3276–3287 *Nucleic Acids Research, 2016, Vol. 44, No. 7* *Published online 09 March 2016*
doi: 10.1093/nar/gkw134

Endonuclease domain of non-LTR retrotransposons: loss-of-function mutants and modeling of the R2Bm endonuclease

Aruna Govindaraju†, Jeremy D. Cortez†, Brad Reveal and Shawn M. Christensen*

Department of Biology, University of Texas at Arlington, Arlington, TX 76019-0498, USA

Received June 25, 2015; Revised February 22, 2016; Accepted February 23, 2016

ABSTRACT

Non-LTR retrotransposons are an important class of mobile elements that insert into host DNA by target-primed reverse transcription (TPRT). Non-LTR retrotransposons must bind to their mRNA, recognize and cleave their target DNA, and perform TPRT at the site of DNA cleavage. As DNA binding and cleavage are such central parts of the integration reaction, a better understanding of the endonuclease encoded by non-LTR retrotransposons is needed. This paper explores the R2 endonuclease domain from *Bombyx mori* using *in vitro* studies and *in silico* modeling. Mutations in conserved sequences located across the putative PD-(D/E)XK endonuclease domain reduced DNA cleavage, DNA binding and TPRT. A mutation at the beginning of the first α-helix of the modeled endonuclease obliterated DNA cleavage and greatly reduced DNA binding. It also reduced TPRT when tested on pre-cleaved DNA substrates. The catalytic K was located to a non-canonical position within the second α-helix. A mutation located after the fourth β-strand reduced DNA binding and cleavage. The motifs that showed impaired activity form an extensive basic region. The R2 biochemical and structural data are compared and contrasted with that of two other well characterized PD-(D/E)XK endonucleases, restriction endonucleases and archaeal Holliday junction resolvases.

INTRODUCTION

R2 is a widely distributed site-specific non-LTR retrotransposon that targets the host genome's 28S ribosomal RNA genes (1,2). R2 elements, like all non-LTR retrotransposons, replicate by inserting into the host's chromosomes using a process called target-primed reverse transcription (TPRT) (3–5). The element encoded endonuclease cleaves the host chromosome to generate a free 3′ OH. The element encoded reverse transcriptase uses the liberated 3′ OH to prime reverse transcription of the element RNA into DNA at the site of insertion (3–6). The protein encoded by the R2 element from *Bombyx mori,* R2Bm, is expressible in bacteria and purified components are readily obtainable—allowing TPRT to be extensively studied *in vitro*. Two R2Bm protein subunits have been shown to be involved in the TPRT reaction: one bound to the 5′ end of the element RNA and one bound to the 3′ end of the RNA (Figure 1B). The protein subunit bound to the element RNA's 3′ protein binding motif (PBM) interacts with target DNA upstream of the insertion site, cleaves the first (bottom) DNA strand and performs TPRT (5). The protein subunit bound to the 5′ PBM RNA, binds to target DNA downstream of the insertion site, cleaves the second (top) DNA strand and is thought to perform second strand synthesis (5,7,8). The upstream and downstream protein subunits bind and cleave different (i.e. non-palindromic) DNA sequences.

R2 elements are divided into subclades based upon sequence homology and number of amino-terminal zinc fingers (ZFs) coded for by the open reading frame (ORF). The two major subclades are the R2-A and R2-D clades. The R2-A clade elements have three amino-terminal ZFs, while the R2-D clade elements have a single ZF. R2 (and R2-like) elements have been found to target various genomic repeats in their host genomes (9–15). The R2 site is located near the middle of the 28S rDNA, the R9 site is near the beginning of the 28S rDNA and the R8 site is in the 18S rDNA (16–18). The ORF of R2Bm, an R2-D clade element, encodes a number of conserved motifs of known and unknown functions (Figure 1A). The ZF and Myb motifs are the major motifs used to bind a protein subunit to a DNA sequence located downstream of the insertion site (7,19,20). The protein motifs used to secure the upstream subunit to the target site remain uncharacterized. The basic region is involved in RNA binding (21). The reverse transcriptase domain (RT) performs the actual TPRT reaction (3,4,6). A cysteine-histidine rich motif (CCHC) of largely unknown function is down-

*To whom correspondence should be addressed. Tel: +1 817 272 0520; Fax: +1 817 272 9722; Email: shawnc@uta.edu
†These authors contributed equally to the paper as first authors.
Present address: Brad Reveal, NICHD, NIH, Bethesda, MD 20814, USA.

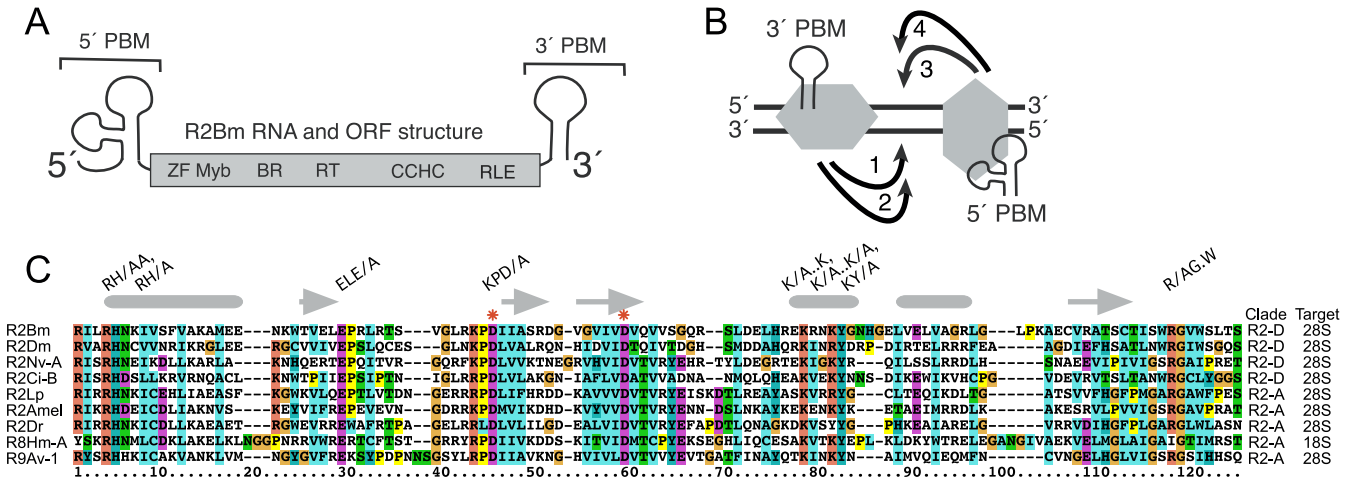

Figure 1. R2Bm structure. (**A**) R2Bm RNA and open reading frame (ORF) structure. The ORF of R2Bm encodes a number of conserved motifs of known and unknown function. Abbreviations: zinc finger (ZF), Myb (Myb), basic region (BR), reverse transcriptase domain (RT), a cysteine-histidine rich motif (CCHC) and a PD-(D/E)XK type restriction-like endonuclease (RLE). RNA motifs present in the 5′ and 3′ untranslated regions bind R2 protein and are marked as 5′ and 3′ protein binding motifs (PBMs), respectively. Brackets indicate the individual segments of the R2Bm RNA (i.e. the 5′ and 3′ PBM RNAs) that were used in this paper. The data for reactions containing the 3′ PBM are presented in the paper. The 5′ PBM RNA data is located in the Supplementary Data. (**B**) The R2 insertion mechanism is depicted: (1) DNA cleavage of the bottom strand, (2) TPRT, (3) DNA cleavage of the top strand and (4) second strand DNA synthesis. (**C**) Clustal alignment of the RLE domain of D-clade and A-clade R2 elements that target the R2, R8 and R9 sites. Abbreviations: *Bombyx mori* R2 (R2Bm) (52), *Drosophila melanogaster* R2 (R2Dm) (16), *Nasonia vitripennis* R2 (R2Nv-A) (53), *Ciona intestinalis* R2 (R2Ci-B) (54), *Limulus polyphemus* R2 (R2Lp) (55), *Apis mellifera* R2 (R2Amel) (2), *Danio rerio* R2 (R2Dr) (54), *Hydra magnipapillata* R8 (R8Hm-A) (17) and *Adineta vaga* R9 (R9Av-1) (18). The red asterisks indicate active site residues that were previously identified and characterized by mutation to alanine in R2Bm (25). The KPD/A mutation is used as a 'wild-type' control in Figures 2D, 3D, 4D and 5B (25). The R2Bm mutants generated for this study were RH/AA, RH/A, ELE/A, K/A..K/A, K/A..K, KY/A and R/AG.W. The rounded gray bars delineate the α-helices as determined by the 3D model presented in Figure 6. The gray arrows delineate the β-strands.

stream of the RT (22–24). Beyond the CCHC motif is a PD-(D/E) restriction endonuclease-like motif (RLE). R2 elements appear to encode an endonuclease variant of the PD-(D/E)XK motif (25). The D-(D/E) residues of the R2 endonuclease have been located. The lysine has not been discovered, and the overall size and structure of the R2 endonuclease is not known (25). As DNA cleavage is such a central part of the integration reaction, a better understanding of the endonuclease encoded by R2 elements is needed.

Endonucleases of PD-(D/E)XK family include most bacterial restriction enzymes, a number of DNA repair enzymes (e.g. MutH), certain group I homing intron endonucleases and certain Holliday junction resolvases (26,27). Members of the PD-(D/E)XK superfamily adopt a common structural core—a restriction endonuclease-like fold (terminology from SCOP2 database)—consisting of a four-stranded, mixed β-sheet flanked by two α-helices on both sides (αββββαβ topology) (26–30). Endonucleases of this family perform various biological roles, including their involvement in DNA repair, replication and even tRNA-intron splicing (26–28,31). The family members share little to no sequence or structural homology beyond the conserved PD-(D/E)XK residues and the restriction endonuclease-like fold. The restriction endonuclease-like fold often contains multiple insertions and extensions, making the core fold difficult to detect (26–29). The similarity of the R2 encoded endonuclease to restriction enzymes has been long noted, and until recently the best hit on structure prediction and threading programs like Phyre2 and HHPRED was the restriction enzyme Fok I, although it modeled only part of the fold (our unpublished data) (25). Like the restriction DNA endonuclease Fok I, the R2 protein is thought to use distinct DNA binding domain(s) coupled to a non-specific DNA endonuclease to cleave the non-palindromic DNA target. The recent top hits in the structural databases now include the Holliday junction resolvases E and C (Hje and Hjc) and allow more extensive modeling of the R2 endonuclease. An article by Mukha *et al*. was published reporting this similarity using the *Drosophila melanogaster* R2 element as a query sequence, corroborating our work-in-progress (32).

In the canonical PD-(D/E)XK arrangement of the catalytic residues within the restriction endonuclease-like fold (αββββαβ), the first aspartate residue is at or near the beginning of the second β-strand (26–29). The second aspartate (or glutamate, D/E) is in the third β-strand and the lysine is one residue away from the aspartate/glutamate in the third β-strand (26–29). These active site residues can sometimes vary with respect to their relative positions within the αββββαβ topology, but regardless of position, they form a conserved active site geometry. The active site residues play various catalytic roles, including coordination of up to three divalent metal ions (26,31,33). This paper maps, models and biochemically demonstrates that sequence motifs beyond the CCHC motif of R2 are part of the DNA endonuclease. The mutations in the RH (α-helix 1), K..KY (α-helix 2) and RG.W (carboxyl terminal to β-strand 4) motifs either knockout or reduce DNA cleavage activity. Mutations in the RH motif decrease DNA binding activity as well as DNA cleavage. Results for both the 3′ PBM RNA (presented in the paper) and the 5′ PBM RNA (provided in the Supplementary Data) *in vitro* reactions were collected. We compare and contrast our biochemical and structural data

for R2 with that of restriction endonucleases and archaeal Holliday junction resolvases.

MATERIALS AND METHODS

Nucleic acid preparation

Target DNA was generated by polymerase chain reaction (PCR) with ^{32}P end labeled primers as previously reported with the target primers listed in Supplementary Table 1 (20,34). The target DNA was 120 bp in length, with 70 bp of upstream DNA and 50 bp of downstream DNA, relative to the insertion dyad. DNA concentration was determined by ethidium bromide dot analysis where DNA is mixed with ethidium bromide (0.1μg/μl) for detection. Dilutions of Lambda DNA (Promega #21674203) were used to generate a standard curve. Pre-nicked target DNA was labeled on the bottom strand and nicked by reacting the target DNA with R2 protein under conditions that nick the bottom strand. The nicked DNA was purified by phenol–chloroform extraction followed by ethanol precipitation.

5′ PBM RNA and 3′ PBM RNA were made essentially as previously reported except that RNA was gel purified on a denaturing 5% polyacrylamide gel and refolded (8,34,35). RNA was quantified by OD260 and additionally checked by staining with SYBR Green (LONZA #50523) and comparing the RNA to a known sample. All quantitations were done using ImageJ software analysis of digital photographs (36).

The protein expression construct contained an *Escherichia coli* codon optimized R2Bm ORF. The ORF contained amino acid residues 70–1114 (i.e. KKSN…GGVG) of the translated GenBank sequence (M16558) and contained a carboxyl-terminal 6X His-tag. A QuikChange site-directed mutagenesis kit (Stratagene #200523–5) was used to generate the RH/AA, RH/A, ELE/E, K/A..K/A, K/A..K, KY/A and R/AG.W mutants. The primers used are listed in Supplementary Table 1. The specific mutations were chosen based on sequence alignments and protein threading models (see Figures 1 and 6). The mutated construct was then transformed into a BL21 strain of *E. coli* (Agilent #200133) by making the cells chemically competent and heat shocking them.

Protein purification

To express the proteins, 250 ml expression cultures containing the appropriate R2Bm expression construct were grown in LB broth supplemented with 50 μg/ml kanamycin in an incubator-shaker (37°C, 240 rpm). At an OD600 of between 0.8–1.0, cells were induced with 0.1 mM IPTG and grown for an additional hour at 37°C. Cells were harvested by centrifugation at 3724 × *g* for 20 min at 4°C. The cells were resuspended in 25 ml 10 mM Tris pH 7.5 and centrifuged again at 400 × *g* for 10 min. The rinsed cells were stored at −80°C.

To purify R2Bm protein, cell pellets were thawed and resuspended in 2.5 ml of Buffer A (100 mM Hepes pH 7.5, 50% glycerol, 5mM β-mercaptoethanol, 2 mg/ml lysozyme). The resuspended cells were incubated at 23°C for 10 min. After incubation, 13.2 ml of Buffer B (100 mM HEPES pH 7.5, 1 M NaCl, 0.2% triton X-100, 5 mM β-mercapto ethanol) was added. The mixture was gently inverted 5–6× and held on ice for 30 min. The cell lysate was cleared of DNA and insoluble material by centrifugation under vacuum for 20 h at 110 000 × *g* at 2°C. The clarified lysate was poured into a 15 ml tube containing 250 μl bed volume of Talon metal affinity resin (Clontech #635501) pre-equilibrated with 3 ml of wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 0.02% triton X-100, 10 mM Imidazole pH 7.5). The lysate plus Talon resin was agitated for 20 min on ice. The Talon resin was spun down at 1000 × *g* for 3 min at 4°C and the supernatant was discarded. The Talon resin was washed 2× quickly with 10 ml aliquots of wash buffer, centrifuging between washes to discard supernatant. Two more identical washes were done with an additional 10 min incubation with agitation prior to centrifugation. The Talon resin was then transferred to a 1.2 ml chromatography column (Biorad 732–6008) and washed with two successive 1 ml aliquots of wash buffer (gravity flow, total wash volume of 2 ml). Proteins were eluted from the resin by addition of 600 μl elution buffer (50 mM HEPES pH 7.5, 100 mM NaCl, 50% glycerol, 0.1% triton X-100, 150 mM imidazole). Proteins were stored in elution buffer supplemented with 0.1 mg/ml bovine serum albumin (BSA) and 2 mM dithiothreitol (DTT) (final concentrations) at −20°C. Proteins were quantified by SYPRO Orange (Sigma #S5692) staining of samples run on sodium dodecyl sulphate-polyacrylamide gel electrophoresis prior to addition of BSA for storage. Dilutions of a BSA standard (Biorad #500–0202) were used to generate a standard curve. All quantitations were done using ImageJ software analysis of digital photographs (36).

R2Bm reactions and analysis

Binding, cleavage and TPRT reactions were performed essentially as described previously (5,8). Reactions were 13 μl and contained 80 fmol labeled substrate DNA, 150 ng of unlabeled poly-dIdC, 360–12 fmol R2Bm protein, 10 mM Tris–HCl (pH 8.0), 200 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 0.1 mg/ml bovine serum albumin, 0.01% Triton X-100 and 12% glycerol. In addition, either 1.2 pmol of R2Bm 5′ PBM RNA or 3′ PBM RNA was present. TPRT reactions contained 2.5 μM of each deoxynucleotide triphosphate (dNTP). The reactions in Figure 4A–C lacked poly-dIdC as these reactions were done prior to poly-dIdC becoming a standard component of the R2 reactions.

Reactions used to determine the relative DNA binding potential of the mutants were assembled as a master mix of all of the components minus protein and aliquoted into 10 μl reactions. Reactions were started with the addition of 3 μl of protein containing equal moles of either wild-type (WT) or mutant protein. The amount of protein added was such that WT protein bound around 40–60% of the labeled DNA. The KPD/A mutant was used as the WT benchmark for mutants that had impaired DNA cleavage ability.

Reactions for determining DNA cleavage and TPRT were assembled as a master mix lacking protein and target DNA. Protein was added to aliquots and allowed to pre-incubate at 25°C for 15 min to allow the RNA to bind to the R2Bm protein. Reactions were started by the addition of substrate

DNA (1 μl) and continued at 37°C for 30 min. The reactions were chilled on ice prior to loading 9 μl of the reactions onto 5% native (1× Tris-borate-EDTA (TBE), 9.5 by 12.5 cm) polyacrylamide gels and run for ∼50 min at 220 V at 4°C. The remainder of the R2Bm reaction was mixed with 95% formamide 0.5× TBE to a final concentration of formamide ≥70%. A 5-μl aliquot of each formamide treated sample was loaded onto denaturing (8M urea) 7% polyacrylamide gels. Typically, four protein concentrations (e.g. spanning from 120 fmol protein to 12 fmol) were used per comparative experiment.

For both approaches, two replicate experiments were performed per protein purification round and usually two independent protein purification rounds were examined. WT and mutant proteins were age matched as the DNA binding and cleavage activities decrease with time post-purification.

All gels were exposed to a phosphorimager screen, which was then scanned on a molecular dynamics STORM 840 phosphorimager. The resulting 16 bit TIFF images were linearly adjusted in Photoshop so that the most intense bands were dark gray. Adjusted TIFF files were quantified using imageJ (36). The fraction of DNA bound by the R2Bm protein in a given reaction was determined using electrophoretic mobility shift assay (EMSA) data. The fraction of DNA cleaved by the R2Bm protein in a given reaction was determined using data from the denaturing gel.

3D modeling

R2 sequences from various R2 elements (especially R2Bm, R2Lp, R8Hm, R9Av) were submitted to Phyre2 structure prediction program as queries (37). Varying lengths upstream and downstream of the KPD-D motif (especially for R2Bm and R2Dm) were submitted for modeling. Molecular graphics and analyses were performed with the UCSF Chimera package. Chimera is developed by the Resource for Biocomputing, Visualization and Informatics at the University of California, San Francisco, CA, USA (38).

RESULTS

R2 alignment

The size and structure of the R2 RLE is not known. To better understand the R2 RLE structure-function, point mutations were generated in conserved regions based upon 3D modeling and amino acid sequence conservation. A Clustal alignment of the RLE domain from representative D-clade and A-clade R2 elements that target the R2, R8 and R9 sites is presented in Figure 1C. The red asterisks indicate active site residues that were previously identified (25). The α-helices and β-strands are also indicated in Figure 1C. The α-helices and β-strands were determined by the 3D model (see the modeling section of the 'Results' section). The R2Bm mutants generated by site directed mutagenesis for this study were located in α-helices 1 and 2 and just after β-strand 4. The mutations were RH/AA, RH/A, ELE/A, K/A..K/A, K/A..K, KY/A and R/AG.W, as noted in Figure 1C.

Mutations in α-helix 1 affect DNA cleavage and DNA binding

To determine if the conserved RH residues located at the beginning of α-helix 1 (see Figure 1C) had an effect on DNA cleavage activity, the cleavage activity of purified RH/AA and RH/A mutant R2Bm protein were compared to the cleavage activity of purified WT R2Bm protein *in vitro*. DNA cleavage reactions were carried out on ^{32}P end-labeled (bottom strand) target dsDNA in the presence of 3′ PBM RNA. Protein was pre-bound to RNA prior to the addition of DNA. After incubation, the aliquots of the DNA cleavage reactions were assayed by EMSAs, run on native polyacrylamide gels (Figure 2A) as well as denaturing polyacrylamide electrophoresis (Figure 2B, full length denaturing gels are in Supplementary Figure 9).

The EMSA allowed a qualitative assessment of the mutant's ability to form native-like protein–DNA–RNA complexes—assayed by migration distance. To assay for DNA cleavage activity, the amount of DNA cleaved per protein-bound unit of DNA was determined by combining information generated from the EMSAs and information generated from denaturing (8M urea) polyacrylamide gels. The EMSA gel allowed the fraction of DNA that had been bound to protein to be calculated. The denaturing polyacrylamide gels allowed the fraction of DNA that had undergone a DNA cleavage event to be calculated. The results are graphed in Figure 2C as a scatter plot of DNA cleavage activity (fraction cleaved) as a function of DNA that had been bound by protein (fraction bound). The RH mutations did not appear to greatly affect the migration pattern of the protein–DNA–RNA complexes (Figure 2A). The RH/AA mutation, however, completely abolished DNA endonuclease activity (Figure 2B and C). The endonuclease activity of RH/A single mutation was much less affected. No cleavages beyond the R2 cleavage site were detected in either WT or mutants (data not shown).

A separate experiment was performed that was tailored to focus on DNA binding activity instead of DNA cleavage activity. As the RH/AA abolished DNA cleavage, a previously characterized endonuclease mutant, KPD/A, was used as the 'wild-type' control (referred to as $WT^{KPD/A}$) in the DNA binding assays (Figure 2D). The KPD/A mutant lacks endonuclease activity but is otherwise WT in form and function (25). In order to reduce errors inherent in pipetting small volumes of DNA, the master mix contained all of the components except R2 protein. The binding reaction was started by the addition of R2 protein. The amount of protein added was such that the $WT^{KPD/A}$ control-reactions resulted in nearly 50% of DNA being bound by protein. The actual fraction bound was determined for each matched set of $WT^{KPD/A}$ and RH/AA mutant lanes, and the results were normalized to $WT^{KPD/A}$ DNA binding activity. $WT^{KPD/A}$ activity was defined to be 100%. Replicate experiments were performed both within and across independent rounds of protein purification, and the results averaged and reported in Figure 2D. The RH/AA mutation was observed to decrease the ability of R2BM protein to bind to target DNA by about 40% in the presence of the 3′ PBM RNA. The RH/A mutation was not directly tested, but appeared to

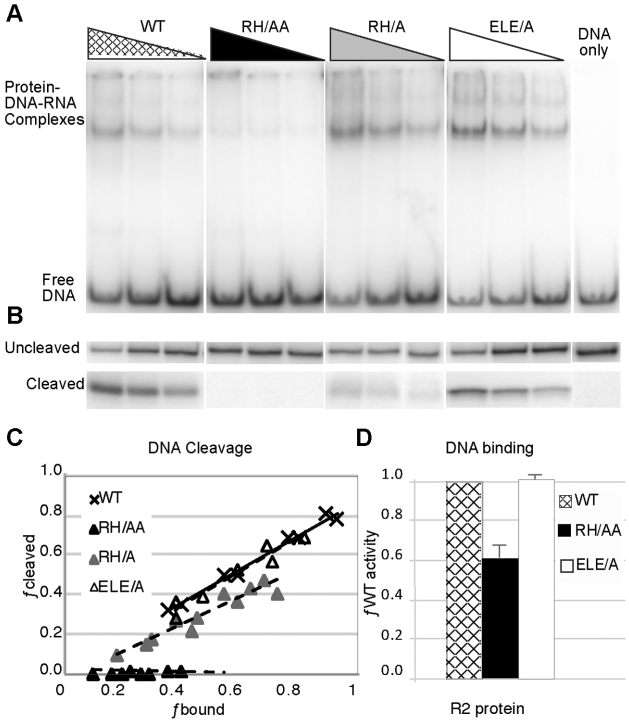

Figure 2. The RH residues in α-helix 1 affect DNA cleavage and DNA binding. DNA cleavage reaction (**A**–**C**) differ from DNA binding reactions (**D**). In the DNA cleavage reactions, purified mutant R2Bm protein and wild-type (WT) R2 protein were pre-bound to 1.2 pmol of 3′ PBM RNA and the reactions were initiated with the addition of a 120-bp segment of 28S rDNA that contained the R2 insertion site. The 120-bp target DNA had been generated by PCR and contained within it the upstream and downstream R2 protein binding sites (8). The antisense 28S PCR primer (i.e. the 'bottom' strand primer) had been 5′ end labeled with ^{32}P to facilitate tracking of bottom strand cleavage events of target DNA bound by R2Bm protein. After an incubation period, each reaction was split into two aliquots and loaded onto native and denaturing polyacrylamide gels for analysis. In the binding reactions, DNA was contained in the master mix and the reactions were initiated by the addition of protein in order to reduce pipetting error. (A) Representative electrophoretic mobility shift assay (EMSA) gel used to calculate the fraction of DNA bound by R2Bm protein across a range of protein concentrations (triangles). (B) Representative denaturing (8M urea) polyacrylamide gel electrophoretic analysis—of the reactions in panel A—used to determine the fraction of cleaved DNA. For graphical and space considerations, the gel has been trimmed to show just the uncleaved and cleaved DNA regions of the gel as R2 is highly site-specific and the mutations did not affect specificity. (C) A scatter plot of the fraction of cleaved DNA (*f*cleaved) as a function of the fraction of protein bound to target DNA (*f*bound). The *f*bound data were derived from EMSA gels, similar to panel A and the *f*cleaved DNA were derived from the corresponding denaturing gels.The black and gray triangles on the graph represent RH/AA and RH/A data respectively, while the 'X' points on the graph represent WT data. The white triangles are the ELE/A data. (D) Bar graph reporting the DNA binding efficiency of the RH/AA and ELE/A mutant proteins calculated from binding reactions assayed on a EMSA gels. The black bar reports the DNA binding activity of RH/AA mutant protein as a fraction of WT$^{KPD/A}$ activity. The white bar reports the DNA binding activity of ELE/A mutant protein as a fraction of WT activity. The standard deviation is indicated above the bar. The DNA binding activity of WT and WT$^{KPD/A}$ was set at 1 and is represented as the hatched bar.

have improved DNA binding over the double mutant in the DNA cleavage reaction EMSAs (Figure 2A).

The RH/AA mutant was also tested for the ability to bind to target DNA in the presence of the 5′ PBM RNA. 5′ PBM RNA drives protein to bind to sequences downstream of the insertion site and to cleave the top strand of the target DNA.The RH/AA mutation reduced DNA binding activity by approximately 30% in the presence of 5′ PBM RNA and abolished top strand cleavage activity (Supplementary Figures 1 and 2, respectively).

Conserved glutamate at the end of β-strand 1 does not appear to be directly involved in either DNA cleavage or DNA binding

Mutating the conserved glutamate residue positioned at the end of the first β-strand to an alanine (ELE/A) does not appear to affect DNA binding (Figure 2D), DNA cleavage (Figure 2A–C), or the migration of the protein nucleic acid complexes in mobility shift gels (Figure 2A).

Mutations in α-helix 2 affect DNA cleavage

Three mutations were generated in the K..KY motif located at the amino-terminal end of α-helix 2: K/A..K/A, K/A..K and KY/A. DNA cleavage reactions were set up to test the mutants for loss of endonuclease function (Figure 3). The DNA cleavage reactions and subsequent analysis were identical to those already described in Figure 2. At least two replicate experiments were performed per protein purification round, and two independent protein purifications were examined. Figure 3A shows a representative EMSA gel. Each of the mutants appeared to be able to form the correct migrating protein–RNA–DNA complexes. The companion denaturing gel is shown in Figure 3B. The scatter plot of DNA cleavage as a function of DNA binding for the cumulative datasets is presented in Figure 3C. The K/A..K/A and the K/A..K mutations virtually abolished DNA cleavage, while the KY/A mutation partially impaired DNA cleavage activity per protein bound unit of DNA. None of the helix 2 mutants affected DNA cleavage specificity.

In order to determine the role, if any, the K..KY motif might have on DNA binding, the DNA binding activity of the K/A..K/A and KY/A mutants were measured relative to the DNA binding activity of the WT$^{KPD/A}$ protein (Figure 3D). Both mutants had WT, or near WT, DNA binding activity. The experimental setup in Figure 3D was identical to the analysis in Figure 2D for the RH/AA mutant.

In addition to the experiments shown in Figure 3, we mutated the RE residues just in front of the K..KY motif at the beginning of α-helix 2 in R2Bm, as an internal control. The RE residues are not conserved across R2 elements and mutating these residues has no observable effect on either DNA binding or DNA cleavage (our unpublished data). Finally, the α-helix 2 mutants were also tested for the ability to bind to target DNA in the presence of 5′ PBM RNA and to cleave the top strand of the target DNA. The α-helix 2 mutations had little to no effect on overall DNA binding activity in the presence of 5′ PBM RNA (Supplementary Figure 3). Both the K/A..K/A and KY/A reduced top strand cleavage activity to below detection levels (Supplementary Figure 4).

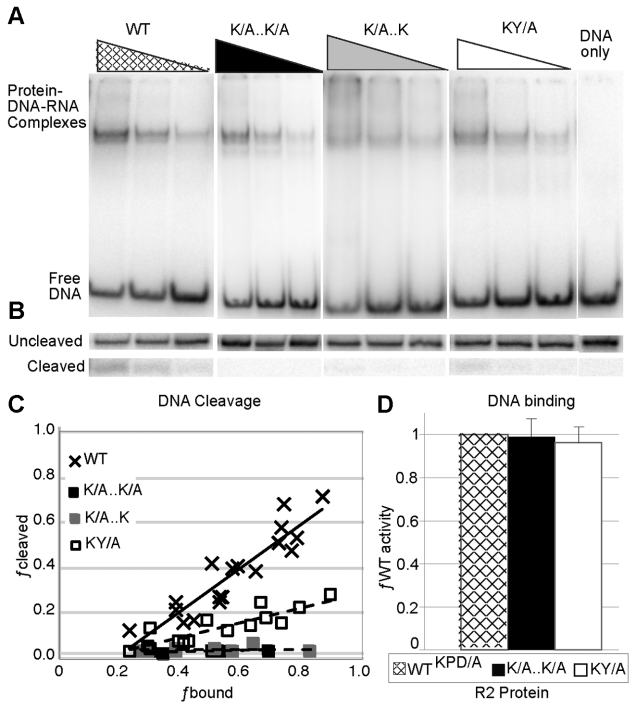

Figure 3. α-Helix 2 K..KY motif residues affect DNA cleavage but not DNA binding. DNA cleavage reactions (**A**–**C**) and DNA binding reactions (**D**) were performed as in Figure 2. (A) Representative EMSA gels used to calculate the fraction of DNA bound by R2Bm protein across a protein concentration series. Triangles represent a protein titration series. (B) Representative denaturing gel of the reactions in panel A. (C) Scatter plot of fraction of cleaved target DNA (*f*cleaved) as a function of fraction of bound target DNA (*f*bound). All reactions, abbreviations and graphs are as in Figure 2 except where noted. The black boxes represent the K/A..K/A data, the gray boxes represent the K/A..K data and the white boxes represent the KY/A data. (D) Bar graph reporting the relative DNA binding efficiency of the K/A..K/A and KY/A mutants relative to $WT^{KPD/A}$ protein. Activity of $WT^{KPD/A}$ was set at 1. All other conditions, abbreviations and symbols are as in previous figures.

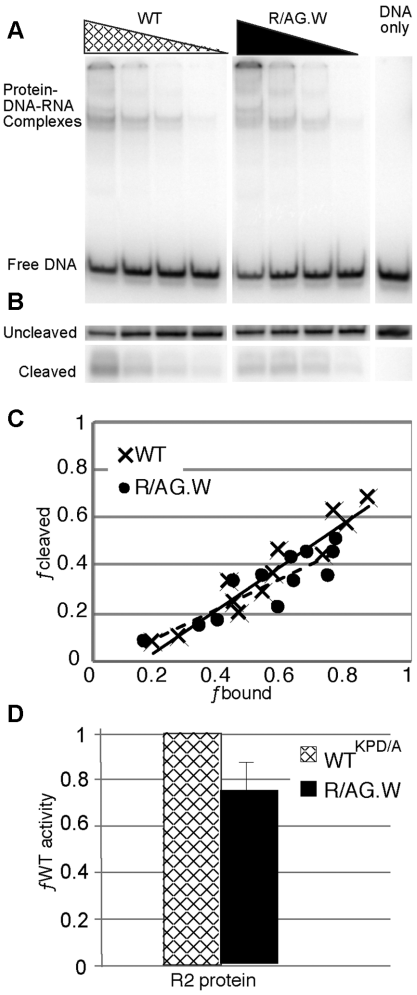

Figure 4. DNA cleavage and DNA binding activity of the R/AG.W mutant. (**A**) Representative EMSA gel used to calculate the fraction of target DNA bound by R2Bm protein across a protein concentration series. (**B**) Representative denaturing gel of the reactions in panel A. (**C**) Scatter plot of fraction cleaved target DNA (*f*cleaved) as a function of fraction target DNA bound by protein (*f*bound). Black circles represent data from the R/AG.W mutant. (**D**) Bar graph reporting the relative DNA binding efficiency of the R/AG.W mutant (black bar) relative to $WT^{KPD/A}$ protein. Activity of $WT^{KPD/A}$ was set at 1. All other abbreviations and symbols are as in previous figures.

Mutations in the loop region following β-strand 4 affect DNA binding and DNA cleavage

As per the previous mutations, DNA cleavage and DNA binding analysis of the R/AG.W R2 mutant were carried out (Figure 4). While the R/AG.W mutant appeared to largely form the correct migrating protein–RNA–DNA complex(es) (Figure 4A), the mutant was partially impaired in DNA cleavage activity per protein bound unit at the higher protein to DNA ratios (Figure 4C). DNA cleavage specificity was maintained. The R/AG.W mutation appeared to have a moderate reduction (25%) on DNA binding activity (Figure 4D).

Again, as with the previous mutations examined in this paper, the R/AG.W mutation was examined for loss-of-function in the presence of 5′ PBM RNA. The results were similar to those observed in the presence of 3′ PBM RNA (see Supplementary Figures 5 and 6).

Target primed reverse transcription

In order to determine if the DNA endonuclease impaired mutants RH/AA, K/A..K/A, KY/A and R/AG.W were capable of performing TPRT, TPRT activity had to be assayed on a pre-nicked target substrate. See Figure 5A for a diagram of the experimental setup. See 'Materials and Methods' section for information on how the nicked DNA (60% nicked) was generated. TPRT activity of the endonuclease impaired mutants was compared to the TPRT activity of $WT^{KPD/A}$ R2Bm protein instead of WT R2Bm. $WT^{KPD/A}$ R2Bm protein is known to have WT TPRT activity given pre-nicked DNA substrates (25). The α-helix 1 mutant, RH/AA, was the only mutant that was found to be deficient in TPRT activity (Figure 5B).

3D modeling of the endonuclease domain

In order to model the endonuclease domain of R2Bm, we submitted C-Terminal sequences of varying lengths up-

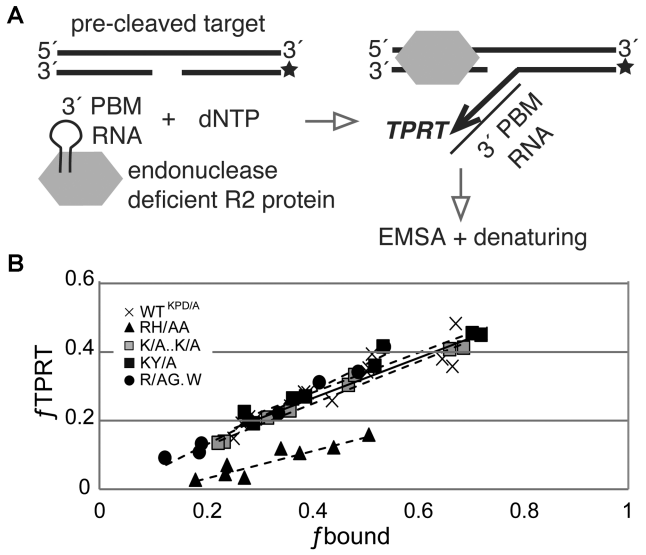

Figure 5. TPRT activity of the endonuclease impaired mutants. The RH/AA, K/A..K/A, KY/A, R/AG.W were tested for the ability to perform TPRT on pre-nicked DNA substrates. (**A**) Experimental design. The master mix contained 3′ PBM RNA, poly-dIdC, ^{32}P end-labeled (asterisk) pre-nicked target DNA (60% nicked). Reactions were started by the addition of R2Bm protein. (**B**) The results are given as a scatter plot of TPRT activity, expressed as the fraction of DNA that had undergone TPRT (*f*TPRT), as a function of the fraction of target DNA bound by protein (*f*bound) at each protein concentration. All other conditions and abbreviations are as in previous figures.

stream and downstream of the known catalytic residues in R2Bm to the Phyre2 structure prediction program (37). The CCHC domain and sequence downstream of the RG.W motif were not modeled well. The presence of the CCHC domain and sequence downstream of the RG.W motif often lead to perturbations of the predicted restriction endonuclease-like fold (αββββαβ) geometry and structure, especially in the extended loop between β-strands 1 and 2, in the placement of the RG.W motif, and the geometry and extent of the two α-helices (data not shown). The sequence block that generated the most reliable model for the R2Bm endonuclease is the block presented in Figure 6: the sequence starting just downstream of the CCHC and ending just past the RG.W motif. The resulting model (Figure 6A) was of generally high confidence (Figure 6G). The mutations that had a measurable effect on DNA binding or cleavage were largely on one face of the modeled protein, forming a positively charged patch (Figure 6B).

Phyre2 generated the models using modern protein threading algorithms. Three of the crystal structures that were highest weighted (Figure 6G) in the modeling of R2Bm were a domain of unknown function (duf 3fov), Hje and Hjc (Figure 6D–F, respectively) (29,39). The individual helices and strands of the core αββββαβ architecture for duf 3fov, Hje and Hjc are roughly equivalently positioned in 3D space except that the two α-helices in Hjc are comparatively less perpendicular. In addition, the loop between the first two β-strands of the αββββαβ core in duf 3fov is more structured and Hje contains an additional β-strand leading into the second helix. Beyond αββββαβ core, both Hjc and Hje contain additional structural motifs.

We have similarly modeled the endonuclease domains for several of the other R2 elements listed in Figure 1 (R2Dm, R2Lp, R8Hm-A, R9Av1) spanning the R2-A and R2-D clades and the R2, R8 and R9 target sites. The models for R2BM and R2Dm superimpose nicely (Figure 6A and C). The other R2 elements modeled also give a similar overall model (see Supplementary File 1). The most variation occurred in the extended loop between β-strands 1 and 2 and the loop between β-strand 3 and α-helix 2. R8Hm was the most poorly modeled, with α-helix 2 not being modeled as well (see Supplementary File 1).

DISCUSSION

PD-(D/E)XK superfamily of endonucleases comprise a diverse set of endonucleases that share an αββββαβ restriction endonuclease-like fold and a PD-(D/E)XK motif (reviewed in (26,28,40)). In this superfamily, the D-(D/E)XK residues of the PD-(D/E)XK motif are known to coordinate the metal ions that are required for DNA cleavage (reviewed in (40)). Many members of the PD-(D/E)XK superfamily have an additional glutamic acid residue located in the first α-helix of the αββββαβ restriction endonuclease-like fold, yielding a motif of E-PD-(D/E)XK (26,31). The α-helix 1 glutamate, at least in the case of EcoRV and BglI, is catalytic and helps to coordinate metal ions in the active site (31,41–44). A small subset of the E-PD-(D/E)XK family members appear to have a conserved histidine (or other residue) of untested function instead of glutamate (26,28). Holliday junction resolvases, like many of the bacterial restriction endonucleases, are of the E-PD-(D/E)XK construction: having a conserved glutamate residue (E10 as numbered in Figure 6H) in α-helix 1 that is important for DNA cleavage (29,45–47). Point mutations in any of the E-PD-(D/E)XK catalytic residues (red asterisks in Figure 6H) of restriction endonucleases, Hjc and Hje virtually abolish catalytic activity but do not greatly affect DNA binding (29,48,49). In addition to the E-PD-(D/E)XK residues, Hjc and Hje have an invariant catalytic serine located between β-strands 1 and 2 that is critical for cleavage of junctions (29).

The R2 endonuclease was modeled using archaeal Hjc and Hje structural data (Figure 6G), as such it shares the αββββαβ core fold and the placement of the PD-(D/E) residues of these enzymes. However, R2 lacks the catalytic serine and the α-helix 1 glutamate of the Holliday junction resolvases. Instead, R2 has a conserved glutamate located between β-strands 1 and 2, near the Hjc/Hje catalytic serine (Figure 6H). Unlike the Hjc/Hje serine, however, the R2 glutamate residue does not appear to be catalytic (Figure 2). The glutamate residue is not a replacement for either the Hjc/Hje serine or the missing α-helix 1 catalytic glutamate, nor is the R2 α-helix 1 histidine serving that catalytic function. The role of the R2 glutamate is as yet undetermined. The role of the RH motif is discussed below. Finally, R2 differs in the placement of the catalytic lysine of the PD-(D/E)XK motif. The location and existence of the catalytic lysine has been an open question for R2 since PD-(D/E) catalytic residues were located (23,25). The lysine residue in R2 is located in α-helix 2 instead of the more canonical position in β-strand 3. The first lysine (K69 of the K69..K72Y73 motif) in α-helix 2 is the catalytic lysine as it is

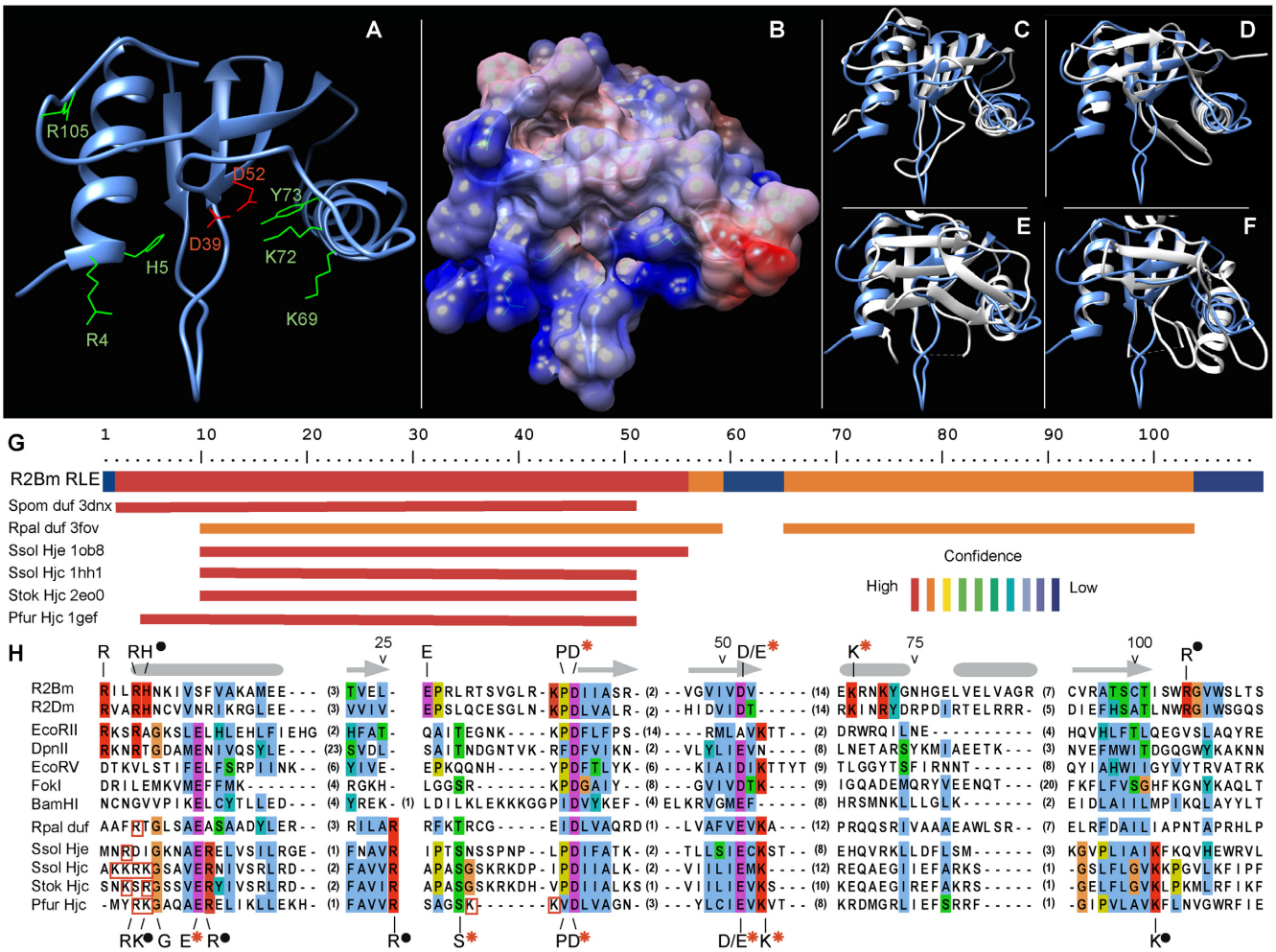

Figure 6. Structural modeling of endonuclease domain. (**A**) R2Bm model. The R2Bm model was generated using the Phyre2 server (37). The side chains in red are the previously characterized active site residues (25). The side chains in green are residues that were changed to alanine in this study. (**B**) Coulombic surface map of R2Bm. (**C**) R2Bm in blue and R2Dm in white overlay. (**D**) Overlay of R2Bm and a domain of unknown function from *Rhodopseudomonas palustris*—Rpal duf, PDB ID: 3fov. (**E**) Overlay of R2Bm and the Holliday junction resolving enzyme from *Sulfolobus solfataricus*—Ssol Hje, PDB ID: 1ob8. (**F**) Overlay of R2Bm and Holliday junction cleavage protein—Ssol Hjc, PDB code 1hh1. (**G**) R2Bm model construction and confidence report from Phyre2. PDB ID: 3dnx is a domain of unknown function (duf) derived from the uncharacterized protein SPO1766 in *Silicibacter pomeroyi*. PDB ID: 3fov is a duf from the uncharacterized protein rpa0323 in *Rhodopseudomonas palustris*. Holliday junction resolvases Ssol Hje 1ob8 and Ssol Hjc 1hh1 are from *Sulfolobus solfataricus*. Holliday junction resolvase Stok Hjc 2eo0 is from *Sulfolobus tokodaii* str. 7 and Pfur Hjc 1gef is from *Pyrococcus furiosus*. (**H**) Structural plus Clustal alignment of the core αββββαβ fold from type II restriction enzymes and from the archaeal Holliday junction resolvases used to build the R2Bm model. The α-helices (rounded bars) and β-strands (arrows) are marked for R2. The amino acid numbers are given for the R2 endonuclease sequence. Known conserved catalytic (red asterisk) and DNA binding (black dot) residues for R2 and for Holliday junction resolvases are listed above and below the alignment, respectively. Arginine and Lysine residues near the start of α-helix 1 of the Holliday junction resolvases are highlighted with a red rectangle. Other conserved residues are marked with the default Clustal coloring scheme.

more highly conserved (Figure 1C) and mutating it knocks out DNA cleavage (Figure 3). The second lysine (K72) is deeper into the cleft than K69. It remains a possibility that both K69 and K72 are catalytic. The side chain of the tyrosine is also oriented toward the cleft. Mutating the tyrosine reduces DNA cleavage (Figure 3C). It is unknown at this point if the tyrosine residue has any role in catalysis or if mutating the residue simply changed the geometry of the catalytic lysine residue(s). Mutations in the K..KY motif have negligible effect on DNA binding (Figure 3D). The R2 D39-(D/E)52 residues are located at the start of β-strand 2 and in β-strand 3, respectively, and are colored red in Figure 6A. These residues are in canonical positions and have been previously tested for catalytic activity in R2Bm (25).

The R2 endonuclease appears to have a large potential DNA-interaction surface—a contiguous electropositive surface surrounding the active site cleft (Figure 6B). The RH of α-helix 1, the K..K of α-helix 2 and the arginine of the RG.W motif form parts of that surface. The RH motif, when mutated, had the greatest effect on DNA binding. In the model, the histidine residue of the RH motif lies inside the active site cleft and points toward the catalytic residues. The arginine residue sits on the surface of the protein at the edge of the active site cleft. Preceding the arginine of the RH is an additional conserved arginine.

There is an arginine rich region consisting of RXXR or NXRXXR found within or in front of the first α-helix of the αββββαβ endonuclease fold of a subset of type II restriction enzymes (e.g. EcoRII, DpnII, MboI, PspGI and Sso II, Figure 6H) (31,40,50,51). Mutation of this 'R-box' region can reduce DNA binding in these enzymes by several orders of magnitude and abolish DNA cleavage (40,50,51). In some instances, the R-Box motif is thought to make contacts with consecutive G residues in the DNA target 3′ of the cleavage site (50,51). Restriction enzymes that do not have an R-box still often use this region as a DNA contacting region. The R-box residues function as DNA recognition and correct positioning of the DNA in the active site for cleavage.

Similar to a restriction endonuclease R-box, arginine and lysine residues are found near the start of α-helix 1 of PD-(D/E)XK family Holliday junction resolvases. In the crystal structure for *Pyrococcus furiosus* Hjc (Figure 6H, Pfur Hjc), the RK residues boxed in red coordinate a sulfate residue (45,46). The sulfate is thought to mimic the placement of a DNA phosphate group in the crystal structure (45,46). The RK residues are generally conserved across other Hjc members (45). Mutating the RK residues to alanines reduces binding to Holliday junctions by increasing the disassociation rate, but retain the ability to disrupt the junction base-pairs and to cleave the DNA (45). The RK residues help form a stable complex with the junction DNA. Deleting the amino terminal MYRKG residues in *P. furiosus* abolishes DNA binding altogether (46). The equivalent residues in *Sulfolobus solfataricus* (Ssol) Hjc crystal structure are disordered, and thought to become ordered upon DNA binding (39). The glycine residue in this region is highly conserved in the Hjc and Hje Holliday junction resolvases. The glycine is also present in some restriction endonucleases, especially those with an R-box. In Holliday junction resolvases, the glycine is thought to aid conformational changes of the N-terminal region upon binding a Holliday junction (45). The protein conformational changes and the local unstacking of DNA bases at the junction point that occur upon binding provide the induced fit required for DNA cleavage (49). R2 lacks this important glycine residue.

In addition to the R and K residues near the beginning of α-helix 1, Hjc and Hje have highly conserved basic residues located within α-helix 1 and at the ends of both β-strands 1 and 4 (highlighted with a black dot in Figure 6H). These basic residues have been biochemically and structurally implicated in interacting with DNA phosphates and positioning the junction for cleavage (29,39,45,46). R2 lacks the basic residues found in α-helix 1 and at the end of β-strand 1 of the Hjc/Hje resolvases. R2, however, does have a conserved basic residue near the end β-strand 4, similar to Hjc/Hje. In the resolvases, the residue following β-strand 4 is a lysine while in R2 the conserved residue is an arginine (the R of the RG.W) motif. As is the case for the Hjc/Hje, the R2's RG.W motif appears to be involved in DNA binding.

Less conserved Hjc/Hje residues located between β-strands 1 and 2, and between β-strand 3 and α-helix 2, have also been implicated in DNA phosphate binding interactions. Point mutants for Ssol and Pfur Hjc basic residues in these regions reduce DNA binding and DNA cleavage (39,45). R2 has several basic residues in the loops between β-strands 1 and 2, especially the lysine (or arginine) next to the PD motif, but largely appears to lack a similar patch of basic residues between β-strand 3 and α-helix 2.

The loop between β-strands 1 and 2 tends to be flanked by proline residues in both R2 and Hjc/Hje. The first proline is at the dimer interface in Hjc and Hje and impacts the geometry of the two subunits relative to each other (29,39). The second proline is presumably important in positioning the catalytic aspartic acid residue.

With the exception of the aforementioned R-box region and basic residues near the PD motif, restriction enzymes, being very diverse with many additions to the αββββαβ core structure, are difficult to make generalizations about when comparing DNA binding motifs to R2 and to Holliday junction resolvases. All three, however, appear to use flexible regions that become more structured upon binding to DNA. The induced fit of these unstructured regions is important for positioning the DNA near catalytic D-(D/E)XK residues.

The RH motif, and likely the preceding arginine residue of R2 (Figure 1C), appears to form an R-box equivalent involved in DNA binding. As in both restriction enzymes and the Holliday junction resolvases Hjc and Hje, major perturbations of the R-box region reduce both cleavage and DNA binding. It is possible that the DNA binding role of the R2 R-box residue(s) might have a site-specific component to it, as the DNA binding defects are greater in the presence of 3′ PBM RNA than they are in the presence of 5′ PBM RNA. In the presence of 3′ PBM RNA, the R2 protein would be expected to bind upstream of the insertion site (5). In the presence of 5′ PBM RNA, the protein would be expected to bind downstream of the insertion site (8). The RH/AA mutation reduced DNA binding activity by 40% in the presence of 3′ PBM RNA (Figure 2D), while in the presence of 5′ PBM RNA DNA binding activity was reduced by 30% (Supplementary Figure 1). However, the fact that the RH motif is present in both A and D clade R2 elements—regardless of the DNA sequence being targeted by the element (Figure 1C)—argues against the reduction in DNA binding being due to site-specific protein–DNA contacts. In addition, the RH/AA mutant shows a similar 40% reduction in DNA binding compared to WT R2 protein when WT and RH/AA R2 proteins are forced to bind to non-target DNA in the presence of non-specific RNA (Supplementary Figure 7), possibly arguing against sequence-specific role for the RH motif.

The other R2 mutation that decreased DNA binding activity was the R/AG.W mutant. The mutation decreased DNA binding activity relative to WT R2 by 25% in the presence of 3′ PBM RNA and 21% in the presence of 5′ PBM RNA, although the reduction in DNA binding might be greater than that indicated in Figure 4D because the R/AG.W mutant retains most of its DNA cleavage activity. The $WT^{KPD/A}$ R2 protein, to which the R/AG.W mutant is being compared for DNA binding, does not cleave target DNA and the R2 protein yields a greater DNA footprint post-DNA cleavage (34). The difference in affinity of the R2 protein for DNA pre- and post-DNA cleavage appears to be small (our unpublished data) at the levels of protein and DNA used in our reactions. The RG.W motif is located at a sharp turn in a coiled region in the R2 models, with the R residue side chain running roughly parallel to the side chain

Table 1. Summary of DNA binding and cleavage results

R2 protein	target DNA binding activity (3′ PBM RNA)	target DNA cleavage activity (3′ PBM RNA)	TPRT activity (3′ PBM RNA)	target DNA binding activity (5′ PBM RNA)	target DNA cleavage activity (5′ PBM RNA)	non-target dsDNA binding activity (non-specific RNA)	non R2 site cleavage activity (3′ PBM RNA)
WT	N.A.	N.A.	N.A.	N.A.	N.A.	non-specific associations	none detected
RH/AA	40% reduction	none detected	reduced	30% reduction	none detected	40% reduction	none detected
RH/A	N.T.	reduced	N.T.	N.T.	N.T.	N.T.	none detected
ELE/A	normal	normal	N.T.	N.T.	N.T.	N.T.	none detected
K/A..K/A	normal	none detected	normal	normal	none detected	N.T.	none detected
K/A..K	N.T.	none detected	normal	N.T.	N.T.	N.T.	none detected
KY/A	normal	reduced	normal	normal	reduced	N.T.	none detected
R/AG.W	25% reduction	marginal reduction	normal	21% reduction	marginal reduction	N.T.	none detected

Not applicable (N.A.). Not tested (N.T.).

of the R in the RH motif (Figure 6A). In Hjc and Hje the equivalent sharp turn to the RG.W turn is located between two β-strands and is near to or including the K residue following β-strand 4 of Hjc/Hje (Figure 6E, F and H). Like the arginine in the α-helix 1 RH motif of R2, the arginine of the RG.W motif likely makes contact with the target DNA. The arginine residues from both motifs are part of the extended basic patch that likely constitutes the DNA contacting surface of the R2 endonuclease (Figure 6B).

Overall the model generated for the R2 endonuclease appears to be a good predictor of form and function. A summary of our biochemical results are presented in Table 1. A large contiguous electropositive surface is present that includes not only the catalytic cleft but also the surface surrounding the cleft (Figure 6B). The arginine of the α-helix 1 RH motif and the arginine of the RG.W motif form part of that surface. The D-(D/E) catalytic residues are properly modeled in the cleft with the correct orientations compared to known crystal structures. The first lysine residue of α-helix 2 appears to be the 'missing' catalytic lysine residue identified for other members of the PD-(D/E)XK endonuclease superfamily. The spacing and placement of the R2 lysine in the restriction endonuclease-like fold, however, is non-canonical, with the lysine being far downstream of the D/E and being located in the second α-helix. We were unable to locate a catalytic residue upstream of the PD motif that might function as the α-helix 1 glutamate of an E-PD-(D/E)XK motif or serine of a Hjc/Hje E-S-PD-(D/E)XK motif.

It is interesting that, of the PD-(D/E)XK endonuclease superfamily with crystal structure data, R2 would be most similar to Holliday junction resolvases; the closest PD-(D/E)XK member previously had been bacterial restriction enzymes (esp., Fok I). If the structural connection to Holliday junction resolvases is more than structural resemblance but also is borne out functionally, we might predict that R2 could look and behave similar to Holliday junction resolvases because transposition intermediates in the integration reaction resemble Holliday junction-like structures. TPRT generates a three-way junction (three-quarters of a Holliday junction). The start of second strand synthesis generates a transient four-way Holliday junction-like nucleic acid structure. We doubt that R2 is a *bona fide* Holliday junction resolvase catalytically as R2 lacks two major catalytic residues (E, S) and several conserved basic residues characteristic of Hjc and Hje. Rather R2 must function like a restriction endonuclease on linear dsDNA yet keep track of and cleave multi-branched DNA structures that form during the integration reaction. If this conjecture is true, the reduced TPRT activity observed when the RH motif is mutated might be the result of a diminished ability to bind and coordinate a three-way junction. Alternatively, the RH motif might form part of an RNA binding surface as well as a DNA binding surface.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Kimberly Bowles, Murshida Mahbub, Monika Pradhan and Brijesh Khadgi for critical reading of the manuscript and for helpful discussions. The authors also wish to thank Micki Christensen for copyediting, Dr. Subhrangsu Mandal for the kind use of his phosphorimager, Monika Pradhan for helping with finalizing some data and Murshida Mahbub for initially generating some of the helix two mutants.

FUNDING

National Science Foundation [0950983]. Funding for open access charge: National Science Foundation [0950983].
Conflict of interest statement. None declared.

REFERENCES

1. Malik,H.S., Burke,W.D. and Eickbush,T.H. (1999) The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.*, **16**, 793–805.
2. Kojima,K.K. and Fujiwara,H. (2005) Long-term inheritance of the 28S rDNA-specific retrotransposon R2. *Mol. Biol. Evol.*, **22**, 2157–2165.
3. Luan,D.D., Korman,M.H., Jakubczak,J.L. and Eickbush,T.H. (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell*, **72**, 595–605.
4. Cost,G.J., Feng,Q., Jacquier,A. and Boeke,J.D. (2002) Human L1 element target-primed reverse transcription in vitro. *EMBO J.*, **21**, 5899–5910.
5. Christensen,S.M. and Eickbush,T.H. (2005) R2 target-primed reverse transcription: ordered cleavage and polymerization steps by protein subunits asymmetrically bound to the target DNA. *Mol. Cell. Biol.*, **25**, 6617–6628.
6. Moran,J.V., Holmes,S.E., Naas,T.P., DeBerardinis,R.J., Boeke,J.D. and Kazazian,H.H.J. (1996) High frequency retrotransposition in cultured mammalian cells. *Cell*, **87**, 917–927.
7. Christensen,S.M., Bibillo,A. and Eickbush,T.H. (2005) Role of the Bombyx mori R2 element N-terminal domain in the target-primed reverse transcription (TPRT) reaction. *Nucleic Acids Res.*, **33**, 6461–6468.

8. Christensen,S.M., Ye,J. and Eickbush,T.H. (2006) RNA from the 5′ end of the R2 retrotransposon controls R2 protein binding to and cleavage of its DNA target site. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 17602–17607.
9. Aksoy,S., Williams,S., Chang,S. and Richards,F.F. (1990) SLACS retrotransposon from Trypanosoma brucei gambiense is similar to mammalian LINEs. *Nucleic Acids Res.*, **18**, 785–792.
10. Teng,S.C., Wang,S.X. and Gabriel,A. (1995) A new non-LTR retrotransposon provides evidence for multiple distinct site-specific elements in Crithidia fasciculata miniexon arrays. *Nucleic Acids Res.*, **23**, 2929–2936.
11. Burke,W.D., Malik,H.S., Rich,S.M. and Eickbush,T.H. (2002) Ancient lineages of non-LTR retrotransposons in the primitive eukaryote, Giardia lamblia. *Mol. Biol. Evol.*, **19**, 619–630.
12. Malik,H.S. and Eickbush,T.H. (2000) NeSL-1, an ancient lineage of site-specific non-LTR retrotransposons from Caenorhabditis elegans. *Genetics*, **154**, 193–203.
13. Volff,J.N., Korting,C., Froschauer,A., Sweeney,K. and Schartl,M. (2001) Non-LTR retrotransposons encoding a restriction enzyme-like endonuclease in vertebrates. *J. Mol. Evol.*, **52**, 351–360.
14. Burke,W.D., Muller,F. and Eickbush,T.H. (1995) R4, a non-LTR retrotransposon specific to the large subunit rRNA genes of nematodes. *Nucleic Acids Res.*, **23**, 4628–4634.
15. Villanueva,M.S., Williams,S.P., Beard,C.B., Richards,F.F. and Aksoy,S. (1991) A new member of a family of site-specific retrotransposons is present in the spliced leader RNA genes of Trypanosoma cruzi. *Mol. Cell. Biol.*, **11**, 6139–6148.
16. Jakubczak,J.L., Xiong,Y. and Eickbush,T.H. (1990) Type I (R1) and type II (R2) ribosomal DNA insertions of Drosophila melanogaster are retrotransposable elements closely related to those of Bombyx mori. *J. Mol. Biol.*, **212**, 37–52.
17. Kojima,K.K., Kuma,K., Toh,H. and Fujiwara,H. (2006) Identification of rDNA-specific non-LTR retrotransposons in Cnidaria. *Mol. Biol. Evol.*, **23**, 1984–1993.
18. Gladyshev,E.A. and Arkhipova,I.R. (2009) Rotifer rDNA-specific R9 retrotransposable elements generate an exceptionally long target site duplication upon insertion. *Gene*, **448**, 145–150.
19. Shivram,H., Cawley,D. and Christensen,S.M. (2011) Targeting novel sites: the N-terminal DNA binding domain of non-LTR retrotransposons is an adaptable module that is implicated in changing site specificities. *Mob. Genet. Elements*, **1**, 169–178.
20. Thompson,B.K. and Christensen,S.M. (2011) Independently derived targeting of 28S rDNA by A- and D-clade R2 retrotransposons: plasticity of integration mechanism. *Mob. Genet. Elements*, **1**, 29–37.
21. Jamburuthugoda,V.K. and Eickbush,T.H. (2014) Identification of RNA binding motifs in the R2 retrotransposon-encoded reverse transcriptase. *Nucleic Acids Res.*, **42**, 8405–8415.
22. Wagstaff,B.J., Barnerssoi,M. and Roy-Engel,A.M. (2011) Evolutionary conservation of the functional modularity of primate and murine LINE-1 elements. *PLoS One*, **6**, e19672.
23. Burke,W.D., Malik,H.S., Jones,J.P. and Eickbush,T.H. (1999) The domain structure and retrotransposition mechanism of R2 elements are conserved throughout arthropods. *Mol. Biol. Evol.*, **16**, 502–511.
24. Doucet,A.J., Hulme,A.E., Sahinovic,E., Kulpa,D.A., Moldovan,J.B., Kopera,H.C., Athanikar,J.N., Hasnaoui,M., Bucheton,A., Moran,J.V. *et al.* (2010) Characterization of LINE-1 ribonucleoprotein particles. *PLoS Genet.*, **6**, 728–741.
25. Yang,J., Malik,H.S. and Eickbush,T.H. (1999) Identification of the endonuclease domain encoded by R2 and other site-specific, non-long terminal repeat retrotransposable elements. *Proc. Natl. Acad. Sci. U.S.A.*, **96**, 7847–7852.
26. Steczkiewicz,K., Muszewska,A., Knizewski,L., Rychlewski,L. and Ginalski,K. (2012) Sequence, structure and functional diversity of PD-(D/E)XK phosphodiesterase superfamily. *Nucleic Acids Res.*, **40**, 7016–7045.
27. 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.
28. Kinch,L.N., Ginalski,K., Rychlewski,L. and Grishin,N.V. (2005) Identification of novel restriction endonuclease-like fold families among hypothetical proteins. *Nucleic Acids Res.*, **33**, 3598–3605.
29. Middleton,C.L., Parker,J.L., Richard,D.J., White,M.F. and Bond,C.S. (2004) Substrate recognition and catalysis by the Holliday junction resolving enzyme Hje. *Nucleic Acids Res.*, **32**, 5442–5451.
30. Andreeva,A., Howorth,D., Chothia,C., Kulesha,E. and Murzin,A.G. (2014) SCOP2 prototype: a new approach to protein structure mining. *Nucleic Acids Res.*, **42**, D310–D314.
31. Pingoud,A., Fuxreiter,M., Pingoud,V. and Wende,W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol. Life Sci.*, **62**, 685–707.
32. Mukha,D.V., Pasyukova,E.G., Kapelinskaya,T.V. and Kagramanova,A.S. (2013) Endonuclease domain of the Drosophila melanogaster R2 non-LTR retrotransposon and related retroelements: a new model for transposition. *Front. Genet.*, **4**, 63.
33. Mones,L., Kulhánek,P., Florián,J., Simon,I. and Fuxreiter,M. (2007) Probing the two-metal ion mechanism in the restriction endonuclease BamHI. *Biochemistry*, **46**, 14514–14523.
34. Christensen,S. and Eickbush,T.H. (2004) Footprint of the retrotransposon R2Bm protein on its target site before and after cleavage. *J. Mol. Biol.*, **336**, 1035–1045.
35. Kierzek,E., Kierzek,R., Moss,W.N., Christensen,S.M., Eickbush,T.H. and Turner,D.H. (2008) Isoenergetic penta- and hexanucleotide microarray probing and chemical mapping provide a secondary structure model for an RNA element orchestrating R2 retrotransposon protein function. *Nucleic Acids Res.*, **36**, 1770–1782.
36. Abramoff,M.D., Magelhaes,P.J. and Ram,S.J. (2004) Image Processing with ImageJ. *Biophotonics Int.*, **11**, 36–42.
37. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
38. 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. Comput. Chem.*, **25**, 1605–1612.
39. Bond,C.S., Kvaratskhelia,M., Richard,D., White,M.F. and Hunter,W.N. (2001) Structure of Hjc, a Holliday junction resolvase, from Sulfolobus solfataricus. *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 5509–5514.
40. Pingoud,V., Sudina,A., Geyer,H., Bujnicki,J.M., Lurz,R., Lüder,G., Morgan,R., Kubareva,E. and Pingoud,A. (2005) Specificity changes in the evolution of type II restriction endonucleases: a biochemical and bioinformatic analysis of restriction enzymes that recognize unrelated sequences. *J. Biol. Chem.*, **280**, 4289–4298.
41. Kostrewa,D. and Winkler,F.K. (1995) Mg2+ binding to the active site of EcoRV endonuclease: a crystallographic study of complexes with substrate and product DNA at 2 A resolution. *Biochemistry*, **34**, 683–696.
42. Newman,M., Lunnen,K., Wilson,G., Greci,J., Schildkraut,I. and Phillips,S.E. (1998) Crystal structure of restriction endonuclease BglI bound to its interrupted DNA recognition sequence. *EMBO J.*, **17**, 5466–5476.
43. 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.
44. Horton,N.C. and Perona,J.J. (2004) DNA cleavage by EcoRV endonuclease: two metal ions in three metal ion binding sites. *Biochemistry*, **43**, 6841–6857.
45. Nishino,T., Komori,K., Ishino,Y. and Morikawa,K. (2001) Dissection of the regional roles of the archaeal Holliday junction resolvase Hjc by structural and mutational analyses. *J. Biol. Chem.*, **276**, 35735–35740.
46. Nishino,T., Komori,K., Tsuchiya,D., Ishino,Y. and Morikawa,K. (2001) Crystal structure of the archaeal holliday junction resolvase Hjc and implications for DNA recognition. *Structure*, **9**, 197–204.
47. Pingoud,A., Wilson,G.G. and Wende,W. (2014) Type II restriction endonucleases–a historical perspective and more. *Nucleic Acids Res.*, **42**, 7489–7527.
48. Daiyasu,H., Komori,K., Sakae,S., Ishino,Y. and Toh,H. (2000) Hjc resolvase is a distantly related member of the type II restriction endonuclease family. *Nucleic Acids Res.*, **28**, 4540–4543.
49. Kvaratskhelia,M., Wardleworth,B.N., Norman,D.G. and White,M.F. (2000) A conserved nuclease domain in the archaeal Holliday junction resolving enzyme Hjc. *J. Biol. Chem.*, **275**, 25540–25546.
50. 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.
51. 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.
52. Burke,W.D., Calalang,C.C. and Eickbush,T.H. (1987) The site-specific ribosomal insertion element type II of Bombyx mori (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. *Mol. Cell. Biol.*, **7**, 2221–2230.
53. Burke,W.D., Eickbush,D.G., Xiong,Y., Jakubczak,J. and Eickbush,T.H. (1993) Sequence relationship of retrotransposable elements R1 and R2 within and between divergent insect species. *Mol. Biol. Evol.*, **10**, 163–185.
54. Kojima,K.K. and Fujiwara,H. (2004) Cross-genome screening of novel sequence-specific non-LTR retrotransposons: various multicopy RNA genes and microsatellites are selected as targets. *Mol. Biol. Evol.*, **21**, 207–217.
55. Burke,W.D., Malik,H.S., Lathe,W.C. and Eickbush,T.H. (1998) Are retrotransposons long-term hitchhikers? *Nature*, **392**, 141–142.