762–770 Nucleic Acids Research, 2009, Vol. 37, No. 3 Published online 11 December 2008
doi:10.1093/nar/gkn988

The structure of M.EcoKI Type I DNA methyltransferase with a DNA mimic antirestriction protein

Christopher K. Kennaway[1], Agnieszka Obarska-Kosinska[2,3], John H. White[4], Irina Tuszynska[2,3], Laurie P. Cooper[4], Janusz M. Bujnicki[2,5], John Trinick[1] and David T. F. Dryden[4,*]

[1]Astbury Centre, Institute of Molecular and Cellular Biology, University of Leeds, UK, [2]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, Trojdena 4, PL-02-109 Warsaw, [3]Institute of Biochemistry and Biophysics PAS, Pawinskiego 5A, 02-106 Warsaw, Poland, [4]School of Chemistry, University of Edinburgh, The Kings' Buildings, Edinburgh, EH9 3JJ, UK and [5]Bioinformatics Laboratory, Institute of Molecular Biology and Biotechnology, Adam Mickiewicz University, Umultowska 89, PL-61-614 Poznan, Poland

Received October 16, 2008; Revised November 20, 2008; Accepted November 21, 2008

ABSTRACT

Type-I DNA restriction–modification (R/M) systems are important agents in limiting the transmission of mobile genetic elements responsible for spreading bacterial resistance to antibiotics. EcoKI, a Type I R/M enzyme from *Escherichia coli*, acts by methylation- and sequence-specific recognition, leading to either methylation of DNA or translocation and cutting at a random site, often hundreds of base pairs away. Consisting of one specificity subunit, two modification subunits, and two DNA translocase/endonuclease subunits, EcoKI is inhibited by the T7 phage antirestriction protein ocr, a DNA mimic. We present a 3D density map generated by negative-stain electron microscopy and single particle analysis of the central core of the restriction complex, the M.EcoKI M_2S_1 methyltransferase, bound to ocr. We also present complete atomic models of M.EcoKI in complex with ocr and its cognate DNA giving a clear picture of the overall clamp-like operation of the enzyme. The model is consistent with a large body of experimental data on EcoKI published over 40 years.

INTRODUCTION

The EcoKI Type I DNA restriction–modification (R/M) enzyme was the first R/M enzyme to be discovered (1) and purified (2). It was determined that EcoKI, like all Type I R/M enzymes (3–8), is a complex oligomer formed from a core methyltransferase (MTase, M.EcoKI) comprised one HsdS DNA *S*pecificity subunit and two HsdM DNA *M*odification subunits, M_2S_1, with a total molecular weight of ~169 kDa. This is complexed with two additional HsdR DNA *R*estriction subunits, giving a complex $R_2M_2S_1$ with a total molecular weight of ~440 kDa that is capable of functioning as both MTase and restriction endonuclease (EcoKI or R.EcoKI) (9,10). Other Type I R/M enzymes display more variable subunit structures with some subunits weakly bound or capable of forming larger aggregates (11–13). Nevertheless, the core functions are performed by the M_2S_1 MTase and the $R_2M_2S_1$ bifunctional MTase/endonuclease. Despite much subsequent work, it is only recently that crystal structures of HsdS from *Methanococcus jannaschi* and *Mycoplasma genitalium* (14,15) and HsdM from *Escherichia coli* and *Bacteroides thetaiotaomicron* (pdb code: 2ar0, New York Structural GenomiX Research Consortium, DOI 10.2210/pdb2ar0/pdb and pdb code: 2okc, Joint Center for Structural Genomics, DOI 10.2210/pdb2okc/pdb, respectively) have become available. The overall structure is not known for either the MTase or the complete R/M complex although several models have been proposed (16–20).

The absence of these structures has hindered study of Type I R/M enzymes but their enduring appeal as research targets (21) is due to their extraordinary complexity of operation. This comprises recognition of a bipartite, asymmetric DNA specificity sequence (e.g. EcoKI recog-

*To whom correspondence should be addressed. Tel: +44 131 650 4735; Fax: +44 131 650 6453; Email: david.dryden@ed.ac.uk

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

nizes A<u>A</u>C[N_6]G<u>T</u>GC), the recognition of methylation of the adenines at the bold underlined positions, and switching between an MTase activity on newly replicated, hemimethylated host DNA and an endonuclease activity on unmethylated foreign DNA (3). The latter reaction requires massive ATP hydrolysis to drive translocation of up to 50 kb of DNA (22,23) at rates of up to 1 kb per second (24–29). The DNA is then cleaved at a random sequence remote from the original specificity sequence, with the generation of variable length single-strand overhangs (28,29). DNA-induced dimerization of EcoKI prior to translocation has been observed using atomic force microscopy (30,31). These varied activities have functional analogies with the operation of other complex DNA-manipulating machines (32,33). Of further note is the presence of two target recognition domains (TRDs) in HsdS that can be easily exchanged for other TRDs to generate novel DNA sequence specificities. The Type I R/M systems can be divided into, at present, five related families, IA to IE, defined by genetic complementation, antibody cross reactivity, DNA hybridization and sequence comparisons (34,35).

Of great topical interest due to the spread of drug resistance by horizontal gene transfer in bacteria (36,37), is the susceptibility of the R/M enzymes to antirestriction measures encoded by a variety of mobile genetic elements such as phage, conjugative plasmids and transposons (38–43). These antirestriction measures include the production of proteins that structurally mimic DNA (44). DNA mimics block the binding sites within R/M enzymes and inactivate them (45–47). The dimeric protein ocr from T7 phage matches the shape and surface potential characteristics of double-stranded DNA bent to the degree necessary for recognition by EcoKI.

We now present a model of M.EcoKI bound to ocr at ~18 Å resolution as determined by negative-stain electron microscopy (EM). Combining the EM model with new atomic models generated from the known crystallographic structures of the subunits allows a detailed structure of the M.EcoKI enzyme to be constructed. This structure is consistent with and rationalizes many experimental results obtained with M.EcoKI and other Type I MTases.

MATERIALS AND METHODS

Protein samples

M.EcoKI and the ocr protein were prepared as previously described (9,10,46). A complex of M.EcoKI and ocr was prepared by mixing the two proteins together at a molar ratio of 1:1.5 (29 μM M_2S_1 and 44 μM dimeric ocr) in 20 mM Tris, 100 mM NaCl, pH 7 buffer and incubating at room temperature for 5 min. The excess of ocr was not visible in the EM due to its small size.

EM

Grids were made by placing M.EcoKI and ocr complexes diluted 400 times onto UV-treated (48) continuous carbon-coated copper mesh grids, then stained with 1% uranyl acetate solution. Grids were viewed in a Jeol 1200EX electron microscope fitted with a LaB_6 electron source operating at 80 kV in low dose mode. Negatives (Kodak SO63) were taken at either 40 000× or 50 000× magnification with defocus ranging from 250 to 850 nm, and digitized at 6.25 μm step size using a Coolscan8000 (Nikon). The defocus and astigmatism were determined with CTFIND3 (49), then 17807 particles were picked from selected micrographs using Boxer (EMAN) (50) in autobox mode. Particles were band-pass filtered using Imagic (ImageScience, Germany), with low frequencies below 1/150 Å and those above the first zero of the contrast transfer function (CTF) being removed. After interpolating the 40 000× and 50 000× magnified particles to a common scale (3.125 Å/pixel), initial class average images without reference bias were generated in Imagic (51) by multi-variate statistical analysis and classification. Six independent starting models consisting of 16 randomly placed Gaussian spheres were made in EMAN and each used in nine rounds of 3D refinement with the dataset. Two of the randomly generated starting models converged on a map similar to others made with some degree of *a priori* knowledge of the structure using the angular reconstitution method (52). These two 3D maps also showed approximate 2-fold symmetry. One was taken for further refinements in EMAN with C2 point group symmetry imposed. At several steps during refinement, classes that had the best match to the reprojections were selected and new 3D maps made. The final map has a nominal resolution of 18 Å at 0.5 Fourier shell cross-correlation using the odd–even test, and was 3D Fourier filtered at 12 Å. Maps were visualized using UCSF Chimera (53).

Model building

Full details of the model building are described in the Supplementary Data and so are described only briefly here. The model of M.EcoKI bound to DNA or to ocr was built using the following crystal structures: HsdS(MjaXIP) from *M. jannaschi* (pdb code: 1yf2) (15), HsdM (EcoKI) (pdb code: 2ar0, New York Structural GenomiX Research Consortium, DOI 10.2210/pdb2ar0/pdb) and HsdM from *B. thetaiotaomicron* VPI-5482 (pdb code: 2okc, Joint Center for Structural Genomics, DOI 10.2210/pdb2okc/pdb), the structure of a Type II MTase bound to its cognate DNA duplex, M.TaqI–DNA (pdb code: 1g38) (54), and the ocr structure (pdb code: 1s7z) (44). The M.TaqI–DNA structure was used for modelling of the HsdS–HsdM orientation. HsdS was modelled on the 1yf2 template by homology modelling and the short C-terminal regions missing from the template were modelled *de novo* using ROSETTA (55). The crystal structure of the EcoKI HsdM has been solved, but we have remodelled its domain–domain contacts based on the crystal structure of a related HsdM from *B. thetaiotaomicron* VPI-5482. The missing C-terminal regions of EcoKI HsdM were modelled *de novo* using ROSETTA and fitted into the EM map for M.EcoKI–ocr described below.

To build the M.EcoKI–DNA model, DNA was added to the HsdS model based on the orientation of DNA in 1g38. HsdM were docked together using HADDOCK (56) to form a dimer via their N-terminal domains. Then, the

HsdS–DNA model was combined with the HsdM dimer model so as to satisfy the orientation between the TRD and the catalytic domains of M.TaqI. To find the best fits, normal mode analysis was performed to generate a range of conformations of HsdS and HsdM.

The M.EcoKI–ocr model was constructed by multi-body flexible docking using HADDOCK (56). During the docking procedure, residues of HsdS and HsdM that form contacts with the DNA in the model of M.EcoKI–DNA were restrained to make contact with ocr, based on the assumption that ocr binds to HsdS and HsdM in a similar manner as DNA (44). The alternative models of M.EcoKI–ocr were fitted to the EM map as single rigid bodies using the COLORES module from the SITUS package (57) and the best scoring model was selected.

RESULTS

EM results

Addition of stoichiometric quantities of the ocr inhibitor was found to stabilize the otherwise labile M.EcoKI complex. Individual molecules of M.EcoKI–ocr complexes embedded in heavy metal stain were imaged by EM and analysed by single particle methods. Particles appeared to be homogeneously sized and were well distributed. Class averages showed a wide variety of views and exhibited considerable detail (Figure 1a) that allowed a unique 3D reconstruction of the complex to a nominal resolution of 18 Å (Figure 1b and c). The atomic model described below has been placed within the EM reconstruction in Figure 1b and has been used to colour the reconstruction surface in Figure 1c. Although at the atomic level, the complex is asymmetric, 2-fold symmetry was imposed since HsdS is known to have strong pseudo-symmetry (14,15,17). The EM map longest dimension was ~120 Å and 100 Å along the base, and the overall shape was an excellent match to the complex modelled *in silico* based on biochemical and X-ray data, as described below. As described below and in the Supplementary Data, when one takes the size, shape and symmetry constraints of the HsdS, HsdM and ocr crystal structures and constraints from biochemical data, only one arrangement of these structures within the EM reconstruction was possible. Given this unique solution, density attributable to the two-helix coiled-coil spacer of HsdS could be clearly identified, as well as regions attributable to the two HsdM. In addition to contacting the TRDs of HsdS, the two HsdM wrap around ocr and make contact with each other. An extension projects from the bottom of each HsdM to either side of the

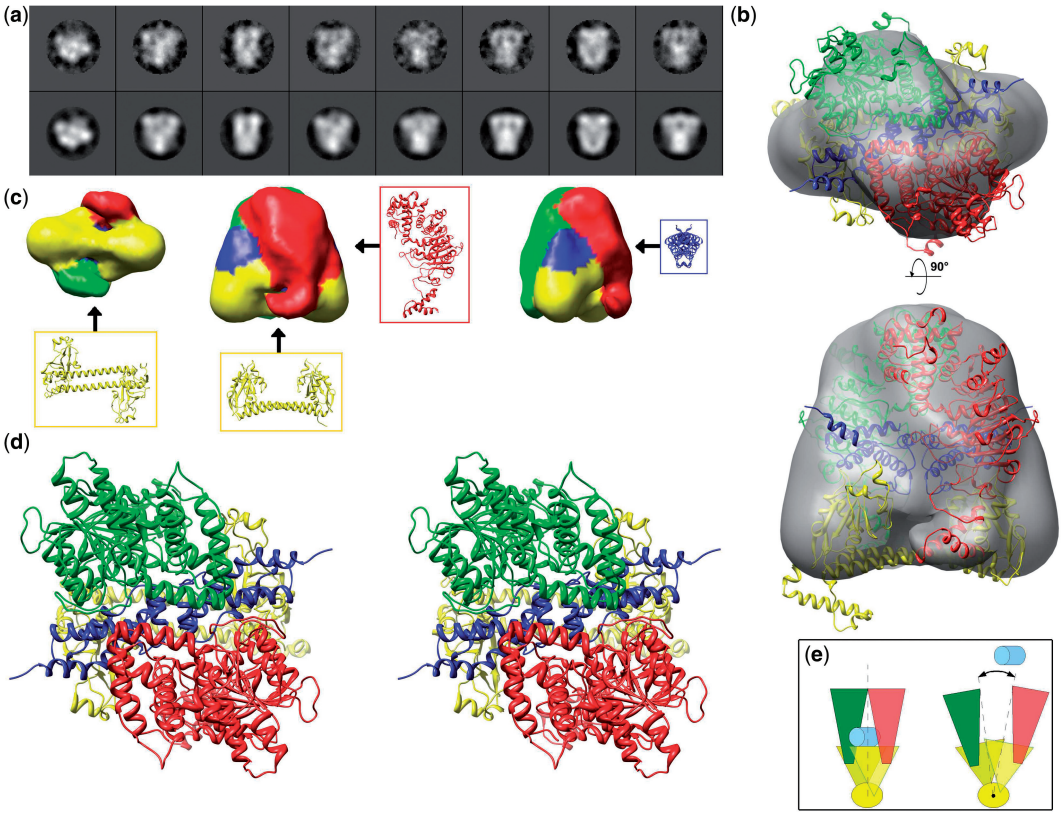

Figure 1. EM data, 3D reconstructions and model for the M.EcoKI–ocr complex. (**a**) Eight selected EM class average images (top row) with their corresponding reprojections of the EM map (second row), showing a range of views. (**b**) Two orthogonal semi-transparent surface representations of the EM 3D reconstruction each with a view of the modelled coordinates fitted as a rigid body (green and red—HsdM; blue—ocr dimer; yellow—HsdS). (**c**) Three surface views of the M.EcoKI–ocr EM 3D reconstruction coloured according to proximity to the fitted coordinates, with the positions of individual protein chains indicated with arrows. (**d**) A stereo view of the M.EcoKI–ocr atomic model. (**e**) Schematic diagram of proposed mechanism of clamping and release of DNA substrate (light blue) facilitated by twisting of the coiled-coil (yellow, viewed end-on).

HsdS coiled-coil spacer. This density returned in subsequent 3D refinements if masked out of the EM map, supporting the presence of these regions. These extensions could be attributed to the C-terminal regions of HsdM that are not present in the X-ray coordinates, as described below. Density in the centre of the map suggested the presence of the bound inhibitor protein ocr as it is known that ocr is fully enfolded by HsdS and HsdM (47). Sufficient detail was not present to show its exact orientation but its general position could be inferred. Features in the 3D map such as the straight coiled-coil spacer, which was ~16 Å wide, could also be seen in many class average images (Figure 1a). That features of this size were visible, just below the limit of resolution as measured by FSC, attests to the quality of the map and suggests that the 18 Å estimate of resolution is conservative.

Model of the M.EcoKI–ocr complex

The general shape of the model of M.EcoKI in complex with ocr dimer is shown in Figure 1d. This atomic model was constructed independently of the EM imaging based upon crystal structures of subunits and biochemical data. The MTase forms a clamp-like structure encircling the ocr dimer. The coiled-coil region of HsdS contains a considerable kink to allow the ocr molecule to fit optimally within this clamp. Each TRD of HsdS contacts one monomer of the ocr molecule and one HsdM. The exact location of the C-terminal helix of the HsdS subunit is speculative but has little effect on the overall outline shape of the molecule. The C-terminal regions of each HsdM, aa ~470–529, project down to contact the coiled-coil region of HsdS. The location of the C-terminal regions of each HsdM is also speculative but the postulated location does bring conserved regions of HsdM and HsdS into close contact as discussed later and does fit with the EM model. This is important as the other region of contact between HsdM and HsdS uses the TRD regions of HsdS; these are highly variable in sequence in Type I R/M enzymes and not expected to provide a good HsdM–HsdS interface. The interface between the two HsdM is formed by their N-terminal domains, aa 1–153. The clamp-like shape of the MTase model suggests that it must open up along the interface between the two HsdM to allow access for ocr or DNA (Figure 1e). The interfaces between the subunits in the MTase–ocr complex are less extensive than in the MTase–DNA complex described below and the whole MTase appears to be strained to accommodate the ocr molecule that is slightly larger than a DNA duplex. The model structure with ocr bound may represent a 'partially open clamp' state, while the model with DNA bound would be a 'fully closed clamp' state. Opening of the clamp to allow ocr (or DNA) binding may be accomplished by simply flexing the coiled-coil region of HsdS. Normal mode calculations indicate that such motions should occur with sufficient amplitude to allow binding of ocr or DNA (data not shown). Having constructed the atomic M.EcoKI–ocr model, it is obvious that this can be simply docked into the EM model as shown (Figure 1b). A comparison of the fit of the EcoKI MTase–ocr complex model to the EM map with the fits obtained for models in which subunit structures are randomly docked in to the EM map indicate that our data driven model fits the EM map significantly better than random models (see Supplementary Data).

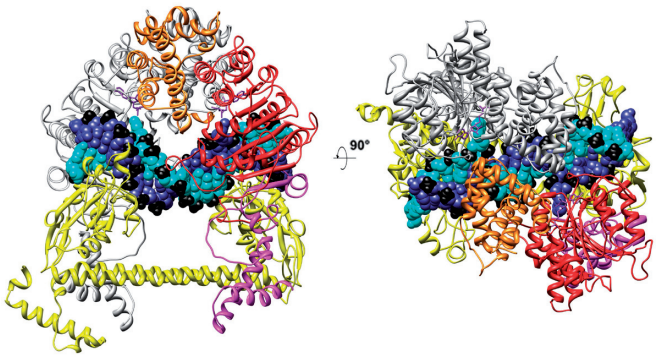

Figure 2. Model of the EcoKI MTase in complex with substrate DNA and AdoMet cofactors. HsdS is shown in yellow, the HsdM at the rear is shown in grey, DNA strands are shown in blue and cyan with phosphates in black, the two AdoMet molecules in purple. The HsdM in the foreground shows the N-terminal domain (aa 1–153) in orange, the catalytic domain (aa 154–469) in red and the C-terminal domain (aa 470–529) in magenta.

Model of the M.EcoKI–DNA complex

The model of M.EcoKI bound to DNA is shown in Figure 2. The two major differences between this and the partial models of Type I R/M enzymes published previously (16–19) are that the locations of the N-terminal and C-terminal regions of the HsdM are now defined within the overall structure. Most importantly, the N-terminal domains of HsdM can be seen to form a bridge over the DNA duplex, thus forming a clamp-like structure encircling the DNA, which is a key feature that was missing in earlier models. The location of the C-terminal regions of each HsdM with respect to HsdS is slightly different from that in M.EcoKI–ocr model, since the relative orientation of the subunits is altered. The coiled-coil region linking the two TRDs is straighter and the subunit interfaces within the MTase are more extensive in area than in the M.EcoKI–ocr model as the DNA duplex appears to fit better within the MTase allowing the clamp to close fully. The DNA bound to the TRDs contains a sharp bend in the non-specific spacer part of the target sequence and the bend angle of ~45° matches that observed experimentally (44,58).

DISCUSSION

The proposed *in silico* models, based upon crystallographic data, fit extremely well with the EM reconstruction and are attractive atomic resolution solutions to the structure of the MTase core of a Type I R/M enzyme. It is important to assess whether the models can accommodate the large body of experimental information concerning M.EcoKI. As most of the published work concerns interaction between EcoKI and DNA, we focus the following

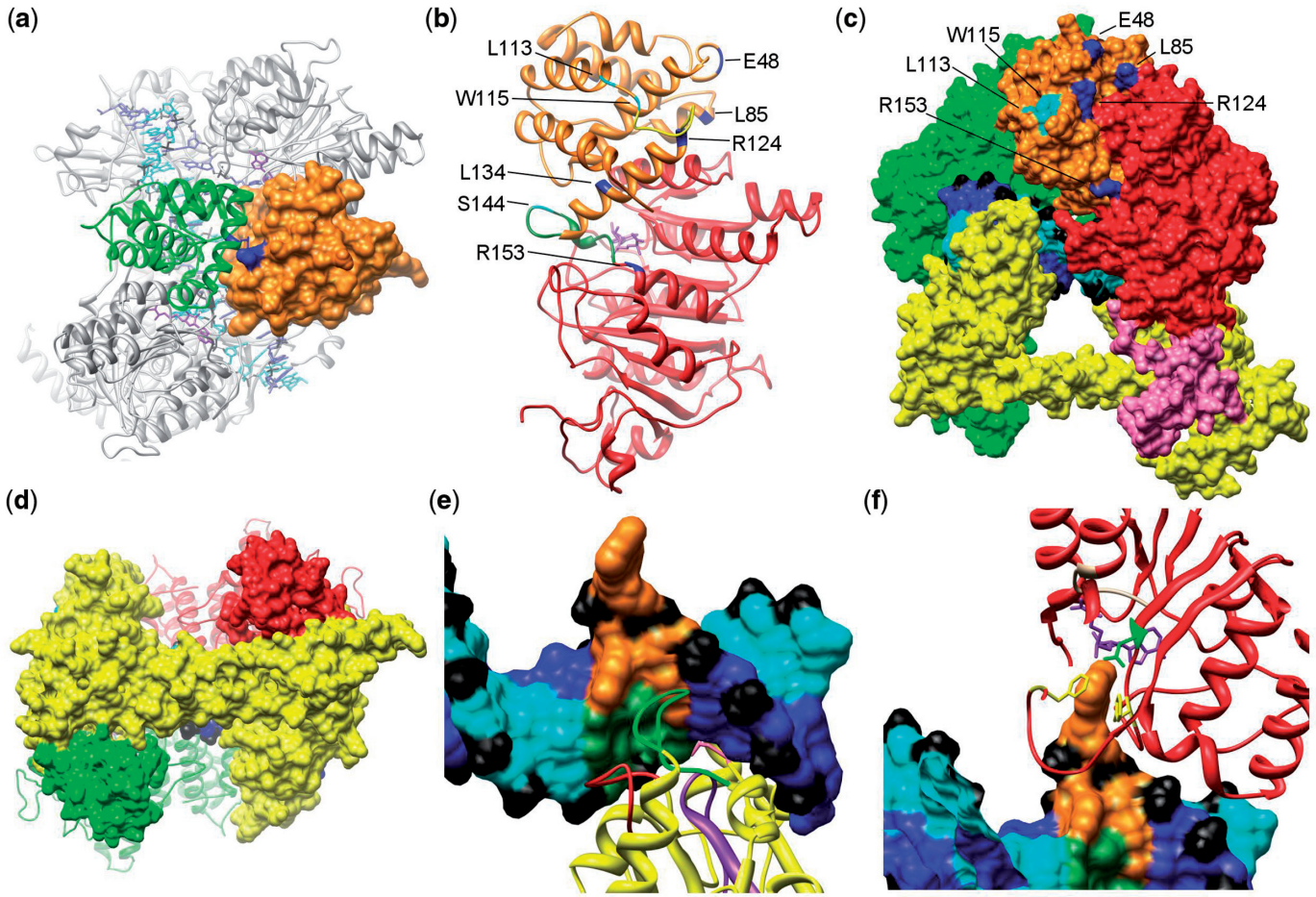

Figure 3. (**a**) A top view of the M.EcoKI–DNA model highlighting the interface between the N-terminal domains of HsdM subunits. One N-terminal domain is shown as an orange surface and the second as a green ribbon. The R72Q mutation giving an r^+m^* phenotype is shown in blue and lies on the HsdM interface. (**b**) The m^* mutations in the HsdM subunit. The N-terminal domain of HsdM is coloured orange and the central methyltransferase catalytic domain in red (the C-terminal domain is omitted). Residues where mutations cause the r^-m^* and r^+m^* phenotypes are shown coloured cyan and blue, respectively. Two flexible loops, disordered in the crystal structure, are shown in green and yellow. Both are associated with several m^* residues. The cofactor AdoMet, in purple, can be seen in the catalytic domain. (**c**) The m^* mutations in N-terminal domain of HsdM subunit presented in the M.EcoKI–DNA model. The HsdS subunit is yellow, the rear HsdM subunit as a green surface. The domains of the HsdM in the foreground are coloured as in Figure 2. The labelled residues in this HsdM subunit that cause r^-m^* and r^+m^* phenotypes are coloured cyan and blue, respectively. DNA is coloured blue and cyan with phosphates in black. (**d**) The contact between coiled-coil region of HsdS subunit and the C-terminal domain of HsdM subunit in M.EcoKI–DNA model. HsdS subunit surface is shown in yellow with the coiled-coil running horizontally across the image, HsdM subunits are coloured red and green with their C-terminal domains shown as a surface and the remainder as a ribbon. The close contact between the ends of the coiled-coil in HsdS with the C-terminal domains of HsdM can be seen. (**e**) Recognition of DNA by the N-terminal TRD of HsdS. Non-specific DNA is coloured blue and cyan with phosphates in black. The sequence AAC with the flipped out second A is shown in orange and the complementary sequence in dark green. The HsdS subunit is shown as a yellow ribbon with the previously identified DNA-contacting loops aa 83–91 in red and aa 96–118 in purple. These loops contact the complementary strand. The previously identified contact by Y27 (pink) can just be seen pointing into the minor groove at the rear of the image. The newly identified loop S_{139}AGANINNIK$_{148}$ is in green and primarily contacts the AAC sequence. (**f**) Contacts between the catalytic site of HsdM and DNA. DNA is shown as in (e) with the flipped out base sandwiched between HsdM F269 in the 'NPPF' motif IV and F345, both shown in yellow. The AdoMet cofactor is shown in purple and the HsdM motif I (beige) containing G177 is contacting the AdoMet. The rest of HsdM is in red. Two phenylalanine residues (F269 and F345, yellow) are available to form stacking interactions with the flipped out DNA base, and N266 (green) is within range for hydrogen bonding.

discussion primarily on the M.EcoKI–DNA model rather than the M.EcoKI–ocr model. Each section of the discussion describes how experimental data can be accommodated by the EM structure and the derived structural model and indicates new regions of interaction, which could be investigated in the future. Lastly, we compare the structural model of M.EcoKI with those recently proposed for the type IC enzyme, EcoR124I (19,20).

Analysis of the M.EcoKI–DNA model

Interface between the two HsdM via their N-terminal domains. Our model fits each HsdM into the EM structure in such a way as to bring their N-terminal domains into contact with each other (Figure 3a). They effectively form a bridge over the DNA and this bridge must swing open to allow the DNA to enter and be recognized by

HsdS. It is known that the ocr protein, which acts as a DNA mimic, is encircled by M.EcoKI (47) supporting the positioning of the N-terminal domains to complete the clamp encircling the DNA. The main interaction between these two domains is via helix G_{64}QEQLQFYRKM LVHL_{78} contacting, in an antiparallel manner, the equivalent helix within the other HsdM. A second contact region is between a loop H_{91}NVSTTIT_{98} with the equivalent loop in the other subunit. The helical axes and loops in this interface run approximately parallel to the axis of the coiled-coil in HsdS.

The N-terminal domain of each HsdM contains the m^* mutations (Figure 3a–c) (59), so called because they change the strong preference of M.EcoKI for methylating hemimethylated DNA target sequences (9) to an ability to additionally methylate unmodified targets. It is noteworthy that the disruption caused by the m^* mutations does not affect the ability of the enzyme to methylate DNA, only its ability to recognize the methylation state of the target sequence. A single m^* mutation, R72Q, lies within the predicted helix G_{64}QEQLQFYRK MLVHL_{78} and contacts the equivalent helix in the other HsdM (Figure 3a). This mutation has an r^+m^* phenotype suggesting a small deformation of the interface affecting communication of the methylation status of the two adenines in the target sequence and closure of the proposed clamp around the DNA but without affecting restriction by a complete EcoKI.

The mutations Q48K, L85Q, R124L, R124C, L134V, R153H and R153S also give rise to an r^+m^* phenotype. These locations in the model lie on the interface between the N-terminal m^* domain (aa 1–153) and the middle catalytic domain (aa 154–469) of the same HsdM. It is possible to envisage how these non-conservative changes will disrupt this interface resulting in the observed r^+m^* phenotype, perhaps by interfering with closure of the clamp.

The remaining mutations, L113R, L113P, L113Q, W115R and S144Y all give an r^-m^* phenotype. L113, W115 and S144 lie on the surface of the modelled structure and do not contact any other part of the MTase (Figure 3b and c). It is possible that they induce a minor folding problem for the N-terminal domain affecting its interaction with the other HsdM although the mutant enzymes can be purified as normal (our unpublished results). The observation that restriction activity is lost because of these mutations may further suggest that this small region interacts with the HsdR subunit in the complete EcoKI.

Interactions between HsdM and HsdS. The EM density at the bottom of HsdM has a protrusion that passes close to the HsdS coiled-coil and we believe this to be the C-terminal region of HsdM comprising residues ~470–529. This region is disordered in the two HsdM crystal structures available suggesting that it is flexible in the absence of HsdS. The EM results support the rigidity of the extensions in the complex, since flexible areas are invariably blurred out by the averaging inherent in the single particle methodology. However, the limited resolution of the EM map, and lack of X-ray data, only allows the computational atomic model to be positioned approximately. Sequence analysis indicates that the C-terminal region of HsdM has a propensity to form α-helices and the model suggests that these reach out and contact the coiled-coil in HsdS, as shown in Figure 3d. These contacts would play a central role in the assembly of the MTase providing an interface between known regions of high sequence conservation in both HsdS and HsdM of the Type IA family (5,8). This high degree of conservation in the interface between HsdS and HsdM would be essential to facilitate the exchange of subunits of different Type IA family systems by complementation. Experimental support for this aspect of the EM and computational model comes from proteolysis experiments. It was found that ~43 residues from the C-terminal region of HsdM could be easily removed, leaving a stable, 55-kDa folded protein (60,61). However, this truncated HsdM was unable to bind to M_1S_1 complexes indicating that a subunit interface had been removed (60).

In the model, a second HsdS–HsdM interface occurs via a helix-loop structure, N_{133}KISSLSAGANIA_{144}, within the first TRD of HsdS and two loops, aa 430–433 and aa 464–470, on the edge of the catalytic domain within HsdM (not shown in Figure 3). Contacting the same region of the second HsdM, one can identify an equivalent helix and loop in the second TRD of HsdS as A_{346}MMNCVKTTSGQK_{358}. It is known that the TRDs within a family of Type I R/M systems can be exchanged easily to generate new sequence specificities (5,8). However, the TRD amino acid sequences are very variable and would not be expected to offer enough conservation to allow the formation of a particularly stable interface with HsdM. Hence, these regions must be of secondary importance to the regions discussed in the previous paragraph.

HsdS to DNA contacts. The N-terminal part of HsdS forms the first TRD, which recognizes the sequence 5′-AAC-3′. Many mutations in this TRD of HsdS from EcoKI have been catalogued (62,63) but the majority of these were silent. Early modelling studies (64) suggested that these were not on the protein–DNA interface. However, two loops were identified in this analysis, aa 83–91 and aa 96–118, as contacting the DNA and this is corroborated in the current model. Within the second of these loops, mutations of aa 91, 95, 105, 107 and 108 led to a loss of activity although the protein was folded and could bind DNA (62,63). In our model, these appear to contact the complementary strand to the 5′-AAC-3′ sequence (Figure 3e). Chen *et al.* (65) identified a further amino acid, Y27, by cross-linking experiments in contact with the DNA target. In our map, Y27 is within 4 Å of the minor groove of DNA (not visible in Figure 3e as it is behind the DNA).

In the first TRD of HsdS, our model has an additional loop to those identified previously, S_{139}AGANINNIK$_{148}$ (Figure 3e). This loop is closest to the major groove DNA bases forming the target sequence 5′-AAC-3′. Mutations S139P, G141V and G141A affected activity but other more conservative substitutions did not alter activity (62,63). N145 is the residue in our model that appears to be able to contact the unpaired thymine base (the partner

of the flipped out adenine). In the second TRD, the equivalent loop is T_{354}SGQKGISGK$_{363}$, and Q357 is able to contact the unpaired thymine in the complementary strand for the second part of the target sequence, 5′-GATC-3′.

HsdM to DNA contacts. Willcock *et al*. (66) substituted amino acids within the well-characterized, conserved MTase motifs within the catalytic domain of HsdM. The substitution in the AdoMet binding motif I, G177D, affected cofactor binding and the model shows motif I in close contact with the AdoMet (Figure 3f). Changes at locations N266 or F269 in the catalytic motif IV led to a loss of activity. In the model of M.EcoKI with DNA bound, these locations can be seen adjacent to the flipped-out adenine (Figure 3f). In addition, F345 is close to the flipped-out adenine. A similar sandwich of a flipped-out adenine between two aromatic side chains was observed in the crystal structure of the M.TaqI MTase (67).

Analysis of the M.EcoKI–ocr model

This atomic model (Figure 1) is more open than the model with DNA (Figure 2) as there are fewer contacts between the three subunits of M.EcoKI. In particular, the interface between the two HsdM is formed only by helix 64–78, and the C-terminal helices of HsdM, 487–507, appear to have minimal contact with the coiled-coil region of HsdS. It appears that ocr is slightly too large to be as easily accommodated as the DNA molecule. As the subunit interfaces are smaller, one might say that the MTase is almost 'bursting at its seams' as it attempts to enwrap the ocr protein (47). This suggests that a clamp-like open and shut mechanism is used by M.EcoKI to bind DNA or ocr.

Mechanism for binding DNA or ocr

The EM structure makes it clear that M.EcoKI can encircle DNA and, to a lesser extent, ocr. Therefore, the structure must open up to allow the DNA (or the ocr protein) access to the TRDs on HsdS and the catalytic site on HsdM. Simple normal mode calculations were performed on HsdS to examine the potential role of the coiled-coil region in allowing such an opening (data not shown). These calculations showed that the lowest mode was a flexing of the coiled-coil moving the TRDs closer together and then further apart. The next lowest mode was a twisting of the coiled-coil around its axis. Noting that the interface between the two HsdM runs approximately parallel to the axis of the coiled-coil in HsdS, then these motions would move the N-terminal domains of the HsdM apart and allow access to circular DNA. A twist of the coiled-coil region of ~12–14° in our model would produce a gap between HsdM N-terminal domains sufficient to allow passage of DNA or ocr. Obviously, the motions of the complete M.EcoKI will be considerably more complex but this analysis is strongly suggestive of a flexible clamp-like enzyme.

Comparison to the Type IC enzyme EcoR124I

Kneale (17) originally recognized the 2-fold symmetry imposed upon Type I enzymes via the structure of the S subunit. This symmetry defines the location of the M subunits and was exploited by Obarska *et al*. (19) to produce a rather opened-out model of the MTase where the N-terminal domains of each HsdM subunit were speculatively placed some distance from both the DNA and any contact with the other HsdM subunit. Although based upon an erroneous structure of an HsdM subunit (pdb 2ar0 is incomplete and partially incorrectly assigned, see Supplementary Data), the model could, if correct, explain the large compaction of M.EcoR124I observed by small angle neutron scattering when DNA is added to the enzyme (67). This compaction was suggested to be the movement of the HsdM subunits swinging in towards the DNA and closing like a clamp to trap the DNA rather like the closed state we have observed in our structure (Figure 1). Thus, it is probable that the structure of Type I MTases fluctuates between an open form as approximately modelled by Obarska *et al*. (19) and the closed form observed here and by Taylor *et al*. (67). Normal mode analyses of our model make it clear that this opening and closing motion, which must open a gap at least 2 nm wide between HsdM subunits to allow double-stranded DNA to pass through, is facilitated by twisting and bending of the coiled-coil region in the HsdS subunit.

More recently, small-angle neutron scattering data and computer modelling of the HsdR subunit of EcoR124I have been presented (20). The HsdR subunit appears to be a disc-shaped protein but a unique atomic model could not be constructed. Presumably, one of these subunits could be placed on either side of the MTase structure so that the ends of the DNA helix would contact the helicase domains in the HsdR subunit. Until a unique structure of the HsdR subunit is presented and more data are acquired on its contacts to the MTase, further speculation on the complete structure of the Type I R/M enzymes seems premature. However, we note that we are currently making progress towards this goal.

CONCLUSION

The new EM structure and computational model of M.EcoKI rationalize, for the first time, a large body of experimental data obtained using many different methods over many years. A mechanistic explanation of the Type I MTase enzymes is suggested by the model, which clearly indicates locations for further analyses such as the HsdM–HsdS and HsdM–HsdM interfaces. When combined with data on M.EcoR124I (67), the model also suggests that a dynamic opening and closing of the protein, driven by a flexing and twisting of the conserved coiled-coil region within HsdS, is required to open up the HsdM–HsdM interface to allow either DNA binding or attack of the ocr antirestriction protein, a protein that 'disguises itself' as DNA. The model also provides a basis for completing or constructing models of other Type I MTases from other

families such as the Type IC enzyme M.EcoR124I (19) and the Type IB enzyme M.EcoAI.

Data depositions

Atomic coordinates from M.EcoKI–ocr and M.EcoKI–DNA complexes are available from the authors or ftp://ftp.genesilico.pl/iamb/models/MTases/EcoKI/. The EM map is deposited with the EMDB (www.ebi.ac.uk/msd) with code 1543.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

This work was initiated at the Isaac Newton Institute for Mathematical Sciences Workshop on 'Statistical Mechanics of Molecular and Cellular Biological Systems', January to July 2004. We thank Professor Noreen Murray and numerous colleagues in her laboratory for countless discussions on the structure of EcoKI.

FUNDING

We gratefully acknowledge support from the Biotechnology and Biological Sciences Research Council (BB/D001870/1 to D.T.F.D. and J.T.), the Wellcome Trust (GR080463MA to D.T.F.D.) and the Royal Society of Edinburgh (International Exchange Programme to A.O.-K. and J.M.B.). A.O.-K. and J.M.B. were also supported by the 6FP grant from the European Union (MRTN-CT-2005-019566) and supplementary funds from the Polish Ministry of Science. Funding for open access charge: Wellcome Trust.

Conflict of interest statement. None declared.

REFERENCES

1. Bertani,G. and Weigle,J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113–121.
2. Meselson,M. and Yuan,R. (1968) DNA restriction enzyme from *E. coli*. *Nature*, **217**, 1110–1114.
3. Yuan,R. (1981) Structure and mechanism of multifunctional restriction endonucleases. *Ann. Rev. Biochem.*, **50**, 285–315.
4. Dryden,D.T.F. (1999) Bacterial DNA methyltransferases. In Cheng,X. and Blumenthal,R.M. (eds), *S-Adenosylmethionine-Dependent Methyltransferases: Structures and Functions*. World Scientific Publishing, Singapore, pp. 283–340.
5. Murray,N.E. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.*, **64**, 412–434.
6. Dryden,D.T.F., Murray,N.E. and Rao,D.N. (2001) Nucleoside triphosphate-dependent restriction enzymes. *Nucleic Acids Res.*, **29**, 3728–3741.
7. Bourniquel,A.A. and Bickle,T.A. (2002) Complex restriction enzymes: NTP-driven molecular motors. *Biochimie*, **84**, 1047–1059.
8. Murray,N.E. (2002) Immigration control of DNA in bacteria: self versus non-self. *Microbiology*, **148**, 3–20.
9. Dryden,D.T.F., Cooper,L.P. and Murray,N.E. (1993) Purification and characterization of the methyltransferase from the Type I restriction and modification system of *Escherichia coli* K12. *J. Biol. Chem.*, **268**, 13228–13236.
10. 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.
11. Eskin,B. and Linn,S. (1972) Deoxyribonucleic acid modification and restriction enzymes of *Escherichia coli* B.2. Purification, subunit structure, and catalytic properties of restriction endonuclease. *J. Biol. Chem.*, **247**, 6183–6191.
12. Suri,B., Shepherd,J.C.W. and Bickle,T.A. (1984) The EcoA restriction and modification system of *Escherichia coli* 15T$^-$ – enzyme structure and DNA recognition sequence. *EMBO J.*, **3**, 575–579.
13. Janscak,P., Dryden,D.T.F. and Firman,K. (1998) Analysis of the subunit assembly of the type IC restriction–modification enzyme EcoR124I. *Nucleic Acids Res.*, **26**, 4439–4445.
14. Calisto,R.M., Pich,O.Q., Pinol,J., Fita,I., Querol,E. and Carpena,X. (2005) Crystal structure of a putative type I restriction–modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.*, **351**, 749–762.
15. Kim,J.S., DeGiovanni,A., Jancarik,J., Adams,P.D., Yokota,H., Kim,R. and Kim,S.H. (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction–modification enzyme and its functional implications. *Proc. Natl Acad. Sci. USA*, **102**, 3248–3253.
16. Burckhardt,J., Weisemann,J., Hamilton,D.L. and Yuan,R. (1981) Complexes formed between the restriction endonuclease EcoK and heteroduplex DNA. *J. Mol. Biol.*, **153**, 425–440.
17. Kneale,G.G. (1994) Symmetrical model for the domain structure of Type-I DNA methyltransferases. *J. Mol. Biol.*, **243**, 1–5.
18. Dryden,D.T.F., Sturrock,S.S. and Winter,M. (1995) Structural modelling of a Type I DNA methyltransferase. *Nat. Struct. Biol.*, **2**, 632–635.
19. Obarska,A., Blundell,A., Feder,M., Vejsadova,S., Sisakova,E., Weiserova,M., Bujnicki,J.M. and Firman,K. (2006) Structural model for the multisubunit Type IC restriction–modification DNA methyltransferase M.EcoR124I in complex with DNA. *Nucleic Acids Res.*, **34**, 1992–2005.
20. Obarska-Kosinska,A., Taylor,J.E., Callow,P., Orlowski,J., Bujnicki,J.M. and Kneale,G.G. (2008) HsdR subunit of the type I restriction–modification enzyme EcoR1241: biophysical characterisation and structural modelling. *J. Mol. Biol.*, **376**, 438–452.
21. Loenen,W.A.M. (2003) Tracking EcoKI and DNA fifty years on: a golden story full of surprises. *Nucleic Acids Res.*, **31**, 7059–7069.
22. 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.
23. Garcia,L.R. and Molineux,I.J. (1999) Translocation and specific cleavage of bacteriophage T7 DNA in vivo by EcoKI. *Proc. Natl Acad. Sci. USA*, **96**, 12430–12435.
24. 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.
25. Seidel,R., van Noort,J., van der Scheer,C., Bloom,J.G.P., Dekker,N.H., Dutta,C.F., Blundell,A., Robinson,T., Firman,K. and Dekker,C. (2004) Real-time observation of DNA translocation by the type I restriction–modification enzyme EcoR124I. *Nat. Struct. Mol. Biol.*, **11**, 838–843.
26. Seidel,R., Bloom,J.G., Dekker,C. and Szczelkun,M.D. (2008) Motor step size and ATP coupling efficiency of the dsDNA translocase EcoR124I. *EMBO J.*, **27**, 1388–1398.
27. Seidel,R., Bloom,J.G.P., van Noort,J., Dutta,C.F., Dekker,N.H., Firman,K., Szczelkun,M.D. and Dekker,C. (2005) Dynamics of initiation, termination and reinitiation of DNA translocation by the motor protein EcoR124I. *EMBO J.*, **24**, 4188–4197.
28. Endlich,B. and Linn,S. (1985) The DNA restriction endonuclease of Escherichia-coli-B.2. Further studies of the structure of DNA intermediates and products. *J. Biol. Chem.*, **260**, 5729–5738.
29. Jindrova,E., Schmid-Nuoffer,S., Hamburger,F., Janscak,P. and Bickle,T.A. (2005) On the DNA cleavage mechanism of Type I restriction enzymes. *FEBS J.*, **272**, 552–552.
30. 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. *Nature Struct. Biol.*, **6**, 15–17.
31. 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.
32. Flaus,A. and Owen-Hughes,T. (2004) Mechanisms for ATP-dependent chromatin remodelling: farewell to the tuna-can octamer? *Curr. Opin. Genet. Dev.*, **14**, 165–173.
33. Singleton,M.R., Dillingham,M.S. and Wigley,D.B. (2007) Structure and mechanism of helicases and nucleic acid translocases. *Ann. Rev. Biochem.*, **76**, 23–50.
34. Titheradge,A.J.B., King,J., Ryu,J. and Murray,N.E. (2001) Families of restriction enzymes: an analysis prompted by molecular and genetic data for type ID restriction and modification systems. *Nucleic Acids Res.*, **29**, 4195–4205.
35. Chin,V., Valinluck,V., Magaki,S. and Ryu,J. (2004) KpnBI is the prototype of a new family (IE) of bacterial type I restriction–modification system. *Nucleic Acids Res.*, **32**.
36. Thomas,C.M. and Nielsen,K.M. (2005) Mechanisms of, and barriers to, horizontal gene transfer between bacteria. *Nature Rev. Microbiol.*, **3**, 711–721.
37. Lindsay,J.A. and Holden,M.T. (2006) Understanding the rise of the superbug: investigation of the evolution and genomic variation of Staphylococcus aureus. *Funct. Integr. Genomics*, **6**, 186–201.
38. Kruger,D.H. and Bickle,T.A. (1983) Bacteriophage survival – multiple mechanisms for avoiding the deoxyribonucleic-acid restriction systems of their hosts. *Microbiol. Rev.*, **47**, 345–360.
39. Bickle,T.A. and Kruger,D.H. (1993) Biology of DNA restriction. *Microbiol. Rev.*, **57**, 434–450.
40. Wilkins,B.M. (2002) Plasmid promiscuity: meeting the challenge of DNA immigration control. *Environ. Microbiol.*, **4**, 495–500.
41. Zavilgelsky,G.B. (2000) Antirestriction. *Mol. Biol.*, **34**, 724–732.
42. Tock,M.R. and Dryden,D.T.F. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.*, **8**, 466–472.
43. Dryden,D.T.F. (2006) DNA mimicry by proteins and the control of enzymatic activity on DNA. *Trends Biotech.*, **24**, 378–382.
44. Walkinshaw,M.D., Taylor,P., Sturrock,S.S., Atanasiu,C., Berge,T., Henderson,R.M., Edwardson,J.M. and Dryden,D.T.F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell*, **9**, 187–194.
45. Bandyopadhyay,P.K., Studier,F.W., Hamilton,D.L. and Yuan,R. (1985) Inhibition of the Type I restriction–modification enzymes EcoB And EcoK by the gene 0.3 protein of bacteriophage-T7. *J. Mol. Biol.*, **182**, 567–578.
46. Atanasiu,C., Byron,O., McMiken,H., Sturrock,S.S. and Dryden,D.T.F. (2001) Characterisation of the structure of ocr, the gene 0.3 protein of bacteriophage T7. *Nucleic Acids Res.*, **29**, 3059–3068.
47. Atanasiu,C., Su,T.J., Sturrock,S.S. and Dryden,D.T.F. (2002) Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res.*, **30**, 3936–3944.
48. Walker,M., Knight,P. and Trinick,J. (1985) Negative staining of myosin molecules. *J. Mol. Biol.*, **184**, 535–542.
49. Mindell,J.A. and Grigorieff,N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.*, **142**, 334–347.
50. Ludtke,S.J., Baldwin,P.R. and Chiu,W. (1999) EMAN: Semiautomated software for high-resolution single-particle reconstructions. *J. Struct. Biol.*, **128**, 82–97.
51. van Heel,M., Harauz,G., Orlova,E.V., Schmidt,R. and Schatz,M. (1996) A new generation of the IMAGIC image processing system. *J. Struct. Biol.*, **116**, 17–24.
52. van heel,M. (1987) Angular reconstitution – a posteriori assignment of projection directions for 3-D reconstruction. *Ultramicroscopy*, **21**, 111–123.
53. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF chimera – a visualization system for exploratory research and analysis. *J. Comp. Chem.*, **25**, 1605–1612.
54. Goedecke,K., Pignot,M., Goody,R.S., Scheidig,A.J. and Weinhold,E. (2001) Structure of the N6-adenine DNA methyltransferase M.TaqI in complex with DNA and a cofactor analog. *Nat. Struct. Biol.*, **8**, 121–125.
55. Simons,K.T., Kooperberg,C., Huang,E. and Baker,D. (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J. Mol. Biol.*, **268**, 209–225.
56. Dominguez,C., Boelens,R. and Bonvin,A. (2003) HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, **125**, 1731–1737.
57. Wriggers,W., Milligan,R.A. and McCammon,J.A. (1999) Situs: A package for docking crystal structures into low-resolution maps from electron microscopy. *J. Struct. Biol.*, **125**, 185–195.
58. 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.
59. Kelleher,J.E., Daniel,A.S. and Murray,N.E. (1991) Mutations that confer de novo activity upon a maintenance methyltransferase. *J. Mol. Biol.*, **221**, 431–440.
60. Cooper,L.P. and Dryden,D.T.F. (1994) The domains of a Type-I DNA methyltransferase – interactions and role in recognition of DNA methylation. *J. Mol. Biol.*, **236**, 1011–1021.
61. Powell,L.M., Lejeune,E., Hussain,F.S., Cronshaw,A.D., Kelly,S.M., Price,N.C. and Dryden,D.T.F. (2003) Assembly of EcoKI DNA methyltransferase requires the C-terminal region of the HsdM modification subunit. *Biophys. Chem.*, **103**, 129–137.
62. O'Neill,M., Dryden,D.T.F. and Murray,N.E. (1998) Localization of a protein-DNA interface by random mutagenesis. *EMBO J.*, **17**, 7118–7127.
63. O'Neill,M., Powell,L.M. and Murray,N.E. (2001) Target recognition by EcoKI: The recognition domain is robust and restriction-deficiency commonly results from the proteolytic control of enzyme activity. *J. Mol. Biol.*, **307**, 951–963.
64. Sturrock,S.S. and Dryden,D.T.F. (1997) A prediction of the amino acids and structures involved in DNA recognition by type I DNA restriction and modification enzymes. *Nucleic Acids Res.*, **25**, 3408–3414.
65. Chen,A., Powell,L.M., Dryden,D.T.F., Murray,N.E. and Brown,T. (1995) Tyrosine-27 of the specificity polypeptide of EcoKI can be UV cross-linked to a bromodeoxyuridine-substituted DNA target sequence. *Nucleic Acids Res.*, **23**, 1177–1183.
66. Willcock,D.F., Dryden,D.T.F. and Murray,N.E. (1994) A mutational analysis of the 2 motifs common to adenine methyltransferases. *EMBO J.*, **13**, 3902–3908.
67. Taylor,I.A., Davis,K.G., Watts,D. and Kneale,G.G. (1994) DNA-binding induces a major structural transition in a type I methyltransferase. *EMBO J.*, **13**, 5772–5778.