Published online 17 February 2009 *Nucleic Acids Research, 2009, Vol. 37, No. 6* **2053–2063**
doi:10.1093/nar/gkp042

Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerization and DNA looping

Kelly J. Neaves[1], Laurie P. Cooper[2], John H. White[2], Stewart M. Carnally[1], David T. F. Dryden[2], J. Michael Edwardson[1] and Robert M. Henderson[1,*]

[1]Department of Pharmacology, University of Cambridge, Tennis Court Road, Cambridge CB2 1PD and [2]School of Chemistry, The King's Buildings, The University of Edinburgh, Edinburgh EH9 3JJ, UK

Received October 29, 2008; Revised January 2, 2009; Accepted January 14, 2009

ABSTRACT

Atomic force microscopy (AFM) allows the study of single protein–DNA interactions such as those observed with the Type I Restriction–Modification systems. The mechanisms employed by these systems are complicated and understanding them has proved problematic. It has been known for years that these enzymes translocate DNA during the restriction reaction, but more recent AFM work suggested that the archetypal EcoKI protein went through an additional dimerization stage before the onset of translocation. The results presented here extend earlier findings confirming the dimerization. Dimerization is particularly common if the DNA molecule contains two EcoKI recognition sites. DNA loops with dimers at their apex form if the DNA is sufficiently long, and also form in the presence of ATPγS, a non-hydrolysable analogue of the ATP required for translocation, indicating that the looping is on the reaction pathway of the enzyme. Visualization of specific DNA loops in the protein–DNA constructs was achieved by improved sample preparation and analysis techniques. The reported dimerization and looping mechanism is unlikely to be exclusive to EcoKI, and offers greater insight into the detailed functioning of this and other higher order assemblies of proteins operating by bringing distant sites on DNA into close proximity via DNA looping.

INTRODUCTION

DNA and DNA–protein constructs have been the subject of frequent investigation by atomic force microscopy (AFM); In fact, since the advent of AFM in the 1980s (1), DNA has been the most studied biological sample (2) and it is now considered to be the reference biosample against which new technological advances are tested (3). There have been numerous AFM investigations involving protein–DNA constructs (4). A few specific examples of DNA-binding proteins that have been investigated by AFM are: the tumour suppressor protein p53 (5,6), the bacterial RNA polymerase (RNAP) (7) and the Type I restriction–modification (R–M) enzymes, EcoKI (8,9) and EcoR124I (10).

Type I R–M systems are the most intricate of the R–M systems (11–14). All their activities are present in one ~440 kDa oligomeric protein comprising one sequence specificity (S) subunit, two modification methyltransferase (M) subunits and two restriction endonuclease (R) subunits. Their operation uses several complex mechanisms, such as interaction between distant sites on DNA, and ATP-dependent DNA translocation. Similar degrees of complexity are also observed in key biological processes, such as DNA replication, repair, recombination and transcription. As such, a more detailed understanding of the Type I R–M systems may provide a better understanding of all such systems. The DNA recognition sequences of the Type I R–M enzymes are asymmetric and comprise two components: one of 3–4 bp and one of 4–5 bp. These components are separated by a non-specific sequence of 6–8 bp, for example, the recognition sequence of EcoKI is $AAC(N)_6GTGC$. Modification occurs via the methylation of an adenine residue, one from each component of the recognition sequence, but on opposite strands of the DNA. This methylation requires the presence of S-adenosylmethionine (AdoMet), which acts as both cofactor and methyl donor. Restriction is preceded by DNA translocation of thousands of base pairs of DNA. The enzyme is believed to pull DNA towards itself, from both sides, whilst remaining bound to the recognition site. Because of this translocation process, the actual DNA cleavage occurs at some distance from the original site and the entire process requires the presence of AdoMet, Mg^{2+} and ATP. Enzyme activity is determined by the

*To whom correspondence should be addressed. Tel: +44 1223 334 053; Fax: +44 1223 334 100; Email: rmh1003@cam.ac.uk

methylation state of the DNA: if the DNA recognition site is fully methylated the enzyme dissociates and no reaction occurs; if the DNA is hemi-methylated, then methylation of the unmodified strand occurs; if the DNA is non-methylated, then the ATP-dependent translocation and restriction pathway is put into action.

Efficient restriction of linear DNA requires two separate recognition sites and it is proposed that restriction is initiated at the collision point when translocation is stalled when the two enzymes meet. If multiple recognition sites are present on the DNA, then cleavage occurs between consecutive sites indicating that translocation occurs in both directions (15–18). However, in contrast to linear DNA substrates, circular substrates with just one recognition site are efficiently cleaved (19–21) and it has been suggested that cleavage simply occurs when the entire substrate has been translocated (16,22), i.e. its closed circular conformation prevents any further translocation and thus causes translocation stalling and cleavage. Whilst biochemical evidence for bidirectional translocation is good, it does predict that two loops of DNA would be produced by each translocating enzyme. However, images produced by electron microscopy [refs (23,24) for EcoKI and ref. (20) for EcoBI] frequently showed only a single loop. It has been argued that the relatively harsh sample preparations required for the production of electron micrographs may not have allowed the preservation of many looped complexes (25) and recent AFM analysis of the EcoR124I Type I enzyme (10) and EcoKI (our unpublished results) show more than one loop being formed.

Most AFM work has been performed on EcoKI. A time-lapse series of images after addition of ATP initially showed a stationary complex surrounded by loops of plasmid DNA. As time progressed, particular loops were seen to contract and were subsequently cleaved to release free ends (9). Further investigations were able to demonstrate that on a linearized plasmid containing two DNA sites, the bound enzyme appeared to dimerize prior to the addition of ATP (8). Such findings appear not to fit the existing model of the restriction reaction, which suggests that the enzymes should only meet at the end of the translocation event (i.e. when their collision promotes cleavage). However, the large lengths of the DNA used meant that that it was not possible to make a conclusive interpretation of the DNA arrangement within the images. Furthermore, the difficulty in interpreting the images may be exacerbated by the sample preparation used to deposit the DNA. The mica surfaces used for both experiments with EcoKI and EcoR124I were pre-treated with poly-L-lysine, a positively charged polymer used to strongly attract negatively charged DNA onto the surface (8–10).

In the research presented here, we have extended the existing AFM studies by using smaller fragments of DNA and by using magnesium ions instead of poly-L-lysine for surface attraction. The results produced support the earlier AFM work (8,9) and also provide evidence for a pre-translocation dimerization model, which involves the formation of DNA loops by a diffusive mechanism as recently found for the Type III R–M enzymes (26–28).

MATERIALS AND METHODS

Calibration

Accurate determination of height and volume is crucial for our experiments so three different AFM scanners were calibrated using gold nanoparticles and proteins of known molecular weight. All three scanners gave identical results (within their margins of error) when using the same mica source, probes and imaging parameters. We first tested gold nanoparticles (size standards from electron microscopy with a diameter of 4.9 ± 0.4 nm). When immobilized upon a polylysine coated mica surface, a height of 4.4 ± 0.9 nm was obtained, comparable to the diameter determined for these particles by EM (it is possible that the particles are embedded in the polylysine layer which will be ~0.4 nm thick). Our AFM's lateral accuracy was confirmed through an extensive calibration procedure using the protocol provided by the manufacturers (Multimode SPM Instruction Manual Version 5, revision B, Veeco Instruments Inc.) Having confirmed the accuracy and precision of our AFM scanners in measuring heights and widths, we were surprised to find that the calculated volumes of our proteins were ~1/3 of their previously determined values (8,9,29). Using a range of proteins of known molecular weight, we compared their expected molecular volumes with their measured volumes as shown in Figure 1. The results showed that the underestimation in volume was a consistent trend and thus we concluded that the discrepancy must be due to AFM method and not the samples.

Our current probes are made of Si and have a higher aspect ratio and a much greater spring constant than the SiN tips previously used. The higher aspect ratio generates superior lateral resolution (less broadening due to tip-convolution), but the greater spring constant means that the probes will 'squash' the biological samples to a greater extent than in earlier experiments (resulting in a reduction in the observed height). The standard calculation used to estimate the volume of proteins from AFM images was derived using the old SiN probes (29). In this calculation the protein was assumed to be a spherical cap and the volume is calculated accordingly. In deriving the calculation the authors overcame the large tip-convolution effect by taking the width measurement at half-height as opposed to at the base of the protein. However we have found that, with the new Si tips, and using the basal width measurement we can achieve an accurate estimation of the protein volume. The finite diameter of the new probes means they will inevitably display some lateral broadening though less than with previous probes, which means that the basal width measurement should still be an over-estimation resulting in high protein volumes. However, it is likely that by using an over-estimated basal width in our equation we are also compensating for the reduction in the vertical dimension due to the higher spring constant of the probes ('squashing').

DNA substrates

The DNA substrate used in this study was pBRsK15; a derivative of the commercially available plasmid pBR322.

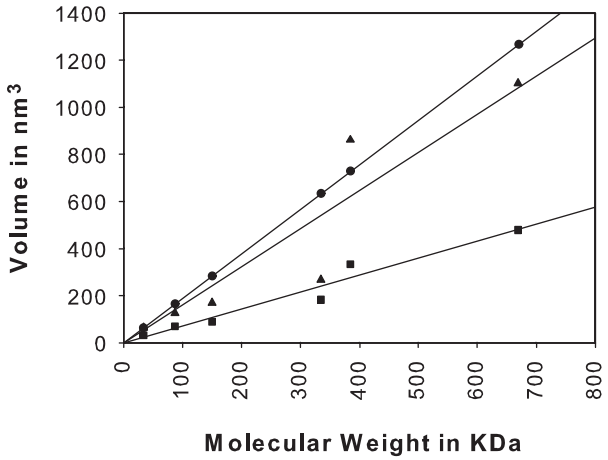

Figure 1. A plot of molecular volume calculated using two different methods from AFM measurements versus molecular weight for a series of proteins: Endophilin (33 kDa); Lactoferrin (87 kDa); IgG (150 kDa); Thyroglobulin (335 kDa) *monomeric*; IgA (385 kDa) *including J-chain and secretory component*; Thyroglobulin (670 kDa) *dimeric*. Circles on the upper line represent the data generated from the expected volumes for each protein. The square symbols (and the bottom line) show the volumes determined from the AFM data using the previously published method (29). The triangular symbols (and the middle line) show the volumes determined using the basal width of the imaged molecules as described in 'Materials and methods' section.

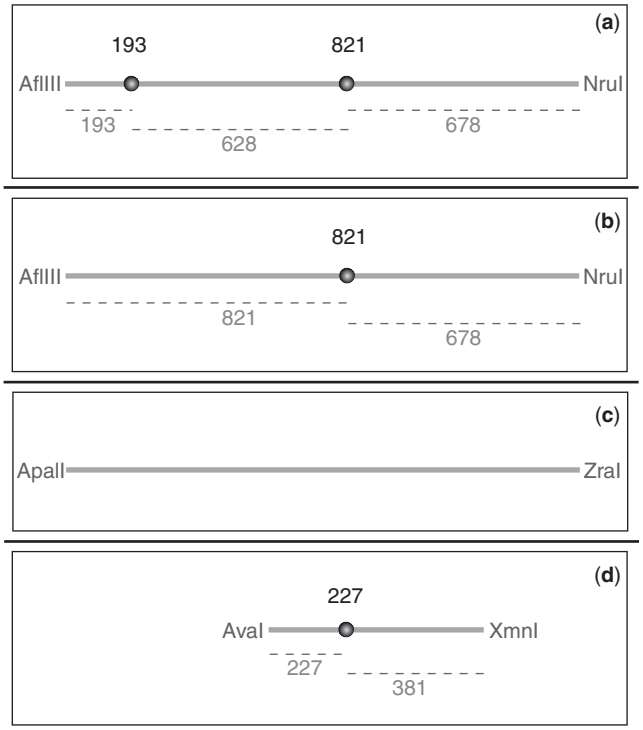

Figure 2. Diagrammatic representation of **(a)** the 1499-bp double-sited DNA substrate **(b)** the 1499-bp single-sited DNA substrate **(c)** the 1499-bp zero-sited fragment and **(d)** the 608-bp single-sited fragment. **(a)** and **(d)** were constructed from the mutated pBRsK15 and **(b)** and **(c)** were constructed from pBRsK15. The restriction enzymes used to produce the individual fragments are written at the end of each fragment. In each diagram the DNA is represented by a line and EcoKI-binding sites are represented by spheres. The centre of the each sphere, and the figure above each sphere, corresponds to the central base pair in the EcoKI-binding sequence. The numbers below each DNA fragment show the distance between binding sites and/or the distance from each binding site to the end of the DNA molecule. The width of each sphere is a proportional representation of the DNA footprint of EcoKI. However, in AFM imaging the widths would appear to be much greater due to tip convolution.

The pBRsK15 plasmid was previously produced by manipulating pBR322 to give a final product of 4.36 kb, containing a single EcoKI site at position 1652 (30). In this study the pBRsK15 plasmid was subjected to further mutagenesis to introduce a second EcoKI site at position 2280 (628 bp from the original site). The original sequence at position 2280, GTA$(N)_6$GTGC, was mutated into the EcoKI recognition sequence, AAC$(N)_6$GTGC using the QuickChange Site-Directed Mutagenesis Kit, from 'Stratagene' (CA, USA).

Two separate linear products were produced using the mutated pBRsK15 plasmid and a further two products were created from the original non-mutated pBRsK15 plasmid. Figure 2 shows a diagrammatic representation of all the linear DNA products.

Protein substrates

Samples of purified EcoKI and methyltransferase were prepared and assayed as described in ref. (31).

AFM sample preparation

Mica discs were prepared for the deposition of DNA and proteins. Disks of ruby muscovite mica (Goodfellow, PA, USA) were attached to 13 mm steel pucks using Aron Alpha high-strength rapid bonding adhesive, alpha cyanocrylate (Agar Scientific Limited, Essex, UK). Immediately prior to protein or DNA deposition, the top layer of the mica was cleaved using Scotch tape to reveal an atomically flat surface.

Protein samples were diluted to their final concentration in 1× buffer solution (10 mM $MgCl_2$, 10 mM KCl, 10 mM Tris–HCl pH 7.9) in the presence of 100 μM *S*-adenosyl-L-methionine (AdoMet, New England Biolabs, MA, USA). Droplets of 20 μl were deposited onto freshly cleaved mica discs and incubated at room temperature for 2 min. After incubation, excess protein was rinsed off with MilliQ water (Millipore System, MA, USA), excess water was wicked from the surface using tissue paper and the discs were dried under a stream of nitrogen gas.

DNA samples were diluted to a final concentration of ~0.75 nM in 1× buffer solution. Droplets of 20 μl were deposited onto freshly cleaved mica discs and incubated at room temperature for 2 min. After incubation the discs were washed and dried as described above.

DNA and purified EcoKI were mixed together at a 2:1 protein to DNA ratio. The reactions were carried out in 1× buffer solution in the presence of 100 μM AdoMet. Samples were incubated for 10 min at 37°C and subsequently diluted to a final DNA concentration of ~0.75 nM. Droplets of 20 μl were immediately deposited onto freshly cleaved mica discs and incubated at room temperature for 2 min. After incubation the discs were washed and dried as described above.

AFM imaging

Imaging was performed with a Multimode Nanoscope IIIa atomic force microscope (Veeco Digital Instruments, Santa Barbara, CA, USA). All samples were imaged in air using tapping mode with root-mean-square amplitude of ~2 V and drive frequency of ~350 kHz. Commercially available silicon probes with a specified spring constant of 42 N m^{-1} were employed (Olympus atomic force OMCL-AC160TS-E).

Data analysis

The length of DNA in a loop was measured using SPIP software (Scanning Probe Image Processor Software—Image Metrology, Denmark). The length in nanometres was measured by drawing a series of very short lines along the contour of the DNA, followed by summation of the line lengths and converted to base pairs by dividing the nanometre length by 0.29 (this conversion factor was derived from the assumption that the DNA was partially hydrated—as expected during AFM imaging in air). This conversion method was first applied to DNA of a known number of base pairs to ensure that it produced accurate results.

Loop angle measurements were also carried out using the SPIP software. The centre point for each angle measurement was the centre of the protein molecule, and the angle measurement was taken between the two points where the DNA exited the protein.

Volume measurements were calculated using an adjusted version of the equation used in the previous AFM investigations on EcoKI (8) (as described above). The height (h) and basal radius (r) of each protein molecule were measured and used to calculate the volume with the following equation (29): Volume = $(\pi h/6)$ $(3r^2 + h^2)$. For individual protein complexes associated with DNA, heights and radii were measured independently using the SPIP software. For the pure EcoKI and methyltransferase molecules an automated measurement was performed using MATLAB software (The MathWorks Inc., MA, USA). The MATLAB automated measurement was checked by hand and a range of random EcoKI and methyltransferase molecules were analysed with both methods to ensure consistency.

Construction of histograms and the fitting of distribution curves with Gaussian functions were both performed in Sigma Plot 10 (Systat Software Inc., CA, USA). Errors quoted are standard errors.

RESULTS

Imaging of pure DNA and protein

The double stranded linear DNA constructs used in this study were 1499-bp fragments; one containing two EcoKI sites, one containing one EcoKI site, and a final one containing zero EcoKI sites. An additional single-sited control fragment of 608 bp was also used. The sites on the double-sited fragment were separated by 628 bp. This choice of site separation was based on evidence that to bring two sites on a DNA molecule together, the ideal separation should be around 550 bp (32). Both the double-sited and single-sited DNA fragments contained asymmetrical tails either side of the EcoKI sites (Figure 2). This asymmetrical tail design makes it easier to determine whether the protein was specifically bound to a target site. Each purified DNA fragment was adsorbed onto mica and imaged. The resultant images (Supplementary Data 1) demonstrated a uniform distribution of linear fragments with about 75% found to be of the expected size. It was also evident that at the concentrations used, the individual DNA molecules were sufficiently dispersed to avoid too many inter-molecular interactions and stiff enough to reduce the occurrence of intra-molecular looping.

Using standard gel-electrophoresis techniques the EcoKI sample was tested for its activity on the double-sited plasmid DNA and the double-sited 1499-bp linear fragment. The experiments revealed effective digestion, thus demonstrating that the enzyme was active on these DNA molecules under the conditions used (results not shown). Samples of the purified EcoKI ($R_2M_2S_1$) and the EcoKI methyltransferase core (M_2S_1) were each separately adsorbed onto mica and imaged. The resultant images showed distributions of globular particles (Supplementary Data 1) as previously shown (8). Analysis of the particle heights and widths were carried out on both samples, and the volumes were calculated and plotted in frequency distribution histograms (Supplementary Data 2). The arrangement of peaks was comparable to the previous experiments (8) as expected. The peaks were used to characterize the estimated volumes for all the potential complexes, as follows: EcoKI ($R_2M_2S_1$) = 646.8 nm^3; M_1S_1 = 170.6 nm^3 – 185.4 nm^3; fully dissociated M and S subunits = 70.5 nm^2. The complete M_2S_1 would not be expected as it has been shown to be relatively unstable under the conditions used (31). Fully dissociated M, S and R subunits were not found in the EcoKI sample due to the stability of the EcoKI complex (31). The volumes obtained were comparable to both the expected molecular volumes and those of previous experiments (8), and out of 257 molecules of EcoKI observed at least 94% were monomers ($R_2M_2S_1$).

Statistics for EcoKI–DNA interactions observed by AFM

Prior to the detailed analysis of the observed complexes, it is worth considering the gross statistics of EcoKI binding in our experiments (Table 1). The EcoKI bound to non-specific sites on our DNA molecules act as internal controls for our analyses of site-specifically bound EcoKI, and the zero-sited fragment acts as an external control. As previously found (8), we observed both EcoKI monomers and EcoKI dimers on our double-sited DNA (8). We have now also observed EcoKI dimers on single-sited DNA and that many of the EcoKI dimers on both single and double-sited DNA were at the base of DNA loops. The EcoKI molecules were found either at a target sequence or a non-specific region of DNA (often at an end of the DNA molecule), and the loops were found to be formed by a dimer of EcoKI bringing together either two target site sequences or one target site with a non-specific

Table 1. Overall statistics for the main DNA–protein constructs analysed

DNA	Number of DNA molecules captured	Average number of DNA molecules per scan	Average number of bound protein molecules per DNA molecule	Percentage of DNA molecules showing looping	Average size of protein at the loop base	Percentage of bound proteins showing dimerization
Double-sited 1499-bp fragment	622	14.5	1.5	16.7	$1094 \pm 48\,nm^3$	64.7
Single-sited 1499-bp fragment	339	13.8	1.2	9.5	$1103 \pm 47\,nm^3$	34.3
Single-sited 608-bp fragment	476	52.9	0.5	0.8	ND	22.7
Zero-sited 1499-bp fragment	109	12.1	0.5	0.9	ND	3.6
Double-sited 1499-bp fragment with ATPγS	270	18	1.1	15.6	$1147 \pm 60\,nm^3$	54.5

ND, not determined. The first four data columns contain statistics generated from the captured data sets obtained by direct counting from the AFM images. The final two columns of data are taken from the detailed analysis of protein volume measurements performed on a random selection of the data.

region of DNA (the complete analysis of these results is shown below).

All sequence-specific DNA-binding proteins such as EcoKI can also bind to other DNA sequences although often with lower affinity. The closer the DNA sequence to the cognate target, the better the binding affinity. The non-specific DNA is in far greater concentration than the concentration of the specific target site when using long pieces of DNA. For example, for our double-sited sample: the target site concentration is 1.5 nM (2×0.75 nM); there are two non-specific ends at ~0.75 nM each (approximate since the enzyme may project beyond the ends and have a partially filled DNA-binding site); and 1.1 mM (i.e. 1497 bp $\times$ 0.75 nM) of non-specific sites as one steps along the DNA substrate one base pair at a time.

In the absence of DNA, EcoKI did not appear to dimerize (257 molecules, Supplementary Data 1) so the association constant for dimerization must be small in this situation and greatly increased when EcoKI binds to DNA. It would appear that a single EcoKI binds to DNA and undergoes some sort of conformational change so that it can recruit a second EcoKI molecule. Once bound to DNA, the proportion of site-specifically bound proteins was considerable given that the sites were greatly outnumbered by non-specific sequences. This indicates that on the DNA substrates used, the association constant for site-specific binding is much greater than the association constant for non-specific binding. Lastly, although the concentration of DNA ends was much lower than the concentration of non-specific sites elsewhere in the DNA molecule, many EcoKI molecules were observed bound non-specifically on DNA ends. This indicates that EcoKI has a higher association constant for DNA ends than for continuous stretches of non-specific DNA. This suggests that EcoKI prefers to have a partially filled DNA-binding site rather than one fully occupied by non-specific DNA. This may reflect a balance between the free energy cost of bending DNA and the free energy cost of having a partially filled DNA-binding groove (33).

EcoKI binding with loop formation on double-sited DNA

The binding of EcoKI to the 1499-bp double-sited linear DNA fragment was carried out with a ratio of 2:1 protein to DNA to give a stoichiometric 1:1 ratio of protein to target site. The reactions were carried out in the presence of buffer containing AdoMet and magnesium ions and in the absence of ATP to halt the reaction prior to translocation. It was found that any higher concentration of protein, with respect to the DNA, caused too much non-specific binding for any clear images [non-specific binding of EcoKI was previously reported in ref. (34)]. AFM imaging revealed the presence of looped DNA structures (Figure 3a–d). A total of 622 protein–DNA complexes were examined from 43 AFM scans acquired from four separate, freshly prepared samples. 104 complexes showed a DNA loop with protein bound at the base of the loop (16.7%), 247 showed protein bound but no DNA loop (39.7%) and there were 271 DNA molecules with no protein bound (43.6%). The value of 16.7% looping is significant when compared to the pure sample of DNA which contained no DNA looping. At the apex of each looped structure there was a large spherical particle suggesting that the looping was directly caused by the EcoKI molecules. The height and width of each particle at the apex of each loop, was measured and the volume calculated. A Gaussian distribution, fitted to a histogram of the estimated volumes, had a maximum at $1094\,nm^3$ ($\pm 47\,nm^3$). In agreement with EcoKI dimerization in previous investigations (8), the volume of EcoKI on these looped DNA structures was, on average, 1.7 times larger than the pure enzyme ($646.8\,nm^3$).

The length of DNA within each loop was measured and the results revealed a Gaussian distribution with an average loop size of 617.5 bp (± 26), a value close to the 628-bp separation between the two specific sites (Figure 3e). Although the distribution had a clear modal peak, close inspection of the histogram and AFM images, showed that there was a range of structures that differed significantly from the average. This suggested that there was a secondary population of looped structures not caused by the joining of the two specific sites. The absence of a bimodal distribution histogram suggested that this secondary population was broadly spread around the main distribution peak. The tail lengths of each looped construct were measured and this in combination with the loop length was used to classify the constructs into those caused by the joining of two specific sites (a double-specific loop) and those that did not join the two specific sites (a random loop). There were approximately equal

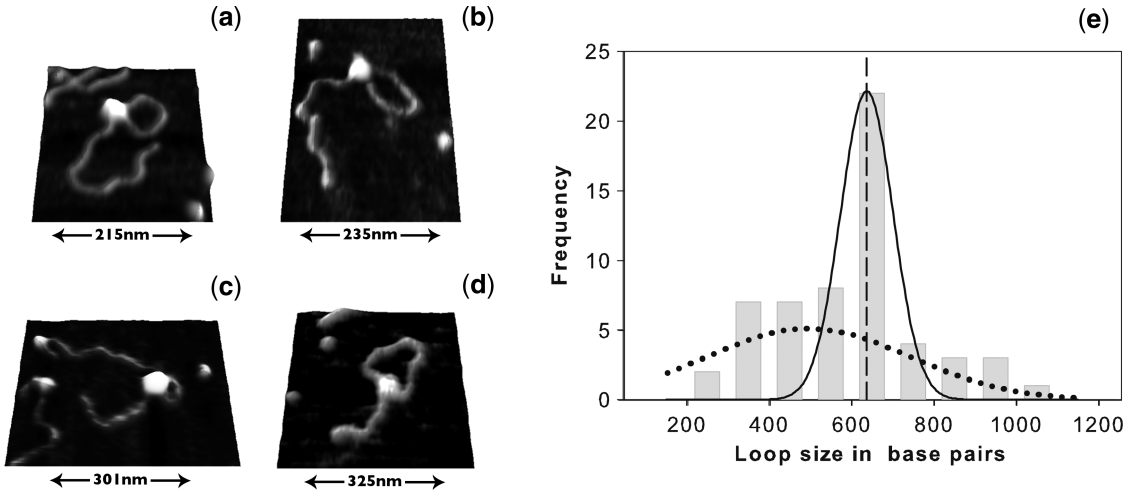

Figure 3. AFM images of EcoKI-DNA looped constructs and distribution of loop sizes on double sited DNA. Parts **(a)** and **(b)** show the EcoKI joining two specific DNA sites, and **(c)** and **(d)** show the EcoKI joining a specific site with a non-specific region of DNA. Part **(e)** show the distribution of loop size measurements for the EcoKI–DNA looped constructs; the bars show the complete distribution and the line shows the peak of this main distribution, whereas the super-imposed dotted curve shows a sub-population distribution comprising of only the loops deemed to be caused by at least one non-specific contact with the DNA (as detailed in the text).

numbers of each type of loop. In this second group, 90% of the complexes contained one tail corresponding to specific binding (i.e. even though one of the protein/DNA contacts was to a non-specific region of DNA, the protein was still making another DNA contact via a specific EcoKI-binding site). Volume measurements of the protein at the apex of both types of loops gave values in close agreement to that produced by the entire data set. This would suggest that dimerization is independent of the loops joining two specific sites. Similarly, measurements of the angle at the apex of each loop for the two types of loop were identical with an estimated average angle of 77.8° (±4°). The consistency in angle could represent the dimerized EcoKI imposing a specific angle on the DNA (i.e. the binding angle between the protein monomers could directly affect the exit angle of the DNA) or it may simply be due to the constraints of DNA looping or forces involved in the binding of the complexes to the mica surface. Figure 3e shows a frequency distribution curve for the random-loops superimposed on the frequency distribution for the entire data set of double-specific and random loops. The random-loop distribution is broad with a maximum of ~550 bp consistent with the flexible properties of the DNA imposing a size distribution on randomly formed loops [i.e. the worm-like chain model for DNA flexibility; (32,35)], whereas the distribution for double-specific loops is narrow and fixed by the distance between the two EcoKI-binding sites.

EcoKI binding without loop formation on double-sited DNA

From the original data set, a random selection of non-looped DNA molecules with clearly bound EcoKI were selected (Figure 4a–d). Of the 247 molecules analysed, 40% showed EcoKI binding and each DNA molecule analysed contained on average 1.5 bound protein molecules. The volume distribution showed a bimodal distribution with a main peak corresponding to monomers and a second peak corresponding to dimers (Figure 4e). However, when just the data corresponding to the site-specifically bound protein were plotted (superimposed on the histogram in Figure 4e) it is clear that the peak corresponding to the dimers is almost entirely represented by site-specifically bound protein. Interestingly, amongst all the double-sited non-looped DNA molecules analysed, only 6% contained enzymes bound at both specific sites.

Analysis of entire data set for EcoKI on double-sited DNA

For the double-sited fragment the entire looped and non-looped data sets were combined and each protein molecule was categorized according to whether it was a dimer or a monomer and whether it was specifically or non-specifically bound. On looped DNA the data were classed as specifically bound when either two specific contacts were made (double-specific), or when one specific and one non-specific contact was made. The results are shown as a bar chart in Figure 4f. It is clear that specifically bound dimers and non-specifically bound monomers dominate the distribution.

EcoKI binding on single-sited DNA

The binding of EcoKI to the 1499-bp single-sited linear DNA fragment was carried out with a 2:1 ratio of protein to DNA in a buffer containing AdoMet and magnesium ions but in the absence of ATP to prevent translocation. 317 DNA molecules were captured from 22 AFM scans (Figure 5). Each DNA molecule contained an average of 1.2 bound protein molecules and there were 30 looped molecules with a protein at the base of the loop (9.5%). The average volume of the protein molecule at the base of each loop was 1103 nm^3 (±47 nm^3) consistent with each complex being formed by a protein dimer. The loops formed on single-sited DNA must be random-loops (as opposed to the double specific loops seen on the double-sited DNA). The same experiment was carried out on the

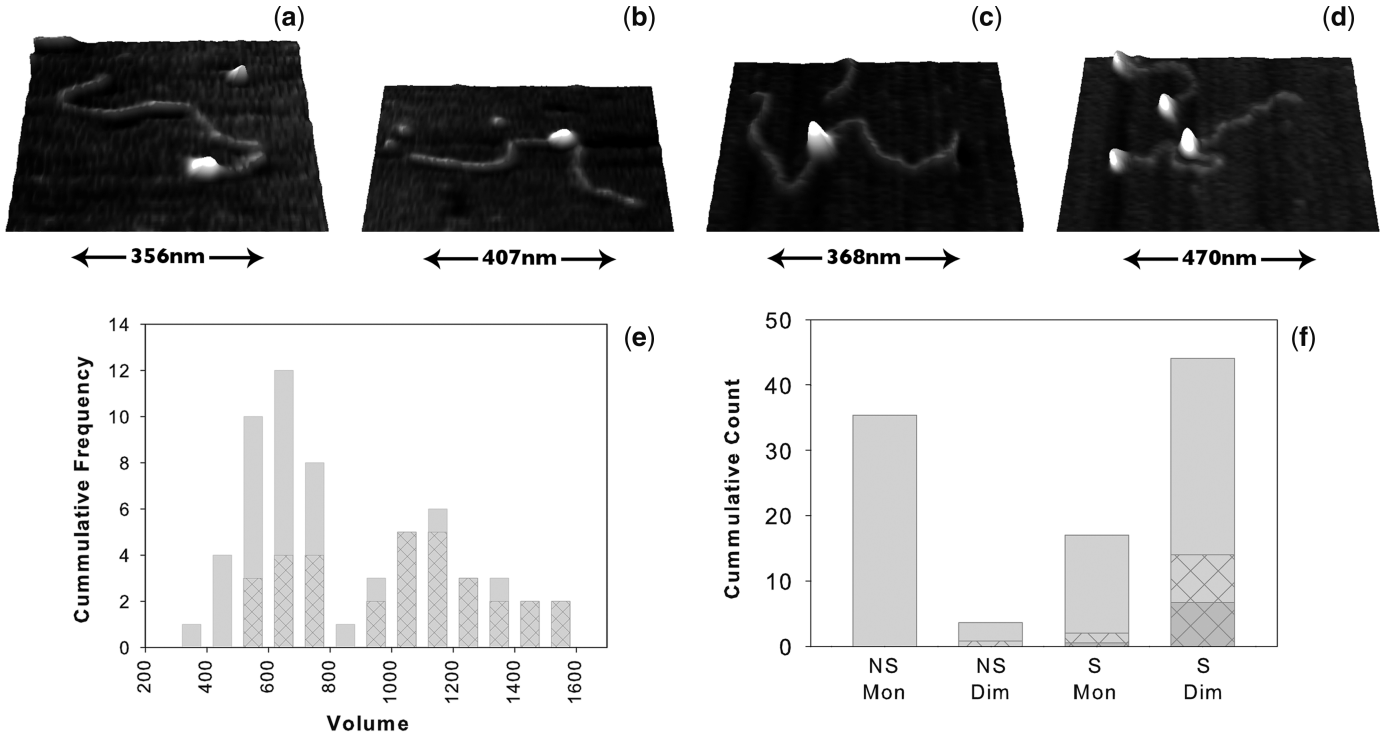

Figure 4. AFM images and protein volume distribution for EcoKI on non-looped double-sited DNA. Part **(a)** shows a non-specific EcoKI monomer, **(b)** shows a specific EcoKI monomer, **(c)** shows a specific EcoKI dimer and **(d)** shows one DNA molecule containing both a specific EcoKI dimer and a non-specific EcoKI monomer. Part **(e)** shows the distribution of protein volume measurements on the non-looped double-sited DNA, where the complete bars represent the entire data set and the hashed areas represent proteins bound only to a specific site. Part **(f)** shows a bar chart of the proportions of all the different protein categories in the combined looped and non-looped double-sited data sets. The abbreviated labels stand for the following: NS Mon = non-specific monomers; NS Dim = non-specific dimers; S Mon = specific monomers; S Dim = specific dimers. The hashed areas represent those protein molecules in looped constructs, and within that population the darkened areas represent those joining two specific sites (double-specific-loops), and the lighter areas represent those joining a specific site with a non-specific region of DNA (random-loop).

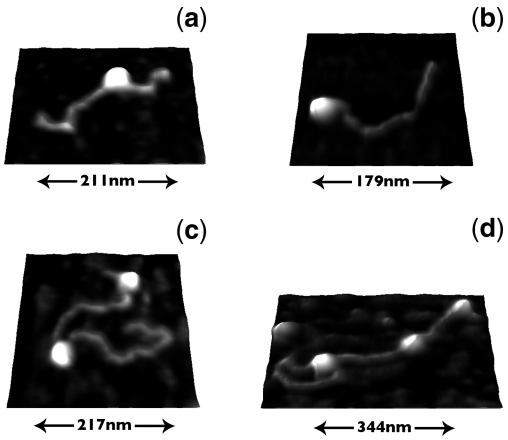

Figure 5. AFM images of EcoKI on single-sited DNA. Part **(a)** shows a specific EcoKI dimer on the 608-bp fragment, **(b)** shows a non-specifically bound EcoKI monomer on the 608-bp fragment, **(c)** shows one specific EcoKI dimer and one non-specific monomer on the 1499-bp fragment and **(d)** shows a specific EcoKI dimer forming a loop on a 1499-bp fragment which also contains two non-specific EcoKI monomers.

608-bp single-sited DNA fragment and negligible loop formation was observed. This was not unexpected as there are only 227 and 381 bp on either side of the target site, equivalent to ~1.5 and ~2.5 persistence lengths in the worm-like chain model (32,35), and the probability of loop formation will be very low.

Detailed analysis of the size and position of EcoKI molecules on the DNA was carried out on a random selection of molecules from the 608- and 1499-bp single-sited data sets and both sets produced comparable results. As with the double-sited DNA the distribution of volume size was bimodal with one peak corresponding to EcoKI monomers and one peak corresponding to specifically bound EcoKI dimers (Figure 6a–b). For both sets each protein molecule was given one of the following classifications: non-specifically bound monomer, non-specifically bound dimer, specifically bound monomer, or specifically bound dimer. In looped complexes, most were formed by joining a specific EcoKI site to a random site and were classed as specific monomers (rare) or specific dimers (the majority). A few loops joined to random locations on the DNA and were classed as non-specific monomers or dimers. From the bar-charts using these classifications (Figure 6c–d), it was clear that a large proportion of the data was made up of non-specific monomers (67% for the long fragment and 73% for the short fragment). The remainder comprised specifically bound dimers (21% for the long fragment and 20% for the short fragment), non-specifically bound dimers (5% on long fragment and 3% on the short fragment) and specifically bound monomers

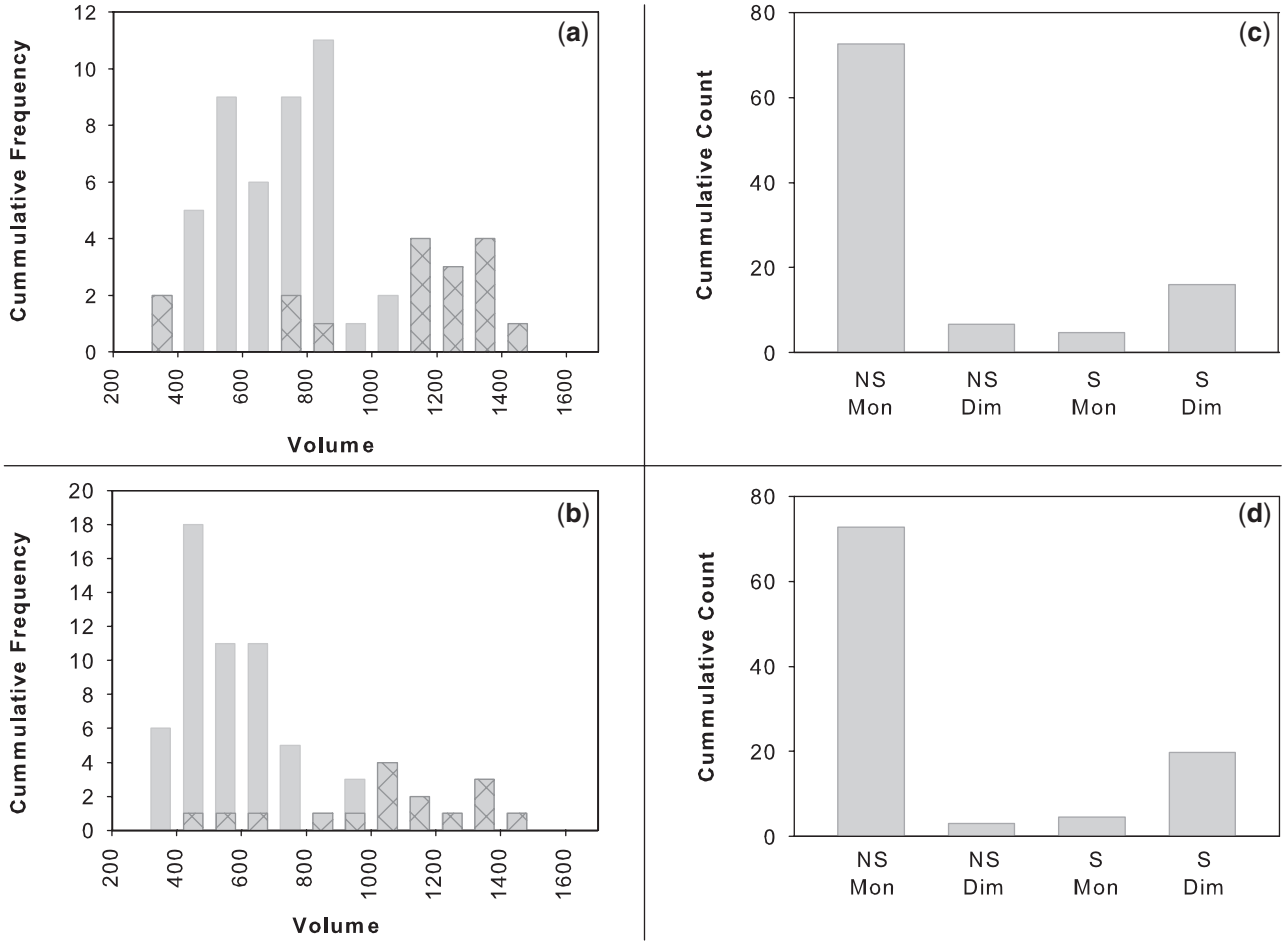

Figure 6. Analysis of EcoKI bound to the single-sited DNA fragments. Parts **(a)** and **(b)** show the distribution of protein volume measurements for the 1499-bp and 608-bp fragments, respectively. The complete bars represent the entire data set and the hashed areas represent proteins bound only to a specific site. Parts **(c)** and **(d)** demonstrate the relative proportions of all different protein categories in the entire analysed single-sited data sets for the 1499-bp and 608-bp fragments respectively. The abbreviated labels stand for the following: NS Mon = non-specific monomers; NS Dim = non-specific dimers; S Mon = specific monomers; S Dim = specific dimers. Note that on the looped constructs, those proteins contacting one specific site and one non-specific site were still classified as 'specific' (as detailed in the text).

(7% on the long fragment and 5% on the short fragment). The occurrence of dimerization on single-sited DNA suggests that dimerization occurs before specific sites are brought together. The relative proportions of all the data suggests that finding the initial specific site is a rate limiting factor. It also suggests that whilst dimerization is rare at non-specific sites, once specific binding occurs, dimerization is relatively straightforward. A further observation was that of the EcoKI molecules bound non-specifically, roughly 50% were bound at the end of a DNA molecule. This is as expected as the EcoKI molecule bends DNA, an energetically costly process, and this energy will not be required if the enzyme sits right on the end of a DNA molecule (33). Comparing Figure 6c and d with Figure 4f, one can see that the presence of two EcoKI sites greatly increases the proportion of EcoKI dimers bound at the EcoKI target site. The proportion of loops was also slightly greater at 16.7% presumably due to the possibility of being able to form stable loops linking the two target sites in addition to loops linking a single target site with some other random sequence. The presence of two types of loops would suggest that the formation of a loop joining two specific sites potentially arises from an intermediate with contacts to one specific and one non-specific site. Once this intermediate has been formed, the high binding affinity of EcoKI for the target sequence enhances the number of complexes with loops joining two specific target sequences.

EcoKI binding on DNA with no target sites

On the zero-sited 1499-bp DNA fragment there was insignificant looping or dimerization (1/109 and 1/28, respectively). This suggests that the presence of a specific binding site is a direct requirement for both loop formation and dimerization.

The M_2S_1 methyltransferase core binding to double-sited DNA

As a control, the EcoKI experiments on double-sited DNA were repeated using the M_2S_1 methyltransferase core. Although a comparatively small amount of

dimerization (when compared with the complete EcoKI enzyme) was observed, no DNA looping was present (Supplementary Data 3). The observation of M_2S_1 as a negative control for looping emphasises that any looping observed with the complete EcoKI complex is genuine, and not due to any artificial limitations in technique or procedure. The observation of some dimerization could suggest that the dimerization domain is present on the methyltransferase core, but the relatively low abundance of dimer complexes and the complete lack of looping would suggest that the whole $R_2M_2S_1$ complex is essential for the entire dimerization/looping phenomenon observed in this report.

EcoKI looping on double-sited DNA in the presence of ATPγS

The experiments on the double-sited DNA fragment were repeated in the presence of 1 mM ATPγS; an analogue of ATP that is hydrolysed at a dramatically slower rate than ATP itself, effectively trapping the structure in the initial ATP-bound state. These experiments showed a significant presence of looped constructs (15.6% over 270 molecules), suggesting that the results presented above are relevant to the complete enzymatic pathway.

DISCUSSION

Our previous observations of DNA looping and enzyme dimerization by EcoKI (8,9) were complicated by the long DNA molecules employed as they could cross over themselves multiple times. By using gentler preparation techniques, shorter DNA fragments and by examining the DNA loops in greater detail than before, the research presented here supports and extends our previous work providing further insight into the behaviour of EcoKI prior to the onset of translocation. In summary, our experiments indicate the following:

- There was no dimerization of EcoKI without DNA indicating that dimerization does not occur before DNA binding.
- There was significant dimerization at EcoKI target sites on unlooped DNA molecules indicating that dimerization takes place on specific sites before any looping.
- Looped DNA molecules with two EcoKI target sites predominantly had an EcoKI dimer at the base of the loop regardless of whether the loop joined two EcoKI target sites or one EcoKI target site with a random site on the DNA indicating that dimerization occurs before the two specific sites are bought together.
- There was significant looping on DNA molecules containing a single EcoKI target site if the molecule was long enough, indicating that looping initially occurs between a target site and a non-specific region of DNA.

Our results suggest the following mechanism for EcoKI. First, an EcoKI monomer binds to DNA at a random sequence and then, in common with other DNA-binding proteins, it may either dissociate or conduct a limited degree of linear diffusion up and down the DNA contour (35). During this diffusion process the enzyme may find its target sequence. A second monomer can bind to the EcoKI already on the DNA to form a dimer and this dimer is more stable if it forms at a target sequence. A dimer bound at the target site then collides with another region of DNA due to the inherent flexibility of the DNA to form a loop. This second location on the DNA can be either a second specificity site or, more probably, a non-specific DNA region. Once the loop is formed, the second EcoKI molecule can scan the DNA contour for another copy of the target sequence. The loop may fall apart during this process so multiple rounds of loop formation may be required before the second target sequence is found. Despite the possibility of repetitive rounds of loop formation, the process is still an efficient means for locating the second site (35). Once both target sequences have been located by the dimer of EcoKI, ATPase-driven translocation and DNA cleavage can be initiated. This dimerization of EcoKI prior to translocation can be incorporated into the existing model for EcoKI translocation and restriction, as shown in Figure 7.

The steps observed in our experiments, namely, initial non-specific binding, specific site location, dimerization and looping may reflect the processes observed biochemically many years ago by Yuan *et al.* (36). They observed an enzyme activation process induced by AdoMet, an 'initial DNA-binding complex' and then a conversion into a site-specific binding 'recognition complex'. Since AdoMet is always present in our reactions, it is possible that the 'initial complex' attributed to non-specific DNA binding could instead have been an EcoKI monomer bound to the DNA target site and the 'recognition complex' could be an EcoKI dimer bound to the target site (with or without looping). To support this, we note that the 'recognition complex' observed by Yuan *et al.* (36) was associated with the large protein form bound to DNA observed by electron microscopy (24). ATP addition led to observation of a small protein form bound to the DNA and this size change was attributed to a loss of subunit from the EcoKI enzyme. Given the sizes of the large and small protein forms, 16–20 nm width and 10–13 nm width respectively, one might equally well assign the large form to be an EcoKI dimer and the small form to be a single EcoKI bound to DNA. We note that if the EcoKI monomer was spherical, it would have a diameter of 11.2 nm. Placing a second monomer alongside the first to form a dimer could account for the size of the large form in a straightforward manner. Thus our data would appear to be consistent with previous data and models but with the addition of a new factor, namely the formation of loops prior to ATP-induced translocation.

The observation of loop formation by the Type I R–M enzyme, EcoKI, is not without precedent as many other DNA-binding proteins such as transcription factors (37), transposases (38), resolvases (39), and many Type II (40) and Type III restriction enzymes (26–28) also feature looping to bring two widely separated DNA segments into close proximity. It is clear the enzyme dimerization and the loop formation by diffusion are not essential prerequisites for ATPase-driven translocation of DNA

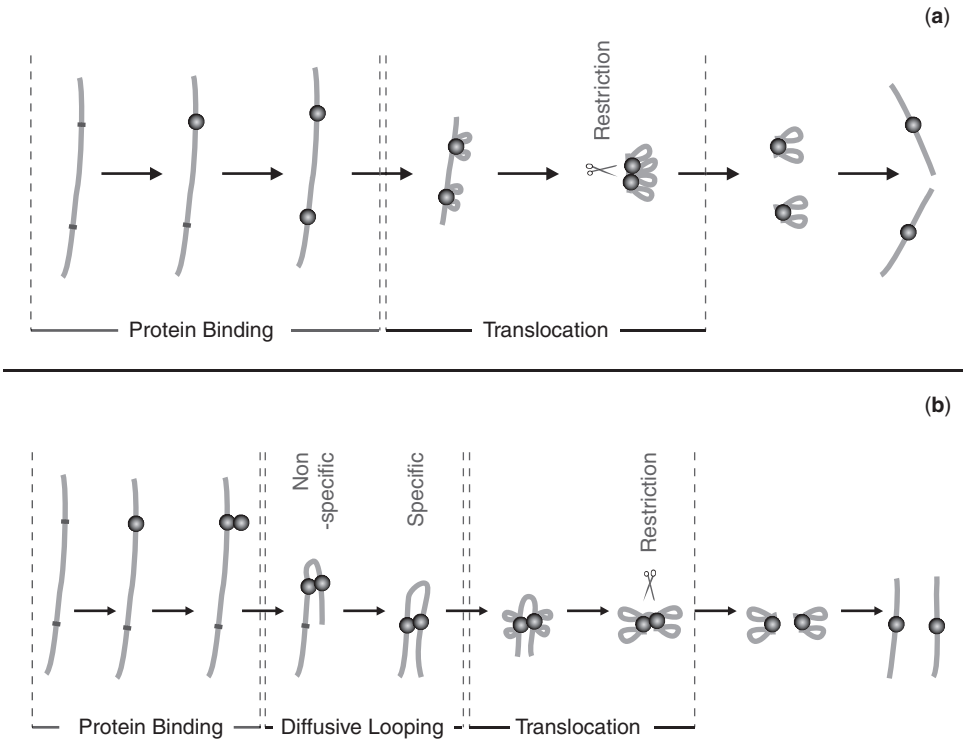

Figure 7. Diagrammatic representations of EcoKI binding, translocation and restriction on DNA. Part **(a)** shows the previous model. In this model EcoKI monomers individually bind to each site and translocation occurs independently from both occupied sites. When the monomers meet, translocation is stalled and restriction occurs. Part **(b)** shows a new model, which is based on the existing model but incorporates the additional information discovered in this paper and in (8,23). In this new model a monomer of EcoKI binds to one site and then a second monomer binds to the same site to form a dimerized complex at that site. This dimerized complex then forms diffusive loops with non-specific regions of DNA until it is stabilized by contact with the secondary EcoKI site. Translocation then occurs from both sides of both monomers and, in agreement with the previous model, restriction occurs when the translocation process is stalled (this time because the diffusive loop between the monomers becomes fully contracted). In the evidence presented here it remains unanswered whether translocation is triggered by the occupation of a secondary site or whether both processes occur concurrently. In both models DNA is represented as a line, specific EcoKI sites are represented by dots on the DNA molecules, and EcoKI monomers are represented as individual spherical objects.

by EcoKI. This raises the question of exactly why EcoKI appears to undergo these two processes when it is clear that EcoKI monomers can hydrolyse ATP and alter their conformation presumably in an attempt to translocate DNA when presented with short DNA duplexes containing a single specificity site (34). There is as yet no obvious answer to this question. It might be suggested that dimerization between two unmodified sites on the host chromosome would be very improbable, and thus the proposed mechanism could be an added safeguard to protect the host DNA from self damage after cellular stress (41). In stress situations, EcoKI can attack unmodified sites occurring on the chromosome. This lethal process is normally kept in check by the ClpXP protease which removes the R subunits. However, an additional requirement for EcoKI dimerization would further reduce the possibility of chromosomal damage. Alternatively, the dimerization mechanism would provide a way of supplying the additional R subunits required for the restriction of a single site on a circular DNA plasmid (22).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank students Sam Shribman and W. L. Chin for their work on the methyltransferase, and Prof. Noreen Murray (Edinburgh) for many helpful discussions.

FUNDING

Biotechnology and Biological Sciences Research Council (BBS/B/1065X to R.M.H., J.M.E., D.T.F.D., BB/D001870/1 to D.T.F.D.); The Wellcome Trust (GR080463MA to D.T.F.D). Biotechnology and Biological Sciences Research Council (grant BBS/B/1065X to R.M.H., J.M.E. and D.T.F.D).

Conflict of interest statement. None declared.

REFERENCES

1. Binnig,G. and Rohrer,H. (1982) Scanning tunneling microscopy. *Helv. Phys. Acta*, **55**, 726–735.
2. Alessandri,A. and Facci,P. (2005) AFM: a versatile tool in biophysics. *Meas. Sci. Technol.*, **16**, R65–R92.
3. Uchihashi,T., Choi,N., Tanigawa,M., Ashino,M., Sugawara,Y., Nishijima,H., Akita,S., Nakayama,Y., Tokumoto,H., Yokoyama,K., Morita,S. and Ishikawa,M. (2000) Carbon nanotube tip for highly-reproducible imaging of deoxyribonucleic acid helical

turns by noncontact atomic force microscopy. *J. Appl. Phys.*, **39**, 887–889.
4. Bennink,M.L., Nikova,D.N., van der Werf,K.O. and Greve,J. (2003) Dynamic imaging of single DNA–protein interactions using atomic force microscopy. *Anal. Chim. Acta*, **479**, 3–15.
5. Jiao,Y., Cherny,D.I., Heim,G., Jovin,T.M. and Schäffer,T.E. (2001) Dynamic interactions of p53 with DNA in solution by time-lapse atomic force microscopy. *J. Mol. Biol.*, **314**, 233–243.
6. Müller,D.J., Amrein,M. and Engel,A. (1997) Adsorption of biological molecules to a solid support for scanning probe microscopy. *J. Struct. Biol.*, **119**, 172–188.
7. Guthold,M., Zhu,X., Rivetti,C., Yang,G., Thomson,N.H., Kasas,S., Hansma,H.G., Smith,B., Hansma,P.K. and Bustamante,C. (1999) Direct observation of one-dimensional diffusion and transcription by *Escherichia coli* RNA polymerase. *Biophys. J.*, **77**, 2284–2294.
8. Berge,T., Ellis,D.J., Dryden,D.T.F., Edwardson,J.M. and Henderson,R.M. (2000) Translocation-independent dimerization of the EcoKI endonuclease visualized by atomic force microscopy. *Biophys. J.*, **79**, 479–484.
9. Ellis,D.J., Dryden,D.T.F., Berge,T., Edwardson,J.M. and Henderson,R.M. (1999) Direct observation of DNA translocation and cleavage by the EcoKI endonuclease using atomic force microscopy. *Nat. Struct. Biol.*, **6**, 15–17.
10. van Noort,J., van der Heijden,T., Dutta,C.F., Firman,K. and Dekker,C. (2004) Initiation of translocation by Type I restriction-modification enzymes is associated with a short DNA extrusion. *Nucleic Acids Res.*, **32**, 6540–6547.
11. Murray,N.E. (2000) Type I restriction systems: Sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.*, **64**, 412–434.
12. Dryden,D.T.F., Murray,N.E. and Rao,D.N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.*, **29**, 3728–3741.
13. Bickle,T.A. and Kruger,D.H. (1993) Biology of DNA restriction. *Microbiol. Rev.*, **57**, 434–450.
14. Yuan,R. (1981) Structure and mechanism of multifunctional restriction endonucleases. *Ann. Rev. Biochem.*, **50**, 285–315.
15. Studier,F.W. and Bandyopadhyay,P.K. (1988) Model for how type-I restriction enzymes select cleavage sites in DNA. *Proc. Natl Acad. Sci. USA*, **85**, 4677–4681.
16. Janscak,P., MacWilliams,M.P., Sandmeier,U., Nagaraja,V. and Bickle,T.A. (1999) DNA translocation blockage, a general mechanism of cleavage site selection by type I restriction enzymes. *EMBO J.*, **18**, 2638–2647.
17. Murray,N.E., Batten,P.L. and Murray,K. (1973) Restriction of bacteriophage-lambda by *Escherichia coli K. J. Mol. Biol.*, **81**, 395–407.
18. Szczelkun,M.D., Dillingham,M.S., Janscak,P., Firman,K. and Halford,S.E. (1996) Repercussions of DNA tracking by the type IC restriction endonuclease EcoR124I on linear, circular and catenated substrates. *EMBO J.*, **15**, 6335–6347.
19. Dreier,J., MacWilliams,M.P. and Bickle,T.A. (1996) DNA cleavage by the Type IC restriction-modification enzyme EcoR124II. *J. Mol. Biol.*, **264**, 722–733.
20. Rosamond,J., Endlich,B. and Linn,S. (1979) Electron microscopic studies of the mechanism of action of the restriction endonuclease of *Escherichia coli*. B. *J. Mol. Biol.*, **129**, 619–635.
21. Janscak,P., Abadjieva,A. and Firman,K. (1996) The type I restriction endonuclease R.EcoR124I: over-production and biochemical properties. *J. Mol. Biol.*, **257**, 977–991.
22. Jindrova,E., Schmid-Nuoffer,S., Hamburger,F., Janscak,P. and Bickle,T.A. (2005) On the DNA cleavage mechanism of Type I restriction enzymes. *Nucleic Acids Res.*, **33**, 1760–1766.
23. Yuan,R., Hamilton,D.L. and Burckhardt,J. (1980) DNA translocation by the restriction enzyme from *Escherichia coli K. Cell*, **20**, 237–244.
24. Bickle,T.A., Brack,C. and Yuan,R. (1978) ATP-induced conformational changes in the restriction endonuclease from *Escherichia coli K-12. Proc. Natl Acad. Sci. USA*, **75**, 3099–3103.
25. Endlich,B. and Linn,S. (1985) The DNA restriction endonuclease of *Escherichia coli* B. I. Studies of the DNA translocation and the ATPase activities. *J. Biol. Chem.*, **260**, 5720–5728.
26. Reich,S., Gossl,D., Reuter,M., Rabe,J.P. and Kruger,D.H. (2004) Scanning force microscopy of DNA translocation by the type III restriction enzyme EcoP15I. *J. Mol. Biol.*, **341**, 337–343.
27. Crampton,N., Roes,S., Dryden,D.T.F., Rao,D.N., Edwardson,J.M., and Henderson,R.M. (2007) DNA looping and translocation provide an optimal cleavage mechanism for the type III restriction enzymes. *EMBO J.*, **26**, 3815–3825.
28. Crampton,N., Yokokawa,M., Dryden,D.T.F., Edwardson,J.M., Rao,D.N., Takeyasu,K., Yoshimura,S.H. and Henderson,R.M. (2007) Fast-scan atomic force microscopy reveals that the type III restriction enzyme EcoP151 is capable of DNA translocation and looping. *Proc. Natl Acad. Sci. USA*, **104**, 12755–12760.
29. Schneider,S.W., Larmer,J., Henderson,R.M. and Oberleithner,H. (1998) Molecular weights of individual proteins correlate with molecular volumes measured by atomic force microscopy. *Pflugers Arch.*, **435**, 362–367.
30. Davies,G.P., Kemp,P., Molineux,I.J. and Murray,N.E. (1999) The DNA translocation and ATPase activities of restriction-deficient mutants of EcoKI. *J. Mol. Biol.*, **292**, 787–796.
31. Dryden,D.T.F., Cooper,L.P., Thorpe,P.H. and Byron,O. (1997) The in vitro assembly of the EcoKI type I DNA restriction/modification enzyme and its in vivo implications. *Biochemistry*, **36**, 1065–1076.
32. Sankararaman,S. and Marko,J.F. (2005) Formation of loops in DNA under tension. *Phys. Rev.*, **71**, 219–211.
33. Su,T.J., Tock,M.R., Egelhaaf,S.U., Poon,W.C.K. and Dryden,D.T.F. (2005) DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Res.*, **33**, 3235–3244.
34. Powell,L.M., Dryden,D.T.F. and Murray,N.E. (1998) Sequence-specific DNA binding by EcoKI, a type IA DNA restriction enzyme. *J. Mol. Biol.*, **283**, 963–976.
35. Halford,S.E. and Marko,J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
36. Yuan,R., Bickle,T.A., Ebbers,W. and Brack,C. (1975) Multiple steps in DNA recognition by restriction endonuclease from E. coli K. *Nature*, **256**, 556–560.
37. Ogata,K., Sato,K. and Tahirov,T. (2003) Eukaryotic transcriptional regulatory complexes: cooperativity from near and afar. *Curr. Opin. Struct. Biol.*, **13**, 262.
38. Pouget,N., Turlan,C., Destainville,N., Salomé,L. and Chandler,M. (2006) IS911transpososome assembly as analysed by tethered particle motion. *Nucleic Acids Res.*, **34**, 4313–4323.
39. Oram,M., Marko,J.F. and Halford,S.E. (1997) Communications between distant sites on supercoiled DNA from non-exponential kinetics for DNA synapsis by resolvase. *J. Mol. Biol.*, **270**, 396–412.
40. Halford,S.E., Welsh,A.J. and Szczelkun,M.D. (2004) Enzyme-mediated DNA looping. *Ann. Rev. Biophys. Biomol. Struct.*, **33**, 1–24.
41. Blakely,G.W. and Murray,N.E. (2006) Control of the endonuclease activity of type I restriction-modification systems is required to maintain chromosome integrity following homologous recombination. *Mol. Microbiol.*, **60**, 883–893.