of the protein. If the stability of the protein, as determined by thermal denaturation, is unchanged (± 0.5 °C) then it is reasonable to assume that the protein tertiary and quaternary structure is not significantly different from the wild type protein. DSC showed that ocr was a thermostable protein with a denaturation temperature, Tm, of 72.19°C when analysed using a model allowing the dimer to dissociate into folded monomers prior to denaturation of the monomers, table 3. A two state model in which the dimer unfolds in one step to denatured monomers gave a poor fit to the DSC scan (data not shown). Four of the mutant proteins and a second form of truncated ocr containing the first 109 amino acids were close in stability to the wild type protein and required the more complex denaturation model. The mutant protein containing the change W94C was significantly less stable than the wild type indicating a role in maintaining the tertiary and/or quaternary structure of ocr. One mutant protein, D25C, was more stable than the wild type perhaps indicating a better packing of the side chain within the protein structure. The truncated form of ocr with only 99 amino acids was the most unstable compared to the wild type protein suggesting that this truncation has encroached upon secondary structure elements in the core of the protein.

Molecular masses were determined by electrospray mass spectrometry, table 3. The mutant proteins all gave masses greater than expected by 76Da corresponding to the attachment of β-mercaptoethanol to the cysteine residue. This modification can apparently survive the ESMS sample preparation method. The β-mercaptoethanol could be removed by treatment with dithiothreitol which was subsequently removed during the normal sample preparation procedure for the ESMS. The correct mass was then found for the N43C mutant and we assumed that a similar procedure would give the correct mass for the other proteins. The N4C mutant protein also gave rise to a species with the expected dimer mass suggesting that an intersubunit disulphide bridge had been formed. This was confirmed by HPLC gel filtration under conditions in which the protein was denatured, figure 7. Gel filtration under denaturing conditions of the unlabelled mutant proteins showed them all to be monomeric except for the N4C mutant protein which showed two peaks, one being a monomer the other being a dimer. The addition of reducing agent converted the disulphide-linked dimer into a monomer. These results were indicative of an intermolecular disulphide bond between the two cysteine residues at position 4 in each polypeptide chain.

Two different fluorescent labels were chosen for the modification of the cysteine residues, I-AEDANS and pyrene maleimide. All six different positions of the cysteine could be modified by these labels to produce a stable modified protein. The level of modification was approximately 10 to 30%. The ability to label all six locations of the cysteine shows that residues are at least partially exposed to the solvent.

The fluorescence emission from the pyrene labelled ocr was particularly informative. Five of the mutant proteins could be labelled by pyrene maleimide in the absence of β-mercaptoethanol, however, N4C could only be labelled when this reducing agent was also present. A further indication that the N4 positions on each subunit lie in close proximity to each other comes from the appearance of pyrene excimer fluorescence emission, figure 8. This can only arise when two pyrene molecules are within approximately 1nm of each other and the conjugated ring systems can stack on top of each other (Lehrer, 1997). One other mutant protein produced pyrene excimer emission, namely, S68C although this protein could be labelled in the absence of reducing agent and produced only a monomer upon denaturing gel filtration, figure 7, indicating that the sulphydryl groups are too far apart for disulplhide bond formation in the absence of chemical modification. Consistent with this is a shift in the pyrene excimer emission relative to that from the N4C protein suggesting that the interaction between the two pyrene labels is altered. The pyrene excimer emission from both N4C and S68C mutant proteins indicates that both N4 and S68 are on the interface between the two subunits of ocr but, since they can be chemically modified without denaturing the protein, they are not buried in the interface.

Discussion

It has been suggested that ocr, by blocking the DNA binding sites of type I R/M enzymes, is structurally mimicking DNA so that the type I R/M enzyme binds to ocr by mistake (Bandyopadhyay *et. al.*, 1985). This mimicry cannot extend to the fine details of DNA sequence recognition as the ocr protein acts against type I R/M enzymes with completely different DNA target sequences. The high proportion of acidic amino acid side chains in ocr suggested instead that the degree of mimicry might only involve the charge distribution on ocr being similar to that of the phosphate backbone of DNA. If ocr does indeed mimic this charge distribution, it also implies that the protein could be elongated in shape so that it could fill a DNA binding groove on the type I R/M enzyme as effectively as a DNA double helix. However, the conformation of the ocr cannot be a perfect mimic of DNA structure as it would then bind to any DNA binding protein. It is probable that the type I R/M enzyme distorts its DNA target and ocr mimics this distorted structure.

Our results from sedimentation equilibrium indicate that ocr is a dimeric protein in both its normal full-length form, in agreement with previous results (Mark & Studier, 1981), and is still a dimer when the C-terminal 17 amino acids have been deleted to give a monomer with only 99 amino acids. Truncated forms of ocr containing 94 or more amino acids have been shown to be active (Dunn *et. al.*, 1981) showing that the C-terminus is not required for activity. As the protein with 99 aa is still a dimer, this implies that the C-terminus is not involved in holding the protein dimer together.

Further hydrodynamic experiments using sedimentation velocity and dynamic light scattering when combined with our knowledge of the quaternary structure of ocr and assuming typical values of protein hydration allow us to model the general shape of the protein. Most proteins can be modelled reasonably well as spheres or as prolate ellipsoids of revolution. Given the mass of ocr, the hydrodynamic data clearly indicate that ocr is not a sphere. Assuming that it is a prolate ellipsoid, we calculate an axial ratio of between 3.8:1 and 5:1 with a small variation due to protein hydration estimation. Using these axial ratios and the protein mass, we calculate that ocr is approximately 9.7nm to 11.6nm in length and 2.3nm to 2.5nm in diameter. These dimensions would clearly allow the protein to mimic the shape of a double-stranded B-form DNA recognition sequence covering 29 to 33 base pairs at 3.4 Å helical rise per base pair and a helix diameter of 2nm. The truncated form of ocr containing only 99 aa is more globular than the full protein suggesting that the C-terminal regions of each monomer lie at the extreme ends of the elongated dimer.

Deletion of 23 aa amino acids from the C-terminus to remove aa from W94 inclusive has been shown to abolish the ability of ocr to inhibit type I R/M enzymes whereas deletion of 22 aa preserves significant activity (Dunn *et. al.*, 1981). W94 could play a role in either inhibiting the type I enzyme or it could merely be essential for structural stability. Our DSC data indicate that truncation of ocr significantly reduces protein stability as does the mutation W94C within the full length protein. This implies that W94 is definitely required for protein stability but does not rule out a role in activity. However, we have found that the W94C mutation is active *in vivo* and *in vitro* (manuscript in preparation) implying that the nature of the side chain is not important for activity.

The fluorescence experiments on W94 and the chemically modified mutant proteins indicate that the 6 aa chosen for investigation are at least partially exposed to the solvent. The aa substitutions and chemical modification do not cause major structural changes. The W94 position is surrounded by a negatively charged patch which is not part of the acidic C-terminus of ocr. Two of the aa investigated, N4 and S68, are located at the interface between the two monomers with the N4 side chains close enough to form a disulphide bond when substituted with cysteine. Modification of the cysteine substitutions at these two locations with pyrene maleimide produced not only the normal structure fluorescence emission of pyrene, as observed for the other four mutant proteins, but also a distinct unstructured emission at longer wavelengths due to excimer emission from a pyrene - pyrene dimer. This emission only occurs when two pyrene molecules stack closely against each other (Lehrer, 1997) and indicates that the N4 position and the S68 position in ocr must lie at the interface of the two subunits forming the dimer. The different excimer emission maxima for N4C and S68C modified with pyrene suggest that the two locations have different hydrophobicities.

The secondary structure prediction for ocr, ignoring the C-terminal tail which is not predicted reliably due to its predominance of acid residues and is not essential for activity, is largely helical with some short beta strands and loops, figure 2. We suggest that the helices in each subunit may form a helical bundle with loops and beta strands protruding from the bundle. The interface between the two monomers would comprise the ends of the helices so that the structure was elongated. The structure would have two-fold rotational symmetry around the interface. Such a two-fold symmetry would be in accord with the approximate two-fold rotational symmetry observed in type I R/M enzymes (Kneale, 1994). If the helical bundles were arranged so that the two bundles were side by side then a more compact structure would be expected and this would not be in accord with the hydrodynamic data.

A model of the methyltransferase core of a type I R/M enzyme containing the DNA specificity, S, subunit and the two modification, M, subunits has been described (Dryden *et. al.*, 1995) and is shown in figure 9. This model is based upon the similarity of domains within these subunits to known tertiary structures of type II DNA methyltransferases in complex with DNA. The DNA sequence specificity subunit of a type I R/M enzyme contains two target recognition domains, each of which recognises one part of the bipartite DNA target sequence characteristic of type I R/M systems. These domains bind in the major groove of the DNA and show a small degree of sequence and secondary structure similarity to the TRDs of DNA methyltransferases from type II R/M systems (Sturrock & Dryden, 1997). This similarity has received strong support from mutational analyses (O'Neill *et. al.*, 1998). This S subunit is flanked by the modification methyltransferase subunits which are responsible for recognising the methylation status of specific adenine nucleotides within the recognition sequence and carrying out the methylation (modification). The M subunits contain a core catalytic domain with similarity to the catalytic domain of methyltransferases from type II R/M systems (Willcock *et. al.*, 1994). Small angle X-ray scattering data suggests that the M subunits wrap around the DNA (Taylor *et. al.*, 1994). The model assumes that the modification process requires the extrusion of the adenine bases out of the DNA double helix by a base flipping mechanism (Roberts & Cheng, 1998). This mechanism has been demonstrated by crystallography for cytosine methyltransferases and DNA repair enzymes and biochemical evidence supports such a mechanism for DNA methylation by type I R/M enzymes (Mernagh *et. al.*, 1998). Mutational analyses of the model support the general disposition of the DNA sequence recognition domains and the methyltransferase catalytic domains around a model DNA substrate (Willcock *et. al.*, 1994; O'Neill *et. al.*, 1998). A comparison of this model and the shape of ocr derived in this paper is shown in figure 9. It is known that ocr displaces DNA from type I R/M enzymes and the methyltransferase core (Bandyopadhyay *et. al.*, 1985). Our data show that ocr could easily mimic the general structure of approximately 30 base pairs of DNA and fill the DNA binding site of the methyltransferase core of a type I R/M enzyme. The distribution of negative charges on the surface of ocr will obviously be crucial for forming a strong complex with the type I R/M enzyme and one such negatively charged group exists around W94.

The behaviour of the ocr antirestriction protein has an interesting parallel with the small UGI protein of *Bacillus subtilis* phage PBS2 (Beger *et. al.*, 1995). The genome of this phage contains uracil instead of thymine hence on infection of a cell it must inactivate the host uracil glycosylase which is the first enzyme required for uracil excision and repair. The action of UNG would obviously be lethal to the phage and UNG would be the equivalent of a restriction system. UNG is a non sequence specific DNA binding enzyme, the structure of which suggests that the uracil base which it excises is flipped out of the DNA helix in a manner similar to that found for DNA methyltransferases. UGI binds along the DNA binding cleft of both human and Herpes virus UNG and covers the active site (Mol *et. al.*, 1995; Savva & Pearl, 1995; Putnam *et. al.*, 1999). Electrostatic contacts between UNG and UGI appear to mimic the UNG-DNA phosphate contacts.

The amino acid sequence of the ocr protein of T7 has no strong homology to any other protein in current databases even to other known antirestriction proteins such as the ocr protein from phage T3 (Davis & Hyman, 1971) or the ard proteins of conjugative plasmids (Belugorov & Delver, 1995). It appears that phage and plasmids have individually evolved different antirestriction proteins from unknown ancestors to inactivate host type I R/M systems. The physical mode of operation of these other proteins is not as well defined as the T7 ocr which blocks the DNA binding site of type I R/M enzymes by apparently being able to mimic the shape of the DNA target sequence.

Materials and Methods

Ocr over-expression plasmids.

Plasmids pAR2993, pAR3786 and pAR3790 were kind gifts from Dr. A. Rosenberg and Prof. W. Studier (Brookhaven National Laboratory). The full length ocr protein is expressed from pAR2993 while pAR3786 and pAR3790 express ocr containing amino acids 1 to 109 and 1 to 99 respectively. pAR2993 was constructed by ligating the *Alu*I fragment of the T7 genome (base pairs 837-1379) into the *Bam*HI site of pET1 using *Bam*HI linkers of sequence CCGGATCCGG. pAR 3786 was constructed by cutting pAR2993 with *Rsa*I at T7 base pair 1258 and pAR3790 by cutting pAR2993 with *Hga*I at T7 base pair 1222. Appropriate palindromic linkers were added to supply a TAA termination codon and a *Bam*HI site for religation. The sequences of the regions containing the ocr gene or the truncations were determined on both strands by manual sequencing using the Sequenase 2.0 kit (Amersham).

Secondary structure prediction.

In selecting amino acids within the ocr protein for alteration by site-directed mutagenesis, we were directed by predictions of protein secondary structure and solvent exposure produced by the PHD sequence alignment and structure prediction computer program (Rost & Sander, 1993; 1994a; 1994b). The reliability of the prediction is less than expected for this program as the amino acid sequence of ocr is unique in current public sequence databases.

Site-directed mutagenesis.

Site-specific mutations in the coding sequence of the ocr gene were created by PCR mutagenesis using the "QuickChange" site-directed mutagenesis kit (Stratagene) following the manufacturers instructions. The oligonucleotide primers used and the amino acid change expected are given in table 4. The DNA sequences of the altered plasmids were checked using the Sequenase kit as for plasmid pAR2993.

Protein production.

The proteins were produced by transformation of *E.coli* strain BL21 (DE3) pLysS with the plasmid pAR2993 or its derivatives. Cells were grown at 37°C in 10 x 500ml of LB broth supplemented with 20μg/ml ampicillin and 25μg/ml chloramphenicol. Once the optical density reached a value of 0.6 at 650nm, the cells were induced by the addition of 5ml 0.4mM isopropyl β-D-thiogalactoside to each flask. The cells were harvested after a further 2 hours. The cell paste was resuspended in 20mM Tris, 0.3M NH_4Cl, 4% glycerol, 10μM phenylmethylsulphonyl fluoride and 10μM benzamidine, pH 8, and sonicated in bursts with intermittent cooling for a total of 1 minute per gram of cell paste with the sample kept on ice. The cell debris was removed by centrifugation for 3 hours at 30,000g. The supernatant was loaded onto a 20cm x 1.6 cm diameter DEAE sepharose fast flow ion exchange column (Pharmacia) equilibrated in 20mM Tris, 300mM NH_4 Cl at pH 8 (buffer A) at 48ml/hr. After washing the column in this buffer to remove unbound proteins, a 500ml gradient from 0.3M to 1M NH_4Cl in buffer A was run at 24ml/hr. Fractions identified by SDS-PAGE as containing ocr were applied to a HiLoad 16/60 Superdex 200 gel filtration column (Pharmacia) in buffer A. This step produced very pure (99%) protein. A slight contamination of nucleic acid could be removed by a subsequent step using a 15cm x 1.6cm diameter phenyl sepharose column (Pharmacia) with a gradient running down from 2.4M to 0M $(NH_4)_2SO_4$ in a buffer of 20mM Tris, pH8. The purified protein was stored at -20°C in buffer A supplemented with glycerol to 50% v/v. Similar procedures were used to purify the mutant forms of ocr although the gradient used for the DEAE chromatography commenced at 0.05M NH_4Cl.

Calculation of extinction coefficient and measurement in guanidinium hydrochloride (GuCl) versus buffer.

The extinction coefficients for pure ocr and its derivatives were calculated by comparing the absorbance of folded native protein and protein denatured in GuCl (Pace & Schmidt, 1997). The concentration of the GuCl stock solution was calculated from the equation

$$M=57.147\Delta N + 38.68(\Delta N)^2 - 91.60(\Delta N)^3 \qquad (1)$$

where ΔN is the refractive index difference between the GuCl solution and the buffer (Pace & Scholtz, 1997). A molar extinction coefficient of 31095 M^{-1} cm^{-1} at 280nm for folded ocr and 31860 M^{-1} cm^{-1} for unfolded ocr were determined. Extinction coefficients for the mutant forms of the proteins were calculated from that of folded wild type ocr accounting for changes in mass and the loss of tryptophan in the W94C variant.

Chemical modification of mutant proteins.

Modification of cysteine with 1,5 I-AEDANS was carried out as follows. 100μl of 50μM protein was incubated overnight at 4°C in the dark with a 50-fold molar excess of N-iodoacetyl-N-(5-sulpho-1-napthyl)ethylenediamine (1,5-I-AEDANS, Fluka) in 20mM Tris, 20mM NH_4Cl, 6mM $MgCl_2$, pH8. The ocr protein containing the N4C mutation could only be labelled in the presence of 7mM β-mercaptoethanol. The reaction was stopped by the addition of β-mercaptoethanol to 100mM and unreacted 1,5--AEDANS removed by gel filtration using a Sephadex G50 column (Pharmacia PD10) equilibrated in the same buffer. The amount of labelling was determined using molar extinction coefficients of 6100M^{-1} cm^{-1} at 337nm and 1060 M^{-1} cm^{-1} at 280nm for AEDANS (Hudson & Weber, 1973). The fluorescence of AEDANS was excited at 330nm and detected at 480nm.

Labelling of cysteine with pyrene maleimide was performed as follows. N-(1-pyrene)maleimide (Molecular Probes) was dissolved in dimethyl sulphoxide. Labelling of ocr proteins was performed as for I-AEDANS except the buffer was 100mM sodium phosphate, pH7. The degree of labelling was determined using an extinction coefficient of 40000 M^{-1} cm^{-1} at 342nm and 25000 M^{-1} cm^{-1} at 280nm (Wu *et. al.*, 1976). Fluorescence was excited at 342nm.

Analytical ultracentrifugation

Sedimentation equilibrium and velocity analysis was performed on a Beckman Optima XL-A analytical ultracentrifuge equipped with scanning absorbance optics tuned to a wavelength of 280 nm. Samples were suspended in 20mM Tris HCl, 0.3 M NH_4Cl, pH 8.0 which was calculated to have a density of 1.0040 g/ml at 20 °C and 1.0056 g/ml at 4 °C using the freeware program SEDNTERP available from ftp://alpha.bbri.org/rasmb/spin/ms_dos/sednterp (Philo, 1994). The value of the partial specific volume (v) was estimated, using the method of Cohn and Edsall (1943) and the consensus partial specific volumes for the constituent amino acids reported by Perkin (1986), from the known amino acid sequence of ocr and the truncated ocr containing only the first 99 amino acids. The partial specific volumes were calculated to be 0.721ml/g and 0.723 ml/g respectively and were assumed to remain constant upon oligomerisation.

Sedimentation equilibrium was attained for solution columns of height 1.5 mm and ascertained by ensuring overlay of scans acquired 3 hours apart. Optical baselines, depleted of macromolecular solute, were obtained by overspeeding to 47,000 rpm. Nine concentrations of ocr from 2.0 to 13.1 μM and eight of the 99 amino acid truncated form of ocr from 2.6 - 15.4 μM were analysed. Equilibrium distributions for the truncated ocr were obtained at rotor speeds of 25,000 and 32,000 rpm and a temperature of 20 °C. The full length ocr was found to be unstable at this temperature, failing to reach equilibrium during a reasonable period of time. Instead, this protein was equilibrated at 4 °C and equilibrium at 25,000 rpm was attained within 30 hours. Equilibrium was not subsequently attained at the higher rotor speed for reasons which are readily apparent.

Equilibrium data were analysed by fitting with the equation describing the solute distribution in a molecular system comprising self-interacting species at thermodynamic equilibrium (Kim *et. al.*, 1977):

$$C_{total}(r) = \delta c + \sum K_{1,i} * (C_{1,0} * \exp[\sigma * \{\xi - \xi_0\} - 2B\sum C_i(r)])^{q(i)} \qquad (2)$$

where $C_{total}(r)$ is the concentration at any radial position in the sample column, δc is proportional to the baseline absorbance; n is the total number of species present in the self-association equilibrium; $K_{1,i}$ is the equilibrium constant for the association of monomer to q(i)-mer; $C_{1,0}$ is the monomer concentration at the radial position of the first data point; σ is the reduced molecular weight ($\sigma = (M\omega^2(1-\nu\rho))/RT$ where M is the monomer mass; ω is the rotor speed in radian/s; ν is the partial specific volume of the protein (ml/g); ρ is the solvent density (g/ml); R is the gas constant (8.31432 x 10-7 erg/mol K) and T is the absolute temperature (K)); ξ and ξ_0 are the values of $0.5r^2$ at any point r in the sample column and at the radial position of the first data point respectively; B is the second virial coefficient (with units of inverse concentration); $C_i(r)$ is the concentration of the ith species at radial position r; q(i) is the degree of association for the ith species. The summations are from i = 1 to n. This equation forms the basis of the program PCNONLIN (Johnson *et. al.*, 1981) which permits the global fitting of up to 15 data sets with one self-association model.

Non-ideal behaviour of the protein was observed at higher concentrations. This behaviour is characterised by the second virial coefficient which describes the reduction in observed mass due to excluded volume and charge repulsion effects. A simplified expression for the complete chemical potential of a solution of charged spherical macromolecules is the sum of the excluded volumes due to shape and charge (Scatchard, 1946),

$$BM = 4\nu + (1000Z^2\nu_1) / (4mM) \qquad (3)$$

where M is the mass of the solute, ν_1 is the specific volume of the solvent and m the molality of salt in the solvent. For rod-shaped particles of length L and width d, the first shape-derived term becomes (L/d)ν. From theoretical pH titration curves (see for example http://www.up.univ-mrs.fr/~wabim/d_abim/compo-p.html) one can calculate the net charge on the protein. The expression for the depression in true molecular mass, M, to the apparent value, M_{app}, measured at sedimentation equilibrium is

$$M_{app} = M / (1+2BMc) \qquad (4)$$

where c is the solute concentration (g/ml).

Sedimentation velocity data were acquired at 40,000 rpm by scanning the 10 mm solution column every 10 minutes. Data were analysed with the program SVEDBERG (Philo, 1994). The boundaries recorded during the sedimentation velocity experiment were fitted with a modified form of the Fujita (Fujita, 1975) equation for the concentration throughout the solution column. This analysis is valid for interacting macromolecular systems only if the kinetics of inter-conversion among species is slow compared to the experimental timescale. The sedimentation equilibrium experiments showed that OCR exists as a dimer, with no trace of monomer present so analysis with SVEDBERG is appropriate.

Dynamic light scattering

Dust contamination, a major problem in DLS experiments, was overcome as follows. Firstly, the 20mM Tris, 0.3M NH_4Cl pH 8buffer was filtered through several 0.2 µm filters and used to rinse the 200µl cylindrical glass DLS cell and a microlitre syringe. The protein solution, made up with the filtered buffer, was filtered once more through a 0.2 µm filter and spun in a microcentrifuge for 45 minutes at 14,000 rpm to force any remaining dust particles to the bottom of the sample. The syringe was used to extract 100µl from the top of the centrifuged solution immediately after spinning and this volume was transferred into the pre-rinsed DLS cell. The cell was then sealed with parafilm. A further small amount of the centrifuged protein solution was extracted from the top of the microcentrifuge tube and used to determine the concentration of the sample by UV absorption. The concentration of protein in the sample was 100µM. The DLS sample was kept at 12°C when not in use and spun for 60 minutes at 6500 rpm at this temperature before each DLS experiment. When this solution was diluted for measurements at lower concentration, another 100 µl of filtered buffer was added with the syringe to the original sample, yielding a protein concentration of 50µM.

DLS measurements were performed using a 514.5 nm wavelength argon ion laser operating at 180 mW and an ALV-5000 correlator. The temperature of the scattering cell and toluene bath during measurements was 19+/-0.3°C. DLS measurements of the autocorrelation function (g2) at both protein concentrations were performed for angles between 30° and 105° in 5° increments. At each angle, five data sets each of five minutes duration were collected and each set analysed individually. Data analysis was performed using the ALV-800 transputer and ALV-5000 software package by second order cumulant fitting, CONTIN fitting and inverse Laplace transform (ILT) (Berne & Pecora, 1976; Ostrowsky *et. al.*, 1981; Provencher, 1982; Stepanek, 1993). All these methods were in agreement with each other, although the ILT method provided the most consistent analysis and results from this method are presented here. The polydispersity index, V(Q)=Avg{decay rate^2}/(Avg{decay rate^2}-1), supplied by the ALV-5000 second order cumulant fitting was used to determine if a particular data set was affected by dust or noise. Sets with a polydispersity index of greater than 0.12 were disregarded.

The absolute scattering intensity average was also recorded for each data set and compared with measurements of the absolute scattering intensity at the same angle and laser power by the toluene bath. While the ratio of the two intensities should be constant, anomalies were occasionally observed due to extra scattering from the scattering cell. When this was observed, the data were discarded as heterodyne scattering could be occurring. As we wish to analyse homodyne scattering, heterodyne scattering would confuse the analysis process.

The numerical results presented in this paper are from a set of four scattering angles between 55^{o} and 70^{o}. The numerical results below 50^{o} and above 75^{o} agree well with the presented data, however, the statistical variation due to the correlator electronics at large angles and extra correlation decays due to dust at small angles were noticeable at these extreme angles. As it is technically difficult to completely eliminate these effects from influencing the analysis, the data from these extreme angles were left out of the final calculations.

Hydrodynamic modelling of spheres and ellipsoids of revolution.

Theory decribes the relation of hydrodynamic parameters to molecular mass and shape (Haschemeyer & Haschemeyer, 1973; Durschlag *et. al.*, 1996; Durschlag & Zipper, 1999). The relationship between sedimentation coefficient (s) and frictional coefficient (f) is

$$f = M(1-v\rho) / (N_A s) \qquad (5)$$

where N_A is Avogadro's number and M is the mass of the macromolecule.
The relationship between diffusion coefficient (D) and frictional coefficient is

$$D = kT/f \qquad (6)$$

where k is Boltzmann's constant and T is the temperature (K).
The frictional coefficient of an anhydrous sphere of radius r_o is

$$f_o = 6\pi\eta r_o \qquad (7)$$

where η is the solvent viscosity in Poise. The frictional ratio (f/f_o) represents the deviation of the macromolecule from an anhydrous sphere which has $f/f_o = 1.0$. A typical protein hydration (δ) lies between 0.2 to 0.4 g water / g protein and the frictional ratio can be expressed as the product of the frictional ratio due to hydration $(f/f_o)_{hyd}$ and due to molecular asymmetry $(f/f_o)_{sh}$.

$$f/f_o = (f/f_o)_{sh} \cdot (f/f_o)_{hyd} = (f/f_o)_{sh} \cdot (1+\delta/v\rho)^{1/3} \qquad (8)$$

This allows the calculation of $(f/f_o)_{sh}$ at a range of assumed values of hydration. The OCR molecule was modelled as a prolate ellipsoid of revolution. Determining an axial ratio (p) from $(f/f_o)_{sh}$ is done using an equation originally derived by Perrin for a prolate ellipsoid of revolution (Haschemeyer & Haschemeyer, 1973; Berne & Pecora, 1975; Durschlag *et. al.*, 1996; Durschlag & Zipper, 1999)

$$(f/fo)_{sh} = [(p^2 - 1)^{1/2}] / [p^{1/3} \cdot \ln\{p + (p^2 - 1)^{1/2}\}] \qquad (9)$$

Fluorescence spectroscopy.

Steady state fluorescence spectroscopy measurements were performed on Perkin-Elmer LS50B or Edinburgh Instruments FS900 fluorimeters. The rotational diffusion coefficient of ocr was estimated using Perrin plots (Weber, 1952; Weltman & Edelman, 1967; Lakowicz, 1999). 1µM samples of ocr were prepared in 20mM Tris, 20mM NH_4Cl, 6mM $MgCl_2$, 7mM β-mercaptoethanol, pH8 and glycerol added by weight to 0%, 30%, 60% and 90% w/w. The fluorescence anisotropy was determined using an excitation wavelength of 295nm and emission at 350nm with slitwidths of 3.6nm and 8.6nm respectively. The cuvette had a 10mm excitation pathlength and a 4mm emission pathlength. The anisotropy was determined as a function of temperature and glycerol by measuring fluorescence intensity with vertically polarised excitation and vertically or horizontally polarised emission and correcting for the instrumental G factor. The G factor was measured with horizontally polarised excitation. The viscosity, η, of the solutions was determined from tabulated data on glycerol/water solutions (Segur, 1953). The data were plotted as 1/(anisotropy) against T/η and fitted to two sets of straight lines, firstly assuming that the data collected at constant temperature and high values of T/η all had the same slope and, secondly, that data collected at constant glycerol had the same intercept on the 1/(anisotropy) axis. These data can be analysed in terms of two rotational times, one time for internal rotation of the fluorophore relative to the rest of the protein and the second time for overall tumbling of the protein. This analysis requires the fluorophore to be small relative to the rest of the protein, a situation which applies to the tryptophan residue in ocr. An estimate of the extent of motion of the fluorophore can be also be made in this case. At constant temperature

$$r/r_a = [f/\{1+(\tau/\phi_s)\} + (1-f)/\{1+(\tau/\phi_f)\}] \qquad (10)$$

where r_a is the intrinsic anisotropy of the fluorophore obtained from the intercept of the lines fitted to constant glycerol concentration data, f is the fraction of anisotropy associated with slow rotation, τ is the average fluorescence lifetime of the fluorophore, ϕ_S is the rotational correlation time for the slow overall tumbling of the protein assuming it is spherical or the harmonic mean rotational correlation time if it is not spherical, and ϕ_f is the rotational correlation time for motion of the fluorophore relative to the rest of the protein. From the values of the intercepts of the constant temperature data, $1/r_{int}$, one can calculate the degree of fluorophore wobbling within in a cone of semi-angle Θ at a particular temperature from

$$1/r_{int} = (1/r_a).(2/\{3\cos^2\Theta - 1\}) \qquad (11)$$

The value of rotational correlation time, ϕ_S, for overall tumbling of the hydrated protein can be compared to that for a hydrated sphere of the same mass and volume at a chosen pair of values for viscosity and temperature from the equation,

$$\phi_{sphere} = \eta V/kT \qquad (12)$$

Fluorescence lifetime measurements for the decay of the emission from Trp94 of ocr were performed using an Edinburgh Instruments 199T spectrofluorimeter. Excitation was at 295nm, bandwidth 20nm, with a hydrogen flashlamp operating at a repetition rate of 65kHz in 0.5 atmospheres of H_2. Emission was collected through a 335nm cutoff filter. The sample was in a 10mm x 4mm cuvette in 20mM Tris, 0.3M NH_4Cl pH 8 buffer at approximately 12μM concentration. Excitation was through the 4mm pathlength side of the cuvette. The lamp profile was collected at 295nm using a dilute solution of Ludox colloidal silica suspension. The half-width of the lamp pulse was around 1.8ns. Data were fitted using multiexponential fitting procedures incorporating lamp profile deconvolution (Birch & Imhof, 1985).

Quenching of tryptophan fluorescence was performed to estimate the degree of solvent exposure of this residue. Fluorescence was quenched by CsCl or KI in 400μl solutions of 0.0185mg/ml ocr protein in 20mM Tris, pH8 at 20°C. The ionic strength of the solutions was kept constant at 0.4M by adding NH_4Cl to each solution. The solutions containing KI were supplemented with 2mM $Na_2S_2O_3$ to prevent the formation of the I_3^- ion. Each sample was excited at 295nm and emission measured at 350nm with 10mm and 2mm pathlengths respectively. The Stern-Volmer equation was used to fit the fluorescence intensity (Eftink & Ghiron, 1981). The equation is

$$F/F_0 = \Sigma f_i / (1+K_iQ) \qquad (13)$$

where the summation is over the number of fluorescent species, i, F is the fluorescence intensity at quencher concentration Q, F_0 is the fluorescence intensity in the absence of quencher and K_i is the Stern-Volmer quenching constant for component i.

Differential scanning calorimetry

Proteins were desalted into 20mM Tris, 20mM NH_4Cl, 6mM $MgCl_2$, pH8 and their concentration determined by UV absorption. The samples were degassed before loading 0.75ml into VP-DSC microcalorimeter. The temperature was raised from 20°C to 100°C at a scan rate of 1°C per minute. Subsequent cooling and reheating of the samples gave the same denaturation profiles indicating a reversible folding transition. The data were analysed using the Origin 5.0 program.

Analytical gel filtration chromatography

Analytical HPLC gel filtration chromatography was performed as described (Dryden *et. al.*, 1997) using 50μl of 20μM ocr in 20mM Tris, 20mM 2-[N-morpholino]ethanesulphonic acid, 0.2M NH_4Cl, 10mM $MgCl_2$, 0.1mM EDTA, 5M GuCl, pH6.5. β-mercaptoethanol, if present, was at a concentration of 7mM. Eluting proteins were detected by UV absorption at 280nm and fluorescence emission at 350nm with excitation at 280 nm.

Mass spectroscopy

10μg of protein was desalted on an Applied Biosystems 130A Microbore Separation System with an aquapore RP-300 30mm x 2.1 mm diameter microbore column eluted at 0.2 ml/min with a 7 to 70% v/v gradient of acetonitrile in 0.1% v/v trifluoroacetic acid. 50μl of this desalted protein was mixed with 200μl of 60% methanol in 1% v/v formic acid and infused by syringe at a flow rate of 3μl/min into a Thermoquest LCQ mass spectrometer. The spectra collected in full scan mode were deconvoluted using BioExplore software.

Ackowledgements

We are grateful to the Leverhulme Trust (grant F/158/BC) for providing funding for this work and supporting S. Sturrock. C. Atanasiu was supported by a Darwin Trust studentship and J. Blackstock by a summer studentship from the Wellcome Trust. H. McMiken contributed to this work as part of her undergraduate degree. D. Dryden thanks the Royal Society for a University Research Fellowship. We also thank Professor Alan Cooper and Mr. Gordon Campbell (Glasgow) for advice on calorimetry and assistance with ultracentrifugation experiments respectively, Dr. Andy Cronshaw (Edinburgh) for running the esms samples and Professor Steve Harding (Nottingham) for useful discussions concerning ellipsoidal modelling. Finally we thank Professor William Studier and Dr. Alan Rosenberg (Brookhaven National Laboratory) for generously supplying the ocr overexpression plasmids.

References

Bandyopadhay, 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 .

Beger, R.D., Balasubramamian, S., Bennett S.E., Mosbaugh, D.W. & Bolton, P.H. (1995). Tertiary structure of uracil-DNA glycosylase inhibitor protein. *J. Biol. Chem.* **270**, 16840-16847.

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. *Nucleic Acids Res.*, **23**, 785-787.

Berne, B. & Pecora, R. (1976). Dynamic Light Scattering with applications to Chemistry, Biology and Physics. John Wiley & Sons Inc., London.

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

Birch, D.J.S. & Imhof, R.E. (1985). Kinetic interpretation of fluorescence decays. *Anal. Instrum.*, **14**, 293-329.

Cohn, E.J., & Edsall, J.T. (1943). Proteins, Amino Acids and Peptides as Ions and Dipolar Ions, Reinhold Publishing Corp., New York.

David, R.W. & Hyman, R.W. (1971). A study in evolution: the DNA base sequence homology between coliphage T7 and T3. *J. Mol. Biol.*, **62**, 287-301.

Dryden D. T. F., Sturrock S. S. and Winter M. (1995). Structural modelling of a type I DNA methyltransferase. *Nature Struct. Biol.*, **2**, 632-635.

Dryden, D.T.F., Cooper, L.P., Thorpe, P.H., Byron, O. (1997). The *in vitro* assembly of the *Eco*KI type I DNA restriction/modification enzyme and its *in vivo* implications. *Biochemistry* **36**, 1065-1076.

Dunn, J.J., Elzinga M., Mark, K.K. & Studier, F.W. (1981). Amino-acid-sequence of the gene 0.3 protein of bacteriophage-T7 and nucleotide sequence of its messenger-RNA. *J. Biol. Chem.* **256**, 2579-2585.

Durchschlag H. & Zipper, P. (1999). Calculation of Structural Parameters from hydrodynamic data. *Progress in Colloid & Polymer Science*, **113**, 87-105.

Durchschlag, H., Zipper, P., Purr, G. & Jaenicke, R. (1996). Comparative studies of structural properties and conformational changes of proteins by analytical ultracentrifugation and other techniques. *Colloid & Polymer Science*, **274**, 117-137.

Eftink, M.R. & Ghiron, C.A. (1981). Fluorescence quenching studies with proteins. *Anal. Biochem.*, **114**, 199-227.

Fujita, H. (1975). Foundations of Ultracentrifugal Analysis, Wiley, New York.

Garcia, L.R. & Molineux, I.J. (1995). Rate of translocation of bacteriophage T7 DNA across the membranes of *Escherichia coli*. *J. Bacteriol.*, **177**, 4066-4076.

Haschemeyer, R.H. & Haschemeyer. A.E.V. (1973). Proteins: a guide to study by physical and chemical methods. J. Wiley & Sons, New York.

Hausmann R. (1988) The T7 group. In *The Bacteriophages*, vol. **1**. Plenum Press ed. R. Callendar, p259-289. New York.

Hudson, E.N. & Weber, G. (1973). Synthesis and characterisation of two fluorescent sulphydryl reagents. *Biochem.* **12**, 4154-4161.

Johnson, M.L., Correia, J.J., Yphantis, D.A., & Halvorson, H.R. (1981). Analysis of data from the analytical ultracentrifuge by nonlinear least squares techniques. *Biophys. J.*, **36**, 575-588.

Kim, H., Deonier, R.C., & Williams, J.W. (1977). The investigation of self association reactions by equilibrium ultracentrifugation. *Chem. Rev.*, **77**, 659-690.

Kneale, G.G. (1994). A symmetrical model for the domain structure of Type I DNA methyltransferases. *J. Mol. Biol.*, **243**, 1-5.

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., Hansen, S. & Reuter. M. (1983). The ocr$^+$ gene function of bacteriophages T3 and T7 counteracts the *Salmonella typhimurium* DNA restriction systems SA and SB. *J. Virol.*, **45**, 1147-1149.

Kruger, D.H., Schroeder, C., Hansen. S. & Rosenthal, H.A. (1977). Active protection by bacteriophages T3 and T7 against. *E.coli* B- and K-specific restriction of their DNA. *Molec. Gen. Genet.*, **153**, 99-106.

Kruger, D.H., Schroeder, C., Reuter. M., Bogdarina, I.G., Buryanov, Y.I. & Bickle, T.A. (1985). DNA methylation of bacterial viruses T3 and T7 by different methylases in *Escherichia coli* K12 cells. *Eur. J. Biochem.*, **150**, 323-330.

Lakowicz, J.R. (1999). Principles of fluorescence spectroscopy. 2nd. edition. Kluwer Academic / Plenum Publishers, New York.

Laue, T.M., Shah, D.D., Ridgeway. T.M., & Pelletier, S.L. (1992). Computer-aided interpretation of analytical sedimentation data for proteins. In Analytical Ultracentrifugation in Biochemistry and Polymer Science (Harding, S.E., Rowe, A.J., and Horton, J.C., Eds.) pp 90-125, Royal Society of Chemistry, Cambridge, UK.

Lehrer, S.S. (1997). Intramolecular pyrene excimer fluorescence: A probe of proximity and protein conformational change. *Meth. Enzymol.*, **278**, 286-295.

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

Mernagh, D.R., Taylor, I.A., Kneale. G.G. (1998). Interaction of the type I methyltransferase M.*Eco*R124I with modified DNA substrates: sequence discrimination and base flipping. *Biochem. J.*, **336**, 719-725.

Moffatt, B.A. & Studier, F.W. (1988). Entry of bacteriophage T7 DNA into the cell and escape from host restriction. *J. Bacteriol.*, **170**, 2095-2105.
Mol, C.D., Arvai, A.S., Sanderson, R.J., Slupphaug, G., Kavli, B., Krokan, H.E., Mosbaugh, D.W. & Tainer, J.A. (1995). Crystal-structure of human uracil-DNA glycosylase in complex with a protein inhibitor - protein mimicry of DNA. *Cell* **82**, 701-708.
Murray, N.E. Microbiol. (2000). Type I restriction systems: sophisticated molecular machines. *Mol. Biol. Rev.*, **64**, 412-434.
O'Neill, M., Dryden, D.T.F., Murray, N.E. (1998). Localization of a protein-DNA interface by random mutagenesis. *EMBO J.* **17**, 7118-7127.
Ostrowsky, N., Sornette, D., Parker, P. & Pike, E.R. (1981). Exponential sampling method for light scattering polydispersity analysis. *Acta Opt.*, **28**, 1059-1070.
Pace, C.N. & Schmidt, F.X. (1997). How to determine the molar absorbance coefficient of a protein. In Protein structure; a practical approach. (Creighton, T.E. Ed.). IRL Press pp 253-259. Oxford.
Pace, C.N. & Scholtz, J.M. (1997). Measuring the conformational stability of a protein. In Protein structure; a practical approach. (Creighton, T.E. Ed.). IRL Press pp 299-348. Oxford.
Perkins, S.J. (1986). Protein volumes and hydration affects. *Eur. J. Biochem.*, **157**, 169-180.
Philo, J.S. (1994). Measuring sedimentation, diffusion and molecular weights of small molecules by direct fitting of sedimentation velocity concentration profiles. In Modern Analytical Ultracentrifugation (Schuster, T. M., and Laue, T. M., Eds.) pp 156-170, Birkhauser, Boston, Basel, Berlin.
Provencher, S.W. (1982). A constrained regularization method for inverting data represented by linear algebraic or integral equations. *Computer Physics Comm.*, **27**, 213-227.
Putnam, C.D., Shroyer, M.J.N., Lundquist, A.J., Mol, C.D., Arvai, A.S., Mosbaugh, D.W. & Tainer, J.A. (1999). Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. *J. Mol. Biol.*, **287**, 331-346.
Roberts, R.J. & Cheng, X. (1998). Base flipping. *Annu. Rev. Biochem.*, **67**, 181-198.
Rost, B. & Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584-599.
Rost, B. & Sander, C. (1994a). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*, **19**, 55-72.
Rost, B. & Sander, C. (1994b). Conservation and prediction of solvent accessibility in protein families. *Proteins*, **20**, 216-226.
Savva, R. & Pearl, L.H. (1995). Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. *Nature Stuct. Biol.*, **2**, 752-757.
Scatchard, G. (1946). Physical chemistry of protein solutions. I. Derivation of the equations for the osmotic pressure. *J. Am. Chem. Soc.*, **68**, 2315-2319.
Segur, J.B. (1953). Physical properties of glycerol and its solutions. In Glycerol (Miner, C.S. & Dalton, N.N. eds). American Chemical Society monographs volume 117, pp 238-334. Reinhold Publishing Corp. New York.
Stepanek, P. (1993). Data analysis in dynamic light scattering. *Monographs on the Physics and Chemistry of Materials*, **49** "Dynamic Light Scattering: The Method and Some Applications" Clarendon Press, Oxford. 177-240.
sturrock
Sturrock, S.S., Dryden, D.T.F. (1997). A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes *Nucleic Acids Res.* **25**, 3408-3414.
Taylor, I.A., Davis, K.G., Watts, D., Kneale, G.G. (1994). DNA binding induces a major structural transition in a type I methyltransferase. *EMBO J.* **13**, 5772-5778.
Weber, G. (1952). Polarisation of the fluorescence of macromolecules. I. Theory and experimental method. *Biochem. J.*, **51**, 145-155.
Weltman, J.K. & Edelman, G.M. (1967). Fluorescence polarization of human γG-immunoglobulins. *Biochem.*, **6**, 1437-1447.
Willcock, D.F., Dryden, D.T.F. & Murray, N.E. (1994). A mutational analysis of the 2 motifs common to adenine methyltransferases. *EMBO J.*, **13**, 3902-3908.
Wilson, G.G., Murray, N.E. (1991). Restriction and Modification Systems. *Annu. Rev. Genet.*, **25**, 585-627.
Wu, C.W., Yarbrough, L.R. & Wu, F.Y.H. (1976). N-(1-Pyrene)maleimide: a fluorescent cross-linking reagent. *Biochem.*, **15**, 2863-2868.

Figure Legends

Figure 1.

The type I R/M enzyme (oval) normally binds to DNA (long rectangle) and recognises its DNA target sequence. If this sequence is unmethylated on both strands then the restriction reaction is triggered resulting in the cleavage of the DNA. The presence of the ocr protein (small rectangle) prevents DNA binding by the enzyme and the DNA is not cleaved.

Figure 2.

The amino acid sequence of the ocr protein and the predicted secondary structure apart from the C-terminal end which has a biased amino acid composition making structure prediction unreliable. Helices and strands are indicated by H and B respectively and the reliability of the prediction is indicated residue by residue with 9 being most reliable. Predictions

of whether a residue is buried (b) or exposed (e) to solvent are also given. The locations of the amino acids selected for substitution with cysteine have a * placed above them. The N-terminal methioine has been removed.

Figure 3a.

Reciprocal of apparent whole-cell weight average molecular mass of wild type ocr (Δ) acquired at 25,000 rpm and the truncated ocr with only the first 99 aa (o) at 25,000 rpm and (●) 32,000 rpm respectively as a function of loading concentration in sedimentation equilibrium studies. Least-squares fits to the data yield values for the whole-cell weight average molecular mass at infinite dilution of $M^0_{w,app}$= 29.4 ± 0.3 kDa for wild type ocr at 25,000 rpm (---) and $M^0_{w,app}$ = 24.7 ± 0.2 kDa and 23.9 ± 0.2 kDa for the truncated ocr at 25,000 rpm (-) and 32,000 rpm (––) respectively. The gradients of these fits yield the second virial coefficient (see text).

Figure 3b.

Apparent weight-average sedimentation coefficient as a function of loading concentration in sedimentation velocity studies for wild type ocr (●) and truncated 99 aa ocr (o). Extrapolation to infinite dilution of the least-squares fit to these data yields sedimentation coefficients of 2.5 ± 0.1 S and 2.2 ± 0.1 S respectively.

Figure 4.

Homodyne autocorrelation functions of representative ocr dynamic light scattering data sets taken at a scattering angle of 60^0. The data clearly depict a single decay process free from dust and other particulate contamination.

Figure 5.

Variation of the anisotropy of tryptophan emission as a function of temperature and viscosity. Data collected at the same temperature have the same symbol (Δ, 45 °C; ■, 35 °C; ❑, 25 °C; ●, 15 °C; o, 10 °C) and are fitted to a set of parallel lines with intercepts $1/r_{int}$ on the y axis. Points collected at the same glycerol concentration (90%, 60%, 30% and 0% from left to right) are joined point to point. Extrapolation of all sets of constant glycerol data to one fixed intercept on the y axis gives the value of $1/r_a$, the intrinsic anisotropy of the fluorophore.

Figure 6.

Stern-Volmer plot of the fluorescence quenching of wild type ocr by Cs^+ (o) and I^- (●) at a constant ionic strength of 0.4M. The Cs^+ data were fitted to a two-component quenching mechanism and the I^- data to a one component mechanism.

Figure 7.

HPLC denaturing gel filtration for the N4C and S68C mutant proteins. In the absence (top trace) of β-mercapoethanol, the S68C protein elutes in a single peak characteristic of the unfolded protein monomer. The protein containing the N4C substitution elutes as two peaks in the absence of β-mercaptoethanol (bottom trace), corresponding to unfolded monomer and unfolded dimer joined covalently by a disulphide bond, and predominantly as a single peak in the presence of the reducing agent (middle trace). Elution was monitored at 280 nm.

Figure 8.

Normalised pyrene emission spectra for chemically-modified ocr containing the N4C and S68C substitutions compared to one of the other mutant proteins, N43C. All pyrene-modified mutant proteins except N4C and S68C gave the same emission spectra. The presence of a broad emission band at long wavelengths for the N4C and S68C proteins indicates the formation of an excimer due to close proximity of the two pyrene moities across the monomer-monomer interface in the ocr dimer.

Figure 9.

The size and asymmetry predicted for ocr, with a hydration of 0.3g/g, by hydrodynamic and light scattering measurements is shown as an yellow ellipsoid of axial ratio 4.3:1 superimposed on top of the structural model of the methyltransferase core of a type I R/M enzyme. This model contains partial structures of two M subunits (blue spacefilling) and one S subunit (grey spacefilling) bound to 24 base pairs of DNA (red wireframe). It is apparent from these views that the ocr is approximately the correct width and length to fill the DNA binding site of the methyltransferase. a, front view; b, top view.

Tables

1. Hydrodynamics of truncated 99 aa ocr.

		Ultracentrifugation analysis			
Hydration $g_{H2O}/g_{protein}$	$(f/f_o)_H$	$(f/f_o)_{sh}$ [a]	a/b	a (nm)	b (nm)
0.2	1.085	1.226 ± 0.09	4.64 ± 1.00	5.22	1.12
0.3	1.123	1.185 ± 0.09	4.03 ± 1.17	4.76	1.18
0.4	1.158	1.148 ± 0.09	3.50 ± 1.31	4.33	1.24

a. $f/f_o = 1.33 \pm 0.11$ for ultracentrifugation, $v = 0.723$ ml/g, $f_o = 3.536 \times 10^{-9}$ kg/s

2. Hydrodynamics for wild type ocr.

		Ultracentrifugation analysis				Dynamic light scattering analysis				Calculated rotational correlation time
Hydration $g_{H2O}/g_{protein}$	$(f/f_o)_H$	$(f/f_o)_{sh}$ [a]	a/b	a (nm)	b (nm)	$(f/f_o)_{sh}$ [b]	a/b	a (nm)	b (nm)	ϕ_{sphere} (ns)
0.2	1.085	1.244 ± 0.05	4.91 ± 0.71	5.74	1.17	1.249 ± 0.008	4.990 ± 0.115	5.80	1.16	10.1
0.3	1.123	1.202 ± 0.05	4.29 ± 0.73	5.24	1.22	1.207 ± 0.008	4.360 ± 0.120	5.30	1.21	11.2
0.4	1.158	1.165 ± 0.05	3.75 ± 0.72	4.79	1.28	1.170 ± 0.008	3.822 ± 0.110	4.85	1.27	12.3

a. $f/f_o = 1.35 \pm 0.06$ for ultracentrifugation, b. $f/f_o = 1.355 \pm 0.008$ for dynamic light scattering, $v = 0.721$ ml/g, $f_o = 3.741 \times 10^{-9}$ kg/s

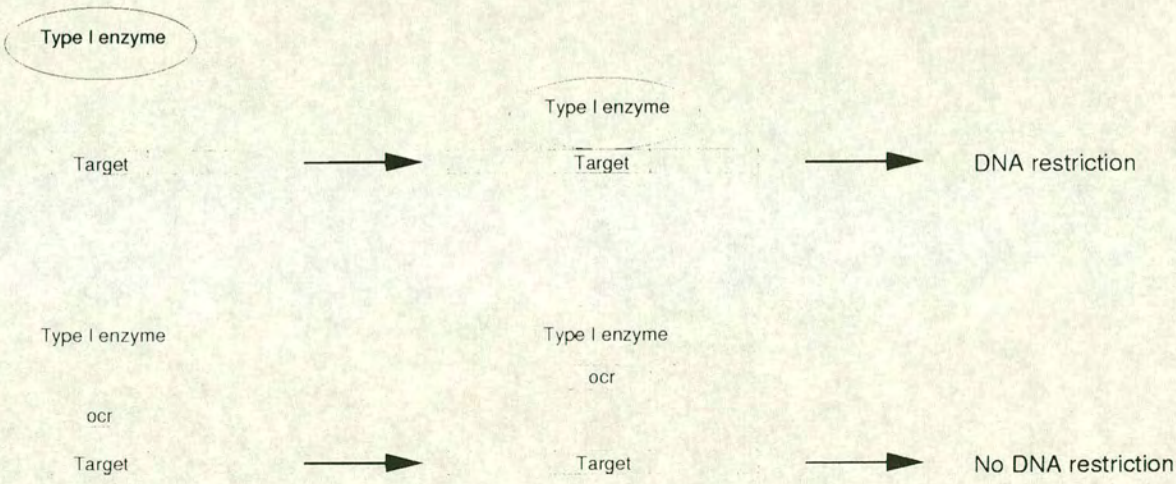
3. Mass spectrometry and differential scanning calorimetry

Mutation	mass by esms	expected subunit mass	Unfolding enthalpy UNITS	Thermal denaturation midpoint, T_m (^oC)	Change in T_m (^oC)
Wild Type	13677	13678	$1.86x10^5 \pm 242$	72.19	0
Truncated ocr (109 aa)		12754	$1.67x10^5 \pm 195$	71.11	-1.08
Truncated ocr (99 aa)		11521	$1.25x10^5 \pm 323$	58.40	-13.79
N4C	13663	13663	$1.74x10^5 \pm 404$	72.86	+0.67
N4C	27326	27326*			
D25C	13739	13665	$2.00x10^5 \pm 426$	77.87	+5.68
N43C	13740	13666	$1.91x10^5 \pm 776$	72.51	+0.32
D62C	13739	13665	$1.84x10^5 \pm 638$	70.64	-1.55
S68C	13769	13693	$1.87x10^5 \pm 258$	71.96	-0.23
W94C	13677	13594	$1.74x10^5 \pm 239$	68.57	-3.52

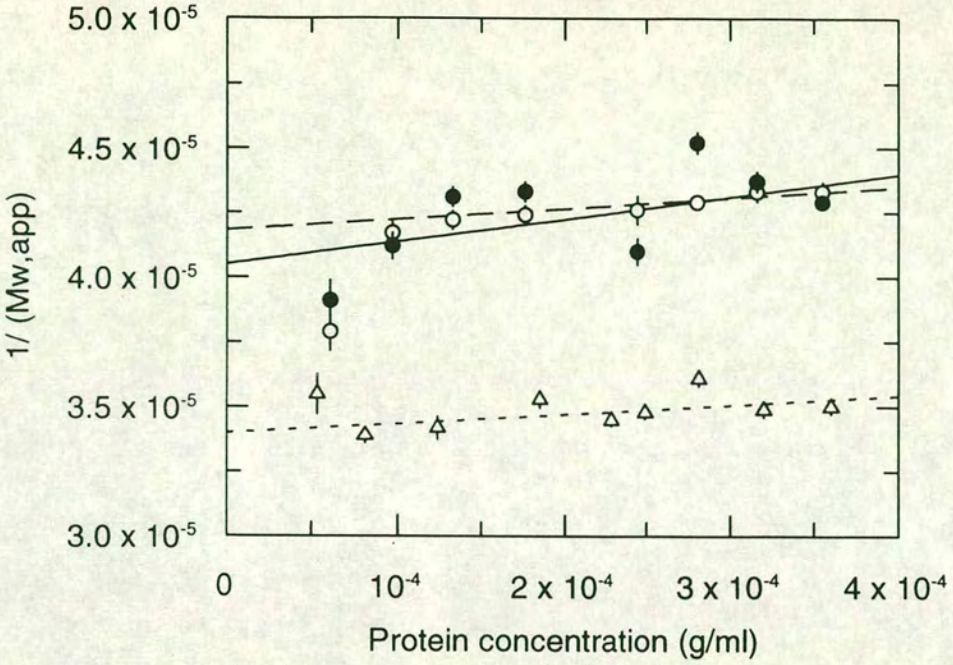
* dimer

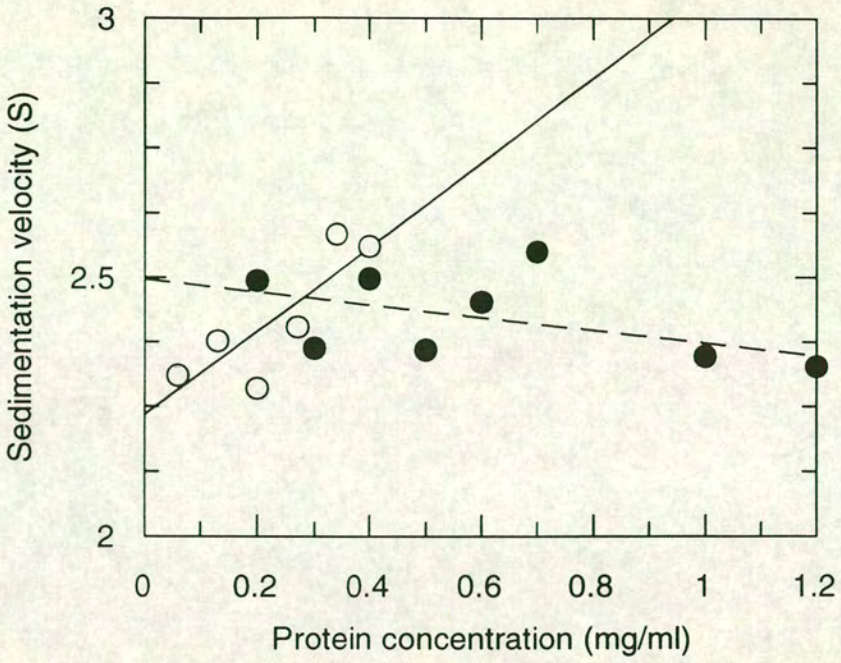
4. Oligonucleotides used for mutagenesis

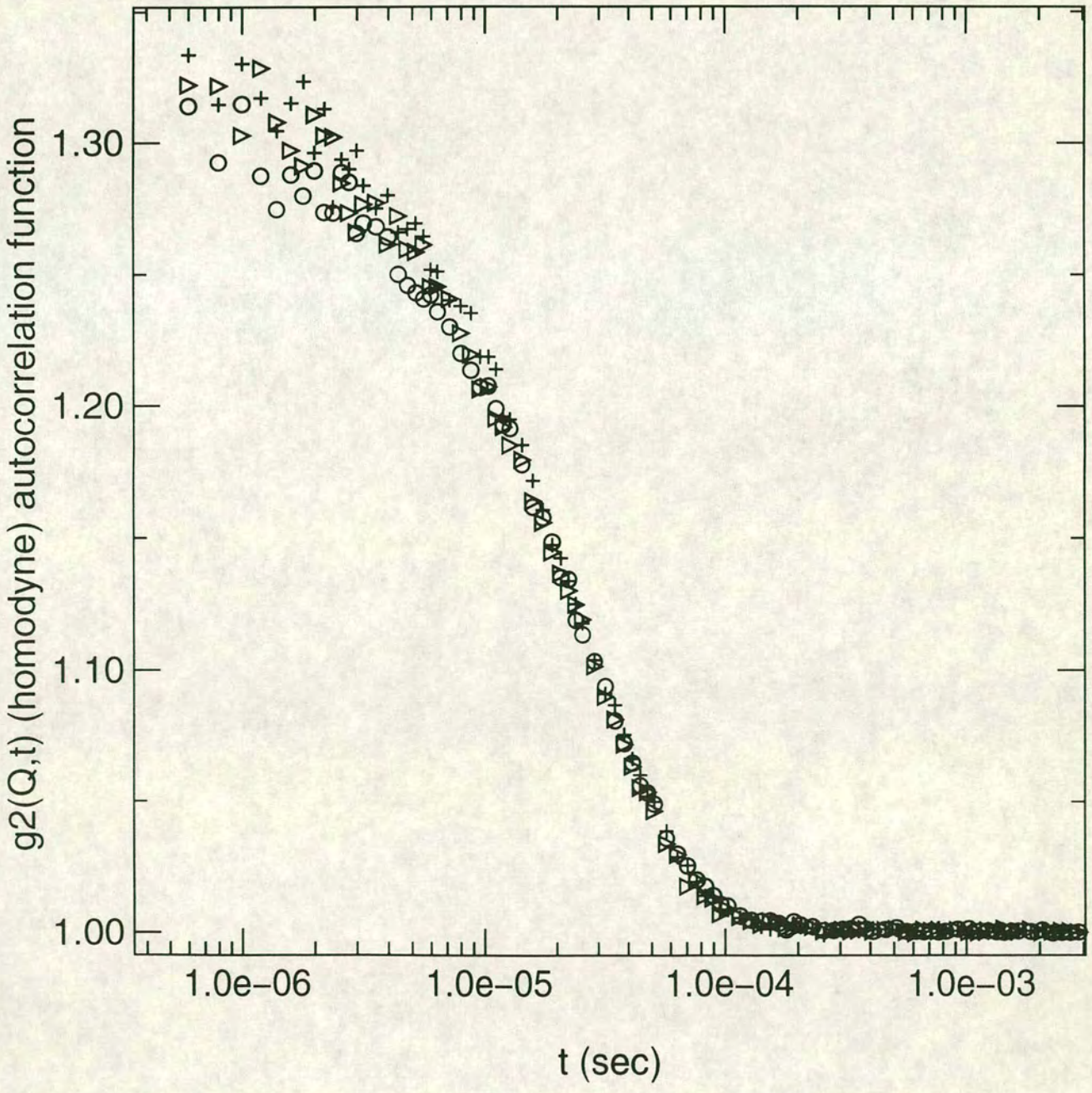
Mutation	Primer length	Strand	Sequence
N4C	44	Forward	5′-CAAGATGGCTATGTCTTGCATGACTTACAACAACGTTTTCGACC-3′
		Reverse	3′-GTTCTACCGATACAGAACGTACTGAATGTTGTTGCAAAAGCTGG-5′
D25C	41	Forward	5′-GCTGAAAGAAAACATCCGTTATTGTGACATCCGTGACACTG-3′
		Reverse	3′-CGACTTTCTTTTGTAGGCAATAACACTGTAGGCACTGTGAC-5′
N43C	44	Forward	5′-GCTATTCACATGGCTGCCGATTGTGCAGTTCCGCACTACTACGC-3′
		Reverse	3′-CGATAAGTGTACCGACGGCTAACACGTCAAGGCGTGATGATGCG-5′
D62C	39	Forward	5′-GGCAAGTGAGGGCATTTGCCTTGAGTTCGAAGACTCTGG-3′
		Reverse	3′-CCGTTCACTCCCGTAAACGGAACTCAAGCTTCTGAGACC-5′
S68C	39	Forward	5′-GACCTTGAGTTCGAAGACTGTGGTCTGATGCCTGACACC-3′
		Reverse	3′-CTGGAACTCAAGCTTCTGACACCAGACTACGGACTGTGG-5′
W94C	37	Forward	5′-GCAATTAACGATTGACCTCTGTGAAGACGCAGAAGAC-3′
		Reverse	3′-CGTTAATTGCTAACTGGAGACACTTCTGCGTCTTCTG-5′

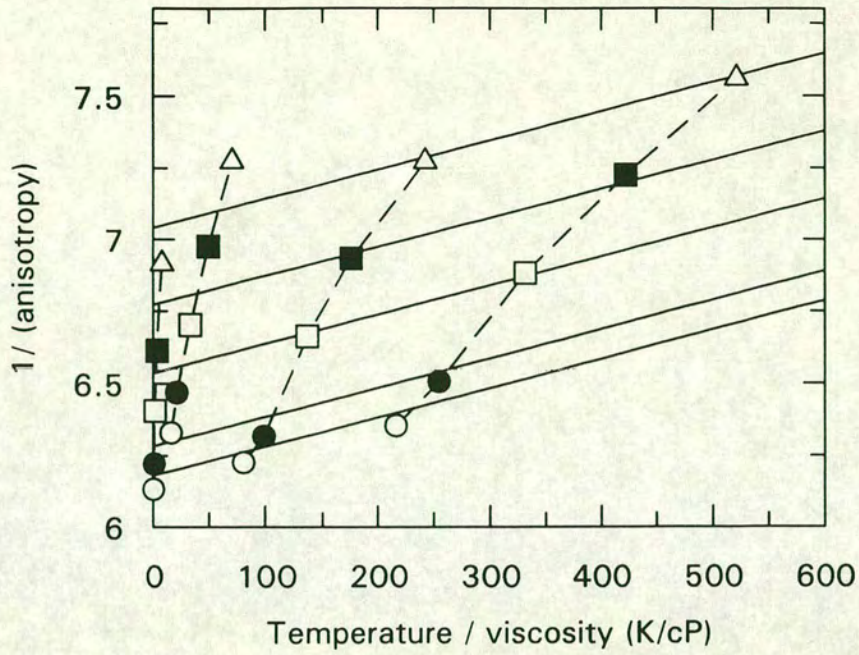

Type I enzyme
Target
Type I enzyme
Target
DNA restriction
Type I enzyme
ocr
Target
Type I enzyme
ocr
Target
No DNA restriction

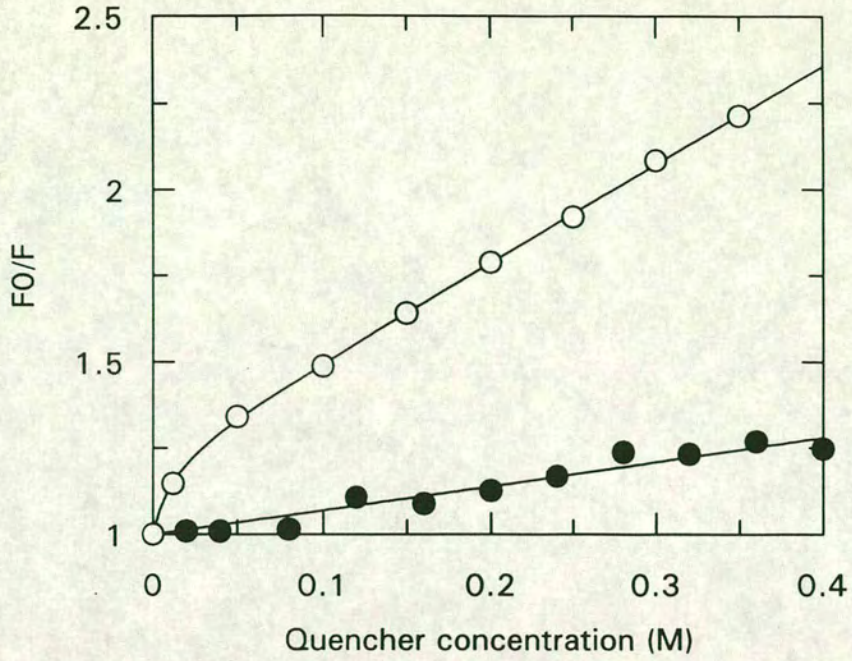
```
                 ...*,....1....,....2....*....3....,....4..*.,....5....,....
        AA       |AMSNMTYNNVFDHAYEMLKENIRYDDIRDTDDLHDAIHMAADNAVPHYYADIFSVMASE|
2ndary structure |       HHHHHHHHHHHHH   BBB    HHHHHHHHHHHHH   HHHHHHHHHHHHH|
Reliability      |9876685378999999988245522135514667778678763363266478999998 1|
Accessibility    |bbbbbbbbebbebbbebbeebbbbbbbbb eebeebbebbbbebbbeb eebbeebbee|

                 6.*..,..*.7....,....8....,....9...*,....10....,...11...,.
        AA       |GIDLEFEDSGLMPDTKDVIRILQARIYEQLTIDLWEDAEDLLNEYLEEVEEYEEDEE|
2ndary structure |   BBBB       HHHHHHHHHHHHHHHHHHHHHH                     |
Reliability      |68744651789999568999999999998765766139                   |
Accessibility    |bbebbb bbeebeeeeebbbbbebbbbeebbbebbeee                    |
```

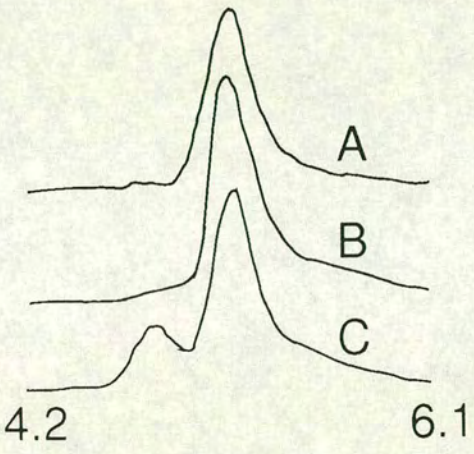

5.0 x 10-5
4.5 x 10-5
4.0 x 10-5
3.5 x 10-5
3.0 x 10-5
1/ (Mw,app)
0
10-4
2 x 10-4
3 x 10-4
4 x 10-4
Protein concentration (g/ml)

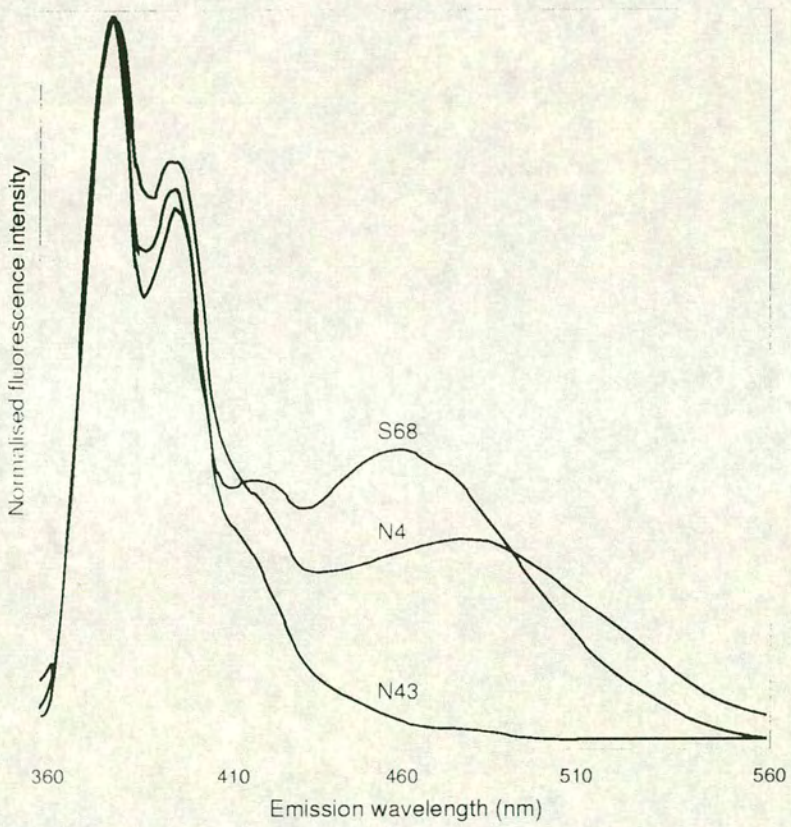

Sedimentation velocity (S)
3
2.5
2
0
0.2
0.4
0.6
0.8
1
1.2
Protein concentration (mg/ml)


g2(Q,t) (homodyne) autocorrelation function
1.30
1.20
1.10
1.00
1.0e−06
1.0e−05
1.0e−04
1.0e−03
t (sec)

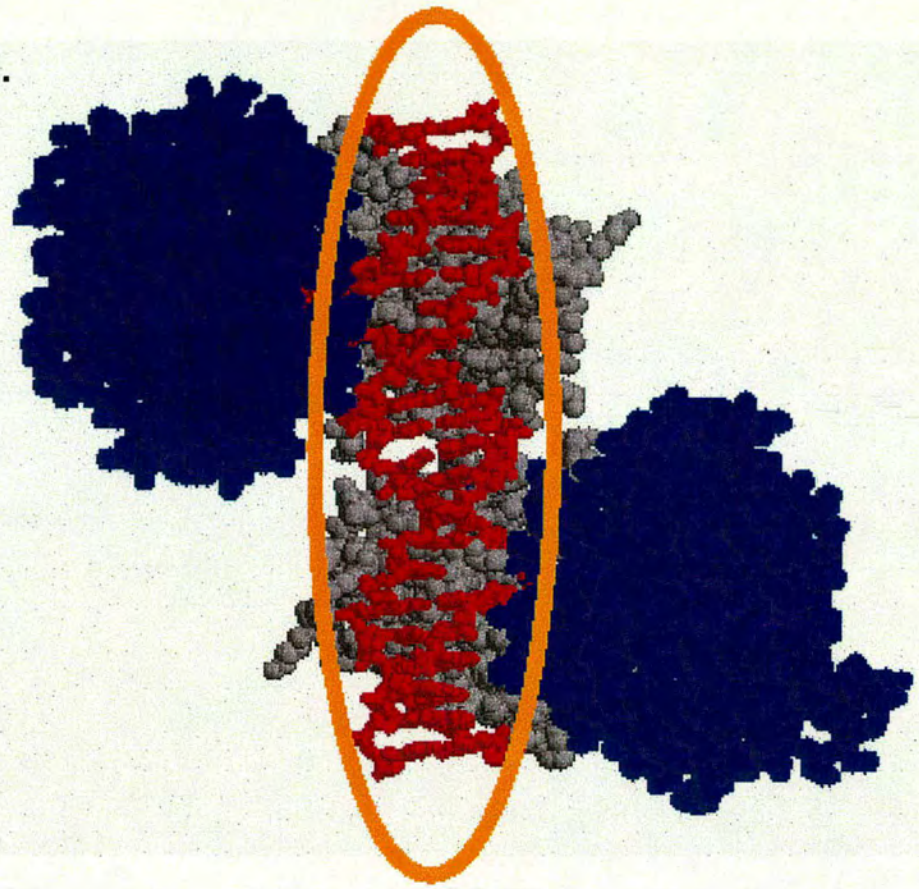

7.5
7
6.5
6
1/ (anisotropy)
0
100
200
300
400
500
600
Temperature / viscosity (K/cP)


2.5
2
1.5
1
FO/F
0
0.1
0.2
0.3
0.4
Quencher concentration (M)


A
B
C
4.2
6.1
Elution time (minutes)


Normalised fluorescence intensity
S68
N4
N43
360
410
460
510
560
Emission wavelength (nm)

a.



b.