

Rational engineering of sequence specificity in R.MwoI restriction endonuclease

Krzysztof Skowronek*, Michal J. Boniecki, Boguslaw Kluge and Janusz M. Bujnicki

Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology, Trojdena 4, 02-109 Warsaw, Poland

Received October 21, 2011; Revised April 27, 2012; Accepted May 19, 2012

ABSTRACT

R.MwoI is a Type II restriction endonuclease enzyme (REase), which specifically recognizes a palindromic interrupted DNA sequence 5'-GCNNNNNNGC-3' (where N indicates any nucleotide), and hydrolyzes the phosphodiester bond in the DNA between the 7th and 8th base in both strands. R.MwoI exhibits remote sequence similarity to R.BglI, a REase with known structure, which recognizes an interrupted palindromic target 5'-GCCNNNNNGGC-3'. A homology model of R.MwoI in complex with DNA was constructed and used to predict functionally important amino acid residues that were subsequently targeted by mutagenesis. The model, together with the supporting experimental data, revealed regions important for recognition of the common bases in DNA sequences recognized by R.BglI and R.MwoI. Based on the bioinformatics analysis, we designed substitutions of the S310 residue in R.MwoI to arginine or glutamic acid, which led to enzyme variants with altered sequence selectivity compared with the wild-type enzyme. The S310R variant of R.MwoI preferred the 5'-GCCNNNNNGGC-3' sequence as a target, similarly to R.BglI, whereas the S310E variant preferentially cleaved a subset of the MwoI sites, depending on the identity of the 3rd and 9th nucleotide residues. Our results represent a case study of a REase sequence specificity alteration by a single amino acid substitution, based on a theoretical model in the absence of a crystal structure.

INTRODUCTION

Restriction endonucleases (REases) of type II, together with cognate DNA modification methyltransferases, has been broadly used in numerous basic techniques of *in vitro*

DNA manipulations for more than 40 years (review: (1)). Type II REases recognize and cleave 4–8 bp DNA sequences with high specificity, and generate defined and predictable fragments. Although more than 3000 REases have been isolated from various bacterial strains, they represent less than 300 different sequence specificities (see eg. the REBASE database (2)). Many possible specificities are therefore still not available in nature. For many applications involving the use of REases, it would be advantageous to have access to a broader spectrum of specificities. Unfortunately, the rate of discovery of new specificities among newly characterized REases has recently decreased to only a few per year.

Substantial efforts have been made to alter the specificity of known REases by various strategies of protein engineering. Initial attempts of rational redesign of protein-DNA contacts based on crystallographic structures were largely unsuccessful (3,4). Some progress was achieved by large scale mutagenesis experiments where selected residues in R.EcoRV and R.BsoBI enzymes were randomized, and the resulting variants screened for altered specificities (5–7). Another approach was developed by Xu and coworkers, in which a three-step *in vivo* screening/selection, based on a non-cognate MTase with a partially different specificity, was used to select for R.BstYI and R.NotI variants with specificity altered toward the specificity of the non-cognate MTase (8,9). This approach, though successful in the described cases, suffers from being dependent on the availability of a MTase with sequence specificity toward which the REase is evolved.

Two most outstanding successes, in which REases with new specificities have been obtained, involved Type IIC REases—enzymes that encode both the MTase and REase activity in one polypeptide chain, and use the same domain to regulate the specificity of both activities. Janulaitis and coworkers obtained a variant of R.Eco57I with a relaxed specificity by inactivating the REase activity with a single substitution in the nuclease domain, error-prone PCR mutagenesis and selection of

*To whom correspondence should be addressed. Tel: +48 22 597 0754; Fax: +48 22 597 0715; Email: krzysiek@genesilico.pl
Correspondence may also be addressed to Janusz M. Bujnicki. Tel: +48 22 597 0750; Fax: +48 22 597 0715; Email: iamb@genesilico.pl

variants with a relaxed methylation specificity (10). In the final step, the REase activity of the modified enzyme was restored by reversing the substitution in the nuclease domain, leading to a bifunctional MTase-REase with a new, relaxed specificity. Morgan and coworkers conducted a comparative analysis of amino acid sequences in a relatively large group of close homologs of R.MmeI with diverged recognition sequences and deciphered a DNA sequence recognition “code” pertinent to these proteins (11). Rational mutagenesis based on this recognition code allowed for engineering of REases with new “mosaic” specificities (12).

The aforementioned successes notwithstanding, sequence specificity engineering of REases remains difficult, and the number of engineered enzymes remains limited. This difficulty stems out of two features of REases. First, experimentally determined structures of REase-DNA complexes reveal that these enzymes use redundant protein-DNA contacts for sequence recognition and often couple recognition to catalysis (13). Therefore, alterations of individual protein-DNA contacts often disrupt the overall network of interactions, and the changes in binding specificity may additionally impair the enzymatic activity. Second, REases are a paradigm for extreme diversity of protein sequences and structures. Even though most of Type II REases possess a catalytic domain from the PD-(D/E)XK superfamily, which groups together domains with a common structural core and a pattern of weakly conserved residues (14), their sequences are rarely conserved. Type II REases often exhibit variations of the active site, including relocation of catalytic residues to different positions in the structure, and they accumulate a variety of sequence insertions and substitutions of secondary structure elements around the minimal structural core (15–17). Typically, their amino acid sequences reveal little or no significant sequence similarities, except for enzymes that exhibit identical or similar recognition and cleavage specificities (18,19).

Our group has developed bioinformatics methodology to predict the structures of type II REases from protein sequence (20). For a number of published predictions, independent crystallographic studies have validated the identification of the 3D-fold and functionally important residues (21–24). Here, we report the development of a structural model for R.MwoI REase, its experimental verification and application to guide a successful engineering of a variant with a new substrate preference.

R.MwoI is type IIP REase isolated from *Methanothermobacter wolfeii* (formerly *Methanobacterium wolfeii*) (25), which recognizes the interrupted palindromic sequence 5'-GCNNNNNNGC-3' (where N indicates any nucleotide) and hydrolyzes the phosphodiester bond in the DNA between the 7th and 8th N base in both strands to generates 3nt long 3' overhangs upon cleavage. We have previously identified the membership of R.MwoI in the PD-(D/E)XK superfamily of nucleases (26), which suggested the identity of its catalytic residues. However, only now we were able to construct a 3D model, which enabled us to predict residues of R.MwoI important for the recognition of its target sequence. The identification of residues close to non-discriminated base pairs

(N) in the R.MwoI recognition sequence suggested a starting point for introducing novel specific contacts with the substrate DNA.

MATERIALS AND METHODS

Structural bioinformatics analyses

Sequence database searches were carried out with PSI-BLAST (27) in the following databases: REBASE (28), the non-redundant database of amino acid sequences, and the DNA sequences from unfinished genomics and metagenomics projects at the National Center for Biotechnology Information (29). A multiple sequence alignment was calculated using PROMALS3D, to take into account the structural information (30). Protein structure prediction was carried out using the GeneSilico MetaServer (31). Fold-recognition (FR) alignments between the sequence of R.MwoI and the structures of the best templates identified by different FR methods (R.BglI and R.SfiI) were used to carry out comparative modeling using the “FRankenstein’s Monster” approach (32,33), which comprises cycles of local realignments in uncertain regions, building of alternative models and their evaluation, realignment in poorly scored regions and merging of the best scoring fragments. Earlier, we used this approach for successful prediction of the structure and accurate identification of protein-DNA contacts of R.SfiI (21), later confirmed by the crystallographic study (34), as well as for many other restriction enzymes and other proteins. For the evaluation of models, we used MetaMQAP (35) and PROQ (36), a tool for predicting the deviation of individual residues in the model from their counterparts in the native structure.

Modeling of the protein-DNA complex

Modeling of R.MwoI in complex with cognate DNA was performed with the assumption that R.MwoI binds DNA in a similar way to its homologs, R.SfiI and R.BglI (whose DNA recognition modes are also similar to each other). We selected R.BglI as the primary modeling template for the protein-DNA complex because its sequence is more similar to R.MwoI (24% vs 14% in the case of R.SfiI); its structure was assessed as a better template for R.MwoI by the FR methods (PCONS score 0.60 compared with 0.49 for R.SfiI), and preliminary models based on this single template exhibited higher MetaMQAP scores (data not shown). The R.BglI DNA substrate is also a valid substrate of R.MwoI.

First, the model of R.MwoI (initially the “core” lacking the non-conserved insertion) was superimposed onto the structure of R.BglI bound to its substrate DNA (37) by minimizing the root mean square deviation of C α atoms in regions of protein-DNA contacts in the R.BglI-DNA complex. The DNA from the R.BglI structure was merged with the R.MwoI model. R.MwoI sequence includes an insertion comprising residues ~183-280, which has no counterpart in the templates. Therefore, it was modelled ‘de novo’ using a modified version of the REFINER program (38), with restraints on predicted secondary structure, and symmetry enforced between the two

subunits in the dimer. A conformation without steric clashes between the remodeled part of the protein and the DNA, and with acceptable MetaMQAP and PROQ scores was regarded one representing the “final” working model. PyMol (39) and Swiss-PDBViewer (40) were used as graphical interfaces for inspection and manipulation of the structures, and to generate rendered figures.

Cloning and mutagenesis of the *mwoIR* gene

The genomic DNA of *Methanothermobacter wolfeii* (formerly *Methanobacterium wolfeii*) (41) was isolated by phenol extraction and ethanol precipitation from actively growing culture supplied by the German Collection of Microorganisms and Cell Cultures (DSMZ). *MwoI* methyltransferase gene (*mwoIM*) was amplified from this genomic DNA in a PCR reaction and cloned into pAlterEx2 vector (Promega) as Mph1103I-PvuI insert in *E. coli* XL1-Blue MRF' Δ (mcrA)183 Δ (mcrCB-hsdSMR-mrr)173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac [F' proAB lacI^q Δ M15 Tn10 (Tet^r)] (Stratagene) strain. The resulting plasmid pMwoMAx provided methylation of the *MwoI* cognate sequences in the host strain DNA. The *mwoIR* gene was amplified in a PCR from the *M. wolfeii* DNA and cloned into pET28a plasmid (Novagene) as the NcoI-XhoI insert. A ligation mixture was transformed to the host strain carrying pMwoMAx plasmid. The resulting plasmid pMwoRHis contained the coding sequence for R.MwoI, with a C-terminal hexahistidine tail, under control of the T7 promoter.

All mutations were introduced by site-directed mutagenesis in PCR reactions, by amplifying the whole pMwoRHis plasmid. Modifications of the gene were introduced by 5' sequences of the primers. PCR reaction mixtures were digested with the R.DpnI enzyme to remove the template DNA, treated with T4 polynucleotide kinase to phosphorylate 5' ends, after which the DNA was circularized with the T4 DNA ligase. When possible, mutagenic primers contained new restriction sites introduced as silent mutations to facilitate identification of mutants. Ligation mixtures were introduced to host strain carrying pMwoMAx plasmid. All recombinant plasmids were sequenced to confirm accordance with the design.

Protein expression and purification

Initially proteins were overexpressed in *E. coli*, strain ER2566 F⁻ fhuA2 [lon] ompT lacZ::T7 gene1 gal sulA11 (mcrC-mrr)114::IS10 R(mcr-73::miniTn10-Tet^S)2 R(zgb-210::Tn10)(Tet^S) endA1 [dcm] (New England Biolabs) carrying both pMwoMAx and pMwoRHis plasmids. Mid-log cultures were induced with 1 mM IPTG, and bacteria were harvested 3–5 hours after induction at 37°C. The expression level in this system was highly variable from experiment to experiment (data not shown). More consistent results were obtained using host strain BL21(DE3) (Novagene) carrying only pMwoRHis plasmid grown for 36 h at 25°C in auto-induction medium (42). In this system, there was no DNA modification enzyme present, but cells were grown at low temperature, where R.MwoI does not exhibit its

endonuclease activity (the optimal reaction temperature for this enzyme is 60°C). According to the New England Biolabs catalog, R.MwoI exhibits only 10% activity at 37°C. We could not detect any λ DNA cleavage *in vitro* at 25°C after 5 h with 30 ng of enzyme and 1 μ g of DNA (data not shown).

Frozen cell pellet from 500 ml culture was resuspended in buffer A (50 mM HEPES pH 8.0, 0.5 M NaCl, 10 mM β -mercaptoethanol, 1 mM imidazole pH 8.0, 1 mM PMSF, 10% glycerol, 0.1% Triton X-100) and lysed by a single passage through French press at 20 000 psi. Lysate was clarified by centrifugation, and recombinant protein was batch bound to 500 μ l HIS-Select affinity gel (Sigma) for 15 min. at 4°C. Resin was washed once with 5 ml of buffer A and applied on a disposable gravity flow column, where it was washed, consecutively, with 5 volumes of buffer B (buffer A with 10 mM imidazole), buffer C (buffer B without Triton X-100), buffer D (50 mM HEPES pH 8.0, 2 M NaCl, 10 mM β -mercaptoethanol), buffer E (50 mM HEPES pH 8.0, 150 mM NaCl, 10 mM β -mercaptoethanol, 10% glycerol) and buffer F (buffer C with 20 mM imidazole). Recombinant protein was eluted with buffer G (buffer C with 250 mM imidazole) in 500 μ l fractions. Fractions 2 and 3, containing most of the eluted protein were pooled, glycerol was added to 50%, and samples were stored at -20°C. REases used in cleavage preference experiments were further purified on Superdex 200 size exclusion chromatography. Protein samples were run in 10% SDS-PAGE to assess purity (homogeneity range from 60 to >90%) and to measure concentration of R.MwoI in each preparation by densitometry.

Construction of substrates with single *MwoI* sites

To characterize specificity toward positions 3 and 9 of the recognition sequence, which are not recognized by the wild-type enzyme, short substrates with single *MwoI* sites were prepared. Double stranded oligonucleotides containing *MwoI* sites with different palindromic sequences at positions 3 and 9 were cloned into pTZ19 vector cleaved with SmaI enzyme. Resulting plasmids pMwoSub_AT, pMwoSub_TA, pMwoSub_GC and pMwoSub_CG contained inserts GCAATCTCTGC, GCTATCTCAGC, GCGATCTCCGC and GCCATCTCGGC, respectively (underlined are positions 3 and 9 of *MwoI* sites). A total of four 316 bp fragments of these plasmids carrying introduced *MwoI* sites in position 140 (1Mwo substrates) were amplified in PCR reaction and used in sequence specificity studies.

Cleavage assay

The λ DNA *in vitro* cleavage assays were set up in 30 μ l volume of buffer O (MBI Fermentas) [50 mM Tris-HCl (pH 7.5), 100 mM NaCl, 10 mM MgCl₂, 0.1 mg/ml BSA] with 2 μ g of λ DNA (dam⁻ dcm⁻) (MBI Fermentas) and various amounts of purified R.MwoI. Reaction was carried out for 2 h at 60°C. Cleavage products were resolved by agarose gel electrophoresis. Cleavage assays with plasmid DNA substrates were performed for 5 hours. Assays with 1Mwo substrates were carried out in the same buffer and temperature with DNA at 10 ng/ μ l and various

amounts of purified enzymes. For kinetic measurements, aliquots were taken for time intervals ranging from 5 minutes to 7.5 hours, and reaction was stopped by addition of EDTA to 20 mM concentration. Cleavage rates were calculated from the slope of reaction progress measured by agarose gels densitometry in the time range where the slope was linear. The measurements were repeated three times.

To characterize sequence preference of R.MwoI variants, products of the presumably complete cleavage (150 ng of enzyme per 1 µg of substrate for 5 hours i.e. amounts of enzyme higher and reaction time longer than sufficient to obtain a stable cleavage pattern) and partial cleavage (7 ng of enzyme per 1 µg of substrate for 1.5 hour i.e. leading to the cleavage of only 'the most easily cleavable sites') were analysed. The products obtained were treated with T4 DNA polymerase to remove 3' overhangs and cloned into SmaI site of pKS II Bluescript vector (Stratagen) for sequencing.

Plasmid cleavage interpretation

To investigate the correlation between the lack of plasmid DNA cleavage by the S310E and S310R enzyme variants and a particular nucleotide sequence at positions 3 and 9 of the MwoI sites, a statistical analysis was conducted that focused on easily interpretable cleavage products. Firstly, the cleavage products obtained by digesting a given plasmid with the R.MwoI310wt enzyme were matched with the expected size of fragments derived from cleaving that plasmid in all MwoI sites. Secondly, the corresponding bands were assigned to cleavage products generated by the S310E and S310R variants. The identified bands were used to create two data sets (for the S310E and S310R variants having 44 and 45 observations, respectively, in total for all three plasmids, i.e. pUC19, ΦX174 and pBR322).

An MwoI site with nucleotides X and Y at positions 3 and 9 is here referred to as an X-Y type site. We do not discriminate between types that are reverse complements (e.g. A-A and T-T), and therefore only 10 types are considered. Each data item in the dataset consisted of the types of the MwoI sites on each end of a particular fragment and a binary variable, indicating whether the fragment was observed. For instance, a triplet (AA, GC, 1) indicates that the fragment is flanked by A-A and G-C type MwoI sites, and that such a fragment was observed on the gel). All 1024 (i.e. 2^{10}) hypotheses corresponding to the subset of site types that the enzyme digests were considered. A penalty was assigned for each hypothesis for observing fragments inconsistent with it, according to the rules in Table 1. For the next step, hypotheses with the lowest penalty were selected, that is, 27 hypotheses for S310E and 14 hypotheses for S310R, below the arbitrarily chosen penalty threshold of 14. For each site type, the fraction of the number of hypotheses that consider the type as cleaved were calculated. Types with frequencies below 0.2 were predicted not to be cleaved.

Table 1. Penalties for observing and not observing a single fragment in the plasmid cleavage analysis

Hypothesis	Fragment observed	Fragment not observed
Fragments flanked by MwoI sites with two different types		
Both sites digested	0	1
Exactly one site digested	1	0
No site digested	2	0
Fragments flanked by MwoI sites with the same types		
Both sites digested	0	1
No site digested	1	0

Each hypothesis implies whether both, one or no MwoI sites flanking a particular fragment are digested. The penalties are summed up for all fragments to produce one value – the penalty for the hypothesis.

DNA-binding assay

The relative specific DNA-binding activity of R.MwoI variants was measured with the nitrocellulose filter binding assay (43). An oligonucleotide with the MwoI target sequence (CGGCTGTAGTCGCAC) was radiolabelled with [γ - 33 P]ATP and T4 polynucleotide kinase, annealed to the complementary oligonucleotide and purified on Sephadex G-25 to remove the unincorporated label. Enzyme—DNA mixtures (50 µl total volume) were set up in the binding buffer (50 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM CaCl₂, 10 µg/ml BSA) containing 1 nM of radiolabelled oligonucleotide with the target sequence, 100 nM of unlabelled double stranded oligonucleotide with a scrambled target sequence (CGCGTG TAGTCCGAC) and 2 nM R.MwoI. Reactions were incubated 30 min. at 25°C and filtered through 0.22 µm nitrocellulose filter (Whatmann) in Dot-Blot apparatus (Bio-Rad). Each well was washed three times with 200 µl of the binding buffer. Dried filters were exposed to a phosphorimager screen overnight. Images were scanned on a Storm Phosphorimager, and the retained radioactivity was quantified using ImageQuant software (Amersham). For each protein variant, triplicate measurements were made.

Equilibrium binding constants were measured for panel of 4 oligonucleotides with MwoI sites with different sequences in positions 3 and 9 (GACAGTGCTATCTCAG CGACAGT, GACAGTGC^AAATCTCT^TGCG^AACAGT, ^GACAGTGC^GGATCTCCGCG^AACAGT and GACAGTGC^AATCTCT^TGCG^AACAGT, where bold indicates residues recognized by the wt R.MwoI). Measurements were performed as described earlier for relative binding, except that reactions were set up with only radiolabelled oligonucleotides at concentrations 6 pM or 0.5 pM in volumes 50 µl or 600 µl, respectively, with enzyme concentrations 1 pM - 500 nM. Equilibrium binding constants were calculated through fitting of the data to one-site specific binding model.

RESULTS

Sequence analysis of R.MwoI and its homologs

A search of the combined non-redundant and env sequence databases (see Methods) identified 22 sequences

similar to R.MwoI, including 15 full length proteins with the characteristic D-X_N-D-X-K variant of the eponymous PD-(D/E)XK motif present in many REases (20). Homologs detected with significant scores include a bona fide REase R.BglI (target site GCCNNNNNGGC) with the e-value of 6e-09, as well as a number of experimentally uncharacterized proteins (putative REases). We have also searched for R.MwoI homologs using the fold-recognition approach, which allows for the identification of homologs among proteins with known structure even in the absence of evident sequence similarity (review: (20); see Methods for details). This search confirmed the match of R.MwoI sequence to R.BglI sequence and structure, as well as identified a structurally related enzyme R.SfiI (target site GGCCNNNNNGGCC) as another, more distantly related homolog.

The analysis of fold-recognition results and the multiple sequence alignment of the aforementioned sequences (Figure 1 and Supplementary File 1) revealed an unambiguous match between the N-terminal region of R.MwoI (including putative catalytic residues D97, D110, and K112) and the PD-(D/E)XK motif of R.BglI and R.SfiI. Additionally, E71 in R.MwoI matched an auxiliary metal-binding residue present in R.BglI and R.SfiI (E87 and E55, respectively). However, two problems made the modelling of R.MwoI structure challenging. First, the central part of R.MwoI (aa ~190–270) corresponds to a long insertion that is absent from other proteins in the

considered protein family. Second, for the C-terminal part of R.MwoI that is expected to include some of the DNA-recognizing residues, different fold-recognition methods generated alternative sequence alignments that disagreed with each other (data not shown). R.BglI and R.SfiI recognize similar DNA sequences and make similar contacts to the common residues in their targets. In particular, a conserved Arg residue (R279 in R.BglI and R220 in RSfiI) interacts with the 5'-terminal G in the GCC half-site recognized by R.BglI, and a conserved Lys residue (K266 in R.BglI and K208 in RSfiI) interacts with a G base-paired with the middle C. The DNA target of R.MwoI includes the corresponding GC dinucleotide and interestingly, most of the alignment variants, despite other differences, included matches of these Lys and Arg residues in the templates with Lys and Arg in R.MwoI, in particular K305A and either R312A or R332.

Experimental test of the importance of residues in R.MwoI that matched functionally relevant residues in R.BglI/R.SfiI

To test the predicted sequence match between R.MwoI and R.BglI/R.SfiI, a series of mutants in the *mwoIR* gene was constructed to introduce substitutions of R.MwoI residues predicted to be involved in either catalysis or DNA binding. To analyze the importance of a

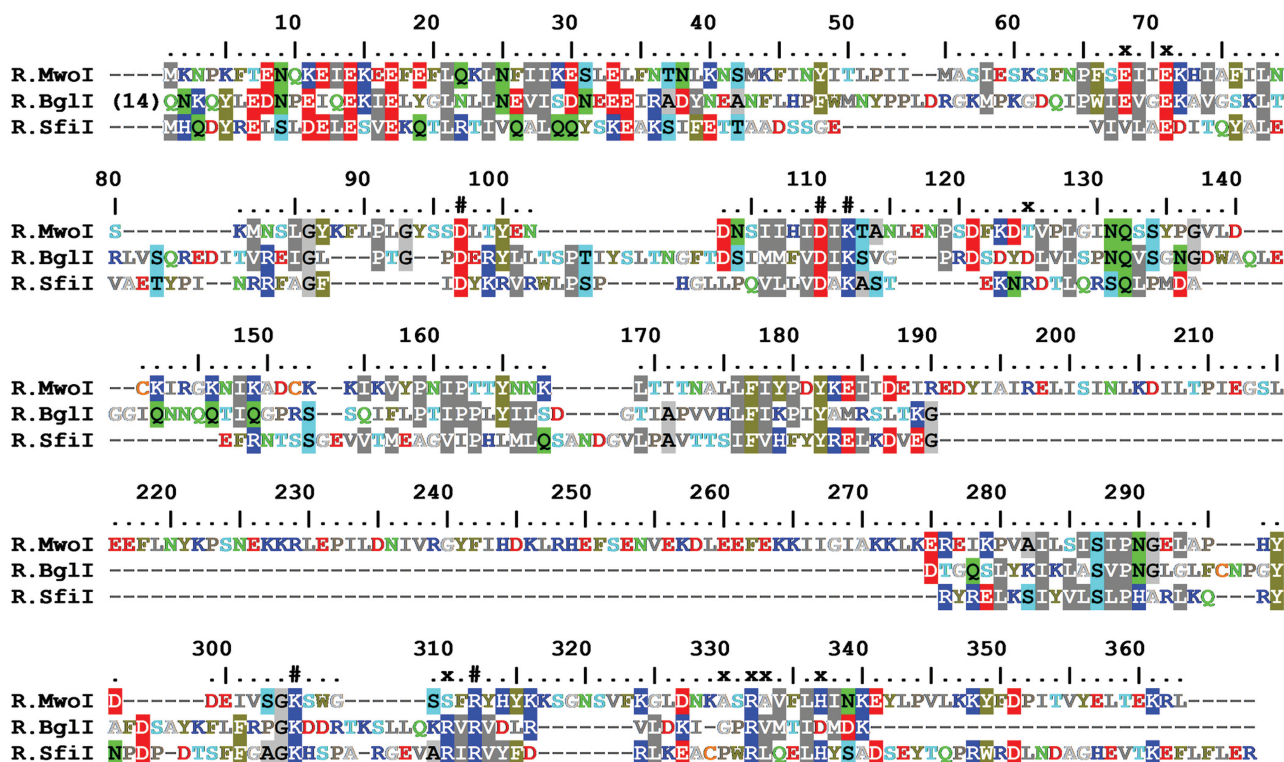


Figure 1. Sequence alignment of R.MwoI, and REases with known structures: R. BglI, and R.SfiI. Residues are colored according to the physico-chemical character of the *r* side-chains (aliphatic: grey, aromatic: yellow grey, Ala and Gly: light grey, basic: blue, acidic: red, amide groups: green, hydroxyl groups: cyan). Residues with a similar type (according to the BLOSUM62 similarity matrix) are highlighted. R.MwoI residues found to be essential for activity are indicated by “#” above the alignment; residues found not to be essential in R.MwoI are indicated by “x”.

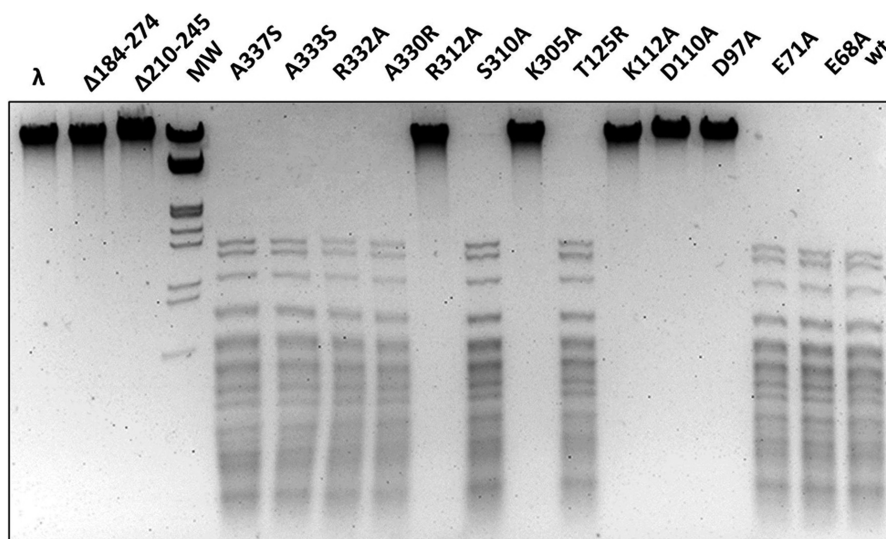


Figure 2. *In vitro* cleavage activity of R.MwoI variants. Bacteriophage λ DNA was digested with partially purified REases (see Methods) and resolved in a 2% TAE-agarose gel. Amino acids substitutions and regions deleted (Δ) are shown above each lane. wt – wild-type enzyme, λ – uncleaved substrate, MW – molecular weight marker: λ DNA digested with EcoRI and HindIII.

region encompassing residues E184-K274, two additional variants were created, one with a deletion of this whole region and another with a smaller deletion of residues P210-D245. Most of point mutations introduced an alanine substitution in the protein variant, which results in truncation of the side chain beyond C- β atom, making the changed residue chemically inert, thereby allowing the determination of importance of particular functional groups for the enzyme activity. We also tested several other substitutions of R.MwoI residues aligned with DNA-binding or otherwise important residues of R.BglI and R.SfiI. The effect of mutations and hence the importance of the selected residues for activity of R.MwoI was analysed using two assays:

Assay 1: The *in vitro* cleavage assay was used to determine the endonuclease activity. The results allowed us to divide R.MwoI variant into two groups: those without obviously decreased cleavage activity (substitutions E68A, E71A, T125R, S310A, A330R, R332A, A333S, H337A), and variants with cleavage activity completely abolished (D97A, D110A, K112A, K305A, R312A and both deletion mutants) (Figure 2). The cleavage patterns of active variants were not discernible from the wt enzyme, indicating that no change in substrate specificity occurred. The set of inactivated variants did not produce any cleavage activity even at enzyme concentration 100 times higher than used in regular assay conditions (data not shown).

Assay 2: The filter-binding assay was used to measure relative binding affinities of R.MwoI variants for DNA with the MwoI target sequence. For the endonucleolytically active variants, the binding ranged from 42% to 370% of the wt level (Figure 3). This group of variants can be divided based on DNA binding assay into three groups: variants with substantially increased DNA binding (>350%; E68A and A330R), those with slightly increased DNA binding (139–181%; E71A, R332A, A333S and

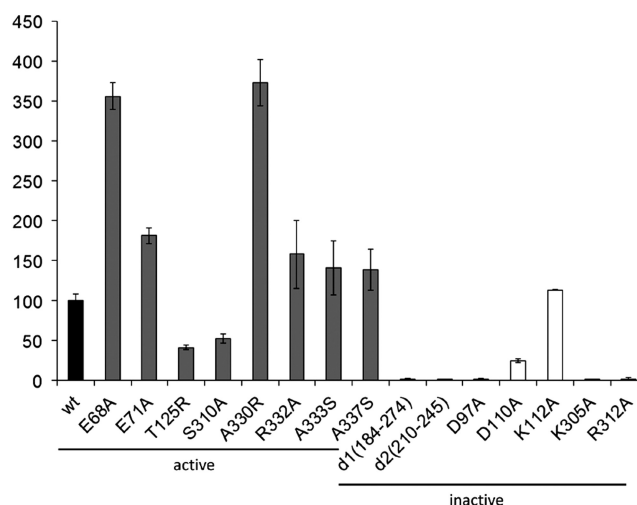


Figure 3. Relative DNA binding activity of R.MwoI variants measured using a filter binding assay (see Methods). The binding level is expressed as percentage of that of the wt enzyme. The cleavage activity of variants is indicated below the plot (active/inactive).

A337S) and two variants with decreased DNA binding (42% and 53%; T125R and S310A, respectively). In a subset of variants defective in DNA cleavage, only two exhibit significant DNA binding: D110A (25%) and K112A (114%).

The aforementioned results validated the predicted active site of R.MwoI: residues D97, D110 and K112 of the PD-(D/E)XK motif are essential for the cleavage activity, and two of them are not essential for DNA binding, similarly to what has been documented for many related REases. On the other hand, residue E71 was not essential for R.MwoI cleavage activity, despite its match with a metal-binding Glu residue in the templates. However, the acidic residue at this position is

conserved in only about half of the known members of the PD-(D/E)XK superfamily (16,44), and it was shown to be dispensable for activity in some REases, such as R.EcoRV (45) or R.HpaI (46). Similarly, E68 predicted to be located on the same face of α -helix, as E71 was also dispensable for R.MwoI activity. The experimental results confirmed the prediction of critical DNA-binding residues, K305 and R312, and falsified the hypothesis that R332 may be responsible for DNA recognition. Finally, we found that the unique insertion in R.MwoI was essential for its activity, despite its absence from R.BglI, R.SfiI and other related sequences.

Modeling of R.MwoI structure

To obtain deeper insight into protein–DNA interactions than possible from sequence alignment alone, we have generated a comparative model of R.MwoI, based on the experimentally validated correspondence between R.MwoI and structures of R.BglI and R.SfiI. The long insertion in R.MwoI could not be modelled using the template-based approach; therefore, it was initially omitted from the analysis and inserted only at the later stage. Models without the insertion were constructed according to the FRankenstein's monster approach (32) (see Methods for details), by iterating the homology modelling procedure (initially based on the raw FR alignments), evaluation of the sequence-structure fit by MetaMQAP, merging of fragments with best scores and local realignment in poorly scored regions. The modelling was initiated with alignments that were compatible with experimental results (in particular with K305A and R312A aligned to DNA-binding residues of the templates). Local realignments were constrained to maintain the overlap between the secondary structure elements found in the template structures, and predicted for R.MwoI. The iterative modelling procedure was stopped when all regions in the protein core obtained acceptable scores. The initial models of R.MwoI without the insertion obtained a PROQ score of 3.3, which indicates a 'very good model', and MetaMQAP predicted that the model's root mean square deviation to the native (presently unknown) structure is about 4Å. These scores indicate that the predicted protein fold and positions of most residues are most likely correct, but the atomic details (in particular conformations of the side chains) must be taken with a grain of salt.

We generated a model of R.MwoI dimer by orienting the two protomers in the same way as in the R.BglI dimer; we have also included a substrate DNA by simply copying the coordinates from the R.BglI-DNA complex structure, followed by minor optimization of side chains at the protein–protein and protein–DNA interface (see Methods for details). As the final step, we added the missing residues 183–280 by modelling them de novo with REFINER while keeping the remaining part of R.MwoI intact. The insertion was predicted to exhibit mostly α -helical conformation, and in according to REFINER, it folded as an extended arm. In some models, the helical arms from two protomers assumed an open conformation, and in others, they formed a

contact with each other, encircling the bound DNA (data not shown). However, we cannot exclude the possibility that R.MwoI may form higher order oligomers, for example, tetramers like R.SfiI, and that the unusual insertion may be involved in their stabilization. Modelling such assemblies is however beyond the limits of the currently available software and out of scope of our present work.

As expected from comparative modelling, the catalytic domain of R.MwoI shares many features of the template structures of R.BglI and R.SfiI, in particular, the overall fold and the residues of the active site motif (Figure 4 and Supplementary Figure 2). The PD-(D/E)XK catalytic domain is characterized by an antiparallel β -sheet, in which the common active site is located, and which serves as a framework for evolutionarily variable decorations. R.MwoI, R.BglI and R.SfiI belong to the β ('EcoRV-like') lineage (47,48) of PD-(D/E)XK nucleases that typically possess two features: an antiparallel β -hairpin (rather than a parallel $\beta\alpha\beta$ motif) positioned C-terminally to the strand that harbors the 'DXK' half-motif in the common β -sheet, and an additional β -sheet used as a scaffold for side-chains involved in protein–DNA recognition (48). As expected from a comparative model, catalytic residues D97, D110 and K112, and DNA-binding residues K305 and R312, assume similar orientations in the model as in the template structures.

Because the target sites of R.MwoI and R.BglI or R.SfiI are not identical, we analysed regions of the R.MwoI model that were in spatial proximity to DNA bases recognized by R.BglI or R.SfiI, but not recognized by

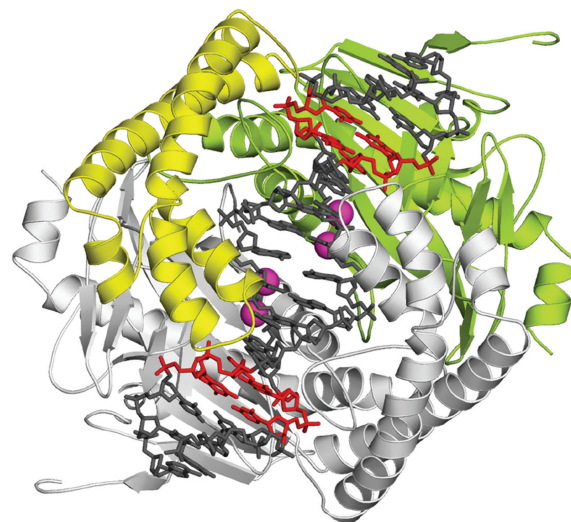


Figure 4. A model of R.MwoI dimer bound to the substrate DNA. Protein backbone is shown in the "cartoon" representation. One of the protomers is shown in light grey, in the other protomer the conserved PD-(D/E)XK domain (modeled "by homology" to R.BglI and R.SfiI) is colored in green, and the unique insertion (modeled "de novo") is shown in yellow. DNA is shown as sticks, nucleotide residues recognized by the wt R.MwoI are indicated in red. Tentative positions of metal ions (corresponding to Ca^{2+} ions in the R.BglI-DNA cocrystal structure) are shown as magenta spheres. Coordinates of the model are available for download from <ftp://genesilico.pl/iamb/models/R.MwoI/>.

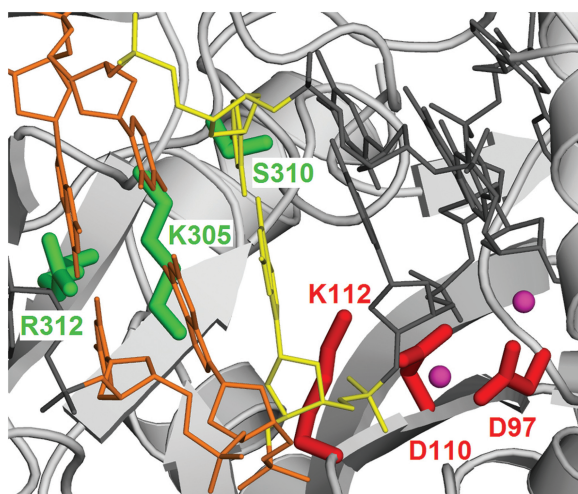


Figure 5. R.MwoI residues found to be functionally relevant. Protein backbone is shown in the “cartoon” representation (a part of the structure is omitted for the clarity of the presentation). Catalytic residues are shown as red sticks. Tentative positions of metal ions (corresponding to Ca^{2+} ions in the R.BglI-DNA cocrystal structure) are shown as magenta spheres. Residues at positions found to be important for DNA binding are shown in green. DNA is shown as thin sticks; nucleotide residues recognized by the wt R.MwoI are indicated in orange. S310 is not directly involved in recognition of the DNA sequence, but R, K, E or D at this position can recognize a G-C or a C-G base pair shown in yellow, which is not recognized by the wt enzyme.

R.MwoI. We found that a residue S310 in R.MwoI, which is dispensable for its cleavage activity, occupied a position analogous to R.BglI R277, which is involved in the recognition of the G base-paired with the 3'-terminal C of the GCC half-site in the BglI target (Figure 5). This correspondence suggested that R.MwoI could be engineered to develop a specific contact to this G-C base pair (corresponding to positions 3 and 9 in the wt R.MwoI target sequence).

Engineering of R.MwoI variants with altered sequence specificity

To test the possibility that position 310 of R.MwoI may be altered to discriminate between different substrates, S310 was substituted with R, K, E and D. These changes were introduced into the A330R variant of R.MwoI (hereafter referred to as 310 wt), which has the same activity as the wt enzyme (the A330R substitution is caused by a mutation that removes the unique MwoI site in the *mwoIR* gene sequence). The resulting four variants of R.MwoI (S310R/A330R, S310K/A330R, S310E/A330R and S310D/A330R) were initially screened in the λ DNA cleavage assay, where they all showed a reduced activity, and an altered banding pattern compared with the wt enzyme (Figure 6). Unfortunately, the observed λ DNA cleavage patterns were inconsistent with any defined specificity toward positions 3/9 (including the selection of purine vs pyrimidine or base pair A/T vs G/C). This result suggested that either the aforementioned substitutions increased the specificity toward some other position in MwoI target sequence or the corresponding enzyme

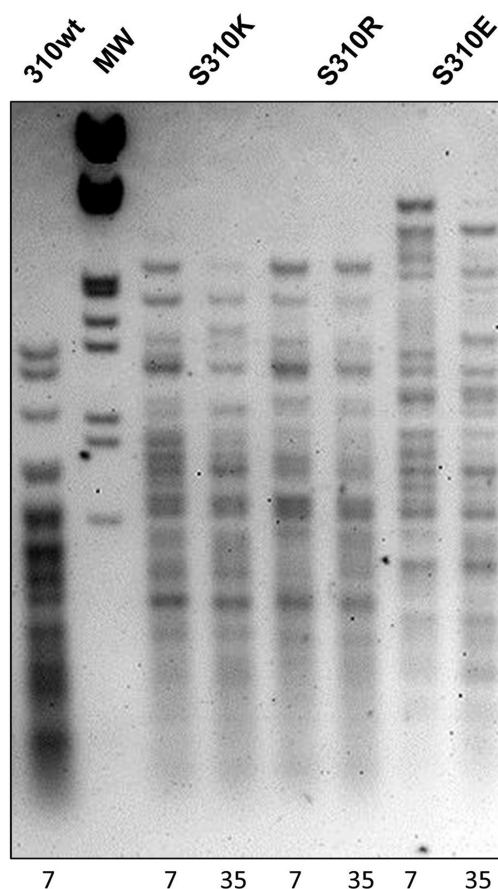


Figure 6. Pattern of λ DNA cleavage by MwoI variants with substitutions at residue 310. 310 wt – R.MwoI A330R variant, MW – molecular weight standard (phage λ DNA cleaved with HindIII and EcoRI); numbers below each lane indicate the amount of enzyme (in ng) in a 30 μ l reaction.

variants had different preferences toward different MwoI sites, resulting in substantial differences in the cleavage rates.

We have also tested other R.MwoI variants with S310 substituted with H, F, Y, N or Q. The S310F variant's activity was severely impaired. All other variants showed a decreased activity comparable with S310R and S310E, but without any changes in the λ DNA cleavage pattern (data not shown).

Further analyses were done for a set of four 1Mwo substrates with palindromic sequences at positions 3 and 9 (see Materials and Methods for details). The cleavage of this set of substrates showed clear discrimination between sequences that differ in the identity of residues at positions 3 and 9 (Figure 7). Variants with substitutions S310R or S310K selectively cleaved a substrate containing the GCCNNNNNGGC sequence, whereas variants with substitutions S310E or S310D preferred the substrate containing the GCGNNNNNCGC sequence. The S310D variant had lower selectivity than S310E, as some cleavage of GCANNNNNTGC sequence was detectable (data not shown). Cleavage rate measurements confirmed high selectivity of S310E and S310R variants for the position

not recognized by the wt R.MwoI (Figure 8). In this assay, a prolonged reaction with high concentration of either of the engineered REases allowed to measure a weak activity of each variant on GCTNNNNNAGC sequences, but not on the two other variants of the original MwoI target (S310E did not cut GCANNNNTGC and GCCNNNNNGGC, and S310R did not cut GCANNNNTGC and GCGNNNNNCGC). The cleavage rate of GCTNNNNNAGC sequence by S310E and S310R variants was 20- and 50-fold slower than the cleavage of substrates preferred by these variants, that is, GCCNNNNNGGC and GCGNNNNNCGC, respectively. Taking into account the estimated sensitivity level of this test, our result indicates that S310E and S310R cleave their preferred variant of the MwoI site (GCGNNNNNCGC and GCCNNNNNGGC respectively) at least 3 orders of magnitude faster than the two other substrates in the set (GCANNNNTGC and GCCNNNNNGGC or GCANNNNTGC and GCGNNNNNCGC, respectively).

According to the analysis of DNA binding in the presence of a nonspecific competitor, R.MwoI variants exhibited binding preferences similar to those observed for cleavage (Figure 9). Variants S310R, S310K and

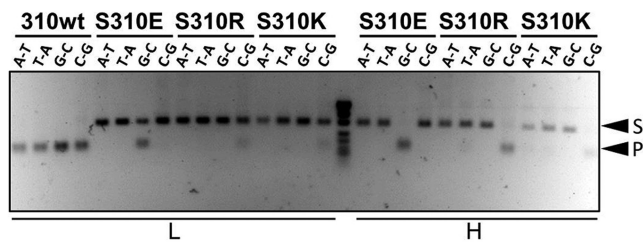


Figure 7. Cleavage of the 1MwoI set of substrates by 310wt R.MwoI and its variants. Sequences of MwoI sites in substrates: A-T = GCAATCTCTGC, T-A = GCTATCTCAGC, G-C = GCGATCTCCGC, C-G = GCCATCTCGGC. S = substrate, P = product. L and H indicates amount of enzyme in reaction (30 μl volume)—7 ng and 35 ng respectively.

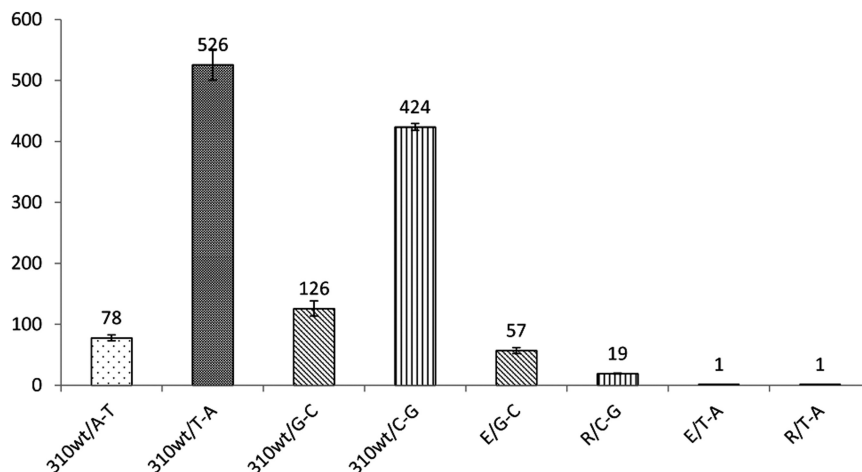


Figure 8. Cleavage rate of the 1MwoI set of substrates by 310wt R.MwoI and its variants. Values are expressed as ng of substrate cleaved per 1 ng of enzyme per 1 minute.

S310E bound their preferred sequence with affinity at least 25 times higher than the three other sequences in the panel. The only exception was the S310D variant, which bound not only its preferred sequence GCGNNNNNCGC, but also GCANNNNTGC. This preference in substrate binding was confirmed by measuring equilibrium binding constants (Table 2). The K_D of variants S310R and S310E for their respective preferred sequences was slightly lower compared with the wt R.MwoI. The wt R.MwoI exhibited similar affinities for the four different substrates; however, the S310R variant exhibited a 100-fold preference for the GCCNNNNNGGC sequence compared with the other three sequences in the set, and the S310E variant exhibits a ~1000 preference for the GCGNNNNNCGC sequence compared with other substrates tested.

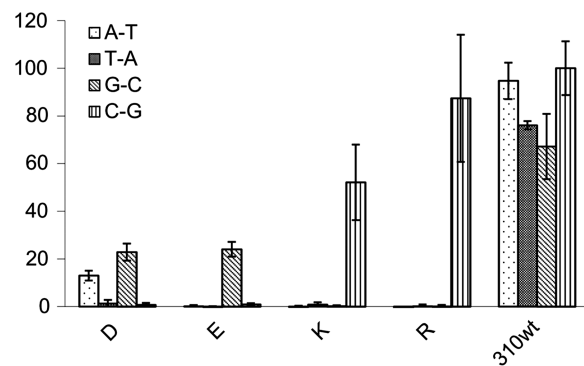


Figure 9. Relative binding of different MwoI sequences by the 310wt R.MwoI enzyme and its variants. Values measured in a filter binding assay in the presence of 100-fold excess of the oligonucleotide without MwoI recognition sequence were normalized against the binding of C/G oligonucleotide by the 310wt R.MwoI (considered as 100%). A/T = GCAATCTCTGC, T/A = GCTATCTCAGC, G/C = GCGATCTCCGC, C/G = GCCATCTCGGC.

Table 2. Equilibrium binding constants measured for selected variants of R.MwoI and panel of MwoI cognate sequences

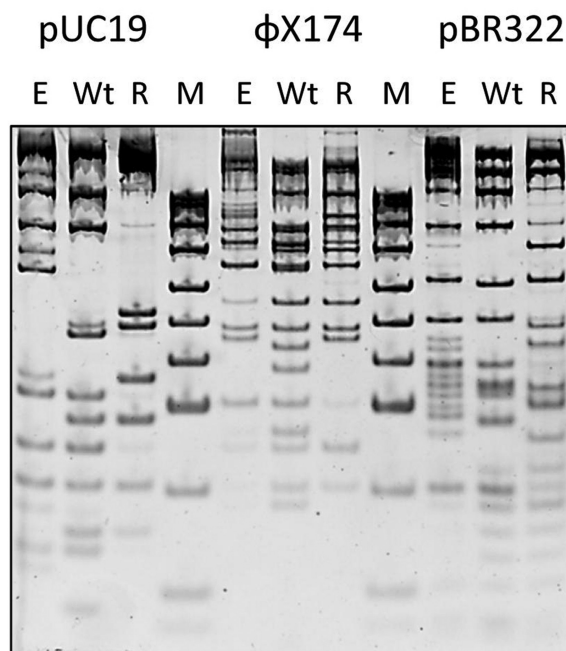
Sequence	310wt	S310E	S310R
GCAN ₅ TGC	5,95 (+/- 0,61 ^a) pM	29,7 (+/-13,1) nM	110,4 (+/- 46,8) nM
GCTN ₅ AGC	6,90 (+/- 0,93) pM	46,0 (+/-4,4) nM	26,2 (+/- 2,7) nM
GCGN ₅ CGC	7,34 (+/-1,66) pM	27,7 (+/-2,6) pM	104,0 (+/- 31,2) nM
GCCN ₅ GGC	2,34 (1,02) pM	17,8 (+/-1,7) nM	305,4 (+/- 56,6) pM

^astandard error of K_D fitting.

Analysis of the sequence preferences of R.MwoI variants

There is an apparent discrepancy between the clear sequence preference of R.MwoI variants observed in cleavage assays with homogenous, unique sequences, and cleavage patterns obtained using λ DNA. To further characterize the sequence preferences of the R.MwoI variants, three small DNA molecules pBR322, pUC19 and Φ X174 were used as substrates (Figure 10). The results obtained confirm that S310R and S310E variants of R.MwoI have preferences toward different subsets of MwoI sites. Some of fragments generated by the wt enzyme are missing, which indicates a substantial decrease in cleavage efficiency. The absence of a certain fragment generated by the wt enzyme can be caused by a failure of the engineered enzyme to cleave the substrate DNA on either or both sides of the fragment. Unequivocal inference of uncleaved sites from the analysis of the size of additional bands is however very difficult. The length of fragments cannot be determined with precision sufficient to clearly discriminate between alternative explanations of the fusion bands by different combinations of adjacent fragments. The inability of the S310E variant to cleave certain MwoI sites can be attributed, in the majority of cases, to the presence of A-T nucleotides at positions 3 and 9. In the case of S310R, the absence of cleavage correlates with MwoI site variants with G-T or G-C nucleotides at positions 3-9.

To obtain quantitative data on cleavage of a large number of MwoI sites with different sequences of the internal 7 bp linker between the GC half-sites, a different approach was used. Products of extensive λ DNA cleavage by R.MwoI 310wt and its S310R and S310E variants (Figure 11A) were cloned and sequenced. Sequences of more than 200 fragments were analysed to detect cleavage sites, as well as sites that were not cleaved and were present inside the cloned products. All sequenced clones had MwoI half-sites at the termini, which indicates that S310R and S310E variants do not cleave additional sites (i.e. that they do not exhibit star activity). Statistical analysis of the observed cleavage frequencies of various MwoI sites (normalized for the sequence frequencies in the λ DNA) indicated that the wt enzyme has no significant preference for or against any variant of the GCNNNNNNNGC site. In the case of S310R and S310E variants, different preferences were observed for

**Figure 10.** Cleavage of plasmid substrates by R.MwoI 310wt and its variants. M – molecular size standard (pUC19 cleaved with MspI). Cleavage reaction products were separated on non-denaturing 6% TBE-PAGE.

GCNNNNNNNGC sequence variants with different nucleotides at positions 3 and 9. The S310E variant exhibited also weak but detectable preference against the combination of A and T nucleotides at positions 4 and 6, respectively (data not shown). No sequence discrimination at other positions in the central fragment of the MwoI site and at positions directly flanking the MwoI site was detected.

Relative cleavage frequencies of 10 possible sequences present at positions 3 and 9 obtained after conversion of each pair of complementary asymmetric combinations into one (i.e. A-G = C-T) were calculated (Supplementary Table 1). After normalization for the site frequencies in the substrate DNA (Figure 11B), distinct sequence preferences for both S310R and S310E variants become apparent. R.MwoI S310R prefers sites with A-G, C-G, G-G and T-G nucleotides at positions 3-9 and cleaves poorly sequences with A-T, G-C and G-T at these positions. Compared with S310R, the S310E variant has a less pronounced preference toward its favorite substrate (C-G at positions 3-9). It cleaves well the MwoI sites with G-C nucleotides at position 3-9. MwoI sites with A-T, A-A and A-G nucleotides at position 3-9 were also cleaved, albeit poorly. Only a few uncleaved MwoI sites were detected inside the cloned fragments of the wt enzyme cleavage products (11 in 483). For both S310R and S310E variants, uncleaved sites were numerous, and their sequences correlated with these enzymes' preference against certain nucleotides at positions 3 and 9 (Figure 11C) inferred from the analysis of cleaved sites.

An analogous analysis was carried out for products of a limited cleavage of the λ DNA (Figure 11D). The S310R

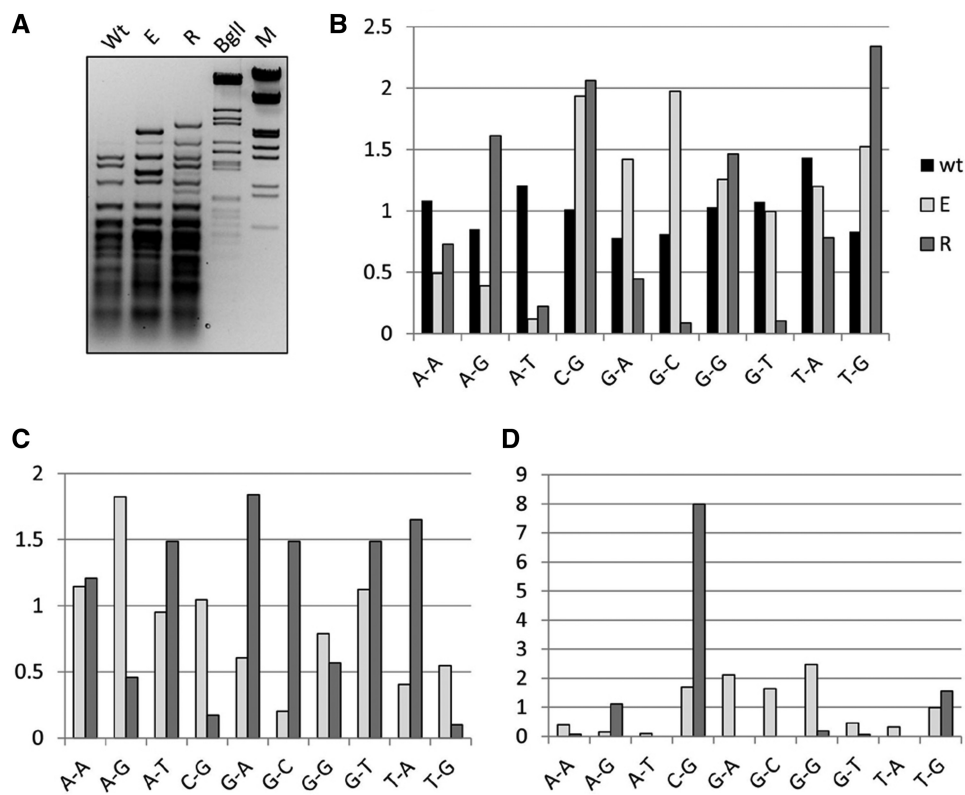


Figure 11. Cleavage preferences of R.MwoI variants assayed on phage λ DNA. (A) samples after extended cleavage (see Methods). BglI – λ DNA cleaved with BglI. M - molecular size standard (λ DNA cleaved with HindIII and EcoRI). (B) Normalized cleavage frequencies measured in products of extended (see Supplementary Table 1). Nucleotides at positions 3/9 are shown below the bars. After normalization, a frequency equal to 1.0 indicates frequency of cleavage proportional to that site's frequency in the substrate. Frequencies higher than 1.0, indicate preference and lower than 1.0, preference against the site. (C) Normalized frequency of sites occurrences inside products of extended cleavage. (D) Normalized cleavage frequencies measured in products of limited cleavage.

variant cleaved 8 times more GCCNNNNNGGC sequences than the wt R.MwoI enzyme that cuts all GCNNNNNNNGC sites without any significant preference. In a limited cleavage of λ DNA by the S310R variant, no cleavage of MwoI sites with sequences A–T, G–A, G–C and T–A at positions 3 and 9 were detected in 267 cleavage products analysed. The S310E variant was less specific than the S310R one. S310E cleaved MwoI sites with C–G, G–A, G–C and G–G nucleotides at positions 3–9 more frequently than expected from their relative occurrence among all MwoI sites in the λ DNA; T–G-containing MwoI sites were cleaved with a relative frequency proportional to their relative occurrence. All other MwoI site variants were cleaved less frequently than expected from their relative occurrence. Compared with the complete cleavage results, the limited cleavage revealed that the S310E variant cleaves G–T and T–A-containing sites rather poorly.

DISCUSSION

Type II REases are known for their extraordinary sequence specificity. The precise recognition of the target DNA sequence is achieved by an extensive network of protein–DNA interactions and, frequently, by the

redundancy of contacts with a particular base. A majority of the reported changes in the sequence specificity of REases led to obtaining a variant, which recognized a subset of the original degenerate targets (7,8) or an extension of the original recognition sequence by introduction of degeneracy in one of the positions (9,10). Our results belong to the first category since new variants of R.MwoI cleave preferentially only a subset of the original recognition sequence. The engineering of enzymes from the R.MmeI family was a distinctive case where deciphering of the “sequence recognition code” led to the construction of several REases with new specificities. Here, we reported the results of structure-guided protein engineering of a thermophilic REase R.MwoI, which has led to the development of a two single position variants with sequence preferences clearly different from the original enzyme.

In the absence of a high-resolution crystal structure for R.MwoI, we used bioinformatics methods to construct a theoretical model of the protein–DNA complex. The model was supported by site-directed mutagenesis of residues predicted to be functionally important, and compared with the crystal structures of two related enzymes with more stringent, but overlapping sequence specificities: R.BglI (GCCNNNNNGGC) and R.SfiI

(GGCCNNNNNCCGG). This comparison suggested that a substitution of the residue S310 of R.MwoI (which is not required for the cleavage of its natural substrate) may lead to the acquisition of selectivity for the positions 3–9 of the MwoI cognate sequence. The model-based prediction was confirmed by the experimental analysis of S310 substitutions in R.MwoI. We obtained four enzyme variants with sequence selectivity different from the wt protein. The sequence preference depends on the chemical character of the amino acid side chain introduced instead of the original Ser residue. Substitution with a positively charged residue Arg (which is present in an analogous position in R.BglI and R.SfiI) or Lys (which is present in some of the uncharacterized homologs, see Supplementary Figure 1) leads to the selectivity for the R.BglI recognition sequence (GCCNNNNNGGC). On the other hand, introduction of a negatively charged residue (not found in naturally occurring homologs) causes a preference toward the GCGNNNNNGCG sequence; this effect is more pronounced in the S310E variant than in the S310D variant. Other substitutions of S310 did not cause a change in the cleavage pattern (data not shown). We predict that the specificity is achieved owing to the specific recognition of Gua by Arg (or Lys) with the direct contact of the positively charged amino acid group with the O6 atom of the DNA base versus the recognition of Cys by Glu or Asp, and either direct (Glu) or water-mediated (Asp) contact of the negatively charged protein side chain with the N4 amino group of the DNA base. It will be interesting to try analogous substitutions in other enzymes of the R.MwoI/R.BglI/R.SfiI family, as they may lead to similar changes in the specificity.

Biochemical characterization of the R.MwoI variants with the altered sequence specificity demonstrated that their selectivity increased substantially. When assayed on set of short substrates with unique MwoI sites, the wt enzyme displays a 5-fold difference in the cleavage rate of the substrates that vary in the positions 3–9. In case of the three most selective variants, we could detect cleavage only for one preferred sequence and at least 20-fold lower cleavage of another sequence (T–A). Cleavage of two other substrates was undetectable, which indicates that these variants exhibit at least 1000-fold increase of the substrate cleavage selectivity.

Patterns of λ DNA cleavage generated by the S310R and S310E variants of R.MwoI indicate, however, that they have more complex sequence preferences than those characterized on the limited panel of short substrates used in the kinetic analysis. Cloning and sequencing of fragments generated by the cleavage of λ DNA provided us with data amenable to statistical analysis of the variants' preferences. The S310R variant cleaves preferentially sequences with A–G, C–G, G–G and T–G at positions 3 and 9 of the GCNNNNNNNGC site (where the discriminated sites are underlined). Therefore, its sequence preference can be summarized as GCNNNNNNNGGC. In partial cleavage conditions, the S310R variant shows clear preference toward the GCNCNNNNNGGC sequence, similarly to the preference revealed in kinetic assays on substrates with a single recognition site.

Cleavage preferences of the S310E variant, according to the analysis of the cloned cleavage products, are broader than the preferences of the S310R variant. The S310E variant cleaves preferentially sites with sequences C–G and G/C at positions 3 and 9 of the of the GCXNNNNNXGC site, but sequences G–A, G–G, T–A and T–G are cleaved only slightly less efficiently. More apparent is the discrimination of the S310E variant against sites with sequences A–A, A–G and A–T at positions 3 and 9.

Further protein engineering efforts are necessary to obtain R.MwoI variants with fully developed specificities toward a single six nucleotide interrupted palindrome. One possibility to explore is the introduction of additional contacts to the base complementary to that contacted by residues at position 310 in R.MwoI. Analysis of R.MwoI model and sequence alignment involving R.BglI and R.SfiI that do form such contacts, reveals however a deletion in a corresponding loop in R.MwoI (see Figure 1, around position 308 in R.MwoI), which hampers the introduction of a suitable contact by a simple amino acid substitution. Thus, optimization of the S310E and S310R specificities will most likely require extensive experiments involving alternative insertions with different sequences and testing for their specificity.

The measurements of the binding affinities of S310R, S310K and S310E R.MwoI variants suggest that discrimination of nucleotide residues in positions 3 and 9 is caused by differences in the substrate binding. R.MwoI variants bind their preferred substrates slightly weaker than the wt enzyme, but they display at least a 25-fold discrimination against other sequences. Equilibrium binding constants show a 100 to 1000-fold difference in the K_D between the preferred sequence and other palindromic extensions of the MwoI cognate sequence. K_D values demonstrate that the variants' affinities for the preferred sequences did not increase in comparison with the wild-type enzyme. The main factor responsible for the sequence preference of the obtained proteins is a significant decrease in the affinities for the substrates with sequences other than the preferred one. We suspect that in the engineered variants new interactions with the 3–9 positions are formed, but they may also weaken other elements of the substrate binding network, compared with the wt enzyme. The answer to such detailed questions would require the availability of a high-resolution crystal structure. Unfortunately, thus far our attempts to crystalize R.MwoI-DNA complexes were unsuccessful.

Our approach was to some extent similar to the one used in one of the first attempts to engineer the R.EcoRV REase, in which the 6 bp EcoRV recognition sequence was extended by the nucleotides in the flanking positions (5,6). In principle, such changes do not require any alteration of the existing interaction network in the native enzyme, but rather involve introduction of new ones, which discriminate nucleobases in additional positions in the immediate vicinity of the wt target sequence. In the case of R.MwoI, which recognizes an interrupted sequence, it was possible to extend the recognition toward the internal positions that are not recognized by the wt enzyme. In the DNA co-crystal

structure of R.NotI (a REase that recognizes an 8-bp sequence), the central six base pairs form 17 interactions with the REase that saturate the interaction potential of this part of the substrate (49). However, distal base pairs of the NotI site are recognized only by two amino acid residues. This observation led to hypothesis that extension (or shortening) of the target sequence may be one of the mechanisms of divergence in REases. Therefore the approach we have used bears some resemblance to the natural evolution.

R.MwoI is a thermophilic REase with an optimum reaction temperature 60°C, 10% activity at 37°C and undetectable activity at the room temperature. Here, we reported an efficient expression of R.MwoI in *E. coli* without a cognate methyltransferase, when grown at the room temperature. Such system brings some new possibilities of REase specificity engineering that are not available in the case of mesophilic REases. The most important application of this system is the potential to express native or engineered thermophilic REases with any sequence specificity because the methyltransferase is not required to protect the host DNA from cleavage. In the case of thermophilic enzymes with detectable activity at temperature 37°C or 42°C, it is possible to create permissive and non-permissive conditions in the same *E. coli* host. In such situation, a screen for desirable mutants of REase would be carried out by combining the replica plating technique with two different growth temperatures. Two obvious exemplification of such strategy would be (i) screening for inactive variants (colonies will grow in a host without the methyltransferase regardless of the incubation temperature); (ii) screening for the variants with sequence specificity not recognized by the cognate methyltransferase (even in the presence of the cognate methyltransferase, colonies will not grow at permissive temperature). In any case, the possibility of expressing a thermophilic REase without a cognate MTase in the host cells, relieves the experimentalist from the necessity of engineering a cognate MTase with the new specificity in the first place. Although R.MwoI variants obtained in this study do not cut sequences not recognized by the wt enzyme, the experimental protocol described herein allows for engineering REase variants that could cut other sites as well.

The results of our study prove the utility of protein modeling as a tool to guide protein engineering. We have successfully demonstrated that a single residue replacement can lead to the development of new, different sequence preferences. This finding explains the extraordinary potential of REases to develop new sequence specificities in the nature. Together with the new procedure for engineering of thermophilic REases, our study provides a toolkit for rational engineering of sequence specificities in Type II REases.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1 and Supplementary Figures 1 and 2.

ACKNOWLEDGEMENTS

We are grateful to Matthias Bochtler, Agata Kamaszewska, Irina Tuszyńska, and Jakub Jopek for critical reading of the manuscript and useful suggestions.

FUNDING

Polish Ministry of Science and Higher Education (MNiSW) [N301 100 31/3043]; K.S. and J.M.B. were additionally supported by the European Research Council [StG grant RNA+P = 123D, grant agreement no: 261351]; J.M.B. was supported by the “Ideas for Poland” fellowship from the Foundation for Polish Science. B.K. was supported by the MNiSW [POIG.02.03.00-00-003/09]. Funding for open access charge: ERC (RNA+P = 123D).

Conflict of interest statement. None declared.

REFERENCES

1. Roberts,R.J. (2005) How restriction enzymes became the workhorses of molecular biology. *Proc Natl Acad Sci USA*, **102**, 5905–5908.
2. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2010) REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res*, **38**, D234–236.
3. Lanio,T., Jeltsch,A. and Pingoud,A. (2000) On the possibilities and limitations of rational protein design to expand the specificity of restriction enzymes: a case study employing EcoRV as the target. *Protein Eng*, **13**, 275–281.
4. Jeltsch,A., Wenz,C., Wende,W., Selent,U. and Pingoud,A. (1996) Engineering novel restriction endonucleases: principles and applications. *Trends Biotechnol*, **14**, 235–238.
5. Schöttler,S., Wenz,C., Lanio,T., Jeltsch,A. and Pingoud,A. (1998) Protein engineering of the restriction endonuclease EcoRV—structure-guided design of enzyme variants that recognize the base pairs flanking the recognition site. *Eur J Biochem*, **258**, 184–191.
6. Wenz,C., Hahn,M. and Pingoud,A. (1998) Engineering of variants of the restriction endonuclease EcoRV that depend in their cleavage activity on the flexibility of sequences flanking the recognition site. *Biochemistry*, **37**, 2234–2242.
7. Zhu,Z., Zhou,J., Friedman,A.M. and Xu,S.Y. (2003) Isolation of BsoBI restriction endonuclease variants with altered substrate specificity. *J Mol. Biol*, **330**, 359–372.
8. Samuelson,J.C. and Xu,S.Y. (2002) Directed evolution of restriction endonuclease BstYI to achieve increased substrate specificity. *J Mol. Biol*, **319**, 673–683.
9. Samuelson,J.C., Morgan,R.D., Benner,J.S., Claus,T.E., Packard,S.L. and Xu,S.Y. (2006) Engineering a rare-cutting restriction enzyme: genetic screening and selection of NotI variants. *Nucleic Acids Res*, **34**, 796–805.
10. Rimseliene,R., Maneliene,Z., Lubys,A. and Janulaitis,A. (2003) Engineering of restriction endonucleases: using methylation activity of the bifunctional endonuclease Eco57I to select the mutant with a novel sequence specificity. *J Mol. Biol*, **327**, 383–391.
11. Morgan,R.D., Dwinell,E.A., Bhatia,T.K., Lang,E.M. and Luyten,Y.A. (2009) The MmeI family: type II restriction-modification enzymes that employ single-strand modification for host protection. *Nucleic Acids Res*, **37**, 5208–5221.
12. Morgan,R.D. and Luyten,Y.A. (2009) Rational engineering of type II restriction endonuclease DNA binding and cleavage specificity. *Nucleic Acids Res*, **37**, 5222–5233.
13. Skowronek,K.J. and Bujnicki,J.M. (2007) In: Polaina,J. and Maccabe,A. (eds), *Industrial enzymes*. Springer, Dordrecht, The Netherlands, pp. 357–378.

14. Pingoud, A., Fuxreiter, M., Pingoud, V. and Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci*, **62**, 685–707.
15. Bujnicki, J.M. (2001) Understanding the evolution of restriction-modification systems: clues from sequence and structure comparisons. *Acta Biochim Pol*, **48**, 935–967.
16. Bujnicki, J.M. (2004) In: Pingoud, A. (ed.), *Restriction Endonucleases*. Springer, Berlin, pp. 63–87.
17. Pingoud, A., Fuxreiter, M., Pingoud, V. and Wende, W. (2005) Type II restriction endonucleases: structure and mechanism. *Cell Mol Life Sci*, **62**, 685–707.
18. Siksnys, V., Timinskas, A., Klimasauskas, S., Butkus, V. and Janulaitis, A. (1995) Sequence similarity among type-II restriction endonucleases, related by their recognized 6-bp target and tetranucleotide-overhang cleavage. *Gene*, **157**, 311–314.
19. Pingoud, V., Sudina, A., Geyer, H., Bujnicki, J.M., Lurz, R., Luder, 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.
20. 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.
21. Chmiel, A.A., Bujnicki, J.M. and Skowronek, K.J. (2005) A homology model of restriction endonuclease SfiI in complex with DNA. *BMC Struct Biol*, **5**, 2.
22. Armalyte, E., Bujnicki, J.M., Giedriene, J., Gasiunas, G., Kosinski, J. and Lubys, A. (2005) MvaI269I: a monomeric type IIS restriction endonuclease from *Micrococcus varians* with two EcoRI- and FokI-like catalytic domains. *J Biol Chem*, **280**, 41584–41594.
23. Kaminska, K.H., Kawai, M., Boniecki, M., Kobayashi, I. and Bujnicki, J.M. (2008) Type II restriction endonuclease R.Hpy188I belongs to the GIY-YIG nuclease superfamily, but exhibits an unusual active site. *BMC Struct Biol*, **8**, 48.
24. Ibryashkina, E.M., Zakharova, M.V., Baskunov, V.B. et al. (2007) Type II restriction endonuclease R.Eco29kI is a member of the GIY-YIG nuclease superfamily. *BMC Struct Biol*, **7**, 48.
25. Lunnen, K.D., Morgan, R.D., Timan, C.J., Krzycki, J.A., Reeve, J.N. and Wilson, G.G. (1989) Characterization and cloning of MwoI (GCN7GC), a new type-II restriction-modification system from *Methanobacterium wolfei*. *Gene*, **77**, 11–19.
26. Orłowski, 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.
27. Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389–3402.
28. Roberts, R.J., Vincze, T., Posfai, J. and Macelis, D. (2005) REBASE—restriction enzymes and DNA methyltransferases. *Nucleic Acids Res*, **33** (Database Issue), D230–D232.
29. Wheeler, D.L., Barrett, T., Benson, D.A. et al. (2005) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*, **33**, D39–D45.
30. Pei, J., Tang, M. and Grishin, N.V. (2008) PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res*, **36**, W30–W34.
31. Kurowski, M.A. and Bujnicki, J.M. (2003) GeneSilico protein structure prediction meta-server. *Nucleic Acids Res*, **31**, 3305–3307.
32. Kosinski, J., Cymerman, I.A., Feder, M., Kurowski, M.A., Sasin, J.M. and Bujnicki, J.M. (2003) A “Frankenstein’s monster” approach to comparative modeling: merging the finest fragments of Fold-Recognition models and iterative model refinement aided by 3D structure evaluation. *Proteins*, **53**(Suppl. 6), 369–379.
33. Kosinski, J., Gajda, M.J., Cymerman, I.A. et al. (2005) FRankenstein becomes a cyborg: the automatic recombination and realignment of fold recognition models in CASP6. *Proteins*, **61**(Suppl. 7), 106–113.
34. Vanamee, E.S., Viadiu, H., Kucera, R., Dorner, L., Picone, S., Schildkraut, I. and Aggarwal, A.K. (2005) A view of consecutive binding events from structures of tetrameric endonuclease SfiI bound to DNA. *EMBO J*, **24**, 4198–4208.
35. Pawlowski, M., Gajda, M.J., Matlak, R. and Bujnicki, J.M. (2008) MetaMQAP: a meta-server for the quality assessment of protein models. *BMC Bioinformatics*, **9**, 403.
36. Wallner, B. and Elofsson, A. (2003) Can correct protein models be identified? *Protein Sci*, **12**, 1073–1086.
37. Newman, M., Lunnen, K., Wilson, G., Greci, J., Schildkraut, I. and Phillips, S.E. (1998) Crystal structure of restriction endonuclease BglII bound to its interrupted DNA recognition sequence. *EMBO J*, **17**, 5466–5476.
38. Boniecki, M., Rotkiewicz, P., Skolnick, J. and Kolinski, A. (2003) Protein fragment reconstruction using various modeling techniques. *J Comput Aided Mol Des*, **17**, 725–738.
39. DeLano. (2002). 0.99rc6 ed. Schrödinger, LLC.
40. Guex, N. and Peitsch, M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.
41. Winter, J., Lerp, C., Zabel, H.P., Wildenauer, F.X., König, H. and Schindler, F. (1984) *Methanobacterium wolfei*, sp.nov., a new tungsten-requiring, thermophilic, autotrophic methanogen. *Syst Appl Microbiol*, **5**, 457–466.
42. Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr Purif*, **41**, 207–234.
43. Conlan, L.H. and Dupureur, C.M. (2002) Multiple metal ions drive DNA association by PvuII endonuclease. *Biochemistry*, **41**, 14848–14855.
44. Bujnicki, J.M. and Rychlewski, L. (2001) Grouping together highly diverged PD-(D/E)XK nucleases and identification of novel superfamily members using structure-guided alignment of sequence profiles. *J Mol Microbiol Biotechnol*, **3**, 69–72.
45. 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.
46. Skowronek, K.J., Kosinski, J. and Bujnicki, J.M. (2006) Theoretical model of restriction endonuclease HpaI in complex with DNA, predicted by fold recognition and validated by site-directed mutagenesis. *Proteins*, **63**, 1059–1068.
47. Huai, Q., Colandene, J.D., Chen, Y., Luo, F., Zhao, Y., Topal, M.D. and Ke, H. (2000) Crystal structure of NaeI—an evolutionary bridge between DNA endonuclease and topoisomerase. *EMBO J*, **19**, 3110–3118.
48. Bujnicki, J.M., Radlinska, M. and Rychlewski, L. (2001) Polyphyletic evolution of type II restriction enzymes revisited: two independent sources of second-hand folds revealed. *Trends Biochem Sci*, **26**, 9–11.
49. Lambert, A.R., Sussman, D., Shen, B., Manus, R., Nix, J., Samuelson, J., Xu, S.Y. and Stoddard, B.L. (2008) Structures of the rare-cutting restriction endonuclease NotI reveal a unique metal binding fold involved in DNA binding. *Structure*, **16**, 558–569.