Biophysical and structural studies of the antirestriction proteins ArdA and KlcA

Dimitra Serfiotis-Mitsa

Ph.D
Biophysical Chemistry

School of Chemistry
University of Edinburgh
October 2009

Declaration

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

Dimitra Serfiotis-Mitsa

Acknowledgments

I would like to thank my supervisor, Dr. David Dryden, for giving me the opportunity to work on the interesting topic of antirestriction.

I would also like to thank all those people who made this thesis possible: Dr. Gareth Roberts, Laurie Cooper and Dr. John White for their continuous support and guidance throughout these three years, Dr. Dusan Uhrin and Dr. Juraj Bella for their help with everything NMR-related, Dr. Andy Herbert for his help with protein structure calculation, Dr. Dinesh Soares for the bioinformatics analysis, and Dr. Garry Blakely for helpful discussions in bacterial genetics.

Finally, many thanks go to my student colleagues, Avgoustino Stephanou, Bansi Sanghvi, Marcel Reuter and Kai Chen with whom I had many enjoyable conversations, and whose lively sense of humour made for a pleasant working atmosphere.

Abstract

Gene *orf18*, which is situated in the conjugative transposon Tn*916* from the bacterial pathogen *Enterococcus faecalis*, encodes a putative ArdA (alleviation of restriction of DNA) protein. ArdA from Tn*916* may be responsible for the apparent immunity of the transposon to DNA restriction and modification (R/M) systems and for ensuring that the transposon has a broad host range. The *orf18* gene was engineered for overexpression in *Escherichia coli* and the recombinant ArdA protein was purified to homogeneity. Biophysical characterisation of ArdA demonstrated tight association between ArdA and the M.EcoKI. Also, ArdA was shown to efficiently inhibit restriction and modification by all four major classes of Type I R/M enzymes *in vivo*. Thus, ArdA can overcome the restriction barrier following conjugation and so helps to increase the spread of antibiotic resistance genes by horizontal gene transfer.

The amino acid sequence of KlcA, from the incompatibility plasmid pBP136 from *Bordetella pertussis*, showed a high degree of similarity with the antirestriction protein ArdB from the IncN plasmid pKM101. In this study the solution structure of KlcA was solved with high-resolution NMR and its antirestriction function demonstrated. The structure of KlcA showed a rigid globular molecule with a novel fold. No antimodification function was observed for KlcA *in vivo* and the antirestriction function of KlcA has been successfully shown *in vivo* but not *in vitro*. Because no direct binding of KlcA to EcoKI was observed *in vitro*, the mechanism of the endonuclease blocking was assumed to be different from that of ArdA. Preliminary experiments including coimmunoprecipitation assays were conducted in order to elucidate the antirestriction mechanism of KlcA.

Abbreviations

BN	Blue Native
CD	Circular Dichroism
CO-IP	Co-Immunoprecipitation
DSC	Differential Scanning Calorimetry
EOP	Efficiency Of Plating
GdmCl	Guanidinium Chloride
HGT	Horizontal Gene Transfer
HPLC	High Pressure Liquid Chromatography
ITC	Isothermal Titration Calorimetry
MALDI	Matrix Assisted Laser Desorption Ionisation
MS	Mass Spectrometry
MTase	Methyltransferase
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
PAGE	Polyacrylamide Gel Electrophoresis
RA	Restriction Alleviation
RDC	Residual Diporal Coupling
REase	Restriction Endonuclease
SAM	S-adenosyl-L-methionine
SDS	Sodium Dodecyl Sulfate
TOF	Time Of Flight
TRD	Target Recognition Domain

Table of Contents

Declaration i

Acknowledgments ii

Abstract iii

Abbreviations iv

1 Introduction

1.1 Horizontal gene transfer and antibiotic resistance 1

1.2 Types of horizontal gene transfer 1

1.3 Conjugative transposons 2

1.3.1 The conjugative transposon Tn*916* 5

1.4 Restriction-Modification systems 7

1.4.1Type I restriction-modification systems 7

1.4.2 Type II restriction-modification systems 14

1.4.3 Type III restriction-modification systems 15

1.4.4 Type IV restriction systems 16

1.5 Antirestriction Strategies 17

1.5.1 Virus-coded DNA modifications 17

1.5.2 Viral stimulation of host modification functions 18

1.5.3 Incorporation of unusual bases in viral DNA 18

1.5.4 Counterselection of restriction sites in phage genomes 19

1.5.5 Coinjection of proteins occluding restriction sites 20

1.5.6 Host-mediated restriction alleviation 20

1.5.7 Production of proteins that inhibit restriction enzymes 23

1.5.7.1 The Ard proteins 26

1.5.7.2 The KlcA proteins: ArdB homologues 28

2 Background theory of protein biophysical and structural methods

2.1 Steady-state fluorescence 32

2.2 Differential Scanning Calorimetry (DSC) 38

2.3 Isothermal Titration Calorimetry (ITC) 42

2.4 Circular Dichroism (CD) 45

2.5 Matrix Assisted Laser Desorption Ionisation- Time Of Flight/ Mass Spectrometry (MALDI-TOF/MS) 48

2.6 Nuclear Magnetic Resonance (NMR) Spectroscopy 51

2.6.1 Chemical shift 52

2.6.2 Indirect spin-spin interaction 54

2.6.3 Nuclear Overhauser Effect (NOE) 56

2.6.4 Spin perturbation 57

2.6.5 Spin pelaxation 60

2.6.6 Residual Dipolar Couplings (RDCs) 63

2.6.7 NMR experiments for protein structure determination 68

2.6.7.1 2D experiments for backbone and side-chain assignment 68

2.6.7.2 3D experiments for backbone assignment 69

2.6.7.3 3D experiments for side-chain assignment 73

2.6.7.4 2D experiments for the assignment of aromatic residues 75

2.6.7.5 3D Experiments for the acquisition of NOE restraints 75

3 Methods

3.1 Expression and purification of ArdA/ KlcA 77
3.2 Size-exclusion HPLC 78
3.3 Circular Dichroism 79
3.4 Fluorescence quenching for the determination of tryptophan accessibility 79
3.5 Determination of the unfolding free energy ΔG using fluorescence spectroscopy 80
3.6 DSC 81
3.7 ITC 82
3.8 *In vivo* Restriction/Modification assays 82
3.9 *In vitro* antirestriction activity of ArdA/ KlcA 84
3.10 *In vitro* assay of DNA displacement from M.EcoKI by ArdA 85
3.11 NMR spectroscopy and structure calculations 85
3.12 KlcA fold recognition and comparison 89
3.13 KlcA conservation and surface analysis 89
3.14 Purification of DNA fragments from PCR reaction 90
3.15 DNA extraction from agarose gels 91
3.16 SDS-PAGE 92
3.17 BN-PAGE 92
3.18 DNA electrophoresis 93
3.19 Cloning Procedure 93
3.20 Plasmid transformation 95
3.21 Plasmid DNA isolation 95
3.22 Western Blot 96

3.23 In-gel protein digestion with trypsin 98
3.24 MALDI-TOF-MS 98
3.25 Co-Immunoprecipitation assay 99
3.26 His-tagged protein purification 100

4 ArdA Results and Discussion

4.1 Cloning, overexpression and purification of *orf18* 102
4.2 Secondary structure analysis of ArdA 105
4.3 Size-exclusion HPLC 108
4.4 Fluorescence quenching experiments for the determination of tryptophan topology 111
4.5 Determination of the free energy ΔG of unfolding 114
4.6 Differential scanning calorimetry 116
4.7 ArdA–M.EcoKI interaction by ITC 117
4.8 Restriction/Modification assays 120
4.8.1 *In vivo* restriction/modification assay 120
4.8.2 *In vitro* antirestriction activity of ArdA 122
4.9 *In vitro* assay of DNA displacement from M.EcoKI by ArdA 125
4.10 Preliminary NMR experiments with ArdA 127
4.11 Truncated ArdA 134
4.12 X-ray crystal structure of ArdA 138
4.13 Conclusion for ArdA experiments 143

5 KlcA Results and Discussion

5.1 ^{1}H, ^{15}N and ^{13}C chemical shift assignments and structure

calculations 147
5.2 The three-dimensional structure of KlcA 155
5.3 Backbone dynamics from ^{15}N relaxation data 161
5.4 Antirestriction activity of KlcA 169
5.5 Expression of KlcA is not lethal to *E. coli* 171
5.6 KlcA does not bind to MTase 172
5.7 KlcA antirestriction function does not involve the ClpXP protease 172
5.8 BN/SDS-PAGE 175
5.9 c-myc tagged KlcA co-IP assay 182
5.10 Conclusion for KlcA experiments 187

6 Final remarks and future work
6.1 Final remarks 189
6.2 Future work 190
6.2.1 BN/SDS-PAGE 190
6.2.2 Far-Western Blotting 191
6.2.3 *In vivo* crosslinking 191
6.2.4 Yeast-two-hybrid method 192

References 194
Appendix 215

Chapter 1: Introduction

1.1 Horizontal gene transfer and antibiotic resistance

The natural transfer of genetic material between bacteria in the environment is necessary for the generation of genetic diversity and provides the raw material for natural selection and evolution. The ability of bacteria to adapt to new environments, to be able to overcome selective pressures or to become pathogenic is often the result of gene transfer, specifically horizontal gene transfer. Horizontal gene transfer between bacterial populations is of particular medical importance since genes that confer resistance to antibiotics can pass from one bacterium to another. The discovery of plasmid-mediated antibiotic resistance in *Shigella* in 1959 (Watanabe, 1963) marked the beginning of a series of discoveries that showed that plasmids and transposons can be important agents for the transfer of antibiotic resistance between bacteria. Antibiotic-resistant bacterial infections are now becoming increasingly common with the appearance of multi-drug resistant strains such as methicilin-resistant *Staphylococci*, vancomycin resistant *Enterococci* and extended-spectrum betalactamase (ESBL) producing Gram- negative bacilli (Dzidic & Bedekovic, 2003).

1.2 Types of horizontal gene transfer

There are three principal mechanisms that allow the horizontal transfer of genes: transformation, transduction and conjugation. In transformation,

prokaryotes can take up free (naked) DNA from their surroundings. In transduction, bacteriophage particles act as vectors and incorporate genes by "injection" into the bacterial chromosome either through the lytic or the lysogenic cycle of the phage. Transduction can be generalised where any bacterial gene may be transferred, or specialised, where only genes located near the site of prophage integration are transferred) (Davison, 1999). The third mechanism of horizontal gene transfer is conjugation (or mating), where bacteria are capable of attaching to each other for the purpose of transmitting DNA from a donor cell to a recipient cell. For this purpose, some bacteria use pili (F pilus in *E. coli*) that adhere to the recipient's cell surface and DNA transfer takes place at the site of contact. Not all bacteria use pili and a notable exception is *Enterococcus faecalis* where the molecules necessary for cell-cell contact (adhesins) are located on the cell surface rather than on pili (White, 2007). Conjugation, can be of several types (Davison, 1999): a) transfer of a self-transmissible conjugative plasmid, b) transfer of a nonself-transmissible plasmid containing an OriT by the action of a conjugative plasmid (mobilisation), c) a non-self-transmissible non-mobilisable plasmid fuses with a self-transmissible partner (cointegration) and d) by conjugative transposons.

1.3 Conjugative transposons

Transposable elements, or transposons, are sequences of DNA that can move from one site to another within a bacterial genome. Transposition does not require extensive regions of DNA homology and this differentiates transposition from other forms of genetic recombination. The enzyme *transposase* acts on a specific DNA sequence at each end of the transposon disconnecting it from the

flanking DNA and then inserting it into the new target site. Transposons can be grouped into three main classes based on their structure and transposition mechanisms: a) DNA-only transposons (move as DNA either excising or following a replicative pathway), b) Non-retroviral retrotransposons (encode reverse transcriptase, move via an RNA intermediate), c) Retroviral-like transposons (encode reverse transcriptase, move via an RNA intermediate, resemble retroviruses). DNA–only transposable elements can be further subdivided into:

- Insertion sequences (simplest form of transposable element, carry no information other than that required for transposition itself, move by conservative transposition)
- Composite transposons (comprise two identical, or nearly identical, insertion sequences that move by conservative transposition)
- Type II transposons (move by replicative transposition increasing the number of copies of the transposon)
- Conjugative transposons (form a covalently closed circular intermediate)

Neither replicative nor conservative transposition involves the generation of a free transposon molecule in the cell. However, conjugative transposons form a circular intermediate that can reinsert into host genomic DNA or it can transfer by conjugation to a new host. The transposon is excised from the donor molecule, it circularises and inserts itself into a target site in the recipient molecule. The following model (Fig. 1) has been proposed for the excision and integration of a conjugative transposon (Scott & Churchward, 1995).

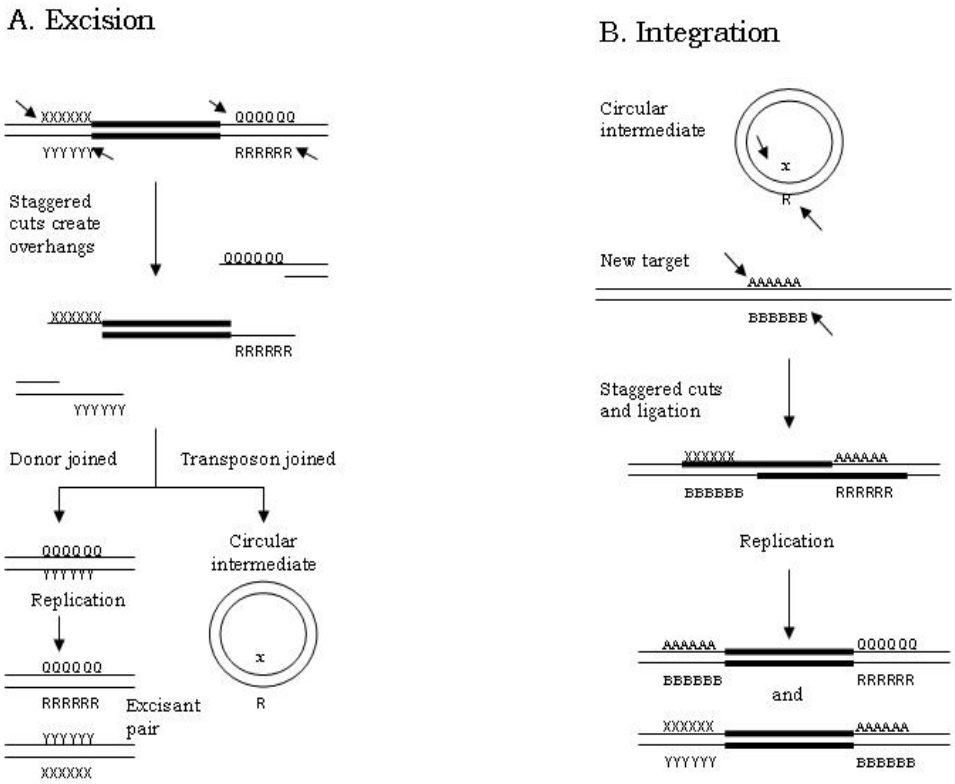


Fig. 1 *A proposed model for the excision and integration of a conjugative transposon according to Scott & Churchward, 1995. The thick lines represent the transposon, the thin lines are the DNA adjacent to the transposon and the letters X-Y, Q-R and A-B represent hypothetical complementary nucleotide pairs. At first, the transposon is excised from the donor molecule. The ends are joined to create an excisant molecule and the circularised transposon (A). Semi-conservative replication produces an excisant pair. Cuts in the circularised transposon and the new DNA target followed by ligation create*

a new insertion of the transposon with a heteroduplex at its end. Replication follows, which resolves the heteroduplexes (B).

1.3.1 The conjugative transposon Tn*916*

Found in *Enterococcus faecalis* DS16 (formerly known as *Streptococcus faecalis* DS16), Tn*916* was the first conjugative transposon identified (Franke & Clewell, 1981), the first to have its nucleotide sequence determined (Flannagan *et al.*, 1994) and the most well studied. Fig. 2 shows a linear map of Tn*916*.

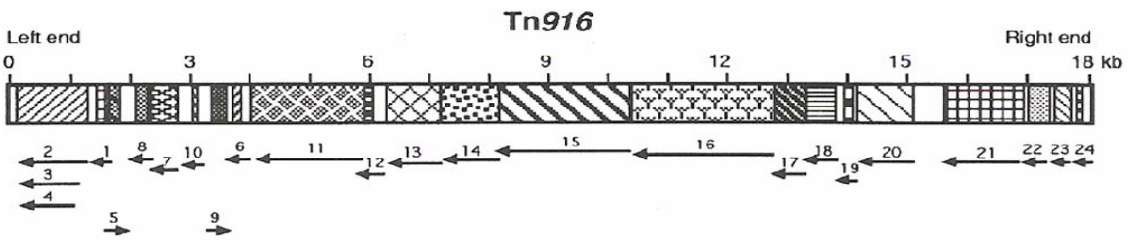


Fig. 2. *Linear map of Tn916 showing open reading frames (ORFs) represented by patterned segments. The assigned number for each ORF is shown above an arrow that indicates the direction of transcription and the approximate extent of the ORF. Reprinted from Flannagan et al., 1994.*

Tn*916* is located chromosomally. Compared to transposons greater in size than 50 kb - as for example Tn*3705* (>52 kb) in *Streptococcus auginosus* (Clermont & Horaud, 1994), Tn*3701* (50 kb) of *Streptococcus pyogenes* (Le Bouguenec *et al.*, 1988), and the multiresistance transposon Tn*5385* (65 kb) also present in *Enterococcus faecalis* (Rice & Carias, 1998) - Tn*916,* being 18.3 kilobases long, is a relatively small conjugative transposon. Tn*916* carries the *tet(M)* gene which provides resistance to both tetracycline-and unlike most other *tet* genes- to minocycline as well. Instead of causing resistance by facilitating the export of the antibiotic as is the case with other *tet* gene products, *tet(M)* encodes a protein that causes the process of protein synthesis to become insensitive to tetracycline (Burdett, 1993). Most Tn*916*-like elements contain *tet(M)* genes. In some cases, however, Tn916-like elements have acquired additional resistance determinants, as in the pneumococcal conjugative transpososon Tn*1545* which encodes kanamycin and erythromycin resistance in addition to tetracycline resistance (Courvalin & Carlier 1987). In other instances, Tn*916*-like elements have been discovered incorporated within larger multiresistance elements, such as in the *Enterococus faecalis* transposon Tn*5385* that encodes resistance to erythromycin, gentamycin, streptomycin, tetracycline and penicillin (Rice & Carias, 1998).

Tn*916* has been shown both *in vitro* and *in vivo* to transfer readily to a wide variety of Gram-positive and Gram-negative bacteria. In one report, Tn*916* was able to transfer from *E. faecalis* to *Alcaligens eutrophus*, *Citrobacter freundii* and *E. coli* (Bertram *et al.*, 1991). Other examples include the transfer of Tn*916* from *E. faecalis* to the Gram-positive *Bacillus licheniformis* (Herzog-Velikonja et al, 1994) and to the Gram-negative *Butyrivibrio fibrisolvens* (Hespell & Whitehead, 1991).

1.4 Restriction-Modification (R-M) systems

R-M systems are widely believed to have evolved as cellular defense mechanisms against infecting foreign DNA identified as alien by lack of cognate methylation. They are designated by a three letter acronym derived from the name of the organism in which they occur. The first letter comes from the genus and the second and third letters come from the species. For example, EcoKI is the first restriction enzyme identified in *E. coli* K-12. R-M systems are classified into four major groups on the basis of composition and cofactor requirements, cleavage position and target sequence characteristics (Bickle & Kruger, 1993; Roberts *et al.*, 2003).

1.4.1 Type I restriction-modification systems

Type I R-M systems are multifunctional enzymes that can catalyse both restriction and modification. They are composed of three subunits encoded by the genes *hsd*R, *hsd*M and *hsd*S (*hsd* denotes "host specificity of DNA") (Glover & Colson, 1969). HsdM (~50-60 kDa) and HsdS (~50 kDa) subunits comprise a methyltransferase (M_2S_1) which in association with HsdR forms an endonuclease ($R_2M_2S_1$). The R (~140 kDa) subunits are required for restriction, the S subunit specifies the DNA target sequence and the M subunits catalyse the methylation reaction.

The S subunit includes two target recognition domains (TRDs) that impart target sequence specificity to both the restriction and modification activities of

the complex. The M subunit includes the binding site of the cofactor and methyl donor SAM (S-adenosyl-L-methionine) and the active site for DNA methylation. Methylation occurs at the N6 position of adenine yielding N6-methyladenine. The R subunits include the active site for ATP hydrolysis and other sequences essential for DNA translocation and endonuclease activity. The restriction reaction requires SAM, ATP and Mg^{2+}. Depending upon the methylation state of the DNA, the R-M complex can function as either a REase or an MTase. If the target sequence is hemi-methylated, DNA is targeted for further methylation. Fully methylated DNA is immune to restriction. If the target sequence is unmodified (unmethylated) then it is targeted for restriction. The enzyme, while bound to its target site, is believed to translocate, or pull, the DNA towards itself simultaneously in both directions in an ATP-dependent manner. DNA translocation allows the cleavage to occur several hundreds to thousands of base pairs away from the recognition sequence. DNA cleavage occurs when translocation is impeded, either by collision with another translocating complex or by the topology of the DNA substrate (Murray, 2000). Recent atomic force microscopy measurements revealed that the EcoKI enzyme goes through an additional dimerisation stage before the onset of translocation (Neaves et *al.*, 2009). The new model (Fig. 3) suggests the following mode of action of EcoKI: an EcoKI monomer binds to DNA at a random sequence and then through linear diffusion it finds its target sequence. A second monomer of EcoKI then binds to the first one. The dimer that is formed collides with another region of the DNA due to the flexibility of DNA and a loop is formed. Once the loop is formed the second EcoKI molecule scans the DNA for the second target

sequence, and once both target sequences are identified translocation and subsequent cleavage of the DNA occurs.

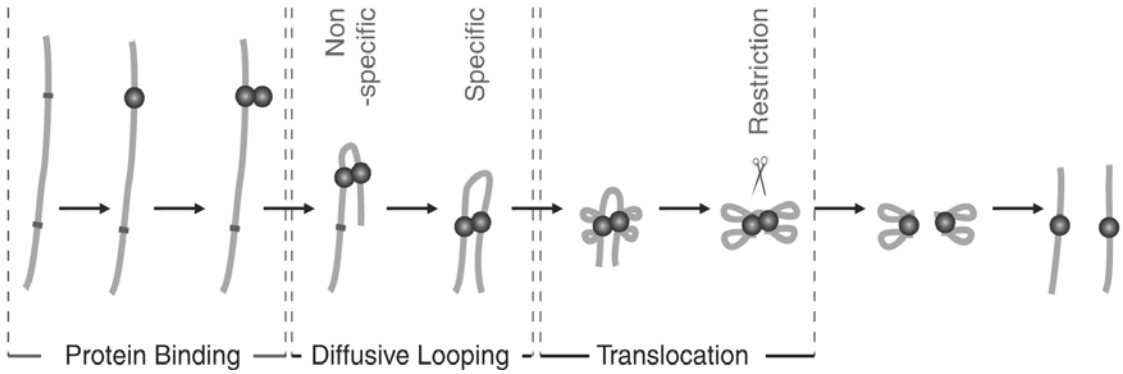


Fig. 3 *Diagrammatic representation of EcoKI binding, translocation and cleavage of DNA. A monomer of EcoKI binds to a DNA site and then a second monomer binds to the first monomer forming a dimer. The dimer forms a loop by colliding with another region of DNA. Translocation occurs from both monomers and the process is stopped when the loop between the monomers becomes fully contracted. After translocation is stalled cleavage of the DNA occurs. Reprinted from Neaves et al., 2009.*

The Type I R-M systems are divided into five families based on sequence similarity, functional complementation and antibody cross-reactivity (Cajthamlova *et al*, 2007): Type IA, e.g EcoKI; Type IB, e.g. EcoAI; Type IC, e.g. EcoR124I; Type ID, e.g. StySBLI and Type IE, e.g. KpnBI.

Two published crystal structures of HsdS from *Methanococcus jannaschi* and *Mycoplasma genitalium* (Calisto *et al*., 2005; Kim *et al*., 2005) have become

available. The monomeric structure of the S subunit from *M. jannaschi* shows four continuous structural motifs: an N-terminal first TRD (TRD1) (Target Recognition Domain), a long-helical CCR (Central Conserved Region), the second TRD (TRD2), and the C-terminal long-helical DCR (Distal Conserved Region). TRD1 and TRD2 have very similar fold and each has extensive interaction with one of two CR helices: TRD1 exclusively interacts with the CCR helix and TRD2 with the DCR helix. CRs form two ten-turn helices running in an antiparallel way and interacting with each other by hydrophobic contacts. The CRs have been suggested to provide a structural motif for protein–protein interaction with the other subunits of type I R-M endonucleases (Kneale, 1994).

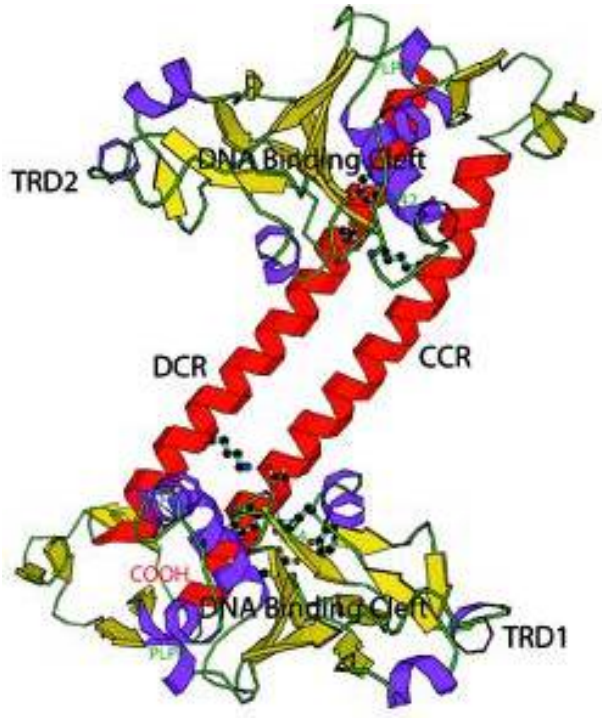


Fig. 4 *Crystal structure of the Type-I S subunit from* Methanococcus jannaschi. *A monomer consists of four successive structural domains: the globular TRD1, a* long

helical CCR domain, the globular TRD2, and a C-terminal DCR helix. (Image reprinted from Kim et al., 2005).

The crystal structure of HsdS from *Mycoplasma genitalium* shows a very similar fold to the HsdS structure from *Methanococcus jannaschi* even though the sequence identity of the two subunits is below 20% (Calisto *et al.*, 2005).

There are no published structures of HsdM, however, there are two unpublished structures deposited in the PDB including the crystal structure of HsdM of EcoKI from *Escherichia coli* (pdb accession number: 2ar0) (Rajashankar *et al*., 2005). An attempt to model the HsdM and HsdS subunits in the MTase has been made by Obarska *et al*., 2006. A structural model of the Type IC M.EcoR124I DNA methyltransferase (MTase), comprising the HsdS subunit, two HsdM subunits, the cofactor AdoMet and the substrate DNA molecule is shown in Fig. 5.

Fig. 5 *Model showing the predicted structure of the M.EcoR124I DNA methyltransferase and substrate. The HsdS subunit is shown in green. The two HsdM subunits are shown in yellow and orange and the substrate DNA in light blue. The adenines of the recognition sequence, which are to be methylated are shown 'flipped-out' coloured red, and the methyl donor is present above these bases (in blue). Reprinted form Obarska et al., 2006.*

The latest contribution to the structure of Type I MTase comes from a combination of structural modelling along with information derived from cryoelectron microscopy. In this model (Fig. 6), the N-terminal domains of HsdM form a bridge over the DNA duplex, thus forming a "clamp" encircling the DNA.

Fig. 6 *Model of EcoKI MTase in complex with substrate DNA and AdoMet. Yellow: HsdS; Grey: HsdM; Blue and cyan: DNA; Black: phosphate groups of DNA; Purple: AdoMet molecules; Orange: N-terminal domain of HsdM; Red: Catalytic domain of HsdM; Magenta: C-terminal domain of HsdM. Image reprinted from Kennaway et al., 2009.*

In this model the N-terminal domains of each HsdM subunit come in contact forming a bridge that must open in order to allow DNA to enter and be recognised by HsdS. Taking both models (Obarska et al., 2006; Kennaway et al., 2009) into consideration, it is probable that the structure of Type I MTases fluctuates between an open form (Fig. 5) and a closed form (Fig. 6). This opening and closing motion is facilitated by twisting and bending of the coiled-coil region of the HsdS subunit (Kennaway *et al*., 2009).

The characterisation of the HsdR subunit of EcoR124I by dynamic light scattering, analytical ultracentrifugation and small angle neutron scattering has indicated that the subunit is monomeric, globular and fairly compact (Obarska *et al*., 2008). Indeed, the recent crystal structure of the HsdR subunit of EcoR124I (Lapkouski, *et al*., 2009) revealed four globular domains in a square-planar arrangement. The four domains are: an endonuclease domain, two RecA-like helicase domains (1 and 2) and a helical domain with no apparent structural relatives (Fig. 7).

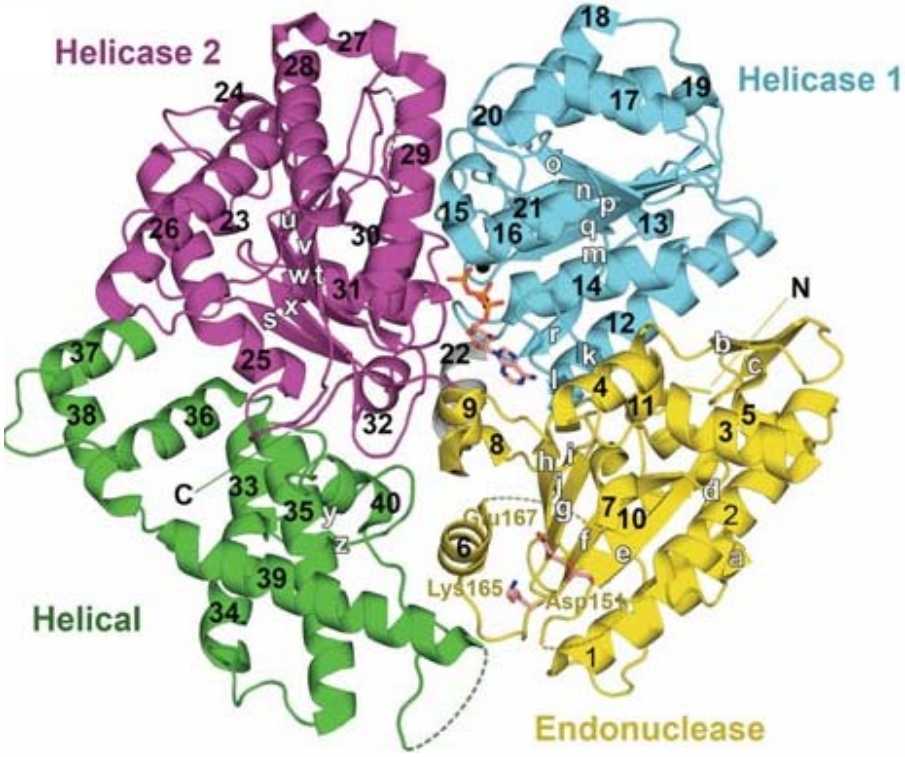


Fig. 7 *Crystal structure of the HsdR subunit of EcoR124I. Four domains can be seen: Endonuclease (residues 13-260), helicase 1 (261-461), helicase 2 (470-731) and helical domain (732-892). Image reprinted from Lapkouski et al., 2009.*

1.4.2 Type II restriction-modification systems

The Type II R-M systems are the simplest of the R-M systems with separate restriction and modification enzymes and no special cofactor requirements for cleavage other than Mg^{2+}. They methylate both strands of DNA at a specific base resulting in N6-methyladenine, 5-methylcytosine or N4-methylcytosine. In

contrast to Type I systems, Type II nucleases recognise and cleave DNA at fixed positions thus making them invaluable tools in molecular biology. Some Type II REases act as homodimers or tetramers thus binding two or more copies of their recognition site before DNA is cleaved (Halford, 2004). Most enzymes can recognise palindromic sequences of 4–8 bp in length, and cleave DNA within this sequence in both strands, producing 3′-hydroxyls and 5′-phosphate ends. Some Type II REases recognize discontinuous palindromes, interrupted by a segment of specified length but unspecified sequence. The DNA fragments produced have 'blunt' or 'sticky' ends with 3′- or 5′-overhangs of up to 5 nucleotides. Most of the restriction enzymes used for recombinant DNA work belong to the subtype that recognises palindromic sequences, which is called Type IIP (P for –palindromic) (Roberts *et al*, 2003).

1.4.3 Type III restriction-modification systems

Type III restriction enzymes consist of two subunits only, Mod (responsible for DNA recognition and modification) and Res (responsible for DNA cleavage). Active nucleases have a Mod_2Res_2 stoichiometry, require ATP and Mg^{2+} for activity and are believed to be stimulated by AdoMet, although recent findings suggest that AdoMet may not be necessary for DNA cleavage (Buchner-Moncke *et al*., 2009). The recognition sequence of Type III enzymes is asymmetric, uninterrupted and 5–6 bp in length. They translocate the DNA in an ATP-hydrolysis dependent manner and cut the DNA close to one recognition site. Typical examples are EcoP1I and EcoP15I (Bourniquel & Bickle, 2002). Studies utilizing fast-scan atomic force microscopy revealed that EcoP15I initially

employs diffusive looping to reduce the inter-site DNA distance, followed by translocation-coupled extruded looping (Crampton *et al*., 2007).

1.4.4 Type IV restriction systems

The Type IV R-M systems are methylation-dependent restriction systems. The first restriction enzyme that was shown to require a methylated substrate was DpnI from *Streptococcus pneumoniae*, which cleaves the sequence GATC only if the A is methylated (Lacks et al, 1986). Type IV R-M systems are REases with no associated methyltransferase and recognize and cleave DNA that has been modified (methylated, hydroxymethylated and glucosyl-hydroxymethylated). Thus foreign DNA is attacked because it does contain methylated bases in a particular sequence, rather than because it does not (Stewart *et al*, 2000). The best studied Type IV system is McrBC from *E. coli* K12. McrBC consists of two polypeptides: one (McrB) that is responsible for GTP binding/hydrolysis as well as DNA binding and another (McrC) that is responsible for DNA cleavage. The enzyme recognizes two methylated or hemimethylated sites at a distance of approximately 30 to more than 2000 base pairs and cleaves the DNA close to one of the two sites. This process is strictly coupled to GTP hydrolysis and involves the formation of high-molecular mass complexes (Pieper *et al*, 2002). A novel Type IV modification dependent restriction nuclease was recently found that targets glucosylated hydroxymethyl cytosine modified DNAs of the bacteriophage T4, T2 and T6 (Bair & Black, 2007).

1.5 Antirestriction Strategies

R-M systems are not a perfect barrier to invasion by foreign DNA. Bacteriophage and conjugative plasmids have developed several strategies to avoid restriction.

1.5.1 Viral-encoded DNA modifications

Some phage genomes encode their own methyltransferases that methylate and thus protect the viral DNA within the bacterial host. Such methylase genes have been found in the *Bacillus subtilis* phage SPβ and ϕ3T . The SPβ genome contains a gene that codes for a modifying DNA-methyltransferase which methylates the central C in the BsuR recognition sequence 5'GGCC (Trautner *et al.*, 1980). Similarly, the DNA of the *Bacillus subtilis* temperate phage ϕ3T is not susceptible to cleavage by the restriction endonuclease HaeIII due to the production of the viral MTase ϕ3TI (Cregg *et al.*, 1980).

Another example of viral DNA modification is the acetimidation of the N^6 position of 15% of the adenine residues in the DNA of bacteriophage Mu rendering the virus immune to the EcoKI, EcoBI, EcoP1 and EcoAI restriction systems (Hattman, 1979).

1.5.2 Viral stimulation of host modification functions

Some lambdoid phage are able to overcome the restriction problem either by stimulating the host MTases to modify phage DNA or by destroying the REase cofactors. The Ral protein of phage lambda can prevent restriction of unmodified phage DNA by the EcoKI and EcoBI enzymes by stimulating the host methyltransferases (Zabeau *et al.*, 1980). Interestingly, it has been shown that Ral enhances modification in a host strain lacking the entire *hsd*R gene, and lambda phage carrying the *hsd*M and *hsd*S genes modify their own DNA inefficiently in the absence of Ral. A model has been proposed in which, as a consequence of the interaction of Ral with either the HsdM or the HsdS polypeptide, the conformation of the enzyme is changed and the efficiency of methylation of unmodified target sites is enhanced (Loenen & Murray, 1986). An analogous function has been found in the Lar protein. The gene encoding Lar lies within Rac, a cryptic prophage sequence present in the genome of most *E. coli* K-12 strains (King & Murray, 1995). The ability to destroy REase cofactors has been found in phage T3 which destroys intracellular S-adenosylmethionine (SAM) necessary for methylation, by the production of a SAM hydrolase (Kruger *et al.*, 1975).

1.5.3 Incorporation of unusual bases in viral DNA

Many *B. subtilis* phage contain unusual bases in their DNA. Unusual bases in this case are defined as the bases which are not adenine, cytosine, guanine or thymine and which are synthesized at the level of nucleotide metabolism, not as

post-synthetic DNA modifications. The T-even phage of *E. coli* contain DNA in which cytosine is completely replaced by the unusual base hydroxymethylcytosine (HMC). HMC-containing DNA is completely resistant to HhaI and HpaII and almost completely resistant to BamIII, HindII and HindIII (Huang *et al.*, 1982). However, *E.coli* has developed mechanisms to deal with these unusual bases, by a) producing the Rlg protein (Rlg for restriction of non-glucosylated DNA) and b) producing a type IV restriction enzyme that is specific for non-glycosylated, HMC-containing DNA (Bair & Black, 2007). The T-even phage have in turn evolved a gene called *arn* (antirestriction of endonucleases) which produces an inhibitor of the Rlg enzymes (Kim *et al*, 1997). Other examples of phage containing unusual bases include phage SPO1, SP8, SP82G where thymine is completely replaced by 5-hydroxymethyluracil, and phage PBS1 and PBS2, where thymine is replaced by uracil (Warren, 1980).

1.5.4 Counterselection of restriction sites in phage genomes

By reducing or completely eliminating the recognition sites in their DNA, phage can avoid restriction. Pseudomonas phage ΦST1 appears to have no recognition sites for some of the restriction enzymes found in its host (Gachechiladze *et al.*, 1990). The same has been found for the streptomycete phage FP22 (Hahn *et al.*, 1990). Increasing the distance between EcoRII restriction sites (two copies of the target sequence are necessary for cleavage to take place) is the strategy adopted by the coliphage T7 and T3 (Kruger *et al*, 1988).

1.5.5 Coinjection of proteins occluding restriction sites

The restriction target sites of phage DNA can be protected by producing virus-encoded proteins which bind to the restriction sites thus creating an obstacle preventing the restriction enzymes from gaining access. Phage P1 is insensitive to restriction due to the production of the head proteins DarA and DarB (dar for defense against restriction). The *dar*A and *dar*B gene products are found in the phage head and protect any DNA packaged into a phage head, including transduced chromosomal markers, from restriction. The proteins must, therefore, be injected into recipient cells along with the DNA. These proteins are coinjected with the phage DNA in the bacteria and bind to the phage DNA recognition sites thus rendering them inaccessible to Type I restriction enzymes (Iida *et al.*, 1987).

1.5.6 Host-mediated restriction alleviation

The host cell's DNA is normally immune to restriction since the unmethylated strand produced by DNA replication is methylated before the next round of replication. However, when the host DNA is damaged by agents such as UV light, nalidixic acid and 2-aminopurine, a temporary loss of restriction proficiency occurs (restriction alleviation) (Thoms, & Wackernagel, 1984). This loss of restriction has been attributed to the ClpXP protease (Makovets *et al*., 1999).

In eukaryotes, proteins destined for proteolysis are ubiquitinated on a specific lysine residue which is recognised by the 26S proteasome (Voges *et al*., 1999). In

prokaryotes, proteins that are destined for degradation are tagged with an 11 amino acid sequence (AANDENYALAA). This specifically directs the tagged proteins to the ClpXP or ClpAP proteases. This 11 amino acid peptide is coded by SsrA, a unique tmRNA which functions as both a messenger and a transfer RNA. The degradation of tagged proteins by ClpXP is markedly enhanced by an adaptor protein SspB which binds specifically to the 11 amino acid tag. SspB is believed to have a role in helping the ClpX recognise the tagged protein. The tag is attached to the C-terminus of nascent polypeptides on stalled ribosomes. When protein synthesis on the ribosome stalls (because it cannot bind the release factors necessary for termination) the SsrA tags the incomplete nascent protein. The alanine-charged SsrA binds to the stalled ribosome mimicking a tRNA and the nascent chain is transferred to the SsrA by transpeptidation. Then, the SsrA functions as a mRNA translating the reading frame coding for the 11 amino acid tag (Keiler *et al*., 1996)

ClpXP is a two component system consisting of ClpP and ClpX (Fig. 8a and 8b). The activator is ClpX and the proteolytic core is ClpP. ClpX is composed of a hexamer which is the ATPase domain. ClpP contains two rings of seven subunits , the rings placed on top of each other. The binding of ClpX to ClpP is mediated by a hydrophobic loop referred to as the IGF loop (Kim *et al*., 2001). ClpP digests the protein translocated from ClpX into uniform peptide fragments of six to eight residues (Choi & Licht, 2005).

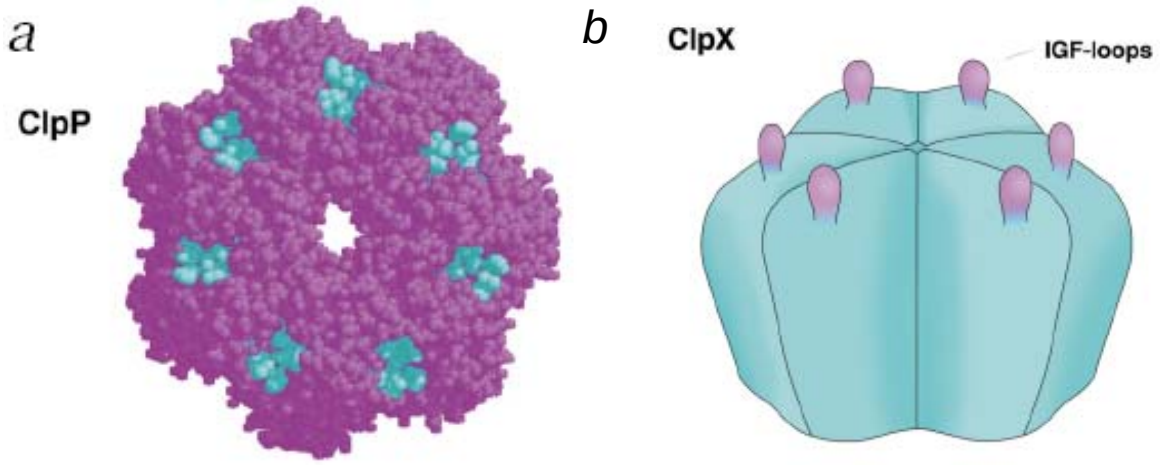


Fig. 8 ***a.*** *The structure of ClpP is viewed face-on with the hydrophobic pocket coloured in cyan, which has been suggested to interact with ClpX and ClpA.* ***b.*** *Cartoon depicting how the ClpX hexamer may dock with the ClpP 14-mer via the exposed IGF loop. Reprinted from Kim et al., 2001.*

ClpXP has been shown to degrade the R subunit from EcoKI so that the latter cannot cleave the host's chromosomal DNA. The host's chromosome is recognized as "self" rather than "foreign" even in the absence of modification. HsdR is recognised by ClpXP only after the EcoKI complex has embarked on its restriction pathway but before any damaged is inflicted on the unmodified chromosomal DNA. Furthermore, HsdR is lost only in cells in which HsdR could produce functional EcoKI (Makovets *et al.*, 1999). It has also been demonstrated that ClpX and ClpP are necessary for the efficient transmission of the genes encoding EcoKI and EcoAI. When the *hsd* genes are transferred to a

recipient cell that lacks a protective modification system, its DNA becomes susceptible to degradation once the endonuclease is formed. It is believed that the ClpXP protease competes with MTase for interaction with HsdR delaying the production of the whole EcoKI complex (Makovets *et al*., 1998). Even though restriction alleviation via ClpXP has been demonstrated for Type IA and Type IB systems, restriction alleviation of Type IC and ID systems does not involve the ClpXP protease (Makovets *et al*., 2004). It is still not clear how unmodified sequences in the resident bacterial chromosome are distinguished from "foreign" DNA that has recently entered the bacterial cell. Perhaps host DNA differs from invading DNA in its location, its association with other proteins and its topology (Murray, 2002).

1.5.7 Production of proteins that inhibit restriction enzymes

The most thoroughly investigated example of active antirestriction is that exerted by the closely related phage T3 and T7. The phage gene responsible is gene 0.3 from its map position that codes for the Ocr protein (overcome classical restriction) (Studier, 1975). T3 and T7 are closely related, and their DNA molecules hybridize extensively; however, there are several functionally important nonhomologous regions. One of the differences between the two phage is the coding of an S-adenosylmethionine hydrolase (SAMase) by gene 0.3 of T3, but not by T7. Even though the T7 gene product only shows the *ocr* phenotype it is nevertheless as resistant to restriction as T3 (Kruger & Schroeder, 1981). Ocr is the first product to be expressed by phage T7 as it enters a bacterium. The phage first injects the early region of its genome which contains the *ocr* gene. Transcription of this region of the genome is a prerequisite for the

injection of the remaining DNA. The delayed injection of the phage DNA may be a mechanism by which the phage prevent exposure of the DNA recognition sites before the appearance of Ocr (Kruger & Schroeder, 1981).

Ocr is a 26 kDa homodimer that protects DNA from restriction by competing with DNA for the DNA-binding site of the Type I R-M enzyme (Bandyopadhyay *et al.*, 1985). The mechanism by which Ocr binds to these enzymes is not target sequence specific. Instead, Ocr mimics the shape and charge distribution of B-DNA. As shown by its crystal structure (Fig. 9), Ocr is highly acidic and its negatively charged groups are superimposable upon the phosphate backbone of DNA. The length of the Ocr dimer corresponds to 24 bp or two to three turns of B-DNA. The bend angle between the fitted DNA duplexes is 33.6 degrees, similar to that induced by Type I R-M systems in their DNA targets (Walkinshaw *et al.*, 2002). A 1:1 strong interaction between Ocr and M. EcoKI suggests that the primary mode of binding is one Ocr dimer per M. EcoKI (Atanasiu *et al.*, 2002). It has been shown through mutational studies that the hydrophobic contact at the interface of two Ocr monomers is critical for the ability of Ocr to inhibit the modification activity of EcoKI and to a lesser extent the antirestriction activity (Zavilgelsky *et al.*, 2009).

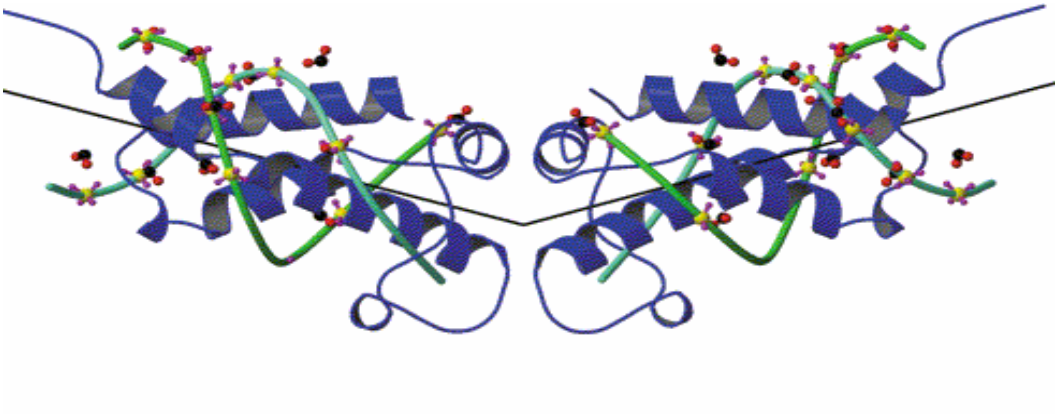


Fig. 9 *Superposition of two 12 bp B-DNA molecules on the Ocr dimer (green strands: DNA; Blue: Ocr; Yellow: phosphorous atoms; Purple: oxygen atoms of the phosphate groups; Red: oxygen atoms of the carboxyl groups; Black: carbon atoms of the carboxyl groups). Reprinted from Walkinshaw et al., 2002.*

Another virus-encoded inhibitor of restriction enzymes is produced by *Bacillus* phage ϕNR2rH and ϕ1rH. These two phage provide the first examples of an active viral block of a type II restriction enzyme. The protein has a molecular weight of 20 kDa and is an inhibitor of the restriction enzyme BamNx (Makino *et al.*, 1980).

Stp is a 26 residue polypeptide encoded by bacteriophage T4 (Depew & Cozzarelli, 1974). Stp activates the latent ACNase (anticodon nuclease) and alleviates the Type IC R-M systems DNA restriction activity. Stp affects the dissociation of the R_2 complex of the R-M system, shifting the equilibrium towards R_1 complex formation which is the restriction-deficient form of the REase (Lisle *et al.*, 2000).

An example of inhibition of restriction of an exonuclease and not an endonuclease is the activity of the Gam protein produced by bacteriophage lambda. The RecBCD enzyme of *E. coli* has several known activities including ATP-dependent dsDNA and exonuclease activity. RecBCD contributes to the degradation of foreign DNA entering the cell. The nuclease activities of RecBCD are involved directly, and its helicase activity is involved indirectly by creating ssDNA which is acted upon by other nucleases (Kowalczykowski *et al.*, 1994). The recently solved crystal structure of Gam suggests that the protein inhibits RecBCD by preventing it from binding DNA (Court *et al*, 2007).

1.5.7.1 The Ard proteins

To overcome the host restriction barrier many self-transmissible plasmids and mobile elements in bacterial chromosomes encode Ard proteins (alleviation of restriction of DNA). Ard proteins fall into three distinct groups, ArdA, ArdB and ArdC based on amino acid sequence homology. All groups have a small region similarity that includes 14 residues designated as the "antirestriction motif" (Belogurov & Delver, 1995). Mutational substitution of negatively charged amino acid residues present in the motif of ArdA from IncN plasmid pKM101, however, did not affect the antirestriction activity, but completely abolished the antimodification activity (Rastorguev & Zavilgelsky, 2004). ArdA proteins seem to be prevalent among Ard proteins with 75 ArdA sequences known today. Sequenced *ardA* genes are predicted to encode small polypeptides of 166-169 amino acids, which are highly acidic with net negative charges ranging from -22 to -29 (Wilkins, 2002). Antirestriction proteins belonging to the

ArdA family are strongly acidic, like Ocr. However, there is no significant sequence similarity between the ArdA protein and Ocr.

ardA genes are expressed rapidly and transiently in the infected cell aiding in the establishment of the immigrant plasmid. Other genes such as *psiB* and *ssb* are also expressed early following plasmid conjugative transfer. The increased production of PsiB is thought to prevent induction of the bacterial SOS response, while the overproduction of SSB may confer the benefit of avoiding SSB starvation after plasmid entry (Bagdasarian *et al*., 1992; Jones *et al*., 1992). Interestingly, *ardA, psiB* and *ssb* genes have the same transcriptional orientation such that the transferred strand is the same as the transcribed strand (Chilley and Wilkins, 1995). This raises the possibility that these genes are transcribed from the entrant strand before the synthesis of the complementary DNA strand. This is achieved by a special type of promoter, *ssi3*, which acts only on single stranded DNA and provides the synthesis of a sufficient amount of ArdA protein during conjugation (Bates *et al*., 1999; Nasim *et al*., 2004). This would allow accumulation of the antirestriction protein before the formation of the double-stranded substrate which would be degraded by endonucleases (Althorpe *et al*., 1999).

ArdAs are produced in the new host in time to prevent destruction of the incoming DNA by restriction endonucleases. Two factors contribute to this: first, *ardA* genes reside in plasmid leading regions (Chilley & Wilkins, 1995). Second, early in conjugation there is a burst of *ardA* transcription. The transcription is initiated within a minute of entry of the leading region (Althorpe *et al*., 1999).

The first plasmid antirestriction system to be discovered was on the enterobacterial plasmid pKM101, a member of the IncN group (Belogurov *et al*., 1985). pKM101 encodes two non-homologous antirestriction proteins, ArdA and ArdB. ArdA from pKM101 inhibits R/M by the Type I R-M enzyme, EcoKI, but the ArdB is ineffective against the modification activity (Belogurov *et al*., 1993). Similarly, ArdA from plasmid R16 preferentially targets restriction activity of the Type I system EcoKI leaving significant levels of methyltransferase activity under conditions in which restriction is almost completely prevented (Thomas *et al*., 2003).

The third type of anti-restriction protein (ArdC) is found in the IncW plasmid pSa, which has a high degree of similarity to the N-terminal region of RP4 TraC1 primase. ArdC can bind to single-stranded DNA. Furthermore, ArdC protects the single-stranded but not double-stranded DNA against the activity of Type II endonuclease HhaI (Belogurov *et al*., 2000).

1.5.7.2 The KlcA proteins : ArdB homologues

In the present work, the product of the *klcA* gene from the incompatibility plasmid pBP136 from *Bordetella pertussis* was characterised and its structure determined. Sequence alignment revealed that KlcA from pBP136 is significantly related (30% identity, 46% similarity) to the 16.5 kDa ArdB from the IncN plasmid pKM101 (Belogurov *et al*., 1993).

Bordetella pertussis is the causative agent of whooping cough, a disease predominantly affecting young children and is spread in aerosols created by infected individuals. The organism adheres to the cilia of the respiratory epithelia and secretes several toxins that kill epithelial cells, causing dysfunction of the central nervous system and inducing hypoglycaemia. The prolonged paroxysms of coughing can lead to apnoea and eventually death (Preston, 2005). Despite the development of several vaccines, immunity is short-lived resulting in reinfection. Thus eradication of *B. pertussis* still remains a challenge with 300,000 deaths occurring each year worldwide (He & Mertsola, 2008).

In 2002, a *B. pertussis* isolate that harboured an IncP-1β plasmid,was identified from a fatal case of pertussis which is believed to be the first case of *B. pertussis* with a naturally occurring plasmid (Kamachi *et al*., 2006). The 41 kb plasmid was named pBP136, contains 46 open reading frames (ORFs) and possesses a classical IncP-1β backbone. Plasmids of the incompatibility group P are found in a wide range of Gram-negative bacteria (Adamczyc & Jagura-Burdzy, 2003). Incompatibility is often a necessary consequence of the cellular mechanism to maintain a defined copy number of a particular plasmid. IncP-1β members carry multiple antibiotic resistance determinants and genes for the degradation of xenobiotic compounds (Burlage *et al*., 1990). An exception is the IncP-1β plasmid pA1 from *Sphingomonas* sp. A1 that does not contain any of the typical accessory genes of IncP-1β plasmids (Harada *et al*., 2006). Similarly, pBP136 was found to contain no accessory genes for antibiotic resistance, remnants of insertion sequences or transposons. The plasmid contains two regions involved in

plasmid conjugation and a third region involved in plasmid replication, central control, stable inheritance and partitioning.

This third region is composed of *incC* (plasmid partitioning), *kfrA* and *kor* (both responsible for transcriptional regulation) and the *kilC* operon. Four *kil* loci have been identified so far in IncP plasmids: *kilA*, *kilB*, *kilC* and *kilE*. These genes have been designated as *kil* because it has been shown that their unregulated expression is lethal to *E. coli* cells, and genes that can override this lethality have been designated " *kor"* (kill override) (Figurski *et al*., 1982). *kilA* consists of three genes *klaA*, *klaB* and *klaC*. Genetic analysis of the locus has demonstrated that each of the three genes expresses a host-lethal phenotype when expressed from the unregulated *kilA* promoter (Goucharoff *et al*., 1991). The *kilB* locus consists of three genes: *klbA*, *klbB*, *klbC*. Deletion analysis established that the host-lethal phenotype requires *klbB*, and that *klbA* is required for conjugal transfer. The repressor *korB* seems to regulate expression of the *kilB* locus (Thomson *et al*., 1993). The *kilE* locus (in RK2) contains two multicistronic operons: the *kleA* operon containing the *kleA* and *kleB* genes, and the *kleC* operon which contains four genes (*kleC*, *kleD*, *kleE* and *kleF*). Each operon is lethal to *E. coli* host cells in the absence of *kor* functions (*korA* and *korC*) (Kornacki *et al*., 1993). The *kilC* operon consists of three genes: *korC*, *klcB* and *klcA*. Genetic analysis has revealed that the host-lethal phenotype of the *kilC* operon of plasmid RK2 is *klcA*. Expression of *klcA* is controlled by *korA* and *korC* (Larsen & Figurski, 1994). However, thus far, no specific function had been attributed to *klcA* genes.

The aim of this project is to characterise with biophysical and structural methods the antirestriction proteins ArdA from *Enterococcus faecalis* and KlcA from *Bordetella pertussis* and understand their mechanism of action against Type I R-M systems.

Chapter 2: Background theory of protein biophysical and structural methods

2.1 Steady-state fluorescence

Following electronic excitation by absorption of a photon, a molecule loses energy as heat by cascading through the closely spaced vibrational levels of all the excited energy levels. The molecule "hesitates" at the lowest vibrational level of the first singlet because a large amount of energy must be lost to go to the ground state. During this period three things may happen:

- the molecule may fluoresce and drop to the ground state by emitting a photon
- it may convert to a triplet state via intersystem crossing
- may go to the ground state without emitting a photon via a non-radiative transition

In the case of the triplet, the molecule may either phosphoresce and drop to the ground state emitting a photon or undergoes a non-radiative transition. The lifetimes for all these processes are illustrated in Fig. 10.

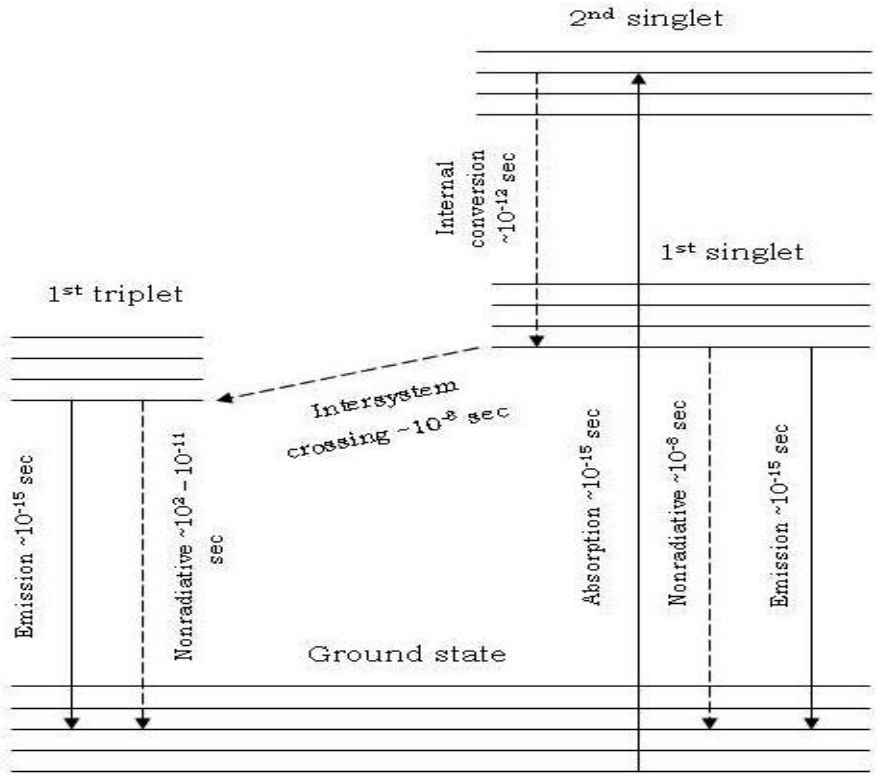


Fig. 10 *One form of a Jablonski diagram (adapted from Van Holde et al., 2006) showing the lifetimes of absorption, fluorescence and phosphorescence at the equilibrium internuclear distance of the ground state.*

Fluorescent light is emitted from samples in an excited electronic state. Fluorescence instruments use light to excite the sample and then observe the light emitted. A monochromator before the sample chooses the wavelength of exciting light, and a monochromator after the sample scans the wavelengths of emitted light. The shape of the fluorescence spectrum as scanned by the emission monochromator will be independent of the wavelength of absorption because excited molecules decay to the ground vibrational level of the first singlet state before fluorescing.

Intensity measured as a function of wavelength comprises the fluorescence spectrum. A plot of emission against wavelength for any given excitation wavelength gives the emission spectrum. When the wavelength of the exciting light is changed and the emission from the sample is plotted against the wavelength of exciting light, the result is the excitation spectrum.

The time-dependence of fluorescence intensity following excitation of the fluorophore is related to the excited state lifetime, τ. This parameter characterises the average time a molecule spends in the excited state before emitting a photon:

$$I(t) \sim e^{-t/\tau}$$

The fluorescence of protein molecules arises from the presence of the aromatic amino acids Trp, Phe and Tyr. The UV absorption maxima of these amino acids in water are 280 nm for tryptophan, 258 nm for phenylalanine and 275 nm for tyrosine. The emission maxima are around 350 nm for tryptophan, 282 nm for phenylalanine and 304 nm for tyrosine (Lakowicz, 2006). The emission of tryptophan is highly sensitive to its local environment and thus is used as a reporter group for protein conformational changes. Generally the emission spectrum of Trp is shifted to shorter wavelengths and becomes more structured with decreased polarity of the surrounding environment. In order to excite Trp fluorescence a wavelength of 295 nm or more is chosen in order to minimise any effects from the Tyr absorbance and eliminate the absorption of phenylalanine. Shifts of the Trp emission spectrum are often used to detect protein structural changes, however, the fluorophore is affected only by its immediate

environment and no information can be obtained regarding global changes in protein structure.

When the intrinsic fluorescence of the molecule is not enough, fluorescence can be obtained by labelling the molecule with fluorescence probes which will preferably have longer excitation and emission wavelengths than the aromatic amino acids. The protein-labeling reagents can be convalently or non-covalently attached to the protein. Covalent fluorophores include dansyl chloride (DNS-CL), fluoresceins and rhodamines and non-covalent probes include 1-anilinonapthalene-6-sulfonic acid (ANS) (Lakowicz, 2006).

In a homogeneous solution the ground state fluorophores are all randomly oriented. When exposed to polarised light (i.e. light which has been filtered to vibrate only in one plane either vertical or horizontal), those fluorophores that have their absorption transition moments oriented along the electric vector of the incident light are excited. The extent of polarisation of the emission is described in terms of the anisotropy (r).

In a fluorescence anisotropy experiment the sample is excited with vertically polarised light. In this way, the electric vector of the excitation light is oriented parallel to the vertical axis. The intensity of the emission is measured through a polariser. When the emission polariser is oriented parallel to the direction of excitation light the observed intensity is called $I_{\parallel}$ and when oriented perpendicular is called $I_{\perp}$ (Lakowicz, 2006). A diagram showing the experimental setup for anisotropy measurements is shown in Fig 11.

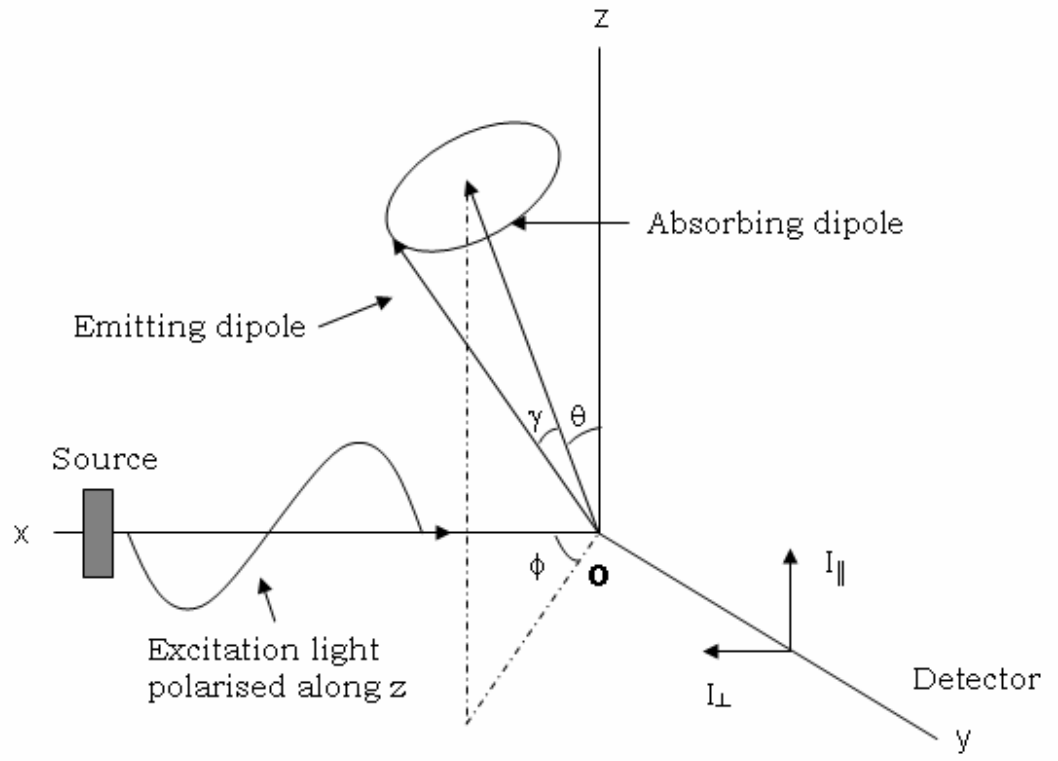


Fig. 11 *Experimental method for measuring the anisotropy of fluorescent emission. Polarised light is sent along the x-axis with the electric vector along the z-axis. The molecule of interest is place at point 0 in the Cartesian coordinate system and has a transition dipole for absorption at an angle θ and an emitting dipole at an angle γ. The polarised light emitted is observed along the y-axis and a polariser is used to separate the intensity of the electric vectors which are parallel or perpendicular to the electric vector of the excited light. In a T-format method the parallel and perpendicular components are observed simultaneously through separate channels. Image adapted from Van Holde et al., 2006.*

The values of $I_{\parallel}$ and $I_{\perp}$ are used to calculate the anisotropy (r):

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

Where the G factor is the ratio of the sensitivities of the detection system of vertically and horizontally polarised light. Measuring the G-factor is important since when the ratio is different than one, the sensitivity of one detection is different from the other and thus the intensity observed is not the true intensity. The G-factor is measured using horizontally polarised excitation.

Factors that affect rotational diffusion such as molecular weight, shape and viscosity will also affect anisotropy. Anisotropy measurements can be useful for measuring any process which increases or decreases the extent of rotational diffusion of the fluorophore. The extent to which a fluorophore rotates during the excited state determines its anisotropy. Thus, the anisotropy measured through steady-state fluorescence can be used to follow the molecular weight changes that result from binding. When a protein binds to a ligand the rotational correlation time decreases and the anisotropy increases. The rotational rate is related to the rotational correlation time (θ) according to the equation:

$$\theta = \frac{\eta V}{R T}$$

Where η is the viscosity of the solution, V the molecular volume, R the gas constant and T the temperature in Kelvin. The relationship between the anisotropy (r) and the rotational correlation time θ is given by the Perrin equation:

$$r = \frac{r_0}{1 + (\tau / \theta)}$$

where r_0 is the anisotropy that would be measured in the absence of rotational diffusion and τ the excited-state lifetime (Lakowicz, 2006).

2.2 Differential Scanning Calorimetry (DSC)

DSC is one of the primary methods for the study of thermal protein denaturation. It is the measurement of the change of the difference in the heat flow rate to the sample and to a reference sample while they are subjected to a controlled temperature program (Fig.12). DSC allows reaction heats and heats of transition to be quickly measured on small sample masses.

In a basic DSC experiment, energy is introduced simultaneously into a sample cell and a reference cell containing only the solvent, raising the temperatures of both identically over time. The difference in the input energy required to match the temperature of the sample to that of the reference at each step of the scan is the amount of excess heat either absorbed (endothermic reaction) or released (exothermic reaction) by the molecule in the sample.

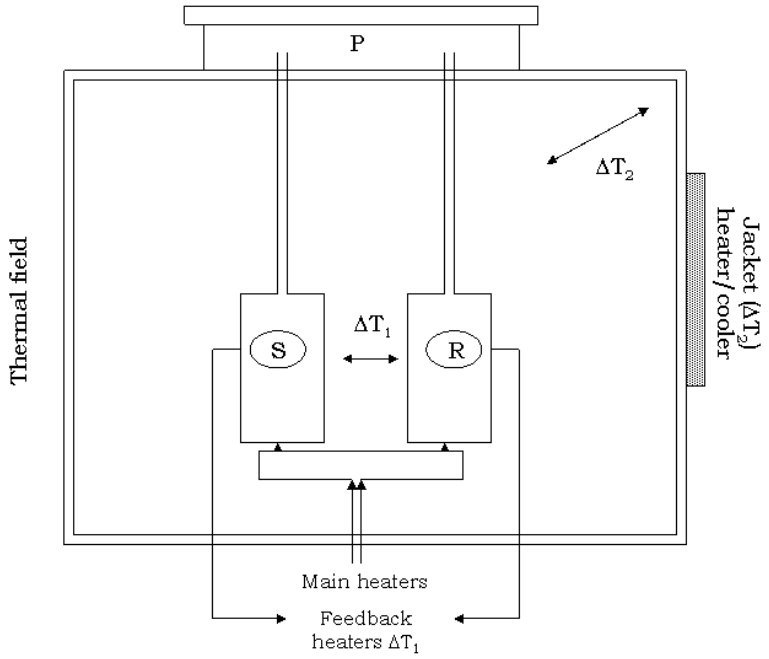


Fig. 12 *Diagram of a typical DSC used for thermal studies of biomolecules. Reproduced from Cooper et al., 2000. The sample (S) and reference (R) cells are held under inert gas pressure to inhibit bubble formation while heating. The main heaters raise the temperature of the cells at a constant rate while monitoring the difference in temperature* ΔT_1 *of the S and R cells and the surrounding adiabatic jacket* ΔT_2. *Feedback heaters compensate for any temperature differences.*

If we consider the sample to be a protein, more energy is required to bring the sample to the same temperature as the reference because of the excess heat absorbed by the protein. The excess heat (dq) absorbed for the incremental

change in temperature (dT) will be a measure of the heat capacity (C_p) of the protein:

$$C_p = dq / dT$$

The heat capacity (Cp) changes from that of the native form (C_N) to that of the denatured form (C_D). After concentration normalisation and baseline correction, DSC data can be fit to a two-state model to yield the enthalpy of denaturation and the melting temperature. The temperature at which half the protein is denatured is the melting temperature (T_m) of the protein. The integrated area under the curve is the calorimetric enthalpy (ΔH_{cal}) of the denaturation which represents the total heat energy uptake by the sample undergoing the transition. Calorimetric enthalpy depends on the total amount of protein in the calorimetric cell.

$$\Delta H_{cal} = \int \Delta C_p \, dT$$

The van't Hoff enthalpy (ΔH_{VH}) is a measure of the cooperativity of transition; the greater the cooperativity, the sharper the transition. ΔH_{VH} depends on the shape of the endotherm and is independent of the amount of protein present.

$$\Delta H_{VH} = \frac{4R\, T^2_m\, C_{p(max)}}{\Delta H_{cal}}$$

Normally ΔH_{cal} should be equal to ΔH_{VH}, however, in cases where the unfolding is not two-state or the protein concentration is under-estimated, $\Delta H_{cal} > \Delta H_{VH}$. On the other hand, if $\Delta H_{cal} < \Delta H_{VH,}$ then either the protein concentration is over-estimated, or aggregation causes distortion of the peak, or not all the protein is initially folded. A typical DSC plot of protein denaturation is shown in Fig. 13.

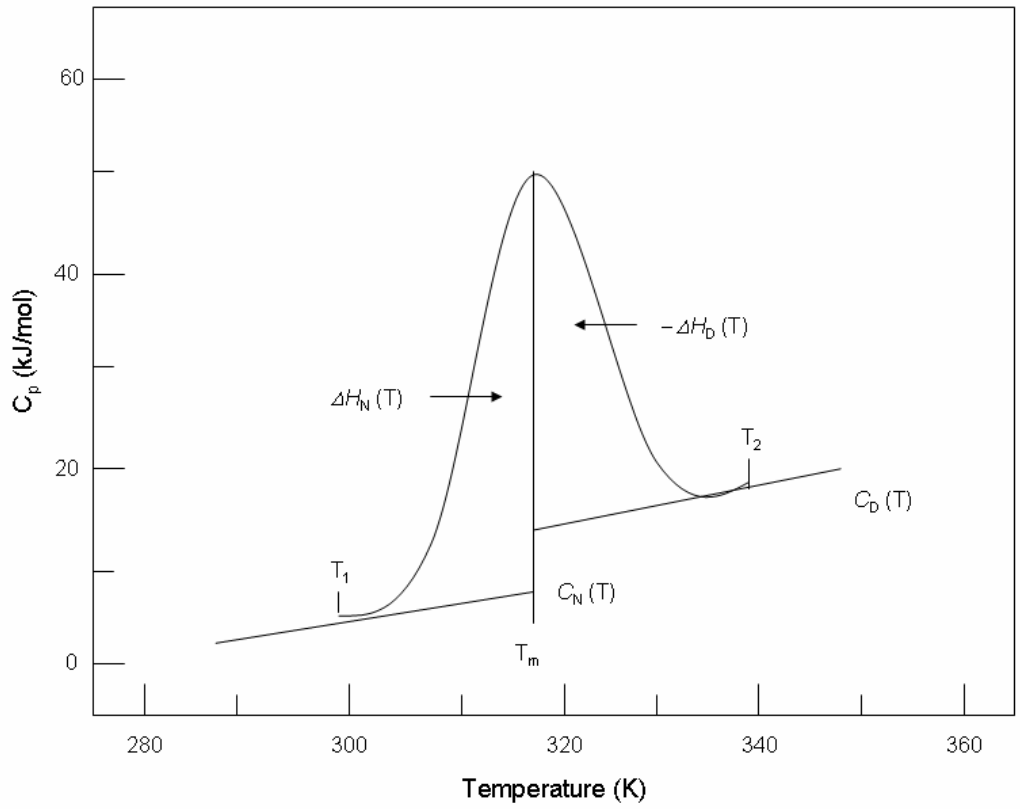


Fig. 13 *DSC plot of protein denaturation. The heat capacity C_p changes from that of the native form $C_N(T)$ to that of denatured $C_D(T)$. The integrated area under the curve from T1 to T2 is the enthalpy of denaturation. T_m denotes the midpoint temperature. The associated enthalpies for the folded state $\Delta H_N(T)$ and unfolded state $\Delta H_D(T)$ are equal and of opposite signs. Image adapted from van Holde et al., 2006.*

When the structure of a protein is "melted" in this way there are no changes to the covalent nature of the molecule. It is only non-covalent interactions that are perturbed and in many cases, if the protein is cooled down again these interactions will reform spontaneously, yielding the active native conformation. To check the reversibility of denaturation, the heated sample is first cooled and then it is heated again with the same scanning rate. If the original curve is completely reproduced, the denaturation is reversible. If no heat absorption is observed at reheating, denaturation is considered to be irreversible.

DSC can also be used to study the thermochemistry of interacting molecules such as protein-protein or protein-ligand interactions (Weber & Salemme, 2003). Some problems for using DSC for such studies are that the fractions of bound and free proteins may be temperature dependent and each component can undergo thermal denaturation during the experiment. For these reasons, the more typical method for studying the thermochemistry of interacting molecules is isothermal titration calorimetry (ITC).

2.3 Isothermal Titration Calorimetry (ITC)

Isothermal titration calorimetry measures the amount of heat absorbed or released by a reaction that results from mixing together two or more components.

ITC involves the incremental addition over time of a ligand into a solution with the molecule of interest (e.g. protein). At each titration point, formation of the

complex results in the release or absorption of heat (q) which is the heat of complex formation. Initially, the amount of free protein is sufficient to bind ligand according to the concentrations of free ligand and protein in solution. Later, however, the amount of free protein becomes limited, until at the end of the titration there is no protein available to form a complex and thus no excess heat is absorbed or released upon addition of the ligand.

The ITC instrument is very similar to the DSC apparatus (see Fig. 12) with the exception that the sample and reference cells are kept at constant temperature (isothermal). A syringe injects small quantities of ligand into the sample cell and a stirrer mixes the solution constantly.

Fig. 14 shows an example of an ITC diagram. The top graph shows the heat change over time. After integration of the area under each peak and substitution of the dilution heat effects and normalisation per mole injected ligand, a titration curve is obtained (bottom graph). The data from the titration curve can be fit with the help of specialised software such as Microcal Origin to a non-linear least squares Marquardt algorithm as described elsewhere (Ladbury & Chowdhry, 1996) to yield three thermodynamic parameters: the stoichiometry n, the enthalpy change (ΔH^0), and the association constant (K_α).

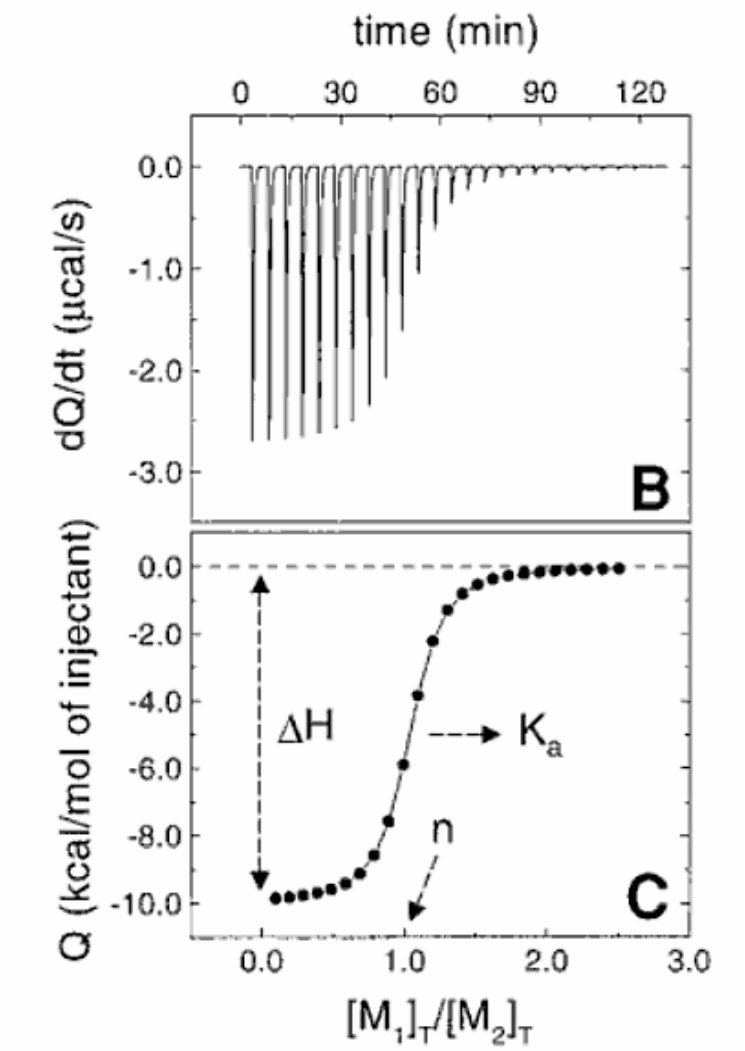


Fig. 14 *ITC plot of protein binding to a ligand. The heat change over time is shown in graph B. After integration of the area under each peak from the top graph, the individual heats are plotted against the molar ratio (C). After non-linear regression analysis, the thermodynamic parameters n, K_α and ΔH can be determined. Reprinted from Haian, 2004.*

One limitation of ITC is that it cannot yield meaningful quantitative results when the binding between two molecules is very strong and the Ka is very large- the practical upper limit is estimated to be $K_a = 10^9\ M^{-1}$ (van Holde *et al*., 2006). The resulting titration will show a sharp transition from complete binding of added ligand to no binding.

2.4 Circular Dichroism (CD)

Dichroism is the phenomenon in which light absorption differs for different directions of polarisation. Linear dichroism involves linearly polarised light where the electric vector is confined to a plane. The direction of the electric vector is constant and its magnitude is modulated. Linear dichroism is a function of wavelength λ and is defined as the difference between absorption of parallel and perpendicular polarised light.

$$LD(\lambda) = A_{\parallel}(\lambda) - A_{\perp}(\lambda)$$

In circularly polarised light, the magnitude is constant and the direction is modulated. The plane polarised radiation is split into in two circularly polarised components. If after passage through the sample, the left and right circularised polarised components are not absorbed or are absorbed to the same extent, then radiation polarised in the original plane will be generated again. However, if one of the two components is absorbed to a different extent, the polarisation of the resultant radiation would form an ellipse. The CD instrument will then display the dichroism at a given wavelength of radiation as either the difference in absorbance of the two components:

$$\Delta A = A_L - A_R$$

or as the ellipticity in degrees (θ, mdeg) which is related to the differential absorption ΔA as follows:

$$\Delta A = \theta/33000$$

A CD signal will only be observed when a chromophore is chiral (optically active). A chromophore can be chiral due to: i) the nature of its structure, ii) being covalently linked to a chiral centre (e.g. a haem group) or iii) being placed in an asymmetric environment (Kelly & Price, 1997).

The CD of a native protein depends on the fractions of its component secondary structures. The study of protein secondary structure by their far-UV CD spectra is based on the suggestion that these spectra are determined mainly by the local conformation of the protein backbone, which can be roughly described by secondary structure. In other words, far-UV CD spectra of proteins are considered as a superposition of spectra of different classes of secondary structure such as α-helix and β-sheet. The presence of secondary structure in a protein will result in a spectrum with positive and negative signals, whereas the absence of regular chiral structure will result in zero CD intensity.

The far-UV absorption region (<250 nm) is dominated by electronic transitions of the peptide backbone of the protein, but transitions from side chains also contribute in this region. The peptide chromophore which gives rise to the transitions observed in the far-UV region has non bonding electrons on the oxygen and also on the nitrogen atoms, π-electrons which are delocalised to

some extent over the carbon, oxygen, and nitrogen atoms and sigma bonding electrons. The lowest energy transition of the peptide chromophore is an n to π* transition and the next transition is π to π*. Even though a number of amino acid side chains also have transitions in the peptide region (and these transitions are often stronger than the peptide π to π* transitions) since the peptide chromophores are in excess, the side chain transitions are generally nearly impossible to detect.

Typical of the α helix is a negative CD band at about 222 nm and a negative and positive couplet at about 208 and 190 nm. The magnitude of the negative 222 nm band is a good measure of α–helix content in a protein. β-sheets exhibit a broad negative band near 218 nm and a large positive band near 195 nm. Disordered chains have a weak broad positive CD band near 217 nm and a large negative band near 200 nm.

The near-UV CD signal (250-310 nm) arises from the environments of aromatic amino acids side-chains as well as disulfide bonds or non-protein cofactors. In the near-UV the aromatic amino acid side-chains (phenylalanine, tyrosine and tryptophan) absorb in the range of 250-300 nm. As an approximation, the aromatic side chains and disulfides absorb at the following frequencies:

-250-270 nm: phenylalanine signal
-270-290 nm: tyrosine signal
-280-300 nm: tryptophan signal
-260 nm: disulfide bond signal

There are several applications of CD to proteins including: determination of secondary structure, analysis of thermodynamics and kinetics of protein folding and characterisation of protein-ligand interactions (Greenfield, 1999).

2.5 Matrix Assisted Laser Desorption Ionisation- Time Of Flight/ Mass Spectrometry (MALDI-TOF/MS)

Matrix-assisted laser desorption/ionisation (MALDI) time-of-flight (TOF) mass spectrometry (MS) produces gas-phase protonated ions by excitation of the sample molecules using the energy of a laser pulse absorbed by an ultraviolet light absorbing matrix.

Protein samples after trypsin digestion, are mixed with an appropriate matrix solvent and spotted onto a stainless steel plate. The solvent is dried and the plate is introduced into the vacuum system of the MALDI-TOF analyser. A laser pulse irradiates the sample which causes ionisation of both the matrix and the sample. Without the matrix, the analytes could only be desorbed at higher energies, but this would result in their fragmentation. Because the matrix is of much lower MW than the ionised samples, the matrix does not interfere with the assay of the protein. Ions are then focused into the mass spectrometer. Larger ions require more time to reach the detector than smaller ions.

Matrix substances are small organic molecules (mainly acids) that have an absorption maximum at the laser wavelength and can co-crystallise with the analyte. The most commonly used method for co-crystallisation is the “dried-

droplet" method where the sample is premixed with the matrix and the mixture is allowed to evaporate slowly at room temperature (Fuchs & Podda, 2004).

In order for desorption to occur, the matrix sample co-crystals are irradiated with a nanosecond laser beam (from a UV laser with a wavelength of 266 or 337 nm). The energy introduced by the laser irradiation causes a structural decomposition of the crystal (in other words, the laser energy is converted to vibrational oscillation of the crystal molecules) and generates a particle cloud. This cloud contains ions that follow a Boltzmann distribution of their initial velocities. Due to the high matrix concentration, the analyte ions are prevented from interacting with each other. Following acceleration, the extracted ions drift through a field-free path and finally reach a detector. The mass-to-charge ratios, (m/z) are calculated by measuring their time-of-flight (TOF) which is longer for larger molecules.

The MALDI-TOF MS utilises the following principle to determine the mass of ions generated in the MALDI process. This is based on the time (t) taken for an ion of mass (m) and known kinetic energy ($eV = ½ mv^2$) to travel a distance (l) in a field-free region. This time t is proportional to the square root of the mass of the ion: $t = l\sqrt{m/2eV}$ (Hoffmann & Stroobant, 2007)

There are two types of TOF analyser, the linear mode and the reflectron mode. The reflectron mode (Fig. 15) was used for the experiments performed for the present work.

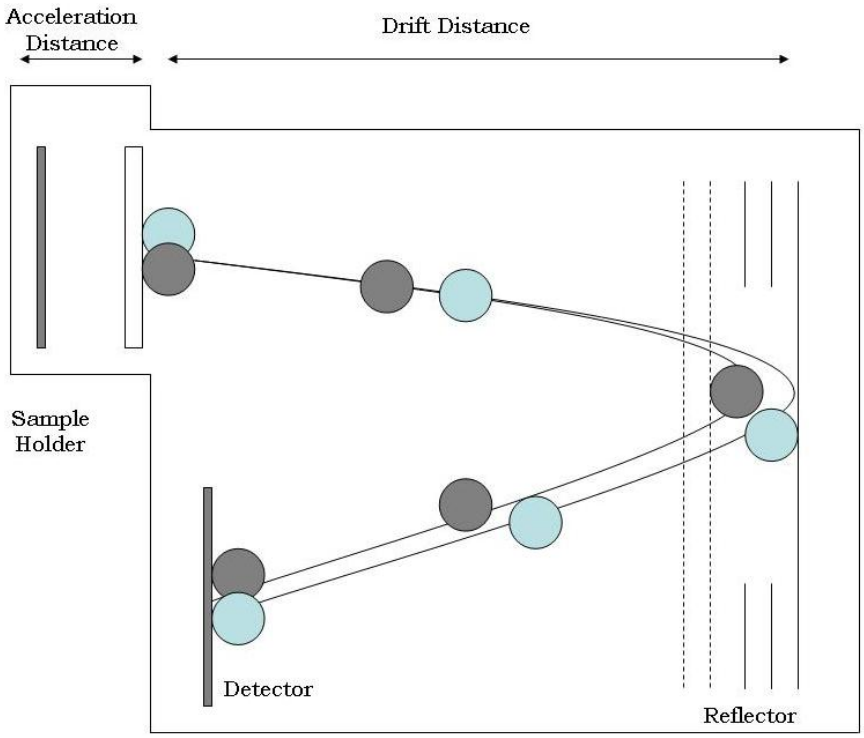


Fig. 15 *A reflectron TOF analyser (image adapted from Manz et al., 2004)*

In a reflector TOF mode, as the ions travel through the field-free tube, they are forced to change direction due to an electrical opposing field which is placed at the end of the tube. Ions that have the same mass but higher velocities get deeper into this field and thus need more time to change direction, but catch up with slower ions at a certain point. By putting the detector at that specific focusing point, very sharp signals can be obtained providing high resolution and mass accuracy (Manz *et al.*, 2004).

MALDI-TOF MS has become the preferred method for high-throughput identification of proteins using peptide mass fingerprinting (PMF). The proteins

are first digested with an endopeptidase such as trypsin. Trypsin cleaves the protein after lysines and arginines generating peptides 8-10 amino acids long. Another endopeptidase commonly in use is Lys-C. Lys-C is used when generation of larger peptides is desired. The resultant fragmented mixture is very specific for a given protein. The peptide mixtures are analysed with MALDI-TOF MS. After peak detection and calibration with internal standards the peak lists are transferred to a peptide mass fingerprint program such as Mascot. The resulting peptide mass fingerprints are compared to theoretical fingerprints from a protein database for identification.

2.6 Nuclear Magnetic Resonance (NMR) Spectroscopy

Nuclear magnetic resonance involves sophisticated manipulation of nuclear spins. Nuclear spin is one of the four physical properties of nuclei, the others being mass, charge and magnetism. Even though mass and charge play fundamental roles in the chemical and physical behaviour of substances, nuclear magnetism and nuclear spin have almost no effect. However, the latter two nuclear properties can be an important tool for elucidating the structures of molecules without disturbing them. The magnetism of a nucleus means that the nucleus can interact with magnetic fields like a tiny bar magnet. Spin can be imagined as the spinning of particles around their axes. However, this oversimplification does not adequately describe the true nature of spin since the concept of the spin is a difficult and highly abstract one, which can only be fully described using quantum mechanical equations.

Most NMR work on biological molecules uses nuclei with the spin number of ½ because the spectra is greatly simplified when there are only two possible quantum states. When an external magnetic field is applied to the sample, the spins start to precess. A compass needle rotates so as to bring the magnetic moment parallel to the field minimising the magnetic energy. However, the response of the spin to the magnetic field is to precess around the magnetic field vector like around a cone, keeping a constant angle (which depends on the initial spin polarisation) between the spin magnetic moment and the field. This type of motion is called precession and is depicted in Fig. 16.

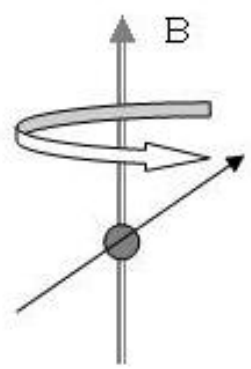


Fig. 16 *Spin precession in the presence of external magnetic field B*

2.6.1 Chemical Shift

The electrons that surround the nuclei in atoms set up tiny magnetic fields of their own. These small magnetic fields act in opposition to the external applied field so that the actual field felt by the nuclei is weaker than the applied field. Since each nucleus in a molecule is in a unique electronic environment, each nucleus is shielded to a different extent from the applied field. Thus nuclei of the same element which are in different chemical environments within a molecule will have different Larmor frequencies. This effect is called chemical

shift. In NMR spectra field strength increases from right to left; left being the "downfield" or "deshielded" part of the spectrum and right being the "upfield" or "shielded" part. Nuclei that require a stronger field strength for resonance will absorb on the upfield side implying that these nuclei are heavily shielded from the external magnetic field.

The chemical shift measured in Hz is field dependent. When the magnetic field increases, the chemical shift increases in a linear proportional way. Since the nuclear Larmor frequency and the chemical shift are both proportional to the applied magnetic field, the ratio of these two quantities is fixed. The field-independent expression of the chemical shift is given by:

$$\delta = \frac{\omega^0 - \omega^0_{ref}}{\omega^0_{ref}}$$

where ω^0 is the Larmor frequency of the nucleus and ω^0_{ref} is the Larmor frequency of the same isotope in a reference compound such as tetramethyl silane for organic molecules. Chemical shifts δ are given in parts per million (ppm) where the dimensionless symbol ppm has the meaning as 10^{-6}. The use of ppm instead of Hz allows for direct comparison of spectra when different magnet strengths are used.

The electron density that a nucleus experiences depends on the atomic number of the atom as well as the electronegativity of neighbouring atoms. However, chemical shift is not affected solely by the electronic cloud surrounding the nucleus. Chemical shift can also be perturbed by additional magnetic fields that arise from the presence of delocalised electrons of aromatic rings or carboxylic

acid groups. This phenomenon is known as the "ring current effect". As an example, aromatic protons because of their close proximity to the centre of the ring experience an increase of ~ 2 ppm due to ring current effects. Other parameters that can influence the chemical shift include paramagnetic ions and hydrogen bonding. Paramagnetic ions such as Fe^{2+}, Mn^{3+} inside a protein molecule may perturb the Larmor frequency of nuclei nearby. Finally, the Larmor frequency of a proton involved in a hydrogen bond is raised slightly.

2.6.2 Indirect spin-spin interaction

Indirect spin-spin interactions are often referred to as through-bond interactions, scalar coupling or *J*-coupling. This type of coupling represents the interaction of the nuclear spins with each other through the involvement of electrons and is independent of molecular orientation. The nuclear spin polarisation of one atom affects the polarisation of the surrounding electrons. The electron polarisation subsequently produces a change in the magnetic field that is sensed by the coupled spin. The result of such an interaction is an observable splitting of the resonance in the NMR spectrum.

The splitting is independent of the magnetic field, and depends only on the nature of the bonding between the two nuclei. The splitting coupling constant *J* gives the splitting in Hz. Since it is independent of the applied magnetic field, this will be the splitting measured on any NMR spectrometer. The size of *J*-coupling depends on several factors, including the gyromagnetic ratio and the number of bonds connecting the coupled spins. In the case of multiple bond

couplings, the value of J-coupling is also affected by the conformation of the intervening bonds. This means that the value of *J* depends on the relative orientation of the magnetic dipoles of the nuclei (the torsion angle) involved in scalar coupling according to the Karplus relationship:

$$^{x}J_{yz} = A \cos^2 \theta + B \cos \theta + C$$

In the case of protons, θ is the angle between the protons for geminal (two-bond) coupling (x=2) or the dihedral angle between protons for vicinal (three-bond) coupling (x=3). The subscripts y and z define the nuclei that are interacting, and A, B and C are empirical constants. The *J*-coupling has a sign. A positive *J* indicates that the spin-spin coupling makes a positive contribution to the energy when the spin polarisations are parallel, and a negative contribution when the spin polarisations are opposite. Typical *J*-coupling values of proteins are shown in Fig. 17.

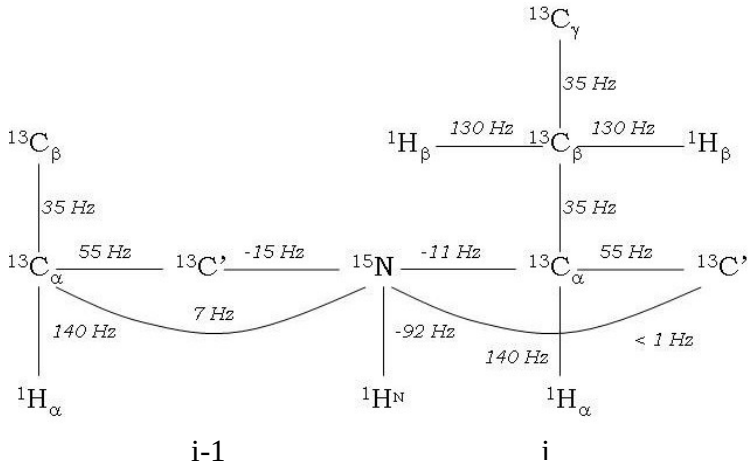


Fig. *17* *Example of J-coupling constants in peptides and proteins.*

2.6.3 Nuclear Overhauser Effect (NOE)

When a weak radiofrequency field is applied at the Larmor frequency of one spin for a sufficiently long time, it has a strong effect on the longitudinal magnetisation of the non-irradiated spins enhancing the magnetisation of those spins. This is known as the Nuclear Overhauser Effect or NOE. The NOE is a dipole-dipole interaction that depends strongly on the distance between the two types of nuclei falling off as the sixth power of internuclear separation (as r^{-6}). It is based on cross-relaxation rather than scalar coupling. If the intensity of a resonance I is I_0, then saturate some neighbouring nucleus and find that intensity I_0 has changed to I, the nuclear overhauser enhancement will be:

$$n_{i(s)} = (I-I_0)/I_0$$

The NOE effect is used in protein structure determination with the acquisition of 2D and 3D NOESY experiments. The basic NOESY pulse sequence is 90^0- t_1- 90^0- t_m - 90^0- t_2. The NOESY experiments still involve cross-relaxation, but there is no saturation of spins. Instead the NOE arises when the magnetisation of a spin precessing at one frequency during a period t_1, is transferred via the cross-relaxation mechanism to another spin during the mixing period known as the mixing time (t_m). The final 90^0 pulse along the x-axis flips the magnetisation from the z-axis to the y-axis where it becomes observable transverse magnetisation. Now the nuclei have a different frequency t_2 and produce a crosspeak revealing the chemical shifts of the two sites.

The second dimension in 2D NOESY (and in every 2D NMR experiment) is achieved by varying the delay t_1 and collecting a series of spectra at various values of this delay. Thus, a second Fourier transform of intensity vs. incremented delay generates a second dimension (frequency axis). The technique can be extended to include another series of delays, thus creating a third dimension and so on. Strong NOE peaks are observed for nuclei which are less then 2.5 Å apart and intermediate ones for 2.5 -3.5 Å apart. Extending the mixture time will reveal crosspeaks from nuclei even 5 Å apart. Normally, for proteins of 10-15 kDa a mixing time of 100 msec is used.

2.6.4 Spin perturbation

Before the pulse, the distribution of magnetic dipoles in the x-y plane is random, i.e. there is no relationship between the transverse (x-y) magnetisation of one spin to another. Thus the transverse magnetisation is termed to be incoherent. Since the sum of a large collection of vectors aligned in random directions is zero, there is no bulk transverse magnetisation at thermal equilibrium.

When a radiofrequency pulse B_1 of the appropriate frequency is applied, the bulk magnetic moment of the sample is tipped from the z-axis to the x-y plane. The extent of precession about B_1 depends on both the strength of the B1 field and the length of time the pulse was applied. Note that the total magnetic field in the rotating frame is now the sum of both the static field (B_o) and the oscillating field (B_1). A pulse of length τ applied at a field of strength ω will rotate the magnetisation through a "flip" angle of $\beta = \omega\tau$. After the application of

the pulse with a flip angle β, a component of the magnetisation cos β will remain along the z-axis and a component in the x-y plane equal to sin β will have been created by the pulse. For example, if a B_1 pulse is applied along the y-axis the individual components of the magnetisation after the pulse are:

$M_z = M_o \cos(\beta)$ $M_x = M_o \sin(\beta)$ $M_y = 0$

When the flip angle is $\pi/2$, the bulk magnetisation is found only along the x-axis. This is referred to as the 90^o pulse. After the B_1 pulse is turned off, the transverse magnetisation precesses in the x-y plane around the B_o field, just as it did before the pulse. The key difference is that the transverse magnetisation is now coherent and gives rise to a non-zero magnetic moment in the x-y plane. Thus the net result of a 90^o pulse is to turn the equilibrium bulk magnetisation from the z-axis and place it in the x-y plane. The effect of a y-pulse on the individual magnetic dipoles is shown in Fig. 18.

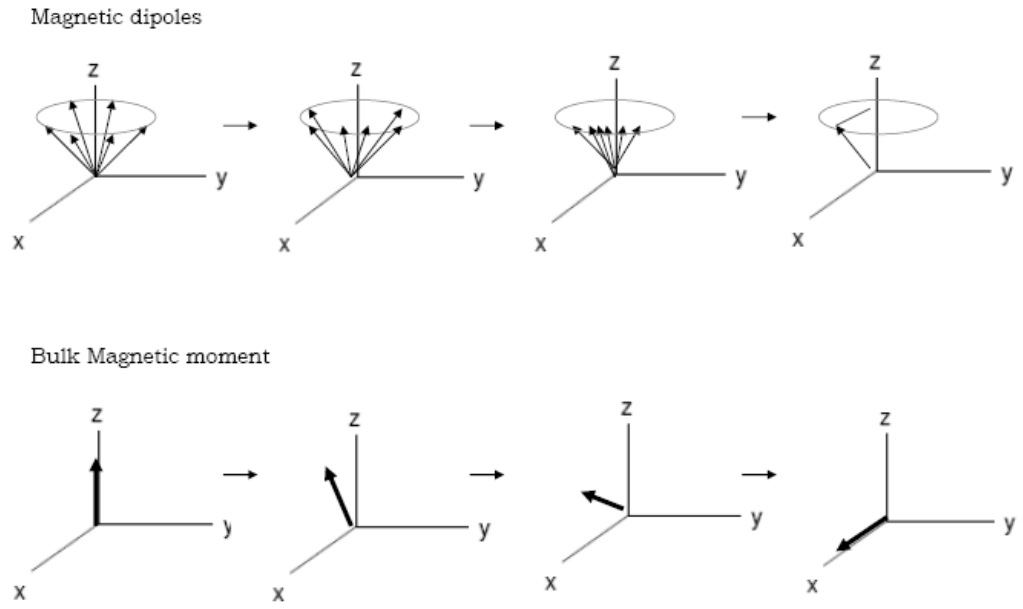


Fig. 18 *Effect of a B_1 y-pulse on the nuclear spins. The upper part of the figure shows a collection of individual magnetic dipoles while the lower part shows the bulk magnetisation. The leftmost part of the panel illustrates the state of the system at thermal equilibrium prior to the pulse. The subsequent sections show the state of the system near the beginning, at the middle and at the end of a 90^o pulse. Image adapted from Rule & Hitchens, 2006.*

When the pulse ends, the nuclei relax and return to their equilibrium positions, and the signal decays. This decaying signal contains the sum of the frequencies from all the target nuclei and it is referred as the Free Induction Decay (FID).

The transverse magnetisation induces a current in the receiver coil of the probe. This induced current is sampled at discrete times (dwell time) and digitised. The

frequency domain spectrum is obtained from the time domain data by Fourier transform (Fig. 19)

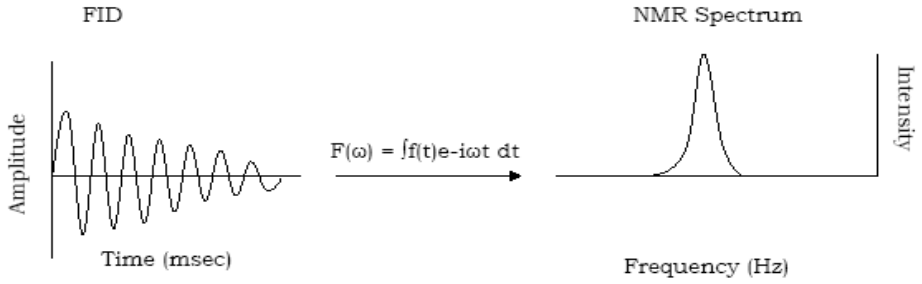


Fig. 19 *The conversion of the free induction decay (FID) signal to an NMR spectrum.*

2.6.5 Spin Relaxation

After perturbation of the nuclear spins by RF pulses, the magnetisation returns to equilibrium via two relaxation pathways (Fig. 20): spin-lattice relaxation and spin-spin relaxation. The first relaxation process arises from a re-alignment of the bulk magnetic moment along the static field transferring energy from the excited state to the surroundings (lattice). The second relaxation process causes reduction of the coherent signal in the x-y plane. The time constants for spin-lattice and spin-spin relaxation are T_1 and T_2 respectively.

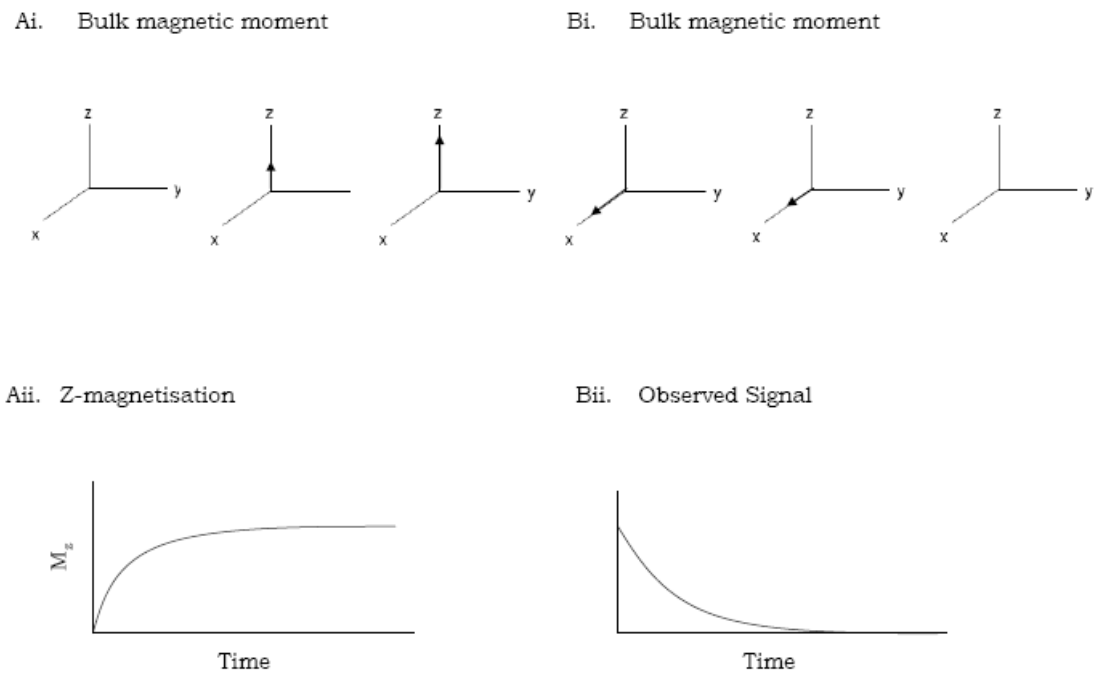


Fig. 20 *Relaxation of nuclear spins. Panel A illustrates spin-lattice relaxation in the absence of spin-spin relaxation and the effect on bulk magnetic moment and z-magnetisation over time (Ai and Aii respectively). Panel B shows spin-spin relaxation in the absence of spin-lattice relaxation. Image adapted from Rule & Hitchens, 2006.*

The experiments used to measure the relaxation parameters (T_1, T_2 and Heteronuclear NOE) are all 2D HSQC-like experiments with pulse sequences that utilize coherence selection to generate quadrature detection and sensitivity enhancement to increase signal to noise. In addition, they are designed to avoid saturation of the water magnetization in order to reduce the saturation of the amide protons that would arise from exchange with the water protons.

HetNOE: The heteronuclear NOE is a measure of the change in the steady-state populations of the heteronuclear (^{15}N) spin when the attached proton (^{1}H) spin is saturated. The pulse sequence of a steady-state heteronuclear NOE experiment utilizes a pulse train to saturate proton equilibrium magnetization prior to the heteronuclear magnetization being excited. Two spectra are collected, one with proton saturation and one without (reference spectrum). The value of the HetNOE is obtained by calculating the ratio of the peak intensities of the spectra taken from the experiment utilizing proton saturation vs the reference experiment. The HetNOE experiment provides information about the motion of individual N-H bond vectors. Those that undergo motion faster than the overall tumbling of the molecule, show a decreased NOE intensity relative to the average observed for the majority of the residues.

T_1/T_2: The T_1 and T_2 times are sensitive to motions on different timescales. T_1 is sensitive to the dynamics on the timescale of ps-μs, whereas T_2 is sensitive to the motions on both ps-μs and μs-ms timescales. For the ^{15}N spins that have a T_1/T_2 ratio below the average value by a difference larger than the standard deviation, local conformational averaging is assumed to be responsible for the shortening of the T_2. When the T_1/T_2 ration is above the average by a difference larger than the standard deviation, the prolongation of T_1 is probably caused by motion comparable to the overall correlation time (τ_m), i.e. ns timescale. The T_1 and T_2 values are also used to define the rotational diffusion properties of the molecule (i.e. whether the molecule tumbles isotropically, axially symmetrically or fully anisotropically).

2.6.6 Residual Dipolar Couplings (RDCs)

A nuclear spin experiences many different sources of local magnetic fields which change as the molecule rotates in solution. Neighbouring nuclear spins also generate magnetic fields of their own. This type of interaction is the magnetic dipole-dipole interaction or dipolar coupling. In an isotropic liquid, the dipole couplings cancel out, but in an anisotropic medium there is incomplete averaging of these magnetic interactions. These residual anisotropic magnetic interactions are called residual dipolar couplings (RDCs) and can be useful in many applications such as the refinement of a 3D structure, characterisation of protein dynamics (Tolman, 2001), and in proteins with multiple domains RDCs can give more information regarding the orientation of these domains (Fischer *et al*., 1999). The principle of residual dipolar couplings is illustrated in Fig. 21.

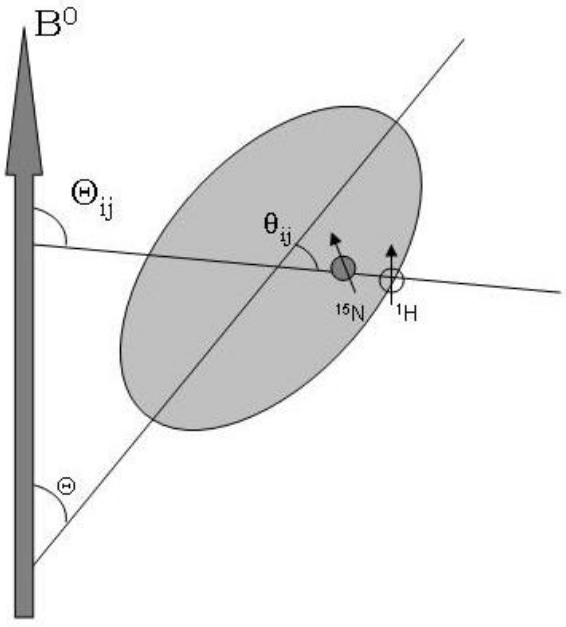


Fig. 21 *The coordinate system for the analysis of residual dipolar couplings (Figure adapted from Levitt, 2001).*

A ^{15}N-^{1}H is depicted in this figure in a weakly oriented molecule. For simplicity, the molecule is depicted as an ellipsoid. By dissolving the molecule in an anisotropic liquid phase, it becomes possible to estimate the angle θ_{ij} which determines the way the ^{15}N-^{1}H vector "points" with respect to the rest of the molecule. The angle Θ is the angle between the molecular long axis and the field and the angle Θ_{ij} is the angle between the internuclear vector and the field. This angular information complements the internuclear distance information revealed by NOESY experiments, to give a more accurate picture of the structure of the molecule.

The average angular dependence of the RDC between a pair of ½ spin nuclei is described by the following equation:

$$D_{ij} = b_{ij} \left\langle \tfrac{1}{2} (3\cos^2 \Theta_{ij} - 1) \right\rangle$$

Where b_{ij} is the dipole-dipole coupling constant:

$$b_{ij} = -\frac{\mu_0}{4\pi} \frac{\gamma_i \gamma_j \hbar}{r^3_{ij}}$$

Here, D_{ij} is the residual dipolar coupling in Hz between nuclei i and j, γ_i and γ_j are nuclear magnetogyric ratios, r_{ij} is the internuclear distance (assumed fixed), and Θ_{ij} angle of the inter-nuclear vector with respect to the external magnetic field. $\hbar$ is Dirac's constant, a physical constant used to describe sizes of quanta. The brackets signify the time average of the quantity.

The term D_{ij} is called the residual dipolar interaction. In a fully anisotropic liquid, D_{ij} vanishes. In a weakly anisotropic liquid, D_{ij} is small but finite. Measurement of the D_{ij} allows the estimation of angle Θ_{ij} which describes the orientation of the internuclear vector with respect to the molecular long axis.

The idea is to induce a weak alignment, because increasing the alignment sharply degrades the resolution of the spectra. If the molecules are oriented too strongly, all of the dipolar couplings become finite, the spin system becomes strongly coupled, and the spectrum becomes intractable. Consequently, the coupling is attenuated by orienting only a small fraction of the molecules of interest, usually one part in 10^4.

Depending on the molecule, a small degree of ordering will give rise to observed dipolar couplings ranging from 1 to 50 Hz. Normally, elongated proteins tend to align more strongly than globular ones.

Information from an RDC measurement is of a "global" nature since it provides the orientation of the inter-nuclear vector with respect to a common axis, the applied magnetic field. Also because this effect depends on $1/r^3$ instead of $1/r^6$ as for the NOE effect, from RDC measurements we can obtain information from longer interatomic distances

The spectrometer's magnetic field has only a very small orientating effect on the protein molecule because normally proteins are diamagnetic. However, it has been possible to measure this tiny alignment (Tjandra & Bax, 1997). Exceptions exist when the protein has a paramagnetic centre, such as a heme group, so that

the anisotropy of the paramagnetic centre is large enough to align the whole molecule to a noticeable extent (Tolman *et al*., 1997).

In order to induce alignment, several alignment media exist such as phospholipid bicelles, cooperative anisotropic membranes, charged polyacrylamide gels or Helfrich phases (Prestegard *et al*., 2004). However, the most popular method for generating an anisotropic environment utilise filamentous bacteriophage (Hansen *et al*., 1998).

One example of filamentous bacteriophage used for RDC measurements is Pf1. Pf1 derived from *Pseudomonas aeruginosa,* is a rod-shaped filamentous bacteriophage, 2000 nm long with a 6.6 nm in diameter, harbouring single-stranded circular DNA. The DNA is packaged with coat protein at a 1:1 nucleotide: coat protein ratio. The coat proteins form a repeating network of carbonyl groups that are believed to be the source of the phage's large anisotropic magnetic susceptibility (Torbet & Maret, 1979). Pf1 is negatively charged at pH 7.4. Pf1 has several advantages as an alignment medium compared to other alignment media. Some characteristics of Pf1 are as follows:

- It aligns spontaneously in the presence of a magnetic field (with the rod lying parallel to the magnetic field).
- Even though the addition of phage makes the protein solution much more viscous, this viscosity has surprisingly no effect on the rotational correlation time of the protein.
- It is stable at pH 6-8
- The aligned sample can be measured even after storage at 4 ^{0}C (although prolonged storage times will increase the alignment) No ^{1}H or ^{31}P

resonances are observed from the large phage particles, thus the phage does not add more peaks in the protein NMR spectrum.

- It is possible to separate the phage from the dissolved protein by ultracentrifugation which pellets the phage while leaving the protein in the supernatant.

Upon addition of the alignment medium to the sample, the degree of alignment is measured by observing the splitting of the D_2O signal. The splitting of D_2O signal arises from the large deuterium quadrupole moment that is not isotropically averaged for water bound to the aligned phage particles. By adding more phage to the protein sample, the alignment increases. A splitting of no more than ~15 Hz is normally used. Higher degrees of alignment cause line broadening from unresolved proton-proton dipolar coupling.

After addition of the alignment medium, two independent ^{15}N-^{1}H IPAP-HSQC spectra are acquired. These spectra are acquired in an interleaved manner without ^{1}H decoupling in the ^{15}N dimension. In one data set, the scalar and dipolar coupling lead to in-phase splitting of the resonances in the nitrogen dimension, while in the other data set, this splitting is anti-phase. The addition of these spectra gives a spectrum that contains one of the doublets while the subtraction of the two spectra gives a spectrum containing the other doublet. This approach is referred to as IPAP (in-phase, anti-phase) (Ottiger *et al*., 1998).

On aligned sample the splitting is the sum of the scalar and dipolar coupling (J+D). The splitting observed in the aligned sample is either either larger than

the J-coupling, or slightly smaller depending on the orientation of the ^{15}N-^{1}H vector with respect to the molecule. If the vector is parallel to the long axis of the molecule, the doublet splitting is slightly smaller.The RDC values are measured as difference in splitting between aligned and isotropic phase.

2.6.7 NMR Experiments for protein structure determination

Because for the present work, no pulse design was attempted, a brief overview of the experiments used for resonance assignments is described below. A full comprehensive experimental setup guidance and pulse sequences used for the following experiments can be accessed online at:
(http://nmr-linux.chem.ed.ac.uk/highfield/highfield.html)

2.6.7.1 2D Experiments for backbone and side-chain assignment:

^{1}H-^{15}N HSQC and ^{1}H-^{13}C HSQC

In order to identify each amino acid in the assignment process, a spectroscopic "handle" is chosen to which all the backbone and side-chain resonances are associated. This is used to catalogue all the chemical shifts of a given amino acid. This kind of "handle" is the backbone amide proton and nitrogen as each amino acid contains only one amide group in the backbone. Thus, the NH group serves as a reference point to which all other backbone and side-chain atoms can be linked. The spectrum that gives information for the dispersion of the NH groups of amino acids in the molecule is the ^{1}H-^{15}N HSQC. The NH groups shown in an HSQC experiment are the backbone amide groups, Trp side-chain N^{ε}-H^{ε} groups, Asn/Gln side-chain N^{δ}-$H^{\delta 2}$/N^{ε}-$H^{\varepsilon 2}$, Arg N^{ε}-H^{ε} peaks (usually

appearing as folded peaks). The Arg N^{η}-H^{η} and Lys N^{ζ}-H^{ζ} are not normally visible due to water exchange unless the experiment is recorded at very low pH. (for the side-chain assignment the ^{1}H-^{13}C HSQC is also used as reference).

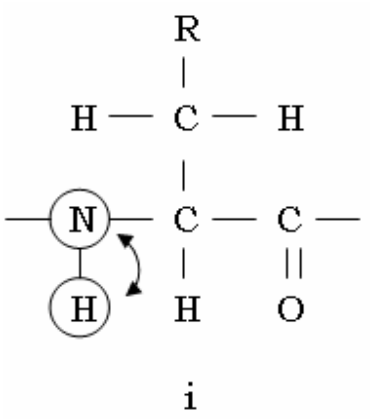


Fig. 21 *Transfer of magnetisation in the ^{1}H-^{15}N HSQC experiment.*

2.6.7.2 3D Experiments for backbone assignment:

CBCANH

This experiment correlates the amide ^{1}H and ^{15}N signals with both the intraresidue (i) C_{α} and C_{β} signals and with the C_{α} and C_{β} of the preceding residue (i-1). The experiment exploits the fact that each amide nitrogen is coupled to its own C_{α} ($^{1}J_{CN}$ ~12 Hz) more strongly than to the C_{α} preceeding it ($^{2}J_{CN}$ ~7 Hz). As a result, the strip for each N-H contains intense peaks for its own C_{α} and C_{β} resonances and weaker peaks for the C_{α} and C_{β} resonances of the previous residue. The way that magnetisation is transferred between C_{α} and C_{β} causes the cross peaks of these carbons to have opposite signs. Aligning the weak peaks in one strip with the strong peaks from another establishes that the corresponding N-H pairs are adjacent in the sequence.

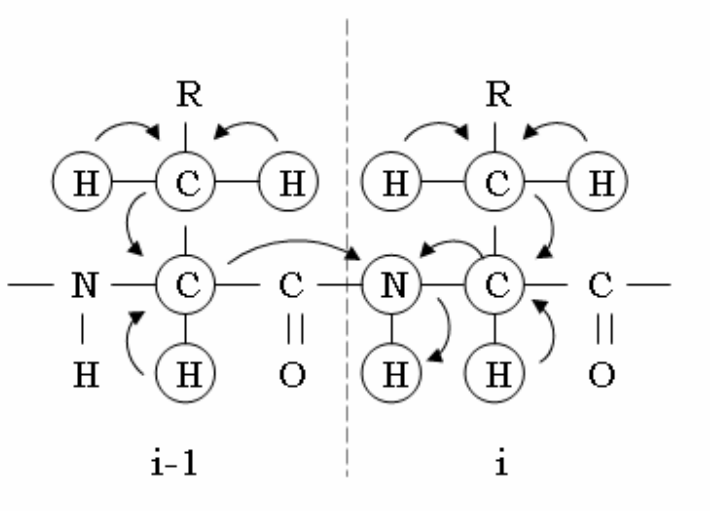


Fig. 22 *Transfer of magnetisation in the CBCANH experiment*

CBCA(CO)NH

This experiment correlates the ^{1}H and ^{15}N amide resonances of one residue with both C_α and C_β resonances of its preceding residue. A more sensitive experiment (less time to acquire data) than its "partner" CBCANH. The interresidue correlations are established by transferring coherence via the intervening ^{13}CO spin. Two peaks are generated in this spectrum instead of four in the CBCANH. The CBCA(CO)NH shows carbon peaks only from the i-1 residue whereas the CBCANH shows both i-1 (these should overlap with the CBCA(CO)NH) and i peaks.

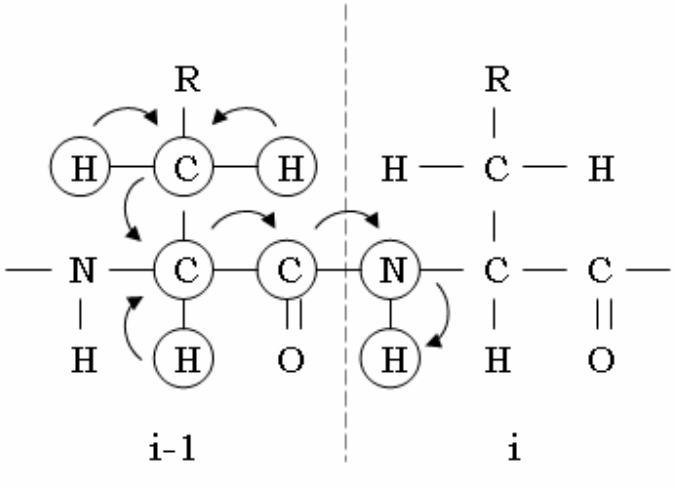


Fig. 23 *Transfer of magnetisation in the CBCA(CO)NH experiment*

HNCO

This experiment correlates amide protons with carbonyl carbons. The magnetisation is transferred from the $H^N_{(i)}$ proton via the $N_{(i)}$ atom to the directly attached $CO_{(i-1)}$ carbon atom and return the same way to the $H^N_{(i)}$ nucleus which is directly detected. The most sensitive experiment of all the backbone assignment experiments, HNCO was among the first triple-resonance experiments to be developed and its high sensitivity makes it one of the most widely used in assignment strategies.

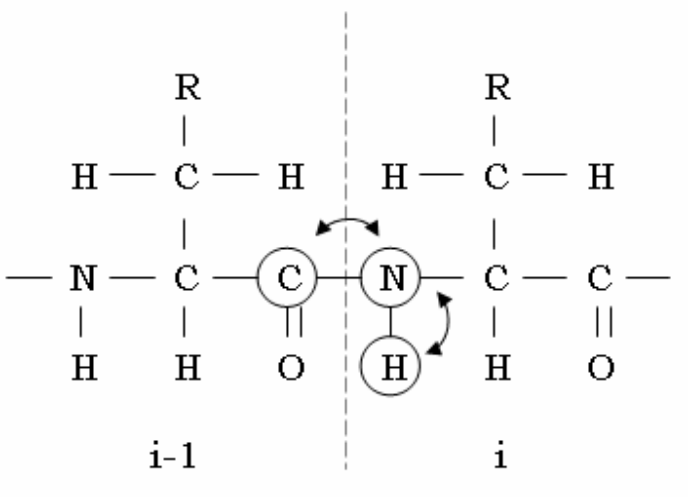


Fig. 24 *Transfer of magnetisation in the HNCO experiment.*

HN(CA)CO

This is a complementary experiment to HNCO. In the HN(CA)CO experiment the magnetisation is transferred from the $H^N_{(i)}$ proton via the $N_{(i)}$ atom and the $C_{\alpha(i)}$ to the $CO_{(i)}$ and $CO_{(i-1)}$ carbon atom and back the same way. The $C_{\alpha(i)}$ and $C_{\alpha(i-1)}$ act only as a relay nuclei, their frequency is not detected. Only the frequencies of H^N, N and CO are detected. The experiment provides

intraresidue and some weaker interresidue correlations, but is much less sensitive than HNCO, as it relies on the small $^1J_{NCa}$ coupling.

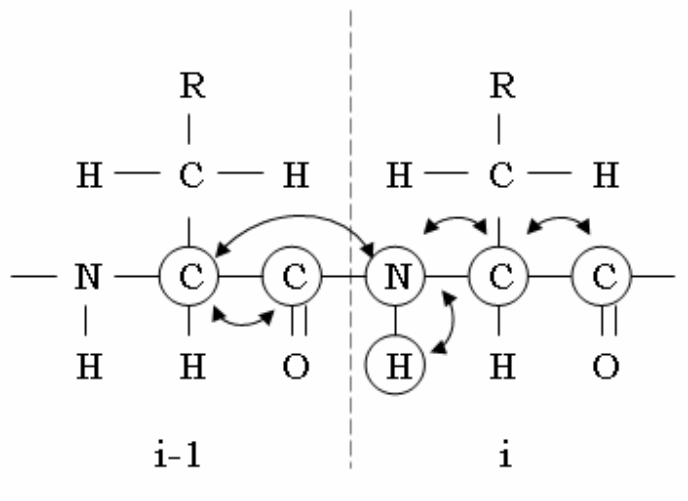


Fig. 25 *Transfer of magnetisation in the HN(CA)CO experiment.*

HBHANH and HBHA(CO)NH

The HBHANH and HBHA(CO)NH experiments are collectively used to assign each $H_{\alpha(i)}$, $H_{\beta/\beta'(i)}$, $H_{\alpha(i-1)}$, $H_{\beta/\beta'(i-1)}$ resonance quartet (or pentet or sextet depending on the multiplicity and degeneracy of the H_β protons) to its amide resonance group. The magnetisation of HBHANH and HBHA(CO)NH is similar to the CBCANH and CBCA(CO)NH accordingly (see Fig. 22 and 23), however, this time the peaks in the spectrum belong to the protons rather than the carbons. H_α and H_β resonances complement the C_α and C_β resonances in establishing unique chemical shift based links between independent amino acid spin systems.

2.6.7.3 3D Experiments for side-chain assignment:

HC(C)H-TOCSY

The HC(C)H-TOCSY experiment correlates all aliphatic ^{1}H and ^{13}C spins within residues and is used to assign aliphatic ^{1}H and ^{13}C resonances and connect the side-chain chemical shifts with the backbone assignments. The reference point for the sidechain assignments is not the ^{1}H-^{15}N HSQC as it is for the backbone assignment, but the ^{1}H-^{13}C HSQC spectrum. All proton shifts of the (i) residue from the HC(C)H-TOCSY are correlated with C^{H}-H^{C} of the ^{1}H-^{13}C HSQC. The HC(C)H-TOCSY experiment spreads a 2D-TOCSY spectrum into a 3rd dimension to reduce signal overlapping. The magnetisation transfer starts from the ^{1}H to its directly attached ^{13}C nucleus via the $^{1}J_{CH}$ coupling (~140 Hz), then from the ^{13}C to neighbouring ^{13}C nuclei via the $^{1}J_{CC}$ coupling (~35 Hz), and finally, from the ^{13}C back to the directly attached ^{1}H.

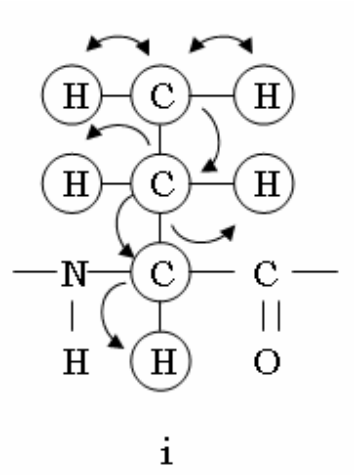


Fig. 26 *Transfer of magnetisation in the HC(C)H-TOCSY experiment.*

H(CCO)NH and C(CO)NH

Two other experiments that help with the assignment of sidechain atoms are the H(CCO)NH and C(CO)NH . They are useful experiments for the identification of C_γ, C_δ and H_γ and H_δ when the information from the HC(C)H-TOCSY is ambiguous. The experiments have similar magnetisation transfer pathways with the H(CCO)NH correlating sidechain proton chemical shifts from i-1 residue with the i residue amide groups and the C(CO)NH correlating the carbon chemical shifts. The transfer of magnetisation starts from the side-chain hydrogen nuclei to their attached ^{13}C nuclei and subsequently to their adjacent carbon nuclei through isotropic mixing. From here, magnetisation is transferred to the carbonyl carbon, on to the amide nitrogen and finally to the amide hydrogen for detection.

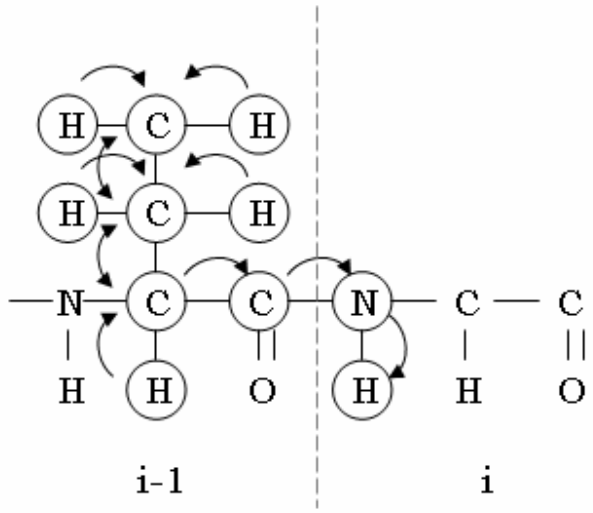


Fig. 27 *Transfer of magnetisation in the H(CCO)NH experiment.*

2.6.7.4 2D Experiments for the assignment of aromatic residues:

Three 2D experiments were used for the assignment of aromatic side-chains. A ^{1}H-^{13}C HSQC optimised for aromatic residues, and aromatic (HB)CB(CGCD)HD / (HB)CB(CGCDCE)HE experiments. The (HB)CB(CGCD)HD correlates C_β with H_δ of aromatic residues, and the (HB)CB(CGCDCE)HE correlates C_β with H_ε of aromatics. These experiments were utilised for the complete assignment of histidine, phenylalanine, tryptophan and tyrosine residues.

2.6.7.5 3D Experiments for the acquisition of NOE restraints:

^{15}N-NOESY-HSQC and ^{13}C-NOESY-HSQC

3D NOESY-HSQC is formed by combining an HSQC spectrum with a NOESY spectrum after removing the acquisition period of the NOESY and the preparation period of the HSQC. The mechanism involves a ^{1}H-^{1}H NOE step followed by heteronuclear transfer via $^{1}J_{NH}$ (or $^{1}J_{CH}$). At the beginning of the experiment, the magnetisation is labelled with the frequency of a proton (H^X). After the evolution of H^X chemical shifts, the magnetisation is transferred to vicinal amide protons by dipolar coupling during the NOESY mixing period τ_m. the magnetisation is then passed via scalar coupling to the bound amide nitrogen. The final step involves the transfer of magnetisation again through scalar coupling from the amide nitrogen to the amide proton. Thus the final signal is recorded with three frequencies: H^X, ^{15}N and H^N. The H^N is detected in the direct dimension while the amide nitrogens and the chemical shifts of the protons close in space are detected indirectly. Each strip in the spectrum

contains NOEs from one N-H (or C-H) group to all other hydrogen atoms close by.

The relatively poor chemical shift dispersion of protons can often lead to multiple assignment possibilities for a given NOE. Thus, in order to help the structure calculation programs reach a fairly correct conformation even from the first round of calculation, some NOEs which are unambiguous are manually assigned.

Chapter 3: Methods

3.1 Expression and purification of ArdA/ KlcA

An almost identical protocol was used to purify ArdA or KlcA. *E. coli* BL21(DE3) was transformed with pDMArdA (pTrc99A containg the *orf18* gene) or pDMKlcA (pET24-a containing the *klcA* gene) and plated onto LB agar supplemented with 100 μg/ml carbenicillin or 25 μg/ml kanamycin. A single colony was picked and grown in LB medium plus 100 μg/ml carbenicillin (or kanamycin) to an optical density at 600 nm (OD600) of 0.5 at 37 ^{0}C. Heterologous gene expression was induced by the addition of IPTG to 1 mM and the growth continued for a further 4 hours at 37 ^{0}C for ArdA and at 30 ^{0}C for KlcA before harvesting. The amount of recombinant protein was assessed by SDS-PAGE (4-12% gradient gel) analysis. Typically 40 g of cells from 10 L of culture broth were obtained. Cells were harvested by centrifugation (6000 *g*, 4 ^{0}C, 15 min) and the cell pellet was stored at -20 ^{0}C until required.

Approximately 20 g of cell pellet was defrosted on ice for 30 min and resuspended in 200 ml of buffer A (20 mM Tris–HCl, pH 8.0, and 7 mM 2-mercaptoethanol). Protease inhibitors benzamidine and PMSF (phenylmethylsulfonyl fluoride) were added to a final concentration of 10 μM. The cells were disrupted by sonication on ice using a Soniprep 150 sonicator (Sanyo, Tokyo, Japan) fitted with a 9-mm-diameter probe (5×1 min bursts with 15 s intervals between each burst). The extract was centrifuged at 20,000g for 1 h at 4 °C, and the supernatant was filtered through a filter unit (0.45 μM; Sartorius

AG, Goettingen, Germany). The clarified extract was then loaded onto a DEAE anion exchange column (30 cm×1.6 cm diameter), equilibrated with buffer A, at a flow rate of 60 ml/h. The column was washed with approximately three column volumes of buffer A to remove unbound material. Bound proteins were then eluted using a 500-ml gradient of 0–1 M NaCl in buffer A at a flow rate of 20 ml/h. Fractions that gave a UV absorbance were subsequently analysed by SDS-PAGE. Fractions containing ArdA or KlcA were pooled and concentrated using a spin concentrator with a 10-kDa cutoff membrane (VivaScience AG, Hanover, Germany) to a final volume of 5 ml. The sample was then loaded on a Superdex-200 column (GE Healthcare, Uppsala, Sweden), equilibrated with buffer A plus 0.2MNaCl, and the protein was eluted at a flow rate of 30 ml/h. The fractions were subsequently analysed by SDS-PAGE, and those containing ArdA or KlcA were pooled and concentrated as before. The extinction coefficient of ArdA was calculated using the ExPASy ProtParam tool to be 30,370 cm^{-1} M^{-1}, assuming ArdA is a monomer. For the KlcA the extinction coefficient was calculated to be 24,410 cm^{-1} M^{-1} assuming KlcA to be a monomer. Purified ArdA or KlcA was stored at −20 °C in buffer A containing 50% (v/v) glycerol.

3.2 Size-exclusion HPLC

The molecular mass of ArdA at different protein concentrations was determined by size-exclusion HPLC. The experiment was conducted using a Superose 12 (10/300) column (GE Healthcare) at a flow rate of 0.5 ml/min. The buffer used was 20 mM Tris–HCl, 20 mM 4-morpholine-ethane sulfonic acid (Na salt), 200 mM NaCl, 10 mM $MgCl_2$, 7 mM 2-mercaptoethanol and 0.1 mM

ethylenediaminetetra-acetic acid adjusted to pH 6.5. Protein elution from the column was monitored by UV absorption at 280 nm and fluorescence emission at 350 nm with excitation at 295 nm. The column was first calibrated with protein standards (bovine serum albumin, alcohol dehydrogenase, carbonic anhydrase and cytochrome c), which were run individually. The ArdA was then analysed at a range of different concentrations from 50 nM to 100 μM.

3.3 Circular Dichroism

CD measurements were carried out on a Jasco model J-180 spectropolarimeter (Jasco Corporation, Tokyo, Japan) in 10 mM Tris–HCl, pH 8.0, 50 mM NaF and 7 mM 2- mercaptoethanol buffer. Far-UV CD spectra (190–260 nm) used a protein concentration of 30 μM and a 0.2-mm pathlength cell. Near-UV CD measurements (260–320 nm) used a protein concentration of 130 μM with a 0.5-cm path-length cell. A bandwidth of 1 nm was used with a scan speed of 10 nm/min. All CD measurements were made at 25 °C, and each spectrum was the average of five individual scans. The spectra were corrected for buffer contribution.

3.4 Fluorescence quenching for the determination of tryptophan accessibility

Fluorescence quenching measurements were performed on an Edinburgh Instruments FS900 spectrofluorometer (Edinburgh, UK). The concentration of ArdA was 3 μM in buffer A, 25 °C. Samples were excited at 295 nm, and the emission spectra were recorded between 300 and 420 nm with 5-nm bandwidth. Tryptophan fluorescence was

quenched by successive additions of acrylamide solution (5 M). Iodide quenching was also performed with successive additions of 5 M KI supplemented with 2 mM $Na_2S_2O_3$ to prevent the formation of I^-_3 (Lakowicz, 2006). The fluorescence of the buffer was subtracted from the emission spectra. The spectra were also corrected for dilution and inner filter effects (acrylamide has a molar extinction coefficient of 0.23 M^{-1} at 295 nm). The data were fitted with the Stern–Volmer equation (Eftink & Ghiron, 1981); $F_0/F = (1+K_{sv}\ [Q])$ where F_0 and F are the fluorescence intensities in the absence and presence of quencher respectively, [Q] is the quencher concentration and k_{sv} the collisional quenching constant. The fitting was performed using Origin 7.5 analysis software (MicroCal, Northampton, MA).

3.5 Determination of the unfolding free energy ΔG using fluorescence spectroscopy

Equilibrium unfolding as a function of GdmCl was monitored by fluorescence spectroscopy. A stock solution of 8M GdmCl was made up, and the precise concentration was determined from the refractive index of GdmCl (Creighton, 1990). Native protein (3.5 μM) in buffer A was incubated with various concentrations of GdmCl at 25 °C and allowed to equilibrate overnight at room temperature. The fluorescence intensity was then measured for each sample using excitation at 295 nm and emission at 330 nm with 5-nm bandwidths. Experimental data were fitted to a two-state unfolding model (Pace *et al.*, 1989) assuming a linear relationship between free energy of unfolding and concentration of GdmCl as follows:

$$S = \frac{\{s_F + m_F D + (s_U + m_U D)\exp[-(\Delta G_w - mD)/RT]\}}{\{1+\exp[-(\Delta G_w - mD)/RT]\}}$$

Where ΔG_w: free energy of unfolding in water ; D: molar concentration of denaturant ; S: observed spectroscopic signal ; s_F: spectroscopic signal of the folded state ; m: slope of transition; s_U: spectroscopic signal of the unfolded state; m_F: slope of the folded state; m_U: slope of the unfolded state; R: gas constant (8.314 $J \cdot K^{-1} \cdot mol^{-1}$) and T: absolute temperature (Kelvin).

3.6 DSC

DSC measurements were done using a VP-DSC calorimeter (MicroCal) with a cell volume of 0.52 ml. The experiments were performed in 20 mM Tris–HCl, pH 8.0, 6 mM MgCl2 and 7 mM 2-mercaptoethanol buffer with a protein concentration of 50 μM. DSC data were corrected for instrument baselines and normalised for scan rate and protein concentration. Scans were from 15 to 100 °C at a rate of 60°/h. Samples were degassed under vacuum for 2 min with gentle stirring before being loaded into the calorimetric cell. Experiments were performed under a constant external pressure of 25 psi to avoid bubble formation. Once the first scan was completed, the sample was cooled and rescanned under the same experimental conditions to test for the reversibility of the denaturation reaction. Fitting of the partial heat capacity (Cp) curves using a two-state unfolding model was performed using MicroCal LLC Origin software.

3.7 ITC

ITC was performed using a VP-ITC calorimeter (Micro-Cal). The stocks of ArdA and M.EcoKI were transferred into 20 mM Tris–HCl, pH 8.0, 6 mM MgCl2 and 7 mM 2-mercaptoethanol buffer using a PD-10 column (GE Healthcare). SAM was added to a final concentration of 100 μM. Protein and buffer were degassed prior to use. A stock solution of ArdA (51.8 μM) was titrated into 5.0 μM M.EcoKI in the VP-ITC cell (active volume=1.4 ml). The heat of dilution was obtained by injection of ArdA into buffer alone or buffer into buffer. These values were subtracted from the ITC titration data. Calorimetric data were converted to differential binding curves by integration of the resultant peaks and fitted to a standard single-site binding model using MicroCal LLC Origin software.

3.8 *In vivo* Restriction/Modification assays

A spot test was first performed to test for alleviation of restriction by ArdA or KlcA. A 200-μl aliquot of an overnight culture of restriction-proficient *E. coli* strains (NM1049, restriction system Type IA; NK354, Type IB; NK402, Type IC; NM1009, Type ID) and restriction-deficient *E. coli* (NM1261 strain) harbouring either pDMArdA or pTrc99A were mixed with 3 ml of molten BBL top agar (<40 °C) and poured over BBL bottom plates containing carbenicillin. A 5-μl aliquot of various dilutions (10^{-1}–10^{-9}) of unmodified virulent phage lambda, λv.0, was then spotted in a defined array onto the plates. After incubation at 37 °C overnight, the number of plaques within each spot was assessed. Based on the results of these spot tests, efficiency of plating (EOP) experiments were then performed. A 200-μl aliquot of an overnight culture from each test strain was

mixed with 50–100 μl of phage at an appropriate dilution to give a total of 100–200 plaques per plate. The mixture was incubated at 37 °C for 15 min to allow phage absorption and then mixed with molten BBL top agar (3 ml) and immediately poured onto a BBL bottom plate. Each experiment was performed in triplicate. The plates were incubated at 37 °C overnight.

EOP was derived by determining the titre of phage on the restriction-proficient strains relative to the nonrestricting strain. Antirestriction was defined as the EOP obtained with strains containing the plasmid pDMArdA divided by the EOP obtained from strains containing the plasmid pTrc99A. For the antimodification assays, plaques grown on each *E. coli* strain in the presence (i.e., strains harbouring pDMArdA) or absence (i.e., strains harbouring pTrc99A) of ArdA were picked and resuspended in 1 ml of phage buffer containing 50 μl of chloroform. The phage suspension was incubated for at least 2 h at room temperature to allow any remaining bacteria to be killed. After spinning the suspension at 6000 g for 1 min, spot tests and EOP measurements were performed as described for the restriction assay experiments using strains that had not been transformed with pDMArdA or pTrc99A. The degree of antimodification was assessed by determining the EOP of each phage stock on the restriction-proficient strain relative to the nonrestricting strain. Antimodification was defined as 1/(EOP).

The same procedure was followed for the *in vivo* restriction/modification assay involving pDMKlcA. In this case, however, since the vector was pET24-a, the r^-m^- and r^+m^+ strains needed to be made lysogenic. The strains were made

lysogenic (courtesy of Dr. Gareth Roberts, Edinburgh University) using the λDE3 lysogenization kit (Novagen) with some modifications. Where necessary, the selection phage provided in the kit was replaced with NM848 phage. Lysogenisation was then verified using a T7 tester phage (Novagen).

3.9 *In vitro* antirestriction activity of ArdA/ KlcA

The *in vitro* assay for antirestriction activity monitored the cleavage of unmethylated pBRsk1 using purified EcoKI in the absence or presence of ArdA or KlcA. The same protocol was used for both ArdA and KlcA. Typical reaction digests were performed in a 50-µl volume at 37 °C and contained 100 mM Tris–acetate, pH 7.5, 100 mM magnesium acetate, 7 mM 2-mercaptoethanol, 50 µg/ml bovine serum albumin, 2 mM ATP, 0.1 mM SAM, 3 nM pBRsk1 and 30 nM EcoKI. Where required, ArdA or KlcA (3, 30 and 300 nM of monomer) was briefly preincubated with EcoKI in the reaction mix prior to initiating the reaction with ATP. Reactions were performed for 10 min and then quenched by incubation at 68 °C for 10 min. The effect of ArdA or KlcA on the linearisation of pBRsk1 with EcoRI was also tested. These reactions were performed using 5 U of EcoRI at 37 °C for 25 min in the following buffer: 50 mM NaCl, 100mMTris–HCl, 10 mM $MgCl_2$ and 0.025% Triton X-100, pH 7.5. The reaction was stopped by incubation at 68 °C for 10 min. Samples were mixed with gel loading buffer and loaded onto 0.8% agarose gel and subjected to electrophoresis in TAE buffer (40 mM Tris acetate and 2 mM ethylenediaminetetraacetic acid) at 100 V/h. DNA was visualised using ethidium bromide (0.1 µg/ml) under UV illumination.

3.10 *In vitro* assay of DNA displacement from M.EcoKI by ArdA

Competition for binding either ArdA or a DNA duplex to M.EcoKI was determined using fluorescence anisotropy. A 21-bp DNA duplex containing the EcoKI target site sequence (AAC(N_6)GTGC), and end-labelled with hexachlorofluorescein was used at a concentration of 2 nM in 20 mM Tris–HCl, pH 8.0, 6 mM $MgCl_2$ and 7 mM 2-mercaptoethanol buffer plus 100 μM SAM. ArdA, if present, was at a monomer concentration of 50 nM. M.EcoKI was titrated to DNA in the presence or absence of ArdA up to a final concentration of 1 μM. The fluorescence anisotropy change caused by the addition of M.EcoKI was measured at 25 ^{0}C on an Edinburgh Instruments FS900 spectrofluorometer (Edinburgh, UK). The excitation and emission pathlengths were 10 and 2 mm respectively. The samples were excited at 530 nm and emission was detected at 560 nm.

3.11 NMR spectroscopy and structure calculations

KlcA was uniformly labelled with $^{15}N/^{13}C$ by growing the recombinant strain in M9-minimal medium containing $^{15}NH_4SO_4$ and ^{13}C-D-glucose as the sole nitrogen and carbon sources, respectively; (per litre: 25.6 gr Na_2HPO_4, 6.0 gr KH_2PO_4, 1.0 gr NaCl, 2.0 gr $^{15}NH_4SO_4$, 4 ml $MgSO_4$ (1M), 1 ml $CaCl_2$ (1M), 40 ml 20 % ^{13}C-D-glucose). The labelled protein was purified as described above. Samples for NMR measurements contained 0.7 mM KlcA, 90% H_2O/10% D_2O and 0.05% NaN_3 in 20 mM deuterated sodium acetate buffer (pH 6.0). All NMR

experiments were carried out at 298 K unless stated otherwise on a Bruker AVANCE 800 MHz with a 5-mm triple resonance cryoprobe. NMR experiments for ^{1}H, ^{15}N, and ^{13}C backbone and sidechain resonance assignments included: 2D ^{1}H-^{1}H TOCSY, ^{1}H-^{1}H NOESY, ^{1}H-^{15}N HSQC, ^{1}H-^{13}C HSQC, 3D HNCO, HN(CA)CO, CBCANH, CBCA(CO)NH, HBHA(CO)NH, HBHANH, (H)CCH-TOCSY, H(CCO)NH, C(CO)NH, and aromatic 2D (HB)CB(CGCD)HD/(HB)CB(CGCDCE)HE using parameters as described previously (Uhrinova *et al*., 1998) 3D ^{15}N- and ^{13}C-edited NOESY-HSQC spectra were acquired to obtain the distance constraints (Muhandiram *et al*., 1993). Aromatic $^{1}H^{\delta}$ and $^{1}H^{\varepsilon}$ frequencies were obtained from the combined use of the (HB)CB(CGCD)HD/(HB)CB(CGCDCE)HE spectra (Yamazaki *et al*., 1993) where the aromatic ^{1}H frequencies are correlated with the assigned $^{13}C^{\beta}$ frequencies of the aromatic residue.

NMR data were processed using the Azara suite of programs provided by Wayne Boucher and the Department of Biochemistry, University of Cambridge, UK and analysed by the CcpNmr Analysis program (Vranken *et al*., 2005). NOE assignments of KlcA were generated using the CANDID module (Herrmann *et al*., 2002) of CYANA 2.1 software (Guntert, 2004). Dihedral constraints were generated using TALOS (Cornilescu *et al*., 1999). The standard protocol was used with seven cycles of combined automated NOE assignment and structure calculation of 100 conformers in each cycle. The NOE assignments obtained by CANDID were checked manually and from the 100 structures calculated, 20 structures with lowest target function were selected for further analysis. The structures were further refined using residual dipolar couplings (RDCs) using

the Crystallography and NMR system (CNS) (Brunger *et al.*, 1998). PROCHECK-NMR (Laskowski *et al.*, 1996) and WHATIF (Vriend, 1990) were used to analyse the quality of the structures. MOLMOL (Koradi *et al.*, 1996) and PYMOL (http://www.pymol.org) were used for visualisation.

A 200 μM ^{15}N-labelled KlcA sample in 20 mM deuterated sodium acetate, 0.05% NaN_3, pH 6.0 in 90% H_2O/10 % D_2O aligned with 6.25 mg/ml of Pf1 filamentous phage (Pf1 protease and RNase free, Profos) was used to measure RDCs. Magnetic alignment was monitored by 1D ^{2}H NMR spectra as a function of phage concentration. A 7 Hz residual quadropolar splitting of the deuterium signal was achieved with 6.25 mg/ml phage. ^{1}H-^{15}N splittings were measured under isotropic and partially aligned conditions using 2D IPAP ^{1}H-^{15}N HSQC experiments (Ottiger *et al.*, 1998). Analysis and back-calculation of the residual dipolar couplings was performed with the program REDCAT (Valafar & Prestegard, 2004).

Relaxation experiments were performed on a 600 MHz Bruker spectrometer at 25 °C with 200 μM ^{15}N labelled KlcA. For the longitudinal (T_1) and transverse (T_2) relaxation data for the backbone ^{15}N nuclei of KlcA, water was kept along the z axis during the entire pulse sequence (Grzesiek & Bax, 1993) and extra delays were used in the T_1 experiment to eliminate the differential heating arising from the variations of the relaxation delay (Yip &, Zuiderweg, 2005). Eight interleaved spectra were recorded for each type of relaxation with different duration of the relaxation delay, T_1 (51 (twice), 401, 601, 801, 901, 1001, 1201 ms) and T_2 (16 (twice), 48, 64, 80, 96, 112, 144 ms).

T_1 and T_2 cross-peak intensities were measured from peak heights and fit to an exponential decay function I (t) = I(o)e (-R* t) where I(o) is the cross-peak intensity at time zero, I(t) is the cross-peak intensity at time t (t is the relaxation delay) and R (1/T) is the relaxation rate. . The fit was performed by non-linear least-squares method using the Analysis software package (Vranken *et al*., 2005). For the 2D ^{1}H-^{15}N heteronuclear NOE experiment (Grzesiek & Bax, 1993) a 3 s pre-saturation period and a 2 s relaxation delay was used; the control experiment had an equivalent 5 s delay. In these experiments, water was returned to the z-axis prior to each scan to avoid saturation transfer. Spectra from the heteronuclear NOE experiment were extended by linear prediction and zero filling prior to apodization with a 5 % shifted mixed Gaussian/exponential function. NOE peak heights were taken from the Gaussian fits of the frequency-domain data. The NOE was calculated as NOE= I_{on}/I_{off} where I_{on} and I_{off} are the peak heights with and without ^{1}H presaturation, respectively.

T_1, T_2 and NOE data were fitted using the "Quadric_diffusion" program, developed in the laboratory of Arthur G. Palmer (Columbia University), in order to obtain the rotational diffusion tensor. Model-free analysis of the relaxation data was performed using the program Tensor2 (Dosset *et al*., 2000). For this analysis a subset of relaxation data was selected using the following criteria. (a) Residues possessing considerable internal motion with heteronuclear NOE values smaller than 0.65 were discarded. (b) residues involved in conformational exchange identified using the following criterion:

[(<T_2> - $T_{2,n}$)/<T_2> - (<T1> - $T_{1,n}$)/<T_1>] > 1.5 x SD, where $T_{2,n}$ is the T_2 value of residue *n* and <T_2> is the average T_2 and SD is the standard deviation of the above formula, were also discarded (Tjandra *et al*., 1995).

3.12 KlcA fold recognition and comparison

The *B. pertussis* KlcA sequence was subjected to a BLAST search (Altschul *et al*., 1990; Altschul *et al*., 1997) against the Protein Data Bank (PDB) (Berman *et al*., 2002) and to fold recognition servers 3D-PSSM (Kelley *et al*., 1996); PHYRE (Bennett-Lovsey *et al*., 2008) and the BioInfo server (Ginalski *et al*., 2003) to check if there was any clear or remote homology to an existing structure or fold. Additionally, the solved KlcA structural fold was compared against fold libraries, SCOP (Murzin *et al*., 1995) and CATH (Orengo *et al*., 1997; Greene *et al*., 2007) with the following methods: (i) DALI version 3.0 (Holm & Sander 1993; Dietmann *et al*., 2001); (ii) CATHEDRAL (Redfern *et al*., 2007); and (iii) SSM (Krissinel &Henrick, 2004).

3.13 KlcA conservation and surface analysis

Related sequences to the *B. pertussis* KlcA sequence (UniProt Accession no: Q08L07) were obtained using a BLAST (blastp) search (Altschul *et al*., 1990; Altschul *et al*., 1997) against the UniProt sequence database (Bairoch *et al*., 2005), *via* the NPS@ server (Combet *et al*., 2000). A total of 182 sequences with an E-value of less than 10^{-6} were obtained and this list was made non-redundant initially by removing 62 identical sequences. The remaining 120 KlcA-like

sequences were retrieved and a multiple sequence alignment was generated using the program PROMALS3D (Pei *et al*., 2008a; Pei *et al*., 2008b). This set of sequences was rendered further non-redundant by removing all 90% pairwise identical sequences in the alignment using the Expasy server (http://www.expasy.ch/tools/redundancy/). A total of 44 non-redundant divergent sequences were thus obtained. Upon analysis, one sequence that lacked the C-terminal helix 5 was discarded and three other sequences, which possessed N-terminal extensions, were trimmed. The remaining 43-aligned sequences were used as input for the Consurf server (http://consurf.tau.ac.il/) (Glaser *et al*., 2003; Landau *et al*., 2005), to map sequence conservation on the surface of the protein. The results were viewed under PyMol version 1.0 (http://www.pymol.org).

The electrostatic surface potential representation was calculated using GRASP (Nicholls *et al*., 1991). The lipophilic surface rendition was generated using MOLCAD under SYBYL version 6.9 (Tripos Associates, St. Louis, MO, USA). Solvent-accessible calculations for each residue were performed using 'NACCESS' computer program (*Department of Biochemistry and Molecular Biology, University College London*).

3.14 Purification of DNA fragments from PCR reaction

The purification of PCR products used for cloning purposes was performed using the QIAquick PCR Purification Kit Protocol (Qiagen). Five volumes of Buffer PB (Qiagen supplied) were added to one volume of the PCR sample. The

sample was applied to a QIAquick column and centrifuged for 30–60 s at 13,000 rpm. The flow-through was discarded and 0.75 ml of the washing Buffer PE (Qiagen supplied) was applied to the QIAquick column and centrifuged for 30–60 s. The flow-through was discarded and the column was centrifuged for an additional 1 min to completely remove residual ethanol. The column was placed in a clean microcentrifuge tube, 30 μl of buffer EB (10 mM Tris·Cl, pH 8.5) was applied at the center of the column membrane, the column was left to stand for 1 min and the final purified sample was collected after centrifugation of the column for 1 min.

3.15 DNA extraction from agarose gels

The extraction and purification of DNA from agarose gels was performed using the QIAquick Gel Extraction Kit Protocol (Qiagen). The gel slice containing the DNA of interest was excised with a scalpel from the agarose gel. The slice was weighed and 3 volumes of buffer QG (Qiagen supplied) were added to 1 volume of gel (100 mg ~100 μl). The slice was incubated for 10 min (or until the agarose had completely dissolved depending on the size of the gel slice) at 50°C with occasional vortexing every 2-3 min. After the agarose had dissolved completely, 1 gel volume of isopropanol was added, mixed, applied the sample to a QIAquick column to bind the DNA and centrifuged for 1 min. at 13,000 rpm. The flow-through was discarded after the run had finished, 0.5 ml of Buffer QG was added to the column and centrifuge for 1 min. in order to remove all traces of agarose. For the washing step, 0.75 ml of Buffer PE (Qiagen supplied) was applied to the column and centrifuged for 1 min. The flow-

through was discarded and the column was centrifuged for an additional 1 min. The column was placed in a microcentrifuge tube, 30 µl of buffer EB (10 mM Tris·Cl, pH 8.5) was applied at the center of the column membrane, the column was left to stand for 1 min and the final purified sample was collected after centrifugation of the column for 1 min.

3.16 SDS-PAGE

For the isolation of proteins in denaturing conditions, Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) was used. Protein samples were mixed with 6x SDS-loading buffer and loaded on pre-cast 4-12% NuPAGE Novex gels (Invitrogen). An X-cell Sure Lock (Invitrogen) apparatus was filled with 1x NuPAGE buffer (Invitrogen provided) and the gels were run at a constant 150 V for 45 min.

3.17 BN-PAGE

For the isolation of proteins in their native state and protein complexes, Blue-Native Polyacrylamide Gel Electrophoresis (BN-PAGE) was used. Protein samples were mixed with NativePage sample buffer (Invitrogen provided) and loaded on pre-cast 4-16% NativePAGE Novex Bis-Tris gels (Invitrogen). The upper buffer chamber was filled with 200 ml 1x NativePage Cathode buffer (Invitrogen) and the lower buffer chamber with 550 ml 1x NativePage Anode buffer (Invitrogen). 8 µl NativeMark (Invitrogen) marker was used as a size

ladder. The electrophoretic apparatus was an X-cell Sure Lock (Invitrogen). A running time of 120 minutes was used with a constant 150 V.

3.18 DNA electrophoresis

Agarose gels for electrophoretic analysis of DNA were prepared with 1x TAE buffer and 0.8 or 1.2% agarose depending on the size of DNA that was to be analysed. The mixture was heated until all the agarose was dissolved, left to cool and ethidium bromide was added (0.5 μg/ml). The mixture was poured in an electrophoretic apparatus, a comb was inserted to create the wells and the gel was left to solidify. After the gel was solid, the comb was removed and the samples containing DNA mixed with loading buffer were loaded in the wells after the immersion of the gel in 1 x TAE buffer. The gel was run at 120 Volts for 1-2 hrs, and the DNA was visualised using a UV transilluminator.

3.19 Cloning Procedure

Orf18 of Tn*916* was amplified by the polymerase chain reaction (PCR) from plasmid pAM120 using the following pair of primers: forward, 5′- AGT CGTCCATGGACGATATGCAAGTCTATATTG- 3′; reverse, 5′- GACACT AAGCTTTTAATAGACGATTTCAAAAATCCC-3′. To facilitate cloning, the forward and reverse primers included NcoI and HindIII recognition sites, respectively. PCR was performed using 1.5 U of Taq DNA polymerase, 10 mM Tris-HCl (pH 9.0 at room temperature), 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM

each deoxynucleoside triphosphate, 40 pmol of each primer and ~50 ng of plasmid DNA. The PCR consisted of the following steps: denaturation for 1 min followed by 30 cycles of denaturation (95ºC for 30 sec), primer annealing (45 ^{0}C for 30 sec) and extension (72 ^{0}C for 4 min). An amplified product of the anticipated size was obtained, digested with NcoI and HindIII and then ligated into the corresponding sites on pTrc99A using the Quick T4 DNA Ligation kit (New England Biolabs) to generate the expression construct pDMArdA. Plasmid DNA was prepared and nucleotide sequencing of the insert DNA was performed to ensure no mistakes had been introduced during amplification.

The same cloning procedure was followed for all the constructs described in this work. The truncated *orf18* was ligated into pET15-b. The primers used were: forward, 5'- CGG TGT CCA TAT GGT ATC GGA ATT AC -3'; reverse, 5'- CAG CTA GGA TCC CTA TTA ATA GAC GAT TTC-3'. *klcA* was synthesised by GeneArt (Germany) using the deposited DNA sequence (Gene ID: 4364084) as a starting point for codon optimisation and synthesis. The synthetic gene was inserted into vector pET24-a. The N-terminus- and C-terminus -myc-tagged *klcA* were ligated into pET24-a. For the N-terminus-myc-tagged *klcA* the primers were: forward, 5'- ATC AGT CCA TAT GGA ACA AAA ACT TAT TAG TGA AGA AGA TCT TAT GAA CAC CGA AGA AC-3'; reverse, 5'-CCG TAC TAA GCT TTC ATC AAT CAA TCG CAC GAT A -3'. For the C-terminus-myc-tagged KlcA : forward, 5'- GAT CAG TCC ATA TGA ACA CCG AAG AAC AG-3'; reverse, 5'- GTA CTA AGC TTT CAA AGA TCT TCT TCA CTA ATA AGT TTT TGT TCA TCA ATC GCA CGA TAA -3'.

3.20 Plasmid transformation

Prior to transformation, bacterial cells were made competent by using the Ca^{2+}-treatment (Sambrook, 2002). Overnight cultures (10 ml) of bacteria incubated at 37°C were recovered by centrifugation at 3500 rpm for 10 min. The medium was decanted and the cell pellets were resuspended by gentle mixing in 20 ml ice cold 80 mM $MgCl_2$ - 20 mM $CaCl_2$ solution. The cells were centrifuged once more, the supernatant was discarded and the pellets were resuspended in 2 ml ice-cold 0.1M $CaCl_2$. For the transformation, 100 µl of competent cells were mixed with 1 µl of plasmid DNA and left on ice for 30 min. The cells were subsequently heat-shocked for 35 sec at 42 °C and immediately put on ice for 2 min. 900 µl of warm LB broth or SOC medium (2% w/v bacto-tryptone, 0.5% w/v bacto-yeast extract, 10mM NaCl, 2.5mM, KCl, 10mM $MgCl_2$, 20mM glucose, pH 7.0) was added to the cells and left to incubate at 37 °C for 1 hr. in order for expression of antibiotic resistance genes to take place. Cells were plated on selective media and incubated overnight at 37 °C to obtain transformants.

3.21 Plasmid DNA isolation

All plasmid isolations were performed with the help of the Qiaprep kit from Qiagen. This protocol (for small volume plasmid extraction) is designed for purification of up to 20 µg of high-copy plasmid DNA from 1–5 ml overnight cultures of *E. coli* in LB medium. The procedure is based on alkaline lysis of bacterial cells followed by adsorption of DNA onto silica in the presence of high

salt. Pelleted bacterial cells were resuspended in 250 μl Buffer P1 (50 mM Tri-HCl, pH 8.0, 10 mM EDTA, 100μg/ml RNaseA) and transfered to a microcentrifuge tube. 250 μl Buffer P2 (200 mM NaOH, 1% SDS) were added and mixed thoroughly by inverting the tube 4–6 times. 350 μl Buffer N3 (4.2 M Gu-HCl, 0.9 M potassium acetate, pH 4.8) were subsequently added and mixed immediately and thoroughly by inverting the tube 4–6 times. The solution was centrifuged for 10 min at 13,000 rpm (~17,900 x g) in a table-top microcentrifuge. The supernatants from the previous step were applied to the QIAprep spin column, centrifuged for 30–60 s and discarded the flow-through. The QIAprep spin column was washed by adding 0.5 ml Buffer PB (5 M Gu-HCl, 30 % ethanol) and centrifuging for 30–60 s. The flow-through was discarded and the QIAprep spin column was washed by adding 0.75 ml Buffer PE (10 mM Tris-HCl, pH 7.5, 80 % ethanol) and centrifuging for 30–60 s. The flow-through was discarded, and the column was centrifuged for an additional 1 min to remove residual wash buffer. To elute DNA, the QIAprep column was placed in a clean 1.5 ml microcentrifuge tube and 50 μl Buffer EB (10 mM Tris·Cl, pH 8.5) were added to the center of each QIAprep spin column which was let to stand for 1 min. and finally and the column was sucentrifuged for 1 min and the flow-through was collected.

3.22 Western Blot

A precast 4-15 % SDS-PAGE gel (Invitrogen) was run and following electrophoresis the bottom 'lip' and 'top comb' of the gel was cut off. The gel was soaked twice for 15 min in 100 ml of chilled Towbin buffer (per litre: 3.03 gr Tris base, 14.4 gr glycine, 100 ml methanol) to remove residual SDS. The

nitrocellulose membrane was soaked for 30 min in 50 ml of chilled Towbin buffer. Two pieces of extra thick blotting paper and both fabric sponges were soaked for 5 min in 50 ml Towbin buffer. The gel sandwich was assembled under a good depth of Towbin buffer. The sandwich holder was placed clear side down. One of the fabric sponges was rolled up under the buffer level and squeezed to remove any residual air bubbles, the sponge was then laid on the clear side of the holder. Next one piece of blotting paper was placed on top of the sponge, followed by the membrane and the gel. Finally, the last piece of blotting paper followed by the last sponge was placed on top. The holder was then closed and placed into the mini-protean blotting cell (Invitrogen). Transfer was achieved by electrophoresis at a constant voltage of 150V for 1 hour and 30 min. The transfer was confirmed by observing the transfer of prestained markers. Following transfer, the membrane was blocked for 1 hour in 1xPBS + 5% skimmed milk powder (Sigma). The membrane was subsequently washed three times in 1xPBS + 0.05% Tween 20 for 5 min each time. The membrane was then washed for a further 20 min in just 1xPBS. The antibody (monoclonal anti-polyhistidine-peroxidase conjugate clone His-1) (Sigma) was prepared according to the manufacturer's description, and used 50 µl of the antibody per 50 ml of 1xPBS + 5% skimmed milk powder and left to react overnight. The next day, the membrane was washed for 1 hour with 3 changes using 1xPBS + 0.05% Tween 20 before rinsing for 20 min with 1xPBS. The membrane was stained using 1 tablet of 3,3'- diaminobenzidine (DAB) (Sigma) and one tablet of urea (Sigma) dissolved in 5 ml dH_2O. Bands appeared within two min. The reaction was stopped by pouring away the development solution and rinsing the membrane with dH_2O.

3.23 In-gel protein digestion with trypsin

Bands of interest from an SDS-PAGE gel were cut tightly with a scalpel. The gel slices were incubated in 300 µl 200 mM NH_4HCO_3 in 50 % acetonitrile (ACN) at room T for 30 min. This step was repeated twice to remove residual SDS. The slices were incubated in 300 µl 20 mM DTT, 200 mM NH_4HCO_3, 50% ACN at room T for 1 h to reduce the protein. The slices were then washed in 300 µl 200 mM NH_4HCO_3 in 50 % ACN three times. The cysteines were alkylated in fresh 100 µl 50 mM iodoacetamide (IAA), 200 mM NH_4HCO_3, 50% ACN at room T in the dark for 20 min. The slices were washed with 500 µl 20 mM NH_4HCO_3, 50% ACN and spinned at 13,000 rpm for 2 min. The slices were then covered with ACN for 5 min. The slices after this step turned white. The ACN was decanted and allowed the slices to dry. A stock solution of trypsin was prepared by adding 50 µl 50 mM NH_4HCO_3 to a new vial of trypsin (Promega). The gel slices were swelled in 29 µl 50 mM NH_4HCO_3 containing 1 µl of trypsin solution prepared previously. During the swelling step, the slices were kept at 4 °C to avoid autodigestion of trypsin. The tops of the tubes containing the gel slices were sealed with Nescofilm and incubated at 32°C overnight. A trypsin blank was also prepared and incubated to be used later as a control. The next day, the samples were sonicated for 10 min to facilitate liberation of the peptides still trapped in the gel. The digests were stored at -80°C.

3.24 MALDI-TOF-MS

A matrix containing a-cyano-4-hydroxycinnamic acid (CHCA) was prepared as follows: weighed 10 mg of CHCA (Sigma), added 400 µl of water, 100 µl of 3%

TFA and 500 µl and mixed. The solution was centrifuged at 5,000 rpm for 1 min. The supernatant was used for spotting along with the sample on the MALDI plate. 0.5 µl of sample was first applied on the wells of the metallic plate and 0.5 µl of matrix were subsequently added. The plate was allowed to dry at room temperature and then loaded on an AppliedBiosystems Voyager DE-STR instrument operating in reflectron mode. The system utilizes a pulsed nitrogen laser, emitting at 337 nm. For each mass spectrum, 200 single laser shots were averaged. The extraction voltage was 20 kV and the low mass gate was set to 500 Da to prevent the saturation of the detector by ions resulting from the matrix. The spectra acquired were calibrated with a trypsin standard (842.5099 Da), baseline corrected and noise filtered and the peak lists were transferred to the Mascot server (Perkins *et al*., 1999) for peptide mass fingerprinting analysis.

3.25 Co-Immunoprecipitation assay

Competent BL21(DE3) cells were transformed with the pET24-a construct harbouring the N-terminus tagged *klcA* and a 500 lt culture in LB + kanamycin (25 µg/ ml) was grown at 37^0C until OD_{600} reached 0.5 - 0.7. The gene was expressed with 1 mM IPTG for 3 hours at 30 ^{0}C. A 3 gr cell pellet was subsequently dissolved in 1 ml of BupH TBS buffer (Pierce provided). The cells were sonicated for 5 min and centrifuged for 30 min at 13,000 rpm. 300 µl of the supernatant were mixed with 10 µl of mouse IgG agarose (Sigma) for the preclearing step and left at 4 ^{0}C for 1 hr. The precleared lysate was subsequently loaded on a Handee Spin Column (Pierce). The same procedure was followed for the controls: uninduced N-terminus tagged *klcA*, pET24-a vector without

insert, and an *E.coli* extract containing c-myc-tagged glutathionine S-transferease (GST) as a positive control. For the GST control, 50 µl of the lysate were mixed with 150 µl BupH TBS buffer. After all the lysates had been applied to the Handee Spin Columns, 10 µl of anti-c-myc agarose slurry was added to each column. The columns were incubated overnight at 4 ^{0}C with gentle end-over-end mixing. Next day the columns were pulse centrifuged for 10 sec. and washed with 0.5 ml of TBS containing 0.05% Tween-20 and centrifuged again. The washing step was repeated for a total of 5 times. For the elution step, 10 µl of Elution Buffer pH 2.8 (Pierce provided) was loaded on the columns and pulse centrifuged for 10 sec. the elution step was repeated for a total of three times. The eluent was neutralised immediately by adding 1 µl of 1M Tris, pH 9.5 per 20 µl of Elution Buffer. For reducing gel analysis, 5 µl of 2-mercaptoethanol were added to 40 µl of 5x Lane Marker Non-Reducing Sample Buffer (Pierce). 7.5 µl of the above described reducing sample buffer were added to 30 µl elution sample. The mixture was heated at 95 ^{0}C for 5 min. and loaded on a SDS-PAGE for analysis.

3.26 His-tagged protein purification

The His-tagged truncated Orf18 was purified with the use of a HisTrap FF crude column (GE Healthcare). The truncated Orf18 was prepared and induced as the wild type Orf18 (see above). The cells were harvested by centrifugation at 3700 rpm for 15 min at 4 ^{0}C. The supernatant was discarded and 5-10 ml of binding buffer were added for each gram of cell paste. The solution was mixed thoroughly and sonicated on ice for 5 min with it was no more viscous. The

stopper of the column was removed and the column was connected to the syringe with the provided adapter "drop-to-drop" to avoid introducing air into the column. The snap-off end was removed and the column was washed with 5 ml of dH_2O. Using a syringe, the column was equilibrated with 10 mls of binding buffer. The sample was loaded on the column as crude lysate. The flow-through fraction was collected for testing later on. The column was washed with 10 ml binding buffer and collected the wash fraction. For the elution step, 5 mls 1x phosphate buffer containing 40 mM imidazole were loaded on the column. The eluent was collected in 1 ml fractions. The elution proceeded with the next imidazole concentration (60 mM) and the same procedure was repeated for the rest of the imidazole concentrations (100, 300 and 500 mM). The different fractions collected were tested with SDS-PAGE.

Chapter 4: ArdA Results and Discussion

4.1 Cloning, overexpression and purification of *orf18*

The gene *orf18* from Tn*916* was inserted into the cloning vector pTrc99A. The pTrc99A expression vector is a derivative of the pKK233-2 expression vector with a strong promoter (*trc*) upstream of the multiple cloning site (MCS) and a strong transcription termination signal (*rrn*B) downstream. The presence of the *laq* I^q gene on the vector allows pTrc99A to be used in *E. coli* hosts lacking the lactose repressor gene (Amann *et al.*, 1988).

In order to test whether insertion of the *orf18* into the expression vector was successful, a PCR reaction was performed containing the *orf18* primers and the plasmid DNA as template. Thus, only plasmids containing the *orf18* insert would yield a PCR product. The expected band size of the *orf18* insert was expected to be 468 bp. A 0.8% agarose gel with the PCR reaction is shown below (Fig. 28). The construct was named pDMArdA.

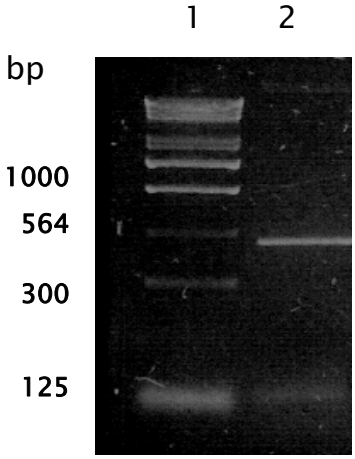


Fig. 28 *Confirmation of recombinant plasmid Lane 1: a λDNA/HindIII marker. Lane 2: the PCR product of pTrc99A plasmid having acquired the orf18 insert.*

Colonies from transformed bacteria were inoculated into LB-carbenicillin medium, and induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) in a fermenter once the cell density reached $O.D_{600nm}$ of 0.5. IPTG induces activity of beta-galactosidase, an enzyme that promotes lactose utilization, by binding and inhibiting the *lac* repressor. In cloning experiments, the *lacZ* gene is replaced with the gene of interest and IPTG is then used to induce gene expression (Brown, 2001). The expression strain BL21 DE3 was used for overexpression purposes as it is deficient in the *lon* and the *ompT* proteases that can degrade proteins during purification (Terpe, 2006).

The overexpressed ArdA eluted from the diethylaminoethyl (DEAE) ion exchange column towards the end of the gradient as a broad peak in ~0.7 M NaCl with a purity of almost 95% (Fig. 29).

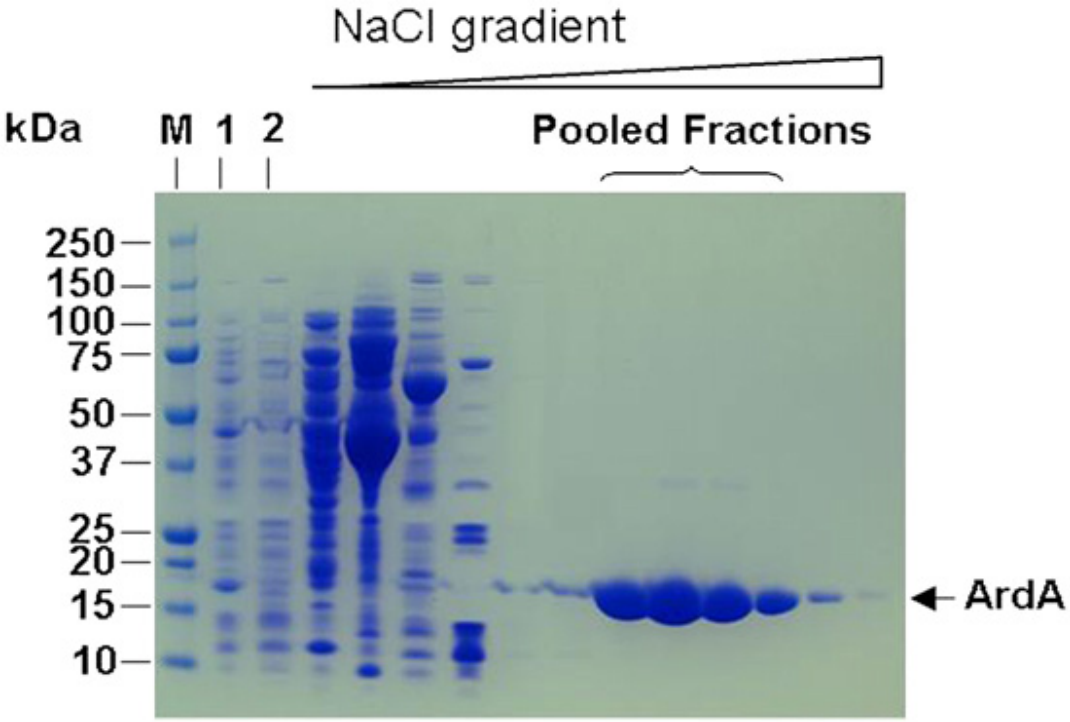


Fig. 29 *SDS-PAGE analysis of fractions obtained from anion exchange (DEAE) chromatography during purification of recombinant Orf18 ArdA. Lane 1, clarified cell-free extract of induced E. coli BL21(DE3) pDMArdA cells as applied to the column; lane 2, flow-through fraction (i.e., unbound material). The next lanes (left to right) represent fractions of eluted protein corresponding to the major peaks obtained from an increasing salt gradient (0–1 M NaCl). Highlighted fractions containing ArdA were pooled and then further purified on a Superdex-200 column. Molecular mass markers are shown in lane M.*

The acidic nature of the ArdA protein, with a predicted pI of 3.91, may account for the observed tight binding to the column matrix. Fractions containing ArdA were subsequently pooled, concentrated and then further purified by gel filtration chromatography using a Superdex-200 column to remove minor protein contaminants. Thus, the surprisingly good resolution obtained from the DEAE column facilitated the purification of ArdA to homogeneity in only two steps. Typically, an overall yield of ~8 mg of purified ArdA per gram (wet weight) of cell paste was obtained.

The identity of the purified protein was confirmed by N-terminal sequencing (courtesy of Dr. Andy Cronshaw, Edinburgh University). Ten unambiguously assigned residues (MDDMQVYIAN) were determined, which matched the anticipated sequence from the sequence database. Furthermore, the accurate mass of the purified protein was obtained by electrospray mass spectrometry. The mass spectrum was obtained on a 12T FT-ICR spectrometer (courtesy of Dr. Logan Mackay, Edinburgh University). The experimentally determined mass

(19,153 Da) was 28 Da higher than that predicted (19,125 Da) from the primary amino acid sequence. This difference was presumably due to the presence of a formyl group on the N-terminal methionine of the protein that had not undergone complete post-translational removal. In bacteria, translation initiates with formyl-methionine; however, the N-terminal formyl group is usually later removed by the enzyme deformylase (Berg *et al.*, 2001). A minor species (12%) corresponding to the exact predicted mass of the protein was also detected, indicating that some deformylation had occurred.

4.2 Secondary structure analysis of ArdA

The far-UV circular dichroism (CD) spectrum of ArdA (Fig. 30) indicated that its secondary structure consisted primarily of α-helices, as determined by the two characteristic minima at 208 nm (part of the π-to-π^* transition) and 222 nm (the n-to-π^* transition) in the peptide bond.

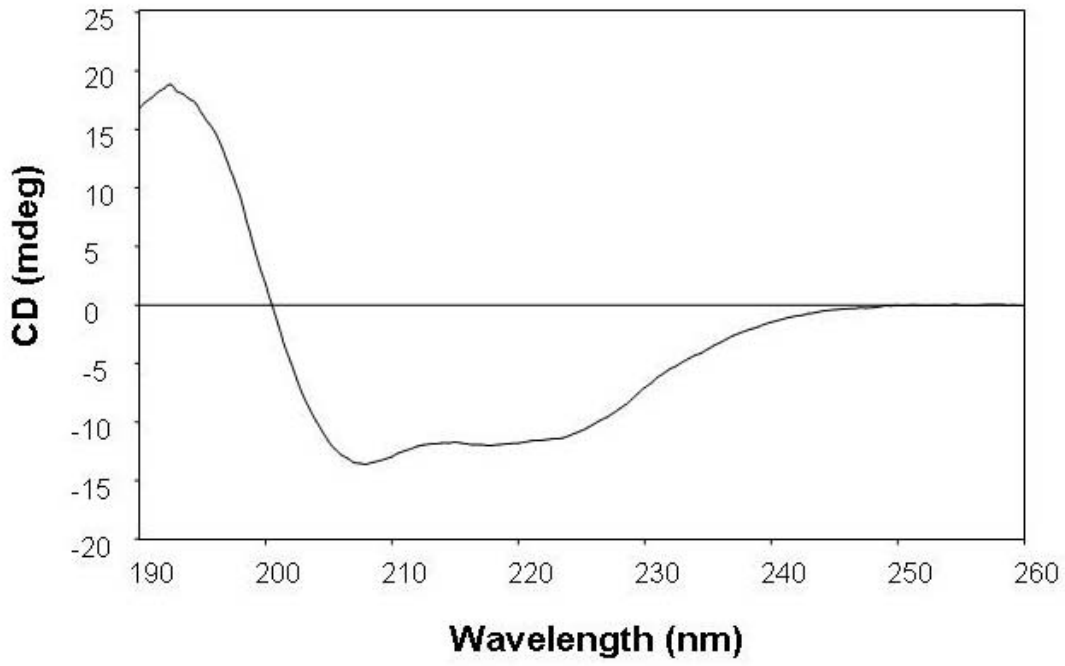


Fig. 30 *Far-UV CD spectrum of 30 μM ArdA.*

The DICHROWEB server (Whitmore *et al.*, 2004) was used to determine the percentage of secondary structure in the protein employing the K2D (Andrade *et al.*, 1993) neural network algorithm with built-in protein reference data. The K2D analysis gave a prediction of 35% α-helix, 15% β-sheet and 49% random coil.

CD spectra in the near-UV region (250–350 nm) reflect the asymmetry of the environment of aromatic amino acid residues. A profound negative band with minima at 278 and 285 nm (Fig. 31) was an indication of aromatic side chains (tyrosine and tryptophan) located in a folded tertiary structure.

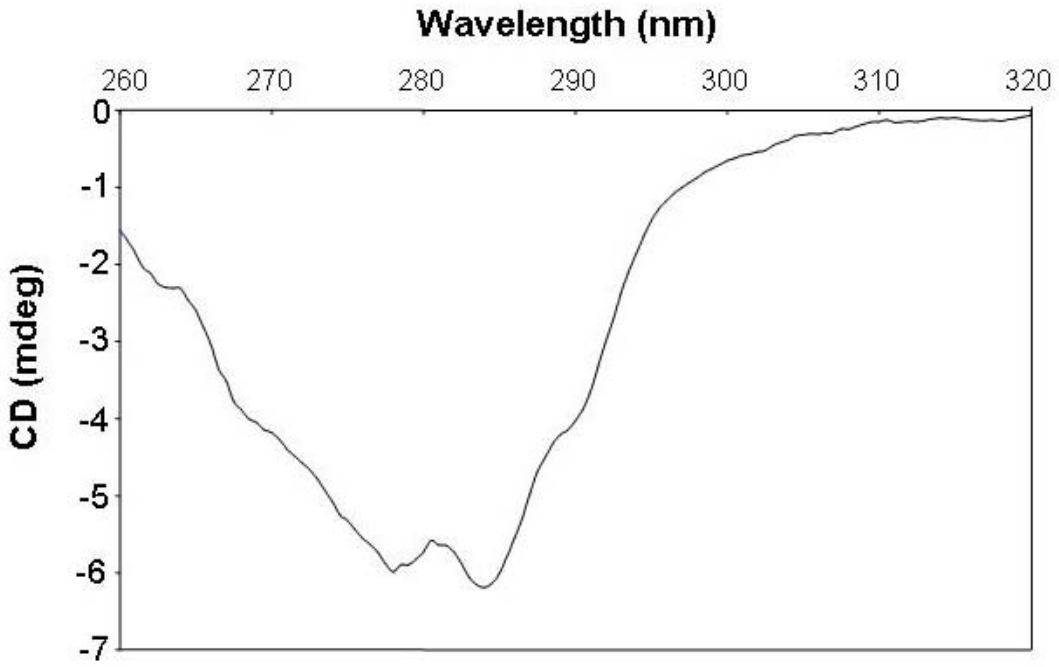


Fig. 31 *Near-UV CD spectrum of 130 μM ArdA*

Analysis of the secondary structure of ArdA using the PSIPRED prediction server (McGuffin *et al.*, 2000) was largely in agreement with the CD experiments in terms of the β-sheet component (50% α-helix, 15% β-sheet), with helices predicted to occur at amino acids 3–14, 30–37, 63–75, 78–90, 94–100, 113–122, 126–127, 131–137 and 139– 148. Strands were predicted at amino acids 18–25, 46–51, 104–106, 151–155 and 158–162.

4.3 Size-exclusion HPLC

Size-exclusion HPLC was used to determine the quaternary structure of ArdA over a range of monomer concentrations. Size-exclusion chromatography (or gel filtration) separates molecules according to their size. The stationary phase consists of porous beads with a well-defined range of pore size. Proteins that are small enough can fit inside all the pores in the beads and are said to be included. These small proteins have access to the mobile phase inside the beads as well as the mobile phase between beads and elute last in a gel filtration separation. Proteins that are too large to fit inside any of the pores are said to be excluded. They have access only to the mobile phase between the beads and, therefore, elute first.

A gel filtration column was first equilibrated with protein standards and the resulting calibration curve is shown in Fig. 32. The ArdA was tested at the following concentrations: 50 nM, 200 nM, 500 nM, 1 μM, 10 μM, 50 μM, and 100 μM. The elution time of each sample was multiplied with the flow rate (0.5 ml/min) to obtain the elution volume. This elution volume was used as the "y" parameter in the equation of the calibration curve. The expected molecular mass of the ArdA as a monomer (19,125 Da) was divided from the antilog of y to obtain the molecular weight of each sample.

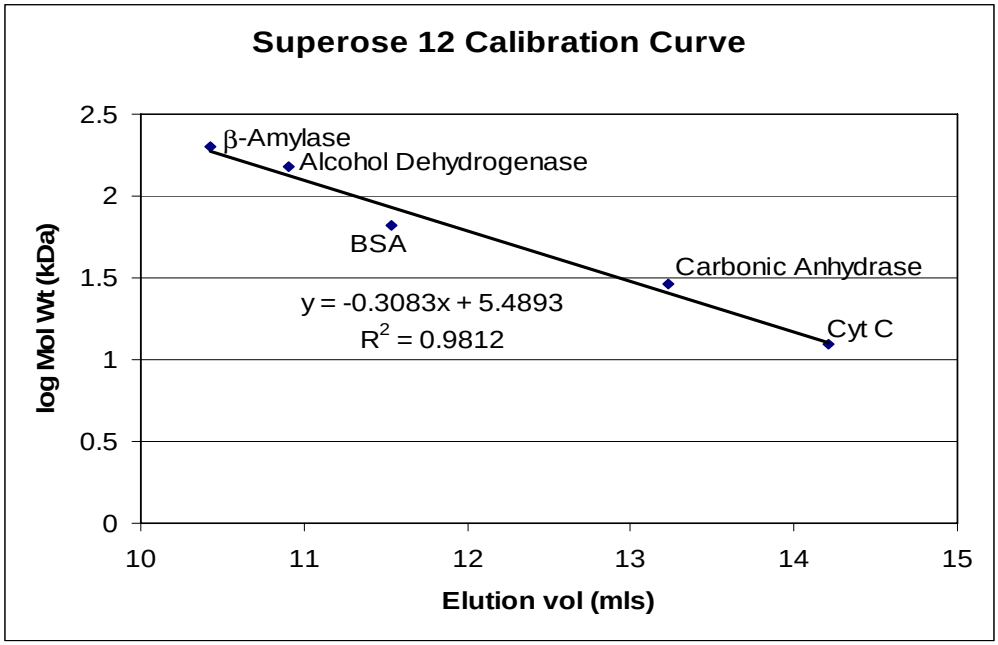


Fig. 32. *Calibration curve of protein standards. A Superose-12 gel filtration column was calibrated with protein standards. The equation derived from plotting the elution volume vs. log molecular weight of standards was used to calculate the molecular weight of ArdA.*

ArdA appeared to exist as a dimer at lower concentrations (50 and 200 nM) but formed larger species, perhaps tetramers and hexamers, at higher concentrations (Fig. 33). The fact that ArdA existed as a dimer at low concentrations agrees with findings for another ArdA from plasmid ColIb-P9 (Nekrasov *et al*., 2007). The data were fitted with a one-binding site equation (Klotz, 1997) and the dissociation constant Kd was found to be 1.4 μM, suggesting a relatively weak interaction for this aggregation.

The width of each elution peak was measured as a single peak can mask overlapping peaks. Theoretically, a peak of the same width should be obtained at all concentrations for a homogeneous protein. In this experiment, a maximum peak width at 1 μM was observed, suggesting that a mixture of species was present in this concentration region.

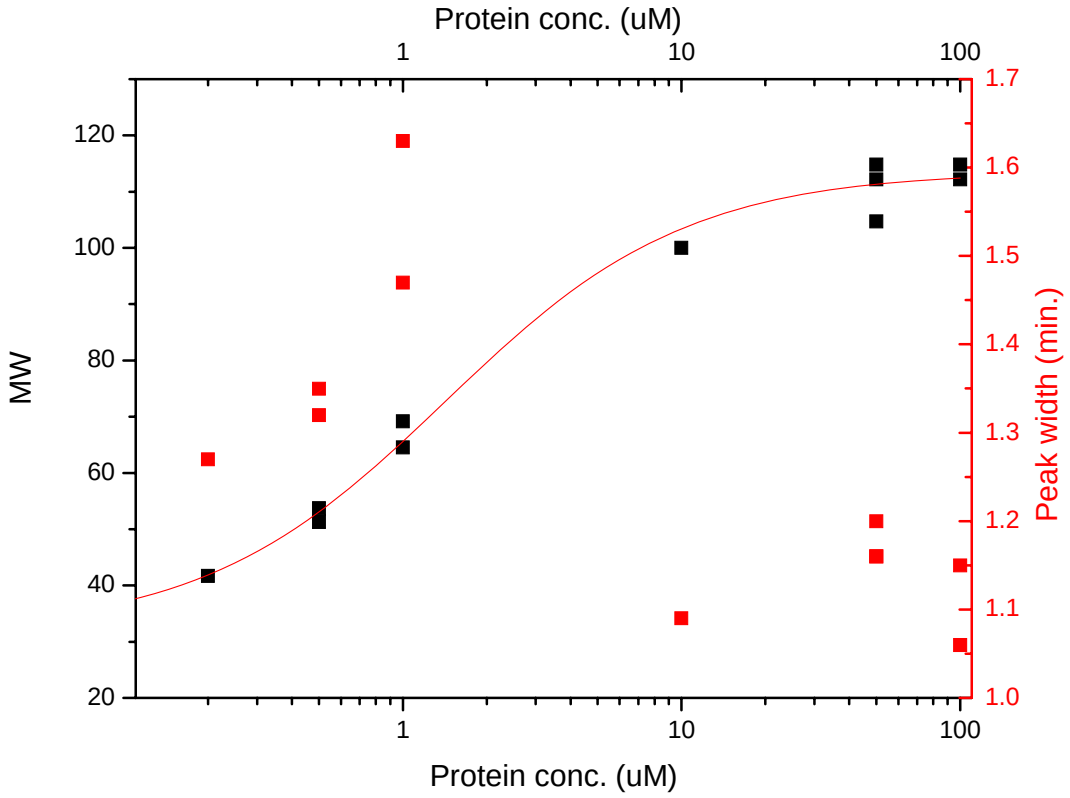


Fig. 33 *Size-exclusion chromatography analysis of ArdA to investigate the solution molecular mass at different protein concentrations. Black squares: molecular mass values of ArdA at different concentrations; red squares: peak width values obtained at different concentrations of ArdA. ArdA forms a dimer at the lower concentration range (50 and 200 nM), a tetramer at 1 μM and a hexamer at 50 and 100 μM. A maximum*

peak width was obtained at 1 μM, suggesting that the protein exists as more than one species at this concentration.

4.4 Fluorescence Quenching Experiments for the Determination of Tryptophan topology

Although fluorescence measurements do not provide detailed structural information, the technique has become quite popular because of its acute sensitivity to changes in the structural and dynamic properties of biomolecules and biomolecular complexes. In a globular protein, the degree to which an amino acid side chain is exposed to the solvent depends on how much it is shielded by the surrounding protein segments. Information regarding the exposure of a residue in a protein can be obtained by crystallographic studies, however, in this way only a static view of the positioning can be obtained. Since proteins in solution are mobile and can change conformation, a simple and effective way to experimentally determine the degree of exposure of tryptophanyl residues in solution is by fluorescence quenching. Acrylamide, oxygen and iodide can act as quenchers. These agents decrease the fluorescence intensity of the trp residues via physical contact with the excited indole ring (Eftink & Ghiron, 1976).

Orf18 ArdA contains two tryptophan residues (W23 and W70). The emission maximum of the ArdA appears to be at 340 nm (Fig. 34). The protein spectrum is slightly blue-shifted from that of aqueous tryptophan (350 nm), suggesting that at least one of the tryptophan residues was partially buried.

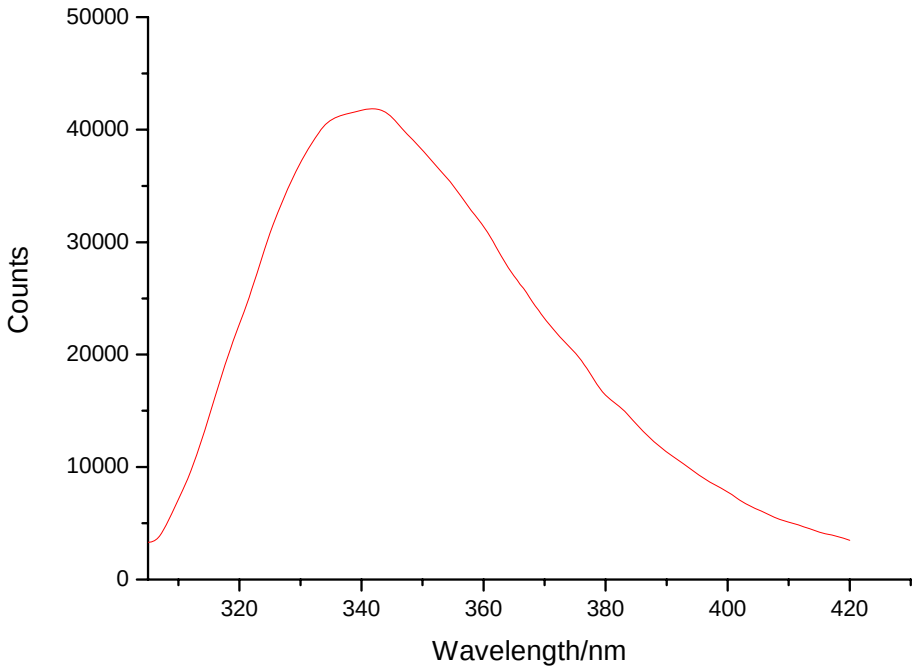


Fig. 34 *Emission maximum of ArdA. Protein sample was excited at 295 nm and the emission spectrum was recorded between 300 and 420 nm. ArdA shows an emission maximum at 340 nm.*

A way to experimentally determine the degree of exposure of tryptophanyl residues in solution is fluorescence quenching by acrylamide, oxygen or iodide via physical contact with the excited indole ring (Eftink &. Ghiron, 1976). Addition of successive aliquots of acrylamide to ArdA caused a gradual reduction in the fluorescence signal intensity. Acrylamide quenching yielded an almost linear response with an effective quenching constant of 3.9 M^{-1} (Fig. 35). The absence of significant curvature implied that the two tryptophan residues differed only slightly in their solvent accessibility (Eftink &. Ghiron,

1981). Quenching was also performed with KI. Although iodide is an effective quencher, it cannot penetrate into the protein structure to quench buried tryptophan residues. The quenching constant for KI quenching was found to be 1.63 M^{-1}. As anticipated, the value of the quenching constant was smaller for iodide quenching than for acrylamide quenching. Interestingly, although ArdA has an overall negative charge at neutral pH (pI=3.91), the results from the iodide quenching experiments appeared to show that the microenvironment surrounding the two tryptophan residues was not highly negatively charged. If it were negatively charged, then this would repel the iodide and no quenching would take place. Indeed, for this reason, the intrinsic fluorescence of the Ocr antirestriction protein could not be quenched by KI (Atanasiu *et al.*, 2001).

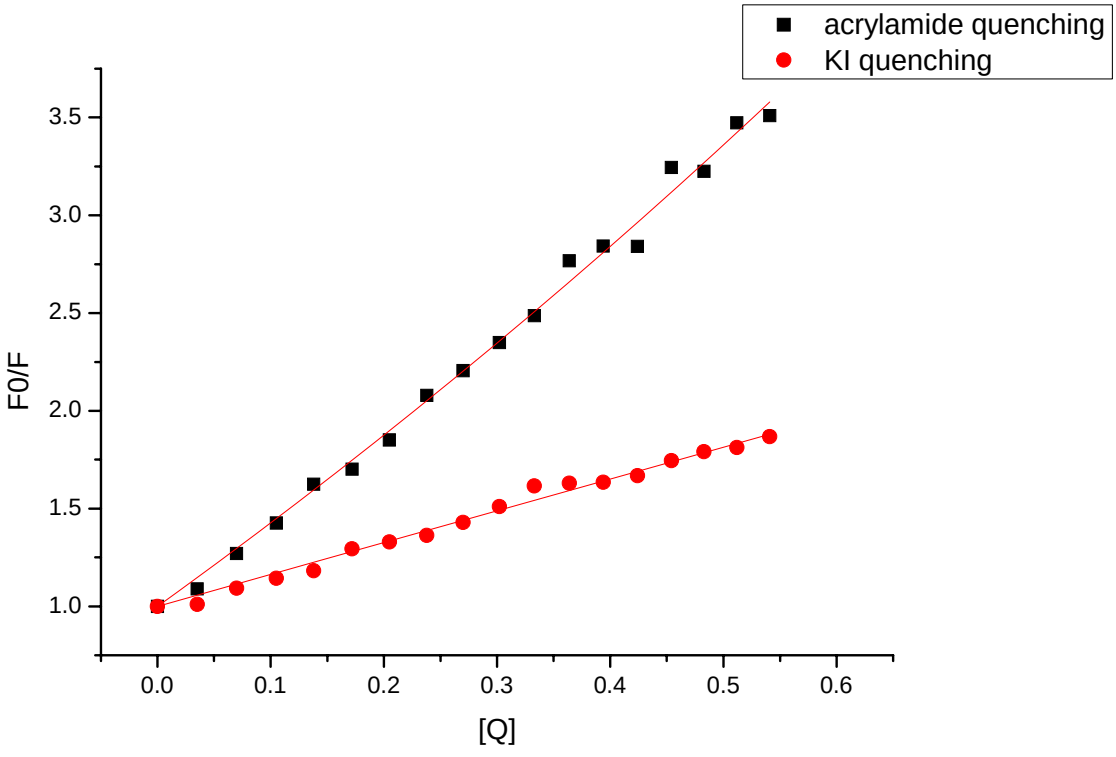


Fig. 35 *Fluorescence quenching of ArdA. 3μM of ArdA were quenched with increasing concentrations of acrylamide (black squares) and iodide (red circles). The data were fitted into a modified Stern-Volmer plot equation.*

4.5 Determination of the free energy ΔG of unfolding

Of special interest in understanding the three dimensional structure of proteins in solution is the detection and identification of stable intermediates between the native and unfolded states by studying the equilibrium denaturation of the protein in the presence of a strong denaturant. Measuring the conformational stability requires determining the equilibrium constant and the free energy ΔG for the reaction:

Folded (F) ⟷ Unfolded (U)

Several chemical agents such as urea and guanidinium chloride are known to induce the unfolding of native proteins when present at high concentrations. Titration of a protein solution with increasing concentration of the denaturant shifts the equilibrium between the native and denatured states of the protein to gradually favour the denatured form. The amount of denaturant needed and the steepness of the titration curve can provide quantitative information on the stability of the protein sample.

Increasing amounts of GdmCl were added to ArdA and fluorescence measurements were performed. The addition of the denaturant causes the protein to unfold. The unfolding causes changes in the quenching of the

tryptophan. The fluorescence intensity was measured as a ratio of the fluorescence intensity at 320nm vs. 370 nm. Upon dilution of the solution, fluorescence recovery was observed following the same transition as the unfolding process, indicating that the refolding of ArdA was fully reversible. The data revealed a two-state transition (Fig. 36) with no intermediate step and were fitted to a two-state model equation (Pace *et al.*, 1989). The midpoint of unfolding occurred at 3 M GdmCl, and the free energy of unfolding, ΔG, was found to be 31.5 kJ/mol. This is less than that found for the Ocr protein, with a ΔG of 64.2 kJ/mol (Atanasiu *et al.*, 2001) indicating that ArdA is less stable.

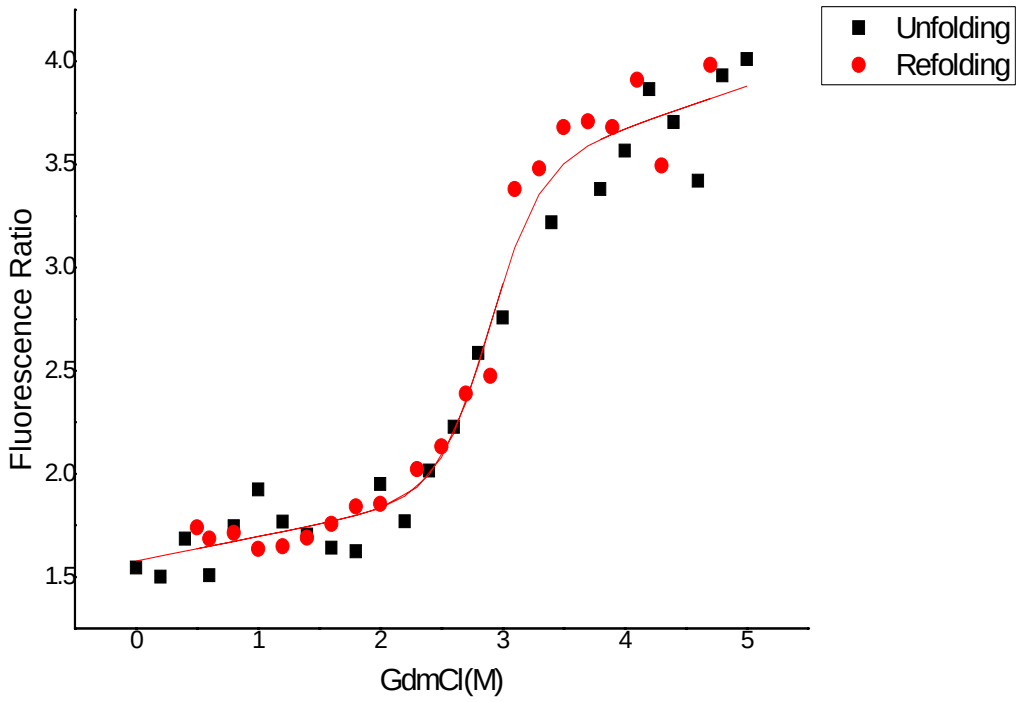


Fig. 36 *Unfolding curve of ArdA with increasing concentration of GdmCl. The fluorescence intensity was measured as a ratio of the fluorescence intensity at 320nm vs. 370 nm. The process of unfolding appeared to be totally reversible (refolding).*

4.6 Differential scanning calorimetry (DSC)

DSC was used to study the thermal stability of 50 μM ArdA. The data show a significant heat uptake consistent with a cooperative endothermic unfolding transition with midpoint temperature T_m= 46.9 °C ±1.9 °C (Fig. 37) Analysis of the DSC endotherm yielded two independent estimates of the unfolding enthalpy of the protein under these conditions. The estimate for the calorimetric enthalpy (ΔH_{cal}=23± 13 kcal/mol) was significantly lower than that for the van't Hoff enthalpy (ΔH_{VH} =110 ± 77 kcal/mol), indicating the presence of more than one species within the solution. These findings were in good agreement with the results from size-exclusion chromatography. A rescan showed that the unfolding of the protein was fully reversible.

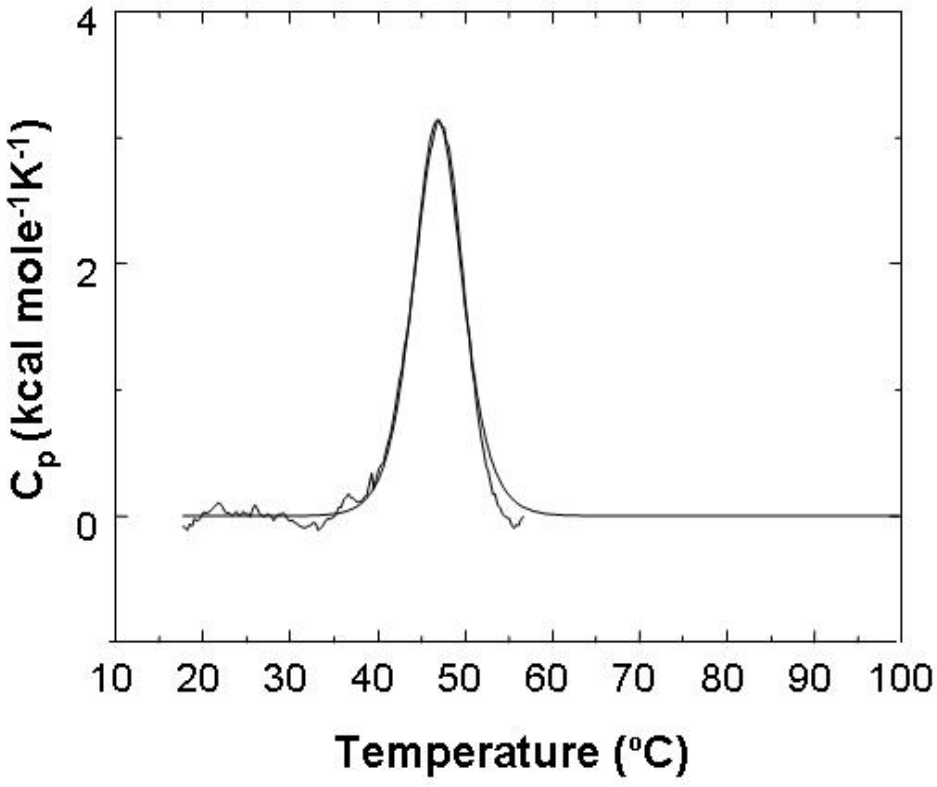


Fig. 37 *DSC of 50 μM ArdA. ArdA has a T_m = 46.9 ±1.9ºC. Fitting of the partial heat capacity (C_p) curves was carried out using a two-state unfolding model.*

4.7 ArdA–M.EcoKI interaction by ITC

Isothermal titration calorimetry (ITC) was used to study the binding of ArdA to M.EcoKI. A typical titration of M.EcoKI with ArdA is shown in Fig. 38. The interaction of M.EcoKI with ArdA is exothermic, with a clear end point corresponding to the concentration of ArdA at which M.EcoKI becomes saturated. After integration of each injection peak and subtraction of the heats of

dilution, the values of the heat of interaction normalised for the ArdA monomer concentration were plotted versus the molar ratio of ArdA to M.EcoKI. The data were fitted to a theoretical titration curve (single-site model) to determine the enthalpy (ΔH) and stoichiometry (n) of interaction. Using this procedure, the ΔH value was calculated to be −17.0 ± 0.2 kcal/mol. The binding between ArdA and M.EcoKI was clearly strong (K_d <1 nM), with a sharp transition that unfortunately precluded reliable determination of the binding affinity using ITC as observed for binding between Ocr and M.EcoKI (Atanasiu *et al.*, 2002).

It is worth noting that the raw ITC data showed a small endothermic event after the end point, which was not evident in the dilution control (i.e., injection of ArdA into buffer). This possibly indicates a shift in equilibrium in the aggregation state of ArdA induced by M.EcoKI binding. The apparent stoichiometry of interaction, *n*, was found to be 1.3±0.01 ArdA monomers per M.EcoKI. It is not clear at this stage whether this non-integer value reflects the true stoichiometry of the ArdA–M.EcoKI complex or is simply a consequence of uncertainties in the quaternary structure of ArdA. However, given the results from the size-exclusion chromatography, it is more likely that several complex equilibria exist with the M.EcoKI binding to ArdA with different quaternary structures and further experiments are required to define the interaction stoichiometry. It was not possible to carry out the ITC experiment using EcoKI instead of M.EcoKI due to the relatively large amount of protein required for each titration.

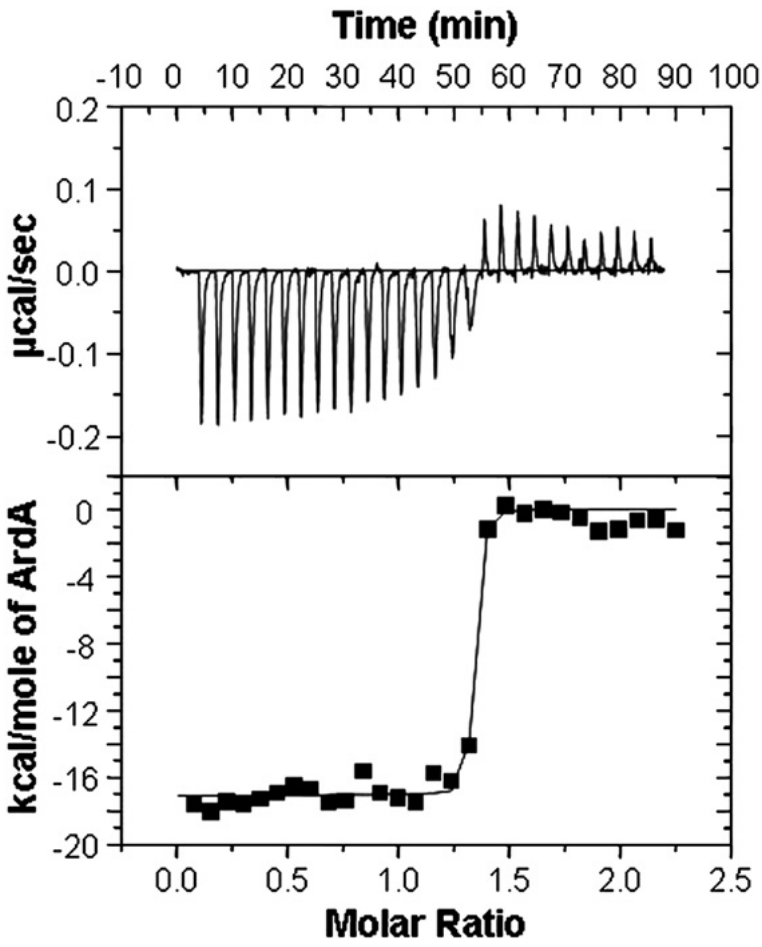


Fig. 38 *Calorimetric titration of ArdA into M.EcoKI. The ITC titration curve is shown in the upper panel. The lower panel shows the calorimetric binding isotherm for the ArdA–M.EcoKI interaction. The x axis shows the molar ratio of ArdA (monomer concentration) to M.EcoKI. The theoretical fit gives a ΔH of −17±0.2 kcal/mol.*

4.8 Restriction/Modification assays

4.8.1 *In vivo* restriction/modification assay

In the restriction assay, *E. coli* strains containing different Type I R/M systems were infected with unmodified virulent λ phage (λv.0). A restriction and modification deficient (r−m−) strain was used as a control (NM1261). Infection using λv.0 of each of the four restriction and modification proficient strains (r+m+) of *E. coli* gave a much reduced titre of phage compared with that obtained using NM1261. No difference in the number of plaque-forming units was observed using the NM1261 strain harbouring either pDMArdA or vector alone (pTrc99A). If ArdA was capable of complete antirestriction, then the r+m+ strains harbouring pDMArdA would give the same number of plaque-forming units as the r−m− strain NM1261. The restriction assay results (Table 1a) showed that orf18 ArdA provided between 2 and 4 orders of magnitude of protection to phage infection using λv.0. The greatest level of antirestriction activity was against the Type IA system.

The ArdA was also tested for possible interference with the modification activity of Type I R/M systems. In the modification assay, phage were recovered from plaques obtained using r+m+ strains harbouring either pDMArdA or pTrc99A. The recovered phage stocks were subsequently used to infect strains lacking the plasmid, either the appropriate r+m+ strain or the NM1261 control strain. If ArdA was capable of inhibiting modification, phage DNA would not have become modified and would therefore remain vulnerable to restriction. Thus, the number of plaque-forming units should be greatly reduced in the

corresponding r+m+ strain compared with the r−m− control strain. However, if ArdA did not inhibit modification, the recovered phage DNA would be modified and rendered immune to restriction. In this case, phage would survive and the numbers of plaque-forming units on the corresponding r+m+ strain and the r−m− control strain would be the same. The results from the antimodification assays (Table 1b) revealed that ArdA was capable of inhibiting the modification activities of all four Type I R/M systems to varying degrees. The greatest antimodification activity was observed against the Type ID R/M system.

Host strain	Restriction system and family	pTrc99A (pfu/ml)	pDMArdA (pfu/ml)	eop pTrc99A (r+m+/r-m-)	eop pDMArdA (r+m+/r-m-)	Restriction Alleviation (RA) (eop pTrc99A/eop pDMArdA)
NM 1261	r-m-	7.50E+09	1.15E+10			
NM 1049	*Eco*KI-IA	9.26E+05	2.86E+10	1.24E-04	2.49E+00	2.02E+04
NM 354	*Eco*AI-IB	6.63E+07	1.56E+10	8.84E-03	1.36E+00	1.54E+02
NK 402	*Eco*R1241-IC	4.26E+06	1.17E+10	5.68E-04	1.02E+00	1.79E+03
NM 1009	*Sty*SBLI-ID	2.53E+06	8.10E+09	3.37E-04	7.06E-01	2.09E+03

Table IA. *Orf18 ArdA antirestriction against Type IA–ID R/M systems as determined in vivo. eop. : (efficiency of plating); pfu: (plaque forming units).*

Host strain	Antimodification Activity {(pfu/ml of r-m-)/(pfu/ml of r+m+)}
NM 1261	1
NM 1049	1.28E+01
NM 354	1.71E+01
NK 402	3.88E+02
NM 1009	4.11E+03

Table 1B *Orf18 antimodification against Type IA–ID R/M systems as determined in vivo.*

4.8.2 *In vitro* antirestriction activity of ArdA

The *in vivo* restriction assays showed that ArdA displayed the greatest level of antirestriction against the EcoKI Type IA R/M system. The inhibition of EcoKI was also analysed *in vitro*. The assay involved monitoring the linearisation of unmethylated plasmid DNA (pBRsk1) using EcoKI in the presence or absence of ArdA. The results are shown in Fig. 39. Incubation of pBRsk1 with a 10-fold excess of EcoKI resulted in complete digestion of the DNA within 10 min (lanes 1 and 2). The experiment was then repeated with three concentrations of ArdA (calculated as monomer) in the reaction mix (i.e., the molar ratio of ArdA–EcoKI was 1:10, 1:1 or 10:1; lanes 3–5). No significant change in the level of digestion could be observed using substoichiometric and stoichiometric amounts of

ArdA. However, almost complete inhibition of the reaction was clearly seen using a 10-fold excess of ArdA over nuclease. Perhaps the ArdA may have required preincubation with nuclease in order to observe full antirestriction activity. The preincubation was performed for 10 min at 37 °C in the reaction mix lacking DNA and ATP. After completion of the preincubation period, the reaction was initiated by addition of DNA plus ATP and allowed to proceed for 10 min. As can be seen in Fig. 39, the highest concentration of ArdA (10-fold excess over EcoKI) caused complete inhibition of digestion (lane 8). No nuclease inhibition was evident at the two lower concentrations of ArdA (lanes 6 and 7).

To test whether ArdA was able to block the nuclease activity of DNA-bound EcoKI, the reaction mixture was made up and incubated for 1 min at 37 °C before addition of the antirestriction protein. No detectable inhibition was observed even at the highest concentration of ArdA (lanes 9–11) using these conditions. Finally, the preincubation of EcoKI with ArdA in the reaction mixture was repeated for 10 min, this time in the presence of pBRsk1. The reaction was then initiated by the addition of ATP. The results of this experiment were essentially the same as those observed when the preincubation of ArdA and EcoKI was performed in the absence of DNA. For the experiments at a 1:1 ratio with 30 nM ArdA monomer, gel filtration indicated that the ArdA protein would be a dimer at a concentration of 15 nM, half that of the EcoKI nuclease. In this situation, one would have expected 50% inhibition of the nuclease if the ArdA–EcoKI association was as strong as that observed for the Ocr protein. Since substantial inhibition was not observed, this implies that ArdA does not interact as strongly as Ocr with EcoKI.

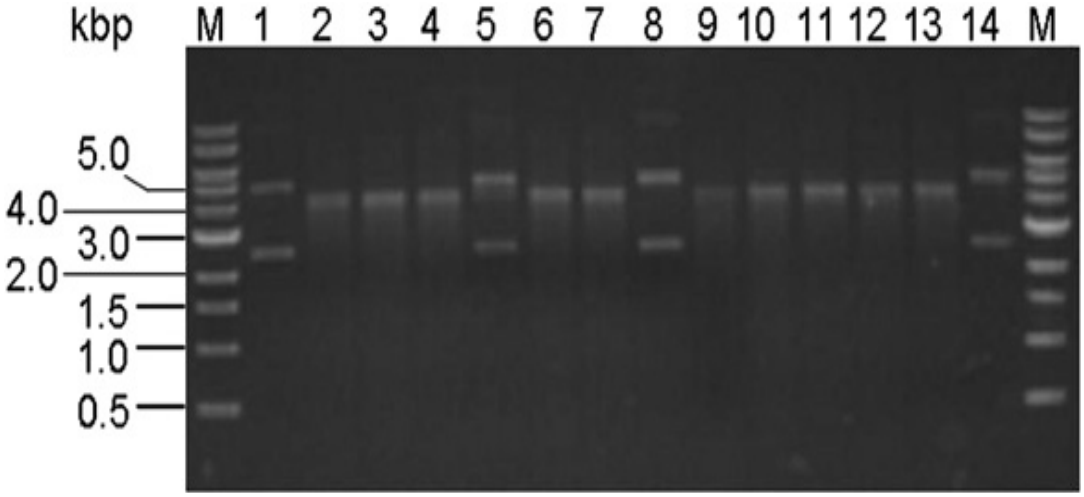


Fig. 39 *In vitro inhibition of EcoKI nuclease by ArdA. pBRsk1, with a single EcoKI recognition site, was used as substrate. Lane 1, uncut pBRsk1; lane 2, EcoKI linearised pBRsk1; lanes 3, 4 and 5, reactions performed in the presence of 0.1-, 1- and 10-fold amounts of ArdA over EcoKI, respectively; lanes 6–8, as for lanes 3–5, except that EcoKI was preincubated with ArdA before initiating the reaction by addition of pBRsk1 and ATP; lanes 9, 10 and 11, plasmid digestion performed for 1 min prior to addition of 0.1-, 1- and 10-fold amounts of ArdA, respectively; lanes 12–14, as for lanes 3–5, except that EcoKI was preincubated with ArdA in the presence of pBRsk1 before initiating the reaction by addition of ATP. Lane M: 1-kbp DNA ladder*

4.9 *In vitro* assay of DNA displacement from M.EcoKI by ArdA

Fluorescence anisotropy of the labelled DNA duplex starts at a low value in the absence of bound M.EcoKI. As M.EcoKI is added, the amount of DNA–M.EcoKI complex increases and since this has a high value of anisotropy, the observed value of the anisotropy for the whole sample increases. If the M.EcoKI is unable to bind to the DNA due to, for instance, the presence of ArdA, then the anisotropy of the whole sample will not increase until all of the ArdA has been bound. Addition of further M.EcoKI then allows the DNA– M.EcoKI complex to be formed, and the anisotropy of the whole sample increases.

Fig. 40 shows that ArdA was an effective competitive inhibitor of DNA binding by the M.EcoKI. Little DNA binding, as indicated by the low anisotropy reading, was observed in the presence of 50 nM ArdA until the M.EcoKI concentration reached ~10 nM, whereas in the absence of ArdA, DNA binding was observed at ~10-fold lower M.EcoKI concentration. This indicated that ArdA bound more strongly to the M.EcoKI and suggested that once this complex was formed, DNA binding could no longer take place. Should ArdA operate in the same manner as the Ocr protein, this effect would be due to a direct binding of ArdA within the DNA binding groove on M.EcoKI. The effect of ArdA was less pronounced than that of Ocr as the onset of DNA binding in the presence of ArdA was not as sharp as observed in the presence of Ocr (Atanasiu *et al*., 2001). This also indicates that ArdA has a weaker affinity for M.EcoKI than Ocr. Addition of further M.EcoKI caused a further increase in the anisotropy signal indicative of multiple M.EcoKI binding to the DNA above ~50

nM M.EcoKI. This has not been observed so strongly in previous work and seems to be specific to this particular preparation of M.EcoKI.

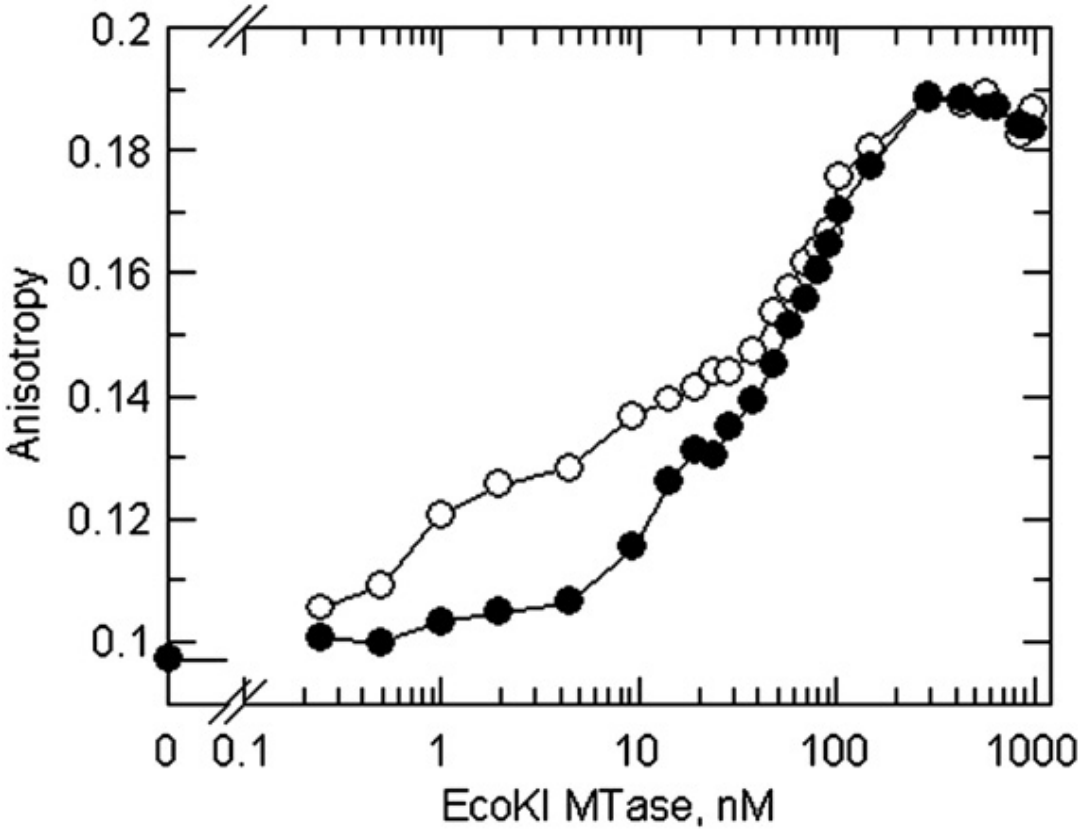


Fig. 40 *Competition between DNA and ArdA for binding to M.EcoKI as determined using the fluorescence anisotropy signal from 2 nM 21-bp fluorescent DNA duplex. Binding of DNA to the M.EcoKI in the absence of ArdA (open circles) or in the presence of 50 nM ArdA (monomer concentration, solid circles) is indicated by an increase in*

anisotropy. ArdA prevents DNA binding until sufficient M.EcoKI has been added to bind the free ArdA.

4.10 Preliminary NMR experiments with Orf18 ArdA

In order to assess whether ArdA would be a suitable candidate for NMR structure determination, a series of preliminary experiments was conducted. First, a 1D ^{1}H spectrum of the protein was collected. A 1D proton spectrum is the simplest of the NMR experiments but it can give a rough idea of the condition of the sample. In the 1D ^{1}H NMR spectra, the dispersion of the NMR signals in the regions of the methyl protons (-0.5 to 1.5 ppm), α-protons (3.5-6 ppm), and amide protons (6-10 ppm) provides the main indicator of folded globular proteins along with the appearance of a very shielded methyl peak at -0.5 ppm. For the ArdA proton spectrum (Fig. 41), there is narrow chemical shift dispersion throughout the spectrum and particularly in the amide proton region. The presence of the large broad signal at the methyl region of the proton spectrum (0.5-1.5 ppm) is a sign of possible multimeric formation or nonspecific aggregation.

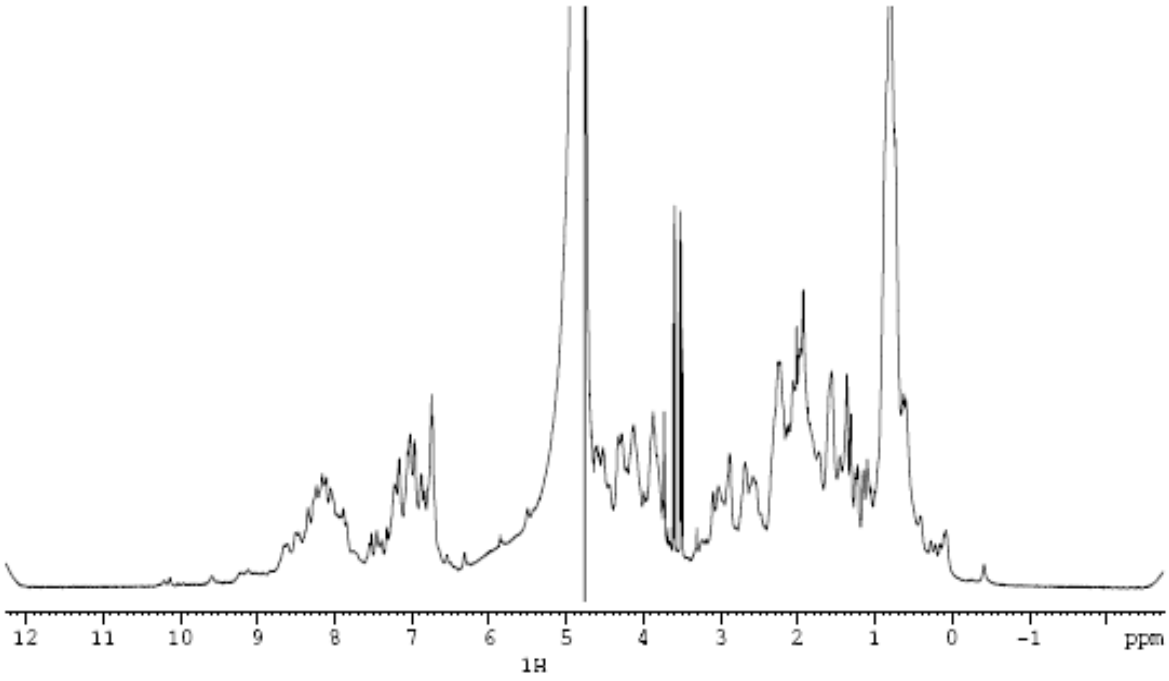


Fig. 41 *1D ^{1}H spectrum with presaturation of 100 μM ArdA in 20 mM phosphate buffer, pH 6.0 at 25 ^{0}C*

A 2D ^{1}H-^{1}H NOESY spectrum (Fig. 42) was also acquired. The experiment transfers magnetisation through space and shows crosspeaks to all protons that are close in space regardless of whether they are in the same spin system or not. The poor dispersion of the peaks in the aliphatic side chain-aliphatic side chain region implies some type of aggregation taking place. Furthermore, the lack of signals in the NH-NH aromatic region was a sign of possible unfolding.

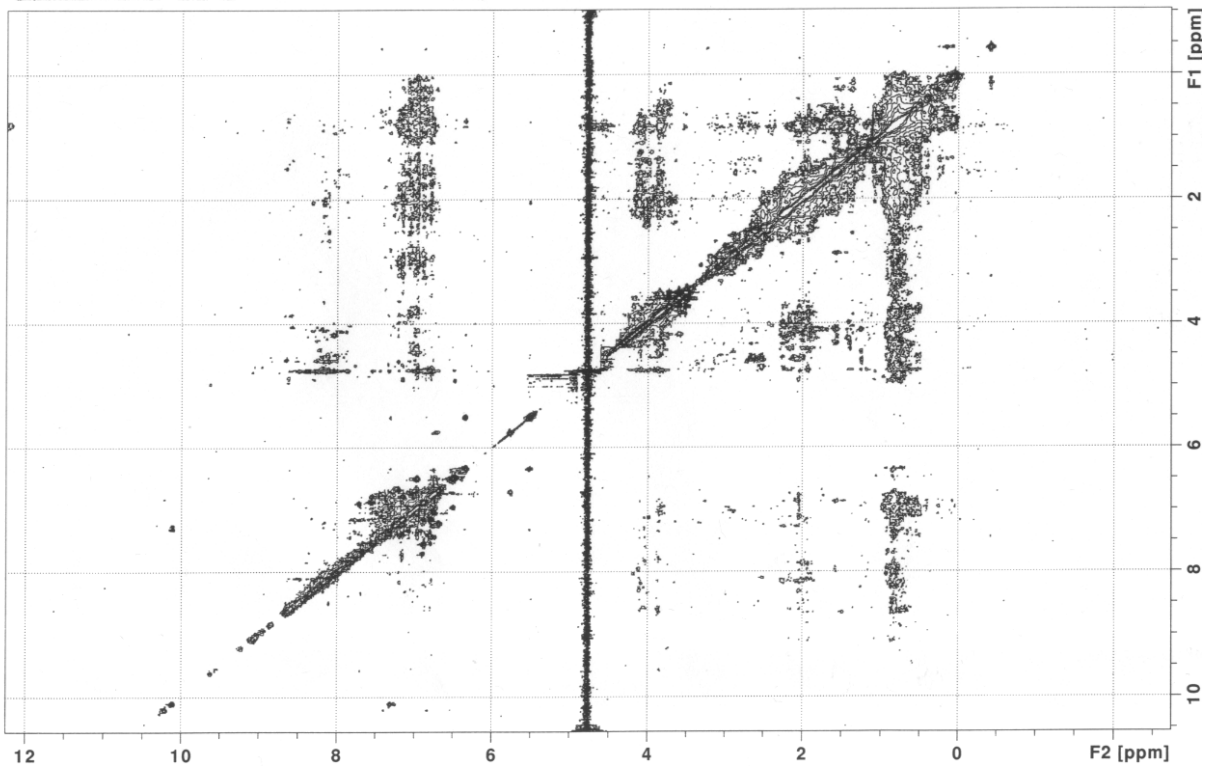


Fig. 42 *^{1}H-^{1}H NOESY spectrum of 230 μM ArdA in 20 mM phosphate buffer, pH 6.0 at 25 ^{0}C.*

From the 1D proton spectrum and the 2D homonuclear NOESY spectrum, however, it cannot be shown unambiguously whether a sample can be a suitable candidate for NMR studies. A uniformly ^{15}N labelled ArdA sample was prepared and a 2D HSQC spectrum was acquired (Fig. 43)

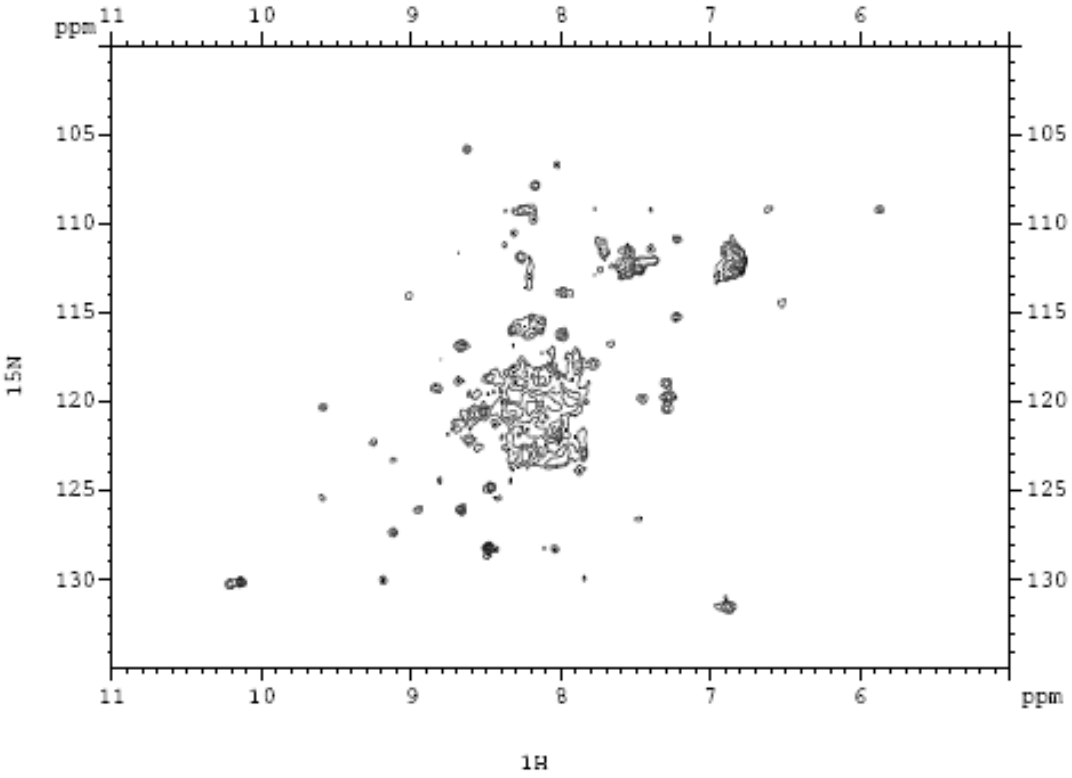


Fig. 43 *^{1}H-^{15}N HSQC spectrum of 140 μM ArdA in 20 mM phosphate buffer, pH 6.0 at 25 ^{0}C*

The spectrum showed very poor dispersion of peaks with most of the signals occupying the centre of the spectrum, a sign of possible aggregation/partial unfolding/oligomerisation of the protein.

In order to improve the quality of the spectra, different conditions were tested such as:

- Buffer: Phosphate buffer is the most commonly used buffer for protein NMR studies, however, phosphate buffer can sometimes make proteins more prone to aggregation than other buffers. Because aggregation was suspected for ArdA, deuterated sodium acetate buffer was also tested, but no improvement was observed in the spectra.

- Temperature: increasing the temperature normally produces sharper peaks in spectra since the higher temperature increases the correlation time of the molecule. A short correlation time implies prolonged relaxation of the spins. The widths of the lines in an NMR spectrum are proportional to 1/T2, T2 being the transverse relaxation time. In other words, fast-moving molecules have long T2 values which in turn will produce sharp peaks in the spectrum. Increasing the temperature to 37 ^{0}C did not produce significant improvement of the spectrum.

- pH : Protein NMR spectra should normally be collected at a pH below neutral. The reason for this is that the amide protons of exposed amide groups are in exchange with the solvent, and this exchange is pH dependent. If amide protons are allowed to exchange with the solvent, the amide proton peaks will disappear from the spectrum. The pH dependence of the amide proton exchange rate is expressed according to Woodward and Hilton (1980):

$$k_{ex} = k_H [H^+] + k_{OH} [OH^-] + k_0$$

Where k_H, k_{OH} and k_0 are the rate constants for acid-catalysed, base-catalysed and direct exchange with water, respectively. Amide proton exchange rates

reach a minimum at pH ~ 3 for model compounds. Different pH values were tested with the minimum pH being 5.0. No significant improvement in the amide region was observed. Below 5.0 the spectra were worsened probably because the pH was approaching the isoelectric point of the molecule (pI = 3.91).

- Ionic strength: Small amounts of NaCl in the protein sample may help to decrease possible aggregation. Increasing amounts of NaCl were titrated to the protein sample up to 200 mM NaCl concentration (high ionic strengths > 0.2 M NaCl is normally avoided because of the generation of ion currents during radiofrequency pulses. These currents decrease the efficacy of the pulses, especially towards the centre of the sample, and lead to sample heating). The addition of even small amounts of NaCl decreased the resolution of the spectra clearly aggregating the protein.

Because the addition of salt worsened the aggregation problem, the type of aggregation was presumed to be hydrophobic in nature. The addition of a non-denaturing detergent is expected to disrupt these hydrophobic intermolecular interactions, alleviating the aggregation problem. The detergent that was chosen was CHAPS (3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate) which is a zwitterionic surfactant commonly used to solubilise membrane proteins as it does not change their secondary and tertiary structure (Kline *et al.*, 1989). The detergent has been successfully used for the structure determination of calcineurin B (Anglister *et al.*, 1993). CHAPS is not commercially available in deuterated form, however, because the protein was isotope-enriched, the signals from the detergent would not interfere with the HSQC experiment.

The effect of CHAPS, however, has been found to be protein dependent as well as concentration dependent. The detergent at 5mM concentration has been shown to have an excellent stabilising effect on beta-lactoglobulin, but the same concentration of surfactant caused unfolding of a different protein, ovalbumin (Sah and Kim, 2006).

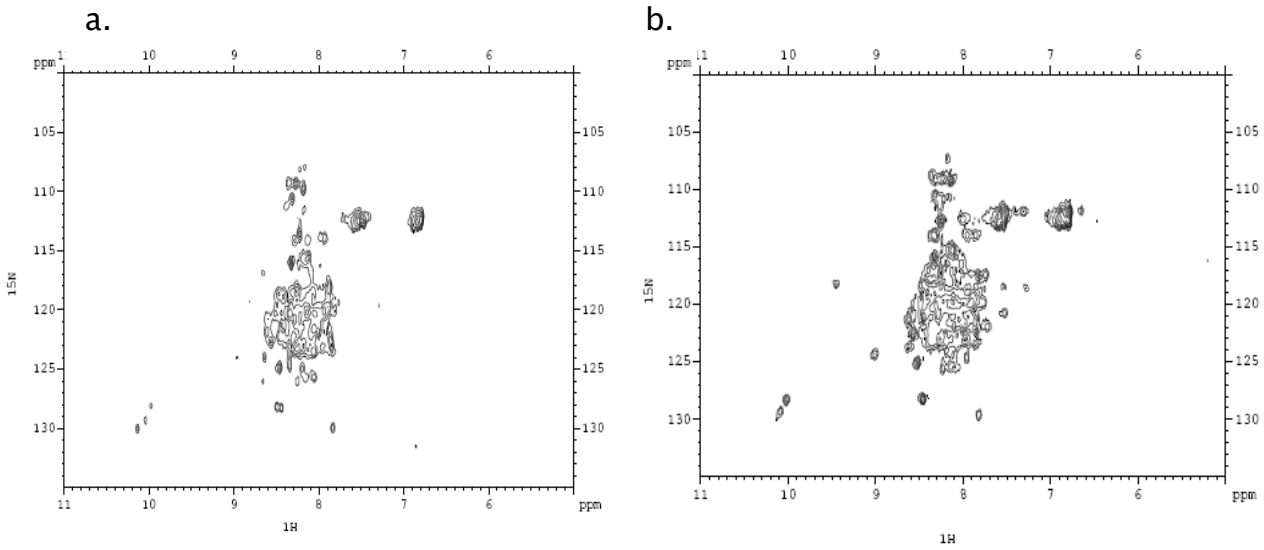


Fig. 44 *^{1}H-^{15}N HSQC spectra of 140 μM ArdA in 20 mM phosphate buffer, pH 6.0 at 25 ^{0}C with the addition of 2 mM CHAPS (a) and 10 mM CHAPS (b).*

The addition of CHAPS had no effect at μM concentrations and at higher (mM) concentrations clearly unfolded the protein as can be seen from the disappearance of most of the peaks and the enlargement of the central "blob" on the spectrum as seen in Fig. 44.

At low surfactant concentrations, protein-surfactant binding involves specific and non-cooperative electrostatic interactions. However, at high surfactant concentrations, massive, cooperative hydrophobic binding occurs which results in protein unfolding and denaturation by disrupting the α-helical network and leading to a more open, random, solvent-exposed protein structure (De *et al*., 2005).

4.11 Truncated Orf18 ArdA

Since the Orf18 ArdA did not prove suitable for NMR studies, an attempt was made to create a truncated version with much lower MW in the hope that this species would not oligomerise or aggregate. Deletion studies that had been performed in the past of an ArdA protein encoded by the IncI plasmid ColIb-P9, revealed that the N-terminal half of the protein does not seem to be essential for its antirestriction activity (Delver *et al*., 1991; Read *et al*., 1992). It was also shown that the C- terminal portion of that ArdA containing 77 residues (coordinates 90 to 166) was able to function against EcoKI restriction as a native ArdA (Nekrasov *et al*., 2007). Thus, a truncated version of Orf18 ArdA was designed where the largest part of the N-terminus was removed (the first 71 amino acids were deleted). The truncated version started with the 72nd amino acid, a methionine. The truncated Orf18 was cloned between the *Nde*I and *Bam*HI recognition sites in the pET-15b vector which carries an N-terminal His-tag sequence followed by a thrombin cleavage site. The cloning procedure was performed as described in the general cloning methods. The following is the whole Orf18 ArdA sequence and the truncated version is highlighted in blue.

MDDMQVYIANLGKYNEGELVGAWFTFPIDFEEVKEKIGLNDEYEEYAIHDYELPFTVDE
YTSIGELNRLWEMVSELPEELQSELSALLTHFSSIEELSEHQEDIIIHSDCDDMYDVAR
YYIEETGALGEVPASLQNYIDYQAYGRDLDLSGTFISTNHGIFEIVY

The following is the truncated Orf18 along with the His tag and the thrombin site:

MGSSHHHHHHSSGLVPRGSHMVSELPEELQSELSALLTHFSSIEELSEHQEDIIIHSDC
DDMYDVARYYIEETGALGEVPASLQNYIDYQAYGRDLDLSGTFISTNHGIFEIVY

In order to confirm that the cloning procedure was successful, a digestion of the construct with NdeI and BamHI was performed. The result is shown in Fig. 45. The linear pET15-b was expected to appear as a band close to 6 kb (5708bp). The expected size of the truncated Orf18 was ~300 bp.

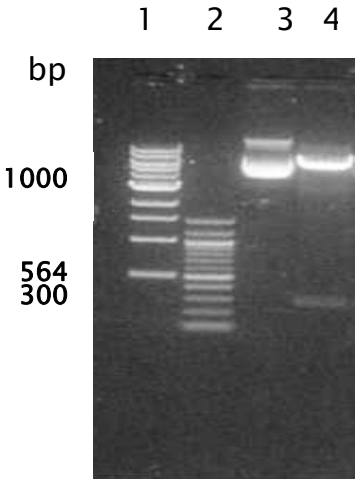


Fig. 45 *Agarose gel showing the result of digestion of the pET15-b construct containing the truncated orf18. Lane 1: 1Kb DNA marke ; Lane 2: 100 bp DNA ladder; Lane 3: pET15-b containing the insert uncut; Lane 4: pET15-b containing the insert digested with NdeI and BamHI.*

An attempt was made to purify the truncated his-tagged Orf18 ArdA. With the HisTrap FF crude columns histidine-tagged proteins can be purified directly from unclarified cell lysates and are recovered from the column under mild elution conditions. The columns are precharged with Ni^{2+}. The matrix of the

column consists of 90 µm particles of highly cross-linked agarose to which Ni^{2+} are bound by a stable chelating group. The imidazole side chain of histidine has an affinity for binding metal ions such as nickel (cobalt is another metal that can be used as an alternative for nickel). The more histidines present, the stronger the binding will be, thus six his residues were used to tag the protein instead of one.

His-tagged proteins were eluted from the nickel rich column with buffers containing imidazole. Imidazole at low concentrations was used in the binding and wash buffers to minimise binding of unwanted host cell proteins. At higher concentrations, imidazole may decrease the binding of the his-tagged protein and thus the optimal concentration is different for different his-tagged proteins. The optimal imidazole concentrations was determined by stepwise elution of the protein with buffers containing increasing concentrations of imidazole.

After running an SDS-PAGE of the eluted his-tagged protein at different imidazole concentrations, no distinct band for the truncated Orf18 was observed. There were however, several very faint bands visible in the gel suggesting non-specific binding of proteins to the column. Some bands of non-specific binding were observed even at the highest recommended (500 mM) imidazole concentration.

The truncated Orf18 was also tested for *in vivo* antirestriction activity, and no antirestriction activity was observed suggesting that the truncation affected the structure of the protein significantly.

Since no band was observed on the SDS-PAGE gel, in order to test whether the protein was successfully tagged and eluted, a Western blot was prepared (Fig 46). Clearly a band can be seen on lane 6 corresponding to the truncated Orf18 ArdA at the expected molecular weight (12859 Da). A positive control (His-tagged P450 Rhf Heme protein) was also used to test efficiency of western blot. The expected MW of the control protein (P450 Rhf Heme) is 48 kDa. The protein seems to be somewhat degraded since two bands appear at around 50 kDa. Also another band at ~ 100 kDa appears (a possible dimer of the control protein).

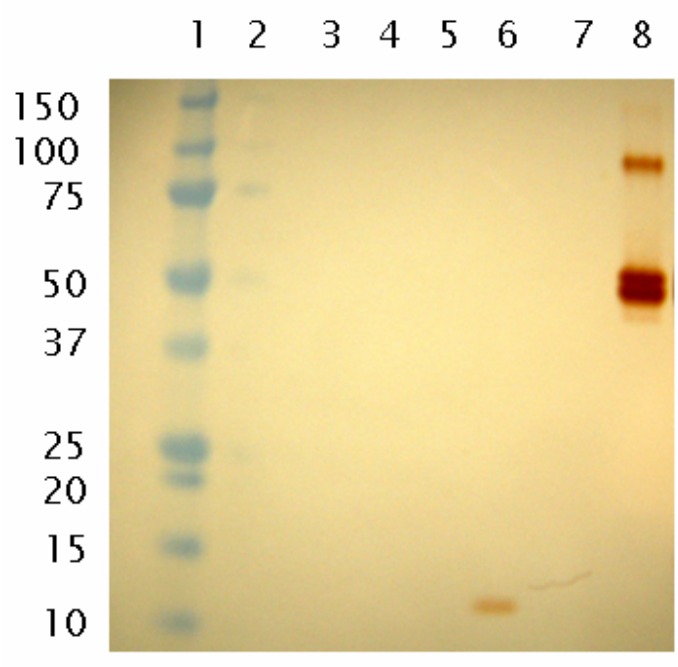


Fig. 46 *Western Blot to test any signs of expression of the truncated Orf18 ArdA. Lane1: protein marker; 2: blank; 3:BL21 (DE3) + pET-15b (0 hrs); 4: BL21 (DE3) + pET-15b (3 hrs); 5: BL21 (DE3) + pET-15b + Truncated Orf18 (0 hrs); 6: BL21* (DE3)

+ pET-15b + Truncated Orf18 (3 hrs) at 30C ; 7: blank 8: (Positive Control) P450 Rhf Heme protein (His-tagged)

The appearance of a band corresponding to the truncated his-tagged Orf18 ArdA indicates that some protein was indeed expressed in the cell, although in very low concentrations which was not enough to purify the protein. The low expression of the protein implies that most of the protein in the cell must have been produced unfolded and thus degraded by proteases. The removal of almost half of the protein seems to have proven detrimental and perhaps one of the causes for causing folding instability may have been that the truncation disrupted helix 3 according to the secondary structure prediction form PSIPRED.

4.12 X-ray crystal structure of Orf18 ArdA

While this thesis was being prepared, the crystal structure of Orf18 ArdA was solved (McMahon *et al*., 2009). The ArdA was solved at a resolution of 2.8 Å (PDB Accession number: 2w82) and appears similar to Ocr (Walkinshaw *et al*., 2002). The crystal structure of ArdA shows an extremely elongated dimer with a highly negatively charged surface (Fig. 47b). Its structure is reminiscent of ~42 bp of B-form DNA making it the largest DNA mimic yet characterised.

ArdA is composed of three domains (Fig. 47a) arranged in such a linear manner that only eight residues are buried form the solvent. The N-terminal domain resembles the B domain in biotin carboxylase from yeast (Shen *et al*., 2004). The

central domain of ArdA shows weak similarity to the anti-termination ANTAR domain found in RNA binding proteins (Morth *et al*., 2004), while the C-terminal domain shows similar fold with the arginine repressor protein (Garnett *et al*., 2008). Even though each of the three domains has similar folding to three known proteins, the overall structure of ArdA has no counterpart in any known structure. The second and third domain of ArdA have similar folding to nucleotide binding proteins, however, the extensive negatively-charged surface of ArdA makes the protein highly unlikely to interact with DNA.

The dimer interface contains the antirestriction motif identified by Belogurov and Delver (1995). Since it belongs to the interface it may be more important for the assembly of the dimer structure rather than having a direct implication in antirestriction activity. However, this is just an assumption and more experiments need to be done to confirm such a hypothesis.

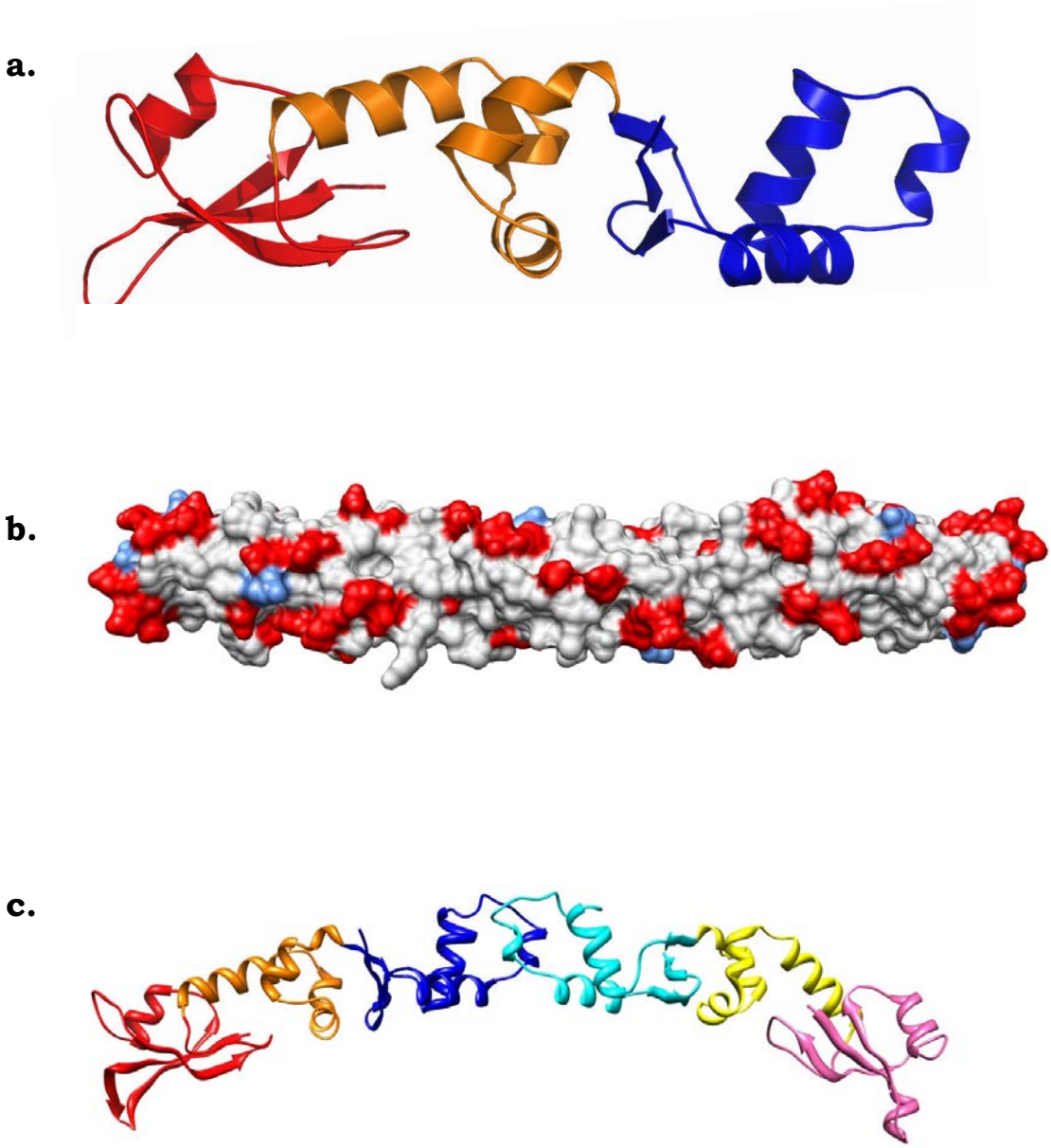


Fig. 47 ***a)*** *Cartoon representation of ArdA monomer. The three domains are coloured red, orange, blue.* ***b)*** *Surface representation of ArdA . Acidic residues are coloured red*

and basic residues are coloured light blue ***c****)Cartoon representation of the dimeric structure of ArdA. (Images reprinted from McMahon et al., 2009).*

By comparing the crystal structure of ArdA with the biophysical analysis of the protein discussed earlier, a few points can be made. The elongated nature of the protein along with its tendency to dimerise can explain the bad quality of the NMR spectra from the preliminary NMR experiments. Furthermore, the secondary structure elements derived from the crystal structure agree with the secondary structure predictions from PSIPRED as well as with the far-UV CD spectrum that the protein is largely composed of α-helices. However, the percentages somewhat differ with the crystal structure showing a higher percentage of β–sheet (46% α-helix and 23 % β–sheet).

The tryptophan quenching experiments showed that two tryptophan residues differed only slightly in their solvent accessibility. When examining the 3D structure (Fig. 48) it is evident that the W23 and W70 are both highly exposed with the indole ring of W23 being somewhat more exposed than that of W70. To further confirm the degree of solvent accessibility of the two residues an analysis was performed with the program GETAREA (Fraczkiewicz & Braun, 1998). The program gave a ratio of side-chain surface area to "random coil" value per residue. The "random coil" value of a residue X is the average solvent-accessible surface area of X in the tripeptide Gly-X-Gly in an ensemble of 30 random conformations. The ratio for W23 was 36.4 and for W70 27.5, suggesting that W23 is slightly more exposed than W70.

Moreover, the fluorescence quenching experiments showed that the microenvironment surrounding the two tryptophan residues was not highly negatively charged since if that was the case, the negative charge would repel the iodide and no quenching would take place. Indeed, the area surrounding the two tryptophans is not highly negatively charged with the exception of the close proximity of D50 and E52 with W23 and E71 with W70.

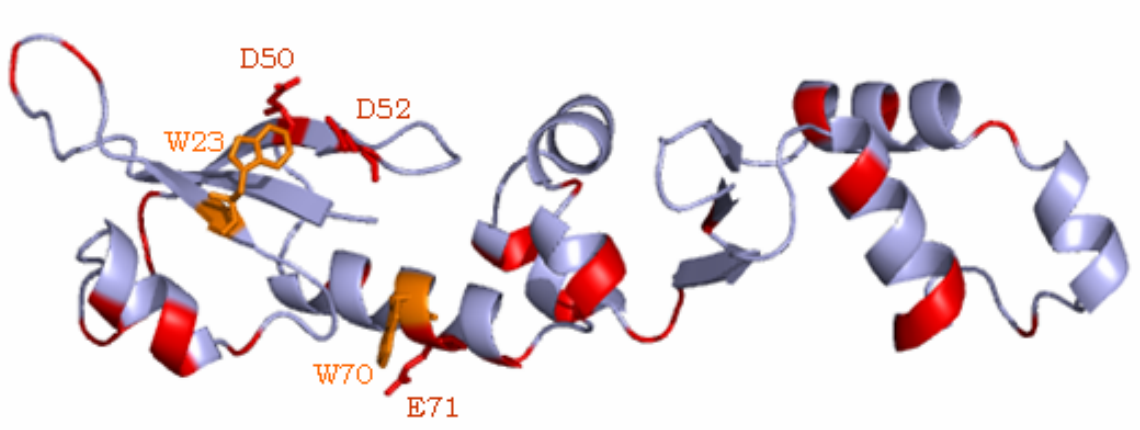


Fig. 48 *Monomer of Orf18 ArdA showing the position of the two tryptophan residues (W23 and W70). Negatively charged residues (Asp, Glu) are coloured red, the tryptophan residues are coloured orange. Image generated using PyMol.*

Gel filtration analysis showed that the protein existed as a dimer up to 200 nM, but formed larger species (tetramers and hexamers) at higher concentrations. The crystal structure, however, shows the protein to exist as a dimer even though the concentration of the protein used to produce the crystal was ~1 mM. This discrepancy could be explained by the elongated shape of the molecule. A tumbling, elongated molecule occupies more volume than a spherical protein

of the same mass. Thus, an elongated molecule will not be able to enter the beads of the matrix in the column and will elute ahead of the spherical protein even though the two proteins are of the same molecular mass (Note also, that the gel filtration column was calibrated with globular protein standards: (BSA, carbonic anhydrase, alcohol dehydrogenase, β-amylase, cytochrome c).

4.13 Conclusion for ArdA experiments

The results of this work show the gene product of *orf18* is an ArdA antirestriction protein and prove that such a function can exist on conjugative transposons in addition to the known antirestriction functions present in phage and conjugative plasmids. The antirestriction activity of Orf18 ArdA presumably aids successful conjugative transposition of Tn*916* and thus assists the spread of antibiotic resistance, in this case tetracycline resistance, within the bacterial population. As was the case for Ocr, it was found that Orf18 ArdA inhibits different Type I R/M systems from different families with different DNA target recognition sequences.

The *in vivo* experiments demonstrate that Orf18 ArdA displays potent antirestriction activity against all 4 classes of Type I R/M systems with the highest activity against the Type IA family. Furthermore, antimodification activity was observed against all four R/M systems with the greatest antimodification activity shown against the Type ID family. Antimodification was not as potent as found for the Ocr protein (Nekrasov *et al*., 2007) perhaps due to the weaker interaction of ArdA with the MTase core. Of further note is the difference in the relative effectiveness of antirestriction versus

antimodification. ArdA is more potent as an antirestriction protein than an antimodification protein versus the IA, IB and IC families. The effect of ArdA on the two activities might be expected to be equal if it only acts on the MTase core of a Type I R/M enzyme. Perhaps ArdA has binding sites on the HsdR restriction subunits in addition to a binding site on the core MTase and that these sites have different binding affinities. It was observed that Ocr also had multiple binding sites on EcoKI (Atanasiu *et al*., 2001).

In vitro assays using EcoKI confirm that ArdA can inhibit Type I nuclease activity. However, this was only observed using an excess of ArdA (monomer concentration) over EcoKI. Stoichiometric and sub-stoichiometric levels of ArdA displayed no detectable inhibitory effects. As is the case for Ocr (Bandyopadhyay *et al*., 1985) ArdA appears to be ineffective in blocking nuclease activity once EcoKI has bound to the DNA.

In this study it is shown that ArdA blocks DNA binding by the M.EcoKI and this should be sufficient to prevent both methylation and DNA cleavage activities by Type I R/M systems but, as noted above, there may be additional binding sites for ArdA on the complete R/M enzyme.The biophysical characterisation of ArdA suggests the protein exists as a dimer in solution at low concentration, as found for ArdA from a conjugative plasmid (Nekrasov *et al*., 2007), but has a tendency to form larger molecular weight species at concentrations above ~1 μM.

The calorimetric data demonstrates tight association between ArdA and the M.EcoKI. However, because the transition was very sharp, it was not possible to accurately determine the association constant K_a. Nevertheless, K_a for ArdA-M.EcoKI interaction is certainly $>10^9$ M^{-1}, a rather stronger interaction than found for the plasmid-encoded ArdA (Nekrasov *et al*., 2007) but weaker than the Ocr-M.EcoKI interaction (Atanasiu *et al*., 2001). The nature of the active form of ArdA is unclear from the apparent stoichiometry of binding (n = 1.3 ArdA monomers per M.EcoKI) perhaps due to the variable quaternary structure of ArdA as suggested by the gel filtration and the discrepancy between ΔH_{cal} and ΔH_{VH} determined by DSC.

The results on this ArdA from a classic conjugative transposon can be compared with those obtained previously with ArdA proteins from conjugative plasmids (Belogurov *et al*., 2000; Belogurov & Delver, 1995; Wilkins, 2002; Belogurov *et al*., 1985; Belogurov *et al*., 1993; Thomas *et al*., 2003). Although these plasmid-borne *ardA* genes have only been tested against the Type IA, IB and IC families, antirestriction and antimodification were clearly observed. Interestingly, the ArdA from plasmid ColIb-P9 had equally strong antirestriction and antimodification activities against EcoKI *in vivo* whereas in this study it was found that this is not the case for ArdA from Tn*916* which displays a weaker antimodification activity. As the two ArdA proteins display limited sequence identity (24% identity, 58% similarity), they may have slightly different interfaces with and different affinities for the restriction and modification parts of the complete R/M enzyme. Whether these differences have a role to play in

the different mechanisms of HGT used by different mobile genetic elements remains to be seen.

Preliminary experiments with a ^{15}N-labelled ArdA sample showed that the protein is unsuitable for structure determination by NMR. However, the protein crystallized and the ArdA structure (McMahon *et al.*, 2009) revealed a structural and electrostatic mimic of DNA in the same way as the Ocr protein. Comparing the biophysical analysis of ArdA with the information obtained from the 3D structure, the circular dichroism data agree well with the secondary structure elements of the crystal structure and the two tryptophans differ slightly in the solvent accessibility as it was predicted form the steady-state fluorescence experiments. The discrepancy between the initial estimation of oligomerisation of the ArdA with the gel filtration method and the apparent dimer from the crystal structure may be attributed to the extreme elongated structure of the molecule.

Chapter 5: KlcA Results and Discussion

5.1 ^{1}H, ^{15}N and ^{13}C chemical shift assignments and structure calculations

klcA was prepared by gene synthesis using the deposited DNA sequences as a starting point for codon optimisation and synthesis (GeneArt). The synthetic gene (Appendix Fig. I) was inserted into vector pET24-a carrying a kanamycin resistance determinant. *E. coli* BL21(DE3) cells were transformed with the recombinant plasmid (pDMKlcA) and *klcA* was overexpressed. KlcA from *B. pertussis* contains 142 amino acid residues (theoretical M_r: 15660) and is rather acidic (calculated pI: 4.76). The acidity of the protein helped the purification process and the KlcA was purified with exactly the same procedure as ArdA as described in Methods. An SDS-PAGE showing the purified KlcA is shown in Fig. 49.

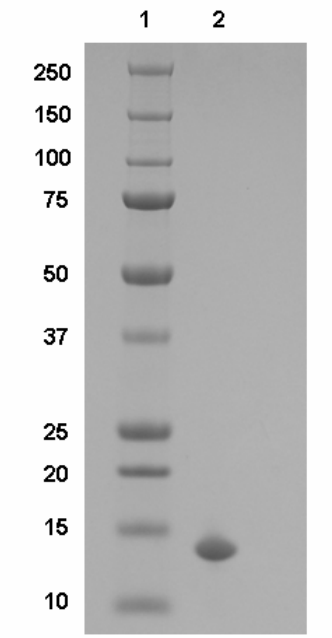
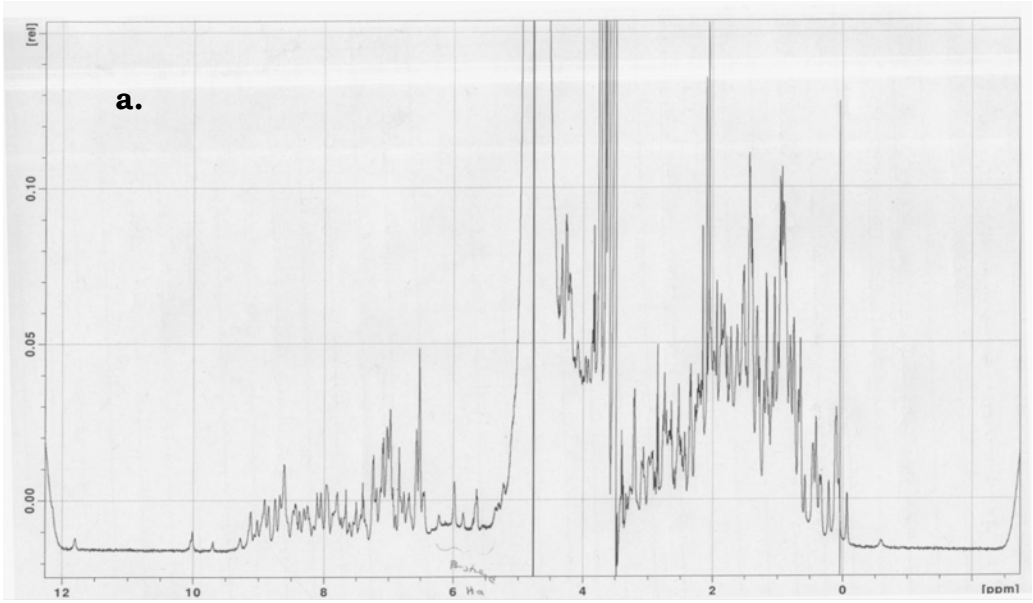


Fig. 49 *SDS-PAGE of purified KlcA. Lane 1: protein marker; Lane 2: purified KlcA after anion exchange chromatography and gel filtration*

In order to assess whether KlcA would be suitable for NMR-based structure determination some preliminary NMR experiments were performed including a 1D presaturated ^{1}H spectrum and a 2D ^{1}H-^{1}H NOESY (Fig. 50) which showed excellent dispersion of peaks.



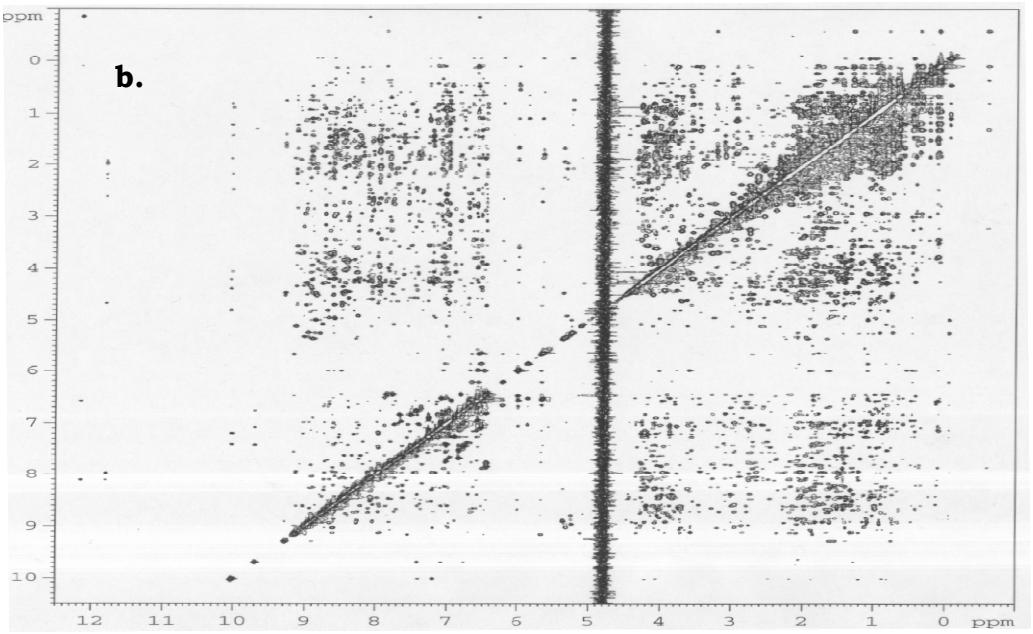


Fig. 50 ***a)*** *1D ^{1}H spectrum with presaturation of 75 μM KlcA in 20 mM sodium acetate buffer, pH 6.0, at 25 ^{0}C. The large contaminating peaks at 3.5-3.8 ppm belong to residual glycerol after dialysis* ***b)*** *^{1}H-^{1}H NOESY of 220 μM KlcA in sodium acetate buffer, pH 6.0 at 25 ^{0}C.*

A ^{15}N labelled KlcA was subsequently prepared and the recorded ^{1}H-^{15}N HSQC spectrum showed a folded protein with well dispersed signals and very little overlap, indicating the feasibility of structure determination of KlcA by NMR spectroscopy. A double labelled ^{15}N-^{13}C KlcA was used for the assignment process. Backbone and side-chain resonances of KlcA were assigned based on 2D and 3D triple-resonance experiments as described in the Methods (Chapter 3). The assigned ^{1}H-^{15}N HSQC spectrum is shown in Fig. 51.

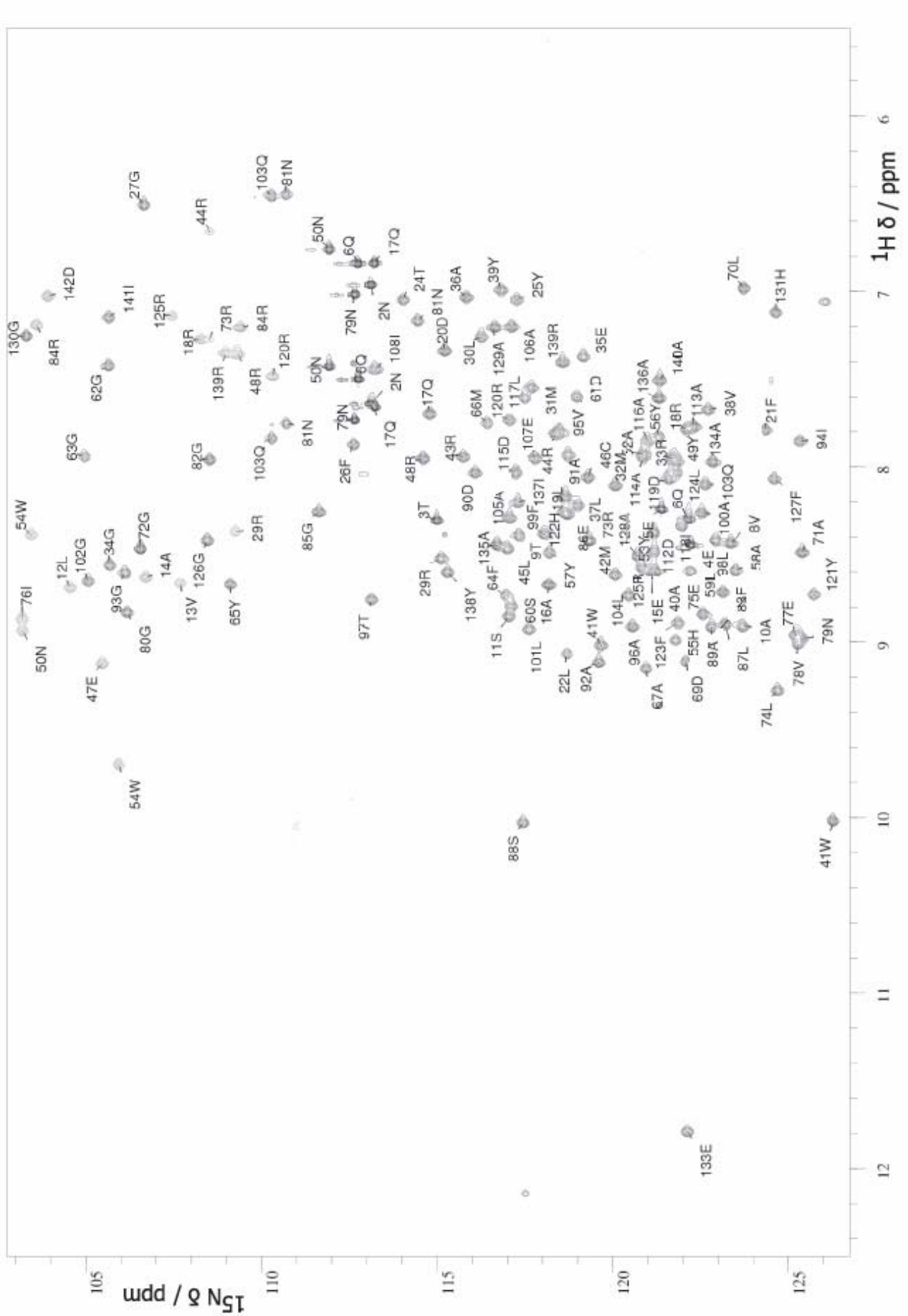

15N δ / ppm
1H δ / ppm

Fig. 51 *2D ^{1}H-^{15}N HSQC spectrum of KlcA at 800 MHz with residue assignments included.*

3D ^{13}C and ^{15}N edited NOESY spectra were carefully peak picked and the NOE cross-peak assignments were obtained from the automated iterative assignment program CANDID which works in conjunction with 3D structure calculations in the program CYANA 2.1. The lowest energy ensemble of 20 structures of KlcA were calculated using NOE restraints, TALOS dihedral restraints and ^{1}H-^{15}N RDCs as detailed in the Methods section. Chemical shift has been shown to have a direct correlation with phi and psi torsion angles and thus with secondary structure (Wishart *et al.*, 1991) The TALOS software tries to predict the most likely backbone angles for a given residue and searches a database that contains $^{13}C_{\alpha}$, $^{13}C_{\beta}$, $^{13}C'$, $^{1}H\alpha$ and ^{15}N chemical shifts for 20 proteins for which a high resolution X-ray structure is available (Cornilescu *et al.*, 1999).

For the RDCs the filamentous phage Pf1 was used as the alignment medium. Since KlcA is a rather acidic protein, Pf1 appeared to be a good choice as an alignment medium for this protein because the negatively charged phage would not normally bind to the acidic KlcA due to electrostatic repulsion. Addition of 6.85 µg/ml of phage appeared to be the optimum phage concentration for this sample which gave a D_2O splitting of 7 Hz. Larger amounts of phage did not increase the splitting. Two 2D IPAP spectra were collected (with and without phage) and the RDCs were measured in Analysis with the help of the RDCcalculator macro, courtesy of Dr. Graeme Ball. The RDC values measured were used as input for REDCAT which calculated the alignment tensor from the

3D structure obtained without RDC refinement and gave the predicted (back-calculated) RDC values. There was significant discrepancy in the back-calculated and the actual experimental values. After addition of the measured couplings in the CNS program and recalculation and refinement of the structure, the new refined structure was again used in REDCAT and this time the measured vs. calculated couplings agreed well (Fig. 52)

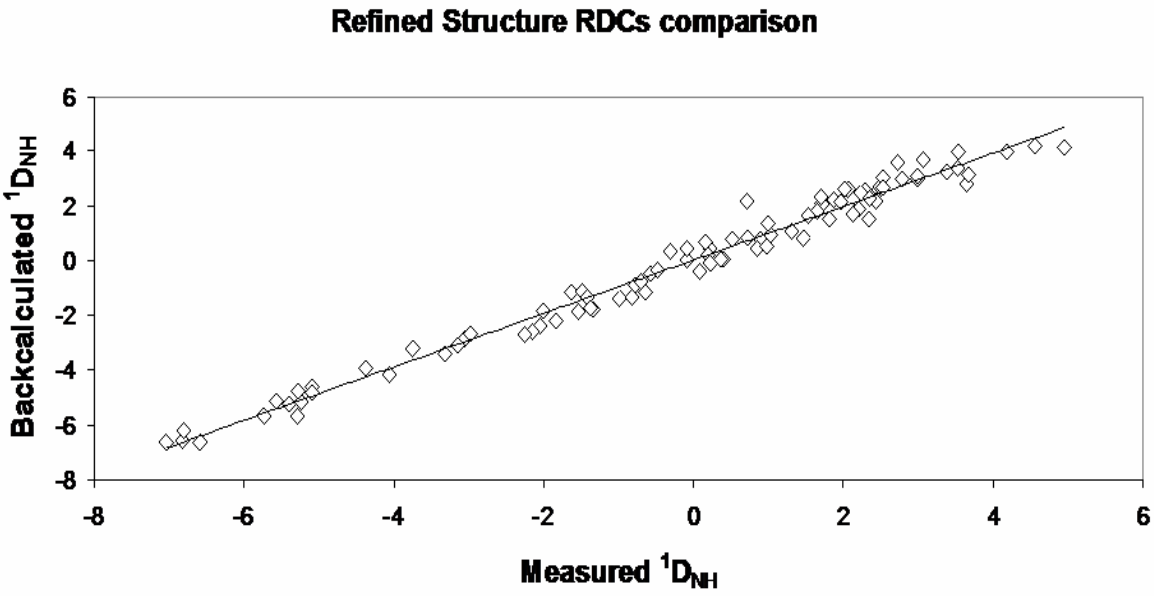


Fig. 52 *Fitting of the $^1D_{NH}$ residual dipolar couplings in the refined KlcA structure. Back-calculation of the couplings was performed with the program REDCAT.*

The final ensemble of 20 structures after RDC refinement is shown in Fig. 53

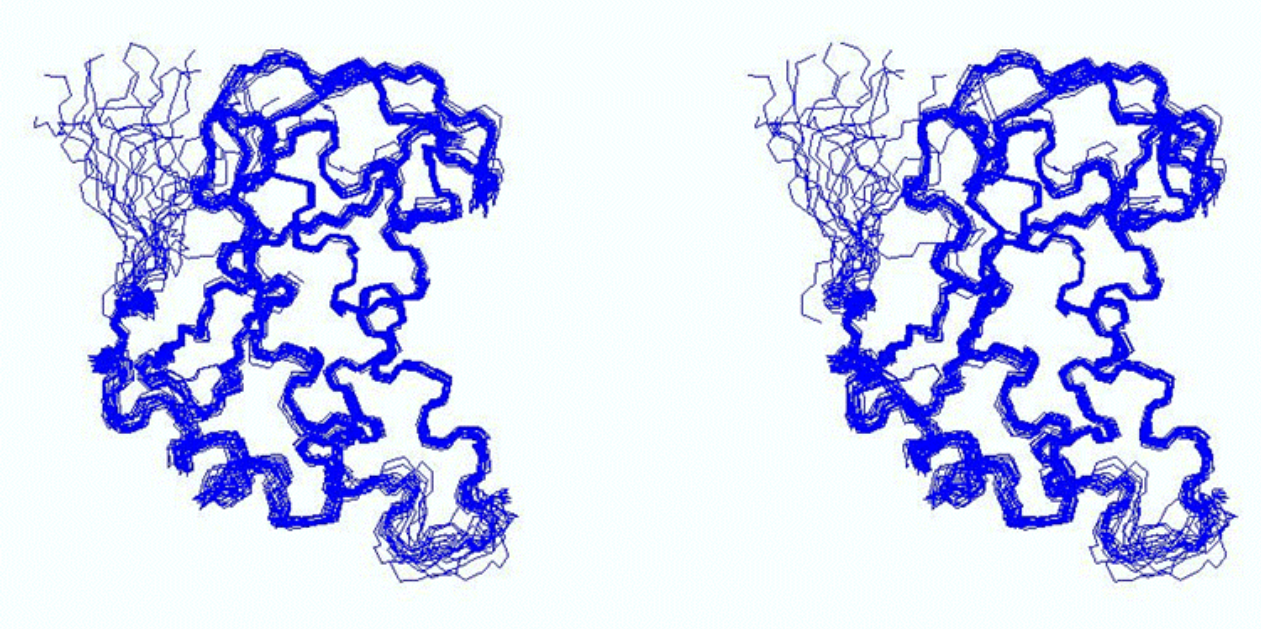


Fig. 53 *Stereoview of the 20 lowest energy structures of KlcA. The steroview was obtained with the program MolMol.*

The calculations converged well, yielding RMSDs of 0.63 and 0.65 Å for backbone and all heavy atoms in structured regions (residues 5-142). The structures were validated using PROCHECK and WHAT IF. 98.5% of residues in the Ramachandran plot appeared in the most favourable and additionally allowed regions. WHATIF analysis yielded an RMS z-score of -1.432 suggesting a good quality structure. A list of the NMR restraints and structural statistics is shown in Table 2**.**

Residual Dipolar Couplings	
$^1D_{NH}$	92
Dihedral Restraints	
Torsion angle (ϕ/ψ)	228
Distance Restraints	
Short-range |i-j| <=1	1117
Medium-range 1 < |i-j| < 5	617
Long-range |i-j| >= 5	604
RSMD (residues 5-142)	
All	1.04
Backbone	0.63
Heavy atoms	0.65
WHATIF score	-1.432
Ramachandran Plot Statistics (%)	
Most favoured regions	83
Additionally allowed regions	14.8
Generously allowed regions	0.7
Disallowed regions	1.6
Assignment percentages (%)	
Amide	94.2
Backbone	95.9
Side-chain H	90.9
Side-chain non-H	66.2
Element C	81.5
Element H	91.9
Element N	78.42

Table 2: *NMR-derived restraints and structural statistics of the 20 lowest energy KlcA structures*

5.2 The three-dimensional structure of KlcA

The KlcA protein structure is made up of five alpha-helices (labeled 1-to-5), a single 3_{10}-helix and five beta-strands (labeled A-to-E) and belongs to the mixed alpha-beta class of proteins (Fig. 54). The five beta-strands form two separate beta-sheets, the first composed of beta-strands A (usually occurring as an isolated beta-bridge from the ensemble consensus), B and C and the other made up of beta-strands D and E. Querying fold libraries and the PBD with the NMR-determined KlcA structure shows that its fold is novel. This is consistent with the lack of sequence identity/significance scores for its sequence against the PDB or remote homology to any potentially related folds, when undertaking fold recognition.

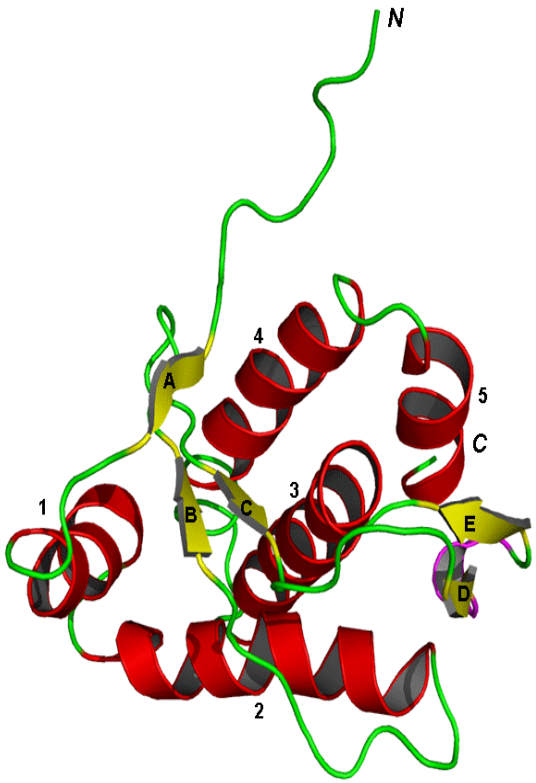


Fig. 54 *A cartoon representation of KlcA using Pymol*

Highlighting the location of all negatively charged side-chain residues on the cartoon schematic representation of KlcA reveals two distinct clusters - an "Aspartate cluster" made up of five linearly arranged residues – i.e. D61 (in the loop between beta-strands B and C), and D110, D112, D115, D119 (in the loop between alpha-helix 3 and 4 and on alpha-helix 4); and a "Glutamate cluster" that contains amino acid residues E47 (in the loop between alpha-helix 2 and beta-strand B), E75 and E77 (on beta-strand D), E86 (on beta-strand E) and E133 (on alpha-helix 5) (Fig. 55)

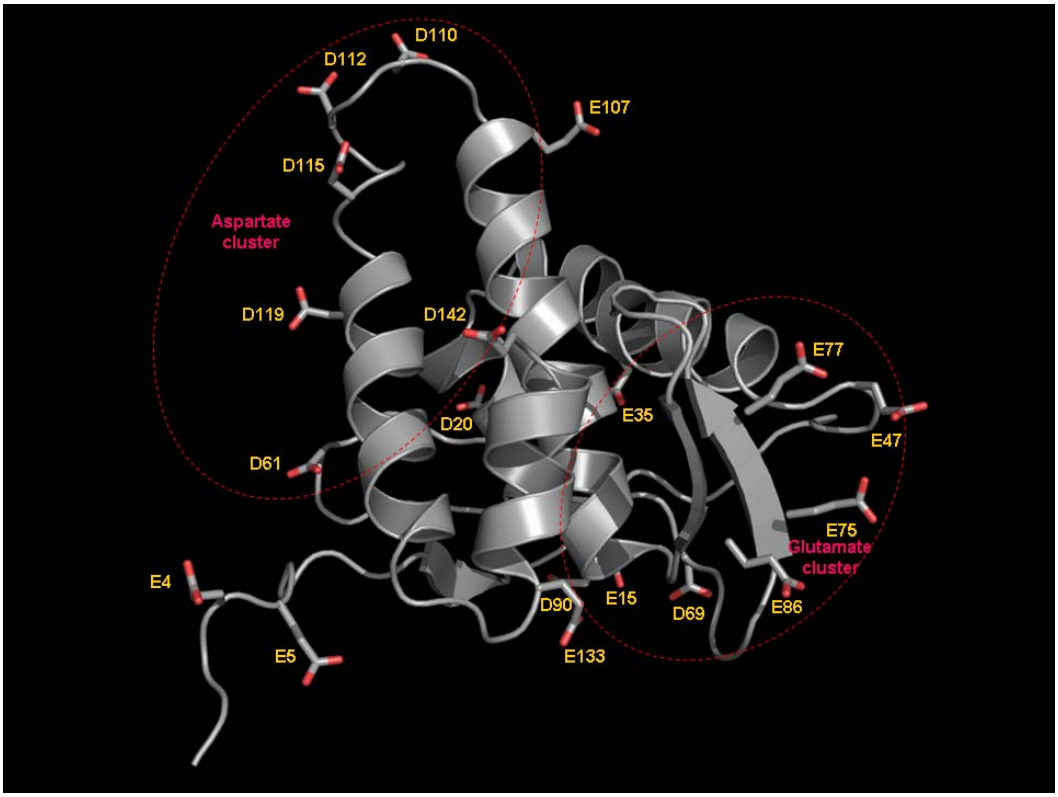
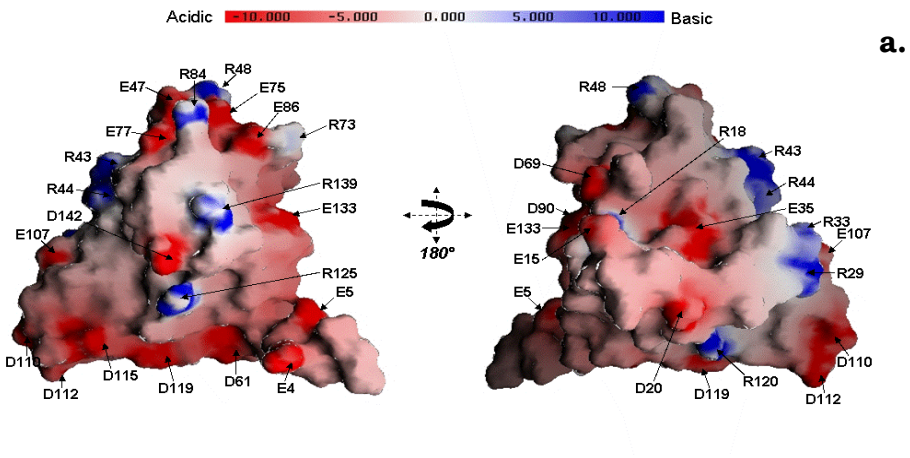
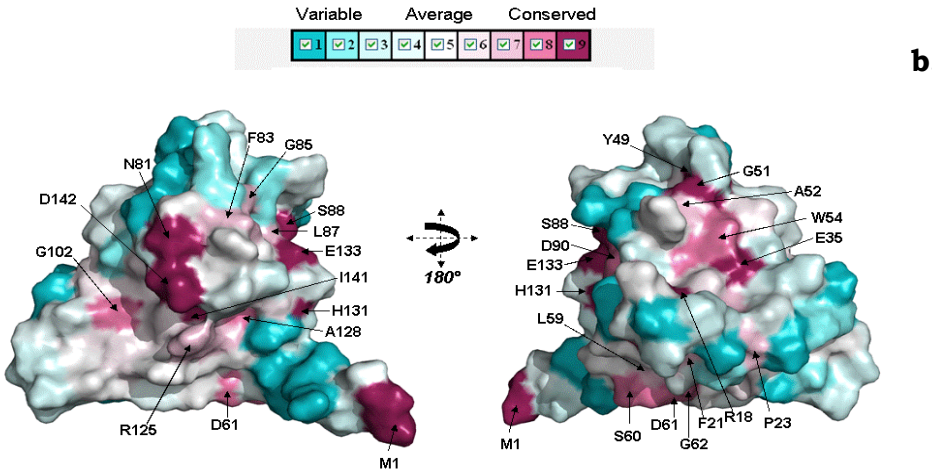


Fig. 55 *Ribbon representation of KlcA highlighting the two negatively charged regions of the protein, the aspartate and glutamate "clusters". Image generated using Pymol*

An electrostatic surface rendition (Fig. 56a) generated by GRASP shows that the surface of KlcA is predominantly negatively charged, consistent with its theoretical pI (4.76). The residues involved in the Aspartate and Glutamate clusters are largely surface exposed. In addition to the five residues present in the 'Glutamate cluster' are residues E15, D69 and D90, which are also located nearby in 3-D space. Intriguingly, a striking linear arrangement of positive charge involving Arginine side-chains is also evident, predominantly on the opposite side of the KlcA protein surface (with respect to its negatively charged clusters). This "Arginine cluster" contains seven residues: R29, R33, R43 and R44 on alpha-helix 2, R48 (in the loop between alpha-helix 2 and beta-strand B; R84 on beta-strand E; and R120 on alpha-helix 4.

Even though three distinct regions of charge are apparent from the electrostatic surface representation, only a handful of charged residues are conserved (Fig. 56b). These include R18, E35, D61, D90, R125, E133 and D142 Additionally, the lipophilic surface rendition generated by MOLCAD (Fig. 56c) reveals two regions of spatially contiguous hydrophobic side-chains, i.e. on the "front-side" (Fig. 56c, left-hand panel) containing F83, L87, I118, Y121, A128, A135, A136, Y138 and I141 and a more prominent hydrophobic region on the "back-side" of the protein (Fig. 56c, right-hand panel) consisting of L19, F21, L22, Y25, P28, L30, M31, M32, A36, Y39, Y53, W54, Y56, L70, L74, I108 and A113. Among these, F21, L22, W54, F83, L87, A128 and I141 are largely conserved and along with other solvent-exposed conserved charged and polar amino acids in their vicinity, could be important for function.


Acidic
-10.000
-5.000
0.000
5.000
10.000
Basic
a.
R84
R48
E47
E75
E86
E77
R73
R43
R44
R139
E133
D142
E107
R125
E5
D110
D115
D112
D119
D61
E4
180°
R48
R18
R43
D69
R44
D90
E35
R33
E107
E133
E15
R29
E5
D110
D20
D119
R120
D112


Variable
Average
Conserved
1
2
3
4
5
6
7
8
9
b.
F83
G85
N81
D142
S88
L87
G102
E133
I141
H131
A128
R125
D61
M1
180°
Y49
G51
A52
S88
W54
D90
E35
E133
H131
L59
M1
S60
D61
F21
R18
P23
G62

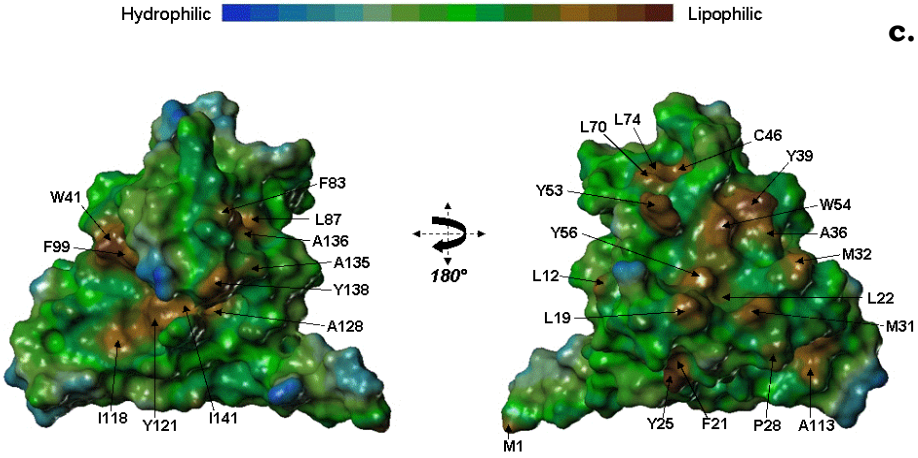

Hydrophilic
Lipophilic
c.
W41
F83
L87
F99
A136
A135
Y138
A128
I118
Y121
I141
180°
L70
L74
C46
Y39
Y53
W54
Y56
A36
M32
L12
L22
L19
M31
M1
Y25
F21
P28
A113

Fig. 56 ***a)*** *Electrostatic representations of the KlcA protein. Two views rotated by 180° along the y-axis of a GRASP (Nicholls et al., 1991) electrostatic surface representation of the KlcA protein. The molecule appears to expose many charged residues, labelled. Negative charge is coloured red and positive charge coloured blue, ranging from -10 kT to +10 kT (k = Boltzmann's constant; T = temperature in Kelvin).* ***b)*** *Sequence conservation mapped onto protein structure and surface. Conservation surface representation based upon the PROMALS3D multiple sequence alignment of KlcA-like sequences where largely and completely conserved residues (labelled in the lower panel, where surface exposed) are shown in lavender/plum/violet colour mapped onto the cartoon secondary structure representation and surface. Images created using PyMol.* ***c)*** *Lipophilic representation of the KlcA protein. Two views rotated by 180° along the y-axis of a MOLCAD (Heiden et al., 1993) lipophilic surface rendition of the protein, where regions of high lipophilicity (hydrophobicity) are coloured brown and labelled and regions of high hydrophilicity are coloured blue.*

Interestingly, in the ^{1}H-^{15}N spectrum of KlcA, two weaker peaks at chemical shifts 72 and 69 ppm are observed. A F_1 ^{1}H-coupled ^{1}H-^{15}N HSQC (Fig. 57) showed that these signals belong to NH_2 moieties. They appear in the spectrum as two outer lines of a triplet with the separation of ca 185 Hz. The ^{15}N chemical shifts are consistent with values reported for the guanidinium moiety of arginine from BioMagResBank (Ulrich *et al*., 2008). These NH_2 groups would normally undergo chemical exchange with water and thus rendered invisible. The fact that they are visible in the spectrum suggests that these are involved in hydrogen bonds. Indeed, by examining the three dimensional structure of the protein, the $N^{\eta 2}$ of R73 (loop, one residue before strand D) and $O^{\varepsilon 2}$ of E86 (strand E) are separated by 2.97 Å, while this distance is 4.47 Å for R48 (loop) and E75

(strand D) (Fig. 58). The former pair meets the criterion (3.5 Å) for H-bonded salt bridge interaction (Kumar & Nussinov, 1999). Although such interactions involving side chains are less common than those of the backbone, they could play a role in maintaining the structural integrity of the protein. The relaxation data indicate that D, E strands are the most flexible on the fast time scale, so an H-bonded salt bridge can stabilise these relatively short beta strands. The second interaction can potentially maintain the relative orientation between the helix 2 and strands D and E, which connect to the rest of the protein via loops.

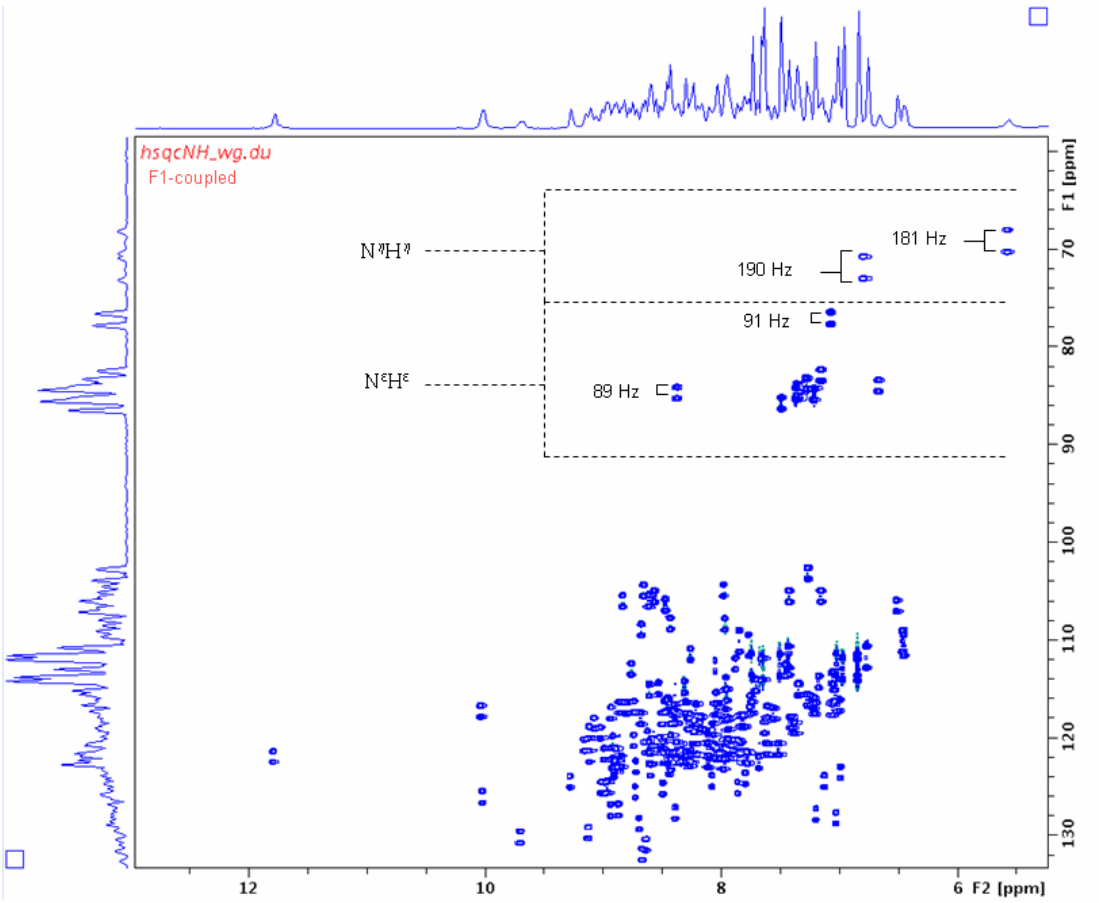


Fig. 57 *A F_1 ^{1}H-coupled ^{1}H-^{15}N HSQC of KlcA (unfolded) showing the $N^{\varepsilon}H^{\varepsilon}$ and $N^{\eta}H^{\eta}$ of arginine residues. Image generated with Topspin NMR.*

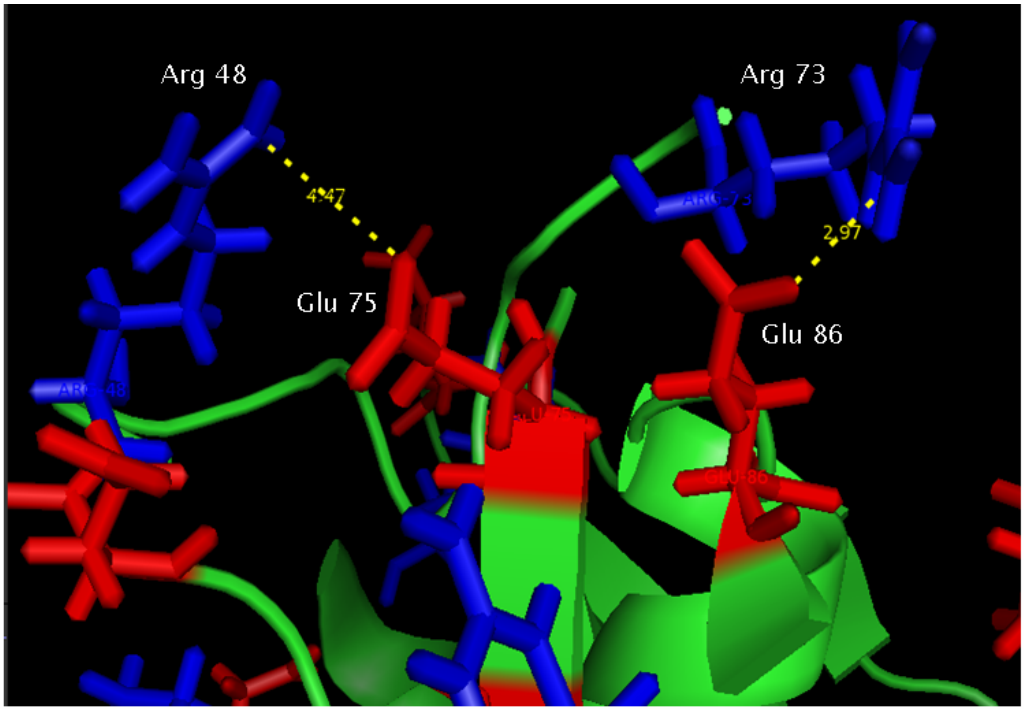
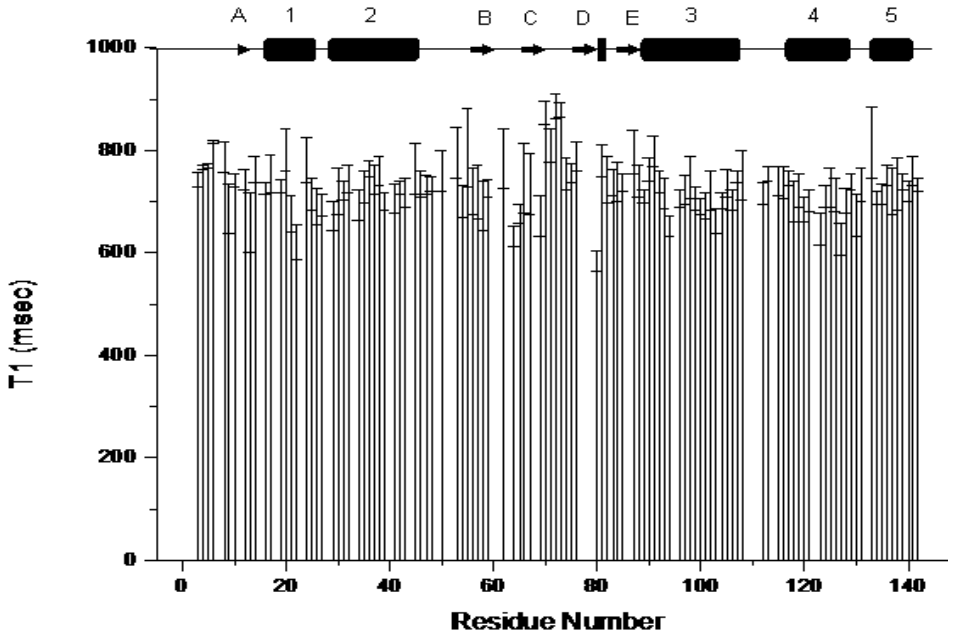


Fig. 58 *Salt Bridges of R48 : E75 and R73 : E86. Image generated with Pymol*

5.3 Backbone dynamics from ^{15}N relaxation data

^{15}N NMR relaxation data are widely used to probe both molecular rotational diffusion and local backbone dynamics of a protein. We have measured ^{15}N longitudinal (T_1) and transverse (T_2) relaxation times and steady state ^{1}H-^{15}N NOEs as detailed in the Methods section (Chapter 3). Plots of T_1, T_2 and heteronuclear NOEs against residue number are shown in Fig 59**.** Because of

partial peak overlap, only 113 of the 137 non-proline residues of the protein were used for the calculation of the relaxation parameters. The residues with higher amplitude of internal motions have simultaneously higher values of T_1 and T_2. By contrast, amide groups with the simultaneously lowest values of T_1 and T_2 values belong to the residues with the most restricted local mobility. The heteronuclear NOE values < 0.68 were used to identify residues of the protein that are dynamic on the ps-ns timescale. As the majority of heteronuclear NOEs are greater than 0.68, the backbone exhibits limited motions on this timescale (Kay *et al.*, 1989). The only exceptions were residues 3 to 6 that exhibited negative NOE values indicating a flexible N-terminus. This correlates well with this region being poorly defined in the structure.



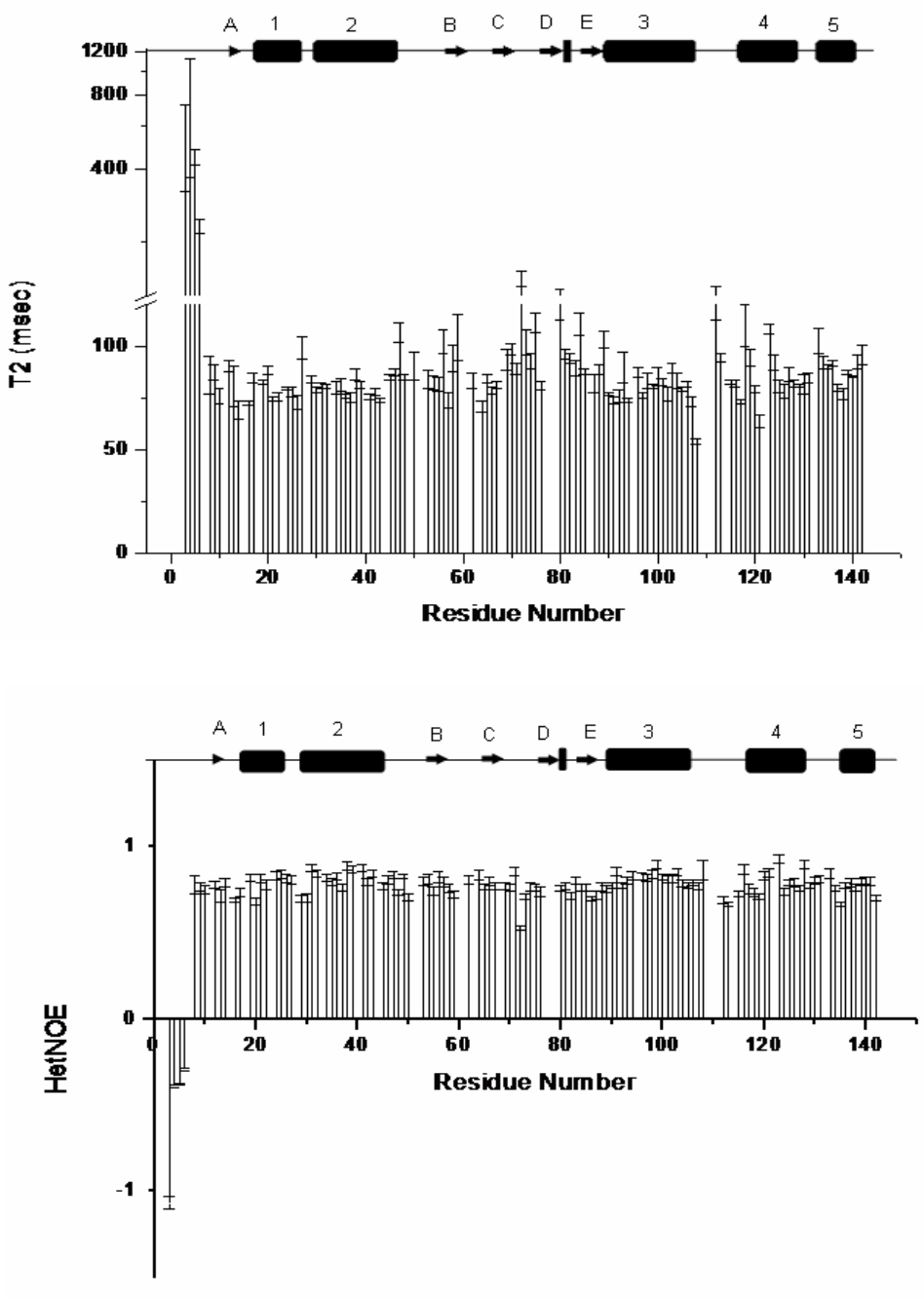


Fig. 59 *Experimental T_1, T_2 and {^1H, ^{15}N} NOE as a function of residue number*

The presence of internal motion and chemical exchange can lead to distortions of the T_1/T_2 ratio, therefore the collection of residues used to determine the diffusion properties of the entire molecule excluded mobile residues or those exhibiting slow exchange broadening. A subset of 97 residues selected as described in the Methods was used in the determination of the diffusion tensor. In this subset the average values of T_1, T_2 and heteronuclar NOEs were 737 and 86 msec respectively. The relative moment of inertia values calculated with the program "pdbinertia" from the Arhtur Palmer Group, Columbia University and found to be: I_z: 1.00 I_y: 0.9858 I_x: 0.8629. I_y and I_x are very close to 1.00, implying that the protein does not deviate significantly from a sphere, having slightly oblate shape.

The program "quadric_diffusion" from the Arthur Palmer Group was used to determine the diffusion tensor. The program fits the relaxation data in an isotropic, axially symmetric and fully anisotropic model using increasing number of fitting parameters. In order to fit the relaxation data to an isotropic model, only one parameter is used, D_{iso}, since $D_z=D_x=D_y$. Fitting the relaxation data to an axially symmetric model requires four parameters: two rotation angles (θ and ϕ) that re-orient the molecule to its principal axis system and two diffusion coefficients $D_{parallel}$ ($D_{\parallel}$) and $D_{perperdicular}$ ($D_{\perp}$). $D_{\parallel}$ corresponds to D_z, and $D_{\perp}$ to $D_x=D_y$. Fitting to a fully anisotropic tensor requires 6 parameters: three rotation angles (θ, ϕ and γ) and three diffusion coefficients (D_z, D_x, D_y) since in anisotropic tumbling $D_z \neq D_x \neq D_y$.

In order to assess the validity of the model, the program calculates the χ^2. The more parameters are used to fit the data, the lower value of χ^2 is expected. The χ^2 is given by:

$$\chi^2 = \sum_{i=1}^{N} \left[\frac{1}{\sigma_i^2} [y_i - y(p)]^2 \right]$$

Where σ^2 is the error associated with the data, y_i is the experimentally derived values, and y(p) is the value calculated from the model, using *n* parameters p= p_1, p_2, …p_n. As more parameters are used to fit the data the χ^2 will naturally decrease. Thus, fitting the data to an axially symmetric model will give a χ^2 value smaller than fitting to an isotropic model. The program quadric_diffusion gave the following results for KlcA:

Type of diffusion	χ^2 value
Isotropic	63293
Axially Symmetric	61977
Anisotropic	61038

The differences in the χ^2 values between isotropic and axially symmetric and axially symmetric *vs*. anisotropic were very small. In order to test whether the difference in the χ^2 values is actually significant, a statistical test called the partial F-test is used. The partial F-test value is derived from the formula:

$$F_{(m-n),N-n} = \frac{(\chi^2_m - \chi^2_n) / (n-m)}{\chi^2_n / (N-n)}$$

In this formula, *N* data points (residues) are initially fit to *m* parameters and then subsequently fit to *n* parameters. Since 92 residues were used for the analysis of relaxation data, the F-test for isotropic vs. axially symmetric fitting will be $F_{(3, 88)}$. Similarly, for axially symmetric vs. anisotropic: $F_{(2, 86)}$. The quadric-diffusion program gave the following partial F-test values: $F_{(3, 88)} = 0.622$ and $F_{(2, 86)} = 0.661$. These F-values derived using a partial F-test with a 95% confidence are statistically non-significant, which implies that adoption of the more complicated diffusion models does not significantly improve the fitting of the data. As a result, the isotropic diffusion tensor with a $D_{iso} = 1.8 \times 10^7\ s^{-1}$ was adopted. The average rotational correlation time (τ_m) calculated using the program "r2r1_tm" from Palmer Group, Columbia University. The τ_m was found to be 8.83 ns ± 0.01, a value that is expected for a protein of this size.

In order to extract microscopic parameters of motion from the relaxation data a model-free approach was used with the help of the program Tensor2 (Dosset *et al*., 2000). In this analysis, the overall and internal motions of the molecule are assumed to be independent. Based on the above analysis of the relaxation data, the isotropic model was selected for further consideration. Relaxation analysis data were fitted using models 1-5, where model 1 fits only the generalised order parameter (S^2); model 2 fits S^2 and the effective internal correlation time (τ_e); model 3 fits S^2 and the chemical exchange term (R_{ex}) and model 5 is an external Lipari-Szabo model, which includes a very fast (S_f^2) and slow (S_s^2) time scales along with τ_e. The program Tensor2 tests the models (1-5) against the experimental relaxation data for each residue selected.

Relaxation data for most residues (73%) were adequately described by the generalised order parameter S^2 (model 1) suggesting an overall rigid molecule. Eleven residues (4Glu, 31Met, 38Val, 41Trp, 47Glu, 56Tyr, 64Phe, 99Phe, 103Gln, 123Phe and 128Ala) did not fit any model. An additional parameter (R_{ex}) was required to fit 17 residues, with residues 14Ala, 16Ala, 91Ala, 107Glu, 108Ile, 117Leu and 121Tyr showing significant R_{ex} values (>2 Hz), suggesting that these N-H sites experience multiple environments interconverting on the micro- to millisecond time scale. The order parameter S^2 as a function of the residue number is shown in Fig. 60**.**

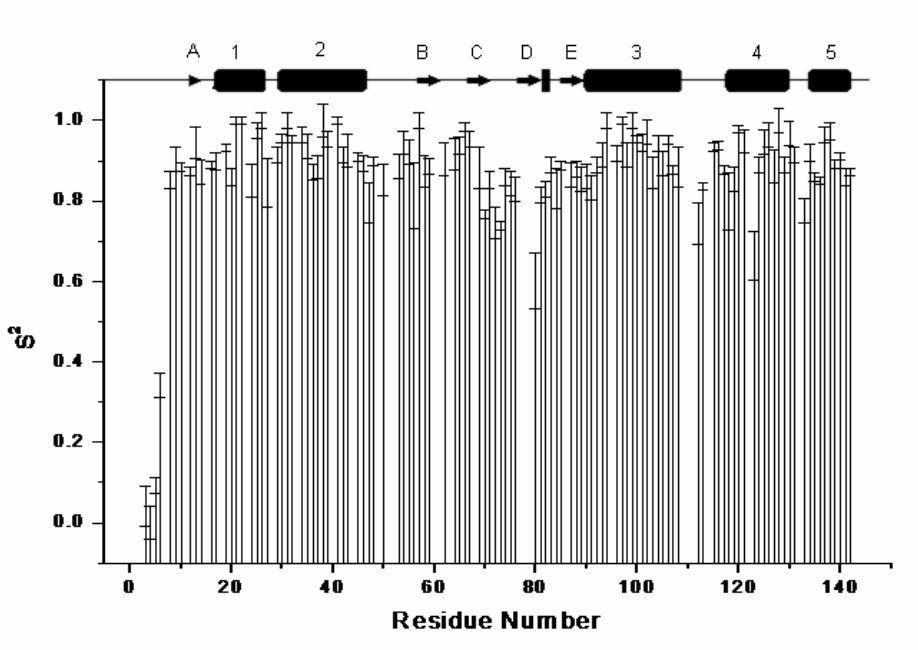


Fig. 60 *S^2 values calculated with the program Tensor2 as a function of residue number*

The order parameter is a measure of the degree of spatial motion of the N-H bond where a value of 0 means that the bond vector has unrestricted motion while 1 implies a completely rigid bond. The distribution of S^2 values visualised on the ribbon structure is shown in Fig. 61**.** This graph shows that the mobility on the nano second time scale is slightly restricted in secondary structure elements when compared to the loops. The largest mobility indicated by the smallest S^2 values (excluding the termini) was observed starting with the residues at the end of strand C up to the beginning of helix 3.

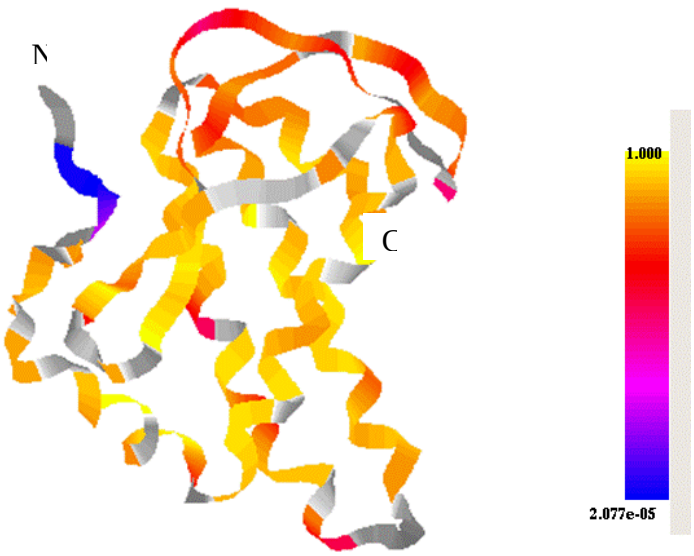
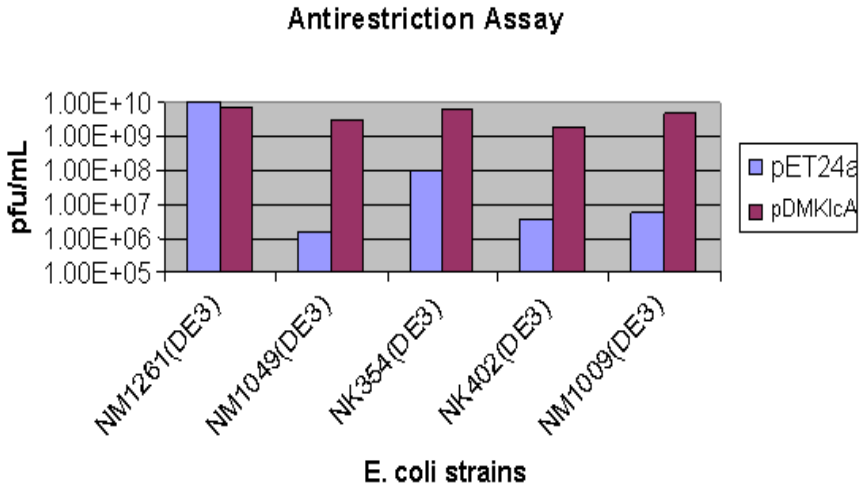


Fig. 61 *The order parameter S^2 visualised on the KlcA ribbon structure (beta-sheets are not shown). Blue denotes high flexibility, whereas residues coloured yellow are the most rigid ones. The image was obtained with Tensor2.*

5.4 Antirestriction activity of KlcA

In the restriction assay, *E. coli* strains harbouring different Type I R-M systems were infected with unmodified, virulent λv.0. phage. A restriction-modification deficient (r^-m^-) strain was also tested as a control (NM1261). For the r^-m^- strain, no difference in plaque formation should be observed in strains harbouring either the KlcA expression construct or vector alone. However, assuming KlcA confers restriction alleviation, r^+m^+ strains harbouring the KlcA expression construct will give a greater number of plaques than r^+m^+ strains harbouring vector alone. The efficiency of plating of unmodified test phage λv.0. was determined by measuring the ratio of phage titre on the restricting strains (r^+m^+) to the titre on the nonrestricting strain (r^-m^-). The ratio of the two e.o.p. measurements shows the restriction alleviation potential of KlcA (Table 3). KlcA expression provided between 1 and 3 orders of magnitude of protection to the unmodified phage. The greatest restriction alleviation was for the Type IA system and the lowest for the Type IB.



Host Strain	Genotype	pET24a (pfu/mL)	pDMKlcA (pfu/mL)	eop pDMKlcA-	eop pDMKlcA+	RA=pDMKlcA+/pDMKlcA-
NM1261(DE3)	r-m-	9.40E+09	7.93E+09			
NM1049(DE3)	EcoKI-1A	1.59E+06	3.26E+09	1.69E-04	4.11E-01	2.43E+03
NK354(DE3)	EcoAI-1B	1.02E+08	6.60E+09	1.09E-02	0.83	76
NK402(DE3)	EcoR124I	3.40E+06	1.80E+09	3.62E-04	0.227	628
NM1009(DE3)	StySBLI	5.90E+06	4.82E+09	6.28E-04	0.608	969

Table 3: *KlcA antirestriction against Type IA-ID Restriction-Modification systems as determined in vivo*

In the modification assay, phage were recovered from plates of restriction-modification proficient strains harbouring the KlcA expression construct or vector only. When these phage were subsequently used to infect the same strains two possible outcomes could be envisaged. If KlcA was capable of inhibiting modification, phage DNA would not be modified and therefore

remain vulnerable to restriction. This would result in a reduced number of plaque forming units. If, however, KlcA was unable to inhibit modification, the phage DNA would be modified and rendered immune to restriction. Thus, the phage would survive and the number of plaque forming units would increase. The modification test revealed that KlcA cannot inhibit modification by any of the four Type I systems. A similar result was reported for the ArdB from pKM101, which showed efficient antirestriction activity of Type I enzymes but no antimodification activity (Belogurov *et al.*, 1993).

5.5 Expression of KlcA is not lethal to *E. coli*

It has been shown that unregulated expression of the *kil* genes leads to killing of the host cells. The cells become elongated with distorted outer membranes and macromolecular synthesis ceases (Saltman *et al*., 1991). Interestingly, previous studies of the *kilC* operon in plasmid RK2 showed that the *klcB* gene is not required for the Kil^+ phenotype and that *klcA* is responsible for the lethality of *E. coli* host cells in the absence of *korA* and *korC* (Larsen & Figurski, 1994). However, as shown in this work, it was possible to overexpress the *klcA* gene in a pET24 vector in the absence of *kor* determinants with no apparent harmful effects. Furthermore, *in vivo* antirestriction experiments performed on the KlcA from RK2 failed to identify any signs of antirestriction function, whereas the KlcA from *B. pertussis* clearly functions as an antirestriction protein as it showed efficient antirestriction activity against all four families of Type I restriction systems. Despite these differences, the two KlcAs are homologous (71% identity).

5.6 KlcA does not bind to MTase

Biophysical characterisation using isothermal titration calorimetry (ITC) and fluorescence spectroscopy performed on two other antirestriction proteins, Ocr and orf18, have shown clear interaction of the MTase subunit of EcoKI with these proteins (Serfiotis-Mitsa *et al*., 2008; Atanasiu *et al*., 2001). Both antirestriction proteins exhibited strong binding with the methyltransferase. However, ITC experiments performed at different temperatures ranging from 10-25°C showed no apparent binding of MTase to KlcA. This result agrees with the *in vivo* modification assay where no modification of the phage DNA was observed.

5.7 KlcA antirestriction function does not involve the ClpXP protease

It has been shown that ClpXP-dependent regulation of the endonuclease activity enables bacteria that acquire unmodified chromosomal target sequences to survive. The restriction subunit of the nuclease (HsdR) is degraded in the presence of ClpXP (Makovets *et al*., 1999). ClpXP is a two-component system consisting of an activator ATPase (ClpX), which recognises, unfolds and translocates the target to be degraded, and a proteolytic core (ClpP), which degrades the substrate (Gottesman *et al*., 1993). When cells are under stress after prolonged exposure to treatments that damage DNA, such as exposure to 2-aminopurine, the ClpXP-dependent pathway is essential in preventing EcoKI from causing DNA damage.

Since no direct interaction of KlcA with the EcoKI was observed *in vitro* (Dr. Gareth Roberts, Edinburgh University, personal communication), the hypothesis that the antirestriction activity of KlcA may be dependent on the ClpXP pathway was tested. A ClpX$^-$ mutant (strain NM1041(DE3), kind gift of Noreen Murray) was plated in the presence and absence of 40 µg/ml 2-AP. The ClpX$^-$ mutant was also transformed with a pET-20b vector carrying the *klcA* gene and tested under the same conditions. The *klcA* was subcloned into the NdeI and HindIII of the polylinker region of pET20-b. The use of a different vector other than pET24-a was necessary for this assay because the ClpX- mutant is kanamycin sensitive. pET20-b provides the necessary second antibiotic selection (ampicillin). As expected, the ClpX$^-$ with no antirestriction determinant present, grew well in the absence of 2-AP, but no growth was observed in the presence of 40 µg/ml 2-AP. (Fig. 62) By contrast, the ClpX$^-$ carrying the *klcA* grew equally well in both the presence and absence of 2-AP. This finding suggests that KlcA is able to block EcoKI *in vivo* without involving ClpX.

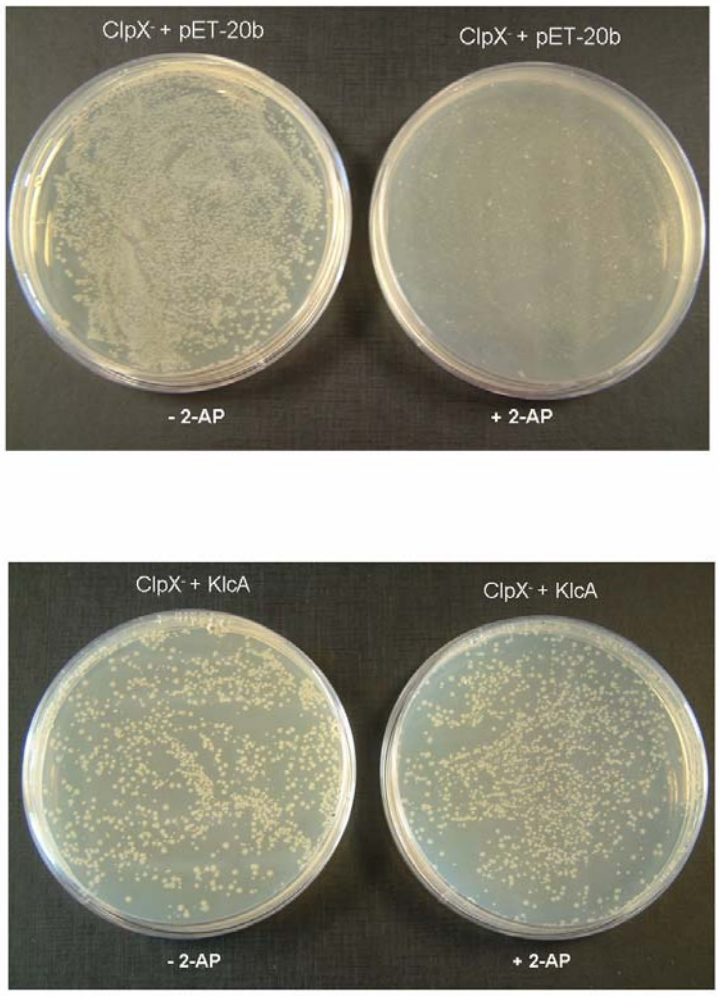


Fig. 62 *The presence of 2-AP hindered the growth of the ClpX- mutant containing only vector (pET20-b) (**a**), whereas in the presence of pET20-b +klcA, growth of the mutant was equal on the plates with or without 2-AP (**b**).*

5.8 Blue Native (BN)/SDS-PAGE

BN/SDS-PAGE is a 2D PAGE technique technique for the identification of protein complexes developed in the early 90's (Schagger & von Jagow, 1991) in order to investigate the individual components of protein complexes forming the respiratory chain of various organisms. The technique has been used primarily to investigate plant proteomes (Eubel *et al*., 2004), however, it has been used to study multiprotein complexes in human HeLa cell cultures (Camancho-Carvajal *et al*., 2004), in yeast (Grandier-Vazeille & Guérin, 1996) and in *E. coli* (Barrett *et al*., 2005).

Blue native polyacrylamide gel electrophoresis is a charge-shift method. In this method, negative charges are not attached to the proteins by detergents such as in the SDS-PAGE, but by addition and binding of the negatively charged Coomassie Blue G250. When the negatively-charged dye is added to the native proteins it induces a charge-shift as it binds to surface hydrophobic patches on proteins. The electrophoretic motility of the proteins is determined by the negative charge due to the binding of the dye making even very basic proteins migrate to the anode. A major difference between BN-PAGE and SDS-PAGE is that the Coomassie dye binds to the surface of the proteins and not the interior and as a consequence, the dye does not denature the proteins.

After the cells are lysed by sonication, the mixture of different protein complexes is separated by BN-PAGE based on molecular weight. The lane of the blue native gel is excised and subjected to a denaturation solution containing 1% SDS and 1% 2-mercaptoethanol. Following denaturation, the native protein

complexes dissociate to their constituent polypeptides. The subunits of the protein complexes remain in position until they are forced electrophoretically into the second dimension gel, this time an SDS polyacrylamide gel. The subunits of the protein complex will form a vertical row on the second dimension gel (Fig. 63).

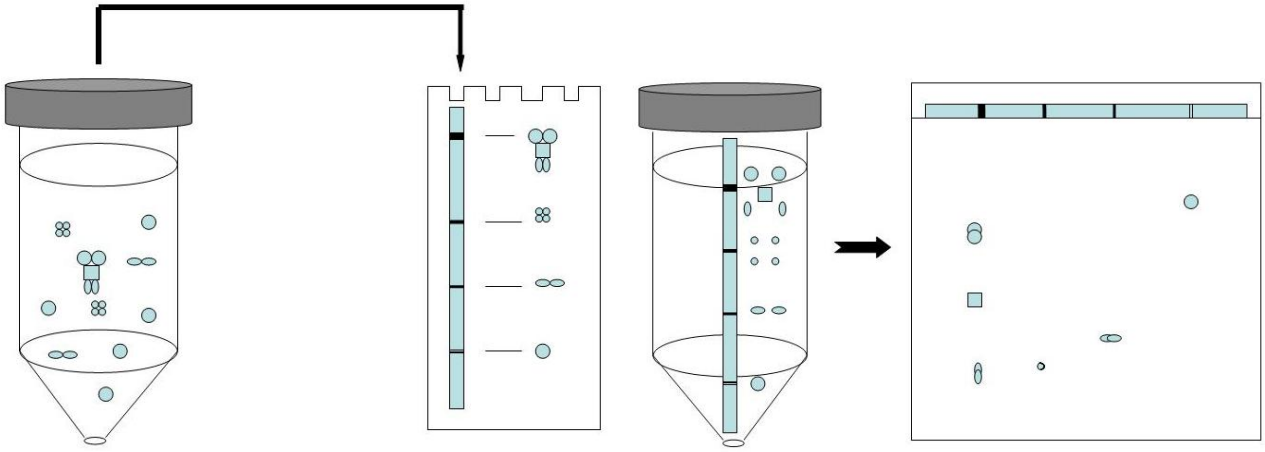


Fig. 63 *The working principle of BN/SDS-PAGE. The solubilised sample is run on a blue native polyacrylamide gel. The non-denaturing condition in the gel allows the various protein complexes to remain intact. The gel slice form the BN polyacrylamide gel is cut and incubated in a denaturation solution (1% SDS and 1% 2-mercaptoethanol) and then placed on top of a SDS polyacrylamide gel. The dissociated complexes now run in the second dimension. Subunits of each protein complex form a vertical column on the second dimension gel. (Image adapted from Eubel et al., 2005).*

A sample of BL21(DE3) cells containing the pDMKlcA vector was induced and subsequently sonicated as described in the Methods. The supernatant was

loaded on a BN gel and after the run the gel slice was excided treated with SDS and 2-meractoethanol and subsequently placed horizontally on top of an SDS gel. The SDS gel along with the slice of the BN gel are shown in Fig. 64.

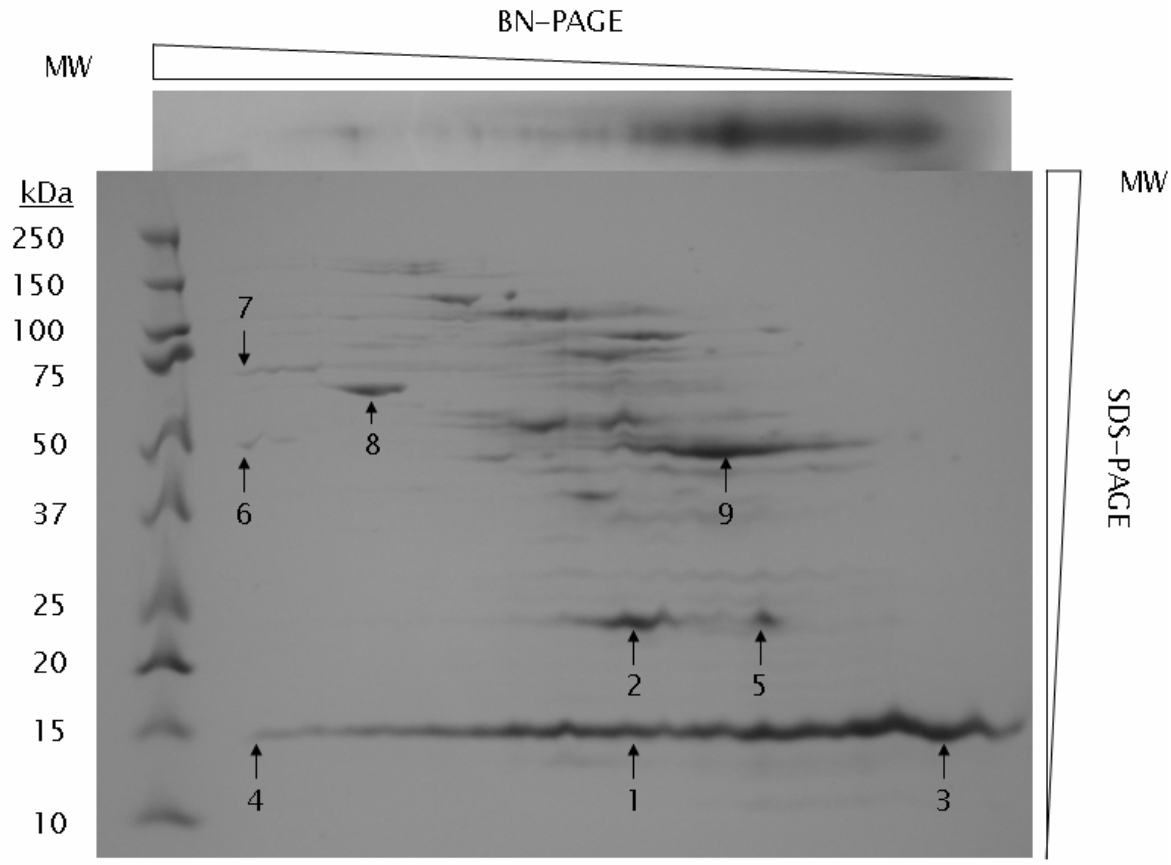


Fig. 64 *BN/SDS-PAGE of cell lysate containing overexpressed KlcA. The BN strip was placed horizontally on top of an SDS polyacrylamide gel with the high to low molecular weight from left to right. The bands with numbers correspond to bands that were analysed by MALDI-TOF-MS.*

All protein spots which are situated vertically below each other after BN/SDS-PAGE are potential subunits of protein complexes. Spots which are located side by side in a horizontal row potentially represent the same protein which is localised in protein complexes of different molecular mass. Caution, however, has to employed because it can be the case that more than one protein complex can separate to the same molecular mass during BN-PAGE and more than one protein subunit may be resolved during SDS-PAGE to the same molecular mass.

A streak of bands corresponding to the expected molecular weight of KlcA (15.6 kDa) appeared on the SDS-PAGE. Some bands appeared in a vertical position of that streak, however, most of the bands on the gel were grouped in a cluster towards the middle of the gel making it difficult to identify which bands had run vertically with the KlcA bands. The fact that a horizontal streak of KlcA bands appeared in the SDS-PAGE implied that the protein may be associated with a very large complex (thus a band of KlcA appearing at position 4, and "free" KlcA at position 3).

The molecular mass of a complex is increasing during assembly of the structural subunits of this complex, thus in the case of a "supercomplex", the analysis of the protein pattern resolves the direction of the stepwise subunit assembly from the lower towards the higher molecular mass of the protein complexes from the right side of the gel towards the left side (Fig. 65). The subunit composition of a protein complex is directly presented in the 2D BN/SDS-PAGE as proteins are arranged in a vertical line. In contrast, the assembly of the protein subunits into one supercomplex can be analysed from the horizontal appearance of one

subunit together with other partner protein subunits at a specific molecular mass position of the corresponding protein precomplexes.

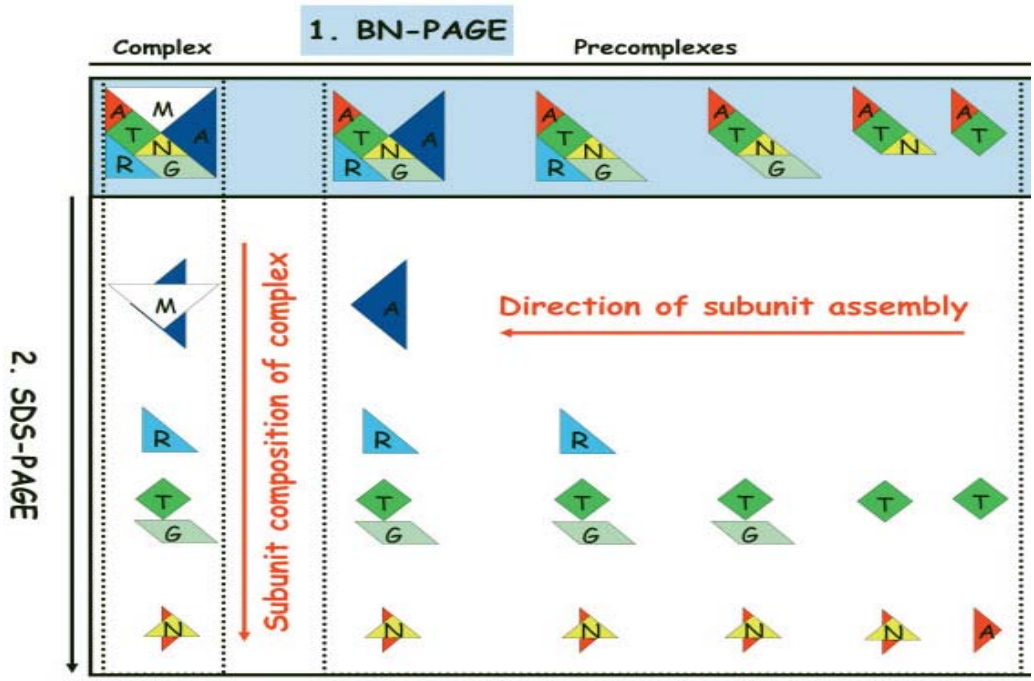


Fig. 65 *When the protein of interest binds to a large complex that involves many protein partners, the analysis of the horizontal "streak" on the SDS-PAGE resolves the direction of the stepwise subunit assembly from the lower to the higher molecular mass of the protein complexes (image reprinted from Reisinger & Eichaker, 2007).*

Two bands (6 and 7) that run vertically with the KlcA band (4) as well as a few other intense bands from the gel were digested with trypsin and analysed by MALDI-TOF-MS as described in the Methods. In order to confirm that the bands in the streak were all KlcA, more than one band was analysed (4, 1, 3). After the MS spectra were collected, peptide mass fingerprinting was performed

with the online server MASCOT. All nine bands came with a significant score and the identities of the bands are summarised in Table 4**.**

Band Number	**Protein Identified**	**Function**
1	KlcA from *B. pertussis*	Antirestriction
2	50 S ribosomal protein L4	Part of the 50 S ribosomal subunit
3	KlcA from *B. pertussis*	Antirestriction
4	KlcA from *B. pertussis*	Antirestriction
5	50 S ribosomal protein L4	Part of the 50 S ribosomal subunit
6	Dihydrolipoamide succinyltransferase	Part of the 2-oxoglutarate dehydrogenase complex, involved in oxidative decarboxylation of2-oxoglutarate, part of Krebs cycle
7	30 S ribosomal protein S1	Part of the 30 S ribosomal subunit
8	Chaperonin 1	Involved in the folding process of nascent proteins
9	Elongation factor Tu 1	Elongates the protein chain during protein synthesis

Table 4. *Results of the peptide mass search (PMS) of selected bands from the BN/SDS-PAGE using MASCOT.*

PMS confirmed that the streak of bands 1, 3 and 4 belonged to the KlcA protein. The bands 6 and 7 which run vertically with band 4 do not belong to the same complex but there was rather a co-migration of proteins from different complexes (2-oxoglutarate dehydrogenase complex and 30 S ribosomal subunit) making it difficult to identify which could be forming a complex with KlcA. The rest of the bands revealed proteins that are involved in the protein machinery of *E. coli* (ribosome, elongation factor, chaperonin). The strong appearance of bands on the gel belonging to proteins involved in the protein machinery of the cell could simply be due to the fact that the lysate used for the experiment contained an overexpressed protein, thus more ribosomes, elongation factors etc. are necessary for the overexpression to take place. As a result, no other assumptions can be made about the result of this experiment apart from the fact that KlcA appeared to be associating with a complex of very high molecular mass. However, co-migration of proteins on BN gels is not a final proof of native association as physically distinct complexes may have similar molecular masses and appear in the same protein bands. One other thing to take into consideration for this experiment is the concentration of KlcA versus the concentration of the rest of the cellular proteins. Overproduction of the protein of interest may cause the protein to adopt a "sticky" nature that favours interactions that would not be normally observed at endogenous levels of abundance. A complementary experiment to the BN/SDS-PAGE technique is the pull-down assay where co-immunoprecipitation of the protein of interest along with the rest of the proteins in the complex may be achieved.

5.9 c-myc tagged KlcA co-IP assay

Identifying novel proteins that could potentially interact with KlcA was also tested by the co-imunoprecipitation (co-IP) method. Co-IP is a method that utilizes antibody-mediated precipitation of antigens to analyse antigen-associated proteins. When an antibody against the protein of interest is not available, co-IP can still be performed using a tagged version of the protein. There are five steps involved in a co-IP process: preparation of protein extract, coupling of the antibody to the antigen, isolation of protein complexes, analysis of the proteins involved in the complexes by SDS-PAGE and identification of the proteins involved in the complex by MALDI-TOF-MS.

A rapid method for capturing an antigen on the antibody-bead matrix is to mix the antigen solution with the beads and rotate or rock the slurry, allowing maximum contact between the antigen and the immobilized antibody. After the binding reaction has been completed, the slurry is passed into a column for collection of the beads, and then the antigen is eluted. Epitope tags are useful for labeling proteins since their small size (5-15 amino acids) is less likely to affect the protein's function compared to larger tags such as the GFP or GST tag for example. For this study, the c-myc epitope was used which is a 10 amino acid sequence (EQKLISEEDL) derived from the human *c-myc* gene. This polypeptide is specifically recognized by the monoclonal antibody Mycl-9E10 (Evan *et al.*, 1985). This allows for sensitive detection and immunoprecipitation of expressed proteins with the anti-c-myc antibody.

Since the region of the protein which is responsible for antirestriction was not known, two constructs were created in case one of the constructs interfered with antirestriction activity: KlcA tagged with c-myc at the N-terminus and KlcA tagged with c-myc at the C-terminus as follows:

N-terminus c-myc tagged KlcA:

EQKLISEEDLMNTEEQPVTASLVAEAQRLDFLPTYFGPRLMMRGEALVYAWMRRLCERYNGAYWHYYALSD
GGFYMAPDLAGRLEIEVNGNGFRGELSADAAGIVATLFALGQLAAEIADTDAADALIDRYHFLRGFAAGHP
EAAAIYRAID

C-terminus c-myc tagged KlcA:

MNTEEQPVTASLVAEAQRLDFLPTYFGPRLMMRGEALVYAWMRRLCERYNGAYWHYYALS
DGGFYMAPDLAGRLEIEVNGNGFRGELSADAAGIVATLFALGQLAAEIADTDAADALIDR
YHFLRGFAAGHPEAAAIYRAIDEQKLISEEDL

To assess if the required recombinant was successful, the DNA was cleaved with NdeI and HindIII in order to release the insert from the vector. An example of such a digestion is shown in Fig. 66. The construct was further confirmed by nucleotide sequencing.

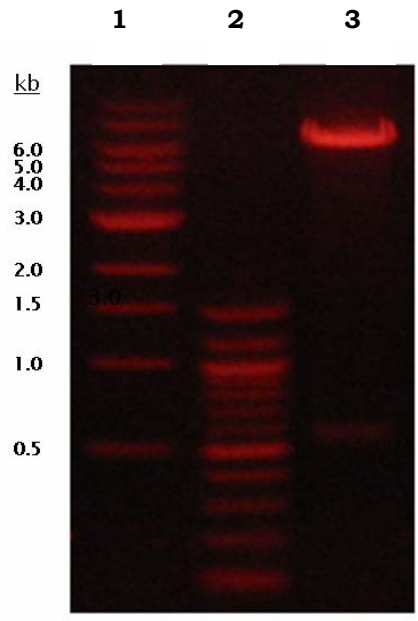


Fig. 66 *The sequence coding for the N-terminus c-myc tagged KlcA was cloned between NdeI and HindIII restriction sites of the pET24-a vector. To test whether ligation was successful, a digestion was performed with NdeI and HindIII enzymes in order to release the insert. Lane 1: 1 kb DNA ladder, lane 2: 100 bp DNA ladder Lane 3: digestion of N-terminus c-myc tagged KlcA construct. A band corresponding to linearised vector (~5.3 kb) and a band corresponding to the insert (~500 bp) can be seen.*

After the two constructs were made, they were tested for antirestriction activity. The N-terminus c-myc tagged KlcA showed the same antirestriction activity as the wild type KlcA, whereas the C-terminus tagged KlcA showed almost no antirestriction, similar to the pET24-a vector alone. The fact that no antirestriction was observed for the C-terminus c-myc tagged KlcA did not necessarily mean that the C-terminus was responsible for antirestriction. In order to assess whether the constructs were expressed in the cell, an induction was set of both constructs (Fig. 67).

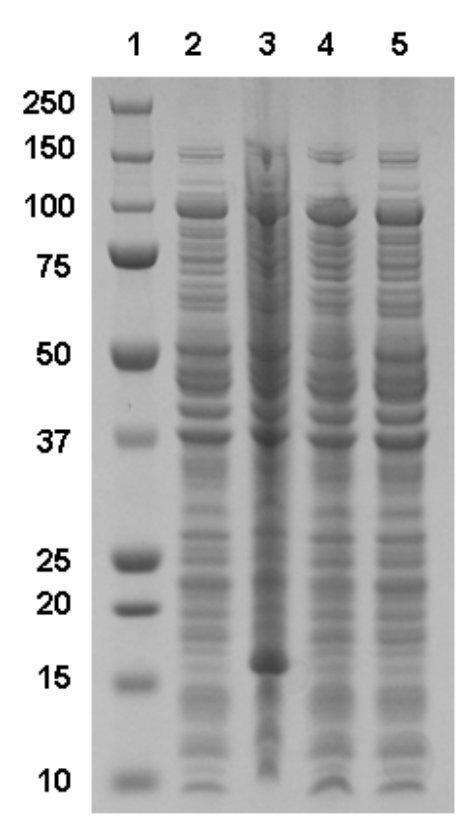


Fig. 67 *Overproduction of N-terminus and C-terminus -myc- tagged KlcA. Lane 1: Protein marker; Lane 2: N-terminus-myc-tagged-KlcA uninduced; Lane 3: N-terminus-myc-tagged-KlcA after 3 hr induction; Lane 4: C-terminus-myc-tagged-KlcA uninduced; Lane 5: C-terminus-myc-tagged-KlcA after 3 hr induction*

Very good expression was observed for the N-terminus c-myc tagged KlcA, whereas no expression can be seen from the gel for the C-terminus c-myc tagged KlcA, suggesting that the protein may not have been able to fold properly in the cell and it was degraded by the cell's proteases.

The N-terminus c-myc tagged KlcA construct was chosen for the co-IP assay. A high affinity anti-c-myc antibody coupled agarose provided in the Profound c-myc tag IP/co-IP kit (Pierce) was used in order to co-purify the tagged KlcA and interacting proteins. As a positive control a c-myc-tagged glutathionine S-transferase was used (Pierce provided), and as a negative control, a lysate containing no tagged KlcA. The result of the co-IP can be seen in Fig. 68.

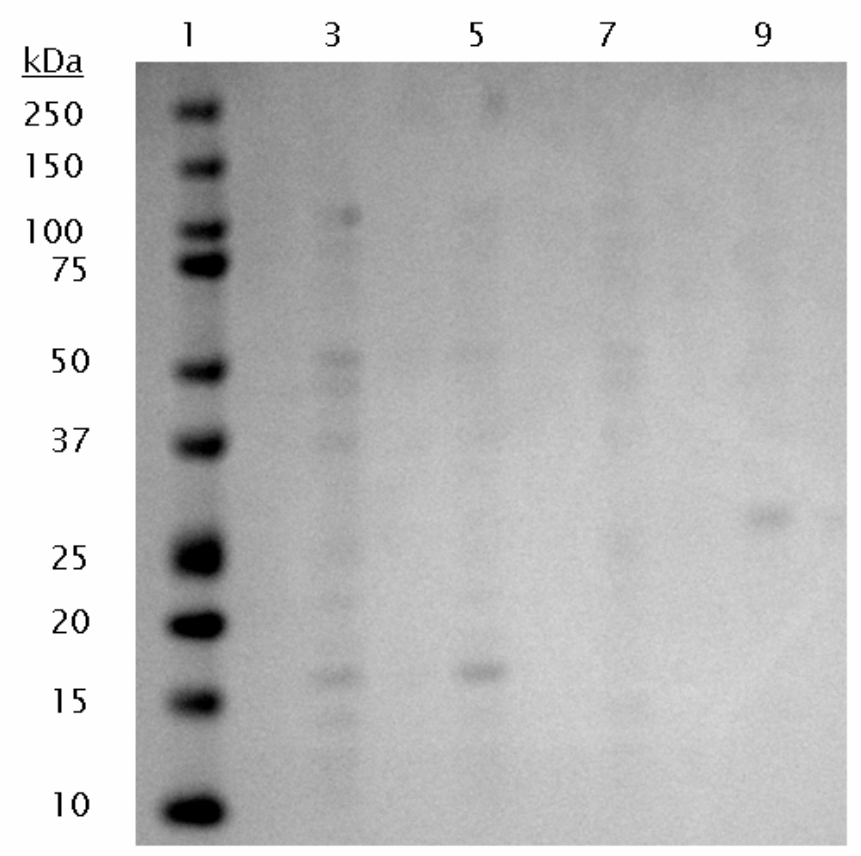


Fig. 68 *An SDS-PAGE showing the result of the co-IP assay. Lane 1: protein marker; lane 3: elution of proteins from sample containing uninduced N-terminus myc tagged KlcA; lane 5: elution of proteins from sample containing induced N-terminus myc tagged KlcA; lane7: elution of proteins from sample containing only vector pEt24-a; lane 9: positive control (c-myc-GST protein)*

For this experiment, a 3 gr pellet was resuspended in only 1 ml of TBS buffer in order to have a lysate as concentrated as possible. Even so, the losses from the column have been so substantial, that only faint bands could be seen on the SDS-PAGE stained with Coomassie blue. The experiment was repeated several times with different incubation times (from 2 hr to overnight) in case longer incubation times would help the binding to the epitope to the antibody, but no difference was observed.

The major problem, however, was the amount of non-specific binding. A band containing a potential binding partner can be masked in the gel by the appearance of non-specific binding. Since no resin is totally inert, several proteins bind to it confusing the interpretation of results. A preclearing step was used where the lysates were first mixed with mouse IgG agarose in order to reduce the background that is due to adhesion of sample components to resin.

For the elution step, an elution buffer (Pierce provided) with a very low pH (2.8) was used instead of boiling the resin in a Laemli-buffer. This precaution prevents contamination of the sample by non-immobilised antibody fragments. Furthermore, extra washing steps were performed.

Ideally, the washing step should break all nonspecific interactions while preserving the specific interaction between antibody and antigen. Increasing the number of washing steps could be one solution as the washing buffer contains mild detergents that can help break nonspecific interactions. Also, increasing the stringency of the buffer can help reduce the background. The TBS buffer used for this experiment is a medium stringency buffer. A higher stringency buffer such as RIPA buffer could be used, however, because this buffer is partly denaturing due to the anionic detergent SDS, it was not considered ideal for co-IP experiments. The non-specific binding, however, persisted and the same bands could be seen on lane 7 where there was no KlcA present as well as on lanes 3 and 5. Also, some non-specific binding can be seen on lane 9 where the c-myc-tagged GST lysate was loaded.

5.10 Conclusion for KlcA experiments

In this study, the solution structure of KlcA, an antirestriction protein belonging to the ArdB family, was solved, and its antirestriction function confirmed. The structure of KlcA is the first ArdB structure reported, and adopts a novel fold. Sequence comparison revealed 43 non-redundant related sequences. Although three distinct regions of charge are apparent from the electrostatic surface representation, only a handful of charged residues are conserved between the related sequences. These conserved residues, as well as residues that display potential chemical exchange and/or higher mobility identified during the relaxation analysis, will be the target for future mutational studies. The antirestriction proteins, such as Ocr (Walkinshaw *et al.*, 2002) and ArdA from

Orf18 of Tn*916* of *Enterococcus faecalis*, have been shown to comprise elongated structures that bind to the methyltransferase subunit of the endonuclease, thereby rendering the nuclease incapable of digesting DNA or methylating the DNA substrate. The structure of KlcA on the other hand, appears to be a rigid globular molecule. Genetic analysis shows that KlcA displays antirestriction but not antimodification activity. However, the antirestriction function of KlcA has only been demonstrated *in vivo* but not *in vitro*. Furthermore, there is no apparent interaction of KlcA with the ClpXP protease. Because no direct binding of KlcA to EcoKI was observed *in vitro*, the mechanism of the endonuclease blocking remains unclear. It is possible KlcA needs another protein partner, or simply a cofactor, for antirestriction to take place. Attempts to find the possible binding partner with BN/SDS-PAGE and co-IP precipitation experiments yielded inconclusive results.

Chapter 6: Final remarks and future work

6.1 Final remarks

The present work dealt with the characterisation of two antirestriction proteins, an ArdA and an ArdB homologue, KlcA. Even though the two proteins appear to have in common the antirestriction function, they differ in many respects. The structures of Ocr and ArdA revealed that these molecules belong to DNA mimics with a characteristic distribution of negatively charged residues imitating distribution of negatively charged phosphate groups along the DNA duplex. KlcA, on the other hand, is a globular protein, less acidic than ArdA and with no characteristic distribution of negative residues. Furthermore, no antimodification activity was observed for KlcA, a characteristic of ArdB proteins. A similarity of ArdA and KlcA however, is noteworthy. It has been shown that the closely linked *ardA* and *ardB* genes on pKM101 are controlled by conserved upstream elements containing a strong promoter. These promoters are regulated by two plasmid proteins, ArdK and ArdR, with at least ArdK acting as a repressor controlling antirestriction gene expression (Delver & Belogurov, 1997). Could this kind of regulation be linked with the *korA* and *korC* regulation of KlcA? Moreover, unlike *ardA* and *ardB* genes, *klcA* is located outside the plasmid leading region. If immediate transcription of the antirestriction genes is necessary for the efficient transfer of the plasmid to a new cell, then how *klcA* genes manage to achieve that?

Future work needs to be focused on the identification of the protein partner(s) of KlcA. This identification will help in the elucidation of its antirestriction mode of action and may give partial answers to the question above.

6.2 Future work

6.2.1 BN/SDS-PAGE

In order test for any "sticky" behaviour of the overexpressed KlcA in the 2D BN-PAGE/SDS-PAGE experiment, a mutant KlcA should be used which has a mutation that abolishes all antirestriction activity and thus no binding with the suspected partners would occur in that case. Furthermore, a BN-PAGE/Western can be done where a blue native gel of the lysate of the uninduced N-terminus – myc-tagged KlcA is followed by a western blot. The significant problem of Coomassie transfer to the membrane (since Coomassie is negatively charged) of a blue native gel can be overcome by changing the cathode dye containing Coomassie blue at about 2/3 of the run by cathode buffer without dye in order to reduce the amount of dye in the gel (the removal of excess dye also helps in protein transfer since the dye tends to migrate first on the membrane blocking some proteins from being transferred), and also destain the gel prior to transfer. The appearance of the tagged KlcA in the lysate can be detected on the membrane by applying the anti-c-myc antibody.

6.2.2 Far-Western Blotting

A different type of western blotting, far-western blotting involves the separation of total cell lysates on an SDS-PAGE, transfer of the protein on a membrane, and then the proteins are first denatured and then renatured with different concentrations of guanidine hydrochloride. The idea is that the denatured proteins after separation in the SDS-PAGE gel, must be renatured in order to adopt their native state and be able for binding. After renaturation of the proteins, the "bait" protein (in this case the N-terminus-myc-tagged KlcA) is applied on the membrane. If the bait protein is involved in a protein complex, it will bind to the relevant proteins on the membrane. The membrane is then blocked as in a normal western and the antibody for the epitope tag is applied in order to visualise the bands where the binding took place.

6.2.3 *In vivo* crosslinking

To stabilize transient complexes and weak interactions for analysis by Co-IP, or BN-PAGE/Western blot, *in vivo* crosslinking can be applied where the protein interaction is "frozen" in the cell. Traditionally, transient interactions were studied using chemical crosslinking with bifunctional NHS-ester crosslinkers. A recent advancement in protein crosslinking involves the incorporation of photoactivatable amino acids that act as zero length crosslinkers (Suchanek *et al*., 2005). Two such amino acids, photo-leucine and photo-methionine containing diazirine rings for photoactivation can substitute for their respective amino acids in endogenously expressed proteins. Photoactivation of the

diazirine rings by UV light creates reactive carbene intermediates that irreversibly crosslink proteins within protein interaction domains containing methionine and/or leucine with a zero-length bond. Photoreactive amino acids do not alter the cell's metabolism and are essentially non-invasive. More importantly, as zero-length linkers they are very specific.

6.2.4 Yeast-two-hybrid method

Novel protein-protein interactions can be easily selected from a pool of potential interaction partners (e.g. a cDNA expression library). Genetic screening systems yield not only information on the interaction itself but also directly provide the cDNA encoding the novel interaction partner. The most widely used methodology of this type is the yeast two hybrid system (Fields & Song, 1989). In the yeast-two-hybrid system protein interactions are detected by their ability to reconstitute the function of a split transcriptional activator complex. The transcription factor can be divided in two domains: a DNA-binding domain and a transcriptional activation domain. These two domains must be in close proximity to each other in order for the transcription factor to active. The sequence coding for the protein of interest is cloned into a vector coding for the isolated DNA-binding domain creating a fusion protein that will act as bait in search for interacting protein partners. An *E. coli* cDNA library must then be created and cloned into a cDNA encoding the isolated transcriptional activation domain of the transcription factor. In this way, a large number of fusion proteins is created that will act as the "prey". The plasmids that are created contain the fusion prey proteins and the fusion bait proteins. A yeat strain transformed with

these plasmids, contains a reporter gene that can only be transcribed in the presence of the whole transcription factor (DNA-binding domain and transcriptional activation domain). When there is an interaction between the bait and one of the prey proteins, the two domains of the transcription factor come together and the reporter gene is transcribed. The measurable output, can be either growth under selective conditions or a colour signal (β-galactosidase assay).

References

Adamczyk, M. & Jagura-Burdzy, G. (2003) Spread and survival of promiscuous IncP-1 plasmids. *Acta Biochim. Pol.* 50, 425-453

Althorpe, N.J., Chilley, P.M., Thomas, A. T., Brammar, W. J. & Wilkins, B. M. (1999) Transient transcriptional activation of the Incl1 plasmid anti-restriction gene (*ardA*) abd SOS inhibition gene (*psiB*) early in conjugating recipient bacteria *Mol. Microbiol*. 31, 133-142

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. (1990) Basic local alignment search tool *J. Mol. Biol*. 215, 403-410

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 25, 3389-3402

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

Andrade, M. A., Chacón, P., Merelo, J. J. & Morán, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism using an unsupervised learning neural network *Protein Eng*. 6, 383–390.

Anglister, J., Grzesiek, S., Ren, H., Klee, C.B. and Bax, A. (1993) Isotope-edited multidimensional NMR of calcineurin B in the presence of the non-deuterated detergent CHAPS *J. Biomol. NMR* 3, 121-126

Atanasiu, C., Byron, O., McMiken H., Sturrock, S. and Dryden, D.T.F. (2001) Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7. *Nucleid Acids Research* 29, 3059-3068

Atanasiu, C., Su, T.J., Sturrock, S.S. & Dryden D.T.F. (2002) Interaction of the *ocr* gene 0.3 protein of bacteriophage T7 with *Eco*KI restriction/modification enzyme. *Nucleic Acids Res.* 30, 3936-3944

Bair, C.L. & Black L.W. (2007) A type IV modification dependent restriction nuclease that targets glucosylated hydroxymethyl cytosine modified DNAs. *J. Mol. Biol.* 366, 768-78

Bairoch, A., Apweiler, R., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N. & Yeah, L.S. (2005) The Universal Protein Resource (UniProt). *Nucleic Acids Res*. 33, D154-159

Bagdasarian, M., Bailone, A., Angulo, J.F., Scholz, P., Bagdasarian, M., & Devoret, R. (1992) PsiB, and anti-SOS protein, is transiently expressed by the F sex factor during its transmission to an Escherichia coli K-12 recipient.*Mol Microbiol.* 6, 885-93

Bandyopadhyay, P.K, Studier, F.W., Hamilton, D.L. & Yuan, R. (1985) Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.* 182, 567-578

Barrett, C.M., Mangels, D. & Robinson, C. (2005) Mutations in subunits of the Escherichia coli twin-arginine translocase block function via differing effects on translocation activity or tat complex structure. *J. Mol. Biol*. 347, 453-63.

Bates, S., Roscoe, R.A., Althorpe, N.J., Brammar, W.J. & Wilkins, B.M. (1999) Expression of leading region genes on IncI1 plasmid ColIb-P9: genetic evidence for single-stranded DNA transcription. *Microbiology* 145, 2655-2662

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

Belogurov, A.A., Delver, E.P., Agafononva, O.V., Belogurova, N.G., Lee, L. & Kado, C. (2000) Antirestriction protein Ard (Type C) encoded by the IncW plasmid pSa has a high similarity to the "protein transport" domain of TraC1 primase of promiscuous plasmid RP4. *J. Mol.Biol.* 296, 969-977

Belogurov, A.A., Delver, E.P. & Rodzevich O.V. (1993) Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose

expression is controlled by homologous regulatory sequences. *J. Bacteriol.* 175, 4843-4850

Belogurov, A.A., Yussifov, T.N., Kotova, V.U. & Zavilgelsky, G.B (1985) The novel gene(s) ARD of plasmid pKM101: alleviation of EcoK restriction. *Mol. Gen. Genet.* 198, 509-13

Bennett-Lovsey, R.M., Herbert, A.D., Sternberg, M.J. & Kelley, L.A. (2008). Exploring the extremes of sequence/structure space with ensemble fold recognition in the program Phyre. *Proteins* 70, 611-625.

Berg, J.M., Tymoczko, J.L. and Stryer, L. (2001) Biochemistry 5th ed. Freeman Publications

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D. & Zardecki, C. (2002) The Protein Data Bank. *Acta Crystallogr. D. Biol Crystallogr*. 58, 899-907

Bertram, J., Stratz, M. and Durre, P. (1991) Natural transfer of conjugative transposon Tn916 between gram-positive and gram-negative bacteria. *J. Bacteriol.* 173, 443-448

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

Bourniquel A. A. and Bickle T. A. (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie* 84, 1047–1059

Brown, T.A. (2001) Gene cloning and DNA analysis. 4th ed. Blackwell Publishing

Buchner-Moncke, E., Rothenberg, M., Reich, S., Wagenfuhr, K., Matsumura, H., Terauchi, R., Kruger, D.H. & Reuter, M. (2009) Functional characterization and modulation of the DNA cleavage efficiency of Type III restriction endonuclease EcoP15I in its interaction with two sites in the DNA target. *J. Mol.Biol*. 387, 1309-1319

Burdett, V. (1993) tRNA modification activity is necessary for *tet(M)*-mediated tetracycline resistance. *J. Bacteriol.* 175, 7209-7215

Burlage, R.S., Bernis, L.A., Layton, A.C., Sayler, G.S. & Larimer, F. (1990) Comparative genetic organisation of incompatibility group P degradative plasmids *J. Bacteriol*. 172, 6818-6825

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

Calisto, R.M., Pich, O.Q., Pinol, J., Fita, I., Querol, E. & Carpena, X. (2005) Crystal structure of a putative type I restriction–modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.* 351, 749–762

Camacho-Carvajal, M.M., Wollscheid, B., Aebersold, R., Steimle, V. & Schamel, W.W.A. (2004) Two-dimensional blue native/SDS gel electrophoresis of multiprotein complexes from whole cellular lysates- a proteomics approach. *Molecular & Cellular Proteomics* 3, 176-182

Chilley, P.M. & Wilkins, B.M. (1995) Distribution of the ardA family of antirestriction genes on conjugative plasmids. *Microbiology* 141, 2157-2164

Choi, K. H. & Licht, S. (2005) Control of peptide product sizes by the energy-dependent protease ClpAP *Biochemistry* 44, 13921–13931

Clermont, D. & Horaud, T. (1994) Genetic and molecular studies of a composite chromosomal element (Tn3705) containing a Tn916-modified structure (Tn3704) in *Streptococcus anginosus* F22. *Plasmid* 31, 40-8

Combet, C., Blanchet, C., Geourjon, C. & Deleage, G. (2000) NPS@: network protein sequence analysis. *Trends Biochem. Sci.* 25, 147-150

Cooper, A., Nutley, M.A. & Wadood, A. Differntial scanning microcalorimetry in Harding, S.E. & Chowdhry, B.Z. Protein-ligand interactions: hydrodynamics and calorimetry. Oxford University Press (2000) p. 287-318

Cornilescu, G., Delaglio, F., Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* 13, 289-302

Court, R., Cook, N., Saikrishnan, K. & Wigley D. (2007) The Crystal Structure of lambda-Gam Protein Suggests a Model for RecBCD Inhibition. *J. Mo.l Biol.* 371, 25-33

Courvalin, P and Carlier, C. (1987) Tn*1545*: a conjugative shuttle transposon. *Mol.Gen. Genet.* 206, 259-264

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

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

Creighton, T.E. (1997) Protein Structure: a Practical Approach 2nd ed. IRL Press

Davison, J. (1999) Genetic exchange between bacteria in the environment. *Plasmid* 42, 73-91

De, S., Girigoswami, A. & Das, S. (2005) Fluorescence probing of albumin-surfactant interaction *J. Colloid Interf. Sci.* 285, 562-573

Delver, E.P., Kotova, V.U., Zavilgelsky, G.B. & Belogurov, A.A. (1991) Nucleotide sequence of the gene (ard) encoding the antirestriction protein of plasmid colIb-P9. *J Bacteriol*. 173, 5887-5892.

Delver, E.P. & Belogurov, A.A. (1997) Organization of the leading region of IncN plasmid pKM101 (R46): a regulation controlled by CUP sequence elements. *J. Mol. Biol.* 271, 13-30

Depew, R.E. & Cozzarelli, N.R. (1974) Genetics and physiology of bacteriophage T4 3'-phosphatase: evidence for the involvement of the enzyme in T4 DNA metabolism. *J. Virol*. 13, 388-897

Dietmann, S., Park, J., Notredame, C., Heger, A., Lappe, M. & Holm, L. (2001) A fully automatic evolutionary classification of protein folds: Dali Domain Dictionary version 3. *Nucleic Acids Res.* 29, 55-57.

Dosset, P., Hus, J.C., Blackledge, M. & Marion, D. (2000) Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data *J. Biomol. NMR* 6, 23-28

Dzidic, S. & Bedekovic, V. (2003) Horizontal gene transfer emerging multidrug resistance in hospital bacteria. *Acta Pharmacol. Sin.* 24, 519-526

Eftink, K. & Ghiron, A.C. (1976) Exposure of tryptophanyl residues in proteins: quantitative determination by fluorescence quenching studies. *Biochemistry* 15, 672-680

Eftink, K. & Ghiron, A.C. (1981) Fluorescence quenching studies with proteins. *Analytical Biochemistry* 144, 199-277

Eubel, H., Braun, H. P. & Millar, H. (2005) Blue-native PAGE in plants: a tool in analysis of protein-protein interactions *Plant Methods* 1, 1-13

Eubel, H., Heinemeyer, J, Sunderhaus, S. & Braun H.P. (2004) Respiratory chain supercomplexes in plant mitochondria *Plant Physiology and Biochemistry* 42, 937-942

Evan, G.I., Lewis, G.K., Ramsay, G. & Bishop, J.M. (1985) Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol Cell Biol*. 5, 3610-3616

Fields S., & Song, O. (1989) A novel genetic system to detect protein– protein interactions *Nature* 340, 245-246.

Figurski, D.H., Pohlman, R.F., Bechhofer, D.H., Prince, A.S. & Kelton , C.A. (1982) Broad host range plasmid RK2 encodes multiple *kil* genes potentially lethal to *Escherichia coli* host cells. *Proc. Natl. Acad. Sci. USA* 79, 1935-1939

Fischer, M.W., Losonczi, J.A., Weaver, J.L. & Prestegard, J.H. (1999) Domain orientation and dynamics in multidomain proteins from residual dipolar couplings. *Biochemistry* 38, 9013-22

Flannagan, S.E., Zitzow, L.A., Su, Y.A. & Clewell, D.B. (1994) Nucleotide sequence of the 18-kb conjugative transposon Tn*916* from *Enterococcus faecalis*. *Plasmid* 32, 350-354

Franke, A.E. & Clewell, B.D. (1981) Evidence for a chromosome-borne resistance transposon (Tn*916*) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the absence of a conjugative plasmid. *J. Bacteriol.* 145, 494-502

Fraczkiewicz, R. & Braun, W. (1998) "Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules" *J. Comp. Chem.*, 19, 319-333.

Fuchs, J & Podda, M. (2004) *Encyclopedia of Diagnostic Proteomics and Genomics* Informa Healthcare Publishing

Gachechiladze, K.K., Balardshishvili, N.S., Adamia, R.S., Chanishvili, T.G. & Kruger, D.H. (1990) Host-controlled modification and restriction as a criterion of evaluating the therapeutical potential of Pseudomonas phage. *J. Basic Microbiol.* 31, 101-106

Garnett, J.A., Marincs, F., Baumberg, S., Stockley, P.G. & Phillips, S.E. (2008) Structure and function of the arginine repressoroperator complex from *Bacillus subtilis*. *J. Mol. Biol.* 379, 284–298

Ginalski, K., Elofsson, A., Fischer, D. & Rychlewski, L. (2003) 3D-Jury: a simple approach to improve protein structure predictions. *Bioinformatics* (Oxford, England) 19, 1015-1018.

Glaser, F., Pupko, T., Paz, I., Bell, R.E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003) ConSurf: identification of functional regions in proteins by surfacemapping of phylogenetic information. *Bioinformatics* 19, 163-164

Glover, S.W. & Colson, C. (1969). Genetics of host-controlled restriction and modification in *Escherichia coli. Genetical Research* 13, 227-240

Gottesman, S., Clark, W.P., de Crecy-Lagard V. & Maurizi, M.R. (1993) ClpX, an alternative subunit for the ATP-dependent Clp protease of *Escherichia coli*. *J. Biol. Chem.* 268, 22618–26

Goucharoff, P., Saadi, S., Chang, C., Saltman, L. H. & Figurski, D.H. (1991) Structural, molecular and genetic analysis of the kilA operon of broad host range plasmid RK2. *J. Bacteriol*. 173, 3463-3477

Grandier-Vazeille, X. & Guérin, M. (1996) Separation by blue native and colorless native polyacrylamide gel electrophoresis of the oxidative phosphorylation complexes of yeast mitochondria solubilized by different detergents: specific staining of the different complexes. *Anal Biochem.* 242, 248-54.

Greene, L.H., Lewis, T.E., Addou, S., Cuff, A., Dallman, T., Dibley, M., Redfern, O., Pearl, F., Nambudiry, R., Reid, A., Sillitoe, I., Yeats, C., Thornton, J.M. & Orengo C.A. (2007) The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution. *Nucleic Acids Res*. 35, D291-297

Greenfield, N. J. (1999) *Applications of circular dichroism in protein and peptide analysis Trends Anal. Chem.* 18, 236–244

Grzesiek, S. & A. Bax, The importance of not saturating H_2O in protein NMR - Application to sensitivity enhancement and Noe measurements (1993) *J. Am. Chem. Soc.* 115, 12593-12594.

Guntert, P. (2004) Automated NMR structure calculation with CYANA .*Methods Mol. Biol*. 278, 353-78.

Hahn, D.R., McHenney, M.A. & Baltz, R.H. (1990) Characterisation of FP22, a large streptomycete bacteriophage with DNA insensitive to cleavage by many restriction enzymes. *J. Gen. Microbiol.* 136, 2395-2404

Haian, Fu (2004) Protein-protein Interactions: Methods and Applications (Methods in Molecular Biology) Humana Press p.35-37

Halford, S.E. & Marko, J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32, 3040-3052.
Hansen, M.R., Mueller, L., & Pardi, A. (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat. Struct. Biol.* 5, 1065–1074

Harada, K.M., Aso, Y., Hashimoto, W., Mikami, B. & Murata, K. (2006) Sequence and analysis of the 46.6 kb plasmid pA1 from *Sphingomonas* sp. A1 that corresponds to the typical IncP-1β plasmid backbone without any accessory genes. *Plasmid* 56, 11-23

Hattman, S. (1979) Unusual modification of bacteriophage Mu DNA. *J. Virol.* 32, 468-475

He, Q. & Mertsola, J. (2008) Factors contributing to pertussis resurgence. *Future Microbiol.* 3, 329-39

Heiden, W., Moeckel, G. & Brickmann, J. (1993) A new approach to analysis and display of local lipophilicity/hydrophilicity mapped on molecular surfaces. *J Comput Aided Mol Des* 7, 503-514

Herrmann, T., Güntert, P. & Wüthrich, K. (2002) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J. Mol. Biol*. 319, 209-27

Herzof-Velikonja, B., Podlesek, Z. & Grabnar, M. (1994) Conjugal transfer of transposon Tn916 from *Enterococcus faecalis* to *Bacillus licheniformis*. *Plasmid* 31, 201-206

Hespell, R.B. & Whitehead, T.R. (1991) Conjugal transfer of Tn*916*, Tn*916*ΔE, and pAMβ1 from *Enterococcus faecalis* to *Butyrivibrio fibrisolvens*. *App. Env. Microbiol.* 57, 2703-2709

Hoffmann, E. & Stroobant, V. (2007) *Mass Spectrometry Principles and Applications* 3rd Ed. Wiley Publications

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol*. 233, 123-138

Huang, L.H., Forret, K., Ehrlich,C. & Ehrlich, M. (1982) Digestion of highly modified bacteriophage DNA by restriction enzymes. *Nucleic Acids Res.* 10, 1579-1591

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

Jones, A.L., Barth, P.T. & Wilkins, B.M. (1992) Zygotic induction of plasmid ssb and psiB genes following conjugative transfer of Incl1 plasmid Collb-P9. *Mol. Microbiol*. 6, 605-13.

Kamachi, K., Sota, M., Tamai, Y., Nagata, N., Konda, T., Inoue T., Top, E.M. & Arakawa, Y. (2006) Plasmid pBP136 from *Bordetella pertussis* represents an ancestral form of IncP-1beta plasmids without accessory mobile elements. *Microbiology* 152, 3477-84

Kay, L.E., Torchia, D.A. & Bax, A. (1989) Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease *Biochemistry.* 28, 8972-8979

Keiler, K.C., Waller, P.R. & Sauer, R.T. (1996) Role of a peptide tagging system in degradation of proteins synthesised from damaged messenger RNA *Science* 271, 990-993.

Kelley, L.A., Gardner, S.P. & Sutcliffe, M.J. (1996) An automated approach for clustering an ensemble of NMR-derived protein structures into conformationally related subfamilies. *Protein Eng.* 9, 1063-1065

Kelly, S.M. & Price, N.C. (1997) The application of circular dichroism to studies of protein folding and unfolding. *Biochim Biophys Acta* 1338, 161-85

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

Kim, J.S., DeGiovanni, A., Jancarik, J., Adams, P.D., Yokota, H., Kim, R. & Kim, S.H. (2005)Crystal structure of DNA sequence specificity subunit of a type I restriction–modification enzyme and its functional implications. *Proc. Natl Acad. Sci. USA* 102, 3248–3253.

Kim, B.C., Kim, K., Park, E.H. & Lim, C.J. (1997) Nucleotide sequence and revised map location of the *arn* gene from bacteriophage T4. *Mol. Cell* 7, 694-6.

Kim, Y.I., Levchenko, I., Fraczkowska, K., Woodruff, R.V., Sauer, R.T. & Baker, T.A. (2001) Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat Struct Biol*. 8, 230-3

King, G. & Murray, N.E. (1995) Restriction alleviation and modification enhancement by the Rac prophage of Escherichia coli K-12. *Mol. Microbiol.* 16, 769-77.

Kline, T., Park, H. & Meyerson, L.R. (1989) CHAPS solubilization of a G-protein sensitive 5-HT1A receptor from bovine hippocampus *Life Sci.* .45, 1997-2005

Klotz, I.M. (1997) Ligand-receptor energetics: a guide for the perplexed. John Wiley publications

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

Koradi, R., Billeter, M., & Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* 14, 51-55

Kornacki, J.A., Chang, C. & Figurski, D.H. (1993) kil-kor regulon of promiscuous plasmid RK2: structure, products and regulation of two operons that constitute the *kilE* locus. *J. Bacteriol*. 175, 5078-5090

Kowalczykowski, S.C., Dixon, D.A., Eggleston, A.K., Lauder, S.D. & Rehrauer, W.M. (1994) Biochemistry of homologous recombination in Escherichia coli. *Microbiol Rev.* 58, 401-65

Krissinel, E. & Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallographica* 60, 2256-2268

Kruger, D.H., Presber, W., Hansen, S. & Rosenthal, H.A. (1975) Biological functions of the bacteriophage T3 SAMase gene. *J. Virol.* 16, 453-455

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

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

Kumar, S. & Nussinov, R. (1999) Salt bridge stability in monomeric proteins. *J Mol Biol.* 293, 1241-1255

Lacks, S.A., Mannarelli, B.M., Springhorn, S.S. & Greenberg. B. (1986) Genetic basis of the complementary *DpnI* and *DpnII* restriction systems of *S. pneumoniae*: an intercellular cassette mechanism. *Cell* 46, 993-1000

Ladbury, J. E. & Chowdhry, B.Z. (1996) Sensing the heat: the application of isothermal titration calorimetry to thermodynamic studies of biomolecular interactions *Chem. Biol.* 3, 791-801

Lakowitz, J.R. (2006) Principles of fluorescence spectroscopy. 3rd ed. Springer Publications

Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res*. 33, W299-302

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

Larsen, M.H. & Figurski, D.H. (1994) Structure, expression and regulation of the kilC operon of promiscuous IncPα plasmids. *J. Bacteriol*. 176, 5022-5032

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R. & Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* 8, 477-86

Le Bouguenec, C. De Cespedes G. & Horaud, T. (1988) Molecular analysis of a composite chromosomal conjugative element (Tn*3701*) of *Streptococcus pyogenes J. Bacteriol.* 170, 3930-36

Levitt, H.M. (2001) Spin dynamics: basics of nuclear magnetic resonance. Wiley Publications

Lisle, W., Dutta, C.F., Penner, M., Amitsur, M., Kaufmann, G. & Firman, K. (2000) Phage T4-encoded Stp alleviates the DNA restriction activity of EcoR124I endonuclease by affecting a critical step in the subunit assembly pathway. *Molecular Biology Today* 1, 57-64.

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

Makino, O., Saito, H. & Ando, T. (1980) *Bacillus subtilis*-phage phi 1 overcomes host-controlled restriction by producing *Bam*Nx inhibitor protein. *Mol. Gen. Genet.* 179, 463-8

Makovets, S., Doronina, V.A. & Murray, N.E. (1999) Regulation of endonuclease activity by proteolysis prevents breakage of unmodified bacterial chromosomes by type I restriction enzymes *Proc. Natl. Acad. Sci. USA* 96, 9757-9762

Makovets, S. Powell, L.M., Titheradge, A. J. B., Blakely, G. W. & Murray, N.E. (2004) Is modification sufficient to protect a bacterial chromosome from a resident restriction endonuclease? *Mol. Microbiol.* 51, 135-147

Makovets, S., Titheradge, A.J.B. & Murray N.E. (1998) ClpX and ClpP are essential for the efficient acquisition of genes specifying type IA and type IB restriction systems *Mol. Microbiol*. 28, 25-35

Manz, A., Pamme N. & Iossifidis D. (2004) *Bioanalytical Chemistry* Imperial College Press

McGuffin, L. J., Bryson, K. & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics,* 16, 404–405

McMahon, S.A., Roberts, G.A., Johnson, K.A., Cooper, L.P., Liu H., White, J.H., Carter, L.G., Sanghvi, B., Oke, M., Walknishaw, M.D., Blakely, G.W., Naismith, J.H. & Dryden, D.T.F. (2009) Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance *Nucleic Acids Res. In Press*

Morth, J.P., Feng, V., Perry, L.J., Svergun, D.I. & Tucker, P.A. (2004) The crystal and solution structure of a putative transcriptional antiterminator from Mycobacterium tuberculosis. *Structure* 12, 1595–1605

Muhandiram, D.R., Farrow, N.A., Xu, G.Y., Smallcombe, S.H., & Kay, L.E. (1993) A gradient ^{13}C NOESY-HSQC experiment for recording NOESY spectra of ^{13}C-labeled protein dissolved in H_2O. *J. Magn. Reson*. 102, 317–321

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

Murray, N.E. (2002) Immigration control of DNA in bacteria: self versus non-self *Microbiology* 148, 3-20

Murzin, A.G., Brenner, S.E., Hubbard, T. & Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol*. *247*, 536-540.

Nasim, M. T., Eperon, I. C., Wilkins, B. M. & Brammar, W. J. (2004) The activity of a single-stranded promoter of plasmid ColIb-P9 depends on its secondary structure *Mol. Microbiol.* 53, 405–417

Neaves, K.J., Cooper, L.P., White, J.H., Carnally, S. M., Dryden, D.T.F., Edwardson, J.M. & Henderson, R. M. (2009) Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerisation and DNA looping *Nucleic Acid Res.* 37, 2053-2063

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

Nicholls, A., Sharp, K.A. & Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11, 281-296.

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

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

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. & Thornton, J.M. (1997) CATH--a hierarchic classification of protein domain structures. *Structure* 5, 1093-1108

Ottiger, M., Delaglio, F. & Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson*. 131, 373–378

Pace, C.N., Shirley, B.A. & Thomson, J.A. (1989) In Creighton, T.E. (2nd ed.) Protein Structure: A practical Approach IRL Press, Oxford pp.311-330

Pei, J., Kim, B.H. & Grishin, N.V. (2008a) PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Re.* 36, 2295-2300.

Pei, J., Tang, M. & Grishin, N.V. (2008b) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36 (Web Server issue)W30-4

Perkins, D.N., Pappin, D.J., Creasy, D.M. & Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* 20, 3551-67

Pieper, U., Groll, D.H., Wünsch, S., Gast, F.U., Speck, C., Mücke, N. & Pingoud, A. (2002) The GTP-dependent restriction enzyme McrBC from *Escherichia coli* forms high-molecular mass complexes with DNA and produces a cleavage pattern with a characteristic 10-base pair repeat *Biochemistry* 41, 5245-5254

Prestegard, J. H., Bougault, C. M. & Kishore, A. I. (2004)Residual Dipolar Couplings in Structure Determination of Biomolecules *Chem. Rev.* 104, 3519-3540

Preston, A. (2005) *Bordetella pertussis*: the intersection of genomics and pathobiology *CMAJ* 173, 55-62

Rajashankar, K.R., Knieewel, R. & Lima, C.D. Crystal structure of Type I restriction enzyme EcoKI M protein (EC 2.1.1.72) (M.EcoKI) *To be published*

Rastorguev, S.M. & Zavilgelsky, G.B. (2004) Role of "Antirestriction" Motif in Functional Activity of Antirestriction Protein ArdA pKM101 (IncN) *Russian Journal of Genetics* 39, 1608-3369

Read, T. D., Thomas, A. T. & Wilkins, B. M. (1992). Evasion of type I and type I1 DNA restriction systems by IncIl plasmid ColIb-P9 during transfer by bacterial conjugation. *Mol. Microbiol.* 6, 1933-1941

Redfern, O.C., Harrison, A., Dallman, T., Pearl, F.M. & Orengo, C.A. (2007) CATHEDRAL: a fast and effective algorithm to predict folds and domain boundaries from multidomain protein structures. *PLoS Comput. Biol*. 3, e232

Reisinger, V. & Eichaker, L.A. (2007) How to analyze protein complexes by 2D blue native SDS-PAGE *Proteomics* 7, 6-16

Rice, L.B. & Carias, L.L. (1998) Transfer of Tn*5385*, a composite, multiresistance chromosomal element from *Enterococcus faecalis. J. Bacteriol.* 180, 714-21

Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.K., Dryden, D.T., Dybvig, K., Firman, K., Gromova, E.S., Gumport, R.I., Halford, S.E., Hattman, S., Heitman, J., Hornby, D.P., Janulaitis, A., Jeltsch, A., Josephsen, J., Kiss, A., Klaenhammer, T.R., Kobayashi, I., Kong, H., Krüger, D.H., Lacks, S., Marinus, M.G., Miyahara, M., Morgan, R.D., Murray, N.E., Nagaraja, V., Piekarowicz, A., Pingoud, A., Raleigh, E., Rao, D.N., Reich, N., Repin, V.E., Selker, E.U., Shaw, P.C., Stein, D.C., Stoddard, B.L., Szybalski, W., Trautner, T.A., Van Etten, J.L., Vitor, J.M., Wilson, G.G. & Xu, S.Y. (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.* 31, 1805–1812

Rule, G.S., & Hitchens, T.K. (2006) *Fundamentals of NMR Spectroscopy* Springer publications

Sah, H. & Kim, K. (2006) Improvement of interfacial protein stability by CHAPS *Biotechnology Letters* 28: 567-570

Saltman, L.H, Kim, K.S. & Figurski, D.H. (1991) The kilA operon of promiscuous plasmid RK2: the use of a transducing phage (lambda pklaA-1) to determine the effects of the lethal klaA gene on *Escherichia coli* cells. *Mol. Microbiol*. 5, 2673-2683

Sambrook, J. & Russell, D.W. (2002) Molecular cloning: a laboratory manual. 3rd ed. Cold Springs Harbor Laboratory Press

Schagger, H. & Vonjagow G. (1991) Blue native electrophoresis for isolation of membrane-protein complexes in enzymatically active form *Anal. Biochem.* 199, 223-231

Scott, J.R. & Churchward, G.G. (1995) Conjugative Transposition. *Annu.Rev.Microbiol.* 49, 367-97

Serfiotis-Mitsa, D. Roberts, G.A., Cooper, L.A., White, J.H., Nutley, M., Cooper, A. Blakely, G.W. & Dryden D.T.F. (2008) The ORF18 Gene Product from Conjugative Transposon Tn916 is an ArdA Antirestriction Protein that Inhibits Type I DNA Restriction-Modification Systems *J. Mol. Biol*. 383, 970- 981

Shen, Y., Volrath, S.L., Weatherly, S.C., Elich, T.D. & Tong, L. (2004) A mechanism for the potent inhibition of eukaryotic acetylcoenzyme A carboxylase by soraphen A, a macrocyclic polyketide natural product. *Mol. Cell* 16, 881–891

Stewart, F.J., Panne, D., Bickle, T.A. & Raleigh, E.A. (2000) Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol.Biol.* 298, 611-622

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

Suchanek, M, Radzikowska, A. & Thiele, C. (2005) Photoleucine and photo-methionine allow identification of protein-protein interactions in living cells *Nat. Methods* 2, 261-267

Terpe, K. (2006) Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Appl. Microbiol. Biotechnol.* 72, 211-22

Thomas, A.T., Brammar, W.J. & Wilkins, B.M. (2003) Plasmid R16 ArdA protein preferentailly targets restriction activity of the Type I restriction-modification system *Eco*KI. *J. Bacteriol.* 185, 2022-2025

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

Thomson, V.J., Jovanovic, O.S., Pohlman, F.R., Chang, C. & Figurski, D.H. (1993) Structure, function, and regulation of the *kilB* locus of promiscuous plasmid RK2. *J. Bacteriol*. 175, 2423-2435

Tjandra, N., & Bax, A. (1997) Measurement of dipolar contributions to $^{1}J_{CH}$ splittings from magnetic-field dependence of *J* modulation in two-dimensional NMR spectra *J. Magn. Reson.* 124, 512-515

Tjandra, N., Feller, S.E., Pastor, R.W. & Bax, A. (1995) Rotational diffusion anisotropy of human ubiquitin from ^{15}N NMR relaxation. *J. Am. Chem. Soc*. 117: 12562-12566

Tolman, J. R. (2001) Dipolar couplings as a probe of molecular dynamics and structure in solution. *Curr. Opin. Struct. Biol*. 11, 532-539

Tolman, J.R., Flanagan J.M., Kennedy M.A. & Prestegard, J.H. (1997). NMR evidence for slow collective motions in cyanometmyoglobin. *Nat. Struct. Biol.* 4, 292-297.

Torbet, J. & Maret, G. (1979) Fibres of highly oriented Pf1 bacteriophage produced in a strong magnetic field *J. Mol. Biol*. 134, 843-846

Trautner T.A., Pawlek, B., Günthert, U., Canosi, U., Jentsch, S. & Freund, M. (1980) Restriction and modification in *Bacillus subtilis*: identification of a gene in the temperate phage SP beta coding for a BsuR specific modification methyltransferase. *Mol. Gen. Genet.* 180, 361-7

Uhrínová, S., Uhrín, D., Denton, H., Smith, M., Sawyer, L. & Barlow, P.N. (1998) Complete assignment of 1H, 13C and 15N chemical shifts for bovine beta-lactoglobulin: secondary structure and topology of the native state is retained in a partially unfolded form. *J. Biomol. NMR.* 12, 89-107

Ulrich, E.L., Akutsu, H., Doreleijers, J.F., Harano, Y., Ioannidis, Y.E., Lin, J., Livny, M., Mading, S., Maziuk, D., Miller, Z., Nakatani, E., Schulte, C.F., Tolmie, D.E., Wenger R, Yao, H., & Markley, J.L. (2008) BioMagResBank. *Nucleic Acids Res.* 36, D402-8

Valafar, H., & Prestegard, G.H. (2004) REDCAT: A residual dipolar couplings analysis tool *J. Magn. Reson.* 167, 228-241

Van Holde, K. E., Johnson, W.C. & Ho Shing, P. (2006) *Principles of Physical Biochemistry* 2nd Ed. Prentice Hall

Voges, D., Zwicki, P. & Baumeister, W. (1999) The 26S proteasome: a molecular machine designed for controlled proteolysis *Annu. Rev. Biochem*. 68, 1015, 1068

Vranken, W.F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E.L., Markley, J.L., Ionides, J. & Laue, E.D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* 59, 687-96

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

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

Watanabe, T. (1963) Infective heredity of multiple drug resistance in bacteria. *Bacteriol. Rev.* 27, 87-115

Weber, P.C & Salemme, F.R. (2003) Applications of calorimetric methods to drug discovery and the study of protein interactions *Curr. Opin. Struct. Biol.* 13, 115-121

White, D. (2007) The physiology and biochemistry of prokaryotes. 3rd ed. Oxford University Press

Whitmore, L. & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res*. 32, W668–W673

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

Wishart, D.S., Sykes, B.D. & Richards, F.M. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol*. 222, 311-33

Woodward, C.K. & Hilton, B.D. (1980) Hydrogen isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. *Biophys. J.* 32, 561-575

Yamazaki, T., Forman-Kay, J. D. & Kay, L. E. (1994) "Two-dimensional NMR experiments for correlating $^{13}C_{\beta}$ and $^{1}H_{\delta/\varepsilon}$ chemical shifts of aromatic residues in ^{13}C-labeled proteins via scalar couplings". *J. Am. Chem. Soc.* 115, 11054-11055

Yip, G.N. & Zuiderweg, E.R (2005) Improvement of duty-cycle heating compensation in NMR spin relaxation experiments *J Magn Reson*. 176, 171-8.

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

Zavilgelsky, G. B., Kotova,V. Y. & Rastorguev, S. M. (2009) Antirestriction and Antimodification Activities of T7 Ocr: Effects of Amino Acid Substitutions in the Interface *Molecular Biology* 43, 93–100

Appendix

Fig I. ArdA and KlcA sequences

ArdA sequence

```
ATGGACGATATGCAAGTCTATATTGCGAATTTAGGCAAATACAATGAGGGCGAATTGGTC
 M  D  D  M  Q  V  Y  I  A  N  L  G  K  Y  N  E  G  E  L  V
GGTGCGTGGTTTACCTTTCCCATTGACTTTGAGGAAGTCAAAGAGAAAATCGGCTTGAAT
 G  A  W  F  T  F  P  I  D  F  E  E  V  K  E  K  I  G  L  N
GATGAATATGAGGAATACGCCATTCATGACTACGAGTTACCCTTTACGGTTGACGAATAC
 D  E  Y  E  E  Y  A  I  H  D  Y  E  L  P  F  T  V  D  E  Y
ACTTCCATTGGCGAACTCAATCGACTATGGGAAATGGTATCGGAATTACCCGAAGAATTA
 T  S  I  G  E  L  N  R  L  W  E  M  V  S  E  L  P  E  E  L
CAATCGGAGCTATCTGCTCTGCTCACTCATTTTTCAAGCATTGAAGAACTAAGCGAACAT
 Q  S  E  L  S  A  L  L  T  H  F  S  S  I  E  E  L  S  E  H
CAAGAGGATATTATCATTCATTCCGATTGTGATGATATGTATGACGTGGCACGCTACTAC
 Q  E  D  I  I  I  H  S  D  C  D  D  M  Y  D  V  A  R  Y  Y
ATTGAAGAAACGGGTGCTTTAGGCGAAGTACCAGCTAGTCTTCAAAACTATATTGATTAT
 I  E  E  T  G  A  L  G  E  V  P  A  S  L  Q  N  Y  I  D  Y
CAAGCCTATGGTCGGGATTTAGACCTTTCAGGAACGTTTATCTCAACCAATCATGGGATT
 Q  A  Y  G  R  D  L  D  L  S  G  T  F  I  S  T  N  H  G  I
TTTGAAATCGTCTAT
 F  E  I  V  Y
```

KlcA sequence (codon optimised)

```
ATGAACACCGAAGAACAGCCGGTGACCGCGAGCCTGGTTGCGGAAGCGCAGCGTCTGGAT
 M  N  T  E  E  Q  P  V  T  A  S  L  V  A  E  A  Q  R  L  D
TTTCTGCCGACCTATTTTGGCCCGCGTCTGATGATGCGTGGCGAAGCGCTGGTGTATGCG
 F  L  P  T  Y  F  G  P  R  L  M  M  R  G  E  A  L  V  Y  A
TGGATGCGTCGTCTGTGCGAACGTTACAACGGCGCGTATTGGCATTACTACGCCCTGAGC
 W  M  R  R  L  C  E  R  Y  N  G  A  Y  W  H  Y  Y  A  L  S
GATGGCGGCTTTTATATGGCGCCGGATCTGGCCGGTCGTCTGGAAATTGAAGTGAACGGC
 D  G  G  F  Y  M  A  P  D  L  A  G  R  L  E  I  E  V  N  G
AACGGCTTTCGTGGCGAACTGAGCGCGGATGCGGCGGGTATTGTGGCGACCCTGTTTGCG
 N  G  F  R  G  E  L  S  A  D  A  A  G  I  V  A  T  L  F  A
CTGGGTCAGCTGGCCGCGGAAATTGCGGATACCGATGCGGCGGATGCGCTGATTGATCGT
 L  G  Q  L  A  A  E  I  A  D  T  D  A  A  D  A  L  I  D  R
TATCATTTTCTGCGTGGCTTTGCGGCGGGTCATCCGGAAGCGGCAGCGATTTATCGTGCG
 Y  H  F  L  R  G  F  A  A  G  H  P  E  A  A  A  I  Y  R  A
ATTGAT
 I  D
```

Fig. II. Secondary structure prediction of ArdA (using PSIPRED)

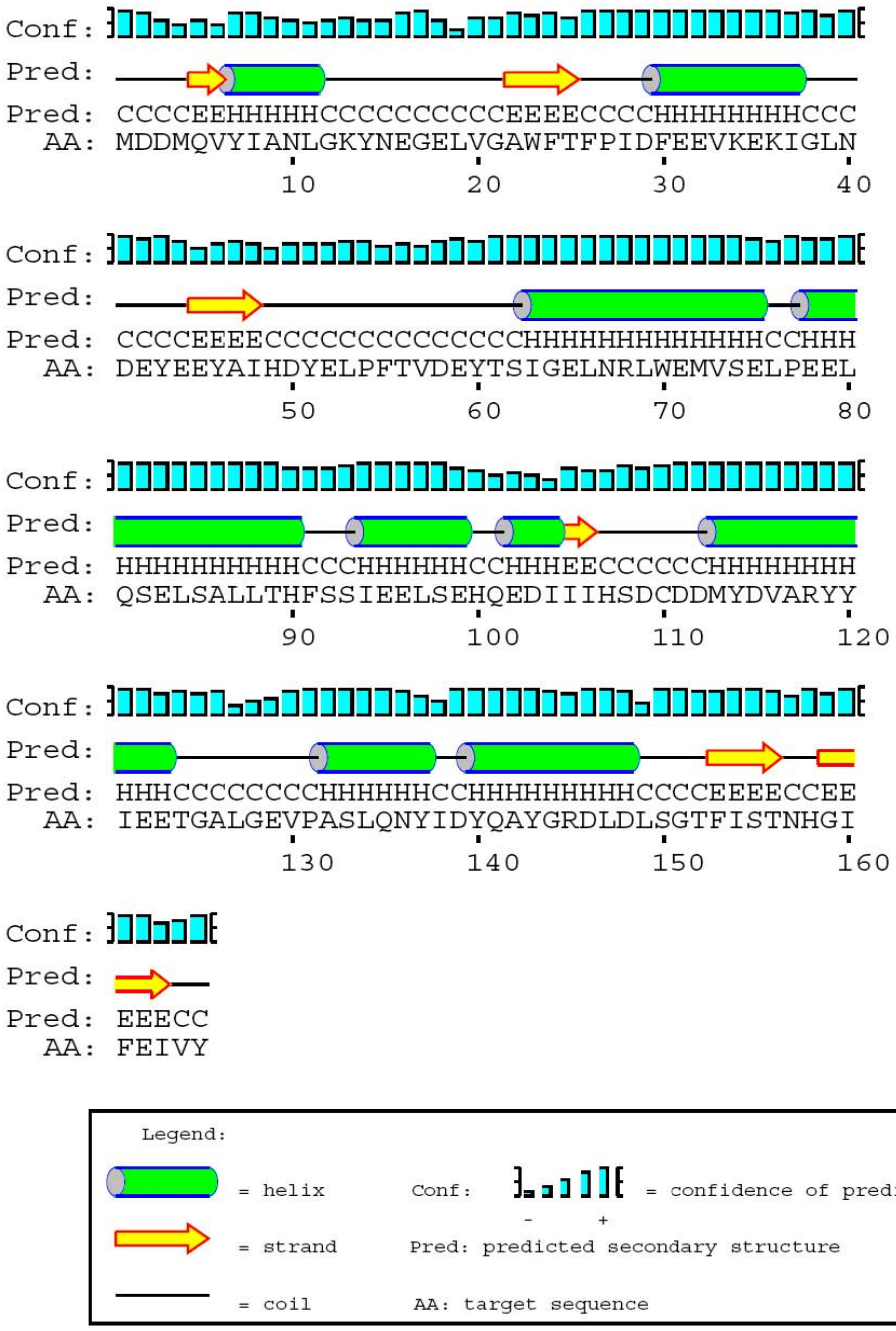


Fig. III. Secondary structure prediction of KlcA (using PSIPRED)

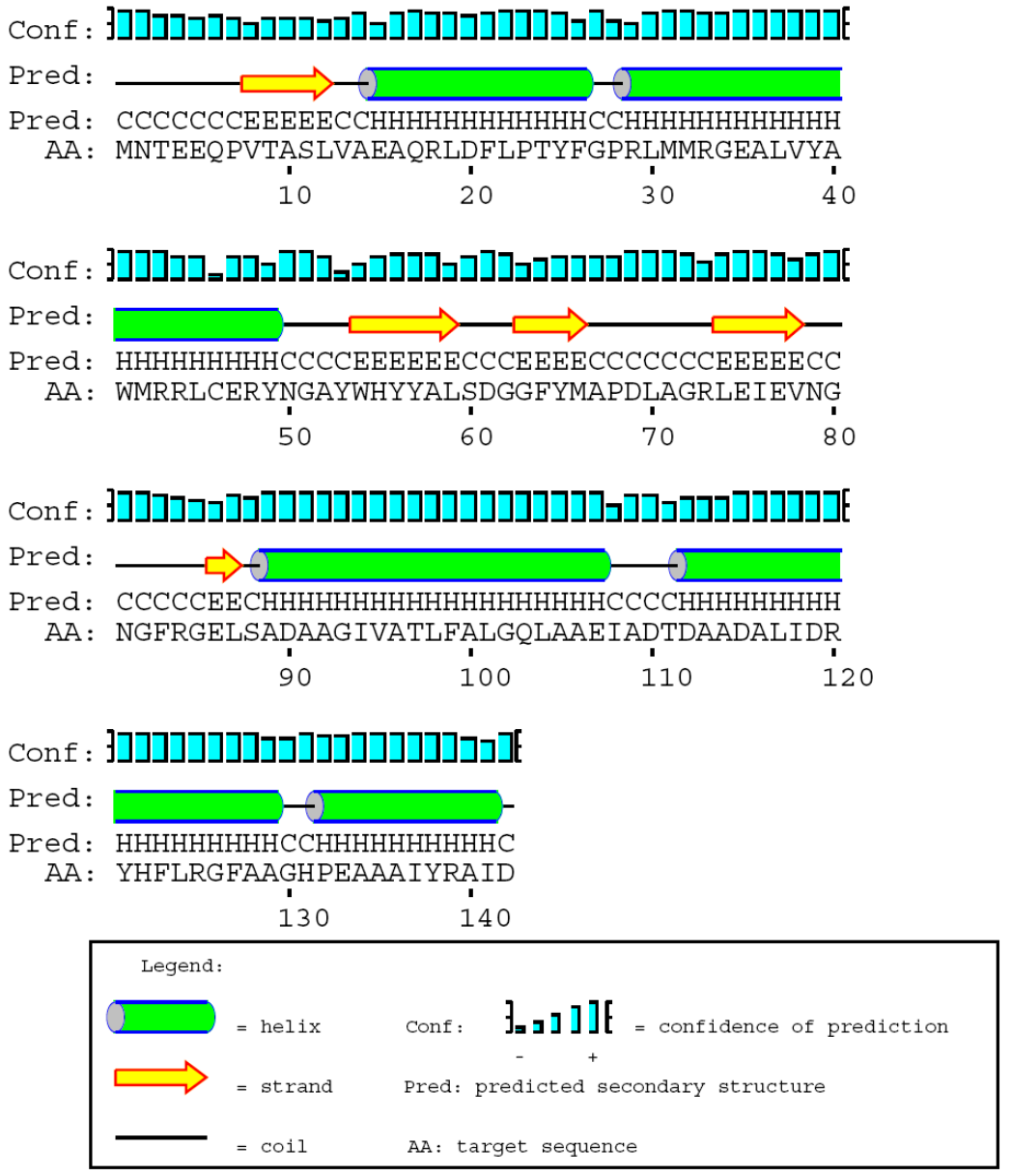


Fig. IV. Multiple sequence alignment of KlcA and ArdB (1/2)

Multiple sequence alignment of 43 non-redundant sequences of ArdB and KlcA using PROMALS3D. The secondary structure (alpha-helices: red; beta-strands: yellow; 3_{10}-helix: purple) from the consensus ensemble of NMR-derived structures is shown above the aligned sequences. Highly conserved amino acid positions that are exposed are shown within a green rectangular box, while highly conserved, buried positions are highlighted in a purple. KlcA from B.pertussis corresponds to Q08L07_BOR.

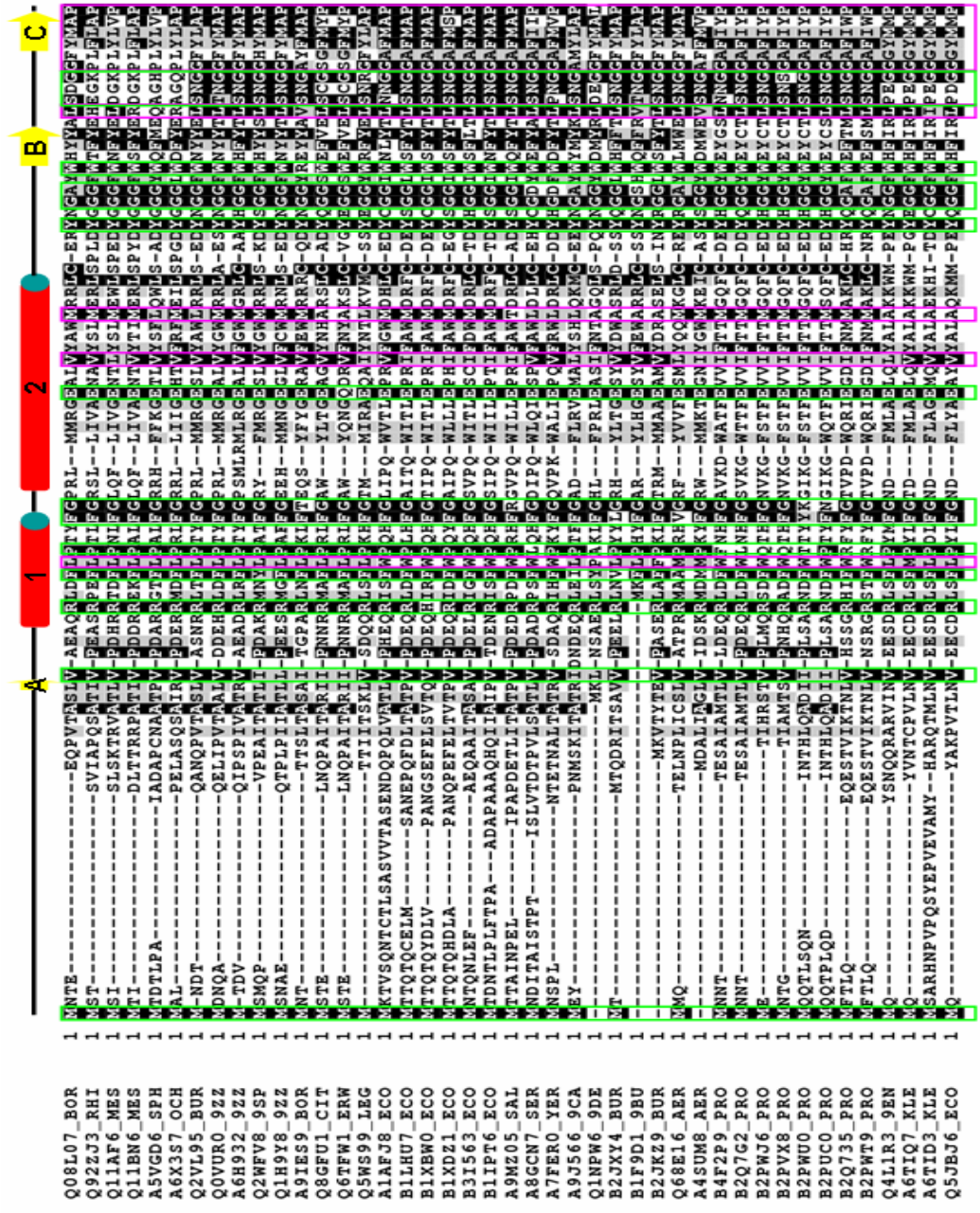


Fig. V. Multiple sequence alignment of KlcA and ArdB (2/2)

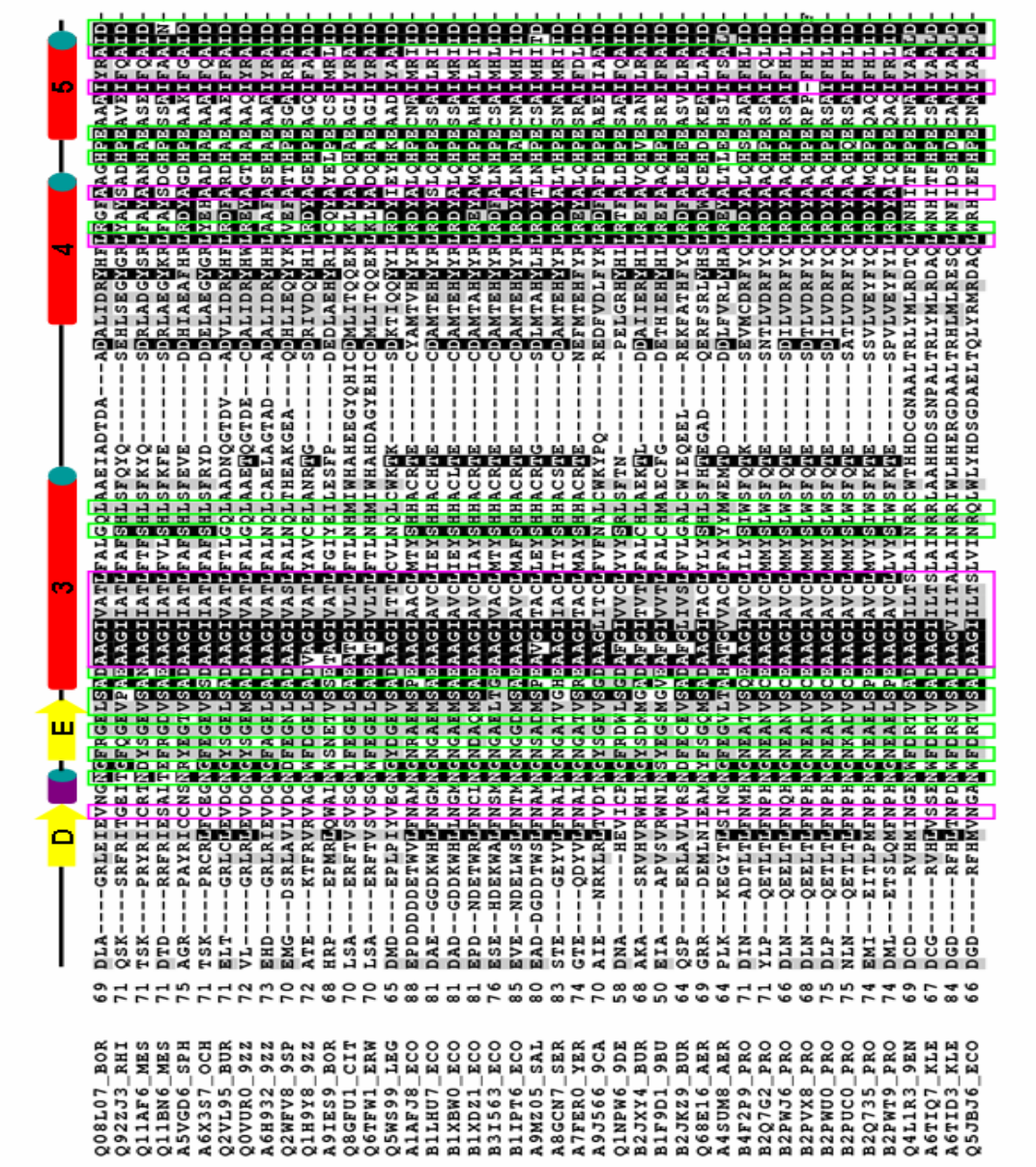


Fig. VI. *Genetic map of the 41 268 bp IncP-1b plasmid pBP136 (Kamachi et al., 2006). Grey arrows indicate directions of transcription. White arrows indicate positions of insertion (mobile genetic elements and phenotypic markers) inother IncP-1b plasmids. The position of klcA and klcB genes is marked by a red box.*