

# **The Interaction of the Ada Protein with DNA: Structure and Thermodynamics.**

**Philip Edward Verdemato**

**Thesis Submitted for the Degree of Doctor of Philosophy  
Department of Biochemistry  
University of Leicester**

**February 2000**

UMI Number: U125665

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U125665

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

# Contents

<b>ACKNOWLEDGEMENTS</b> .....	<b>1</b>
<b>KINEMAGE TUTORIAL</b> .....	<b>2</b>
<b>ABBREVIATIONS</b> .....	<b>3</b>
<b>ABSTRACT</b> .....	<b>4</b>
<b>CHAPTER 1</b> .....	<b>5</b>
<b>DNA REPAIR</b> .....	<b>5</b>
<b>1.1 INTRODUCTION TO DNA REPAIR</b> .....	<b>5</b>
1.1.1 DNA damage – based on (Lindahl, 1993), (Mol et al., 1999). .....	5
1.1.2 Consequences of DNA damage.....	9
1.1.3 Repair strategies and mechanisms in general .....	9
1.1.3.1 Excision Repair-Based on (Sancar et al., 1988) and (Mol et al., 1999). ....	10
1.1.3.2 Mismatch repair -based on (Schar et al., 1998) .....	13
1.1.3.4 Direct repair .....	15
<b>1.2 REPAIR OF ALKYLATED DNA BY THE <i>E.coli</i> ADA PROTEIN</b> .....	<b>15</b>
1.2.1 The Role of Ada as Transcription Factor- The Adaptive Response.....	16
1.2.1.1 The N-terminal domain of Ada provides the ‘On-Switch’ for the adaptive response .....	18
1.2.1.2 The C-terminal domain provides the key to differential regulation of adaptive response genes.....	20
1.2.1.3 An alternative pathway to upregulation of the adaptive response genes.....	21
1.2.2 Structures of Ada and Relation to Catalytic and Transcriptional Roles.....	21
1.2.2.1 The N-terminal Methylphosphotriester Repair Domain of Ada.....	21
1.2.2.2 Mechanism of methylation-dependent switch to transcriptional regulator: .	22
1.2.2.3 Mechanism of methyl abstraction .....	22
1.2.2.4 The C-terminal O <sup>6</sup> -methylguanine-DNA methyltransferase domain of Ada .....	23
1.2.2.5 Enzyme structure and postulated conformational change upon binding DNA .....	23
1.2.2.6 Roles of conserved residues .....	25
1.2.2.7 Mechanism of action .....	26
1.2.2.8 Sequence specificity and target site identification .....	28

1.3 AIMS OF THIS THESIS .....	30
<b>CHAPTER 2.....</b>	<b>32</b>
<b>METHODS .....</b>	<b>32</b>
2.1 OPTIMISING OVEREXPRESSION OF ADA-C IN <i>E.Coli</i> .....	32
2.1.1 The Expression System .....	33
2.1.2 Enhancing expression in BL21(DE3).....	34
2.1.2.1 Testing effects of different induction times on expression in BL21(DE3) ..	34
2.1.2.2 Testing effects of different lengths of induction on expression in BL21(DE3)	35
2.1.2.3 Testing effects of different media on growth of BL21(DE3):pET22b: <i>adaC35</i>	35
2.1.3 Expression in other <i>E.coli</i> strains .....	36
2.1.4 Examining expression in B834(DE3).....	36
2.1.5 Enhancing growth and expression in B834(DE3)- “Cellular resuspension” .....	37
2.1.6 The Isotope label-dependent protocols for expressing Ada-C .....	38
2.1.6.1 Expressing unlabeled Ada-C .....	38
2.1.6.2 Expressing Ada-C labeled with low cost isotopes .....	38
2.1.6.3 Expressing Ada-C labeled with high cost isotopes .....	39
2.1.7 Purification of Ada-C .....	41
2.1.8 Quality control of Ada-C.....	42
2.1.8.1 Electrospray mass spectrometry and N-terminal sequencing of Ada-C.....	42
2.1.8.2 Activity Assay of Ada-C .....	42
2.1.9 Synthesising and purifying DNA oligomers .....	43
2.2 EXAMINING THERMODYNAMICS OF ADA-C/DNA INTERACTIONS .....	45
2.2.1 Attempting to obtain binding constants for Ada-C:DNA using Fluorescence and	
Circular Dichroism .....	46
2.2.1.1 Fluorescence .....	46
2.2.1.2 Circular Dichroism .....	46
2.2.2 Isothermal Titration Calorimetry- DNA binding and demethylation energetics.	47
2.3 ASSIGNING THE BACKBONE OF ADA-C.....	50
2.3.1 Attempts to assign the backbone using protonated <sup>13</sup> C/ <sup>15</sup> N-labeled Ada-C.....	50
2.3.2 The Ada-C backbone assignment using 70% deuterated <sup>13</sup> C/ <sup>15</sup> N-labeled Ada-C	51
2.4 EXAMINING INTERACTIONS BETWEEN ADA-C AND DNA BY NMR- .....	52
2.4.1 General NMR methods.....	52
2.4.2 Calculation of binding constant between Ada-C and single-stranded DNA by	
NMR .....	54

2.4.3 In Vitro methylation reactions.....	55
2.4.4 pKa determination for the active site Cys 144 .....	55
<b>2.5 EXAMINING THE EFFECT OF ADA-C BINDING ON THE DNA.....</b>	<b>56</b>
2.5.1 Energetics of duplex melting.....	56
2.5.2 Examining imino protons of DNA upon binding by Ada-C .....	56
2.5.3 Probing the structure of methylated and unmethylated DNA using <sup>19</sup> F NMR....	57
<b>2.6 MODELLING DNA-BINDING.....</b>	<b>58</b>
<b>CHAPTER 3.....</b>	<b>59</b>
<b>PROTEIN EXPRESSION, PURIFICATION AND LABELING, DNA SYNTHESIS AND</b>	
<b>PURIFICATION.....</b>	<b>59</b>
<b>3.1 INTRODUCTION.....</b>	<b>59</b>
<b>3.2 RESULTS.....</b>	<b>61</b>
3.2.1 Enhancing expression in BL21(DE3).....	61
3.2.2 Expression in other strains.....	63
3.2.3 Further investigation of expression in B834(DE3) .....	64
3.2.4 Enhancing growth and expression in B834(DE3)- Cellular resuspension .....	65
3.2.5 Production of unlabeled Ada-C .....	68
3.2.6 Production of selectively <sup>15</sup> N-labeled Ada-C .....	69
3.2.7 Production of uniformly labeled deuterated Ada-C .....	69
3.2.8 Purification of Ada-C .....	71
3.2.9 Quality control of Ada-C.....	71
3.2.10 Assaying Ada-C.....	71
<b>3.3 DISCUSSION .....</b>	<b>71</b>
3.3.1 Increasing the yield of Ada-C.....	72
3.3.1.1 Increasing the yield of Ada-C through media design .....	72
3.3.1.2 Increasing the yield of Ada-C through strain alteration .....	72
3.3.1.3 Cellular resuspension to further enhance Ada-C yield.....	73
3.3.2 Selectively labeling Ada-C and monitoring crosslabeling .....	74
3.3.3 Purification of Ada-C .....	76
3.3.4 Observations regarding Ada-C stability .....	77
<b>3.4 SUMMARY .....</b>	<b>77</b>
<b>CHAPTER 4.....</b>	<b>77</b>
<b>THERMODYNAMICS OF ADA-C : DNA INTERACTIONS.....</b>	<b>77</b>
<b>4.1 INTRODUCTION.....</b>	<b>77</b>
4.1.1 Measuring binding by Fluorescence.....	78

4.1.3 An overview of ITC- How it works .....	82
4.1.4 Interpreting thermodynamic information .....	84
4.2 <i>RESULTS</i> .....	85
4.2.1 Fluorescence .....	85
4.2.2 Circular dichroism .....	86
4.2.3 Thermodynamics of binding interactions by ITC .....	86
4.2.4 Thermodynamics of DNA demethylation .....	90
4.3 <i>DISCUSSION</i> .....	93
4.3.1 Fluorescence and circular dichroism analysis of DNA binding .....	93
4.3.2 ITC analysis of Ada-C binding unmethylated single-stranded DNA.....	94
4.3.3 ITC analysis of Ada-C binding unmethylated duplex DNA .....	94
4.3.4 General considerations of Ada-C binding DNA.....	95
4.3.5 Energetics of <i>in vitro</i> DNA demethylation by Ada-C .....	97
4.4 <i>SUMMARY</i> .....	98
<b>CHAPTER 5.....</b>	<b>99</b>
<b>ASSIGNING THE BACKBONE OF ADA-C.....</b>	<b>99</b>
5.1 <i>INTRODUCTION TO NMR</i> .....	99
5.1.1 Exciting and relaxing spins .....	99
Generating energy level transitions .....	101
5.1.2 Chemical shift.....	103
5.1.3 Relaxation.....	104
5.1.6 Two-dimensional NMR.....	108
5.1.7 NMR of Ada-C .....	111
5.1.8 The Backbone Assignment Strategy.....	111
5.2 <i>RESULTS</i> .....	115
5.2.1 Limited backbone assignment through the use of protonated <sup>13</sup> C/ <sup>15</sup> N-labeled Ada-C .....	115
5.2.2 Successful backbone assignment employing 70% deuterated <sup>13</sup> C/ <sup>15</sup> N-labeled Ada-C .....	116
5.3 <i>DISCUSSION</i> .....	120
5.5 <i>SUMMARY</i> .....	122
<b>CHAPTER 6.....</b>	<b>123</b>
<b>ANALYSIS OF DNA BINDING BY NMR .....</b>	<b>123</b>
6.1 <i>INTRODUCTION</i> .....	123
6.2 <i>RESULTS</i> .....	125
6.2.1 Mapping the DNA-binding site on wild type Ada-C .....	125

6.2.2	Ada-C binding duplex DNA.....	132
6.2.3	Effects of binding DNA on the tautomeric state of His 145 .....	133
6.2.4	<i>In vitro</i> methylation reactions- suicidal inactivation of wild type Ada-C .....	135
6.2.5	A Comparison of binding to methylated and unmethylated DNA by inactive C144S mutant Ada-C .....	137
6.2.6	pK <sub>a</sub> determination of active site Cys 144 .....	141
6.2.7	Structural homology of Ada-C to other DNA-binding proteins.....	144
6.3	<b>DISCUSSION</b> .....	146
6.3.1	Involvement of a helix-turn-wing motif .....	146
6.3.2	Selective broadening of the HTH chemical shifts, and protein dynamics.....	147
6.3.3	Examining the hydrogen bonding network of the active site .....	148
6.3.4	Effect of self methylation on Ada-C stability- implications for adaptive response termination.....	149
6.3.5	Inactive C144S mutant Ada-C binding to methylated and unmethylated DNA	150
6.3.6	Relationship between Ada-C, Integrase and helix-turn-helix proteins .....	151
6.4	<b>SUMMARY</b> .....	153
<b>CHAPTER 7.....</b>		<b>154</b>
<b>EFFECTS OF METHYLATION AND ADA-C BINDING ON DNA .....</b>		<b>154</b>
7.1	<b>INTRODUCTION</b> .....	154
7.2	<b>RESULTS</b> .....	155
7.2.1	Effect of O <sup>6</sup> meG on duplex stability .....	155
7.2.2	Effect of Ada-C binding on DNA imino protons .....	158
7.2.3	Probing the DNA using <sup>19</sup> F NMR .....	159
	Unmethylated DNA .....	161
7.3	<b>DISCUSSION</b> .....	164
7.4	<b>SUMMARY</b> .....	165
<b>CHAPTER 8.....</b>		<b>166</b>
<b>MODELLING HOW ADA-C BINDS DNA &amp; GENERAL DISCUSSION.....</b>		<b>166</b>
8.1	<b>INTRODUCTION</b> .....	166
8.2	<b>THE DNA-BINDING MODEL</b> .....	167
8.3	<b>MECHANISM OF LESION LOCALISATION</b> .....	174
8.4	<b>MECHANISM OF BASE-FLIPPING</b> .....	176
8.5	<b>MECHANISM OF DEMETHYLATION</b> .....	177
<b>CONCLUSIONS.....</b>		<b>179</b>
<b>FURTHER WORK .....</b>		<b>181</b>

<b>APPENDICES.....</b>	<b>182</b>
<i>APPENDIX 1-PLASMIDS, MEDIA AND OLIGOS .....</i>	<i>182</i>
Appendix 1.1- Plasmid preparation.....	182
Appendix 1.2- The cellular transformation and glycerol storage protocol.....	184
Appendix 1.3 SDS-PAGE .....	184
Appendix 1.4- Media.....	185
Appendix 1.5: Calculation of Extinction coefficients for linear oligonucleotides .....	186
Appendix 1.6 Native PAGE for DNA Annealing assessment .....	187
<i>APPENDIX 2- ITC .....</i>	<i>187</i>
2.1 Data Analysis for ITC .....	187
<b>CHAPTER 9.....</b>	<b>191</b>
<b>BIBLIOGRAPHY.....</b>	<b>191</b>

**ADDENDA**

**One 3.5” floppy computer disk**

# Acknowledgements

## *People outside the NMR Centre*

I would like to thank Dr. Barbara Briggs from the Dept. Chemistry for running the DSC experiments on the DNA. Drs. Kathryn Lilley and Sharrad Mistry of PNACL synthesised all the oligonucleotides used here, performed all the mass spectrometry and protein sequencing. Dr. G. Major and Ms. S. Charlton of the University of Newcastle kindly assessed the activity of Ada-C with their assay. Dr. J. Brannigan at the University of York provided the plasmids of the mutant Ada-C. I am also grateful to Prof. S. Ealick at Cornell for providing the coordinates of the Ada-C/DNA complex model by Vora *et al.* (1998). Drs. H. Trayer and J. Eads (University of Birmingham) allowed me to use their ITC apparatus. Finally, Prof. Tony Maxwell (Leicester), and Dr. Bill Primrose (PanTherix) helped me with plentiful advice on many aspects of the project.

In the x-ray crystallography lab. here at Leicester, Dr. Jane Wibley provided useful advice on many aspects of the project, particularly the structural aspects of DNA-binding. Dr. Geoff Briggs helped out with information on CD and other techniques. Dr. Peter Moody, my supervisor, and alkyltransferase guru helped direct, and infuse much enthusiasm into the project. I am very grateful to him for always being available to dispense his wide knowledge on many matters.

## *People in the NMR Centre*

I am enormously grateful to members of the NMR Centre who helped me during this Thesis, these include: Dr. Fabio Zuccotto, who did the minimisation of the CAP-based DNA-flipped model. Dr. Igor Barsukov, who spent ages teaching me about NMR, computer programming, skiing and tennis whilst his wife, Lena, ensured that supplies never ran out. Andy Prescott aided me in producing Ada-C, and taught me the ropes when I first arrived. Dr. Ramin Badii helped me with molecular biology and spiritual enlightenment. Dr. Christian Damblon helped me with everything from protein purification to 3D NMR. He provided much supervision and I am proud to have been taught by him. Richard Ward and Marcus Tillet helped with relaxation analysis and together with Samantha Jones provided support. Prof. Lu-Yun Lian, supervised the project, and provided directionality and funding for the research, as well as much helpful personal advice and support for which I am most grateful.

Finally, I would like to thank my parents for their love and support, and Lucy for financial assistance whilst writing up.

# **Kinemage Tutorial**

Accompanying this Thesis is a kinemage tutorial consisting of one 3.5" floppy disk containing three files for use on a PC using the operating system Windows 95 or above. The 'readme.txt' file should be read first, and is self-explanatory. The purpose of this disk is to provide a small molecular visualisation program called 'kinemage' with a tutorial about the structure of the model of the Ada-C/DNA complex. This tutorial allows more detailed examinations of three dimensional molecular structures than are possible to obtain on a page. The tutorial was designed as an extra, to complement what is written in this Thesis.

# Abbreviations

A<sub>595</sub>: Absorbance at 595nm used for assessing cellular growth

Ada-C: C-terminal domain of the Ada protein (residues 93-177)

Ada-N: N-terminal domain of the Ada protein (residues 12-92)

BFC: DNA oligonucleotide: 5'-GTGGCAACFTGTGGCAAFCT

(where F is 5-fluorocytosine)

BIGTAR: DNA oligonucleotide: 5'-AGGTTGCCACAGGTTGCCAC

BIGCOMP: DNA oligonucleotide: 5'-GTGGCAACCTGTGGCAACCT

BMT: DNA oligonucleotide: 5'-AGGTTGCCACA(O<sup>6</sup>meG)GTTGCCAC

(where O<sup>6</sup>meG is O<sup>6</sup>-methylguanine)

dH<sub>2</sub>O: Deionised water

DSC: Differential Scanning Calorimetry

dsDNA: double-stranded DNA

HSQC: Heteronuclear Single Quantum Coherence (a 2D NMR spectrum often involving <sup>15</sup>N & <sup>1</sup>H chemical shifts)

HMQC: Heteronuclear Multiple Quantum Coherence (a 2D NMR spectrum often involving <sup>15</sup>N & <sup>1</sup>H chemical shifts)

Hz: Hertz

ITC: Isothermal Titration Calorimetry

mwco: Molecular Weight Cut Off

NMR: Nuclear Magnetic Resonance

TFA: Trifluoroacetic acid

ssDNA: Single-stranded DNA

# Abstract

The C-terminal domain of the *E.coli* Ada protein (Ada-C) aids the maintenance of genomic integrity by efficiently repairing pre-mutagenic O<sup>6</sup>-alkylguanine lesions in DNA. The aim of this thesis was to discern the manner in which Ada-C binds and repairs DNA. The research was pursued from both structural and thermodynamic perspectives, to obtain a model of the DNA-binding process.

In order for NMR structural studies to be practically feasible, the production of recombinant Ada-C from *E.coli* culture was enhanced through a combination of rational media design, *E.coli* strain choice, and the employment of a growth strategy for maximizing cell density prior to induction of protein expression. A purification protocol was then devised for Ada-C to maximise yield and efficiency.

Nuclear Magnetic Resonance (NMR) studies mapped the DNA-binding site to the recognition helix of the helix-turn-helix motif and a loop region (residues 149-155) known as the 'wing'. Using this binding interface, and in the absence of a large conformational change in the protein upon DNA-binding, it was found that an O<sup>6</sup>meG lesion was inaccessible to active site nucleophile Cys 144 when the lesion remained stacked within the DNA duplex. This lesion could enter the active site by being rendered extrahelical, or "flipping". A previously published model of how Ada-C bound DNA involved "base-flipping" (Vora *et al.*, 1998), and was found to have good agreement with the experimentally determined DNA-binding face of Ada-C.

The DNA-binding process was shown to be entropically driven, whilst the demethylation reaction provoked an exothermic heat change. At millimolar concentrations, methylation of Ada-C led to a loss of structural integrity.

Duplex DNA containing an O<sup>6</sup>meG lesion had a lower enthalpy of melting than unmethylated DNA. This, along with <sup>19</sup>F NMR work on modified DNA, led to the postulation that O<sup>6</sup>meG might induce localised duplex melting. This would certainly aid the base-flipping process.

# Chapter 1

## *DNA Repair*

### **1.1 INTRODUCTION TO DNA REPAIR**

Alterations to the genomic DNA sequence, although necessary for the evolution of a species, often occur to the detriment of individual lifeforms. It was once calculated that a mutation occurs in a bacterial cell once every  $10^8$  cell divisions, whilst in higher organisms this rate is increased (Brown, 1992). A number of these mutations occur at random, through errors in the proofreading involved with DNA replication. However, this mutation rate can be dramatically enhanced by exposure of the DNA to mutagenic agents. Consequences of the induced DNA damage range from silent point mutations, which have no effect on the gene product, to lethal mutations.

The cell contains an abundance of DNA repair mechanisms which cope with myriad types of damage. This section briefly reviews general DNA damage and its repair, as a prelude to the more specific repair of O<sup>6</sup>-methylguanine by the Ada protein, to which this thesis was dedicated. Due to the enormity of the DNA repair field it was only possible to outline the DNA damage and repair mechanisms here.

#### **1.1.1 DNA damage – based on (Lindahl, 1993), (Mol et al., 1999).**

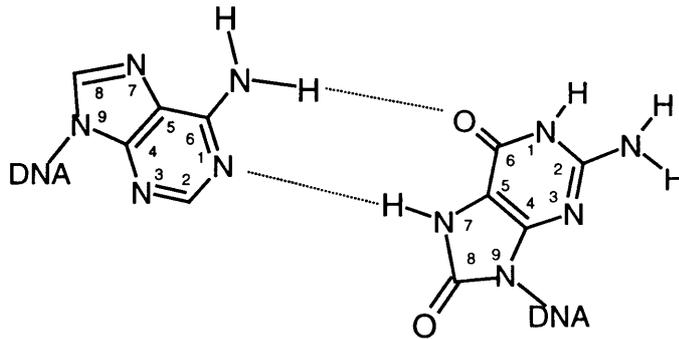
The chemical stability of DNA is limited, and the genomic integrity threatened, by alterations to its structure. DNA can be modified by attack from myriad agents ranging from external radiation to endogenously generated alkylating agents. The five major types of DNA base damage which occur are: oxidation, deamination, alkylation, ultraviolet (uv) radiation-induced damage (such as base dimerisation), and the formation of AP sites (AP stands for apurinic or apyrimidinic sites caused by base loss.)

#### **Oxidation:**

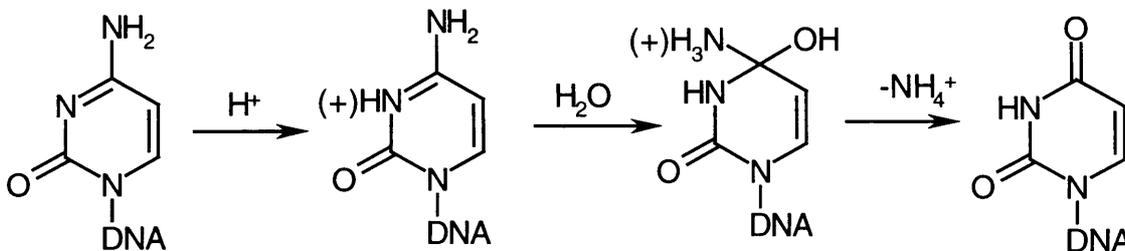
Intracellular oxidative processes lead to the generation of reactive oxygen species which attack cellular infrastructure including DNA. Hydroxyl radicals often react with the C<sup>5</sup>-C<sup>6</sup> double bond in thymine bases, leading to saturation, concomitant loss of planarity of the pyrimidine ring and coding properties. Hydroxyl addition frequently occurs at the C<sup>8</sup> position of the guanine ring. The 8-hydroxyguanine formed can tautomerise to 8-oxoguanine, which may assume the *syn* conformation in the DNA duplex. As a result this altered base can base pair with Adenine during DNA replication, causing transversion mutations (Figure 1.1).

## Deamination:

DNA bases with exocyclic amino groups (adenine, guanine and cytosine) can be deaminated. As with oxidative damage, this can lead to altered base pairing properties and subsequent mispairs at DNA replication. Deamination of cytosine renders uracil, whilst that of 5-methylcytosine renders thymine. These altered bases are capable of pairing with adenine, and causing GC→AT transition mutations if left unrepaired prior to the next round of replication. Potential mechanisms of deamination are shown in Figure 1.2.



**Figure 1.1** The adenine:8-oxoguanine mispair. Adenine is orientated in the *anti* conformation, 8-oxoguanine is *syn*.

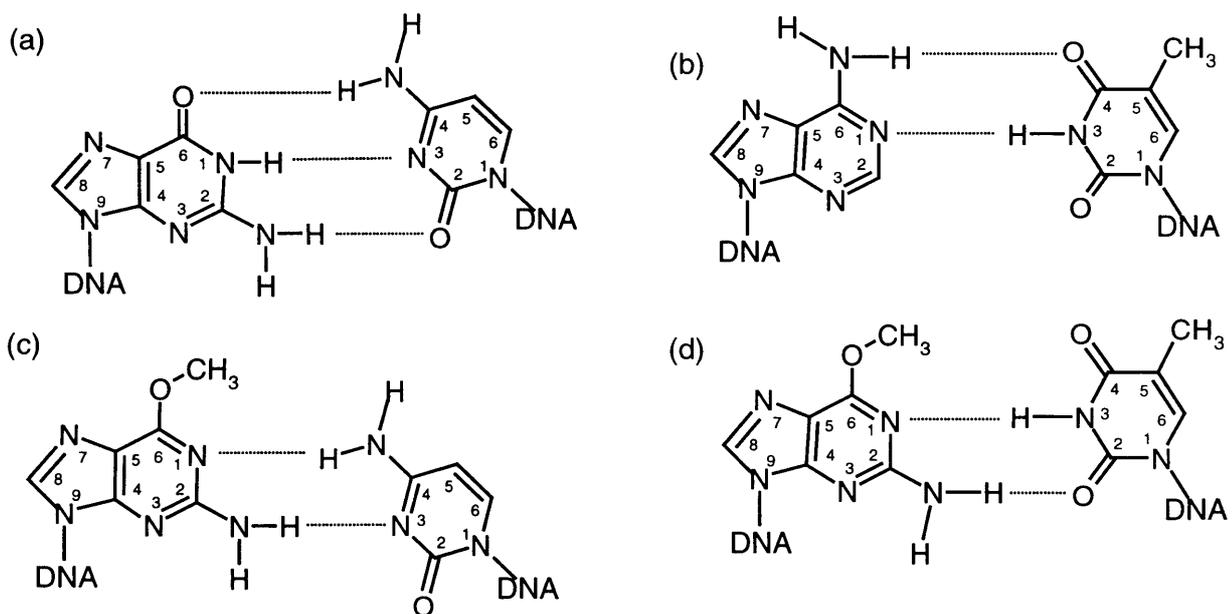


**Figure 1.2** Potential mechanism for the deamination of Cytosine to Uracil at neutral pH, obtained from Shapiro *et al.* (1981).

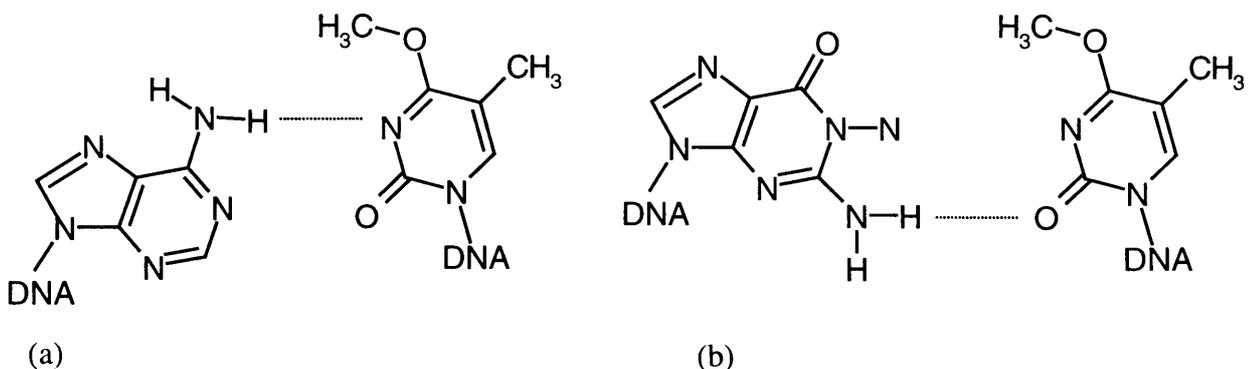
## Alkylation:

Along with oxygen species, living cells contain other agents which cause DNA damage and may be considered genotoxic. These include alkylating agents such as S-adenosylmethionine (SAM) and methylnitrosourea. Alkylating agents react with nucleophilic centres present in the DNA bases and backbone. Major alkylation sites include the N<sup>3</sup> atoms of purines, and the exocyclic O<sup>4</sup> and O<sup>6</sup> oxygen atoms on thymine and guanine, respectively. SAM is a methyl group donor used during transmethylation reactions (eg. the methylation of C<sup>5</sup> in cytosine by *HhaI* methyltransferase). It can also methylate DNA in a nonenzymatic manner, its main targets being the purine ring nitrogens, resulting in the formation of 3-methyladenine and 7-

methylguanine (Rydberg *et al.*, 1982). Since the latter base does not have altered base pairing properties, it is not mutagenic. By contrast, 3-methyladenine can block DNA replication and is a major threat to the cell. Methylnitrosourea may also be generated endogenously during normal metabolism (Vaughan *et al.*, 1991). This compound can methylate at the exocyclic oxygens of guanine and thymine. This alkylation locks the bases irreversibly into the keto tautomer, allowing them to form stable mispairs with thymine and guanine, respectively (Figures 1.3 and 1.4).



**Figure 1.3** (a) A Watson-Crick Guanine:Cytosine base pair. (b) A Watson-Crick Adenine:Thymine base pair. (c) A wobble  $O^6$ -methylguanine:Cytosine base pair. (d) A Watson-Crick  $O^6$ -methylguanine:Thymine mispair. These depictions of base pairing are based on Figure 1 from Moody *et al.* (1995). The exact orientation of the alkyl group *syn* or *anti* with respect to the  $N^1$  atom varies, as does the number of hydrogen-bonds involved in the  $O^6$ -methylguanine (recently reviewed in more detail elsewhere, Verdemato *et al.* (1998). Here all alkyl groups are drawn *syn* to  $N^1$  for simplicity.



**Figure 1.4** The possible base pairing arrangements for  $O^4$ -methylguanine. (a)  $O^4$ meT:A wobble base pair as depicted in Kalnik *et al.* (1988a). (b)  $O^4$ meT:G modified Watson-Crick base pair Kalnik *et al.* (1988b).

The products of such mispairing are transition mutations following the second round of DNA replication (Loechler *et al.*, 1984). Thymine is inserted opposite O<sup>6</sup>-methylguanine (O<sup>6</sup>meG) more readily than cytosine during DNA replication, hence the mutagenicity of O<sup>6</sup>meG. Since O<sup>6</sup>meG:T base pairs are less stable than O<sup>6</sup>meG:C (Gaffney *et al.*, 1989); (Gaffney *et al.*, 1984) thermodynamic stability of base pairing is not the reason for the mispairing observed. Instead, it is believed that cytosine is less readily incorporated opposite O<sup>6</sup>meG because DNA polymerase has difficulty in aligning the 3'OH of the replicating DNA strand with the  $\alpha$ -phosphorous of the dCTP due to the wobble arrangement of the base pairing (Tan *et al.*, 1994). This wobble arrangement causes slight kink in the phosphodiester backbone either side of the cytosine, which is not observed in the O<sup>6</sup>meG:T mispair (Kalnik *et al.*, 1989b); (Kalnik *et al.*, 1989a).

Alkylating agents can also cause the formation of Rp and Sp-methylphosphotriesters on the backbone of DNA, although these lesions are not pre-mutagenic. In the Sp-diastereoisomer the methyl group protrudes from the backbone of DNA whilst in the Rp-diastereoisomer the methyl group is accommodated within the major groove. Although the Sp-diastereoisomer is thought to cause some slight impediment to DNA polymerase I, the lesions are not dangerous as neither is pre-mutagenic (Weinfeld *et al.*, 1985).

### **UV radiation:**

Ultraviolet radiation is another common damage-inducing agent. It can damage DNA indirectly through the production of oxygen radicals, or directly through the formation of cyclobutane dimers. Here adjacent pyrimidines are joined through a cyclobutane ring structure involving their C<sup>5</sup> and C<sup>6</sup> atoms.

#### **1.1.2 Consequences of DNA damage**

Whether it be altered base structure, base or nucleotide deletion, or the formation of dimers, damage to DNA ultimately results in an altered DNA sequence if replication is allowed to proceed without repair. DNA damage therefore gives rise to spontaneous mutagenesis, and by inference to carcinogenesis. Consequences of DNA damage manifest themselves readily in DNA repair-deficient phenotypes. Defects in the nucleotide excision repair apparatus associated with the genetic disorder xeroderma pigmentosum, for example, reveal the effects of DNA damage. The disease is characterised by a hypersensitivity to uv radiation, and high predisposition to skin cancers, demonstrating directly what may happen to living organisms if DNA damage is left unrepaired. Another example is provided in the formation of O<sup>6</sup>-methylguanine, which causes G:C to A:T transition mutations in the second position of codon 12 in the *H-ras* oncogene. This point mutation is sufficient to cause the onset of mammary

and skin tumours in response to cellular treatment with the alkylating agent methylnitrosourea (Georgiadis *et al.*, 1991). DNA alkylation is also used in tumour treatment by chemotherapy. Chemotherapeutic agents function by alkylating DNA. The human enzyme O<sup>6</sup>-methylguanine-DNA methyltransferase is responsible for the removal of these alkyl groups from exocyclic oxygens on guanine. It counters the effects of chemotherapy, and its inhibition is desirable in order to reduce the dosage of chemotherapy administered (Pegg *et al.*, 1995). Along with disease, DNA damage has also been implicated in the ageing process (Lindahl, 1993).

### **1.1.3 Repair strategies and mechanisms in general**

Multiple repair mechanisms of varying complexities and specificities exist to reverse the effects of DNA damage (Sancar *et al.*, 1988), (Vassilyev *et al.*, 1997), (Mol *et al.*, 1995). There are four categories of DNA repair: excision repair, mismatch repair, recombinational repair, and direct repair. In the first two categories, damaged nucleotides or bases are removed and the information in the complementary strand used to rebuild an undamaged duplex. This frequently requires multiple steps and a number of enzymes. In direct repair, often only a single enzyme is employed to reverse the effects of the DNA damage. One example would be removal of the alkyl adduct on O<sup>6</sup>meG by O<sup>6</sup>-methylguanine-DNA methyltransferase. In recombinational repair single-stranded postreplicative gaps in DNA (which may result when a dithymidine dimer blocks DNA replication) are filled in by strand transfer from an intact sister duplex. Recombinational repair will not be discussed further. Brief descriptions of the DNA repair mechanisms not involving recombination, and therefore of more relevance to this thesis are given below.

#### **1.1.3.1 Excision Repair**-Based on (Sancar *et al.*, 1988) and (Mol *et al.*, 1999).

Two types of excision repair exist: base excision repair and nucleotide excision repair. The former deals with the removal and replacement of single damaged bases, whilst nucleotide excision repair involves the removal and replacement of an oligonucleotide encompassing the lesion.

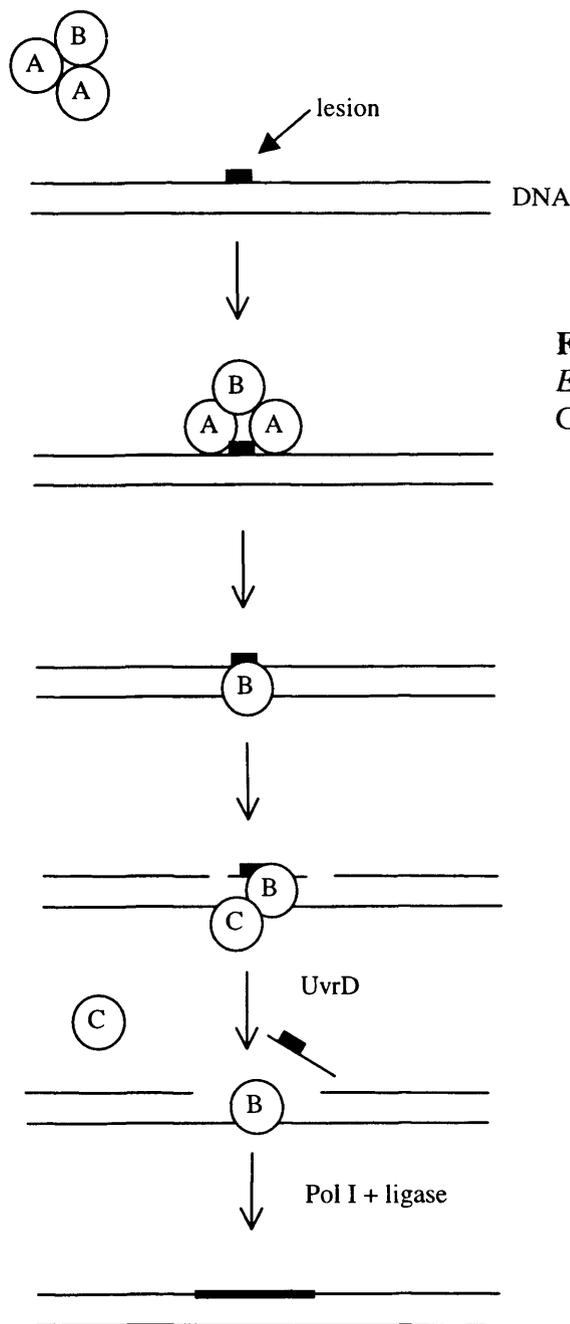
### **Nucleotide excision repair (NER)**- based on (Goosen *et al.*, 1998)

This type of repair mechanism exists in prokaryotes and eukaryotes and is responsible for the repair of lesions which cause major helical distortions to duplex DNA. Such lesions include pyrimidine dimers, sites of base loss, and bulky base adducts (those containing more than 2 carbon atoms such as O<sup>6</sup>-butylguanine (Todd *et al.*, 1983)). A damaged DNA site is thought to be recognised by virtue of the distortions it causes to base pairing and the phosphate backbone of DNA. In *E.coli* the NER apparatus consists of just 6 proteins (UvrA, UvrB, UvrC, UvrD, DNA polymerase I and ligase) compared with eukaryotic NER which needs about 30 proteins. The UvrA protein exists as a dimer, and also binds a monomer of UvrB. UvrA locates the lesion of DNA and then the UvrA<sub>2</sub>B-DNA complex undergoes a conformational change concomitant with UvrB hydrolysing ATP. As a result UvrA is released and UvrB remains bound to the lesion. UvrA acts as a 'molecular matchmaker' bringing the UvrB protein to the site of the lesion. UvrB kinks the DNA when bound, UvrC joins the complex, and as a result an incision is made in the 4<sup>th</sup> or 5<sup>th</sup> phosphodiester bond on the 3' side of the lesion. Which protein makes this cut is unknown at the present time. This first backbone break is followed directly by another incision at the 8<sup>th</sup> phosphodiester on the 5' side of the damage. UvrD (helicase II) unwinds the cleaved DNA so that it can be removed, which subsequently results in the loss of UvrC. UvrB remains associated with the gapped DNA and is released after the gap has been filled by DNA polymerase I. The final 3' nick is closed by DNA ligase (Figure 1.5).

### **Base excision repair (BER)** –based on (Mol *et al.*, 1999)

BER occurs in two stages, an initial damage-specific stage involves individual DNA glycosylases excising distinct base lesions. This is followed by a general stage where the generated abasic site (which may also arise through spontaneous depurination and depyrimidation caused by oxygen radical attack) is subjected to phosphodiester bond hydrolysis on the 5' side of the AP site by an AP endonuclease. This strand nick provides a free 3' –OH for DNA repair synthesis by DNA polymerase (DNA polymerase I in *E.coli*, DNA polymerase  $\beta$  in mammals). The abasic nucleotide is removed and replaced with the correct nucleotide using the 5'→3' exonuclease and synthesis activities of DNA polymerase. DNA ligase then seals the backbone nick (Figure 1.6).

The lesion-specific glycosylase activities come from a number of separate enzymes. Products of base deamination, alkylation, uv radiation damage and oxidation can be removed by specific glycosylases. For example, 3-methyladenine is cleaved from the sugar moiety by AlkA in *E.coli*, whilst the deamination product of cytosine, uracil is removed by Uracil-DNA glycosylase.



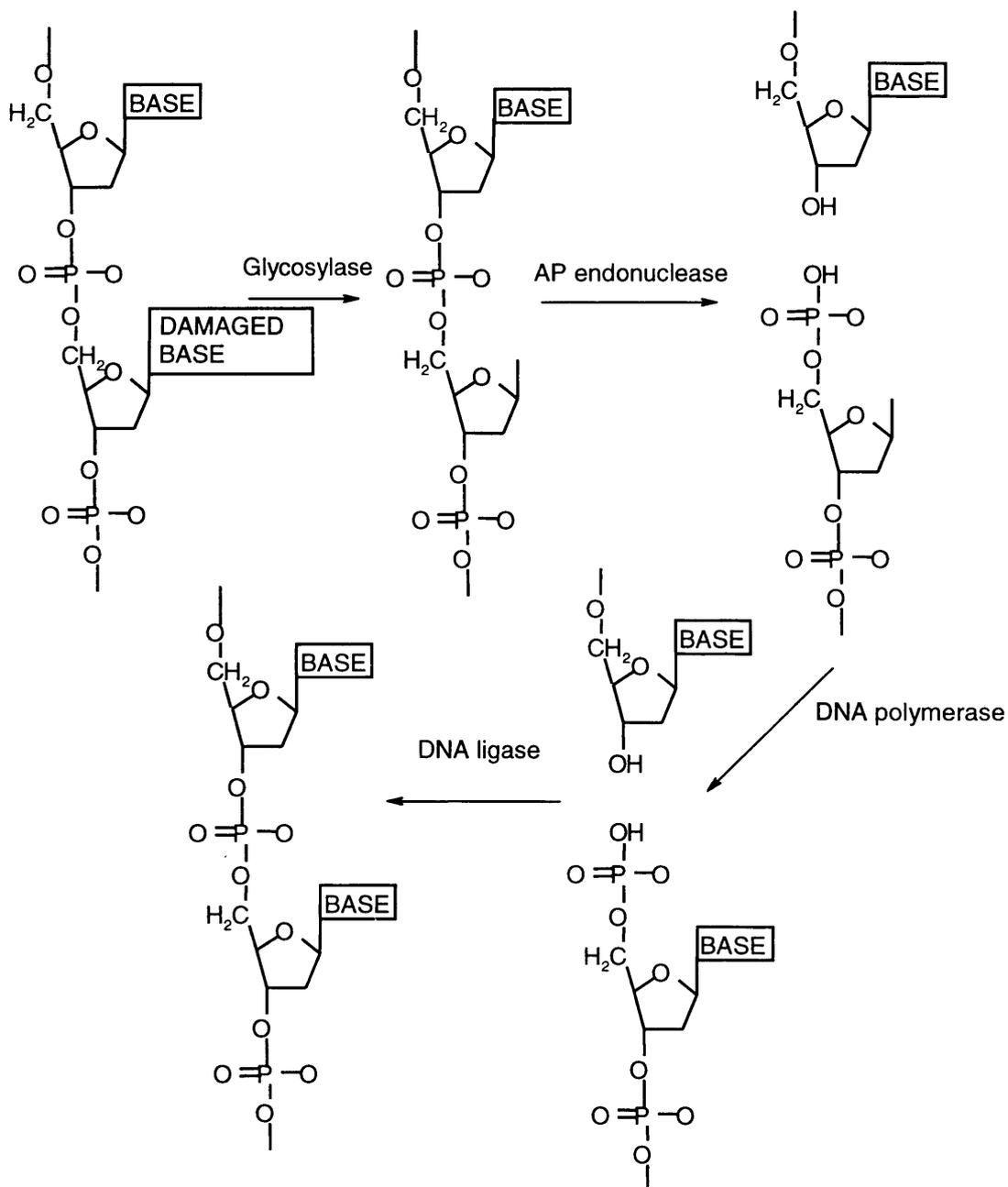
**Figure 1.5** A model of NER by the *E. coli* UVR system, based on Goosen *et al.* (1998). See text.

Many of the lesions repaired by this pathway cause minor deformations of the DNA backbone and base stacking. It is believed that the glycosylases use a rapid damage-scanning mechanism to probe for these deviations and locate the lesion. The deformability of the base stack aids the base-flipping process, which has been proven to occur for many of these enzymes and postulated to occur for many others. In this process the damaged base is usually flipped out of the duplex into the active site cavity where the glycosyl bond can be cleaved (or in the case of T4 endonuclease V, the base opposite the lesion flipped out).

Base-flipping improves specificity beyond direct recognition of DNA damage. The base can only be excised if it can be rendered extrahelical, form correct interactions within the enzyme active site, and be stabilised in its extrahelical conformation. This adds extra levels of

selectivity to the repair process during flipping in comparison with simple one step lesion recognition.

The exact mechanism of scanning for the lesion is unknown, several possibilities have been suggested. The protein may capture a spontaneously flipped-out base and exercise repair (Panayotou *et al.*, 1998), or may induce DNA strand separation (Vassilyev *et al.*, 1996) to flip the base. In the case of Uracil-DNA glycosylase an active flipping mechanism has been proposed (Slupphaug *et al.*, 1996). The enzyme is thought to flip the uracil base out of the duplex by compressing the DNA backbone, and inserting a leucine sidechain into the base stack. The base is pushed out of the helix by the leucine and pulled into the enzyme by



**Figure 1.6** The pathway for base excision repair. Only the initial glycosylase is lesion-specific. The rest of the pathway is general for all abasic sites.

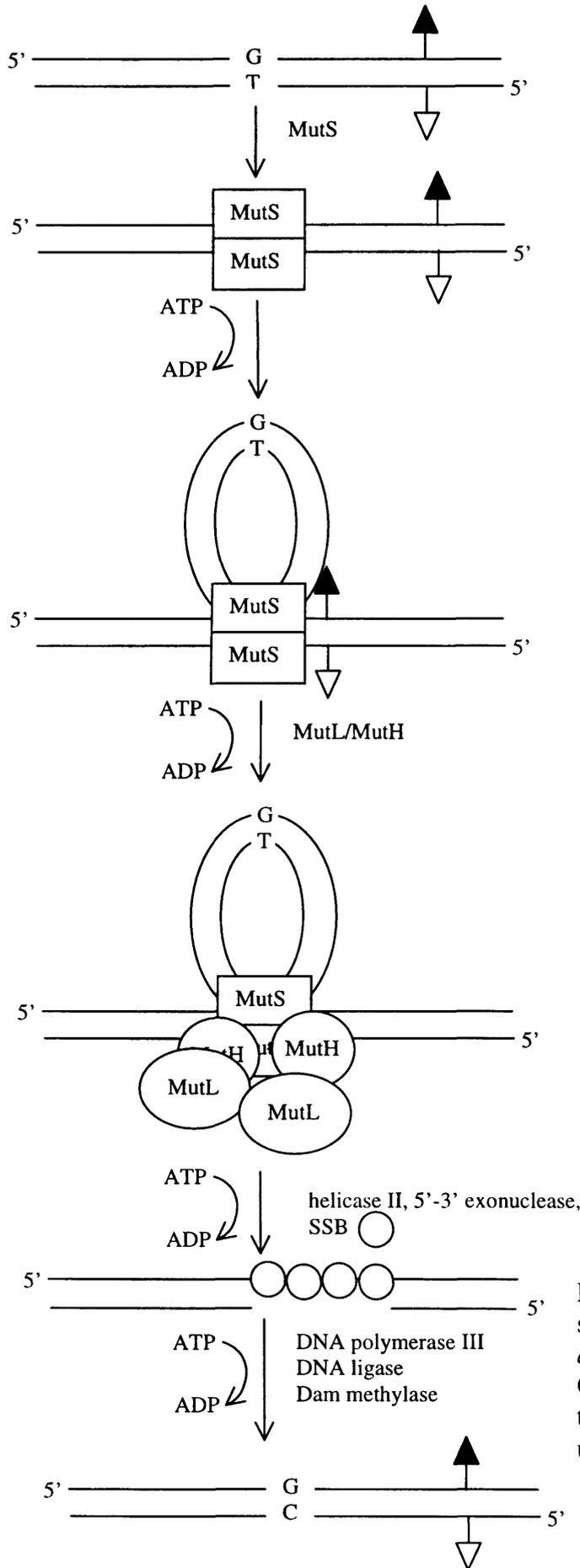
complementary active site residues in a so called 'push-pull' mechanism. Uracil-DNA glycosylase is thought to locate the lesion by scanning along the DNA, with the Leucine sidechain probing the minor groove. Distortions in the base pairing caused by the presence of uracil may be detected by this leucine, and also by the deformability of the DNA backbone surrounding the lesion. These factors would enable flipping to occur once the lesion was located.

An alternative scanning mechanism termed 'processive extrusion' has been proposed (Verdine *et al.*, 1997). Here a glycosylase enzyme binds DNA and randomly extrudes a base. The protein then processes along the DNA, alternately flipping out consecutive bases. Bases are sampled in the active site until a lesion is found where upon catalysis occurs. Such a mechanism would almost certainly be slower at locating lesions than if base-flipping only occurred for the damaged base.

#### **1.1.3.2 Mismatch repair** -based on (Schar *et al.*, 1998)

Mismatched base pairs such as G:T may be repaired by the mismatch repair pathway. In *E.coli* the system responsible is known as MutHLS, since it consists of the proteins MutH, MutS, and MutL. This system relies on the methylation of DNA at specific sites following replication in order to identify the newly synthesised strand, and hence the one containing the erroneous base of the mismatch. The scheme of repair is shown in Figure 1.7. Briefly, the mismatch is bound by a MutS homodimer, ATP is hydrolysed and the DNA is looped through the protein. An  $\alpha$ -loop forms whose limits are defined by a hemimethylated *dam* sequence (dGATC, where the parent strand contains N<sup>6</sup>-methyladenine, the newly synthesised daughter strand is not yet methylated). MutL and MutH then bind MutS in the complex. The function of MutL is not fully understood. It seems to aid the formation of the  $\alpha$ -loop, and the final MutS-DNA-MutL-MutH complex, making it another example of a 'molecular matchmaker'. Upon complexation, the endonucleolytic activity of MutH becomes activated and the protein incises the unmethylated DNA strand at the GATC site. An exonuclease then removes all the nucleotides from the GATC sequence to the mismatch and beyond. In Figure 1.7, the GATC sequence is 5' to the mismatch so a 5'-3' exonuclease such as exonuclease VII is used. Were the GATC sequence 3' to the mismatch, a 3'-5' exonuclease, such as the proofreading function of DNA polymerase III, could be used. The degradation process is facilitated by duplex unwinding by helicase II (UvrD), and the stretch of single-stranded DNA formed is stabilised by the binding of single-stranded binding protein (SSB). DNA polymerase III then fills in the gap, and DNA ligase seals the final nick. The corrected daughter strand is then methylated in the GATC sequence by Dam methylase.

Defects in the homologous mismatch repair system in humans have been linked to the onset of cancers in the hereditary non-polyposis colorectal carcinoma (HNPCC) syndrome. This was recently reviewed (O'Driscoll *et al.*, 1998).



**Figure 1.7** Mismatch repair by the MutHLS system in *E. coli* based on Figure 1 in Schar *et al.* (1998). For details see text. Methylated G(meA)TC sequences are shown as black triangles, unmethylated GATC shown as unshaded triangles.

### **1.1.3.4 Direct repair**

Direct repair mechanisms usually involve single enzymes which cause a direct reversal of the DNA modification without a requirement for resynthesis. One such example is DNA photolyase which repairs dipyrimidine dimers by harvesting light energy to break the cyclobutane ring. Repair of such dimers is effected in both single and double-stranded DNA. It is believed that the enzyme recognises the lesion on the basis of the disturbance it causes to the phosphodiester backbone, and also by direct contacts to the cyclobutane ring.

The repair of the premutagenic mispairing lesions O<sup>6</sup>-methylguanine and O<sup>6</sup>-ethylguanine is effected through the ubiquitous enzyme O<sup>6</sup>-methylguanine-DNA methyltransferase (MGMT) (EC 2.1.1.63) (Olsson *et al.*, 1980). Larger adducts are also repaired by MGMT, although as the size increases, the efficiency of removal by this enzyme decreases, and the *uvr* system in *E.coli* takes over the repair (Morimoto *et al.*, 1985), (Todd *et al.*, 1983). O<sup>6</sup>-methylguanine-DNA methyltransferase transfers the alkyl adduct from the O<sup>6</sup>-alkylguanine to an active site cysteine in a single-step, stoichiometric reaction, during which it is irreversibly inactivated (Robins *et al.*, 1979) ; (Lindahl *et al.*, 1982). This inactivation has implications for the cell, since during times of heavy alkylation damage, more MGMT will have to be synthesised to cope with the repair. In *E.coli* this is achieved through a positive feedback pathway known as the 'Adaptive Response' and is outlined in greater detail in the next section.

The exact mechanism by which MGMT locates and removes an alkyl group from a damaged base was unknown and is the subject of this thesis.

### **1.2 REPAIR OF ALKYLATED DNA BY THE *E.coli* ADA PROTEIN**

The *E.coli* Ada protein contains an O<sup>6</sup>-methylguanine-DNA methyltransferase repair activity for the repair of premutagenic O<sup>6</sup>-methylguanine and O<sup>4</sup>-methylthymine lesions, and also a methylphosphotriester repair activity. Two methylphosphotriester diastereoisomers are possible, Sp and Rp. In the Sp diastereoisomer the methyl group protrudes from the backbone of DNA whilst in the Rp diastereoisomer the methyl group is accommodated within the major groove. Ada only repairs the Sp diastereoisomer. This Sp-diastereoisomer might serve as a marker for more dangerous alkylation damage within the DNA, it is repaired in preference to the Rp-diastereoisomer, which is less accessible to repair proteins. No great distortion of the backbone by the methyl group is seen and this lesion is not mutagenic (Weinfeld *et al.*, 1985). By contrast, O<sup>6</sup>-methylguanine does cause minor distortions to the structure of DNA. The methylated base was noted to have reduced stacking interactions in the duplex (Wong *et al.*, 1992). These may give rise to kinks in duplex DNA of sufficient length (Voigt *et al.*, 1990). Perturbations of the phosphodiester backbone in O<sup>6</sup>meG:C base pairs were also observed (Patel *et al.*, 1986a), (Kalnik *et al.*, 1989b). Such disturbances were less obvious in the case

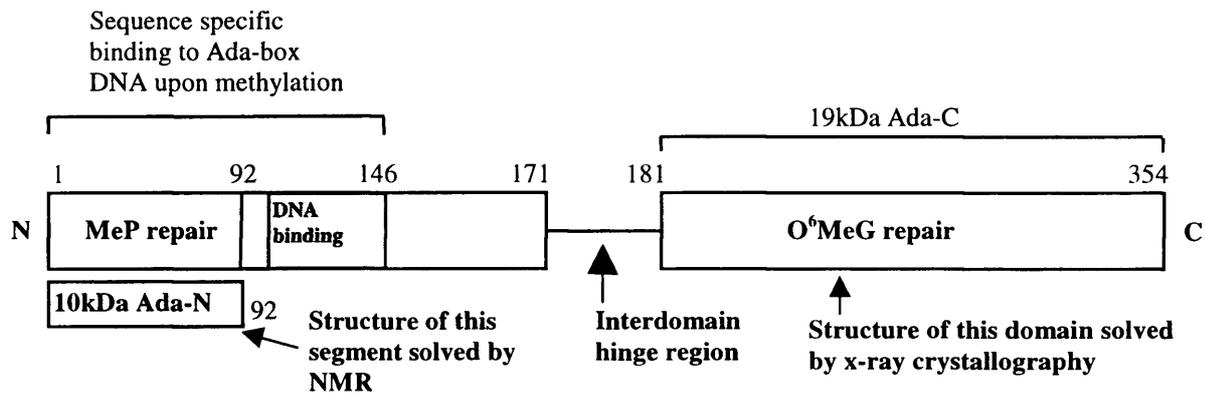
of the O<sup>6</sup>meG:T base pair (Patel *et al.*, 1986b), (Kalnik *et al.*, 1989a), suggesting that if such distortions are used as a guidance system for lesion localisation, repair by MGMT would be most efficient prior to DNA replication. Following replication and the incorporation of a mismatched thymine opposite O<sup>6</sup>meG, the mismatch repair apparatus of the cell would engage in futile cycles of removing the mispaired thymine, whilst leaving the O<sup>6</sup>meG unrepaired.

The reduced base stacking, and hydrogen-bonding potential lower the thermodynamic stability of base pairs involving O<sup>6</sup>-methylguanine, compared with unmethylated bases (Gaffney *et al.*, 1989); (Gaffney *et al.*, 1984). This ease of localised duplex melting surrounding the lesion may also contribute to the DNA repair mechanism, especially if base flipping were employed by MGMT. The protein may use the reduced stacking potential in the form of kinked DNA, and/or the phosphodiester backbone distortions to localise the lesions amidst a mass of competing genomic DNA.

### **1.2.1 The Role of Ada as Transcription Factor- The Adaptive Response**

The irreversible inactivation of MGMT following repair of O<sup>6</sup>meG or O<sup>4</sup>meT has implications for repair of further lesions. Turnover of MGMT in response to low levels of DNA alkylation is not expected to exert a great strain on the cell. However, under exceptional circumstances when the cell is exposed to high doses of alkylating agent *de novo* synthesis of MGMT occurs to comply with cellular demand (Samson *et al.*, 1977), (Robins *et al.*, 1979). *E.coli* contains two known forms of MGMT, a constitutive 19kDa Ogt protein, which is capable of repairing O<sup>6</sup>meG and O<sup>4</sup>meT lesions (Rebeck *et al.*, 1988) and the inducible 39kDa Ada protein which also contains a methylphosphotriester repair activity (Figure 1.8), (McCarthy *et al.*, 1985). An exponentially growing *E.coli* cell is estimated to possess two molecules of Ada and ten times more Ogt (Potter *et al.*, 1989). Due to its initially higher cellular concentration, the Ogt protein is probably the first line of defense against low levels of O<sup>6</sup>meG in DNA (Shevell *et al.*, 1988). As the level of alkylation damage increases, the Ogt molecules are expended and the expression of *ada* is induced. This can lead to an increase of up to 1000 times the original concentration of Ada and is termed the "adaptive response" (Robins *et al.*, 1979), (Figure 1.9).

Cells pretreated with low doses of methylating agent are more resistant to a further challenge with alkylating agent than non-"adapted" cells (Samson *et al.*, 1977). Thus the Ada molecule confers protection against chronic exposure to alkylating agents whilst Ogt deals with transient challenges.

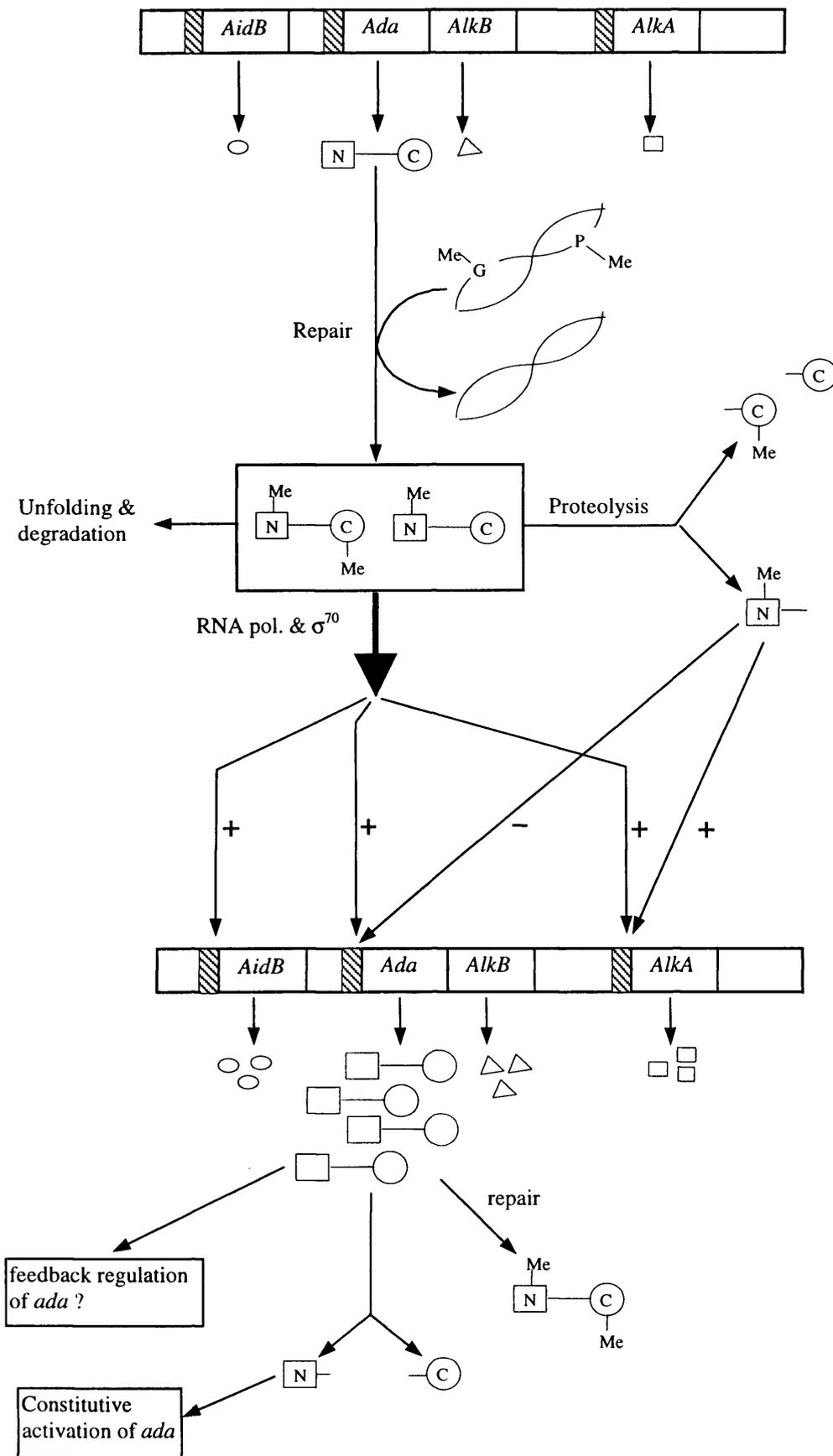


**Figure 1.8** The *E. coli* Ada protein. Zinc binding and Sp-methylphosphotriester repair activity resides within the first 92 residues of the 20kDa N-terminal domain (Ada-N). The structure of this N-terminal, 92 residue, 10k Da fragment was solved by NMR and is depicted as a separate rectangle Myers *et al.* (1993b). The predicted helix-turn-helix motif, and basic 2kDa segment implicated in DNA binding is shown in grey. O<sup>6</sup>-methylguanine and O<sup>4</sup>-methylthymine repair activities reside in the 19kDa C-terminal domain, separated from the 20kDa N-terminal domain by a protease-sensitive hinge region. This domain structure has been solved by x-ray crystallography.

The increase in Ada concentration is accompanied by a rise in expression of *alkB*, which forms an operon with *ada* (Kondo *et al.*, 1986). *AlkA* and *aidB*, whose products have roles in the response to alkylation damage, are also upregulated (Figure 1.9), (Lindahl *et al.*, 1988). *AlkA* is a 3-methyladenine glycosylase whose crystal structure has recently been solved (Yamagata *et al.*, 1996). *AlkB* encodes a 24kDa protein which has been isolated but remains of undetermined function (Wei *et al.*, 1995). *AidB* encodes a protein of ca. 60kDa. The predicted sequence showed significant homology to human isovaleryl-coenzyme A dehydrogenase. Its expression reduced the mutagenic effect of N-methyl-N'-nitro-N-nitrosoguanidine (MNNG) and thus it has been proposed to have a role in the detoxification of alkylating agents (Landini *et al.*, 1994).

### **1.2.1.1 The N-terminal domain of Ada provides the 'On-Switch' for the adaptive response**

The Ada protein irreversibly transfers methyl groups from the premutagenic lesions O<sup>6</sup>-methylguanine and O<sup>4</sup>-methylthymine to an active site Cys 321 residue located within the C-terminal domain of the protein (Olsson *et al.*, 1980), (Dempfle *et al.*, 1982). Methyl groups from methylphosphotriester Sp-diastereoisomers are also abstracted and irreversibly transferred to Cys 69 in the N-terminal domain of the protein (McCarthy *et al.*, 1985). Upon methylation at Cys 69, Ada is converted from weak to strong transcriptional activator of *alkB*, *alkA*, *aidB*, and also its own gene (Figure 1.9). This expression enhancement is promoted by specific



**Figure 1.9** Adaptive response to alkylating agents (adapted from Lindahl *et al.* 1988). Whilst methylation at the N-terminal domain alone is sufficient to transform the protein into a transcriptional activator, the C-terminal domain provides the key to differential regulation of the adaptive response genes. *Ada* methylated at the N-terminal domain specifically binds to the “Ada box” sequence (shown here as *hatched boxes*) in the promoter regions of adaptive response genes, enhancing their transcription by facilitating the binding of RNA polymerase (*pol*).

binding of the methylated Ada molecule to a DNA sequence known as the 'Ada box'. Footprinting studies showed that in the *ada* gene this is an eight nucleotide sequence (5'-AAAGCGCA) which lies upstream of the -35 box (Teo *et al.*, 1986), (Nakamura *et al.*, 1988). Another binding sequence (5'-AAAGCAA), was proposed for the *AlkA* gene based on the observation that the *ada* Ada box overlaps its -35 box (Akimaru *et al.*, 1990). Further studies on transcriptional activation of the *AidB* gene by Ada suggested only a broader consensus sequence 5'-AATnnnnnGCAA could account for the specific binding to all three promoters (Landini *et al.*, 1995). This recognition sequence is found immediately upstream of the -35 box. Its position is important for correct interactions between the Ada protein and RNA polymerase, whose binding to the *ada* promoter is known to be facilitated by the presence of the former protein (Sakumi *et al.*, 1989). Suggestions have been made that the C-terminal domain of Ada is responsible for mediating these contacts from Ada to RNA polymerase (Demple, 1986). It is however the methylated N-terminal domain of Ada which binds the Ada box, whether free, or as part of the full length Ada molecule. Activation of the *ada* and *alkA* genes occurs regardless of the methylation state of Cys 321, (the residue responsible for accepting the methyl moiety from O<sup>6</sup>meG), provided Cys 69 is methylated (Teo *et al.*, 1986). Recently it was revealed that *in vitro* transcription of *ada* and *alkA* is increased by high concentrations of unmethylated Ada (200-500 molecules per cell), as well as by low concentrations of Ada methylated at Cys 69. Exceeding these concentrations of unmethylated Ada leads to an unexpected inhibition of *ada* transcription by methylated Ada both *in vitro* and *in vivo*. The *alkA* gene is unaffected by these high levels of unmodified Ada (Saget *et al.*, 1994). Such seemingly high Ada concentrations are still physiologically relevant, as levels of 3000 Ada molecules per cell are believed to be attained during the adaptive response (Robins *et al.*, 1979).

#### **1.2.1.2 The C-terminal domain provides the key to differential regulation of adaptive response genes**

The C-terminal domain of Ada plays an important role in regulating the adaptive response. Unmethylated Ada derivatives lacking 10-20% of their C-termini, (missing the O<sup>6</sup>-methylguanine-DNA methyltransferase activity) are constitutive activators of *ada*, (albeit less effective activators than methylated Ada) but remain inducible activators of the *alkA* gene (Shevell *et al.*, 1991), (Saget *et al.*, 1994). Ada lacking the entirety of the C-terminal domain, (Ada-N) fails to bind the Ada box in *AlkA*, and does not stimulate RNA polymerase binding (Akimaru *et al.*, 1990). This suggests that the first half of the C-terminal domain (first 92 residues) are important in enhancing transcription, perhaps by initiating contacts to

RNA polymerase. Methylated Ada-N binds to the Ada box, but fails to induce transcription of the *ada* gene. It acts instead as an inhibitor which down-regulates the response, again indicating the requirement for the presence of the C-terminal domain for transcription activation. Ada-N also binds the Ada box of *AlkA*, but in this case the transcription enhancing capabilities are retained (Figure 1.9). Changing the hinge region (Figure 1.8) Lys-178 to Pro also resulted in an Ada derivative deficient in *ada*, but not *alkA* transcription (Saget *et al.*, 1995). This demonstrates the differential regulation of *ada* and *alkA* by Ada, and indicates that the activation of *ada* requires the two domains, whilst that of *alkA* is more predominantly dependent on the N-terminal domain alone. The different response to Ada regulation by *ada* and *alkA* has been attributed to differences in Ada box, and promoter organisation in the two genes (Saget *et al.*, 1994).

The mechanism of Adaptive response termination is unknown. Originally it was suspected that cleavage of Ada at the Lys178-Gln179 bond in the hinge region and subsequent separation of the two domains shut off the adaptive response. This hypothesis has so far been unsupported, as no intracellular protease capable of performing this proteolysis has been identified (Sedgwick, 1989). Observations that methylated Ada-N fragments were all more prone to irreversible unfolding than the non-methylated forms, led to the suggestion that as well as proteolysis, unfolding of the transcription-enhancing N-terminal domain may serve to down regulate transcription from the *alkA* promoter (Myers *et al.*, 1994). This may also be applicable to Ada-C which has been observed to spontaneously lose tertiary structure upon methylation (Chapter 6). An alternative hypothesis, that unmethylated Ada reaches an inhibitory threshold which prevents further *ada* transcription, has also been entertained (Saget *et al.*, 1994).

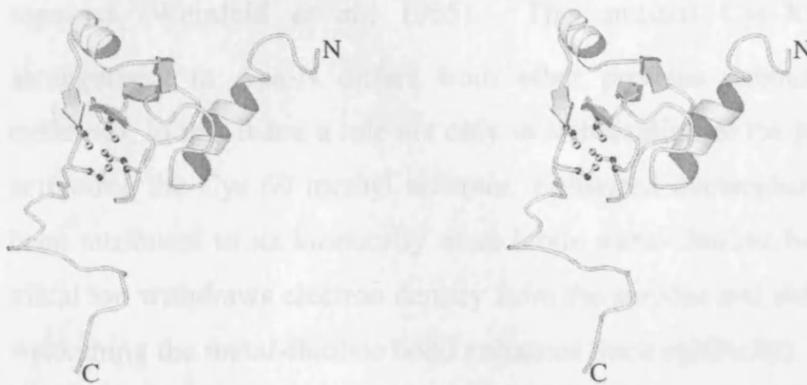
### **1.2.1.3 An alternative pathway to upregulation of the adaptive response genes**

It has been demonstrated that in addition to transcription induction in response to methylated Ada, expression of the *aidB* and *ada* genes may be enhanced upon entry into stationary phase. This enhanced transcription makes use of the alternative sigma factor RpoS (Taverna *et al.*, 1996), instead of the sigma 70 subunit of RNA polymerase used for *ada* and *alkA* transcription during exponential cell growth (Sakumi *et al.*, 1989). Transition to stationary phase and RpoS thus provide an alternative route to increased cellular concentrations of Ada and AidB.

## 1.2.2 Structures of Ada and Relation to Catalytic and Transcriptional Roles

### 1.2.2.1 The N-terminal Methylphosphotriester Repair Domain of Ada

Structural and functional studies have been done on the isolated 20kDa N-terminal domain of Ada, (Ada-N) which retains the Sp-methylphosphotriester repair activity and the capability to bind the Ada Box in its methylated form. This domain was shown to bind a single zinc ion, the presence of which was a requirement for the correct folding and maintenance of the folded state, and which is tetrahedrally coordinated to four conserved Cysteine residues (Myers *et al.*, 1992). One of these is the methyl acceptor Cys 69 (Myers *et al.*, 1993a). The solution structure of the first 92 residues, comprising a 10kDa fragment of Ada-N (Myers *et al.*, 1993b), revealed a novel overall fold comprising two  $\alpha$  helices and a  $\beta$  sheet with mixed parallel and antiparallel strands (Figures 1.8 & 1.10).



**Figure 1.10** Stereo image of the solution structure of the N-terminal 10kDa fragment of Ada-N solved by Myers *et al.* (1993b). A single  $Zn^{2+}$  ion is coordinated by four Cysteine residues, one of which is the methyl acceptor Cys 69.

Whilst the Sp-methylphosphotriester repair activity is located in the first 92 residues of Ada-N, the capacity to bind specifically to the Ada-box following methylation is absent. Through work with fragments of Ada-N, a basic region spanning residues 130 to 146 was implicated in specific binding to the Ada-box, following repair and self-methylation (Sakashita *et al.*, 1993). However, the fact that normal rates of DNA repair could occur without this basic segment means that some capacity to bind DNA non-specifically must be retained in the first 92 residues. Consistent with this, NMR evidence has implicated four residues in this region of Ada-N in binding non-specifically to DNA. Two of these residues were solvent-exposed arginines which presumably interact with the phosphodiester backbone of DNA during generalised binding for repair (Myers *et al.*, 1993b).

### **1.2.2.2 Mechanism of methylation-dependent switch to transcriptional regulator:**

The mechanism of the Ada protein's methylation-dependent switch to sequence-specific DNA-binding is still under investigation. Methylation of Cys 69 is required. It is apparent that zinc coordination is maintained upon methylation, and that no large conformational change in protein structure occurs. The sequence-specific switch is believed to be due to direct interactions between the Cys 69 methyl group and the Ada box DNA (Myers *et al.*, 1994), (Ohkubo *et al.*, 1994). A helix-turn-helix motif is thought to exist in a basic segment of Ada-N (Figure 1.8). This is believed to be required for binding DNA both specifically (to the Ada box) and non-specifically (to random DNA).

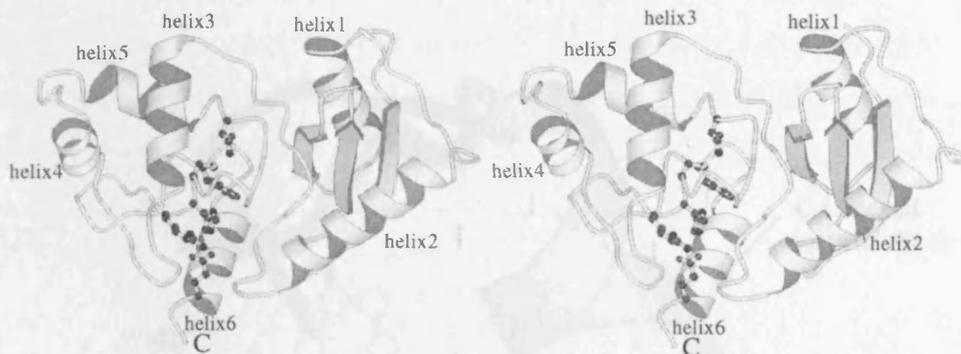
### **1.2.2.3 Mechanism of methyl abstraction**

The Cys 69 nucleophile of Ada-N is inaccessible to the Rp-diastereoisomer which projects inwards towards the major groove of DNA which is why only the Sp-diastereoisomer is repaired (Weinfeld *et al.*, 1985). The unusual Cys-X<sub>3</sub>-Cys-X<sub>26</sub>-Cys-X<sub>2</sub>-Cys ligand arrangement in Ada-N differs from other proteins containing zinc coordinated by four cysteines, in that it has a role not only in stabilisation of the protein conformation, but also in activating the Cys 69 methyl acceptor. Enhanced nucleophilicity of the Cys 69 thiolate has been attributed to its kinetically more labile metal-thiolate bond (Wilker *et al.*, 1997). (The metal ion withdraws electron density from the sulphur and reduces thiolate nucleophilicity, so weakening the metal-thiolate bond enhances nucleophilicity). The proposed model is that the coordination to zinc of the Cys 69 thiol initially serves to aid its deprotonation by lowering the pK<sub>a</sub> (Myers *et al.*, 1995a). The thiolate nucleophile is then protected from protonation by zinc coordination, whilst in the transiently dissociated state it is ready to perform the alkyl transfer reaction.

### **1.2.2.4 The C-terminal O<sup>6</sup>-methylguanine-DNA methyltransferase domain of Ada**

This 19k Da domain spans the region Met 175 - Arg 354 of the Ada protein and contains the O<sup>6</sup>-methylguanine-DNA methyltransferase activity, (MGMT; EC 2.1.1.63). As for the N-terminal domain, methyl transfer is a stoichiometric, single-step and irreversible reaction that leaves the DNA in its original unmodified state and inactivates the enzyme. The reaction proceeds without the requirement for additional cofactors (Lindahl *et al.*, 1988). Methyl transfer occurs to the highly conserved Cys 321 residue which lies in the conserved PCHR sequence (Dempfle *et al.*, 1985). This sequence is highly conserved between different alkyltransferases (Pegg *et al.*, 1995), and is similar to the sequence surrounding the acceptor cysteine in the N-terminal domain of Ada (PCKR (Sedgwick *et al.*, 1988). In each case this sequence is preceded by a short stretch of hydrophobic residues. Given such striking

conservation it was suggested that the PCH sequence reflected a common mechanism for generating the reactive cysteine nucleophile (Dempfle *et al.*, 1985). However no structural similarity has been found.



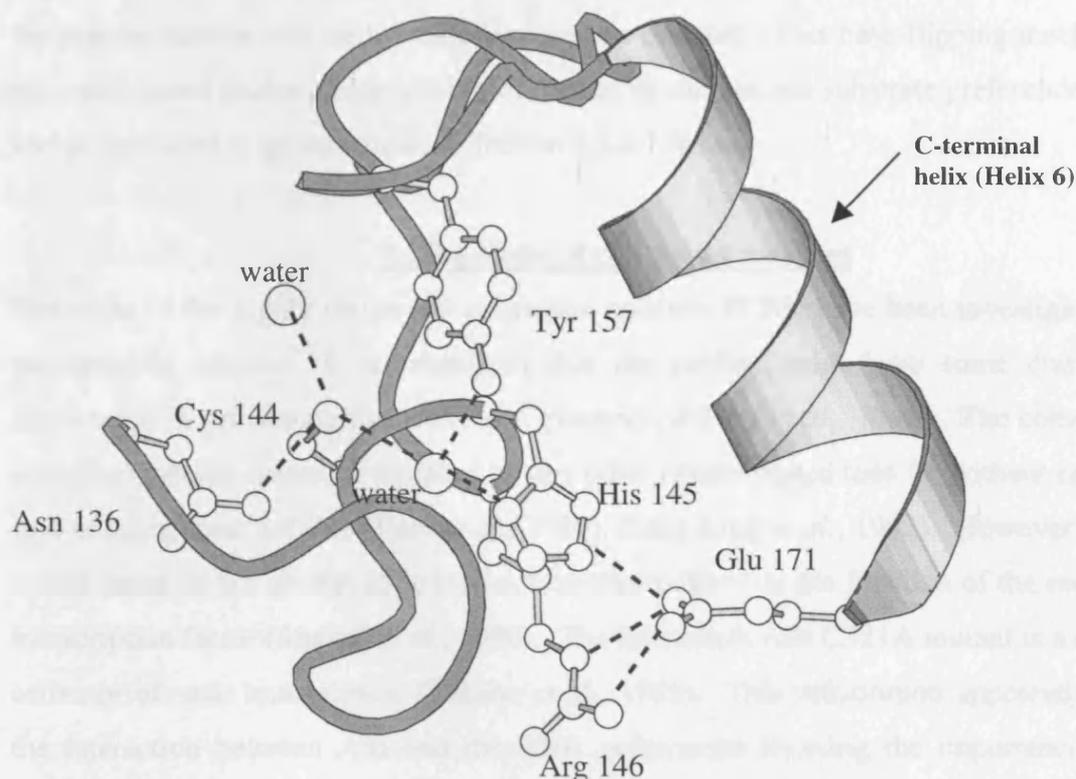
**Figure 1.11** Stereo image showing the crystal structure of Ada-C (Moore, *et al.* 1994). Sidechains of active site residues discussed are displayed and helices labeled in accordance with usage in the text. Helices 3, 4 and 5 form the Helix-turn-Helix motif, with helix 5 acting as the recognition helix.

#### **1.2.2.5 Enzyme structure and postulated conformational change upon binding DNA**

The high resolution crystal structure of the C-terminal fragment corresponding to residues Thr-176 to Arg-354, (renumbered Thr-1 to Arg-178 by Moore *et al.*, 1994) revealed two distinct sub-domains. The N-terminal sub-domain comprising the first 92 residues was noted to have a fold reminiscent of part of ribonuclease H (Moore *et al.*, 1994). However, the high structural homology to integrase and transposase enzymes was only discovered during this work, and is detailed in Chapter 6. The C-terminal sub-domain has novel topology consisting exclusively of helix and coil (Figure 1.11). Helices 3, 4 and 5 form a helix-turn-helix, (HTH) motif similar to those noted in a variety of prokaryotic and eukaryotic DNA-binding proteins (Schwabe *et al.*, 1993). Were it to bind DNA in an analogous manner to other HTH variant proteins one might expect helix 5 to act as a recognition helix and lie within the major groove of DNA.

Surprisingly, the crystal structure showed that the methyl acceptor residue, Cys 144 was buried within the protein, rendering it inaccessible to stacked alkylated bases in DNA (Figure 1.11). Only the sidechain of Asn 136 and two ordered water molecules are within hydrogen bonding distance to the Cys 144 sulfhydryl group (Figure 1.12). It was suggested that one of the water molecules may represent the position occupied by the exocyclic oxygen of the O<sup>6</sup>meG-DNA substrate. Indeed, one water molecule is hydrogen bonded to both the N $\delta$  of the conserved His 145 and the O $\eta$  of conserved Tyr 157. In turn the His 145 N $\epsilon$  is hydrogen bonded to the sidechain of the conserved Glu 171 on the C-terminal helix. This hydrogen bonding network involving the tetrad Cys 144 - water - His 145 - Glu 171, is reminiscent of

that found in the active sites of cysteine and serine proteases and lipases where the acid and histidine stereochemistry is used to generate the reactive nucleophile (Blow, 1990). Due to its inaccessibility, a large conformational change in Ada-C was postulated to allow exposure of the active site to the substrate (Moore *et al.*, 1994).



**Figure 1.12** The active site region of Ada-C. Sidechains of conserved residues and possible hydrogen bonds are shown (Moore, *et al.* 1994).

The suggestion was that the C-terminal helix of Ada-C could swivel out and be accommodated within the major groove, thus exposing the active site residues and bringing the Cys 144 thiol into close proximity with the O<sup>6</sup>meG lesion. In such a scheme the protein would cover ~8bp of DNA which is consistent with the results of spectroscopic analysis of the DNA-protein complex (Takahashi *et al.*, 1990); (Chan *et al.*, 1993). Such a movement would necessarily break the salt link between Glu 171 and His 145. This would enable the His 145 to rotate about its  $\chi_1$  and  $\chi_2$  angles such that the N $\delta$  could replace an ordered water molecule, only 3.2Å from the S $\gamma$  of Cys 144, making it ideally located to generate the reactive thiolate anion. In this scheme a change in DNA structure from B-DNA is not required, although it was not ruled out by the authors.

There are several failings of this model. The first is that the HTH motif is uninvolved in DNA-binding. Analysis of the DNA binding and repair properties of the homologous human MGMT indicated that the HTH region of the protein was necessary for correct DNA-binding (Kanugula *et al.*, 1995). Swinging out of Helix 6 would expose the hydrophobic core of the

protein just prior to docking with DNA. This would not be energetically favorable, and is therefore also unlikely.

It is therefore believed that the protein does not undergo this large conformational change. Instead Ada-C was postulated to access the O<sup>6</sup>meG base by rendering it extrahelical from the duplex of DNA (Dempfle, 1995). The base could then enter the active site through a hole in the protein surface and methyl transfer could be effected. This base-flipping mechanism was proposed based on the inaccessibility of the active site and the substrate preferences of Ada-C and is discussed in greater depth in Section 1.2.2.7 below.

### **1.2.2.6 Roles of conserved residues**

The roles of the highly conserved active site residues PCHR have been investigated through mutagenesis studies. It is postulated that the proline must have some conformational importance in orientating the active site cysteine (Wibley *et al.*, 1995). The conserved alkyl acceptor cysteine cannot be replaced by any other residue tested thus far without concomitant loss of transferase activity (Tano *et al.*, 1989), (Ling-Ling *et al.*, 1992). However, alterations at this point do not always have such a deleterious effect on the function of the molecule as a transcription factor (Shevell *et al.*, 1990). The full-length Ada C321A mutant is a constitutive activator of *ada* transcription (Takano *et al.*, 1988). This substitution appeared to enhance the interaction between Ada and the RNA polymerase showing the importance of the C-terminal domain as well as the N-terminal domain for transcription enhancement (Sakumi *et al.*, 1989).

The suggested role of the His 145 is to deprotonate the thiol (Dempfle *et al.*, 1985). Examination of the crystal structure revealed that His 145 and Arg 146 are directed away from the active site due to the formation of a hydrogen bond to Glu 171, interactions which force Cys 144 into the active site (Wibley *et al.*, 1995). Given that the orientation of the active site residues may alter upon interaction with DNA, the position of this histidine may rearrange as described in Section 1.2.2.5, making it more favourably located to abstract a proton from the thiol. Alternatively, sidechain rotations of His 145 may not be necessary, since a proton might be abstracted from the Cys 144 thiol by acting through an intermediate water molecule (Figure 1.12). The thiol may even be deprotonated without the presence of His 145 since substitutions for this histidine in human MGMT, by residues whose sidechains would be unable to generate the thiolate anion, (Phe, Met, Asn, Gln) lead to proteins that retained a low level of activity (Ling-Ling *et al.*, 1992). His 145 probably serves to enhance the rate of thiol deprotonation through the water intermediate, and hence enhances the demethylation rate.

Mutating the conserved Glu 172 in human MGMT, (Glu 171 in Ada-C), to Gln lead to deficient DNA binding and alkyl transfer activity (Rafferty *et al.*, 1994). This residue hydrogen-bonds to the active site His and Arg, and may play a role in orientating the acceptor Cys correctly (Wibley *et al.*, 1995). Upon alkylation of the acceptor cysteine in the human MGMT, an epitope lying on what would be the region between helices 3 and 4 in the analogous Ada-C, is exposed. The protein is also rendered sensitive to cleavage at Glu 70 and Glu 172 by protease V8 (Ayi *et al.*, 1994), (Oh *et al.*, 1996). It was postulated that this alkylation led to disruption of the Glu-172 to His-146 salt link (numbering for human alkyltransferase, equivalent residues in Ada-C are H145 and E171) suggesting its importance in maintaining the protein in an active conformation.

### **1.2.2.7 Mechanism of action**

Reaction rates of different O<sup>6</sup>-alkylguanines in DNA with various MGMTs support an S<sub>N</sub>2 mechanism of alkyltransfer to active site cysteine (Spratt *et al.*, 1994); (Spratt *et al.*, 1992). As the alkyl group size diminishes, the rate of reaction with both Ada and rat liver alkyltransferase increases (methyl>ethyl>*n*-propyl>*iso*-propyl>*iso*-butyl) (Morimoto *et al.*, 1985), (Graves *et al.*, 1989). During an S<sub>N</sub>1 mechanism the alkyl group must stabilise a positive charge which forms on the carbocation intermediate. This made an S<sub>N</sub>2 mechanism more probable given the reactive order of the different alkyl groups. Steric hinderance at the active site is another major factor affecting the reaction rates of MGMTs with O<sup>6</sup>-substituted guanines (Pegg *et al.*, 1993). Reaction with the free base O<sup>6</sup>-benzylguanine by human MGMT and “humanised” mutant Ada-C (lacking the steric hinderance to the active site of wild type Ada-C) could be enhanced by the addition of DNA to the reaction mixture (Goodtzova *et al.*, 1997); (Goodtzova *et al.*, 1994). This demonstrated that steric hinderance could be partially relieved by a change concomitant with DNA-binding, perhaps a small conformational change.

Neutralization of the guanyl moiety succeeding methyl transfer would create a better leaving group. Work on the repair of O<sup>6</sup>meG analogues is consistent with the proposal that following methyl transfer, the protein protonates the O<sup>6</sup> oxygen to neutralize the leaving group (Spratt *et al.*, 1994); (Spratt *et al.*, 1992). The residue involved in the protonation is as yet unidentified but has been postulated to be the conserved Tyr 157 in Ada-C (Wibley *et al.*, 1995). The N<sup>1</sup> ring nitrogen is involved in hydrogen bonding interactions with the protein as is the exocyclic amino group in the free nucleoside (N<sup>2</sup>). When in duplex DNA, this amino group is involved in hydrogen bonding to the opposite base. The finding that removal of this group lead to a repair rate increase, suggested that base pair separation maybe part of the mechanism (Spratt *et al.*, 1994). Such separation may occur upon base flipping.

MGMTs bind single-stranded DNA preferentially to duplex (Lindahl *et al.*, 1982), (Fried *et al.*, 1996). However alkylated duplex DNA reacts more rapidly with MGMT than single-stranded DNA (Lindahl *et al.*, 1982), (Scicchitano *et al.*, 1986), suggesting that if strand separation does occur it is unlikely to be the rate-limiting step (Fried *et al.*, 1996). It is reported that the R128A human MGMT mutant (the equivalent residue in Ada-C is R127 located in Helix 5) shows reduced repair efficiency and attenuated binding to single-stranded DNA, but retains the capability to bind duplex DNA. This implicates a greater role for this conserved Arginine (on the recognition helix of the HTH) in binding single-stranded than duplex DNA (Kanugula *et al.*, 1995).

Base flipping of the target O<sup>6</sup>meG or other bases in the region of the lesion is also consistent with experimental data and has been suggested as a potential mechanism by which Ada-C may act (Demple, 1995), (Kanugula *et al.*, 1995). This well documented phenomenon occurs with a number of DNA modification and repair proteins, such as the DNA (cytosine-5-) methyltransferases M.HhaI and M.HaeIII, human Uracil-DNA glycosylase, and T4 endonuclease V (Klimasauskas *et al.*, 1994), (Reinisch *et al.*, 1995), (Slupphaug *et al.*, 1996). The most common reason to flip is to render the base accessible to the active site residues. This may mean flipping the target base into a deep cavity containing the active site as in the case of Uracil-DNA glycosylase (Mol *et al.*, 1995), or flipping the base opposite the target base to allow entry into the duplex by the active site residues as for T4 endonuclease V (Vassylyev *et al.*, 1995). Thus Ada-C may flip the O<sup>6</sup>meG into its active site, or may flip out the base opposite the lesion and to enable a proteinaceous invasion of the duplex.

Recently a model of how Ada-C might bind DNA and flip out O<sup>6</sup>meG was devised (Vora *et al.*, 1998). This model was based on how Mu transposase (MuA) (which bore some sequence homology to Ada-C) bound and flipped DNA. Ada-C was positioned such that the recognition helix of the helix-turn-helix motif (Helix 5) lay in the major groove. O<sup>6</sup>meG was flipped out of the double helix, through an opening in the surface of Ada-C and into the active site. This was made possible without the requirement for a large conformational change in the protein. DNA was also contacted by a loop region of the protein termed the 'wing' which was made up of residues 149-155. Ada-C was therefore defined as employing a 'helix-turn-wing' motif in the binding of DNA. This model was attractive since it allowed the active site to gain access to O<sup>6</sup>meG without the need for a large, and energetically unfavorable conformational change in the protein. There was however, no structural experimental evidence to back up this postulated binding model.

### **1.2.2.8 Sequence specificity and target site identification**

How Ada and homologous MGMTs *recognise* and repair the alkylated base lesions lying amongst the mass of unaltered cellular DNA with such efficiency remains unclear.

Differential repair rates for single-stranded compared with duplex substrate DNA, and the failure to repair O<sup>6</sup>-methylguanine lesions in Z-DNA (Boiteux *et al.*, 1985) suggest that the conformation of substrate DNA affects repair rates of MGMT. All O<sup>6</sup>-methylguanine lesions in B-form DNA are repaired faster than the free O<sup>6</sup>-methylguanine base (Yarosh *et al.*, 1986). This supports the notion that DNA-binding aids methyl transfer, perhaps by enhancing binding affinity for the base, or by inducing a conformational change in the protein or DNA. There appears to be some dependence of repair rates upon oligonucleotide substrate length with the optimum being a pentanucleotide (Liem *et al.*, 1993). This is in accordance with the recently discovered increase in binding density for human MGMT as the length of single-stranded oligonucleotides decreases (Fried *et al.*, 1996). Indeed, binding of the human MGMT to single-stranded DNA is now believed to be cooperative (Fried *et al.*, 1996), although both Ada and human MGMT have been shown to be monomeric in solution (Takahashi *et al.*, 1990), (Bhattacharyya *et al.*, 1990). This contrasts to the earlier finding of non-cooperativity for Ada binding duplex DNA (Takahashi *et al.*, 1990).

Repair efficiency has been noted to have some slight dependence on the nucleotide sequence surrounding the lesion (Topal *et al.*, 1986), (Dolan *et al.*, 1988), (Georgiadis *et al.*, 1991). Attempts have been made to link this weak sequence specificity with stacking interactions between O<sup>6</sup>meG and neighbouring bases (Wong *et al.*, 1992), (Liem *et al.*, 1993). It has been demonstrated that human DNA-(cytosine-5-) methyltransferase preferentially methylates cytosine in the O<sup>6</sup>meG:C base pair which has been shown to be more unstacked from the duplex than in a G:C base pair (Tan *et al.*, 1990), (Wong *et al.*, 1992). This ties in well with the known base flipping mechanism of this enzyme. There are also indications that the O<sup>6</sup>meG lesion causes an unstacking of its neighbouring bases, the degree of which is sequence dependent. Increased unstacking appears to enhance the repair efficiency by human MGMT (Wong *et al.*, 1992), (Liem *et al.*, 1993) lending support to a suggested mechanism involving local duplex melting (which would include base flipping). Any sequence specificity caused by altered stacking interactions is slight, and no true discrimination based on nucleotide sequence surrounding the lesion occurs (Bender *et al.*, 1996); (Fried *et al.*, 1996). This is to be expected if Ada is to repair alkylated DNA from all parts of the genome with high efficiency. The rate of repair is almost at the limit for a diffusion controlled reaction *in vitro*, (Graves *et al.*, 1989), (Pegg *et al.*, 1995).

Exactly how the protein locates the target lesion buried within the mass of competing genomic DNA with such high efficiency is at present unknown. Reducing the dimensionality

of diffusion would enhance localisation rates. Many DNA-binding proteins are believed to achieve rapid rates of localisation of their cognate DNA sequences through one-dimensional diffusion along the DNA (von Hippel *et al.*, 1989). This 'sliding' mechanism along the DNA, although theoretically attractive, is believed not to occur in the case of human MGMT, and by inference also Ada-C (Bender *et al.*, 1996). It is possible that instead Ada-C reduces dimensionality by 'hopping' on and off the DNA in a series of minicollisions, during which it could translocate along the DNA length. Although not as effective at enhancing the lesion localisation rate, this method would certainly increase the speed with which an O<sup>6</sup>meG lesion was found in the genomic DNA relative to three-dimensional diffusion within the cellular (or nuclear) compartment.

The question remains as to how Ada-C recognises a lesion. Ada has a higher affinity for methylated than non-methylated DNA (Takahashi *et al.*, 1990), suggesting that the protein recognises some structural aspect of the O<sup>6</sup>meG and O<sup>4</sup>meT lesions. Direct recognition of O<sup>6</sup>meG within the DNA duplex would seem unlikely since there is insufficient distinction between this, and normal bases to attract Ada-C. Indirect recognition of the lesion through structural deviations is more probable. This recognition would almost certainly involve the phosphodiester backbone of DNA. Evidence suggests that slightly more distorted backbones caused by wobble O<sup>6</sup>meG:C base pairs allows more efficient repair than Watson-Crick O<sup>6</sup>meG:T pairs (Graves *et al.*, 1989). Altered stacking interactions surrounding the lesion may give rise to transiently flipped out bases or local duplex melting, features which may also be recognised by Ada.

Alternatively, lesion recognition may occur within the active site of Ada-C. This would enable the lesion to be directly probed, but would necessitate it to be flipped out of the DNA duplex first. Verdine and Bruner ((Verdine *et al.*, 1997)) proposed a model for recognition of lesions that cause little DNA destabilisation by DNA glycosylases. The enzyme binds to the DNA randomly and extrudes a base. Both protein and extrahelical base then process along the DNA. Alternate bases are sampled in the active site until a lesion is found where upon catalysis occurs. This model, termed "processive extrusion" may also be applicable to Ada-C when scanning the genome for O<sup>6</sup>meG lesions.

Yet another mechanism of lesion localisation was proposed by Fried *et al.* (Fried *et al.*, 1996). These workers suggested that cooperative binding of human MGMT occurred to the single-stranded regions of DNA which form in advance of the replication complex. This would enable a scanning of the genome without the protein molecules diffusing along the DNA.

The work presented in this thesis has helped to clarify the mechanism by which this remarkable protein recognises and repairs alkylated DNA within the bacterial genome.

### **1.3 AIMS OF THIS THESIS**

At the time of commencement of this research, the structure of Ada-C had been solved, but the manner in which it bound DNA, localised an O<sup>6</sup>meG lesion, and abstracted the methyl group was unknown. A model of DNA-binding involving a large conformational change in the protein had been proposed. This was not however based on any experimental evidence.

The aim of the PhD research summarised in this thesis was to determine the mechanism by which Ada-C bound DNA. This main aim can be dissected into how Ada-C binds unmethylated DNA, and how it binds DNA containing O<sup>6</sup>meG lesions, with a view to deciphering the demethylation mechanism.

Efforts were mainly devoted to the structural side of the protein-DNA interaction. Valuable thermodynamic information was also obtained to complement the structural work. Using structural and thermodynamic data a model of how Ada-C interacts with DNA was built up.

To achieve this goal Nuclear Magnetic Resonance (NMR) spectroscopy was employed to map the DNA-binding site on Ada-C, whilst Isothermal Titration Calorimetry (ITC) revealed the thermodynamics of the interaction. A high resolution picture of the interaction could have been obtained by solving the co-crystal structure of Ada-C complexed with DNA. Attempts were made without success to obtain such co-crystals, hence the use of NMR, and model building instead.

In order to perform any such biophysical analyses pure protein must be readily available. Due to the requirement to incorporate expensive isotopic labels, such as <sup>13</sup>C into the protein when using NMR, the yield of protein, per litre of labeled media must be maximised. In the case of Ada-C this was achieved through a combination of rational media design, strain choice and cellular growth conditions. A novel method termed 'cellular resuspension' was employed to maximise the biomass of an *E.coli* culture prior to addition to isotopically enriched media. This method led to an increased production of Ada-C per litre of labeled (isotopically enriched) media. The original purification protocol for Ada-C produced unacceptable yield losses and hence a new protocol designed. The results of yield enhancements made with Ada-C are covered in the Chapter 3. All methods are detailed in Chapter 2.

Chapter 4 covers the thermodynamics of the interactions between Ada-C and both unmethylated DNA, and DNA containing O<sup>6</sup>meG. Information such as binding constants, and enthalpy changes of binding, and demethylation were obtained using Isothermal Titration Calorimetry.

Chapter 5 reveals the results of the backbone chemical shift assignment of Ada-C. These backbone assignments were necessary in order to map the DNA-binding site on Ada-C using NMR. Assignments of backbone amides represent a map of the protein. Binding Ada-C to DNA alters the chemical environment of certain backbone amides. These may be directly

involved in contacts to the DNA, or altered as a result of secondary interactions from other parts of the protein.

Monitoring which chemical shifts alter upon DNA-binding, enables the binding site to be located. Obtaining the chemical shifts of the Ada-C backbone must therefore occur if such mapping is to be carried out.

Chapter 6 continues the mapping theme, revealing the results of the actual DNA-binding studies by NMR. Along with the chemical shift assignments obtained in Chapter 5, these results enabled the mapping of the DNA-binding site on Ada-C. Also included in this chapter are interactions between inactive C144S mutant Ada-C and DNA containing O<sup>6</sup>meG, as well as demethylation reactions between wild type Ada-C and methylated DNA.

Chapter 7 examines the interactions between Ada-C and DNA from the perspective of the DNA rather than the protein (which was focused on in Chapters 5 and 6). To this end <sup>19</sup>F NMR was employed with DNA oligonucleotides containing 5-fluorocytosine which acted as a probe of the DNA environment. Differential Scanning Calorimetry was also used to gain thermodynamic information on the destabilising effect of an O<sup>6</sup>meG lesion on a DNA duplex. This was done with a view to obtaining insights into potential mechanisms by which Ada-C might locate a lesion.

Finally, Chapter 8 introduces various models of how Ada-C binds DNA and discusses potential mechanisms for lesion localisation and the final demethylation reaction, based on the available experimental evidence. Implications for termination of the Adaptive response are also discussed.

# Chapter 2

## *Methods*

### **2.1 OPTIMISING OVEREXPRESSION OF ADA-C IN *E. COLI***

Initially BL21(DE3) was chosen as the strain to produce Ada-C in minimal media as it was prototrophic and maintained good plasmid stability. Expression levels in this strain were optimised in minimal media by inducing the culture at very high densities, a strategy made possible by addition of extra carbon source (glucose) to the media. However, yields were still insufficient for NMR work and it was found that the strain B834(DE3) was capable of producing twice the amount to BL21(DE3). B834(DE3) was used for subsequent selective labeling, deuteration and  $^{13}\text{C}/^{15}\text{N}$  labeling Ada-C. Further enhancements of yield, were made by inoculating labeled media with very high densities of cells prior to induction, made possible by a novel method termed “*cellular resuspension*”.

All solid chemicals used were of analytical grade unless stated otherwise. Compounds incorporating  $^{15}\text{N}$  or  $^{13}\text{C}$  isotopes were purchased from Isotec Inc. Cell growth absorbance readings were always taken at a wavelength of 595nm using a Philips PU 8620 spectrophotometer, SDS-PAGE gels were run on a Bio-Rad mini gel system at 200V (see Appendix 1.3). Large scale centrifugations (>1ml) were carried out in a Beckman J2-MC centrifuge unless otherwise stated and absorbance values in the UV range were determined using a Unicam UV2 spectrophotometer. Dialysis tubing was purchased from Spectrapor and pressure cell filters from Diaflo.

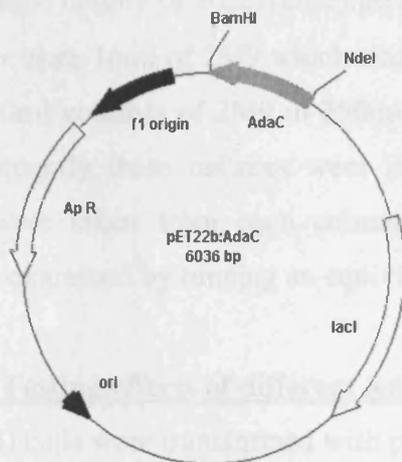
#### **2.1.1 The Expression System**

The plasmid and *Ada-C* gene insert were the generous gift of Dr. J. Brannigan<sup>1</sup>. The plasmid is pET22b, (Novagen) which employs an efficient *T7lac* promoter and an ampicillin resistance marker. The *Ada-C* gene is inserted in between the NdeI and BamHI restriction sites in the multiple cloning site, (Fig.2.1), thus omitting the *pelB* leader sequence, (for periplasmic localization) often associated with this pET plasmid. Expression will only occur in a host strain containing a T7 RNA polymerase gene, ( $\lambda$ DE3 lysogen). BL21(DE3) is such a strain, having the added advantages that it is deficient in the *lon* and *ompT* proteases. Its full genotype is: *F*<sup>-</sup>, *OmpT*<sup>-</sup>, *lon*<sup>-</sup>, *gal*<sup>-</sup>, *dcm*<sup>-</sup>, *hsdSB*(*rB*<sup>-</sup>*mB*<sup>-</sup>) (DE3). The system is inducible by IPTG. This removes the *lac* repressor from the *lac* operator (*lac o*) adjacent to the *lacUV5* promoter (which controls the transcription of T7 RNA polymerase). This promoter is then

---

<sup>1</sup> Present Address: Department of chemistry, University of York, UK.

freed, so cellular RNA polymerase may transcribe the lysogenic T7 RNA polymerase gene. Additional control is provided at the plasmid level since the T7 promoter also contains *lac o*, from which lac repressor, (coded for by the plasmid) must be released before the T7 polymerase may catalyse transcription.



**Figure 2.1** pET22b plasmid with *AdaC* inserted between the NdeI and BamHI restriction sites.

All plasmid stocks were amplified using a commercially available Wizard Mini Prep kit, (Promega), the protocol for this preparation is described in Appendix 1.1.

All cell transformations were carried out using the Calcium chloride method together with selection on LB plates containing 100mg/l ampicillin. The method for making cells competent and transforming is described in Appendix 1.2.

### **2.1.2 Enhancing expression in BL21(DE3)**

Levels of expression were assessed by SDS-PAGE electrophoresis on 15% gels (see Appendix 1.3). Up to 1ml sample volumes of cell culture were pelleted on a bench top microfuge, the supernatant was discarded, and then 50µl of SDS-PAGE loading buffer and 50µl water were mixed into the cell pellet. This was then boiled for a few minutes and then 5µl loaded to the gel which was run at 150V. Levels of Ada-C expression could be crudely assessed by comparing the size of the Ada-C band with those of background proteins. Naturally, this method is only suited to finding significant increases or decreases in expression. All cell cultures were grown at 37°C with 200rpm shaking. Results from this section and subsections 2.1.2.1 to 2.1.2.3 are given in section 3.2.1, Chapter 3.

### **2.1.2.1 Testing effects of different induction times on expression in BL21(DE3)**

Using the above method, but taking into account the differences in cell density such that approximately the same number of cells were used for each sample point, the effects of inducing cells with IPTG at various points in the culture growth curve were examined. The media used for the effect of different induction time was 2M9A with 4g/l glucose (Appendix 1.4). A single colony of BL21(DE3):pET22b:adaC plate containing 100mg/l ampicillin was used to inoculate 16ml of 2M9 which was grown up overnight at 37°C with 200rpm shaking. Separate 50ml volumes of 2M9 in 250ml flasks were inoculated from the overnight cultures and subsequently these cultures were induced with 1mM IPTG at different time points. Samples were taken from each culture 2 hours following induction and the effect on expression examined by running an equivalent number of cells on a 15% SDS-PAGE gel.

### **2.1.2.2 Testing effects of different lengths of induction on expression in BL21(DE3)**

BL21(DE3) cells were transformed with pET22b:AdaC plasmid as described above. A single colony was picked and added to 1ml of sterile 1% NaCl solution, 200µl of the resulting cellular suspension was then added to 8ml of the following media: 2M9-G, 2M9-P, 2M9, celltone –U<sup>2</sup> (Martek) and the culture allowed to grow overnight. 50ml of the respective media in 250ml conical flasks was then inoculated using these overnight cultures. The cultures were set to grow at 37°C and 200rpm shaking. 1ml samples of culture were taken from each flask prior to induction with 1mM IPTG at  $A_{595} = 0.5$ , and again at selected time points up to 21 hours after induction. These samples were spun down and the cells boiled in 50µl dH<sub>2</sub>O and 50µl SDS-PAGE loading buffer. 5µl of the resulting solution was then electrophoresed on a 15% SDS-PAGE gel.

### **2.1.2.3 Testing effects of different media on growth of BL21(DE3):pET22b:adaC**

50ml volumes of media in 250ml flasks were inoculated from resuspended scrapes of cultures of BL21(DE3):pET22b:adaC colonies off LB ampicillin plates as above to ensure starting inoculum was identical for all cultures. The effects of three media types on cell growth, in addition to 2M9, were assessed. Each media was based on 2M9A (Appendix 1.4), but had extra nutrients (Table 2.1.1).

---

<sup>2</sup> Celltone –9U is a commercial media made up of dipeptides, glucose and other carbohydrates, it is a richer media than 2M9, but not as rich in nutrients as 2YT. This particular Celltone media is unlabelled, but equivalent media containing <sup>15</sup>N and <sup>13</sup>C isotopes is also available.

<b>MEDIA NAME</b>	<b>ADDITIONS TO MEDIA IN COMPARISON TO 2M9A (Appendix 1.4)</b>
2M9-G	10g/l glucose as opposed to 4g/l in standard 2M9
2M9-P	Twice the concentration of Phosphate compared to 2M9
2M9-V	Contained following trace supplement additions: Nicotinamide 2mg Folic acid 2mg ZnSO <sub>4</sub> .7H <sub>2</sub> O 2mg Thiamine hydrochloride 20mg/l Biotin 5ug

**Table 2.1.1** Media types used to examine growth of cultures of BL21(DE3):pET22b:*adaC*

2M9-V was constructed after a report that addition of extra vitamins and trace elements to minimal media led to yields of up to 200mg/l of a <sup>13</sup>C/<sup>15</sup>N enriched fusion protein in *E.coli* (Jansson *et al.*, 1996).

The three media in Table 2.1.1, along with standard 2M9, were allowed to grow at 37°C, 200rpm shaking. Samples for A<sub>595</sub> measurements, were taken at regular time intervals. Using these A<sub>595</sub> absorbance readings, growth curves were constructed as shown in Chapter 3 (Section 3.2.1).

### **2.1.3 Expression in other *E.coli* strains**

Even after optimising expression in BL21(DE3), yields of Ada-C in minimal media were still only 7mg/l, insufficient for economic isotopic labeling. The effect on Ada-C expression of changing the host strain was examined in *E.coli* B834(DE3) and JM109(DE3).

JM109(DE3) is a prototroph and can use 2M9 media. B834(DE3) is a methionine auxotroph, and thus minimal media for this strain was 2M9 supplemented with 250mg/l L-methionine.

A comparison of Ada-C expression in B834(DE3), BL21(DE3) and JM109(DE3) in 2YT media (Appendix 1.4) was made by SDS-PAGE. Effects of induction and expression in minimal media were also examined. All culture media contained 100mg/l ampicillin, and all culture volumes were 5ml in 50ml tubes. Growth temperature was always 37°C, and 200rpm shaking was used. Results are given in section 3.2.2, Chapter 3.

#### **2.1.4 Examining expression in B834(DE3)**

The dramatic increase in Ada-C expression in B834(DE3) in relation to other tested strains warranted further investigation. Both the effect of B834(DE3) on expression of other proteins, and the effect of the DE3 insert (lysogenic  $\lambda$  T7 RNA polymerase) on Ada-C expression in B834(DE3) were examined. Expression of a  $\beta$ -lactamase enzyme BCII, and *L.casei* protein Dihydrofolate reductase (DHFR<sup>3</sup>) from pET plasmids were studied in strains B834(DE3), BL21(DE3) and JM109(DE3) by growing overnight cultures in 2YT and examining the leaky expression levels by SDS-PAGE.

The *Ada-C* gene is also present on a plasmid which uses a *ptac* promoter and thus does not rely on the T7 polymerase, (Moore *et al.*, 1994). This *ptac* plasmid was transformed into the non-DE3 strains of B834, (called WA834) and BL21. Leaky expression overnight in 2YT was then compared for the two strains, in order to observe whether enhanced expression occurred with WA834 relative to BL21 when the T7 polymerase system was not employed. Results are given in section 3.2.3, Chapter 3.

#### **2.1.5 Enhancing growth and expression in B834(DE3)- “Cellular resuspension”**

Growth curves for B834(DE3) in celltone-U unlabeled media (Martek), 2M9 with 4g/l glucose and 2M9 with 10g/l glucose media were determined by measuring the  $A_{595}$  of 5ml cultures in 50ml tubes at regular time intervals. Growth occurred at 37°C and 200rpm shaking. Cultures were inoculated from glycerol stock of B834(DE3):pET22b:*adaC*. The effects of 1mM IPTG induction on growth and expression were examined using SDS-PAGE. In contrast to BL21(DE3) the biomass of a B834(DE3) culture could not be increased by addition of more glucose to the media. Enhancements of biomass, and protein yield should be possible by inoculating labeled media to high cell densities with cells grown in unlabeled media. In this manner less of the label, (eg. <sup>15</sup>N) is used to generate cell biomass and more is channeled into production of the desired protein. To test this hypothesis a 250ml volume of 2M9 media was inoculated from glycerol stock and the culture grown to stationary phase, (36 hours) in a 2L flask. The cells were then aseptically transferred to a sterile Beckman 500ml centrifuge pot and centrifuged at 3000 rpm for 10 minutes using the JA10 rotor on a Beckman J2-MC2 centrifuge at 4°C. Cells were then gently resuspended into 350ml of unlabeled celltone -U media (Martek) in a 2 litre flask and allowed to grow for 25 hours. Samples were taken at 2, 6, 7 and 25 hours after the resuspension.  $A_{595}$  measurements and SDS-PAGE analysis to test expression were made. This method led to an increase in Ada-C yield (Section

---

<sup>3</sup> Plasmids for BCII and DHFR were kindly provided by C. Damblon and R. Badii of the Biological NMR Centre, University of Leicester, respectively.

3.2.4, Chapter 3). Resuspension into labeled 2M9 media, in place of unlabeled Celltone, and use of this method for the production of labeled Ada-C protein was also possible (Section 2.1.6). Once yields of Ada-C had been augmented by strain choice, media design and use of the cellular resuspension method described above, the labeled protein was produced in sufficient quantities for biochemical and NMR studies. This was achieved using the protocols described in section 2.1.6 below.

### **2.1.6 The Isotope label-dependent protocols for expressing Ada-C**

It was found that freshly transformed B834(DE3): pET22b:*adaC* could be grown to  $A_{595}$  of 0.6-0.8 in 2M9. The cells were then pelleted by centrifugation at 7000 rpm on a bench top microfuge, and resuspended in filter sterilised LB containing 20% glycerol. They were then snap frozen at  $-80^{\circ}\text{C}$ . This glycerol stock of cells, once frozen could be stored at  $-80^{\circ}\text{C}$  for many months without loss of expression capability. Effectively three different protocols exist for producing Ada-C. Which protocol was employed depended on the labeling strategy and the cost of the isotope to be incorporated into Ada-C. Unlabeled Ada-C was produced in a unique manner which took advantage of leaky expression in 2YT media and the inherently high yields. The route to production of labeled Ada-C made use of the cellular resuspension method (outlined in section 2.1.5) in minimal media, if isotopic costs were large. If the isotope was cheap (as in the case of  $^{15}\text{N}$ ) the time-consuming cellular resuspension method was not employed. These protocols were equally applicable to the production of wild type and mutant Ada-C, and are outlined in sections 2.1.6.1 to 2.1.6.3 below.

#### **2.1.6.1 Expressing unlabeled Ada-C**

Unlabeled Ada-C was used for all non-NMR work, such as Isothermal Titration Calorimetry studies. 500ml of sterile 2YT in 2l flasks was inoculated with 10ul of B834(DE3):pET22b:*adaC* glycerol stock in the evening and left to grow overnight at  $37^{\circ}\text{C}$  and 200rpm shaking. Cells were harvested 15 to 17 hours later by centrifugation at 8000 rpm for 10 minutes and  $4^{\circ}\text{C}$  in a Beckman J2-MC centrifuge with a JA10 rotor. The supernatant was discarded and cell pellet frozen at  $-20^{\circ}\text{C}$  until purification.

#### **2.1.6.2 Expressing Ada-C labeled with low cost isotopes**

'Low cost isotopes' are defined as costing less than £100 for the quantity required for 1 litre of media. One example is  $^{15}\text{N}$ , either in the form of  $(^{15}\text{NH}_4)_2\text{SO}_4$  for uniform (ubiquitous) labeling, or in the form of a labeled  $\alpha$  amino group in an amino acid for selective labeling. In producing wild type Ada-C labeled in this way, high enough yields ( $>15\text{mg/l}$ ) were obtained

when using B834(DE3) to render the yield-enhancing cellular resuspension method unnecessary.

The difference in protocols between making uniformly  $^{15}\text{N}$ -labeled, and selectively labeled protein is in the type of media used. Uniformly labeled Ada-C is produced with 2M9 media containing  $(^{15}\text{NH}_4)_2\text{SO}_4$  as the unique nitrogen source and supplemented with 250mg/l unlabeled L-Met, since B834(DE3) is a Met auxotroph (Appendix 1.4). In  $^1\text{H}$ - $^{15}\text{N}$  HSQC NMR spectra the three methionine signals will therefore be missing (hence the term 'uniform labeling' is not strictly true). This was the sacrifice made for using auxotrophic B834(DE3) to gain higher protein yields. Selectively labeled protein production used 2M9 media supplemented with all the amino acids, one of which contains a  $^{15}\text{N}$ -labeled  $\alpha$  amino group (2M9A media, Appendix 1.4). In both cases the protocol for protein expression was as follows:

Glycerol stock of B834(DE3):pET22b:*adaC* was used to inoculate 50ml of unlabeled media, in a 250ml flask. This was grown overnight and then used to inoculate one litre of labeled media to an  $A_{595}$  of 0.1. 250-500ml of media was contained in a 2L flask. IPTG was added to 1mM when this main culture reached an  $A_{595}$  of 0.6 for selectively labeled protein and  $A_{595}$  of 2.0 for uniformly labeled protein. Three hours later cells were pelleted by centrifugation at 8000 rpm for 10 minutes at  $4^\circ\text{C}$  and the pellet frozen  $-20^\circ\text{C}$  until purification. Samples of wild type Ada-C labeled individually with  $^{15}\text{N}$ -Leu, Val, Ala, Ile, Tyr, Phe, Glu, Gly, and uniformly with  $^{15}\text{N}$  were prepared in the above manner.

### **2.1.6.3 Expressing Ada-C labeled with high cost isotopes**

'High cost isotopes' are defined as costing more than £100 for the quantity required for 1 litre of media. Preparation of deuterated<sup>4</sup> or  $^{13}\text{C}$ -labeled protein falls into this category. The cost of 1 litre of  $\text{D}_2\text{O}$  is approximately £200, whilst the cost of 4g of glucose required for 1 litre of 2M9 media is £800, at the time of writing. It is in these productions that Ada-C yield must be as high as possible. Hence when labeling wild type Ada-C or C144S mutant with  $^2\text{H}$  or  $^{13}\text{C}$ , B834(DE3) cells are used along with the yield-enhancing cellular resuspension method. An exception to this rule was the production of wild type Ada-C selectively labeled with  $^{15}\text{N}$ -Cys. This protein was produced using the cysteine auxotroph BL21(DE3)*cys*<sup>-</sup> and hence required the cellular resuspension method to enhance yields. In this case 2M9A media was used. For the production of uniformly labeled protein, 2M9 media was used. The media only contained the labeled isotopes in the final stage following cellular resuspension. Where protein deuteration was required,  $\text{D}_2\text{O}$  was used in place of  $\text{H}_2\text{O}$ , and the media was filter sterilised

using 250ml, 0.2µm pore size, sterilising vessels (Nalgene) to minimise the risk of  $^1\text{H}$ - $^2\text{H}$  exchange with atmospheric  $\text{H}_2\text{O}$  (which may be incurred by autoclaving). The procedure for the cellular resuspension method or Ada-C production was as follows:

B834(DE3):pET22b:adaC cells were grown overnight in 5ml of unlabeled media from glycerol stock in a 50ml tube. 1ml of the subsequent overnight culture was used to inoculate 250ml of unlabeled media per 2l flask. (To produce 52-62% deuterated Ada-C the media at this stage was 100%  $\text{H}_2\text{O}$ . To produce a final protein deuteration level of 70%, the media at this stage had to consist of 100%  $\text{D}_2\text{O}$ ). The cells were allowed to grow to an  $A_{595}$  of 2 and were then aseptically pelleted using sterile 500ml Beckman centrifuge pots and centrifuging the cultures at 3000 rpm, at  $4^\circ\text{C}$  for 5 minutes. (In the case of 70% deuteration levels, the slow growth rate in  $\text{D}_2\text{O}$  made it unfeasible to commence cellular resuspension at  $A_{595}$  of 2, so the cultures were left for 24 hours instead). The unlabeled media was then discarded and the cells gently resuspended into 250ml of media containing the required isotopes (eg.  $^{13}\text{C}$ -glucose,  $(^{15}\text{NH}_4)_2\text{SO}_4$ ,  $\text{D}_2\text{O}$ , or a selectively labeled amino acid). This culture was allowed to adapt at  $37^\circ\text{C}$ , 200rpm shaking for one hour, after which IPTG was added to 1mM. Three to four hours later cells were pelleted by centrifugation at 8000 rpm for 10 minutes, at  $4^\circ\text{C}$  and the cell pellet frozen at  $-20^\circ\text{C}$  until purification. When making triple labeled  $^2\text{H}/^{15}\text{N}/^{13}\text{C}$ -labeled wild type Ada-C with 70% deuteration for example, the first media into which the overnight culture of cells was inoculated was 2M9 made of 100%  $\text{D}_2\text{O}$ . The second batch of media into which the cells were aseptically resuspended was also 100%  $\text{D}_2\text{O}$  2M9, but also contained  $^{13}\text{C}$ -glucose and  $(^{15}\text{NH}_4)_2\text{SO}_4$  as sole carbon and nitrogen sources. The level of deuteration could be reduced to 52-62%, if the first stage media was  $\text{H}_2\text{O}$  in place of  $\text{D}_2\text{O}$ , as outlined above. Production of  $^{15}\text{N}$ -cysteine labeled wild type Ada-C used 2M9A media containing  $^{15}\text{N}$ -cysteine as the second round media, and unlabeled 2M9A as first round media. 70% deuterated, uniformly  $^{15}\text{N}$ -labeled, and selectively  $^{15}\text{N}$ -labeled C144S mutant Ada-C was also produced using the above method. A summary of how some labeled proteins were produced is found in Table 2.1.2.

---

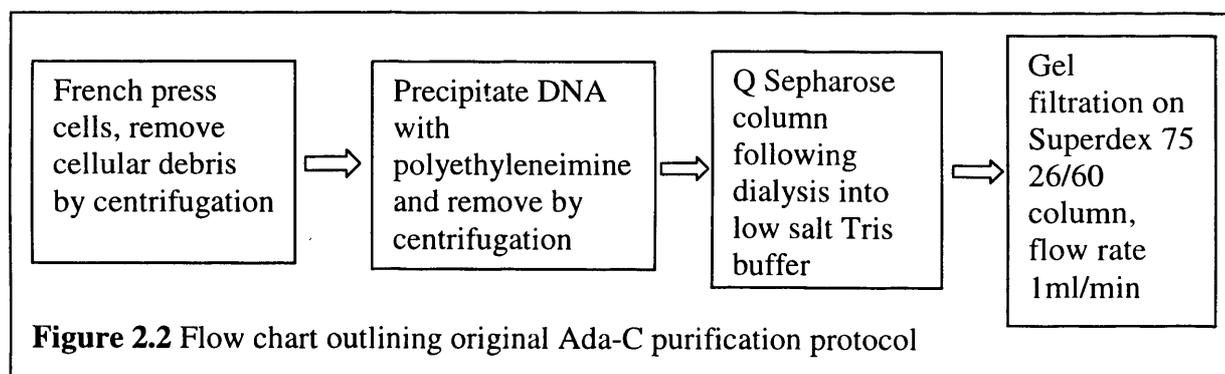
<sup>4</sup> 'Deuterated' refers to random fractional deuteration, where not all non-exchangeable protons are replaced by deuterons.

Protein produced	Strain used	Isotope incorporation	Cellular resuspension used	Media used	
				Before resuspension	After resuspension
Ada-C WT	B834(DE3)	Ubiquitous $^{15}\text{N}$	No	n/a	$^{15}\text{N}$ 2M9
Ada-C C144S	B834(DE3)	Ubiquitous $^{15}\text{N}$	No	n/a	$^{15}\text{N}$ 2M9
Ada-C WT	B834(DE3)	Ubiquitous $^{15}\text{N}$ / $^{13}\text{C}$ 70% $^2\text{H}$	Yes	$^{14}\text{N}$ / $^{12}\text{C}$ 2M9, 100% $\text{D}_2\text{O}$	$^{15}\text{N}$ / $^{13}\text{C}$ 2M9, 100% $\text{D}_2\text{O}$
Ada-C WT	B834(DE3)	Ubiquitous $^{15}\text{N}$ / $^{13}\text{C}$ 52-62% $^2\text{H}$	Yes	$^{14}\text{N}$ / $^{12}\text{C}$ 2M9, 100% $\text{H}_2\text{O}$	$^{15}\text{N}$ / $^{13}\text{C}$ 2M9, 100% $\text{D}_2\text{O}$
Ada-C WT	B834(DE3)	Ubiquitous $^{15}\text{N}$ / $^{13}\text{C}$	Yes	2M9 unlabeled	$^{15}\text{N}$ / $^{13}\text{C}$ Celltone (rich media)
Ada-C WT	BL21(DE3) <i>cys<sup>-</sup></i>	Selectively $^{15}\text{N}$ Cysteine	Yes	2M9A unlabeled	2M9A $^{15}\text{N}$ - Cys

**Table 2.1.2** Some typical isotopic labeling strategies for Ada-C. n/a = not applicable, WT = wild type, C144S = Cys144Ser mutant Ada-C. 2M9A media is 2M9 plus amino acids, one of which may be labeled with an isotope. Note that the term ‘uniform labeling’ is not strictly correct when B834(DE3) was employed to express the protein in minimal media supplemented with unlabeled Met.

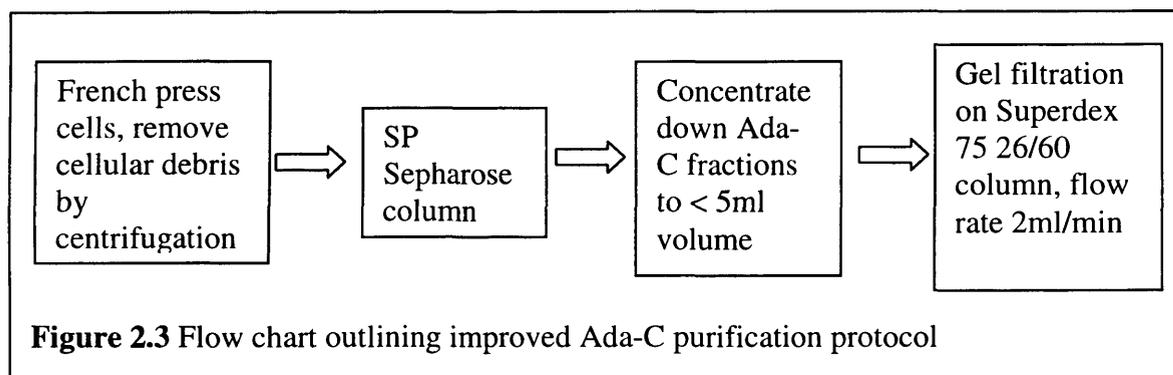
### 2.1.7 Purification of Ada-C

The original purification protocol outlined in Figure 2.2 proved unsatisfactory in terms of efficiency and was thus changed to the one detailed below, and outlined in Figure 2.3



Pelleted cells from a 1l culture were resuspended into 30ml of lysis buffer (50mM  $\text{NaH}_2\text{PO}_4$ , 2mM DTT, 1mM EDTA, 4mM Pefabloc<sup>™</sup> (Boehringer), pH 7.8). One Complete inhibitor tablet<sup>™</sup> (Boehringer) was added and the cells lysed using a French press. After addition of  $\text{MgCl}_2$  to render a final concentration of  $\text{Mg}^{2+}$  ions of 7mM, 2.6mg of Dnase I, (Bovine pancreas, Type II, Sigma) was added to the resulting solution. Cellular debris was then removed by centrifugation at 4°C, (18K rpm, 45 mins, Beckman JA 20 rotor). The supernatant is then filtered using a 0.2 $\mu\text{m}$  acrodisk, (Gelman Sciences) and chromatographed through a home-made SP-Sepharose (Pharmacia) column. The protein was eluted with a gradient of 0 to 0.6M NaCl in FPLC-A buffer, (50mM  $\text{NaH}_2\text{PO}_4$ , 5mM DTT, 1mM EDTA,

pH 7.8) over 210ml, at a flow rate of 2ml/min. Fractions containing Ada-C, (as determined by SDS-PAGE) were pooled, concentrated down to a volume less than 5ml using a 50ml pressure cell, (Amicon) with a 10kDa mwco filter, (Diaflo) and chromatographed through a Superdex 75 26/60 column, (Pharmacia) at a flow rate of 2ml/min, employing FPLC-A buffer containing 400mM NaCl to prevent interactions with the column matrix. For both columns 5ml fractions were collected. Ada-C eluted at 180ml as confirmed by SDS-PAGE. Fractions containing pure Ada-C were pooled and concentrated down to 10ml using a 50ml pressure cell with a 10kDa mwco filter. Subsequently the Ada-C solution was transferred into a 10ml pressure cell with 10kDa mwco filter and buffer exchanged into high salt NMR buffer (10% D<sub>2</sub>O, 50mM Na<sub>2</sub>HPO<sub>4</sub>, 10mM DTT, 100mM NaCl, 100mM NaBr, 1mM EDTA, 0.01% NaN<sub>3</sub>, pH 6.7). Further concentration occurred in this pressure cell and was completed using 10kDa mwco centricon concentrators and centrifuging for periods up to 40 minutes at 7000 rpm in a Sorvall RC-5B centrifuge employing an SS-34 rotor at 7°C.



### **2.1.8 Quality control of Ada-C**

Three sets of quality control were performed on Ada-C: Mass spectrometry to confirm the correct mass, N-terminal sequencing as an extra check to the correct sequence and activity assaying as a function test.

#### **2.1.8.1 Electrospray mass spectrometry and N-terminal sequencing of Ada-C**

Samples of pure Ada-C at mM concentrations were dialysed exhaustively against deionised H<sub>2</sub>O prior to mass spectrometry on a Micromass Platform electrospray mass spectrometer<sup>5</sup>. Both wild type and C144S and C144A mutants were subjected to mass spectrometry analysis. The N-terminal 10 residues of a sample of <sup>13</sup>C/<sup>15</sup>N labelled Ada-C were subjected to sequencing on a Perkin Elmer 476 automated protein sequencer which employs Edman degradation chemistry.

<sup>5</sup> Mass spectrometry was carried out by Dr. K. Lilley, PNAACL, University of Leicester, UK.

### **2.1.8.2 Activity Assay of Ada-C**

The activities of a sample of wild type Ada-C and C144S mutant Ada-C were determined by Dr. G. Major<sup>6</sup> using a published method (Major *et al.*, 1991). Briefly, a sample of unlabelled Ada-C was made up to 200µl with Buffer A, (50mM Tris/HCl pH 8.3, 0.5mM EDTA, 1mM DTT) and incubated with 10µl of substrate DNA, (duplex DNA containing a tritylated methyl group on the exocyclic oxygen in guanines at 2000-5000 d.p.m./10ul), for 90 minutes at 37°C. The reaction was stopped by addition of 300µl of 80mM EDTA pH 6.0 containing 200µg calf thymus DNA, and mixing on ice. All DNA was then precipitated by addition of 200µl 3% Cetyltrimethylammonium bromide. Any protein remaining in association with DNA was solubilized by addition of 10µl of 1mM CaCl<sub>2</sub> containing 50µg Proteinase K, (Sigma type XI) and incubation at 37°C for 60 minutes. DNA was then pelleted by centrifugation, (15K rpm, 15 minutes), 650µl of supernatant was added to 2ml deionised H<sub>2</sub>O and mixed with 10ml of Pico-Aqua scintillation cocktail, (Canberra-Packard). Radioactivity was counted in a scintillation counter giving a measurement of d.p.m. A blank assay reaction containing no Ada-C was also counted and the result subtracted from the test d.p.m. to provide a "net d.p.m." Protein concentrations were determined using the extinction coefficient at 280nm of 1.095 mg cm ml<sup>-1</sup> as determined by Mr. P. Sharratt<sup>7</sup>.

### **2.1.9 Synthesising and purifying DNA oligomers**

All oligonucleotides were synthesised on a 1µM scale by Dr. K. Lilley on an Applied Biosystems 394 automated DNA synthesizer using phosphoramidite chemistry.

The following 20-mer oligonucleotides were synthesised:

BMT: 5'-AGGTTGCCACA(O<sup>6</sup>meG)GTTGCCAC (where O<sup>6</sup>meG is O<sup>6</sup>-methylguanine)

BIGTAR: 5'-AGGTTGCCACAGGTTGCCAC

BIGCOMP: 5'-GTGGCAACCTGTGGCAACCT

BFC: 5'-GTGGCAACFTGTGGCAAFCT (where F is 5-fluorocytosine)

5-Fluoro-2'-deoxyuridine phosphoramidite was obtained from (Glen Research<sup>8</sup>) and during standard deprotection yielded 5-Fluorocytosine.

The modified phosphoramidite for incorporation of O<sup>6</sup>-methylguanine is 5'-dimethoxytrityl-N-isobutyryl-O<sup>6</sup>-methyl-2'-deoxyguanosine,3'-[(2-cyanoethyl)-N,N-diisopropyl]-phosphoramidite, (Glen Research). Final dimethoxytrityl protecting groups were removed by standard methods following the final phosphoramidite base addition. Ammonia was removed

---

<sup>6</sup> Medical Molecular Biology Group, The Medical school, University of Newcastle

<sup>7</sup> Protein and Nucleic acid Chemistry Facility, Dept. Biochemistry, University of Cambridge

<sup>8</sup> 22825 Davis Drive, Sterling, Virginia, USA

from the oligonucleotide solutions by vacuum centrifugation and the oligonucleotides then lyophilised down to solid form. The BMT oligonucleotide required deprotection of the O<sup>6</sup>-methylguanine base which remained coupled to a cyanoethyl group. This was removed in accordance with manufacturers instructions using 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU). Briefly, the process was as follows:

1µmol of lyophilized oligonucleotide was re-dissolved in 1ml 10% DBU in anhydrous methanol (Aldrich) in a clean glass vial. This was done under anhydrous conditions using anhydrous N<sub>2</sub>(g) to fill syringes required to extract the methanol and DBU. The septum-capped vial was left in the dark at room temperature for 5 days and was then lyophilized to a small volume. Lyophilised 1µmole aliquots of BIGTAR and BIGCOMP or the small volume representing 1µmole of BMT solution were made up to 1.5ml with 10mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.0. 10µmole amounts of oligo solution were then dialysed overnight against 1l of 10mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.0 at 4°C, using a 1000 Da mwco membrane. Following this they were purified by HPLC using an adaptation of a published method (Arghavani *et al.*, 1995) outlined below. A preparative reverse phase Phenomenex primesphere 5 C18 HC column was used connected to a Gilson HPLC system. Optical detection of the samples was set at 215nm and 260nm and the following buffer system was employed: buffer A: 10mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.0, buffer B: methanol:acetonitrile 1:7 mix. Oligonucleotides had retention times from 36 to 40 minutes using a gradient of 0 to 50% buffer B over 50 mins with a flow rate of 3ml/min. HPLC chromatograms indicated that oligos were ~95% pure prior to HPLC, and ~99% pure following this treatment. Fractions of 3.6ml were collected and those containing oligonucleotide were pooled. The organic solvent mix was removed from the pure oligonucleotide solution by flushing with N<sub>2</sub>(g), and the solutions were subsequently dialysed against 2l dH<sub>2</sub>O at 4°C, lyophilised and stored at -20°C until required.

BIGTAR tagged with fluorescein was made in the same manner as the oligonucleotides detailed above, with the fluorescein incorporated by standard synthesis to the 5' end of the DNA (Glen Research).

The DNA was quantified using the extinction coefficients at A<sub>260</sub> listed in Table 2.1.3. These were calculated using the nearest neighbor method as described in Appendix 1.5.

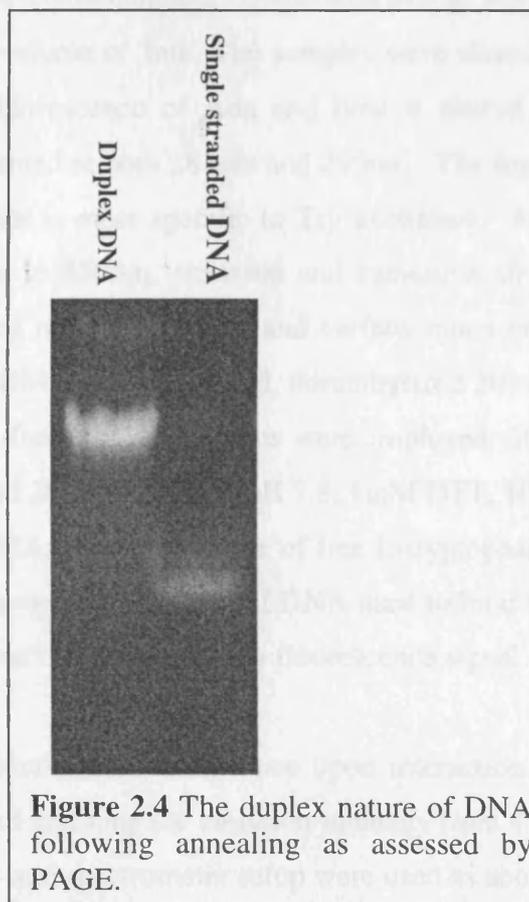
Where double-stranded DNA was required, this was made by adding equimolar amounts of both complementary strands, heating at 80°C for 5 minutes, and then allowing to cool slowly to room temperature. Annealing was checked by native 20% PAGE (Appendix 1.6). Volumes equivalent to 0.7 nmol DNA were added to one fifth volume STEB (40% sucrose, 0.5mg/ml bromophenol blue, 10mM Tris, 1mM EDTA, pH 7.5). 5µl of the subsequent solution was loaded to the gel, which was run cold at 4°C at 100V, keeping the current down

to below 6 Amp. The gel was stained using 0.01% ethidium bromide for 5 minutes and then examined under UV light. Annealing was assessed by running single-stranded DNA next to duplex DNA on the gel and comparing the migration distances, as shown in Figure 2.4.

DNA	Extinction coefficient at 260nm ( $\text{mM}^{-1} \text{cm}^{-1}$ )
BIGTAR+Fluorescein	200.2
BIGTAR / BMT	188.2
BIGCOMP / BFC	192.4

**Table 2.1.3** The calculated extinction coefficients for 20-mer oligonucleotides., The extinction coefficient for fluorescein ( $12\text{mM}^{-1}\text{cm}^{-1}$ , Glen research) was added to  $188.2 \text{ mM}^{-1}\text{cm}^{-1}$  to obtain the final extinction coefficient.

Lyophilised samples of HPLC purified and unpurified oligonucleotides were redissolved into 7:3 acetonitrile:50mM imidazole mix and the mass and purity assessed using a Micromass Platform electrospray mass spectrometer.



**Figure 2.4** The duplex nature of DNA following annealing as assessed by PAGE.

## **2.2 EXAMINING THERMODYNAMICS OF ADA-C/DNA INTERACTIONS**

A number of attempts involving different techniques were made to obtain binding constants between Ada-C and 20-mer single-stranded DNA. These included isotropic fluorescence (with and without fluorescein tagged DNA), circular dichroism (CD), and Isothermal Titration Calorimetry (ITC). ITC proved to be the most suitable for the current system, but the other attempts made are also included below.

### **2.2.1 Attempting to obtain binding constants for Ada-C:DNA using Fluorescence and Circular Dichroism**

Readily available techniques for defining binding constants were fluorescence and CD, both of which were attempted.

#### **2.2.1.1 Fluorescence**

All fluorescence work was carried out using a Perkin Elmer LS50B Luminescence Spectrometer at 25°C. Pure unlabeled Ada-C was prepared and purified as described in sections 2.1.6.1 and 2.1.7, DNA was prepared as described in section 2.1.9. All solutions were filtered through a 0.2µm pore size filter prior to use. Use of plasticware was minimised to avoid fluorescence contamination. Concentrations of protein and DNA used were in the µM range in a final volume of 3ml. The samples were stirred during spectrometric scanning. When testing the fluorescence of Ada and how it altered upon binding untagged DNA, excitation was performed at both 281nm and 295nm. The former wavelength excites Trp and Tyr residues, the latter is more specific to Trp excitation. An emission scan was performed ranging from 290nm to 450nm, emission and excitation slit widths were both 3nm. Such scans were performed on Ada-C, DNA and various ratios of the two mixed together. Both BIGTAR and BIGCOMP single-stranded, unmethylated 20-mer oligonucleotides were tested for binding. Two different buffer systems were employed: 50mM Phosphate, pH 7.8, 10mM NaCl, 1mM DTT, and 20mM Tris-HCl pH 7.8, 1mM DTT, 100mM NaCl. To check for inner filter effects from DNA, the fluorescence of free L-tryptophan was examined in the presence and absence of the same concentration of DNA used to bind to Ada-C. The concentration of Trp used was such that the sample gave a fluorescence signal of similar intensity to that of the free protein.

Monitoring of the fluorescein fluorescence upon interaction with Ada-C was performed by exciting at 497nm and scanning the emission intensity from 450nm-600nm. The same buffers concentration ranges and spectrometer setup were used as above.

### **2.2.1.2 Circular Dichroism**

CD experiments were performed using a Jasco J-715 spectropolarimeter using a Starna 1mm pathlength cuvette (type 37.10). Both Ada-C, methylated single-stranded DNA (BMT), unmethylated, single-stranded DNA (BIGTAR) and unmethylated duplex DNA were dialysed extensively against 50mM Phosphate, pH 7.8, 30mM NaF (used in place of NaCl for higher transparency in the far-uv (Federwisch *et al.*, 1997)). The concentrations of protein and DNA were determined by uv spectroscopy using the appropriate extinction coefficients (sections 2.1.8.2 and 2.1.9).

Scans from 190nm to 450nm were performed on 0.1mg/ml samples of Ada-C with and without unmethylated duplex and single-stranded DNA, or methylated single-stranded DNA at ratios from 1:1 to 1:3 protein : DNA. CD spectra of the DNA samples used were also recorded. In all cases two accumulations of the spectrum were taken. The setup of the other parameters is revealed in Table 2.2.1.

<b>Parameter</b>	<b>Setup</b>
Response time	2 secs
Number of accumulations (scans)	2
Scan speed	50 nm/min
Resolution	0.5nm
Sensitivity	50mdeg

**Table 2.2.1** The Jasco J-715 CD spectropolarimeter setup.

### **2.2.2 Isothermal Titration Calorimetry- DNA binding and demethylation energetics**

The 20-mer oligonucleotides BIGTAR, BMT, unmethylated duplex DNA, and Ada-C were prepared as previously described. The successful annealing of BIGTAR and BIGCOMP to create double-stranded DNA was ascertained by PAGE analysis as described in section 2.1.9. Titrations were carried out in four buffers so that corrections could be made to  $\Delta H$  for buffer ionisations that may occur. The buffer compositions were as follows:

Buffer (50mM)	NaCl concentration (not accounting for salt formed during buffer pH correction)	DTT concentration
Phosphate-NaOH	30mM	5mM
Tris-HCl	67.5mM	5mM
MOPS-NaOH	77mM NaCl	5mM
HEPES-NaOH	78mM NaCl	5mM

**Table 2.2.2** The composition of the four buffers used for ITC experiments. All buffers are at pH 7.8 and 9.8mS conductivity at 20°C as measured using a hand held conductivity meter. Different salt concentrations were employed to attain identical conductivity.

Both DNA and Ada-C were dialysed extensively into the required buffers using 1kDa and 10 kDa mwco dialysis membrane respectively. Dialysis of these two components of the titration was allowed to proceed in the same container such that buffers were exactly matched. A sample of the external dialysate solution was kept for control titrations.

Calorimetric measurements were carried out using a microcalorimetry system isothermal titration calorimeter (Microcal)<sup>9</sup>. Just prior to a titration experiment all solutions were carefully degassed using the Thermolyne degasser supplied with the ITC unit. The 1.34ml sample cell was filled with DNA. Ada-C was drawn into a 250µl syringe which was then mounted into the unit's stepped motor for delivery into the cell. All titrations were carried out at 25°C, with the outer jacket of the unit cooled to 20°C by a Neslab RTE-111 waterbath. The stirring rate was always set to 400rpm. The injection routine varied depending upon the titration performed.

Both binding interactions between Ada-C and unmethylated DNA, and demethylation reactions between Ada-C and methylated DNA were monitored. With binding titrations, an injection routine involved 20 injections of 12.5µl with a 210 second delay between injections. Occasionally this routine was fine-tuned if problems were encountered in obtaining points at the ends of the titration curve, or just to minimise sample wastage. (One example of this tuning, was using a lower volume, 6.25µl first injection. The first point of the titration is always discarded, since there is usually air at the end of the syringe which introduces volume errors in the first injection. Use of a lower volume initial injection decreases sample wastage).

For binding titrations between Ada-C and unmethylated single-stranded (BIGTAR), or duplex DNA, concentrations of Ada-C in the syringe were 0.6 to 0.7mM, whilst those of DNA in the

<sup>9</sup> Thanks to Drs. H. Trayer and J. Eads at University of Birmingham for the use of their ITC apparatus.

cell were on average ten times lower. Titrations of wild type Ada-C into unmethylated single-stranded DNA were carried out in tris, phosphate, MOPS and HEPES buffer systems.

When using the tris or HEPES buffer during binding titrations, it was found that 20 injections were insufficient to complete the titration curve. Another 20 injections were therefore made by reloading the syringe, allowing equilibration, and then proceeding to titrate into the contents of the sample cell remaining from the first titration. These two datasets were then concatenated after correcting for the altered equilibration baseline. Analysis of calorimetric data was carried out with the Origin software provided with the equipment. Thermodynamic parameters were extracted by fitting data to a single-site model, the details of data processing are fully described in Appendix 2.1. Prior to processing the first datapoint was always removed.

Titrations of Ada-C into duplex DNA were carried out in these four buffer systems. Only the titration in MOPS buffer was successfully fitted. The titrations in other buffers could not be fitted because either the initial or the final plateau of the titration was insufficiently flat. A summary of the successfully fitted titrations from which thermodynamic information was obtained is outlined in Table 2.2.3.

When demethylation reactions were carried out initially, similar concentrations of single-stranded, methylated DNA (BMT) and Ada-C were used as with binding titrations to unmethylated DNA, outlined above. By employing a large (~ 300 fold) excess of BMT in the cell (concentration around 0.19mM) and adding small volumes of Ada-C (6.25µl injections at a concentration of 0.14mM) the heat emitted with each injection could be kept constant and thus quantified in terms of heat emitted per mole of demethylation reaction. Several injections of Ada-C (<10) were employed, just to ensure that the heat evolved from the system was constant with each injection. Fitting analysis was unnecessary in the case of demethylation reactions involving a vast excess of BMT, as the molar heat of reaction could be directly read from the integrated data plot.

<b>Contents of sample cell (into which Ada-C was titrated)</b>	<b>Buffer systems data successfully fitted for</b>
unmethylated ssDNA	MOPS, Tris, Phosphate, HEPES
unmethylated duplex DNA	MOPS
excess methylated ssDNA	MOPS, Tris, Phosphate, HEPES

**Table 2.2.3** The successfully fitted titration data for wild type Ada-C binding unmethylated single-stranded (ssDNA), and double-stranded DNA. Thermodynamic data for the *in vitro* demethylation reactions between Ada-C and methylated, single-stranded DNA (BMT), were also performed in all 4 buffers. These datasets were not fitted to the single-site model since heat changes occurred as a result of demethylation, as well as binding and these could not be deconvoluted.

As a control to the binding titrations, the heats of buffer dilution were also measured by titrating Ada-C into buffer. An alternative control of titrating buffer into DNA was also performed.

## **2.3 ASSIGNING THE BACKBONE OF ADA-C**

Attempts were made to assign the backbone of wild type Ada-C employing a protonated  $^{13}\text{C}/^{15}\text{N}$ -labeled sample of Ada-C along with data from selectively  $^{15}\text{N}$ -labeled Ada-C. Due to exceptionally broad line widths (as a result of protein aggregation and conformational flexibility), and high degree of peak overlap in the data, triple resonance experiments were carried out again at higher magnetic field strength on a 70% deuterated  $^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C sample.

### **2.3.1 Attempts to assign the backbone using protonated $^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C**

The six triple resonance experiments listed in Table 2.3.1 were carried out on a  $^{13}\text{C}/^{15}\text{N}$ -labeled sample of wild type Ada-C (labeled uniformly except for the Met residues) on a Bruker DMX 500 spectrometer with a 5mm triple resonance ( $^{13}\text{C}/^{15}\text{N}/^1\text{H}$ ) probe equipped with a single-axis actively-shielded gradient, at 298K. The concentration of Ada-C was 31mg/ml (1.5mM) in 50mM Phosphate, pH 6.7, 1mM deuterated EDTA, 10mM deuterated DTT, 100mM NaCl, 100mM NaBr, 0.01%  $\text{NaN}_3$ , 10%  $\text{D}_2\text{O}$ . Spectral sizes and acquisition parameters are shown in Table 2.3.1. 32 scans were used for each triple resonance experiment except HNCO where, with increased sensitivity 16 scans were sufficient. All 1D and 2D data was processed using XwinNMR software, whilst 3D data sets were analysed using Felix 97 software on a Silicon Graphics Indigo workstation. In processing both 2D and 3D experiments a  $65^\circ$  phase-shifted sine-bell window function and zero filling were applied before Fourier transformation.  $^1\text{H}$  chemical shifts were referenced to 2,2-dimethyl-2-silapentane-5-sulfonate, (DSS).  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shifts were referenced using the frequency obtained for protons and the gyromagnetic ratios for  $^{15}\text{N}$  and  $^{13}\text{C}$  (Cavanagh *et al.*, 1996).

Experiment	Sample and conditions used	Time domain sizes (points)			Spectral widths (Hz)			Size of processed spectrum (points)		
		t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>	ω <sub>1</sub>	ω <sub>2</sub>	ω <sub>3</sub>	t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>
HNCO	<sup>13</sup> C/ <sup>15</sup> N Ada-C, pH 6.7, 298K	451 (H)	44 (C)	40 (N)	6775	1667	1667	1024	256	128
HN(CA)CO	<sup>13</sup> C/ <sup>15</sup> N Ada-C, pH 6.7, 298K	451 (H)	44 (C)	40 (N)	6775	1667	1667	1024	256	128
HN(CO)CA	<sup>13</sup> C/ <sup>15</sup> N Ada-C, pH 6.7, 298K	451 (H)	56 (C)	36 (N)	6775	3333	1667	1024	256	128
HNCA	<sup>13</sup> C/ <sup>15</sup> N Ada-C, pH 6.7, 298K	451 (H)	42 (C)	40 (N)	6775	3333	1667	1024	256	128
HCBCA(CO)NH	<sup>13</sup> C/ <sup>15</sup> N Ada-C, pH 6.7, 298K	451 (H)	48 (C)	32 (N)	6775	8000	1667	1024	256	128
HCBCANH	<sup>13</sup> C/ <sup>15</sup> N Ada-C, pH 6.7, 298K	451 (H)	48 (C)	32 (N)	6775	8000	1667	1024	256	128
NOESY-HSQC	<sup>15</sup> N Ada-C, pH 7.8, 293K	902 (H)	320 (H)	72 (N)	7936	7936	1960	1024	512	128
TOCSY-HSQC	<sup>15</sup> N Ada-C, pH 7.8, 293K	902 (H)	200 (H)	52 (N)	7506	7002	2200	1024	512	256
<sup>1</sup> H- <sup>15</sup> N HSQC	Selectively <sup>15</sup> N-labeled Ada-C, pH 7.8 or 6.7, 298K	1024 (H)	100 (H)	-	7000	2500	-	2048	512	-

**Table 2.3.1** The NMR experiments performed on the protonated Ada-C. All points are complex. The triple resonance and HSQC experiments were recorded at 500MHz. The NOESY-HSQC and TOCSY-HSQC were recorded using a Bruker AMX 600 spectrometer at 600MHz.

Assignment of protein backbone resonances was achieved in a semi-automated fashion using a series of in-house designed macros interfacing with Felix 97. Final sequential alignment of dipeptide spin systems utilised a simulated annealing algorithm. Selective <sup>15</sup>N-labeling of Cys, Ala, Phe, Tyr, Gly, Val, Leu, Ile, Glu, allowed identification of these backbone resonances and created more constraints for the simulated annealing approach of resonance assignment. Ambiguities in dipeptide spin linkage could sometimes be resolved through the use of the NOESY and TOCSY experimental data. Using this method 44 out of 170 backbone amide resonances were assigned.

### **2.3.2 The Ada-C backbone assignment using 70% deuterated <sup>13</sup>C/<sup>15</sup>N-labeled Ada-C**

The six triple resonance experiments listed in Table 2.3.2 were carried out on a 26mg/ml (1.3mM) sample of 70% deuterated, <sup>13</sup>C/<sup>15</sup>N-labeled Ada-C in 50mM Phosphate, pH 6.7,

1mM deuterated EDTA, 10mM deuterated DTT, 100mM NaCl, 100mM NaBr, 0.01% NaN<sub>3</sub>, 10% D<sub>2</sub>O. The experimental temperature was 298K and the instrument used was a Bruker DRX 600 spectrometer. 16 scans were employed in each case and the data were processed as above employing a 65° phase-shifted sine-bell window function prior to Fourier transformation, with linear prediction to reduce linewidth in the nitrogen frequency dimension.

Experiment	Time domain sizes (points)			Spectral widths (Hz)			Size of processed spectrum (points)		
	t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>	ω <sub>1</sub>	ω <sub>2</sub>	ω <sub>3</sub>	t <sub>1</sub>	t <sub>2</sub>	t <sub>3</sub>
HNCO	442 (H)	50 (C)	54 (N)	7788	2272	2212	512	256	256
HN(CA)CO	442 (H)	50 (C)	54 (N)	7788	2272	2212	512	256	256
HN(CO)CA	442 (H)	48 (C)	54 (N)	8388	4200	2212	512	256	256
HNCA	442 (H)	50 (C)	54 (N)	8388	4200	2212	512	256	256
HN(COCA)CB	442 (H)	64 (C)	54 (N)	7788	9600	2212	512	256	256
HN(CA)CB	442 (H)	64 (C)	54 (N)	7788	9600	2212	512	256	256

**Table 2.3.2** The triple resonance experiments performed on 70% deuterated, <sup>13</sup>C/<sup>15</sup>N-labeled Ada-C, 26mg/ml at 298K and 600MHz field strength. All points shown are complex points.

Following final assignment, chemical shift indexing of the carbon frequencies was performed using the method and software developed as described by (Wishart *et al.*, 1994).

## **2.4 EXAMINING INTERACTIONS BETWEEN ADA-C AND DNA BY NMR-**

### **2.4.1 General NMR methods**

All NMR experiments were recorded either at a  $^1\text{H}$  frequency of 500MHz using a Bruker DMX 500 spectrometer, or at 600MHz using a Bruker DRX 600. All spectra of the protein were acquired at 298K. Two-dimensional proton-detected HSQC<sup>10</sup> experiments were used to monitor proton and  $^{15}\text{N}$  chemical shift changes upon DNA-binding. Two dimensional HMQC experiments were performed to assess the tautomeric states of the active site histidine (His 145). This was assigned on the basis that it was the only visible His imidazole NH resonance, thus correlating with its buried nature (the other histidines in Ada-C are solvent exposed and thus unlikely to provide a detectable imidazole NH signal due to exchange with water protons). Interactions with DNA were monitored using both selectively-labeled wild type Ada-C, and deuterated, uniformly  $^{15}\text{N}$ -labeled wild type and C144S mutant Ada-C. Binding was monitored both at pH 7.8 and pH 6.7. All Ada-C protein and DNA was prepared and purified as described in section 2.1 above. The high salt buffer used for NMR of the protein alone was: 10%  $\text{D}_2\text{O}$ , 50mM Phosphate pH 6.7 or pH 7.8, 10mM DTT, 1mM EDTA, 100mM NaCl, 100mM NaBr, 0.01%  $\text{NaN}_3$ . For a DNA addition to be made, the purified solid DNA was first made up in  $\text{dH}_2\text{O}$  and the pH corrected to that of the Ada-C using  $\text{NaOD}_{(\text{aq})}$  or  $\text{DCl}_{(\text{aq})}$ . The solution was filtered through a 0.2 $\mu\text{m}$  microspin filter (Amicon) and the concentration determined using the absorbance at 260nm and the calculated extinction coefficient (Appendix 1.5). The correct volume for the desired amount of DNA was then lyophilised to solid form. The solid was then directly dissolved in a volume of Ada-C of known concentration. The pH was checked, and if necessary small adjustments were made to ensure it was identical to that of the free protein solution. This complex was then filtered as above, placed in an NMR tube and degassed with Argon. Alternatively the complex could be dialysed into low salt buffer (10%  $\text{D}_2\text{O}$ , 50mM Phosphate pH 6.7 or pH 7.8, 10mM DTT, 1mM EDTA, 10mM NaCl, 0.01%  $\text{NaN}_3$ ) to enhance binding prior to degassing. Dialysis employed a 1000 Da mwco membrane and was carried out against 20ml of buffer for a 0.5ml sample. Interactions of Ada-C with methylated and unmethylated, single- and double-stranded DNA were examined. To make double-stranded DNA, known concentrations of the component single strands were made up in low salt buffer (10%  $\text{D}_2\text{O}$ , 50mM Phosphate pH 6.7 or pH 7.8, 10mM DTT, 1mM EDTA, 10mM NaCl) and then mixed in a 1:1 ratio. The ensuing solution of DNA was annealed by heating to 80°C for 5 minutes and then allowing to cool slowly to room temperature and filtered. A defined amount of this duplex DNA was

---

<sup>10</sup> Don't freak out ! Explanations of NMR terminology are found in an introduction to the technique in Chapter 5.

then added to Ada-C in solution and the buffer exchanged into either low salt buffer using small scale dialysis with 1000Da mwco. The solution was then degassed and NMR experiments carried out as above. Binding of the inactive C144S mutant with methylated 20-mer single and double-stranded DNA was also assessed.

1D and 2D data were processed using XwinNMR software, on a Silicon Graphics workstation. In processing 2D experiments a 60° phase-shifted sine-bell window function and zero filling was applied prior to Fourier transformation. 1D experimental data was subjected to gaussian multiplication with a line broadening of -7 Hz and gaussian broadening of 0.1 Hz.

Experiment	Reason for doing experiment	Time domain sizes (points)		Spectral widths (Hz)		Number of scans
		t <sub>1</sub>	t <sub>2</sub>	ω <sub>1</sub>	ω <sub>2</sub>	
<sup>15</sup> N HSQC (2D)	Examining polypeptide backbone resonances of selectively or uniformly <sup>15</sup> N labeled Ada-C at 298K.	2048	200	7000	2200	64
<sup>15</sup> N HSQC (2D)	Examining backbone and Arg and His sidechain resonances from uniformly <sup>15</sup> N labeled Ada-C at 298K.	2048	400	9000	4156	128
<sup>15</sup> N HMQC (2D)	Examining the tautomeric state of His at 298K.	2048	250	10000	10000	256
1D proton spectrum	Examining protein to check folding, and interactions with DNA at 298K.	4096	-	10000	-	128 or 1024

**Table 2.4.1** Some typical NMR experiments performed at a spectral frequency of 500MHz.

### **2.4.2 Calculation of binding constant between Ada-C and single-stranded DNA by NMR**

A titration of 70% deuterated wild type Ada-C with single-stranded unmethylated DNA (BIGCOMP) was performed in high salt buffer (10% D<sub>2</sub>O, 50mM Phosphate pH 6.7, 10mM DTT, 1mM EDTA, 100mM NaCl, 100mM NaBr, 0.01% NaN<sub>3</sub>). <sup>1</sup>H-<sup>15</sup>N HSQC spectra were acquired at 600MHz on a Bruker DRX 600 following each addition of solid DNA as described in Section 2.4.1. The chemical shift changes were then plotted against amount of DNA added, and the binding curve fitted to the standard equation:

$$\delta_{obs} - \delta_A = \frac{(\delta_{DA} - \delta_A) \left\{ (D_T + A_T + K_d) - \sqrt{(D_T + A_T + K_d)^2 + 4D_T A_T} \right\}}{2D_T}$$

(where K<sub>d</sub> = dissociation constant, δ<sub>A</sub> = chemical shift of free Ada-C, δ<sub>DA</sub> = chemical shift of fully bound Ada-C, δ<sub>obs</sub> = chemical shift of partially bound Ada-C, D<sub>T</sub> = total concentration of DNA, A<sub>T</sub> = total concentration of Ada-C, (Lian *et al.*, 1993)), using the program Graftit.

### **2.4.3 In Vitro methylation reactions**

The methylation of wild type Ada-C by *in vitro* reaction with 20-mer methylated DNA (BMT) was studied by making additions of BMT to Ada-C as described above (both cysteine selectively labeled Ada-C, and 70% and 52%-62% deuterated, uniformly <sup>15</sup>N-labeled Ada-C samples were methylated). Following NMR experiments, protein and DNA were separated from each other by passage down an SP-sepharose column. Briefly, the 0.5ml mixture was diluted into 30ml FPLC-A buffer (50mM NaH<sub>2</sub>PO<sub>4</sub>, 5mM DTT, 1mM EDTA, pH 7.8) and loaded to the SP-sepharose column. Separation was carried out employing the gradient outlined in section 2.1.7. The protein bound to the column and was eluted with increasing salt, the DNA did not bind. Both DNA and methylated Ada-C were collected and their masses assessed by electrospray mass spectrometry. In the cases where deuterated Ada-C had been used, mass spectrometry was only performed on the DNA due to the inherent spread of mass in deuterated protein.

### **2.4.4 pKa determination for the active site Cys 144**

Ada-C selectively labeled with <sup>13</sup>C/<sup>15</sup>N Cysteine was dialysed into high salt D<sub>2</sub>O buffer to minimise the water signal. This buffer was: 50mM Phosphate, 100% D<sub>2</sub>O, 10mM DTT, 1mM EDTA, 100mM NaCl, 100mM NaBr, 0.01% NaN<sub>3</sub>. <sup>1</sup>H-<sup>13</sup>C HSQC spectra were run at 298K and 600MHz on a Bruker DRX 600 spectrometer for a sample at different pD ranging from 6.3 to 11.0. The pD was altered between spectra using NaOD<sub>(aq)</sub>.

## **2.5 EXAMINING THE EFFECT OF ADA-C BINDING ON THE DNA**

### **2.5.1 Energetics of duplex melting**

Solid single-stranded oligonucleotides were dissolved in 50mM Phosphate, pH 7.0, 10mM NaCl. The oligonucleotide solutions were then dialysed into two times 2 litres of the same buffer. All dialysis took place in the same beaker such that buffers for all oligonucleotides would be exactly matched. Following dialysis, the concentrations of oligonucleotides were determined using  $A_{260}$  measurements and the complementary DNA strands added in an equimolar ratio to obtain a final concentration of duplex DNA of 0.10mM. BIGCOMP and BMT oligonucleotides were mixed and annealed to create methylated duplex DNA, whilst BIGCOMP and BIGTAR were annealed to create unmethylated 20-mer duplex DNA. The annealing procedure is described in section 2.1.9. Annealed oligonucleotides were then subjected to Differential Scanning Calorimetry<sup>11</sup> (DSC) to determine the thermodynamic parameters of duplex melting. All DSC experiments were performed on a Microcal MC-2 differential scanning microcalorimeter (Microcal, Amherst, USA). The DNA samples were scanned at a rate of 60K/hr in a sample cell of volume 1.2 cm<sup>3</sup>. Subtraction of the baseline from the raw data, allowed data analysis using established procedures (Blandamer *et al.*, 1994) with Origin software (Microcal, Amherst, USA).

### **2.5.2 Examining imino protons of DNA upon binding by Ada-C**

All spectra were recorded on a Bruker DRX 600 spectrometer at a <sup>1</sup>H operating frequency of 600 MHz. One-dimensional spectra of single- and double-stranded methylated and unmethylated DNA were recorded at 298K using 128 and 1024 scans. Superior spectra of single-stranded DNA were recorded at 283K with 2048 scans. Spectra of wild type Ada-C complexed with single-stranded, and duplex unmethylated DNA were recorded at 298K with 2048 scans. Spectra of inactive mutant C144S Ada-C alone and complexed with methylated and unmethylated DNA were recorded at 298K and 283K. A sweep width of 12531 Hz was employed in each case.

All spectra were processed with XwinNMR software on a Silicon Graphics Indy workstation. Prior to Fourier transformation, the data was subjected to gaussian multiplication with a line broadening of -7 Hz and gaussian broadening of 0.15 Hz.

---

<sup>11</sup> The DSC experiments were performed by Dr. B. Briggs in the Dept. of Chemistry, University of Leicester.

### **2.5.3 Probing the structure of methylated and unmethylated DNA using $^{19}\text{F}$ NMR**

DNA was designed containing two 5-fluorocytosines, 5'-GTGGCAACF<sup>2</sup>TGTGGCAAF<sup>1</sup>CT where F<sup>1</sup> and F<sup>2</sup> represent 5-fluorocytosine. This DNA strand was designed to basepair F<sup>2</sup> with O<sup>6</sup>meG on the 20-mer known as BMT, or with normal guanine on the BIGTAR oligonucleotide. F<sup>1</sup> would base pair with normal guanine in both cases and serves as an internal reference, providing data on the environment of nucleotides far from the O<sup>6</sup>meG lesion. F<sup>2</sup> probes the environment of the bases immediately opposite the O<sup>6</sup>meG lesion. The fluorinated oligonucleotide was base paired with both methylated (BMT) and unmethylated (BIGTAR) 20-mer strands by the standard annealing method (Section 2.1.9). One-dimensional  $^{19}\text{F}$  NMR experiments were run on the two types of DNA alone, and in complex with equimolar 70% deuterated,  $^{15}\text{N}$ -labeled Ada-C C144S mutant. All experiments were done on samples in low salt NMR buffer (10% D<sub>2</sub>O, 50mM Phosphate, 10mM DTT, 1mM EDTA, 10mM NaCl, pH 6.5) on a Bruker DRX 600 equipped with a  $^{19}\text{F}\{^1\text{H}\}$  probehead, and operating at a  $^{19}\text{F}$  resonance frequency of 564.69 Hz. A sample of ratio 1.9 Ada-C C144S : 1 duplex methylated DNA was also used. Attempts to get this ratio with unmethylated duplex DNA caused protein precipitation and sample loss. One-dimensional  $^{19}\text{F}$  spectra were collected using DNA concentrations from 0.22 to 0.84 mM at 298K and 308K (this higher temperature was omitted when studying the DNA in complex with Ada-C). 4096 scans were collected over a sweep width of 11261 Hz. Prior to Fourier transformation, the data were multiplied by an exponential multiplication function with a line broadening of -5 Hz.

Transverse relaxation time,  $T_2$ , for the free  $^{19}\text{F}$  duplex DNA and the 1.9 Ada-C C144S : 1 methylated duplex complex were measured at 298K with nine 1D spin echo spectra [ $90^\circ(^{19}\text{F})\text{-}\{T_r/2\text{-}180^\circ(^{19}\text{F})\text{-}T_r/2\}_n\text{-Acq}$ ], with spin echo delays  $T_r$  dictated by  $n = \{T_r/2\text{-}180^\circ(^{19}\text{F})\text{-}T_r/2\} = 1.6, 9.7, 14.5, 19.4, 29.1, 40.4, 50.1, 61.4$  ms. Other acquisition parameters are 1344 scans and sweep width of 7062 Hz. Attempts to measure  $T_2$  times for the 1.9 Ada-C C144S : 1 unmethylated DNA complex were thwarted by the precipitation of this sample.  $^{19}\text{F}$  signal  $T_2$  relaxation experiments were also run on duplex  $^{19}\text{F}$  DNA alone at 308K. In this case eight experiments were performed with  $n = 1.6, 3.2, 4.8, 6.4, 9.7, 14.5, 19.4, 25.9$  ms. The sweep width was 7062 Hz, and the number of scans collected 2048. All data were processed using XwinNMR on a Silicon Graphics Indy workstation. Prior to Fourier transformation the data were exponentially multiplied by a line broadening factor of -10 Hz. All  $^{19}\text{F}$  chemical shifts were referenced relative to trifluoroacetic acid (TFA) calculated using the ratio  $^{19}\text{F}/^1\text{H} = 0.940866982$  (Klimasauskas *et al.*, 1998).

## **2.6 MODELLING DNA-BINDING**

The PDB code for Ada-C, 1sfe was put through the automated structural alignment program Dali located at URL: <http://www2.ebi.ac.uk/dali/> (Holm *et al.*, 1994); (Holm *et al.*, 1996).

This was done with the hope that finding structurally similar proteins may provide clues to the function of the middle domain of Ada-C (residues 12-92), and also provide insight into the mechanism of DNA-binding. A structurally related protein of known DNA-binding mechanism may reveal mechanisms by which Ada-C interacts with DNA.

The assigned amino acids which underwent exchange broadening upon binding DNA were found to belong to the helix-turn-helix type motif. Based on the findings of this search, models of how Ada-C binds DNA were made by docking Ada-C on to the structurally homologous protein in complex with DNA. A model of how Ada-C docks into the major groove of DNA was generated by superimposing the helix-turn-helix type region (determined by NMR to be involved in DNA-binding) on to the three DNA-binding helices of Catabolite gene activator protein (CAP) using Insight II software (MSI) on a Silicon Graphics workstation and the coordinates of the CAP/DNA complex, 2cgp (Schultz *et al.*, 1991). Coordinates of Ada-C were taken from 1sfe, (Moore *et al.*, 1994). The proteins were superimposed manually, and the CAP complex DNA then substituted for a DNA structure containing O<sup>6</sup>-methylguanine (coordinate file 1d27), (Leonard *et al.*, 1990). This model provided no insight into how the active site C144 of Ada-C gained access to the lesion. Base flipping was postulated and an alternative model of how Ada-C docks into the major groove of DNA and flips out the methylated base was generated. The DNA coordinates were taken from the PDB file 1mht depicting the complex structure of *EcoRV* methyltransferase with DNA containing a flipped target base (Klimasauskas *et al.*, 1994). The extrahelical cytosine base and its guanine pair from this structure were altered to an O<sup>6</sup>-methylguanine, and a cytosine respectively using the builder module in Insight II. The Ada-C was docked into the DNA with its recognition helix lying into the major groove, running almost parallel to the planes of the base-pairs. Only one available hole existed through which the active site cysteine could be observed from the molecule exterior. This hole lay between the helix-turn-helix and the 'wing' region (residues 149-158), and it was through this that the methylated base to gained access to the active site. The Ada-C was positioned such that arginine sidechains near the DNA-binding interface were aptly located to form salt links to the phosphate backbone of DNA. Energy minimisation was performed using the Macromodel/Batchmin package (version 5.5) (Mohamadi *et al.*, 1990). Only the region within 8Å of the C5 atom of the extrahelical guanine were minimised. Within this region different restraining potentials were applied to parts of protein and DNA. Unless otherwise stated a force restraint of 100kJ/Å was applied to the atoms to constrain them to their

crystallographic coordinates. The sidechains of residues Y113, V126 and L138-V148 were left without force restraint, and were free to move. So was the extrahelical guanine base, its segment of the DNA backbone, the three nucleotides 5' of this guanine. Nucleotides A405, G406, C407 (numbering from PDB file 1MHT) on the complementary DNA strand were unconstrained. The unpaired cytosine base C408, opposite the extrahelical guanine was constrained by 50kJ/Å. The protein backbones of residues Y113, I140-V147 were constrained only by 25kJ/Å. Energy minimisation was performed using the Steepest descent gradient procedure, and was terminated after 1000 iterations. Minimisation was then continued using a PRCG gradient procedure which was terminated after the energy gradient rms fell below 0.01 kJ/mol Å.

This model was then compared with a recently published another model of Ada-C in complex with flipped DNA, (Vora *et al.*, 1998). This model was based on the manner on which MuA transposase bound DNA (Clubb *et al.*, 1994). Results of this modelling are discussed in detail in Chapter 8.

# Chapter 3

## *Protein expression, purification and labeling, DNA synthesis and purification.*

### 3.1 INTRODUCTION

Biochemical and structural studies on proteins require that large quantities of homogeneous protein be produced as efficiently, in terms of time and resources as possible. For the purposes of NMR studies, which frequently involve the use of expensive isotopic labeling such as  $^{13}\text{C}$  or  $^2\text{H}$ , maximising protein yields is of paramount importance (Table 3.1.1). Levels of protein overexpression in bacterial systems may be enhanced by manipulation of the host strain and plasmid genetic structure, cell growth conditions and protein purification strategy.

Protein labeling strategy	Use of protein	Cost of production per litre culture
Selective $^{15}\text{N}$ (eg. $^{15}\text{N}$ Ala)	Mapping binding sites, backbone assignment	varies- around £100
Ubiquitous $^{15}\text{N}$	Mapping binding sites	£20
Ubiquitous $^{15}\text{N}$ $^2\text{H}$	Mapping binding sites	£420
Ubiquitous $^{15}\text{N}$ $^{13}\text{C}$	Backbone assignment	£820
Ubiquitous $^{15}\text{N}$ $^{13}\text{C}$ $^2\text{H}$	Backbone assignment	£1220

**Table 3.1.1** Costs and uses of differently labeled Ada-C (the cost given in each case is that used in the production of Ada-C using minimal media). The effect of deuteration ( $^2\text{H}$ ) is simply to reduce signal loss and therefore aid acquisition of useful NMR spectra.

Genetic manipulations include optimising the codon bias on the expressed gene and employing a suitable plasmid vector for over expression. In the case of Ada-C production, the pET expression system, which makes use of the highly efficient T7 RNA polymerase was used. All Ada-C pET plasmids (wild type and C144S, C144A mutants) were a generous gift of Dr. J. Brannigan<sup>1</sup>. This chapter deals with the enhancements in yield of Ada-C made through rational media design, *E.coli* strain choice, growth conditions, and final purification strategy.

For isotopic labeling, defined minimal media is frequently used. This ensures that the source of nitrogen and carbon is unique and can be manipulated in accordance with the labeling

<sup>1</sup> Dept. of Chemistry, University of York, UK

strategy. A case example is the use of 2M9 media, (Maniatis *et al.*, 1989) where  $(^{15}\text{NH}_4)_2\text{SO}_4$  replaces the more usual  $^{14}\text{N}$  isotope as the sole source of nitrogen to make uniformly  $^{15}\text{N}$ -labeled protein. The effect of addition of various supplements to minimal media was explored with regard to Ada-C yield, since certain additions have been previously shown to enhance protein production in other systems (Jansson *et al.*, 1996). Effects of varied points of induction and cell harvest were explored in this vein, as were alternate strains and various strategies to augment biomass, and hence protein production. Through changing the *E.coli* strain from BL21(DE3) to B834(DE3), and increasing the cell density prior to induction, yields of Ada-C were enhanced ten-fold from their starting level of 6mg/l media. This made the NMR experiments economically viable. Consequently, these studies into protein yield amplification were of great importance to the feasibility of the entire project.

## **3.2 RESULTS**

### **3.2.1 Enhancing expression in BL21(DE3)**

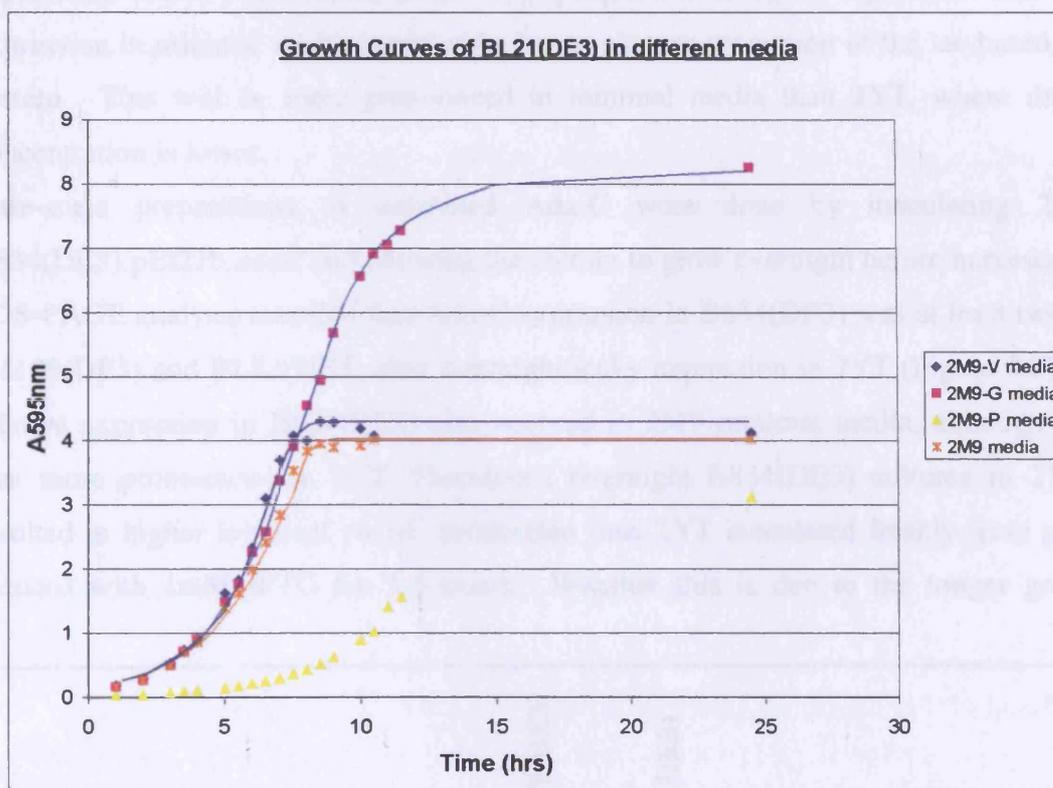
Initially, Ada-C was produced in minimal media using the prototrophic strain BL21(DE3). Harvest and induction points were examined with regard to protein expression levels, using SDS-PAGE analysis. 1mM IPTG was required in minimal media to obtain expressed Ada-C. The point on the growth curve, prior to stationary phase at which 1mM IPTG was added did not alter the Ada-C expression level. Induction following attainment of stationary phase however, led to decreased Ada-C levels. Harvest times up to 6 hours following induction were equally efficient, only after 21 hours induction did the level of cellular Ada-C visibly decrease. Induction with 1mM IPTG at any point in the growth curve led to a complete cessation of growth.

The effects on non-expressing cell growth obtained by altering the media were also monitored. As expected, cells grew to higher densities and produced more overexpressed protein in richer media, such as 2YT, than in minimal media. This was probably because they were under less pressure and had more resources. Three components of minimal media were increased in an attempt to ascertain which was the most useful for the cell. Growth curves are shown in Figure 3.1. Growth in media with double the amount of phosphate to normal 2M9 (2M9-P) proved slightly toxic to the cells and resulted in slower growth. Adding a few extra trace supplements, such as biotin to the media had no effect on growth (2M9-V in Figure 3.1). However, increasing the glucose concentration from 4g/l to 10g/l resulted in an increase in the culture density at stationary phase (2M9-G).

The  $A_{595}$  of stationary phase with 4g/l glucose in the media was 4.0, whilst with 10g/l glucose this value doubled to 8.0. Cessation of growth upon induction, and the observation that induction at high cell densities (still below stationary phase) had no deleterious effects on cellular expression levels, led to the postulation that induction at a higher cell density would render a higher yield of Ada-C. Mid log phase was chosen as a safe point of induction. Increasing the cell density of mid log phase should increase the Ada-C yield. This hypothesis was tested by comparing final yields of purified Ada-C from cells grown in media containing 10g/l glucose, with cells grown in media containing 4g/l glucose. Both cultures were induced with 1mM IPTG at mid log phase ( $A_{595}$  of 2.0 for 4g/l glucose media, and  $A_{595}$  of 4.0 for 10g/l glucose media). Cultures were harvested 3 hours after induction. There was a 50% increase in yield of Ada-C when induction occurred at  $A_{595}$  of 4.0 compared with  $A_{595}$  of 2.0. Results are summarised in Table 3.2.1

Concentration of glucose in 2M9 media used	A <sub>595</sub> at induction (equivalent to culture density)	Final yield of purified Ada-C from 1 litre of culture
4g/l (2M9 media)	2.0	6mg
10g/l (2M9-G media)	4.0	9mg

**Table 3.2.1** The resulting increase in Ada-C yield achieved by inducing a culture at twice the cell density.



**Figure 3.1** Growth curves for BL21(DE3):pET22b:adaC in various 2M9-based media types. 2M9-V = 2M9 with extra trace supplements; 2M9-G = 2M9 with 10g/L glucose (as opposed to usual 4g/L glucose used in 2M9); 2M9-P = 2M9 with additional phosphate.

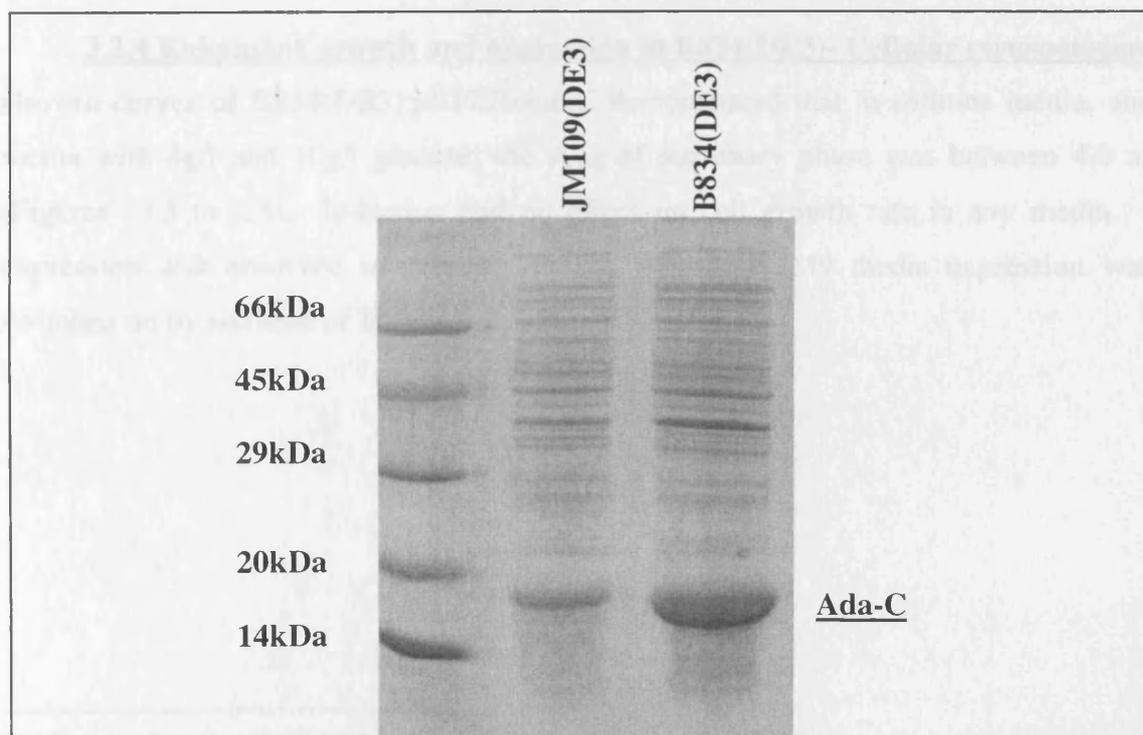
A test to establish the economic viability of using 1g/l and 2g/l of glucose and as opposed to 4g/l was also performed. A one litre culture grown in 2M9 media containing 1g glucose per litre, resulted in the production of 1mg Ada-C. As the expression levels and induction A<sub>595</sub> for the 2g glucose per litre media was identical to those of 1g/l glucose media it is unlikely that the yields from this would have been any different. Thus, obtaining <sup>13</sup>C labeled Ada-C from one litre of 4g glucose per litre of media is more economical than purifying Ada-C from 4 litres of 1g glucose/l media, or 2 litres of 2g glucose/l media. This finding is in agreement with the results obtained by Jansson *et al.*, (1996) who also noted that at 4g/l the cellular glucose utilization was most efficient.

### 3.2.2 Expression in other strains

In 2YT media, but not minimal media, leaky expression was observed for Ada-C in both B834(DE3), JM109(DE3) and BL21(DE3). Following overnight leaky expression in 2YT, B834(DE3) cells could not be subcultured to produce further Ada-C. Over-expression capacity was lost. Presumably this is due to plasmid loss from the culture, or selective loss of the Ada-C gene insert from the plasmid (through recombination), as unexpressing cells exerted their selective advantage, divided more rapidly and dominated the culture. This meant that any overnight cultures which were required for subculturing followed by Ada-C expression, had to use minimal media where expression is tightly regulated. The non-leaky expression in minimal media is probably due to glucose repression of the lac-based promoter system. This will be more pronounced in minimal media than 2YT, where the glucose concentration is lower.

Litre-scale preparations of unlabeled Ada-C were done by inoculating 2YT with B834(DE3):pEt22b:*adaC* and allowing the culture to grow overnight before harvesting.

SDS-PAGE analysis revealed that Ada-C expression in B834(DE3) was at least twice that of JM109(DE3) and BL21(DE3) after overnight leaky expression in 2YT (Figure 3.2). Greater relative expression in B834(DE3) also occurred in 2M9 minimal media, although the effect was more pronounced in 2YT. Uninduced overnight B834(DE3) cultures in 2YT media resulted in higher levels of Ada-C production than 2YT inoculated freshly from plates, and induced with 1mM IPTG for 3.5 hours. Whether this is due to the longer growth, and



**Figure 3.2** Comparison of the expression of Ada-C in B834(DE3) and JM109(DE3) after overnight leaky expression in 2YT. Expression in BL21(DE3) is identical to that of JM109(DE3).

expression time of the overnight culture, or that leaky expression exerts less cellular strain, and is more tolerated than IPTG expression was not investigated further.

### **3.2.3 Further investigation of expression in B834(DE3)**

The reason for the enhanced expression in B834(DE3) in comparison with other strains is unknown but had been noted previously to be the case with both Ada-C and T7 Ligase, (Doherty *et al.*, 1995). To investigate protein expression in B834(DE3) further, the expression of other proteins of interest to the laboratory was investigated. These included Dihydrofolate reductase (DHFR) and BCII  $\beta$ -lactamase.

Expression in 2YT of both DHFR and BCII  $\beta$ -lactamase in B834(DE3) was not enhanced in comparison to the expression of these proteins in BL21(DE3). Both proteins however displayed lower expression levels in JM109(DE3). The Ada-C production in WA834 and BL21 cells using the ptac promoter-based plasmid was much lower than expression in the pET system, demonstrating the efficiency of the T7 RNA polymerase. Expression levels in 2YT did not vary between these two strains. However, the expression of both rac and rho GDI in non-pET expression systems was enhanced in WA834 compared to BL21 (personal communication with R.Badii and D.Hawkins respectively).

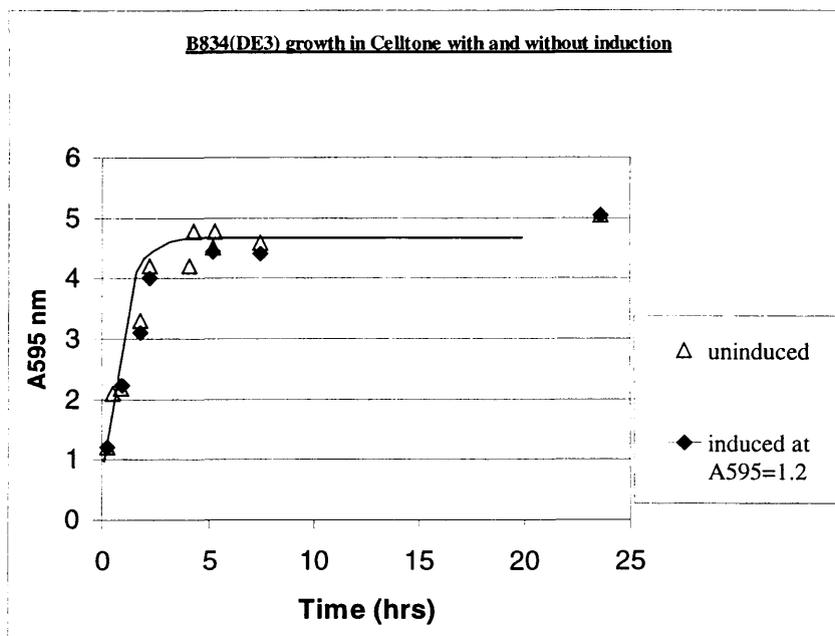
Taken together these results suggest that expression enhancement in B834(DE3) compared to other strains is protein dependent and does not occur in all cases. The indistinguishably low levels of Ada-C expression observed in WA834 and BL21 (using the ptac system) suggest that no expression enhancement in WA834 occurred relative to BL21.

### **3.2.4 Enhancing growth and expression in B834(DE3)- Cellular resuspension**

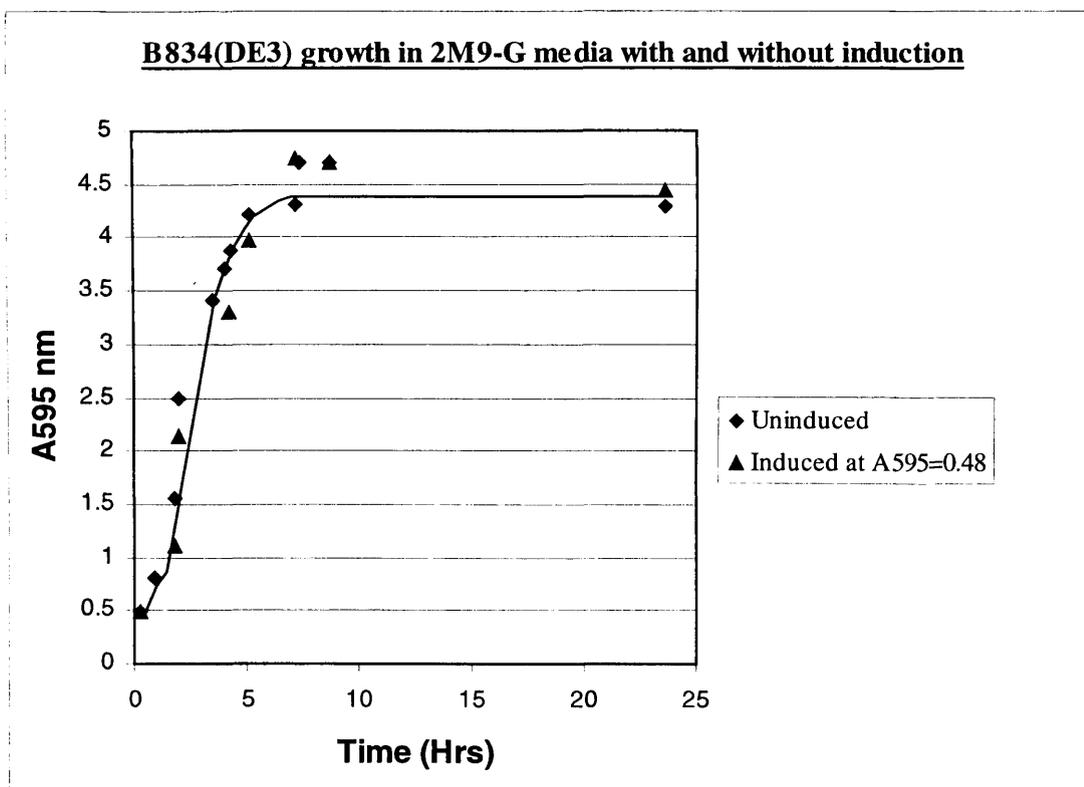
Growth curves of B834(DE3):pET22b:*adaC* demonstrated that in celltone media, and 2M9 media with 4g/l and 10g/l glucose, the  $A_{595}$  of stationary phase was between 4.0 and 5.0 (Figures 3.3 to 3.5). Induction had no effect on cell growth rate in any media. Leaky expression was observed in celltone<sup>2</sup> media, whilst in 2M9 media expression was only switched on by addition of IPTG.

---

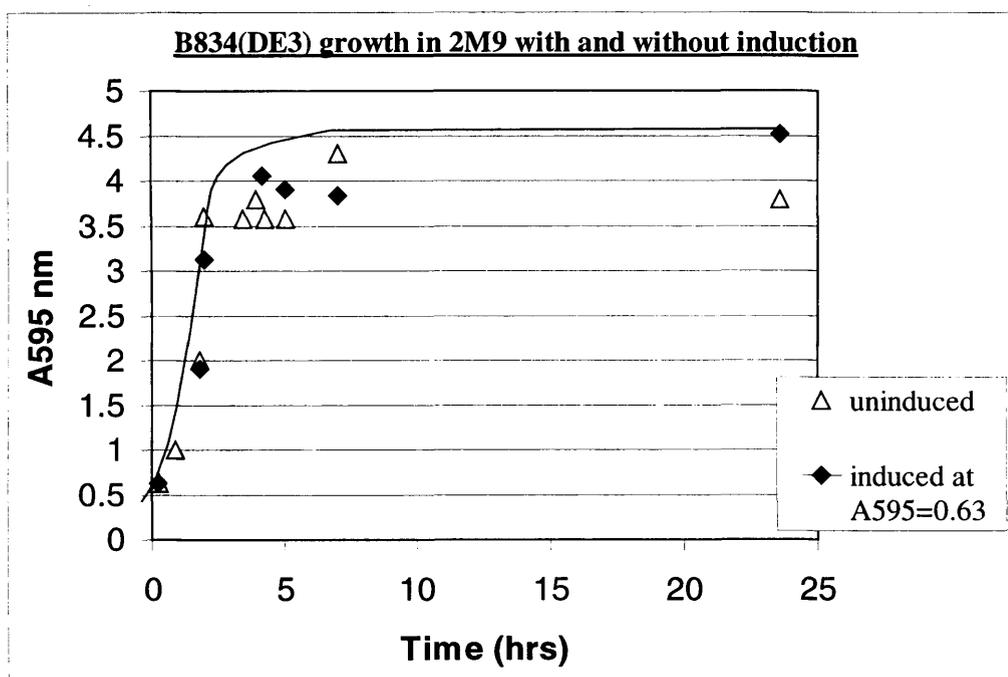
<sup>2</sup> Celltone is a commercially available media containing dipeptides, carbohydrates and cofactors. It is available in <sup>15</sup>N and <sup>13</sup>C isotopically enriched forms for uniformly labeling proteins.



**Figure 3.3** Growth curve of B834(DE3):pET22b:AdaC in celltone media at 37°C. Higher cell densities were induced without effect on expression level or growth. Leaky expression was observed in absence of IPTG.



**Figure 3.4** Growth of B834(DE3):pET22b:AdaC in 2M9-G media (2M9-G media is 2M9 media with 10g/l glucose) at 37°C. Induction had no effect on cell growth. Different induction cell densities had no visible effect on Ada-C expression levels provided IPTG was added prior to stationary phase.



**Figure 3.5** The growth of B834(DE3):pET22b:AdaC in 2M9 media at 37°C. Other cultures were induced at higher cell densities with no effect on cell growth. Higher induction densities had no effect on expression level, as observed by SDS-PAGE, provided IPTG addition occurred before stationary phase.

Cell growth in celltone media was used as the initial test for biomass enhancement by cellular resuspension. In this method, cells were grown to stationary phase in a media where they do not exhibit leaky expression (ie. 2M9). The cellular mass was then separated from this media and resuspended aseptically into fresh celltone media. Cells are allowed to grow, and leaky expression to proceed. Finally, the final biomass of the culture is assessed by  $A_{595}$  measurements and the expression by SDS-PAGE. The idea behind this 'cellular resuspension' method was that increased biomass might enhance the yield of Ada-C.

In the test experiment B834(DE3):pET22b:*adaC* cells were grown to stationary phase in 2M9 without induction. These cells were then separated from the used 2M9 by centrifugation, resuspended into unlabeled celltone media and allowed to grow for 25 hours. The final  $A_{595}$  of the culture at stationary phase was 7.9. This is an two fold increase in final cell density relative to celltone inoculated from glycerol stock (Figure 3.3). Correspondingly, the final levels of Ada-C produced were also higher.

Further tests were carried out where B834(DE3):pET22b:*adaC* cells were initially grown in unlabeled 2M9 and then resuspended into isotopically-enriched 2M9 media. Here they were allowed to adapt for up to one hour prior to induction with IPTG for up to 4 hours. A two fold enhancement of Ada-C production could also be obtained from growing resuspended cells in 2M9 in this manner. This method of 'cellular resuspension' was extensively used to enhance yields of Ada-C for high cost isotope incorporation, where protein yield per litre of

labeled media was critical. It frequently led to a doubling of cellular mass and hence yield of Ada-C (Table 3.2.2).

Type of wild type Ada-C labeling	Strain used	Media used	Use of cellular resuspension method	Typical yield (mg per litre media)	Isotope cost per litre media
None	B834(DE3)	2YT	No	100	£0
Ubiquitous <sup>15</sup> N	BL21(DE3)	2M9-G	No	9	£20
Ubiquitous <sup>15</sup> N	B834(DE3)	2M9	No	18	£20
Ubiquitous <sup>15</sup> N, 68-72% fractionally deuterated	B834(DE3)	2M9 / D <sub>2</sub> O	No	16	£220
Ubiquitous <sup>15</sup> N, 52-62% fractionally deuterated	B834(DE3)	2M9 / H <sub>2</sub> O used before resuspension, after used 2M9 / D <sub>2</sub> O	Yes	54	£220
Ubiquitous <sup>15</sup> N, 70% fractionally deuterated	B834(DE3)	2M9 / D <sub>2</sub> O used before and after resuspension	Yes	50	£420
Ubiquitous <sup>15</sup> N/ <sup>13</sup> C	B834(DE3)	Celltone	Yes	54	£1600
Ubiquitous <sup>15</sup> N/ <sup>13</sup> C, 70% fractionally deuterated	B834(DE3)	2M9 / D <sub>2</sub> O used before and after resuspension	Yes	63	£1220
Selective <sup>15</sup> N (Aromatic or aliphatic residue type)	B834(DE3)	2M9A	No	25	Varies according to residue type: eg. £20 for Ile
Ubiquitous <sup>15</sup> N, 70% fractionally deuterated	B834(DE3)	2M9 / D <sub>2</sub> O	Yes	54	£420
Selective <sup>15</sup> N / <sup>13</sup> C Cys	B834(DE3)	2M9A	Yes	100	£400
Selective <sup>15</sup> N Cys	BL21(DE3)	2M9A	Yes	33	£100 for Cys

**Table 3.2.2.** Summary of yields and costs of labeled wild type Ada-C (Prices are correct at the time of writing). Increased yields of Ada-C were obtained through the use of “cellular resuspension”, switching strains from BL21(DE3) to B834(DE3), and increasing the nutritional value of the media. 2M9A is 2M9 media with all amino acids added; 2M9-G is 2M9 media with 10g/l as opposed to the usual 4g/l glucose; celltone is richer media, commercially available from Martek, USA.

### 3.2.5 Production of unlabeled Ada-C

Leaky expression of Ada-C occurred in 2YT, and so cells were simply left overnight uninduced. Approximately 6g of wet cells were obtained from one litre of this overnight 2YT culture. Once purified this yielded an average of 100mg wild type Ada-C per litre 2YT.

### 3.2.6 Production of selectively <sup>15</sup>N-labeled Ada-C

The final yield of wild type Ada-C selectively labeled with one amino acid from <sup>15</sup>N-Leu, Val, Ala, Ile, Tyr, Phe, Glu, Gly was 25mg per litre media used. The protein production method employed B834(DE3) grown in 2M9A, without the use of cellular resuspension (section 2.1.6.2 for method). When labeling with Cys, the cost was greater and thus the cellular resuspension method was used (Table 3.2.2)

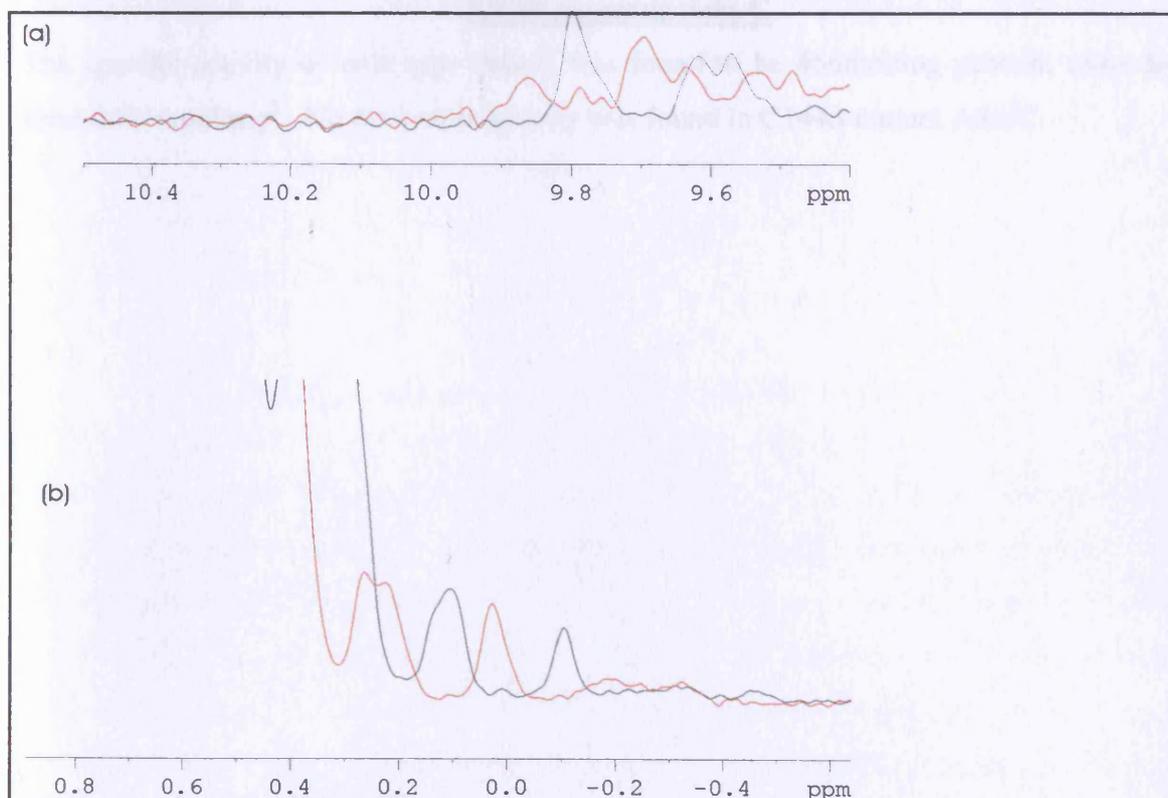
Since B834(DE3) was not auxotrophic for any of the amino acids to be labeled, crosslabeling was occasionally observed. The reason this strain was chosen was due to its high expression levels of Ada-C. The strain DL39(DE3) is polyauxotrophic for aliphatic amino acids, but this strain failed to express Ada-C and thus could not be used. Crosslabeling was observed by NMR on the purified protein. HSQC experiments that revealed more peaks than expected, indicated crosslabeling. From examining biochemical pathways it was possible to postulate the identity of the amino acid type causing the extra peaks. In many cases the postulated destinations of the <sup>15</sup>N-amino group were confirmed by labeling Ada-C with the final <sup>15</sup>N-labeled amino acid itself. The results of crosslabeling are given in Table 3.2.3 below.

<sup>15</sup> N-Labeled Amino Acid	Cross-labeling observed	Product of cross-labeling
Leu	No	-
Ile	No	-
Val	Yes	Ala
Ala	No	-
Tyr	No	-
Phe	No	-
Gly	Yes	Ser
Glu	Yes	Gln, Ala
Cys	Yes	Ala, Glu, Met, Ser

**Table 3.2.2** Results of selective labeling of wild type Ada-C in B834(DE3) in terms of crosslabeling. Wild type Ada-C was labeled with <sup>15</sup>N-Cysteine using the auxotroph BL21(DE3)*cys*<sup>-</sup> which prevented any crosslabeling. When labeling the protein with more expensive <sup>15</sup>N / <sup>13</sup>C-Cysteine, and labeling the lower yielding mutant C144S Ada-C with <sup>15</sup>N-Cysteine, B834(DE3) was used to increase yields. Use of B834(DE3) resulted in crosslabeling. This crosslabeling of Cys was of a much lower scale than with some aliphatic residues, such as Val when B834(DE3) was used. This is because Cys is not a direct precursor to the synthesis of other residue types.

### 3.2.7 Production of uniformly labeled deuterated Ada-C

Yields of mutant C144S Ada-C were always 2 to 3 times lower than those of the wild type protein produced in the same manner. A similar decrease in yield was also observed for the homologous human alkyltransferase when the methyl acceptor cysteine was mutated to a serine (Hazra *et al.*, 1997). This may reflect the lower intrinsic stability of this protein with respect to the wild type in the cell (Chapter 6). The cellular resuspension method was utilised extensively to enhance yields of both mutant and wild type Ada-C during deuteration and uniform labeling (Table 3.2.2). NMR decoupling analysis revealed that resuspension of cells from unlabeled media into media containing the desired isotopes, did not result in noticeable isotopic dilution (Figure 3.7). Conversely, deuteration levels were affected by this method. When the cellular resuspension method was not employed, and cells were inoculated directly into 2M9 D<sub>2</sub>O media, mass spectrometry revealed deuteration levels to lie between 68% and 72%. Yields were low using this method, so the cells were grown up first in unlabeled 2M9 H<sub>2</sub>O media and then resuspended into labeled 2M9 D<sub>2</sub>O media (containing <sup>13</sup>C or <sup>15</sup>N) where expression was induced. This method led to a three fold yield enhancement, but caused a decrease in deuteration levels to 52%-62%. At increased expense, cells could be inoculated directly into unlabeled 2M9 D<sub>2</sub>O media and grown up without expression. Cellular resuspension into labeled 2M9 D<sub>2</sub>O media followed by induction of expression led to similarly high yields and an increase in deuteration levels to 70%.



**Figure 3.7** (a) 1D proton spectrum with <sup>15</sup>N-<sup>1</sup>H decoupling on (black) and off (red). (b) 1D proton spectrum with <sup>13</sup>C-<sup>1</sup>H decoupling on (red) and off (Black). The difference in signal intensities between the red and black reveal good incorporation of <sup>15</sup>N and <sup>13</sup>C.

For examining protein-DNA interactions, lower deuteration levels are acceptable. However, when using the  $^{13}\text{C}/^{15}\text{N}$ -labeled protein for triple resonance experiments the level of deuteration should be as high as possible. For this reason when making the  $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C, the method used involved cell growth in 100%  $\text{D}_2\text{O}$   $^{12}\text{C}/^{14}\text{N}$  2M9 media prior to cellular resuspension into  $^{13}\text{C}/^{15}\text{N}$ -labeled 100%  $\text{D}_2\text{O}$  media for induction.

### **3.2.8 Purification of Ada-C**

The improved purification protocol resulted in higher efficiency than the original protocol in terms of yield and time. Ada-C preparations following this purification technique were judged to be 95% - 99% homogeneous by SDS-PAGE analysis. No problems were encountered with proteolysis of Ada-C during or following its purification. This was not the case when the original protocol was employed.

### **3.2.9 Quality control of Ada-C**

The mass of wild type Ada-C as determined from electrospray mass spectrometry was 19442 Da. The sequence of the first ten N-terminal residues of wild type Ada-C were confirmed by N-terminal sequencing. Masses of C144S and C144A mutants were also confirmed as being 19426 Da and 19410 Da respectively.

### **3.2.10 Assaying Ada-C**

The specific activity of wild type Ada-C was found to be 46nmol/mg protein, close to the limit of 51nmol/mg<sup>3</sup>. No detectable activity was found in C144S mutant Ada-C.

---

<sup>3</sup> The activity limit is reached when all Ada-C molecules undergo a single demethylation reaction when reacting with methylated DNA. This limit is 51nmol/mg because 1mg protein = 51nmol.

### **3.3 DISCUSSION**

#### **3.3.1 Increasing the yield of Ada-C**

Due to the expense involved in labeling proteins for NMR work, it is of paramount importance to be able to produce a protein at high yield per litre of media. The purpose of the work described thus far in this chapter was to enhance final yields of Ada-C, by maximising cellular expression, and minimising loss through design of an efficient purification protocol.

##### **3.3.1.1 Increasing the yield of Ada-C through media design**

Initially, Ada-C was produced in BL21(DE3), a prototrophic strain which could easily grow in the minimal media used for introducing isotopes into protein. BL21(DE3) suffered from low expression levels of Ada-C and thus attempts were made to enhance these by manipulating the times of induction and harvest, and the composition of the media. Apart from the detrimental effect of harvesting 24 hours after induction, or inducing in stationary phase, the exact points of harvest and IPTG addition had little effect on cellular levels of Ada-C. The only significant factor in the media composition to have an effect on cell growth was the glucose concentration. Increasing this from 4g/l to 10g/l resulted in a doubling of cell density in stationary phase. Induction at mid log phase in 2M9 media with 10g/l glucose allowed the yield of Ada-C obtained per litre of culture to be increased by 50% in comparison to when 4g/l glucose was used (where mid log phase occurred at half the cell density (table 3.2.1). Use of media containing 10g/l of glucose is suitable for producing <sup>15</sup>N-labeled protein but not <sup>13</sup>C where the increased cost caused by glucose utilisation outweighs any gains in yield. This is a clear example of how the media composition directly affects the protein yield. The high induction cell density was necessitated by the fact that IPTG arrests the growth of BL21(DE3) in minimal media.

##### **3.3.1.2 Increasing the yield of Ada-C through strain alteration**

IPTG induction did not arrest cell growth in B834(DE3), which is more tolerant of high levels of Ada-C production than its daughter strain BL21(DE3). In B834(DE3), stationary phase in 2M9 media occurred at the same cell density ( $A_{595} = 4.5$  regardless of whether there was 4g/l or 10g/l glucose present (Figures 3.3-3.5). This strain also produced twice the expression level of Ada-C in rich media (2YT) than BL21(DE3). The disadvantage of using B834(DE3) over BL21(DE3) is that in minimal media L-Met must be added since B834(DE3) is a Met auxotroph. This means methionine signals will be missing from NMR spectra unless labeled Met is added to the media.

Such an impressive difference in protein overexpression by B834(DE3) compared with BL21(DE3) has been previously noted for both Ada-C and T7 Ligase, (Doherty *et al.*, 1995). However, there are indications that such expression enhancement does not always occur. BCII  $\beta$ -lactamase and DHFR showed no detectable increase in expression level in B834(DE3) compared to BL21(DE3).

No increase in yield of Ada-C was observed in the non-DE3 strain WA834 compared with BL21 (when using the *ptac* expression system with Ada-C). However, increases have been observed in WA834 compared to BL21 with the proteins rac and rho GDI. The ability of B834(DE3) to enhance expression levels depends on a complex interplay between protein, strain and plasmid expression system.

In the case of Ada-C, why B834(DE3) is able to withstand high expression of certain proteins better than BL21(DE3) is a mystery. BL21(DE3) was created from B834(DE3) by p1 transduction to *Met*<sup>+</sup>, (Studier *et al.*, 1986). In theory the only difference between the two strains is that B834(DE3) is *MetB*<sup>-</sup> whilst BL21(DE3) is a prototroph. However, possibly during the transduction process other alterations were made to the genome, since it is unclear how a deficiency in Cystathionine  $\gamma$ -synthase<sup>4</sup> (EC 4.2.99.9, *MetB*<sup>-</sup>) would confer the high expression capability to a cell.

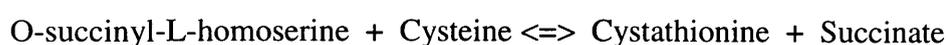
Expression of Ada-C in both B834(DE3) and BL21(DE3) was leaky in rich media such as 2YT, but tightly regulated in 2M9 minimal media. The absence of leaky expression in 2M9 is attributed to the high glucose content of the media, which leads to lower cAMP levels through inhibition of adenylate cyclase, and hence lower Catabolite Activator Protein activity. Such glucose levels are absent in 2YT.

### **3.3.1.3 Cellular resuspension to further enhance Ada-C yield**

High expression levels in B834(DE3) could be further enhanced through the use of the cellular resuspension method. Here, cells are grown in unlabeled media to a high cell density ( $A_{595} = 2$  when the unlabeled media was H<sub>2</sub>O-based, or  $A_{595} = 4$  when D<sub>2</sub>O-based media was used), they are then separated from this used media and resuspended into fresh media containing the isotope labels for incorporation into Ada-C. The theory behind this, is that enhanced biomass produced increased yields, provided the cells were not in stationary phase upon induction. Transfer to fresh media enabled higher cell densities to be reached without the cells being in stationary phase when induced. Cells would use the unlabeled media to

---

<sup>4</sup> NB- Cystathionine  $\gamma$ -synthase catalyses:



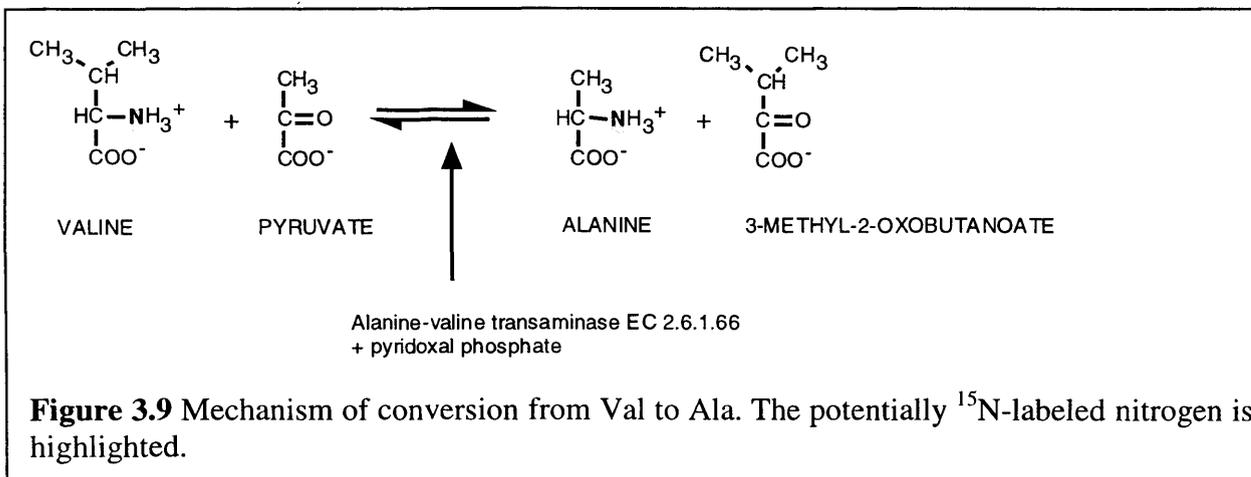
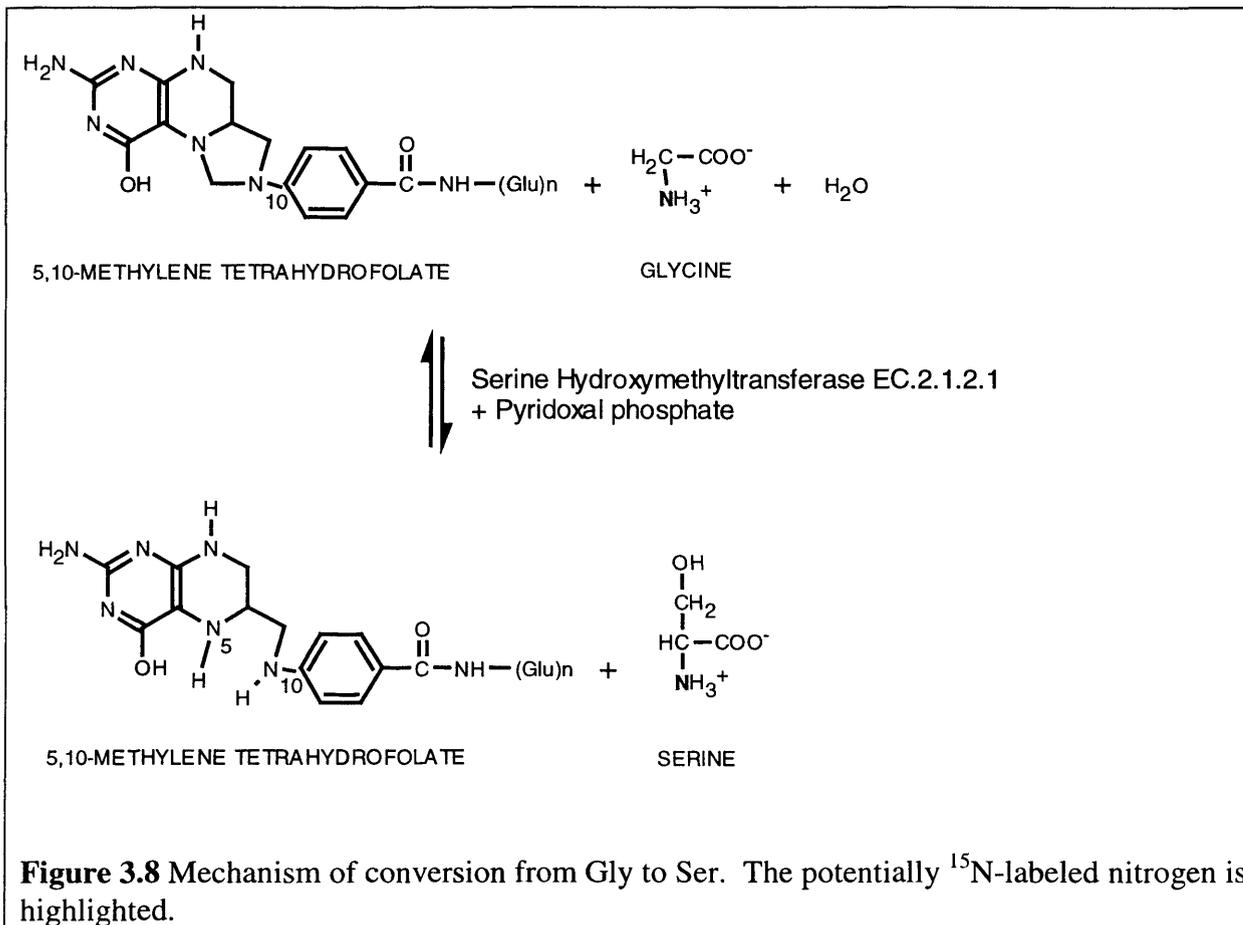
grow, and the labeled media purely for protein production, since expression would only be induced in the labeled media. Transfer into fresh media would also remove any toxins which had built up as a result of normal metabolism, or attainment of stationary phase in the first media, which may hamper protein production. Clearly, this thinking proved correct, and cellular resuspension often led to a tripling of protein yield (Table 3.2.2).  $^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C was produced at levels of 54mg/l media by resuspending cells grown to mid-log phase in 2M9 into labeled celltone media, and then inducing with IPTG. Isotopic incorporation into the protein was judged to be very good (Figure 3.7). Deuteration presented more of a problem. Triple resonance experiments, required for NMR assignment of the backbone resonances, are enhanced by a level of 70-90% deuteration (Shan *et al.*, 1998), (Zhou *et al.*, 1998). Deuteration levels of 68-72% were obtainable from inducing B834(DE3):pet22b:adaC in 100% D<sub>2</sub>O 2M9 media. However, without cellular resuspension, this technique rendered low Ada-C yields (16mg/l). Introducing a cellular resuspension step in which cells are initially grown in unlabeled 100% H<sub>2</sub>O 2M9 media, and subsequently introduced to labeled 100% D<sub>2</sub>O 2M9 media, resulted in an increase in yield (54mg/l), but a decrease in deuteration levels to 52-62%. Clearly, growth of cells for long periods of time in H<sub>2</sub>O-based media results in a reduction in deuteration levels. The solution was to grow cells initially in unlabeled 100% D<sub>2</sub>O 2M9 media, then resuspend them in 100% D<sub>2</sub>O 2M9 media containing the required label (either  $^{15}\text{N}$  or  $^{13}\text{C}$ , or both). This lead to a high yield (63mg/l) and higher deuteration levels of 70%. It was using this latter technique that triple-labeled Ada-C was produced that was 70% deuterated, and  $^{13}\text{C}/^{15}\text{N}$ -labeled for triple resonance experiments (Chapter 5).

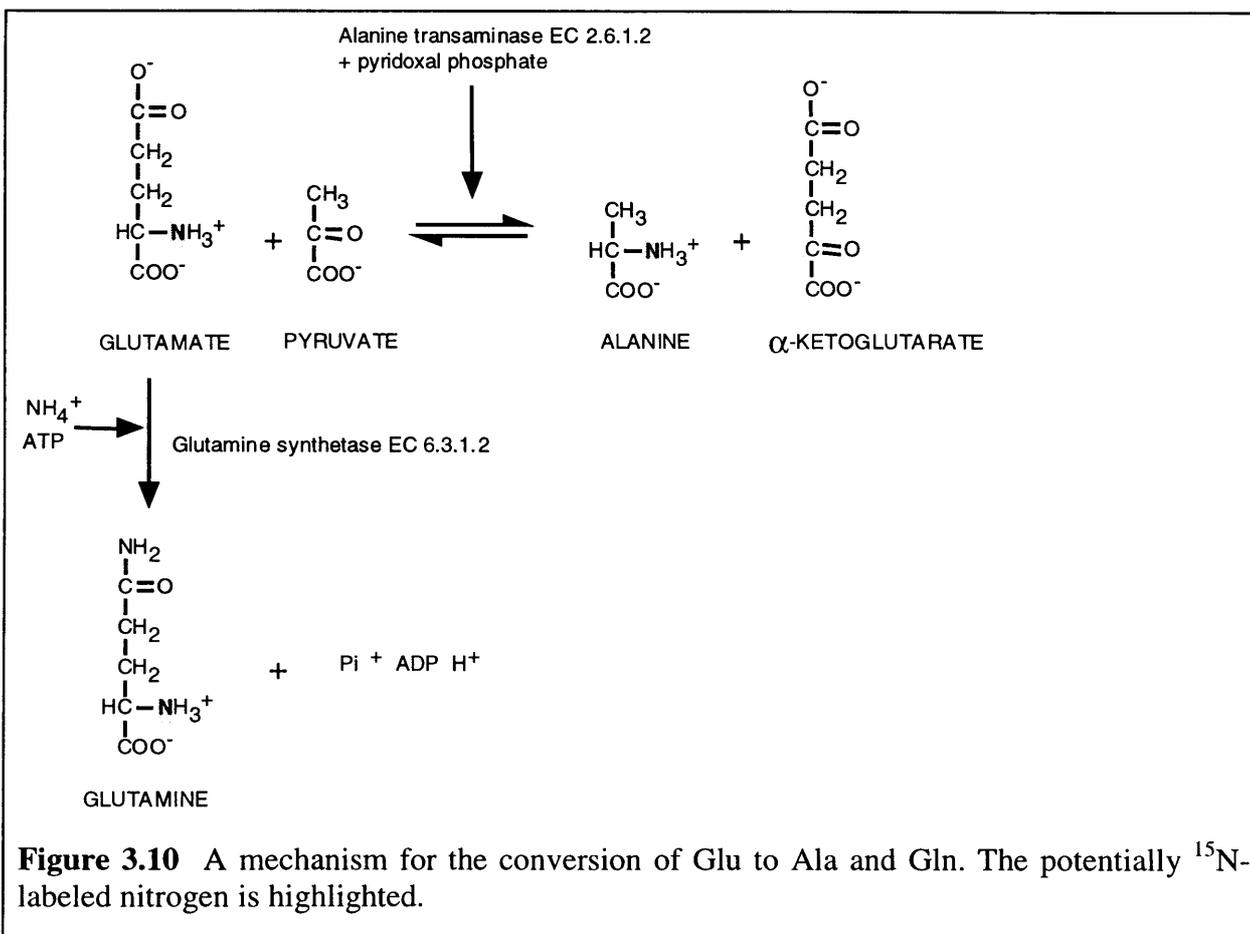
The cellular resuspension technique proved invaluable for enhancing yields of Ada-C expressed in *E.coli*. As a consequence of these increased yields, sufficient protein was produced to enable the backbone assignment of Ada-C. However, although this method worked exceptionally well for Ada-C, it was not universally applicable. Cellular resuspension had no effect on yield enhancement with rac for instance (R.Badii, personal communication).

### **3.3.2 Selectively labeling Ada-C and monitoring crosslabeling**

The selective labeling of Ada-C with aliphatic and aromatic residues in B834(DE3) was highly successful in terms of yield and stable isotope incorporation. Only for Gly, Val and Glu was there a significant level of crosslabeling observed. Crosslabeling occurs between amino acids during selective labeling in B834(DE3) because transaminase enzymes enable the amino  $^{15}\text{N}$  to be switched from one carbon skeleton to another. Val and Ala are both in the same pyruvate family of amino acids. Gly and Ser belong to the serine family whilst Glu and

Gln belong to the glutamate family. The simple transaminations which have lead to observed crosslabeling occur both within, and between families of amino acids (Figures 3.8 to 3.10).





Attempts were made to prevent crosslabeling of Val to Ala by increasing the amount of unlabeled Ala in the media. However crosslabeling was still observed. Stopping crosslabeling of Glu should prove immensely difficult due to its use in many transamination reactions, and was not attempted.

Cys labeling through the use of BL21(DE3)*cys<sup>-</sup>* was very successful in that no crosslabeling was observed. Use of the non-Cys auxotrophic B834(DE3) strain to label Ada-C with <sup>13</sup>C and <sup>15</sup>N-Cys resulted in small amounts of crosslabeling observed to Ala, Glu, Met, Ser. This crosslabeling appeared to be more random, and in low quantities. Cys is not a direct precursor in the synthesis of another amino acid type. This is why crosslabeling levels were lower than for residues such as Glu (which can be used directly to make Gln).

### 3.3.3 Purification of Ada-C

The design of a superior purification protocol involved removal of a DNA precipitation step, and its replacement with Dnase treatment, and the substitution of anion exchange for cation exchange column chromatography prior to gel filtration. These two changes augmented both yield and time efficiency of the purification protocol.

### **3.3.4 Observations regarding Ada-C stability**

Both wild type and mutant Ada-C proteins exhibited identical behaviour during purification, although the yields of C144S mutant were always 2-3 times lower than wild type Ada-C. This maybe because the lower stability of C144S (Chapter 6) results in diminished cellular levels. The C144A mutant Ada-C was expressed and purified normally, but precipitated upon concentration. When repeated the same phenomenon occurred. This C144A mutant of Ada-C is unstable in concentrated form. A change of the crucial methyl acceptor cysteine to an alanine may cause disruption to the stabilising hydrogen-bonding network in the active site, involving the cysteine sulfhydryl group (Figure 1.12, Chapter 1). This may result in localised instability. Such instability may spread (perhaps by breaking the salt link between the C-terminal helix glutamate and active site histidine) and cause unfolding and precipitation. This has connotations for the theory into how the adaptive response is terminated and will be discussed further in Chapter 6.

### **3.4 SUMMARY**

To a small degree the level of Ada-C production in the strain BL21(DE3) was enhanced by increasing the glucose concentration of the media, and inducing the cells at higher cell density. Larger Ada-C yield enhancements were obtained by switching strains from BL21(DE3) to B834(DE3). Yields of Ada-C were then further increased also by employing the cellular resuspension method. This enabled the removal of toxins from the media and the increase in cell density prior to induction. Labeled media could then be employed specifically for production of Ada-C, with less wasted on culture growth. The culmination of these advances in protein yield was the production of 63mg of 70% deuterated,  $^{13}\text{C}/^{15}\text{N}$ -labeled wild type Ada-C, and 54mg of protonated  $^{13}\text{C}/^{15}\text{N}$ -labeled wild type Ada-C (for triple resonance studies), using  $^{13}\text{C}$ -glucose and  $^{15}\text{N}$ -ammonium sulphate as unique carbon and nitrogen sources.

Although wild type Ada-C is active, no activity could be detected in the C144S mutant, and the C144A mutant is unstable leading to implications for termination of the active response (discussed in Chapter 6).

# Chapter 4

## *Thermodynamics of Ada-C : DNA interactions*

### **4.1 INTRODUCTION**

Structural information on protein-DNA complexes only provides limited information on the physical basis for specificity and stability of the interaction. NMR data can be used to obtain binding constants (provided the components are in fast exchange), along with structural information, whilst high resolution structures generated by x-ray crystallography provide merely a static picture with no thermodynamic basis for the interaction. The general problem is that structures of complexes reveal little about the entropy or free energy of complex formation. Equilibrium binding analyses are vital to complement the structural data, if our understanding of how these complexes form is to advance. A great deal of work has been done in characterising the thermodynamics of protein-DNA interactions. Enthalpic changes upon binding reveal information about bond formation such as hydrogen bonding, Van der Waals interactions, proton linkage, and along with entropic factors, the involvement of ions and water in binding processes.

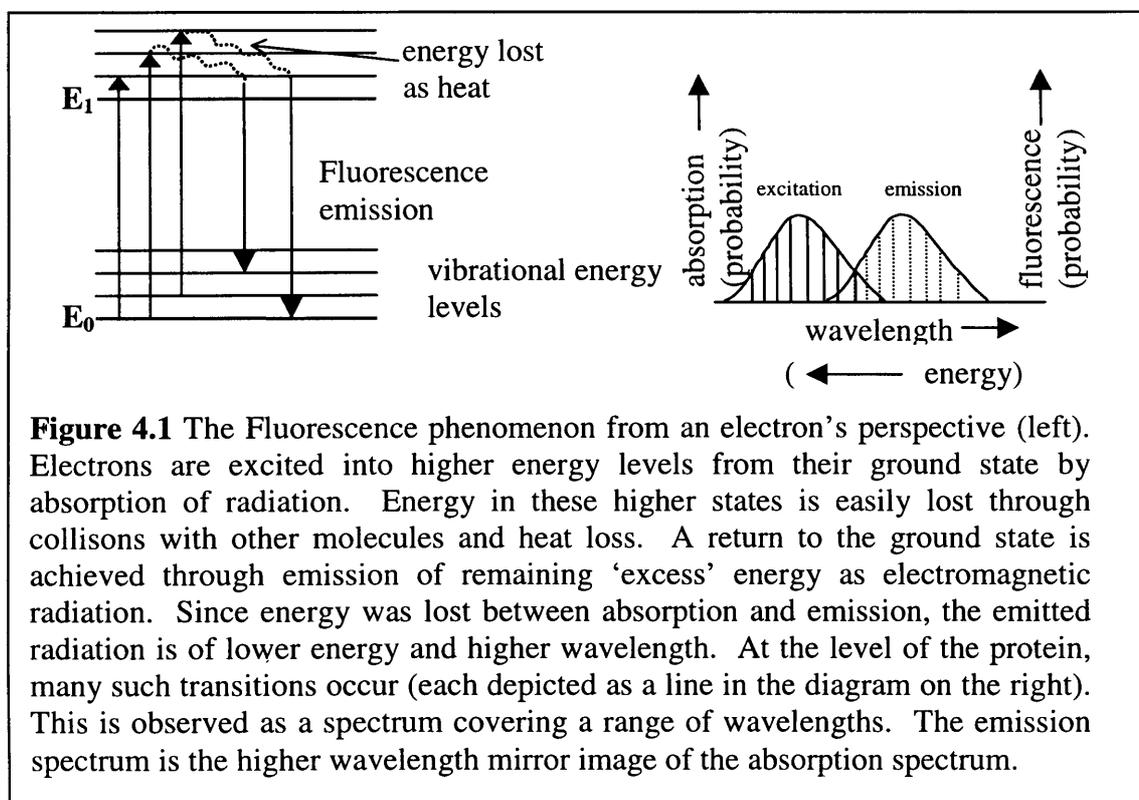
Various techniques exist which can be used to derive thermodynamic information about complex formation. Anisotropic and isotropic fluorescence studies, circular dichroism, analytical ultracentrifugation and surface plasmon resonance are all capable of producing binding constants between macromolecules. Where Van't Hoff analysis has been used, further information, such as enthalpy and entropy of binding can be determined.

However the most definitive method is that of Isothermal Titration Calorimetry (ITC). The advantage of this technique over others is that direct measurement of enthalpy changes involved in a macromolecular interaction can be obtained with precision. The heat signal is a nearly universal property of binding reactions, and thus ITC is applicable to many ligand-macromolecule or macromolecule-macromolecule interactions (Wiseman *et al.*, 1989). Fluorescence binding assays, on the other hand rely on spectroscopic signaling from fluorophores whose properties become perturbed upon binding. Such signals may not be present in every system, making fluorescence less universally applicable than ITC. Surface Plasmon Resonance (SPR) requires immobilisation of one component (Chaiken *et al.*, 1992). The ITC method does not rely on any intervention to the system under study and therefore requires fewer controls.

Along with NMR, other techniques employed in this thesis to examine binding of Ada-C to DNA were Fluorescence, Circular dichroism, and ITC. A brief description of these techniques is given in the following pages.

#### 4.1.1 Measuring binding by Fluorescence

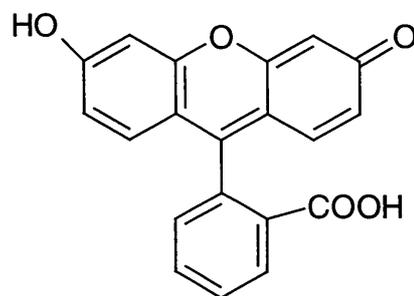
Fluorescence is an emission phenomenon, where transitions from a higher energy state to a lower state are measured by detection of emitted radiation. In order for such a transition to occur, an earlier excitation event must first have occurred. This event entails the absorption of electromagnetic radiation. In proteins, aromatic sidechains from Trp, Phe and Tyr absorb at the uv wavelengths, as do DNA bases. The absorption of energy quanta allows electron transitions to occur from their ground state,  $E_0$  to a state of elevated energy  $E_1$  (Figure 4.1).



**Figure 4.1** The Fluorescence phenomenon from an electron's perspective (left). Electrons are excited into higher energy levels from their ground state by absorption of radiation. Energy in these higher states is easily lost through collisions with other molecules and heat loss. A return to the ground state is achieved through emission of remaining 'excess' energy as electromagnetic radiation. Since energy was lost between absorption and emission, the emitted radiation is of lower energy and higher wavelength. At the level of the protein, many such transitions occur (each depicted as a line in the diagram on the right). This is observed as a spectrum covering a range of wavelengths. The emission spectrum is the higher wavelength mirror image of the absorption spectrum.

In aromatic systems the excitation transition involves electrons moving from  $\pi$  to higher energy  $\pi^*$  orbitals. Energy in this higher level state is rapidly dissipated as heat to the surrounding environment. This partially explains why the best sensitivity and accuracy is achieved at low protein concentrations, because there are fewer molecules and hence collisions (less 'quenching'). The release of energy allows drops to lower vibrational energy levels (smaller energy differences than between  $E_0$  and  $E_1$ ). A subsequent decay back to ground state  $E_0$  occurs through emission of the remaining energy. This emission peak will be at a higher wavelength than the peak of absorbed radiation, since less energy is emitted than originally absorbed. The emission spectrum appears as a red-shifted mirror image of the absorption spectrum (Figure 4.1, right).

The wavelength of emitted radiation is sensitive to the environment of the aromatic group in the protein or DNA. An alteration in this environment caused by a binding interaction may lead to a measurable shift in emitted wavelength (so called 'Stokes shift') or a decrease in quantum yield at the same wavelength. This change in fluorescence can be quantified and, when plotted against concentration of ligand added, can yield a binding curve from which a  $K_d$  can be obtained. Such isotropic fluorescence measurements only yield binding information if such an emission shift can be observed. Emission shifts are also sensitive to changes in solvent, pH and temperature, so all experiments have to be carefully controlled, and buffers of reactants matched prior to mixing. Care must also be taken to avoid the use of certain plasticware which may release contaminating fluorophores. Concentrations of protein normally employed are in the  $\mu\text{M}$  range as are the measured  $K_d$  values. Binding interactions with millimolar  $K_d$  values are probably not best measured using fluorescence due to the mM concentrations that would have to be employed (more "quenching"). Should internal aromatics fail to provide a measurable signal upon ligand binding use might be made of a fluorescent probe. These can be attached to protein or DNA, and may provide a signal which differs upon ligand binding. In this research Fluorescein was attached the end of DNA in an attempt to monitor binding to Ada-C (Figure 4.2).



**Figure 4.2** The structure of the fluorescent probe Fluorescein.

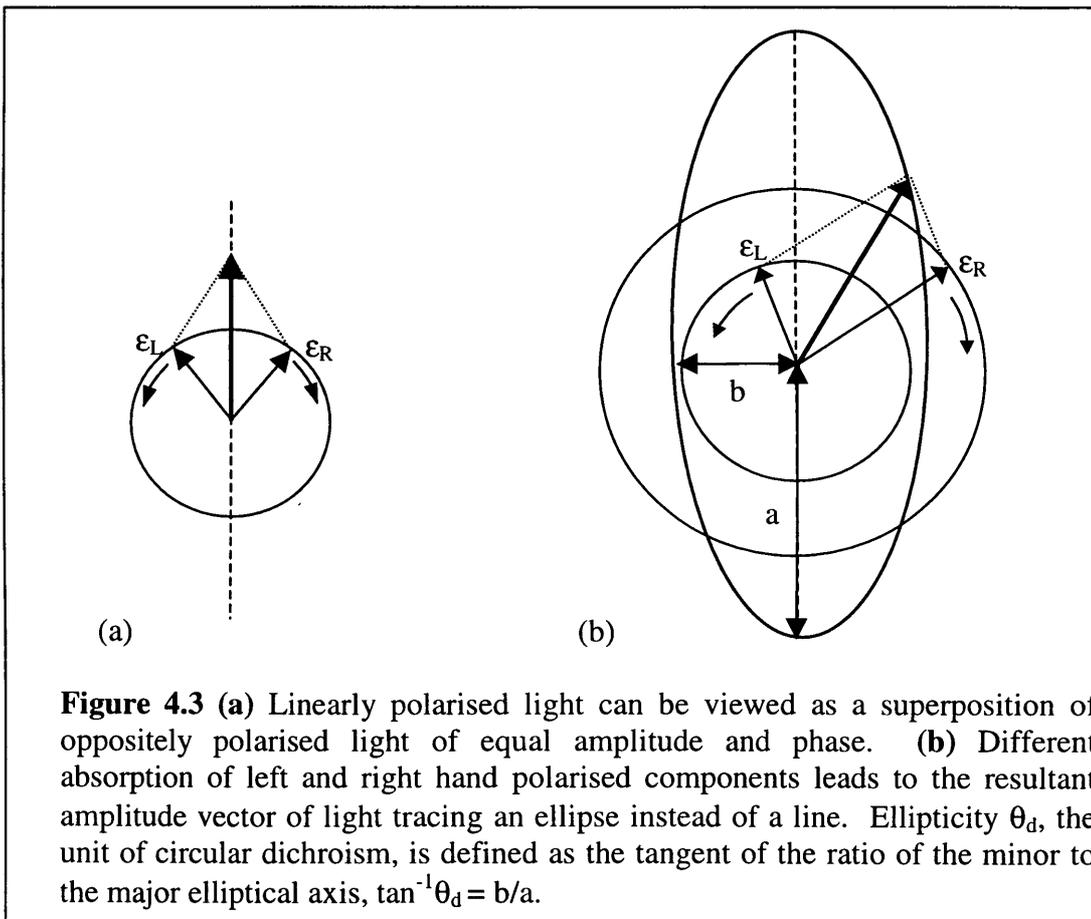
Failing any isotropic change in fluorescence upon ligand-binding, binding constants can be obtained by virtue of changes in fluorescence depolarisation, so called anisotropic fluorescence changes. A molecule absorbs light with greatest efficiency when the wave vibrates along certain directions in the molecule. When the light used for excitation is plane polarised, it only excites molecules in the appropriate orientation. The light subsequently emitted from these molecules is also polarised. The average polarisation of the light passing through the solution will be decreased by the extent to which the molecules have rotated following absorption, but prior to emission of light. This depolarisation can be measured by examining the intensities of horizontally and vertically polarised light. When a ligand such as DNA is bound to the original protein molecule, the mass and consequently the rotation time in

solution increases. Provided the measured fluorophore is stiffly attached to the macromolecules, there will be a concomitant increase in anisotropy (ie. less depolarisation of the originally plane polarised light). The anisotropy can be plotted as a function of ligand concentration to obtain a binding constant. This technique was successfully used to obtain the binding constant of Ada to DNA (Takahashi *et al.*, 1990).

#### 4.1.2 Examining binding by Circular Dichroism

Based on (Campbell *et al.*, 1984) and (Berndt, 1996).

Circular Dichroism (CD) is another technique based on the polarisation of light. In linearly polarised light as used in anisotropic fluorescence, the light wave oscillates in a set plane, only its amplitude varies. In circularly polarised light, the direction of oscillation alters (Figure 4.3). Circular dichroism is defined as the difference in molar extinction coefficients for left- and right-handed circularly polarised light: molar CD =  $\epsilon_L - \epsilon_R = \Delta\epsilon$ .



CD is measured as the difference in absorbance of right- and left- circularly polarised light as a function of wavelength. Despite this, for historical reasons ellipticity is the unit of CD (Figure 4.3). The molar ellipticity,  $\theta_m$  is related to the difference in extinction coefficients by:

$\theta_m = 3298 \Delta\epsilon$  (units: degree  $\text{cm}^2 \text{dmol}^{-1}$ ). The raw data output for the JASCO spectropolarimeter is ellipticity  $\theta_d$  measured in units of degrees.  $\theta_d$  is defined as:

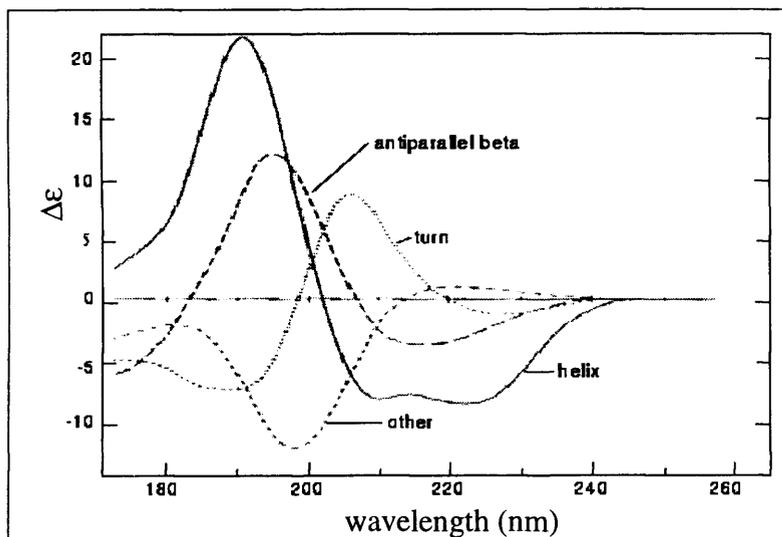
$$\theta_d = \frac{2.303}{4} \cdot (A_L - A_R) \cdot \frac{180}{\pi}$$

This can be converted to molar ellipticity by multiplication of  $\theta_d$  by the following relationship:

$$\theta_m = \frac{(\theta_d M_w)}{100Lc}$$

Where L is the path length in decimeters (1cm = 0.1dm), and c is the concentration of protein in g/ml. (Conversion to the standard unit of molar ellipticity  $\theta_m$  is only required if secondary structure analysis is to be carried out. In the case of Ada-C, such analysis was not performed since there were no visible differences in the CD spectra ( $\theta_d$  versus wavelength) of Ada-C free and bound to DNA.)

A negative CD is observed if  $A_R > A_L$ , and a positive CD observed if the opposite is true. CD bands of proteins occur in two regions; the far uv (170-250nm) which is dominated by the peptide bonds, and the near uv region (250-300nm) attributed to aromatic sidechains. A CD spectrum is created by a number of different  $n \rightarrow \pi^*$  and  $\pi \rightarrow \pi^*$  transitions. An  $n \rightarrow \pi^*$  transition gives rise to the 220nm band in an  $\alpha$ -helix for example. Optical activity only arises when the environment in which a transition occurs is asymmetric. One such environment is a peptide bond which shows CD near 220nm and 190nm. When part of secondary structure, these amide bonds become ordered, and consequently their electronic transitions can interact. The arrangement of peptide bonds dictates the CD spectral pattern observed (Greenfield, 1996). Different secondary structure elements give rise to typical CD spectra (Figure 4.4). An observed CD spectrum of a protein is the linear combination of the sub spectra of secondary structure elements weighted according to their abundance. This forms the basis of secondary structure prediction by CD. Spectra of 'pure' secondary structure elements are combined in different percentages until the observed protein spectrum is reproduced. The secondary structure content can then be defined. Both near and far uv CD spectroscopy are useful tools in identifying conformational changes in protein structure which may occur upon ligand binding. Far uv CD has the added advantage that secondary structure information may also be derived.



**Figure 4.4** Typical CD spectra of various structural elements within proteins. This picture was reproduced from Berndt *et al.* (1996).

When measuring CD spectra of proteins care must be taken not to employ buffers which may also absorb in the uv region. NaF is frequently used in place of NaCl for increasing ionic strength due to the lower absorbance of uv light by the F<sup>-</sup> ion.

### **4.1.3 An overview of ITC- How it works**

ITC is the only technique which allows simultaneous determination of all binding parameters ( $K_d$ ,  $\Delta H$ ,  $\Delta S$ , stoichiometry) in a single experiment. The signal to noise ratio when using a microcalorimeter is very high, enabling high dilutions to be used to measure binding constants approaching nanomolar (Wiseman *et al.*, 1989), (Doyle, 1997). Measurable weak binding affinities are in the millimolar range (Doyle, 1997). With its very precise computer-controlled, motor-driven, stepped-injection regime, the ITC unit is capable of determining binding stoichiometry accurately. High reproducibility of results is also possible because reactants are analysed in their native forms. There is no need for chemical modification such as in fluorescence or immobilisation as in SPR methods which are inherently irreproducible and lead to error in determining the stoichiometry.

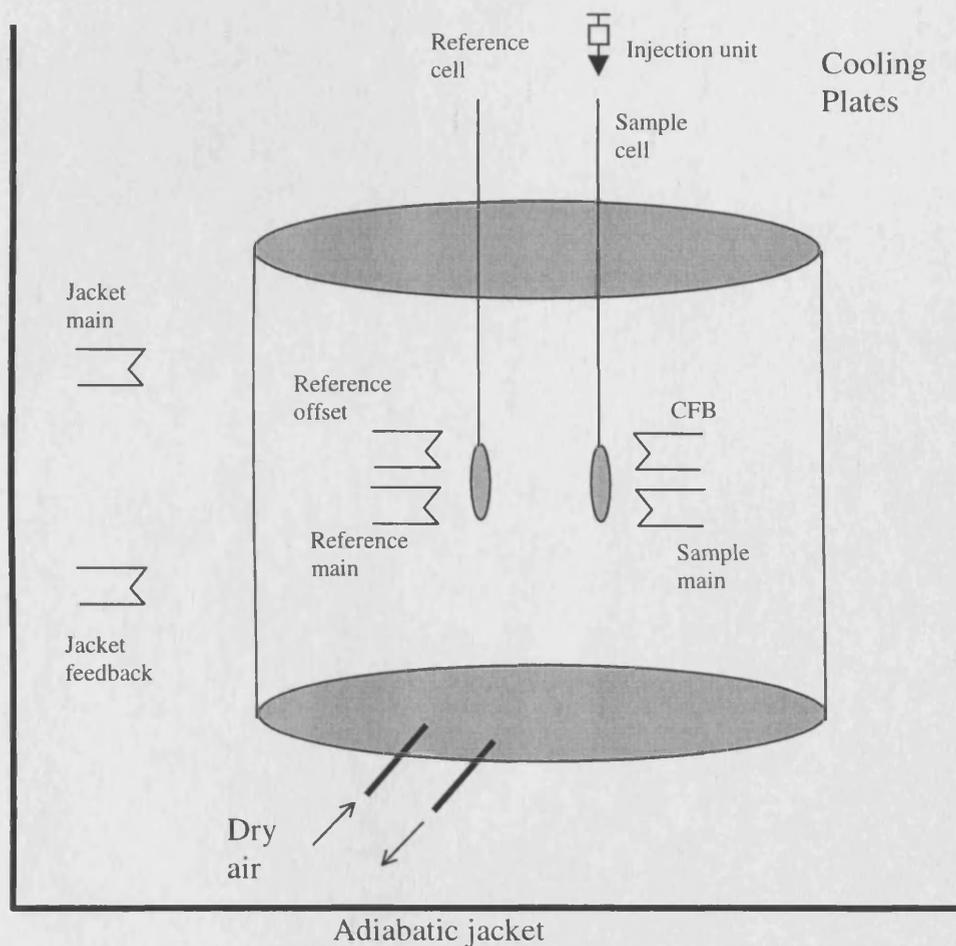
The ITC unit directly measures heat evolved or absorbed as a result of injecting precise amounts of one reactant which binds to the other reactant in the sample cell. A spinning syringe is used for injection and subsequent mixing. There is another identical coin-shaped 1.34 ml volume cell which is the reference cell. This is filled with buffer or water and acts purely as a reference (nothing is injected into this). These two cells are kept at thermal equilibrium by cooling the adiabatic jacket in which they are housed, such that heat must be continuously added to keep the cells and their contents at the experimental temperature.

During an experiment the reference cell is heated by a very small constant power, the reference offset (Figure 4.1).

The temperature difference between the two cells is continuously measured. By increasing or reducing the amount of heating power applied to the sample cell, the unit tries to maintain a constant temperature in that cell. A signal proportional to this cell feedback power (CFB) is generated, which along with the instrument temperature, and time, makes up the raw data for the experiment. CFB is calibrated in units of  $\mu\text{cal}/\text{sec}$ . When one component of the binding interaction is injected into the sample cell (typically in aliquots from 5-25 $\mu\text{l}$ ), binding may give rise to an enthalpic change. If the binding reaction is endothermic more heat will have to be applied to the sample cell to maintain thermal equilibrium with the reference. Thus the CFB will show a positive spike. The opposite is true for exothermic reactions where a negative spike is observed. The experimental volumes are set up such that over a series of additions the binding sites of the molecule in the sample cell become saturated, such that the heat observed with last few additions is simply the heat of dilution of the second molecule. The heat input per injection is determined by integrating the CFB peaks with respect to time. This processed data can then be plotted as kcal/mole of heat absorbed, or emitted per molar ratio of the two interacting components.

The result is a sigmoidal binding curve as shown in Figure 4.3 (results section 4.2.3), from which  $K_a$ ,  $\Delta H$ , and the stoichiometry can be determined using the relationships and fitting procedures described in Appendix 2.1.

## Isothermal Titration Calorimetry Unit



**Figure 4.1** An Isothermal Calorimetry Unit

### 4.1.4 Interpreting thermodynamic information

When performing the ITC titrations it is necessary to repeat them using different buffers. This stems from the fact that buffer ionisation may occur which would provide a misleading binding enthalpy. By performing the experiment in different buffers this can be corrected if the enthalpies of buffer ionisation are known. If a binding reaction involves a change in protonation, the measured enthalpy  $\Delta H_{\text{meas}}$  is the sum of the reaction binding enthalpy  $\Delta H_{\text{bind}}$ , which is independent of the buffer used, and a term proportional to the enthalpy change of the buffer such that:

$$\Delta H_{\text{meas}} = \Delta H_{\text{bind}} + N_{\text{H}^+} \Delta H_{\text{ion}}$$

where  $N_{\text{H}^+}$  is the number of protons released by the buffer, and  $\Delta H_{\text{ion}}$  is the ionisation heat of the buffer. The slope of a plot of  $\Delta H_{\text{ion}}$  for buffers against  $\Delta H_{\text{meas}}$  yields information about the thermodynamic linkage between ligand binding and protonation. A positive slope of +1.0 indicates that the one proton is taken up by the complex upon binding, conversely a negative

slope indicates protons are released upon binding. The intercept yields  $\Delta H_{\text{bind}}$ , (Murphy *et al.*, 1993), (Eads *et al.*, 1998). Along with proton linkage, enthalpy of binding obtained from ITC data provides information about bond formation, and other contacts between binding interfaces. Together with entropy considerations, a picture of the thermodynamic processes involved in binding can be built up.

## **4.2 RESULTS**

### **4.2.1 Fluorescence**

No shifting in the wavelength of emission maximum was observed when exciting either fluorescein tagged on to the 5' end of unmethylated, single-stranded DNA, or exciting aromatic residues in Ada-C. In the case of Ada-C fluorescence, addition of unmethylated, single-stranded DNA caused a decrease in maximal fluorescence intensity. The same decrease in signal intensity was observed when an identical concentration of DNA was added to free tryptophan, thus attributing the decrease to the inner filter effect of DNA (DNA absorbs light at the excitation frequency and thus reduces the overall emission intensity) Figure 4.1.

No binding constants could be determined from these data sets, as no change in intensity or shift of the fluorescence signal occurred as a result of binding.

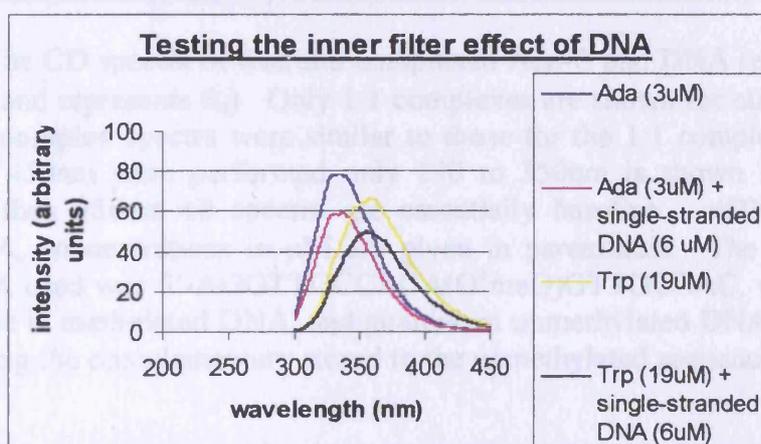
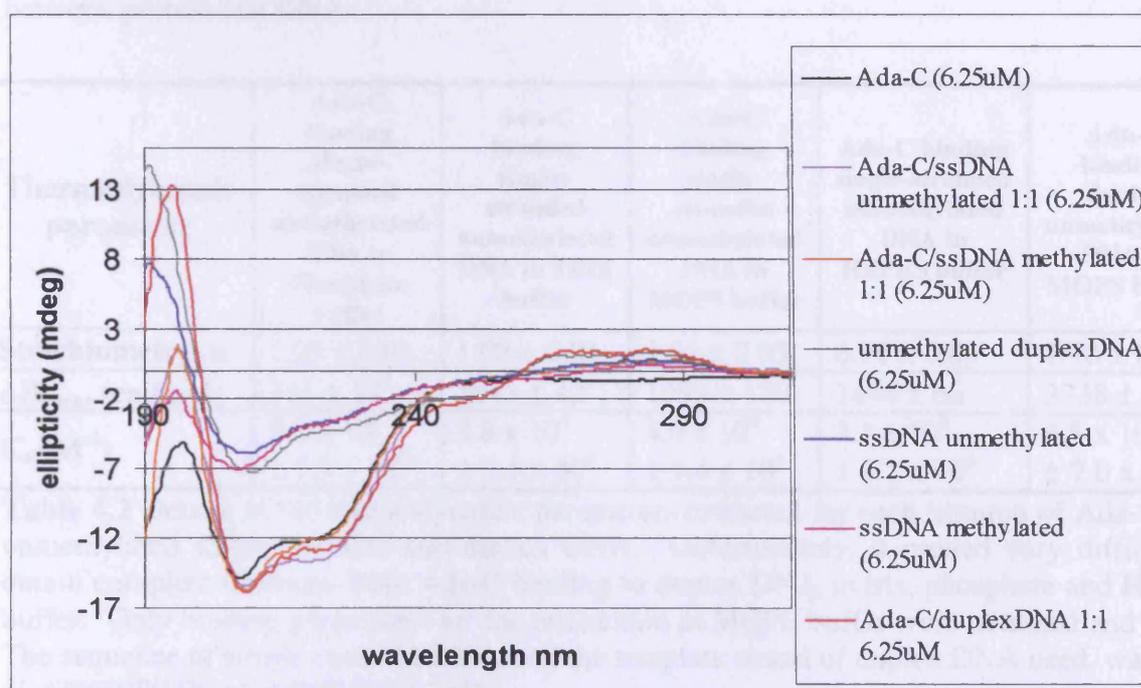


Figure 4.1 Excitating at 295nm proves the inner filter effect of DNA. The fluorescence signal of Trp decreases with similar magnitude to that of the Ada-C when single-stranded, unmethylated DNA is added. The DNA sequence used was: 5'-GTGGCAACCTGTGGCAACCT. Alterations to the signal intensity cannot therefore be attributed to binding.

### **4.2.2 Circular dichroism**

Near uv spectra (250-300nm) reveal information on the aromatic residues in proteins and DNA. In the case of Ada-C negative molar ellipticity was observed, whilst DNA had positive ellipticity in the same wavelength range (Figure 4.2). The far uv (170-250nm) spectral

regions revealed virtually identical spectra for Ada-C free, and bound to methylated single-stranded DNA, duplex unmethylated DNA, and unmethylated single-stranded DNA. The 1:1 complexes of Ada-C:DNA are shown in Figure 4.2 along with spectra of free protein and DNA. Further increases in DNA concentration to 2 and 3 fold excess over Ada-C left the spectrum unaltered (for clarity, these are not plotted).



**Figure 4.2** The CD spectra of free and complexed Ada-C and DNA (ellipticity is shown in millidegrees, and represents  $\theta_d$ ). Only 1:1 complexes are shown for clarity (the 1:2 and 1:3 Ada-C/DNA complex spectra were similar to those for the 1:1 complex). Although scans from 190 to 450nm were performed only 190 to 350nm is shown here since at higher wavelengths than 350nm all spectra are essentially baseline. ssDNA refers to single-stranded DNA, concentrations in  $\mu\text{M}$  are given in parenthesis. The sequence of single-stranded DNA used was 5'-AGGTTGCCACA(O<sup>6</sup>meG)GTTGCCAC, where O<sup>6</sup>meG is O<sup>6</sup>-methylguanine in methylated DNA, and guanine in unmethylated DNA. Duplex DNA was made by adding the complementary strand to the unmethylated sequence shown.

#### **4.2.3 Thermodynamics of binding interactions by ITC**

The control titrations of Ada-C into buffer, and reverse control of buffer into DNA revealed no heat of dilutions and no deviation from baseline. One such titration is shown in Figure 4.3 adjacent to the actual titration of Ada-C into DNA in phosphate buffer. The titration into Tris buffer is shown in Figure 4.4. This contains 40 injections and is the result of concatenating two 20 injection titrations, as described in Section 2.2.2. The thermodynamic parameters extracted from the titrations of wild type Ada-C into single- and double-stranded unmethylated DNA are given in Table 4.2. Taking into account the error ranges, the median  $K_a$  value obtained with MOPS, HEPES and Tris buffers lay between  $3.9 \times 10^5 \text{ M}^{-1}$  and  $4.0 \times$

$10^5 \text{ M}^{-1}$ . For simplicity the average of these two values,  $3.95 \times 10^5 \text{ M}^{-1}$  is taken as the  $K_a$  for binding single-stranded DNA in these buffers. Phosphate buffer caused an increase in  $K_a$  to  $9.0 \times 10^5$ . Reasons for this increased affinity between Ada-C and single-stranded DNA in phosphate buffer are unknown. Empirically, one would have expected the reverse trend, that phosphate weakens affinity through binding-site competition with the Phosphate backbone of DNA. Perhaps the phosphate ions allow bridged hydrogen bonds or salt links to form between protein and DNA.

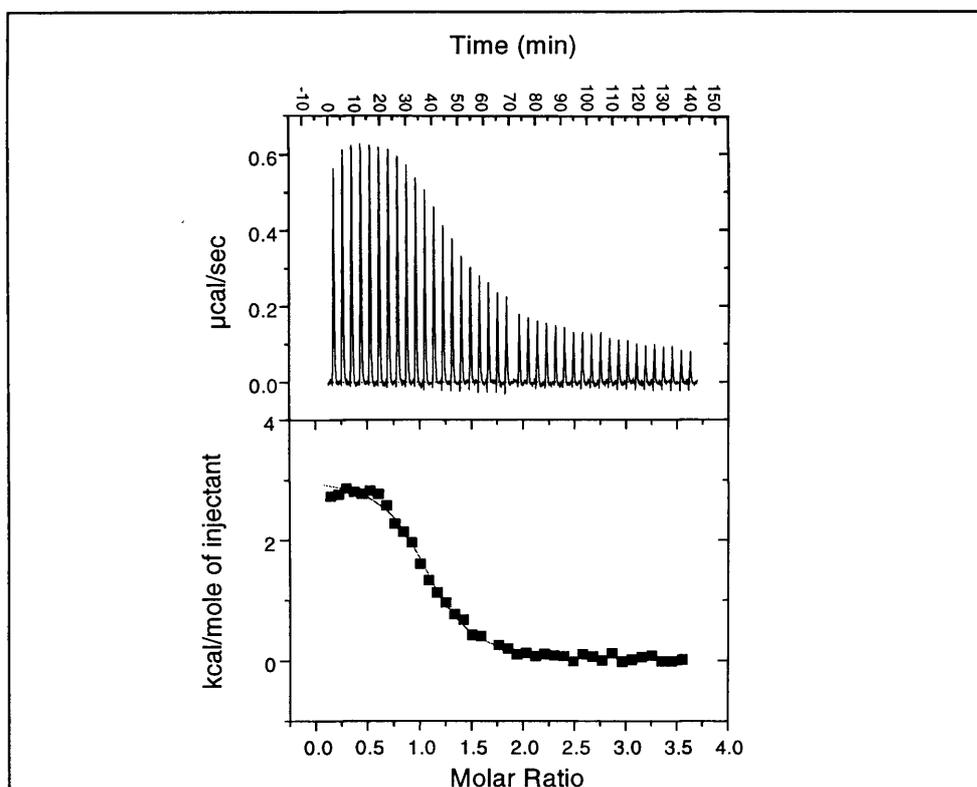
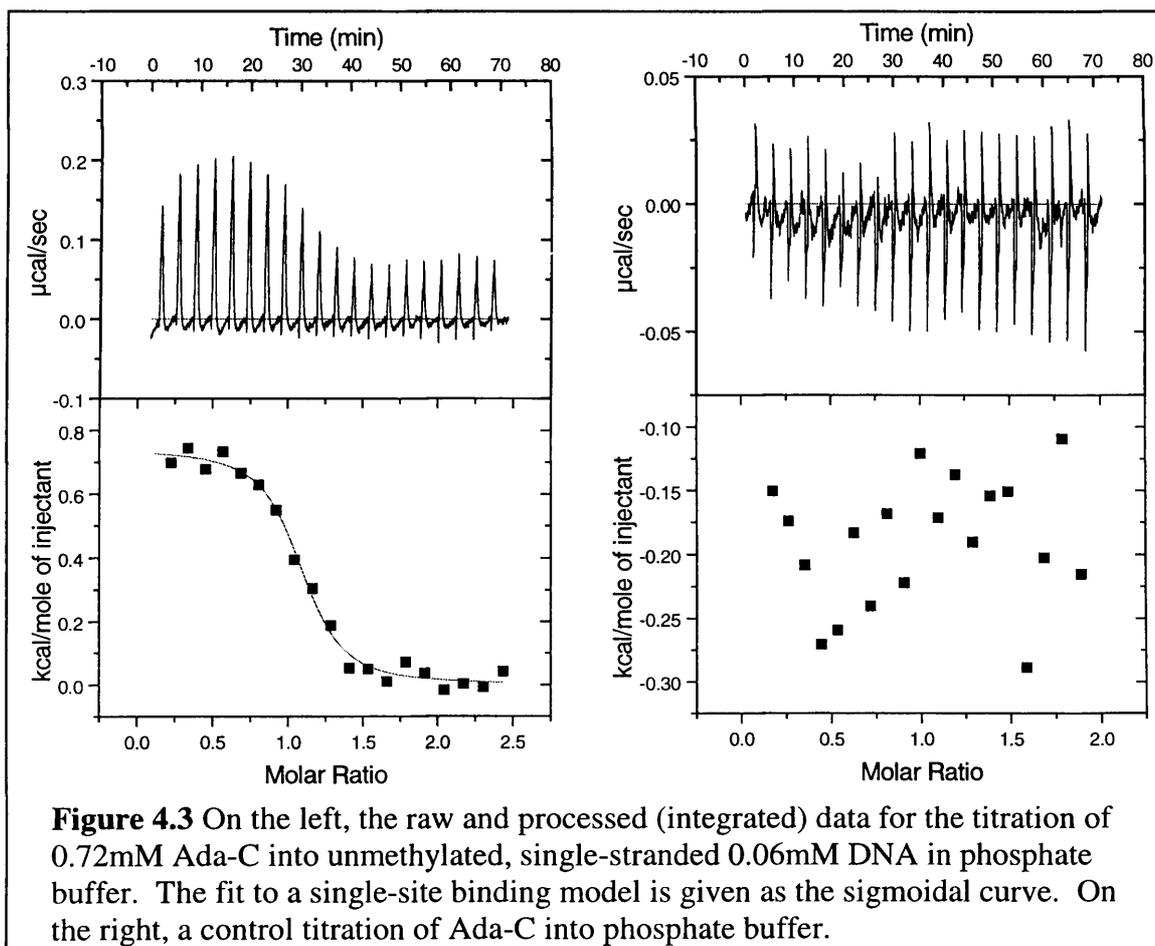
Thermodynamic parameter	Ada-C binding single-stranded unmethylated DNA in Phosphate buffer	Ada-C binding single-stranded unmethylated DNA in TRIS buffer	Ada-C binding single-stranded unmethylated DNA in MOPS buffer	Ada-C binding single-stranded unmethylated DNA in HEPES buffer	Ada-C binding duplex unmethylated DNA in MOPS buffer
<b>Stoichiometry, n</b>	$1.05 \pm 0.01$	$1.08 \pm 0.01$	$1.01 \pm 0.05$	$0.99 \pm 0.03$	$1.98 \pm 0.03$
<b><math>\Delta H_{\text{meas}}</math> (cal/mol)</b>	$741 \pm 17$	$3043 \pm 42$	$1999 \pm 120$	$1494 \pm 66$	$3738 \pm 67$
<b><math>K_a</math> (<math>\text{M}^{-1}</math>)</b>	$9.0 \times 10^5 \pm 1.9 \times 10^4$	$3.8 \times 10^5 \pm 3.5 \times 10^4$	$4.0 \times 10^5 \pm 1.4 \times 10^4$	$3.1 \times 10^5 \pm 9.1 \times 10^4$	$4.8 \times 10^5 \pm 7.0 \times 10^4$

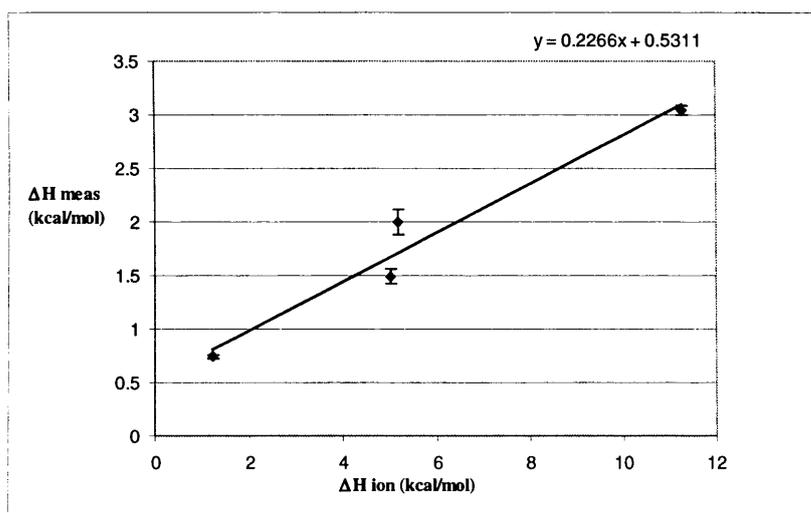
**Table 4.2** Details of the thermodynamic parameters extracted for each titration of Ada-C into unmethylated single-stranded and duplex DNA. Unfortunately, it proved very difficult to obtain complete titrations from Ada-C binding to duplex DNA in tris, phosphate and HEPES buffer. Only binding parameters for the interaction in MOPS buffer were obtained and fitted. The sequence of single-stranded DNA, and the template strand of duplex DNA used, was: 5'-AGGTTGCCACAGGTTGCCAC.

As previously explained (Section 4.1.2) titrating the same reactants in different buffers produces a different measured enthalpy ( $\Delta H_{\text{meas}}$ ). This is due to the ionisation of the buffer itself producing a measured heat change. Fortunately, the true enthalpy of complexation ( $\Delta H_{\text{bind}}$ ) can be obtained along with information concerning the number of protons given off or taken up by the complex, from a plot of  $\Delta H_{\text{ion}}$  against  $\Delta H_{\text{meas}}$  values. The values for  $\Delta H_{\text{ion}}$  were obtained from Microcal Inc. and are listed in Table 4.3 together with the  $\Delta H_{\text{meas}}$  values plotted against them in Figure 4.5.

Buffer used	$\Delta H_{\text{ion}}$ (kcal/mol)	$\Delta H_{\text{meas}}$ (kcal/mol)
TRIS	11.3	3.043
MOPS	5.2	1.999
HEPES	5.0	1.494
Phosphate	1.2	0.741

**Table 4.3** Showing the measured  $\Delta H_{\text{meas}}$  together with corresponding  $\Delta H_{\text{ion}}$  for the buffer type utilised for the binding interaction between Ada-C and single-stranded unmethylated DNA.





**Figure 4.5** The plot of  $\Delta H_{meas}$  vs.  $\Delta H_{ion}$ , the equation of the line is shown in the top right hand corner. This is related to  $\Delta H_{meas} = \Delta H_{bind} + N_H^+ \Delta H_{ion}$ . The slope of the line indicates 0.23 protons are taken up by the complex, whilst the intercept reveals the heat of binding,  $\Delta H_{bind}$  to be 531 cal/mol upon complexation. Vertical error bars are depicted on the points.

From the  $K_a$  determined by the ITC titration and the  $\Delta H_{bind}$  value obtained from Figure 4.5 it is possible to derive the other thermodynamic parameters  $T\Delta S$ , and Gibb's free energy  $\Delta G$  in the following manner:

$$R = 8.314 \text{ J K}^{-1} \text{ mol}^{-1}$$

$$T = 298 \text{ K}$$

$$K_a = 3.95 \times 10^5 \text{ M}^{-1} \text{ for single-stranded DNA (this is an average value obtained as described in the text above)}$$

By inputting these values into the following equation,  $\Delta G = -RT \ln K_a$ ,  $\Delta G$  is obtained:

$$\begin{aligned} \Delta G &= -8.314 \times 298 \times \ln 3.95 \times 10^5 \\ \Delta G &= -31927.6 \text{ J/mol} \\ &= -7.6 \text{ kcal/mol (1cal} = 4.184\text{J)} \end{aligned}$$

$\Delta H$  is known to be 531 cal/mol (obtained from the y-intercept in Figure 4.5) this converts to:

$$\Delta H = 2221.7 \text{ J/mol}$$

Using the equation  $\Delta G = \Delta H - T\Delta S$ , and the values of  $\Delta G$ ,  $\Delta H$  and temperature,  $\Delta S$  can be obtained:

$$-31927.6 = 2221.7 - 298 \times \Delta S$$

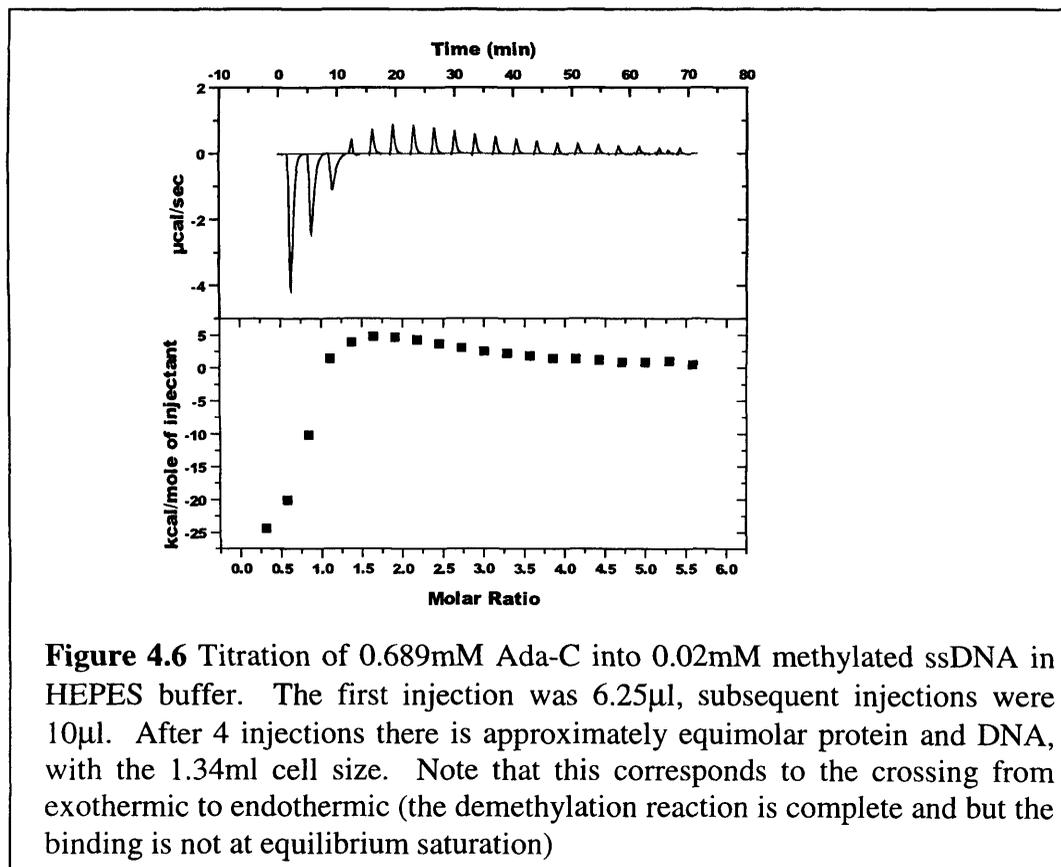
$\Delta S = 114.6 \text{ J mol}^{-1} \text{ K}^{-1}$  This can then be used to obtain the frequently quoted thermodynamic parameter  $T\Delta S$ . A greater value of  $T\Delta S$  relative to  $\Delta H$  (ignoring the sign), indicates a greater reliance on entropy, than enthalpy, as the driving force for the binding reaction.

$$T\Delta S = +8.2 \text{ kcal/mol}$$

The majority of DNA-binding proteins exhibit enthalpy changes of greater magnitude than +0.53 kcal/mol observed for Ada-C binding DNA. Most proteins use an exothermic enthalpy change as the driving force for binding DNA, one such example is the yeast transcription factor GCN-4 when binding the AP-1 site for which an enthalpy change of  $-36.8$  kcal/mol was measured ( $T\Delta S = -26.6$  kcal/mol, Berger *et al.*, (1996)). As the enthalpy change gets more endothermic the entropic contribution to  $\Delta G_{\text{binding}}$  increases. An example of this is the Trp repressor binding its cognate DNA sequence, where  $\Delta H = -1.6$  kcal/mol,  $T\Delta S = +11.8$  kcal/mol (Jin *et al.*, 1993). Glucocorticoid receptor exhibits endothermic binding ( $\Delta H = +5.5$  kcal/mol) and the entropic contribution to the binding driving force is again higher ( $T\Delta S = +21.7$  kcal/mol, Lundback *et al.*, (1996)).

#### 4.2.4 Thermodynamics of DNA demethylation

The demethylation reaction between Ada-C and methylated single-stranded DNA proved to be highly exothermic. A titration of Ada-C into this DNA revealed how the demethylation reaction proceeded to completion and then how the Ada-C continued to bind to saturation on the DNA. The point at which the demethylation reaction is complete is marked by the transition to endothermic heat change (Figure 4.6).

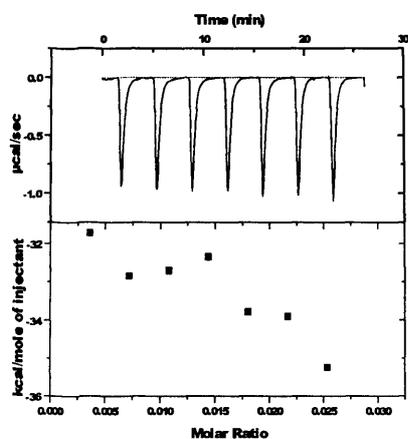


**Figure 4.6** Titration of 0.689mM Ada-C into 0.02mM methylated ssDNA in HEPES buffer. The first injection was 6.25µl, subsequent injections were 10µl. After 4 injections there is approximately equimolar protein and DNA, with the 1.34ml cell size. Note that this corresponds to the crossing from exothermic to endothermic (the demethylation reaction is complete and but the binding is not at equilibrium saturation)

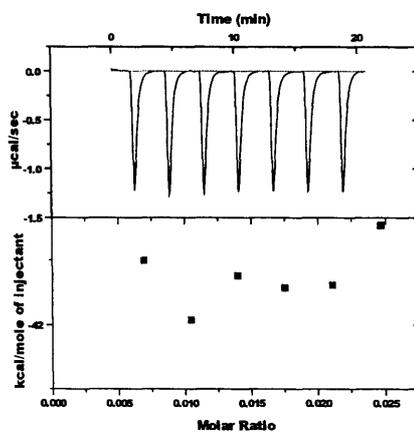
To analyse the exothermic demethylation reaction in more detail, methylated ssDNA at a high concentration was introduced into the cell. Ada-C was injected at a low concentration and low volume, ensuring that the methylated DNA would be in vast excess (~300 fold) to the protein in the cell. The integral value of the exothermic heat spike was almost identical for each successive injection. Since demethylation is a unidirectional reaction, and no equilibrium is set up (ignoring the binding equilibria), a single heat spike represents the heat change occurring when the amount of Ada-C injected reacts with methylated DNA. Integration with respect to time, and correction for the amount of Ada-C involved, should therefore render the magnitude of the heat change per mole of Ada-C reacted. An average of these integrated heat values was taken to be the heat involved in demethylation of single-stranded DNA in the particular buffer. The data for these experiments is shown in Figure 4.7, and the results in Table 4.4 below.

<b>Buffer</b>	<b>Heat of reaction for demethylation of ssDNA (kcal/mol)</b>
Phosphate	-41
Tris	-33
MOPS	-25
HEPES	-29

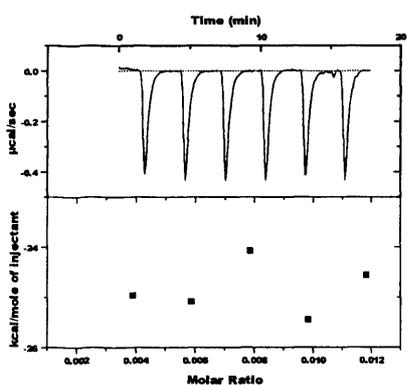
**Table 4.4** The heats of demethylation of methylated ssDNA by Ada-C in four buffers. These  $\Delta H$  values were obtained by integrating the area defined by the exothermic peaks resulting from additions of small amounts of Ada-C into 300 fold excess of methylated single-stranded DNA (BMT). The values obtained were then averaged to arrive at the figures shown above.



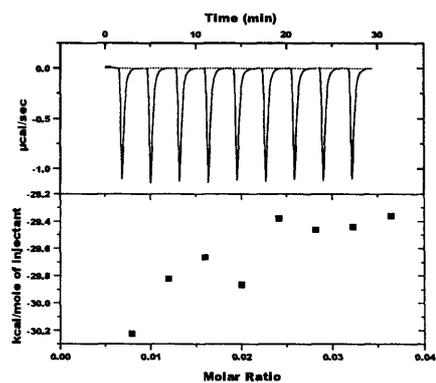
(a)



(b)



(c)



(d)

**Figure 4.7** Assessing the thermodynamics of demethylation of 300 fold excess single-stranded, methylated DNA (BMT) by Ada-C in (a) tris buffer, (b) phosphate buffer, (c) MOPS buffer and (d) HEPES buffer. Both raw and integrated data are shown. An average value of  $\Delta H$  was taken for each buffer, this is shown in Table 4.4.

## **4.3 DISCUSSION**

### **4.3.1 Fluorescence and circular dichroism analysis of DNA binding**

No changes indicative of binding were observed when anisotropic fluorescence studies were carried out using either the intrinsic protein fluorescence or fluorescence of a fluorescein tag on the 5' end of DNA. The observed decrease in signal intensity of the protein fluorescence upon addition of DNA also occurred to the signal of free Trp when DNA was added. This signal intensity decrease was attributed to the inner filter effect of DNA. (This is where DNA absorbs in the excitation wavelength region thus diminishing the intensity of excitation). There are three possibilities to account for the unaltered fluorescence signal of the protein upon binding, since NMR indicated that Ada-C does bind DNA under the conditions used (Chapter 6):

- 1) The environments of the Ada-C Trp and Tyr residues do not change as a result of binding, possibly because they are far from the binding site.
- 2) There is a change in the environment of one or more of these residues but it is too small to give a significantly different fluorescence signal and is not observed.
- 3) There is a change in one or more environments and this creates an altered fluorescence signal. However the complex interplay between absorbing and emitting groups means this altered emission signal goes unobserved.

In light of NMR evidence (Chapter 6), which revealed that no conformational change occurred in Ada-C upon binding DNA, it was plausible that no significant alteration in environment of Tyr or Trp residues occurred to cause an isotropic fluorescence change. The fluorescein-tagged DNA also revealed no alteration in the environment of the tag upon addition of Ada-C. This may be because binding could not occur with the presence of this large fluorescing group attached to the end of the oligonucleotide. This was considered unlikely, as the DNA is sufficient in length to accommodate Ada-C far from its tagged end. At the end of the DNA however, the fluorescein group might remain in an identical environment although protein had bound to the rest of the DNA molecule. This is a more likely explanation for the unaltered DNA fluorescein signal when free and complexed to Ada-C.

Circular dichroism analysis revealed qualitatively that the secondary structure of Ada-C did not alter dramatically upon binding unmethylated duplex, or ssDNA, or methylated ssDNA. This was also the case with human alkyltransferase where similar CD spectra were obtained for the protein free, and bound to methylated duplex DNA (Chan *et al.*, 1993).

### **4.3.2 ITC analysis of Ada-C binding unmethylated single-stranded DNA**

Isothermal titration calorimetry proved to be the most suitable technique for determining the thermodynamic parameters for the interaction of Ada-C with unmethylated single-stranded 20-mer DNA. The observed monomeric binding to single-stranded DNA agrees well with the findings of others for Ada (Bhattacharyya *et al.*, 1990), (Takahashi *et al.*, 1990). No evidence of cooperative binding was observed, contrasting to the recently proposed cooperative interactions of human alkyltransferase on single-stranded DNA (Fried *et al.*, 1996). The average  $K_a$  value for binding single-stranded DNA was  $3.95 \times 10^5 \text{ M}^{-1}$  in tris, MOPS or HEPES buffer. However, binding to single-stranded DNA in phosphate buffer occurred with a higher  $K_a$  than when other buffers were used. Exact reasons for this are unknown. Certainly, less importance is attributed to knowledge of the exact binding constant, rather than the range it lies in. Binding constants will differ depending on buffer conditions. Buffer ionic strength is likely to play the most significant role in dictating the binding constant. Although here buffers of similar ionic strength were employed, the buffering substances themselves may also play a role in interfering with, or enhancing binding. For instance utilisation of phosphate in the buffer may result in binding of positively charged protein residues to phosphate ions, and their consequent displacement upon interaction with DNA. This may not occur with tris buffer, thus resulting in slightly differing binding constants observed between these two buffers.

The fact that  $\Delta H_{\text{meas}}$  significantly differed between the four buffers indicated that some buffer ionisation may be occurring during the binding process. Use of the relationship  $\Delta H_{\text{meas}} = \Delta H_{\text{bind}} + N_{\text{H}^+} \Delta H_{\text{ion}}$  revealed 0.23 protons to be taken up by the complex. This partial protonation of the complex can be explained in terms of something lying near its pKa at the experimental pH. Some individual complexes get protonated, others do not. The net result is the fractional protonation observed.

### **4.3.3 ITC analysis of Ada-C binding unmethylated duplex DNA**

Unfortunately, only data for Ada-C binding 20-mer duplex DNA in MOPS buffer was obtained. Titrations in the other three buffer systems were of insufficient quality to fit the data correctly. Reasons for this are unknown, but may have something to do with the self-aggregating nature of Ada-C (Chapter 5).

The stoichiometry for binding duplex DNA in MOPS buffer was 2 Ada-C : 1 duplex DNA, whilst the binding energy measured ( $\Delta H_{\text{meas}}$ ) was approximately twice that for binding single-stranded DNA. This suggests that 2 molecules of Ada-C bind a single molecule of 20-mer duplex DNA, with a similar binding constant to binding single-stranded DNA (the  $K_a$  for this

interaction was  $4.8 \times 10^5 \pm 7.0 \times 10^4 \text{ M}^{-1}$ ). Such an increased Ada-C binding density on duplex relative to single-stranded DNA suggests that Ada-C binds predominantly to a single DNA strand when binding duplex DNA. Polyacrylamide gel electrophoresis of single-stranded and duplex DNA suggested no lengths of duplex structure longer than 2-3 base pairs occurred in single-stranded DNA. This makes binding to regions of duplex DNA in single-stranded DNA unlikely. NMR evidence suggests that binding single-stranded and duplex unmethylated DNA occurs in an identical manner with respect to Ada-C (Chapter 6). The single-stranded DNA, being flexible, should be able to accommodate a structure similar to that of duplex DNA, enabling it to bind Ada-C in a similar manner.

Attempts were made to demonstrate the stoichiometry of Ada-C bound to duplex DNA using gel filtration. However, the results did not reveal anything with a mass indicative of 2 molecules of Ada-C and one molecule of duplex 20-mer. This may simply be because the complex dissociated during the length of the gel filtration run. Assessment of multimeric states by gel filtration is usually carried out on proteins whose affinities for one another are higher than the recorded affinity of Ada-C for duplex 20-mer.

#### **4.3.4 General considerations of Ada-C binding DNA**

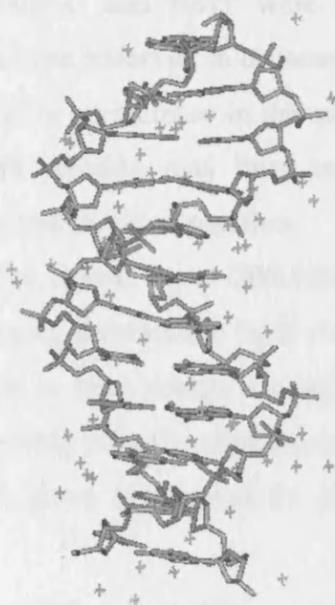
The  $K_a$  values detailed here are consistent with those obtained for full length Ada and human alkyltransferases with various lengths of DNA, (Takahashi *et al.*, 1990). To my knowledge these results are the first example of binding enthalpies obtained for an alkyltransferase repair enzyme binding to DNA. The binding of Ada-C to unmethylated single- or double-stranded DNA occurs endothermically ( $\Delta H = +0.53\text{kcal/mol}$ ), and must therefore be entropically driven ( $T\Delta S = +8.2\text{kcal/mol}$ ).

Binding of the majority of proteins to DNA occurs exothermically (Berger *et al.*, 1996); (Jin *et al.*, 1993); (Merabet *et al.*, 1995); (Lohman *et al.*, 1996). However, other endothermically-binding systems have been defined. These include glucocorticoid receptor binding specifically to its cognate DNA sequence (Lundback *et al.*, 1996), and MetJ repressor binding non-specifically to random DNA (Hyre *et al.*, 1995).

It is well established that stabilities and specificities of protein nucleic acid interactions, just as for all macromolecular interactions, are sensitive to the solution conditions such as salt type, concentration and pH. As nucleic acids are effectively highly charged polyanions, a sequestering of cations such as  $\text{K}^+$ ,  $\text{Mg}^{2+}$ ,  $\text{Na}^+$  occurs along the sugar-phosphate backbone. Evidence of this is available from the mass spectrometry results obtained in this work. This clearly demonstrated  $\text{K}^+$  ions bound to the DNA. Disturbances in this linear charge density caused by protein binding to the nucleic acid will result in cation release from the backbone. As a direct consequence of this release, any increase in cation concentration in the bulk

solution will lower the binding constant for most protein-nucleic acid interactions, (for review see (Lohman *et al.*, 1992) and references therein). This weakening of binding is both through loss of favourable entropic and enthalpic interactions, (Lohman *et al.*, 1996); (Lohman *et al.*, 1992); (Mascotti *et al.*, 1997); (Mascotti *et al.*, 1993); (Kozlov *et al.*, 1998).

As well as ion effects, ordered water molecules around the DNA also play a role in protein-DNA interactions. These ordered water molecules often trace around the major groove of B-form DNA and other DNA forms (Leonard *et al.*, 1990), (Laughlan *et al.*, 1994), Figure 4.8). These may be released into bulk solution upon protein binding, thus causing a favourable entropy factor.



**Figure 4.8** The structure of the 12-mer duplex of DNA containing two O<sup>6</sup>-methylguanine lesions by Leonard *et al.* (1990). Ordered water molecules run along the major groove and the phosphate backbone. These could be displaced to bulk solution upon binding Ada-C, thus lending an entropic contribution to the binding driving force.

The fact that Ada-C aggregates readily in the absence of salt suggests that it may require ions to shield electrostatic interactions. These ions bind the protein and are released upon interaction with another protein or DNA. The potential of Ada-C to bind anions so readily would provide an entropic contribution to binding DNA through release of such ions upon complexation. The loss of anions upon DNA binding has been noted previously with full length Ada, (Takahashi *et al.*, 1990).

When Ada-C binds DNA, possible enthalpic contributions come from:

- 1) Salt links formed between Arg sidechains and the phosphate backbone.
- 2) Stacking interactions with bases and hydrophobic groups from the protein.
- 3) Van Der Waals and H-bonds that may form.

However, the driving force for binding is entropic. Favourable entropy factors which may occur to aid binding are:

- 1) Loss of anions from the protein upon binding
- 2) Loss of any ordered cations from the phosphate backbone
- 3) Loss of any ordered waters from the major groove of DNA or the protein binding interface.

#### **4.3.5 Energetics of *in vitro* DNA demethylation by Ada-C**

The demethylation of single-stranded methylated DNA (BMT) *in vitro* was exothermic. When similar amounts of Ada-C and BMT were reacted together in the ITC cell the magnitude of this heat release was observed to decrease with each injection of protein (Figure 4.6). This could be explained by an increase in the amount of unmethylated DNA after each Ada-C addition, to which the protein may bind endothermically, thus counteracting the exothermicity of the overall demethylation reaction.

Additions of small amounts of Ada-C to a ~300 fold excess of methylated single-stranded DNA resulted in an unchanging exothermic heat change (Figure 4.7). This heat change represents the overall change in heat energy caused by Ada-C binding methylated DNA, demethylating it and then possibly Ada-C unfolding (a discussion on the instability of Ada-C following self-methylation is given in Chapter 6), plus any buffer effects. The measured  $\Delta H_{\text{measured}}$  might equate to:

$$\Delta H_{\text{measured}} = \Delta H_{\text{binding}} + \Delta H_{\text{ion}} + \Delta H_{\text{unfolding}} + \Delta H_{\text{demethylation}}$$

where  $\Delta H_{\text{binding}}$  represents the heat associated with Ada-C binding methylated DNA,  $\Delta H_{\text{ion}}$  represents any heat changes due to the buffer ionisation, and  $\Delta H_{\text{unfolding}}$  represents any heat change associated with the reduced stability of methylated Ada-C.

It is possible that the binding reaction of Ada-C to methylated DNA is itself exothermic. An exothermic heat change might be associated with base flipping, which may only occur with methylated DNA. It seems unlikely however that the heat change should be so exothermic based on binding and flipping alone. It is more probable that the large exothermic factor observed in the heat change upon interaction between methylated DNA and Ada-C emanates from the methyl transfer reaction itself.

The degree of heat change observed was buffer dependent, with phosphate buffer showing the largest exothermic change of  $-41$  kcal/mol (Table 4.4). These heats of reaction did not correlate simply with the  $\Delta H_{\text{ion}}$  for the buffers used, as they did for the  $\Delta H_{\text{meas}}$  (Figure 4.5). It is possible that the different salt concentrations in the buffers, as well as the buffering substances themselves altered the structure and degree of base stacking of single-stranded

DNA in solution. The degree of base stacking (and defined structure) may influence how easily the DNA binds Ada-C and perhaps also the methyl transfer chemistry itself. The reaction system in question is very complex, there are too many variables to draw any conclusions concerning the cause for the discrepancy in demethylation heat change observed with different buffers.

#### **4.4 SUMMARY**

Fluorescence changes were not detected upon DNA binding. CD analysis suggests no gross change in secondary structure upon binding single-stranded, duplex or methylated DNA at low protein concentrations.

The  $K_a$  for binding Ada-C to single-stranded 20-mer DNA is  $4 \times 10^5 \text{ M}^{-1}$ , this binding constant is very similar to those determined for the whole protein and for the human form, (Chan *et al.*, 1993), (Takahashi *et al.*, 1990), (Bender *et al.*, 1996) Binding occurs with an endothermic  $\Delta H$  and is therefore entropically driven, probably by loss of ordered water and bound ions to the bulk solution. No protons are taken up or given up to bulk solvent upon complex formation.

*In vitro* demethylation of single-stranded methylated DNA occurred with an exothermic energy change, the magnitude of which was dependent on the buffer used.

# Chapter 5

## *Assigning the backbone of Ada-C*

### 5.1 INTRODUCTION TO NMR

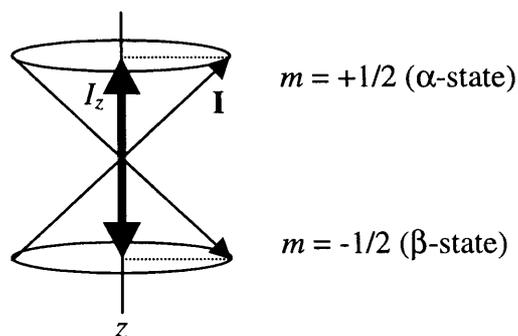
This section is devoted to analysis of the end result of an NMR experiment, with relevance to protein structure and dynamics.

#### 5.1.1 Exciting and relaxing spins

The theory of NMR spectroscopy is based on the quantum mechanics of nuclear spin angular momentum. *Spin*, is a purely quantum mechanical property and does not have a classical analog. Spin angular momentum is characterised by the nuclear spin quantum number,  $I$ . Only nuclei with non-zero  $I$  values are NMR-active. Nuclei with integral values of  $I$ , such as deuterium, usually have shorter lifetimes in magnetic states than  $I = 1/2$  nuclei, giving rise to correspondingly broader lines and are therefore more difficult to study. NMR studies usually make use of  $I = 1/2$  nuclei, of which  $^{13}\text{C}$ ,  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{19}\text{F}$  and  $^{31}\text{P}$  are the most relevant to biomolecular NMR. The nuclear spin angular momentum,  $\mathbf{I}$ , is a vector quantity whose magnitude is given by:  $|\mathbf{I}^2| = \hbar^2 [I(I+1)]$ , where  $\hbar$  is Planck's constant divided by  $2\pi$ . Due to quantum mechanical restrictions, only one of the three Cartesian components of  $\mathbf{I}$  can be specified with  $\mathbf{I}^2$ . The components along the other axes are uncertain. By convention the  $z$  axis lies parallel to the external magnetic field,  $B_0$ . The  $z$  component of  $\mathbf{I}^2$  is specified by the equation  $I_z = \hbar m$ , where  $m$  is the magnetic quantum number which can take any of the values  $m = I, (I-1), (I-2) \dots -I$ . So  $I_z$  has  $2I+1$  possible values, eg. for  $I = 1/2$  nuclei there are two values. Because the magnitude of the angular momentum is constant, and the  $z$  component has two discrete values (for  $I = 1/2$  nuclei), it is the orientation of the spin angular momentum vector which is quantised. For an isolated spin in the absence of an external magnetic field, the two quantum states defined by  $m = +1/2$  and  $m = -1/2$  have the same energy, and the spin angular momentum vector has no preferred orientation.

Nuclei with nonzero spin angular momentum also have a nuclear magnetic moment whose vector is colinear with the nuclear spin angular momentum vector  $\mathbf{I}$ . This magnetic moment,  $\boldsymbol{\mu}$  is defined by  $\boldsymbol{\mu} = \gamma \mathbf{I}$ ; where  $\mu_z = \gamma I_z = \gamma \hbar m$ , and the gyromagnetic ratio  $\gamma$  is a constant, which is characteristic of nucleus type. The greater the magnitude of  $\gamma$ , the greater the sensitivity of the nucleus in NMR spectroscopy. Since angular momentum is quantised, so is the nuclear magnetic moment. In the presence of an external magnetic field, the spin states of

the nucleus have energies defined by:  $E = -\mu B_0$ , where  $B_0$  is the magnetic field vector of the external magnetic field. Minimum energy is obtained when the projection of  $\mu$  on to  $B_0$  is maximised (when the two vectors are made as parallel as possible). As  $|\mathbf{I}|^2 > I_z^2$ , and  $\mu$  is always colinear with the  $\mathbf{I}$  vector, it is not possible for  $\mu$  to be colinear with  $B_0$ . The  $m$  spin states become quantised instead with energies proportional to their projections on to  $B_0$ , which by definition lies along the  $z$  axis (Figure 5.1).



**Figure 5.1** The angular momentum vectors of a spin  $\frac{1}{2}$  nucleus. The angular momentum vector  $\mathbf{I}$ , and its projection on to the  $z$  axis,  $I_z$  in the two spin states  $m = +\frac{1}{2}$  and  $-\frac{1}{2}$  which are also referred to as  $\alpha$ - and  $\beta$ -states respectively. At any instant in time the vector  $\mathbf{I}$  could be found somewhere inside the cone of precession about the  $z$  axis. The exact position of this vector on the surface of the cone however cannot be specified because of quantum mechanical uncertainties in the  $I_x$  and  $I_y$  components in Cartesian space.

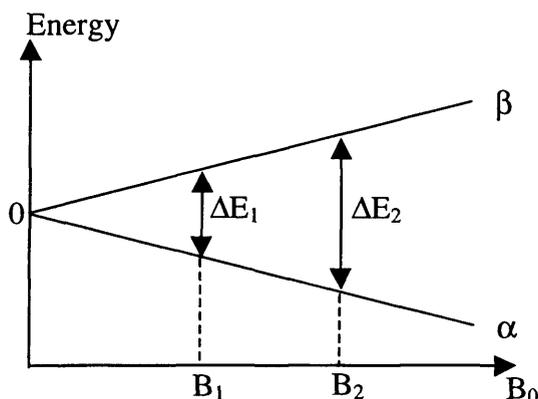
In the presence of an external magnetic field the projection of angular momentum on the  $z$  axis,  $I_z$  can lie parallel or antiparallel to the  $B_0$  field whose vector also lies along  $z$ . The energy level of each state is given by  $E_m = -\gamma I_z B_0 = -\gamma \hbar m B_0$ , where  $B_0$  is the external magnetic field strength.

When the angular momentum vector projection,  $I_z$ , is parallel to the  $B_0$  field, the nucleus is in a lower energy state than when  $I_z$  is antiparallel to  $B_0$ . These energy levels are of very similar energy, and at equilibrium there is only a slight excess of spins in the lower energy  $\alpha$ -state than the higher energy  $\beta$ -state. Indeed in a field strength of 14 Tesla the energy difference between  $\alpha$  and  $\beta$ -states, can be calculated to be  $3.9 \times 10^{-25}$  J for protons. At this energy difference for proton spin states, the relative populations of the spins in the upper ( $N_\beta$ ) and lower ( $N_\alpha$ ) energy states at room temperature is  $N_\beta/N_\alpha = 0.9999$ . Increasing the external magnetic field strength  $B_0$  enhances sensitivity, because it increases the population difference between the two energy levels, as does decreasing the temperature (Figure 5.2).

Transitions between the two energy levels can be stimulated by application of electromagnetic radiation (as with other spectroscopic techniques). The energy needed for such transitions between  $m$  and  $m+1$  states is  $\Delta E = \gamma \hbar B_0$  (the difference between these energy levels).

According to Planck's law ( $\Delta E = h\nu$ ) the frequency of radiation,  $\nu$  needed to cause these

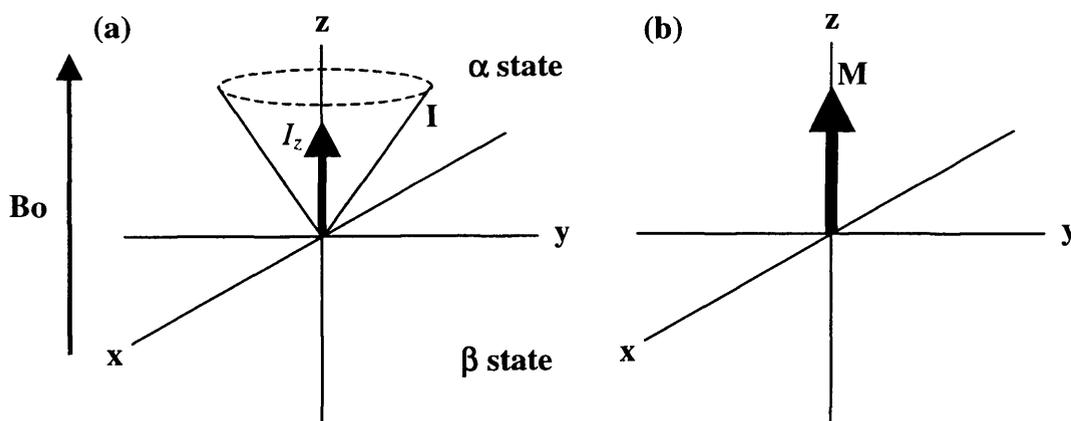
transitions is  $\Delta E = (h/2\pi)\gamma B_0 = h\nu$ , so  $\nu = \gamma B_0/2\pi$  (Hz). This frequency,  $\nu$ , is known as the Larmor frequency.



**Figure 5.2** In the presence of a magnetic field  $B_0$ , a spin  $\frac{1}{2}$  nucleus can exist in two different energy levels,  $\alpha$  and  $\beta$ . As the strength of the external field is increased, the energy difference between these two states is also increased.

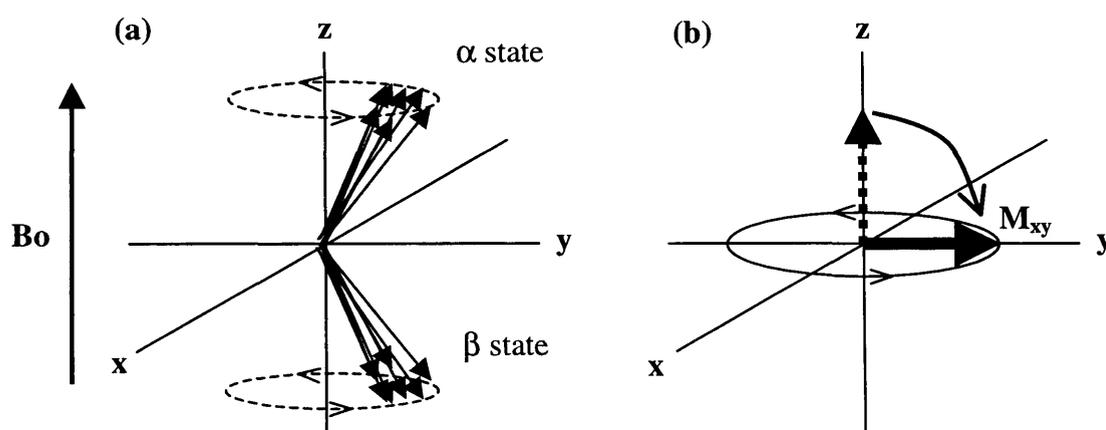
### Generating energy level transitions

A single spin, at equilibrium in the presence of an external magnetic field could lie in the  $\alpha$  or  $\beta$  state; for simplicity, in the following discussion we define it as being in a purely  $\alpha$  state. The angular momentum vector,  $\mathbf{I}$  lies somewhere within the probability cone centred about the  $z$  axis, but due to quantum mechanical uncertainties of the  $I_x$  and  $I_y$  components, the exact position of the vector on this cone remains undefined. Only the  $z$ -axis projection,  $I_z$  is defined to lie along  $+z$ . In terms of bulk magnetisation (the resultant magnetisation vector from all identical spins in the sample), there is a slight excess of spins in the lower energy  $\alpha$  state than the  $\beta$  state, hence a net magnetisation along the  $+z$  axis (Figure 5.3).



**Figure 5.3** (a) Magnetic moment vector  $\mathbf{I}$  of a spin in a purely  $\alpha$  state is allowed anywhere on the surface of the cone. Only the projection on the  $z$  axis,  $I_z$  is of defined magnitude. (b) In a population of identical spins the resultant bulk magnetisation vector  $\mathbf{M}$  also lies along the  $+z$  axis.

When an electromagnetic pulse at the correct frequency,  $\nu$  is applied to the spin, energy is absorbed and a transition to the  $\beta$  state can be induced. (Similarly, were the spin initially in the  $\beta$  state an emission of energy to the final  $\alpha$  state would be induced). Prior to measurement, quantum mechanics dictates that this spin now exists simultaneously in both the  $\alpha$  and  $\beta$  states. The  $\alpha$  and  $\beta$  states have individual wavefunctions, which during this period of “duality” interfere with each other in a manner dependent on the energy difference between the  $\alpha$  and  $\beta$  states. The resultant wavefunction acts as an oscillating vector precessing about the  $z$  axis at the Larmor frequency of the spin. In a classical model one might envisage the angular momentum vectors of all identical spins vectors precessing around the  $z$  axis in phase (tracing the probability cones as before). The bulk magnetisation vector therefore also precesses about  $z$  at the Larmor frequency (Figure 5.4). Exactly where the



**Figure 5.4** (a) Following an RF pulse at the Larmor frequency, individual spins are in both  $\alpha$  and  $\beta$  states, the angular momentum vectors for each state precess about the  $z$  axis in phase. (b) Depending on the strength of the RF pulse and the length of time it is applied, the equilibrium bulk magnetisation (dotted arrow) can be tipped different degrees towards the  $xy$  plane. In this case, the resultant bulk magnetisation lies totally in the  $xy$  plane (solid arrow), populations of spins in and  $\alpha$  state are equal with those in the  $\beta$  state (following a  $90^\circ$  pulse). In a more extreme case, the populations could be inverted, such that the majority of spins occupy the  $\beta$  state. In this case bulk magnetisation lies along  $-z$  (after a  $180^\circ$  pulse).

equilibrium net magnetisation vector (originally along  $+z$ ) ends up, depends on the power of the RF pulse applied to the spin. Following a  $90^\circ$  tip angle, the bulk magnetisation lies in the  $xy$  plane (the population of spins in the  $\alpha$  state equals that of spins in the  $\beta$  state). A  $180^\circ$  pulse results in net magnetisation along  $-z$ , here the equilibrium populations of the spins different energy states has been inverted. It is of course possible to generate any tip angle desired, but for the purposes of most NMR experiments the  $90^\circ$  and  $180^\circ$  pulses are the most commonly employed. Only bulk magnetisation in the  $xy$  plane is detected by the spectrometer. Thus, following a certain sequence of pulses to stimulate certain spins, the

magnetisation from those spins must lie in the xy plane for detection. The bulk magnetisation vector precesses around the z axis in the xy plane at the Larmor frequency. It is this frequency which is measured in an NMR experiment.

### **5.1.2 Chemical shift**

The Larmor frequency of a spin is dependent on the local magnetic field strength experienced. The net magnetic field at the location of the nucleus depends on  $B_0$ , and the induced secondary fields (which can either enhance or oppose the main field, and are caused by the motions of electrons). This net local magnetic field strength is defined by  $(1-\sigma) B_0$ , where  $\sigma$  is the nuclear shielding factor (determined by the secondary magnetic fields). Thus the equation relating Larmor frequency to local magnetic field strength,  $\nu = \gamma B_0/2\pi$  can be rewritten as  $\nu = (1-\sigma) \gamma B_0/2\pi$ . These differences in resonance frequencies, referred to as *chemical shifts*, allow the possibility to distinguish between nuclei in different chemical environments. The chemical shift is measured in terms of  $\delta$ , which is defined as:

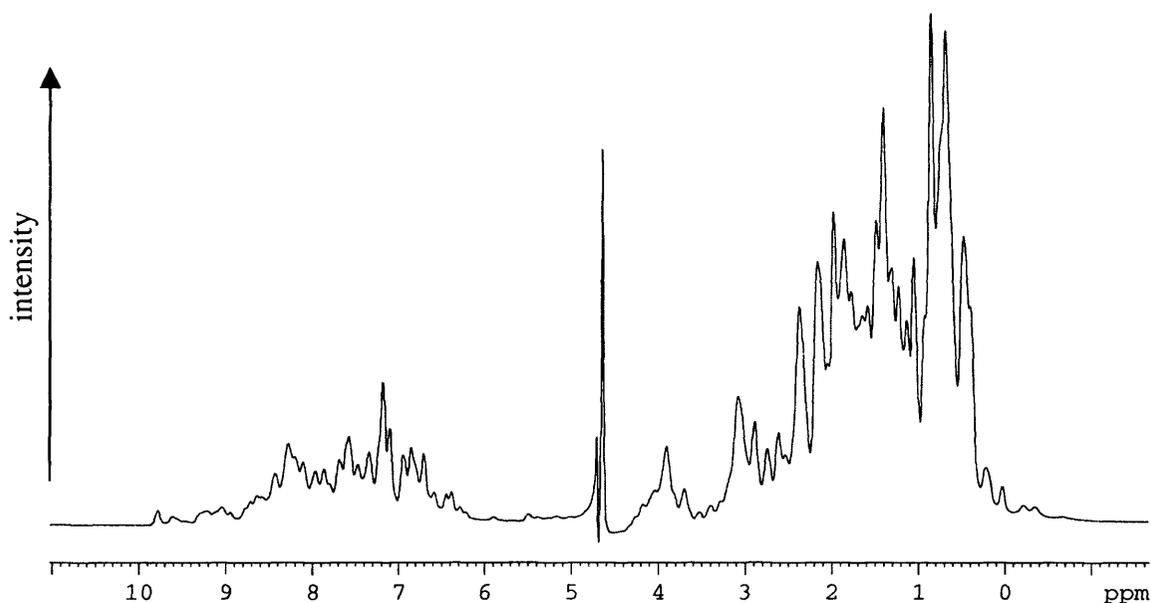
$$\delta = \frac{\Omega - \Omega_{ref}}{\omega_0} \times 10^6 = (\sigma_{ref} - \sigma) \times 10^6$$
 where  $\Omega$  and  $\Omega_{ref}$  are the frequencies of the signal of

interest and the reference signal respectively, and  $\omega_0$  is the Larmor frequency at field strength  $B_0$  in the absence of shielding (the operating frequency of the spectrometer).

Because  $\sigma$  represents a very small fraction of the external magnetic field, it is expressed in parts per million (ppm or  $\delta$ ) relative to a reference resonance signal from a standard molecule (usually 2,2-dimethyl-2-silapentane-5-sulfonic acid, DSS, which is taken to be 0ppm).

It is convenient to characterise signal positions by chemical shift rather than absolute frequency, since the former is independent of  $B_0$ .

A one-dimensional NMR spectrum is a plot of chemical shift against the amplitude of a resonance signal (Figure 5.5). The linewidth of an NMR signal peak will never be infinitely small due to the inherent uncertainty in the measurement (from the Uncertainty Principle). However, the phenomenon of relaxation also contributes to linewidth, and is of particular importance to protein NMR.



**Figure 5.5** An example of a one-dimensional NMR spectrum of Ada-C. The chemical shift relative to DSS (0ppm) is shown on the horizontal axis. The intensity of each signal increases along the vertical axis and is in arbitrary units. Many of the resonance signals overlap, and it is not possible to resolve individual proton frequencies.

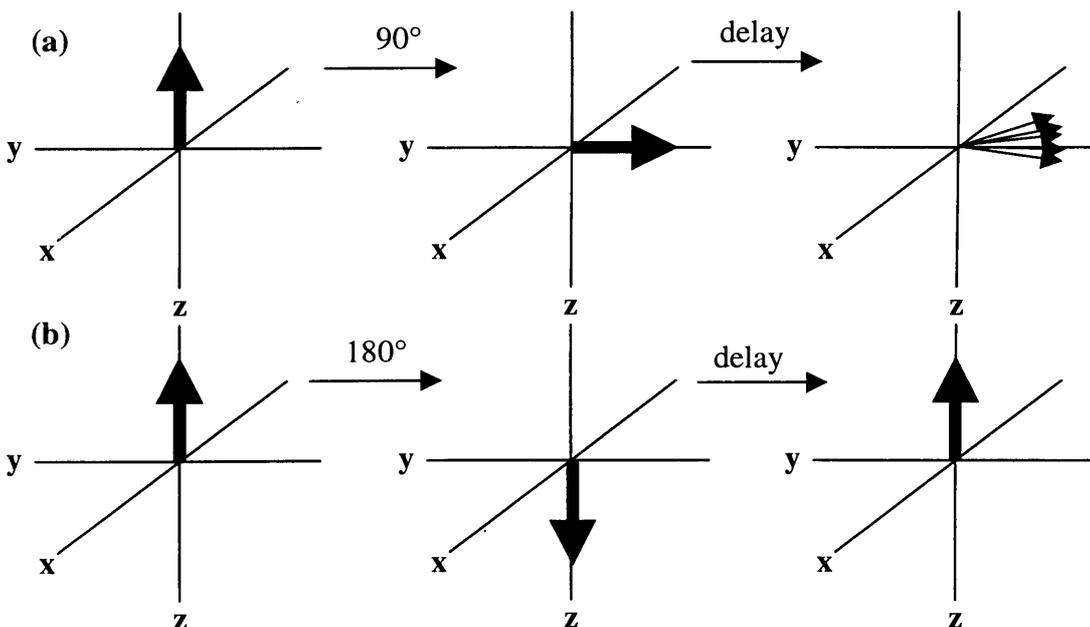
### 5.1.3 Relaxation

Once a radiofrequency (RF) pulse has been applied, the populations of  $\alpha$  and  $\beta$  states are disturbed from their equilibrium values. This perturbed net magnetisation will relax back to its equilibrium position through two types of relaxation process, known as  $T_1$  and  $T_2$  relaxation.  $T_1$  is the time taken for the populations of the energy states to return to the original Boltzmann distribution, and hence the time taken for the net  $z$  magnetisation to return to its original value along  $+z$ . This is *spin-lattice* relaxation, and arises when the spin exchanges energy with the lattice. Spins in the  $\beta$  state can emit energy to the lattice, whilst those in the  $\alpha$  state can absorb energy from the lattice and make the transition to  $\beta$  state. All transitions are caused by irradiation at the Larmor frequency which enables the spin to either absorb or emit energy. This irradiation comes from fluctuating local magnetic fields which can be envisaged as acting like localised RF pulses. The “lattice” constitutes the various motions of molecules which lead to a locally fluctuating magnetic field at the site of the spin involved. The main causes of these local magnetic field changes are nuclear *dipole interactions* and *chemical shift anisotropy* (CSA). The dipolar interactions are the result of the influence of nuclear magnetic moments with each other. The size of such interactions depends on the inter-nuclear distance, and the orientation of the inter-nuclear vector in the  $B_0$  field. In general, the electron cloud causing the nuclear magnetic shielding is asymmetric, so the degree of shielding depends on its orientation in  $B_0$ . In practice this manifests itself as chemical shift anisotropy, which usually averages to zero in small molecules in solution. For

larger proteins in solution, the effects of CSA become more apparent and give rise to fluctuating local magnetic fields which can cause relaxation.

$T_2$  is a measure of the time taken for the coherence in the  $xy$  plane (generated after a  $90^\circ$  pulse on  $+z$  magnetisation) to dephase to its equilibrium value of zero. There is a loss of phase coherence and a “fanning out” of the net magnetisation in the  $xy$  plane (Figure 5.6a). With  $T_2$ , or *spin-spin* relaxation there are no energy state transitions involved, the effect is an entropic one, rather than enthalpic.  $T_2$  relaxation is caused by irradiating the spins at any frequency (not just the Larmor frequency). As the molecule tumbles in solution, CSA and dipolar interactions lead to local alterations in the magnetic field, and hence the magnitude of the  $z$  axis magnetisation vector. This in turn causes the magnetisation vector component in the  $xy$  plane to precess faster or slower about the  $z$  axis. Consequently, there is an observed dephasing of formerly coherent magnetisation vectors precessing in the  $xy$  plane as their Larmor frequencies are slightly perturbed. Spin-spin exchange, also accounts for the dephasing of precessing spins. Here a spin in the  $\alpha$  state flips to the  $\beta$  state upon receiving energy from another spin which has undergone the reverse transition. The result is a loss of coherence-  $T_2$  relaxation.

The rate of relaxation is a function of the rate of molecular motion. The rate of  $T_2$  relaxation increases with increasing molecular mass, and consequent slowing of molecular tumbling times in solution. This is important because the widths of lines in NMR spectra are inversely proportional to  $T_2$ - slow moving molecules have larger line widths than smaller, faster moving molecules (line widths are measured in Hz, at half peak height).



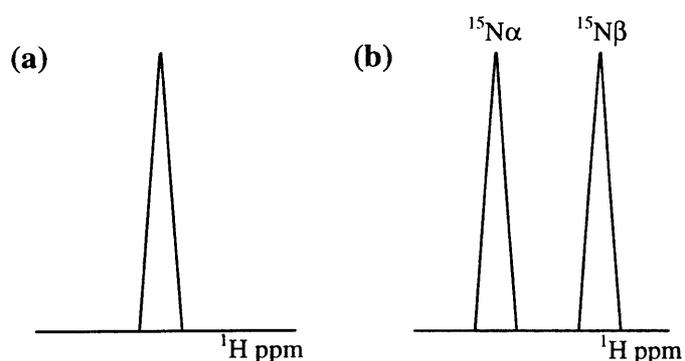
**Figure 5.6 (a)**  $T_2$  relaxation taking place in  $xy$  plane magnetisation. The spins dephase and net magnetisation in the  $xy$  plane decreases.

**(b)**  $T_1$  relaxation where the number of spins in the  $\alpha$  and  $\beta$  states return to their Boltzmann distribution equilibrium values. This leads to a return of magnetisation to the  $+z$  axis.

Broad linewidths in NMR spectra are undesirable, because as a line broadens out, its intensity is decreased. Since protons have the largest magnetic moment, and are the most abundant nucleus type in proteins, they contribute greatly to locally fluctuating magnetic fields and hence increasing the rate of relaxation. The magnetic moment of deuterons is one sixth that of protons. Deuterons result in less fluctuation in local magnetic fields than protons. Replacement of protons with deuterons lowers contributions to linewidth from relaxation, and hence sharpens spectral lines.

#### 5.1.4 Coupling

Another NMR phenomenon which has been made use of in this thesis is *coupling*, which leads to multiplicity of NMR signals. The magnetic moment of one nucleus weakly polarises the bond electrons, which in turn cause an alteration in the local magnetic field experienced by the second nucleus. Thus through bonding electrons, one nucleus can affect the transition energy between  $\alpha$  and  $\beta$  states, and hence the Larmor frequency of the second nucleus. In a sample containing a large number of molecules, a particular  $^{15}\text{N}$  spin will be in the  $\alpha$  state for approximately half the molecules, and the  $\beta$  state for the other half. Consider a theoretical case where a single proton is bonded to a  $^{15}\text{N}$  nucleus. The local magnetic field experienced by half the proton population coupled to an  $\alpha$  state  $^{15}\text{N}$  will be different to that experienced by the other half of the proton population (coupled to a  $\beta$  state  $^{15}\text{N}$ ). This effect is known as scalar coupling. The result is that two Larmor frequencies, separated by the coupling constant  $J$ , are experienced for one particular proton spin (one for when the coupled  $^{15}\text{N}$  spin exists in the  $\alpha$  state, and one when it is in the  $\beta$  state, Figure 5.6).



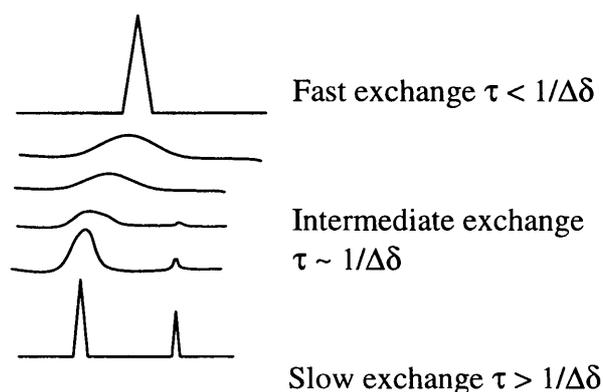
**Figure 5.6** Proton spectra of (a) an uncoupled proton, and (b) a proton coupled to a  $^{15}\text{N}$  spin through a single bond. The populations of proton spins experience two local magnetic fields depending on whether the coupled  $^{15}\text{N}$  spin is in the  $\alpha$  or  $\beta$  state. This manifests itself as two Larmor frequencies detected.

The one-bond coupling constant between an amide proton and nitrogen in a protein,  $^1J_{\text{NH}}$  is about 90Hz. This means that the difference in frequency between the two signals arising due to this coupling is 90Hz. Longer range couplings are also observed. A 3-bond coupling between an amide proton and an  $\alpha$  carbon proton,  $^3J_{\text{NH}\alpha\text{H}}$  is 4-10 Hz.

### **5.1.5 Chemical exchange**

Another phenomenon which has been observed during the course of this work has been that of *chemical exchange*. Protein structures are often in a state of dynamic equilibrium. Loop regions which contain no secondary structure, may for example move between different environments. This means that the spins in that loop will also experience different magnetic environments, and consequently different chemical shifts. The effect of this on the NMR spectrum is dictated by the chemical shift difference, and the rate of exchange between the environments (chemical exchange). If the rate of exchange is very slow compared with the chemical shift difference between two environments (in Hz), then a spin will spend enough time in each environment alone to produce two separate chemical shift signals. If the relative rate of exchange is very fast, then the spins do not spend enough time in each environment to produce a separate signal for each. Instead they appear to precess at an average frequency. The spectral peak produced is a measure of the relative time spent in each environment, and the intensity of this peak is usually not greatly diminished relative to when the spins exchange slowly between environments.

If  $\Delta\delta$  is the difference in chemical shift between the spin in each environment in Hz, then a spin occupying a particular environment for around  $1/\Delta\delta$  seconds will exchange at an intermediate rate between the two environments. If for example, the difference in chemical shift between environment A and environment B is 100Hz, and the spin exists in environment A for 10 milliseconds then an intermediate exchange regime will be observed. If the chemical shift difference was only 10Hz, then 100 milliseconds in one environment would render the intermediate exchange regime. One way of thinking about it is that the more similar the magnetic environments, the more similar the chemical shifts, and the more time a spin has to spend in one of them to be in intermediate exchange. With intermediate exchange rates spectra vary between the two separate signals observed with slow exchange, and the single averaged signal observed with fast exchange (Figure 5.7). This leads to signal linewidth increases and consequent intensity decreases. For this reason, intermediate exchange can be the bane of a spectroscopist's life when attempting to assign a protein !



**Figure 5.7** The effect of different exchange rates between alternate magnetic environments. The spectrum top shows fast exchange, the middle section intermediate exchange and the lower spectrum slow exchange regimes between two environments.  $\tau$  is the lifetime of the spin in one state,  $\Delta\delta$  is the difference in chemical shift between the two states (in Hz).

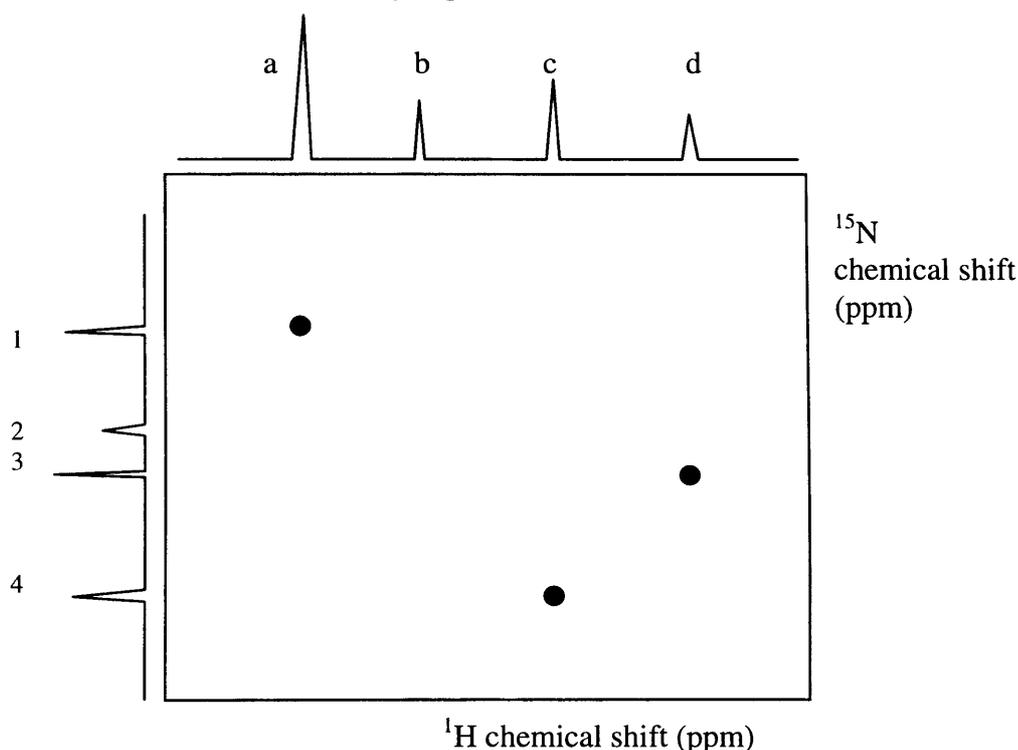
### 5.1.6 Two-dimensional NMR

The problem with one-dimensional proton NMR spectra of proteins is that due to the similar chemical shifts, there is extensive signal overlap. It is difficult to assign proton resonances due to this poor signal resolution. Two-dimensional NMR provides a means of simplifying spectra of one group of spins (eg. protons) and also gives extra information about another group spins (eg.  $^{15}\text{N}$ ). This second group of spins is linked either through space, or through bond, to the first group of spins. One such example is the HSQC (Heteronuclear Single Quantum Coherence) which in this thesis has been mainly used to obtain the chemical shifts of protons directly bonded with  $^{15}\text{N}$  atoms (eg. in the peptide backbone, or Asn and Gln sidechains).

The final two dimensional NMR spectrum is usually represented as a contour plot, with signals containing an x and y coordinate. The chemical shift of proton is plotted along the x axis, and that of  $^{15}\text{N}$  along the y axis (Figure 5.8). Such a spectrum may be thought of as equivalent to having two one dimensional spectra up each axis, with signals only showing up in the two dimensional spectrum when the proton and  $^{15}\text{N}$  spins are coupled.

Exactly how multidimensional NMR works from the perspective of spins is too complicated to be covered in this short introduction. Essentially, such experiments work because phase coherence can be exchanged between different spins. The essence of the HSQC experiment for looking at backbone amides is that the proton spins are excited and their phase coherence is transferred to the  $^{15}\text{N}$  spins to which they are coupled. Parameters can be set up such that this coherence transfer occurs only to the  $^{15}\text{N}$  spin directly coupled to a proton (ie. through one bond). The Heteronuclear Multiple Quantum Coherence (HMQC) experiment provides

the same information as the HSQC. It achieves this end through a different transfer of coherence than in the HSQC (essentially in the HMQC both  $^{15}\text{N}$  and  $^1\text{H}$  net magnetisation vectors lie in the  $xy$  plane in the same time, whilst with the HSQC only  $^1\text{H}$  or  $^{15}\text{N}$  lies in this plane at any one time). The HMQC experiment contains fewer RF pulses than the HSQC, and so it should be less sensitive to pulse imperfections. However, the  $T_2$  relaxation rate which determines the linewidth in the second dimension ( $F_1$  dimension) is faster for multiple-quantum coherence than for single-quantum coherence.



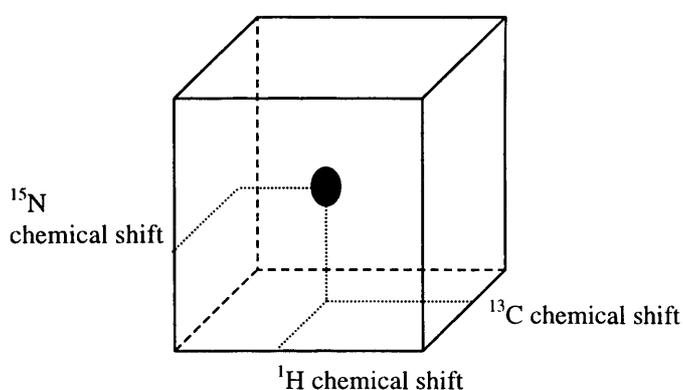
**Figure 5.8** A  $^1\text{H}$ - $^{15}\text{N}$  correlation two-dimensional HSQC (or HMQC) spectrum (boxed) flanked by the proton and  $^{15}\text{N}$  one-dimensional spectra. Only protons coupled to  $^{15}\text{N}$  spins by one bond exhibit peaks in the HSQC.  $^{15}\text{N}$  peak 4 is correlated to  $^1\text{H}$  peak c for example. Proton peak b is not correlated to any  $^{15}\text{N}$  signals because it is not coupled to an  $^{15}\text{N}$  nucleus by one bond.

The HMQC also exhibits proton-proton splitting in the  $F_1$  dimension which is absent in the HSQC. These factors often lead to narrower lines in the  $F_1$  dimension, and consequently higher sensitivity in the HSQC. If however, the species measured relaxes slowly, then it is less relevant whether the HSQC or HMQC is employed to measure it. The imidazole nitrogen magnetisation relaxes more slowly than that of a backbone amide nitrogen. Consequently, the HSQC was used to examine the backbone amides of Ada-C, whilst the HMQC was employed to examine the tautomeric and charged state of the histidine sidechain imidazole (Chapter 6). In this experiment, couplings between protons and  $^{15}\text{N}$  atoms separated by more than one bond were detected. This differs from the HSQC used to examine the backbone amide of Ada-C, which detected protons and  $^{15}\text{N}$  that were coupled via a single bond. Theory may

dictate that the HSQC would have been a superior choice of experiment to examine the histidine tautomeric states than the HMQC, but in practise there was little observed difference in the sensitivity for the two types of experiment (C. Damblon, personal communication).  $^1\text{H}$ - $^{15}\text{N}$  correlation HSQC spectra are useful for locating the binding site on a protein when it binds to a ligand. An HSQC spectrum of the protein alone is recorded, and then re-recorded in the presence of a ligand. Any chemical shift changes might occur as a result of changes in the magnetic environment due to ligand binding. The ligand-binding site on the protein can therefore be mapped, provided that the chemical shift assignments of the protein resonances have been established. Usually mapping can be carried out after assigning the protein backbone amide  $^1\text{H}$  and  $^{15}\text{N}$  spins. Sidechain assignments render further information, but take more time to gather. In this thesis chemical shift mapping experiments relied on the backbone assignments of Ada-C alone.

By recording HSQC spectra on proteins in the presence of different ratios of protein:ligand, it is possible to obtain an approximate binding constant if the free and bound protein are in fast exchange. The binding constant is measured by monitoring the magnitude of chemical shift movement against ratio of protein:ligand. This method was also applied to Ada-C to yield an approximate binding constant to single-stranded DNA.

In order to obtain the chemical shift assignments of the protein backbone atoms (amide proton and nitrogen, carbonyl carbon (CO),  $\text{C}\alpha$ , and  $\text{C}\beta$  atoms, three-dimensional NMR experiments need to be performed. These experiments serve two purposes, they provide extra chemical shift information on a third spin type (in the experiments used here,  $^{13}\text{C}$ ) and in so doing reduce the spectral overlap which may have been incurred in the proton and  $^{15}\text{N}$  dimensions. A 3D-NMR experiment is like a two-dimensional HSQC experiment with a third dimension tacked on. Figure 5.9 shows a schematic of a 3D data set.



**Figure 5.9** A three-dimensional heteronuclear NMR spectrum. This is an extension of the two-dimensional HSQC experiment with another dimension tacked on. This dimension provides chemical shift information on a  $^{13}\text{C}$  atom from the protein backbone, eg.  $\text{C}\alpha$ . The advantage of this is that any signal overlap in two-dimensions may be resolved in the  $^{13}\text{C}$  dimension, and  $^{13}\text{C}$  chemical shift information is required to help trace along the backbone when performing the assignment.

### **5.1.7 NMR of Ada-C**

The aim of this thesis was to establish the manner by which Ada-C bound DNA. Since co-crystallisation experiments with Ada-C and DNA failed to yield crystals, full structure determination by x-ray crystallography was not possible. Full structure determination of the complex by NMR would require much time, given the final size of the complex and the poor spectral dispersion of Ada-C. Since the structures of Ada-C and DNA are already known, the strategy employed was to localise the DNA-binding site on Ada-C, and then model the two molecules together.

Chemical shift assignment of the Ada-C polypeptide backbone resonances was necessary before a monitoring of interactions between the protein and its DNA substrate could begin. These assignments should be sufficient for identifying the DNA binding site on Ada-C using the “chemical shift mapping” strategy. Any gross conformational change predicted for Ada-C upon binding DNA, (Moore *et al.*, 1994) should be visible, and the identity of the parts of the protein backbone involved definable.

### **5.1.8 The Backbone Assignment Strategy**

The classical approach was to identify amino acid spin systems using experiments such as COSY, (COrrrelation SpectroscopY) and TOCSY, (TOtal Correlation SpectroscopY). Links between the spin systems are identified using  $^{15}\text{N}$ -edited NOESY, (Nuclear Overhauser Effect SpectroscopY) experiments which give through-space connections along the backbone, (eg. linking  $\text{C}\alpha$  proton of residue  $i$  to the amide proton of residue  $i+1$ ). This approach is very difficult for larger proteins such as Ada-C due to the large spectral overlap. The heteronuclear multidimensional methods described below rely on the scalar coupling, (magnetisation transfer through-bond) as opposed to the dipolar, (through-space) coupling of NOESY experiments, to make the inter-residue connections. These scalar couplings are more efficient than through-space magnetisation transfers and so the final signal is stronger. Heteronuclear experiments mean using more than one NMR-active nuclei type in the protein. In the case of Ada-C,  $^{13}\text{C}$  and  $^{15}\text{N}$  nuclei were used along with the ubiquitous  $^1\text{H}$ .

A variety of different heteronuclear triple resonance experiments were performed on Ada-C. Each experiment provided different chemical shift information from the different through-bond couplings for magnetisation transfer used, as shown in Figure 5.10. For instance, in the HNC0 experiment magnetisation was transferred from the amide proton, through the amide  $^{15}\text{N}$  to the carbonyl  $^{13}\text{C}$  and then back to the proton, after which the signal was detected. Only protons coupled in this way to carbonyl carbons showed up as crosspeaks. The position of a

single crosspeak in the 3D spectrum gave the chemical shifts of an amide proton and amide nitrogen of residue  $i$  and of the carbonyl carbon of residue  $i-1$ . Other triple resonance experiments showed the chemical shifts of different atoms, eg.  $C\alpha$  or  $C\beta$  together with amide proton, (HN) and nitrogen, (N), (Figure 5.10 and Table 5.1 below).

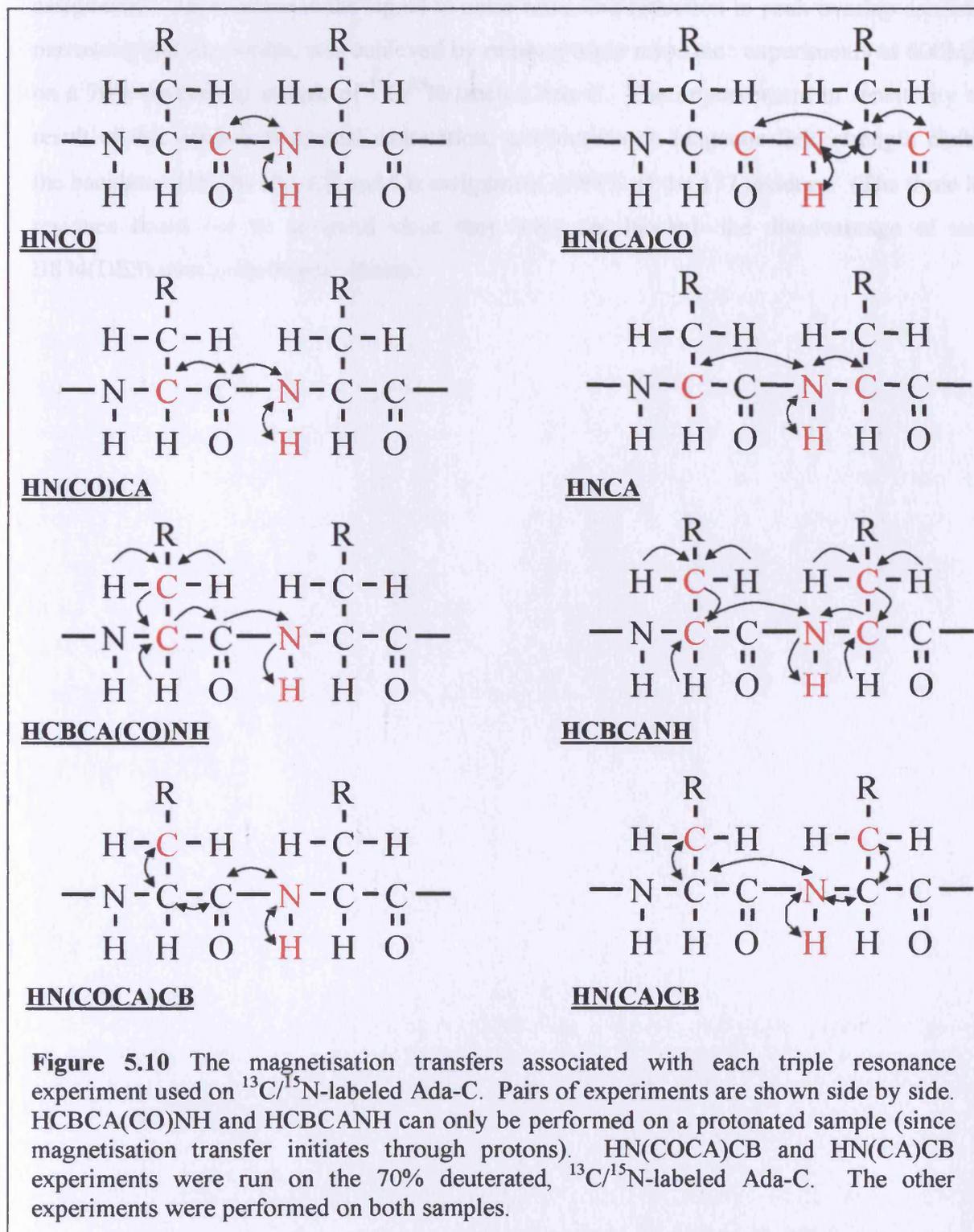
Spin identification was the first part of the assignment procedure following data collection. A particular chemical shift peak was classified as belonging to a particular atom type and the peaks then grouped into dipeptide spin systems. In each of these systems there was chemical shift information about HN, N, carbonyl C, (CO),  $C\alpha$ , and  $C\beta$  from the current residue and the previous residue in the sequence.

<b>Triple Resonance Experiment</b>	<b>Chemical shift information obtained for residue <math>i</math></b>	<b>Chemical shift information obtained for residue <math>i-1</math></b>
HNCO	HN, N	CO
HN(CA)CO	HN, N, CO	CO
HN(CO)CA	HN, N	$C\alpha$
HNCA	HN, N, $C\alpha$	$C\alpha$
HCBCA(CO)NH	HN, N	$C\alpha$ , $C\beta$
HCBCANH	HN, N, $C\alpha$ , $C\beta$	$C\alpha$ , $C\beta$
HN(COCA)CB	HN, N	$C\beta$
HN(CA)CB	HN, N, $C\beta$	$C\beta$

**Table 5.1** Chemical shift information obtained from each triple resonance experiment. HN: amide proton, N: amide nitrogen, CO: carbonyl carbon,  $C\alpha$ : backbone alpha carbon,  $C\beta$ : sidechain beta carbon. Note that in the case of Gln and Asn sidechains crosspeaks may also be observed in experiments whose magnetisation transfer originates from through the amide proton such as HNCO. These peaks can usually be identified as sidechain amide by virtue of their  $C\beta$  and  $C\alpha$  shifts (often antiphase with respect to those emanating from backbone amides)

The next step is the sequential assignment, (ordering these dipeptide spin systems in a manner consistent with the primary sequence of the protein, and chemical shift information for residue type). For this, a global optimisation procedure, which employs a simulated annealing algorithm was used (K.H. Sze, I.L. Barsukov, L-Y. Lian, G.C.K. Roberts, unpublished work). This determined the sequential arrangement of spins which had the lowest energy in terms of both chemical shift linkage between adjacent spin systems, and the agreement of these systems with the known protein sequence. Information from the selective labeling of Ada-C

could be incorporated into the simulated annealing program to 'lock' certain spins to a defined residue type and thus impose needed constraints on the program to increase the accuracy of the output.



Initially triple resonance experiments were carried out on a protonated  $^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C sample at 500MHz field strength, but the data was of insufficient quality to complete the assignment. An increase in the signal to noise ratio, and reduction in peak overlap caused by narrowing the line width, was achieved by running triple resonance experiments at 600MHz<sup>1</sup> on a 70% deuterated sample of  $^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C. The improvement in sensitivity as a result of this random fractional deuteration, and increase in magnetic field strength enabled the backbone HN, N, CO, C $\beta$  and C $\alpha$  assignment of 94% of the 177 residues. (The three Met residues could not be assigned since they were not labeled- the disadvantage of using B834(DE3) over prototrophic strains).

---

<sup>1</sup> The upgraded Bruker DRX 600 spectrometer was only installed in February 1999. The improved electronics enabled better quality data to be obtained which enabled triple resonance experiments to be performed routinely at 600MHz.

## 5.2 RESULTS

Analysis of rapidly obtainable 2D versions of the 3D triple resonance experiments performed on a number of concentrations of Ada-C, revealed that at about 36mg/ml aggregation became more acute than at lower concentrations. The  $T_2$  relaxation and linewidth were adversely affected by this aggregation, resulting in a concomitant loss of signal intensity. Thus for the protonated  $^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C sample 31mg/ml was used, whilst for the 70% deuterated  $^{13}\text{C}/^{15}\text{N}$ -labeled sample 26mg/ml was chosen as the concentration to record the triple resonance data.

### 5.2.1 Limited backbone assignment through the use of protonated $^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C

Even using 3D experiments there are difficulties resolving peaks due to poor spectral dispersion in the  $^{13}\text{C}$  dimension, as well as in  $^{15}\text{N}$  and  $^1\text{H}$  dimensions. Many  $\text{C}\alpha$  links were found between dipeptide spin systems, but the high level of ambiguity meant these systems could not be successfully ordered along the polypeptide backbone.  $\text{C}\beta$  links were present for most Alanine resonances due to the extra intensity of these peaks. However they were lacking for other residue types. Carbonyl carbon links were most frequently absent also due to the poor sensitivity of the  $\text{HN}(\text{CA})\text{CO}$  experiment (see Table 5.2). This poor sensitivity, due to the inherent flexibility of the protein, led to the assignment of only 44 out of 170 backbone resonances. These assignments occurred in short stretches, not totaling more than 5 residues, and were frequently involved in secondary structure.

EXPERIMENT	SIGNAL:NOISE RELATIVE TO HNCO
HNCO	1
HN(CA)CO	0.1
HNCA	0.5
HN(CO)CA	0.6
HCBCA(CO)NH	0.2
HCBCANH	0.1

**Table 5.2** Information concerning the magnetisation transfers, and relative sensitivities of the different experiments used on protonated Ada-C. These values were estimated by considering the details of the magnetisation pathways and were obtained from Prof. S. Grzesiek (personal communication).

### 5.2.2 Successful backbone assignment employing 70% deuterated $^{13}\text{C}/^{15}\text{N}$ -labeled Ada-C

Through the use of the higher 600MHz field strength and the 70% deuterated  $^{13}\text{C}/^{15}\text{N}$ -labeled sample of Ada-C, a successful assignment of 164 backbone HN and N resonances was made. Omitting the unlabeled methionines, and the unassignable imino prolines, only 3 residues (Q2, K137 and W159) remain unassigned (Figure 5.11). In most cases it was possible to get

the CO, C $\beta$  and C $\alpha$  assignments of the proline residues. All assignments are listed in Table 5.3 below.

Residue Number	Amino Acid type	HN chemical shift (ppm)	N chemical shift (ppm)	C $\alpha$ chemical shift (ppm)	CO chemical shift (ppm)	C $\beta$ chemical shift (ppm)
1	M	np	np	np	np	np
2	Q	0	0	55.642	175.088	29.442
3	F	8.462	123.002	57.714	175.497	39.522
4	R	8.324	123.641	55.806	175.753	30.499
5	H	8.450	121.475	56.085	175.754	30.279
6	G	8.556	111.460	45.430	174.688	np
7	G	8.412	109.807	45.286	174.225	np
8	E	8.505	121.136	56.914	176.264	29.845
9	N	8.554	120.242	53.242	174.637	38.742
10	L	8.002	123.753	54.834	176.268	41.004
11	A	8.316	126.830	52.716	177.396	18.551
12	V	7.932	121.686	61.227	174.452	34.715
13	R	8.469	126.549	53.001	174.850	31.702
14	Y	8.084	116.055	56.054	173.467	43.463
15	A	9.150	122.042	51.300	174.734	23.958
16	L	8.411	121.163	53.916	176.607	43.630
17	A	9.186	124.306	51.235	174.924	22.393
18	D	8.189	118.725	54.518	175.548	42.675
19	C	8.748	115.556	56.310	175.486	31.993
20	E	9.287	120.706	59.075	176.561	29.255
21	L	7.721	117.673	53.965	175.033	41.803
22	G	7.693	110.679	44.964	172.700	np
23	R	8.329	125.205	54.201	174.641	29.749
24	C	9.083	121.520	55.143	171.935	31.239
25	L	8.757	127.090	54.089	174.805	45.104
26	V	8.882	126.522	61.090	173.912	34.214
27	A	9.236	129.255	49.390	175.874	23.496
28	E	9.398	123.103	55.402	173.991	35.518
29	S	9.291	124.530	56.283	175.447	66.921
30	E	9.293	118.889	58.527	176.801	28.966
31	R	8.318	117.193	56.228	175.718	30.894
32	G	7.339	106.508	44.490	172.457	np
33	I	8.855	121.554	60.418	176.379	35.310
34	C	9.267	124.605	56.283	173.914	31.245
35	A	7.555	123.084	53.463	175.503	21.119
36	I	9.364	122.957	61.894	173.139	40.662
37	L	8.922	122.330	52.174	175.131	43.844
38	L	5.353	118.580	52.225	176.462	44.699
39	G	8.786	108.183	46.208	185.516	np
40	D	8.958	118.122	54.510	175.392	42.991
41	D	7.655	118.299	53.238	174.676	44.216
42	D	8.679	126.975	57.903	177.218	39.831
43	A	8.382	120.635	55.475	181.283	17.525
44	T	8.048	116.773	66.568	176.619	68.460
45	L	7.511	123.619	58.180	178.082	41.505
46	I	8.444	119.916	65.748	178.015	37.876
47	S	7.908	115.046	61.777	177.547	62.981
48	E	8.047	122.884	59.554	179.807	28.938
49	L	8.207	121.158	57.782	179.586	39.901
50	Q	8.565	118.040	57.869	177.398	27.738
51	Q	7.410	116.472	57.787	177.134	28.333
52	M	np	np	np	np	np
53	F	8.503	117.429	54.954	172.620	39.865
54	P	np	np	64.654	177.760	31.746

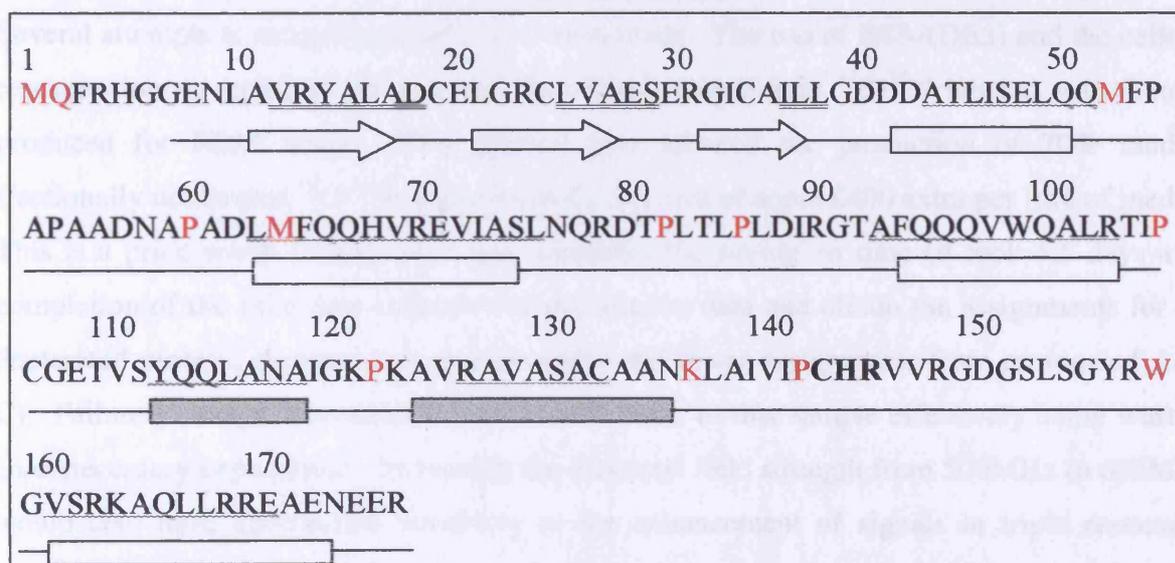
Residue Number	Amino Acid type	HN chemical shift (ppm)	N chemical shift (ppm)	C $\alpha$ chemical shift (ppm)	CO chemical shift (ppm)	C $\beta$ chemical shift (ppm)
55	A	8.545	120.766	51.255	177.039	17.845
56	A	7.830	123.819	52.608	176.510	20.364
57	D	8.504	121.301	52.796	174.583	44.408
58	N	8.551	121.207	52.561	174.537	39.155
59	A	8.423	125.303	50.720	174.836	19.247
60	P	np	np	64.167	177.050	31.637
61	A	8.382	120.635	51.490	176.480	18.618
62	D	7.200	120.542	53.242	175.793	41.478
63	L	8.333	126.720	58.184	179.937	40.992
64	M	np	np	np	np	np
65	F	8.065	121.193	59.324	177.520	38.237
66	Q	8.360	116.425	58.419	179.198	26.760
67	Q	7.747	120.278	59.077	178.594	27.009
68	H	8.130	120.283	57.263	176.971	27.824
69	V	8.314	120.390	67.537	177.370	30.662
70	R	7.542	117.591	59.876	179.476	29.172
71	E	7.919	121.486	59.656	179.551	28.884
72	V	8.376	122.029	66.928	178.014	30.644
73	I	8.480	120.167	64.229	177.267	35.989
74	A	7.761	120.544	55.401	180.507	17.361
75	S	7.441	113.512	62.604	176.104	63.897
76	L	8.171	120.146	57.262	178.834	40.185
77	N	7.980	116.309	53.681	175.170	38.881
78	Q	7.413	119.217	54.433	175.316	30.428
79	R	8.620	121.290	59.236	176.427	30.235
80	D	8.595	117.441	53.810	174.526	40.919
81	T	8.349	119.032	59.702	172.443	70.836
82	P	np	np	62.566	175.866	31.641
83	L	8.656	124.670	54.435	178.216	42.366
84	T	9.294	118.743	61.717	174.905	68.763
85	L	6.347	122.860	52.793	173.203	41.707
86	P	np	np	0	176.613	0
87	L	7.880	121.402	53.287	176.381	42.947
88	D	8.641	124.855	51.727	173.913	41.025
89	I	8.002	123.753	61.914	175.868	37.764
90	R	8.443	128.110	54.078	175.189	31.203
91	G	8.202	109.821	45.216	173.614	np
92	T	8.615	112.174	61.351	174.774	70.729
93	A	8.960	124.494	55.817	180.493	17.262
94	F	8.473	118.854	61.289	177.101	39.183
95	Q	7.473	118.042	59.097	176.775	25.561
96	Q	8.188	113.330	59.814	178.743	27.930
97	Q	7.644	120.355	59.061	179.183	27.942
98	V	7.546	121.754	66.929	177.433	30.748
99	W	8.806	120.151	58.292	178.371	28.309
100	Q	8.380	117.594	58.734	178.832	27.707
101	A	7.581	122.993	54.768	181.982	17.254
102	L	8.539	120.980	57.543	177.894	41.219
103	R	7.271	112.507	58.586	177.125	30.011
104	T	7.565	109.492	62.158	174.641	70.051
105	I	7.694	127.628	59.138	174.739	37.485
106	P	np	np	62.713	175.288	32.371
107	C	8.511	120.770	61.476	175.986	26.427
108	G	8.171	119.204	45.586	173.388	np
109	E	7.947	120.353	54.568	175.450	31.741

Residue Number	Amino Acid type	HN chemical shift (ppm)	N chemical shift (ppm)	C $\alpha$ chemical shift (ppm)	CO chemical shift (ppm)	C $\beta$ chemical shift (ppm)
110	T	7.741	107.909	59.385	174.144	73.493
111	V	8.543	114.781	58.814	173.818	35.454
112	S	8.592	117.741	56.157	177.041	66.025
113	Y	7.693	118.697	62.522	178.211	37.594
114	Q	8.433	118.564	58.407	178.117	29.014
115	Q	8.233	121.359	59.065	179.586	28.395
116	L	8.273	123.751	57.713	177.497	41.460
117	A	8.373	121.291	55.207	179.319	17.000
118	N	8.410	116.926	55.871	179.394	37.435
119	A	8.291	125.089	55.049	180.089	17.539
120	I	7.428	109.444	61.050	175.811	37.197
121	G	7.799	109.787	45.617	174.369	np
122	K	7.984	119.041	53.035	173.493	33.039
123	P	np	np	64.758	178.281	31.964
124	K	8.436	114.897	56.333	177.893	30.976
125	A	8.077	125.723	51.554	176.952	17.983
126	V	7.156	119.483	67.148	176.878	31.268
127	R	8.590	120.747	59.514	179.086	28.624
128	A	7.772	123.891	54.710	181.666	17.854
129	V	8.750	121.700	67.216	178.232	31.124
130	A	8.364	124.498	56.305	181.131	16.601
131	S	8.303	116.169	61.777	176.987	62.795
132	A	8.355	124.936	55.260	181.593	17.567
133	C	8.447	120.053	65.748	176.911	27.292
134	A	7.819	121.848	54.286	176.342	17.777
135	A	7.332	118.056	51.237	176.773	19.358
136	N	7.238	119.029	54.826	174.208	37.982
137	K	0	0	55.880	174.426	34.386
138	L	7.451	123.395	52.906	173.081	42.157
139	A	8.081	126.675	54.765	177.972	16.812
140	I	8.671	114.715	68.323	178.340	34.271
141	V	9.698	123.281	64.891	177.536	31.298
142	I	7.431	120.195	58.835	176.274	35.260
143	P	np	np	63.662	175.689	27.632
144	C	9.668	127.651	61.744	176.676	29.710
145	H	8.254	118.820	57.015	173.990	29.598
146	R	7.319	116.622	57.104	173.162	28.728
147	V	6.836	117.200	58.752	174.879	32.166
148	V	8.528	120.546	58.917	175.093	34.900
149	R	8.328	119.617	57.131	179.188	31.075
150	G	8.934	112.293	46.288	173.875	np
151	D	7.895	118.353	53.191	177.154	40.179
152	G	8.324	109.774	45.065	174.376	np
153	S	8.138	116.637	58.858	175.054	64.199
154	L	8.629	123.356	56.061	177.227	41.763
155	S	8.083	114.476	59.492	174.645	64.333
156	G	8.563	110.469	45.188	173.218	np
157	Y	8.122	120.994	57.045	177.673	40.396
158	R	8.457	130.268	59.328	0	0
159	W	0	0	0	175.330	0
160	G	7.249	107.668	45.413	174.957	np
161	V	8.780	124.856	66.383	177.388	31.242
162	S	8.689	116.701	61.695	178.022	62.564
163	R	7.946	122.483	59.896	177.319	29.595
164	K	7.489	121.203	60.267	177.485	31.998
165	A	8.263	118.944	55.127	180.453	17.517
166	Q	7.818	119.434	59.004	179.262	28.218

Residue Number	Amino Acid type	HN chemical shift (ppm)	N chemical shift (ppm)	C $\alpha$ chemical shift (ppm)	CO chemical shift (ppm)	C $\beta$ chemical shift (ppm)
167	L	8.203	122.295	57.697	178.664	42.148
168	L	8.355	118.467	58.089	180.753	40.267
169	R	8.126	119.623	59.336	179.036	29.433
170	R	7.959	118.853	57.824	178.131	28.311
171	E	7.714	118.668	58.448	176.197	28.665
172	A	7.506	121.968	52.983	178.033	18.869
173	E	7.707	119.355	56.389	175.900	29.795
174	N	8.372	120.207	53.250	175.035	39.290
175	E	8.455	122.092	56.480	176.200	30.079
176	E	8.441	123.416	56.511	175.497	29.794
177	R	7.971	128.365	57.336	181.031	31.050

**Table 5.3** Amide proton (HN), amide nitrogen (N), C $\alpha$ , C $\beta$  and carbonyl carbon (CO) chemical shifts for wild type Ada-C. An entry of zero indicates that chemical shift could not be assigned due to insufficient data. An entry of 'np' indicates that the chemical shift could not be assigned because it did not exist eg. Met residues were not labeled, so could not be assigned.

Chemical Shift Indexing (CSI) analysis was carried out on the carbon chemical shifts in the manner described by, and employing the software obtained from (Wishart *et al.*, 1994); (Wishart *et al.*, 1992). The results of this analysis are also depicted in Figure 5.11.



**Figure 5.11** The sequence of Ada-C, and the secondary structure obtained from the crystal structure (PDB accession code 1sfe). Helices 4 and 5 of the helix-turn-helix motif are shaded in grey, the conserved PCHR motif is shown in bold. Residues with unassigned backbone amides are shown in red. The secondary structure determined by CSI of NMR data is as follows: Double underlined residues are in an extended chain conformation, often indicative of  $\beta$ -sheet; wave underlined residues are in an  $\alpha$ -helix.

The secondary structure predicted using this method was in excellent agreement with that displayed by the crystal structure of Ada-C, (Moore *et al.*, 1994). This lends extra credence to the assignments made.

All peaks in the HSQC spectrum were assigned except two. One of these is found at 5.2ppm, 117ppm and is too weak and too close to the water peak to appear in the 3D triple resonance data. The other peak near the C-terminal Arg resonance appears in the triple resonance experiments but does not correlate to anything, it may be an artefact caused by minor truncation of the protein (it still appears in larger sweepwidth experiments).

Several sidechain amino groups from Gln and Asn residues were also assigned and are detailed in Table 5.4 below.

Residue sidechain amino group	Proton chemical shifts (ppm)	Nitrogen chemical shift (ppm)
Q2	7.545, 6.897	112.925
Q51	6.831, 7.377	112.793
Q78	7.578, 6.959	112.789
N77	7.439, 6.801	110.552
Q114	7.769, 7.406	116.936

**Table 5.4** Chemical shift assignments of sidechain amino groups from five Gln and Asn residues.

### **5.3 DISCUSSION**

Several attempts at assigning Ada-C have been made. The use of B834(DE3) and the cellular resuspension method has allowed suitable yields of uniformly  $^{13}\text{C}/^{15}\text{N}$  labeled Ada-C to be produced for NMR work. This method also allowed the production of 70% random fractionally deuterated  $^{13}\text{C}/^{15}\text{N}$  labeled Ada-C, at a cost of some £400 extra per litre of media. This is a price worth paying when one considers the saving on time (it took 5.5 days after completion of the final data collection to process the data and obtain the assignments for the deuterated protein, compared to several weeks and fewer assignments from protonated Ada-C). Failure to assign a protonated sample will result in that sample effectively being wasted, an unnecessary expenditure. Increasing the magnetic field strength from 500MHz to 600MHz would also have contributed positively to the enhancement of signals in triple resonance experiments.

Deuteration levels of 90% were required for the backbone assignment of the *Trp* repressor (Shan *et al.*, 1998), and are considered superior to 70% deuteration<sup>2</sup>. Attaining 90% deuteration would be difficult unless deuterated glucose was employed as the sole carbon source during protein expression (at considerable extra expense) or deuterated acetate used as the carbon source (often to the detriment of protein yield). 70% deuteration is easily obtainable with protonated glucose, and the cellular resuspension method used for the

<sup>2</sup> All references to deuteration are to random fractional deuteration (where not all non-exchangeable protons are replaced by deuterons)

production of Ada-C (Chapter 3). This level of deuteration was sufficient to assign the backbone of Ada-C, and has also sufficed to determine the solution structures of DNA/protein complexes (Zhou *et al.*, 1998).

In comparison with other proteins of similar molecular weight protonated Ada-C has broad linewidths and an abnormally low signal to noise ratio. This maybe accounted for by the unusually high level of conformational flexibility in the protein backbone, together with the tendency of Ada-C to aggregate. Such aggregation has been demonstrated empirically by analysing spectral linewidths at different protein concentrations, and also by diffusion methods developed and carried out by M.L. Tillet (M.L. Tillet, University of Leicester, PhD Thesis 1999).

An analysis of the temperature factors of a number of DNA-binding protein structures revealed that the C $\alpha$  B-factors of Ada-C were almost twice the average (Table 5.5). The average C $\alpha$  B-factor in the 1sfe structure of Ada-C is 45, much higher than those of other DNA-binding protein structures, suggesting that Ada-C is an inherently flexible protein.

PDB code	Protein	Average C $\alpha$ B-factor
1sfe	Ada-C	45.2
1bco	Bacteriophage muA tranposase core domain	15.5
1vsd	Avian Sarcoma virus integrase core domain	21.1
1itg	Hiv-1 integrase	37.2
1smt-A	Metallothionein repressor	21.3
1bisA	BirA biotin biosynthetic operon repressor	28.0
1rnl	nitrate-nitrite response regulator protein, NarL	33.3

**Table 5.5** The average B-factors for crystal structures of other DNA-binding proteins (complexed or free) for comparison with that of Ada-C. The mean of the average C $\alpha$  B-factors, excluding that of Ada-C is 26.0.

These high B-factors, suggest a higher degree of structural mobility in solution than other DNA-binding proteins. A consequence of this flexibility is a reduction in signal intensity, leading to a loss of the links between the CO and the C $\beta$  carbons in the triple resonance data obtained from the protonated Ada-C sample. Coupled with the poor spectral dispersion (caused by the high degree of helicity of the protein), only 44 out of 177 backbone amide resonances could be unambiguously assigned using a non-deuterated sample.

Replacement of protons with deuterons removes contributions to proton linewidth from proton-proton spin-spin relaxation, and  $^1\text{H}$ - $^1\text{H}$  scalar coupling, thus reducing linewidths. Deuteration levels of 50-75% make up for the reduced sensitivity resulting from the limited

number of protons with this reduction in linewidth (Gardner *et al.*, 1998). Deuteration of Ada-C caused a reduction of linewidth and hence an enhancement of broad signals, making the backbone assignment possible. Deleterious effects of aggregation and flexibility on signal intensity were counteracted by deuterating this protein. *i* to *i-1* Links for all carbon atoms were found for most residues with the deuterated sample, and the data lent itself to easy, and rapid backbone assignment.

Deuterium isotope effects on  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts could be compensated for by referencing the deuterated spectra identically to the fully protonated spectra. There were then found to be no random  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shift differences between identical signals. Chemical shift and assignment information could therefore be transferred between deuterated and protonated spectra (to the selectively-labeled spectra for instance). Such chemical shift transferability is a prerequisite if NOE-crosspeaks recorded using a protonated sample were to be assigned from the chemical shifts obtained with a deuterated sample. Such events might take place were one determining the three dimensional structure of Ada-C by NMR.

Selective labeling was of paramount importance to the assignment strategy of the protonated Ada-C, as it enabled simplification of the spectra and amino acid type identification. This data was also used to sort out ambiguities in the deuterated protein data, although it was less vital, and the assignment could have been completed without it.

## **5.5 SUMMARY**

A tendency to aggregate, the inherent conformational flexibility and high degree of helicity allowed the assignment of only 44 out of 177 backbone amides using  $^{13}\text{C}/^{15}\text{N}$ -labeled protonated Ada-C. 70% deuteration levels, and use of higher field strength caused a reduction in signal loss and a consequent assignment of 94% of backbone amides in Ada-C. These assignments were sufficient to utilise a chemical shift mapping strategy to determine the DNA-binding site on Ada-C (Chapter 6).

# Chapter 6

## *Analysis of DNA binding by NMR*

### 6.1 INTRODUCTION

Nuclear Magnetic Resonance (NMR) spectroscopy is a suitable tool to characterise the dynamic and structural events involved in DNA recognition and binding. Recent advances in NMR have enabled the determination of solution structures for a number of protein/DNA complexes. Such works usually require the use of 70%-90% deuterated,  $^{15}\text{N}/^{13}\text{C}$  labeled protein samples, 3D experiments and extensive use of NMR machine time (Zhou *et al.*, 1998), (Clare *et al.*, 1997). When the protein structure is already known, simpler 2D NMR can be conveniently used to map its DNA-binding interface more rapidly than obtaining a 3D structure of the complex. The most frequently used diagnostic probes are the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts of the protein backbone. These resonances are generally well resolved and once assigned, binding studies can be performed with inexpensive uniformly or selectively  $^{15}\text{N}$ -labeled protein (although expense must be incurred to make the  $^{13}\text{C}/^{15}\text{N}$  protein sample to assign the backbone in the first place).  $\text{H}^{\text{N}}$  and  $\text{N}$  chemical shifts are not sensitive to the same environmental and structural parameters as are the  $\text{H}\alpha$ ,  $\text{C}\alpha$ , or  $\text{C}\beta$  resonances. Changes in  $\text{C}\alpha$  chemical shift often indicate alterations in the secondary structure of the protein (Wishart *et al.*, 1994); (Wishart *et al.*, 1992). Any such changes upon binding DNA might be indicative of changes to secondary structure and hence a conformational change in the protein, (Holmbeck *et al.*, 1998), (Foster *et al.*, 1998). Thus, if a  $^{13}\text{C}$ -labeled protein is available, monitoring  $\text{C}\alpha$  shifts upon binding would provide added information.

The basis of the mapping technique with  $^{15}\text{N}$ -labeled protein is the perturbation of chemical shifts caused by altered magnetic environments of amino groups. Such shifts could be caused by altered protein conformation, hydrogen bonding variations, and/or novel contacts induced by DNA-binding. When monitoring the binding of ligand to a protein, the largest chemical shift changes observed upon ligand binding are most likely to emanate from parts of the protein involved in direct contacts to the ligand. In such cases, chemical shift changes can be mapped to the protein to localise the site of ligand-binding. This mapping technique is widely used to locate DNA-binding sites on proteins (Foster *et al.*, 1998), (Holmbeck *et al.*, 1998), (Shan *et al.*, 1998), (Artz *et al.*, 1996), (Ramesh *et al.*, 1994), (Baleja *et al.*, 1994). If no large, undefined conformational change occurs in the protein, this can be docked on to DNA in a manner consistent with the defined DNA-binding interface. In this way a model of the protein/DNA complex can be formed.

For proteins such as Ada-C where there is extremely poor spectral dispersion, obtaining the solution structure of the protein/DNA complex would be a Herculean task. Mapping the DNA-binding interface of the known Ada-C structure is a far more amenable process. Although it will reveal less detailed information than a high resolution structure obtained by NMR or x-ray crystallography, valuable information on the mechanism of binding and repair can still be elucidated. This process was successfully applied to the N-terminal domain of Ada to evaluate its binding mechanism to the Ada-Box (Sakashita *et al.*, 1995).

The accuracy of the DNA-binding model can be enhanced by the interpretation of NOE measurements to render distance constraints. In the case of Ada-C, such constraints were not gathered because of the poor spectral dispersion of NOESY spectra, and finally a lack of time. Binding of methylated and unmethylated DNA to the inactive Ada-C C144S mutant was also monitored using this technique. This mutant is less stable than the wild type protein, but maintains the same fold, and is therefore suitable for examining how Ada-C binds methylated DNA (something which cannot be achieved with the active wild type protein).

Together with structural details obtained by mapping, NMR has been used to probe the mechanistic details of the repair of methylated DNA by Ada-C. The effect of methylation and DNA binding on the tautomeric state of the active site histidine, and pK<sub>a</sub> of active site cysteine (C144) were assessed to provide insight into the mechanism of DNA repair. Observations from the NMR data provided useful information on the dynamics of the DNA-binding reaction. *In vitro* methylation reactions examined the stability of self-methylated Ada-C; the implications this posed for termination of the adaptive response are discussed.

A summary of the Ada-C complexes formed with DNA and studied by NMR are given in Table 6.1.

Ada-C	single-stranded unmethylated DNA	double-stranded unmethylated DNA	single-stranded methylated DNA	double-stranded methylated DNA
wild type	complex formed	complex formed	reaction performed	not done
C144S inactive Ada-C	complex formed	complex formed	complex formed	complex formed

**Table 6.1** Ada-C/DNA complexes formed and reactions between methylated DNA and wild type Ada-C carried out. All interactions were examined by <sup>1</sup>H-<sup>15</sup>N HSQC experiments to assess the effect of DNA-binding or demethylation on the structure of Ada-C. A number of differently labeled Ada-C samples were employed. These are elaborated on in the results section below.

## **6.2 RESULTS**

### **6.2.1 Mapping the DNA-binding site on wild type Ada-C**

Spectral changes upon addition of unmethylated single-stranded DNA which delineated binding, were clearly visible from the use of uniformly  $^{15}\text{N}$ -labeled Ada-C samples. However, due to the poor spectral dispersion, the exact nature of these changes was further examined using 52%-62% and 70% deuterated<sup>1</sup>, and uniformly  $^{15}\text{N}$ -labeled Ada-C, and selectively  $^{15}\text{N}$ -labeled Ada-C samples. The effect of deuteration on reducing linewidth was clearly visible, and enabled the resonances changing upon DNA-binding to be identified with greater ease. Selectively  $^{15}\text{N}$ -labeled Ada-C was employed to monitor chemical shift changes in regions of extreme signal overlap (around 8.4 ppm, 121 ppm). These spectra were vastly simplified versions of those generated by uniformly-labeled protein, and hence peak changes were more readily identifiable. Resonances of Ada-C selectively  $^{15}\text{N}$ -labeled at the amino position of either Cys, Ile, Gly, Val, Glu or Ala were monitored upon addition of single-stranded 20-mer DNA.

The degree of complexation was altered by changing the ratio of Ada-C:DNA and manipulating the ionic strength of the buffer. In high salt buffer (100mM NaCl and 100mM NaBr, 50mM phosphate pH 6.7, 10mM DTT, 1mM EDTA, 0.01%  $\text{NaN}_3$ ) a DNA-binding titration was carried out using 70% deuterated  $^{15}\text{N}$ -labeled Ada-C. A  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Ada-C was recorded at various ratios to single-stranded 20-mer unmethylated DNA (Figure 6.1). Chemical shift changes were observed upon addition of DNA. These were placed into two categories:

- (1) those resonances involving a change of less than 0.15ppm in proton or 0.5ppm in  $^{15}\text{N}$ -dimensions.
- (2) those involving greater chemical shift changes than category 1 upon complexation with a 2.5 fold excess of DNA.

These categories were chosen as they represented the two categories observed when Ada-C complexed with 2.5 fold excess of single-stranded unmethylated DNA in low salt buffer (10mM NaCl, 50mM phosphate pH 6.7, 10mM DTT, 1mM EDTA, 0.01%  $\text{NaN}_3$ ). Here the first category of resonances shifted, but were still visible (eg. A93, Figure 6.4), whilst category 2 resonances were line-broadened so extensively that they were lost below the level of the noise (eg. V126 and N136 Figure 6.4). By plotting the movement of the latter category of chemical shifts with respect to the amount of added DNA in high salt buffer, and fitting the curve to a standard binding equation, a  $K_d$  was determined (Figure 6.2).

---

<sup>1</sup> In this work, 'deuterated' refers to random fractional deuteration (where not all non-exchangeable protons are replaced by deuterons).

The equation employed was:

$$\delta_{obs} - \delta_A = \frac{(\delta_{DA} - \delta_A) \left\{ (D_T + A_T + K_d) - \sqrt{(D_T + A_T + K_d)^2 + 4D_T A_T} \right\}}{2D_T}$$

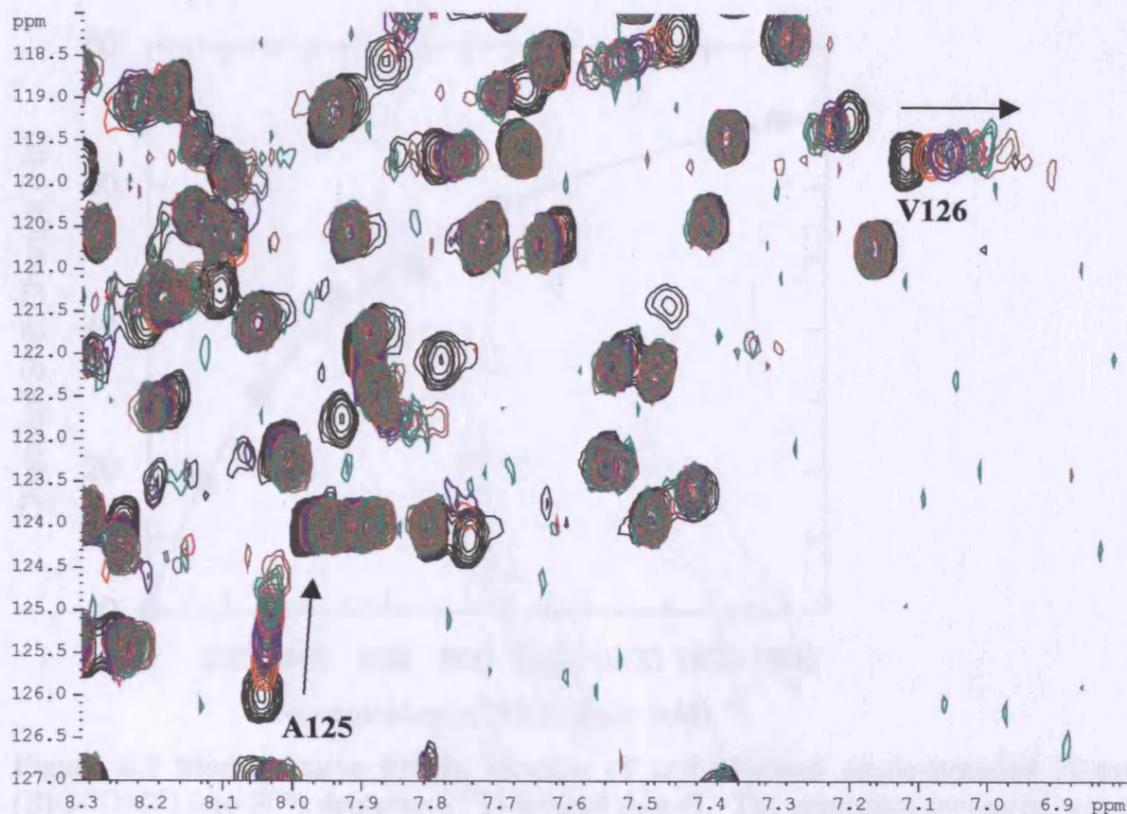
where  $K_d$  = dissociation constant,  $\delta_A$  = chemical shift of free Ada-C,  $\delta_{DA}$  = chemical shift of fully bound Ada-C,  $\delta_{obs}$  = chemical shift movement of bound Ada-C,  $D_T$  = total concentration of DNA,  $A_T$  = total concentration of Ada-C, (Lian *et al.*, 1993).

Following data fitting of 7 points gathered from the chemical shift of the Ala 125 amide peak, a  $K_d$  of  $595 \pm 54 \mu\text{M}$ , and  $\delta_{DA} = 95 \pm 3 \text{ Hz}$  were obtained. A  $K_d$  value of  $905 \pm 120 \mu\text{M}$  ( $\delta_{DA} = 114 \pm 8 \text{ Hz}$ ) was obtained from the movement of the V126 peak (the data fitting curve for V126 is not shown), showing that  $K_d$  measurement using this technique is consistent within an order of magnitude.

A gradual titration of resonance position suggested that Ada-C existed in a fast exchange regime between free and bound forms in high salt NMR buffer. In low salt NMR buffer (of similar ionic strength to the low salt ITC buffer in which the  $K_d$  of  $2.5 \times 10^{-6} \text{ M}$  was measured), the correspondingly higher affinity observed between Ada-C and DNA, and the extensive line-broadening observed for some resonances suggested the complex had moved into a more intermediate exchange regime.

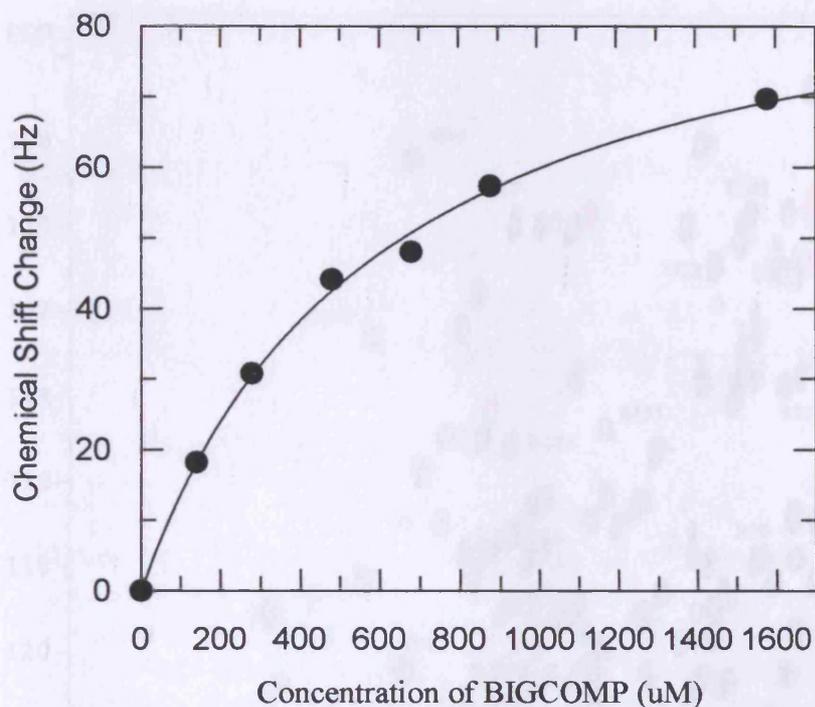
Figures 6.3 and 6.4 display the two categories of chemical shift alteration observed upon complexation of Ada-C with 2.5 fold excess DNA in low salt buffer (containing 10mM NaCl in place of 100mM NaCl, 100mM NaBr). Some minor chemical shift changes were observed in the control experiment where Ada-C was transferred into low salt buffer (eg. S155), but in most cases these did not include line-broadening, and were taken into account when analysing changes induced by DNA-binding. Such DNA-induced changes, as placed into the two categories of chemical shift movement, are depicted in Figures 6.5 and 6.6.

Larger chemical shift changes suggest greater changes in the chemical and magnetic environment of the nucleus involved. Areas of the protein backbone affected most upon binding DNA were the recognition helix of the helix-turn-helix motif (Helix 5), and the random coil spanning residues 149 to 158 (known as the 'wing').

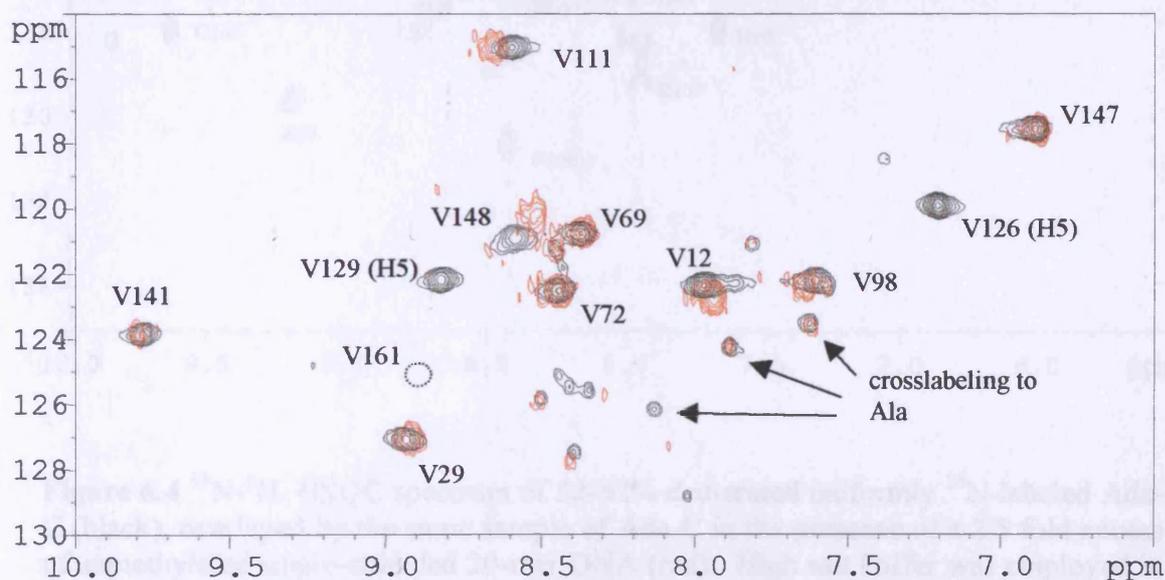


**Figure 6.1** Part of an HSQC spectra of 10mg/ml 70%  $^2\text{H}$ ,  $^{15}\text{N}$ -labeled Ada-C titrated with unmethylated ssDNA. In black: Ada-C; in red: Ada-C/DNA 3.7:1; in blue: Ada-C/DNA 1.8:1; in dark green: Ada-C/DNA 1:1, in pink: Ada-C/DNA 1:1.3; in light green: Ada-C/DNA 1:1.7, in brown: Ada-C/DNA 1:3.1. The titrating peaks (whose direction of movement is demarked by an arrow) are from the backbone amide of Ala 125 and Val 126, from the recognition helix of the helix-turn-helix.

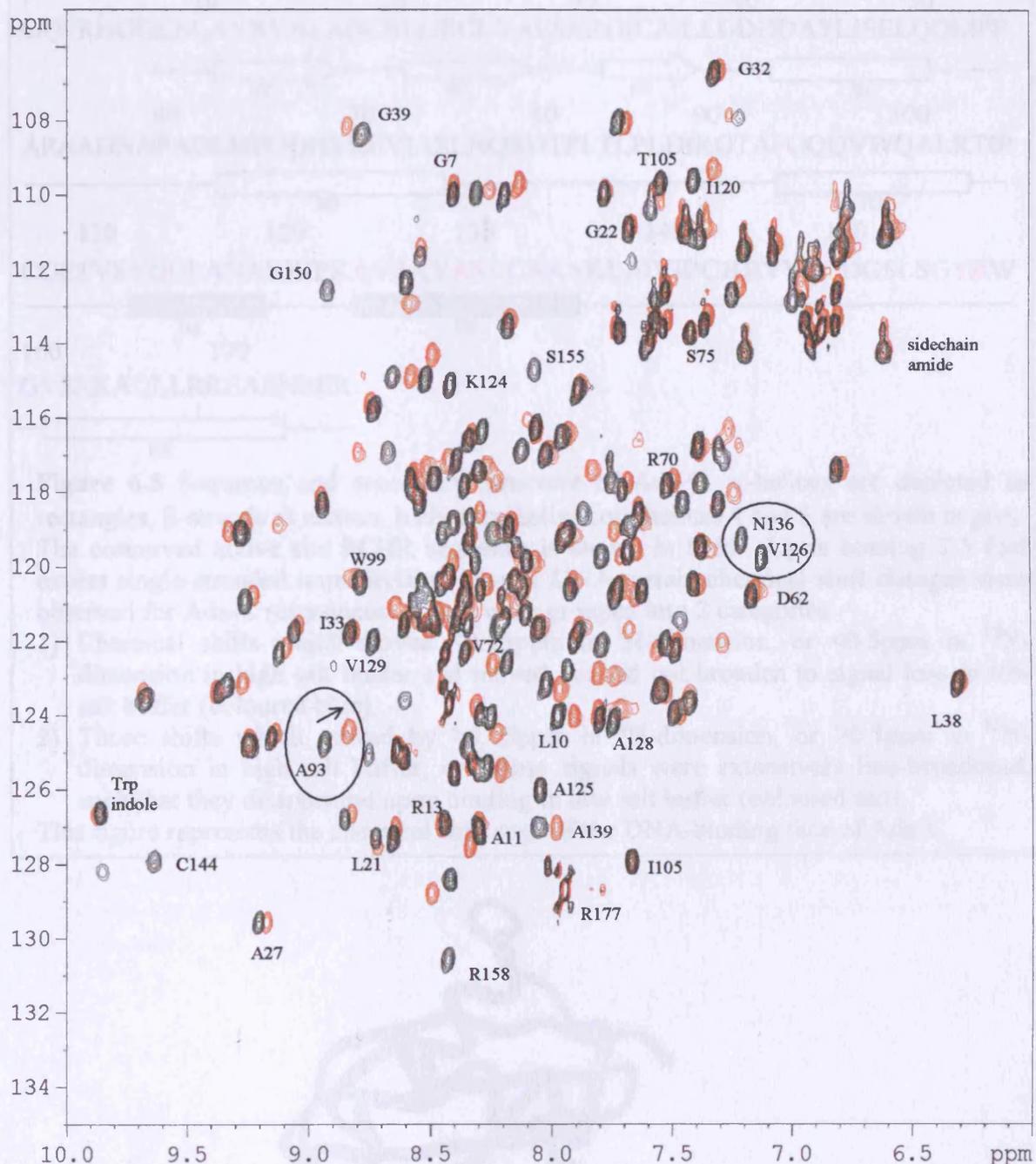
Other chemical shifts affected upon DNA-binding do not exhibit line broadening in low salt buffer, and were therefore unlikely to represent regions directly contacting the DNA. These regions (coloured blue in Figures 6.5 and 6.6) underwent minor alterations to their chemical environments probably as a result of secondary interactions transmitted from the binding interface.



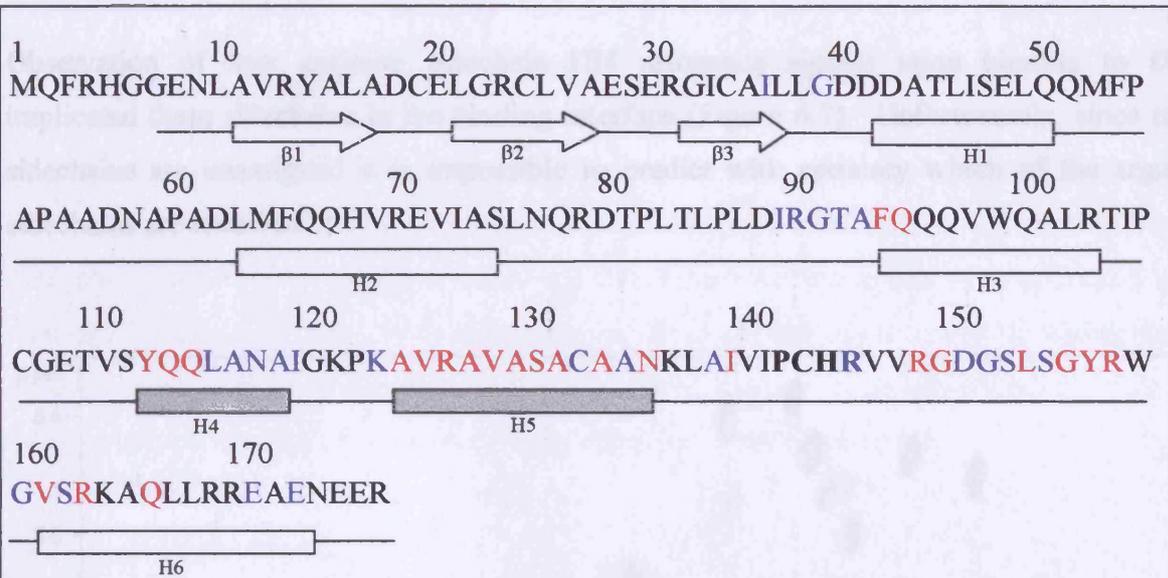
**Figure 6.2** Binding curve for the titration of unmethylated single-stranded 20-mer DNA (BIGCOMP) into 70% deuterated  $^{15}\text{N}$ -labeled Ada-C. The resonance measured was Ala 125. Following curve fitting to the standard binding equation described in the text, a  $K_d$  of  $\sim 0.6\text{mM}$  was determined. The  $K_d$  measured for the V126 peak shift was determined to be  $\sim 0.9\text{mM}$  (results not shown here).



**Figure 6.3** A  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum of Ada-C selectively labeled with Valine in presence (red) and absence of 2 fold excess of single-stranded unmethylated 20-mer DNA (black). The spectrum of the protein alone was taken in high salt NMR buffer containing 100mM NaBr and 100mM NaCl. When DNA was added, the buffer salt concentration was also dropped to 10mM to enhance binding. Both spectra were taken at pH 7.6 and 298K. Helix 5 from the HTH motif is labeled as H5. The concentration of Ada-C used was 13mg/ml. Note the crosslabeling to alanine (less intense peaks), and that the V161 amide signal is not visible (due to a combination of high pH, and line-broadening caused by conformational flexibility; it is depicted here as a dotted circle). Both categories of chemical shift change upon binding are observable here: category 1: small shifts like V111; category 2: resonance disappears eg. V129.



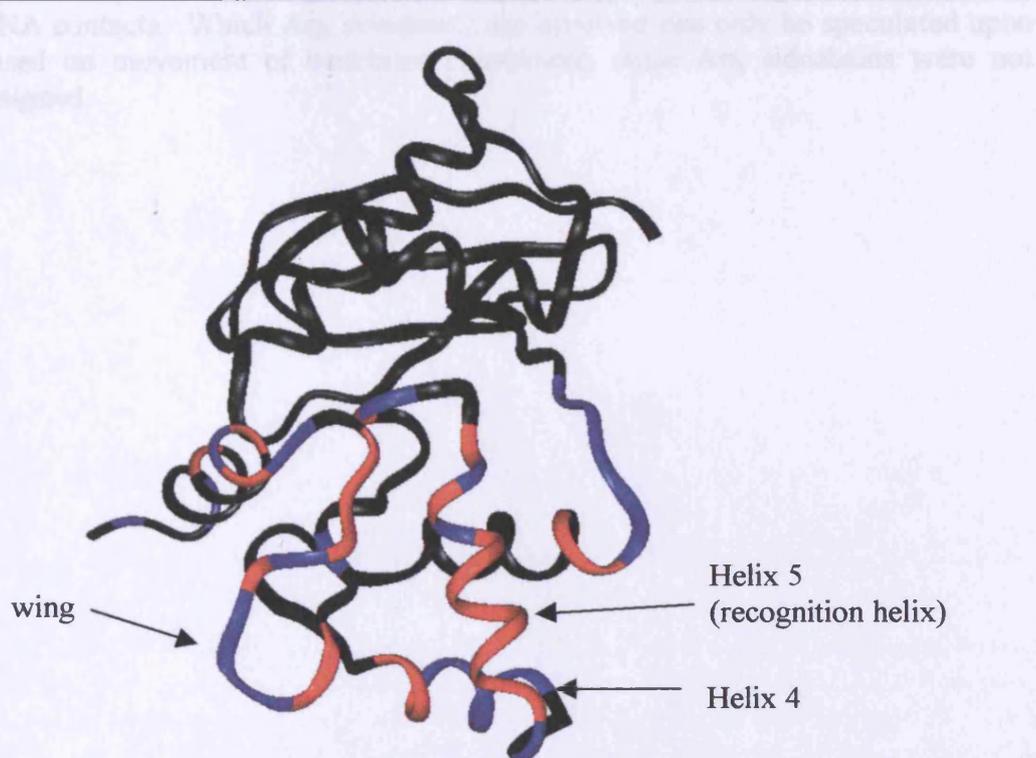
**Figure 6.4**  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum of 52-62% deuterated uniformly  $^{15}\text{N}$ -labeled Ada-C (black), overlaid by the same sample of Ada-C in the presence of a 2.5 fold excess of un methylated single-stranded 20-mer DNA (red). High salt buffer was employed in the solution containing Ada-C alone, whilst low salt buffer constituted the DNA/Ada-C complex solution to ensure tight binding. Both spectra were taken at 298K and pH 6.7, the concentration of Ada-C was 10mg/ml. A control spectrum of Ada-C in low salt buffer demonstrated that a few peaks shifted as a result of the buffer change, but none of these were assignable to the HTH region. Some assignments have been annotated to the spectrum, due to poor spectral dispersion, not all assignments peaks could be labeled. Examples of each category of chemical shift change have been circled; category 1: small shifts upon binding such as A93; category 2: resonance is broadened upon binding and disappears eg. V126, N136.



**Figure 6.5** Sequence and secondary structure of Ada-C.  $\alpha$ -helices are depicted as rectangles,  $\beta$ -strands as arrows, helix-turn-helix motif helices 4 and 5 are shown in grey. The conserved active site PCHR sequence is shown in bold. Upon binding 2.5 fold excess single-stranded unmethylated 20-mer DNA certain chemical shift changes were observed for Ada-C resonances. These were grouped into 2 categories:

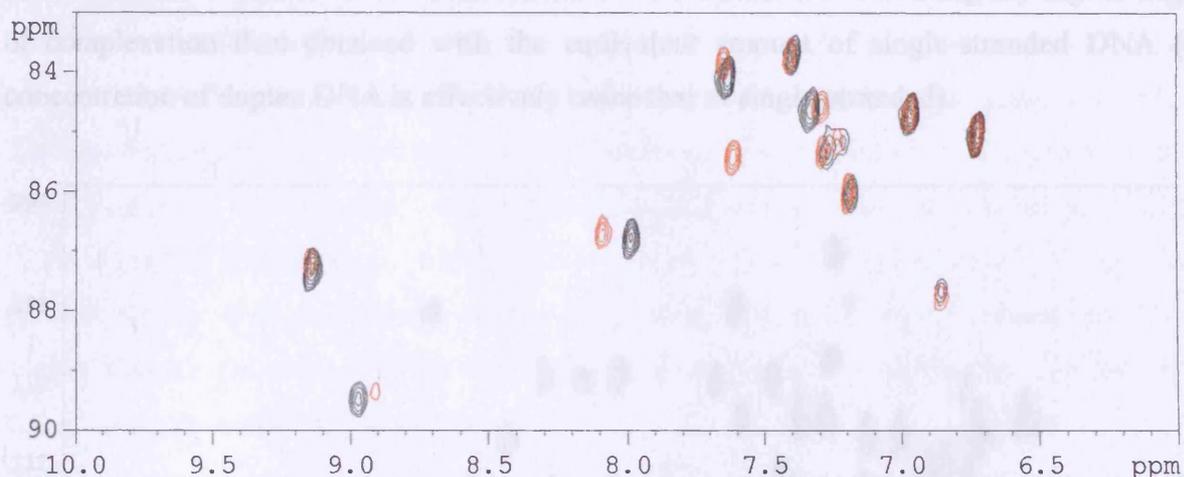
- 1) Chemical shifts which moved  $<0.15\text{ppm}$  in  $^1\text{H}$ -dimension, or  $<0.5\text{ppm}$  in  $^{15}\text{N}$ -dimension in high salt buffer and moved but did not broaden to signal loss in low salt buffer (coloured blue).
- 2) Those shifts which moved by  $>0.15\text{ppm}$  in  $^1\text{H}$ -dimension, or  $>0.5\text{ppm}$  in  $^{15}\text{N}$ -dimension in high salt buffer, or whose signals were extensively line-broadened, such that they disappeared upon binding in low salt buffer (coloured red).

This figure represents the chemical shift map of the DNA-binding face of Ada-C



**Figure 6.6** Ribbon diagram of Ada-C. Regions of the backbone whose amide resonances are unaffected by DNA-binding are coloured black. The regions whose chemical shifts are affected are coloured red and blue, in accordance with the colour scheme of the 2 categories of chemical shift change observed and depicted in Figure 6.5.

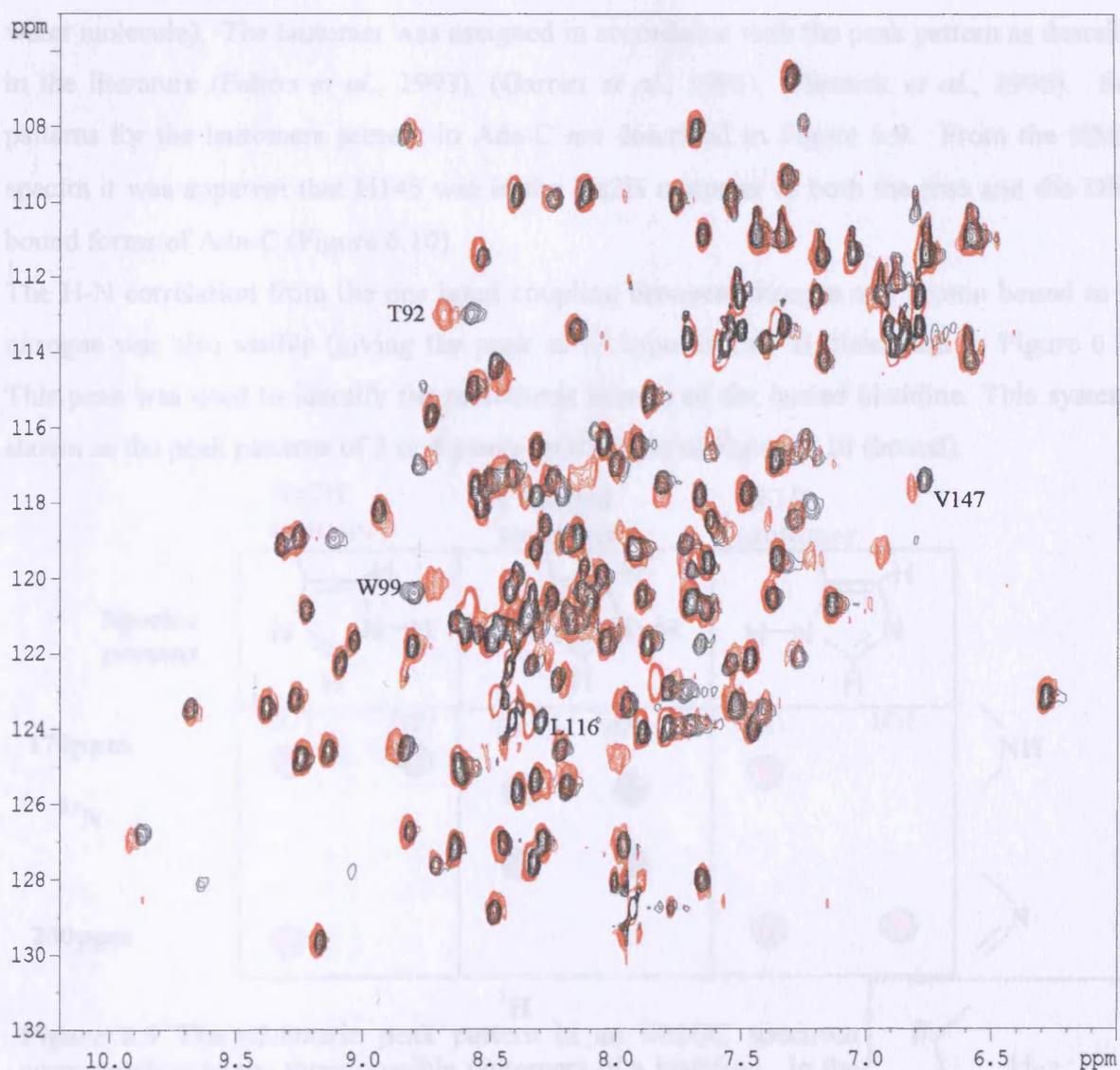
Observation of new arginine sidechain HN resonance signals upon binding to DNA implicated these sidechains in the binding interface (Figure 6.7). Unfortunately, since these sidechains are unassigned it is impossible to predict with certainty which of the arginine sidechains are involved.



**Figure 6.7** A  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of 52-62% deuterated uniformly  $^{15}\text{N}$  labeled Ada-C (black); in the presence of a 2.5 fold excess of unmethylated single-stranded 20-mer DNA (red). These spectra were recorded under the same conditions as those in Figure 6.4. This section depicts the Arg sidechain resonances, note the changes upon DNA-binding implicate arginine sidechains in DNA contacts. Which Arg sidechains are involved can only be speculated upon based on movement of backbone resonances, since Arg sidechains were not assigned.

### 6.2.2 Ada-C binding duplex DNA

70% deuterated  $^{15}\text{N}$ -labeled Ada-C was bound to duplex unmethylated 20-mer DNA in an equimolar, and two fold molar excess of DNA. The effect on the protein HSQC spectrum was similar to that of binding an equivalent ratio of single-stranded DNA, with a few extra peak shifts (Figure 6.8). These chemical shift changes were too minor to be caused by extra contacts made to the DNA. Such alterations are more attributable to a slightly higher degree of complexation than obtained with the equivalent amount of single-stranded DNA (the concentration of duplex DNA is effectively twice that of single-stranded).



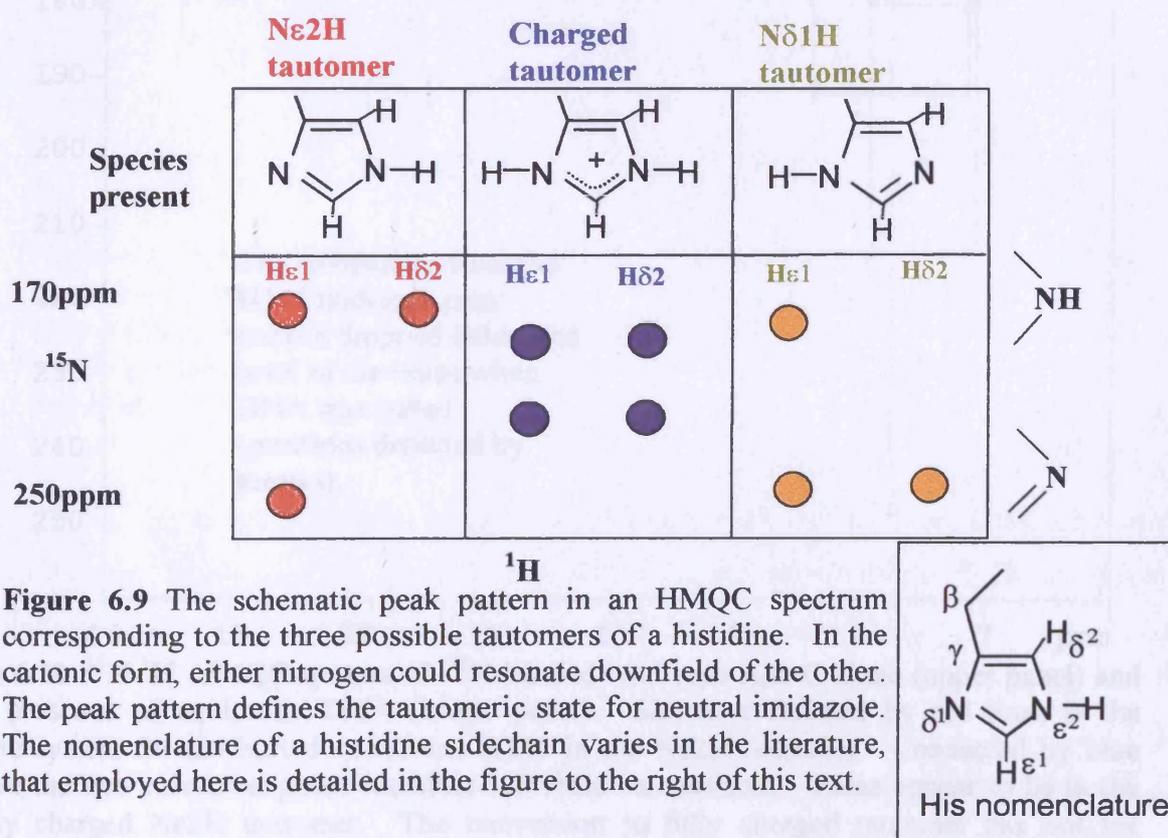
**Figure 6.8**  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of 70% deuterated wild type Ada-C bound to two fold excess single-stranded unmethylated DNA (black), and bound to two fold excess duplex DNA (red). Both spectra were recorded at 298K, pH 6.7 in low salt buffer. Although there are some differences between the two spectra, these could be explained by an increased degree of complexation with duplex DNA than single-stranded DNA.

From these results it is apparent that both double and single-stranded DNA are bound in the same manner with respect to Ada-C.

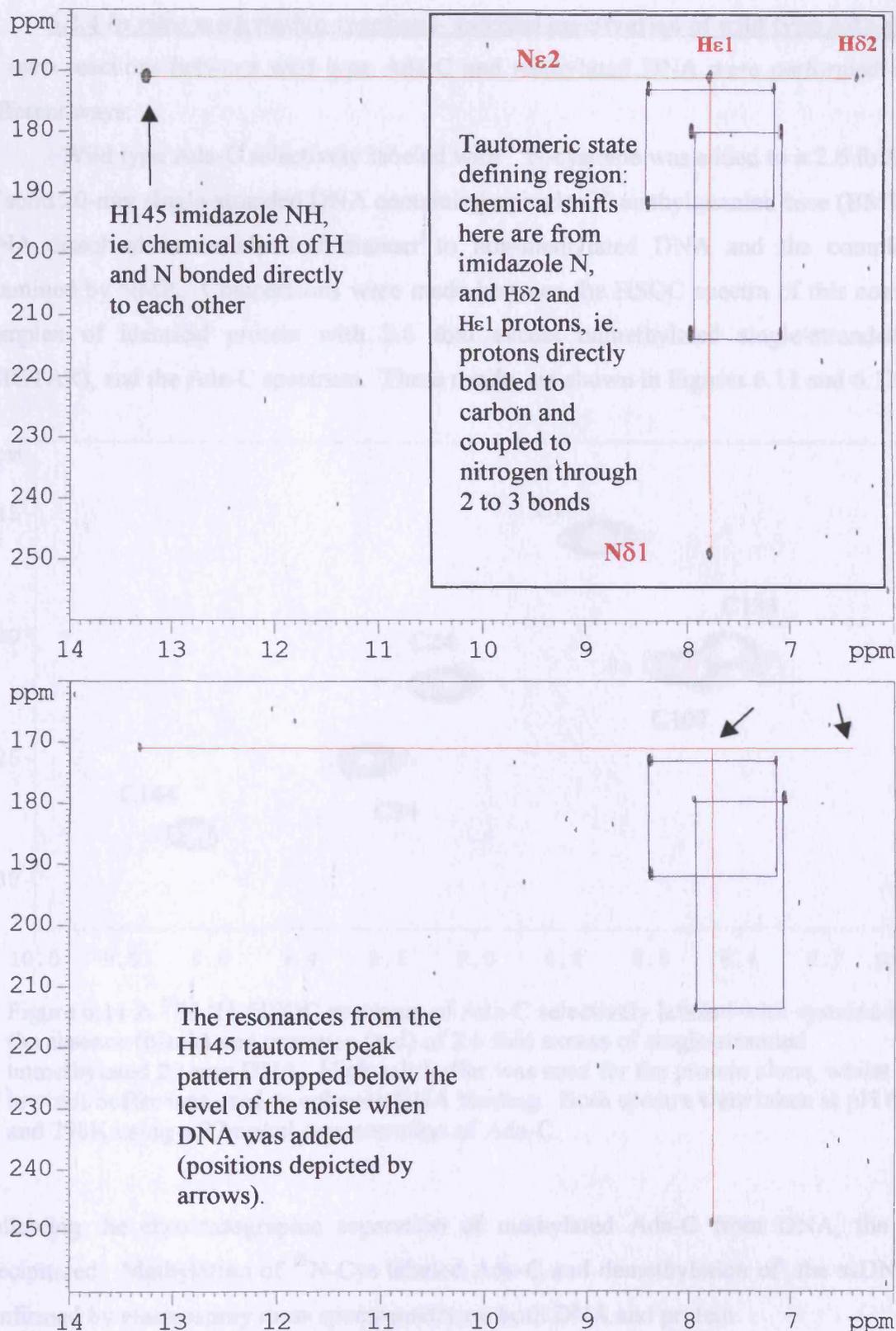
### 6.2.3 Effects of binding DNA on the tautomeric state of His 145

A series of  $^1\text{H}$ - $^{15}\text{N}$  HMQC experiments were performed on Ada-C with equimolar and 2.5 fold excess of single-stranded unmethylated DNA, or equimolar duplex 20-mer unmethylated DNA. Only one His sidechain imidazole NH was visible, indicative that this signal emanated from a buried sidechain (solvent exposed His imidazole NH would have undergone proton exchange with water, such that  $^1\text{H}$  directly bonded to  $^{15}\text{N}$  would not have been detectable). Examination of the structure of Ada-C revealed that of the three histidine residues present only the active site one is buried. This buried imidazole NH was thus assigned to H145, the histidine forming part of the catalytic tetrad (made up of C144, H145, E171, and an ordered water molecule). The tautomer was assigned in accordance with the peak pattern as described in the literature (Pelton *et al.*, 1993), (Garrett *et al.*, 1998), (Plesniak *et al.*, 1996). Such patterns for the tautomers present in Ada-C are described in Figure 6.9. From the HMQC spectra it was apparent that H145 was in the N $\epsilon$ 2H tautomer in both the free and the DNA-bound forms of Ada-C (Figure 6.10).

The H-N correlation from the one bond coupling between nitrogen and proton bound to this nitrogen was also visible (giving the peak at 13.3ppm in the  $^1\text{H}$  dimension in Figure 6.10). This peak was used to identify the tautomeric system of the buried histidine. This system is shown as the peak patterns of 3 or 4 peaks on the right of Figure 6.10 (boxed).



**Figure 6.9** The schematic peak pattern in an HMQC spectrum corresponding to the three possible tautomers of a histidine. In the cationic form, either nitrogen could resonate downfield of the other. The peak pattern defines the tautomeric state for neutral imidazole. The nomenclature of a histidine sidechain varies in the literature, that employed here is detailed in the figure to the right of this text.



**Figure 6.10**  $^{15}\text{N}$ - $^1\text{H}$  HMQC spectra of  $^{15}\text{N}$ -labeled wild type Ada-C alone (upper panel) and in the presence of equimolar DNA (lower panel). Shown connected by red lines is the histidine system for the buried active site H145 in the N $\epsilon$ 2H tautomer. Connected by blue lines are the two solvent-exposed histidine sidechain resonances. These appear to be in the partially charged N $\epsilon$ 2H tautomer. The conversion to fully charged tautomer has not yet occurred for these sidechains, although they exhibit a greater degree of charge than the sidechain of H145. The tautomeric state of H145 does not alter upon interaction with unmethylated ssDNA, or with binding duplex 20-mer DNA (results not shown). Spectra were run at 298K, pH 6.7 using 14mg/ml Ada-C.

#### 6.2.4 *In vitro* methylation reactions- suicidal inactivation of wild type Ada-C

*In vitro* reactions between wild type Ada-C and methylated DNA were performed in three different ways:

1) Wild type Ada-C selectively labeled with  $^{15}\text{N}$ -cysteine was added to a 2.6 fold excess of solid 20-mer single-stranded DNA containing a single  $\text{O}^6$ -methylguanine base (BMT). The DNA dissolved in an identical manner to non-methylated DNA and the complex was examined by NMR. Comparisons were made between the HSQC spectra of this complex, a complex of identical protein with 2.6 fold excess unmethylated single-stranded DNA (BIGTAR), and the Ada-C spectrum. These results are shown in Figures 6.11 and 6.12.

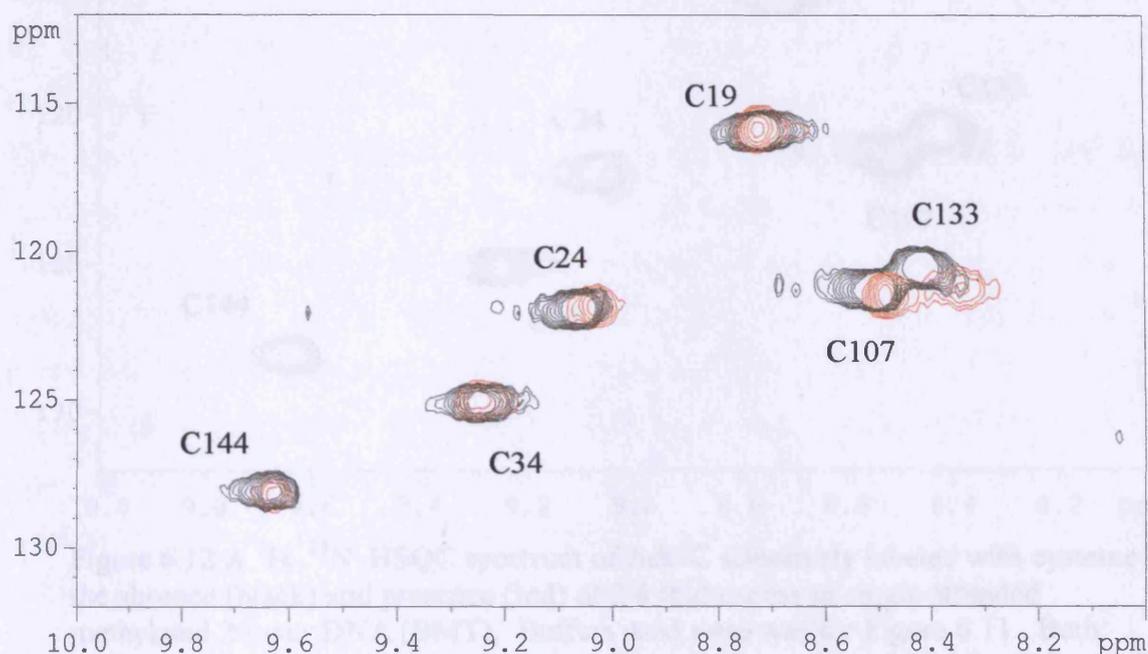


Figure 6.11 A  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum of Ada-C selectively labeled with cysteine in the absence (black) and presence (red) of 2.6 fold excess of single-stranded unmethylated 20-mer DNA. High salt buffer was used for the protein alone, whilst low salt buffer was used to enhance DNA binding. Both spectra were taken at pH 6.7 and 298K using a 37mg/ml concentration of Ada-C.

Following the chromatographic separation of methylated Ada-C from DNA, the protein precipitated. Methylation of  $^{15}\text{N}$ -Cys labeled Ada-C and demethylation of the ssDNA were confirmed by electrospray mass spectrometry on both DNA and protein.

2) To further test the effect of methylating Ada-C, a 52-62% deuterated, uniformly  $^{15}\text{N}$ -labeled wild type Ada-C was used. Single-stranded methylated BMT DNA was reacted in an equimolar ratio rather than at 2.6 fold excess. Upon addition of Ada-C to the required amount of solid BMT, instant precipitation of the protein occurred. Factors such as pH had remained unchanged, and methylation of the protein was confirmed by separating the DNA using chromatographic techniques and then subjecting this DNA to mass spectrometry.

3) It seemed probable that excess DNA was needed to keep methylated Ada-C in solution. By first adding equimolar unmethylated ssDNA and then equimolar methylated ssDNA to a 70% deuterated, uniformly  $^{15}\text{N}$ -labeled sample of Ada-C, it was possible to methylate the protein and keep it in solution. The reactions were all done in high salt buffer to further prevent precipitation. This method allowed stabilisation of methylated Ada-C such that an HSQC spectra could be recorded.

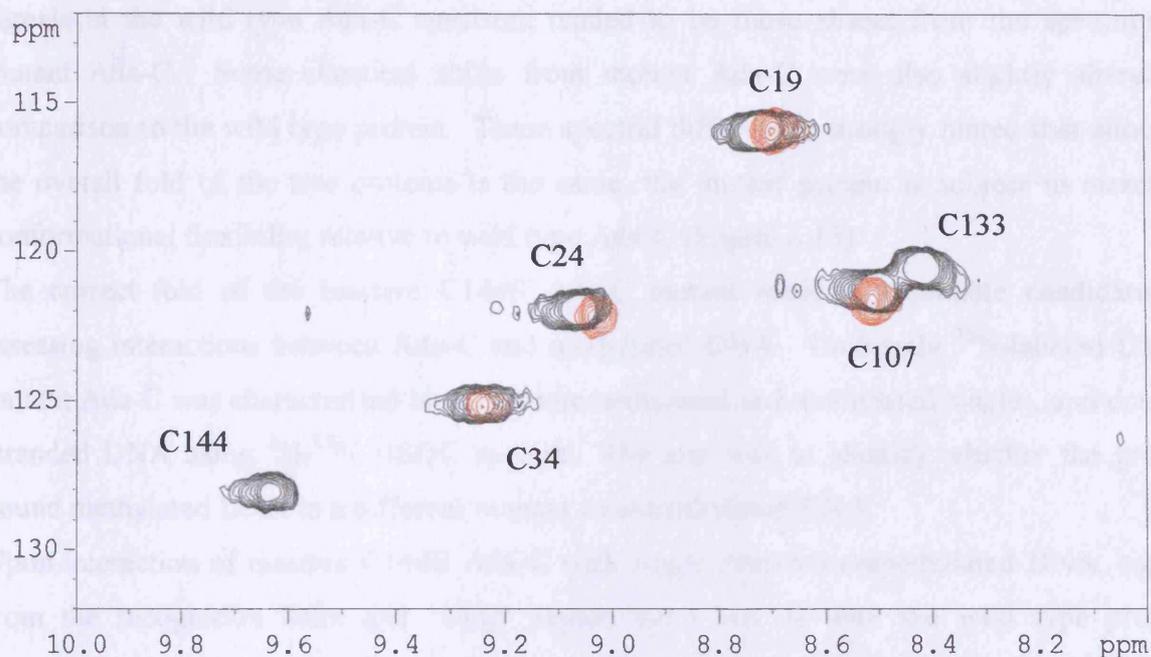


Figure 6.12 A  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of Ada-C selectively labeled with cysteine in the absence (black) and presence (red) of 2.6 fold excess of single-stranded methylated 20-mer DNA (BMT). Buffers used were as for Figure 6.11. Both spectra were taken at pH 6.7 and 298K using a concentration of Ada-C of 37mg/ml. Note the disappearance of backbone amide signals from C144 and C133.

The main conclusions drawn from the above experiments were that methylation of Ada-C introduces structural instability which easily led to precipitation. The protein could be maintained in solution in the presence of an excess of DNA, but was prone to precipitation in its absence. A  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of methylated Ada-C in complex with DNA revealed a loss of weaker signals from all regions of the protein. This suggested heightened conformational flexibility, again implying that methylation of Ada-C resulted in structural instability. A large protein conformational change, such as the movement of Helix 6 away from the protein core, should manifest itself as chemical shift alterations mapped to a particular region of the Ada-C. No such observation was made, and so it is believed that whilst maintained in solution by binding DNA, methylated Ada-C exists in a similar fold to unmethylated DNA. Methylation at C144 does however cause a weakening in the maintenance of the tertiary structure.

### **6.2.5 A Comparison of binding to methylated and unmethylated DNA by inactive C144S mutant Ada-C**

$^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were recorded on uniformly  $^{15}\text{N}$ -labeled C144S inactive mutant Ada-C alone, and in the presence of a two fold excess of methylated (BMT) or unmethylated single-stranded DNA (BIGTAR). The fold of the mutant was judged to be similar to that of the wild type, since their HSQC spectra show a high degree of identity, and the majority of chemical shift assignments can be transferred from wild type protein (Figure 6.13). The less intense signals in the wild type Ada-C spectrum tended to be those absent from the spectrum of mutant Ada-C. Some chemical shifts from mutant Ada-C were also slightly altered in comparison to the wild type protein. These spectral differences strongly hinted that although the overall fold of the two proteins is the same, the mutant protein is subject to increased conformational flexibility relative to wild type Ada-C (Figure 6.13).

The correct fold of the inactive C144S Ada-C mutant made it a suitable candidate for assessing interactions between Ada-C and methylated DNA. Uniformly  $^{15}\text{N}$ -labeled C144S mutant Ada-C was characterised binding to unmethylated and methylated single-, and double-stranded DNA using  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra. The aim was to identify whether the protein bound methylated DNA in a different manner to unmethylated DNA.

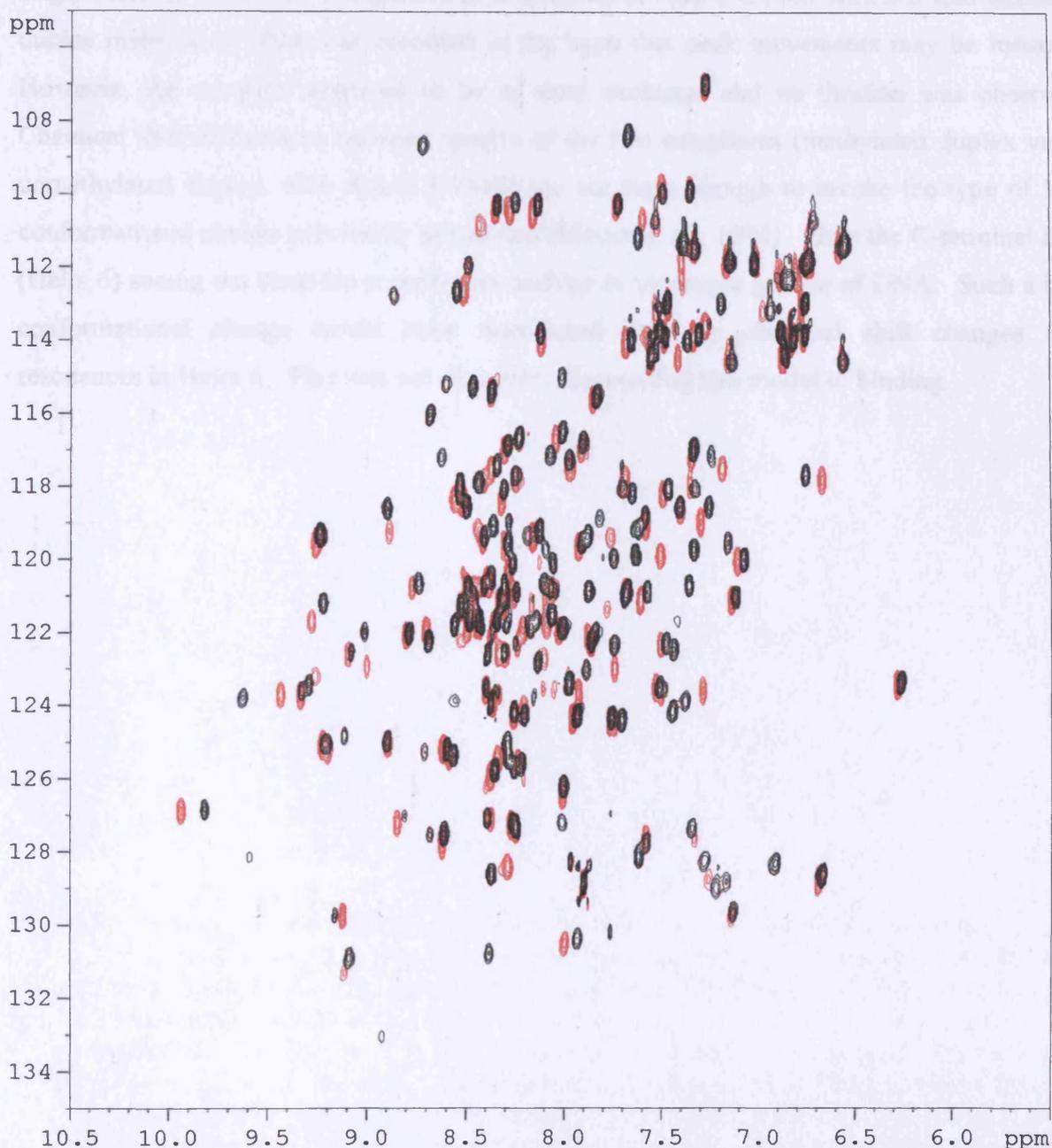
Upon interaction of inactive C144S Ada-C with single-stranded unmethylated DNA, signals from the recognition helix and 'wing' region were lost as with the wild type protein, suggesting the involvement of the same area in DNA-binding. Some extra signals, which had similar chemical shifts to wild type resonances uninvolved in the DNA-binding site also appeared. This suggested that DNA-binding stabilised the protein structure and reduced the conformational flexibility, allowing previously line-broadened signals to become visible.

The DNA-induced stabilisation of the protein structure occurred with both C144S mutant and wild type Ada-C. It therefore constitutes another observation which demonstrates the validity of using the mutant protein to study interactions with DNA.

In terms of chemical shift change, similar spectra were observed when mutant Ada-C C144S bound single-stranded methylated and unmethylated DNA. This implied that the manner of binding in both cases involved similar regions of the protein. An observed increase in line width occurred upon binding methylated ssDNA relative to unmethylated ssDNA (Figure 6.14). This could be explained by an increase in the conformational flexibility on an intermediate exchange timescale (milliseconds) of the protein complexed with methylated ssDNA, relative to when complexed with unmethylated ssDNA.

Binding of inactive C144S Ada-C to duplex methylated and unmethylated DNA was also examined. Here a 70% deuterated, uniformly  $^{15}\text{N}$ -labeled protein was employed along with higher field strength for enhanced signal:noise. As with ssDNA-binding, signals from the

recognition helix and 'wing' were extensively broadened. However, unlike with ssDNA, there was no increase in line broadening of the protein signals when in complex with

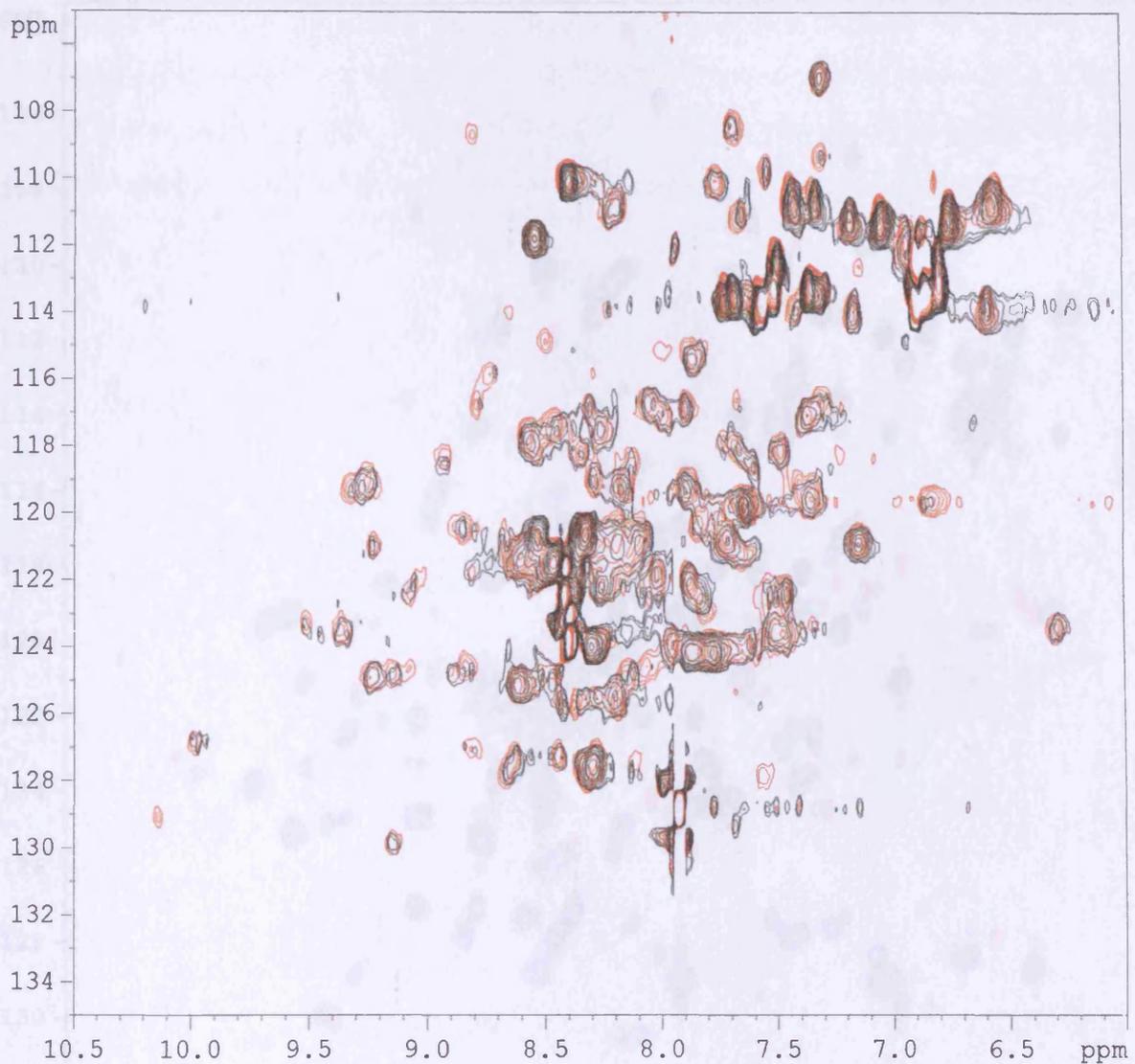


**Figure 6.13**  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of  $^{15}\text{N}$ -labeled, 70% deuterated Ada-C C144S mutant (red) and Ada-C wild type (black) for comparison. Generally the spectra are very similar, indicating that the C144S mutant maintains the correct fold. Resonances missing from the mutant spectrum probably stem from the increased conformational flexibility of the mutant relative to the wild type protein. Spectra were recorded at 298K, pH 6.7, and processed in an identical manner.

equimolar methylated duplex DNA relative to when bound to unmethylated duplex DNA.

This suggested the use of different dynamics in binding single and double-stranded methylated DNA. There were chemical shift differences between the spectra of Ada-C C144S bound to methylated, relative to unmethylated duplex DNA (Figure 6.15). Such

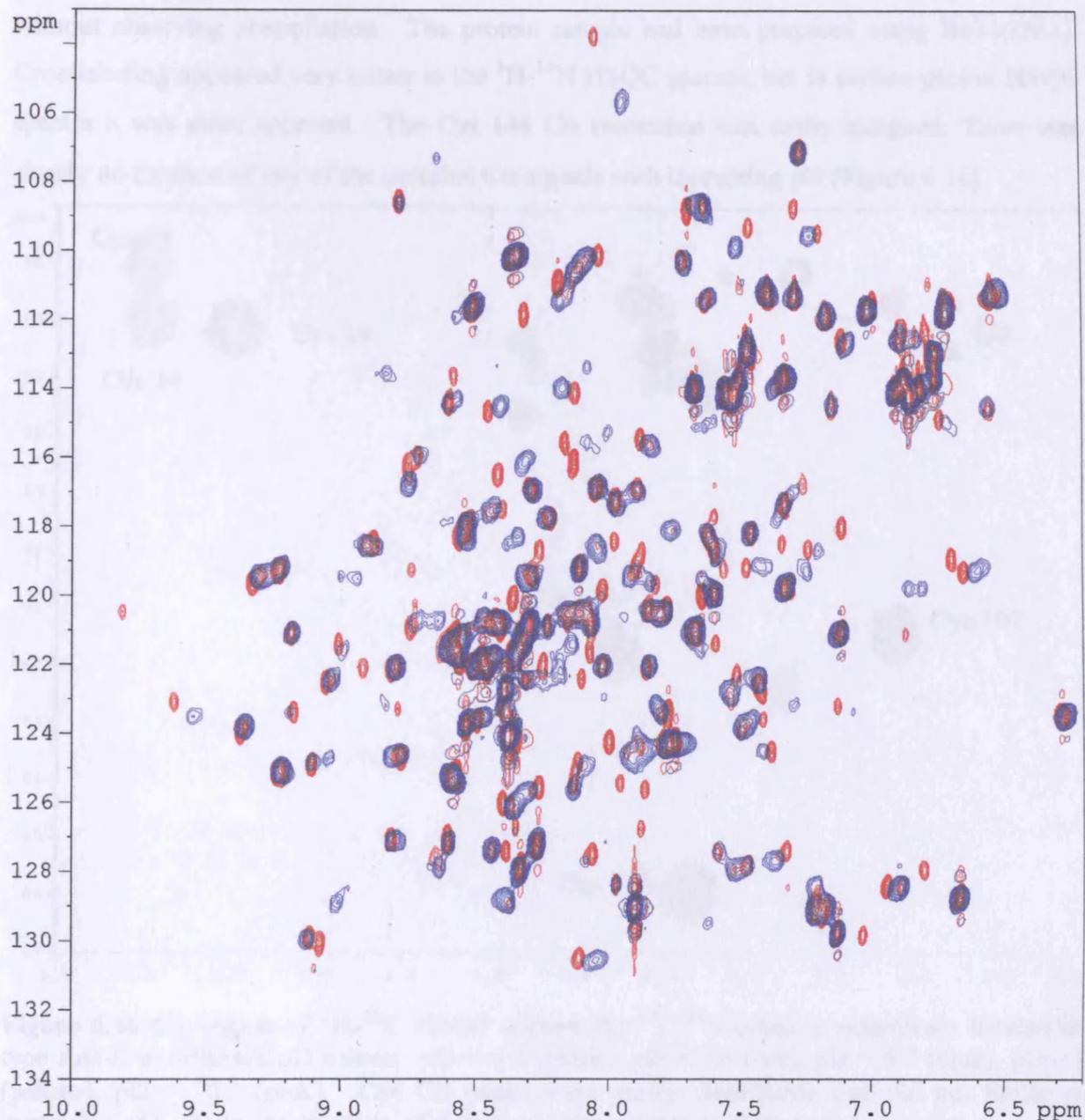
differences mainly involve the occurrence of previously unseen signals. These extra peaks cannot be assigned through transfer of assignments from the wild type protein, since their origin remains unknown. (A spectra of a complex of Ada-C C144S with a 2 fold excess of duplex methylated DNA was recorded in the hope that peak movements may be followed. However, the complex appeared to be in slow exchange and no titration was observed). Chemical shift differences between spectra of the two complexes (methylated duplex versus unmethylated duplex, with Ada-C C144S) are not large enough to invoke the type of large conformational change previously postulated (Moore *et al.*, 1994). Here the C-terminal helix (Helix 6) swung out from the protein core and lay in the major groove of DNA. Such a large conformational change would have manifested itself as chemical shift changes from resonances in Helix 6. This was not observed, discounting this model of binding.



**Figure 6.14**  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectrum of uniformly  $^{15}\text{N}$ -labeled protonated Ada-C C144S mutant in the presence 2 fold excess of: (a) single-stranded 20-mer unmethylated DNA (red). (b) single-stranded methylated 20-mer DNA (black). Low salt buffer was employed in both cases to ensure tight binding. Both spectra were recorded at 298K and pH 6.7. The concentration of Ada-C was 8mg/ml. Globally the two spectra appear identical. However the Ada-C/methylated DNA complex (black) appears to give broader linewidths, suggesting that either the complex is not as tightly bound as with unmethylated DNA, or there is enhanced protein flexibility coupled to binding methylated DNA. Either way, the structural manner in which Ada-C binds methylated single-stranded DNA appears very similar to the way it binds single-stranded unmethylated DNA.

#### 6.2.5.15. Selective labeling of Ada-C C144S

Wild type Ada-C selectively labeled with  $^{13}\text{C}$  Nucleosides was subjected to a pH titration in high salt 100%  $\text{D}_2\text{O}$  buffer (50mM phosphate, 150mM NaCl, 100mM sodium acetate, 1mM EDTA, 100%  $\text{D}_2\text{O}$ ).  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra were run at 298K and 600MHz for the same sample at different pH values, ranging from 6.5 to 11.0. Protein precipitation occurred instantly at pH values of 10 and over. pH = 9.3 was the most alkaline value measured

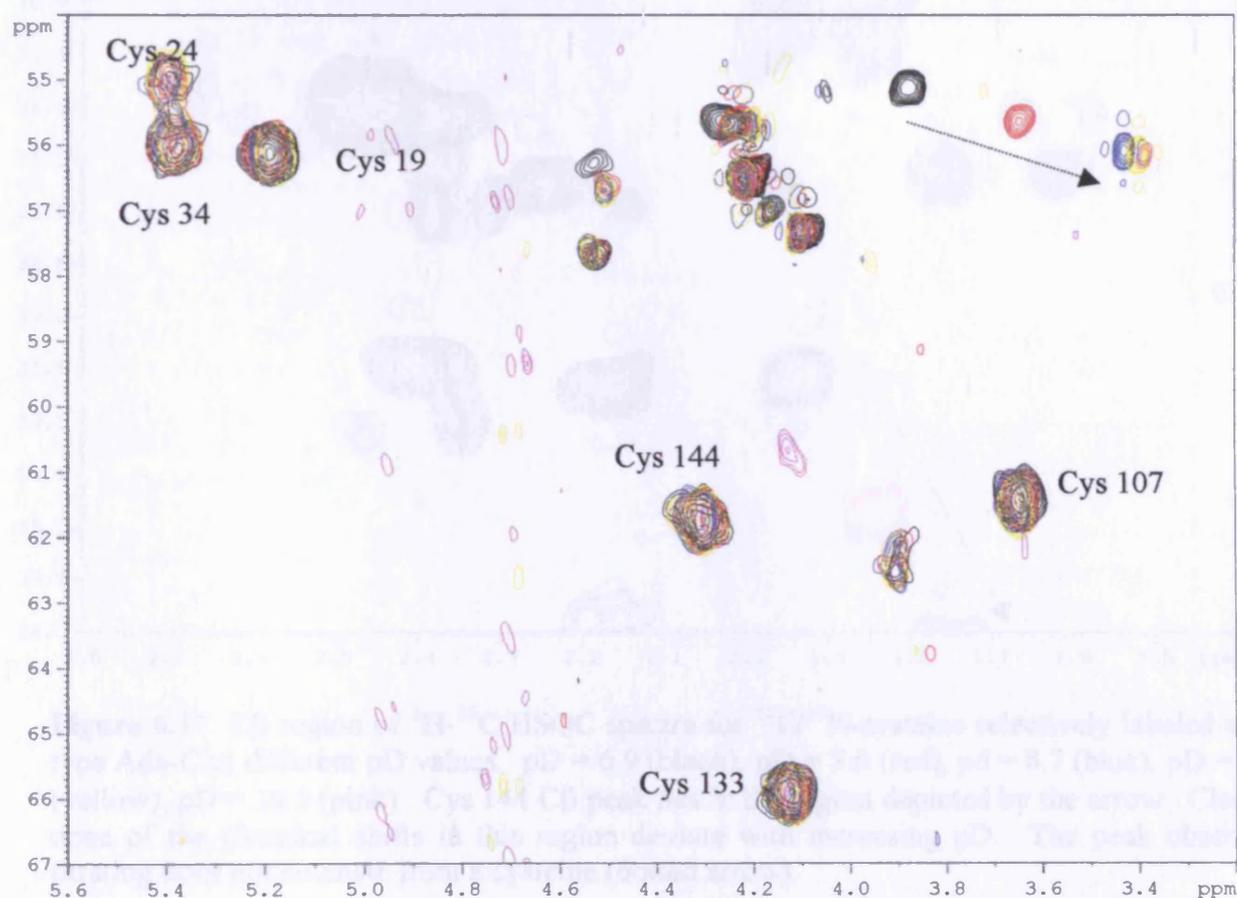


**Figure 6.15**  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra of 70% deuterated,  $^{15}\text{N}$ -labeled Ada-C C144S mutant in complex with equimolar methylated duplex DNA (red) and equimolar unmethylated duplex DNA (blue). Both spectra were recorded at 600MHz and 298K, pH 6.7 in low salt buffer, spectral processing in each case was identical. The complex of Ada-C with methylated DNA appeared to be in an slower exchange regime than the complex with unmethylated DNA.

#### **6.2.6 pK<sub>a</sub> determination of active site Cys 144**

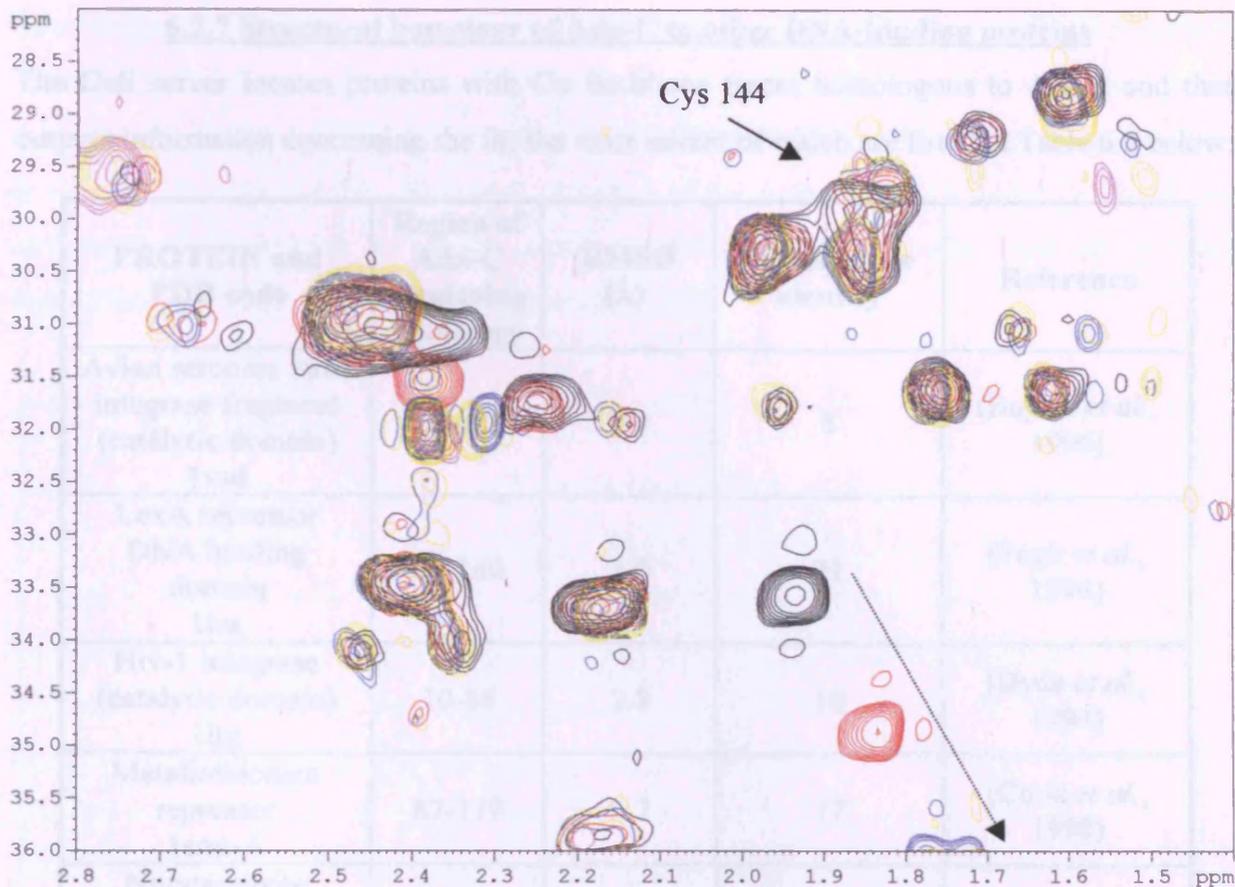
Wild type Ada-C selectively labeled with  $^{13}\text{C}/^{15}\text{N}$ -cysteine was subjected to a pD titration in high salt 100%  $\text{D}_2\text{O}$  buffer (50mM phosphate, 100mM NaCl, 100mM NaBr, 10mM DTT, 1mM EDTA, 100%  $\text{D}_2\text{O}$ ).  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra were run at 298K and 600MHz for the same sample at different pD values, ranging from 6.3 to 11.0. Protein precipitation occurred instantly at pD values of 10 and over. pD = 9.3 was the most alkaline value maintained

without observing precipitation. The protein sample had been prepared using B834(DE3). Crosslabeling appeared very minor in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra, but in carbon-proton HSQC spectra it was more apparent. The Cys 144  $\text{C}\alpha$  resonance was easily assigned. There was clearly no titration of any of the cysteine  $\text{C}\alpha$  signals with increasing pD (Figure 6.16).



**Figure 6.16**  $\text{C}\alpha$  region of  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra for  $^{13}\text{C}/^{15}\text{N}$ -cysteine selectively labeled wild type Ada-C at different pD values. pD = 6.9 (black), pD = 7.6 (red), pD = 8.7 (blue), pD = 9.3 (yellow), pD = 10.3 (pink). Cys  $\text{C}\alpha$  peaks were easily identifiable and did not titrate with increasing pD. Note the titration of a non-cysteine chemical shift occurs, serving as a control (dotted arrow).

The  $\text{C}\beta$  signal could not be unambiguously assigned due to extensive signal crowding in the  $\text{C}\beta$  region due to crosslabeling. However no resonance titrated near the suspected  $\text{C}\beta$  signal, thus precluding any titration of the active site Cys 144 sidechain in this pD range (Figure 6.17).



**Figure 6.17** C $\beta$  region of  $^1\text{H}$ - $^{13}\text{C}$  HSQC spectra for  $^{13}\text{C}/^{15}\text{N}$ -cysteine selectively labeled wild type Ada-C at different pD values. pD = 6.9 (black), pD = 7.6 (red), pD = 8.7 (blue), pD = 9.3 (yellow), pD = 10.3 (pink). Cys 144 C $\beta$  peak lies in the region depicted by the arrow. Clearly none of the chemical shifts in this region deviate with increasing pD. The peak observed titrating does not emanate from a cysteine (dotted arrow).

Other, non-cysteine C $\beta$  and C $\alpha$  chemical shifts clearly titrated with increasing pD, serving as internal controls. That the chemical shift of cysteine C $\beta$  and C $\alpha$  signals did not alter suggests that the pK $_a$  most likely lies above 9. Low pK $_a$  values of between 7 and 8 observed in other protein systems reveal themselves by clearly titrating the  $^{13}\text{C}$  chemical shift between pD 7 and 8 (Chivers *et al.*, 1997) (Wilson *et al.*, 1995).

Table 6.3 lists the results from the structural homology search of Ada-C from the Dal. Server. Rv320 is the most similar sequence (identity of the C $\alpha$  atoms in rigid body superposition), its sequence identity and region of Ada-C for which the homology refers are also given. In summary, the N-terminal region of Ada-C (the 'middle domain' of the Ada protein, consisting of the  $\beta$ -sheet and five  $\alpha$ -helices) are structurally homologous to tetragonal and tetragonal proteins. The helix-turn-helix region consisting of  $\alpha$ -helices 3, 4 and 5 is homologous to these motifs found in gene regulator proteins such as CAP.

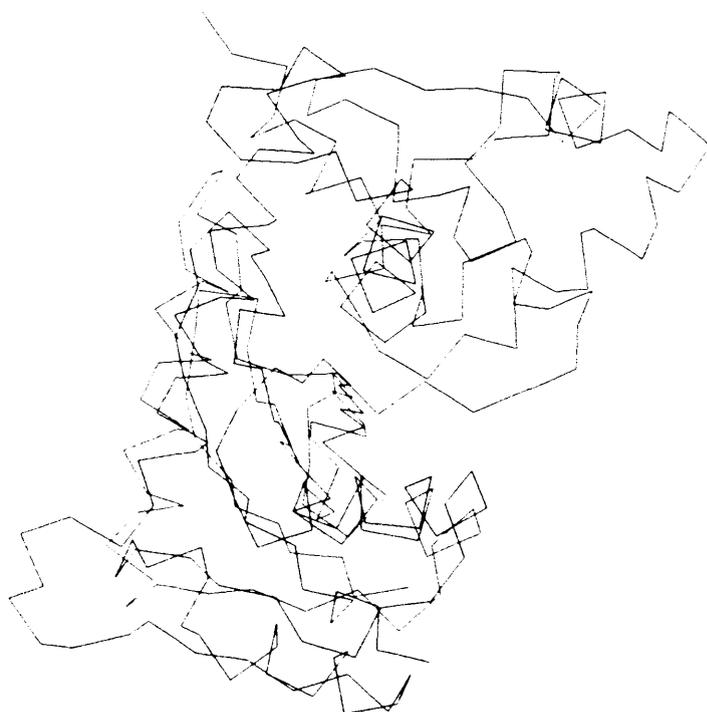
### 6.2.7 Structural homology of Ada-C to other DNA-binding proteins

The Dali server locates proteins with C $\alpha$  backbone traces homologous to Ada-C and then outputs information concerning the fit, the most salient of which are listed in Table 6.2 below.

PROTEIN and PDB code	Region of Ada-C displaying homology	RMSD (Å)	% Sequence identity	Reference
Avian sarcoma virus integrase fragment (catalytic domain) 1vsd	10-80	3.0	8	(Bujacz <i>et al.</i> , 1996)
LexA repressor DNA binding domain 1lea	88-140	2.9	21	(Fogh <i>et al.</i> , 1994)
Hiv-1 integrase (catalytic domain) 1itg	10-86	2.8	10	(Dyda <i>et al.</i> , 1994)
Metallothionein repressor 1smt-A	87-139	3.7	17	(Cook <i>et al.</i> , 1998)
Nitrate-nitrite response regulator protein, NarL 1rnl	90-143	2.1	6	(Baikalov <i>et al.</i> , 1996)
BirA biotin biosynthetic operon repressor 1bia	92-140	2.8	15	(Wilson <i>et al.</i> , 1992)
DNA catabolite gene activator protein (CAP) 2cgp-C	92-140	10.7	14	(Schultz <i>et al.</i> , 1991)
Bacteriophage muA transposase (catalytic domain) 1bco	10-80	3.1	7	(Rice <i>et al.</i> , 1995)

**Table 6.2** Results from the structural homology search of Ada-C from the Dali Server. RMSD = root mean square deviation of the C $\alpha$  atoms in rigid body superimposition, % sequence identity and region of Ada-C for which the homology refers are also given. In summary, the N-terminal region of Ada-C (the 'middle domain' of the Ada protein, consisting of the  $\beta$ -sheet and first two  $\alpha$ -helices) are structurally homologous to transposase and integrase proteins. The helix-turn-helix region consisting of  $\alpha$ -helices 3, 4 and 5 is homologous to these motifs found in gene regulator proteins such as CAP.

The integrase and transposase enzymes superimposed with a strikingly low RMSD to the  $\beta$  sheet and helices 1 and 2 of Ada-C (this region of Ada-C shall henceforth be known as the “middle sub-domain” of the Ada protein). A visual representation of one such superposition is depicted in Figure 6.18. Helices 3, 4 and 5 of the helix-turn-helix motif in Ada-C are structurally homologous to similar motifs in DNA-binding proteins involved in gene regulation, such as Catabolite activator protein (CAP).



**Figure 6.18** An illustration of the superimposition of the Avian sarcoma virus integrase fragment  $C\alpha$  trace onto that of Ada-C. The RMSD for this superimposition is 3.0 Å. Ada-C is shown in black.

It is the CAP protein which was initially chosen as a starting point on which to base a DNA-binding model for Ada-C. The reason for this was that it was the only helix-turn-helix motif protein whose DNA-bound structure had been solved. The helix-turn-helix motif of Ada-C was superimposed onto that of CAP in complex with DNA before the model was further refined. These modelling procedures and results are discussed fully in Chapter 8.

## **6.3 DISCUSSION**

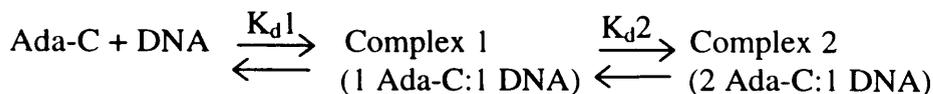
### **6.3.1 Involvement of a helix-turn-wing motif**

Chemical shift mapping strategies have been widely used to localise intermolecular binding surfaces in other DNA/protein systems (Foster *et al.*, 1998), (Holmbeck *et al.*, 1998), (Shan *et al.*, 1998), (Artz *et al.*, 1996). In the case of Ada-C, chemical shift movements upon single-stranded DNA (ssDNA) binding were placed into two categories, based on whether the resonance line-broadened so extensively as to be lost below the level of noise, or whether the chemical shift peak position simply moved but remained visible, in low salt buffer. This line-broadening occurred due to the existence of an intermediate exchange regime, in turn brought about by the affinity between Ada-C and ssDNA ( $K_d = 2.5 \times 10^6 \text{ M}^{-1}$ ).

In high salt buffer, Ada-C bound ssDNA with a lower affinity ( $K_d \sim 0.6 \text{ mM}$ ), and a fast exchange regime was observed. On this exchange timescale (sub-millisecond) no line-broadening was observed. Instead, the two categories of chemical shift change were defined as changes of either less than, or greater than 0.15ppm changes in proton, and 0.5ppm changes in  $^{15}\text{N}$  dimensions upon complexation of Ada-C with a 2.5 fold excess of DNA in high salt buffer (Figure 6.5). It was this latter category of chemical shifts which were subject to line broadening and signal loss upon complexation with 2.5 fold excess ssDNA in low ionic strength buffer.

This selective line-broadening occurred for resonances from the recognition helix of the helix-turn-helix (Helix 5), and the loop region made up of residues 149-158 (the 'wing'). It is these two regions which are therefore inferred to make up the DNA-binding surface of Ada-C. To my knowledge this is the first time a helix-turn-helix motif has been implicated in binding to single-stranded DNA. This experimental evidence lends credence to the previous postulation that this 'helix-turn-wing' motif is involved in direct contacts with DNA (Vora *et al.*, 1998). Binding of Ada-C to single and double-stranded DNA involved a similar pattern of backbone amide chemical shift changes. This implied that Ada-C utilised the same binding surface in each case. It is therefore likely that when binding to duplex DNA, Ada-C binds more predominantly to a single DNA strand. This hypothesis is backed by thermodynamic data, which indicated that two molecules of Ada-C bound duplex 20-mer DNA, whilst half the number bound single-stranded 20-mer DNA (Chapter 4). However, HSQC spectra of Ada-C complexed with duplex DNA do not display linewidths typical of a 56kDa complex (2 molecules of Ada-C and one 20-mer duplex DNA).

The binding reaction between duplex DNA and Ada-C could be summarised as follows:



$K_d2$  may differ from  $K_d1$  such that at the concentrations of Ada-C and DNA present in the NMR tube, only complex 1 forms. Complex 2 forms in the ITC sample cell because there was a higher ratio of Ada-C to DNA, thus allowing all binding sites to be filled.

A few resonances from regions of Ada-C not believed to be involved in direct contacts with DNA underwent minor chemical shift alterations (less than 0.2 ppm in  $^1\text{H}$ ) upon binding duplex DNA relative to binding ssDNA. These regions included the first helix of the helix-turn-helix motif (Helix 4), and the loop region spanning residues 89-95 (Figure 6.5). Spectral changes upon binding to DNA are insufficient to implicate any large conformational change in the protein, and are reminiscent of a simple docking interaction. Docking to DNA creates large chemical and magnetic environment changes in the binding site of Ada-C, and also transmits such changes to other regions of the protein, incurring small chemical shift perturbations. The chemical shift changes observed for Helix 4 and residues 89-95 result from structural environment perturbations transmitted from the direct DNA-binding site of Ada-C. Such transmission effects have also been observed in the case of *Trp* repressor, (Ramesh *et al.*, 1994) and *lac* repressor headpiece, (Artz *et al.*, 1996).

### **6.3.2 Selective broadening of the HTH chemical shifts, and protein dynamics**

The switch from high to low salt buffer resulted in the substitution of a fast exchange environment for one of intermediate exchange for resonances in the recognition helix (Helix 5). Typically such intermediate exchange occurs on the millisecond timescale (Kay *et al.*, 1996), (Foster *et al.*, 1997). Intermediate exchange line-broadening for signals emanating from the recognition helices of Helix-turn-helix motifs is a recurrent theme in protein/DNA complexes (Ramesh *et al.*, 1994), (Dekker *et al.*, 1993), (Vuister *et al.*, 1994) and has also been noted to occur with Ada-N (Sakashita *et al.*, 1995). Such selective signal line-broadening has been attributed to flexibility within the protein-DNA interface (Foster *et al.*, 1997), (Slijper *et al.*, 1997). Flexibility enables fine control over the stability of a molecular complex. A protein must be flexible to accommodate itself on to the DNA in an energetically favorable manner and conformational fluctuations in the binding site may confer an entropic advantage to the binding interaction (Foster *et al.*, 1997). There are many examples of mobility within protein-protein or protein-DNA interfaces. Two recent examples include flexibility of *lac* repressor headpiece (Slijper *et al.*, 1997), and *zfl-3*, (Foster *et al.*, 1997),

during interaction with DNA. Such observations enforce the notion that DNA recognition is a dynamic, not static process. A retention of some mobility upon DNA-binding would ensure some plasticity of fit, and perhaps also enable Ada-C to sense the DNA surface, and search for the O<sup>6</sup>-methylguanine lesion with greater ease than rigid docking might. Protein flexibility at the DNA-binding site would certainly render the protein more amenable to enveloping an extrahelical base than rigid docking during base-flipping.

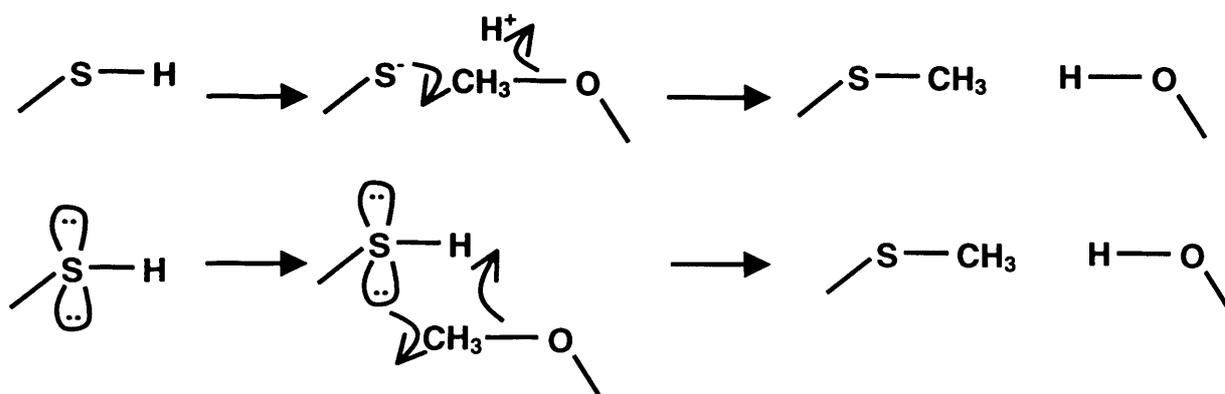
### **6.3.3 Examining the hydrogen bonding network of the active site**

Active site His 145 was in the Ne2H tautomer in the free, single, and double-stranded DNA bound states. In this tautomer, the imidazole proton is capable of involvement in a hydrogen bond with the acidic sidechain of conserved Glu 171, and forms part of a hydrogen bonding network which involves the thiol of Cys 144 (Figure 1.12, Chapter 1) (Moore *et al.*, 1994). The large conformational change postulated to occur upon DNA-binding (involving the swinging out of Helix 6) would have enabled the sidechain of H145 to rotate into a location apt to abstract the proton from the thiol of C144, to generate the reactive thiolate anion (Moore *et al.*, 1994). Were this the case, the sidechain imidazole NH chemical shift would be visibly altered as the histidine sidechain occupied an altered chemical environment. This was not the case, hence no such rotation or disruption of the H145 sidechain environment occurred upon binding unmethylated DNA.

No alteration was observed in the Ne2H tautomeric state of His 145 upon DNA-binding either. Maintenance of the H145 Ne2H tautomer, implied that no disruption to the active site hydrogen-bonding network occurred upon DNA-binding. This suggests that the sidechain of C144 remains hydrogen-bonded to H145 via an ordered water molecule, and implies that the thiol remains protonated upon binding unmethylated DNA. The pK<sub>a</sub> of the C144 thiol was above 9, since no titration was observed at pD values lower than 9.3. A consequence of this for the mechanism of methyl abstraction, is that unless the hydrogen-bonding network is altered upon binding methylated DNA, the thiol group would have to initiate nucleophilic attack whilst partially protonated.

Involvement of the C144 thiol proton in a hydrogen-bond via water to H145 would allow a partial build up of negative charge on the sulphur atom. Orientation of the thiol in the crystal structure is such that the thiol sulphur is aptly poised to abstract a methyl group from an O<sup>6</sup>meG base flipped into the active site. The negative charge build up, and orientation of the active site thiol is aided by the hydrogen-bonding network and the Ne2H tautomer of H145. There is no a priori reason why the hydrogen-bonding network should be disrupted prior to base-demethylation. The nucleophilicity of the sulphur atom may be sufficient that methyl abstraction occurs in concert with deprotonation of the sulphur. This is opposed to the

sequential scheme where the sulphur atom is deprotonated and then subject to methyl abstraction (Figure 6.19).



**Figure 6.19** (a) The sequential mechanism and (b) the concerted mechanism for the deprotonation and methylation of the C144 sulphur atom. The concerted mechanism seems more probable as it does not break the H-bonding network of the active site or introduce negative charge.

The tautomeric state of H145 upon C144 methylation could not be ascertained due to the instability of the methylated Ada-C sample. An absence of H145 imidazole NH resonance signal also meant that the tautomeric state of H145 could not be established in C144S mutant Ada-C. This absence is due to the more flexible nature of the mutant, and the concomitant line broadening and signal intensity depletion.

#### **6.3.4 Effect of self methylation on Ada-C stability- implications for adaptive response termination**

Self-methylation of wild type Ada-C resulted in structural instability. In the absence of excess DNA, this instability was sufficient to cause protein aggregation and precipitation. This precipitation was concentration dependent, as no alteration in secondary structure content, or precipitation of Ada-C was observed when methylation of Ada-C was monitored by CD (Chapter 4).

Methylation at the active site cysteine would cause a disruption of the active site hydrogen-bonding network. This ultimately results in further instability in an already conformationally flexible molecule. Disruption of the hydrogen-bonding network culminating in the breakage of the salt link between Glu 171 on Helix 6 and His 145 is believed to occur also for the human homologue of Ada-C (Ayi *et al.*, 1994), (Oh *et al.*, 1996). Cysteine 144 not only acts as methyl abstracter but also as an anchor for the structural integrity of Ada-C. Mutating this residue to alanine resulted in a mutant unstable, and prone to aggregation in concentrated conditions. Mutating it to a less hydrogen-bond-disrupting serine caused less instability. The

C144S mutant folded correctly, however the broader resonance line-widths suggested this protein had a greater conformational flexibility relative to wild type Ada-C.

Interestingly, Ada-N was also subject to structural instability once methylated (Myers *et al.*, 1994). Such increases in structural instability may render both N and C-terminal domains of Ada more susceptible to proteolytic cleavage. The adaptive response may be switched off by degradation of methylated Ada by cellular proteases. Previous suggestions of adaptive response down-regulation have included proteolysis at the hinge region between the N and C-terminal domains of Ada (Sedgwick, 1989). However, such a mechanism would not differentiate between methylated, and unreacted Ada. Proteolytic susceptibility might be enhanced by localised instability in either Ada domain. Heightened degradation of methylated Ada (or even simple domain separation) relative to unmethylated Ada would eventually result in termination of the adaptive response. This termination would occur in a methylation-dependent manner, thus not incurring unnecessary degradation of unreacted Ada which could still be used to repair DNA.

When stabilised in solution and complexed with DNA, methylated Ada-C was found not to have undergone a large conformational change. The recognition helix and wing regions were again implicated in forming the DNA-binding interface of this methylated protein as with the unmethylated form of Ada-C.

### **6.3.5 Inactive C144S mutant Ada-C binding to methylated and unmethylated DNA**

The recognition helix (Helix 5) and the wing (residues 149-158) of C144S mutant Ada-C were involved in binding both unmethylated DNA, and that containing an O<sup>6</sup>meG base. The protein/methylated duplex DNA complex appeared to be in slow exchange, contrasting to the intermediate exchange regime observed in the complex with unmethylated duplex DNA. An increased affinity of the protein for methylated DNA, and increased time in complex would account for this observation. Indeed, tighter complexation with methylated DNA is expected if cellular DNA repair is to occur efficiently amidst a mass of competing unmethylated genomic DNA. Strangely, upon binding methylated single-stranded DNA, linewidths in HSQC spectra appeared to broaden relative to binding unmethylated single-stranded DNA. This implies a difference in the dynamics by which the protein binds duplex and single-stranded methylated DNA. Perhaps the affinity for single-stranded methylated DNA is less than for duplex methylated DNA, pushing the dynamics more towards an intermediate exchange regime. The methylated duplex DNA/Ada-C C144S complex may also be more conformationally restricted than the equivalent complex with single-stranded methylated DNA. This may stem from a greater difficulty in flipping the O<sup>6</sup>meG base into the active site in single-stranded, than from more conformationally restrained duplex DNA. This

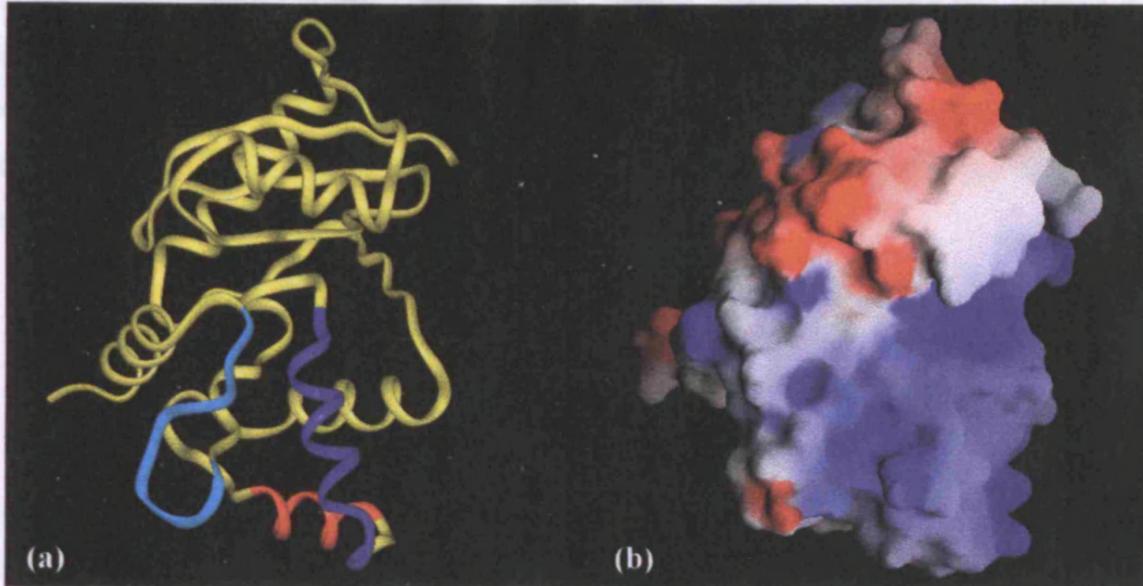
explanation would account for the fact that both Ada and its human homologue repair single-stranded DNA with lower efficiency than duplex DNA (Lindahl *et al.*, 1982), (Bhattacharyya *et al.*, 1990), (Fried *et al.*, 1996).

Differences were observed in the protein chemical shifts when C144S Ada-C bound methylated and unmethylated duplex DNA. Most of these differences entailed the emergence of signals whose origin was unknown. These signals were therefore unassignable, given the current data. Hence, it was not possible to obtain detailed structural information regarding the exact differences between the manner of binding methylated and unmethylated duplex DNA.

A large conformational change in either case can however be ruled out. Flipping the O<sup>6</sup>meG base into the active site when C144S Ada-C bound methylated duplex DNA, but not flipping bases when bound to unmethylated duplex DNA would probably be consistent with the observed differences in protein chemical shift between the two complexes.

### **6.3.6 Relationship between Ada-C, Integrase and helix-turn-helix proteins**

Use of the Dali server (Holm *et al.*, 1998) revealed a striking relationship between the 3D structure of the middle sub-domain (residues 12-92) of Ada-C and a family of integrase core domains. Helices 3, 4 and 5 in the C-terminal half of Ada-C bore structural homology to helix-turn-helix motifs in other transcription factors such as CAP. Sequence alignment searches revealed no significant sequence homology between Ada-C and integrases or helix-turn-helix proteins for the structurally homologous regions. This hints at distant evolutionary relationships between Ada-C, and the HTH proteins, and also the integrases, possibly revealing the formation of structurally equivalent features through convergent evolution. Clearly, the helix-turn-helix motif has been conserved as DNA-binding motif. Such motifs are usually involved in docking into the major groove in duplex DNA (Harrison, 1990), (Brennan, 1993). In the case of Ada-C, it was the recognition helix of this motif which was involved in binding DNA. The abundance of positive charge located on the helix-turn-helix motif makes it ideally suited to interactions with the oppositely charged phosphate backbone of DNA (Figure 6.20).



**Figure 6.20** (a) A ribbon diagram of the Ada-C polypeptide backbone coloured in yellow. Coloured in light blue is the 'wing' (residues 149-158), helix 4 is red, the recognition helix of the helix-turn-helix is dark blue. Both the wing and recognition helix are involved in binding single and double-stranded DNA. (b) A GRASP electrostatic surface potential representation of Ada-C shown in the same orientation as in the left frame. Positively charged sidechains are coded blue, negatively charged are red. Note that the helix-turn-helix motif at the bottom of each picture is highly positively charged, making it ideally suited to interactions with the sugar-phosphate backbone of DNA.

Reasons for the structural homology between Ada-C and integrase core domains were less apparent. Integrase and the homologous transposase core domains catalyse the DNA nicking and ligation through a series of steps, cofactors such as  $Mg^{2+}$  are required and the enzymes function as dimers. The D,D-35E motif conserved throughout integrases (Dyda *et al.*, 1994), (Goldgur *et al.*, 1998), (Bujacz *et al.*, 1996) and transposases (Rice *et al.*, 1995) is absent in Ada-C, as is any nicking and ligating activity. The superimposable part of the integrase core domain is the dimer interface region. Dimer-dimer interfaces frequently consist of  $\beta$ -sheets, and the possibility that the middle sub-domain of Ada-C might mediate protein-protein interactions was examined. Both human alkyltransferase and Ada are monomeric in solution (Takahashi *et al.*, 1990), (Bhattacharyya *et al.*, 1990), (Fried *et al.*, 1996) (also from observation of the elution profile of Ada-C during gel filtration chromatography in this work). However, suggestions have been made that human alkyltransferase binds in a cooperative manner along single-stranded DNA (Fried *et al.*, 1996). Although no evidence of cooperativity has been observed for Ada-C it is possible that it occurs in a similar fashion to its human homologue. Were this to be the case, inter-Ada-C contacts might be made through the middle sub-domain (residues 12-92). An alternative explanation is that the middle sub-

domain is used to mediate protein-protein contacts to the RNA polymerase  $\sigma^{70}$  subunit. Molecular biology studies indicate that Ada-C plays a role in contacting this subunit during the Adaptive response (Landini *et al.*, 1998). This, along with the Ada protein's inherent ability to promote localised duplex melting (through base-flipping), would enhance transcription. A third, more mundane possibility is that the middle domain exists simply to stabilise the C-terminal domain (residues 93-177) of Ada-C.

#### **6.4. SUMMARY**

- A helix-turn-wing motif in Ada-C was involved in binding both duplex and single-stranded DNA. Binding to duplex DNA occurred predominantly through a single strand of DNA. To my knowledge this is the first demonstration of the use of part of a helix-turn-helix motif in binding single-stranded DNA.
- In the inactive C144S mutant Ada-C, this motif was involved in binding methylated and unmethylated DNA. Some dynamic differences existed between the complexes of inactive C144S Ada-C with methylated and unmethylated DNA. A base flipping mechanism in the case of methylated, but not in the case of unmethylated DNA might explain these dynamic differences.
- The active site His 145 was protonated in the Ne2H tautomer in the free, single-stranded and double-stranded DNA bound states. The active site Cys 144 did not have a  $pK_a$  lower than 9 in the free protein. These facts suggest that the active site hydrogen-bonding network is maintained upon binding DNA. It is most probable that Cys 144 acts as a nucleophile during methyl abstraction whilst still partially protonated, and that deprotonation of the thiol is concerted with methyl transfer.
- Cys 144 a structural epicentre of Ada-C. Methylation of this residue, or mutation to a residue type incapable of hydrogen bonding, such as alanine, resulted in a disruption of the active site hydrogen bonding network. This resulted in structural instability and may be a mechanism by which the Adaptive response is terminated.
- The middle sub-domain of Ada-C (residues 12-92) showed remarkable structural homology to parts of the core domain of integrases. Reasons for this convergent evolution are unknown. It may be that the middle sub-domain is responsible for protein-protein interactions and could be the contact point to the  $\sigma^{70}$  subunit of RNA polymerase during transcription activation in the adaptive response.

# Chapter 7

## *Effects of methylation and Ada-C binding on DNA*

### 7.1 INTRODUCTION

Elucidation of the DNA repair mechanism of Ada-C requires study not only into the effects of DNA-binding on Ada-C, but also the effects of protein-binding on the structure of DNA itself. With the discovery, and recent popularisation of base-flipping (Roberts, 1995), (Cheng *et al.*, 1996) there have been suggestions that this phenomenon might also occur with Ada-C (Goodtzova *et al.*, 1997), (Verdemato *et al.*, 1998). Co-crystallisation experiments with DNA and Ada-C have so far failed, preventing high resolution structure determination of the complex. The protein's aggregated and flexible nature precludes rapid solution of the NMR structure. Alternative, less direct approaches have thus been made to study the characteristics of methylated DNA, and how it might be affected by Ada-C binding.

Alkylation at the O<sup>6</sup> position of guanine has been observed to cause minor distortions to the phosphodiester backbone of DNA structure (Patel *et al.*, 1986a), (Kalnik *et al.*, 1989b), and also to base pair buckle and twist parameters, (Sriram *et al.*, 1992), (Vojtechovsky *et al.*, 1995). Although this methylated base is, on average stacked into the double helix (Leonard *et al.*, 1990), (Patel *et al.*, 1986b), there is evidence to suggest that this stacking is weaker than with usual guanine bases (Wong *et al.*, 1992). Any weakening in stacking, combined with reduced hydrogen bonding potential in the alkylated base should lead to a reduction in stability of the duplex. The effect of methylation on duplex stability has been examined using differential scanning calorimetry (DSC) for the 20-mer oligonucleotides used in this research. Information on differential DNA stability may shed light on how Ada-C localises and binds to the lesion amidst a mass of competing genomic DNA.

The imino protons of DNA bases were probed using one-dimensional NMR to try to identify any changes that might be consistent with base flipping upon Ada-C binding. <sup>19</sup>F probes were also introduced into the DNA in the form of 5-fluorocytosine bases in order to further examine the effects of Ada-C binding. <sup>19</sup>F NMR would only examine DNA signals, and hence there would be no problems with overlapping signals from protein as with imino proton signals. <sup>19</sup>F NMR has previously been used to probe docking interactions between *lac* and  $\lambda$  *cro* repressors and their cognate operators (Rastinejad *et al.*, 1993), (Metzler *et al.*, 1989). More recently base-flipping of 5-fluorocytosine by *HhaI* Methyltransferase has been observed in the form of a ~11 ppm shift of the respective <sup>19</sup>F signal (Klimasauskas *et al.*, 1998).

Unfortunately,  $^{19}\text{F}$ -labeled  $\text{O}^6\text{meG}$  was not readily available. 5-fluorocytosine was thus used instead to probe the region of DNA directly opposite the lesion. The hope was that any flipping of the base opposite the 5-fluorocytosine might manifest itself as an observable  $^{19}\text{F}$  resonance change. Duplex 20-mer DNA was designed with a 5-fluorocytosine at positions 3 ( $\text{F}^1$ ) and 12 ( $\text{F}^2$ ) on one strand 5'-GTGGCAACF<sup>2</sup>TGTGGCAAF<sup>1</sup>CT. This could then either be base-paired with a second unmodified strand ('BIGCOMP', Section 2.1.9), or a strand containing  $\text{O}^6\text{meG}$  at position 8 ('BMT', thus base-pairing  $\text{F}^2$  with  $\text{O}^6\text{meG}$ ).  $\text{F}^1$  serves as an internal reference, providing data on the environment of nucleotides far from the  $\text{O}^6\text{meG}$  lesion.  $\text{F}^2$  probes the environment of the bases immediately opposite the  $\text{O}^6\text{meG}$  on the complementary strand.

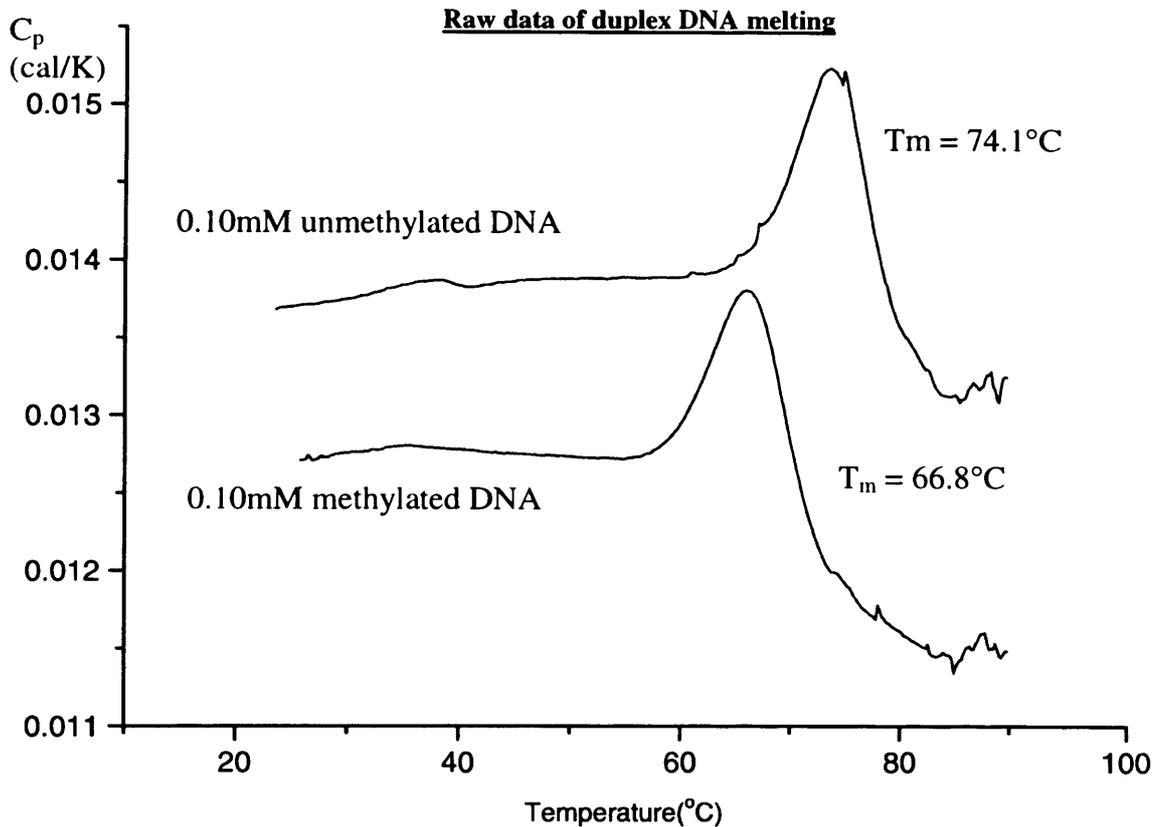
## 7.2 RESULTS

### 7.2.1 Effect of $\text{O}^6\text{meG}$ on duplex stability

A differential scanning calorimeter records the energy input required to increase the temperature of the contents of the sample cell by a given fraction. This required energy input is compared with that required by a reference cell containing buffer with no DNA. When DNA melts, extra energy input is needed to raise the temperature, since energy is utilised in the melting process. This is recorded as an increase in  $C_p$  (heat capacity at constant pressure, defined as  $\delta H/\delta T$ ). The endothermic melting process reveals itself as a positive spike of  $C_p$  as demonstrated in Figure 7.1. These melting curves revealed the melting temperature ( $T_m$ ) of methylated DNA to be  $66.2^\circ\text{C}$ , whilst that of unmethylated DNA was  $74^\circ\text{C}$ . Subtraction of the chemical baseline in this raw data gave the positive endothermic curve which was found to fit a 2-state melting model ( $X \rightleftharpoons Y$ ). Integration of this curve with respect to temperature revealed the molar  $\Delta H$  of melting. This, along with  $T_m$  could be used to determine a number of thermodynamic parameters which are summarised in Table 7.1.

	<b>Unmethylated duplex DNA</b>	<b>Methylated duplex DNA</b>
<b><math>\Delta H</math> melting (kcal/mol)</b>	120	94
<b><math>T_m</math> (<math>^\circ\text{C}</math>)</b>	$74^\circ\text{C}$ (347K)	$66.8^\circ\text{C}$ (339.8K)
<b>Number of states involved</b>	2	2
<b><math>\Delta S</math> kcal/ mol/ K</b>	0.346	0.277
<b><math>\Delta G</math> melting (kcal/mol) at 310K</b>	11.6	8.2
<b>Equilibrium constant K at 310K</b>	$1.5 \times 10^8$	$6.0 \times 10^5$

**Table 7.1** Thermodynamic parameters deduced from knowledge of  $T_m$  and  $\Delta H$ , measured for melting methylated and unmethylated 20-mer duplex DNA. Equilibrium constant  $K = [\text{duplex DNA}] / [\text{melted DNA}]$ .



**Figure 7.1** Raw data for the melting of 0.1mM methylated and unmethylated DNA. Melting temperature  $T_m$  is given

The  $\Delta G$  of melting was calculated from the  $T_m$  and  $\Delta H$  in the following way:

$$\Delta G = \Delta H - T\Delta S$$

$$\Delta G = 0 \text{ at } T_m$$

$$\text{So } \Delta H = T_m\Delta S$$

$$\text{For methylated DNA: } \Delta H/T_m = \Delta S = 94/339.8 = 0.277 \text{ kcal/ mol/ K}$$

The Gibbs Helmholtz equation:  $\Delta G_2/T_2 - \Delta G_1/T_1 = -\Delta H ( (T_2 - T_1) / (T_1 T_2) )$ , was used to determine the  $\Delta G$  at 37°C:

At 37°C for methylated DNA:

$$0/339.8 - \Delta G_1/310 = -94 ( (339.8 - 310) / (339.8 \times 310) )$$

$$\Delta G_1 = 8.2 \text{ kcal/mol -For melting DNA at 37°C}$$

The equation  $\Delta G = -RT \ln K$  was then be used to determine the equilibrium constant  $K$  between melted and duplexed forms of DNA at 37°C,

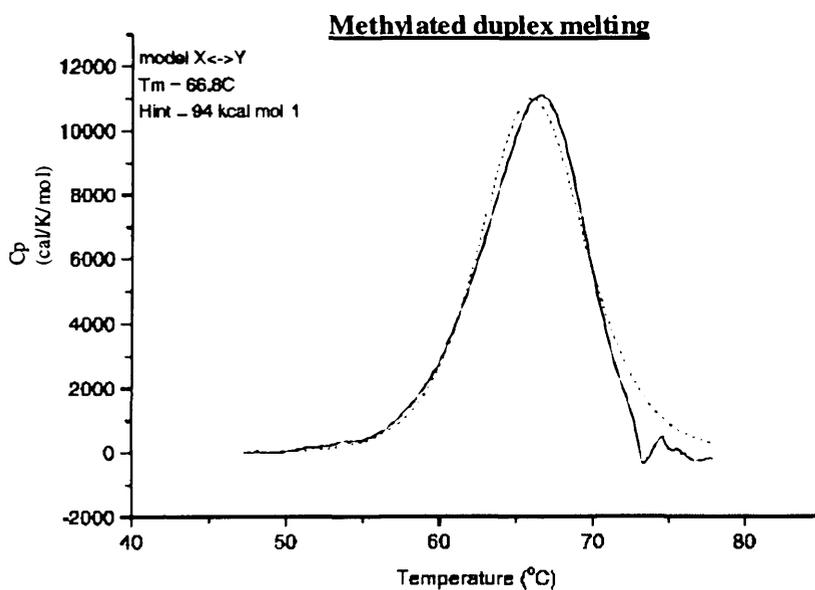
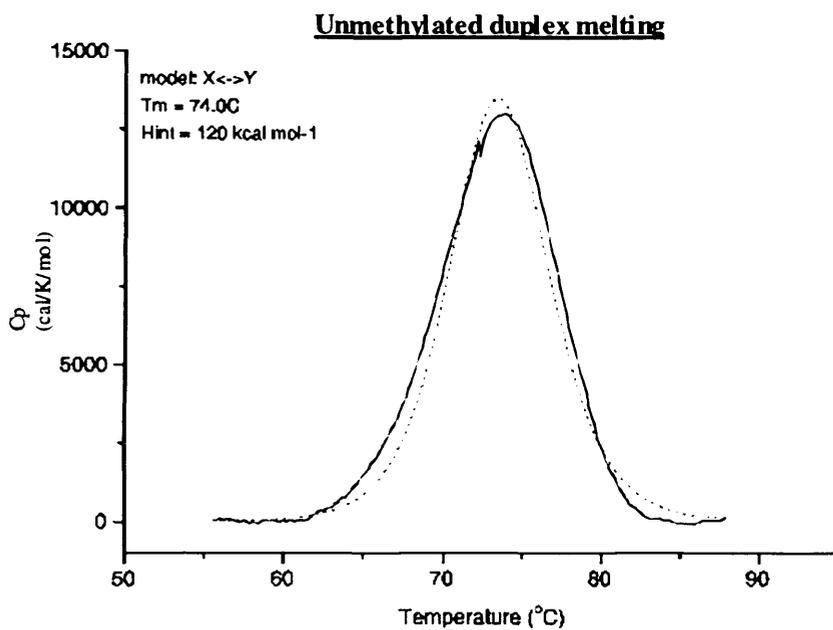
$$K = [\text{duplex DNA}] / [\text{one strand of melted DNA}]$$

$$R = 8.31451 \text{ J/K mol} = 1.9872 \text{ cal/K mol}$$

So for methylated DNA:

$$\ln K = \Delta G / -RT = -8200 / -1.9872 \times 310 = 13.3$$

$$K = 6.0 \times 10^5$$



**Figure 7.2** Following subtraction of the chemical baseline from the raw data, and correction for amount of DNA present (Figure 7.1), the endothermic curve was found to fit a 2-state melting model (X $\leftrightarrow$ Y, depicted as a dotted line). Integration of this fit with respect to temperature (the dotted line) reveals the molar  $\Delta H$  of melting (since  $C_p = \delta H/\delta T$ ). For unmethylated DNA this  $\Delta H$  was 120 kcal/mol, whilst for methylated DNA  $\Delta H = 94$  kcal/mol.

### **7.2.2 Effect of Ada-C binding on DNA imino protons**

Numerous one-dimensional spectra of single and double-stranded DNA with and without wild type and C144S mutant Ada-C were recorded. Single-stranded DNA had poor signal:noise ratios for the imino protons at 298K, superior spectra of these regions were obtained by acquiring the data at 283K. Additions of Ada-C to single-stranded DNA did not reveal any observable significant changes in the imino resonances, and so for the sake of clarity these spectra are not included.

Imino resonances of duplex DNA were more readily observable than those of single-stranded DNA, and there was no requirement to run these spectra at 283K. Spectra of methylated and unmethylated DNA revealed the subtle differences in the imino proton resonances brought about by methylation (Figure 7.3). Addition of Ada-C led to a broadening of many signals in the imino proton region. Some extra signals appeared at 13.3ppm, 11.4ppm, and around 10.5ppm upon binding equimolar Ada-C C144S mutant to methylated duplex DNA. These signals may however emanate from the protein. This is either because they were visible in the  $^1\text{H}$ - $^{15}\text{N}$  HSQC, or in the proton spectrum of the free wild-type protein. No evidence of imino protons in dramatically altered environments (which may indicate base-flipping) were observed (Figure 7.3).

ADA-C C144S

Ada-C wild type

Ada-C wild type + 2 fold excess  
unmethylated duplex DNA

Ada-C C144S with equimolar duplex  
unmethylated DNA

Ada-C C144S with equimolar duplex  
methylated DNA

Duplex unmethylated DNA

Duplex methylated DNA

14.0 13.5 13.0 12.5 12.0 11.5 11.0 10.5 ppm

**Figure 7.3** Imino proton regions of Ada-C and duplex DNA. All spectra were recorded at 298K using DNA concentrations from 0.2 to 0.6mM. The same processing parameters were applied for each spectrum.

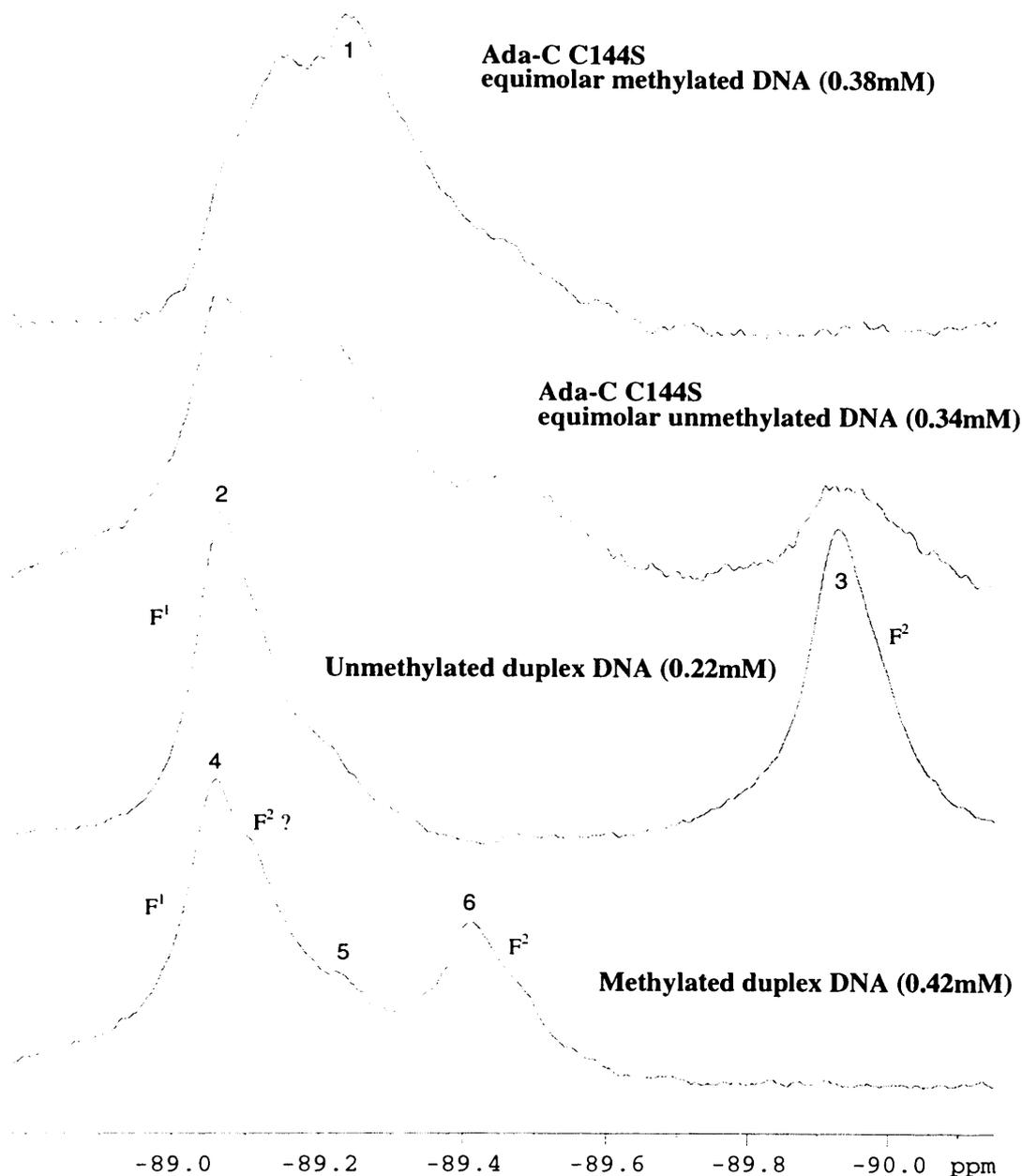
### 7.2.3 Probing the DNA using $^{19}\text{F}$ NMR

$^{19}\text{F}$  NMR proved a greater asset for studying the effects of Ada-C binding on DNA than proton NMR. The reason for this is that all signals emanated from DNA, there were no contaminating signals from the protein. The oligonucleotide used, possessed two 5-fluorocytosines,  $\text{F}^1$  and  $\text{F}^2$ .  $\text{F}^2$  base paired with  $\text{O}^6\text{meG}$  in the methylated duplex and normal

guanine in the unmethylated duplex.  $F^1$  paired with guanine in both DNA duplexes (Figure 7.4). One-dimensional  $^{19}\text{F}$  spectra of methylated and unmethylated duplex DNA enabled the assignment of the  $F^1$  and  $F^2$  signals, based on the assumption that only the chemical shifts of  $F^2$  should differ between the two DNA types (Figure 7.4). Interestingly, the  $F^2$  signal in methylated DNA was approximately half the intensity of  $F^2$  in unmethylated DNA. The line width at equal intensity for both these peaks was identical, suggesting that in methylated DNA, the  $F^2$  signal observed was in slow exchange between two chemical environments. The signal representing the second environment of the  $F^2$  fluorine was believed to be the shoulder of the  $F^1$  signal (shoulder on peak 4, Figure 7.4). Increasing the temperature from 298K to 308K allowed the resolution of this postulated extra  $F^2$  signal (peak 2, Figure 7.5). The lifetime that  $F^2$  exists in each environment for was likely to be milliseconds or longer (Klimasauskas *et al.*, 1998).

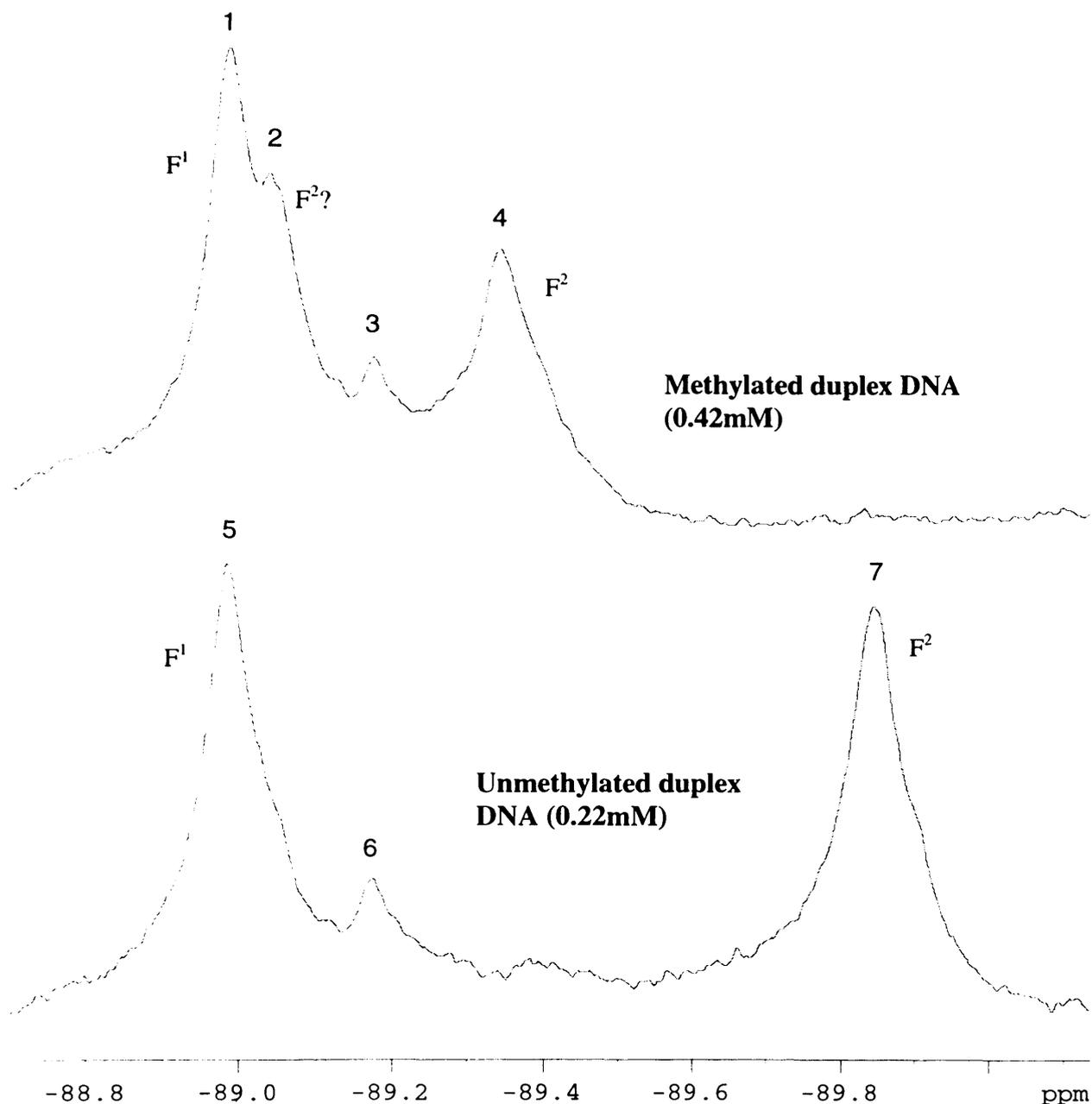
$T_2$  relaxation times obtained at 308K indicated that the three most intense signals from methylated DNA, and the two most intense signals from unmethylated DNA had similar relaxation times of approximately 13 milliseconds (Figure 7.5). This suggested that these main signals all originate from 5-fluorocytosines that formed part of a DNA duplex. The less intense signal at  $-89.18\text{ppm}$  (peaks numbered 3 and 6 in Figure 7.5) had an increased  $T_2$  time and was therefore believed to be an impurity. In the case of free methylated duplex DNA, this reduced  $T_2$  relaxation rate is visually displayed in Figure 7.6. This impurity may be single-stranded fluorinated DNA which failed to anneal correctly. Unfortunately, no spectra of single-stranded fluorinated DNA were recorded due to material and time constraints, so this could not be verified.

Upon binding equimolar inactive C144S Ada-C, both methylated and unmethylated DNA exhibited  $F^2$  chemical shift changes, and smaller  $F^1$  chemical shift changes. The complex with methylated DNA revealed a different  $^{19}\text{F}$  spectrum to the complex with unmethylated DNA (Figure 7.4). This may be partially explained by weaker binding to unmethylated DNA, and hence the incomplete transfer of the  $F^2$  signal to its new chemical shift. Such weaker binding was also deduced to occur from monitoring the protein chemical shifts (Chapter 6). Attempts to shift the spectra of unmethylated DNA into that for the fully bound form were made by increasing the Ada-C concentration to 1.9 times that of DNA. This however, resulted in protein precipitation and sample loss. An identical increase in Ada-C amount compared to methylated DNA did not result in protein precipitation, again indicating that binding to methylated DNA occurred with higher affinity.  $T_2$  relaxation parameters were not measured for the equimolar complex of Ada-C with unmethylated DNA, and subsequently could not be measured for the 1.9 Ada-C : 1 unmethylated DNA complex (due to precipitation).



Methylated DNA		Unmethylated DNA	
Peak Number	T <sub>2</sub> (milliseconds)	Peak number	T <sub>2</sub> (milliseconds)
1	6.08 ± 1.11	2	5.80 ± 0.25
4	4.60 ± 0.08	3	4.60 ± 0.31
5	25.95 ± 8.42		
6	5.69 ± 0.49		

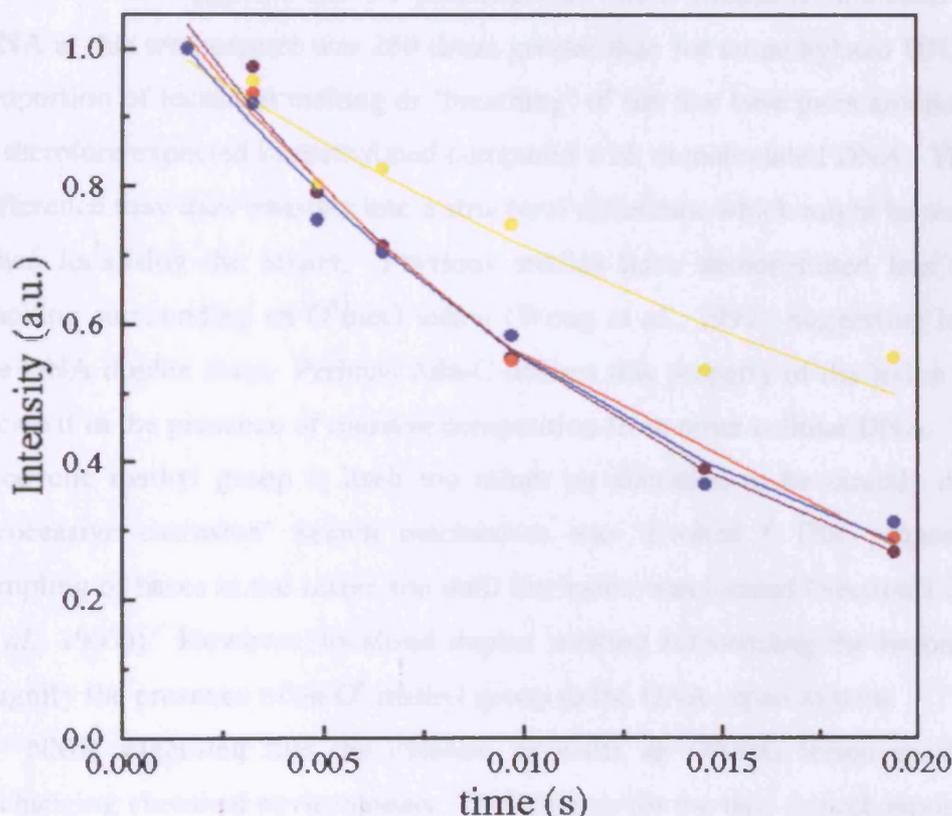
**Figure 7.4** One-dimensional <sup>19</sup>F spectra of methylated and unmethylated duplex DNA. T<sub>2</sub> relaxation times for labeled signals are tabulated. Peak 1 must represent a signal from unbound DNA, in a sample of equimolar Ada-C C144S with duplex methylated DNA (because its T<sub>2</sub> value is the same as for free DNA). T<sub>2</sub> relaxation times from complexed DNA signals could not be measured due to the rapid relaxation rate. Peaks 2 and 3 are from F<sup>1</sup> and F<sup>2</sup> signals respectively in free unmethylated duplex DNA. Peaks 4-6 are from F<sup>1</sup> and F<sup>2</sup> signals in free methylated duplex DNA. All spectra were recorded at 298K on a Bruker DRX 600 spectrometer. The oligonucleotide sequence is: 5'-TGGCAACF<sup>2</sup>TGTGGCAAF<sup>1</sup>CT  
3'-ACCGTTGMACACCGTTGGA  
where F<sup>1</sup> and F<sup>2</sup> represent 5-fluorocytosine. F<sup>2</sup> base pairs with O<sup>6</sup>meG (M) in the methylated duplex, and G in unmethylated DNA. The buffer used was: 10% D<sub>2</sub>O, 50mM Phosphate, 10mM DTT, 1mM EDTA, 10mM NaCl, pH 6.5.



Methylated DNA		Unmethylated DNA	
Peak Number	T <sub>2</sub> (milliseconds)	Peak number	T <sub>2</sub> (milliseconds)
1	13.18 ± 0.78	5	14.64 ± 1.81
2	13.71 ± 0.73	6	33.38 ± 7.21
3	26.11 ± 3.49	7	12.49 ± 2.50
4	14.11 ± 0.87		

**Figure 7.5** One-dimensional <sup>19</sup>F spectra of free methylated and unmethylated duplex DNA at 308K. T<sub>2</sub> relaxation times in milliseconds are tabulated for each numbered signal. Since this T<sub>2</sub> time is approximately twice that for the main signals, these signals may emanate from single-stranded DNA. Note the resolution of peak 2 by increasing the temperature to 308K from 298K (compare to shoulder on peak 4 in Figure 7.4). Peak 2 is thought to be the extra F<sup>2</sup> signal brought about by the slow exchange between two distinct chemical environments within the DNA duplex. This manifests itself as two signals, peaks 2 and 4. Peaks 3 and 6 are believed to be an impurity, perhaps single-stranded DNA due to their slower relaxation times.

The 1.9 Ada-C : 1 methylated DNA complex was subjected to  $T_2$  relaxation experiments at 298K.  $T_2$  relaxation times for complexed DNA signals were unobtainable due to the high relaxation rate. Only the  $T_2$  value of one peak was obtained, and this was approximately equivalent to the  $T_2$  values measured from the free DNA signals, suggesting it represented a signal from uncomplexed DNA (Figure 7.4).



**Figure 7.6**  $T_2$  relaxation plots of signal intensity against relaxation delay  $\tau$  for methylated duplex DNA at 308K. The signals, numbered in accordance with Figure 7.5 are coloured: 1 = purple; 2 = red; 3 = yellow; 4 = blue. The yellow signal clearly has a reduced  $T_2$  relaxation rate in comparison with the other signals present. It is believed that this signal arises from an impurity, the most likely form of which is unannealed single-stranded DNA.

### 7.3 DISCUSSION

Differential Scanning Calorimetry demonstrated that methylated 20-mer duplex DNA (containing one O<sup>6</sup>meG) melted at 7.2°C lower than unmethylated DNA of the same sequence. This trend is identical to that obtained from thermodynamic studies performed on other oligonucleotides (Gaffney *et al.*, 1989); (Gaffney *et al.*, 1984). The  $\Delta H$  of melting was 26 kcal/mol greater for unmethylated DNA. Combining this data through thermodynamic analysis allowed the calculation of equilibrium constants between melted and duplex DNA at 37°C. These suggested that the proportion of totally melted to unmelted methylated duplex DNA at this temperature was 250 times greater than for unmethylated DNA. An even higher proportion of localised melting or 'breathing' of the few base pairs around the O<sup>6</sup>meG lesion is therefore expected in methylated compared with unmethylated DNA. This thermodynamic difference may thus translate into a structural difference which might be recognised by Ada-C when localising the lesion. Previous studies have demonstrated less than average base stacking surrounding an O<sup>6</sup>meG lesion (Wong *et al.*, 1992), suggesting less maintenance of the DNA duplex state. Perhaps Ada-C utilises this property of the lesion in order to rapidly locate it in the presence of massive competition from other cellular DNA. The presence of an exocyclic methyl group is itself too minor an alteration to be directly detectable, unless a 'processive extrusion' search mechanism was invoked. This proposed the individual sampling of bases in the active site until the lesion was located (Section 1.2.2.8, and (Verdine *et al.*, 1997)). However, localised duplex melting surrounding the lesion would effectively magnify the presence of an O<sup>6</sup> methyl group to the DNA repair system.

<sup>19</sup>F NMR suggested that the cytosine opposite an O<sup>6</sup>meG lesion existed in two slowly exchanging chemical environments. The lifetime for the base in each environment is likely to be milliseconds or more (Klimasauskas *et al.*, 1998). This lifetime in each environment would fall within the allowed range of ~10msecs lifetime for individual G-C base pairs in unmethylated DNA (Gueron *et al.*, 1995). One might therefore speculate that these two environments are created by alternate stacking arrangements of the O<sup>6</sup>meG base caused by spontaneous DNA breathing. The presence of two arrangements certainly suggests a localised loosening of the duplex structure surrounding the O<sup>6</sup>meG compared with unmethylated DNA. Such unstacking would certainly aid rendering the base extrahelical, were such a process to be employed by Ada-C during the repair mechanism.

Although the methylated duplex appeared to be bound with higher affinity than unmethylated, both DNA types exhibited altered F<sup>2</sup> chemical shifts of less than 1ppm. The magnitude of such chemical shift changes is consistent with those observed during the docking of proteins to DNA (Metzler *et al.*, 1989), (Rastinejad *et al.*, 1993). The large (~11ppm) chemical shift changes associated with flipping a base out of the duplex (Klimasauskas *et al.*, 1998) were not

observed in the case of Ada-C. This was expected since the environment of the 5-fluorocytosine opposite the O<sup>6</sup>meG will not alter that drastically even if the lesion itself is rendered extrahelical. Attempts were made to convert all unmethylated DNA to the complexed form by addition of excess Ada-C. These however caused the protein to precipitate. It thus remains to be established whether the completely complexed unmethylated DNA would have exhibited an identical <sup>19</sup>F spectrum to that of complexed methylated DNA. Although they do not provide direct evidence for the flipping of O<sup>6</sup>meG, such differences in the <sup>19</sup>F spectra observed for Ada-C with methylated and unmethylated DNA do not preclude base flipping activity with methylated DNA. Flipping O<sup>6</sup>meG, but not unmodified bases would explain the subtle differences observed between the protein spectra of Ada-C C144S complexed with methylated and unmethylated duplex DNA. From a biological stand point, repair would be more rapidly effected if Ada-C targeted the lesion directly, without the excessive base-flipping invoked by a 'processive extrusion' mechanism (Verdine *et al.*, 1997). (Here the protein binds to DNA and randomly flips out a base, it then slides along the DNA, alternately flipping bases out until the lesion is located).

#### **7.4 SUMMARY**

Methylated DNA is less thermodynamically stable in duplex form, than unmethylated DNA. There is evidence to suggest localised conformational distortions surrounding the O<sup>6</sup>meG lesion which may be caused by less stacking into the duplex. Such structural differences may be targeted by Ada-C when localising the lesion for repair. From an empirical view point, base-flipping appears necessary in order to repair the O<sup>6</sup>meG lesion without activating a large conformational change in Ada-C. Base-flipping O<sup>6</sup>meG, although not proven, is not precluded from the data obtained here. Flipping O<sup>6</sup>meG, but not other bases would be consistent with the differences observed in the protein environment between Ada-C C144S complexed with methylated and unmethylated DNA (described in Chapter 6).

Whether base-flipping is employed when scanning the genome for lesions is unknown. This latter strategy would be less efficient at lesion localisation than a strategy of locating the lesions based on their less-stacked nature, and subsequently flipping them.

# Chapter 8

## *Modelling how Ada-C binds DNA & general discussion*

### 8.1 INTRODUCTION

The main aim of this thesis was to determine the manner by which Ada-C binds DNA. Two models had previously been proposed which allowed Ada-C to bind DNA with close proximity between active site methyl acceptor Cys 144 and the O<sup>6</sup>meG lesion. The first model invoked an opening up of the protein core, by swinging the C-terminal helix (Helix 6) away from the protein into the major groove of DNA (Moore *et al.*, 1994). This allowed the exposure of the ordinarily buried Cys 144 to the O<sup>6</sup>meG stacked into DNA such that methyl transfer could be affected. More recently an alternative model was devised which employed base flipping as a means of rendering the O<sup>6</sup>meG lesion accessible to the Cys 144 thiol (Vora *et al.*, 1998). In this latter model, Ada-C uses a helix-turn-wing motif to bind the DNA. The recognition helix of the helix-turn-helix (Helix 5) lies in the major groove of DNA, whilst the loop region formed from residues 149-155 (the 'wing') contacts the DNA backbone. No large conformational change occurs in the protein, and the O<sup>6</sup>meG lesion is flipped out of the double helix and into the buried active site through a hole in the surface of Ada-C.

Neither of these models were based on experimentally determined structural evidence. This evidence was gathered during this thesis work and has led to the exclusion of the first model by Moore *et al.* The regions predicted to be involved in binding DNA by Vora *et al.* were found to be correct. However, prior to the publication of the latter paper attempts were made to model Ada-C on to DNA using the helix-turn-helix motif of Catabolite Activator Protein (CAP) as a structural guide. CAP shared a high degree of structural homology with Ada-C in the helix-turn-helix motif, and the structure of this protein had been solved in complex with DNA. It thus provided an excellent start point for examining how Ada-C might dock to DNA. A non-productive binding model was made, where Ada-C bound DNA in a manner analogous to CAP, but the stacked O<sup>6</sup>meG was inaccessible to Cys 144. This was turned into a productive binding model by flipping O<sup>6</sup>meG out of the duplex and into the active site. This model was then compared to the Vora model, and found to be very similar. The Vora model fitted the experimental data slightly better however, since in this model Ada-C only contacts a single-strand of duplex DNA. NMR revealed no differences in the manner by which Ada-C bound single and double-stranded DNA, suggesting that binding in both cases occurred predominantly to a single-strand. The model based on CAP had some minor interactions with the second DNA strand of the duplex, and was therefore less satisfactory.

What follows is a description of how the DNA-binding model based on CAP was built up, and a more detailed comparison with the Vora model. Mechanisms of lesion localisation, flipping and demethylation are also discussed.

## **8.2 THE DNA-BINDING MODEL**

NMR revealed no displacement of Helix 6 in inactive C144S mutant Ada-C upon binding either unmethylated or methylated DNA, or in wild type Ada-C when it bound unmethylated DNA. This precludes the postulated large conformational change in Ada-C, where Helix 6 swings away from the protein core upon DNA-binding (Moore *et al.*, 1994). A novel model was thus required. Structural analysis of binding suggested the involvement of the recognition helix (Helix 5) of the helix-turn-helix motif, and a 'wing' region (defined as residues 149-158, and therefore longer than the wing of Vora *et al.*). This NMR analysis also indicated that identical regions of Ada-C were involved in binding to both duplex and single-stranded DNA. ITC data revealed a doubling in stoichiometry when Ada-C bound duplex as opposed to single-stranded DNA. Taken together, this ITC and NMR data suggested that Ada-C binding occurred predominantly to a single strand of the DNA duplex. Any DNA-binding model should therefore employ the recognition helix and the 'wing' in the binding interface and should involve docking to a single DNA strand only in duplex DNA. Uracil-DNA glycosylase, which repairs both duplex and single-stranded DNA also interacts predominantly with a single DNA strand in the duplex (Slupphaug *et al.*, 1996). To my knowledge Ada-C is the first example of involvement of a helix-turn-helix motif in the binding of single-stranded DNA.

Due to the high degree of structural homology between Ada-C and other helix-turn-helix proteins, it was anticipated that the binding mechanism to DNA may be similar. Helix-turn-helix proteins employ the recognition helix to make contacts with the major groove of DNA (Harrison, 1990) (Brennan, 1993). A flexible 'wing' often makes contacts to the phosphate backbone and minor groove, although recently it has also been observed to form protein-protein contacts and not contact DNA (Littlefield, 1999).

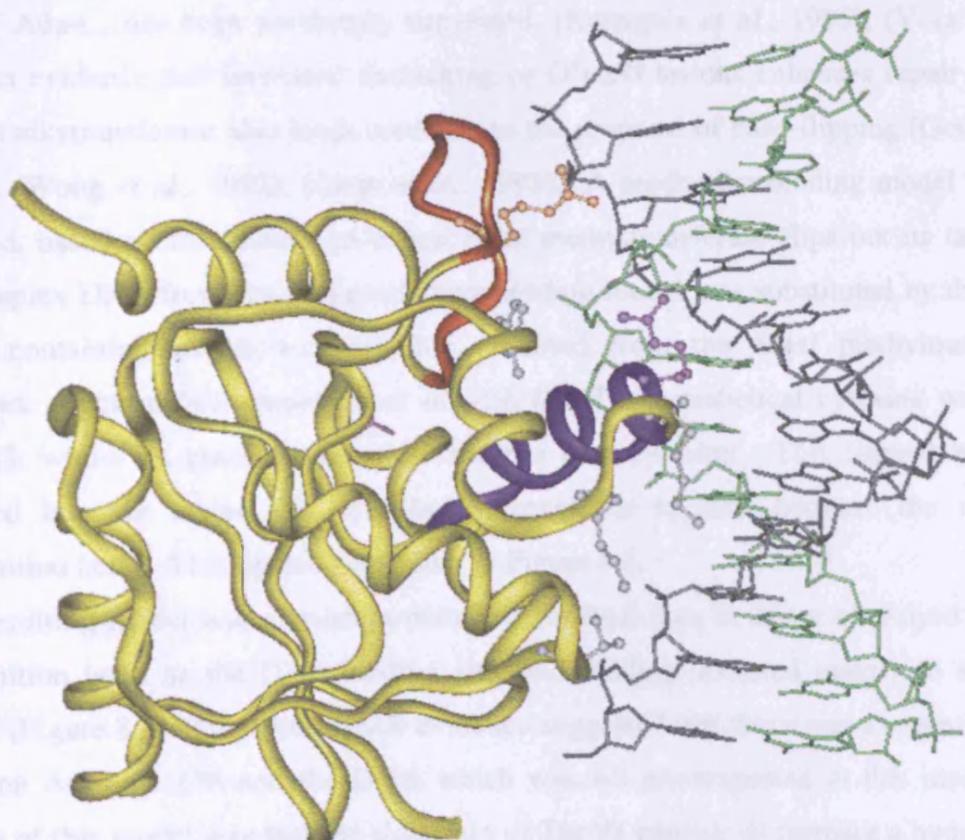
### **8.2.1 A non-productive binding model**

A preliminary model of Ada-C docking into the major groove of duplex DNA was generated by superimposing the structurally homologous helix-turn-helix regions of Ada-C and CAP, from the structure of the CAP/DNA complex<sup>1</sup> (Schultz *et al.*, 1991). CAP was chosen over

---

<sup>1</sup> Please note that a similar model of Ada-C binding non-productively to DNA, based on CAP, was designed by Dr. J. Wibley prior to the making of the model described here (Wibley *et al.*, (2000) *Nucleic Acids Research* Vol. 28, 393-401).

other helix-turn-helix proteins since its structure in complex with DNA had been determined. The bent CAP DNA was then substituted for the structure of a non-distorted duplex dodecamer containing two O<sup>6</sup>meG lesions (Leonard *et al.*, 1990). This was done, since there is no reason to believe Ada-C should bend the DNA in an identical manner to CAP. The resulting model revealed Ada-C docking on to duplex DNA utilising both wing and recognition helix to make contacts with the phosphate backbone (Figure 8.1). In this model, binding is non-productive since the active site Cys 144 remains buried within the protein, and the methylated guanine stacked within the duplex of DNA. This model also fails in that, although the majority of contacts from Ada-C form to a single-strand of DNA, there are also contacts from the region between Helices 4 and 5 and the second DNA strand. More specifically Lys 124 is capable of forming a salt link with the phosphate backbone of the second DNA strand (shown in green in Figure 8.1). The amide chemical shift of Lys 124 did not exhibit signal loss through extensive line-broadening when complexed with 2.5 fold excess DNA (single or double-stranded) in low salt buffer. It was therefore unlikely that this residue was involved in direct contacts with the DNA (amide chemical shifts from residues believed to be involved in direct contacts did exhibit such line-broadening as described in Chap



**Figure 8.1** Non-productive binding model of Ada-C with duplex DNA. This model was based on the mechanism employed by CAP to bind its cognate DNA sequence, and fails to fulfill experimental observations due to interactions between protein and both strands of DNA. The recognition helix is blue, the wing red. Sidechains from Thr 92, Lys 122, Lys 124, Arg 127 and Arg 149 make potential contacts with the DNA phosphate backbone (these sidechains are drawn in). The sidechain of Tyr 113 does not contact the DNA.

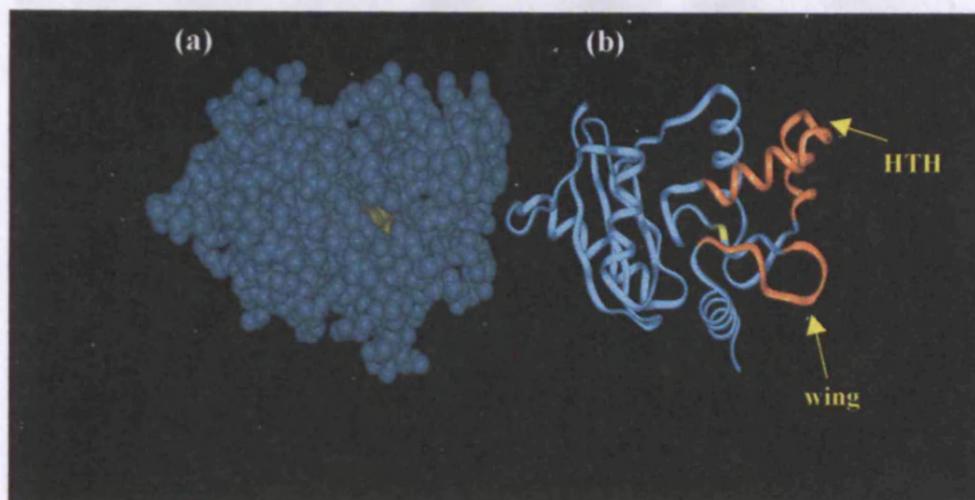
Rotation of Ada-C about the DNA would enable it to bind solely to a single strand of the duplex utilising the wing and recognition helix. The region between Helices 4 and 5 would then not contact the second strand of DNA. However, such a model would still be non-productive since O<sup>6</sup>meG would remain stacked into the DNA double helix, inaccessible to the buried methyl acceptor Cys 144.

### **8.2.2 The productive binding model**

Inactive C144S Ada-C bound methylated and unmethylated DNA in subtly different manners. Such differences may suggest two modes of binding, one 'productive mode' with methylated DNA, and a non-productive 'scanning mode' with unmethylated DNA. This hypothesis has also been made on the basis of kinetic studies with homologous human alkyltransferase (Hazra *et al.*, 1997), (Spratt, 1999). The productive mode differed from the scanning mode in a manner which would be consistent with base-flipping methylated but not unmethylated bases from DNA. Although base-flipping was not proven, it appears to be a functional necessity in the absence of a large conformational change in Ada-C. Such a base-flipping mechanism, whereby the methylated guanine lesion protrudes extrahelically from the DNA into the active site of Ada-C, has been previously suggested, (Kanugula *et al.*, 1995), (Vora *et al.*, 1998). Indirect evidence that increased unstacking of O<sup>6</sup>meG lesions enhances repair efficiency by human alkyltransferase also lends credence to the proposal of base-flipping (Georgiadis *et al.*, 1991), (Wong *et al.*, 1992), (Liem *et al.*, 1993). A productive binding model was therefore devised, based on the manner in which *HhaI* methyltransferase flips out its target cytosine. The duplex DNA from the non-productive binding model was substituted by the structure of DNA containing extrahelical cytosine obtained from the *HhaI* methyltransferase/DNA complex structure (Klimasauskas *et al.*, 1994). The extrahelical cytosine was mutated to O<sup>6</sup>meG, whilst its guanine pair was changed to a cytosine. This flipped base was then inserted into the active site of Ada-C through an opening between the wing and the recognition helix. This opening is shown in Figure 8.2.

The resulting model was consistent with experimental data in that it employed the wing and recognition helix as the DNA-binding site, and binding occurred mainly to a single DNA strand (Figure 8.3). However, NMR evidence suggested that there was a potential interaction between Arginine 158 and the DNA which was not encompassed in this model. Another failing of this model was that the sidechain of Thr 92 capable of forming a hydrogen-bond to the phosphate backbone of the second strand of DNA. Structural evidence gathered indicated this should not be the case. When CAP binds DNA, the recognition helix lies in the major groove, approximately parallel to the base pairs. The CAP-based models described above for

Ada-C are based on similar positioning of the recognition helix in the major groove. This, although it employs the correct binding surface of Ada-C, also leads to contacts with both strands of DNA, which are inconsistent with NMR data.

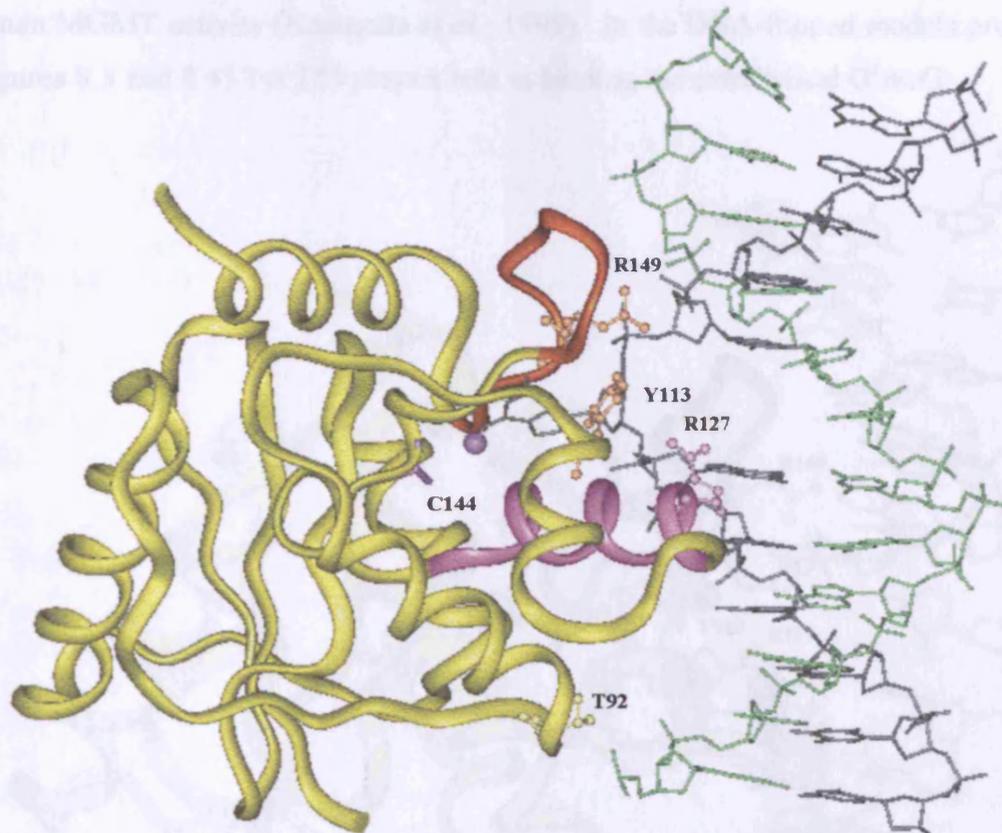


**Figure 8.2** Ada-C viewed from the unique angle which permits visibility of the buried active site Cys 144 sidechain. **(a)** A spacefilled depiction of Ada-C with Cys 144 coloured yellow. **(b)** A ribbon diagram of Ada-C in the same orientation, recognition helix and 'wing' are coloured red. In the model of Ada-C flipping O<sup>6</sup>meG, the lesion was inserted through this hole.

The previously published model of Vora *et al.* (Figure 8.4) (Vora *et al.*, 1998) was more consistent with the NMR data. This model was based on Ada-C homology to the DNA-binding domain of Mu transposase (MuA), and employed the same binding mode as determined for the latter enzyme (Clubb *et al.*, 1994). Again, the wing and recognition helix are used in the binding interface, the recognition helix lies in the major groove, but is tilted more along the axis of the DNA backbone than in the model based on CAP-DNA binding (Figure 8.3). As a consequence there are no interactions to the second strand of DNA, and potential interactions between Arginine 158 and the phosphate backbone of a single DNA strand exist. To all intents and purposes then, it appears as though the model of Vora *et al.* fits the available experimental data better than a model based on the binding mechanism of CAP.

Interestingly, Ada-C was related to MuA via a sequence similarity search (Vora *et al.*, 1998), whilst it was related to CAP using structural homology. This demonstrates how the type of homology used to find a protein/DNA complex on which to base the DNA-binding model can affect the final outcome of that model. In this case two subtly different variations of the same binding mechanism between Ada-C and DNA were created by basing the models on sequence or structural homology.

Although here it is believed that the Vora model (Figure 8.4) fits the experimental data more closely than the flipped model based on CAP (Figure 8.3), the differences in these two models are very minor. In both models a helix-turn-wing motif is employed to bind the DNA. Exactly what orientation the recognition helix assumes in the major groove of DNA, and whether there are any very minor interactions to the second DNA strand will best be determined by a high resolution structure of the complex.

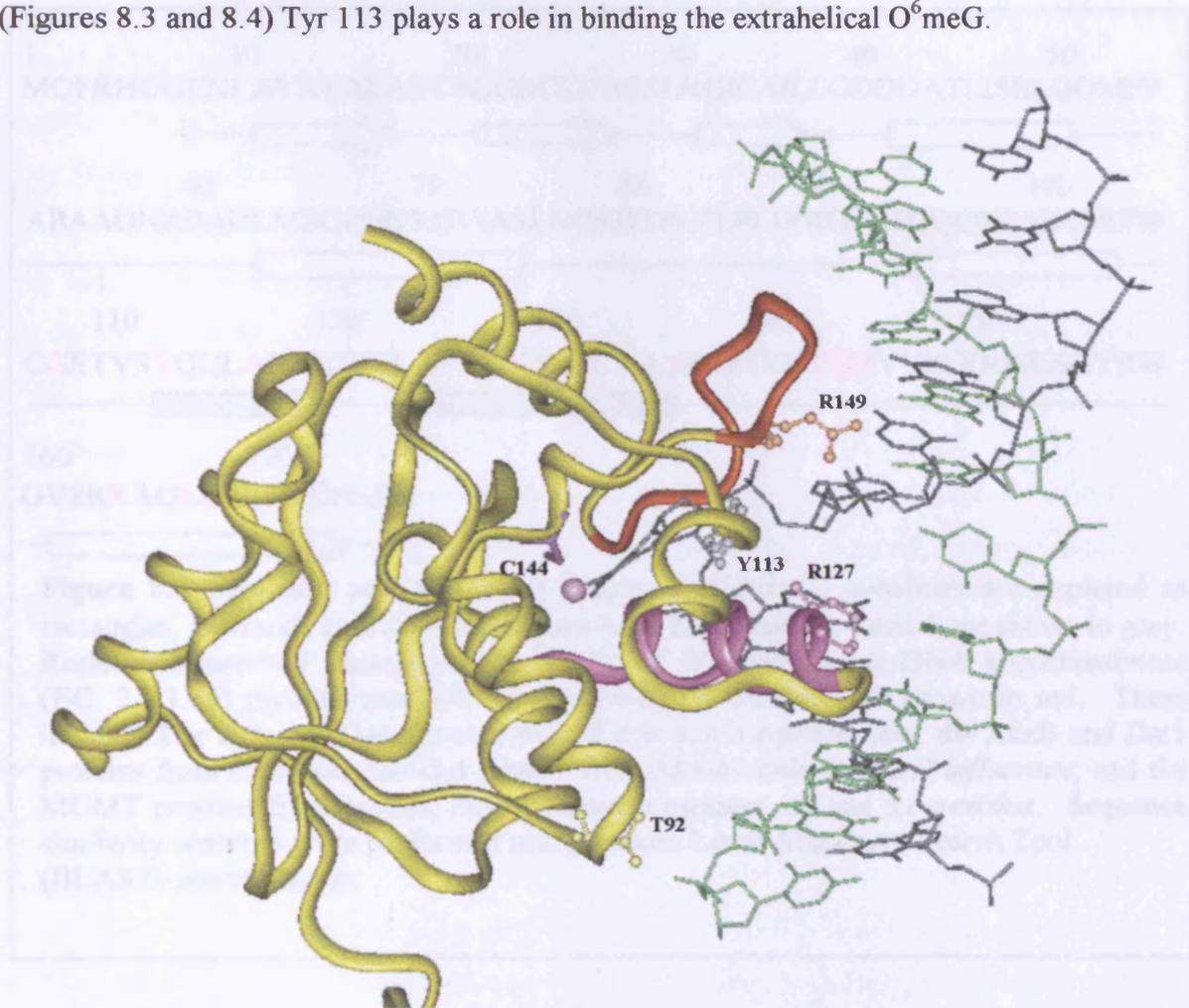


**Figure 8.3** DNA-binding model of Ada-C with O<sup>6</sup>meG flipped into the active site through the hole depicted in Figure 8.2. The orientation of the recognition helix of Ada-C with respect to the DNA is based on the CAP/DNA complex Schultz *et al.* (1991). The recognition helix is coloured blue, the wing is red. Note the potential interactions from the sidechains of Tyr 113, Arg 149 and Arg 127 to the phosphate backbone of one strand, and the interaction with Thr 92 and the second DNA strand (sidechains of K122 and K124 do not contact the DNA in this model, and are not depicted). The exocyclic methyl group from O<sup>6</sup>meG is shown as a blue ball in the active site, juxtaposed with Cys 144.

A non-productive 'scanning mode' of binding, may also involve identical relative orientations of Ada-C and DNA as depicted in the Vora model. The only difference being that no bases are extrahelical, thus making binding non-productive. This may be the mode by which Ada-C searches the DNA for lesions prior to flipping them.

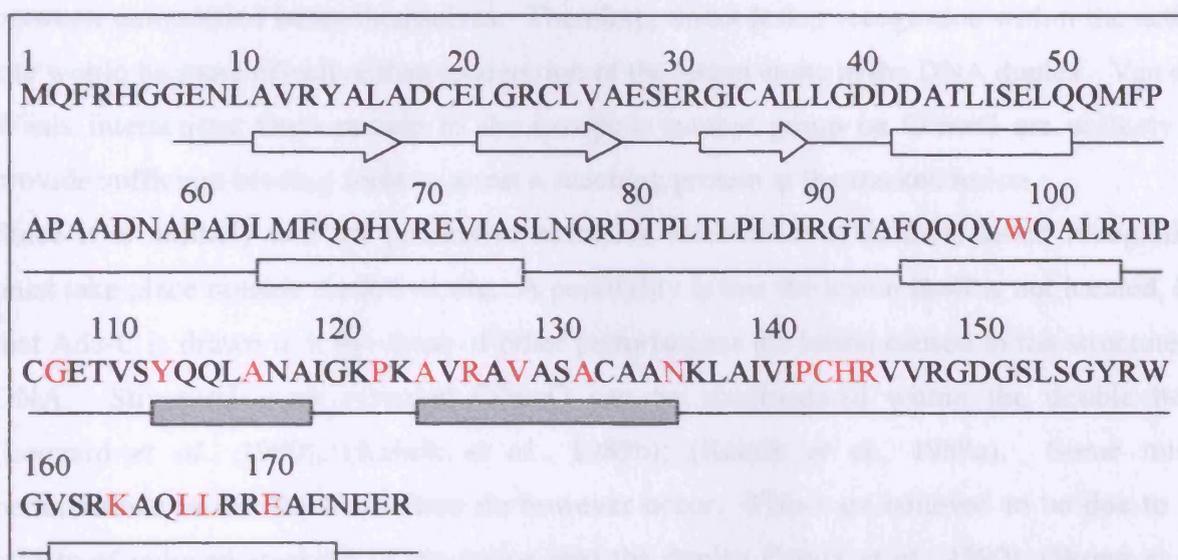
In both flipped models, potential salt links form between the sidechains of arginine 127 in the recognition helix, arginine 149 in the 'wing', and the phosphate backbone of DNA. That only arginine 127 is highly conserved suggests that this residue is of greater importance (Figure

8.5). The homologous residue in the human alkyltransferase has been implicated in binding single-stranded DNA (Kanugula *et al.*, 1995). This residue may be of some importance in stabilising the conformation of single-stranded DNA upon binding, to allow intrusion of O<sup>6</sup>meG into the active site. The less conformationally flexible duplex DNA has less need of this arginine, since binding of the human R127A mutant to duplex DNA was less impaired than binding to single-stranded. Conserved Tyr 113 has also been shown to be critical for human MGMT activity (Kanugula *et al.*, 1995). In the DNA-flipped models presented above (Figures 8.3 and 8.4) Tyr 113 plays a role in binding the extrahelical O<sup>6</sup>meG.



**Figure 8.4** DNA-binding model by Vora *et al.*, depicting Ada-C flipping O<sup>6</sup>meG into the active site. As in Figure 8.3, the recognition helix (blue) and wing (red) make up the DNA-binding interface. The orientation of the recognition helix differs to the CAP-based flipped model in Figure 8.3. Here the angle between the plane of the base-pairs and the recognition helix axis is greater than in Figure 8.3. Effectively, the recognition helix lies more along the axis of phosphate backbone. Tyr 113, Arg 127 and Arg 149 whose sidechains are displayed, make salt links to the phosphate backbone of a single DNA strand, as does Arg 158 (not drawn). Thr 92 does not contact the second strand of DNA and is uninvolved in binding. The methyl group of O<sup>6</sup>meG is shown as a blue ball in the active site.

On the whole, there are not many obviously strong points of enthalpic interaction, such as salt links between Ada-C and DNA in either productive or non-productive binding modes. This would tie in with the known entropically-driven binding mechanism. This mechanism would rely heavily on the displacement of ordered water molecules and ions from DNA and protein surfaces to bulk solution as a driving force for binding.



**Figure 8.5** Sequence and secondary structure of Ada-C.  $\alpha$ -helices are depicted as rectangles,  $\beta$ -strands as arrows, helix-turn-helix motif helices 4 and 5 are shown in grey. Residues conserved throughout the family of  $O^6$ -alkylguanine-DNA alkyltransferase (EC. 2.1.1.63) proteins available in the Swissprot database are shown in red. These include: The Ada and Ogt proteins from *E.coli* and *S.typhimurium*, the AdaB and Dat1 proteins from *B.subtilis*, the Ogt protein from *M.tuberculosis* and *H.influenzae*, and the MGMT proteins from human, mouse, chinese hamster, rat and *S.cerevisiae*. Sequence similarity searches were performed using a Basic Local Alignment Search Tool (BLAST) search engine.

### 8.3 MECHANISM OF LESION LOCALISATION

DNA repair by the alkyltransferase family of proteins is almost at the limit for a diffusion controlled reaction *in vitro* (Graves *et al.*, 1989), (Pegg *et al.*, 1995). The *in vivo* second order repair rate constant is lower ( $8.5 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ ) due to competitive binding by excess unmethylated DNA (Graves *et al.*, 1989), (Yarosh *et al.*, 1986). The exact mechanism used by Ada-C to locate the  $O^6$ meG lesion with such efficiency is unknown. However, from the results presented here, and those from other studies on alkyltransferase repair proteins, certain hypothesis can be made.

Inactive C144S mutant Ada-C bound methylated DNA in a subtly different manner to unmethylated DNA. These differences might be consistent with base-flipping  $O^6$ meG in the

methylated DNA, but not flipping unmodified bases in unmethylated DNA. This would preclude the 'processive extrusion' mechanism for lesion localisation where the protein randomly binds DNA and flips out a base (Verdine *et al.*, 1997). Ada-C would then scan along the DNA alternately flipping bases into the active site where they could be sampled until a lesion was located. This localisation mechanism is theoretically attractive as it allows lesion recognition within the active site. The presence of an extra methyl group protruding into the major groove on O<sup>6</sup>meG provides no greater distinction from other bases than between unmodified bases themselves. Therefore, direct lesion recognition within the active site would be more effective than recognition of the lesion alone in the DNA duplex. Van der Waals interactions from protein to the exocyclic methyl group on O<sup>6</sup>meG are unlikely to provide sufficient binding force to arrest a scanning protein at the stacked lesion.

Since it is unlikely that the processive extrusion mechanism is utilised, lesion recognition must take place outside the active site. A possibility is that the lesion itself is not located, but that Ada-C is drawn to it by virtue of other perturbations the lesion caused to the structure of DNA. Structural work revealed O<sup>6</sup>meG can be accommodated within the double helix (Leonard *et al.*, 1990), (Kalnik *et al.*, 1989b); (Kalnik *et al.*, 1989a). Some minor perturbations to the DNA structure do however occur. These are believed to be due to the effects of reduced stacking of the lesion into the duplex (Voigt *et al.*, 1990), (Wong *et al.*, 1992). <sup>19</sup>F NMR work indicated environmental changes of the cytosine paired with O<sup>6</sup>meG occurred on the millisecond, or longer timescale. This would be consistent with the O<sup>6</sup>meG lesion stacking and unstacking from the DNA duplex, which may also be seen as localised duplex melting. The suspected timescale of melting is consistent with that measured for localised melting or 'breathing' in unmodified B-form DNA (Gueron *et al.*, 1995). The ratio of melted to duplexed 20-mer DNA containing O<sup>6</sup>meG is 250 times greater than for unmethylated DNA of the same sequence at 37°C (Chapter 7). Localised melting of just two or three base pairs around the lesion would occur with an even greater ratio relative to unmethylated DNA, due to the cooperative forces involved in duplex stability. Taken together, these factors strongly suggest that O<sup>6</sup>meG introduces local instability into a DNA duplex.

Through these perturbations, O<sup>6</sup>meG could advertise its presence more dramatically to DNA repair proteins such as Ada-C. Locating a region of reduced base stacking (which frequently introduces kinks into duplex DNA (Voigt *et al.*, 1990)) would certainly be more rapid than finding an exocyclic methyl group on a base. This mechanism of recognition would explain why single-stranded DNA is repaired more slowly than duplex DNA, even though the two types are bound with similar affinity (Bhattacharyya *et al.*, 1990), (Fried *et al.*, 1996). A lesion in the less structured single-stranded DNA might be less readily located than in B-form

duplex DNA, because there is less induced structural perturbation in the former. The exact nature of the structural perturbation recognised by Ada-C remains unclear. Both kinks in the phosphate backbone and unstacked bases are themes more associated with duplex DNA. Some degree of base stacking is expected in single-stranded DNA from an energetic viewpoint. Perhaps the presence of O<sup>6</sup>meG causes unstacking and distortion to the DNA backbone in single-stranded DNA, and this is also recognised by Ada-C.

The question remains as to whether Ada-C uses one-dimensional diffusion to slide along the DNA looking for the structural perturbation, or whether it is located by simple bimolecular collision. One-dimensional *sliding* is a form of facilitated diffusion by which the protein diffuses along the DNA driven by thermal fluctuations in the solvent. The reduced dimensionality of this diffusion allows an increase in the rate of DNA scanning. The efficiency with which certain regulatory proteins such as *lac* repressor locate their operator sequences has been attributed to their use of one-dimensional sliding (von Hippel *et al.*, 1989). No evidence for one-dimensional sliding of human alkyltransferase along DNA was observed (Bender *et al.*, 1996). It is therefore unlikely that the homologous Ada-C should slide when searching for the lesion.

Bimolecular collisions between Ada-C and DNA might be made more efficient by electrostatic steering of the positively charged DNA-binding face of Ada-C to polyanionic DNA. Such macromolecular collisions are long-lived (inelastic) and can be further subdivided into a series of minicollisions once the protein is in the vicinity of the DNA. These reiterated minicollisions might give rise to short protein excursions along the DNA, the process being termed 'hopping' (von Hippel *et al.*, 1989). Each '*hop*' is thought to cover 4-8 base pairs. This *hopping* phenomenon, although less efficient than one-dimensional sliding, would certainly reduce the dimensionality of the search for lesions.

There are cooperative interactions between the human alkyltransferase proteins in binding single-stranded DNA (Fried *et al.*, 1996). The oligonucleotides used to monitor interactions with Ada-C in this thesis may have been too short to allow any cooperativity. Hence such interactions cannot be ruled out *in vivo*. Since only a small amount of DNA free from condensing proteins is available in prokaryotic and eukaryotic genomes DNA repair proteins should not have to scan great segments of DNA. Indeed, it was recently shown that human alkyltransferase is targeted to transcriptionally active genes (Thomale *et al.*, 1994), (Ali *et al.*, 1998). Such small lengths of naked DNA could probably be efficiently scanned by a hopping mechanism, or by cooperative binding of alkyltransferase.

#### **8.4 MECHANISM OF BASE-FLIPPING**

Upon localisation of the lesion via the distortion it caused to the local duplex structure, base-flipping could occur to effect the repair. The question then is: does Ada-C actively flip out the O<sup>6</sup>meG lesion, or does it rely on serendipitous capture of a spontaneously extrahelical base ?

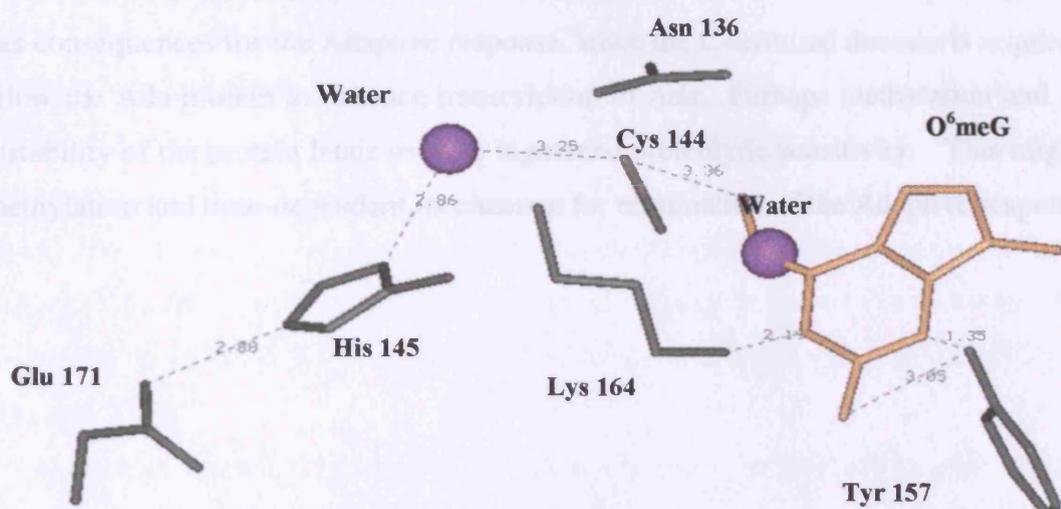
As with many of the enzymes employing base-flipping to gain access to target bases, the exact mechanism of rendering the base extrahelical is poorly understood (Mol *et al.*, 1999). In the case of Uracil-DNA glycosylase for example, three mechanisms of flipping have been proposed: An active 'push-pull' mechanism where the protein forces the uracil to become extrahelical by inserting a leucine into the duplex (Slupphaug *et al.*, 1996), contrasts to the passive capture of spontaneously flipped Uracil (Panayotou *et al.*, 1998). A mechanism of protein-induced DNA strand separation has also been put forward as an aid to base-flipping (Vassylyev *et al.*, 1996). From the model of how Ada-C binds DNA it is possible that the sidechain of Arg 127 is be able to fill the hole left by the flipped base. This arginine may be involved in a similar push-pull mechanism to that believed to occur in Uracil-DNA Glycosylase. Arg 127 may also flip the lesion by exerting strain on the phosphate backbone together with Arg 149. However, given the known localised duplex instability surrounding an O<sup>6</sup>meG lesion it is also possible that no great effort has to be put into flipping the base. The binding of Ada-C to the phosphate backbone might be sufficient to enable it to capture a spontaneously flipped lesion. Without further information on the intermediates involved in base flipping, all potential mechanisms remains speculative.

#### **8.5 MECHANISM OF DEMETHYLATION**

During the modelling of Ada-C base-flipping O<sup>6</sup>meG, the extrahelical O<sup>6</sup>meG nucleotide was orientated in the active site to minimise steric clashing, and bring the exocyclic methyl group proximal to the sulphur of Cys 144. It was found that these criteria could be fulfilled with the exocyclic oxygen of O<sup>6</sup>meG superimposed exactly onto an ordered water molecule in the active site. This ordered water molecule is present in the structure of the human homologue (J.E.A. Wibley, personal communication), suggesting important hydrogen-bonds occur to this point (in the free enzyme such bonds can form from Lys 164 and Cys 144 to this water). The active-site containing the extrahelical lesion is shown in Figure 8.6.

Studies with O<sup>6</sup>meG analogues have delineated the importance of the O<sup>6</sup>, N<sup>1</sup> and exocyclic amino N<sup>2</sup> positions of the base as contact points from Ada-C (Spratt *et al.*, 1994); (Spratt *et al.*, 1992). More recently it was also noted that interactions to the 3- and 7-positions of the base, although less important, could not be precluded (Spratt, 1999). The positively charged sidechain of conserved K164 was capable of forming a hydrogen-bond to either O<sup>6</sup> or N<sup>1</sup> of

the base. This may aid stabilisation of the base in a position with the O<sup>6</sup> methyl proximal to the thiol of Cys 144, such that nucleophilic attack can occur. Conserved Y113 can form a potential hydrogen-bond to the N<sup>3</sup> or N<sup>2</sup> amino group of the base, again helping to hold it in position. The Hydrogen-bonding network of the catalytic tetrad, Glu 171-His 145-water-Cys 144 is undisrupted by entry of O<sup>6</sup>meG into the active site. The tautomeric state of His 145 is unaltered upon binding to unmethylated DNA, and is such that the Histidine imidazole proton is shared with Glu 171. As a consequence of this, it is believed that the thiol proton of Cys 144 is shared with the ordered water molecule also hydrogen-bonded with His 145 (Figure 8.3).



**Figure 8.6** The active-site of Ada-C containing extrahelical O<sup>6</sup>meG. This is taken from the model depicted in Figure 8.3. Note the extensive hydrogen-bonding network involved, and the displacement of an ordered water molecule (shown as blue spheres) by the exocyclic oxygen of O<sup>6</sup>meG. Distances between atoms are given in Å, with potential Hydrogen bonds represented by dotted lines (note hydrogen atoms are not depicted).

NMR pH titrations revealed that in the free form of the protein, the thiol of Cys 144 did not have an abnormally low pK<sub>a</sub> (a pK<sub>a</sub> below 9). Based on these observations it is not unreasonable to speculate that deprotonation of the thiol occurs concurrently with nucleophilic attack by the sulphur atom on the methyl group of the base. The hydrogen-bond between the cysteine thiol and ordered water would cause charge polarisation, with a resultant build up of electronegativity occurring on the sulphur. This sulphur atom could then attack the methyl group of O<sup>6</sup>meG, with concurrent deprotonation of the thiol. A build up of negative charge would then have to be stabilised on the recently demethylated guanine. This could be partially achieved by resonance delocalisation around the aromatic ring of the base. The positively charged sidechain of conserved Lys 164 is also perfectly located to aid in stabilising this negative charge until protonation of the O<sup>6</sup> is achieved. Both Lys 164 and Tyr

113 are suitably located to act as proton donors. However, their sidechain  $pK_a$  values would have to be significantly lower than the  $pK_a$  of free residues ( $pK_a = 10$  for free Tyr and Lys). Alternatively, the protonated water might translocate and reprotonate the base.

Upon methylation of Cys 144, Ada-C loses structural integrity. Mutation of this residue to alanine left a structurally unstable protein. Mutation to serine gave rise to a protein which, although correctly folded, was more conformationally flexible than the wild type. It would thus appear that the ability of the residue in the position of Cys 144 to form a sidechain hydrogen-bond is important for maintenance of structure. Methylation of Cys 144, is therefore believed to cause a disruption of the network of hydrogen-bonds involving Glu 171, His 145 and Cys 144. This then reduces the stability of the folded form of the protein. Cys 144 is not only the nucleophilic centre of Ada-C, but also a structural epicentre. *In vivo* this has consequences for the Adaptive response, since the C-terminal domain is required intact to allow the Ada protein to enhance transcription of *Ada*. Perhaps methylation and subsequent instability of the protein lends itself to heightened proteolytic sensitivity. This might lead to a methylation and time-dependent mechanism for termination of the Adaptive response.

## Conclusions

NMR evidence has shown that Ada-C binds to single- and double-stranded DNA in a similar manner. Coupled with the observed doubling of Ada-C stoichiometry upon binding duplex as opposed to single-stranded DNA, this suggests Ada-C binds predominantly to a single strand of duplex DNA. The binding interface of Ada-C is made up of the recognition helix of the helix-turn-helix (Helix 5) and a loop region known as the 'wing' (residues 149-158). This experimentally mapped DNA-binding surface is consistent with the DNA-binding model proposed by Vora *et al.*, (1998). Here the recognition helix lies in the major groove of duplex DNA and there are potential interactions to the phosphate backbone from Arg 149 on the wing, and Arg 127 on Helix 5, as well as Y113 and R158. The O<sup>6</sup>meG base is believed to be flipped out of the double helix in a manner similar to that observed with the *HhaI* methyltransferase/DNA complex (Klimasauskas *et al.*, 1994). Although base-flipping was not proven, it is consistent with the available results obtained by NMR, which examined how inactive Ada-C C144S bound in a subtly different manner to methylated and unmethylated DNA. Base-flipping is the only means by which O<sup>6</sup>meG can gain access to the buried methyl acceptor Cys 144, without a large conformational change in Ada-C (now known not to occur). Rendering the O<sup>6</sup>meG lesion extrahelical is therefore a functional necessity if repair is to be effected.

An alternative model of how Ada-C binds DNA was derived through examining how CAP binds DNA. The difference between this model and the one of Vora *et al.* was the angle between the recognition helix and the plane of bases in the major groove of DNA. This angle allowed Ada-C to contact only a single DNA strand, and allowed contacts between the sidechain of Arg 158 and the phosphate backbone in the Vora model. In the CAP-based model Ada-C forms minor contacts to the second DNA strand, and there are no contacts between Arg 158 and DNA. In these two respects the CAP-based model fits the experimental data less well than the Vora model. Differences in the two models are however very minor. Determination of the exact angle with which Ada-C lies in the major groove of DNA awaits the elucidation of a high resolution structure of the complex.

An endothermic enthalpy change was observed when Ada-C bound single and double-stranded DNA. Binding to DNA was therefore defined as being 'entropically driven'.

Demethylation of O<sup>6</sup>meG in single-stranded DNA was however highly exothermic (the magnitude of this exothermic change was buffer-dependent). Upon methylation of Cys 144, Ada-C loses structural integrity, delineating the importance of maintaining the hydrogen-bonding network in the active site.

The method of lesion localisation amidst the mass of competing genomic DNA is unknown. However, it seems likely that Ada-C recognises perturbations caused by the lesion, rather than direct recognition of the lesion itself. The precise mechanism of base-flipping is also unknown, as with other base-flipping enzymes. Flipping O<sup>6</sup>meG lesions but not unmodified bases, although not proven, is consistent with available NMR evidence. This led to the speculative mode of action where by Ada-C locates the lesion on the basis of its perturbation to the DNA structure, and then flips it out and effects repair. Perhaps Ada-C even captures a flipped base, although it is more likely that the protein recognises the unstacked nature of the target base and then converts it to fully extrahelical form.

## Further work

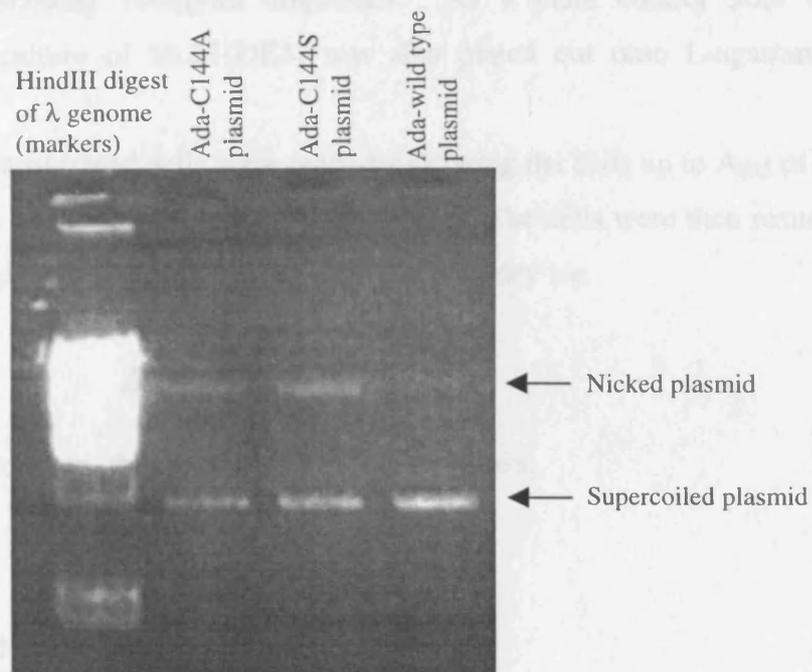
There is still much work to be done on the mechanism of DNA-repair by Ada-C. High resolution detail of how the protein binds DNA can only be obtained by solving the solution of crystal structure of the Ada-C/DNA complex. In my opinion this would most easily be achieved through x-ray crystallography. It would be of great interest to discern whether the protein flips both normal bases and lesions. In this vein, co-crystallisation of inactive mutant C144S Ada-C with duplex DNA containing O<sup>6</sup>meG should be attempted. Solution of this structure would define the productively bound complex. Solution of wild type Ada-C with unmethylated duplex DNA would define the non-productive 'scanning mode' of binding. I suspect that it might be easier to crystallise the productive complex because, there may be less heterogeneity involved between the population of complexes than for non-productive binding. It is a matter of some interest also as to whether Ada-C binds DNA containing O<sup>6</sup>meG with higher affinity to unmethylated DNA. ITC titrations using inactive C144S Ada-C and DNA could rapidly determine this. This might provide valuable clues as to the mechanism by which Ada-C is able to so rapidly localise lesions amidst a mass of competing genomic DNA. Fluorine NMR work could be continued on the interactions between inactive Ada-C and DNA. Perhaps an oligonucleotide could be custom synthesised with a <sup>19</sup>F on an O<sup>6</sup>meG base (eg. O<sup>6</sup>-3-fluoromethylguanine). This would provide a direct probe for base-flipping as used by Klimasauskas *et al.* (1998).

# Appendices

## APPENDIX 1-PLASMIDS, MEDIA AND OLIGOS

### Appendix 1.1- Plasmid preparation

Plasmids were transformed into non-expressing strains in accordance with the transformation protocol described below. pET22b:*adaC* wild type plasmid was grown up in DH5 $\alpha$ , whilst pET22d:*adaC* C144S and C144A mutant plasmids were amplified in the *E. coli* strain XL-1. 5ml volume of LB media was inoculated from cell glycerol stock and the culture allowed to attain stationary phase overnight by incubating at 37°C and 200rpm shaking. 3ml of the ensuing stationary phase culture was spun down on a bench top microfuge at 6500rpm for 40 seconds. The cell pellets were collectively resuspended into 200ul of cell resuspension buffer (50mM Tris-HCl, pH 7.5, 10mM EDTA, 100ug/ml RNaseA). 200ul of cell lysis solution (0.2M NaOH, 1% SDS) was then added and the tube inverted until the solution became cloudy. 200ul of neutralization buffer solution (1.32M Potassium Acetate) was then added and the cellular suspension observed to clear. The precipitate was removed by centrifugation at 13000 rpm for 5 minutes in a microfuge and the supernatant transferred to a clean eppendorf tube. 1ml of DNA purification resin (Promega) was added and the suspension mixed by inverting the tube. The suspension was pipetted into a syringe attached to a wizard minicolumn (Promega) which is in turn linked to a vacuum manifold. The vacuum was engaged and the solution allowed to drain through.



**Figure 9.1** The degree of plasmid nicking and purity of plasmid preparations from the wizard miniprep kit (Promega).

2mls of wash solution (8.4mM Tris-HCl, pH 7.5, 2.1mM EDTA, 84mM NaCl, 55% EtOH) was then passed through the syringe and the vacuum left on for a further 2 minutes to allow column drying. The minicolumn was then transferred into a 1.5ml microcentrifuge tube and centrifuged for 20 seconds at 13000 rpm to remove any residual wash solution. The minicolumn was then transferred into a clean microcentrifuge tube and 50ul of heat sterilised water added. After 2 minutes the solution was spun through the column and the plasmid quantified using the standard of 1mg/ml DNA creates an A<sub>260nm</sub> of 20.3. Degree of plasmid nicking was assessed to be acceptable by agarose gel electrophoresis (Figure 9.1).

### **Appendix 1.2- The cellular transformation and glycerol storage protocol**

An overnight culture of BL21(DE3) in 2YT media was used to inoculate a further 10mls of 2YT. This was allowed to grow at 37°C and 200rpm shaking until the absorbance at 595nm was between 0.3 and 0.5. The cells were centrifuged in a Sovall TC6 centrifuge, (Dupont) at 3000 rpm for 2 minutes and then resuspended in 5ml of ice cold 100mM CaCl<sub>2</sub>. After leaving for 1 hour on ice the cell solution was re-centrifuged as above and the pellet gently resuspended into 200ul of ice cold 100mM CaCl<sub>2</sub>. 10ul of 16ng/μl plasmid DNA was added and the solution left on ice for 1.5 hours. The cells were then heat shocked at 42°C for 90 seconds and returned briefly to ice. 1ml of 2YT was added and the cells allowed to grow for 1 hour at 37°C with 200rpm shaking. 10ul of transformed BL21(DE3) was then plated out onto L-agar plates containing 100ug/ml ampicillin. As a plate control 50ul of non-transformed overnight culture of BL21(DE3) was also plated out onto L-agar/ampicillin plates.

Glycerol stocks of all transformed cells were made by growing the cells up to A<sub>595</sub> of 0.6 and then gently pelleting the cells using a bench top microfuge. The cells were then resuspended gently into sterile 20% glycerol/LB media and snap frozen on dry ice.

### **Appendix 1.3 SDS-PAGE**

SDS-PAGE 15% gels were run, the recipe for these is as follows:

#### **Separating gel:**

Water 814μl  
acrylamide/bisacrylamide 37.5:1 1764μl  
1.5M Tris pH 8.8 875μl  
10% SDS 35μl  
TEMED 2μl  
10% Ammonium persulphate 17μl

### Stacking gel:

Water 1500µl	acrylamide/bisacrylamide 37.5:1 325µl
0.5M Tris pH 6.8 625µl	TEMED 2µl
10% SDS 25µl	10% Ammonium persulphate 12µl

### Running buffer:

Tris 30g/L, Glycine 144g/L, SDS 10g/L

### SDS loading buffer:

63mM Tris, 10% glycerol, 2% SDS, 5% 2-β-mercaptoethanol,  $1.25 \times 10^{-3}$  % bromophenol blue

## Appendix 1.4- Media

### 2M9 Media composition

Solution A (made up to 900ml with dH<sub>2</sub>O, brought to pH 7.2 with 5M NaOH and autoclaved):

Na<sub>2</sub>HPO<sub>4</sub> (anh) 12.5g  
KH<sub>2</sub>PO<sub>4</sub> (anh) 7.5g  
(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> 1g

Solution B (made up to 100ml and filter-sterilised using 0.2µm acrodisk):

D-Glucose 4g  
MgSO<sub>4</sub> 240mg  
5g/L MnSO<sub>4</sub> 0.5ml  
37.5g/L CaCl<sub>2</sub>·2H<sub>2</sub>O 0.5ml  
27g/L F<sub>3</sub>Cl<sub>3</sub> 0.5ml

Ampicillin added to 100mg/L in final media from concentrated, filter-sterilised stock.

Note: For growth of B834(DE3) 250mg L-Met is added to Solution B

### **2M9A Media composition**

Solution A as for 2M9 but also contains the following L-amino acids:

L-Amino Acid	Mass added when unlabelled (mg)	Mass added when <sup>15</sup> N labelled (mg)
Ala	500	600
Glu	650	550
Gly	550	500
His	100	not labelled
Ile	230	55
Leu	230	55
Lys	420	not labelled
Met	250	not labelled
Tyr	170	35
Val	230	100
Pro	100	not labelled
Arg	400	not labelled
Cys	50	50
Asp	400	not labelled
Asn	200	not labelled

**Table 1.4** Amino acids found in Solution A of media 2M9A.

Solution B is similar to that of 2M9, but has the additional amino acids added:

L-Amino Acid	Mass added when unlabelled (mg)	Mass added when <sup>15</sup> N labelled (mg)
Phe	130	130
Ser	1500	not labelled
Thr	230	not labelled
Trp	50	not labelled
Gln	400	not labelled

**Table 1.5** Amino acids found in Solution B of media 2M9A.

### **2YT Media composition**

16g Bacto-tryptone  
10g Bacto-yeast extract  
5g NaCl

Make up to 1L with deionised water.  
pH to 7.0 with NaOH

### **LB Media composition**

10g Bacto-tryptone  
5g Bacto-yeast extract  
10g NaCl

Make up to 1L with deionised water.  
pH to 7.0 with NaOH

### Appendix 1.5: Calculation of Extinction coefficients for linear oligonucleotides

This data is obtained from Handbook of Biochemistry and Molecular Biology, Volume 1: Nucleic acids, Fasman, G.D. editor, page 589, 3<sup>rd</sup> edition, CRC Press, 1975. The extinction coefficient at 260nm, 25°C and neutral pH for single stranded DNA is determined by the nearest-neighbour method  $E = 2E_{dNpdN} - E_{pdN}$ .

For example, the extinction coefficient of oligomer 5'-ATGCTTC-3' is:

$$E_{ATGCTTC} = 2(E_{dApdT} + E_{dTpdG} + E_{dGpdC} + E_{dCdT} + E_{dTpdT} + E_{dTpdC}) - E_{pdT} - E_{pdG} - E_{pdC} - E_{pdT} - E_{pdT}$$

STACK OR MONOMER	EXTINCTION COEFFICIENT 1/(mmol.cm)
pdA	15.4
pdC	7.4
pdG	11.5
pdT	8.7
dApdA	13.7
dApdC	10.6
dApdG	12.5
dApdT	11.4
dCpdA	10.6
dCpdC	7.3
dCpdG	9.0
dCpdT	7.6
dGpdA	12.6
dGpdC	8.8
dGpdG	10.8
dGpdT	10.0
dTpdA	11.7
dTpdC	8.1
dTpdG	9.5
dTpdT	8.4

**Table 1.4** Showing the extinction coefficients of the stack or monomeric units in DNA.

### Appendix 1.6 Native PAGE for DNA Annealing assessment

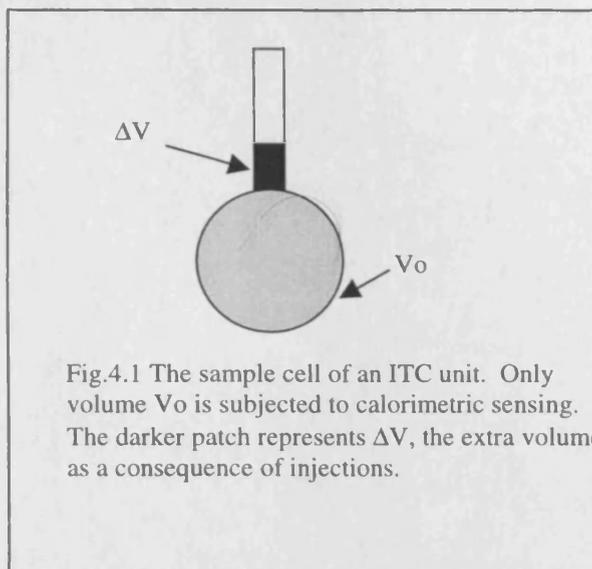
20% Separating gel:

Water 3.9ml  
 acrylamide/bisacrylamide 19:1 5ml  
 10mM Tris, 1mM EDTA 1ml  
 TEMED 10µl  
 10% Ammonium persulphate 100µl

## APPENDIX 2- ITC

### 2.1 Data Analysis for ITC

At the beginning of an ITC experiment the sample cell is filled with macromolecule solution. Only the liquid within volume  $V_0$  is sensed calorimetrically. When an



injection occurs, the total volume in the cell increases by  $dV$  and solution is driven up the tube (fig.4.1). The concentration of macromolecule in  $V_0$  is diluted each time an injection adds more ligand to the mixture. The average bulk concentration of the macromolecule in  $\Delta V$  (the total volume of injections) is the mean of the initial concentration  $M_i^0$  and the present concentration  $M_t$  (these concentrations represent both free and bound macromolecule). During an injection the ligand solution enters into  $V_0$  and is mixed with the macromolecule. Assuming mixing is rapid a more dilute concentration of macromolecule is then displaced into the tube connecting to the cell. Following the subsequent injection another injection volume of further diluted macromolecule is pushed into the connector tube. Thus the final concentration of macromolecule in this tube after all the injections are completed is:

$$\frac{1}{2}(M_t + M_i^0)\Delta V$$

The first injection will result in a solution of macromolecule slightly diluted from the initial concentration  $M_i^0$ , to be displaced into the connector tube, we shall call this concentration  $M_1$ . The second injection will occur, and will again result in an identical volume of macromolecule at a lower concentration  $M_2$  to end up in the connector tube and so on for the rest of the injections. At the end of the titration the tube will be filled with a number of aliquots of identical volume but different concentration. Thus the concentration of the entire

volume  $\Delta V$  in the connector tube is the average of the concentration of the initial,  $M_i^o$  and final concentration  $M_i$  in the cell. (In reality the final concentration in the tube is

$$\frac{1}{2}(M_t + M_i^o)\Delta V$$

because the first aliquot to be pushed into the connector tube is  $M_i$  not  $M_i^o$ . However in ITC only small volumes are injected relative to the total cell volume (20ul relative to 1700ul) so the error derived from this will be negligible and this equation is used by the Origin ITC software for simplicity.)

Thus, the initial amount of macromolecule in  $V_o$  is represented by the final amount of macromolecule in  $V_o$ , ( $M_i V_o$ ) plus the amount in  $\Delta V$  which is the average of starting and final concentrations times the volume  $\Delta V$ . This is expressed mathematically in eqn. (1).

$$M_i^o V_o = M_i V_o + \frac{1}{2}(M_t + M_i^o)\Delta V \quad (1)$$

So that :

$$M_t = M_i^o \left[ \frac{1 - (\Delta V / 2V_o)}{1 - (\Delta V / 2V_o)} \right] \quad (2)$$

$X_t$  , the concentration of ligand in  $V_o$  after an injection is related to the hypothetical total ligand concentration in  $V_o$  (which would occur if we assumed all ligand remained in  $V_o$  and none escaped into the  $\Delta V$  space)  $X_t^o$ . The relationship is shown in eqn. (3). The  $\frac{1}{2} X_t^o \Delta V$  term is the average of start and final ligand concentrations in  $\Delta V$ , using similar logic to that for macromolecule concentration in  $\Delta V$ .

$$X_t^o V_o = X_t V_o + \frac{1}{2} X_t^o \Delta V \quad (3)$$

$$X_t = X_t^o \left( 1 - \frac{\Delta V}{2V_o} \right) \quad (4)$$

The concentrations of ligand and macromolecule,  $X_t$  and  $M_t$  in the cell after a certain number of injections are calculated from the above expressions in eqns. (2) and (4), which take into account displaced volume effects. The next four equations deal with the heat change in the cell as a consequence of binding, and how it relates to stoichiometry, binding constant and enthalpy of binding. All subsequent equations deal with a single binding site, models for multiple sets of sites are not included here.

In the following equations,

**K** = Binding constant ( $K_a$ )

**n** = number of binding sites (stoichiometry)

$V_o$  = cell volume

$X_t$  and  $[X]$  are total and free concentration (in moles/L) of ligand in  $V_o$ .

$M_t$  and  $[M]$  are total and free concentration (in moles/L) of macromolecule in  $V_o$ .

$\Theta$  = fraction of sites occupied by ligand, (equal to  $[MX]/M_t$  where  $[MX]$  is concentration of complex).

So

$$K = \frac{\Theta}{(1 - \Theta)[X]} \quad (5)$$

$$X_t = [X] + n\Theta M_t \quad (6)$$

and

where  $n\Theta M_t = [MX]$  if stoichiometry  $n = 1$ .

Combining eqns. (5) and (6) gives:

$$\Theta^2 - \Theta \left[ 1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} \right] + \frac{X_t}{nM_t} = 0 \quad (7)$$

The total heat content **Q** of the solution contained in  $V_o$  at fractional saturation  $\Theta$  is:

$$Q = n\Theta M_t \Delta H V_o \quad (8)$$

Assuming  $n=1$ :

$$dQ = d[MX] \cdot \Delta H \cdot V_o \quad (9)$$

Where  $\Delta H$  is the molar enthalpy of binding,  $dQ$  is the change in heat evolved or absorbed in  $V_o$  upon binding and  $d[MX]$  is the change in the complex concentration.

By solving the quadratic equation (7) for  $\Theta$  and then substituting this into eqn. (8) we get:

$$Q = \frac{nM_t \Delta H V_o}{2} \left[ 1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} - \sqrt{\left( 1 + \frac{X_t}{nM_t} + \frac{1}{nKM_t} \right)^2 - \frac{4X_t}{nM_t}} \right] \quad (10)$$

It is the change in heat content  $\Delta Q$  occurring during an injection which is the experimental parameter of interest (more succinctly  $\Delta Q$  is the difference in  $Q(i-1)$  which occurred at the end of the last injection and  $Q(i)$  which occurred at the end of the present injection). The expression for  $Q$  above is only related to the change in heat within the volume  $V_o$ . Again we must take into account the volume displacement caused by the injection. The solution which

is displaced into the connector tube does contribute towards the heat effect before it is removed from  $V_o$ . However the displaced solution only contributes 50% as much heat as that in  $V_o$ .<sup>2</sup>

This is taken into account in eqn (11) below which is that the Origin software uses to turn the measured heat change  $\Delta Q$  back into values of  $Q$  before and after each injection.

$$\Delta Q(i) = Q(i) + \frac{dV_i}{V_o} \left[ \frac{Q(i) + Q(i-1)}{2} \right] - Q(i-1) \quad (11)$$

The process of fitting the experimental data to obtain the values of  $K_a$ ,  $\Delta H$  and stoichiometry involve initial estimations of these variables their use to calculate  $\Delta Q$  which can then be compared with the experimentally determined value for each injection. Improvement in the fit of these variables to the data are obtained by standard Marquardt methods. Iterations of the above procedure are then carried out until no significant improvement in fit is observed. The data are usually reported as a least squares fit with an inherent statistical error.

---

<sup>2</sup> The first infinitesimal volume element of the volume to be displaced from the cell into the connector tube contributes no heat as it is already equilibrated at existing concentrations that were present after completion of the previous injection. The last volume to be secreted into the tube contributes heat effects equal to the liquid remaining in  $V_o$  after the injection, (it has had time to start binding reactions that alter  $Q$ ). Assuming linearity over this tiny volume change, then the liquid in the displaced volume is only 50% as effective as producing heat relative to the liquid in  $V_o$ .

## Chapter 9

### *Bibliography*

H. Akimaru, K. Sakumi, T. Yoshikai, M. Anai, M. Sekiguchi (1990). Positive and negative regulation of transcription by a cleavage product of Ada protein. *Journal of Molecular Biology* 216, 261-273.

R.B. Ali, K.-C. Teo, H.-K. Oh, L. S.-L. Chuang, T.-C. Ayi, B.N.F.L. Li (1998). Implication of localization of human DNA repair enzyme O<sup>6</sup>-methylguanine-DNA methyltransferase at active transcription sites in transcription-repair coupling of the mutagenic O<sup>6</sup>-methylguanine lesion. *Molecular and Cellular Biology* 18, 1660-1669.

M.B. Arghavani, L.J. Romano (1995). A method for the purification of oligonucleotides containing strong intra- or intermolecular interactions by reversed-phase high-performance liquid chromatography. *Analytical Biochemistry* 231, 201-209.

P.G. Artz, K. G. Valentine, S. J. Opella, P. Lu (1996). Lac repressor-operator interaction: N-terminal peptide backbone H-1 and N-15 chemical shifts upon complex formation with DNA. *Journal of Molecular Recognition* 9, 13-24.

T.-C. Ayi, H-K. Oh, T. K-Y. Lee, B.F.L. Li (1994). A method for simultaneous identification of human active and active-site alkylated O<sup>6</sup>-methylguanine-DNA methyltransferase for monitoring human exposure to alkylating carcinogens. *Cancer Research* 54, 3726-3731.

I. Baikalov, I. Schroder, M. KaczorGrzeskowiak, K. Grzeskowiak, R. P. Gunsalus, R. E. Dickerson (1996). Structure of the Escherichia coli response regulator NarL. *Biochemistry* 35, 11053-11061.

J.D. Baleja, T. Mau, G. Wagner (1994). Recognition of Dna By Gal4 in Solution - Use of a Monomeric Protein-DNA Complex For Study By NMR. *Biochemistry* 33, 3071-3078.

K. Bender, M. Federwisch, U. Loggen, P. Nehls, M.F. Rajewsky (1996). Binding and repair of O<sup>6</sup>-ethylguanine in double-stranded oligodeoxynucleotides by recombinant human O<sup>6</sup>-alkylguanine-DNA alkyltransferase do not exhibit significant dependence on sequence context. *Nucleic Acids Research* 24, 2087-2094.

C. Berger, I. Jelesarov, H.R. Bosshard (1996). Coupled folding and site-specific binding of the GCN4-bZIP transcription factor to the AP-1 and ATF/CREB DNA sites studied by microcalorimetry. *Biochemistry* 35, 14984-14991.

K.D. Berndt (1996). Protein secondary structure. Uniform Resource Locator <http://www-structure.llnl.gov/cdtutorial.htm>.

D. Bhattacharyya, R.S. Foote, A.M. Boulden, S. Mitra (1990). Physicochemical studies of human O<sup>6</sup>-methylguanine-DNA methyltransferase. *European Journal of Biochemistry* 193, 337-343.

M. Blandamer, B. Briggs, P.M. Cullis, A.P. Jackson, A. Maxwell, R.J. Reece (1994). Domain structure of *Escherichia coli* Gyrase as revealed by differential scanning calorimetry. *Biochemistry* 33, 7510-7516.

D. Blow (1990). More of the catalytic triad. *Nature* 343, 694-695.

B. Boiteux, R. Costa Da Oliveira, J. Laval. (1985). The *Escherichia coli* O<sup>6</sup>-Methylguanine-DNA Methyltransferase Does Not Repair Promutagenic O<sup>6</sup>-Methylguanine Residues When Present in Z-DNA. *Journal of Biological Chemistry* 260, 8711-8715.

R.G. Brennan (1993). The winged-helix DNA-binding motif: another helix-turn-helix takeoff. *Cell* 74, 773-776.

T.A. Brown (1992). *Genetics: a molecular approach*. Chapman and Hall publishing.

G. Bujacz, J. Alexandratos, Q. ZhouLiu, C. ClementMella, A. Wlodawer (1996). The catalytic domain of human immunodeficiency virus integrase: Ordered active site in the F185H mutant. *Febs Letters* 398, 175-178.

I.D. Campbell, R.A. Dwek (1984). *Biological Spectroscopy*. Published by Benjamin Cummings, 156-158.

J. Cavanagh, W.J. Fairbrother, A.G. Palmer, N.J. Skelton (1996). *Protein NMR spectroscopy principles and practice*. Academic Press Inc.

I. Chaiken, S. Rose, R. Karlsson (1992). Analysis of macromolecular interactions using immobilized ligands. *Analytical Biochemistry* 201, 197-10.

C.-L. Chan, Z. Wu, T. Ciardelli, A. Eastman, E. Bresnick (1993). Kinetic and DNA-Binding Properties of Recombinant Human O6-Methylguanine-DNA Methyltransferase. *Archives of Biochemistry and Biophysics* 300, 193-200.

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

P.T. Chivers, K.E. Prehoda, B.F. Volkman, B-M. Kim, J.L. Markley, R. R.T. (1997). Microscopic pKa values of *Escherichia coli* Thioredoxin. *Biochemistry* 36, 14985-14991.

G.M. Clore, A.M. Gronenborn (1997). NMR structures of proteins and protein complexes beyond 20,000 Mr. *Nature Structural Biology* 4, 849-853.

R.T. Clubb, J. G. Omichinski, H. Savilahti, K. Mizuuchi, A. M. Gronenborn, G. M. Clore (1994). A Novel Class of Winged Helix-Turn-Helix Protein - the DNA-Binding Domain of Mu Transposase. *Structure* 2, 1041-1048.

W.J. Cook, S. R. Kar, K. B. Taylor, L. M. Hall (1998). Crystal structure of the cyanobacterial metallothionein repressor SmtB: A model for metalloregulatory proteins. *Journal of Molecular Biology* 275, 337-346.

N. Dekker, M. Cox, R. Boelens, C. P. Verrijzer, P. C. Vandervliet, R. Kaptein (1993). Solution Structure of the Pou-Specific Dna-Binding Domain of Oct-1. *Nature* 362, 852-855.

B. Demple (1995). DNA repair flips out. *Current Biology* 5, 719-721.

B. Demple (1986). Mutant *Escherichia coli* Ada protein simultaneously defective in the repair of O<sup>6</sup>-methylguanine and in gene activation. *Nucleic Acids Research* 14, 5575-5589.

B. Demple, A. Jacobsson, M. Olsson, P. Robins, T. Lindahl (1982). Repair of Alkylated DNA in *Escherichia coli* - Physical properties of O<sup>6</sup>-methylguanine-DNA methyltransferase. *Journal of Biological Chemistry* 257, 13776-13780.

B. Demple, B. Sedgwick, P. Robins, N. Totty, M.D. Waterfield, T. Lindahl (1985). Active site and complete sequence of the suicidal methyltransferase that counters alkylation mutagenesis. *Proceedings of the National Academy of Sciences USA* 82, 2688-2692.

A.J. Doherty, S.R. Ashford, J.A. Brannigan, D.B. Wigley (1995). A superior host strain for the over-expression of cloned genes using the T7 promoter based vectors. *Nucleic Acids Research* 23, 2074-2075.

M.E. Dolan, M. Oplinger, A.E. Pegg (1988). Sequence specificity of guanine alkylation and repair. *Carcinogenesis* 9, 2139-2143.

M.L. Doyle (1997). Characterisation of binding interactions by isothermal titration calorimetry. *Current opinion in Biotechnology* 8, 31-35.

F. Dyda, A. B. Hickman, T. M. Jenkins, A. Engelman, R. Craigie, D. R. Davies (1994). Crystal-Structure of the Catalytic Domain of Hiv-1 Integrase - Similarity to Other Polynucleotidyl Transferases. *Science* 266, 1981-1986.

J.C. Eads, N.M. Mahoney, S. Vorobiev, K-K. Wen, P.A. Rubenstein, B.K. Haarer, S.C. Almo (1998). Structure determination and characterisation of *Saccharomyces cerevisiae* profilin. *Biochemistry* 37, 11171-11181.

M. Federwisch, U. Hassiepen, K. Bender, M.F. Rajewsky, A. Wollmer (1997). Recombinant human O<sup>6</sup>-alkylguanine\_DNA alkyltransferase induces conformational change in bound DNA. *FEBS lett.* 407, 333-336.

R.H. Fogh, G. Otteleben, H. Ruterjans, M. Schnarr, R. Boelens, R. Kaptein (1994). Solution Structure of the Lexa Repressor Dna-Binding Domain Determined By H-1-Nmr Spectroscopy. *EMBO Journal* 13, 3936-3944.

M.P. Foster, D. S. Wuttke, K. R.Clemens, W. Jahnke, I.Radhakrishnan, L.Tennant, M.Reymond, J. Chung, P. E. Wright (1998). Chemical shift as a probe of molecular interfaces: NMR studies of DNA binding by the three amino-terminal zinc finger domains from transcription factor IIIA. *Journal of Biomolecular NMR* 12, 51-71.

M.P. Foster, D.S Wuttke, I. Radhakrishnan, D.A. Case, J.M. Gottesfeld, P.E. Wright (1997). Domain packing and dynamics in the DNA complex of the N-terminal zinc fingers of TFIIIA. *Nature Structural Biology* 4, 605-608.

M.G. Fried, S. Kanugula, J.L. Bromberg, A.E. Pegg (1996). DNA binding mechanism of O<sup>6</sup>-Alkylguanine-DNA Alkyltransferase: Stoichiometry and effects of DNA base composition and secondary structure on complex stability. *Biochemistry* 35, 15295-15301.

B.L. Gaffney, R.A. Jones (1989). Thermodynamic comparison of the base pairs formed by carcinogenic lesion O<sup>6</sup>-Methylguanine with reference both to Watson-Crick pairs and to mismatched pairs. *Biochemistry* 28, 5881-5889.

B.L. Gaffney, L.A. Marky, R.A. Jones (1984). Synthesis and Characterization of a Set of Four Dodecadeoxyribonucleoside Undecaphosphates Containing O<sup>6</sup>-Methylguanine opposite Adenine, Cytosine, Guanine and Thymine. *Biochemistry* 23, 5686-5691.

K.H. Gardner, L.E. Kay (1998). The use of <sup>2</sup>H, <sup>13</sup>C, <sup>15</sup>N, multidimensional NMR to study the structure and dynamics of proteins. *Annual Review of Biophysics and Biomolecular Structure* 27, 357-406.

D.S. Garrett, Y. J. Seok, A. Peterkofsky, G. M. Clore, A. M. Gronenborn (1998). Tautomeric state and pK(alpha) of the phosphorylated active site histidine in the N-terminal domain of enzyme I of the Escherichia coli phosphoenolpyruvate: sugar phosphotransferase system. *Protein Science* 7, 789-793.

P. Georgiadis, C.A. Smith, P.F. Swann (1991). Nitrosamine-induced Cancer: Selective Repair and Conformational Differences between O<sup>6</sup>-Methylguanine Residues in Different Positions in and around Codon 12 of rat H-ras. *Cancer Research* 51, 5843-5850.

Y. Goldgur, F. Dyda, A. B. Hickman, T. M. Jenkins, R. Craigie, D. R. Davies (1998). Three new structures of the core domain of HIV-1 integrase: An active site that binds magnesium. *Proceedings of the National Academy of Sciences of the United States of America* 95, 9150-9154.

K. Goodtzova, S. Kanugula, S. Edara, G.T. Pauly, R.C. Moschel, A.E. Pegg (1997). Repair of O<sup>6</sup>-benzylguanine by the *Escherichia coli* Ada and Ogt and the human O<sup>6</sup>-alkylguanine-DNA alkyltransferase. *Journal of Biological Chemistry* 272, 8332-8339.

K. Goodtzova, T.M. Crone, A.E. Pegg (1994). Activation of human O<sup>6</sup>-alkylguanine-DNA alkyltransferase by DNA. *Biochemistry* 33, 8385-8390.

N. Goosen, G.F. Moolenaar, R. Visse, P.V.d. Putte (1998). Functional domains of the *E.coli* UvrABC proteins in nucleotide excision repair. *Nucleic Acids and Molecular Biology* 12, 103-123.

R.J. Graves, B.F.L. Li, P.F. Swann (1989). Repair of O<sup>6</sup>-methylguanine, O<sup>6</sup>-ethylguanine, O<sup>6</sup>-isopropylguanine and O<sup>4</sup>-methylthymine in synthetic oligonucleotides by *Escherichia coli* ada gene O<sup>6</sup>-alkylguanine-DNA alkyltransferase. *Carcinogenesis* 10, 661-666.

N.J. Greenfield (1996). Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Analytical Biochemistry* 235, 1-10.

M. Gueron, J.L. Leroy (1995). Studies of base pair kinetics by NMR measurement of proton exchange. *Methods in Enzymology* 261, 383-413.

S.C. Harrison, Aggarwal, A.K. (1990). DNA recognition by proteins with the helix-turn-helix motif. *Annual Review of Biochemistry* 59, 933-969.

T.K. Hazra, R. Roy, T. Biswas, D.T. Grabowski, A.E. Pegg, S. Mitra (1997). Specific recognition of O<sup>6</sup>-methylguanine in DNA by active site mutants of human O<sup>6</sup>-methylguanine-DNA methyltransferase. *Biochemistry* 35, 5769-5776.

L. Holm, C. Sander (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Research* 26, 316-319.

L. Holm, C. Sander (1994). Searching protein structure databases has come of age. *Proteins* 19, 165-173.

S. Holm, C. Sander (1996). Mapping the protein universe. *Science* 273, 595-602.

S.M.A. Holmbeck, H. J. Dyson, P. E. Wright (1998). DNA-induced conformational changes are the basis for cooperative dimerization by the DNA binding domain of the retinoid X receptor. *Journal of Molecular Biology* 284, 533-539.

D.E. Hyre, L.D. Spicer (1995). Thermodynamic evaluation of binding interactions in the methionine repressor system of *Escheichia coli* using isothermal titration calorimetry. *Biochemistry* 34, 3212-3221.

M. Jansson, Y-C. Li, L. Jendeberg, S. Anderson, G.T. Montelione, B. Nilsson (1996). High-level production of uniformly <sup>15</sup>N- and <sup>13</sup>C-enriched fusion proteins in *Escherichia coli*. *Journal of Biomolecular NMR* 7, 131-141.

L. Jin, J. Yang, J. Carey (1993). Thermodynamics of ligand binding to *trp* repressor. *Biochemistry* 32, 7302-7309.

M.W. Kalnik, M. Kouchakdjian, B.F.L. Li, P.F. Swann, and D.J.Patel (1988a). Base Pair Mismatches and Carcinogen-Modified Bases in DNA: An NMR Study of A.C and A.O<sup>4</sup>meT Pairing in Dodecanucleotide Duplexes. *Biochemistry* 27, 100-108.

M.W. Kalnik, M. Kouchakdjian, B.F.L. Li, P.F. Swann, and D.J.Patel (1988b). Base Pair Mismatches and Carcinogen-Modified Bases in DNA: An NMR Study of G.T and G.O<sup>4</sup>meT Pairing in Dodecanucleotide Duplexes. *Biochemistry* 27, 108-115.

M.W. Kalnik, B.F.L. Li, P.F. Swann, D.J. Patel (1989b). O<sup>6</sup>-Ethylguanine carcinogenic lesions in DNA: An NMR study of O<sup>6</sup>etG:C pairing in dodecanucleotide duplexes. *Biochemistry* 28, 6182-6192.

M.W. Kalnik, B.F.L. Li, P.F. Swann, D.J. Patel (1989a). O<sup>6</sup>-Ethylguanine Carcinogenic Lesions in DNA: An NMR Study of O<sup>6</sup>etG:T Pairing in Dodecanucleotide Duplexes. *Biochemistry* 28, 6170-6181.

S. Kanugula, K. Goodtzova, S. Edara, A.E. Pegg (1995). Alteration of Arginine-128 to Alanine Abolishes the Ability of Human O<sup>6</sup>-Alkylguanine-DNA Alkyltransferase To Repair Methylated DNA but Has No Effect on Its Reaction with O<sup>6</sup>-Benzylguanine. *Biochemistry* 34, 7113-7119.

L.E. Kay, D.R. Muhandiram, N.A. Farrow, Y. Aubin, J.D. Forman-Kay (1996). Correlation between dynamics and high affinity binding in an SH2 domain interaction. *Biochemistry* 35, 361-368.

S. Klimasauskas, S. Kumar, R.J. Roberts, X. Cheng (1994). HhaI Methyltransferase Flips Its target Base Out of the DNA Helix. *Cell* 76, 357-369.

S. Klimasauskas, T. Szyperski, S. Serva, K. Wuthrich (1998). Dynamic modes of the flipped-out cytosine during HhaI methyltransferase-DNA interactions in solution. *EMBO Journal* 17, 317-324.

H. Kondo, Y. Nakabeppu, H. Kataoka, S. Kuhara, S-I. Kawabata, M. Sekiguchi (1986). Structure and expression of the *alkB* gene of *Escherichia coli* related to the repair of alkylated DNA. *Journal of Biological Chemistry* 261, 15772-15777.

A.G. Kozlov, T.M. Lohman (1998). Calorimetric studies of *E. coli* SSB protein-single stranded DNA interactions. Effects of monovalent salts on binding enthalpy. *Journal of Molecular Biology* 278, 999-1014.

P. Landini, L.I. Hajec., M.R. Volkert (1994). Structure and Transcriptional Regulation of the *Escherichia coli* Adaptive Response Gene *aidB*. *Journal of Bacteriology* 176, 6583-6589.

P. Landini, J. A. Bown, M. R. Volkert, S. J. W. Busby (1998). Ada protein-RNA polymerase sigma subunit interaction and alpha subunit-promoter DNA interaction are necessary at different steps in transcription initiation at the *Escherichia coli* *ada* and *aidB* promoters. *Journal of Biological Chemistry* 273, 13307-13312.

P. Landini, M.R. Volkert (1995). Transcriptional Activation of the *Escherichia coli* Adaptive Response Gene *aidB* is Mediated by Binding of Methylated Ada protein. *Journal of Biological Chemistry* 270, 8285-8289.

G. Laughlan, A.I.H. Murchie, D.G. Norman, M.H. Moore, P.C.E. Moody, D.M.J. Lilley, B. Luisi (1994). The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science* 265, 520-524.

G.A. Leonard, J. Thomson, W.P. Watson, T. Brown (1990). High-resolution structure of a mutagenic lesion in DNA. *Proceedings of the National Academy of Sciences USA* 87, 9573-9576.

L-Y. Lian, G.C.K. Roberts (1993). Effects of chemical exchange on NMR spectra. *NMR of Macromolecules- A Practical Approach*, Ed. G.C.K. Roberts, IRL Press.

L.-K. Liem, C-W. Wong, A. Lim, B.F.L. Li (1993). Factors influencing the repair of the mutagenic lesion O<sup>6</sup>-methylguanine in DNA by Human O<sup>6</sup>-methylguanine-DNA Methyltransferase. *Journal of Molecular Biology* 231, 950-959.

T. Lindahl (1993). Instability and decay of the primary structure of DNA. *Nature* 362, 709-715.

- T. Lindahl, B. Demple, P. Robins (1982). Suicide inactivation of the *E.coli* O<sup>6</sup>-methylguanine-DNA methyltransferase. *EMBO Journal* 1, 1359-1363.
- T. Lindahl, B. Sedgwick, M. Sekiguchi, Y. Nakabeppu (1988). Regulation and Expression of the Adaptive Response to Alkylating Agents. *Annual Review of Biochemistry* 57, 133-157.
- C. Ling-Ling, T. Nakamura, Y. Nakatsu, K. Sakumi, H. Hayakawa, M. Sekiguchi (1992). Specific amino acid sequences required for O<sup>6</sup>-methylguanine-DNA methyltransferase activity: analyses of three residues at or near the methyl acceptor site. *Carcinogenesis* 13, 837-843.
- O. Littlefield, Nelson, C.M. (1999). A new use for the 'wing' of the 'winged' helix-turn-helix motif in the HSF-DNA cocystal. *Nature Structural Biology* 6, 464-470.
- E.L. Loechler, C.L. Green, J.M. Essigmann (1984). *In vivo* mutagenesis by O<sup>6</sup>-methylguanine built into a unique site in a viral genome. *Proceedings of the National Academy of Sciences USA* 81, 6271-6275.
- T.M. Lohman, L.B. Overman, M.E. Ferrari, A.G. Kozlov (1996). A highly salt-dependent enthalpy change for *Escherichia coli* SSB protein-nucleic acid binding due to ion-protein interactions. *Biochemistry* 35, 5272-5279.
- T.M. Lohman, D.P. Mascotti (1992). Thermodynamics of ligand-nucleic acid interactions. *Methods in Enzymology* 212, 400-424.
- T. Lundback, T. Hard (1996). Sequence-specific DNA-binding dominated by dehydration. *Proceedings of the National Academy of Sciences USA* 93, 4754-4759.

G.N. Major, E.J. Gardner, P.D. Lawley (1991). Direct assay for O<sup>6</sup>-methylguanine-DNA methyltransferase and comparison of detection methods for the methylated enzyme in polyacrylamide gels and electroblots. *Biochemical Journal* 277, 89-96.

T. Maniatis, E.F. Fritsch, J. Sambrook (1989). *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press 1.

D.P. Mascotti, T.M. Lohman (1997). Thermodynamics of oligoarginines binding to RNA and DNA. *Biochemistry* 36, 7272-7279.

D.P. Mascotti, T.M. Lohman (1993). Thermodynamics of single-stranded RNA and DNA interactions with oligolysines containing tryptophan. Effects of base composition. *Biochemistry* 32, 10568-10579.

T.V. McCarthy, T. Lindahl (1985). Methyl phosphotriesters in alkylated DNA are repaired by the Ada regulatory protein of *E.coli*. *Nucleic Acids Research* 13, 2683-2698.

E. Merabet, G.K. Ackers (1995). Calorimetric analysis of  $\lambda$  cI repressor binding to DNA operator sites. *Biochemistry* 34, 8554-8563.

W.J. Metzler, L. Ponzy (1989).  $\lambda$  *cro* repressor complex with O<sub>R</sub>3 operator DNA <sup>19</sup>F nuclear magnetic resonance observations. *Journal of Molecular Biology* 205, 149-164.

F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson (1990). An integrated software system for modeling organic and bioorganic molecules using molecular mechanics. *Journal of Computational Chemistry* 11, 440-467.

C.D. Mol, A.S. Arvai, G. Slupphaug, B. Kavli, I. Alseth, H.E. Krokan, J.A. Tainer (1995). Crystal Structure and Mutational Analysis of Human Uracil-DNA Glycosylase: Structural Basis for Specificity and Catalysis. *Cell* 80, 869-878.

C.D. Mol, S.S. Parikh, C.D. Putnam, T.P. Lo, J.A. Tainer (1999). DNA repair mechanisms for the recognition and removal of damaged DNA bases. *Annual Review of Biophysics and Biomolecular Structure* 28, 101-128.

M.H. Moore, J.M. Gulbis, E.J. Dodson, B. Dimple, P.C.E. Moody (1994). Crystal structure of a suicidal DNA repair protein: the Ada O<sup>6</sup>-methylguanine-DNA methyltransferase from *E.coli*. *EMBO Journal* 13, 1495-1501.

K. Morimoto, M.E. Dolan, D. Scicchitano, A.E. Pegg (1985). Repair of O<sup>6</sup>-propylguanine and O<sup>6</sup>-butylguanine in DNA by O<sup>6</sup>-alkylguanine-DNA alkyltransferase from rat liver and *E.coli*. *Carcinogenesis* 6, 1027-1031.

K.P. Murphy, D. Xie, C. Garcia, L. M. Amzel, E. Freire (1993). Structural energetics of peptide recognition: angiotensin II/antibody binding. *Proteins: Structure, Function and Genetics* 15, 113-120.

L.C. Myers, F. Jackow, G.L. Verdine (1995a). Metal Dependence of Transcriptional Switching in *Escherichia coli* Ada. *Journal of Biological Chemistry* 270, 6664-6670.

L.C. Myers, G.L. Verdine, G. Wagner (1993b). Solution structure of the DNA methyl phosphotriester repair domain of *Escherichia coli* Ada. *Biochemistry* 32, 14089-14094.

L.C. Myers, M.P. Terranova, A.E. Ferentz, G. Wagner, G.L. Verdine (1993a). Repair of DNA Methylphosphotriesters Through a Metalloactivated cysteine Nucleophile. *Science* 261, 1164-1167.

L.C. Myers, M.P. Terranova, H.M. Nash, M.A. Markus, G.L. Verdine (1992). Zinc Binding by the Methylation Signaling Domain of the *Escherichia coli* Ada Protein. *Biochemistry* 31, 4541-4547.

L.C. Myers, T.D. Cushing, G. Wagner, G.L. Verdine (1994). Metal-Coordination sphere in the methylated Ada protein-DNA co-complex. *Chemistry and Biology* 1, 91-97.

T. Nakamura, Y. Tokumoto, K. Sakumi, G. Koike, Y. Nakabeppu, M. Sekiguchi (1988). Expression of the *ada* gene of *Escherichia coli* in response to alkylating agents. *Journal of Molecular Biology* 202, 483-494.

M. O'Driscoll, O. Humbert., P. Karran (1998). DNA mismatch repair. *Progress in nucleic Acids research and Molecular Biology* 12, 173-197.

H.-K. Oh, A. K.-C. Teo, R.B. Ali, A. Lim, T.-C. Ayi, D.B. Yarosh, B.F.L. Li (1996). Conformational change in human DNA repair enzyme O<sup>6</sup>-methylguanine-DNA methyltransferase upon alkylation of its active site by SN1 (indirect-acting) and SN2 (direct-acting) alkylating agents: breaking a "salt-link". *Biochemistry* 35, 12259-12266.

T. Ohkubo, H. Sakashita, T. Sakuma, M. Kainosho, M. Sekiguchi, K. Morikawa (1994). Methylation Dependent Functional Switch Mechanism Newly Found in the *Escherichia coli* Ada Protein. *Journal of American Chemical Society* 116, 6035-6036.

M. Olsson, T. Lindahl (1980). Repair of Alkylated DNA in *Escherichia coli*. *Journal of Biological Chemistry* 255, 10569-10571.

P. Panayotou, T. Brown, L.H. Pearl, R. Savva (1998). Direct measurement of the substrate preference of Uracil-DNA Glycosylase. *Journal of Biological Chemistry* 273, 45-50.

D.J. Patel, L. Shapiro, S.A. Kozlowski, B.L. Gaffney, R.A. Jones (1986a). Structural Studies of the O<sup>6</sup>meG.C Interaction in the d(C-G-C-G-A-A-T-T-C-O<sup>6</sup>meG-C-G) Duplex. *Biochemistry* 25, 1027-1036.

D.J. Patel, L. Shapiro, S.A. Kozlowski, B.L. Gaffney, R.A. Jones (1986b). Structural Studies of the O<sup>6</sup>meG.T Interaction in the d(C-G-T-G-A-A-T-T-C-O<sup>6</sup>meG-C-G) Duplex. *Biochemistry* 25, 1036-1042.

A. Pegg, M.E. Dolan, R.C. Moshel (1995). Structure, Function, and Inhibition of O<sup>6</sup>-Alkylguanine-DNA Alkyltransferase. *Prog. Nuc. Acids. Res. and Mol. Biol.* 51, 157-223.

A.E. Pegg, M. Boosalis, L. Samson, R.C. Moschel, T.L. Byers, K. Swenn, E. Dolan (1993). Mechanism of inactivation of human O<sup>6</sup>-alkylguanine alkyltransferase by O<sup>6</sup>-benzylguanine. *Biochemistry* 32, 11998-12006.

J.G. Pelton, D. A. Torchia, N. D. Meadow, S. Roseman (1993). Tautomeric States of the Active-Site Histidines of Phosphorylated and Unphosphorylated Iii(Glc), a Signal-Transducing Protein From *Escherichia-Coli*, Using 2-Dimensional Heteronuclear Nmr Techniques. *Protein Science* 2, 543-558.

L.A. Plesniak, G. P. Connelly, W. W. Wakarchuk, L. P. McIntosh (1996). Characterization of a buried neutral histidine residue in *Bacillus circulans* xylanase: NMR assignments, pH titration, and hydrogen exchange. *Protein Science* 5, 2319-2328.

P.M. Potter, K. Kleibl, L. Cawkwell, G.P. Margison (1989). Expression of the *ogt* gene in wild-type and *ada* mutants of *E.coli*. *Nucleic Acids Research* 17, 8049-8060.

J.A. Rafferty, J. Tumelty, M. Skorvaga, R.H. Elder, G.P. Margison, K.T. Douglas (1994). Site-directed mutagenesis of glutamic acid 172 to glutamine completely inactivated human O<sup>6</sup>-alkylguanine-DNA-alkyltransferase. *Biochemistry and Biophysics Research Communications* 199, 285-291.

V. Ramesh, R. O. Frederick, S. E. H. Syed, C. F. Gibson, J. C. Yang, G. C. K. Roberts (1994). The Interactions of *Escherichia-Coli* Trp Repressor With Tryptophan and With an

Operator Oligonucleotide Nmr-Studies Using Selectively N- 15-Labeled Protein. European Journal of Biochemistry 225, 601-608.

F. Rastinejad, P. Artz, L. Ponzy (1993). Origin of the asymmetrical contact between *lac* repressor and *lac* operator DNA. Journal of Molecular Biology 233, 389-399.

G.W. Rebeck, S. Coons, P. Carroll, L. Samson (1988). A second DNA methyltransferase repair enzyme in *Escherichia coli*. Proceedings of the National Academy of Sciences USA 85, 3039-3043.

K.M. Reinisch, L. Chen, G.L. Verdine, W.N. Lipscomb (1995). The Crystal Structure of HaeIII Methyltransferase Covalently Complexed to DNA: An Extrahelical Cytosine and Rearranged Base Pairing. Cell 82, 143-153.

P. Rice, K. Mizuuchi (1995). Structure of the Bacteriophage-Mu Transposase Core - a Common Structural Motif For Dna Transposition and Retroviral Integration. Cell 82, 209-220.

R.J. Roberts (1995). On Base Flipping. Cell 82, 9-12.

P. Robins, J. Cairns (1979). Quantitation of the adaptive response to alkylating agents. Nature 280, 74-76.

R. Rydberg, T. Lindahl (1982). Nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. EMBO Journal 1, 211-216.

B.M. Saget, D.E. Shevell, G.C. Walker (1995). Alteration of lysine 178 in the hinge region of the *Escherichia coli* Ada protein interferes with activation of *adabut* not *alkA*, transcription. Journal of Bacteriology 177, 1268-1274.

B.M. Saget, G.C. Walker (1994). The Ada protein acts as both a positive and a negative modulator of *Escherichia coli*'s response to methylating agents. Proceedings of the National Academy of Sciences USA *91*, 9730-9734.

H. Sakashita, T. Sakuma, T. Ohkubo, M. Kainosho, K. Sakumi, M. Sekiguchi, K. Morikawa (1993). Folding topology and DNA binding of the N-terminal fragment of Ada protein. FEBS Letters *323*, 252-256.

H. Sakashita, T. Sakuma, T. Ohkubo, Y. Akitomo, M. Kainosho, M. Sekiguchi, K. Morikawa (1995). Sequence-Specific DNA Recognition of the *Escherichia coli* Ada Protein Associated with the Methylation-Dependent Functional Switch for Transcriptional Regulation. Journal of Biochemistry *118*, 1184-1191.

K. Sakumi, M. Sekiguchi (1989). Regulation of expression of the *ada* gene controlling the adaptive response. Journal of Molecular Biology *205*, 373-385.

L. Samson, J.Cairns (1977). A new pathway for DNA repair in *Escherichia coli*. Nature *267*, 281-282.

A. Sancar, G.B. Sancar (1988). DNA repair enzymes. Ann. Rev. Biochem. *57*, 29-67.

P. Schar, J. Jiricny (1998). Eukaryotic mismatch repair. Nucleic Acids and Molecular Biology *12*, 199-247.

S.C. Schultz, G.C. Shields, T.A. Steitz (1991). Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. Science *253*, 1001-1007.

J.W.R. Schwabe, A.A. Travers (1993). What is evolution playing at ? Current Biology *3*, 628-630.

D. Scicchitano, R.A. Jones, S. Kuzmich, B.Gaffney, D.D. Lasko, J.M. Essigmann, A.E. Pegg (1986). Repair of oligodeoxynucleotides containing O<sup>6</sup>methylguanine by O<sup>6</sup>-alkylguanine DNA-alkyltransferase. *Carcinogenesis* 7, 1383-1386.

B. Sedgwick (1989). In Vitro Proteolytic Cleavage of the *Escherichia coli* Ada Protein by the *ompT* Gene Product. *J. Bacteriol.* 171, 2249-2251.

B. Sedgwick, P. Robins, N. Totty, T. Lindahl (1988). Functional Domains and Methyl Acceptor Sites of the *Escherichia coli* Ada protein. *Journal of Biological Chemistry* 263, 4430-4433.

X. Shan, K. H. Gardner, D. R. Muhandiram, L. E. Kay, C. H. Arrowsmith (1998). Subunit-specific backbone NMR assignments of a 64 kDa trp repressor DNA complex: A role for N-terminal residues in tandem binding. *Journal of Biomolecular NMR* 11, 307-318.

D.E. Shevell, A.M. Abou-Zamzam, B. Demple, G.C. Walker (1988). Construction of an *Escherichia coli* K-12 *ada* Deletion by Gene Replacement in a *recD* Strain Reveals a Second Methyltransferase That Repairs Alkylated DNA. *J. Bacteriol.* 170, 3294-3296.

D.E. Shevell, B.M. Friedman, G.C. Walker (1990). Resistance to alkylation damage in *Escherichia coli*: Role of the Ada protein in induction of the adaptive response. *Mutation Research* 223, 53-72.

D.E. Shevell, G.C. Walker (1991). A region of the Ada DNA-repair protein required for the activation of *ada* transcription is not necessary for activation of *alkA*. *Proceedings of the National Academy of Sciences USA* 88, 9001-9005.

M. Slijper, R. Boelens, A. L. Davis, R. N. H. Konings, G. A. vanderMarel, J. H. vanBoom, R. Kaptein (1997). Backbone and side chain dynamics of lac repressor headpiece (1-56) and its complex with DNA. *Biochemistry* 36, 249-254.

G. Slupphaug, C.G. Mol, B.Kavli, A.S. Arvai, H.E. Krokan, J.A. Tainer (1996). A nucleotide-flipping mechanism from the structure of human uracil-DNA glycosylase bound to DNA. *Nature* 384, 87-92.

T.E. Spratt, Wu, J.D., Levy, D.E., Kanugula, S., Pegg, A.E. (1999). Reaction and binding of oligodeoxynucleotides containing analogues of O<sup>6</sup>-methylguanine with wild-type and mutant human O<sup>6</sup>-alkylguanine-DNA alkyltransferase. *Biochemistry* 38, 6801-6806.

T.E. Spratt, C.R. Cambell (1994). Synthesis of Oligodeoxynucleotides containing analogs of O<sup>6</sup>-methylguanine and reaction with O<sup>6</sup>-alkylguanine-DNA alkyltransferase. *Biochemistry* 33, 11364-11871.

T.E. Spratt, H.d.l. Santos (1992). Reaction of O<sup>6</sup>-Alkylguanine-DNA Alkyltransferase with O<sup>6</sup>-methylguanine analogues: Evidence that the oxygen of O<sup>6</sup>-methylguanine is protonated by the protein to effect methyl transfer. *Biochemistry* 31, 3688-3694.

M. Sriram, G.A. Van der Marel, H.L.P.F. Roelen, J.H. Van Boom, A.H.-J. Wang (1992). Conformation of B-DNA containing O<sup>6</sup>-ethyl-G-C base pairs stabilized by minor groove binding drugs: molecular structure of d(CGC[e<sup>6</sup>G]AATTCGCG complexed with Hoechst 33258 or Hoechst 33342. *EMBO Journal* 11, 225-232.

W.F. Studier, B.A. Moffatt (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology* 189, 113-130.

M. Takahashi, K. Sakumi, M. Sekiguchi (1990). Interaction of Ada protein with DNA examined by Fluorescence anisotropy of the protein. *Biochemistry* 29, 3431-3436.

K. Takano, Y. Nakabeppu, M. Sekiguchi (1988). Functional sites of the Ada regulatory protein of *Escherichia coli*. *Journal of Molecular Biology* 201, 261-271.

H.-B. Tan, P.F. Swann, E.M. Chance (1994). Kinetic Analysis of the Coding Properties of O<sup>6</sup>-Methylguanine in DNA: The Crucial Role of the Conformation of the Phosphodiester Bond. *Biochemistry* 33, 5335-5346.

N.-W. Tan, B.F.L. Li (1990). Interaction of Oligonucleotides Containing 6-O-Methylguanine with Human DNA (Cytosine-5-)-methyltransferase. *Biochemistry* 29, 9234-9240.

K. Tano, D. Bhattacharyya, R.S. Foote, R.J. Mural, S. Mitra (1989). Site-directed mutation of the *Escherichia coli ada* gene: effects of substitution of methyl acceptor cysteine-321 by histidine in Ada protein. *J. Bacteriol.* 171, 1535-1543.

P. Taverna, B. Sedgwick (1996). Generation of an Endogenous DNA-Methylating Agent by Nitrosation in *Escherichia coli*. *Journal of Bacteriology* 178, 5105-5111.

I. Teo, B. Sedgwick, M.W. Kilpatrick, T.V. McCarthy, T. Lindahl (1986). The intracellular signal for induction of resistance to alkylating agents in *E.coli*. *Cell* 45, 315-324.

J. Thomale, K. Hochleitner, M.F. Rajewsky (1994). Differential formation and repair of the mutagenic DNA alkylation product O<sup>6</sup>-ethylguanine in transcribed and nontranscribed genes of the rat. *Journal of Biological Chemistry* 269, 1681-1686.

M.L. Todd, P.F. Schendel (1983). Repair and Mutagenesis in *Escherichia coli* K-12 After Exposure to Various Alkyl-Nitrosoguanidines. *J. Bacteriol.* 156, 6-12.

M.D. Topal, J.S. Eadie, M. Conrad (1986). O<sup>6</sup>-methylguanine Mutation and Repair Is Nonuniform. *Journal of Biological Chemistry* 261, 9879-9885.

D.G. Vassilyev, K. Morikawa (1997). DNA-repair enzymes. *Current Opinion in Structural Biology* 7, 103-109.

D.G. Vassilyev, K. Morikawa (1996). Precluding uracil from DNA. *Structure* 4, 1381-1385.

D.G. Vassilyev, T. Kashiwagi, Y. Mikami, M. Ariyoshi, S. Iwai, E. Ohtsuka, K. Morikawa (1995). Atomic Model of a Pyrimidine Dimer Excision Repair Enzyme Complexed with a DNA Substrate: Structural Basis for Damaged DNA Recognition. *Cell* 83, 773-782.

P. Vaughan, B. Sedgwick, J. Hall, J.Gannon, T. Lindahl (1991). Environmental mutagens that induce the adaptive response to alkylating agents in *Escherichia coli*. *Carcinogenesis* 12, 263-268.

P.E. Verdemato, P.C.E. Moody (1998). Repair of alkylated DNA by the *E.coli* ada protein. *Nucleic Acids and Molecular Biology* 12, 1-27.

G.L. Verdine, S.D. Bruner (1997). How do DNA repair proteins locate damaged bases in the genome ? *Chemistry & Biology* 4, 329-334.

J.M. Voigt, M.D. Topal (1990). O<sup>6</sup>-Methylguanine and A.C and G.T Mismatches Cause Asymmetric Structural defects in DNA That Are Affected by DNA Sequence. *Biochemistry* 29, 5012-5018.

J. Vojtechovsky, M.D. Eaton, B. Gaffney, R. Jones, M. Berman (1995). Structure of a New Crystal Form of a DNA Dodecamer Containing T.(O<sup>6</sup>Me)G Base Pairs. *Biochemistry* 34, 16632-16640.

P.H. von Hippel, O.G. Berg (1989). Facilitated target location in biological systems. *Journal of Biological Chemistry* 264, 675-678.

R.A. Vora, A.E. Pegg, S.E. Ealick (1998). A new model for how O<sup>6</sup>-methylguanine-DNA methyltransferase binds DNA. *Proteins: Structure, Function and Genetics* 32, 3-6.

G.W. Vuister, S. J. Kim, C. Wu, A. Bax (1994). Nmr Evidence For Similarities Between the Dna-Binding Regions of *Drosophila-Melanogaster* Heat-Shock Factor and the Helix-Turn-Helix and Hnf-3/Forkhead Families of Transcription Factors. *Biochemistry* 33, 10-16.

Y.-F. Wei, B.J. Chen, S. Samson (1995). Suppression of *Escherichia coli alkB* mutants by *Saccharomyces cerevisiae* genes. *J. Bacteriol.* 177, 5009-5015.

M. Weinfeld, A.F. Drake, J.K. Saunders, M.C. Paterson (1985). Stereospecific removal of methyl phosphotriesters from DNA by an *Escherichia coli ada+* extract. *Nucleic Acids Research* 13, 7067-7077.

J.E.A. Wibley, K. McKie, J.H. Embrey, D.S. Marks, K.T. Douglas, M.H. Moore, P.C.E. Moody (1995). A homology model of the three-dimensional structure of human O<sup>6</sup>-alkylguanine-DNA alkyltransferase based on the crystal structure of the C-terminal domain of Ada protein from *Escherichia coli*. *Anti-cancer Drug Design* 10, 75-95.

J.J. Wilker, S.J. Lippard (1997). Alkyl Transfer to Metal Thiolates: Kinetics, Active Species Identification, and relevance to the DNA Methyl Phosphotriester Repair Center of *Escherichia coli* Ada. *Inorganic Chemistry* 36, 969-978.

K.P. Wilson, L. M. Shewchuk, R. G. Brennan, A. J. Otsuka, B.W. Matthews (1992). *Escherichia-Coli* Biotin Holoenzyme Synthetase Biorepressor Crystal- Structure Delineates the Biotin-Binding and Dna-Binding Domains. *Proceedings of the National Academy of Sciences of the United States of America* 89, 9257-9261.

N.A. Wilson, E. Barbar, J.A. Fuchs, C. Woodward (1995). Aspartic acid 26 reduced *Escherichia coli* thioredoxin has a pKa > 9. *Biochemistry* 34, 8931-8939.

T. Wiseman, S. Williston, J.F. Brandts, L-N Lin (1989). Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Analytical Biochemistry* 179, 131-137.

D.S. Wishart, B.D. Sykes (1994). The C-13 Chemical-Shift Index - a Simple Method For the Identification of Protein Secondary Structure Using C-13 Chemical- Shift Data. *Journal of Biomolecular NMR* 4, 171-180.

D.S. Wishart, B.D. Sykes, F.M. Richards (1992). The Chemical-Shift Index - a Fast and Simple Method For the Assignment of Protein Secondary Structure Through Nmr-Spectroscopy. *Biochemistry* 31, 1647-1651.

C.-W. Wong, N-W. Tan, B.F.L. Li (1992). Structure-related properties of the mutagenic lesion O<sup>6</sup>-methylguanine in DNA. *Journal of Molecular Biology* 228, 1137-1146.

Y. Yamagata, M. Kato, K. Odawara, Y. Tokuno, Y. Nakashima, N. Matsushima, K. Yasumura, K-I. Tomita, K. Ihara, Y. Fujii, Y. Nakabeppu, M. Sekiguchi, S. Fujii (1996). Three-Dimensional Structure of a DNA Repair Enzyme, 3-Methyladenine DNA Glycosylase II, from *Escherichia coli*. *Cell* 86, 311-319.

D.B. Yarosh, S. Hurst-Calderone, M.A. Babich, R.S. Day III (1986). Inactivation of O<sup>6</sup>-methylguanine-DNA methyltransferase and sensitization of human tumor cells to killing by chloroethylnitrosourea as a free base. *Cancer Research* 46, 1663-1668.

P. Zhou, L. J. Sun, V. Dotsch, G. Wagner, G.L. Verdine (1998). Solution structure of the core NFATC1/DNA complex. *Cell* 92, 687-696.