

# Extensive DNA mimicry by the *ArdA* anti-restriction protein and its role in the spread of antibiotic resistance

Stephen A. McMahon<sup>1</sup>, Gareth A. Roberts<sup>2</sup>, Kenneth A. Johnson<sup>1</sup>, Laurie P. Cooper<sup>2</sup>, Huanting Liu<sup>1</sup>, John H. White<sup>2</sup>, Lester G. Carter<sup>1</sup>, Bansi Sanghvi<sup>2</sup>, Muse Oke<sup>1</sup>, Malcolm D. Walkinshaw<sup>3</sup>, Garry W. Blakely<sup>3</sup>, James H. Naismith<sup>1,\*</sup> and David T. F. Dryden<sup>2,\*</sup>

<sup>1</sup>Centre for Biomolecular Science, The University, St Andrews, KY16 9ST, <sup>2</sup>School of Chemistry and

<sup>3</sup>School of Biological Sciences, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3JJ, UK

Received March 2, 2009; Revised May 13, 2009; Accepted May 18, 2009

## ABSTRACT

The *ardA* gene, found in many prokaryotes including important pathogenic species, allows associated mobile genetic elements to evade the ubiquitous Type I DNA restriction systems and thereby assist the spread of resistance genes in bacterial populations. As such, *ardA* contributes to a major healthcare problem. We have solved the structure of the *ArdA* protein from the conjugative transposon Tn916 and find that it has a novel extremely elongated curved cylindrical structure with defined helical grooves. The high density of aspartate and glutamate residues on the surface follow a helical pattern and the whole protein mimics a 42-base pair stretch of B-form DNA making *ArdA* by far the largest DNA mimic known. Each monomer of this dimeric structure comprises three alpha-beta domains, each with a different fold. These domains have the same fold as previously determined proteins possessing entirely different functions. This DNA mimicry explains how *ArdA* can bind and inhibit the Type I restriction enzymes and we demonstrate that 6 different *ardA* from pathogenic bacteria can function in *Escherichia coli* hosting a range of different Type I restriction systems.

## INTRODUCTION

Sequence analysis of bacterial genomes has demonstrated that horizontal gene transfer (HGT) is a fundamental mechanism for driving diversity and evolution. The transmission of DNA to bacterial cells that are not direct

descendants of the donor is often achieved via mobile genetic elements such as bacteriophage, plasmids and conjugative transposons. Mobilization of these elements can lead to the spread of antimicrobial resistance in clinical environments and in the wider community (1–5). Human morbidity and mortality as a result of this trend is mirrored by the economic cost of, for example, the spread of herbicide resistance genes in crop pathogens (6).

Over 50% of eubacteria and archaea contain the genes for one or more of the four classes of known DNA restriction and restriction-modification (RM) systems (7–10). In laboratory experiments, it is clear that the function of these RM systems is to protect the host cell from invasion by foreign DNA. RM systems work by recognizing specific DNA sequences and triggering an endonuclease activity which rapidly cleaves the foreign DNA allowing facile destruction by exonucleases.

Despite the demonstrated efficacy of the RM systems, genome analysis of pathogenic bacteria from both clinical and environment settings make it abundantly clear that HGT by transduction, transformation and conjugation is extremely common. The process is not only widespread within species but also between species where RM systems are operative. It is established that HGT is directly responsible for the spread of resistance genes (11). It is therefore important in understanding and tackling antibiotic resistance to ascertain the mechanism by which HGT circumvents such an apparently effective defence. One most likely explanation has come from the identification of potential anti-RM genes within mobile genetic elements such as phage, plasmids and transposons (12,13). These anti-RM systems seem to have been acquired and maintained by the host organism and their occasional activation of such genes weakens or negates the RM defence system allowing further HGT (14,15).

\*To whom correspondence should be addressed. Tel: +44 (0)131 650 4735; Email: david.dryden@ed.ac.uk

Correspondence may also be addressed to James H. Naismith. Tel: +44 (0)1334 463792; Email: naismith@st-and.ac.uk

The first two authors should be regarded as joint First Authors.

© 2009 The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Characterized anti-RM systems include the gene 0.3 protein, *Ocr*, of phage T7 (16–20) and the *ArdA* and *ArdB* proteins very commonly found on conjugative plasmids and transposons in a large range of prokaryotes (13,21–27). The *Ocr* protein is relatively restricted in its distribution (28,29) but is the most fully characterized of anti-RM proteins. Structural analysis suggests it to be a structural mimic of DNA and it, similar to most anti-RM systems characterized thus far, targets the Type I RM systems (17,18,30). Other mechanisms described for evading host RM systems include loss of DNA target sequences, modification of the DNA, proteolysis of RM systems and the hydrolysis of RM cofactors and have been extensively reviewed (10,12).

Type I RM systems are a very widespread and effective defence system against foreign DNA (31) perhaps explaining their selection as targets for anti-RM systems. Biochemically characterized Type I RM enzymes comprise two restriction (*HsdR*, *R*) subunits, two modification (*HsdM*, *M*) subunits and one DNA sequence specificity (*HsdS*, *S*) subunit in a single ~440-kDa complex (32,33). The methyltransferase (MTase) core  $M_2S_1$  modifies hemimethylated target ‘specificity’ sequences and triggers the *R* subunits to act when unmethylated targets are recognized. The restriction reaction involves extensive ATP hydrolysis to drive DNA translocation by the helicase motifs in each *R* subunit until an endonuclease activity is activated at a site on the DNA molecule distant from the initial specificity sequence. These RM systems can be grouped into different families, IA to IE so far, defined by subunit complementation *in vivo*, DNA hybridization, antibody cross reactivity and, to some extent, amino acid sequence comparison (34,35).

Despite the wide distribution of *ardA* genes in many important pathogens (24), there is very little biochemical data on their mode of action and behaviour in solution (26,27) and no atomic structure. We now demonstrate that *ArdA* from a number of pathogens is very effective against the archetypal Type I RM systems of *Escherichia coli*. We also report the first structure of an *ArdA* protein and its implications for the structure of Type I RM enzymes. *ArdA* forms an extremely elongated molecule with a highly charged surface. Its structure is reminiscent of ~42 bp of B-form DNA making it the largest DNA mimic yet characterized. It has already been demonstrated that *ArdA* binds to the core MTase of a Type I RM enzyme (26,27). The *ArdA* structure allows a rationalization of this behaviour.

## MATERIALS AND METHODS

### *Escherichia coli* strains

JM109 was used as a general cloning strain (Promega Madison, WI, USA). The expression strain BL21(DE3) was purchased from Invitrogen (Groningen, The Netherlands). NM1261 (restriction and modification negative;  $r^-m^-$ ), NM1049 (EcoKI RM system, Type IA), NK354 (EcoAI RM system, Type IB), NK402 (EcoR124I RM system, Type IC) and NM1009 (StySBLI RM system, Type ID) were a kind gift of Professor Noreen E. Murray (School of Biology, University of Edinburgh, UK).

The NK and NM strains were converted to DE3 lysogens using the  $\lambda$ DE3 lysogenization kit (Invitrogen) with some modifications. Where necessary, the Selection phage provided in the kit was replaced by NM848 ( $h^{82}imm^{21}cI$ , a kind gift from Noreen Murray) which possesses the appropriate host range. Integration of  $\lambda$ DE3 was verified by assessing their ability to support growth of the T7 Tester phage (supplied in the kit).

### Genes for production of *ArdA* protein

Four *ardA* genes were identified in the public sequence database held by the National Center for Biotechnology Information (Bethesda, MD, USA) (Table 1). The gene sequences were then synthesized by GeneArt (Regensburg, Germany) using an optimized codon usage pattern for expression in *E. coli*. To facilitate cloning, an *NdeI* restriction site was engineered to overlap the ATG start codon of each synthetic gene and a *HindIII* site was also included 6 bp downstream of the stop codon. The synthesized genes were ligated into pET24a at the *NdeI* and *HindIII* sites of the expression vector. The anticipated sequence was subsequently verified by DNA sequencing on both strands. Two further *ardA* genes from transposon Tn916 (open reading frame 18 in Tn916) and *Bacteroides fragilis* NCTC9343 (open reading frame BF1222) (25,36) were amplified and ligated into pTrc99a (37) as previously described (27). Amino acid substitutions were constructed by site-directed mutagenesis using a QuikChange kit (Stratagene, La Jolla, CA, USA) as per the manufacturer's instructions. Substitutions were confirmed by sequencing.

*Escherichia coli* BL21(DE3) were transformed with the pET24a or pTrc99a-based expression constructs and plated onto LB agar containing 100  $\mu$ g/ml kanamycin or 100  $\mu$ g/ml carbenicillin, respectively. Single colonies were picked and grown in LB medium supplemented with 100  $\mu$ g/ml of the appropriate antibiotic to an  $OD_{600}$  of

**Table 1.** The source of the *ardA* genes investigated

Reference	Organism	Gene ID	Amino acid length	Predicted pI	Optimal protein expression temperature, °C
25	<i>Enterococcus faecalis</i>	Tn916 (ORF18)	165	3.91	37
36	<i>Enterococcus faecalis</i> V583	EF2335	166	3.91	30
37	<i>Staphylococcus aureus</i> Mu50	SAV0405	166	3.92	30
38	<i>Clostridium difficile</i> 630	CD0376	167	3.78	37
39	<i>Streptococcus agalactiae</i> 2603V/R	SAG2011	160	4.05	30
40	<i>Bacteroides fragilis</i> NCTC 9343	BF1222	177	3.98	37

0.5. Heterologous gene expression was induced by addition of IPTG to a final concentration of 1 mM. Growth was continued for up to 4 h before harvesting. The amount of recombinant protein was initially assessed by SDS-PAGE (4–15% gradient gel) analysis. In each case, a prominent band of the anticipated molecular mass was clearly visible in the induced cell extract, which was absent from the control experiments conducted using the corresponding plasmid vector without insert (data not shown). In order to purify the recombinant proteins, cells were grown in 10-l cultures and harvested 4 h post IPTG induction. Typically we obtained ~40 g wet cell paste which was stored at  $-20^{\circ}\text{C}$  until required.

### Assessment of *in vivo* activity of ArdA

The methods employed for assaying ArdA activity *in vivo* used the efficiency of plating of phage lambda (virulent)  $\lambda v$  on the various strains of *E. coli* have been previously described (27). All assays were performed in triplicate and at least 50 phage plaques per plate per experiment were counted. Experiments were performed on numerous days with fresh samples and control experiments performed each day. Little variation was observed during the replicate experiments. The standard deviations for the anti-restriction and anti-modification results are  $\pm 25\%$  or less depending on the particular *ardA* investigated (Supplementary Table 1). The previously published results (27) for the activity of *ardA* from Tn916 are included here for completeness. Anti-restriction was defined as the efficiency of plating (eop) obtained with strains containing the plasmid expressing the *ardA* gene divided by the eop obtained from strains containing the vector plasmid. The degree of anti-modification was assessed by determining the eop of each phage stock on the restriction proficient strain relative to the non-restricting strain. Anti-modification was defined as  $1/(\text{eop})$ .

### Structural methods

The ArdA protein from transposon Tn916 was purified as described previously (27). Purified protein was concentrated to 15 mg/ml in 10 mM Tris, pH 7.5, 150 mM NaCl and 2 mM DTT in preparation for crystallization. Initial protein crystals were attained after screening purified ArdA against 384 different precipitants in 0.1  $\mu\text{l}$  protein + 0.1  $\mu\text{l}$  precipitant hanging drops set up on the Hamilton-Rhombix-Thermo integrated crystallization system at 293 K. Initial crystallization conditions were stochastically optimized and optimum crystals were found to grow in a precipitant range between 2.27 and 2.31 mM Na-K<sub>2</sub>-phosphate, 0.1 M sodium acetate, pH 4.0 at 293 K. To phase the structure a single ArdA crystal was soaked in 50 mM dipotassium tetrachloroplatinate (II) for ~2 min before being back soaked in 2.4 mM Na-K<sub>2</sub>-phosphate, 0.1 M sodium acetate, pH 4.0, 25% ethylene glycol which acted as cryoprotectant. Native and derivative data were collected at beamline ID-29 at the ESRF, Grenoble, France. All data were processed with XDS (38), scaled in XSCALE and converted to scalepack format with XPREPX. Heavy atom sites (6  $\times$  Pt) were located using SHELXD from the SHELXC/D/E (39) suite of programs

**Table 2.** The X-ray crystallography details (numbers in brackets refer to the highest-resolution shell)

Data collection	Native	Platinum derivative
Wavelength		
Resolution ( $\text{\AA}$ )	12.5–2.8 (2.87–2.8)	17.72–3.29 (3.38–3.29)
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Temperature	100K	100K
Cell dimensions	$a = 63.8$ $b = 103.4$	$a = 63.8$ $b = 104.5$
$a, b, c$ ( $\text{\AA}$ )	$c = 173.0$	$c = 171.5$
$V_m$ ( $\text{\AA}^3/\text{Da}$ )	3.73	3.74
Solvent (4 mol/asu%)	67	67
Total number reflections	68 208	140 499
Unique reflections	27 101	33 192
$I/\sigma I$	12.6 (1.8)	9.31 (2.31)
Redundancy	2.5 (2.5)	4.2 (4.1)
Completeness (%)	93.3 (93.6)	99.2 (95.4)
Anomalous correlation (%)		40
$R_{\text{merge}}$	6.4 (56.5)	10.6 (65.3)
Refinement		
$R_{\text{work}}/R_{\text{free}}$	20.7 (30.9)/24.9 (33.8)	
No. atoms		
Protein	10 203	
Water	51	
R.m.s deviations		
Bond lengths ( $\text{\AA}$ )	0.006	
Bond angles ( $^{\circ}$ )	0.817	

using native and derivative data with anomalous signal (SIRAS) to 3.5  $\text{\AA}$ . Located sites were entered into SOLVE/RESOLVE (40,41) and 2-fold averaging was automatically detected. Electron density maps were produced at 3.5  $\text{\AA}$  and a partial model built. The C $\alpha$  coordinates from this model were used to automatically detect and employ 4-fold averaging at 3  $\text{\AA}$  in RESOLVE after which the asymmetric unit model was built by hand. The structure was refined using REFMAC5 (42). TLS parameters were introduced for each domain and non-crystallographic restraints applied throughout the refinement. XFIT (43) was used for manual model building. The quality of the structure was checked with MOLPROBITY (44). The final refined structure and data have been deposited with PDB accession code 2w82. Table 2 shows the data collection and refinement details.

### Phylogenetic analysis

Ninety-eight sequences of ArdA homologues were derived from the Pfam family PF07275 (45). A distance matrix was generated by ProtDist (bootstrap analysis used a random seed number of 3 with 100 replicates) and analysed using Quicktree (46) before output in PhyloWidget (47).

## RESULTS

### Distribution of ArdA in sequenced genomes

Genes encoding putative ArdA homologues are widespread within the bacterial kingdom and can be identified

by sequence homology in the Actinobacteria, the Cyanobacteria, the Proteobacteria and the Firmicutes (data not shown). This is consistent with the known promiscuity of conjugative plasmids and conjugative transposons such as Tn916. Interestingly the genus *Bacillus* does not, as yet, appear to contain any identifiable homologues of ArdA.

### ArdA activity *in vivo*

Previously, *in vivo* assays have been performed for a few ArdA proteins from conjugative plasmids specific for *E. coli* against the IA, B and C families of Type I RM systems (13,21–24,26,48,49). Similar anti-restriction and anti-modification activities of the Ocr protein have been determined against the IA to ID families (34,35). Ocr was found to act as an anti-restriction protein against all four families but was only effective as an anti-modification protein against the IA and ID families (16,19,30).

We have tested the *in vivo* activity of ArdA proteins from six bacterial species (25,50–53,36) (Table 1) against representatives of the four major Type I RM families (Table 3, Supplementary Table 1). We arbitrarily define that the anti-restriction and anti-modification activities should have a value >2 for the ArdA protein to be considered truly active *in vivo*. The effectiveness varies greatly between different ArdA proteins and between different Type I RM families. All of the ArdA proteins were active as both anti-restriction and anti-modification proteins against the IA family, the IB family and the ID family. ArdA showed anti-restriction activity against the IC family, although the anti-modification activity of the ArdA from *S. agalactiae* and from *B. fragilis* was unmeasurable against the IC family. Despite these variations, the fact that all the ArdA were active in blocking restriction to a measurable extent in an unnatural host without induction of high levels of protein expression suggests that these ArdA will function efficiently as anti-restriction proteins if expressed in their normal hosts and be capable of modulating HGT.

### ArdA crystal structure

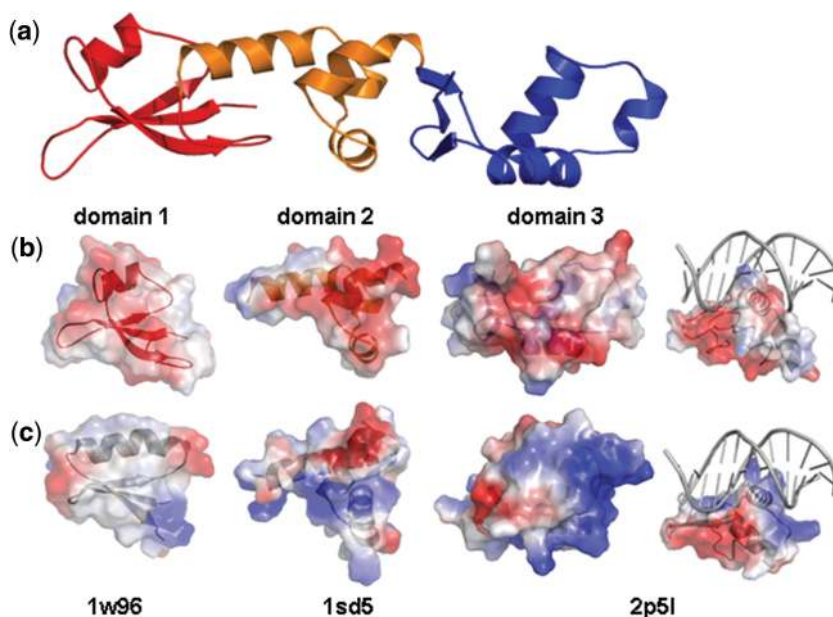
*The monomer structure.* The asymmetric unit contains four monomers of the protein forming two independent dimers. Each monomer can be decomposed into three domains: the N-terminal domain 1 (residues 3–61), the central domain 2 (62–103) and the C-terminal domain 3 (residues 104–165) (Figure 1). The three domains of the ArdA monomer are arranged in an approximately linear manner giving a very elongated molecular shape (70 Å × 20 Å) with a definite curve. The arrangement of the three domains has no counterpart in any known structure. As a consequence of the elongated arrangement, only eight residues are buried from solvent. In effect the protein lacks the conventional hydrophobic core. Superimposing the four monomers in the asymmetric unit reveals very small variations in the precise arrangement of the domains with respect to each other despite the limited contact area between the domains.

The N-terminal domain consists of a three-stranded anti-parallel  $\beta$  sheet and one short  $\alpha$  helix interspersed

**Table 3.** Effect of expression of 6 different *ardA* on the restriction and modification of bacteriophage  $\lambda$  V by 4 different Type I RM systems, each representing one of the 4 families (IA to ID) of Type I systems

Source organism	Gene expressed in host <i>E. coli</i>	Chromosomal restriction system in host <i>E. coli</i>							
		NM1049 Type IA EcoKI		NK354 Type IB EcoAI		NK402 Type IC EcoR124I		NM1009 Type ID StySBLI	
		Anti-restriction	Anti-modification	Anti-restriction	Anti-modification	Anti-restriction	Anti-modification	Anti-restriction	Anti-modification
<i>Enterococcus faecalis</i>	Tn916 ( <i>ORF18</i> )	30000	13	130	17	1500	390	1700	4200
<i>Enterococcus faecalis</i> V583	EF2335	6200	12	100	11	430	3.4	14000	160
<i>Staphylococcus aureus</i> Mu50	SAY0405	4600	4.8	100	2.8	430	6600	11000	200
<i>Clostridium difficile</i> 630	CD0376	5000	360	100	60	1900	2.5	8500	52
<i>Streptococcus agalactiae</i> 2603Y/R	SAG2011	3400	11	40	2.6	6.9	1.2	12000	1600
<i>Bacteroides fragilis</i> NCTC 9343	BF1222	68000	10000	140	13	1200	1.0	2100	1500
Plasmid vector alone	None	1.0	0.9	1.0	1.7	1.0	1.7	1.0	1.1

The host *E. coli* contained a chromosomal copy of the RM system and a plasmid expressing *ardA* at low level. An anti-restriction or anti-modification value >2 was taken to indicate ArdA activity *in vivo* and the standard deviation in these values is no more than +/– 25%.



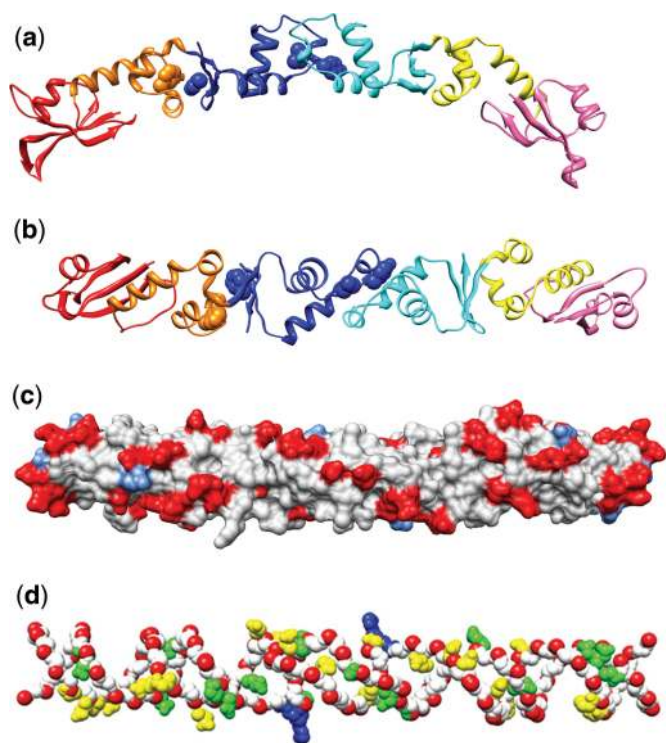
**Figure 1.** Domains in the Tn916 Orf18 ArdA monomer. (a) The N-terminal domain 1 is red, the central domain 2 is orange and the C-terminal domain is blue. (b) An electrostatic representation of each ArdA domain demonstrating their negatively charged surface. The panel on the extreme right has DNA from the structure 2p5l (DNA bound to the winged helix–turn–helix domain) modelled on to the ArdA domain 3. As can be seen, the charges on ArdA are not consistent with binding of DNA despite its fold being similar to that of 2p5l. (c) Electrostatic representations of domains homologous to ArdA, including the DNA binding motif of 2p5l. Although the overall folds of the biotin carboxylase B domain (1w96), the ANTAR domain (1sd5) and the winged helix–turn–helix domain (2p5l) are similar to those in ArdA, the charge distributions are very different.

with three large loops of 10 or more residues. The arrangement of the secondary structure elements is reminiscent of, but distinct from, the B domain in biotin carboxylase (54) (PDB 1w96), residues 248–293, from yeast (rmsd 2.1 Å over 36 residues). The central domain of ArdA is a four-helix bundle and shows weak similarity to the anti-termination ANTAR domain found in RNA-binding proteins (55). A superposition of the central domain with the closest structural match (PDB 1sd5) has rmsd of 3.2 Å for 40 C $\alpha$  atoms. The N-terminal and central domain pack against each other burying around 1700 Å<sup>2</sup> of surface area. The junction of the two domains creates a cleft with approximate dimensions 15 × 10 Å, which is partially occluded by the 12-residue loop connecting the two domains. The C-terminal domain has a three-stranded  $\beta$  sheet and three  $\alpha$  helices packed together in a manner that creates a groove in the structure ~11 Å wide. The C-terminal residue of the protein (Y165) is at the very end of the middle  $\beta$  strand. The helical arrangement in the C-terminal domain contains a component of the winged helix–turn–helix motif. Superposition of the C-terminal domain with arginine repressor protein (56) (PDB code 2p5l) reveals 42 C $\alpha$  atoms superimpose with an rmsd of 2.1 Å. The arrangement of domains 2 and 3 occludes only 600 Å<sup>2</sup> and no cleft is observed at the juncture of domains 2 and 3.

Although domains 2 and 3 from ArdA have folds reminiscent of nucleotide binding proteins, a comparison of the electrostatic surface of ArdA shows these domains have a profoundly negative potential (the pI of ArdA is 4), the opposite of the profoundly positive potential on nucleotide binding proteins. This reversal in surface

potential would seem to rule out ArdA as a DNA- (or RNA)-binding protein. It remains an open question whether the similarity is due to evolutionary divergence or simply that the robust structural frameworks of DNA-binding proteins and ArdA have been arrived at by convergent evolution.

*The dimer structure and its mimicry of DNA.* Analysis of the structure with the program PISA (57) identifies a head-to-head dimeric arrangement as the functional biological unit (Figure 2a and b). The presence of two such independent dimers in the asymmetric unit is additional evidence for the biological significance of the arrangement. A small  $\alpha$ -helix in one monomer (residues L11 to E16) is replaced with a short loop in the other subunit. The dimer, like the monomer, is highly elongated and curved. The chord that connects the extreme ends has a length of 140 Å. There is no evidence in the crystal structure for higher order oligomers such as those observed for a DNA mimic encoded by a eukaryotic virus (58). This would appear to be inconsistent with our previous gel filtration analysis in which the molecular weight of the protein increased on going from low concentration (~0.1  $\mu$ M) to high concentration (~10  $\mu$ M) (27). The chromatography calibration relied upon globular protein standards. As the structure of ArdA is now revealed to be highly elongated such a calibration is not valid and the gel filtration data, originally interpreted as a change from a dimer at low concentration to possibly a hexamer at high concentration, should now perhaps be interpreted as a change from monomer to dimer. This would be consistent with the observation



**Figure 2.** The dimeric structure of ArdA and its distribution of negative charge clearly show mimicry of the DNA double helix. (a) The ArdA dimer with domain 1 coloured red and salmon, domain 2 coloured orange and yellow and domain 3 coloured blue and cyan. Residues F91, Y137, I138 and F161 are shown space-filled. F91 is orange, Y137 and I138 are blue and located at the dimer interface and F161 is blue and at the interface with domain 2. (b) The ArdA dimer rotated 90° from (a) looking at the convex face. The residues F91, Y137, I138 and F161 are highlighted as above. (c) The surface of ArdA in the same orientation as in (b) with acidic residues coloured red and basic residues coloured light blue. The negatively charged residues form a helical pattern across the surface. (d) An overlay of a selection of the acidic residues in the dimer, viewed from the same angle as in (b) and (c), onto the DNA duplex from the model structure of the MTase bound to DNA (32). Only the phosphate backbone of DNA is shown (red and white but with the nucleotides flipped out by the MTase for methylation shown in blue) with acidic residues from ArdA matching one strand shown in yellow and matching the other strand shown in green.

(27) that the change in molecular weight as a function of concentration followed a simple one step binding equilibrium equation. The high protein concentration used for crystallization is near the upper limit of the concentration range studied by the gel filtration experiment and the crystallization sample eluted as a single peak as expected (27).

The two ArdA dimers in the unit cell when compared to each other show minor variation along the molecular axis consistent with a flexible dimer interface. The highly elongated dimer has a very large solvent surface area of 20 000 Å<sup>2</sup>. Examining the protein surface with a 1.5-Å probe shows a wide and narrow groove running the length of the structure, akin to the major groove in DNA. In addition to the general negative potential of the surface, there are two ribbons of negative charge on a raised surface that

entwine around the entire length of the molecule (Figure 2c). This striking pattern is reminiscent of the phosphate backbone of a polynucleotide. Many of the carboxyl groups can be superimposed upon the DNA (32) structure derived from the structural model of a Type I MTase core bound to DNA (Figure 2d). The pattern of negative charge even extends across the dimer interface through the conserved residues D109, D111, D112, D115, E122, E123 and E129. This distribution and conservation of charged residues is evidence for the necessity of dimer formation for protein function and suggests that ArdA across all species will have similar structural requirements.

The residues, F91, Y137, I138 and F161 were previously identified as essential (26). The crystallographic data suggest their role is not functional *per se* but rather they maintain the fold of the protein at the dimer interface (Y137 and I138) and at the interface of domains 2 and 3 (F91 and F161) (Figure 2a and b).

The essential VF-motif (I160 F161 in the Tn916 ArdA protein) is located in the domain 3 of ArdA (26) and is part of the middle β-strand in this domain. Removal of these residues (26) would cause the β-sheet to collapse destroying the structural integrity of domain 3 and the interface with domain 2. F161 and I163, on the interface between domains 2 and 3, contact E79, L80, E83 and F91 (another essential residue) in domain 2.

The dimer interface (Figure 3) contains the anti-restriction motif (amino acids 126–140 in the Tn916 ArdA protein) identified by Belogurov and Delver (59) and the two essential residues Y137 and I138 identified by molecular genetics (26). The structure makes it clear that this motif serves a structural purpose rather than being involved directly in the anti-restriction activity. The dimer buries just under 600 Å<sup>2</sup> per monomer (<6% of the total surface area) but is predominantly hydrophobic with L127, L134 and Y137 providing the bulk of the contacts. There are four hydrogen bonds; two involving main chain to main chain contacts (Y137 to D139) and two side chain to side chain contacts (Y137 to D146). Y137 and I138 are at the core of the dimer interface with Y137 making two H-bond contacts with D139 and D146 of the opposing monomer and I138 acting to stabilize the loop regions either side of the interface helix through interaction with V130 on the same monomer.

### The structure of ArdA compared to other ArdA sequences

Figure 4 shows an alignment of the amino acid sequences of the ArdA investigated in this study and by others (13,21–27). As can be seen, there is considerable variability in sequence in this set but the pattern of charged residues and the two motifs identified previously (26,59) are well conserved, indicating that all ArdA are very likely to have the same structure and to operate in the same manner. Biochemical data on the ArdA listed in Table 1 (Roberts, G.A. and Sanghvi, B., unpublished data) when compared to that from Tn916 ArdA (27), support this assertion.

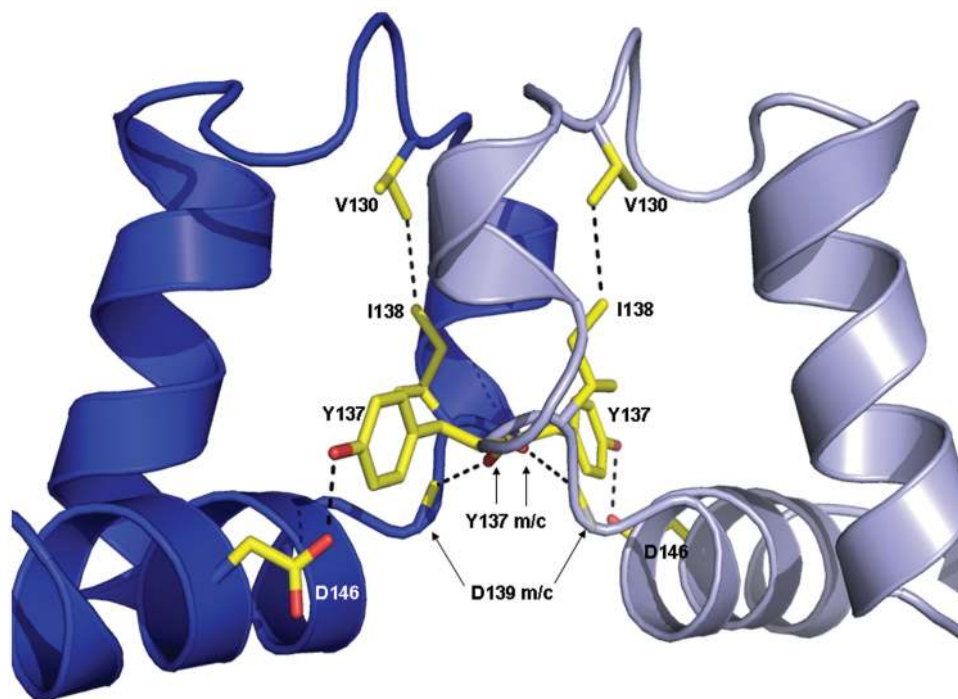


Figure 3. The ArdA dimer interface showing the residues involved in H-bonding (m/c indicates a main chain H-bond).

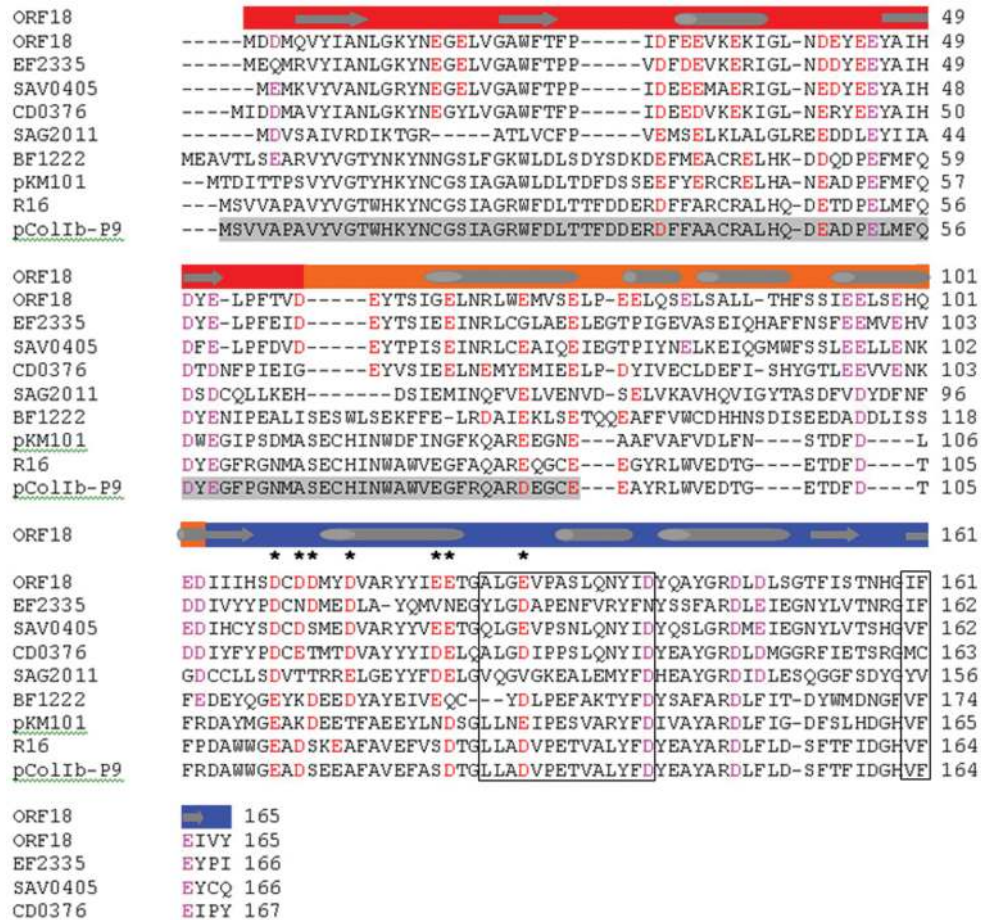
## DISCUSSION

### The interaction of ArdA with the EcoKI Type I RM enzyme: comparison of ArdA with the DNA footprint of the enzyme

The ArdA dimer appears to mimic ~42 bp of bent B-form DNA. This is comparable in length to the footprint of the EcoKI Type IA RM enzyme, without its cofactors, on DNA (60,61). In comparison, the Ocr dimer from phage T7 mimics only ~24 bp, similar in length to the ~30-bp footprint of the Type I RM enzyme in the presence of its cofactors and to the footprint of the MTase core, M.EcoKI, of the Type I RM enzyme (60–62). Binding experiments for the interaction of Ocr with M.EcoKI indicated that Ocr was engulfed by the MTase core (63) and this has recently been confirmed by image reconstruction from electron micrographs of an M.EcoKI–Ocr complex (32). The typical DNA target for a Type I RM enzyme is ~14-bp long and bipartite, e.g. EcoKI recognizes 5'-AACNNNNNGTGC-3', and lies centrally in the experimental DNA footprints. Figure 5a shows the M.EcoKI–DNA model built using the EM reconstruction (32) with the DNA extended in length to 42 bp. Figure 5b shows the ArdA structure within the M.EcoKI structure. This is derived by taking advantage of the superposition of carboxylates in ArdA and phosphates in DNA (Figure 2d). A correspondence between the domains of ArdA and the regions of DNA is obvious: domain 3 overlaps the EcoKI target sequence, domain 2 contacts the extremities of the DNA-binding groove in M.EcoKI and domain 1 projects beyond the M.EcoKI structure. The footprint of the entire EcoKI RM enzyme is ~42 bp (61), hence it would appear that domain 1 mimics the

region of DNA in contact with the HsdR restriction subunits of EcoKI (these HsdR subunits are of course absent from the MTase structure). This would also be consistent with the model proposed recently from the structure of HsdR (33). This model approximately placed the HsdR against the ends of the DNA-binding groove in the MTase. Their placement would be consistent with the ArdA–MTase model in Figure 5b and earlier footprinting data. However, the exact orientation of the two HsdR with respect to the MTase is still unknown and large changes in the DNA footprint of the EcoKI nuclease (61) suggest large conformational changes. In addition, the 12-residue loop linking domains 1 and 2 may indicate that domain 1 can move relative to the rest of the protein to better contact the HsdR subunits. Domain 1 is not essential for anti-restriction as it can be deleted (26) indicating that the key aspect of anti-restriction by ArdA is the binding to the MTase core using domains 2 and 3.

The degree of anti-restriction observed *in vivo* varied considerably between different ArdA and different Type I families (from 7 to 70 000). The Type I RM systems are all expressed from a single chromosomal copy of their genes but because of the different stabilities of the Type I enzymes (63–65), the number of active enzyme molecules in the different host strains may vary considerably. The number of EcoKI Type I enzyme has been estimated as <100 per cell (66,67). The number of target sequences on phage lambda for the RM enzymes also varies with 5, 1, 14 and 3 targets for the IA, IB, IC and ID systems, respectively. In addition, the expression levels of the different *ardA* genes may vary. These factors may account for the unpredictable differences in anti-restriction. The IA, ID and, taking into account the single target site, IB systems



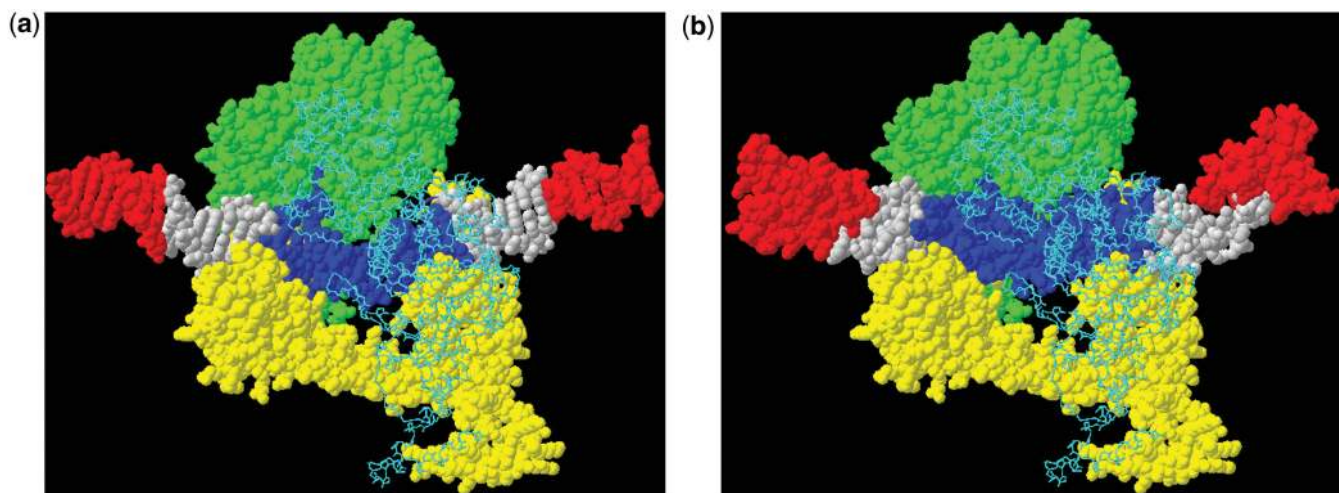
**Figure 4.** CLUSTALW multiple sequence alignment of ArdA proteins in Table 1 plus ArdA from several previously studied conjugative plasmids from different incompatibility groups. The colours above each residue in the sequence alignment indicate the domain of the Tn916 Orf18 ArdA; red is domain 1, orange is domain 2 and blue is domain 3. Secondary structure from the crystallographic analysis is shown above the sequence as cylinders for helix and block arrows for strands. The secondary structure assignment is from the first monomer of the dimer. The second monomer lacks the first helix. Acidic residues that entwine ArdA are highlighted in red and magenta. Those conserved across the dimer interface are highlighted with asterisk. Highlighted in an open box are the ‘anti-restriction motif’ (59) and the VF-motif both essential for activity (26). The region of ArdA from the conjugative plasmid pColIb-P9 that can be deleted without loss of anti-restriction activity is highlighted in grey (26).

seem most susceptible to anti-restriction by ArdA. Taking into account the number of potential target sequences for the IC system, the IC system seems rather resistant to ArdA. This is perhaps because even a single interaction of the IC enzyme with one of the numerous target sequences on the phage DNA would lead to restriction and there is insufficient ArdA to knock out every enzyme molecule. However, it is also noticeable that the IC enzyme is most able to carry out its methylation function as shown by a low anti-modification value for most of the ArdA (anti-modification varies between 1 and 10000). This difference in the ability of ArdA to inhibit methylation of phage DNA by Type I families may reflect different binding affinities of the ArdA proteins for the Type I enzymes. At present, binding affinity is not well characterized but for the ArdA–EcoKI interaction it has been estimated that the dissociation constant,  $K_d$ , is  $<1$  nM *in vitro* (27) and  $\sim 170$  nM *in vivo* (68). This is not as strong as the interaction between Ocr and EcoKI, where the  $K_d$  is  $\sim 50$  pM (69); therefore, perhaps ArdA is just not such a strong competitive inhibitor as Ocr. A weaker equilibrium

binding affinity would allow a greater fraction of the RM enzymes to remain active *in vivo*. These questions will only be addressable with extensive *in vitro* analysis of different ArdA/RM enzyme combinations.

The ArdA structure and that of other DNA mimics, e.g. the phage lambda gam protein which inhibits the RecBCD exonuclease (70), has implications for HGT. The ArdA structure is a novel fold comprising three small domains, each of a known fold, arranged in a row in each monomer. The dimer interface is very small and the overall protein is extremely elongated and decorated with negative side chains arranged to mimic the phosphate backbone of DNA. The evolution of ArdA is curious as two of its domains show a related fold but the opposite charge to oligonucleotide binding proteins. The sequential acquisition of single-point mutations to code for a predominance of acidic amino acids and DNA mimicry is certain to be a rare event. This may be the reason that few mimics of extended DNA structures have been found. This rarity may also mean that whenever the mimics do arise, they spread rapidly throughout their ecological niche, for





**Figure 5.** The domains of ArdA correspond to different parts of the DNA bound by EcoKI. (a) Forty-two base pairs of DNA within the EcoKI MTase (HsdS subunit yellow, HsdM subunits in green space-fill and cyan backbone trace) (32). The DNA contains a substantial bend matching the bend in ArdA. The DNA target sequence (central 14 bp) in blue plus the regions shown in white (7 bp on either side of the target sequence), totalling 28 bp, are protected in footprinting assays by EcoKI MTase or the EcoKI nuclease in the presence of ATP (46–48). The regions of DNA shown in red are the additional 7-bp regions protected in the footprinting assay by the EcoKI nuclease when no cofactors are present. The protection of these red regions of the DNA is attributed to the presence of the additional two HsdR subunits in the nuclease. The protection of these red regions of the DNA is attributed to the presence of the additional two HsdR subunits in the nuclease. (b) The ArdA domains of the dimer coloured red, white and blue for domains 1, 2 and 3 (each is approximately equivalent in length to 7 bp of DNA) replacing the DNA in the EcoKI MTase structural model. A clear correspondence between the domains of ArdA and the regions of DNA defined by footprinting experiments is visible.

example, ArdA is now widely spread in the eubacteria on plasmids and transposons (24,26,27,59), phage T7 Ocr is found in other related phage (28,29) and the MfpA pentapeptide repeat protein (71) involved in quinolone resistance is spreading rapidly (4). Recently, a DNA mimic has been found in a highly infectious eukaryotic virus of crustaceans (58). One may anticipate that this mimic will also spread rapidly.

All ArdA tested here are active against restriction by Type I RM systems, so if they are expressed in their natural host then they will certainly function as anti-restriction proteins to assist HGT allowing, for example, the acquisition of resistance genes. HGT would be rampant without the existence of RM systems resulting in disruption of the host genome. An interplay between anti-restriction systems and RM systems is clearly beneficial for the evolution of bacterial species.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank Professor Noreen Murray and numerous colleagues in her laboratory for countless discussions and Dimitra Serfiotis-Mitsa for the expression plasmid and assistance with early crystallization trials.

## FUNDING

The Biotechnology and Biological Sciences Research Council (BB/D001870/1 to D.T.F.D.); the Wellcome Trust (GR080463MA to D.T.F.D., M.D.W., J.H.N. and

G.B.); the Darwin Trust of Edinburgh (a PhD studentship to B.S.); and the BBSRC (BB/S/B14450) and the Scottish Funding Council (SULSA and SSPF). Funding for open access charge: Wellcome Trust.

*Conflict of interest statement.* None declared.

## REFERENCES

- Bennett, P.M. (2008) Plasmid encoded antibiotic resistance: acquisition and transfer of antibiotic resistance genes in bacteria. *Br. J. Pharmacol.*, **153**, S347–S357.
- Lindsay, J.A. and Holden, M.T. (2006) Understanding the rise of the superbug: investigation of the evolution and genomic variation of *Staphylococcus aureus*. *Funct. Integr. Genomics*, **6**, 186–201.
- Kennedy, A.D., Otto, M., Braughton, K.R., Whitney, A.R., Chen, L., Mathema, B., Mediavilla, J.R., Byrne, K.A., Parkins, L.D., Tenover, F.C. *et al.* (2008) Epidemic community-associated methicillin-resistant *Staphylococcus aureus*: recent clonal expansion and diversification. *Proc. Natl Acad. Sci. USA*, **105**, 1327–1332.
- Robicsek, A., Jacoby, G.A. and Hooper, D.C. (2006) The worldwide emergence of plasmid-mediated quinolone resistance. *Lancet Infect. Dis.*, **6**, 629–640.
- Wright, G.D. (2007) The antibiotic resistome: the nexus of chemical and genetic diversity. *Nat. Rev. Microbiol.*, **5**, 175–186.
- Vedler, E., Vahter, M. and Heinaru, A. (2004) The completely sequenced plasmid pEST4011 contains a novel IncP1 backbone and a catabolic transposon harboring *tfd* genes for 2,4-dichlorophenoxyacetic acid degradation. *J. Bacteriol.*, **186**, 7161–7174.
- Roberts, R.J. (2005) How restriction enzymes became the workhorses of molecular biology. *Proc. Natl Acad. Sci. USA*, **102**, 5905–5908.
- Roberts, R.J., Belfort, M., Bestor, T., Bhagwat, A.S., Bickle, T.A., Bitinaite, J., Blumenthal, R.M., Degtyarev, S.Kh., Dryden, D.T.F., Dybvig, K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
- Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2007) REBASE—enzymes and genes for DNA restriction and modification. *Nucleic Acids Res.*, **35**, D269–D270.

10. Tock, M.R. and Dryden, D.T.F. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
11. Thomas, C.M. and Nielsen, K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nat. Rev. Microbiol.*, **3**, 711–721.
12. Bickle, T.A. and Krüger, D.H. (1993) Biology of DNA restriction. *Microbiol. Rev.*, **57**, 434–450.
13. Wilkins, B.M. (2002) Plasmid promiscuity: meeting the challenge of DNA immigration control. *Env. Microbiol.*, **4**, 495–500.
14. King, G. and Murray, N.E. (1995) Restriction alleviation and modification enhancement by the Rac prophage of *Escherichia coli* K-12. *Mol. Microbiol.*, **16**, 769–777.
15. Webb, J.L., King, G., Ternent, D., Titheradge, A.J. and Murray, N.E. (1996) Restriction by EcoKI is enhanced by co-operative interactions between target sequences and is dependent on DEAD box motifs. *EMBO J.*, **15**, 2003–2009.
16. Bandyopadhyay, P.K., Studier, F.W., Hamilton, D.L. and Yuan, R. (1985) Inhibition of the Type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.*, **182**, 567–578.
17. Dryden, D.T.F. (2006) DNA mimicry by proteins and the control of enzymatic activity on DNA. *Trends Biotechnol.*, **24**, 378–382.
18. Dryden, D.T.F. and Tock, M.R. (2006) DNA mimicry by proteins. *Biochem. Soc. Trans.*, **34**, 317–319.
19. Krüger, D.H., Schroeder, C., Hansen, S. and Rosenthal, H.A. (1977) Active protection by bacteriophages T3 and T7 against *E. coli* B- and K-specific restriction of their DNA. *Mol. Gen. Genet.*, **153**, 99–106.
20. Stephanou, A.S., Roberts, G.A., Tock, M.R., Pritchard, E.H., Turkington, R., Nutley, M., Cooper, A. and Dryden, D.T.F. (2009) A mutational analysis of DNA mimicry by Ocr, the gene 0.3 antirestriction protein of bacteriophage T7. *Biochem. Biophys. Res. Commun.*, **378**, 129–132.
21. Belogurov, A.A., Yussifov, T.N., Kotova, V.U. and Zavilgelsky, G.B. (1985) The novel gene(s) ARD of plasmid pKM101: alleviation of EcoK restriction. *Mol. Gen. Genet.*, **198**, 509–513.
22. Belogurov, A.A., Delver, E.P. and Rodzevich, O.V. (1993) Plasmid pKM101 encodes two nonhomologous antirestriction proteins (ArdA and ArdB) whose expression is controlled by homologous regulatory sequences. *J. Bacteriol.*, **175**, 4843–4850.
23. Belogurov, A.A., Delver, E.P., Agafonova, O.V., Belogurova, N.G., Lee, L. and 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.
24. Chilley, P.M. and Wilkins, B.M. (1995) Distribution of the arDA family of antirestriction genes on conjugative plasmids. *Microbiology*, **141**, 2157–2164.
25. Flannagan, S.E., Zitzow, L.A., Su, Y.A. and Clewell, D.B. (1994) Nucleotide sequence of the 18-kb conjugative transposon Tn916 from *Enterococcus faecalis*. *Plasmid*, **32**, 350–354.
26. Nekrasov, S.V., Agafonova, O.V., Belogurova, N.G., Delver, E.P. and 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.
27. Serfiotis-Mitsa, D., Roberts, G.A., Cooper, L.P., White, J.H., Nutley, M., Cooper, A., Blakely, G.W. and 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.
28. Garcia, E., Elliott, J.M., Ramanculov, E., Chain, P.S., Chu, M.C. and Molineux, I.J. (2003) The genome sequence of *Yersinia pestis* bacteriophage phiA1122 reveals an intimate history with the coliphage T3 and T7 genomes. *J. Bacteriol.*, **185**, 5248–5262.
29. Scholl, D., Kieleczawa, J., Kemp, P., Rush, J., Richardson, C.C., Merril, C., Adhya, S. and Molineux, I.J. (2004) Genomic analysis of bacteriophages SP6 and K1-5, an estranged subgroup of the T7 supergroup. *J. Mol. Biol.*, **335**, 1151–1171.
30. Walkinshaw, M.D., Taylor, P., Sturrock, S.S., Atanasiu, C., Berge, T., Henderson, R.M., Edwardson, J.M. and Dryden, D.T.F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell*, **9**, 187–194.
31. Murray, N.E. (2002) 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self. *Microbiology*, **148**, 3–20.
32. Kennaway, C.K., Obarska-Kosinska, A., White, J.H., Tuszyńska, I., Cooper, L.P., Bujnicki, J.M., Trinick, J. and Dryden, D.T.F. (2009) The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein. *Nucleic Acids Res.*, **37**, 762–770.
33. Lapkouski, M., Panjikar, S., Janscak, P., Smananova, I.K., Carey, J., Ettrich, R. and Csefalvay, E. (2009) Structure of the motor subunit of type I restriction-modification complex EcoR124I. *Nat. Struct. Mol. Biol.*, **16**, 94–95.
34. Titheradge, A.J., King, J., Ryu, J. and Murray, N.E. (2001) Families of restriction enzymes: an analysis prompted by molecular and genetic data for type I restriction and modification systems. *Nucleic Acids Res.*, **20**, 4195–4205.
35. Chin, V., Valinluck, V., Magaki, S. and Ryu, J. (2004) KpnBI is the prototype of a new family (IE) of bacterial type I restriction-modification system. *Nucleic Acids Res.*, **32**, e138.
36. Cerdano-Tarraga, A.M., Patrick, S., Crossman, L.C., Blakely, G., Abratt, V., Lennard, N., Poxton, I., Duerden, B., Harris, B., Quail, M.A. *et al.* (2005) Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science*, **307**, 1463–1465.
37. Amann, E., Ochs, B. and Abel, K.J. (1988) Tightly regulated tac promoter vectors useful for the expression of unfused and fused proteins in *Escherichia coli*. *Gene*, **69**, 301–315.
38. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.*, **26**, 795–800.
39. Sheldrick, G.M. (2008) A short history of SHELX. *Acta Cryst. A.*, **64**, 112–122.
40. Terwilliger, T.C. and Berendzen, J. (1999) Automated MAD and MIR structure solution. *Acta Cryst. D.*, **55**, 849–861.
41. Terwilliger, T.C. (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Cryst. D.*, **59**, 38–44.
42. CCP4 (1994) Collaborative computational project number 4. *Acta Cryst. D.*, **50**, 760–763.
43. McRee, D.E. (1999) XtalView/Xfit – a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.*, **125**, 156–165.
44. Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B. III, Snoeyink, J., Richardson, J.S. *et al.* (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, **35**, W375–W383.
45. Finn, R.D., Tate, J., Mistry, J., Coghill, P.C., Sammut, J.S., Hotz, H.R., Ceric, G., Forslund, K., Eddy, S.R., Sonnhammer, E.L. *et al.* (2008) The Pfam protein families database. *Nucleic Acids Res.*, **36**, D281–D288.
46. Howe, K., Bateman, A. and Durbin, R. (2002) QuickTree: building huge Neighbour-Joining trees of protein sequences. *Bioinformatics*, **18**, 1546–1547.
47. Jordan, G.E. and Piel, W.H. (2008) PhyloWidget: web-based visualizations for the tree of life. *Bioinformatics*, **24**, 1641–1642.
48. Thomas, A.T., Brammar, W.J. and Wilkins, B.M. (2003) Plasmid R16 ArdA protein preferentially targets restriction activity of the Type I restriction-modification system EcoKI. *J. Bacteriol.*, **185**, 2022–2025.
49. Zavil’gel’skii, G.B., Letuchaja, T.A. and Rastorguev, S.M. (2006) Antirestriction and antimodification activities of the ArdA protein encoded by the IncII transmissible plasmids R-64 and ColIb-P9. *Russ. J. Genet.*, **42**, 252–258.
50. Paulsen, I.T., Banerjee, L., Myers, G.S., Nelson, K.E., Seshadri, R., Read, T.D., Fouts, D.E., Eisen, J.A., Gill, S.R., Heidelberg, J.F. *et al.* (2003) Role of mobile DNA in the evolution of vancomycin-resistant *Enterococcus faecalis*. *Science*, **299**, 2071–2074.
51. Kuroda, M., Ohta, T., Uchiyama, I., Baba, T., Yuzawa, H., Kobayashi, I., Cui, L., Oguchi, A., Aoki, K., Nagai, Y. *et al.* (2001) Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet*, **357**, 1225–1240.
52. Sebailhia, M., Wren, B.W., Mullany, P., Fairweather, N.F., Minton, N., Stabler, R., Thomson, N.R., Roberts, A.P., Cerdeño-Tarraga, A.M.,

- Wang, H. *et al.* (2006) The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.*, **38**, 779–786.
53. Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S. *et al.* (2005) Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc. Natl Acad. Sci. USA*, **102**, 13950–13955.
54. Shen, Y., Volrath, S.L., Weatherly, S.C., Elich, T.D. and Tong, L. (2004) A mechanism for the potent inhibition of eukaryotic acetyl-coenzyme A carboxylase by soraphen A, a macrocyclic polyketide natural product. *Mol. Cell*, **16**, 881–891.
55. Morth, J.P., Feng, V., Perry, L.J., Svergun, D.I. and Tucker, P.A. (2004) The crystal and solution structure of a putative transcriptional antiterminator from *Mycobacterium tuberculosis*. *Structure*, **12**, 1595–1605.
56. Garnett, J.A., Marincs, F., Baumberg, S., Stockley, P.G. and Phillips, S.E. (2008) Structure and function of the arginine repressor-operator complex from *Bacillus subtilis*. *J. Mol. Biol.*, **379**, 284–298.
57. Krissinel, E. and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.*, **372**, 774–797.
58. Wang, H.C., Wang, H.C., Ko, T.P., Lee, Y.M., Leu, J.H., Ho, C.H., Huang, W.P., Lo, C.F. and Wang, A.H. (2008) White spot syndrome virus protein ICP11: A histone-binding DNA mimic that disrupts nucleosome assembly. *Proc. Natl Acad. Sci. USA*, **105**, 20758–20763.
59. Belogurov, A.A. and 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 anti-restriction functions. *Nucleic Acids Res.*, **23**, 785–787.
60. Davies, G.P., Martin, I.V., Sturrock, S.S., Cronshaw, A., Murray, N.E. and Dryden, D.T.F. (1999) On the structure and operation of type I DNA restriction enzymes. *J. Mol. Biol.*, **290**, 565–579.
61. Powell, L.M., Dryden, D.T.F. and Murray, N.E. (1998) Sequence-specific DNA binding by EcoKI, a type IA DNA restriction enzyme. *J. Mol. Biol.*, **283**, 963–976.
62. Powell, L.M., Dryden, D.T.F., Willcock, D.F., Pain, R.H. and Murray, N.E. (1993) DNA recognition by the EcoK methyltransferase. The influence of DNA methylation and the cofactor S-adenosyl-L-methionine. *J. Mol. Biol.*, **234**, 60–71.
63. Suri, B., Shepherd, J.C. and Bickle, T.A. (1984) The EcoA restriction and modification system of *Escherichia coli* 15T<sup>-</sup>: enzyme structure and DNA recognition sequence. *EMBO J.*, **3**, 575–579.
64. Dryden, D.T.F., Cooper, L.P., Thorpe, P.H. and Byron, O. (1997) The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications. *Biochemistry*, **36**, 1065–1076.
65. Janscak, P., Dryden, D.T.F. and Firman, K. (1998) Analysis of the subunit assembly of the typeIC restriction-modification enzyme EcoR124I. *Nucleic Acids Res.*, **26**, 4439–4445.
66. Heip, J., Rolfe, B. and Schell, J. (1974) Abolition of host cell restriction by high multiplicity of phage infection. *Virology*, **59**, 356–370.
67. Kelleher, J.E. and Raleigh, E.A. (1994) Response to UV damage by four *Escherichia coli* K-12 restriction systems. *J. Bacteriol.*, **176**, 5888–5896.
68. Zavilgelsky, G.B., Kotova, V.Y. and Rastorguev, S.M. (2008) Comparative analysis of anti-restriction activities of ArdA (Collb-P9) and Ocr (T7) proteins. *Biochemistry*, **73**, 906–911.
69. Atanasiu, C., Su, T.-J., Sturrock, S. S. and Dryden, D.T.F. (2002) Interaction of the Ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res.*, **30**, 3936–3944.
70. Court, R., Cook, N., Saikrishnan, K. and Wigley, D. (2007) The crystal structure of lambda-Gam protein suggests a model for RecBCD inhibition. *J. Mol. Biol.*, **371**, 25–33.
71. Hegde, S.S., Vetting, M.W., Roderick, S.L., Mitchenall, L.A., Maxwell, A., Takiff, H.E. and Blanchard, J.S. (2005) A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA. *Science*, **308**, 1480–1483.