An Mrr-family nuclease motif in the single polypeptide restriction–modification enzyme LlaGI

Rachel M. Smith[1], Jytte Josephsen[2] and Mark D. Szczelkun[1,*]

[1]DNA–Protein Interactions Unit, Department of Biochemistry, School of Medical Sciences, University of Bristol, Bristol, BS8 1TD, UK and [2]Department of Dairy and Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark

ABSTRACT

Bioinformatic analysis of the putative nuclease domain of the single polypeptide restriction–modification enzyme LlaGI reveals amino acid motifs characteristic of the *Escherichia coli* methylated DNA-specific Mrr endonuclease. Using mutagenesis, we examined the role of the conserved residues in both DNA translocation and cleavage. Mutations in those residues predicted to play a role in DNA hydrolysis produced enzymes that could translocate on DNA but were either unable to cleave the polynucleotide track or had reduced nuclease activity. Cleavage by LlaGI is not targeted to methylated DNA, suggesting that the conserved motifs in the Mrr domain are a conventional sub-family of the PD-(D/E)XK superfamily of DNA nucleases.

INTRODUCTION

The nucleolytic cleavage of DNA is central to a great many biological events and a wide diversity of enzyme folds and motifs have evolved to undertake this task (1–9). Many of these nuclease domains are encountered in the bacterial and archaeal restriction–modification (RM) enzymes (10), studies of which have played a key role in understanding nuclease mechanism (11). Amongst the different classes of RM enzyme, the Type II RM enzymes show the widest diversity with at least five unrelated structural folds identified (5): the PD-(D/E)XK superfamily (also known as Superfamily I), which accounts for the majority of RM enzymes; the HNH superfamily; the phospholipase D (PLD) superfamily of phosphodiesterases; the GIY-YIG superfamily; and a novel 'half-pipe' (HP) fold. Whilst the PD-(D/E)XK,

HNH and GIY-YIG superfamilies are all dependent on divalent metals, the PLD and HP families are not (12,13). The Type I and III RM nuclease domains are less diverse, reflecting in part the classification of these enzymes which groups them principally on the basis of similar domain/subunit architecture: the Type I enzymes contain variations on the RecB nuclease-family (14,15) and the Type III enzymes contain variations of the Archaeal Holliday Junction Resolvase (AHJR) family (16), both being sub-families of Superfamily I (Ref. 1). All Type I and III nucleases are believed to be Mg^{2+}-dependent. ATP hydrolysis is required for DNA cleavage but this is linked to the structurally distinct helicase domains in these proteins (17,18).

The LlaGI RM enzyme is a single polypeptide enzyme that comprises four domains (Figure 1A) (19): a nuclease; a helicase; an adenine methyltransferase (MTase); and target recognition domain (TRD). DNA cleavage occurs at a random non-specific site and requires the interaction of at least two LlaGI enzymes bound at distant DNA sites (20). Long-range interaction is facilitated by a 1-D loop translocation mechanism (21). Previous bioinformatic analysis suggested that LlaGI contains a putative classical PD-(D/E)XK motif typical of the nuclease domain of classical Type I enzymes (19). We show here that LlaGI and related Type ISP enzymes actually contain a nuclease domain most related to a sub-group of the AHJR family as exemplified by the *Escherichia coli* methylated DNA-specific Mrr endonuclease.

MATERIALS AND METHODS

LlaGI sequence analysis

Amino acid sequences were obtained from REBASE (10) and were selected from the list of Type IIG enzymes on the basis of the presence of helicase motifs and on their size (all sequences >150 kDa being selected). (Note: all

the LlaGI-like enzymes are currently categorized in REBASE as being Type IIG due to their RM gene arrangement). Of these, two distinct groupings were noted on the basis of characteristic Walker A box sequences in their helicase domains: those with a GTGKT sequence (56 members including LlaGI) and those with an RFGKT sequence (14 members) (see Table 1 in ref. 20). The RFGKT group were not used further in this analysis and will be considered elsewhere. Sequence alignments of 54 of the GTGKT group were carried out using ClustalW [(22); http://www.ebi.ac.uk/Tools/clustalw/index.html; two sequences were discarded because of extensive N-terminal deletions]. Manual realignments were made using Jalview (23) before the final alignments were submitted to JPred (24) to predict average secondary amino acid structure. The LlaGI sequence was also submitted separately to JPred. Consensus sequences were calculated from the full alignment at: http://coot.embl.de/Alignment//consensus.html.

Mutagenesis and protein purification

Plasmids pRSFLlaGI-E38A, -D74A, -D78A, -E82A, -Q92A, -K94A, -K102A and -E104A were generated by QuikChange mutagenesis using as the template pRSFLlaGI (Ref. 20). The oligonucleotide sequences used are available upon request. The LlaGI mutants were purified following the wild-type protocol as described in Smith *et al* (20). The protein concentration was determined by the Bradford dye-binding procedure (Bio-Rad Protein Assay, Bio-Rad, Hemel Hempstead, UK) using bovine serum albumin (BSA) as a standard (20). Each mutant was then compared with the wild type by densitometry following SDS-gel electrophoresis and staining with colloidal protein stain.

DNA translocation and cleavage assays

Translocation and cleavage assays were carried out as described in the accompanying papers (20,21).

RESULTS AND DISCUSSION

Identification of an Mrr nuclease motif in LlaGI

To identify the nuclease fold in LlaGI, we aligned the amino acid sequences of 54 LlaGI-related enzymes (a subset of which are shown in Figure 1B) (see 'Materials and Methods' section). This identified a three motif consensus sequence and predicted secondary structure most similar to that found in the methylation-specific endonuclease Mrr (Figure 1C) (8,9) which is defined as part of a larger AHJR subfamily (1). This classification of LlaGI-like enzymes into the Mrr-family has been noted previously on the basis of randomized alignment searches (1,8,9). It contrasts with the classification of the LlaGI nuclease domain made on the basis of alignment with Type I and III enzymes which identified a putative classical PD-(D/E)XK motif (Figure 1D) (19). Whilst this motif is found in enzymes with close similarity to LlaGI, it is not conserved across the whole group. Moreover, the position of the PD-(D/E)XK motif is not in the correct register

with the predicted secondary structure elements that make up the nuclease α−ββα core (Figure 1D) (2). The register of the putative Mrr motifs in LlaGI matches more closely the predictions for *E. coli* Mrr made on the basis of the homology models (Figure 1C) (8,9). Both motifs of LlaGI are tested below.

What is the role of each of the residues in the Mrr motif? Surprisingly little is known of the molecular mechanism of Mrr, and the only mutagenesis studies of the nuclease domain have used randomised *in vivo* screening (9,25). The residues identified by these screens coincide with a subset of the key residues identified in the alignments, as well as more weakly conserved features that are likely to have structural rather than catalytic roles (Figure 1C and E) (9). Molecular models based on sequence alignments, homology modelling and threading onto BglII, suggest that Mrr has the characteristic α−ββα fold that arranges the nuclease motifs into a metal ion–DNA binding pocket (Figure 1E) (8,9). We can therefore use knowledge gained from the study of other PD-(D/E)XK family RM enzymes to predict the roles of the Mrr motifs.

In the prototypical version of the PD-(D/E)XK fold, two acidic residues and a lysine from Motifs II and III form a catalytic triad involved in divalent metal ion binding and/or catalysis (11). The aspartate residue of Motif II (corresponding to D78 of LlaGI, Figure 1B and D) is the only invariant residue across Superfamily I (1). It is preceded by a hydrophobic residue (I202 in Mrr, Figure 1C and E; V77 in LlaGI, Figure 1B and D, but more commonly Proline in Type II RM enzymes). LlaGI and the Mrr-related enzymes contain a supplementary aspartate that is 100% conserved, four amino acids N-terminal to the catalytic aspartate and linked via a conserved glycine (Figure 1A–C) (8,9). This residue does not superimpose with any other conserved residue within the superfamily (1), and in the structural model appears to point away from the catalytic triad (D199 in Mrr, Figure 1C and E, D74 in LlaGI, Figure 1B and D) (9). The lysine residue of Motif III (K221 in Mrr, Figure 1C and E; K94 in LlaGI, Figure 1B and D) corresponds to the catalytic residue typical of Superfamily I [although it is exchanged with aspartate or glutamate in some Type II RM enzymes; (11)]. However, the signature Motif III glutamate acidic residue is replaced in all LlaGI- and Mrr-like enzymes with a glutamine (Q219 in Mrr, Figure 1C and E, Q92 in LlaGI, Figure 1B and D).

A number of different non-canonical versions of the PD-(D/E)XK fold can be categorized by the replacement of the acidic residue in Motif III (11,26–34). In these cases, the 'missing' residue is substituted by the presence of alternative carboxylate residues from elsewhere in the primary sequence that contribute to the catalytic/metal binding site (2,11,29). This highlights that spatial conservation of residues is important rather than sequence conservation *per se* (29). In Type II RM enzymes Cfr10I, NgoMIV, and Bse634I, where a serine or glycine replaces the conserved acidic residue, a distal glutamate found in the α-helix C-terminal to Motif III plays the substituting role (26–28,30). In Type II RM enzymes SsoII and PspGI, where a serine or threonine replaces the conserved acidic
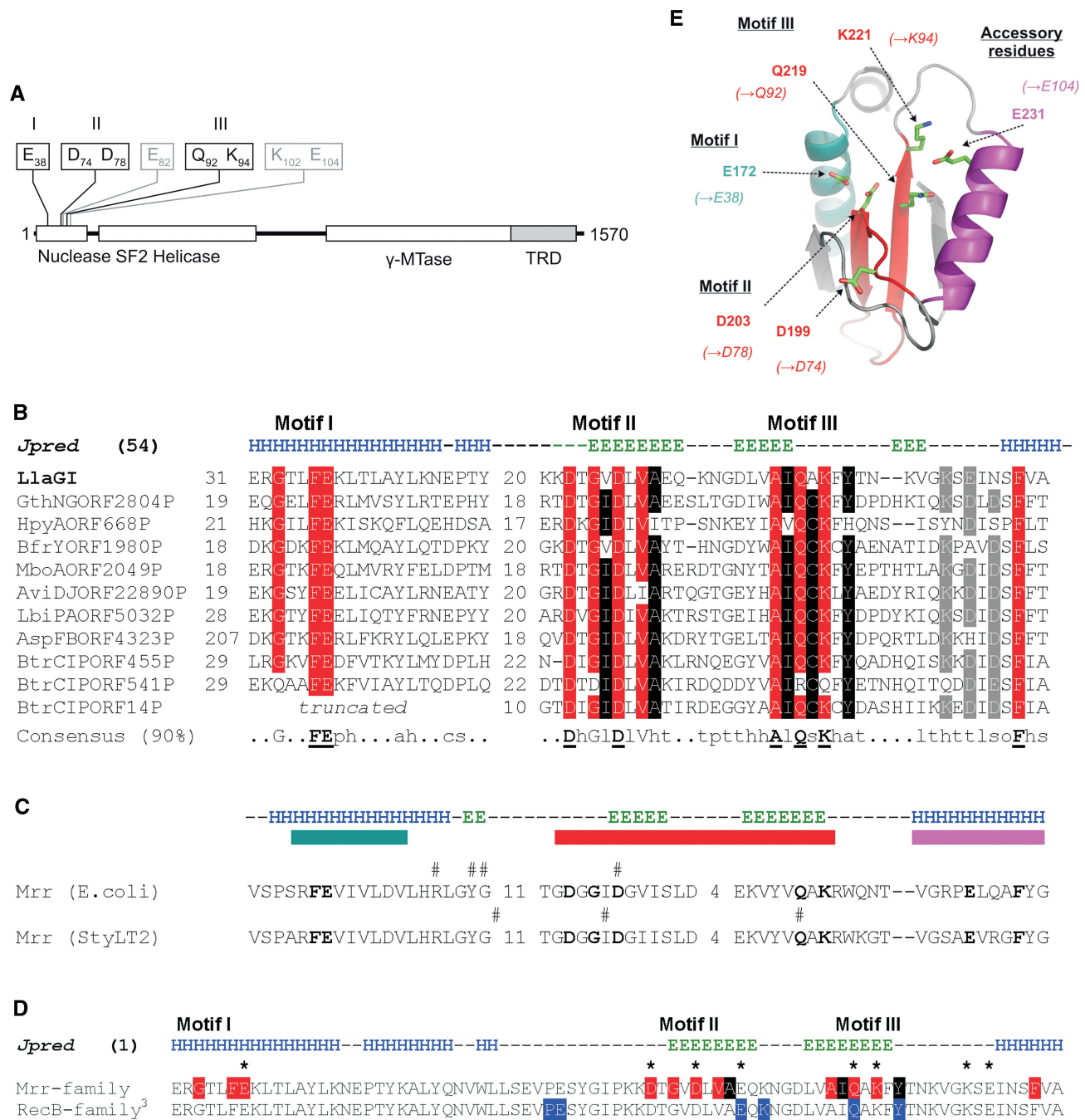
Figure 1. Mrr nuclease motifs in the LlaGI RM enzyme. (**A**) Cartoon of LlaGI with domains indicated. The locations of the Mrr nuclease Motifs I, II and III are shown in Black. Additional residues tested here are shown in grey. SF2 is Superfamily 2. (**B**) Multiple alignment of amino acid sequences of the LlaGI-like family (11/54 sequences shown). Alignments were carried out using ClustalW (22) and secondary structure predictions were made by JPred (24) using the full alignment. Consensus sequences were also calculated using the full alignments. For further details, please refer to 'Materials and Methods' section. Each protein is identified by the gene name as given in REBASE (10). Numbers to the left indicate the number of N-terminal residues truncated. Poorly conserved regions are replaced by numbers indicating the residues omitted. Red highlighting indicates sequences with >95% identity and black highlighting indicates sequences with >85% identity. Grey highlighting indicates sequences with 70–81% identity that could be accessory catalytic residues. The predicted secondary structure is shown above the alignments with H for an α-helix and E for a β-strand. The ≥90% consensus is shown below the alignment with the following key: *t*, turnlike (ACDEGHKNQRST); *h*, hydrophobic (ACFGHIKLMRTVWY); *l*, aliphatic (ILV); *a*, aromatic (FHWY); *p*, polar (CDEHKNQRST); *c*, charged (DEHKR); *o*, alcohol (ST); and, *s*, small (ACDGNPSTV). Residues in bold and underlined are 100% conserved. Three of the twelve LlaGI-like sequences from Bartonella tribocorum CIP 105476 are shown as examples of inactivation of the nuclease motif by mutation of Motif III (BtrCIPORF541P) or truncation of Motif I (BtrCIPORF14P). (**C**) Mrr sequences from *E. coli* and *S. typhimurium* LT2. The synaptomorphic Mrr residues are shown in bold. The secondary structure was calculated from the *E. coli* Mrr model (9), and the numbering is shown as in (B). The green, red and purple boxes refer to the colouring of secondary structure elements in (E). Residues identified from *in vivo* screens as important (9,25) are indicated by hash marks. (**D**) LlaGI sequences showing the Mrr motifs identified here (upper sequence, highlighted in red and black) and the PD-(D/E)XK motifs identified previously (19) (lower sequence, highlighted in blue). Mutations made in this study are marked with an asterisk. The secondary structure was predicted for the LlaGI sequence alone using JPred (24) and is labelled as in (B). (**E**) Amino acids 160–250 of the structural model of *E. coli* Mrr (ftp://genesilico.pl/iamb/models/Mrr/) (9). Amino acids in the brackets refer to the equivalent residues in LlaGI. Secondary structure colouring refers to (C).

residue in Motif III, two potential glutamates are present in the adjacent α-helix, each of which support cleavage with both required for maximal activity (31,32). In the homology model structure of Mrr (Figure 1E), E231 from the α-helix C-terminal to Motif III points towards the D203 and K221 residues of Motifs II and III and it has been suggested this replaces the 'missing' carboxylate at residue Q219 (9). Whilst this residue shows some conservation amongst enzymes most closely related to *E. coli* Mrr (8,9), the LlaGI-family show much greater variability in this region (Figure 1B), apart from a strongly conserved phenylalanine that most likely plays a structural role (F108 in LlaGI, Figure 1B). LlaGI contains a single glutamate (E104) similarly placed to the Mrr E231 residue. A nearby lysine residue (K102) is also present in 81% of the LlaGI-like sequences. Either or both of these residues may play a role in the catalytic site similar to the suggestions for the Type II enzymes described above (29).

In the PD-(D/E)XK fold, the glutamate at Motif I is strongly conserved. However, the exact function of this acidic residue is still a matter of debate, with suggested roles, for example, in stabilizing metal ion binding (1,11), in acting as a general base in catalysis (35) or in facilitating conformational transitions on-pathway to the coordination of the catalytic triad (36). Mutations at this residue in some enzymes knock out nuclease activity to a similar extent as mutations in the catalytic triad (37,38). In contrast, in other enzymes comparable mutations cause only a partial reduction in activity, suggesting a more peripheral role (39–41). In many of the Superfamily I enzymes the helix containing Motif I also has a small residue (G or P) N-terminal to the conserved glutamate that most likely allows flexibility/a kink that is structurally important (1). In the Mrr homology model, E172 from Motif I is located close to D203 of Motif II, consistent with the roles described above (9). The equivalent residue in LlaGI is E38 (Figure 1B). An adjacent phenylalanine residue is conserved across all Mrr-like enzymes and is likely to play a structural role (1,8,9). Glycine and tyrosine residues at the C-terminus of the Motif I helix have been identified as important by *in vivo* screens of *E. coli* and *Salmonella typhimurium LT2* Mrr (9,25). However, equivalent residues are not conserved across the LlaGI-like enzymes.

Mutagenesis of the LlaGI nuclease motifs and analysis of DNA translocation activity

To find out which of the conserved residues of LlaGI contribute to nuclease activity, alanine substitutions were made in Mrr Motif I (E38A), Mrr Motif II (D74A and D78A), putative PD-(D/E)XK Motif III (E82A), Mrr Motif III (Q92A and K94A) and the putative accessory residues (K102A and E104A) (Figure 1D). The nuclease residues were mutated in pRSFLlaGI using the QuikChange mutagenesis protocol (see 'Materials and Methods' section). The wild type and mutant LlaGI enzymes were purified as described in the accompanying paper (Figure 2) (20); the mutants behaved as wild type during expression and purification (data not shown), with the exception of K102A which gave a reduced yield.
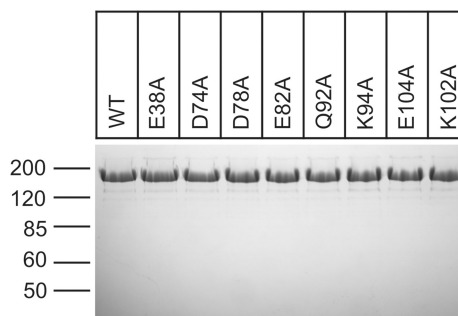

Figure 2. Purification of LlaGI mutants. SDS polyacrylamide gel showing 1.5 μg of LlaGI wild type and mutants following the gel filtration purification step. See 'Materials and Methods' section for full details.

We first compared the ability of the mutant LlaGI enzymes to translocate on linear DNA. This was carried out using a triplex displacement assay as described in Smith *et al.*, (21) with a saturating protein concentration. Figure 3A shows the triplex displacement profiles for the 1977-bp spacing for wild-type LlaGI and all eight nuclease mutants. The profiles are those expected for dsDNA translocation, with a lag phase and an exponential displacement phase. Displacement profiles were obtained for a further three triplex spacings and the lag times calculated (see 'Materials and Methods' section) (21,42). From these we obtained the linear relationship between distance and lag time from which we could calculate translocation rates (Figure 3B). All of the mutant LlaGI enzymes showed translocation kinetics consistent with the wild-type enzyme, indicating that the mutations had not had a detrimental effect on gross structure and that the nuclease domain is unlikely to play a direct role in translocation. However, we have shown previously for the Type I RM enzyme EcoR124I that whilst nuclease mutations can have no apparent effect on translocation as judged by the ensemble triplex assay, increased static disorder in the mutant enzyme population can cause stalling events that are only observed in single molecule assays (43). Examination of the exponential phases of the triplex displacement profiles showed variations that could be interpreted as changes in the distribution of the enzyme population during translocation (42). However, the differences were not consistent between different DNA spacings and did not show any relationship with subsequent kinetic data. That is, some mutants had wild-type-like cleavage properties despite having slower second phase exponential profiles on some DNA. We therefore suggest that the differences in the profiles most likely represent experimental error and that all mutants have wild-type-like translocation activity under the saturating enzyme concentrations used here.

DNA cleavage properties of the LlaGI nuclease domain mutants

In the subsequent assays, we tested DNA cleavage activity using a DNA substrate containing two indirectly repeated sites [pHH-3, (20)]. Figure 4 shows the extent
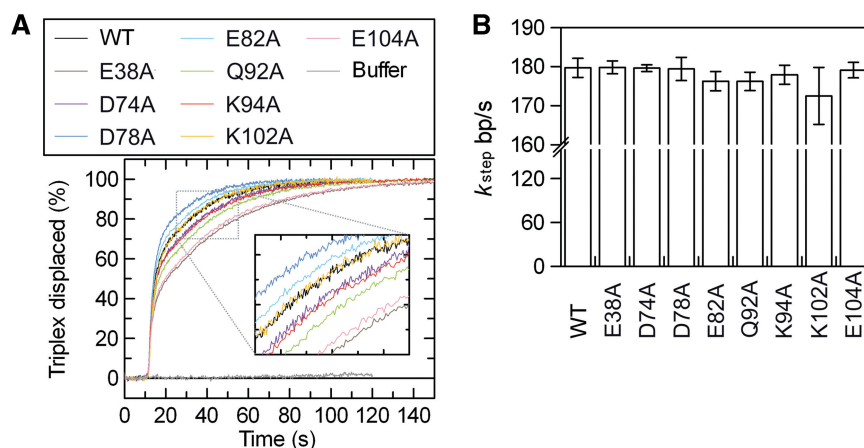
Figure 3. DNA translocation by the nuclease mutants of LlaGI. (**A**) Representative triplex displacement profiles. DNA (RMA03F at 1 nM) with the 1977-bp spacing triplex bound (0.5 nM) were pre-incubated with LlaGI (10 nM) and translocation initiated by mixing with 4 mM ATP. (Inset) Expanded portion of the exponential phase as indicated. Profiles at three other triplex spacings were also collected for each protein (data not shown). See text for further details. (**B**) Translocation rates (k_{step}) determined from the lag times of the triplex displacement profiles [(A) and data not shown]. Error bars represent the standard error of the linear fit of the lag times to distance. Reactions were carried out and analysed as in the accompanying paper (21).
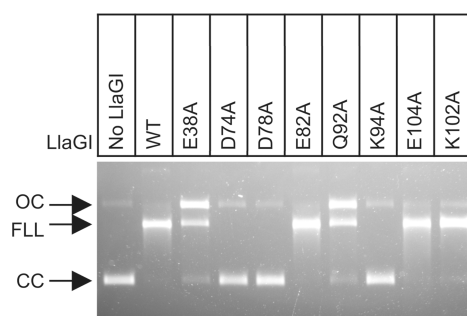

Figure 4. Summary of the DNA cleavage properties of LlaGI and the nuclease mutants; 4 nM pHH-3 was incubated with 200 nM LlaGI (or 100 nM $LlaGI_{K102A}$) as indicated, plus 4 mM ATP for 10 min. DNA substrates (CC, covalently closed circular DNA), intermediates (OC, open circle) and products (FLL, full-length linear) were separated by agarose gel electrophoresis.

of DNA cleavage after extensive incubation with an excess of each LlaGI enzyme. On the basis that, under these conditions, the wild-type enzyme cleaves close to 100% of the substrate to produce >95% linear DNA and that all proteins are active for translocation, we can identify three different enzyme groups with either: (i) full nuclease activity (>95% linear produced), including wild type (Figure 4, lane 2), E82A (lane 6), E104A (lane 9) and K102A (lane 10); (ii) no nuclease activity, including D74A (lane 4), D78A (lane 5) and K94A (lane 8); or, (iii) partial/reduced nuclease activity (significant amounts of nicked DNA remaining at the end of the reaction), including E38A (lane 3) and Q92A (lane 7). To assess further the endonuclease mechanism of the active and partially active mutants, we measured the extent of DNA cleavage at the reaction 'end point' as a function of the LlaGI:DNA molecular ratio and measured the rate of DNA cleavage at a saturating concentration of LlaGI.

LlaGI nuclease mutants with wild-type activity. The 'classical' PD-(D/E)XK motif identified by alignment of LlaGI with representative Type I and III enzymes (Figure 1D) (19) was tested by mutating residue E82 that corresponds to the archetypical essential acidic residue of Motif III. Both the cleavage titration (Figure 5A, upper panel) and cleavage time-course (lower panel) are characteristic of wild-type activity (compare to Figure 6A and D in ref. 20), indicating that it is unlikely that E82 plays any role in the nuclease active site. This is consistent with the full alignment data which does not identify the PD-(D/E)XK motif as conserved (Figure 1B) or correctly aligned with the predicted secondary structure elements (Figure 1D). The correct identification of the nuclease motifs is relatively challenging as the sequence similarity between active sites is relatively low. A similar case of mistaken identity was also made with the homing endonuclease I-PpoI, which displays a sequence resembling a PD-(D/E)XK motif but which actually plays no role, the enzyme instead using an unrelated His–Cys box motif (44,45).

A residue that could act as a catalytic replacement for the missing acidic residue in Motif III is E104. However, E104A had titration and kinetic profiles very similar to the wild-type protein and distinct from those of Q92A or E38A (Figures 4 and 5B). We therefore suggest that E104 plays little or no role in the LlaGI cleavage activity. The partially conserved lysine residue K102 in the α-helix C-terminal to Motif III was also tested and shown to have wild type activity (Figure 4; data not shown). An equivalent conserved lysine residue is not observed in the *bona fide* Mrr protein alignment (8,9) and this region shows virtually no sequence conservation across Superfamily I (1). Although individual nuclease sub-families are likely to have important synaptomorphic residues, in this study we could not find any specific role for K102 or E104 in LlaGI activity.
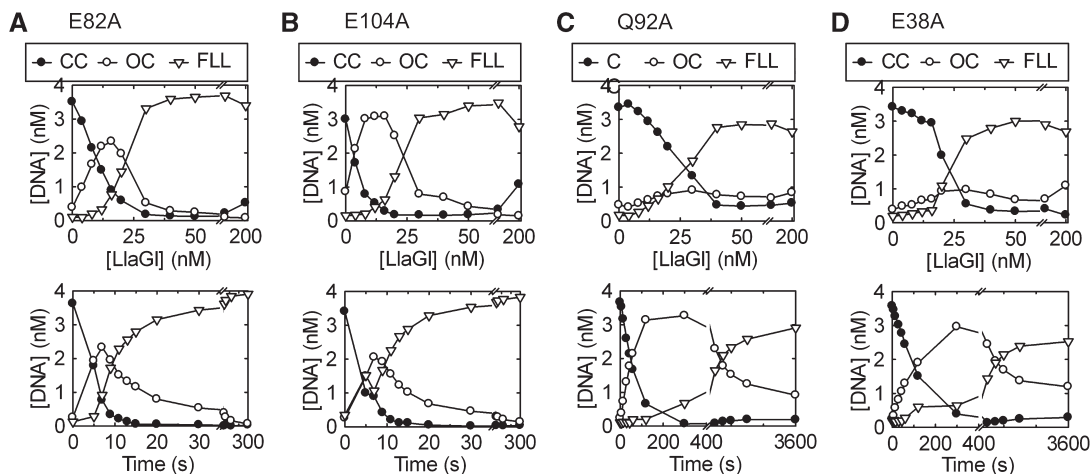
Figure 5. Stoichiometry and rate of DNA cleavage by LlaGI mutants. Points are the mean of at least two repeat experiments. Note that where the FLL product becomes smeared due to end-processing, this can cause an overestimation of the CC band which overlaps with the smear. (Upper panels) pHH-3 was incubated with the LlaGI mutants at the concentrations indicated for 1 h. (Lower panels) Time course of cleavage of pHH-3 with a saturating concentration (200 nM) of the LlaGI mutants. Reactions were carried out and analysed as in the accompanying paper (20).

LlaGI nuclease mutants with no nuclease activity. The principal residues of Motifs II and III that have equivalents across Superfamily I are D78 and K94. Alanine substitutions at either position were completely inactive over one hour incubation, irrespective of the protein concentration used (Figure 4, data not shown). These observations are consistent with the important metal ion binding and catalytic roles identified for these residues in Superfamily I proteins (11). It also corroborates the *in vivo* screening data for Mrr that identified the I-D residues of Motif III (V-D in LlaGI) as being critical (Figure 1C) (25).

Strikingly, we also found that alanine substitution at D74 in Motif II was completely inactive (Figure 4, data not shown). This conserved residue is a synaptomorphic marker of the Mrr enzymes that does not appear in any other sub-group of Superfamily I (1). The structural models for Mrr have not identified a direct role for this residue in catalysis (Figure 1E) (8,9). Since the LlaGI cleavage pathway is complex, preceded by many different steps [binding, loop initiation, ATPase, translocation, collision, activation, etc, (21)], we cannot rule out that the nuclease domain retains some weak activity but that the collision complex collapses before a strand break can occur.

LlaGI nuclease mutants with partial nuclease activity. The Q92A mutation in Motif III showed reduced activity in both the titration and kinetic analysis (Figure 5C). An elevated concentration of Q92A was still unable to completely cleave the CC DNA compared to WT enzyme (Figure 5C, upper panel). Very little nicked DNA intermediate accumulated during the titrations. During the DNA cleavage time course under saturated conditions, a first strand cleavage rate that was more than 10-fold slower than wild type was followed by a significantly slower second strand cleavage resulting in the accumulation of nicked DNA to >80% of the total DNA (Figure 5C, lower panel). Both these observations

may reflect a weaker association of the nuclease active site with the DNA. Our data suggests that whilst Q92 does not play a critical catalytic role, it may have a secondary structural role in stabilizing the active site on the DNA. A similar role has been suggested for a conserved glutamine that is part of the RecB-family nuclease fold of classical multi-subunit Type I RM enzymes (15).

The E38A mutation in Motif I also showed reduced activity with titration and kinetic profiles very similar to Q92A (Figure 5D). Again the data are consistent with a non-critical yet important secondary role in the active site (11). Partial nuclease activity of Motif I mutants has also been seen with Type II RM enzymes [e.g. E45A of EcoRV (39,40) and E55A of PvuII (41)], reflecting this secondary role. We can however rule out the glutamate of Motif I as being the replacement residue in LlaGI for the missing acidic residue of Motif III (34).

CONCLUSIONS

The *in vitro* analysis of the LlaGI nuclease mutants is consistent with the bioinformatic analysis; i.e. residues in the Mrr-family Motifs I, II and III that are >90% conserved are important for cleavage activity, with D74, D78 and K94 being critical. The changes in DNA cleavage activity cannot be put down to inefficient translocation as all the mutants had triplex displacement properties similar to wild type. We cannot, however, state whether these mutants had altered DNA-binding properties as our assays were all undertaken using saturating enzyme concentrations. The previously identified PD-(D/E)XK motif (19), was shown to be both non-essential and not conserved across the LlaGI-like enzyme family. Previous studies of Type II RM enzymes with a (D/E)(S/T/A/G) substitution at Motif III have identified putative replacement acidic residues in the α-helical region immediately C-terminal to Motif III (26–34). The structural model of Mrr predicts that a similar residue in a similar location

plays a similar role (E231, Figure 1E) (8,9). However, we were unable to identify a residue in this structural region that could play a critical role in the cleavage mechanism of LlaGI. This suggests that either: (i) A currently unidentified residue from elsewhere in the sequence plays an important role; or (ii) the Motif I, II and III residues are sufficient for activity. It is notable that our study highlights D74 as absolutely required for catalysis. This Motif II residue is 100% conserved across Mrr-family enzymes but is not seen in any other sub-group of Superfamily I (1). It may be that the secondary aspartate of Motif II supplements for the glutamine in Motif III and forms the catalytic triad (in apparent contradiction with the Mrr homology models).

FUNDING

Conflict of interest statement. None declared.

REFERENCES

1. Aravind,L., Makarova,K.S. and Koonin,E.V. (2000) SURVEY AND SUMMARY: Holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.*, **28**, 3417–3432.
2. Bujnicki,J.M. (2004) Molecular phylogenetics of restriction endonucleases in restriction enzymes. In Pingound,A. (ed.), *Nucleic Acids and Molecular Biology*, Vol. 14. Springer, Germany, pp. 63–93.
3. Bujnicki,J.M. (2003) Crystallographic and bioinformatic studies on restriction endonucleases: inference of evolutionary relationships in the "midnight zone" of homology. *Curr. Protein. Pept. Sci.*, **4**, 327–337.
4. Kosinski,J., Feder,M. and Bujnicki,J.M. (2005) The PD-(D/E)XK superfamily revisited: identification of new members among proteins involved in DNA metabolism and functional predictions for domains of (hitherto) unknown function. *BMC Bioinformatics*, **6**, 172.
5. Orlowski,J. and Bujnicki,J.M. (2008) Structural and evolutionary classification of Type II restriction enzymes based on theoretical and experimental analyses. *Nucleic Acids Res.*, **36**, 3552–3569.
6. Knizewski,L., Kinch,L.N., Grishin,N.V., Rychlewski,L. and Ginalski,K. (2007) Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC Struct. Biol.*, **7**, 40.
7. Dunin-Horkawicz,S., Feder,M. and Bujnicki,J.M. (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics*, **7**, 98.
8. Bujnicki,J.M. and Rychlewski,L. (2001) Identification of a PD-(D/E)XK-like domain with a novel configuration of the endonuclease active site in the methyl-directed restriction enzyme Mrr and its homologs. *Gene*, **267**, 183–191.
9. Orlowski,J., Mebrhatu,M.T., Michiels,C.W., Bujnicki,J.M. and Aertsen,A. (2008) Mutational analysis and a structural model of methyl-directed restriction enzyme Mrr. *Biochem. Biophys. Res. Commun.*, **377**, 862–866.
10. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2007) REBASE–enzymes and genes for DNA restriction and modification. *Nucleic Acids Res.*, **35**, D269–D270.
11. Galburt,E.A. and Stoddard,B.L. (2002) Catalytic mechanisms of restriction and homing endonucleases. *Biochemistry*, **41**, 13851–13860.
12. Zaremba,M., Urbanke,C., Halford,S.E. and Siksnys,V. (2004) Generation of the BfiI restriction endonuclease from the fusion of a DNA recognition domain to a non-specific nuclease from the phospholipase D superfamily. *J. Mol. Biol.*, **336**, 81–92.
13. Miyazono,K., Watanabe,M., Kosinski,J., Ishikawa,K., Kamo,M., Sawasaki,T., Nagata,K., Bujnicki,J.M., Endo,Y. and Tanokura,M. (2007) Novel protein fold discovered in the PabI family of restriction enzymes. *Nucleic Acids Res.*, **35**, 1908–1918.
14. 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 EcoR124I: biophysical characterisation and structural modelling. *J. Mol. Biol.*, **376**, 438–452.
15. Sisáková,E., Stanley,L.K., Weiserová,M. and Szczelkun,M.D. (2008) A RecB-family nuclease motif in the Type I restriction endonuclease EcoR124I. *Nucleic Acids Res.*, **36**, 3939–3949.
16. Janscak,P., Sandmeier,U., Szczelkun,M.D. and Bickle,T.A. (2001) Subunit assembly and mode of DNA cleavage of the type III restriction endonucleases EcoP1I and EcoP15I. *J. Mol. Biol.*, **306**, 417–431.
17. Gorbalenya,A.E. and Koonin,E.V. (1991) Endonuclease (R) subunits of type-I and type-III restriction-modification enzymes contain a helicase-like domain. *FEBS Lett.*, **291**, 277–281.
18. McClelland,S.E. and Szczelkun,M.D. (2004) Molecular motors that process DNA in restriction enzymes. In Pingound,A. (ed.), *Nucleic Acids and Molecular Biology*, Vol. 14. Springer, Germany, pp. 111–135.
19. Madsen,A. and Josephsen,J. (2001) The LlaGI restriction and modification system of Lactococcus lactis W10 consists of only one single polypeptide. *FEMS Microbiol. Lett.*, **200**, 91–96.
20. Smith,R.M., Diffin,F.M., Savery,N.J., Josephsen,J. and Szczelkun,M.D. (2009) DNA cleavage and methylation specificity of the single polypeptide restriction modification enzyme LlaGI.
21. Smith,R..M., Josephsen,J. and Szczelkun,M.D. (2009) The single polypeptide restriction modification enzyme LlaGI is a self-contained molecular motor that translocates DNA loops.
22. Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R. *et al.* (2007) ClustalW and ClustalX version 2. *Bioinformatics*, **23**, 2947–2948.
23. Waterhouse,A.M., Procter,J.B., Martin,D.M.A., Clamp,M. and Barton,G.J. (2009) Jalview Version 2 – a multiple sequence alignment editor and analysis workbench. *Bioinformatics*, doi: 10.1093/bioinformatics/btp033.
24. Cole,C., Barber,J.D. and Barton,G.J. (2008) The Jpred 3 secondary structure prediction server. *Nucleic Acids Res.*, **36**, W197–W201.
25. Aertsen,A., Tesfazgi Mebrhatu,M. and Michiels,C.W. (2008) Activation of the Salmonella typhimurium Mrr protein. *Biochem. Biophys. Res. Commun.*, **367**, 435–439.
26. Bozic,D., Grazulis,S., Siksnys,V. and Huber,R. (1996) Crystal structure of Citrobacter freundii restriction endonuclease Cfr10I at 2.15 A resolution. *J. Mol. Biol.*, **255**, 176–186.
27. Skirgaila,R., Grazulis,S., Bozic,D., Huber,R. and Siksnys,V. (1998) Structure-based redesign of the catalytic/metal binding site of Cfr10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. *J. Mol. Biol.*, **279**, 473–481.
28. Deibert,M., Grazulis,S., Sasnauskas,G., Siksnys,V. and Huber,R. (2000) Structure of the tetrameric restriction endonuclease NgoMIV in complex with cleaved DNA. *Nat. Struct. Biol.*, **7**, 792–799.
29. Tamulaitis,G., Solonin,A.S. and Siksnys,V. (2002) Alternative arrangements of catalytic residues at the active sites of restriction enzymes. *FEBS Lett.*, **518**, 17–22.
30. Grazulis,S., Deibert,M., Rimseliene,R., Skirgaila,R., Sasnauskas,G., Lagunavicius,A., Repin,V., Urbanke,C., Huber,R. and Siksnys,V. (2002) Crystal structure of the Bse634I restriction endonuclease: comparison of two enzymes recognizing the same DNA sequence. *Nucleic Acids Res.*, **30**, 876–885.
31. Pingoud,V., Kubareva,E., Stengel,G., Friedhoff,P., Bujnicki,J.M., Urbanke,C., Sudina,A. and Pingoud,A. (2002) Evolutionary relationship between different subgroups of restriction endonucleases. *J. Biol. Chem.*, **277**, 14306–14314.
32. Pingoud,V., Conzelmann,C., Kinzebach,S., Sudina,A., Metelev,V., Kubareva,E., Bujnicki,J.M., Lurz,R., Lüder,G., Xu,S.Y. *et al.* (2003) PspGI, a type II restriction endonuclease from the extreme thermophile Pyrococcus sp.: structural and functional studies to

investigate an evolutionary relationship with several mesophilic restriction enzymes. *J. Mol. Biol.*, **329**, 913–929.

33. Tamulaitis,G., Mucke,M. and Siksnys,V. (2006) Biochemical and mutational analysis of EcoRII functional domains reveals evolutionary links between restriction enzymes. *FEBS Lett.*, **580**, 1665–1671.

34. Dunten,P.W., Little,E.J., Gregory,M.T., Manohar,V.M., Dalton,M., Hough,D., Bitinaite,J. and Horton,N.C. (2008) The structure of SgrAI bound to DNA; recognition of an 8 base pair target. *Nucleic Acids Res.*, **36**, 5405–5416.

35. Stanford,N.P., Halford,S.E. and Baldwin,G.S. (1999) DNA cleavage by the EcoRV restriction endonuclease: pH dependence and proton transfers in catalysis. *J. Mol. Biol.*, **288**, 105–116.

36. Horton,N.C., Otey,C., Lusetti,S., Sam,M.D., Kohn,J., Martin,A.M., Ananthnarayan,V. and Perona,J.J. (2002) Electrostatic contributions to site specific DNA cleavage by EcoRV endonuclease. *Biochemistry*, **41**, 10754–10763.

37. Xu,S.Y. and Schildkraut,I. (1991) Isolation of BamHI variants with reduced cleavage activities. *J. Biol. Chem.*, **266**, 4425–4429.

38. Pawlak,S.D., Radlinska,M., Chmiel,A.A., Bujnicki,J.M. and Skowronek,K.J. (2005) Inference of relationships in the 'twilight zone' of homology using a combination of bioinformatics and site-directed mutagenesis: a case study of restriction endonucleases Bsp6I and PvuII. *Nucleic Acids Res.*, **33**, 661–671.

39. Selent,U., Rüter,T., Köhler,E., Liedtke,M., Thielking,V., Alves,J., Oelgeschläger,T., Wolfes,H., Peters,F. and Pingoud,A. (1992)

A site-directed mutagenesis study to identify amino acid residues involved in the catalytic function of the restriction endonuclease EcoRV. *Biochemistry*, **31**, 4808–4815.

40. Groll,D.H., Jeltsch,A., Selent,U. and Pingoud,A. (1997) Does the restriction endonuclease EcoRV employ a two-metal-Ion mechanism for DNA cleavage? *Biochemistry*, **36**, 11389–11401.

41. Nastri,H.G., Evans,P.D., Walker,I.H. and Riggs,P.D. (1997) Catalytic and DNA binding properties of PvuII restriction endonuclease mutants. *J. Biol. Chem.*, **272**, 25761–25767.

42. McClelland,S.E., Dryden,D.T. and Szczelkun,M.D. (2005) Continuous assays for DNA translocation using fluorescent triplex dissociation: application to type I restriction endonucleases. *J. Mol. Biol.*, **348**, 895–915.

43. Sisáková,E., Weiserová,M., Dekker,C., Seidel,R. and Szczelkun,M.D. (2008) The interrelationship of helicase and nuclease domains during DNA translocation by the molecular motor EcoR124I. *J. Mol. Biol.*, **384**, 1273–1286.

44. Wittmayer,P.K. and Raines,R.T. (1996) Substrate binding and turnover by the highly specific I-PpoI endonuclease. *Biochemistry*, **35**, 1076–1083.

45. Flick,K.E., Jurica,M.S., Monnat,R.J. Jr and Stoddard,B.L. (1998) DNA binding and cleavage by the nuclear intron-encoded homing endonuclease I-PpoI. *Nature*, **394**, 96–101.