Screening for catalytically active Type II restriction endonucleases using segregation-induced methylation deficiency

Mindaugas Ukanis[1,2], Rimantas Sapranauskas[2] and Arvydas Lubys[2,*]

[1]Institute of Biotechnology, Vilnius University and [2]Fermentas UAB (part of Thermo Fisher Scientific), V.A. Graiciuno 8, Vilnius LT-02241, Lithuania

ABSTRACT

Type II restriction endonucleases (REases) are one of the basic tools of recombinant DNA technology. They also serve as models for elucidation of mechanisms for both site-specific DNA recognition and cleavage by proteins. However, isolation of catalytically active mutants from their libraries is challenging due to the toxicity of REases in the absence of protecting methylation, and techniques explored so far had limited success. Here, we present an improved SOS induction-based approach for _in vivo_ screening of active REases, which we used to isolate a set of active variants of the catalytic mutant, Cfr10I^{E204Q}. Detailed characterization of plasmids from 64 colonies screened from the library of ∼200 000 transformants revealed 29 variants of _cfr10IR_ gene at the level of nucleotide sequence and 15 variants at the level of amino acid sequence, all of which were able to induce SOS response. Specific activity measurements of affinity-purified mutants revealed >200-fold variance among them, ranging from 100% (wild-type isolates) to 0.5% (S188C mutant), suggesting that the technique is equally suited for screening of mutants possessing high or low activity and confirming that it may be applied for identification of residues playing a role in catalysis.

INTRODUCTION

Restriction-modification (R-M) systems are widespread among prokaryotic organisms (1,2). They are composed of two enzymatic activities. One of them, DNA methylation activity, ensures modification of A or C base within the specific DNA sequence. This site-specific modification protects host DNA from the action of the other component, namely, the endonucleolytic activity of the same

specificity (3). The biological function of R-M systems is generally thought to be the protection of the host genome against foreign DNA, in particular bacteriophage DNA. However, two other hypotheses for the biological function of R-M systems have been proposed. One hypothesis proposes that R-M enzymes are modulators of genetic variation frequency (4), while another considers R-M genes to be selfish mobile genetic elements, like viruses or transposons that invade genomes without necessarily providing selective advantages (5,6).

Based on enzyme composition, requirement for cofactors and mode of action, restriction enzymes are classified into one of four types (7), of which Type II restriction endonucleases (REases) are the most abundant and the best studied. In general, Type II enzymes recognize specific DNA sequences of 2–8 bp and cleave DNA at fixed positions within or in close vicinity to the target (8). These unique features make them indispensible in DNA manipulations _in vitro_. On the other hand, the relatively simple structure of Type II REases offers an attractive model for studying of protein–DNA interactions and for investigation of DNA cleavage catalysis. Two _in vivo_ selection/screening systems based either on transcriptional interference or on visualization of induced SOS response were used in order to identify catalytic amino acid residues of Type IIP REases (9–16). Transcriptional interference was first adapted to select for catalytically deficient but binding-proficient variants of BamHI (11) and then to identify substitutions of BamHI amino acid residues directly involved in target recognition that maintain site-specific recognition of the cognate target 5'-GGA TCC (12). Successful selection of BamHI mutants of wild-type (wt) specificity from the collection of 8000 variants in just one selection step led to the idea that this same technique could be used for isolation of mutants of altered specificity (12). In contrast, SOS response-based screening of REases relies on their catalytic activity (9). The SOS response is visualized by using a genetic fusion between DNA damage-inducible promoter and reporter gene _lacZ_ which codes for β-galactosidase.

*To whom correspondence should be addressed. Tel: +370 5 2602105; Fax: +370 5 2602142; Email: arvydas.lubys@thermofisher.com

If REase introduces either single-, or double-stranded DNA breaks that provoke SOS response, this is manifested by the blue color of colonies on X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galacto-pyranoside) plates (9,17). However, the host cells can tolerate only a limited number of DNA breaks. For instance, in the absence of protective methylation, the level of detection for BamHI activity by using induced SOS response was very low, 0.01% of the wt activity (10), while 4% activity was lethal (12). Toxicity of REases under constant expression is a major drawback of SOS response-based screening and results in isolation of mutants of low specific activity (9,10,13,14).

Here, we describe a new approach for *in vivo* screening for active mutants of REases, which uses segregation-induced methylation deficiency (SIMD) and overcomes limitations described above. SIMD is based on previous observations that some Type II R-M systems increase stability of plasmids that carry them by inducing death of plasmid-free segregants (6,18). This so-called post-segregational killing is determined by gradually lowering intracellular concentration of protecting methyltransferase (MTase) through dilution during cell growth and division. At certain point the concentration of MTase becomes too low to protect all genomic targets of the newly replicated chromosome. MTase deficiency results in the cleavage of chromosome at unmodified sites, induction of SOS response and cell death if there are too many nicks or double-stranded DNA breaks (19–22). As a proof of principle, we used this approach to isolate catalytically active variants of REase Cfr10I which recognizes the sequence 5′-Pu↓CCGGPy and cleaves it as indicated by the arrow. Based on the crystal structure of Cfr10I (23) and mutational analysis of amino acid residues that are implicated in catalysis (24), the active site of Cfr10I is formed from the acidic residues D134, E71 and E204 as well as the basic residue K190. Remarkably, S188 from the 188SVK sequence motif is a spatial equivalent of the acidic residue from the (E/D)XK-part of the active site motif conserved in many Type II REases. However, mutational analysis revealed that the S188 residue in Cfr10I is not so important for catalysis, while E204 located 2.8 Å away serves the role of acidic residue from the (E/D)XK motif rather than S188. It was shown that S188E replacement compensates for a sharp drop in activity observed in the case of E204S mutant, restoring it from <0.002% to 10% of the wt level in double mutant S188E/E204S (24). In this work, codons of all five amino acid residues were semi-randomly mutated in the Cfr10I^{E204Q} mutant which possesses 0.004% of wt activity (24), and blue colonies screened using a SIMD approach were further characterized for both spectrum of *cfr10IR* gene mutations and activity of mutant Cfr10I enzymes.

MATERIALS AND METHODS

Bacterial strains and reagents

Escherichia coli strain ER2267 (e14⁻ (*mcrA*⁻) *endA1* *supE44* *thi-1* Δ(*mcrC-mrr*)114::IS*10* Δ(*argF-lac*)U169 *recA1*/F' *proA*⁺*B*⁺ *lacI*^{q} Δ(*lacZ*)M15 *zzf*::mini-Tn*10*

(Km^r)) was used as a host for cloning experiments. The *dinD*::*lacZ* SOS indicator strain ER1992 (25) (F⁻ λ⁻ Δ(*argF-lac*)U169 *supE44* e14⁻ (*mcrA*⁻) *dinD1*::Mu dI1734 (Km^r, LacZ$^+$) *endA1* *thi-1* Δ(*mcrC-mrr*) 114::IS*10*) was used as a host for SIMD experiments. ER2566 (F⁻ λ⁻ *fhuA2* [*lon*] *ompT* *lacZ*::T7 gene1 *gal* *sulA11* Δ(*mcrC-mrr*)114::IS*10* R(*mcr-73*::miniTn*10*– Tet^s)2 R(*zgb-210*::Tn*10*)(Tet^s) *endA1* [*dcm*]) was used for overproduction of His-tagged Cfr10I variants. All three strains were obtained from New England Biolabs (Ipswich, MA). *E. coli* cells were grown in Luria–Bertani (LB) broth or on LB agar at appropriate temperatures and supplemented, where needed, with X-Gal (0.032 mg/ml) and antibiotics ampicillin (Ap; 0.1 mg/ml) and/or chloramphenicol (Cm; 0.03 mg/ml). Cells were transformed using standard CaCl$_2$-heat-shock method or electroporation (26).

All enzymes, DNA molecular mass markers, kits, disposable protein purification columns, nickel-nitriloacetic acid (Ni-NTA) resin and other reagents (if not mentioned elsewhere) were obtained from Thermo Scientific and used following recommendations of manufacturers. Oligonucleotides were obtained from Metabion (Martinsried, Germany).

DNA manipulations

All DNA manipulations followed standard procedures (26) and recommendations of manufacturers. Sequencing was performed using BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA), sequencing data were collected on Genetic Analyzer 3130*xl* (Applied Biosystems).

Construction of plasmids

pTSA29 is a low copy number plasmid carrying temperature-sensitive (ts) replication initiator of pSC101 and encoding resistance to Ap (27). It is compatible with both pMB1 and p15A replicons. pTSA29-Cfr10IM was constructed by cleaving DNA of pCfr10IRM-1 (Thermo Scientific collection) with Mph1103I and PvuII, filling of cohesive ends using T4 DNA polymerase, gel purifying the resulting fragment of 1.56 kb encompassing the gene for Cfr10I methyltransferase, ligating the fragment into Eco32I-cleaved and FastAP-treated vector pTSA29 and screening for recombinant plasmid carrying cloned *cfr10IM* gene under the control of vector-encoded P$_{lac}$ promoter. pLATE31 is a pMB1-derived T7 expression vector from aLICator^{TM} LIC Cloning and Expression System (Thermo Scientific). It was used for placing of poly-His affinity tag at the C terminus of both Cfr10I^{wt} and Cfr10I^{E204Q}. DNAs of plasmids coding for Cfr10I^{wt} and for Cfr10I^{E204Q} were kindly provided by M. Zaremba and served as templates for polymerase chain reaction (PCR) using specially designed primers LIC Dir and LIC Rev (Table 1), which have 5′-terminal extensions (underlined) for ligation-independent cloning. Details on the construction of pLATE-Cfr10Iwt and pLATE-Cfr10IE204Q are available upon request. In order to ensure constitutive expression of His-tagged Cfr10I^{wt} and Cfr10I^{E204Q} in *E. coli* strains lacking T7 RNA

Table 1. Primers used for construction of plasmids and for mutagenesis

Primer	Sequence (5′→3′)
LIC Dir	<u>AGAAGGAGATATAACT</u>ATGGATATAATTTCAAAGTCTGG
LIC Rev	<u>GTGGTGGTGATGGTGATGGCC</u>ACTAGATAAAATGTGACTCAGGC
E71 Dir	GGGTTGAAGGCTCTCAAGGGC
E71 Rev*	CCCTTATTCCAATTAACCA(G50%C50%)(G15%A15%T55%C15%)(G15%A15%T15%C55%) ATACCAATCACCATTACAC
D134 Dir*	CTGATAACTTCGAATCCG(G55%A15%T15%C15%)(G15%A55%T15%C15%)(G50%C50%)TTCTCCATAATA GATATTAGA
S188/K190 Rev*	ATCAGGTCTGAAAGTCGT(G50%C50%)(G15%A15%T55%C15%)(G15%A15%T55%C15%)TAC (G50%C50%)(G55%A15%T15%C15%)(G15%A55%T15%C15%)TAAAAAAGACTTTATATGTTCG
E204 Dir*	AAGATTACAACTTGCTCAT(G55%A15%T15%C15%) (G15%A55%T15%C15%)(G50%C50%)GGGAGTCTTATGAAGGC
E204 Rev	GCGCTCATCGTCATCCTCGGC

Degenerate primers are marked by asterisk.
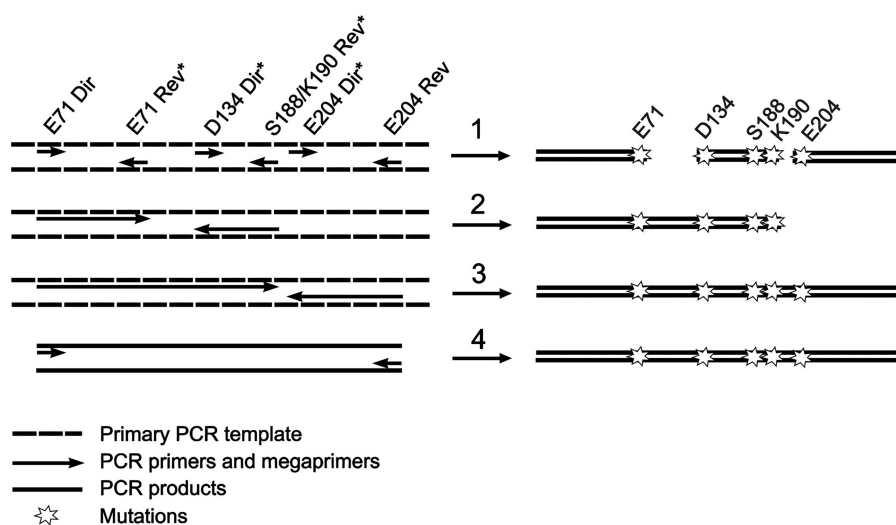
Figure 1. Schematic representation of four PCR steps used to introduce mutations into five *cfr10IR*^{E204Q} codons. Left panel shows primers and DNA templates used in the individual steps of mutagenesis, right panel displays amplified DNA fragments and relative positions of codons targeted for mutagenesis.

polymerase, corresponding genes together with the upstream T7 promoter were transferred from pLATE-Cfr10Iwt and pLATE-Cfr10IE204Q to the Eco32I-SphI cleaved p15A derivative pACYC184 (28) under the control of constitutive promoter P$_{tet}$. Both donor plasmids were cleaved with Acc65I, treated with T4 DNA polymerase and then, after heat inactivation of polymerase, were cleaved with SphI. Recombinant DNA molecules resulting after ligation of gel-purified fragments of 1 kb with DNA of pACYC184 vector were used to transform ER2267 (pTSA29-Cfr10IM). Restriction analysis and sequencing were used to confirm the structure of resulting plasmids pAC-Cfr10Iwt and pAC-Cfr10IE204Q both conferring resistance to Cm.

Mutagenesis and library construction

A library of Cfr10I^{E204Q} variants mutated at amino acid residues E71, D134, S188, K190 and Q204 was generated using the multistep 'megaprimer' PCR technique (29) which is outlined in Figure 1, primers listed in Table 1

and DNA of pAC-Cfr10IE204Q as a template. Three individual DNA fragments carrying semi-random mutations at all three codon positions coding for five aforementioned amino acid residues were amplified during the first amplification round using Pfu polymerase and following PCR conditions: initial denaturation at 95°C for 3 min and 35 cycles of amplification (95°C, 30 s; 60°C, 30 s and 72°C, 60 s) followed by 10 min extension at 72°C. DNA fragments of 454, 207 and 442 bp were gel-purified and two of them, 454 bp and 207 bp in length, were used in the second PCR round using High Fidelity enzyme mix. PCR conditions were as follows: initial denaturation 94°C, 3 min; 35 cycles of amplification (94°C, 60 s; 55°C, 60 s and 72°C, 60 s) and final extension at 72°C for 10 min. Gel-purified product of 810 bp was used in the third PCR round together with the DNA fragment of 442 bp from the first PCR step. The full-length product of 1254 bp was amplified using High Fidelity enzyme mix under the following PCR conditions: initial denaturation 94°C, 5 min; 35 cycles of amplification (94°C, 90 s; 50°C, 90 s

and 72°C, 90 s) and final extension at 72°C for 10 min. In order to provide more material for library construction, the resulting gel-purified DNA fragment of 1254 bp was re-amplified using outward primers E71 Dir and E204 Rev and High Fidelity enzyme mix under the following PCR conditions: initial denaturation 94°C, 3 min; 35 cycles of amplification (94°C, 60 s; 60°C, 60 s and 72°C, 90 s) and final extension at 72°C for 10 min. The gel-purified PCR product was cleaved with PstI and SphI, and the resulting fragment of 1025 bp, which encompasses the mutagenized *cfr10IR* gene was isolated from the agarose gel and ligated to the vector backbone of 3893 bp obtained by cleavage of pAC-Cfr10IE204Q DNA with PstI and SphI. Ligation mixture was column-purified and used to electrotransform ER2566 (pTSA29-Cfr10IM) cells. A small fraction of transformed cells was plated onto LB agar plate supplemented with Ap and Cm and grown at 30°C in order to evaluate the total number of transformed cells. The remaining part of transformed cells was grown at 30°C overnight in liquid LB medium in the presence of Ap and Cm, and then the total plasmid DNA representing the library of 10^6 mutants was column-purified and used to transform ER1992 (pTSA29-Cfr10IM) cells by $CaCl_2$-heat-shock method. Heat shock and subsequent recovery of transformed cells before plating was performed at 30°C temperature, while growth of colonies on agar plates supplemented with Ap, Cm and X-Gal was performed by incubating plates first at 30°C for 10 h and then at 42°C for additional 12 h.

Protein synthesis and purification

ER2566 (pTSA29-Cfr10IM) cells carrying plasmids which encode His-tagged Cfr10I^{wt} and its mutant variants were grown in 200 ml of LB medium at 30°C to ∼0.6 OD (λ = 600 nm). Expression of REase gene was induced by adding isopropyl-β-D-thiogalactopyranoside to 1 mM concentration, and then growth was continued for 4 h at the same temperature. Cells were harvested by centrifugation at 4°C and 15 000g for 15 min. Centrifuged cells were re-suspended in 5 ml of lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 5 mM 2-mercapthoethanol and 1× ProteoBlock Protease Inhibitor Coctail; pH 8.0), disrupted by sonication and centrifuged at 4°C and 20 000g for 30 min to remove cell debris. Supernatants were applied onto 5 ml volume disposable plastic columns packed with 2 ml of HisPur^{TM} Ni-NTA resin and pre-equilibrated with 5 ml of lysis buffer. After gravity flow-through of supernatants, 10 ml of wash buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, 5 mM 2-mercapthoethanol and 0.1% Triton X-100; pH 8.0) was applied to the columns. His-tagged proteins were then eluted with 1 ml of elution buffer (50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, 5 mM 2-mercapthoethanol and 0.1% Triton X-100; pH 8.0). This procedure was repeated twice, and the second fraction was dialyzed overnight against the storage buffer containing 10 mM Tris, pH 7.4, 100 mM KCl, 1 mM ethylenediaminetetraacetic acid (EDTA), 50% glycerol and 1 mM DTT. Concentration and physical purity of purified wt and mutant Cfr10I REase

was determined using Agilent 2100 Bioanalyzer and Protein 230 Kit (Agilent Technologies) according to the manufacturer's recommendations.

Restriction endonuclease activity assay

The specific activity of wt and mutant Cfr10I REases was determined by incubating different amounts of partially purified proteins with 1 μg phage λ DNA in 50 μl reaction buffer, containing 10 mM Tris–HCl (pH 8.0), 5 mM $MgCl_2$, 100 mM NaCl, 0.02% Triton X-100 and 1 mg/ml bovine serum albumin (BSA). Enzyme preparations were initially pre-diluted with dilution buffer (20 mM potassium phosphate (pH 7.4), 200 mM KCl, 1 mM EDTA, 7 mM 2-mercaptoethanol, 10% glycerol and 0.2 mg/ml BSA), if needed. DNA cleavage reactions were carried out for 1 h at 37°C and then stopped by adding 25 μl of stop solution (60 mM EDTA, pH 8.2, 30% glycerol, 0.005% bromophenol blue, 0.12% sodium dodecyl sulphate). Products were analyzed by agarose gel electrophoresis, and completeness of λ DNA cleavage was evaluated visually. One unit of enzyme activity is considered as the amount of protein required to completely digest 1 μg of λ DNA in 1 h at 37°C.

RESULTS

Rationale for new screening approach

The idea of direct visual screening of transformants expressing catalytically active REases is outlined in Figure 2. Such screening is based on the ability of active mutants to induce SOS response and thus requires the use of *E. coli* strains suitable for direct SOS response visualization. ts replication plasmid encodes the gene for protecting methyltransferase, and mutants of cognate restriction endonuclease to be screened are located on a compatible stably maintained plasmid. Under permissive conditions (30°C), the ts plasmid replicates and is stably maintained; at non-permissive temperature (42°C) replication is blocked and the plasmid is quickly lost from descendants of dividing cells. The left part of Figure 2 shows descendants of a cell expressing inactive REase. In contrast, the right part of the figure describes the scenario of events for active REase. In order to ensure stable replication of ts plasmid during the entry process of the second plasmid coding for REase mutant, heat shock and recovery of cells as well as the initial growth of colonies were performed at permissive temperature (Figure 2, Stages I and II). The plates were incubated at 30°C for the time period sufficient to form microcolonies (Figure 2, Stage II), and then their growth was continued at non-permissive temperature (Figure 2, Stage III). Incubation at 42°C results in the arrest of ts plasmid replication and asynchronous plasmid loss from some number of dividing cells within the colony. As a consequence, genomic DNA of such cells becomes hypomethylated due to reduced intracellular MTase concentration. At Stage III restriction-deficient colonies remain white (Figure 2, left part), while hypomethylated genomic DNA targets of some cells within restriction-proficient colonies are cleaved (Figure 2, red arrows)
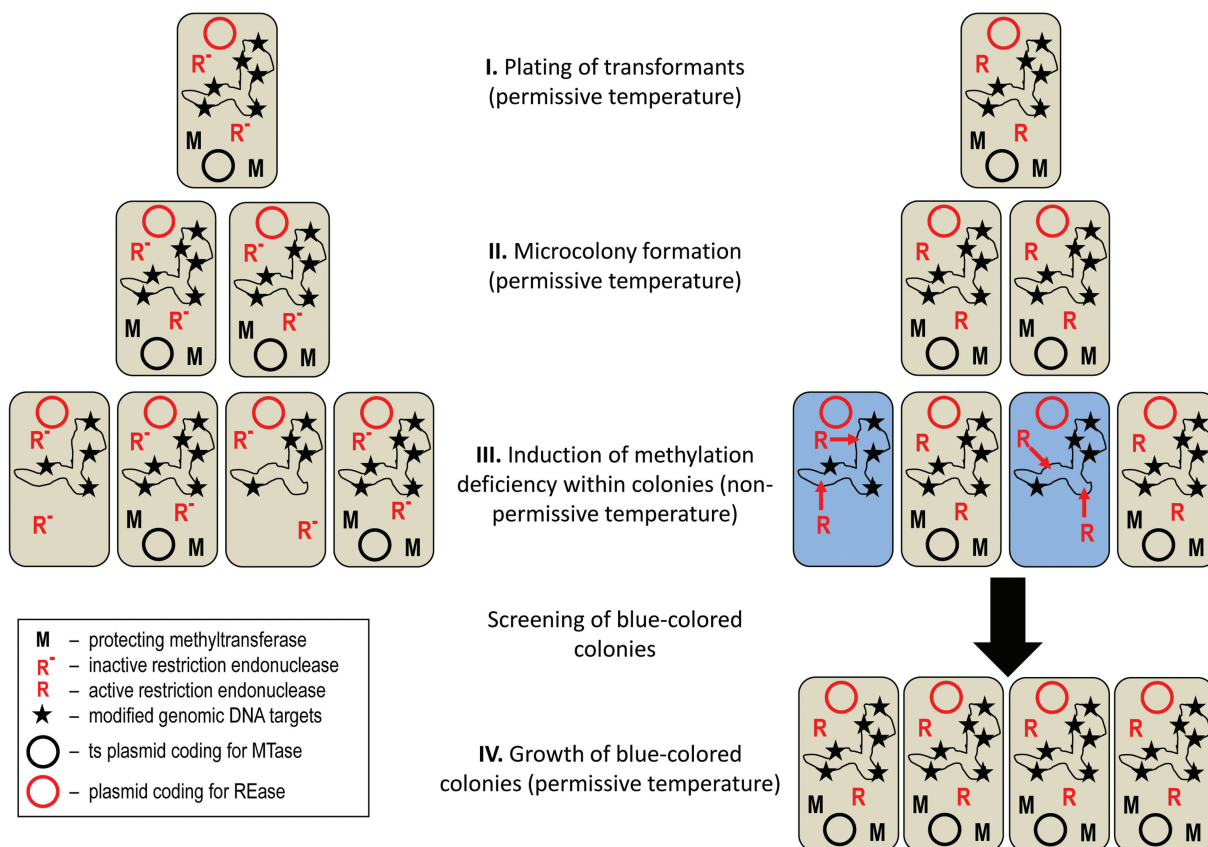
Figure 2. Schematic representation of segregation-induced methylation deficiency approach used in this work. Left and right panels show presumable scenarios of events in colonies expressing inactive and active REases, respectively. Red arrows indicate cleavage of unprotected DNA targets in the individual cells within the colony leading to induction of SOS response. The latter is manifested in blue color of individual cells, making the whole colony blue as well.

thus inducing SOS response. In the presence of X-Gal, such cells (and respective colonies) become blue as shown in Figure 2. It was assumed that some fraction of cells forming the blue colony would retain the ts plasmid, stay viable and thus could be recovered by transferring the colony onto a fresh LB agar plate at the permissive temperature (Figure 2, Stage IV).

Growth requirements for screening of cells expressing active *cfr10IR* gene

Previous studies with EcoRI and some other R-M systems revealed that in case when R-M system resides on a plasmid with ts replication, it induces SOS response under conditions of blocked plasmid replication (6,18–21,30). SOS induction was manifested in liquid cultures and on agar plates, which were antibiotic-free, and it was dependent on the activity of EcoRI (20). We asked ourselves whether it would be possible to combine both the antibiotic-based selection of transformants that received plasmid coding for Cfr10I[wt] restriction endonuclease and the programmable loss of ts replication plasmid carrying gene for protecting M.Cfr10I as depicted in Figure 2. Set-up of growth conditions optimal for SIMD-mediated SOS response induction was performed by CaCl$_2$-heat shock transformation of

dinD::*lacZ* indicator strain ER1992 (pTSA29-Cfr10IM) with the empty vector, pACYC184; the pACYC184 derivative, pAC-Cfr10Iwt, which codes for His-tagged Cfr10I[wt] and another derivative pAC-Cfr10IE204Q, encoding His-tagged catalytic mutant Cfr10I[E204Q] (24), all three conferring resistance to Cm. Transformation mixtures were plated onto different sectors of LB agar plates supplemented with X-Gal and antibiotics Ap and Cm, plates were incubated for varying periods of time (7–10 h) at 30°C, and then the growth was continued for 7–12 h at 37°C or 42°C before visual inspection of colonies. It has to be mentioned that the loss of ts replication plasmid from bacterial cells abolishes *de novo* synthesis of all proteins encoded by the plasmid, including β-lactamase, which ensures resistance to Ap. Ampicillin and some other β-lactams are known to induce SOS response (31), suggesting that growth of colonies under conditions mentioned above may result in SOS induction in case when ts replication plasmid encodes β-lactamase and Ap is added to the LB agar. However, we have found that incubation of plates at 30°C for 10 h followed by 10–12 h at 42°C resulted in dark blue color of colonies expressing active REase, whereas pACYC184-containing colonies remained white and pAC-Cfr10IE204Q were pale blue (see Table 3). The difference in color was even more
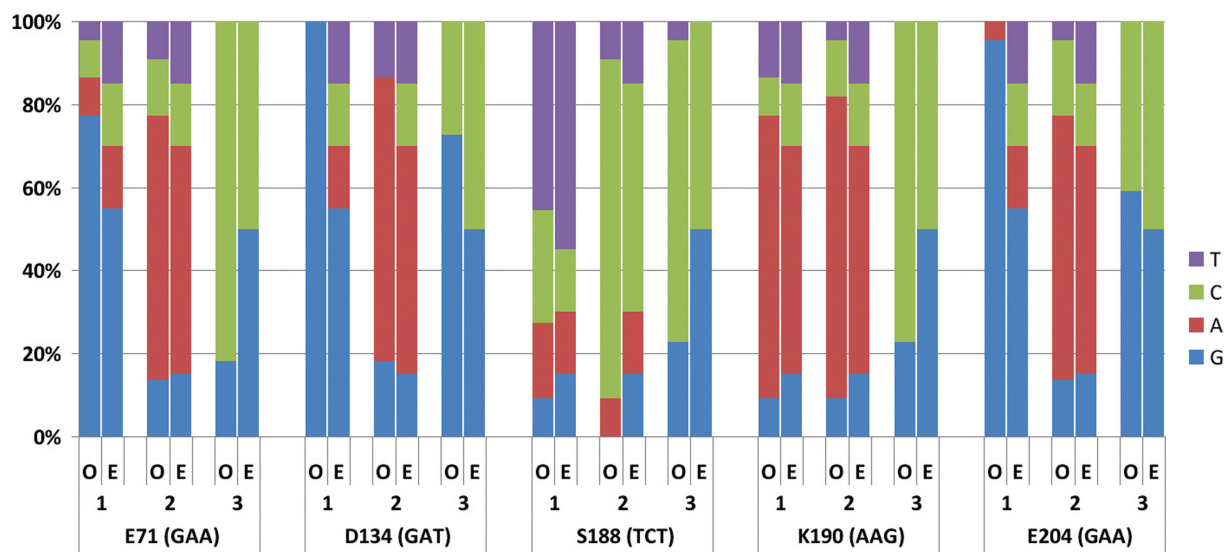
Figure 3. Comparison of expected and observed frequencies of G, A, T and C nucleotides at each position of five mutated codons. Numbers 1, 2 and 3 stand for first, second and third position of codons. Three letters next to the individual amino acid residues show triplets, which encode them in *cfr10I*^{wt}. Expected (E) frequencies represent distribution of four nucleotides at individual positions of degenerate oligonucleotides used for mutagenesis. The oligonucleotides were designed to have a 55% probability of containing a wild-type nucleotide at the first and the second position of codons and one of the three remaining nucleotides with equal 15% probability. The third codon position was designed to have a 50% probability of containing cytosine and 50% guanine. Observed (O) frequencies were calculated from results of sequencing of 22 plasmids isolated from randomly picked colonies of mutant library.

prominent after subsequent storage of plates for 24–72 h at 4°C. The absence of any Ap effect under SIMD conditions could be explained either by high intracellular β-lactamase concentration that remains high enough even after several rounds of division of plasmid-free cells, by accumulation of secreted β-lactamase within the growing colony thus lowering the local concentration of Ap within and in the vicinity of growing colony, or by both these factors.

Construction of library of Cfr10I^{E204Q} mutants

Codons of E71, D134, S188, K190 and E204 amino acid residues were semi-randomly mutated following the approach presented in Figure 1, and PCR-amplified full-length genes cloned back to the backbone of pAC-Cfr10IE204Q. The oligonucleotides used for mutagenesis (Table 1) were designed to have a 55% probability of containing a wt nucleotide at the first and the second position of codons and one of the three remaining nucleotides with equal 15% probability, while the third codon position was designed to have a 50% probability of containing cytosine and 50% guanine. The theoretical frequency of appearance of variants featuring wt amino acid sequence of Cfr10I was calculated to be 0.016%. The library of 10^6 mutants was created in ER2566 (pTSA29-Cfr10IM) using electroporation. Sequence analysis of plasmids isolated from 27 randomly picked colonies revealed three rearranged plasmids and two plasmids carrying exactly the same sequence as the PCR template used during three initial steps of mutagenesis (Figure 1), while sequences of remaining 22 plasmids featured mutations at expected positions. Collectively, sequencing results revealed that >80% of plasmids from

the library encode mutated genes. The frequency of appearance of individual nucleotides at each position of five mutated codons (Figure 3) revealed relatively good correlation between expected (Figure 3, E columns) and observed (Figure 3, O columns) values at many positions. The analysis of distribution of mutations was not comprehensive due to rather limited number of analyzed clones; however, for unknown reasons, the wt nucleotide G clearly dominated in first positions of D134 and Q204 codons.

Screening for Cfr10I^{E204Q} mutants able to induce SOS response

The library of plasmids coding for Cfr10I^{E204Q} mutants was used to transform ER1992 (pTSA29-Cfr10IM) by CaCl$_2$-heat-shock method, and the growth of transformants was performed following temperature regime described above: plates supplemented with Ap, Cm and X-Gal were incubated at 30°C for 10 h and then at 42°C for 12 h. Our initial screening trials revealed that efficiency of SOS repair induction (visualized by blue color development) depends on the number of colonies present on the LB agar plate, and that blue color-based screening works very well with up to 25 colonies per cm^2 of plate (~2000 colonies per standard Petri dish). Following these precautions 200 000 transformants were analyzed, and 64 blue/dark blue colonies (0.032%) were identified among them. Individual colonies were used to inoculate LB medium (5 ml), grown overnight at 30°C in the presence of Ap and Cm, and cultures were used for isolation of plasmids. All 64 plasmids were individually introduced back to ER1992 (pTSA29-Cfr10IM) to check whether the SOS response in isolated colonies was indeed

Table 2. Frequency and distribution of mutations in genes isolated during the screening procedure

Frequency[a]	Amino acid residue position and codon in wt gene					Additional mutations	Isolate
	E71 GAA[b]	D134 GAT[b]	S188 TCT[b]	K190 AAG	E204 GAA		
8×	GAG	GAT	TCT	AAG	GAG		Wild type
8×	GAG	GAC	TCT	AAG	GAG		
5×	GAG	GAC	TCC	AAG	GAG		
4×	GAA	GAT	TCT	AAG	GAG		
3×	GAG	GAT	TCC	AAG	GAG		
1×	GAG	GAT	TCG	AAG	GAG		
1×	GAA	GAT	TCG	AAG	GAG		
1×	GAA	GAC	TCT	AAG	GAG		
1×	GAG	GAC	TCG	AAG	GAG		
2×	GAG	GAC	TCT	AAG	GAG	L180P	L180P
1×	GAG	GAC	TCT	AAG	GAG	S26L	S26L
2×	CAG (Q)	GAC	TCC	AAG	GAG		E71Q
2×	CAG (Q)	GAT	TCT	AAG	GAG		
1×	CAG (Q)	GAT	TCC	AAG	GAG		
2×	CAG (Q)	GAT	TCT	AAG	GAG	R143G	E71Q/R143G
3×	TAG (Q[a])	GAC	TCT	AAG	GAG		E71Q[a] (supE44)
2×	TAG (Q[a])	GAT	TCT	AAG	GAG		
1×	TAG (Q[a])	GAC	TCC	AAG	GAG		
1×	CAG (Q)	GAC	TCG	ACC (T)	CTG (L)		E71Q/K190T/E204L
2×	GCC (A)	GAC	TCG	ACC (T)	GTG (V)		E71A/K190T/E204V
1×	GCC (A)	GAC	TCC	ACC (T)	GTG (V)		
1×	GCG (A)	GAT	TCC	ACC (T)	GTC (V)		
1×	GCC (A)	GAC	TCG	ACC (T)	ATC (I)		E71A/K190T/E204I
1×	GCG (A)	GAC	GCC (A)	ACC (T)	CAG (Q)		E71A/S188A/K190T/E204Q
1×	GAG	AAC (N)	TCG	AAC (N)	GTG (V)		D134N/K190N/E204V
1×	AAC (N)	GAC	TCT	AAG	GAG	D254G	E71N/254G
1×	GAG	GAC	GCG (A)	AAG	GAG		S188A
1×	GAG	GAC	GCC (A)	AAG	GAG	A78V	S188A/A78V
1×	GAG	GAC	TGC (C)	AAG	GAG		S188C

[a]Grey background shows isolates used in protein purification and/or transformation experiments.
[b]Boxed letters indicate nucleotides originating from the wild-type *cfr10IR* gene.

induced by the plasmids. As a result, 61 plasmids of 64 tested induced SOS response upon re-introduction into ER1992 (pTSA29-Cfr10IM) after growth under the above described conditions, indicating that the ability to induce SOS response in 95% of cases was determined by plasmids.

Sequence analysis of SOS inducing plasmids

In order to define the nature of mutations able to induce SOS response, cloned DNA fragments of all 61 plasmids were sequenced. Analysis of sequence data revealed that 60 plasmids encode *cfr10IR* mutants (Table 2), while the remaining plasmid contains *E. coli* genes *ybaE*, *ybaX* and *ybaW*. Twenty-nine distinct *cfr10IR* variants were detected at the nt level, from which 12 were found in multiple isolates, while the others were unique. Surprisingly, three variants contained the wt nucleotide A at third position of E71 codon, 10 variants—wt nucleotide T at third position of D134 codon and 11 variants— wt nucleotide T at third position of S188 codon (Table 2; respective nucleotides are highlighted on a grey background). Taking into account that mutagenic oligonucleotides were designed to contain either G or C in third position (see Table 1) it seems that PCR primers were not completely removed during purification of megaprimers after the first round of mutagenesis,

introducing additional level of variability. As expected, the wt sequence was the most common sequence to emerge from the screened library at the amino acid sequence level (35 of 60 sequenced clones, or 58%). Of these, two mutants contained additional L180P replacement and one—S26L mutation, both probably resulted from PCR errors (Table 2). Analysis of 25 remaining mutants revealed 12 distinct Cfr10I variants having up to four amino acid replacements (Table 2). It is noteworthy that six mutants contained the translation termination codon TAG at the position of E71 codon. ER1992 possesses mutation *supE44*, which ensures translation of some fraction of TAG codons into glutamine (Q), suggesting that the translated REase has Q at 71st position. The same E71Q mutation was also found in few other clones as well, but in those cases Q was encoded by CAG codon (Table 2). To distinguish between the two types of E71Q mutants, supE44-dependent E71Q replacements are marked with asterisk.

Characterization of Cfr10I mutants able to induce SOS response

To examine whether the catalytic activity of Cfr10I indeed is a prerequisite for SOS response induction, affinity-purified preparations of His-tagged Cfr10I[wt], Cfr10I[E204Q] and nine other mutants (Figure 4) were
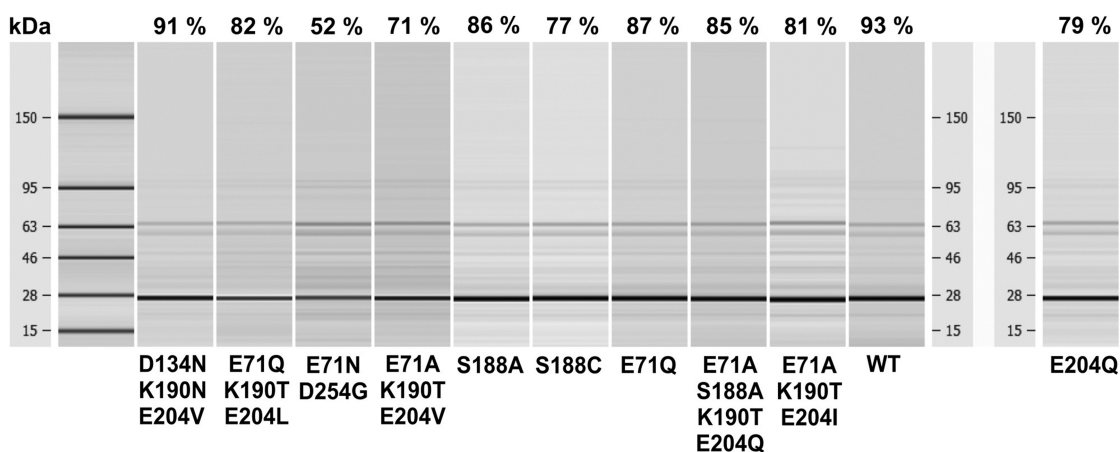
Figure 4. Gel-like image of a Protein 230 chip run of purified wt and mutant Cfr10I REase proteins generated with Agilent 2100 Bioanalyzer. Numbers on the top indicate physical purity (% of total proteins) of target R.Cfr10I proteins ~27 kDa in size.

examined in phage λ DNA cleavage assay. In parallel, plasmids carrying respective genes were used to transform ER1992 cells deficient in Cfr10I MTase, and transformants were grown either at 30°C or 42°C in the presence of X-Gal in order to investigate if bacterial cells may tolerate mutant proteins at any of these temperatures. λ DNA cleavage activities of mutant proteins, their impact on morphology and color of ER1992 (pTSA29-Cfr10IM) colonies grown under SIMD conditions as well as their requirements for MTase protection are summarized in Table 3. In accordance with previous observations (24), E71Q and S188A mutants retained significant residual activities (10% and 4% of His-tagged Cfr10I^{wt} activity, respectively) and both were toxic in the absence of protecting methylation. It should be noted that residual activities of E71Q and S188A mutants measured in this study differed slightly from those published previously (E71Q, 3.3% and S188A, 12%). These dissimilarities most probably arose from the differences in proteins used for measurements (His-tagged enzymes versus untagged) and from different temperatures at which activities were analyzed (37°C versus 55°C). Interestingly, S188 replacement to cysteine reduced the activity down to 0.5%, resulting in 8-fold activity decrease compared with S188A. In addition, ER1992 (pTSA29-Cfr10IM) cells expressing S188C formed larger colonies after their growth under SIMD conditions, while ER1992 cells survived at both tested temperatures in the absence of protecting methylation and formed blue colonies under conditions of constitutive S188C synthesis. The same was true for E71Q* mutant where E71 replacement to glutamine was encoded by translation termination codon TAG. A detailed analysis of the impact of individual mutations on the activity of Cfr10I was out of the scope of this work. Moreover, possible reasons why E71Q* and S188C behave so differently remain to be elucidated in future studies. However, taking into account the fact that there are three cysteines in Cfr10I, one could speculate that the appearance of additional cysteine in S188C may interfere with the proper folding of mutant protein thus resulting in decrease of the specific activity and/or

effective REase concentration within the cell. Likewise, the lowered toxicity of E71Q* could be explained by the fact that only a small fraction of TAG codons are translated into glutamine in the strain possessing the *supE44* mutation, leading to the lowered intracellular concentration of Cfr10I^{E71Q*}.
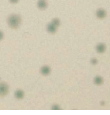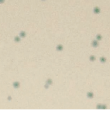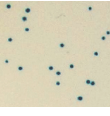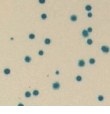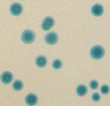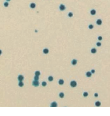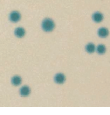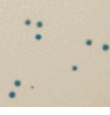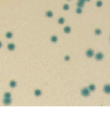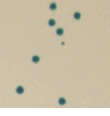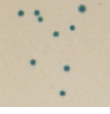
Investigation of catalytic properties of Cfr10I^{E204Q} and six other mutants isolated in this study (Table 3, four mutants shown on the bottom in the left part and remaining two on the right part) revealed that their *in vitro* cleavage activities are not detectable under assay conditions used. Of those, only a double mutant E71N/D254G formed colonies at 42°C in the absence of protecting methylation and yielded large blue colonies when introduced into ER1992 (pTSA29-Cfr10IM) and grown under SIMD conditions. In contrast, Cfr10I^{E204Q} and five other mutants were toxic in the absence of protecting methylation at both temperatures, but only Cfr10I^{E204Q} formed pale blue colonies after their growth at 30°C followed by incubation at 42°C, while colonies of other five mutants were blue/dark blue. However, these results do not suggest why the aforementioned mutants behaved differently, and further experiments are needed to explain why mutants isolated in this study induce SOS response to a higher extent compared with E204Q.

DISCUSSION

Even low activity of Type II REases leads to the killing of host cells in the absence of protective methylation and SOS-response-based visual screening can be used only for isolation of mutants possessing extremely low activities, which can be tolerated by unprotected cells. On the other hand, cognate methylation protects intracellular DNA from the cleavage by REase, but at the same time prevents the use of induced SOS response for isolation of catalytically active REases. Here, we demonstrated that SIMD could be used for direct visual screening of highly catalytically active revertants of Type II REase Cfr10I^{E204Q} when combined with the use of SOS-reporter strain. Screening was carried out by placing the
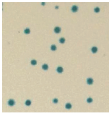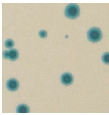
Table 3. Properties of mutants isolated during the screening procedure

Isolate	Relative specific activity (%)	Colony growth without MTase		Picture of colonies (1.5 × 1.5 cm)[a]
		30°C	42°C	
pACYC184 (control)	NA	NA	NA	
E204Q (pAC-Cfr10IE204Q)	ND	−	−	
Wild type (pAC-Cfr10Iwt)	100	−	−	
E71Q	10	−	−	
E71Q[a]	NA	+	+	
S188A	4	−	−	
S188C	0.5	+	+	
E71A/K190T/E204I	ND	−	−	
E71A/K190T/E204V	ND	−	−	
E71Q/K190T/E204L	ND	−	−	
E71A/S188A/K190T/E204Q	ND	−	−	

(continued)

Table 3. Continued

Isolate	Relative specific activity (%)	Colony growth without MTase		Picture of colonies (1.5 × 1.5 cm)[a]
		30°C	42°C	
D134N/K190N/E204V	ND	−	−	
E71N/D254G	ND	−	+	

[a]Competent ER1992 (pTSA29-Cfr10IM) cells were transformed by CaCl$_2$-heat-shock approach at 30°C. Colonies of transformants were grown for 10 h at 30°C and 12 h at 42°C, plates incubated for additional 48 h at 4°C and then photographed.
NA, not analyzed; ND, not detected.

gene of the protecting MTase into ts-replication plasmid pTSA29, inserting the *cfr10IR^{E204Q}* gene, which was semi-randomly mutated at five key amino acid residue positions, into stably maintained plasmid pACYC184 and then applying SIMD. Unexpectedly, electroporation of *dinD::lacZ* indicator strain ER1992 with empty vector pACYC184 resulted in some fraction of blue colonies on agar plates supplemented with Cm and X-Gal, suggesting that electrotransformation induces SOS response in some fraction of transformed cells and, thus, limited the use of electroporation for construction and screening of the library in the same experiment. Therefore, the experiment was accomplished in two steps: electroporation was used in the first step to generate the library of 10^6 mutants, and then pooled plasmids were used to transform ER1992 (pTSA29-Cfr10IM) by the CaCl$_2$-heat-shock method. Small scale sequencing of plasmids isolated from randomly picked transformants from the library revealed that >80% of plasmids encode mutated genes. However, distribution of mutations at first positions of D134 and E204 codons did not conform to the distribution expected based on structure of degenerate oligonucleotides used for mutagenesis, clearly indicating G dominance in both positions. Such discrepancy could be explained either by drawbacks in oligonucleotide synthesis or by mutagenesis conditions favorable for amplification of DNA fragments possessing G at indicated positions.

The frequency of blue/dark blue colonies among transformants analyzed by SIMD approach was found to be 0.032% (64 of 2 × 10^5). Of these, 95% (61 of 64) retained the ability to induce SOS response after re-transformation, demonstrating very high accuracy of this screening method. Surprisingly, sequencing of isolated plasmids revealed that one of plasmids carried *E. coli* genes *ybaE*, *ybaX* and *yba*, which were inserted into the cloning vector during construction of mutant

library. The cloned genes for unknown reasons induced SOS response and most probably originated from genomic DNA contaminating plasmid pAC-Cfr10IE204Q DNA used for both cloning and mutagenesis. Sixty other plasmids encoded 29 variants of *cfr10IR* gene differing at nucleotide sequence level and 12 among them were found in multiple isolates (Table 2). Taking into account the fact that only 20% of representatives of the library (2×10^5 of 10^6) were screened for SOS response induction, it seems likely that plasmids featuring identical sequences are 45 independent mutants rather than multiple isolates of 12 initial mutants. Amino acid sequence analysis revealed that 35 of 60 isolates encode amino acid residues of the wild type at all five mutated positions. Assuming that 80% of plasmids in the library encode the mutated *cfr10IR* gene, the observed frequency of isolation of wt gene from the library (at the level of amino acid sequence) is 0.019%. It is very close to the theoretical frequency of 0.016%, which was calculated based on probabilities of individual nucleotides at discrete positions within codons. On the other hand, this coincidence is surprising due to the finding that expected and observed distribution of nucleotides at individual positions in mutant library does differ and, in some cases, remarkably (Figure 3), and due to observation that several analyzed mutants inherited codons for individual amino acid residues from PCR template but not from primers used for mutagenesis (Table 2). Nevertheless, it may be concluded that the SIMD technique worked perfectly and allowed us to isolate true revertants of Cfr10I.

Purification and characterization of nine other mutants isolated during the screening procedure revealed various levels of relative specific activity. The highest was manifested by E71Q (10%), the lowest was observed in case of S188C (0.5%) and was undetectable in assay conditions for six isolated mutants as well as in the E204Q, which served as a template for mutagenesis (Table 3). Of note, the latter formed pale blue colonies under SIMD conditions, while colonies of isolated mutants were either blue or dark blue (Table 3). Finally, all of them except E71N/ D254G were toxic in the absence of protective methylation at 42°C, raising a question about the relationship between mutant activity and their ability to kill host cells and induce SOS response of certain strength in the absence of protective methylation. Based on more intense blue color of colonies, one could speculate that activities of four mutants shown on the bottom in the left part of Table 3 and remaining two on the right part are higher than E204Q thus resulting in a stronger SOS response. However, even subtle changes in the structure of mRNA or protein leading to altered expression level or stability of mutant enzyme within ER1992 (pTSA29-Cfr10IM) cells may have effect on its intracellular concentration, thus leading to different outcome under SIMD conditions. On the other hand, investigating why isolated mutants induced a higher SOS response or what the impact of individual replacements on the enzyme activity was remains a topic for future studies.

It is important to note that although efficiency of SIMD approach was demonstrated only for the screening of catalytically active Cfr10I variants, it may be easily adapted for other applications. For instance, pTSA29-Cfr10IM may be directly used for SIMD-mediated isolation of genes coding for active REases, which do not cleave DNA targets, modified by M.Cfr10I, from cloning libraries. However, the use of two MTases may be required for SIMD-mediated isolation of REase mutants of altered specificity: one for protection of cognate DNA targets (MTase gene located on a stable plasmid) and the second for protection of genomic DNA targets from cleavage by altered specificity mutants (MTase gene located on a ts-replication plasmid). For instance, screening of Cfr10I mutants acting on 5′-TCCGGA, 5′-CCCGGG, 5′-PyCCGGPu or 5′-CCGG targets could be performed by inserting the gene for M.Cfr10I into the plasmid, which is stably maintained (next to the mutated *cfr10IR* gene) and placing the gene for M.MspI (which protects 5′-CCGG targets) into the ts-replication plasmid. Under SIMD conditions, only Cfr10I mutants of altered specificity should be able to induce SOS response; and based on results described in this work, one may expect to isolate active mutants of altered specificity regardless of their activity level.

REFERENCES

1. Roberts,R.J. and Halford,S.E. (1993) In: Linn,S.M., Lloyd,R.S. and Roberts,R.J. (eds), *Nucleases*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 35–88.
2. Raleigh,E. and Brooks,J.E. (1998) In: De Bruijn,F.J., Lupski,J.R. and Weinstock,G.M. (eds), *Bacterial Genomes: Physical Structure and Analysis*. Springer, Chapman&Hall, New York, pp. 78–92.
3. Wilson,G.G. (1991) Organization of restriction-modification systems. *Nucleic Acids Res.*, **19**, 2539–2566.
4. Arber,W. (2000) Genetic variation: molecular mechanisms and impact on microbial evolution. *FEMS Microbiol Rev.*, **24**, 1–7.
5. Kobayashi,I. (2001) Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nucleic Acids Res.*, **29**, 3742–3756.
6. Naito,T., Kusano,K. and Kobayashi,I. (1995) Selfish behavior of restriction-modification systems. *Science*, **267**, 897–899.
7. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S., Dryden,D.T., Dybvig,K. *et al.* (2003) A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Res.*, **31**, 1805–1812.
8. 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–D236.
9. Heitman,J. and Model,P. (1990) Mutants of the EcoRI endonuclease with promiscuous substrate specificity implicate

residues involved in substrate recognition. *EMBO J.*, **9**, 3369–3378.

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

11. Dorner,L.F. and Schildkraut,I. (1994) Direct selection of binding proficient/catalytic deficient variants of BamHI endonuclease. *Nucleic Acids Res.*, **22**, 1068–1074.

12. Dorner,L.F., Bitinaite,J., Whitaker,R.D. and Schildkraut,I. (1999) Genetic analysis of the base-specific contacts of BamHI restriction endonuclease. *J. Mol. Biol.*, **285**, 1515–1523.

13. 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.

14. 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.

15. 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.

16. Elledge,S.J., Sugiono,P., Guarente,L. and Davis,R.W. (1989) Genetic selection for genes encoding sequence-specific DNA-binding proteins. *Proc. Natl. Acad. Sci. USA.*, **86**, 3689–3693.

17. Heitman,J., Ivanenko,T. and Kiss,A. (1999) DNA nicks inflicted by restriction endonucleases are repaired by a RecA- and RecB-dependent pathway in Escherichia coli. *Mol. Microbiol.*, **33**, 1141–1151.

18. Kulakauskas,S., Lubys,A. and Ehrlich,S.D. (1995) DNA restriction-modification systems mediate plasmid maintenance. *J. Bacteriol.*, **177**, 3451–3454.

19. Kusano,K., Naito,T., Handa,N. and Kobayashi,I. (1995) Restriction-modification systems as genomic parasites in competition for specific sequences. *Proc. Natl. Acad. Sci. USA.*, **92**, 11095–11099.

20. Handa,N., Ichige,A., Kusano,K. and Kobayashi,I. (2000) Cellular responses to postsegregational killing by restriction-modification genes. *J. Bacteriol.*, **182**, 2218–2229.

21. Ichige,A. and Kobayashi,I. (2005) Stability of EcoRI restriction-modification enzymes in vivo differentiates the EcoRI restriction-modification system from other postsegregational cell killing systems. *J. Bacteriol.*, **187**, 6612–6621.

22. Asakura,Y. and Kobayashi,I. (2009) From damaged genome to cell surface: transcriptome changes during bacterial cell death triggered by loss of a restriction-modification gene complex. *Nucleic Acids Res.*, **37**, 3021–3031.

23. Bozic,D., Grazulis,S., Siksnys,V. and Huber,R. (1996) Crystal structure of *Citrobacter freundii* restriction endonuclease Cfr10I at 2.15 A resolution. *J. Mol. Biol.*, **255**, 176–186.

24. Skirgaila,R., Grazulis,S., Bozic,D., Huber,R. and Siksnys,V. (1998) Structure-based redesign of the catalytic/metal binding site of Cfr10I restriction endonuclease reveals importance of spatial rather than sequence conservation of active centre residues. *J. Mol. Biol.*, **279**, 473–481.

25. Fomenkov,A., Xiao,J.P., Dila,D., Raleigh,E. and Xu,S.Y. (1994) The 'endo-blue method' for direct cloning of restriction endonuclease genes in *E. coli.. Nucleic Acids Res.*, **22**, 2399–2403.

26. Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology*. John Wiley & Sons, Inc, New York.

27. Phillips,G.J. (1999) New cloning vectors with temperature-sensitive replication. *Plasmid*, **41**, 78–81.

28. Chang,A.C. and Cohen,S.N. (1978) Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.*, **134**, 1141–1156.

29. Brøns-Poulsen,J., Nøhr,J. and Larsen,L.K. (2002) Megaprimer method for polymerase chain reaction-mediated generation of specific mutations in DNA. *Methods Mol. Biol.*, **182**, 71–76.

30. Liu,Y. and Kobayashi,I. (2007) Negative regulation of the EcoRI restriction enzyme gene is associated with intragenic reverse promoters. *J. Bacteriol.*, **189**, 6928–6935.

31. Miller,C., Thomsen,L.E., Gaggero,C., Mosseri,R., Ingmer,H. and Cohen,S.N. (2004) SOS response induction by β-lactams and bacterial defense against antibiotic lethality. *Science*, **305**, 1629–1631.