

# Organization of the Bcgl restriction-modification protein for the cleavage of eight phosphodiester bonds in DNA

Rachel M. Smith<sup>1</sup>, Jacqueline J. T. Marshall<sup>1</sup>, Alistair J. Jacklin<sup>1</sup>, Susan E. Retter<sup>1</sup>, Stephen E. Halford<sup>1,\*</sup> and Frank Sobott<sup>2</sup>

<sup>1</sup>The DNA-proteins Interaction Unit, School of Biochemistry, University of Bristol, University Walk, Bristol BS8 1TD and <sup>2</sup>Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

Received September 11, 2012; Revised October 2, 2012; Accepted October 3, 2012

## ABSTRACT

**Type IIB restriction-modification systems, such as Bcgl, feature a single protein with both endonuclease and methyltransferase activities. Type IIB nucleases require two recognition sites and cut both strands on both sides of their unmodified sites. Bcgl cuts all eight target phosphodiester bonds before dissociation. The Bcgl protein contains A and B polypeptides in a 2:1 ratio: A has one catalytic centre for each activity; B recognizes the DNA. We show here that Bcgl is organized as A<sub>2</sub>B protomers, with B at its centre, but that these protomers self-associate to assemblies containing several A<sub>2</sub>B units. Moreover, like the well known FokI nuclease, Bcgl bound to its site has to recruit additional protomers before it can cut DNA. DNA-bound Bcgl can alternatively be activated by excess A subunits, much like the activation of FokI by its catalytic domain. Eight A subunits, each with one centre for nuclease activity, are presumably needed to cut the eight bonds cleaved by Bcgl. Its nuclease reaction may thus involve two A<sub>2</sub>B units, each bound to a recognition site, with two more A<sub>2</sub>B units bridging the complexes by protein–protein interactions between the nuclease domains.**

## INTRODUCTION

Restriction-Modification (RM) systems use two enzyme activities to defend bacteria against foreign DNA, a restriction endonuclease (REase) and a modification methyltransferase (MTase) (1,2). The MTase transfers a

methyl group from S-adenosylmethionine (SAM) to a cytosine or an adenine within a particular DNA sequence, the recognition site for that system, while the REase cleaves DNA with unmethylated sites (3). The majority of RM systems fall into either the Type I or Type II categories (4). Most Type I systems are oligomeric (R<sub>2</sub>M<sub>2</sub>S) proteins containing subunits for DNA cleavage (R), methylation (M) and sequence specificity (S) (2,5,6), though in some cases the R, M and S functions are all carried as domains within a single polypeptide (7). The cleavage activity is triggered by unmodified sites, though it often occurs distant from the site, due to the R subunits possessing an ATP-dependent DNA translocase. The main reaction of the MTase is to convert hemi-methylated sites, methylated in one strand only, to fully modified products methylated in both strands. Neither the fully nor the hemi-methylated DNA is cut by the REase. Consequently, DNA fully modified once by the MTase is never cleaved by the REase even after its semi-conservative replication.

Most Type II RM systems use separate proteins for restriction and modification that both bind to the same DNA sequence (1,8,9). The REase is commonly a homodimer that recognizes a palindromic sequence and cuts both strands at specified loci within the site: one subunit attacks the target phosphodiester bond in one strand, while the other cuts the scissile bond in the opposite strand (9,10). The MTase is usually a monomer whose main role is to transfer one methyl group onto DNA already methylated in one strand (8). But many Type II systems deviate considerably from these classical systems. They have been categorized into numerous subtypes—IIG, IIE, IIF, IIS and IIB, among others—on the basis of their genetic organization, their mode of action and their recognition sites and cleavage loci (4).

\*To whom correspondence should be addressed. Tel: +44 1173 312156; Fax: +44 1173 312168; Email: s.halford@bristol.ac.uk

Present addresses:

Jacqueline Marshall, CRUK, London Research Institute, Clare Hall Laboratories, Blanche Lane, South Mimms, Potters Bar EN6 3LD, UK.  
Frank Sobott, Department of Chemistry and Center for Proteomics, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerpen, Belgium.

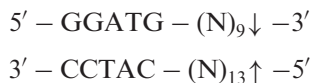
The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Perhaps the only aspect common to all Type II REases, and the one that differentiates them from Type I systems, is that they cleave DNA at fixed positions relative to their recognition sites.

Variations in genetic organization include the Type IIG systems that, instead of separate REase and MTase proteins, possess a single polypeptide with both activities (11–14). The IIG systems require SAM not only as a co-factor for their methylation reactions but also to activate the nuclease.

Variations in mode of action are illustrated by the Type IIE and IIF systems (15,16). While the orthodox enzymes act at individual sites, the IIE and IIF REases need two copies of their target sequence for full activity. The IIE enzymes, often dimers, proceed to cut both DNA strands at one of the two sites; the other site acts as an allosteric activator (17,18). In contrast, the Type IIF enzymes tend to cut both strands at both sites before leaving the DNA (18,19): most operate as tetramers of identical subunits, with the four active sites, one in each subunit, each positioned to cleave one of the four target bonds (20–22). Enzymes that require two DNA sites can be identified by comparing their activities on DNA substrates with one or two cognate sites (19,22–24). Proteins that interact with two sites prefer sites *in cis*, where they loop out the intervening DNA, over sites *in trans*, where they must hold together two separate DNA molecules (16). Consequently, both Type IIE and IIF REases can cleave two-site substrates more rapidly than a one-site DNA. However, IIE enzymes cut two-site substrates one site at a time, whereas IIF enzymes cut both sites together (18).

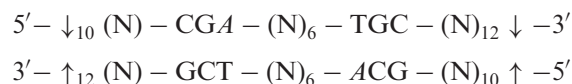
Variations in recognition and cleavage loci are shown by the Type IIS REases, which bind asymmetric rather than palindromic sites and which cut the DNA at fixed loci away from the sequences, on one side of it (25,26). For example, FokI recognizes



and cuts both strands at the positions indicated ( $\downarrow, \uparrow$ ), 9 and 13 nt away (3). An asymmetric sequence cannot be recognized in the same way as a dimeric enzyme recognizing a palindromic site, where each subunit contacts one half of the site (9,10). Instead, either a monomeric protein contacts the entire site, as is the case with FokI (27), or different subunits in an oligomeric protein each contact discrete parts of the sequence, as with BbvCI (28). If the nuclease is a monomer, it may have only one active site (26), yet it needs to cut both 5'–3' and 3'–5' strands. The use of the same active site to cut both strands is often—but not always—prohibited, as it requires the site to rotate not only around the DNA, from one strand to the other, but also perpendicular to the DNA, to match their antiparallel nature (29). To cut both strands, the monomer of FokI bound to the DNA associates with a second monomer (30–32) to give a dimeric unit with two active sites (33). The second active site can alternatively come from the isolated catalytic domain of the FokI protein, without its DNA recognition domain (30). The two catalytic centres each cut one

specified strand: the DNA-bound monomer, the bottom strand 13 nt away and the recruited unit, the top strand 9 nt distant (34). However, the enzyme at the site associates more readily with another monomer bound elsewhere in the same DNA than with one in free solution, as proteins *in cis* cannot escape from each other (31,35). FokI thus attacks two-site substrates more rapidly than one-site DNA, although the two-site substrate is initially cut at just one site: the residual site is then cleaved at the same slow rate as the one-site DNA (23,31). The need for two recognition sites is common among the Type IIS REases: some, like FokI, act as IIE enzymes; others, like the tetramer BspMI, as IIF enzymes (36).

Another group that cleave outside their recognition sites are the Type IIB systems (37). Unlike the IIS enzymes that cut the DNA on one side of the site, the nucleases from the Type IIB systems make two double-strand breaks (DSBs) at each site, one either side, to excise a short fragment that carries the recognition sequence. Moreover, their recognition sequences usually consist of two specified but divergent blocks of sequence separated by a non-specific segment of fixed length, like many Type I sites (2,5). For example, BcgI, an archetype of the IIB enzymes (12,38–40), recognizes the sequence,



Its REase cuts both strands on both sides of this sequence, 10 and 12 nt away, to leave the site on a fragment 34 nt long in both strands (noted below as the 34-mer): its MTase targets the adenines marked above in italics. However, the BcgI nuclease has no activity when bound to a solitary site (41). Instead, like most Type IIB REases (24), it needs two sites (40), ideally *in cis* but alternatively *in trans*. After bridging the sites, it then proceeds to make two DSBs at both sites before releasing the final products, thus cutting a total of eight phosphodiester bonds within a single complex: hardly any of the partially cleaved intermediates accumulate during the course of the reaction (41).

BcgI, like many but not all Type IIB systems (37), is composed of two subunits: an A subunit of 71.6 kDa that carries both MTase and REase catalytic functions; a B subunit of 39.2 kDa that recognizes the target sequence (12,39). Like the Type IIG systems that have both activities in the same subunit, BcgI requires SAM not only for its MTase but also for its REase function: the nuclease also needs  $\text{Mg}^{2+}$  (38). Densitometry of coomassie-stained SDS gels yielded a 2:1 ratio of A:B subunits (12), albeit with the assumption—not always valid (42)—that the two polypeptides stain to the same extent. However, size exclusion chromatography of native BcgI gave apparent MW values of 250–300 kDa (39,43), larger than the 182 kDa expected for a protein with one B and two A subunits. Nevertheless, the size exclusion  $\text{MW}_{\text{app}}$  can be related to an  $\text{A}_2\text{B}$  protomer if that unit has a large frictional ratio or if it self-associates (44).

The A subunit of BcgI resembles a fusion of R and M subunits from a Type I system without the ATP motor functions of R, while B is strikingly similar to an

S subunit: it likewise possesses separate domains for recognizing each section of its bipartite target site (39). An A<sub>2</sub>B assembly for BcgI is thus a direct analogue of the R<sub>2</sub>M<sub>2</sub>S organization of Type I systems. Other Type IIB systems contain a single subunit with all three functions (13,45), akin to the single-polypeptide Type I enzymes (7). But while Type I REases generally make one DSB, usually about halfway between two recognition sites, BcgI makes four DSBs per reaction (41). How the BcgI RM protein is organized to hydrolyse eight phosphodiester bonds in a complex with two recognition sites is addressed here. In the following article (46), we show that the BcgI MTase transfers to an individual site in the DNA just one methyl group at a time. Hence, this protein may be organized differently for its REase and MTase activities.

## MATERIALS AND METHODS

### DNA

The plasmid pARG3-BcgI (a gift from New England Biolabs) over-expresses the native forms of *bcgIA* and *bcgIB* to yield the wild-type (WT) BcgI protein. The plasmid was subjected to site-directed mutagenesis (QuikChange II: Agilent) to replace Glu53 in the A subunit with Ala, to give pARG3-BcgIE53A [the E53A mutant has no nuclease activity (39)]. To construct the plasmid pAJJ1, the section of pARG3-BcgI encompassing *bcgIA* was amplified by PCR and the resultant fragment cloned between the NcoI and XhoI sites of pET28a (Merck).

The supercoiled (SC) forms of the plasmids pUC19 and pDG5 were <sup>3</sup>H-labelled and purified as described previously (19,24). The following duplexes, all with HEX (hexachlorofluorescein) at the 5'-end of the top strand, were as before (41): HEX-42S, a 42 bp specific DNA with the recognition sequence for BcgI and both upstream and downstream cleavage loci; HEX-42NS, a 42 bp non-specific DNA that is identical to HEX-42S apart from single bp changes in each segment of the recognition sequence. HEX-24S and HEX-24NS are truncated versions containing the central 24 bp of HEX-42S and HEX-42NS respectively: HEX-24S thus has the recognition sequence but not the cleavage loci.

### Proteins

Transformants of *Escherichia coli* XL1-Blue MFR with either pARG3-BcgI or pARG3-BcgIE53A, or *E. coli* Rosetta 2 with pAJJ1, were grown in 20 L L-broth with the appropriate antibiotics at 37°C in a Applikon Biotechnology bioreactor to an OD<sub>600</sub> of 0.5–0.7. IPTG was added to 1 mM and the culture continued for 3–4 h before it was passed through a GEA Westphalia separator prior to cell harvesting by centrifugation. WT BcgI, the E53A mutant and the A protein were all purified by essentially the same procedure as described before for the WT protein (41), except that the initial extract—after cell disruption and the removal of the debris—was fractionated, first with polyethyleneimine and then with ammonium sulphate (47), prior to the column

chromatography steps. As neither BcgIE53A nor the A subunit alone possess nuclease activity, the purification of these proteins was monitored by SDS-PAGE: the final preparations of all three proteins contained no detectable impurities (48). The identities of the purified proteins were confirmed by N-terminal amino-acid sequencing (W. Mawby, personal communication). Approximately 50% of the A subunit carried methionine at its N-terminus, as specified by its gene sequence, and 50% valine, the second residue in the sequence: the latter presumably arises from post-synthetic processing. Protein concentrations were assessed from A<sub>280</sub> measurements using molar extinction coefficients calculated from amino acid compositions. Molarities of WT BcgI and E53A refer to the A<sub>2</sub>B assembly and that for the A protein to its monomeric state. Molecular masses encompass four possible states of the A protein (46): with and without the N-terminal methionine, and with or without bound SAM.

### Native mass spectrometry

Aliquots of BcgIE53A were dialysed overnight against 200 mM AmAc (ammonium acetate), 10% glycerol, pH 8.4, and flash-frozen. When required, the samples were thawed on ice and buffer-exchanged twice with 200 mM AmAc using Micro Bio-Spin P6 columns (Bio-Rad) before being diluted in AmAc to the requisite concentration. (In this buffer, E53A still bound DNA specifically and the WT enzyme cleaved DNA when MgCl<sub>2</sub> was added: data not shown.)

The protein sample was loaded into the mass spectrometer using nano-electrospray ionization (nano-ESI) in positive ion mode with an automated Advion Triversa inlet system (5 μm nozzle diameter, spray voltage 1.75 kV, gas pressure 0.25 psi). Mass to charge values were determined using a Waters Synapt 1 T-wave ion mobility quadrupole/time-of-flight tandem mass spectrometer with an 8 k quadrupole for medium-to-high m/z transmission. Experimental parameters were chosen to effect desolvation of high m/z sample ions while maintaining conditions that promote the transmission of intact protein complexes, with minimal disruption of native conformations (49): backing pressure, 5.4 mbar; source temperature, 25°C; sampling and extraction cone, 180 and 0 V, respectively; trap and transfer collision energies, 60 and 40 V; trap DC bias, 80 V. Data analysis was conducted using MassLynx 4.1 software (Waters) without background subtraction using smoothing by the Savitzky–Golay method (± 50 channels).

### Analytical ultracentrifugation

Sedimentation equilibrium experiments were carried out at 20°C in six-channel centrepieces for a Beckman XL-A ultracentrifuge. Sample chambers contained the protein—either WT BcgI, E53A or A protein—in 110 μl analytical ultracentrifugation (AUC) buffer (20 mM Tris–HCl, pH 8.4, 66 mM NaCl, 2 mM CaCl<sub>2</sub>, 20 μM SAM and 10% glycerol) and reference chambers 120 μl AUC buffer. During centrifugation, absorbancies were recorded along the centrifugal radius at 230, 280 or 292 nm, depending on



the protein concentration. An initial scan was made directly on reaching 3000 rpm and the horizontal trace used to determine the protein concentration in the cell. Further scans were made after  $\geq 16$  h at speeds between 6000 and 15 000 rpm: at least two scans,  $\geq 4$  h apart, were taken at each speed. A final scan was recorded after 6 h at 40 000 rpm, to yield from the meniscus a baseline absorbance ( $B$ ). Data sets were fitted in ORIGIN to Equation 1 for the distribution of a single ideal species after sedimentation to equilibrium:

$$A_r = A_{r_0} \cdot e^{\left[\frac{\omega^2}{2RT} M(1-\bar{v}\rho)(r^2 - r_0^2)\right]} + B \quad (1)$$

where  $A_r$  and  $A_{r_0}$  are the absorbance at radius  $r$  and at the reference  $r_0$ ,  $\omega$  the centrifugal speed,  $\bar{v}$  the partial specific volume and  $\rho$  the solvent density: when applied to a single species,  $M$  denotes the molecular mass of that species but when applied to a mixture of species of varied mass but uniform  $\bar{v}$ ,  $M$  denotes the mean weight-average MW ( $M_w$ ) of all of the species in the cell (44,50). Multiple species were considered to be present whenever the residuals between the best fit to Equation 1 varied systematically from the experimental data. Further AUC experiments were carried out as above, but monitored at 539 nm, on samples containing either the HEX-42S or the HEX-42NS duplex and excess BcgI.

## Reactions

DNA cleavage rates were generally measured by adding the requisite concentration of BcgI protein to 200  $\mu$ l  $^3$ H-labelled DNA (5 nM) in buffer R (AUC buffer supplemented with 1 mM DTT and 200  $\mu$ g/ml BSA, and with 10 mM  $MgCl_2$  in place of the  $CaCl_2$ ) at 37°C: the DNA was the SC form of either pUC19 or pDG5, which carry, respectively, one or two BcgI sites (24,41). Samples (10  $\mu$ l) were taken before and at varied times after adding the enzyme, and mixed immediately with an EDTA quench prior to electrophoresis through agarose to separate the following DNA species: the intact SC substrate; the nicked open circle (OC) form; the linear (LIN) species with at least one DSB at one site and, for the two-site substrate, the final products with at least one DSB at both sites (L1 and L2). The concentration of each form was evaluated by scintillation counting. The reactions at each BcgI concentration were repeated at least three times, although occasionally with different sampling times for each test. For repeats with the same time points, the data at each time were averaged and the averages fitted to a single exponential decay with offset. For reactions sampled at different times, the decrease in SC DNA in each reaction was fitted to an exponential and the apparent rate constants then averaged.

Reactions that were too fast to measure by manual sampling were carried out by mixing, in a Hi-Tech RQF-63 quench-flow apparatus (TgK Scientific), equal volumes (80  $\mu$ l) of  $^3$ H-labelled pDG5 (10 nM) and BcgI (at twice the required concentration), both in buffer R at 37°C. After the requisite delay time, the reaction was mixed with 80  $\mu$ l EDTA (500 mM). Samples (100  $\mu$ l) of the quenched reaction were then analysed by

electrophoresis through agarose as above, to yield the residual concentration of the SC DNA.

For BcgI reactions in the presence of additional A protein, plasmid cleavage and the release of the 34-mer were measured in parallel (41). The reactions, in 200  $\mu$ l buffer R\* (as Buffer R but 5% [w/v] Ficoll 400 instead of the glycerol) at 37°C, contained 5 nM  $^3$ H-labelled pUC19 and specified concentrations of both native BcgI and BcgIA proteins. Samples (20  $\mu$ l) were taken from the reactions at various times, mixed immediately with quench supplemented with 250  $\mu$ g/ml proteinase K and incubated at 37°C for  $\geq 30$  min. Half of each sample was analysed by electrophoresis through agarose, as above, to evaluate the concentrations of the DNA species that could be resolved on agarose (the SC, OC and LIN forms of the plasmid). The other half was loaded onto a native 20% acrylamide gel, alongside an equivalent amount (5 nM) of HhaI-digested pUC19. The acrylamide gel was stained with SYBR Safe (Invitrogen) for 20 min, washed and its fluorescence recorded in a Typhoon PhosphorImager (Molecular Dynamics) using its 488 nm laser and 520 BP 40 emission filter. Images were analysed using ImageQuant software. The concentration of the 34-mer was determined relative to the known concentration of the 28 and 33 bp fragments from the HhaI-digested pUC19 control.

## DNA binding

Tests to measure DNA binding by BcgI were carried out at 37°C in 10  $\mu$ l buffer R\*C (as Buffer R\* but with 2 mM  $CaCl_2$  instead of the  $MgCl_2$ ) containing a HEX-labelled oligoduplex (20 nM) and varied concentrations of native BcgI. After adding the protein, samples were incubated for 5 min and then loaded directly onto 6% acrylamide gels in TBC (as TBE but with 2 mM  $CaCl_2$  in place of the EDTA) running at 4 V/cm. After loading, the voltage was increased to 12.5 V/cm. Gels were imaged in their plates in the PhosphorImager using its 532 nm laser and 555 BP 20 emission filter. Images were analysed in ImageQuant and the concentration of free DNA in each lane assessed relative to that in the lane without BcgI. The background was taken at an equivalent position in a lane lacking DNA.

## RESULTS

Glu53 in the A subunit of the BcgI RM protein forms part of the PD...EXK motif common to many nuclease active sites (9) and its replacement with Ala abolishes the nuclease activity of BcgI (39). Hence, to enable studies of BcgI without its cleavage reaction, an over-producing strain for the E53A mutant of BcgIA was constructed (see 'Materials and Methods' section). Preparations of the E53A mutant from this strain yielded about twice the amount of protein as parallel preparations of the active WT enzyme (48). Over-expression of the WT nuclease may result in cell death due to cleavage of the chromosome at non-cognate sites, as seen with other restriction enzymes (51,52), although the BcgI RM protein carries the MTase that ought to protect the recognition sites on the cellular

DNA (12,39). On the other hand, cells over-producing the inactive mutant might remain viable and so generate more of the protein. Experiments requiring high concentrations of the BcgI protein were thus facilitated by using the E53A protein rather than WT BcgI.

### Mass spectrometry

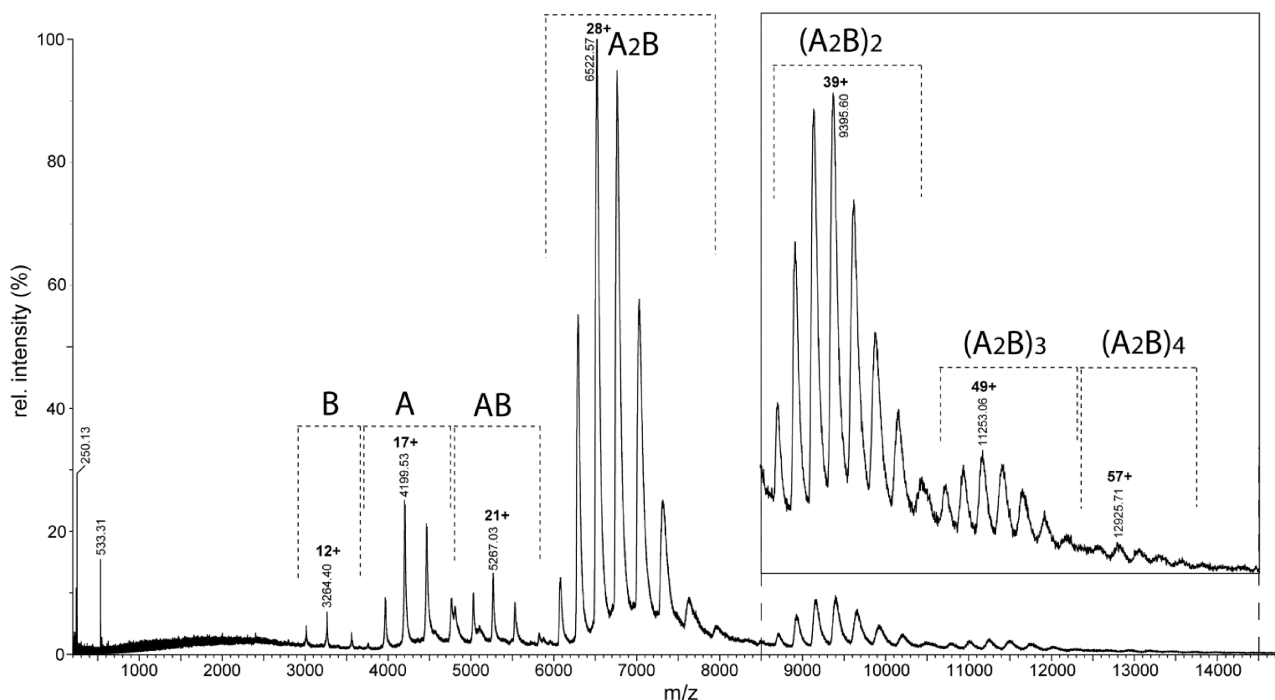
Native mass spectrometry (MS) was carried out primarily on the E53A mutant, although equivalent spectra were obtained with WT BcgI albeit with lower signal-to-noise ratios (data not shown). Under conditions where non-covalent interactions are preserved, the spectrum in Figure 1 shows a series of peaks, which are assigned to the A and B subunits and their complexes with different stoichiometries. The predominant species (~75% of the total intensity) is the  $A_2B$  complex with an  $m/z$  of around 7000, with principal charge states of 28+ and 29+. The mass calculated for the  $A_2B$  complex is ca. 182.4 kDa, as averaged over the four different species ( $\pm$ N-terminal methionine and  $\pm$ bound SAM; see above), which are not resolved here. The major peak in Figure 1 yields an experimental mass of 182434, which matches closely the  $A_2B$  form of BcgI. In addition, further multiples of  $A_2B$  assemblies were observed up to at least four such units, an  $(A_2B)_4$  complex at around  $m/z$  13000 (Figure 1, insert). The relative ratios of these higher-order complexes in the sample cannot be directly measured from the spectrum, as the sensitivity of the ion detector decreases with increasing  $m/z$  in a non-linear way. The time-of-flight system has in theory no upper limit on the size of molecule it can detect but the transmission and

the detection of ions depends on their mass. It is therefore difficult to determine the relative amounts of each species in the solution.

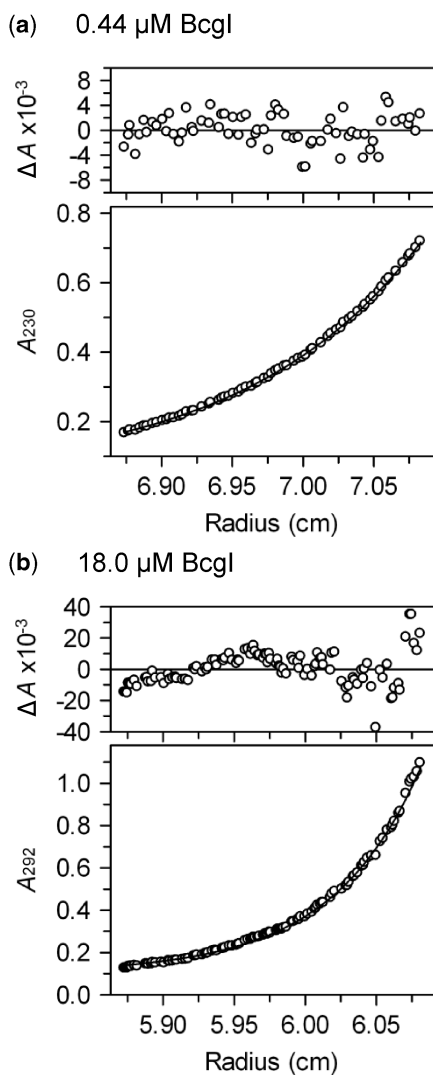
In addition to the presence of these larger species, some separated components of the complexes were also observed. The masses of these smaller species, up to  $m/z$  6000, correspond to the BcgIA and BcgIB subunits by themselves and to an AB assembly (Figure 1). These species are likely to be subunits that had dissociated from the BcgI endonuclease in solution, as suggested by the low charge states of these species (gas-phase dissociation products would have had disproportionately high charge states, similar to denatured proteins). In an  $A_2B$  protomer, the BcgIB subunit would normally be expected to have the weakest non-covalent linkage to the assembly as its mass, and therefore surface area, is less than that of the A subunit: this ought to make it easier for BcgIB to leave the complex than BcgIA. Hence, the presence of an AB species, but not an  $A_2$  dimer, suggests that the two A subunits in the active BcgI protein both interact with B but not with each other: the A subunits presumably flank the B subunit, one on either side. If this suggestion is correct, then the first step in the disassembly of the protein should be the dissociation of one BcgIA subunit from the  $A_2B$  unit to leave AB, as is seen in the spectra.

### AUC

Native MS identified several assemblies of the BcgI protein containing multiple copies of the  $A_2B$  protomer but it does not provide quantitative information about the amount of each species. Moreover, the native MS was

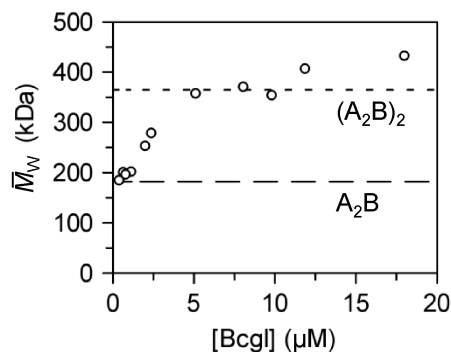


**Figure 1.** Native MS. A nano-ESI mass spectrum of the E53A form of BcgI (27.5  $\mu$ M) in 200 mM AmAc was recorded under non-denaturing conditions (see 'Materials and Methods' section). The insert shows the signals for the oligomeric states in the  $m/z$  range from 8500 to 14500 magnified 5-fold. The charge state series assigned to each molecular species (in bold) are indicated with dashed brackets, and the charge and  $m/z$  value for the primary peak in each series is given.



**Figure 2.** Sedimentation equilibrium. Samples of BcgI in AUC buffer were centrifuged at 7500 rpm at 20°C for the times noted below. The main panels show the absorbance (white circles) as a function of the centrifugal radius: (a) 0.44 μM BcgI at 230 nm after 30 h; (b) 18 μM BcgI at 292 nm after 44 h. The lines through the data points in the main panels are the best fits of the following data sets to Equation 1: (a) the records after 26 and 30 h; (b) after 32, 38 and 44 h. The best fits were obtained with values for  $M$  of 183 and 431 kDa in (a) and (b) respectively. The plots above the main panels display the residuals between these fits and the experimental data: the random variations in (a) show that the record can be accounted for by a single ideal species with a unique value of  $M$ ; the systematic deviations in (b) demonstrate that more than one species is present, which gives rise to a  $\bar{M}_w$  value.

done in a buffer designed to permit volatilization of the sample rather than that for optimal enzyme activity. To address these issues, AUC experiments were performed on WT BcgI under conditions similar to those used for its DNA cleavage assays, though in  $\text{Ca}^{2+}$  rather than  $\text{Mg}^{2+}$  so as to allow for DNA binding without cleavage (53). The inactive E53A mutant of BcgI behaved similarly to the WT in both the absence and presence of DNA and, in this case, with either  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$  at hand (data not shown).



**Figure 3.** Self-association of BcgI. Varied concentrations of BcgI in AUC buffer were loaded into the AUC and sedimented to equilibrium at 7500 rpm at 20°C. For each sample, the absorbance was measured as a function of centrifugal radius and fitted to Equation 1. Apart from the data at the lowest BcgI concentrations (*viz.* Figure 2a), the residuals between the fits and the experimental data displayed non-random deviations (*viz.* Figure 2b), so the best fit at each initial concentration denotes an  $\bar{M}_w$  value. The  $\bar{M}_w$  values are plotted against the loading concentrations of BcgI. The values expected for the 182.4 kDa  $\text{A}_2\text{B}$  protomer and for the 365 kDa  $(\text{A}_2\text{B})_2$  dimer are marked with long and short dashes, respectively.

Sedimentation equilibrium experiments were carried out with varied concentrations of WT BcgI. After centrifugation for  $\geq 20$  h to achieve equilibrium (as evidenced by overlaying traces from successive scans  $\geq 4$  h apart), the radial distribution of the protein was measured by absorbance at 280 nm for protein concentrations between 1 and 10 μM; or at 230 or 292 nm for lower and higher concentrations, respectively (Figures 2a and 2b). At each concentration, the distribution was fitted to Equation 1, for the sedimentation of a single ideal species: the fits encompassed several data sets from different times. At the lowest concentration of BcgI tested (0.4 μM), the residuals between the experimental data and the best fit to Equation 1 were both small and random, showing that the data can indeed be accounted for by a single species (Figure 2a). The AUC data at low BcgI concentrations yielded a MW of 183 kDa, which corresponds closely to the 182.4 kDa expected for the  $\text{A}_2\text{B}$  protomer. In contrast, the fit of the data at high concentrations to Equation 1 gave large and non-random residuals (Figure 2b), suggesting either the presence of multiple species and/or non-ideality (44). In this situation, the best fit to Equation 1 does not yield the MW of any individual species but rather the mean weight-average MW ( $\bar{M}_w$ ) of all of the species in the sample (44,50).

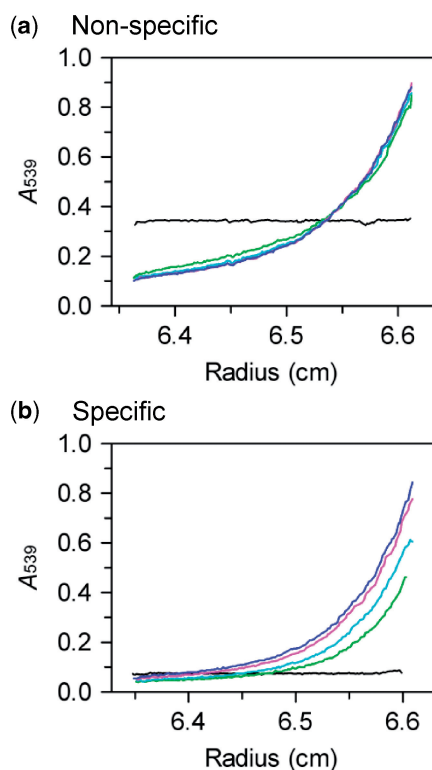
Values for  $\bar{M}_w$  were obtained at each loading concentration of BcgI tested (Figure 3). The  $\bar{M}_w$  values increased with increasing levels of BcgI: from that corresponding to a single  $\text{A}_2\text{B}$  unit at low concentrations, then to values consistent with dimers of  $\text{A}_2\text{B}$  and finally to levels indicating assemblies with more than two protomers. At the highest concentration tested (18 μM: Figure 2b), the  $\bar{M}_w$  came to 431 kDa. As these mean values for  $\bar{M}_w$  encompass the full range of protein concentrations present in the centrifuge cell after sedimentation to equilibrium, from low concentrations at the inside edge to high at the outside, a significant fraction of the protein in this



sample must be present in assemblies containing more than two protomers.

AUC was also used to analyse complexes of WT BcgI with either a specific duplex carrying both recognition and cleavage sites (HEX-42S) or a non-specific DNA that lacked the recognition sequence (HEX-42NS). Sedimentation was monitored at 539 nm, to detect the HEX moiety on the DNA, either in its free state or when enzyme-bound. In the absence of BcgI, MW values of 27.6 and 26.6 kDa were obtained for HEX-42S and HEX-42NS, respectively (data not shown), close or equal to the expected MW, 26.6 kDa. When the non-specific duplex (2  $\mu$ M) was centrifuged in the presence of excess BcgI (7.5  $\mu$ M), the initial uniform distribution of the HEX label throughout the cell (Figure 4a, black trace) was converted into the characteristic profiles for an approach to sedimentation equilibrium (Figure 4a, coloured traces). The records at equilibrium (>22 h) yielded a  $\overline{M}_w$  of 349 kDa, much larger than that for the free DNA so all of the DNA must be bound to the

protein. The  $\overline{M}_w$  for the non-specific complex is similar to that for the same concentration of free protein (Figure 3), which indicates that either two or more protomers can bind to one duplex or that the DNA–protein complexes associate with themselves. Conversely, after adding BcgI (7.5  $\mu$ M) to the specific duplex (2  $\mu$ M) and spinning the sample at low velocity, the solution lacked virtually any absorbance at 539 nm (Figure 4b, black trace). Subsequent scans after centrifugation at a higher speed revealed a progressive recovery in the  $A_{539}$  signal (Figure 4b, coloured traces) but even after 54 h, the total absorbance in the cell containing specific DNA was still lower than that containing non-specific DNA and had yet to reach equilibrium. The complex of BcgI with specific DNA must precipitate out of solution as soon as it is formed and then slowly redissolves back into solution. An 18 bp duplex that contains the recognition site but not the cleavage loci also precipitated from solution on the addition of WT BcgI, as did the specific 42 bp duplex with E53A in the presence of either  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$  (data not shown).



**Figure 4.** AUC of BcgI–DNA complexes. The samples, in AUC buffer at 20°C, contained WT BcgI protein (7.5  $\mu$ M) and a DNA duplex (2  $\mu$ M) as follows: (a) HEX-42NS, a non-specific DNA that lacks the recognition sequence; (b) HEX-42S, a specific DNA with both recognition and cleavage loci. The samples were centrifuged in the AUC and the absorbance across the cell monitored at 539 nm at the following stages after initiating the run: the instrument was initially set to spin at 3000 rpm and an absorbance scan recorded immediately on reaching that speed (black trace). The velocity was then raised to 7500 rpm and scans taken after 14 h and after further 8 h delays. The records shown are after 14 h (green trace), 22 h (cyan trace), 38 h (pink trace) and 54 h (blue trace). The data sets with the non-specific duplex (a) recorded after  $\geq 30$  h were fitted to Equation 1: the best fit gave a  $\overline{M}_w$  of 349 kDa.

#### DNA cleavage

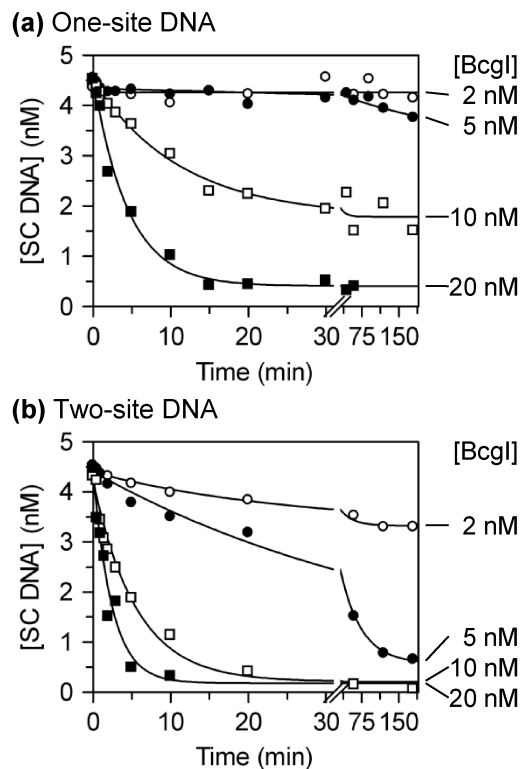
During its DNA cleavage reaction, the BcgI protein interacts with two copies of its recognition sequence and then proceeds to make four DSBs, one either side of each site, thus cutting a total of eight phosphodiester bonds in a single DNA–protein complex (41). It is highly unlikely that one A<sub>2</sub>B protomer can cut all eight bonds because it contains only two active sites for phosphodiester hydrolysis, one in each A subunit. Instead, it seems more likely that DNA cleavage is carried out by an assembly of A<sub>2</sub>B units, maybe like those seen by native MS and AUC. This possibility was addressed by analysing how the extent of cleavage of plasmids with one or two BcgI sites varied as the protein concentration was raised from below to above the level of BcgI sites in the DNA (Figure 5). BcgI can act with two recognition sites *in cis* or *in trans*, although, as is usual for interactions bridging two loci (16), it prefers sites *in cis* over those *in trans*. Two-site substrates are thus cleaved more readily than one-site DNA (24,40).

In previous studies (40), the extent of cleavage of plasmids with one or two BcgI sites was measured at varied BcgI concentrations after a fixed time (15 min): complete cleavage of the DNA within this time required a molar excess of protein over DNA, with more protein needed for the one-site than for the two-site DNA. As the extent of cleavage after a fixed time interval may reflect either the end point or the rate of the reaction, reactions at varied BcgI concentrations were monitored in this study over extended time bases (Figure 5). On the one-site plasmid, no—or little—cleavage was detected at BcgI levels below or equal to that of the DNA, even after prolonged incubation (Figure 5a). Instead, the reaction on the one-site substrate required enzyme concentrations in excess of the DNA. The extent of cleavage increased as the ratio of enzyme to recognition sites was raised above 1:1 but complete cleavage was only observed with a substantial excess of protein over DNA. The plasmid with

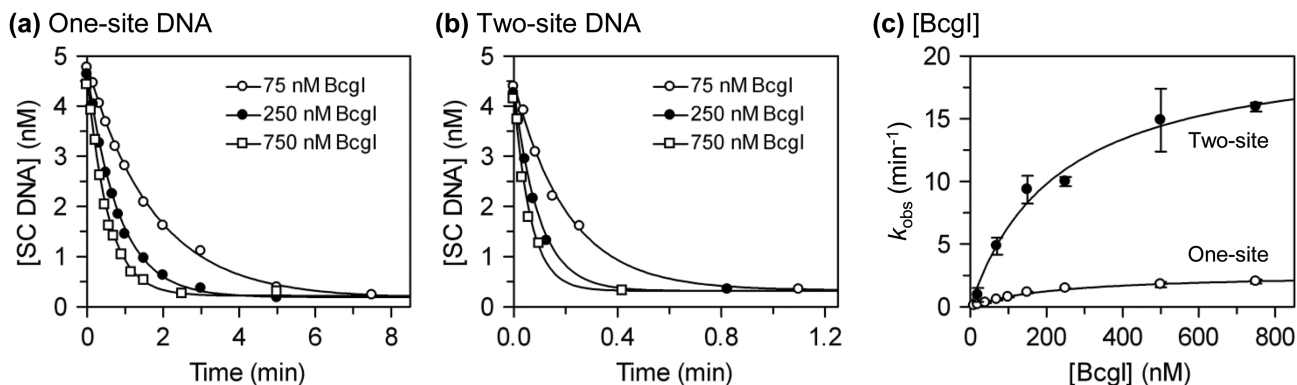
two cognate sites might have been expected to require more enzyme per mole DNA to fill both sites but instead almost all of the two-site DNA was cleaved at equal levels of enzyme and DNA: i.e. with BcgI at half the concentration of recognition sites (Figure 5b).

Enzyme concentrations that gave complete cleavage of the DNA did not necessarily give maximal reaction rates: for instance, the two-site plasmid (present at 5 nM) was cleaved fully by both 10 and 20 nM BcgI protein but the latter gave a faster rate (Figure 5b). Cleavage of both one- and two-site substrates was therefore studied across a range of BcgI concentrations far above that of the DNA (Figure 6). During these reactions, the concentrations of the SC DNA substrates fell exponentially to end points where essentially all of the DNA had been cleaved. With enzyme in excess over the DNA, the rate constants ( $k_{\text{obs}}$ ) from the exponential decline of the SC DNA denote the cleavage of at least one phosphodiester bond in the enzyme–DNA complex rather than the entire reaction ending with the release of the DNA cleaved at eight bonds. Phosphodiester hydrolysis by BcgI bound to the one-site substrate took  $\geq 2$  min to reach completion and could be monitored by manually mixing the reagents (Figure 6a). In contrast, BcgI bound to the two-site plasmid could take  $<0.2$  min to cut all of the DNA, so in these cases the experiments were done in a rapid quench-flow device (Figure 6b). At BcgI concentrations that permitted both hand-mixing and quench-flow approaches, the two techniques yielded comparable results (data not shown).

The values for  $k_{\text{obs}}$  (Figures 6a and 6b: other data not shown) were plotted against the enzyme concentration (Figure 6c). For both one- and two-site plasmids, the rates increased with increasing BcgI concentrations in a hyperbolic manner (Figure 6c). The data with each substrate were fitted to an equation for a rectangular hyperbola to yield values for  $k_{\text{max}}$  (the rate constant for DNA cleavage at saturation of the DNA with protein) and for

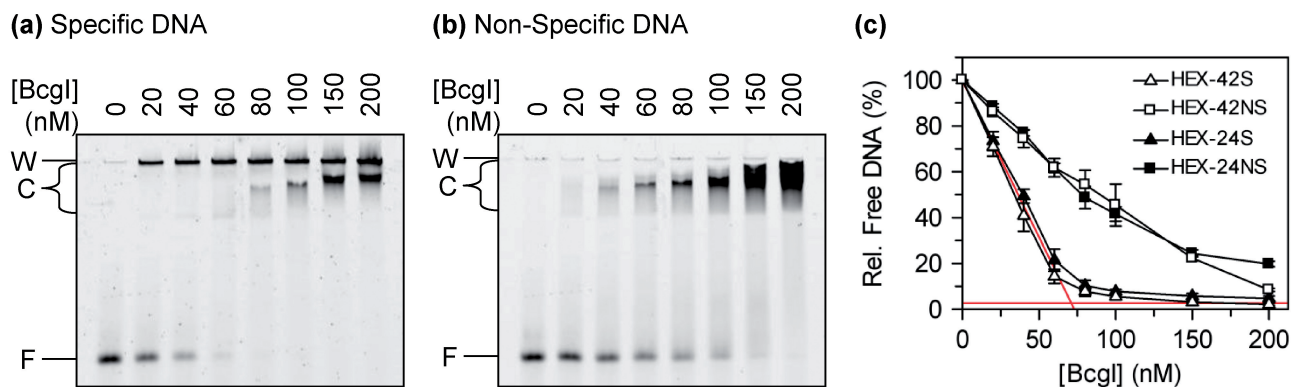


**Figure 5.** DNA cleavage at low [BcgI]. Reactions in buffer R at 37°C contained one of the following plasmids ( $^3\text{H}$ -labelled, initially  $\sim 90\%$  SC monomer) at 5 nM and BcgI protein at the concentrations noted below. The plasmids were (a) pUC19, which has one BcgI site; (b) pDG5, which has two sites. The enzyme concentrations for each reaction are marked on the right: 2 nM, white circles; 5 nM, black circles; 10 nM, white squares; 20 nM, black squares. Samples were taken at the indicated times and analysed by electrophoresis to separate the SC substrate from the cleaved products. The residual concentrations of the SC DNA were determined and are plotted as a function of time. The lines are the best fits to single exponential decays with floating offsets.



**Figure 6.** DNA cleavage at high [BcgI]. (a) The reactions, in buffer R at 37°C, contained 5 nM  $^3\text{H}$ -labelled SC pUC19 and BcgI protein at 75 nM, white circles; 250 nM, black circles; 750 nM, white squares. After manually mixing the reagents, samples were taken from the reactions by pipette and mixed immediately with an EDTA quench. The samples were then subjected to electrophoresis through agarose, to enable the concentration of SC DNA to be determined. The amount of SC DNA present in each sample is plotted against the time. (b) As (a) except that the DNA was SC pDG5 and that the reactions were carried out by quench-flow. (c) Apparent rate constants ( $k_{\text{obs}}$ ) from  $\geq 3$  independent experiments at each BcgI concentration were averaged and plotted against the concentration: for reactions on pUC19, white circles; for reactions on pDG5, black circles. Error bars denote standard deviations. The line drawn through each data set is the best fit to a hyperbolic function,  $k_{\text{obs}} = (k_{\text{max}} \times [\text{BcgI}]) / (K_{1/2} + [\text{BcgI}])$ . The best fits were obtained with the following: for pUC19,  $k_{\text{max}} = 2.7 \text{ min}^{-1}$  and  $K_{1/2} = 232 \text{ nM}$ ; for pDG5,  $k_{\text{max}} = 21 \text{ min}^{-1}$  and  $K_{1/2} = 230 \text{ nM}$ .





**Figure 7.** DNA binding. (a and b) The mixtures contained, in 10  $\mu$ l buffer R\*C, 20 nM DNA, either HEX-42S (a) or HEX-42NS (b), and BcgI at one of the following levels, from left to right: 0, 20, 40, 60, 80, 100, 150 or 200 nM. After 5 min at 37°C, the samples were subjected to electrophoresis through polyacrylamide and the fluorescence from each band recorded. On the left of gel images, F marks the position of the free DNA; C, the DNA-protein complexes that entered the gel; and W, the bottom of the wells. (c) From titrations of both HEX-42S and HEX-42NS with increasing amounts of BcgI protein, and from additional experiments with 24 bp duplexes (HEX-24S and HEX-24NS: gels not shown), the amount of free DNA in each sample was determined relative to the amount in a parallel lane without enzyme. The % DNA in the free form is shown as a function of the BcgI concentration: HEX-42S, white triangle; HEX-42NS, white squares; HEX-24S, black triangles; HEX-24NS, black squares. Mean values from three repeats are given; error bars show standard deviations. The red lines through the data with the specific duplexes denote stoichiometric binding where all of the added BcgI is bound to DNA until saturation of the DNA is achieved.

$K_{1/2}$  (the protein concentration for half maximal rate). The two-site plasmid gave nearly a 10-fold higher value for  $k_{max}$  than the one-site DNA, 21  $\text{min}^{-1}$  compared with 2.7  $\text{min}^{-1}$ . In contrast, the one- and two-site substrates gave the same value for  $K_{1/2}$ , 230 nM in both cases. The faster rate of cutting the two-site over the one-site plasmid is thus due solely to the different  $k_{max}$  values. The DNA cleavage step in the BcgI reaction on sites *in cis* is thus faster than that on sites *in trans*.

The values for  $K_{1/2}$  merit two comments. Firstly, they indicate that remarkably high BcgI concentrations are required for its maximal reaction rate: cleavage of 5 nM DNA at 90% of its maximal rate requires an enzyme concentration of >2000 nM. Secondly, though proteins that bind two DNA sites nearly always have higher affinities for sites *in cis* than sites *in trans* (16), the equal  $K_{1/2}$  values show that reactions with two sites *in trans* feature the same affinity for BcgI protein as that during reactions with sites *in cis*. The  $K_{1/2}$  values may therefore reflect a weak protein-protein association in the assembly of the reaction complex rather than the DNA-protein association.

### DNA binding

An alternative explanation for why DNA cleavage by BcgI requires such large amounts of protein is that the enzyme simply has a low affinity for its recognition sequence. To test this possibility, the binding of the BcgI protein to specific and non-specific DNA was examined directly by gel retardation (Figure 7). The binding studies were initially done with two 42 bp oligoduplexes, both labelled with HEX at one 5' terminus. One, HEX-42S, is a potential substrate, as it carries both the recognition sequence and the cleavage loci both sides of the site. The other, HEX-42NS, differs only at 2 bp, one in each segment of the bipartite site so is effectively a non-specific sequence for BcgI (41). The buffer used here contained  $\text{Ca}^{2+}$  rather than  $\text{Mg}^{2+}$ , to allow WT BcgI to bind to

HEX-42S without cutting off its ends. Increasing concentrations of BcgI were added to a fixed amount of each DNA and the free DNA then separated by electrophoresis from the protein-bound species with retarded mobilities (Figures 7a and 7b). The extent of binding at each enzyme concentration was then determined by assessing the fluorescence from the band corresponding to the free DNA relative to that in samples without enzyme (Figure 7c). Similar results (not shown) were obtained with the inactive E53A mutant with either  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$  and with no metal present.

When increasing concentrations of BcgI were added to the specific duplex, HEX-42S, the concentration of the free DNA decreased linearly with the protein concentration until it had all been converted to DNA-protein complexes (Figure 7c). BcgI thus binds to DNA carrying its recognition sequence in a stoichiometric fashion, where all of the added BcgI is bound to the DNA until saturation of the DNA is achieved, as opposed to a dynamic equilibrium between free and bound proteins. However, saturation of the specific duplex seems to require 3–4 mole BcgI per mole DNA, as it took ~70 nM BcgI to saturate 20 nM DNA (Figure 7c, intersection of red lines). BcgI also bound to the non-specific DNA, HEX-42NS, although with a lower affinity than the specific sequence (Figures 7b and 7c). Moreover, the specific and non-specific sequences gave rise to different complexes. The specific DNA was initially transformed to species that failed to enter the polyacrylamide gel and remained in the wells (Figure 7a): species that entered the gel were observed only after sufficient enzyme had been added to convert all of the DNA to complexes. The nature of the complexes that remain in the wells and those that enter the gel have yet to be established. They may be related to the complexes seen in the AUC, where the addition of BcgI to cognate DNA led initially to an insoluble complex which then slowly redissolved back into solution (Figure 4b). In contrast, all of the complexes formed with the non-specific

DNA entered the gel (Figure 7b) but, as with the specific DNA, the species within the gel displayed progressively reduced mobilities with increasing amounts of BcgI, probably due to the binding of additional molecules of protein to each molecule of DNA.

Further experiments were carried out with 24 bp duplexes with or without the recognition sequence, HEX-24S and HEX-24NS: the former contains the recognition sequence but lacks the cleavage loci 10 and 12 nt away from the site. Nevertheless, BcgI bound to both of these 24 bp duplexes in the same manners (gels not shown) and with the same affinities (Figure 7c) as their 42 bp counterparts. The cleavage loci are therefore not needed for specific binding to the site.

A striking feature of the equilibrium binding of BcgI to its cognate sequence is that a much lower concentration of the protein is required to saturate binding to the site (Figure 7c) compared with the amount needed to saturate the DNA cleavage rate (Figure 6c): 70 *cf.* >2000 nM. A scheme that can account for this behaviour is a mechanism of the type seen with the FokI restriction enzyme (30–32),



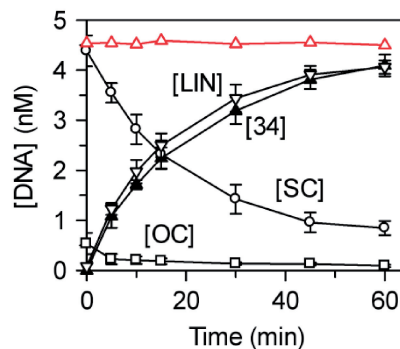
In this scheme, one molecule of the enzyme (E) first binds to its DNA site (S) with an equilibrium dissociation constant  $K_{DNA}$  but the resultant E.S complex has no activity until the DNA-bound protein associates with a second protein, with a constant  $K_{prot}$  for the protein–protein assembly: the complex containing two molecules of enzyme then proceeds to cleave the DNA at the  $k_{max}$  rate. In the case of FokI (31,32), the  $K_D$  for the binding of the enzyme to its site (4 nM) is much smaller than that for the protein–protein association (in the range from 100 to 166 nM), so the level of FokI needed for its maximal DNA cleavage rate is much higher than that required to fill its binding sites on the DNA. The same seems to apply to BcgI, though the stoichiometry of the binding reaction (Figure 7c) implies that even its initial E.S complex must contain more than one  $A_2B$  unit, and that it then proceeds to bind yet more  $A_2B$  units. Indeed, if the  $K_{1/2}$  from its maximal cleavage rate (230 nM) corresponds to its value for  $K_{prot}$ , then the E.S complexes for BcgI and for FokI display similar affinities for their additional enzyme.

### Activation by BcgIA

The monomer of FokI at its recognition site can be activated by associating not only with intact FokI protein but also with just its catalytic domain (30). The association of the intact protein and the catalytic domain generates a unit with two active sites that can then proceed to cut both DNA strands (33). The nuclease function of the BcgI protein resides in its A subunit while the sole function of B is DNA recognition (12,39). Hence, it might be possible to activate the BcgI REase at its recognition site by excess A protein, isolated from the B subunit. To test this idea, the *bcgIA* gene was cloned and over-expressed in *E. coli* cells that lacked *bcgIB* and the A protein purified to homogeneity (see ‘Materials

and Methods’ section). In solution, BcgIA was found by AUC and native MS to exist primarily as a 71.5 kDa monomer (48). The fact that the isolated BcgIA protein exists as a monomer concurs with the native MS on the native BcgI protein, which had shown A monomers, AB and  $A_2B$  assemblies, but no  $A_2$  dimers (Figure 1).

When the purified A protein was added by itself to the one-site plasmid pUC19, no DNA cleavage was observed even after long incubations with high BcgIA concentrations (Figure 8, red line). BcgIA was then added to pUC19 together with intact BcgI protein. However, the latter was present at the same concentration (5 nM) as the DNA, which is too low for it to cleave the DNA unaided (Figure 5a). Samples were taken from the mixture and a portion of each analysed by electrophoresis through agarose to separate the SC, OC and LIN forms of pUC19. Though neither protein alone was capable of cutting this DNA, together they rapidly converted most of the SC plasmid directly to LIN DNA, without accumulating nicked OC forms en route (Figure 8). The LIN DNA could have been cut in both strands on just one side of the BcgI site, to give a LIN species of 2686 bp, or on both sides to release the 34-mer product and a LIN DNA of 2652 bp. As these LIN forms are inseparable on agarose, a second portion from each sample was applied to a polyacrylamide gel to capture the 34-mer (41). The increase in the concentration of the 34-mer matched that for the LIN DNA (Figure 8). The A protein must therefore enable the intact protein to make DSBs on both sides of its site. Hence, both A subunits in the  $A_2B$  protomer at the recognition site interact with free BcgIA monomers to give an assembly with two pairs of nuclease active sites, each of which makes a DSB. If the addition of A to  $A_2B$



**Figure 8.** Activation of BcgI by BcgIA. The reactions, in 200  $\mu$ l buffer  $R^*$  at 37°C, contained 100 nM BcgIA protein and 5 nM  $^3H$ -labelled pUC19 (initially 90% SC monomer), in either the absence or presence of 5 nM WT BcgI. Aliquots (20  $\mu$ l) were removed at the times shown and immediately quenched before analysis by electrophoresis: one-half through agarose, to separate the SC, OC and LIN forms of the plasmid; the other half through polyacrylamide, to capture the 34-mer. The concentrations of the various forms of pUC19 were evaluated from the agarose gel and that for the 34-mer from the polyacrylamide gel (see ‘Materials and Methods’ section). For the reaction with BcgIA alone, only the level of SC DNA is shown: red unfilled triangles and line. For the reaction with both BcgIA and native BcgI, the following were assessed: intact SC DNA, white circles; nicked OC DNA, white squares; LIN DNA, with at least one DSB at the BcgI site, white inverted triangles; the 34-mer, black triangles. Mean values from three repeats are shown: error bars denote standard deviations.

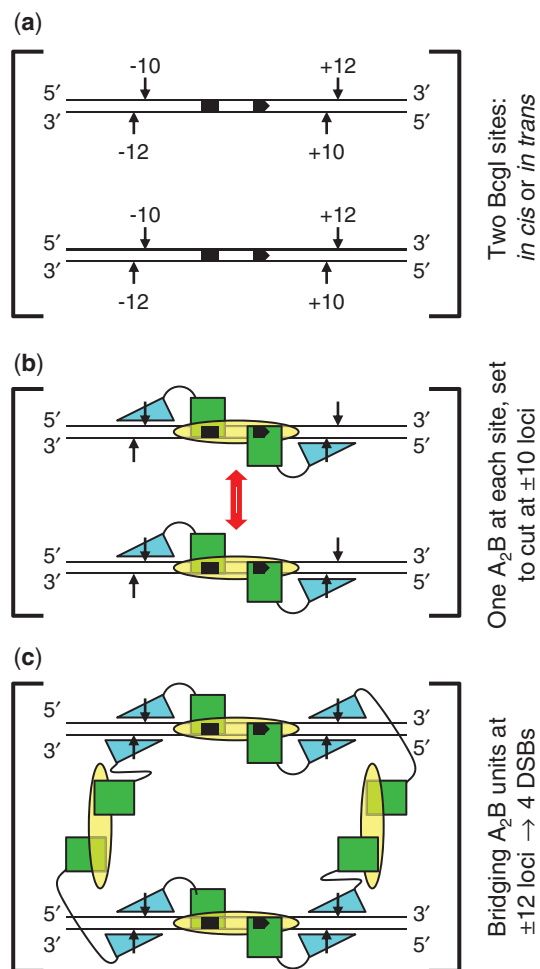
had resulted in only one unit for DSBs, it would have first cut one side of the site and only later the other side: the 34-mer would then have been liberated not at the same time as the LIN species but instead at some later time. Strikingly, extra A subunits also enhance the MTase activity of the BcgI protein (46) but, as the MTase methylates only one adenine at a time, the activation of the MTase almost certainly does not require the catalytic functions for methyl transfer in the free A subunit. In contrast, the liberation of the 34-mer shows that the extra A subunits contribute to the BcgI REase functional active sites for phosphodiester hydrolysis.

The effect of the A protein was also examined at higher concentrations of native BcgI to see whether or not additional BcgIA could still enhance the cleavage rate. In reactions containing 25 nM WT BcgI and 5 nM pUC19, all of the DNA was cleaved readily (Figure 5a). When these reactions were supplemented with 100 nM BcgIA (as in Figure 8), no difference was observed in either the extent or the rate of the reaction relative to that in the absence of the A protein (data not shown). Hence, the additional catalytic domains that the REase at the recognition site needs for its nuclease reaction can come from either free A subunits or excess A<sub>2</sub>B protomers. Nevertheless, the activation requires a lower concentration of A<sub>2</sub>B units than free A subunits.

## DISCUSSION

The substrate for the REase activity of the BcgI RM protein (Figure 9a) consists of two copies of its recognition sequence (24,40,41). In contrast to the many restriction enzymes that are most active when bound to two sites but which still have some activity when bound to a single site (22,54–56), the BcgI REase has an absolute requirement for two sites. It has no detectable activity when bound to a single site on a DNA held in physical isolation from any other DNA, but that isolated site is cleaved when a second specific DNA is added (41). At each site, it cuts both strands at two separate loci, one 10/12 nt upstream and the other 12/10 nt downstream (38). After binding to its substrate, the cleavage reaction yields none of the nicked species cut in one strand and only low levels of the intermediates with one DSB. Instead, most of the DNA is converted directly to the final products with two DSBs at both sites. Hence, after spanning two sites, the REase can cut all eight bonds within the lifetime of the synaptic complex (41). To achieve this feat, the REase activity of BcgI most likely requires an assembly containing eight catalytic centres for phosphodiester hydrolysis. Given the highly concerted nature of its reaction, it is unlikely that it sequentially moves its catalytic centres from bond to bond, especially not from a bond in the 5'–3' strand to one in the 3'–5' strand.

The A subunit of the BcgI RM system carries both REase and MTase functions, while B resembles a DNA specificity subunit (HsdS) from a Type I RM system (12,39). HsdS subunits recognize bipartite sequences via separate domains for each segment, with each domain linked to a conserved region that associates with a



**Figure 9.** Scheme for action at eight phosphodiester bonds by BcgI. (a) The substrate for the REase activity of BcgI consists of two copies of its recognition sequence. The sites, shown here aligned in parallel, can be located either on the same DNA molecule, *in cis*, or on separate molecules, *in trans*. The bipartite recognition site is indicated by black segments, with an arrowhead to mark its 5'–3' orientation: strand polarities are also indicated at DNA termini. Cleavage loci 10 and 12 nt distant from the site are marked with vertical arrows: upstream and downstream loci are noted with – or + signs, respectively. (b) The two sites in (a) are both shown bound to one A<sub>2</sub>B protomer of BcgI: an interaction between the protomers is suggested by the double-headed arrow (in red). The B subunit (yellow oval) is responsible for DNA specificity and is positioned spanning both segments of the recognition sequence. The A subunits carry both REase and MTase activities and are illustrated as separate domains connected by a flexible linker. The MTase domains (green squares) overlay the sites of methylation in the specified segments of the recognition sequence. The REase domains (cyan triangles) are placed against the target bonds at the –10 and +10 positions, upstream and downstream of the site in top and bottom strands, respectively: to mark their orientation on the DNA, the triangle points towards the 5'–end of each strand. At this stage, BcgI has yet to engage the scissile bonds at the –12 and +12 positions. (c) As (b) except that two additional A<sub>2</sub>B units from free solution now bridge the A<sub>2</sub>B units bound to each site via interactions between the nuclease domains of the free and the DNA-bound units. The two nuclease domains from one of the extra A<sub>2</sub>B units engage the scissile bonds at the –12 positions upstream of both recognition sites (on the left, as drawn in [c]) while those from the other occupy the downstream +12 positions in both sites (on the right). The scheme thus results in a dimer of nuclease domains, with two catalytic centres for phosphodiester hydrolysis at all four loci where BcgI makes a DSB.



MTase subunit in a pseudo-symmetric fashion (6,57). BcgIB appears to be organized similarly (39). Following SDS-PAGE, it appeared from protein staining that BcgI contained two mole A per mole B (K94), but gel stains are often far from quantitative (42). Nevertheless, native MS of the intact protein under conditions that minimize the disruption of non-covalent complexes (49) validated the earlier proposal and showed that BcgI is indeed composed of  $A_2B$  units (Figure 1). Under these conditions, however, we also find some lone A and B subunits and an AB assembly but no  $A_2$  dimers. In all likelihood, the B subunit occupies the central position in the  $A_2B$  assembly with A subunits symmetrically disposed on either side, both interacting with B but not with each other. This configuration is the same as that for the  $M_2S$  structure of a Type I MTase (6,57) and a similar arrangement has been proposed for the AhdI MTase, a system that has a Type II REase but a Type I-like MTase (58). Native MS also revealed the presence of species containing multiple copies of the  $A_2B$  protomer, from an  $(A_2B)_2$  dimer to at least up to a  $(A_2B)_4$  tetramer, albeit with progressively decreasing peak heights at higher masses (Figure 1). Yet, while native MS can identify these larger species, peak heights do not provide quantitative information about the amount of each species present in the sample as it was injected from the nano-ESI, since the sensitivity of the ion detector declines with increasing  $m/z$  values. Nonetheless the spectra indicated no preference for one oligomer over another, as the intensity of the peaks from the aggregates declined steadily with size.

AUC was therefore used to investigate the self-association of the BcgI protein (Figures 2 and 3). At sub-micromolar concentrations, BcgI existed in solution as a single ideal species with a MW matching that for an individual protomer (Figure 2a). But at micromolar concentrations, of the order of those required for maximal nuclease activity (Figure 6c), multiple aggregates were present (Figure 2b). The  $\overline{M}_w$  values increased with increasing concentrations of BcgI, to levels above that for a dimeric  $(A_2B)_2$  unit (Figure 3). As the  $\overline{M}_w$  values reflect all of the species in the cell, the mixtures must contain some of greater mass than the dimer. BcgI thus readily associates with itself to give large assemblies containing several copies of the  $A_2B$  protomer. Its complex with non-specific DNA also contained multiple protomers (Figure 4a). In contrast, the addition of BcgI to a specific duplex gave an insoluble complex that precipitated immediately from solution, as judged by its removal from solution on centrifugation at low velocity (Figure 4b) and by its inability to enter a polyacrylamide gel on electrophoresis (Figure 7a). The binding of BcgI to specific DNA thus generates highly aggregated states that possibly link together several DNA molecules, but these then slowly dissociate to release individual DNA-protein complexes back into solution.

The cleavage of a plasmid with one BcgI site required excess protein over DNA (Figure 5a), which raises the possibility that this REase may carry out just one turnover. It had been suggested that further cleavage reactions were blocked by the excised 34-mer remaining bound to the enzyme while it was being methylated (12).

Yet the fact that nuclease reactions on a plasmid with two sites proceed to completion when the enzyme is present at lower concentrations than sites on the DNA shows that the BcgI REase is capable of multiple turnovers, with each enzyme molecule participating in more than one cleavage event. BcgI is therefore not a suicide enzyme (59,60), one that is inactivated by its own reaction product. The need for excess enzyme for its reaction *in trans* on one-site DNA, despite its reactions *in cis* featuring multiple turnovers, can be accounted for if the BcgI proteins bound to both recognition sites have to recruit additional molecules of protein before they become active. In this situation, a reaction *in trans* will require two DNA molecules to be both loaded with enzyme at the recognition site and with the extra enzyme. But if the enzyme binds tightly to its cognate site as seems to be the case here (Figure 7c), fully loaded molecules of the one-site DNA carrying the extra enzyme will appear in solution only after adding a stoichiometric excess of enzyme over sites. Conversely, for a reaction *in cis*, the BcgI molecules bound to the two recognition sites in the same DNA molecule may interact with each other cooperatively (16), which could lead directly to the capture of the additional molecules of protein required for the cleavage reaction.

When bound to its recognition site, the  $A_2B$  protomer is in a position to place its two nuclease domains against two of the four target phosphodiester bonds at that site (Figure 9b). This would give a structure similar to that for M.AhdI (58), an  $M_2S$  MTase but which is thought from neutron scattering data to have arms extending from both M subunits. In the scheme shown in Figure 9b, the nuclease domains of the A subunits are placed against the -10 and +10 targets in the two strands, one upstream and one downstream of the site, but an equally plausible model would put these domains against the -12 and +12 bonds. On the other hand, it is improbable that an  $A_2B$  unit places one nuclease centre against a bond 10 nt distant from the site and the other 12 nt away, as this would break the intrinsic symmetry of the system. Moreover, the positioning of one nuclease domain at a 10 position and the other at a 12 position calls for an implausible reconfiguration of one A subunit: an attack on both scissile bonds in the same strand (say -10 and +12) requires a 180° rotation of the nuclease domain in one A subunit relative to the other, to keep register with strand polarity; alternatively, to attack both strands at one cleavage locus (say -10 and -12), the distance between the nuclease and MTase domains in one A subunit would have to be massively different from that in the other. The synaptic complex of two  $A_2B$  protomers and two recognition sites (Figure 9b) would thus seem to be incapable of making a DSB at any single cleavage locus. Indeed, it is probably incapable of cutting even one bond since no nicked species appear during DNA cleavage by BcgI (Figure 8: 41).

The monomer of the FokI REase binds to its asymmetric site via its DNA recognition domain (27) and then moves its catalytic domain to its target bond, 13 nt away in the bottom strand, but the REase remains inactive, not even nicking the DNA (32). It becomes active only after a second catalytic centre engages the target bond in the top

strand. The second centre can come from another monomer of the FokI protein, either in free solution or bound elsewhere to the same DNA (31), or from just its catalytic domain (30). A similar scheme may also apply to the BcgI REase (Figure 9c). Much higher concentrations of BcgI protein are required to cleave DNA (Figure 6c) than to bind to specific DNA (Figure 7c) so the complex formed at the cognate site must remain inactive until it recruits additional molecules of the protein from free solution. BcgI bound to its site can also be activated by the A subunit alone (Figure 8). The function of either the additional A<sub>2</sub>B protein, or BcgIA alone, is presumably to associate transiently with the A subunits of the DNA-bound enzyme to position two catalytic centres for phosphodiester hydrolysis at each cleavage locus: the two centres can then cut both strands at that locus. Moreover, as the activation of the BcgI REase by the A protein alone leads directly to the excision of the 34-mer (Figure 8), both A subunits in the A<sub>2</sub>B protomer on the DNA must interact with free A protein before any reaction.

Figure 9 shows a proposal for the juxtaposition of eight catalytic centres against the eight target bonds in the BcgI substrate. In this scheme, one A<sub>2</sub>B protomer binds to each recognition site and positions its two centres for nuclease activity, one in each A subunit, against one particular set of phosphodiester bonds, illustrated here (Figure 9b) as the -10 and +10 targets. Two additional A<sub>2</sub>B protomers from free solution then become incorporated into the synaptic complex through interactions between their nuclease domains and the nuclease domains of the DNA-bound proteins (Figure 9c). The additional protomers bridge the two DNA sites, so that (in the configuration drawn) one of the extra protomers engages the -12 targets upstream of both sites and the other the +12 targets downstream of both sites. Alternative configurations for the bridging protomers are also possible: for example, they could span “diagonally” the upstream locus at one site and the downstream locus at the other site. Indeed, the insolubility of the specific DNA-protein complex for BcgI seen in the AUC (Figure 4b) could be due to the two bridging protomers spanning not the same two DNA-bound protomers to form the closed structure shown in Figure 9c but rather an open-ended structure with one DNA-bound protomer being linked via bridging units to two different DNA-protein complexes.

## ACKNOWLEDGEMENTS

We thank New England Biolabs for the over-producing plasmid for the native BcgI R-M system, and James McCullagh (Mass Spectrometry facility, Department of Chemistry, University of Oxford) for use of the Synapt MS.

## FUNDING

Biotechnology and Biological Sciences Research Council [BB/C513077/1]; Wellcome Trust [078794]. Funding for open access charge: Wellcome Trust.

*Conflict of interest statement.* None declared.

## REFERENCES

- Wilson,G.G. and Murray,N.E. (1991) Restriction and modification systems. *Annu. Rev. Genet.*, **25**, 585–627.
- Murray,N.E. (2002) Immigration control of DNA in bacteria: self versus non-self. *Microbiology*, **148**, 3–20.
- Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.*, **38**, D234–D236.
- Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S.K., 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.
- Dryden,D.T.F., Murray,N.E. and Rao,D.N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.*, **29**, 3728–3741.
- Kennaway,C.K., Taylor,J.E., Song,C.F., Potrzebowski,W., Nicholson,W., White,J.H., Swiderska,A., Obarska-Kosinska,A., Callow,P., Cooper,L.P. *et al.* (2012) Structure and operation of the DNA-translocating Type I DNA restriction enzymes. *Genes Dev.*, **26**, 92–104.
- Smith,R.M., Diffin,F.M., Savery,N.J., Josephsen,J. and Szczelkun,M.D. (2009) DNA cleavage and methylation specificity of the single polypeptide restriction-modification enzyme LlaGI. *Nucleic Acids Res.*, **37**, 7206–7218.
- Cheng,X. and Roberts,R.J. (2001) AdoMet-dependent methylation, DNA methyltransferases and base flipping. *Nucleic Acids Res.*, **29**, 3784–3795.
- Pingoud,A., Fuxreiter,M., Pingoud,V. and Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.*, **62**, 685–707.
- Perona,J.J. (2002) Type II restriction endonucleases. *Methods*, **28**, 353–364.
- Janulaitis,A., Petrusyte,M., Maneliene,Z., Klimasauskas,S. and Butkus,V. (1992) Purification and properties of the Eco57I restriction endonuclease and methylase - prototypes of a new class. *Nucleic Acids Res.*, **20**, 6043–6049.
- Kong,H., Roemer,S.E., Waite-Rees,P.A., Benner,J.S., Wilson,G.G. and Nwankwo,D.O. (1994) Characterization of BcgI, a new kind of restriction-modification system. *J. Biol. Chem.*, **269**, 683–690.
- Piekarowicz,A., Golaszewska,M., Sunday,A.O., Siwinska,M. and Stein,D.C. (1999) The HaeIV restriction modification system of *Haemophilus aegyptius* is encoded by a single polypeptide. *J. Mol. Biol.*, **293**, 1055–1065.
- Morgan,R.D., Bhatia,T.K., Lovasco,L. and Davis,T.B. (2008) Mmel: a minimal Type II restriction-modification system that only modifies one DNA strand for host protection. *Nucleic Acids Res.*, **36**, 6558–6570.
- Mucke,M., Kruger,D.H. and Reuter,M. (2003) Diversity of Type II restriction endonucleases that require two DNA recognition sites. *Nucleic Acids Res.*, **31**, 6079–6084.
- Halford,S.E., Welsh,A.J. and Szczelkun,M.D. (2004) Enzyme-mediated DNA-looping. *Annu. Rev. Biophys. Biomol. Struct.*, **33**, 1–24.
- Yang,C.C. and Topal,M.D. (1992) Nonidentical DNA-binding sites of endonuclease NaeI recognize different families of sequences flanking the recognition site. *Biochemistry*, **31**, 9657–9664.
- Embleton,M.L., Siksnys,V. and Halford,S.E. (2001) DNA cleavage reactions by type II restriction enzymes that require two copies of their recognition sites. *J. Mol. Biol.*, **311**, 503–514.
- Wentzell,L.M., Nobbs,T.J. and Halford,S.E. (1995) The SfiI restriction endonuclease makes a four-strand DNA break at two copies of its recognition sequence. *J. Mol. Biol.*, **248**, 581–595.
- Deibert,M., Grazulis,S., Sasnauskas,G., Siksnys,V. and Huber,R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.*, **7**, 792–799.
- Vanamee,E.S., Viadiu,H., Kucera,R., Dorner,L., Picone,S., Schildkraut,I. and Aggarwal,A.K. (2005) A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. *EMBO J.*, **24**, 4198–4208.

22. Zaremba, M., Owsicka, A., Tamulaitis, G., Sasnauskas, G., Shlyakhtenko, L.S., Lushnikov, A.Y., Lyubchenko, Y.L., Laurens, N., van den Broek, B., Wuite, G.J.L. *et al.* (2010) DNA synapsis through transient tetramerization triggers cleavage by Ecl18kI restriction enzyme. *Nucleic Acids Res.*, **38**, 7142–7154.
23. Bath, A.J., Milsom, S.E., Gormley, N.A. and Halford, S.E. (2002) Many type IIs restriction endonucleases interact with two recognition sites before cleaving DNA. *J. Biol. Chem.*, **277**, 4024–4033.
24. Marshall, J.J.T., Gowers, D.M. and Halford, S.E. (2007) Restriction endonucleases that bridge and excise two recognition sites from DNA. *J. Mol. Biol.*, **367**, 419–431.
25. Szybalski, W., Kim, S.C., Hasan, N. and Podhajski, A.J. (1991) Class-IIS restriction enzymes - a review. *Gene*, **100**, 13–26.
26. Chan, S.H., Stoddard, B.L. and Xu, S.Y. (2011) Natural and engineered nicking endonucleases - from cleavage mechanism to engineering of strand-specificity. *Nucleic Acids Res.*, **39**, 1–18.
27. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. and Aggarwal, A.K. (1997) Structure of the multimodular endonuclease FokI bound to DNA. *Nature*, **388**, 97–100.
28. Heiter, D.F., Lunnen, K.D. and Wilson, G.G. (2005) Site-specific DNA-nicking mutants of the heterodimeric restriction endonuclease R.BbvCI. *J. Mol. Biol.*, **348**, 631–640.
29. Sasnauskas, G., Zakrys, L., Zaremba, M., Cosstick, R., Gaynor, J.W., Halford, S.E. and Siksnys, V. (2010) A novel mechanism for the scission of double-stranded DNA: BfiI cuts both 3'-5' and 5'-3' strands by rotating a single active site. *Nucleic Acids Res.*, **38**, 2399–2410.
30. Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 10570–10575.
31. Catto, L.E., Ganguly, S., Milsom, S.E., Welsh, A.J. and Halford, S.E. (2006) Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res.*, **34**, 1711–1720.
32. Pernstich, C. and Halford, S.E. (2012) Illuminating the reaction pathway of the FokI restriction endonuclease by fluorescence resonance energy transfer. *Nucleic Acids Res.*, **40**, 1203–1213.
33. Wah, D.A., Bitinaite, J., Schildkraut, I. and Aggarwal, A.K. (1998) Structure of FokI has implications for DNA cleavage. *Proc. Natl Acad. Sci. USA*, **95**, 10564–10569.
34. Sanders, K.L., Catto, L.E., Bellamy, S.R. and Halford, S.E. (2009) Targeting individual subunits of the FokI restriction endonuclease to specific DNA strands. *Nucleic Acids Res.*, **37**, 2105–2115.
35. Rusling, D.A., Laurens, N., Pernstich, C., Wuite, G.J.L. and Halford, S.E. (2012) DNA looping by FokI: the impact of synapse geometry on loop topology at varied site orientations. *Nucleic Acids Res.*, **40**, 4977–4987.
36. Gormley, N.A., Hillberg, A.L. and Halford, S.E. (2002) The type IIs restriction endonuclease BspMI is a tetramer that acts concertedly at two copies of an asymmetric DNA sequence. *J. Biol. Chem.*, **277**, 4034–4041.
37. Marshall, J.J.T. and Halford, S.E. (2010) The Type IIB restriction endonucleases. *Biochem. Soc. Trans.*, **38**, 410–416.
38. Kong, H., Morgan, R.D., Maunus, R.E. and Schildkraut, I. (1993) A unique restriction endonuclease, BcgI, from *Bacillus coagulans*. *Nucleic Acids Res.*, **21**, 987–991.
39. Kong, H. (1998) Analyzing the functional organization of a novel restriction modification system, the BcgI system. *J. Mol. Biol.*, **279**, 823–832.
40. Kong, H. and Smith, C.L. (1998) Does BcgI, a unique restriction endonuclease, require two recognition sites for cleavage? *Biol. Chem.*, **379**, 605–609.
41. Marshall, J.J.T., Smith, R.M., Ganguly, S. and Halford, S.E. (2011) Concerted action at eight phosphodiester bonds by the BcgI restriction endonuclease. *Nucleic Acids Res.*, **39**, 7630–7640.
42. Gauci, V.J., Wright, E.P. and Coorsen, J.R. (2011) Quantitative proteomics: assessing the spectrum of in-gel protein detection methods. *J. Chem. Biol.*, **4**, 3–29.
43. Marshall, J.J.T. (2006) BcgI and other Type IIB restriction endonucleases, Ph.D. Thesis. University of Bristol.
44. Cantor, C.R. and Schimmel, P.R. (1980) *Biophysical Chemistry Part II: Techniques for the Study of Biological Structure and Function*. W.H. Freeman, New York, NY.
45. Cesnaviciene, E.E., Petrusyte, M.M., Kazlauskienė, R.R., Maneliene, Z., Timinskas, A., Lubys, A. and Janulaitis, A. (2001) Characterization of AolI, a restriction-modification system of a new type. *J. Mol. Biol.*, **314**, 205–216.
46. Smith, R.M., Jacklin, A.J., Marshall, J.J.T., Sobott, F. and Halford, S.E. (2012) Organisation of the BcgI restriction-modification protein for the transfer of one methyl group to DNA. *Nucleic Acids Res.*, **41**, 405–417.
47. Burgess, R.R. (2009) Protein precipitation techniques. *Methods Enzymol.*, **463**, 331–342.
48. Jacklin, A.J. (2011) Structure and assembly of BcgI, a Type IIB restriction-modification system, Ph.D. Thesis. University of Bristol.
49. Sobott, F., McCammon, M.G., Hernández, H. and Robinson, C.V. (2005) The flight of macromolecular complexes in a mass spectrometer. *Philos. Transact. A: Math. Phys. Eng. Sci.*, **363**, 379–389.
50. Tanford, C. (1961) *Physical Chemistry of Macromolecules*. John Wiley, New York, NY.
51. Heitman, J., Zinder, N.D. and Model, P. (1989) Repair of the *Escherichia coli* chromosome after *in vivo* scission by the EcoRI endonuclease. *Proc. Natl Acad. Sci. USA*, **86**, 2281–2285.
52. Taylor, J.D., Goodall, A.J., Vermote, C.L. and Halford, S.E. (1990) Fidelity of DNA recognition by the EcoRV restriction/modification system *in vivo*. *Biochemistry*, **29**, 10727–10733.
53. Vipond, I.B. and Halford, S.E. (1995) Specific DNA recognition by EcoRV restriction endonuclease induced by calcium ions. *Biochemistry*, **34**, 1113–1119.
54. Zaremba, M., Sasnauskas, G., Urbanke, C. and Siksnys, V. (2005) Conversion of the tetrameric restriction endonuclease Bse634I into a dimer: Oligomeric structure-stability-function correlations. *J. Mol. Biol.*, **348**, 459–478.
55. Bellamy, S.R., Milsom, S.E., Kovacheva, Y.S., Sessions, R.B. and Halford, S.E. (2007) A switch in the mechanism of communication between the two DNA-binding sites in the SfiI restriction endonuclease. *J. Mol. Biol.*, **373**, 1169–1183.
56. Laurens, N., Rusling, D.A., Pernstich, C., Brouwer, I., Halford, S.E. and Wuite, G.J.L. (2012) DNA looping by FokI: the impact of twisting and bending rigidity on protein-induced looping dynamics. *Nucleic Acids Res.*, **40**, 4988–4997.
57. Kneale, G.G. (1994) A symmetrical model for the domain structure of Type I DNA methyltransferases. *J. Mol. Biol.*, **243**, 1–5.
58. Callow, P., Sukhodub, A., Taylor, J.E. and Kneale, G.G. (2007) Shape and subunit organisation of the DNA methyltransferase M.AhdI by small-angle neutron scattering. *J. Mol. Biol.*, **369**, 177–185.
59. Lindahl, T., Demple, B. and Robins, P. (1982) Suicide inactivation of the *E. coli* O6-methylguanine-DNA methyltransferase. *EMBO J.*, **1**, 1359–1363.
60. Walsh, C.T. (1984) Suicide substrates, mechanism-based enzyme inactivators: recent developments. *Annu. Rev. Biochem.*, **53**, 493–535.