

## ArdA proteins from different mobile genetic elements can bind to the EcoKI Type I DNA methyltransferase of *E. coli* K12

Kai Chen<sup>a,1</sup>, Marcel Reuter<sup>a,1,2</sup>, Banshi Sanghvi<sup>a,1,3</sup>, Gareth A. Roberts<sup>a,1</sup>, Laurie P. Cooper<sup>a</sup>, Matthew Tilling<sup>a</sup>, Garry W. Blakely<sup>b</sup>, David T.F. Dryden<sup>a,\*</sup>

<sup>a</sup> EaStCHEM School of Chemistry, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3JJ, UK

<sup>b</sup> Institute of Cell Biology, School of Biological Sciences, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3JR, UK

### ARTICLE INFO

#### Article history:

Received 30 September 2013

Received in revised form 11 December 2013

Accepted 13 December 2013

Available online 22 December 2013

#### Keywords:

DNA methyltransferase

ArdA protein

DNA mimic

Horizontal gene transfer

### ABSTRACT

Anti-restriction and anti-modification (anti-RM) is the ability to prevent cleavage by DNA restriction–modification (RM) systems of foreign DNA entering a new bacterial host. The evolutionary consequence of anti-RM is the enhanced dissemination of mobile genetic elements. Homologues of ArdA anti-RM proteins are encoded by genes present in many mobile genetic elements such as conjugative plasmids and transposons within bacterial genomes. The ArdA proteins cause anti-RM by mimicking the DNA structure bound by Type I RM enzymes. We have investigated ArdA proteins from the genomes of *Enterococcus faecalis* V583, *Staphylococcus aureus* Mu50 and *Bacteroides fragilis* NCTC 9343, and compared them to the ArdA protein expressed by the conjugative transposon Tn916. We find that despite having very different structural stability and secondary structure content, they can all bind to the EcoKI methyltransferase, a core component of the EcoKI Type I RM system. This finding indicates that the less structured ArdA proteins become fully folded upon binding. The ability of ArdA from diverse mobile elements to inhibit Type I RM systems from other bacteria suggests that they are an advantage for transfer not only between closely-related bacteria but also between more distantly related bacterial species.

© 2013 The Authors. Published by Elsevier B.V. Open access under [CC BY license](http://creativecommons.org/licenses/by/3.0/).

## 1. Introduction

DNA mimics are a diverse group of proteins expressed by many mobile genetic elements (MGEs) such as bacteriophage, conjugative plasmids encoding multidrug resistance, transposons and even a shrimp virus [1–4]. They can also be found incorporated in bacterial genomes on prophages and other integrative elements. Their function is to bind to DNA-binding sites on target proteins and thereby prevent the target protein from binding to DNA, its normal substrate. Thus these proteins can be considered to be competitive inhibitors.

The best understood DNA mimics are those which inhibit DNA restriction–modification (RM) enzymes and these “anti-RM” proteins allow the DNA of the MGE to successfully invade the new host bacterial cell [5]. The RM system would usually destroy the invading DNA if it lacks the appropriate pattern of DNA methylation in the DNA sequence recognised by the RM system but the rapid transcription and translation of the DNA mimic overwhelm the RM system [6,7]. Crystal structures of two DNA mimics, the Ocr protein from phage T7 and the ArdA protein from the Tn916 conjugative transposon, show elongated dimeric proteins whose surfaces are decorated with a great number of aspartate and glutamate side chains in locations corresponding to the phosphate groups on the surface of B-form DNA [8,9], Fig. 1. The mimics fit closely into the DNA binding groove of their targets, the Type I DNA RM enzymes [10,11], and prevent their cleavage of the phage or transposon DNA. Very little sequence variation between Phage T7 Ocr and its few homologues [12] is apparent even though the anti-RM activity is extraordinarily robust to extensive mutagenesis or chemical modification [13–15]. In contrast, putative *ardA* genes are very widespread on MGE within a broad range of bacteria [7,16–19]. ArdA genes are expressed from a novel single-stranded promoter as soon as the conjugative plasmid or transposon enters a new host [7,19]. The predicted ArdA proteins show considerable sequence variation and the genes often encode long N-terminal and/or C-terminal extensions [20]. These differences between Ocr and ArdA may relate to the different life styles of the parent MGE, namely a lytic phage versus a conjugative plasmid or transposon.

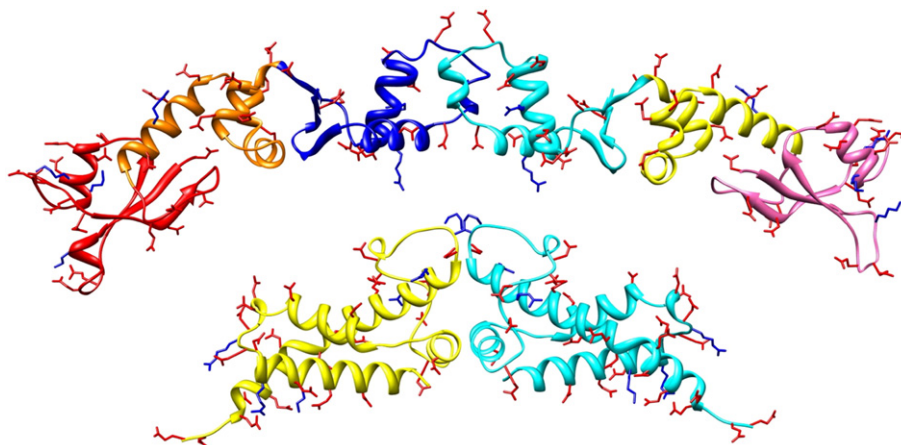
**Abbreviations:** RM, restriction–modification; anti-RM, antirestriction/antimodification; MGE, mobile genetic element; MTase, modification methyltransferase; M subunit, modification subunit; S subunit, sequence specificity subunit; Orf, open reading frame; CD, circular dichroism; GuCl, guanidinium chloride; 2-ME, 2-mercaptoethanol; SEC, size exclusion chromatography;  $K_d$ , dissociation constant

\* Corresponding author.

<sup>1</sup> Joint first authors.

<sup>2</sup> Current address: Department of Genetics, Cancer Genomics Center, Erasmus University Medical Center, PO BOX 2040, 3000 CA Rotterdam, The Netherlands.

<sup>3</sup> Current address: Sartorius Stedim Australia Pty Ltd, Unit 5, 7-11 Rodeo Drive, Dandenong South, Vic 3175, Australia.



**Fig. 1.** The upper structure shows the dimeric Orf18 ArdA protein with domain 1 coloured red and salmon, domain 2 coloured orange and yellow and domain 3 coloured blue and cyan [9]. The lower structure shows the dimeric structure of the phage T7 Ocr protein (yellow and cyan subunits) [8]. Aspartate and glutamate side chains are shown as red sticks and arginine and lysine side chains as blue sticks. The chord joining the extreme ends of the Orf18 ArdA dimer is ~140 Å in length while that joining the ends of the Ocr dimer is ~85 Å in length.

It has been shown that ArdA homologues from several MGE infecting *Bacteroides fragilis*, *Enterococcus faecalis* and *Staphylococcus aureus* can operate effectively *in vivo* against the EcoKI Type I RM system of *Escherichia coli* K12 when expressed in *E. coli* [9]. These ArdA differ considerably in amino acid sequence and essentially can be considered to be rather extreme “variants” of the Tn916 ArdA whose crystal structure is known [9]. Furthermore their bacterial hosts are classified differently. *E. coli* and *B. fragilis* are Gram-negative members of the Gammaproteobacteria and Proteobacteria Bacteroidetes groups respectively. *E. faecalis* and *S. aureus* are Gram-positive members of the Firmicutes (Lactobacillales and Bacillales groups respectively).

In this paper we explore the interaction *in vitro* of these ArdA proteins with the core modification methyltransferase (MTase, a complex of two modification, M, subunits and one sequence specificity, S, subunit) of the EcoKI RM system [10,11,21,22] and find that they all bind approximately equally well to the MTase. Binding to a partially assembled but inactive MTase composed of one M and one S subunit [21,22] is also demonstrated. In addition we find that the ArdA proteins display different structural stability ranging from a fully folded form to a partially folded form suggesting that the partially folded ArdA completes its folding upon binding to the MTase. Overall our data suggest that ArdA would assist the spread of MGE between unrelated bacterial species encoding different Type I RM systems.

## 2. Materials and methods

The *ardA* genes selected for overexpression and the molecular biological procedures have been described previously and shown to be active against Type I RM systems *in vivo* [9,16,20]. The genes selected were open reading frame (Orf) EF2335 from *E. faecalis* V583, Orf SAV0405 from *S. aureus* Mu50 and Orf BF1222 from *B. fragilis* NCTC 9343. The ArdA proteins were overexpressed and purified using DEAE anion exchange chromatography and size exclusion chromatography as described previously for the Orf18 ArdA protein expressed by the conjugative transposon Tn916 [16]. Extinction coefficients for monomer forms of the ArdA proteins were calculated from their amino acid sequences (Bfr ArdA 39,400 M<sup>-1</sup> cm<sup>-1</sup>; Mu50 ArdA 28,020 M<sup>-1</sup> cm<sup>-1</sup>; V583 ArdA 23,610 M<sup>-1</sup> cm<sup>-1</sup>; Orf18 ArdA 28,020 M<sup>-1</sup> cm<sup>-1</sup>) and used to calculate protein concentration. These coefficients are accurate to +/- 5% [23]. EcoKI MTase was prepared as described previously [21,22]. All experiments were carried out at 25 °C unless otherwise stated. Sequence alignments and determination of amino acid identity and similarity were calculated by PROMALS [24], EMBOSS Needle Alignment

([http://www.ebi.ac.uk/Tools/psa/emboss\\_needle/](http://www.ebi.ac.uk/Tools/psa/emboss_needle/)) and CLUSTAL W2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>). Sequences of Type I RM systems were obtained from REBASE [25].

Circular dichroism (CD) measurements were carried out on a Jasco Model J-180 spectropolarimeter (Jasco Corporation, Tokyo, Japan). All measurements were conducted in 20 mM sodium phosphate, 50 mM NaF, and 0.5 mM 2-mercaptoethanol (2-ME) pH 8.0. NaF has much lower absorption in the far UV than other salts. Far-UV CD spectra were measured in the range of 190–260 nm at protein concentrations of 5.6 to 5.8 μM. All the CD measurements were made at 20 °C using a 1.0 mm pathlength cell and each spectrum was the average of three individual scans run at 20 nm/min, 1 s response time and 1 nm bandwidth. The spectra were corrected for buffer contribution. Secondary structure analysis was performed using CONTIN [26] via the Dichroweb server [27].

Equilibrium unfolding as a function of guanidinium chloride (GuCl) was monitored by tryptophan fluorescence spectroscopy. A stock solution of GuCl was made up in buffer and the precise concentration was determined from the refractive index [28]. Protein (3.5 μM) in 20 mM Tris-HCl 10 mM MgCl<sub>2</sub> 7 mM 2-ME pH 8.0 was incubated with various concentrations of GuCl at 25 °C and allowed to equilibrate overnight. The fluorescence intensity was then measured for each sample using excitation at 295 nm and emission at 350 nm and 380 nm with 5 nm bandwidths on an Edinburgh Instruments FS900 fluorimeter (Edinburgh Instruments, Livingston, UK). The ratio of intensity at 350 to 380 nm was then fitted to a two-state unfolding model assuming a linear relationship between free energy of unfolding and concentration of GuCl [29]. This ratio compensates for slight differences in the protein concentration between samples.

Size exclusion chromatography (SEC) with a 30 cm long × 0.46 cm diameter Biosep-SEC-S3000 gel filtration column (Phenomenex) was used to characterise the ArdA proteins and their interaction with EcoKI MTase as described previously [30]. The buffer contains 20 mM Tris-HCl, 20 mM MES, 200 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.1 mM EDTA, and 7 mM 2-mercaptoethanol, pH 6.5. The flow rate was set to 0.5 ml/min and the injected sample volume was 40 μl. The column eluate was excited at 295 nm and the fluorescence emission continuously monitored at a wavelength of 350 nm. It is important to note that the pH and sodium chloride concentration used in the chromatography experiments were 6.5 and 0.2 M respectively. The pH is required for the stability of the silica column material. The presence of salt in the buffer is required to prevent non-specific interaction between the protein and the column matrix. Elution profiles were fitted to sums of Gaussian functions using ORIGIN (Originlab Corp., Northampton, MA, USA).

### 3. Results

#### 3.1. Sequence comparison

The ArdA proteins from *E. faecalis* V583, *S. aureus* Mu50 and *B. fragilis* NCTC 9343 will be referred to as V583 ArdA, Mu50 ArdA and Bfr ArdA, respectively. The ArdA from Orf18 of Tn916 will be referred to as Orf18 ArdA. Three of the four proteins examined, Orf18 ArdA, V583 ArdA and Mu50 ArdA, show 74 to 81% sequence similarity and between 56 and 63% identity to each other (Table 1). The Bfr ArdA shows ~40% similarity and ~25% identity to the other three ArdA sequences and is therefore more distantly related. The variation in sequence is spread throughout the length of the polypeptides but is predominantly located in the second domain (amino acids 62 to 103 of Orf18 ArdA) where identity drops to 50% or even lower (down to 11%) (Supplementary Fig. S1).

#### 3.2. Protein purification and secondary structure content of Bfr, V583 and Orf18 ArdA proteins

The Bfr, V583 and Mu50 ArdA proteins could all be purified to homogeneity using the same procedure as used for purifying Orf18 ArdA (size exclusion chromatography profiles for each are described later, SDS-PAGE gels show a single band for each ArdA when stained with Coomassie Blue—data not shown). The crystal structure of Orf18 ArdA shows a mixture of alpha helix (44%), beta sheet (19%) and random coil regions (37%), Table 1. CD spectroscopy was used to assess the secondary structure content of the four ArdA proteins and was found to be surprisingly variable, Fig. 2 and Table 2, given the levels of sequence similarity shown in Table 1. The Orf18 ArdA CD analysis showed the same amount of alpha helix as observed in the crystal structure [9]. Bfr ArdA was almost identical to Orf18 ArdA. V583 ArdA showed more irregular structure than Bfr and Orf18 ArdA proteins with a particularly noticeable drop in alpha helical content. The Mu50 ArdA was the least structured of the proteins showing very low helical content and a large proportion of irregular structure. Domain 2 in the crystal structure is mostly alpha helical or loops; hence the drop in helical content observed with V583 and Mu50 ArdA may reflect the loss of structure of this domain in particular.

#### 3.3. Stability of ArdA proteins to denaturation with guanidinium hydrochloride

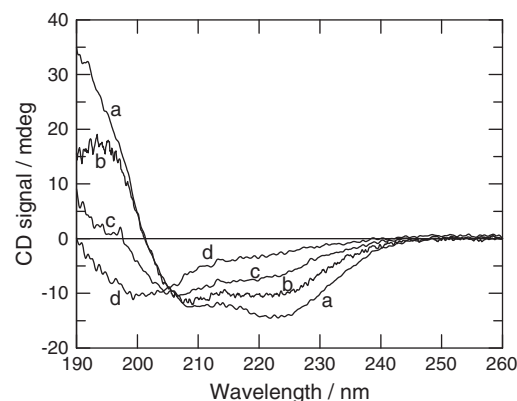
Fluorescence emission spectra showed a fluorescence maximum due to tryptophan emission at 340 nm for Orf18 ArdA and Bfr ArdA and at 350 nm for V583 ArdA and Mu50 ArdA when excited at 295 nm. These four ArdA proteins have only one conserved tryptophan residue in domain 1 (W23 in Orf18 ArdA and V583 ArdA, W22 in Mu50 ArdA and W28 in Bfr ArdA). This is the only tryptophan in V583 ArdA. An additional tryptophan is found in Orf18 ArdA (W70) and Mu50 ArdA (W91). Both of these tryptophans are in domain 2 but not at conserved positions. Bfr ArdA has four tryptophans in total (the conserved one in domain 1, two unconserved ones in domain 2, W73 and W99, and the fourth in domain 3, W167).

The addition of guanidinium chloride as a denaturant changed the fluorescence intensity and shape of the spectra as the proteins unfolded.

**Table 1**

Percentage similarity (above diagonal) and identity (below diagonal) for the four ArdA proteins.

	Orf18 ArdA	V583 ArdA	Mu50 ArdA	Bfr ArdA
Orf18	100	74	81	40
V583	56	100	78	43
Mu50	63	62	100	43
Bfr	24	30	26	100



**Fig. 2.** CD spectra of Bfr (a), Orf18 (b), V583 (c) and Mu50 (d) ArdA proteins. The protein concentration was 5.6 to 5.8  $\mu$ M for each sample.

As none of the tryptophan residues are located near to the ArdA dimer interface, the change in fluorescence emission observed is due to changes in secondary and tertiary structure only rather than in quaternary structure. Taking the ratio of fluorescence intensity at two emission wavelengths removed minor sample-to-sample variation in protein concentration and allowed calculation of the free energy of unfolding in the buffer in the absence of denaturant, Table 3 and Fig. 3. The folding curves determined by diluting out the denaturant from a sample with fully unfolded protein were identical within experimental error to the unfolding curves (data not shown). This indicates that the transitions are fully reversible. The data were analysed by a two state model. It can be seen that despite the similarity in amino acid sequence between the four proteins their stability varies considerably with Orf18 ArdA being the most stable and Mu50 ArdA the least. All three parameters determined from the unfolding curves; midpoint of the transition, slope of the transition and free energy, showed the same trend as one progressed from the least to the most stable ArdA. There was no evidence for any stable intermediate states in the transition, even with Bfr ArdA protein which contains tryptophan reporter groups in each of the three domains comprising the monomer, such as those observed with multi-domain fragments of Factor H where steps in the transition were obvious [31]. This indicates that the folding/unfolding process is cooperative despite the domain structure. The transition midpoint for Mu50 ArdA was not well defined as there was no plateau region prior to the unfolding transition. The Orf18 ArdA and the Mu50 ArdA proteins are 63% identical and 81% similar in sequence and both have two tryptophan residues, one conserved in domain 1 and the second being unconserved in domain 2, yet they show the most extreme differences in stability. The amino acid sequence between W70 in Orf18 ArdA and W91 in Mu50 ArdA (H90 in Orf18 ArdA) is the most poorly conserved in a pairwise alignment, therefore, if the tryptophan fluorescence from domain 2 in these two proteins is greater than that from domain 1, then the large difference in stability could primarily reflect differences in stability of domain 2 in these two proteins. The difference in stability correlates well with the observed variability in secondary structure content revealed by CD spectroscopy, Fig. 2, and the low levels of sequence identity between the proteins in domain 2 (Supplementary Fig. S1).

**Table 2**

Secondary structure content of the ArdA proteins determined by analysis of the CD spectra using CONTIN [26,27]. Secondary structure content of the crystallised Orf18 ArdA is from [9].

Protein	Helix %	Sheet %	Turns and irregular %
Orf18 ArdA crystal	44	19	37
Orf18 ArdA	39	12	49
V583 ArdA	17	29	54
Mu50 ArdA	9	33	58
Bfr ArdA	42	11	47

**Table 3**  
Analysis of unfolding as a function of guanidinium chloride concentration using a two-state model. The free energy of unfolding in the absence of denaturant is calculated by multiplying the slope of the transition by the midpoint of the transition. The temperature was 298 K.

Protein	Midpoint of transition/M	Slope of transition/RT	Free energy of unfolding in buffer/kJ mol <sup>-1</sup>
Orf18 ArdA <sup>a</sup>	2.88 +/- 0.06	5.24 +/- 1.37	37.2
V583 ArdA	1.22 +/- 0.16	2.92 +/- 0.46	8.8
Mu50 ArdA	0.30 +/- 0.75	2.31 +/- 0.85	1.7
Bfr ArdA	2.32 +/- 0.08	3.36 +/- 0.65	19.3

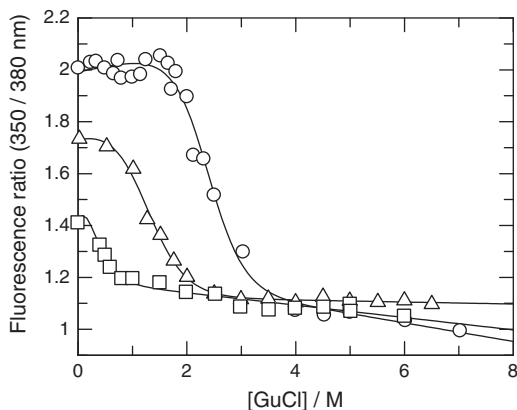
<sup>a</sup> Analysis from [16].

### 3.4. Size exclusion chromatography to analyse the stability of the Bfr, V583 and Mu50 ArdA proteins

Analytical SEC of Bfr, V583 and Mu50 ArdA was performed and compared to the behaviour of Orf18 ArdA, Fig. 4a. Each protein eluted essentially as a single peak at all concentrations investigated except Orf18 ArdA which showed a small amount of a lower molecular mass species (Fig. 4a) but the peaks broadened at lower concentrations due to the presence of a mixture of species differing in quaternary structure (data not shown).

The elution time and apparent molecular mass for each ArdA were determined and plotted against protein concentration injected onto the column, Fig. 4b. The apparent molecular masses at low and high protein concentrations for all three proteins were the same within experimental error and similar to that observed for Orf18 ArdA. Previously, Orf18 ArdA has been found to exist in an equilibrium mixture of monomeric and dimeric forms in solution with a dissociation constant estimated to be ~1 μM [16,20]. Both forms are active in antirestriction but the monomeric form has lost the ability to inhibit the MTase modification function of the EcoKI RM system [20]. The Bfr, V583 and Mu50 ArdA proteins appear to behave in the same way as Orf18 ArdA as their molecular mass changes with protein concentration.

An apparent dissociation constant,  $K_d$ , for the dimer was calculated to be 0.36 +/- 0.8, 3.18 +/- 0.14, 2.20 +/- 0.35 and 0.98 +/- 0.13 μM for Bfr, Mu50, Orf18 and V583 ArdA respectively but this value is approximate as firstly, the protein concentration on the column is diluted with respect to the injected concentration and secondly, the dissociation on the column is a non-equilibrium process. The most stable dimer was Bfr ArdA and the least stable were Orf18 ArdA and Mu50 ArdA. The change in apparent  $K_d$  was less than 10-fold so the difference in free energy between the different proteins is only a few kJ mol<sup>-1</sup>. The apparent molecular mass of each species determined by SEC, monomer at low concentration and dimer at high concentration, is very different from the masses calculated from the amino acid sequences. This is due to the highly

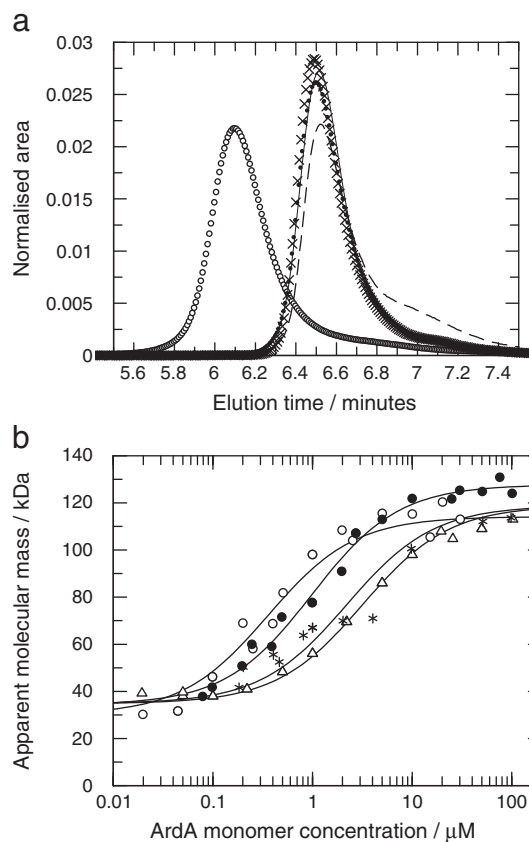


**Fig. 3.** Stability of ArdA proteins in solutions of guanidinium chloride (GuCl). Denaturation curves were determined using the ratio of fluorescence emission intensity at 350 nm to that at 380 nm when excited at 295 nm for Bfr ArdA (circles) V583 ArdA (triangles) and Mu50 ArdA (squares) as a function of GuCl concentration. The fitted lines are from a two-state model of folding [29].

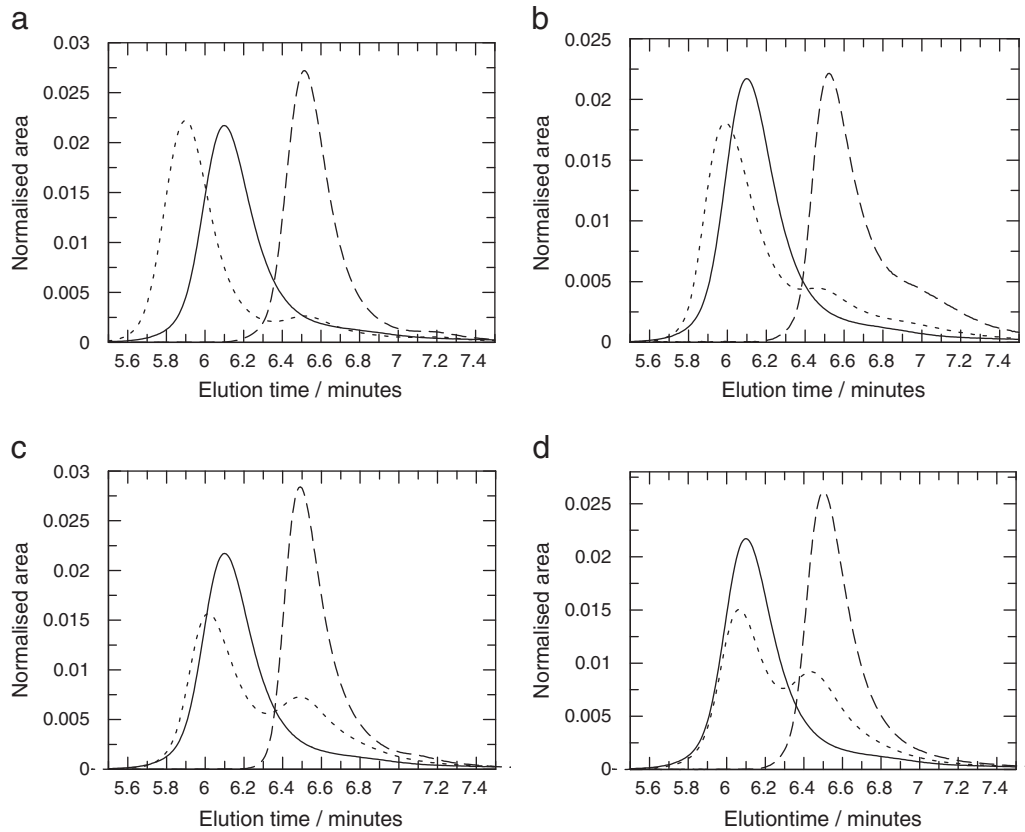
elongated shapes of the monomer and dimer giving larger than expected hydrodynamic radii [16,20,32].

### 3.5. Interaction of ArdA proteins with EcoKI MTase examined via size exclusion chromatography

Mixtures of one ArdA dimer (2 μM monomer) with 2 μM MTase were made and subjected to SEC, Fig. 5. All of the mixtures showed the formation of a complex eluting ahead of the MTase elution peak time suggesting binding between the ArdA proteins and the MTase. The mixtures also showed a second peak eluting near the elution time of the free ArdA and this must correspond to unbound ArdA. Fitting two Gaussian curves to these elution profiles allowed a rough estimate of how much ArdA had been bound in each mixture (data not shown). The amount of ArdA bound to the MTase decreased in the order Bfr



**Fig. 4.** Size exclusion chromatography of ArdA proteins and EcoKI MTase. a. Elution profiles of Bfr ArdA (solid line), Mu50 ArdA (crosses), Orf18 ArdA (dashed line), V583 ArdA (dots) and MTase (open circles) normalised to the same overall area. All proteins were injected onto the column at 2 μM concentration (ArdA monomer concentration or M<sub>2</sub>S<sub>1</sub> MTase concentration). b. The apparent molecular mass of Bfr ArdA (open circles), Mu50 ArdA (open triangles), Orf18 ArdA (crosses), and V583 ArdA (solid circles) plotted against the monomer concentration. The injected concentrations refer to that of the sample when loaded and ignore dilution by diffusion and mixing during the chromatography. Fitted lines are assuming a monomer–dimer equilibrium and give apparent dissociation constants of 0.36 +/- 0.8, 3.18 +/- 0.14, 2.20 +/- 0.35 and 0.98 +/- 0.13 μM respectively.



**Fig. 5.** Size-exclusion chromatography analysis of binding between MTase and ArdA to investigate the solution molecular weight. 2  $\mu\text{M}$  concentrations (ArdA monomer or  $\text{M}_2\text{S}_1$  MTase) were injected. The concentrations refer to that of the sample when loaded and ignore dilution by diffusion and mixing during the chromatography. Elution profiles were normalised to the same area. Dotted lines are the mixture of MTase and ArdA; solid lines are MTase; dashed lines are ArdA. a. Bfr ArdA. b. Orf18 ArdA. c. Mu50 ArdA. d. V583 ArdA.

ArdA > Orf18 ArdA > Mu50 ArdA > V583 ArdA. The elution time of the major peak in the complex of ArdA and MTase could also be used as an indicator of the strength of the ArdA:MTase interaction. This showed the same trend as the Gaussian fitting analysis with the most stable complex eluting at the earliest time. The interaction between ArdA proteins and the MTase is much weaker than the previously characterised interaction between Ocr and the MTase as in such mixtures, no free Ocr was observed; it was all bound by the MTase [30].

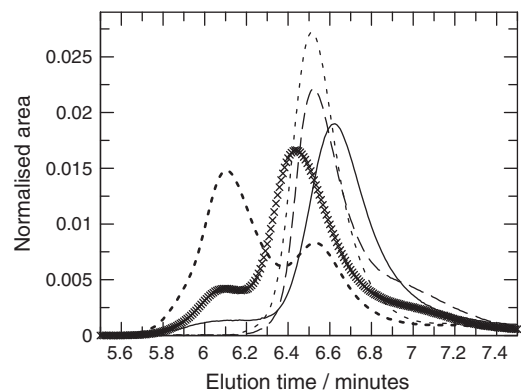
### 3.6. Interaction of Bfr and Orf18 ArdA proteins with $\text{M}_1\text{S}_1$ examined via size exclusion chromatography

The MTase has previously been shown to dissociate at low concentrations as it passes through a size exclusion column [21,22,30]. It dissociates into an  $\text{M}_1\text{S}_1$  form and an M subunit with a dissociation constant of between 15 and 90 nM. The behaviour of mixtures of the  $\text{M}_1\text{S}_1$  form and two of the ArdA proteins on SEC was investigated. Mixtures of 2  $\mu\text{M}$   $\text{M}_1\text{S}_1$  with 2  $\mu\text{M}$  monomer of Bfr ArdA or Orf18 ArdA were made and subjected to SEC, Fig. 6. The Bfr ArdA bound well to  $\text{M}_1\text{S}_1$  (main peak at 6.1 min with a smaller peak at 6.5 min due to uncomplexed material) whereas the Orf18 ArdA bound less well with the bulk of the material eluting at 6.4 min ahead of the individual ArdA and  $\text{M}_1\text{S}_1$  peaks. A small amount of material eluted at 6.1 min. The elution of material ahead of the individual ArdA and  $\text{M}_1\text{S}_1$  peaks indicated that  $\text{M}_1\text{S}_1$  can bind ArdA but the affinity for  $\text{M}_1\text{S}_1$  was higher for Bfr ArdA than for Orf18 ArdA. We assume that Mu50 and V583 ArdA can also interact in a similar manner with  $\text{M}_1\text{S}_1$ .

## 4. Discussion

The size exclusion data clearly show that the ArdA proteins exist in two forms in agreement with previous data on Orf18 ArdA [16,20].

Ultracentrifugation has previously defined these as monomer and dimer forms of ArdA [20]. Furthermore, the size exclusion data show that ArdA proteins can interact with the core MTase of EcoKI and the  $\text{M}_1\text{S}_1$  partially assembled core. Preliminary data obtained using both isothermal titration calorimetry and quartz crystal microbalance support the existence of these interactions under a range of pH and ionic strength conditions (G.A. Roberts and M. Reuter, unpublished data). All of the data lead to the conclusion that a complex set of multiple equilibria between the different quaternary structure forms of EcoKI and of ArdA will exist. The determination of the dissociation constants between all of these different species will be impossible with the current



**Fig. 6.** Interaction of ArdA proteins (2  $\mu\text{M}$  monomer injected) with the  $\text{M}_1\text{S}_1$  partially assembled form of the EcoKI MTase (2  $\mu\text{M}$   $\text{M}_1\text{S}_1$  injected) assessed by size exclusion chromatography. The areas under each curve have been set equal. The concentrations refer to that of the sample when loaded and ignore dilution by diffusion and mixing during the chromatography.  $\text{M}_1\text{S}_1$  (solid line), Bfr ArdA (dotted line), Orf18 ArdA (dashed line),  $\text{M}_1\text{S}_1$  mixed with Bfr ArdA (bold dotted line) and  $\text{M}_1\text{S}_1$  mixed with Orf18 ArdA (crosses).

methods and would require a technique such as mass spectrometry to quantify every species present in a mixture [33]. However, from the presented data, we conclude that the strength of the binding between ArdA and the MTase decreases in the order Bfr ArdA > Orf18 ArdA > V583 ArdA ~ Mu50 ArdA. This correlates with the stability and secondary structure content of the proteins (Bfr ArdA ~ Orf18 ArdA > V583 ArdA > Mu50 ArdA).

The affinity of ArdA for EcoKI MTase is poorer than that of the DNA target sequence for the MTase where affinities of the order of 1 to 10 nM were observed [34,35]. Nor do the ArdA proteins bind as tightly to the EcoKI MTase as the Ocr antirestriction protein [30]. EcoKI MTase binds very tightly ( $K_d$  of ~50 pM) to Ocr and the interaction is only weakened when Ocr is extensively neutralised by chemical modification [15] or large-scale mutagenesis [14]. Single or double mutations of Ocr had no effect on the interaction between Ocr and EcoKI [13]. Assuming that no additional protein component of ArdA is missing in our protein preparations, an assumption which is reasonable given that the *ardA* genes are not part of any operon and are transcribed from a single promoter [7,19], then the ArdA proteins are not such good mimics of the DNA bound by the MTase as the Ocr protein. This is in agreement with data on ArdA from the conjugative plasmid Collb-P9 [17,18]. Our conclusion is perhaps unexpected as ArdA has approximately the same number of negative charges per unit length as Ocr. Orf18 ArdA, for example, has 14 aspartate residues and 24 glutamate residues but only 3 arginine and 3 lysine residues per monomer and each monomer is ~7 nm in length. Ocr has 17 aspartate and 17 glutamate residues but only 4 arginine and 2 lysine residues per monomer and each monomer is ~4 nm in length so one might expect ArdA to be as effective as Ocr if only electrostatic interactions were involved in the DNA mimicry. Comparison of the shape of Ocr [8] and ArdA [9] implies that Ocr is a better fit into the MTase than the ArdA proteins. Thus shape complementarity in addition to electrostatic attraction is required for very tight binding as previously concluded from a study of chemical modification of Ocr [15].

The observation that the secondary structure and stability of the ArdA proteins varied greatly was unexpected given the considerable levels of sequence similarity particularly between the Mu50, V583 and Orf18 ArdA proteins. The level of similarity is lowest in the predominantly alpha helical Domain 2 which would correlate with the low level of alpha helical structure observed in the least stable Mu50 ArdA. These data may indicate a varying amount of flexibility in the different proteins with Mu50 ArdA being the most flexible DNA mimic. Given that all of the proteins bind similarly to the EcoKI MTase, we speculate that the binding event induces unstructured or weakly structured parts of ArdA to fold. These differences in structure and dynamics could be investigated further by, for example, ion-mobility mass spectrometry [36].

The amino acid sequences of the subunits of Type I RM enzymes usually differ considerably even when isolated from the same bacterial species. The levels of identity are around ~30% at best and confined to conserved motifs required for substrate and cofactor binding and catalysis [5,37,38]. The only exceptions to this are when the enzymes belong to the same “family” and exhibit subunit complementation, DNA hybridisation and antibody cross reactivity. In a family, sequence identity is nearly 100% for the M subunits and the R subunits. The Type I RM systems encoded by the bacterial species used as the sources of the ArdA proteins examined here, namely, *E. faecalis*, *B. fragilis* and *S. aureus*, do not belong to the same family when assessed by sequence conservation (data not shown, sequences of Type I RM systems obtained from REBASE [25]). Given that the ArdA proteins examined here are from unrelated species but all work against the EcoKI system of *E. coli* K12 [10,11], then it would seem to be a reasonable prediction that they will also work against the Type I RM systems from *E. faecalis*, *B. fragilis* and *S. aureus* and, indeed, against any Type I RM system. Thus the source of the ArdA does not seem to be important for their function as anti-RM proteins. This retention of function could aid

horizontal gene transfer of mobile genetic elements containing *ardA* genes across species boundaries [39].

## Acknowledgements

We gratefully acknowledge financial support from the Wellcome Trust (GR080463MA and 090288/Z/09/ZA) and the Biotechnology and Biological Sciences Research Council (BB/D001870/1 and BB/C511599/1). Bansi Sanghvi was supported by a Ph.D. studentship from the Darwin Trust, Kai Chen by an EaStChem Ph.D. studentship and Marcel Reuter by the Marie-Curie Network “From FLIM to FLIN” (MRTN-CT-2005-019481). We thank Sarah Dunn and Kasia Katzarwylis for assistance with early experiments.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.bbapap.2013.12.008>.

## References

- [1] D.T.F. Dryden, DNA mimicry by proteins and the control of enzymatic activity on DNA, *Trends Biotechnol.* 24 (2006) 378–382.
- [2] D.T.F. Dryden, M.R. Tock, DNA mimicry by proteins, *Biochem. Soc. Trans.* 34 (2006) 317–319.
- [3] D.C. Putnam, A.J. Tainer, Protein mimicry of DNA and pathway regulation, *DNA Repair (Amst)* 4 (2005) 1410–1420.
- [4] H.C. Wang, H.C. Wang, T.P. Ko, Y.M. Lee, J.H. Leu, C.H. Ho, W.P. Huang, C.F. Lo, A.H. Wang, White spot syndrome virus protein ICP11: a histone-binding DNA mimic that disrupts nucleosome assembly, *Proc. Natl. Acad. Sci. U. S. A.* 105 (2008) 20758–20763.
- [5] M.R. Tock, D.T.F. Dryden, The biology of restriction and anti-restriction, *Curr. Opin. Microbiol.* 8 (2005) 466–472.
- [6] F.W. Studier, Gene 0.3 of bacteriophage T7 acts to overcome the DNA restriction system of the host, *J. Mol. Biol.* 94 (1975) 283–295.
- [7] M.T. Nasim, I.C. Eperon, B.M. Wilkins, W.J. Brammar, The activity of a single-stranded promoter of plasmid Collb-P9 depends on its secondary structure, *Mol. Microbiol.* 53 (2004) 405–417.
- [8] M.D. Walkinshaw, P. Taylor, S.S. Sturrock, C. Atanasiu, T. Berge, R.M. Henderson, J.M. Edvardson, D.T.F. Dryden, Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA, *Mol. Cell* 9 (2002) 187–194.
- [9] S.A. McMahon, G.A. Roberts, K.A. Johnson, L.P. Cooper, H. Liu, J.H. White, L.G. Carter, B. Sanghvi, M. Oke, M.D. Walkinshaw, G.W. Blakely, J.H. Naismith, D.T.F. Dryden, Extensive DNA mimicry by the ArdA anti-restriction protein and its role in the spread of antibiotic resistance, *Nucleic Acids Res.* 37 (2009) 4887–4897.
- [10] C.K. Kennaway, A. Obarska-Kosinska, J.H. White, I. Tuszynska, L.P. Cooper, J.M. Bujnicki, J. Trinick, D.T.F. Dryden, The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein, *Nucleic Acids Res.* 37 (2009) 762–770.
- [11] C.K. Kennaway, J.E. Taylor, C.F. Song, W. Potrzebowski, W. Nicholson, J.H. White, A. Swiderska, A. Obarska-Kosinska, P. Callow, L.P. Cooper, G.A. Roberts, J.B. Artero, J.M. Bujnicki, J. Trinick, G.G. Kneale, D.T.F. Dryden, Structure and operation of the DNA-translocating Type I DNA restriction enzymes, *Genes Dev.* 26 (2012) 92–104.
- [12] D. Scholl, J. Kieleczawa, P. Kemp, J. Rush, C.C. Richardson, C. Merrill, S. Adhya, I.J. Molineux, Genomic analysis of bacteriophages SP6 and K1-5, an estranged subgroup of the T7 supergroup, *J. Mol. Biol.* 335 (2004) 1151–1171.
- [13] A.S. Stephanou, G.A. Roberts, M.R. Tock, E.H. Pritchard, R. Turkington, M. Nutley, A. Cooper, D.T.F. Dryden, A mutational analysis of DNA mimicry by Ocr, the gene 0.3 antirestriction protein of bacteriophage T7, *Biochem. Biophys. Res. Commun.* 378 (2009) 129–132.
- [14] G.A. Roberts, A.S. Stephanou, N. Kanwar, A. Dawson, L.P. Cooper, K. Chen, M. Nutley, A. Cooper, G.W. Blakely, D.T.F. Dryden, Exploring the DNA mimicry of the Ocr protein of phage T7, *Nucleic Acids Res.* 40 (2012) 8129–8143.
- [15] A.S. Stephanou, G.A. Roberts, L.P. Cooper, D.J. Clarke, A.R. Thomson, C.L. MacKay, M. Nutley, A. Cooper, D.T.F. Dryden, Dissection of the DNA mimicry of the bacteriophage T7 Ocr protein using chemical modification, *J. Mol. Biol.* 391 (2009) 565–576.
- [16] D. Serfiotis-Mitsa, G.A. Roberts, L.P. Cooper, J.H. White, M. Nutley, A. Cooper, G.W. Blakely, D.T.F. Dryden, The Orf18 gene product from conjugative transposon Tn916 is an ArdA antirestriction protein that inhibits Type I DNA restriction-modification systems, *J. Mol. Biol.* 383 (2008) 970–981.
- [17] S.V. Nekrasov, O.V. Agafonova, N.G. Belogurova, E.P. Delver, A.A. Belogurov, Plasmid-encoded antirestriction protein ArdA can discriminate between Type I methyltransferase and complete restriction-modification system, *J. Mol. Biol.* 365 (2007) 284–297.
- [18] G.B. Zavilgelsky, V.Y. Kotova, S.M. Rastorguev, Comparative analysis of anti-restriction activities of ArdA (Collb-P9) and Ocr (T7) proteins, *Biochemistry (Moscow)* 73 (2008) 906–911.
- [19] B.M. Wilkins, Plasmid promiscuity: meeting the challenge of DNA immigration control, *Environ. Microbiol.* 4 (2002) 495–500.

- [20] G.A. Roberts, K. Chen, E.K.M. Bower, J. Madrzak, A. Woods, A.M. Barker, L.P. Cooper, J.H. White, G.W. Blakely, I. Manfield, D.T.F. Dryden, Mutations of the domain forming the dimeric interface of the ArdA protein affect dimerisation and antimodification activity but not antirestriction activity, *FEBS J.* 280 (2013) 4903–4914.
- [21] D.T.F. Dryden, L.P. Cooper, N.E. Murray, Purification and characterization of the methyltransferase from the Type I restriction and modification system of *Escherichia coli* K12, *J. Biol. Chem.* 268 (1993) 13228–13236.
- [22] D.T.F. Dryden, L.P. Cooper, P.H. Thorpe, O. Byron, The in vitro assembly of the EcoKI Type I DNA restriction/modification enzyme and its in vivo implications, *Biochemistry* 36 (1997) 1065–1076.
- [23] S.C. Gill, P.H. von Hippel, Calculation of protein extinction coefficients from amino acid sequence data, *Anal. Biochem.* 182 (1989) 319–326.
- [24] N.V. Grishin, J. Pei, PROMALS: towards accurate multiple sequence alignments of distantly related proteins, *Bioinformatics* 23 (2007) 802–808.
- [25] R.J. Roberts, T. Vincze, J. Posfai, D. Macelis, REBASE—a database for DNA restriction and modification: enzymes, genes and genomes, *Nucleic Acids Res.* 38 (2010) D234–D236.
- [26] S.W. Provencher, J. Glockner, Estimation of globular protein secondary structure from circular dichroism, *Biochemistry* 20 (1981) 33–37.
- [27] L. Whitmore, B.A. Wallace, DICHROWEB: an online server for protein secondary structure analyses from circular dichroism spectroscopic data, *Nucleic Acids Res.* 32 (2004) W668–W673.
- [28] C.N. Pace, B.A. Shirley, J.A. Thomson, Measuring the conformational stability of a protein, in: T.E. Creighton (Ed.), *Protein Structure. A Practical Approach*, IRL Press, Oxford, 1989, pp. 311–330.
- [29] J. Clarke, A.R. Fersht, Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation, *Biochemistry* 32 (1993) 4322–4329.
- [30] C. Atanasiu, T.-J. Su, S.S. Sturrock, D.T.F. Dryden, Interaction of the Ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme, *Nucleic Acids Res.* 30 (2002) 3936–3944.
- [31] M.D. Kirkitadze, M. Krych, D. Uhrin, D.T.F. Dryden, B.O. Smith, A. Cooper, X. Wang, R. Hauhart, J.P. Atkinson, P.N. Barlow, Independently melting modules and highly structured intermodular junctions within complement receptor type 1, *Biochemistry* 38 (1999) 7019–7031.
- [32] R.H. Haschemeyer, A.E.V. Haschemeyer, *Proteins: A Guide to Study by Physical and Chemical Methods*, Wiley-Interscience, New York, 1973.
- [33] A. Dyachenko, R. Gruber, L. Shimon, A. Horovitz, M. Sharon, Allosteric mechanisms can be distinguished using structural mass spectrometry, *Proc. Natl. Acad. Sci. U. S. A.* 110 (2013) 7235–7239.
- [34] T.-J. Su, M.R. Tock, S.U. Egelhaaf, W.C.K. Poon, D.T.F. Dryden, DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping, *Nucleic Acids Res.* 33 (2005) 3235–3244.
- [35] L.M. Powell, B.A. Connolly, D.T.F. Dryden, The DNA binding characteristics of the trimeric EcoKI methyltransferase and its partially assembled dimeric form determined by fluorescence polarisation and DNA footprinting, *J. Mol. Biol.* 283 (1998) 947–961.
- [36] E. Jurneczko, F. Cruickshank, M. Porrini, P. Nikolova, I.D. Campuzano, M. Morris, P.E. Barran, Intrinsic disorder in proteins: a challenge for (un)structural biology met by ion mobility-mass spectrometry, *Biochem. Soc. Trans.* 40 (2012) 1021–1026.
- [37] N.E. Murray, Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle), *Microbiol. Mol. Biol. Rev.* 64 (2000) 412–434.
- [38] N.E. Murray, 2001 Fred Griffith review lecture. Immigration control of DNA in bacteria: self versus non-self, *Microbiology* 148 (2002) 3–20.
- [39] C.M. Thomas, K.M. Nielsen, Mechanisms of, and barriers to, horizontal gene transfer between bacteria, *Nat. Rev. Microbiol.* 3 (2005) 711–721.