

**VILNIUS UNIVERSITY
INSTITUTE OF BIOTECHNOLOGY**

Rimantas Šapranauškas

**INVESTIGATION OF TYPE IIS RESTRICTION
ENDONUCLEASE BfiI DOMAIN ORGANIZATION BY USING
A NEW RANDOM GENE DISSECTION APPROACH**

Summary of doctoral dissertation

Physical science, biochemistry (04P), nucleic acids, protein biosynthesis (P320)

Vilnius 2008

This study has been carried out at the Institute of Biotechnology during 1998–2008.

Dissertation is maintained by extern.

Scientific consultant:

Dr. Arvydas Lubys (“Fermentas” UAB, physical science, biochemistry – 04P, nucleic acids, protein biosynthesis P320)

Evaluation board of dissertation of Biochemistry trend:

Chairman:

Dr. Gintautas Žvirblis (Institute of Biotechnology, physical science, biochemistry – 04P, nucleic acids, protein biosynthesis – P320)

Members:

Prof. Dr. Donaldas Čitavičius (Vilnius University, biomedical science, biology – 01B, microbiology, bacteriology, virusology, mycology – B230)

Dr. Rolandas Meškys (Institute of Biochemistry, physical science, biochemistry – 04P, nucleic acids, protein biosynthesis – P320)

Prof. Dr. habil. Kęstutis Sasnauskas (Institute of Biotechnology, physical science, biochemistry – 04 P, nucleic acids, protein biosynthesis – P320)

Dr. Saulius Serva (“Fermentas” UAB, physical science, biochemistry – 04 P, nucleic acids, protein biosynthesis – P320)

Official opponents:

Dr. Alma Gedvilaitė (Institute of Biotechnology, physical science, biochemistry – 04 P, nucleic acids, protein biosynthesis – P320)

Dr. Violeta Pependikytė (“SORPO” UAB, physical science, biochemistry – 04 P, nucleic acids, protein biosynthesis – P320)

The dissertation defend will take place at 10 a.m. on 28th of June, 2008, at the Institute of Biotechnology.

Address: V.A.Graičiūno 8, Vilnius LT-02241, Lithuania.

Summary of doctoral dissertation was distributed on 28th of May 2008.

The dissertation is available at the Library of Institute of Biotechnology and at the Library of Vilnius University.

**VILNIAUS UNIVERSITETAS
BIOTECHNOLOGIJOS INSTITUTAS**

Rimantas Šapranauskas

**IIS POTIPIO RESTRIKCIJOS ENDONUKLEAZĖS BfiI
DOMENINĖS ORGANIZACIJOS TYRIMAS PANAUDOJANT
NAUJĄ GENO ATSITIKTINIO PADALINIMO METODĄ**

Daktaro disertacijos santrauka

Fiziniai mokslai, biochemija (04P), nukleorūgštys, baltymų sintezė (P320)

Vilnius 2008

Disertacija rengta 1998 – 2008 metais Biotechnologijos institute.

Disertacija ginama eksternu.

Mokslinis konsultantas:

dr. Arvydas Lubys („Fermentas“ UAB, fiziniai mokslai, biochemija – 04P, nukleorūgštys, baltymų sintezė – P320)

Disertacija ginama Vilniaus Universiteto Biochemijos mokslo krypties taryboje:

Pirmininkas:

Dr. Gintautas Žvirblis (Biotechnologijos institutas, fiziniai mokslai, biochemija - 04P, nukleorūgštys, baltymų sintezė - P320)

Nariai:

Prof. dr. Donaldas Čitavičius (Vilniaus Universitetas, biomedicinos mokslai, biologija – 01 B, mikrobiologija, bakteriologija, virusologija, mikologija – B230)

Dr. Rolandas Meškys (Biochemijos institutas, fiziniai mokslai, biochemija - 04P, nukleorūgštys, baltymų sintezė - P320)

Prof. habil. dr. Kęstutis Sasnauskas (Biotechnologijos institutas, fiziniai mokslai, biochemija – 04 P, nukleorūgštys, baltymų sintezė - P320)

Dr. Saulius Serva („Fermentas“ UAB, fiziniai mokslai, biochemija – 04 P, nukleorūgštys, baltymų sintezė - P320)

Oponentai:

Dr. Alma Gedvilaitė (Biotechnologijos institutas, fiziniai mokslai, biochemija - 04P, nukleorūgštys, baltymų sintezė - P320)

Dr. Violeta Popendikytė („SORPO“ UAB, fiziniai mokslai, biochemija – 04 P, nukleorūgštys, baltymų sintezė - P320)

Disertacija bus ginama viešame Biochemijos mokslo krypties tarybos posėdyje 2008 m. Birželio mėn. 28 d. 10 val. Biotechnologijos instituto konferencijų salėje. Adresas: V.A.Graičiūno g. 8, Vilnius 02241, Lietuva

Disertacijos santrauka išsiuntinėta 2008 m. gegužės mėn. 28 d.

Disertaciją galima peržiūrėti Biotechnologijos instituto ir Vilniaus Universiteto bibliotekose.

Contents

Contents	1
Abbreviations.....	2
Introduction.....	3
1. Materials and methods	6
1.1. Reagents, enzymes and kits.	6
1.2. Bacteria strains.....	6
1.3. Plasmids	6
1.4. Oligonucleotides	7
1.5. Methods.....	8
1.5.1. Purification of genomic DNA, plasmids and DNA fragments	8
1.5.2. Construction of recombinant plasmids.....	8
1.5.3. Construction of RR1-FokIM and RR1-BfiIM strains	9
1.5.4. Random insertional mutagenesis.....	10
1.5.5. Transformation of <i>Escherichia coli</i>	11
1.5.6. Determination of the endonuclease activities	11
1.5.7. Western blotting	11
1.5.8. DNA binding studies	12
2. Results and discussion.....	13
2.1. Analysis of genetic organization of BfiI restriction–modification system.....	13
2.1.1. Cloning of BfiI restriction-modification genes	13
2.1.2. Deletion analysis of BfiI R-M system.....	13
2.1.3. Nucleotide sequence analysis of BfiI restriction–modification system	15
2.2. Random Gene Dissection technique and its use for investigation of protein domain organization.....	17
2.2.1. Principles of Random Gene Dissection method.....	18
2.2.2. Random dissection of the model gene encoding FokI REase	19
2.2.3. Screening and analysis of endonucleolytically active FokI mutants.	20
2.3. Investigation of BfiI restriction endonuclease domain organization.....	24
2.3.1. Random dissection of the <i>bfiIR</i> gene.....	24
2.3.2. BfiI mutants of wild-type specificity: isolation and structure.....	24
2.3.3. BfiI mutants of relaxed specificity: isolation and structure	26
2.3.4. Investigation of functions of putative BfiI domains	27
Discussion.....	29
Conclusions.....	34
List of publications	35
Acknowledgements.....	36
Reziümé	37
References.....	39

Abbreviations

Ap ^r	resistance to ampicillin
aa	amino acid
ORF	open reading frame
ATP	adenosine triphosphate
bp	base pair
BSA	bovine serum albumin
<i>cat</i>	gene of chloramphenicol acetyltransferase
CIAP	calf intestinal alkaline phosphatase
Cm ^r	resistance to chloramphenicol
DNase I	deoxyribonuclease I
EDTA	ethylene diamine tetraacetic acid
IPTG	izopropyl-β-D-tiogalactopyranoside
Km ^r	resistance to kanamycine
LB	Luria-Bertani broth
MTase	DNA methyltransferase
SDS	sodium dodecylsulfate
SDS-PAGE	SDS - polyacrylamide gel electrophoresis
nt	nucleotide
PCR	polymerase chain reaction
REase	restriction endonuclease
RGD	Random Gene Dissection
R-M	restriction-modification
Sm ^r	resistance to streptomycin
Tc ^r	resistance to tetracycline

Introduction

The proteins of restriction–modification (R-M) systems perform two enzymatic reactions: restriction endonuclease (REase) recognizes and cleaves specific 4-8 bp long sequences in the DNA, while methyltransferase (MTase) modifies the same sequences by adding the methyl group to cytosine or adenine, and thereby rendering cellular DNA resistant to cleavage by REase. In contrast, exogenous DNA is not modified and therefore it is susceptible to REase cleavage. It is suggested that in this way restriction-modification systems protect bacterial cells from invading DNA such as bacteriophages.

R-M systems comprise a large group of enzymes that are classified into four major types according to their differences in subunit composition, cleavage position or cofactor requirement (Roberts et al., 2003). Type II REases cleave target DNA inside of recognition sequence or close to it. Due to this property, these enzymes have become widely used tools in molecular biology, biotechnology and diagnostics. The intensive use of restriction endonucleases has stimulated search of R-M systems with new specificities, resulting in more than 3600 type II restriction endonucleases with over 270 different specificities which have been identified to date (www.rebase.neb.com). Type II REases recognize several kinds of nucleotide sequences (symmetric/asymmetric, continuous/discontinuous), they may have one or more functions (monofunctional/bifunctional) and act as monomers, dimers or even tetramers. Due to such big variety type II REases are subdivided in 11 subtypes. Type II REases differ also in their structural domain organization - e.g. orthodox type II enzymes are composed of a single domain where recognition and catalysis functions are tightly coupled and require the high level of coordination between them; in contrast, enzymes of types IIE and IIS have a two-domain organization. Representatives of type IIE REases, NaeI and EcoRII, have two domains, both of which specifically interact with the recognition sequence. One of these domains, in addition to the DNA recognition function, catalyses DNA cleavage (Huai et al., 2001; Mucke et al., 2002). In contrast, two domains of type IIS enzymes differ in their function – DNA-binding domain recognizes specific targets, while catalytic domain cleaves DNA. When separated from the DNA-binding domain, the catalytic domain acts as a non-specific nuclease (Li et al., 1992; Kriukiene, 2006).

Analysis of protein domain organization plays a significant role in understanding protein structure, its activity mechanism or in deducing evolutionary origin. Furthermore, knowledge of domain organization has also found practical application in protein engineering experiments. For instance, the nucleolytic domains of FokI and BmrI REases have been successfully used in construction of chimeric restriction endonucleases (Kim et al., 1996; Chan et al., 2007; Zhang et al., 2007).

To date, limited proteolysis is the most popular method for analysis of protein domain organization. This approach is based on limited digestion of protein by proteases, which cleave flexible, surface-exposed interdomain linkers faster than the intradomain protein parts that are either buried in the protein core or comprise regular secondary structural elements. Protein fragments produced by this procedure are then isolated and individually tested for their biochemical properties, a task that is sometimes complicated and labor intensive. Furthermore, limited proteolysis not always accurately indicates the interdomain part of the protein. Correctly folded native proteins may have

locally occurring unfolded stretches up to 12 amino acid residues that can serve as a substrate for a protease used (Hubbard et al., 1994). Also, some interdomain regions may be too short, may form secondary structures or may have no specific amino acid sequences required for interaction with the protease used. These drawbacks put forward the need of developing alternative methods for protein domain organization analysis, which could possibly supplement or even replace the approach of limited proteolysis.

In this work, both the genetic organization of BfiI restriction–modification system and the domain structure of REase BfiI were investigated. The type IIS REase BfiI recognizes asymmetric sequence 5'-ACTGGG and cleaves the upper DNA strand 5 nucleotides and the bottom DNA strand 4 nucleotides away from the recognition site in the absence of magnesium ions. A new technique - the Random Gene Dissection (RGD) was developed and used to study domain architecture of BfiI restriction endonuclease.

The aims of this study were:

1. Cloning of BfiI restriction-modification system genes and characterization of proteins comprising the system.
2. Developing the new method, Random Gene Dissection (RGD), for analysis of protein domain organization and testing it on a type IIS restriction endonuclease FokI with known two-domain structure.
3. Investigation of BfiI restriction endonuclease domain organization by using RGD.

Scientific novelty:

The genes coding for the BfiI restriction-modification system were cloned and sequenced. The BfiI R-M system consists of three proteins: two N4-methylcytosine methyltransferases and a restriction endonuclease. The BfiI REase recognizes asymmetric sequence, cleaves outside the target and thus belongs to the type IIS group of restriction enzymes, suggesting that it might be composed of two structural domains like the archetypal type IIS REase FokI. The new method, Random Gene Dissection (RGD), was developed for analysis of protein domain organization. The performance of RGD was first assayed on a well-studied type IIS REase FokI. The RGD technique enabled precise separation of target recognition and catalytic domains of FokI, at the same time the interdomain region was accurately located. RGD was then applied to investigate the domain organization of BfiI restriction endonuclease. Analysis revealed that BfiI is composed of two domains. The N-terminal domain catalyses DNA cleavage, while the C-terminal one is responsible for target recognition. In addition, complementary protein fragments were identified with both FokI and BfiI by using RGD. These results show that the RGD technique is a powerful method for analyzing a protein domain structure, producing complementary fragments of desirable protein and optimization of functional domain size in protein engineering works.

The new findings presented for defense:

1. The genes of BfiI restriction–modification system were cloned in *Escherichia coli* cells and their functions and nucleotide sequences were determined.

2. A new approach for analysis of protein domain organization – Random Gene Dissection (RGD) was proposed.
3. RGD suitability for protein domain analysis was tested on the model protein - FokI REase.
4. Domain organization of the BfiI restriction endonuclease was analyzed by using the RGD approach. The position of interdomain region was determined.
5. Functions of BfiI domains were confirmed – the N-terminal domain is responsible for DNA cleavage, whereas the C-terminal domain recognizes and binds DNA target.
6. Complementary protein fragments of FokI and BfiI were generated by RGD.

1. Materials and methods

1.1. Reagents, enzymes and kits.

All reagents used in this study were reagent-grade commercial products. Bovine pancreatic RNase and DNase I were obtained from Sigma, porablot PVDF membrane, [α -³³P] dATP and [γ -³³P] dATP were obtained from GE Healthcare. All other enzymes, protein and DNA molecular weight markers and kits were obtained from Fermentas.

1.2. Bacteria strains

<i>Bacillus firmus</i> S8120	wild type strain, encoding enzymes of BfiI restriction-modification system;
<i>E. coli</i> ER2267	F' <i>proA</i> ⁺ <i>B</i> ⁺ <i>lacI</i> ^q Δ (<i>lacZ</i>) <i>M15</i> <i>zzf::mini-Tn10</i> (Km ^r) / <i>e14</i> ⁻ (<i>mcrA</i> ⁻) <i>endA1</i> <i>glnV44</i> <i>thi-1</i> Δ (<i>mcrC-mrr</i>) <i>114::IS10</i> Δ (<i>argF-lac</i>) <i>U169</i> <i>recA1</i> (New England Biolabs); strain was used as a host for subcloning procedures;
<i>E. coli</i> ER1992	F ⁻ λ Δ (<i>argF-lac</i>) <i>U169</i> <i>glnV44</i> <i>e14</i> ⁻ <i>dinD1::Mu</i> <i>dI1734</i> (Km ^r , LacZ ⁺) <i>endA1</i> <i>thi-1</i> Δ (<i>mcrC-mrr</i>) <i>114::IS10</i> (Fomenkov et al., 1994); strain was used for selection of SOS response inducing mutants;
<i>E. coli</i> ER1992/F'	F' ⁺ <i>Tn10</i> (Tc ^r) <i>proA</i> ⁺ <i>B</i> ⁺ <i>lacI</i> ^q Δ (<i>lacZ</i>) <i>M15</i> / Δ (<i>argF-lac</i>) <i>U169</i> <i>glnV44</i> <i>e14</i> ⁻ <i>dinD1::Mu</i> <i>dI1734</i> (Km ^r , LacZ ⁺) <i>endA1</i> <i>thi-1</i> Δ (<i>mcrC-mrr</i>) <i>114::IS10</i> ; strain was used for analysis of insertional mutants (G. Mitkaitė, unpublished);
<i>E. coli</i> RR1	F ⁻ <i>leuB6</i> <i>proA2</i> <i>thi-1</i> <i>araC14</i> <i>lacY1</i> <i>galK2</i> <i>xyl-5</i> <i>mtl-1</i> <i>rpsL20</i> (Sm ^r) <i>glnV44</i> Δ (<i>mcr-mrr</i>) (Bolivar et al., 1977); was used for construction of strains with F' factors encoding BfiI and FokI methyltransferases;
<i>E. coli</i> XL1-Blue	F' ⁺ <i>Tn10</i> (Tc ^r) <i>proA</i> ⁺ <i>B</i> ⁺ <i>lacI</i> ^q Δ (<i>lacZ</i>) <i>M15</i> / <i>recA1</i> <i>endA1</i> <i>gyrA96</i> <i>thi-1</i> <i>hsdR17</i> <i>glnV44</i> <i>relA1</i> <i>lac</i> (Bullock et al., 1987); was used for construction of strains with F' factors encoding BfiI and FokI methyltransferases.

1.3. Plasmids

pBR322	cloning vector; Tc ^r Ap ^r ; ColE1 replicator (Bolivar et al., 1977);
pFokIRM	plasmid encoding FokI R-M system; Ap ^r ; ColE1 replicator (K. Stankevičius, unpublished);
pUC19	cloning vector; Ap ^r ; ColE1 replicator (Yanisch-Perron et al., 1985);
pUC57	cloning vector; Ap ^r ; ColE1 replicator (Fermentas);
pACYC184	cloning vector; Tc ^r Cm ^r ; p15A replicator (Chang and Cohen, 1978);

pZA34-MCS-1	expression vector; Cm ^r ; p15A replicator (Lutz and Bujard, 1997);
pBRmTn10dKm-NotI	vector encoding miniTn10 transposon; Tc ^r Ap ^r ; ColE1 replicator (G. Mitkai�, unpublished).

1.4. Oligonucleotides

- Universal sequencing primers (Fermentas):

M13/pUC dir	5'-GTAAAACGACGGCCAGT
M13/pUC rev	5'-CAGGAAACAGCTATGAC
pBR322 EcoRI cw	5'-GTATCACGAGGCCCT
pBR322 Sall ccw	5'-AGTCATGCCCCGCGC

- Primers used for sequencing of insertion sites of DNA cassette (MWG Biotech):

Cmol3	5'-GTTCTTTACGATGCCATTGGG
Cmol4	5'-GTGATGGCTTCCATGTCGGC

- Primers for construction of recombinant plasmids (MWG Biotech):

Primers used for amplification of DNA cassette:

Cmol1	5'- TAAAAAGTCTTC AGGAGCTAAGGAAGC - STOP codon (TAA) is in bold, BpiI site (GTCTTC) is in bold-italic;
Cmol2	5'- <u>CATAAAGTCTTC</u> <u>CTCCTTACGCCCCGCCCTGCC</u> - initiation codon and SD sequence (on complementary strand) are underlined, BpiI site (GTCTTC) is in bold-italic.

Primers used for amplification of wild type *bfiIR* gene and DNA sequences encoding N- and C-terminal BfiI polypeptides:

Bfi-N	5'-GTCCGAGGTTTTTTAAATGAATTTTTTCTCTTTACA – primer corresponding to the 5'-end of <i>bfiIR</i> gene; SD sequence and initiation codon is underlined;
Bfi-C	5'-CCAGCTATAACTTTGCTTAAAATG – primer complementary to the 3'-end of <i>bfiIR</i> gene; STOP codon is underlined;
Bfi130	5'-CTACTAAAGATTCTCCTAAG – complementary primer, designed for amplification of the N-terminal part of <i>bfiIR</i> gene;
BfiIN1-185	5'- <u>TTAATCATAAAGTAGATTCCAGCC</u> – 3'-terminal complementary primer, designed for amplification of DNA sequence encoding the BfiIN ¹⁻¹⁸⁵ oligopeptide; STOP codon is underlined;
BfiIN1-192	5'- <u>TTATGTCAAGTTTGTGTACGCTC</u> - 3'-terminal complementary primer, designed for amplification of DNA sequence encoding the BfiIN ¹⁻¹⁹² oligopeptide; STOP codon is underlined;
BfiIN1-204	5'- <u>TTAAGTAACAATTAAGTAACTCTCTC</u> - 3'-terminal complementary primer, designed for amplification of DNA sequence encoding the BfiIN ¹⁻²⁰⁴ oligopeptide; STOP codon is underlined;
BfiIC186-358	5'- <u>GGAGGATAACATATGGAGCGTACAACAACTTGAC</u> – 5'-terminal primer designed for amplification of DNA sequence encoding the BfiIC ¹⁸⁶⁻³⁵⁸ oligopeptide; SD site and initiation codon is underlined;

BfiIC193-358 5'-GGAGGATAACATATGTTAGATGAAACAGAGAGAG -
5'-terminal primer designed for amplification of DNA sequence encoding the BfiIC¹⁹³⁻³⁵⁸ oligopeptide; SD site and initiation codon is underlined;

BfiIC205-358 5'-GGAGGATAACATATGTTAGGTCACGCTGATACAG -
5'-terminal primer designed for amplification of DNA sequence encoding the BfiIC²⁰⁵⁻³⁵⁸ oligopeptide; SD site and initiation codon is underlined.

• Oligonucleotides used for protein-DNA binding experiments (MWG Biotech):

Bfispc+ 5'-CAGCACTGGGGGCCAGsATGGT

Bfispc- 5'-ACCATCsTGGCCCCAGTGCTG

Bfinsp+ 5'-CAGCACAGGGGGCCAGsATGGT

Bfinsp- 5'-ACCATCsTGGCCCCTGTGCTG

BfiI recognition sequence underlined, “s” indicates the position of phosphorothioate.

1.5. Methods

1.5.1. Purification of genomic DNA, plasmids and DNA fragments

Isolation of genomic DNA from *Bacillus firmus* S8120 cells pretreated with lysozyme was carried out as described (Marmur, 1961). Plasmids were prepared by using the alkaline-lysis procedure (Sambrook et al., 1989). Low molecular mass RNA was removed using silica powder suspension in 6 M NaClO₄ solution (Marko et al., 1982). DNA fragments from agarose gels were purified using “DNA Extraction Kit” (Fermentas).

1.5.2. Construction of recombinant plasmids

Construction of genomic library, deletion analysis, subcloning experiments, restriction mapping and preparing of nested deletions by using BAL31 were made according standard protocols (Sambrook et al., 1989; Ausubel et al., 1992).

Construction of pUCm plasmid. The cassette for insertional mutagenesis was constructed by PCR amplification of *cat* gene using *Taq* DNA polymerase, DNA of pACYC184 as template and primers Cm11 and Cm12. PCR fragment of 0.7 kb was phosphorylated using T4 kinase and blunted using T4 DNA polymerase. Then the DNA cassette was cloned into Eco32I-digested and CIAP-dephosphorylated pUC57 vector.

Construction of pACFokIM plasmid. The Bsu15I DNA fragment of 3.0 kb which encodes FokI methyltransferase was cleaved out from pFokIRM and ligated with the PvuII-Eco88I DNA fragment of 2.9 kb from pACYC184. Bsu15I and Eco88I-generated cohesive DNA ends were blunted using T4 DNA polymerase.

Construction of pTn-FokIM plasmid. The same Bsu15I DNA fragment from pFokIRM was subcloned into NotI-digested pBRmTn10dKm-NotI vector. Bsu15I and NotI-generated cohesive DNA ends were blunted using T4 DNA polymerase.

Construction of pUCFokI plasmid. The SspI-Bpu10I DNA fragment of 1.85 kb from pFokIRM encompassing the *fokIR* gene was subcloned into pUC57 vector, digested with SmaI, under the control of P_{lac} promoter. Bpu10I-generated cohesive DNA ends were blunted using T4 DNA polymerase.

Construction of pACBfiM plasmid. The KspAI-Eco72I DNA fragment of 3.18 kb encoding both BfiI methyltransferases was cleaved out from pBfiIRM14 and ligated into PvuII-Eco88I cleaved pACYC184. Eco88I-generated cohesive DNA ends were blunted using T4 DNA polymerase.

Construction of pTn-BfiIM plasmid. The same KspAI-Eco72I DNA fragment of 3.18 kb from pBfiIRM14 was subcloned into Eco32I-cleaved pACYC184. The resulting recombinant plasmid was digested with Acc65I-Cfr42I and the DNA fragment of 3.98 kb was ligated into NotI-digested pBRmTn10dKm-NotI vector (cohesive DNA ends produced by Acc65I, Cfr42 and NotI were blunted).

Construction of pUCBfiI plasmid. The DNA fragment, encoding the N-terminal fragment of *bfiIR* was amplified by PCR using two primers - Bfi-N (which has the optimized Shine-Dalgarno (SD) sequence and the translation start codon ATG which replaced TTG present in the wild-type gene) and Bfi130. DNA of plasmid pBfiIRM14 was used as a template. The PCR fragment was cloned into Eco32I-digested pUC57 vector and verified by dideoxy sequencing (Sanger et al., 1977). The obtained construct was digested with XbaI-Eco130I and the full-length *bfiIR* gene was reconstituted by ligating the resulting DNA fragment of 2.9 kb with the Eco130I-Van91I DNA fragment of 0.93 kb which was cleaved out from pBfiIRM14. XbaI and Van91I-generated cohesive DNA ends were blunted using T4 DNA polymerase.

Construction of pBfiIN¹⁻¹⁸⁵, pBfiIN¹⁻¹⁹² ir pBfiIN¹⁻²⁰⁴ plasmids. Three DNA fragments encoding N-terminal regions of BfiI of various sizes - BfiIN¹⁻¹⁸⁵, BfiIN¹⁻¹⁹² and BfiIN¹⁻²⁰⁴ were amplified using the same 5'-primer Bfi-N (containing SD sequence and start codon) and three different 3'-primers - BfiIN1-185, BfiIN1-192 and BfiIN1-204 for each mutant, respectively. DNA of pBfiIRM14 was used as a template. PCR fragments were cloned into Acc65I-digested, blunted pZA34-MCS-1 vector and sequence was verified by dideoxy sequencing.

Construction of pBfiIC¹⁸⁶⁻³⁵⁸, pBfiIC¹⁹³⁻³⁵⁸ bei pBfiIC²⁰⁵⁻³⁵⁸ plasmids. Three DNA fragments encoding C-terminal regions of BfiI of various size – BfiIC¹⁸⁶⁻³⁵⁸, BfiIC¹⁹³⁻³⁵⁸ and BfiIC²⁰⁵⁻³⁵⁸ were amplified using the same 3'-primer Bfi-C (which is complementary to the 3'-end of *bfiIR*), and three different 5'-primers - BfiIC186-358, BfiIC193-358 and BfiIC205-358 for each mutant, respectively. DNA of pBfiIRM14 was used as a template. PCR fragments were cloned into Eco32I-digested pUC57 vector and verified by dideoxy sequencing.

1.5.3. Construction of RR1-FokIM and RR1-BfiIM strains

RR1-FokIM and RR1-BfiIM are derivatives of *Escherichia coli* RR1 strain producing F' factor-encoded FokI and BfiI methyltransferases, respectively. The strains were constructed using the following scheme. The genes encoding FokI and BfiI methyltransferases were inserted into the Tn10 derivative, so-called 'mini-kan' (Km^r) transposon (Way et al., 1984) resulting in plasmids pTn-FokIM and pTn-BfiIM (see **Construction of recombinant plasmids**). Strains XL1-Blue [pTn-FokIM] and XL1-Blue [pTn-BfiIM] were grown to mid-log phase, and then transcription of the transposase gene (which is located outside the transposon under the control of P_{tac}) was induced by adding of IPTG to a final 1 mM concentration. The strains were allowed to grow overnight and after that F' plasmid conjugative DNA transfer into recipient strain RR1 (F⁻ Sm^r) was carried out in order to select for those F' variants which acquired a

copy of modified 'mini-kan' transposon. The liquid mating assay protocol was as follows. Overnight culture of the donor strain grown in the presence of appropriate antibiotics was diluted 1:50 into fresh antibiotic-free LB and allowed to grow to $OD_{600} \sim 0.1$ at 37 °C. Then 200 μ l of the donor culture were mixed with 10 μ l of overnight culture of recipient RR1 and incubated at 37°C. After 30 min, 10-fold serial dilutions of mating mixture were prepared using LB and plated onto LB-agar containing streptomycin, tetracycline and kanamycin to counter-select donor strain and unmated recipient cells while selecting for transconjugants which contain F' plasmid (Tc^r) with the inserted transposon (Km^r). A few resulting transconjugants carrying the expected F':mTn10 (*fokIM*⁺) or, respectively, F':mTn10 (*bfiIMC1*⁺ *bfiIMC2*⁺) were colony-purified. In order to test the ability to protect host's DNA from restriction endonuclease activity, individual colonies were transformed either with pUCFokI or pUCBfiI which express FokI or BfiI restriction endonuclease, respectively. One strain from every experiment which tolerated the introduction of either pUCFokI or pUCBfiI was called RR1-FokIM or RR1-BfiIM, respectively.

1.5.4. Random insertional mutagenesis

The random insertional mutagenesis of genes for FokI and BfiI restriction enzymes was done using the technique similar to that described by Biondi et al. (Biondi et al., 1998). Fifty micrograms of DNA under investigation (pUCFokI or pUCBfiI) and the same amount of ethidium bromide were diluted in 1 ml of G⁺ buffer (Fermentas). In a control reaction, fifty microliters of this solution were incubated with 100 μ g of DNase I. Aliquots of control reaction were removed at different time points and analyzed in order to evaluate the kinetics of appearance of open circular plasmids. Incubation for 5 minutes at 30°C resulted in 90% of nicked DNA. Therefore, 1.9 ng of DNase I was added to remaining DNA solution and reaction was terminated after 5 min of incubation at 30°C by adding EDTA to a final concentration of 50 mM. Nicked circular plasmids were separated from other forms by electrophoresis on 0.8% agarose gel. Purified DNAs were then treated with S1 nuclease in S1 buffer (Fermentas) in order to cleave the second DNA strand at the position of nick. Incubation time with the S1 nuclease was chosen such that no more than 40% of nicked DNA was transformed to linear form in order to avoid degradation of ends of linear DNA. Resulting randomly linearized plasmids were gel-purified, dephosphorylated using CIAP and ligated with the specially designed DNA cassette which codes for chloramphenicol resistance and which was excised from pUCm with BpiI and blunted using T4 DNA polymerase. Electrocompetent RR1-FokIM and RR1-BfiIM cells were transformed with pUCFokI and pUCBfiI ligation mixtures, respectively. Transformants were selected on LB-agar supplemented with Ap and Cm. More than 10⁵ recombinant clones were obtained in both cases. Plasmid DNA isolated from the pooled transformants was used in enrichment procedure in order to remove these plasmids in which the cassette is inserted outside the restriction endonuclease gene. In pUCFokI, Eco81I and PstI have unique targets which are located on both sides of *fokIR* and are separated by 1.86 kb. Therefore, total plasmid DNA of pUCFokI insertional mutants was cleaved with Eco81I-PstI and then a DNA fragment of 2.56 kb (the size of this fragment is a sum of Eco81I-PstI fragment of 1.86 kb and the DNA cassette of 0.7 kb) was gel-purified and ligated with the larger (2.69 kb) Eco81I-PstI fragment of pUCFokI which represents the vector backbone. Ligation mixture was

introduced into RR1-FokIM by electroporation. Plasmid DNA was isolated from the pooled Ap^rCm^r transformants and used in subsequent experiments aimed to find out endonucleolytically active FokI mutants. The same approach was used to enrich the library of pUCBfiI. However, in the latter case BamHI and Mph1103I were used to cleave out the *bfiIR* gene carrying the inserted cassette. Resulting fragment of 1.85 kb was inserted into BamHI-Mph1103I cleaved vector backbone of 2.7 kb. Ligation mixture was transformed back into RR1-BfiIM strain and total plasmid DNA was isolated from resulting Ap^rCm^r transformants. This collection was used in all subsequent screening procedures.

1.5.5. Transformation of *Escherichia coli*

Transformations of *E. coli* were carried out using the CaCl₂-heat shock method (Ausubel et al., 1992) or electrotransformation using Gene Pulser (Bio-Rad). DNA in ligation mixture was precipitated with ethanol to remove salts and dissolved in deionized water before electrotransformation. Transformed cells were selected on LB medium supplemented where needed with the antibiotics (Ap – 0.1 mg/ml, Kn – 0.05 mg/ml, Cm – 0.02 mg/ml, Tc – 0.01 mg/ml). All *E. coli* strains were grown in LB medium at 37°C 16 h.

1.5.6. Determination of the endonuclease activities

The endonuclease activity *in vitro* was assayed by incubation of 1-3 µl of cell-free extracts (prepared using technique described by Whitehead and Brown (Whitehead and Brown, 1985)) or purified protein with 1 µg of λ DNA at 37°C for 1 h in a 40 µl reaction volume containing 33 mM Tris-CH₃COO, pH 7.9 (at 37°C) 10 mM Mg(CH₃COO)₂ 2.66 mM KCH₃COO, and 0.1 mg/ml BSA. Products of reaction were analyzed by electrophoresis in 0.8% agarose gels.

1.5.7. Western blotting

Protein samples were separated by SDS-PAGE. Then proteins were electroblotted onto PVDF membrane (Macherey-Nagel) in 12 mM Tris-base, 96 mM glycine and 20% methanol (pH 8.3), using Semi-dry Blotting Unit V10-SDB (Fermentas) at 0.8 mA/cm² for 30 minutes. Further incubations were performed on a shaker at low speed in room temperature. The membranes were blocked for one hour in TBST buffer (0.15 M NaCl, 50 mM Tris-HCl, pH 7.4, 0.1% Tween-20) containing 1% skim milk, incubated with primary antibodies (mouse polyclonal anti-BfiI diluted 1:1000 in blocking solution) for one hour and washed three times for 10 minutes with TBST. Then membranes were incubated with secondary antibodies (goat anti-mouse linked to alkaline phosphatase (Quattromed) diluted 1:5000 in TBST with 1% skim milk) for 30 min and washed three times for 5 min in TBST. Immunoreactive bands were detected by incubating membranes in 10 ml of alkaline phosphatase buffer (0.1 M Tris-HCl, pH 9.5, 0.1 M NaCl, 5 mM MgCl₂) containing 0.33 mg/ml nitro-blue-tetrazolium and 0.165 mg/ml 5-bromo-4-chloro-3-indolyl-phosphate.

1.5.8. DNA binding studies

DNA binding by wild-type BfiI and by mutant BfiI¹⁹³⁻³⁵⁸ was assayed using the gel-mobility-shift-assay. Two DNA oligoduplexes were used for DNA binding studies. The specific oligoduplex contains BfiI target and was obtained by annealing of oligonucleotides Bfisp⁺ and Bfisp⁻. Phosphorothioate substitutions at the BfiI cleavage positions were introduced in order to avoid cleavage of oligoduplex by wt BfiI. The nonspecific oligoduplex was obtained by annealing of oligonucleotides Bfinsp⁺ and Bfinsp⁻. It differs from the specific one by a single base pair within the BfiI target:

```
Bfisp+/- 5'-CAGCACTGGGGCCAGsATGGT
          GTCGTGACCCCGGTsCTACCA
```

```
Bfinsp+/- 5'-CAGCACAGGGGCCAGsATGGT
          GTCGTGTCCCCCGGTsCTACCA
```

(BfiI recognition sequence is boxed; phosphorothioate substitutions at the BfiI cleavage site are denoted as “s”). The upper strand of both oligoduplexes was radioactively labeled by phosphorylation of 5'-hydroxyl group, using [γ -³³P] ATP. In DNA binding experiments, either specific or nonspecific DNA oligoduplexes at 1 nM concentration were incubated for 20 min at 37°C with increasing amount of protein under investigation in 20 μ l of 33 mM Tris-CH₃COO (pH 7.9 at 37°C), 66 mM KCH₃COO, 0.1mg/ml BSA and 10 % (v/v) glycerol. Samples were loaded onto 8 % polyacrylamide gels (29:1 acrylamide / bisacrylamide) and run in TAE buffer (40 mM Tris-CH₃COO, pH 7.9, 1 mM EDTA) for 2 h at ~ 7 V/cm. After electrophoresis, radioactive bands were visualised by CycloneTM Storage Phosphor System and analyzed by using OptiQuantTM Image Analysis Software from Packard Instruments. The K_d values for cognate oligonucleotide binding by BfiI and by the mutant BfiI¹⁹³⁻³⁵⁸ were calculated by fitting binding data to:

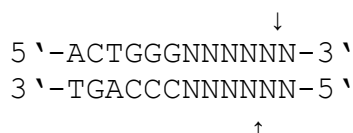
$$y = (s_0 - x - K_d + ((s_0 + x + K_d)^2 - 4s_0x)^{0.5})/2$$

(where y is the free DNA concentration (in terms of nM) at each protein concentration x , s_0 is the DNA concentration in binding mixture, and K_d is the dissociation constant of protein–DNA complex. Data were analysed using the KYPLOT 2.0 software (Yoshioka, 2002).

2. Results and discussion

2.1. Analysis of genetic organization of BfiI restriction–modification system

BfiI restriction endonuclease, isolated from *Bacillus firmus* S8120 strain, is a member of the type IIS restriction enzymes (Vitkute et al., 1998). The enzyme recognizes non-palindromic nucleotide sequence 5'-ACTGGG and cleaves complementary DNA strands 5 and 4 nucleotides away from the recognition sequence:



After the first biochemical assays performed with BfiI it was noted that the enzyme, in contrast to all restriction endonucleases known to date, cleaves DNA in the absence of Mg^{2+} ions (Vitkute and Petrusyte, unpublished). In order to investigate the organization and mechanism of DNA recognition and catalysis employed by this unusual restriction endonuclease, the genes of the BfiI restriction-modification system had to be cloned first.

2.1.1. Cloning of BfiI restriction-modification genes

The genes coding for the BfiI R-M system were cloned using a strategy based on the selection of self-modifying recombinant plasmids that become resistant to the restriction enzyme cleavage *in vitro* (Szomolanyi et al., 1980). The gene library of *Bacillus firmus* S8120 was constructed by partially digesting genomic DNA with Bsp143I and ligating the fragments into BamHI-cleaved and dephosphorylated pBR322. After the second round of BfiI digestion of plasmid population from partial Bsp143I library (300,000 clones) and retransformation of surviving plasmids into *E. coli* cells, several thousand transformants were obtained. Individual plasmids were isolated from 18 randomly picked transformants. Of these, 17 were BfiI-resistant, suggesting that they contain the cloned gene for BfiI methyltransferase. Restriction mapping of resistant plasmids identified six different types. Analysis of the crude cell lysates prepared from representatives selected from each group revealed that three of them exhibit BfiI REase activity. Among these, pBfiIRM14 contained the shortest cloned DNA fragment (5.6 kb) and was selected for further analysis.

2.1.2. Deletion analysis of BfiI R-M system

To determine the location of BfiI R-M genes on pBfiIRM14 plasmid, a set of deletion mutants was constructed and tested for the methylation and restriction phenotypes. First, restriction map of the cloned region was established (see Fig. 1A), followed by deletion analysis designed according to the restriction map. The opposite ends of the cloned region were named KspAI and Kpn2I according to locations of unique restriction sites of KspAI and Kpn2I REases in the region.

Kpn2I-terminal deletions (Fig. 1B) were obtained by cleaving pBfiIRM14 plasmid with following REases:

1. Kpn2I;
2. Kpn2I-MunI;
3. Kpn2I-Eco72I;
4. NdeI;
5. Kpn2I-NcoI;
6. Kpn2I-Bpu1102I.

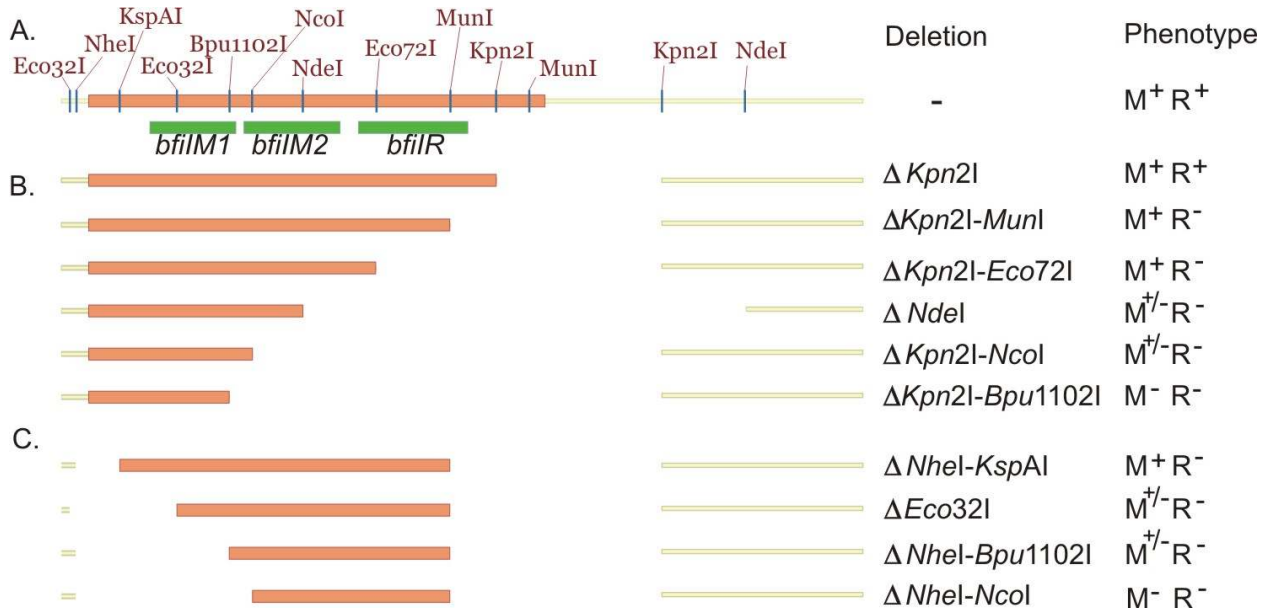


Fig.1. Deletion mapping of genes encoding the BfiI R-M system. (A) Restriction map of pBfiIRM14 plasmid. Cloned region is marked in red, approximate locations of BfiI R-M system genes are marked in green. (B) Kpn2I-terminal deletions and their phenotypes. (C) KspAI-terminal deletions and their phenotypes. R⁺ and R⁻ indicate sublones exhibiting and lacking restriction endonuclease phenotype, respectively. Methyltransferase protection against digestion by BfiI restriction enzyme is indicated as follows: M⁺ - full protection, M^{+/}- - partial protection, M⁻ - no protection.

Plasmid carrying the $\Delta Kpn2I$ deletion was the smallest that conferred restriction- and methylation-positive phenotypes. The absence of restriction endonuclease activity in cells carrying plasmids with $\Delta Kpn2I-MunI$ or $\Delta Kpn2I-Eco72I$ deletions alongside the unaltered modification phenotype suggests that MunI and Eco72I cleavage sites are situated within the regulatory or structural part of the BfiI REase gene. Plasmids carrying $\Delta NdeI$ and $\Delta Kpn2I-NcoI$ deletions were only partially protected from digestion by BfiI REase. This could mean that either these deletions disrupt regulatory sites of BfiI methyltransferase gene, or, more likely, that BfiI R-M system, like many other type IIS systems, has two methyltransferases. Two methyltransferases (either separate or genetically fused) are necessary to recognize and modify bases on opposite strands of the asymmetric target, recognized by type IIS restriction endonuclease (Lubys et al., 1996; Gunn and Stein, 1997). Deletion variant $\Delta Kpn2I-Bpu1102I$ had no BfiI-protection which could possibly occur due to destruction of the second BfiI methyltransferase gene.

Results of the analysis of KspAI-terminal deletions agreed well with the idea of presence of two independently acting methyltransferases. In order to avoid toxicity of the BfiI REase, these deletions were constructed using the deletion derivative $\Delta Kpn2I-MunI$ which lacked the endonuclease activity. KspAI-terminal deletions (Fig. 1C) were obtained by cleaving DNA of the aforementioned plasmid with following REases:

1. NheI-KspAI;
2. Eco32I;
3. NheI-Bpu1102I;
4. NheI-NcoI.

DNA of plasmid with the deletion Δ NheI-KspAI was completely protected from BfiI cleavage. Having in mind that Δ Kpn2I deletion does not disrupt any gene, it was concluded that the BfiI R-M system is located within the 4.7 kb KspAI-Kpn2I fragment. Plasmids carrying deletions Δ Eco32I and Δ NheI-Bpu1102I were only partially BfiI-protected similar to above-mentioned plasmids with Δ NdeI and Δ Kpn2I-NcoI deletions. Deletion variants Δ Kpn2I-NcoI and Δ NheI-Bpu1102I carried DNA inserts from *B. firmus* that overlapped by just 300 nucleotides. Obviously, such region is too short to encode a full-length methyltransferase. Based on these observations it was concluded that methylation of the BfiI recognition sequence is most probably accomplished by two methyltransferases, and a possible boundary between the two genes is located within the Bpu1102I-NcoI fragment.

2.1.3. Nucleotide sequence analysis of BfiI restriction–modification system

Deletion analysis revealed that the cloned KspAI-Kpn2I DNA fragment harbors all genes of the BfiI restriction modification system. Sequencing of the KspAI-Kpn2I fragment was done in both directions using series of nested deletions generated by Bal31 nuclease (Sambrook et al., 1989) (EMBL accession no. AJ290970). Three large open reading frames (ORFs) were identified within the sequenced region of 4662 bp (Fig. 1). They were in perfect agreement with the results of the deletion mapping, indicating that genes *bfiM1* and *bfiM2* code for two independent BfiI methyltransferases, M1.BfiI and M2.BfiI, respectively, whereas the third one, *bfiIR*, codes for the BfiI restriction enzyme. Putative promoter sequences were found upstream of all three genes.

In the ORF assigned for the first MTase gene, *bfiM1*, alternative translational start codons ATG appeared at nucleotide positions 294 and 309 and the termination codon at nt position 1434. Putative ribosome binding sites AGGA (nt 286–289) and AGGTG (nt 299–303) were detected upstream of both ATG sequences, respectively. Therefore, the position of a translational start codon of *bfiM1* could not be unambiguously determined based on the nucleotide sequence analysis alone. The ORF for the second MTase, *bfiM2*, started 21 bp downstream of the *bfiM1* termination codon, ended at nt position 2969 and encoded a protein of 504 amino acid residues with a calculated mass of 57.9 kDa. The *bfiM2* gene was preceded by a strong putative ribosome binding site AAGGAGGT (nt 1444-1451). Analysis of the deduced protein sequences revealed that both M1.BfiI and M2.BfiI contain conserved motifs F-G-G and TSPPY, typical for N4-methylcytosine methyltransferases (data not shown). The mutual location of the conserved motifs suggests that both methyltransferases could be assigned to the S₁₂ class of N-methyltransferases (Timinskas et al., 1995). Pairwise comparison of protein sequences of BfiI methyltransferases demonstrated only marginal resemblance (21% identity). The alignment of the predicted protein sequences of BfiI MTases to the protein sequences of other N4-methylcytosine methyltransferases belonging to the S₁₂ class revealed similar levels of homology. The most similar proteins to BfiI methyltransferases were MTases from other R-M system BtsI. M1.BfiI has 26% identical residues with M1.BtsI and M2.BfiI has 25% identical residues with M2.BtsI.

Similar to BfiI enzymes, BtsI enzymes recognizes a nucleotide sequence that is asymmetric - 5'-GCAGTG (Roberts et al., 2007).

Presumption that both BfiI methyltransferases modify cytosine was confirmed by sodium bisulfite treatment assay. These experiments were conducted by G. Vilkaitis in Laboratory of Biological DNA Modification at the Institute of Biotechnology. It was shown that M1.BfiI modifies the second C within the strand 5'-CCCAGT, whereas M2.BfiI methylates the unique C of the 5'-ACTGGG strand.

The *bfiIR* gene is located on the DNA strand opposite to that bearing the genes for cognate methyltransferases. *bfiIR* ends with the termination codon TAA at nt 3018. Two possible translation start codons TTG at nt 4092 and ATG at nt 4029 were identified. Despite the presence of putative Shine-Dalgarno sequences upstream of both initiation codons (TAAGGGGG is located 7 bp upstream of TTG, and AAGGA is 9 bp upstream of ATG), the TTG was assumed to be the initiation codon. This assumption relies on the experimental observation that cloning and expression of an amplified DNA fragment starting at the TTG but not at ATG codon yielded an active BfiI enzyme (data not shown). The ORF starting at TTG encoded a protein of 358 amino acid residues with a calculated mass of 40 kDa.

Alignment of the BfiI REase amino acid sequence with sequences available in protein databases was performed in Laboratory of Protein-DNA Interactions at the Institute of Biotechnology. Analysis revealed that the N-terminal part of the BfiI REase exhibits marginal but significant similarities to the EDTA-resistant Nuc nuclease (Fig. 2), a member of the phospholipase D superfamily (Ponting and Kerr, 1996). Similarly to BfiI REase, Nuc nuclease does not use Mg^{2+} ions for DNA cleavage. It was noted that amino acid residues which formed an active site in Nuc nuclease (His94, Lys96, Ser109 and Asn111) had amino acid analogs in BfiI REase (His105, Lys107, Ser123 and Asn124). Site-specific mutagenesis of these amino acids confirmed assumption that they could form the active site of BfiI REase which resembles that of Nuc nuclease (Lagunavicius et al., 2003).

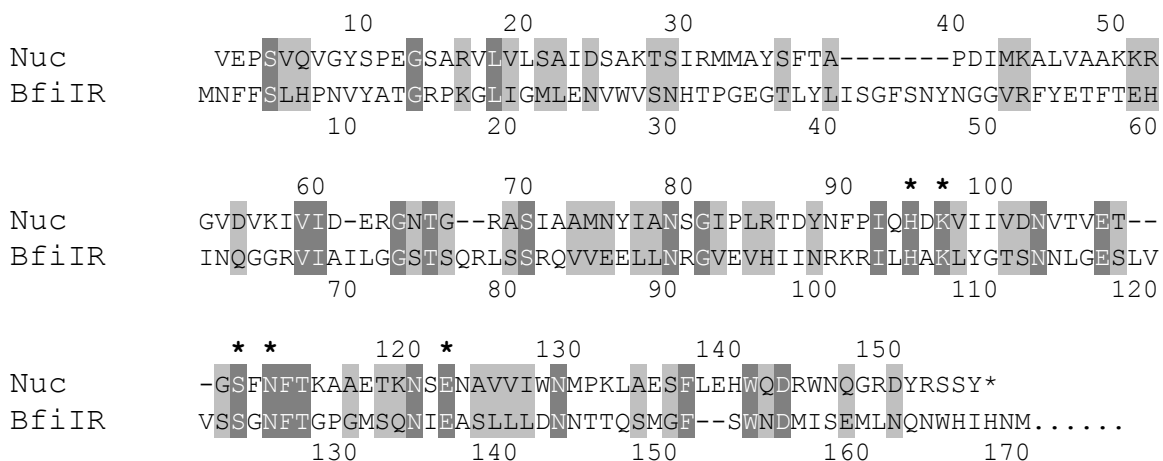


Fig. 2. Comparison of amino acid sequences of EDTA-resistant nuclease Nuc of *S. typhimurium* and N-terminal part of BfiI (residues 1–170) restriction enzyme. White letters on the dark gray background indicate identical residues; black letters on the light gray background indicate similar residues. Amino acid residues forming the active site of Nuc are marked by asterisk.

BfiI	MNFFSLHPNVYATGRPKGLIGMLENVVWSNHTPGEGLYLIISGFSNYNGCVRFYETFFTEH	60
BmrI	MNYFSLHPNVYATGRPKGLINMLESVWVISNPKPGDGTMYLISGFANYNGCIRFYETFFTEH	60
		* *
BfiI	INQGGRVIAAILGGSTSQRLLSSRQVVEELLNRGVEVHIINRKRI LHAKLYGTSNNL GESLV	120
BmrI	INHGGKVIAAILGGSTSQRLLSSKQVVAELVSRGVDVYIINRKRI LHAKLYGSSNS GESLV	120
		* * *
BfiI	VSSGNFTGPGMSQNIIEASLLLDNNTTQSMGFSWNDMISEMLNQNWHIHNMTNATDASP GW	180
BmrI	VSSGNFTGPGMSQNVVEASLLLDNNTTSSMGFSWNGMVNSMLDQKWQIHNLSNSNPTSPSW	180
BfiI	NLLYDERTTNLTLDDETERVTLIVTLGHADTARIQAAPGTTAGQGTQYFWLSKDSYDFFPP	240
BmrI	NLLYDERTTNLTLDDETQKVTLIILTLGHADTARIQAAPKSKAGEGSQYFWLSKDSYDFFPP	240
BfiI	LTIRNRRGTKATYSSLINMNYIDINYTDTCRVTFEAENNFDFRLGTGKLRYSVAKSND	300
BmrI	LTIRNKRGTATYSSCLINMNYLDIKYIDSECRVTFEAENNFDFRLGTGKLRYSNVAASDD	300
BfiI	IAAITRVGDSYELRIKQGTPEHSQLDPYAVSFIGNRGKRFYISNEEFGRICVTF	358
BmrI	IAAITRVGDSYELRIKKGSSNYDALSAAVNFIGNRGKRYGYIPNDEFGRICAKE	358

Fig. 3. Comparison of amino acid sequences of BfiI and BmrI restriction enzymes. White letters on the dark gray background indicate identical residues; black letters on the light gray background indicate similar residues. Amino acid residues forming the active site of BfiI are marked by asterisk.

Very recently a new REase, BmrI, was found, which recognizes the same sequence and cuts DNA at the same position as BfiI (Chan et al., 2007). Another relatedness of these two restriction endonucleases – they do not require Mg^{2+} ion for DNA hydrolysis. Sequence alignment revealed high level of homology (79% identical amino acid residues). Moreover, the amino acid residues of the BfiI active site (His105, Lys107, Ser123, Asn124 and Glu136) have counterparts in corresponding positions of BmrI (see Fig. 3). Thus BmrI, like Nuc and BfiI, is a member of phospholipase D superfamily.

2.2. Random Gene Dissection technique and its use for investigation of protein domain organization

As mentioned above BfiI REase is a type IIS restriction endonuclease that recognizes asymmetric sequence and cleaves DNA outside of the target sequence. The enzyme hydrolyses DNA in the absence of Mg^{2+} ions. Protein sequence alignment followed by site-specific mutagenesis revealed that BfiI is a member of the phospholipase D superfamily. Amino acid residues of the active site are located in the N-terminal part of protein, suggesting that this part of the protein might form catalytic domain. A new approach, Random Gene Dissection (RGD), was proposed for analysis of domain organization of BfiI REase. In order to prove its suitability for this application, the method was first applied on FokI REase – a well-studied two-domain protein with known tertiary structure (Li et al., 1992; Wah et al., 1997).

2.2.1. Principles of Random Gene Dissection method

It is generally assumed that multidomain proteins have evolved from fusion of two or more ancestral single-domain proteins or as a result of recombination events such as domain duplication or swapping (Bjorklund et al., 2005). The type IIS REases comprised of two domains might serve as examples of this kind of evolution - supposedly they originated by fusion of DNA recognition and DNA cleavage domains. Therefore it is possible, that after separation from each other the two domains would retain their function. For example, in case of FokI and MnlI functional target recognition and catalytic domains were obtained after their separation by limited proteolysis (Li et al., 1992; Kriukiene, 2006). In this work the Random Gene Dissection method is proposed as a novel, presumably more convenient way to separate two domains, where actual separation of domains is performed at the genetic level.

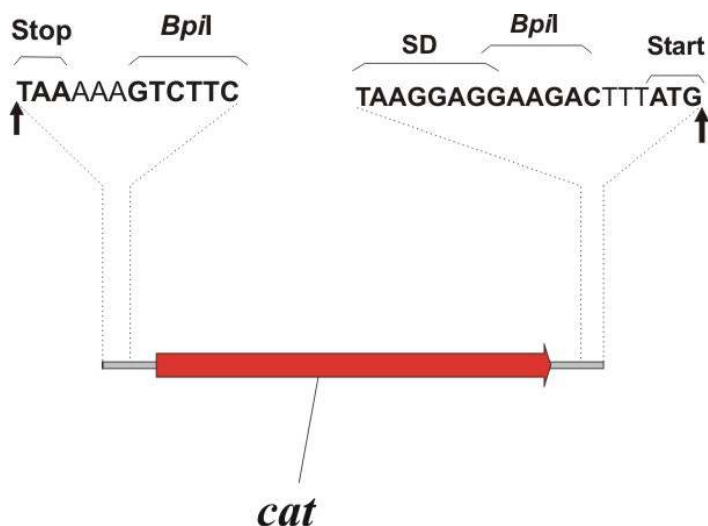


Fig. 4. The structure of the cassette used for Random Gene Dissection. The locations of the translation termination codon (Stop), the Shine-Dalgarno sequence (SD), and translation initiation (Start) codon are shown.

The RGD technique is based on the random insertion of specially designed DNA cassette (Fig. 4) into the gene of interest. This cassette, when inserted into the reading frame of a target gene, results in the premature termination of translation of its messenger RNA (mRNA) due to the presence of a TAA stop codon at the 5'-end of the cassette. As the 3'-end of the cassette contains a ribosome binding site followed by the initiation codon ATG, the remainder of the target gene is translated as a separate polypeptide. Such a construct encodes pieces of the target protein without any additional amino acids except for the initiator formyl-methionine.

The presence of the *cat* gene coding for chloramphenicol resistance in the cassette facilitates selection for insertional mutants. Because the *cat* gene has no promoter of its own it can be transcribed only if the cassette is correctly inserted into a transcriptionally active DNA region. In its both ends the DNA cassette harbours two targets of BpiI REase. They are used to cut the DNA cassette from the vector plasmid. BpiI targets are designed in the way that after cleavage by BpiI REase, followed by blunting with T4 DNA polymerase, the DNA cassette retains all necessary genetic elements (SD sequence, start and stop codons).

When such DNA cassette is inserted into interdomain region-encoding part of the gene, the two domains will be synthesized independently from each other. As most of individual domains possess autonomous structure and are associated with a particular function, which can be performed even in the absence of adjacent parts of a protein, separate domains can remain functional. In order to isolate catalytically active insertional mutants, the genetic screening approach developed by Heitman et al. can be applied

(Heitman et al., 1989). The DNA damage is the primary reason for the induced SOS response which is responsible for DNA repair in the cell, thus, the phenotype of SOS induction has a potential to be applied in all experiments where intracellular endonucleolytic activity has to be identified. SOS response can be visualized by using *Escherichia coli* ER1992 strain in which the SOS regulon gene *dinD1* is fused to the reporter gene *lacZ* coding for β -galactosidase (Fomenkov et al., 1994). The development of blue colour in these cells, when plated on media containing 5-bromo-4-chloro-3-indolyl- β -d-galactopyranoside (X-gal), is a convenient evaluation of the SOS response. One can expect that if endonucleolytic domain of certain type IIS REase remains functional when separated from a target recognition domain, it will cleave chromosomal DNA and thus cause SOS response. As a result, *E. coli* cells containing REase mutants with such activity will form blue colonies on media with X-gal.

2.2.2. Random dissection of the model gene encoding FokI REase

Restriction endonuclease FokI, a two-domain protein with known tertiary structure (Wah et al., 1997; Wah et al., 1998) was used as a model in experiments aimed to test the feasibility of RGD technique. The methodology used for random insertion of blunt-ended DNA cassette into DNA of pUCFokI, the plasmid which contains the gene for FokI restriction endonuclease, was essentially the same as described (Biondi et al., 1998) (see Fig. 5). Single-stranded nicks at random positions were introduced into supercoiled pUCFokI by incubating plasmid DNA with DNase I in presence of Mg^{2+} ions and ethidium bromide. In the next step nicks were converted into double-stranded breaks by the use of S1 nuclease. The resulting linear molecules of near to full-length size were then ligated with the DNA cassette and introduced into cells of *E. coli* RR1-FokIM strain by electro-transformation, yielding 10^5 Cm-resistant colonies. Noteworthy, the size of individual transformants varied, indicating either the toxicity of some fraction of mutants or differences in individual levels of chloramphenicol resistance that most likely depended on the position of inserted cassette. Restriction mapping of the total plasmid DNA isolated from pooled transformants has indicated that half of plasmids contained the DNA cassette inserted outside of restriction endonuclease gene. These were eliminated by

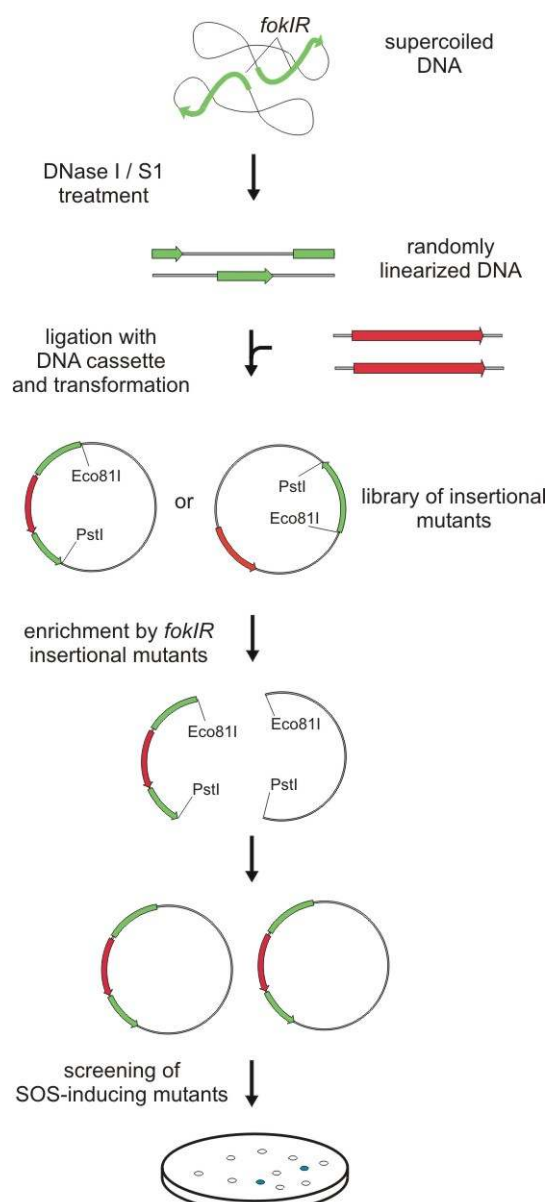


Fig. 5. Construction of random insertions.

subcloning of the DNA fragment, which corresponded to the *fokIR* gene with inserted cassette, back into the appropriately cleaved pUCFokI and selecting for Cm-resistant transformants (for details see Materials and Methods).

2.2.3. Screening and analysis of endonucleolytically active FokI mutants.

The genetic screening strategy by using ER1992 (Fomenkov et al. 1994), an *Escherichia coli* strain that carries *dinDI::lacZ* fusion was used to search for those *fokIR* mutants which induce SOS response but do not kill the host. As mentioned above cells expressing *dinDI::lacZ* can be easily visualized by their blue colour in the presence of X-gal, the substrate for *lacZ*-encoded β -galactosidase. Screening of *fokIR* insertional mutants that induce SOS response in the absence of cognate methylation was carried out using the strain ER1992. Transformation of ER1992 with the library of *fokIR* insertional mutants resulted in very high fraction of blue colonies (approximately 30 % of colonies formed). Plasmids from 20 dark-blue colonies representing a limited fraction of hundreds of positives were isolated and transformed back into ER1992 and also into ER1992 [pAC-FokIM] in order to confirm that introduction of selected plasmids results in SOS response and whether they code for endonuclease with FokI specificity, respectively. We found that all positives tested induce SOS response in ER1992 but not in ER1992 [pAC-FokIM], indicating that insertional mutants retained their wild-type specificity and could not cleave genomic DNA methylated by M.FokI.

To determine the exact location of the DNA cassette insertion, boundaries of N- and C-terminal fragments encoded by insertional mutants of all 20 clones were sequenced (Table 1). Insertion points were dispersed in the N-terminal part which covers amino acids 7-99 of FokI. The extent of nucleotide deletions varied from 0 to 14 bp in 70 % of the cases (14/20); in the remaining mutants, deletions were in the range from 17 to 39 bp. These values are nearly identical to those that were obtained in other experiment in which random insertions were also generated using DNase I/S1 treatment (Biondi et al., 1998). The results of analysis of FokI N- and C-terminal fragments encoded by selected mutants were somewhat unexpected as just one mutant (FWTS28) was found to contain N- and C-terminal fragments, which were in frame with the cassette-donated translation termination codon TAA and translation initiation codon ATG, respectively, and thus had a potential to produce a protein of nearly wild-type structure. Five out of 19 remaining clones (FWTS13, FWTS34, FWTS12, FWTS35 and FWTS16) contained the C-terminal fragment starting with the cassette-donated ATG, however, the reading frame of their N-terminal fragments did not coincide with the TAA codon of the cassette and thus contained extra amino acids. In contrast, the reading frame of N-terminal fragment in FWTS26, FWTS29 and FWTS22 was stopped at TAA codon from the cassette, however, C-terminal fragments in these mutants were out of frame of the cassette-donated ATG and thus resulted in fragments of unpredictable size. Finally, 11 remaining mutants were out of frame at both ends and for that reason their N-terminal fragments had additional amino acids whereas C-terminal fragments were of uncertain size. Nevertheless, all these mutants induced SOS response in the absence but not in the presence of FokI methylation, indicating that functionally active FokI enzyme of wild-type specificity was produced in each case. Moreover, all mutants were analyzed by deletion analysis, which showed that both N- and C-terminal fragments were required for

Table 1. The structure of FokI mutants of wild-type specificity. Numbers in columns refer to the first and the last amino acid residue of the respective oligopeptide relative to the amino acid sequence of the wild-type FokI.

Number of mutant	N-terminal polypeptide*	C-terminal polypeptide*
FWTS28	1-7	9-579
FWTS13	1-7 (+1)	9-579
FWTS23	1-28 (+2)	39-579 (+1)
FWTS34	1-35 (+1)	48-579
FWTS12	1-38 (+2)	50-579
FWTS9, FWTS19	1-42 (+2)	45-579 (+1)
FWTS35	1-44 (+2)	54-579
FWTS24	1-45 (+2)	60-579 (+2)
FWTS26	1-46	51-579 (+2)
FWTS1	1-47 (+2)	51-579 (+2)
FWTS29	1-49	51-579 (+2)
FWTS22	1-51	54-579 (+2)
FWTS3	1-54 (+2)	56-579 (+2)
FWTS25	1-57 (+2)	60-579 (+2)
FWTS16	1-60 (+1)	67-579
FWTS32	1-72 (+2)	75-579 (+1)
FWTS30	1-74 (+2)	81-579 (+1)
FWTS31	1-95 (+2)	99-579 (+1)
FWTS4	1-95 (+2)	99-579 (+2)

* - the numbers in parentheses indicate the reading frame of *fokIR* in which the cassette-encoded translation initiation codon ATG or translation termination codon TAA was inserted.

activity as removal of any fragment resulted in disappearance of SOS-inducing properties. This finding indicates that the two fragments presumably complement each other *in vivo* and form functional FokI heterodimer. The regions in which cut-points occurred are mapped onto the crystal structure of FokI (Fig. 6).

The recognition domain of *FokI* is made of three smaller subdomains D1 (1-159 aa), D2 (160-302 aa) and D3 (303-372 aa) (Wah et al., 1997). We found that the DNA cassette disrupts the integrity of D1 in all selected mutants. In FokI-DNA complex, the subdomain D1 covers the DNA major groove, recognizing three 3'-terminal base pairs of the recognition sequence 5'-GGATG. D1-DNA interactions involve hydrogen bonds between Gln12, Asp13 (both are within the N-terminal arm), Arg79 (loop L1), Gln95 (loop L2) and DNA base pairs. Cut-points in 65% of analyzed mutants (13/20) occur in surface-located D1 helices $\alpha 2$ or $\alpha 3$; more importantly, in nearly all of these mutants, deletions of various sizes that destroy these secondary structure elements in putative heterodimers were found. Taken together with the observation that the residual FokI-specific activity is high enough to induce SOS response, it seems that $\alpha 2$ and $\alpha 3$

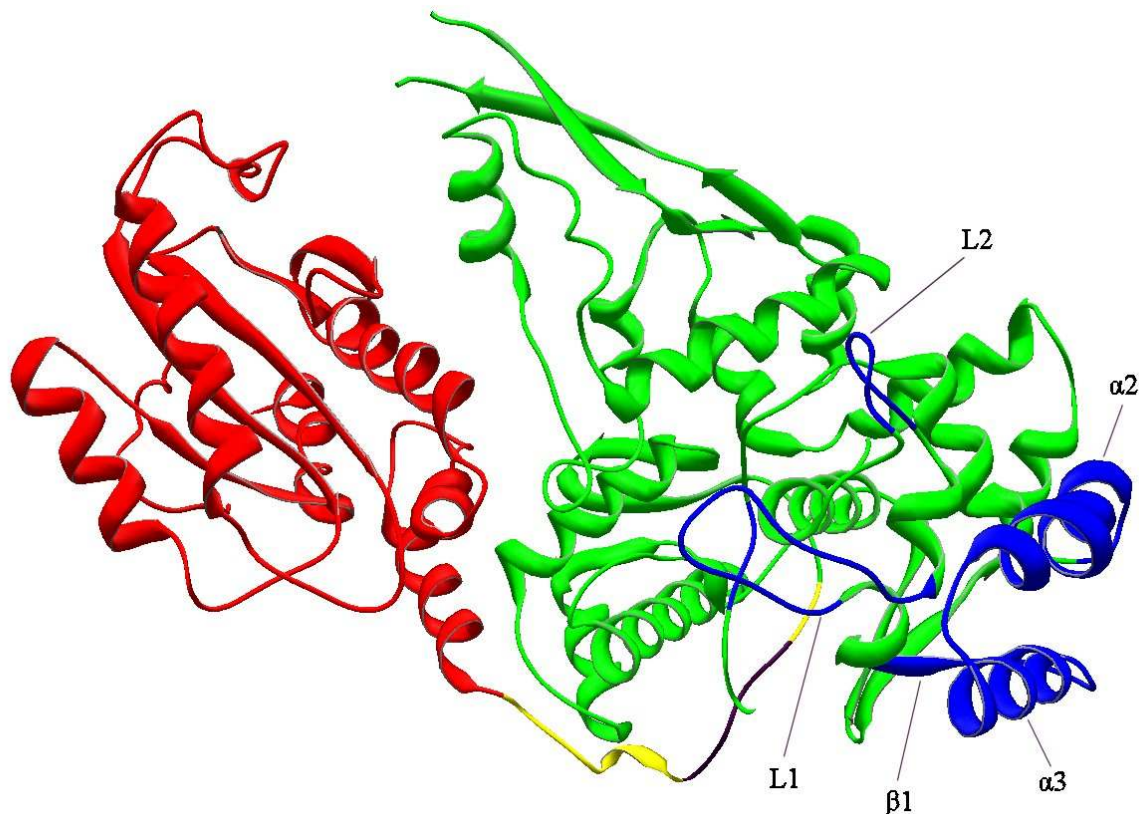


Fig 6. Insertion sites within the FokI monomer. Nucleolytic domain is in red, target recognition domain – in green, interdomain linker - in yellow, interdomain linker located by RGD is in black, sites where protein was dissected in two complement fragment is in blue.

helices are not crucial for both the correct heterodimer folding and target recognition. Interestingly, two mutants (FWTS28 and FWTS13) were dissected beyond first seven amino acid residues, and deletion analysis confirmed that these seven N-terminal amino acids are necessary for FokI functional activity *in vivo*. In one mutant (FWTS16) the β -strand $\beta 1$ (66-67 aa) which is also located on the protein surface is disrupted. Four remaining mutant genes encode FokI which is dissected either within the D1 loop L1 (FWTS30 and FWTS32) or within the loop L2 (mutants FWTS4 and FWTS31). All four mutants have extra amino acids at the end of their N-terminal fragments. In general, loops are secondary structure elements of choice for large in-frame insertions (Ostermeier et al., 1999). Surprisingly, residue Arg79 from loop L1 which makes base-specific contacts in FokI-DNA complex is absent in FWTS30. It was observed that L1 is disordered in free enzyme (Wah et al., 1998) and becomes ordered only in the presence of DNA (Wah et al., 1997). Taken together with the residual FokI-specific activity of mutant FWTS30 *in vivo* it might be proposed that the contact between Arg79 and DNA is not necessary for target recognition. It should be noted, however, that this statement is based solely on properties of the mutant FWTS30 in which synthesis of the C-terminal polypeptide starts at unknown position, and for that reason needs additional experiments.

In parallel experiment screening of *fokIR* insertional mutants that code for endonuclease of relaxed specificity and thus are able to induce SOS response in the presence of cognate methylation was carried out. The library of *fokIR* insertional mutants was introduced into ER1992 [pAC-FokIM], the derivative of ER1992 which produces

the FokI MTase and for that reason tolerates the presence of wild-type FokI restriction endonuclease. The cells were plated on LB-agar plates containing X-gal and a fraction of colonies that were blue (positives; about 2 % of the colonies formed) were examined further. Plasmids from 8 positives were isolated and retransformed back into the same strain to confirm that the blue colour of colonies was the result of expression of gene(s) located on the plasmid. Determination of the exact position of insertions was performed by sequencing (Table 2).

Table 2. The structure of FokI mutants of relaxed specificity. Numbers in columns refer to the first and the last amino acid residue of the respective oligopeptide relative to the amino acid sequence of the wild-type FokI.

Number of mutant	N-terminal polypeptide*	C-terminal polypeptide
FREL43	1-286 (+2)	380-579
FREL2, FREL14, FREL41, FREL42, FREL46	1-369 (+1)	378-579
FREL45	1-370 (+2)	375-579
FREL44	1-372 (+1)	375-579

* - the numbers in parentheses indicate the reading frame of *fokIR* in which the cassette-encoded translation translation termination codon TAA was inserted.

Previous FokI trypsin cleavage studies have demonstrated that all information necessary for sequence-specific recognition of DNA is encoded within the 41 kDa N-terminal fragment of FokI, which corresponds to amino acids 2-383, whereas the C-terminal fragment of 25 kDa (amino acids 384-579) cleaves nonspecifically both M.FokI-methylated and non-methylated DNA substrates in the presence of MgCl₂ (Li et al., 1992). The crystal structure of FokI restriction endonuclease, solved in the presence of specific DNA (Wah et al., 1997), supported these findings and established that the linker that connects two major domains is 15 aa long (amino acids 373–388). Both the FokI architecture and properties of individual domains allowed to expect that genetic dissection of domains would result in impaired specificity of restriction endonuclease leading to double-stranded DNA breaks and, subsequently, to the SOS response. Indeed, eight mutants containing the cassette inserted within the *fokIR* gene were found (Table 2). Five of them had identical structure, suggesting that all are re-isolates of the same initial mutant and thus reducing the total number of unique isolates to four. In all four variants the translation initiation codon ATG, present at the end of the DNA cassette, is in frame with the sequence for FokI C-terminal region and most probably serves to initiate the translation of the latter. For this series of positives, all C-terminal fragments start at residues located within the FokI linker region and thus should code for the full-length DNA cleavage domain. In contrast, the 5'-terminal TAA codon of the DNA cassette is out of frame with N-terminal segments of *fokIR* due to deletions varying from 5 to 277 nucleotides that could have occurred during the preparation of randomly cleaved linear molecules of pUCFokI. In order to evaluate the impact of each part of the dissected *fokIR* gene on the induction of SOS response, either N- or C-terminal fragments were deleted from all four mutants and resulting plasmids were introduced

back into ER1992 [pAC-FokIM]. As expected, only C-terminal fragments were required to induce SOS response, most probably due to their endonucleolytic activity.

FokI dissection experiments, described above, clearly indicated that their outcome depends on the nature of the screening used. Screening in ER1992 lacking the cognate methyltransferase resulted in isolation of insertional mutants which encode complementing fragments. Protein dissection points in these mutants are located on protein surface structure elements, and this observation is in good agreement with findings of others (Ostermeier et al., 1999). On the other hand, it should be noted that, in theory, screening in the absence of protecting methylation is suitable for isolation of both the FokI-specific mutants with the reduced activity (cells must survive) and non-specific mutants. However, all 20 mutants isolated in the absence of protecting methylation belonged to the first group. Therefore, it appears that the frequency of generation of mutants with complementing fragments is much higher compared to that of variants where domains are separated from each other and active, and screening of the latter group of mutants should be done under conditions where mutants of the FokI specificity are ignored. As expected, screening of endonucleolytically active mutants using the FokI modification-positive ER1992 enabled identification of mutants where target recognition and nucleolytic domains of FokI REase are separated. Separation position of the C-terminal domain (375-380 aa) corresponds perfectly with the interdomain linker detected earlier by limited proteolysis and confirmed by X-ray studies (see Fig. 6).

2.3. Investigation of BfiI restriction endonuclease domain organization

2.3.1. Random dissection of the *bfiIR* gene

Random dissection of the *fokIR* gene followed by screening and analysis of functionally active mutants revealed that this method is suitable not only for determining of protein domain organization, but also for construction of complementing protein fragments. In the next step RGD was applied to investigate structural organization of type IIS restriction endonuclease BfiI which is expected to be composed of the N-terminal endonucleolytic domain and the C-terminal DNA recognition domain. The cassette was inserted into the randomly linearized plasmid pUCBfiI, which contains the gene for BfiI restriction endonuclease, using the same technique as in previous experiment. Introduction of ligation mixture into *E. coli* RR1-BfiIM (with genes for both BfiI methyltransferases located on F') resulted in 10^5 Cm-resistant colonies. Restriction mapping of the total plasmid DNA isolated from pooled transformants indicated that about 40% of all insertions occurred outside of the restriction endonuclease gene. These were eliminated essentially in the same way like in experiment with FokI by agarose-purifying the BamHI-Mph1103I DNA fragments of a size matching the *bfiIR* gene with the inserted cassette, subcloning these fragments into the appropriately cleaved pUCBfiI backbone, and then selecting for Cm-resistant transformants.

2.3.2. BfiI mutants of wild-type specificity: isolation and structure

Functionally active insertional mutants of BfiI were identified by their ability to induce SOS response. Transformation of ER1992 with the enriched library of insertional mutants yielded 2 % blue colonies. Eighteen dark-blue colonies were selected for further

analysis. Plasmid DNAs isolated from selected mutants were introduced into both ER1992 and ER1992 [pACBfiIM] in order to check if mutant enzymes retained wild-type specificity (these should be unable to induce SOS response in cells expressing BfiI MTases) or became relaxed, the feature which could be distinguished by methylation status-independent SOS induction. One plasmid was discarded after the observation that it does not induce SOS response in any strain. Of remaining 17 plasmids, sixteen induced SOS response in ER1992 but not in ER1992 [pACBfiIM], clearly indicating that they all coded for BfiI restriction endonuclease of wild-type specificity (BWTS). Also, it seems that the activity of these mutants is greatly reduced as recipient cells survive without protecting DNA methylation. In contrast, one mutant induced SOS response in both strains like other mutants that were isolated using BfiI modification-proficient ER1992; its analysis is described in the next section. Sequencing of mutants of wild-type specificity (Table 3) revealed three regions of BfiI where disruption of the enzyme is tolerated. The largest one is the region between Asp263 and Gln270, in which deletions of up to 5 amino acids are possible (mutant BWTS25). The second one is limited by Pro217 and Thr220, and the third is in fact the boundary between amino acids Gly248 and Thr249. DNA fragments coding for either N- or C-terminal oligopeptides were removed from mutants and deletion variants were tested for their ability to induce SOS response. As expected, both oligopeptides were required to induce the SOS response, suggesting that they complement each other *in vivo*.

Table 3. The structure of BfiI mutants of wild-type specificity. Numbers in columns refer to the first and the last amino acid residue of the respective oligopeptide relative to the amino acid sequence of the wild-type BfiI.

Number of mutant	N-terminal polypeptide*	C-terminal polypeptide
BWTS31	1-217	218-358
BWTS37	1-219	220-358
BWTS39	1-220	221-358
BWTS5	1-248	249-358
BWTS29	1-263	267-358
BWTS25	1-263	269-358
BWTS7; BWTS27	1-264	266-358
BWTS23, BWTS33	1-264	267-358
BWTS3	1-264 (+2)	269-358
BWTS17, BWTS21	1-265	268-358
BWTS19	1-266	268-358
BWTS1	1-267	269-358
BWTS11	1-270	271-358

* - the numbers in parentheses indicate the reading frame of *bfiIR* in which the cassette-encoded translation termination codon TAA was inserted.

Investigation of FokI structural organization by Random Gene Dissection has revealed that FokI can be dissected into two complementing polypeptides and that location of cut-points in mutants of wild-type specificity is within surface-located α

helices and loops. By analogy with FokI, it was proposed that all three aforementioned BfiI regions could be also surface-located.

2.3.3. BfiI mutants of relaxed specificity: isolation and structure

The enriched library of *bfiIR* insertional mutants was introduced into ER1992 [pACBfiIM], the derivative of ER1992 which produces BfiI MTases, and screened for dark-blue colonies which were relatively rare (0.2 %). Plasmids of 18 identified mutants were isolated and transformed back into ER1992 [pACBfiIM] in order to check if the phenotype of selected mutants depends on the genotype of plasmid. All plasmids resulted in SOS response, suggesting that they all encoded the mutant BfiI endonuclease with the relaxed (BREL) specificity. One mutant from previous selection (BREL15) falls into this group, because like other mutants of the group it induces SOS response independent from presence of BfiI MTases and thus has a relaxed specificity. Sequencing of gene dissection points has revealed that in all cases the insertion occurred in a very narrow region of restriction enzyme (Table 4). In total, eight mutants differing in their structure were isolated. DNA sequence analysis has indicated that N-terminal oligopeptides of all but one mutant terminate at the translation termination codon which was derived from the DNA cassette, whereas the putative translation initiation codon remained unpredictable in four isolated mutant variants due to small deletions resulting in shift of reading frame (Table 4). In one mutant (BREL15) the orientation of the cassette was found to be of opposite orientation. This finding indicates that the *E. coli* promoter-like structure is located within the non-coding strand of *bfiIR* gene which can

Table 4. The structure of BfiI mutants of relaxed specificity. Numbers in columns refer to the first and the last amino acid residue of the respective oligopeptide relative to the amino acid sequence of the wild-type BfiI.

Number of mutant	N-terminal polypeptide	C-terminal polypeptide*
BREL7, BREL21	1-185	189-358 (+2)
BREL15**	1-192	194-358
BREL3, BREL39	1-197	203-358 (+1)
BREL5	1-200	205-358
BREL33	1-203	204-358
BREL31	1-203	205-358 (+2)
BREL1, BREL11, BREL13, BREL17, BREL19, BREL23, BREL25, BREL29, BREL35, BREL37	1-204	205-358
BREL16	1-204	207-358 (+2)

* - the numbers in parentheses indicate the reading frame of *fokIR* in which the cassette-encoded translation initiation codon ATG was inserted. In these mutants the putative translation initiation codon of the C-terminal oligopeptide is unpredictable.

** - BREL15 differs from the other mutants by the orientation of the inserted cassette. Because of this, the cassette-encoded amino acid residue Ser was added to the N-terminal oligopeptide, whereas the putative translation initiation codon of the C-terminal oligopeptide cannot be predicted.

ensure transcription of cassette-located chloramphenicol acetyl transferase gene. A deletion study of mutants has indicated that in all cases the N-terminal oligopeptide is enough to induce SOS response and thus it should represent the endonucleolytic domain. This finding suggests that the C-terminal domain should be responsible for specific DNA recognition. Previous FokI studies have indicated that endonucleolytically active mutants with relaxed specificity appear after the precise separation of two functional FokI domains at the linker region connecting them. By analogy with FokI, it may be suggested that two BfiI domains are separated by a linker which comprises amino acids 185-204.

2.3.4. Investigation of functions of putative BfiI domains

Deletion analysis of mutants of relaxed specificity has indicated that the N-terminal domain of BfiI is sufficient to induce SOS response, suggesting that it is responsible for DNA cleavage. To demonstrate the DNA cleavage function directly, we attempted to

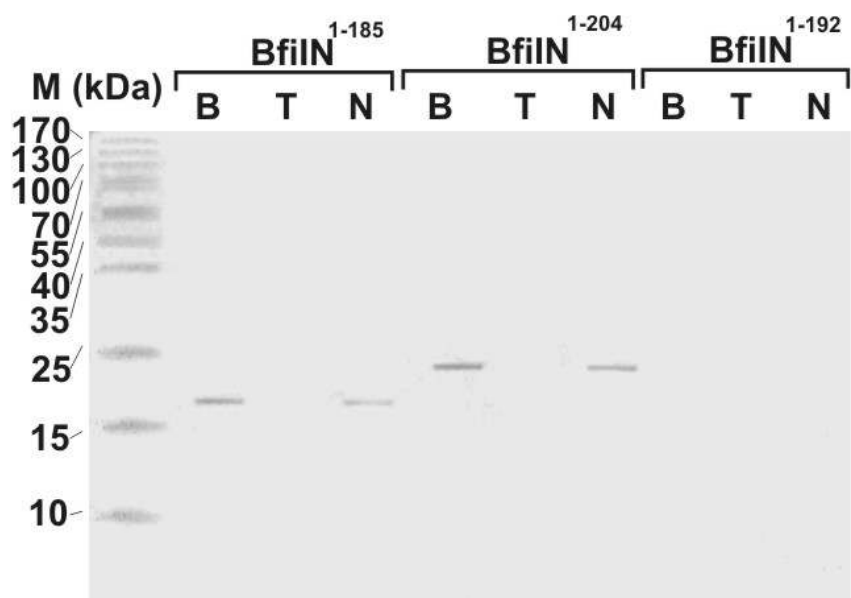


Fig. 7. Analysis of solubility of N-terminal BfiI polypeptides. T – total protein, S - soluble protein fraction, I – insoluble protein fraction. M – protein molecular weight marker.

purify the N-terminal polypeptide. In order to reduce the toxicity of this oligopeptide before induction as much as possible, the vector pZA34-MCS-1 was used, in which tight regulation of a gene of interest can be ensured by recombinant promoter $P_{lac/ara-1}$ (Lutz and Bujard, 1997). DNA fragments coding for N-terminal oligopeptides of mutants with relaxed specificity BREL21 (residues from 1 to 185), BREL15 (1-192) and BREL17 (1-204) were PCR-amplified and cloned under the control of $P_{lac/ara-1}$. The resulting plasmids were named according to the size of N-terminal fragments they encode (pBfiIN¹⁻¹⁸⁵ encodes BfiIN¹⁻¹⁸⁵ polypeptide, pBfiIN¹⁻¹⁹² - BfiIN¹⁻¹⁹² and pBfiIN¹⁻²⁰⁴ - BfiIN¹⁻²⁰⁴, respectively). Introduction of pBfiIN¹⁻¹⁹² into the strain ER1992/F' induced SOS response even in the absence of $P_{lac/ara-1}$ inducers L(+)-arabinose and IPTG, whereas addition of both inducers resulted in complete arrest of cell growth. In contrast, two other plasmids had a very mild effect on the induction of SOS response without inducers but induced a strong SOS response when both inducers were added. To understand the reason for diverse behavior of three oligopeptides differing in just a few C-terminal amino acids, their synthesis in *E. coli* cells was induced, the yield of N-terminal oligopeptides and their distribution in fractions of soluble and insoluble proteins was compared by using immunoblotting and polyclonal antibodies developed against BfiI (see Fig. 7). It was observed that cells carrying

pBfiIN¹⁻²⁰⁴ and pBfiIN¹⁻¹⁸⁵ produce very low amounts of corresponding oligopeptides which are found mostly in fraction of insoluble proteins. In contrast, no reacting proteins were found when cells carrying pBfiIN¹⁻¹⁹² were induced. Basing on all these observations it can be stated that the endonucleolytic activity of Bfil¹⁻¹⁹² is highest and for that reason it is the most detrimental to the cell, whereas the expression of two other derivatives is not so toxic due to their lowered solubility. Also, it was concluded that yields of soluble N-terminal oligopeptides were not sufficient for purification experiments.

In order to analyse function of the C-terminal part of Bfil REase, three polypeptides, which are different in length and correspond the previously analysed three N-terminal polypeptides BfiIN¹⁻¹⁸⁵, BfiIN¹⁻¹⁹² and BfiIN¹⁻²⁰⁴, were chosen. PCR-amplified *bfiIR* gene fragments coding for BfiIC¹⁸⁶⁻³⁵⁸ (residues from 186 to 358), BfiIC¹⁹³⁻³⁵⁸ (193-358) and BfiIC²⁰⁵⁻³⁵⁸ (205-358) were inserted into pUC57 under the control of P_{lac} promoter. Resulting plasmids were named according to Bfil regions encoded - pBfiIC¹⁸⁶⁻³⁵⁸, pBfiIC¹⁹³⁻³⁵⁸ and p BfiIC²⁰⁵⁻³⁵⁸. The cells transformed with pBfiIC¹⁸⁶⁻³⁵⁸ and pBfiIC¹⁹³⁻³⁵⁸ were less viable compared to those carrying pBfiIC²⁰⁵⁻³⁵⁸ or vector pUC57, however, none of the plasmids induced SOS response (data not shown). In addition, co-introduction of pACBfiIM, the plasmid coding for two Bfil methyltransferases, together with pBfiIC¹⁸⁶⁻³⁵⁸ or pBfiIC¹⁹³⁻³⁵⁸ increased greatly the viability of transformants (data not shown). These results indicate that site-specific DNA-protein interaction is the most probable reason of the observed toxicity which can be reduced by methylation of Bfil targets.

Analysis of expression levels and solubility using Bfil-specific antibodies has indicated that oligopeptides BfiIC¹⁸⁶⁻³⁵⁸ and BfiIC¹⁹³⁻³⁵⁸ are soluble, whereas BfiIC²⁰⁵⁻³⁵⁸ was found in fraction of insoluble proteins only (Fig. 8). It should be noted, however, that the yield of Bfil²⁰⁵⁻³⁵⁸ is greatly reduced, suggesting that either this oligopeptide for unknown reasons is expressed very poorly or it is very susceptible to intracellular proteases and is very quickly degraded. Nevertheless, these results may explain why pBfiIC²⁰⁵⁻³⁵⁸ had no detectable effect on the viability of host cells. The shorter protein, BfiIC¹⁹³⁻³⁵⁸, was selected for further analysis. Purification of this oligopeptide by four consecutive steps of chromatography has resulted in BfiIC¹⁹³⁻³⁵⁸ of >99 % purity (Fig. 9). DNA binding studies of wt Bfil and its truncated form BfiIC¹⁹³⁻³⁵⁸ were carried out using specific and non specific oligoduplexes as substrates and gel-mobility-shift-assay. It should be noted that Bfil REase cleaves specific DNA even in the absence of metal ions. In order to minimize

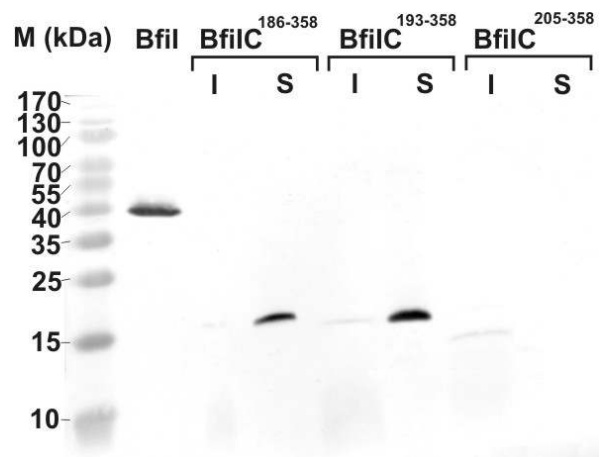


Fig. 8. Analysis of solubility of C-terminal Bfil polypeptides. I – insoluble protein fraction, S – soluble protein fraction. M - protein molecular weight marker.

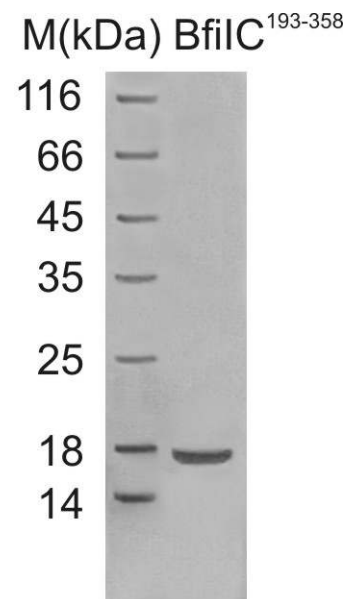


Fig 9. SDS-PAGE of purified BfiIC¹⁹³⁻³⁵⁸ polipeptide.

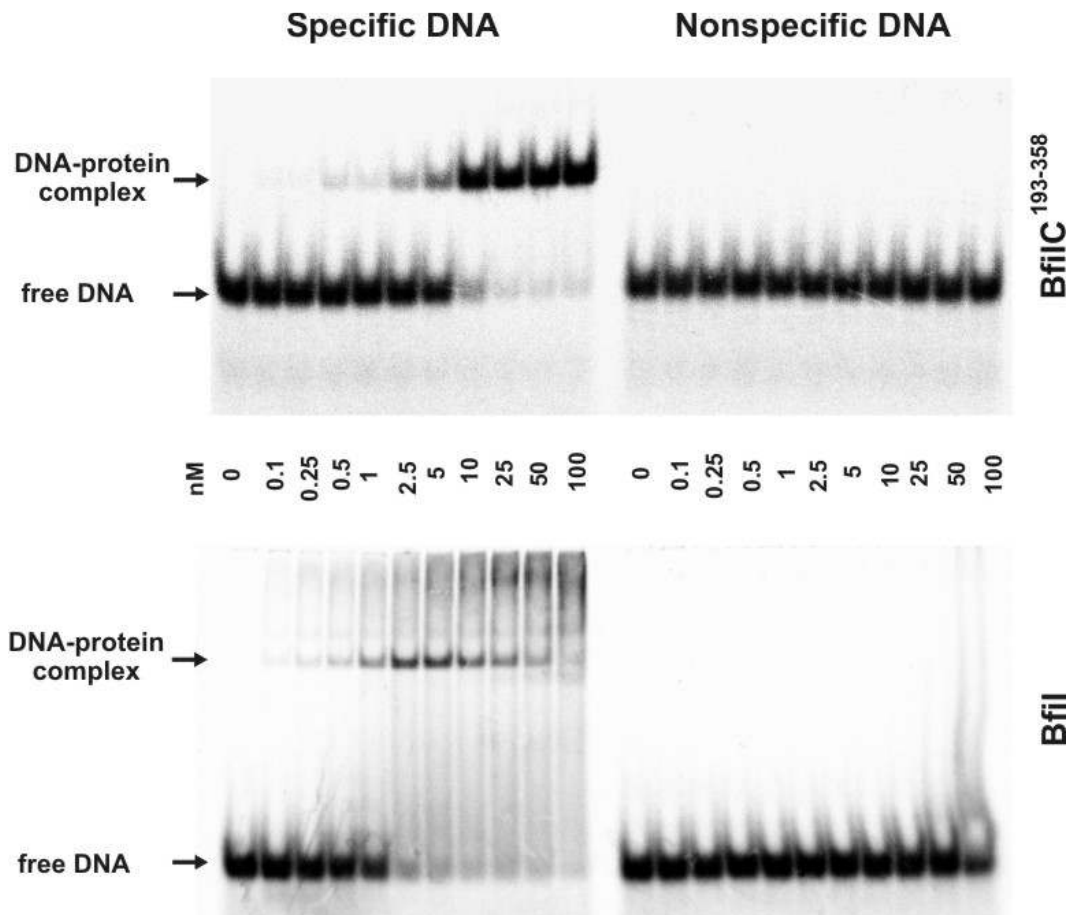


Fig. 10. Gel-shift assays comparing affinity of wild type BfiI and deletion mutant BfiIC¹⁹³⁻³⁵⁸ for specific and nonspecific 20 bp oligoduplexes (1 nM of each). The protein concentrations (in nM) are indicated above each lane. Gel-shift conditions are described in Materials and Methods.

DNA digestion during gel-mobility-shift-assay, the modified oligonucleotides with phosphorothioates introduced at BfiI cleavage positions were used. DNA binding studies (Fig. 10) were carried out using constant substrate DNA concentration (1 nM) and protein concentrations in range of 0-100 nM (concentrations of proteins are given for a monomer form of protein). Protein-DNA complexes with both wt BfiI and BfiIC¹⁹³⁻³⁵⁸ using specific DNA were observed, however, no complexes were formed with nonspecific DNA. Also, it seems that wild-type BfiI has a bit higher affinity for specific DNA than BfiIC¹⁹³⁻³⁵⁸ polypeptide (K_d values $3,2 \pm 0,6$ nM and $6,4 \pm 1,3$ nM respectively). These results indicate that all information necessary for sequence-specific DNA recognition is included in C-terminal fragment of BfiI and therefore it may be considered as target recognition domain of BfiI REase.

Discussion

Type IIS restriction endonuclease BfiI recognizes the asymmetric hexanucleotide sequence 5'-ACTGGG and cleaves the top DNA strand 5 nt and bottom strand 4 nt away from the sequence, producing a single base 3'-protruding end. The genes of BfiI R-M system were cloned and sequenced. Deletion analysis and amino acid sequence alignment revealed that BfiI R-M system contains three proteins: two N4-methylcytosine

methyltransferases and a restriction endonuclease. The amino acid sequence of the N-terminal part of BfiI has some similarities to Nuc of *Salmonella typhimurium*, an EDTA-resistant nuclease, which is a member of phospholipase D superfamily. Like Nuc nuclease BfiI REase does not require Mg²⁺ ions for DNA cleavage.

Domain organization of the only one type IIS restriction enzyme, FokI, was known at the starting point of this work. The N-terminal polipeptide of FokI functions as target recognition domain whereas the C-terminal domain is involved in DNA cleavage. BfiI recognizes asymmetric sequence and cuts DNA outside the target like FokI, suggesting that it may be composed of two domains of different function as well. In addition, amino acid residues of the active site are located on the N-terminal part of BfiI suggesting that this part of the protein could form the nucleolytic domain. Based on these assumptions the aim to investigate the domain organization of BfiI was set.

Various protein fragmentation approaches are used to study structural organization of proteins. The most popular among them is limited proteolysis. However, during the recent years some alternative strategies have been developed. In one of them, insertions into the gene under investigation are generated using either transposons (Manoil and Bailey, 1997) or random DNA breaks introduced by concerted DNase I and S1 nuclease action (Biondi et al., 1998). Screening for functionally active mutants carrying in-frame insertions helps to identify loops or other surface elements which are able to accommodate small (Manoil and Bailey, 1997) or large (Biondi et al., 1998) insertions. The other approach uses incremental N- and C-terminal deletions of the same gene which are generated using DNA degrading enzymes. Shortened variants are then co-introduced into cells in order to find out complementing pairs (Ostermeier et al., 1999). As alternative, DNA regions coding for truncated proteins can be joined together in all possible combinations by ligation and then resulting chimeras can be selected for catalytically active variants (Vilkaitis et al., 2002). The information gained from such experiments can be used for domain swapping or other protein engineering strategies as well as for development of techniques which require interdomain or intradomain interactions resulting in restored functional activity of dissected protein.

A new protein fragmentation technique, Random Gene Dissection, has been developed in this work, which in fact mimics reverse evolution, the process in which a two-domain protein is dissected into two individual domains. The technique is based on a random DNase I/S1-mediated disruption of a gene of interest followed by insertion of a specially designed DNA cassette. The cassette ensures premature termination of the N-terminal protein fragment and independent translation initiation of the C-terminal fragment if inserted in frame. It should be noted, however, that DNase I has some sequence preferences (Herrera and Chaires, 1994). As a result, some bias towards cassette insertion into preferred regions should exist. A prove-of-principle experiment was carried out using the well-studied restriction endonuclease FokI as a model. Screening for endonucleolytically active mutants of relaxed specificity, i.e. mutants which induce SOS response regardless of protecting methylation, and their analysis indicated that C- but not N-terminal fragments of selected mutants are responsible for DNA cleavage activity. These results agree well with data from other laboratories (Li et al., 1992; Wah et al., 1997). More important, in all cases C-terminal fragments start within the short linker which connects FokI domain responsible for DNA recognition with that for DNA cleavage. It should be mentioned, however, that this result was somewhat surprising as it was expected that additional amino acids derived from the

N-terminal domain wouldn't have very strong effect on the activity of endonucleolytic domain and thus should be tolerated. Therefore, it seems that localization of cut-point within the very narrow protein region which coincides with the interdomain linker may be the indication that these extra amino acids either greatly reduce the activity of FokI C-terminal domain or make it insoluble or unstable. On the other hand, these results have demonstrated that Random Gene Dissection can be used for the identification of functionally active endonucleolytic domain.

In the next experiment the RGD technique was applied to investigate the domain structure of BfiI restriction endonuclease. As mentioned above, protein comparison studies have allowed to predict that, in contrast to FokI, the BfiI catalytic domain is located at the protein N-terminus. Random insertions of the DNA cassette into the gene for BfiI were generated, and then screening for functionally active BfiI mutants both in the presence and in the absence of protecting methylation was performed. Cut-points of 18 mutants, screened in the presence of protecting methylation, were determined. It was observed that translation of the N-terminal fragment in all cases stops within the narrow region of BfiI starting with amino acid Asp185 and ending with Thr204. Deletion of gene fragments coding for C-terminal oligopeptides had no effect on the ability of mutants to induce SOS response, clearly indicating that the N-terminal oligopeptide possesses endonucleolytic activity. In order to investigate the presumable function of domains experimentally, overexpression and purification of corresponding protein regions was attempted. BfiI gene regions coding for C-terminal amino acid residues 186-358, 193-358 and 205-358 were expressed and solubility of respective oligopeptides evaluated using BfiI-specific polyclonal antibodies. It was found that BfiIC¹⁸⁶⁻³⁵⁸ and BfiIC¹⁹³⁻³⁵⁸ are soluble. In contrast, the yield of BfiIC²⁰⁵⁻³⁵⁸ was reduced sharply and all reacting protein was found in insoluble fraction. Also, some cytotoxicity by two former oligopeptides is observed which was reduced significantly by introduction of protecting BfiI methyltransferases. Purification and DNA binding studies of both wild type BfiI and its mutant BfiIC¹⁹³⁻³⁵⁸ showed that both have similar affinities for specific DNA. Basing on these results it can be concluded that the C-terminal oligopeptide represents DNA binding domain and that the observed toxicity most probably is the result of domain binding to specific DNA targets which can be prevented by their methylation. Unfortunately, attempts to express and purify N-terminal oligopeptides BfiIN¹⁻¹⁸⁵, BfiIN¹⁻¹⁹² and BfiIN¹⁻²⁰⁴ failed. Western blotting using BfiI-specific antibodies showed that small amounts of BfiIN¹⁻²⁰⁴ and BfiIN¹⁻¹⁸⁵ are synthesized, however, both oligopeptides are found in insoluble fraction only. In contrast, no detectable traces of BfiIN¹⁻¹⁹² were detected. Worthy of note, the plasmid coding for BfiIN¹⁻¹⁹² was most detrimental to the host as judged by induction of sharp SOS response in the absence of inducers and immediate arrest of culture growth after their addition. Taken together, it is concluded that the N-terminal domain of BfiI is responsible for DNA cleavage. It should be noted that the smallest functional polypeptides are BfiIN¹⁻¹⁸⁵ and BfiIC¹⁹³⁻³⁵⁸, thus the region between these parts (from Asp185 to Leu193) could correspond to the linker which connects two domains. Also, experimental data indicate that the solubility of domains depends strongly on the position of cut-point, and even few additional amino acids (including those which come from the linker) can decrease the solubility greatly.

During investigation of domain organization of BfiI using the RGD approach, the parallel research on the structure of BfiI was performed in Laboratory of Protein-DNA Interactions at the Institute of Biotechnology by using limited proteolysis, and

three-dimensional structure of the protein was determined by X-ray crystallography (Zaremba et al., 2004; Grazulis et al., 2005). BfiI was dissected into N-terminal and C-terminal domains by using protease thermolysin that cleaved the protein at Leu191. Functions of individual purified domains were demonstrated: the N-terminal domain has non-specific nucleolytic activity, while C-terminal domain binds specifically to the recognition sequence (Zaremba et al., 2004). These results correlate well with those received by RGD technique. Furthermore, the X-ray analysis revealed that the interdomain region is located from Thr171 to Arg198 (Grazulis et al., 2005), and it includes the linker region which was predicted using RGD (region from Asp185 to Leu193) (Fig. 11). The results of parallel research confirm the suitability of RGD approach for detecting of interdomain linkers in two-domain proteins.

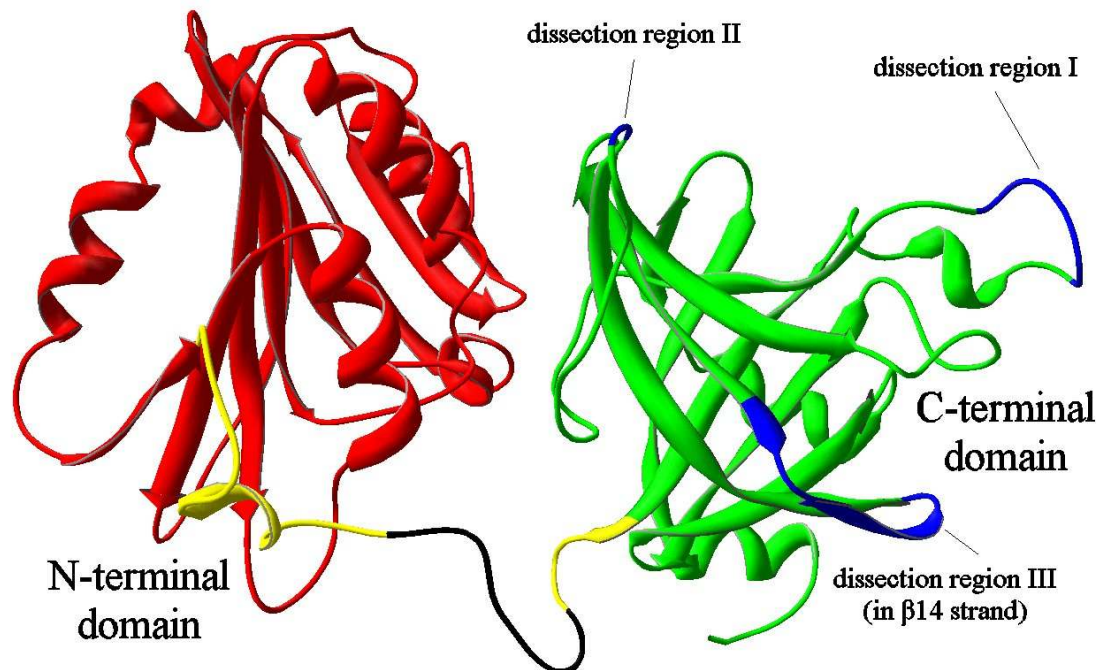


Fig. 11. Position of dissection regions in BfiI tertiary structure. Endonucleolytic domain is marked in red, target recognition domain – in green, interdomain region determined by X-ray crystallography is in yellow, interdomain region determined by RGD – in black, regions where BfiI was dissected into complementary fragments are marked in blue.

Screening for functionally active FokI and BfiI mutants in the absence of protecting methylation resulted in different protein dissection regions. These insertional mutants possess the wild-type specificity as judged by loss of their ability to induce SOS response in cells expressing FokI or BfiI MTases. Deletion analysis of isolated insertional mutants indicated that in all cases both oligopeptides of dissected proteins are required to induce SOS response, thus they complement each other *in vivo* resulting in functional enzyme. Analysis of FokI dissection points revealed that they are always located on surface-exposed α helices ($\alpha 2$, $\alpha 3$) and loops (L1, L2) in the target recognition domain. In case of BfiI three dissection regions were identified (I - from 217 to 221 amino acid residue, II - from 248 to 249, and III – from 263 to 271). Similarly to FokI, BfiI dissection regions are located on the surface of the BfiI target recognition domain: the first two are in loops, and the third one – on $\beta 14$ strand (Fig. 11). The observation

that dissection points are located in elements of the protein surface is in good agreement with results of other researchers (Ostermeier et al., 1999).

The results obtained with FokI and BfiI demonstrate that RGD could be a method of choice not only for investigation of domain organization of proteins, but also for isolation of complementing protein fragments. Protein fragment complementation is used to explore theories of protein evolution (Bertolaet and Knowles, 1995) or protein folding (Ladurner et al., 1997), it has also proved useful for selection of recombinant DNA molecules (Ullmann et al., 1967), for targeting of proteins (Cabantous et al., 2005) and for the development of enzyme-based two-hybrid systems (Remy and Michnick, 1999).

Complementary fragments of proteins can be generated in several different ways: they have been constructed by synthesis of randomly chosen proteins fragments (Shiba and Schimmel, 1992), by using limited proteolysis (Tasayco and Carey, 1992), after analysis of three-dimensional structure model (Ghosh et al., 2000) or by construction of incremental truncation libraries (Ostermeier et al., 1999). The latter method is based on construction of N- and C-terminal protein deletion sets, followed by combining of deletion variants and selection of complementing parts. The procedures performed and information obtained from RGD partially resembles this combinatorial protein engineering, however, RGD approach has some advantages over it. First, incremental truncation results in overlapping gene fragments which can recombine and restore functional activity of gene under investigation resulting in background of false-positives, the problem that does not exist in RGD approach. Second, in order to find out all possible complementing pairs of truncated gene which codes for a protein of 300 amino acid residues (900 nt) one needs to combine all possible mutants sequentially shortened starting from the 5'-terminus (at least 900 mutants) with the collection of mutants shortened from the 3'-terminus (at least 900 mutants). It means that at least $900 \times 900 = 810\,000$ mutants are required to analyze. However, taking into account the statistical distribution of sizes of deleted regions one needs to combine both libraries with at least 5-fold redundancy. This results in necessity to analyze $5 \times 900 \times 5 \times 900 = 2 \times 10^7$ transformants, a task which can be carried out only in case if powerful selection technique exists. In contrast, it is necessary to analyze only a very limited number of mutants generated by RGD as each mutant represents a unique bisection point. For example, to create a collection of all possible insertional mutants of 900-nt long gene one needs to analyze a collection of 900 insertional mutants. Having in mind the preference of DNase I for some sequences and tendency to produce small deletions at the cassette insertion point (Biondi et al., 1998) some level of redundancy is required to be sure that all possible variants were created. Nevertheless, even a 20-fold redundancy requires analysis of just 1.8×10^4 of mutants, a number that is more than 1000 times smaller compared to the number of variants required to analyze using the approach of incremental truncations. Furthermore, each of 1.8×10^4 mutants can be individually screened for functional activity using high-throughput systems. Finally, each possible cut-point is represented by only one mutant in RGD approach. In contrast, combining of incrementally truncated proteins can result in a lot of complementing pairs which are in fact variants of the same mutation. Thus RGD approach can be considered as the most efficient and convenient method for construction of complementary fragments of protein.

Conclusions

1. The genes of BfiI restriction-modification system were cloned. The system comprises two cytosine-N4-methyltransferases and a restriction endonuclease.
2. A new method for investigation of protein domain organization, Random Gene Dissection, was proposed. The interdomain region of the chosen model protein, FokI REase, was determined using the new method, and it was in perfect agreement with the linker region predicted from the tertiary structure of FokI.
3. Interdomain region of BfiI REase was determined by Random Gene Dissection method.
4. Indirect data (induction of SOS response, toxicity) indicate that the N-terminal domain of BfiI acts as a nuclease.
5. The DNA binding assay using purified C-terminal domain of BfiI indicates that it is responsible for target recognition.
6. The complementary fragments of FokI and BfiI were isolated using Random Gene Dissection. Therefore the Random Gene Dissection method can be used not only for identification of interdomain regions, but also for isolation of complementing fragments of proteins.

List of publications

The doctoral dissertation is based on the following original publications:

1. **Šapranauškas R.**, Sasnauskas G., Lagunavičius A., Vilkaitis G., Lubys A., Šikšnys V. 2000. Novel subtype of type IIs restriction enzymes. BfiI endonuclease exhibits similarities to the EDTA-resistant nuclease Nuc of *Salmonella typhimurium*. J Biol Chem. 2000. 275:30878-85.
2. **Šapranauškas R.**, Lubys A. 2005. Random gene dissection: a tool for the investigation of protein structural organization. Biotechniques. 2005. 39:395-402.

Part of material of doctoral dissertation was presented in poster presentations at conferences:

1. **Šapranauškas R.**, Janulaitis A. and Lubys. Investigation of protein domain organization by insertional mutagenesis. The 3th Genetical Congress of Baltic States. 2002, Vilnius, Lithuania.
2. **Šapranauškas R.**, Janulaitis A. and Lubys. Random insertional mutagenesis: an efficient tool for investigation of protein structural organization. 5th New England Biolabs Meeting on Restriction/Modification. 2004, Bristol, United Kingdom.

Acknowledgements

I am grateful to my supervisor dr. Arvydas Lubys for suggestion of theme of my work, consistent supervision, critical notes and help in preparing dissertation. I thank Jolanta Giedriene for the help with protein purification and Margarita Leckiene for production of antibodies. I am grateful to my colleagues from Laboratory of Prokaryote Gene Engineering for their assistance, discussions and moral support. I am thankful to Giedre Tamulaitiene and Gintas Tamulaitis for valuable discussions when preparing the doctoral thesis.

I would like to thank my family for their patience and understanding during the preparation of the dissertation.

I am grateful to “Fermentas” UAB for providing enzymes and kits and access to laboratory equipment. This work was supported in part by contract with “Fermentas” UAB.

CURRICULUM VITAE

Name: Rimantas Šapranauskas

Date of birth: 17.06.1974

Address: Laboratory of Protein - DNA Interactions
Institute of Biotechnology
V.A. Graičiūno 8, Vilnius, LT-02241
(+370 5) 2602698
sapranauskas@ibt.lt

Education: 1996 Bachelor of Science in Biology, Faculty of Natural Sciences, Vilnius University.
1998 Master of Science in Biology, Faculty of Natural Sciences, Vilnius University.

Employment: Currently Institute of Biotechnology, Laboratory of Protein - DNA Interactions, junior scientist.

2002-2008 Institute of Biotechnology, Laboratory of Prokaryote Gene Engineering, junior scientist.

1998-2002 Ph.D student of Biochemistry.

1997-1998 Institute of Biotechnology, Laboratory of Prokaryote Gene Engineering, research assistant.

Reziümė

Restrikcijos endonukleazės (REazės) atpažįsta trumpas, 4-8 bp ilgio DNR sekas ir kerpa DNR atpažinimo sekos viduje arba už jos ribų. Kartu su metiltransferazėmis jos yra sudedamoji restrikcijos-modifikacijos (R-M) sistemų dalis. Manoma, jog pagrindinė R-M sistemų funkcija - ląstelių apsauga nuo egzogeninės DNR. Pagal struktūrinę baltymų organizaciją, kirpimo pobūdį, o taip pat pagal kofaktorių panaudojimą šiuo metu restrikcijos fermentai yra skirstomi į keturis tipus. Plačiausiai ištyrinėtos yra II potipio restrikcijos endonukleazės, kurios yra nepakeičiami molekulinės biologijos įrankiai konstruojant ir analizuojant rekombinantines DNR molekules. II tipo fermentai skiriasi atpažįstamomis sekomis (simetrinės, asimetrinės), funkcijomis (monofunkciniai, bifunkciniai), oligomerine būsena (monomerai, dimerai, tetramerai) bei evoliucine kilme (gali priklausyti PD-(D/E)XK, PLD, HNH, arba GIY-YIG superšeimoms). Be to, II tipo REazės skiriasi ir savo domenine organizacija. Pavyzdžiui, ortodoksinės IIP, o taip pat į jas struktūriškai panašios tetramerus formuojančios IIF potipio REazės yra monodomeniniai fermentai, kuriuose taikinio atpažinimo bei DNR fosfodiesterinės jungties hidrolizės procesuose dalyvauja persipynusios baltymų sritys, o šių procesų mechanizmai yra glaudžiai susiję. Tuo tarpu IIE ir IIS REazių tyrimai parodė, jog šiems fermentams būdinga domeninė organizacija.

Šiame darbe buvo tyrinėta BfiI restrikcijos-modifikacijos sistemos genetinė organizacija bei BfiI REazės domeninė sandara. BfiI REazė atpažįsta asimetrinę 5'-ACTGGG seką, kerpa DNR viršutinę grandinę už 5, o apatinę – už 4 nukleotidų nuo atpažinimo sekos ir pagal šiuos požymius yra priskiriama IIS potipiui. Darbo metu buvo klonuoti BfiI restrikcijos-modifikacijos sistemos genai. Nukleotidinės sekos bei iškritų analizės metu buvo nustatyta, jog šią R-M sistemą sudaro dvi m4C metiltransferazės bei viena restrikcijos endonukleazė. BfiI REazė yra homologiška Nuc nukleazei iš *Salmonella typhimurium* kuri priklauso PLD/Nuc fosfolipazių superšeimai. Kaip ir Nuc nukleazė, BfiI REazė DNR karpymui nereikalauja Mg^{2+} jonų. Tyrimo metu buvo žinoma vienintelio IIS potipio atstovo FokI struktūra. Šis baltymas pasižymi dvidomenine organizacija: fermento N-galinė dalis yra atsakinga už taikinio atpažinimą, o C-galinė dalis – už DNR kirpimą. Atsižvelgiant į tai, kad abi REazės, FokI ir BfiI, priklauso tam pačiam potipiui, buvo padaryta prielaida, jog BfiI REazė taip pat gali pasižymėti domenine sandara. Šiai prielaidai neprieštaravo ir aminorūgščių sekos palyginimo rezultatai, kurie parodė, kad potencialaus katalitinio centro aminorūgštys yra lokalizuotos N-galinėje baltymo dalyje.

BfiI REazės domeninei organizacijai tirti buvo pasiūlytas naujas geno atsitiktinio padalinimo (GAP) metodas. GAP remiasi prielaida, jog baltymo domenai yra struktūriniai-funkciniai baltymo vienetai, kurie pasižymi savarankiška funkcija bei sugeba nepriklausomai vienas nuo kito susivynioti. Taikant šį metodą tikslinį baltymą koduojantis genas atsitiktinėse vietose padalinamas į dvi dalis įterpiant specialiai sukonstruotą DNR kasetę, kurios 5'-gale lokalizuotas transliacijos terminacijos signalas TAA, o pačiame 3'-gale išdėstyti transliacijos iniciacijos signalai (Shine-Dalgarno seka bei iniciatorinis kodonas ATG). Tokiai kasetei reikiama orientacija įsistačius į atviro skaitymo rėmelį turėtų būti užtikrinama nepriklausoma N- ir C-galinių polipeptidų sintezė. Tie mutantai, kuriuose kasetė bus padalinusi baltymą tarpdomeninėje srityje turėtų sintetinti funkcionalius N- ir C-galinius polipeptidus. IIS potipio REazių atveju

nukleazinis domenas įprastai veikia kaip nespecifinė nukleazė, ir ląstelėse turėtų sudaryti DNR trūkius bei iššaukti SOS atsaką. Todėl mutantus su atskirtais taikinio ir katalitiniais domenais buvo nuspręsta atrinkti panaudojant SOS atsako indikatorinio *E. coli* kamieno ER1992 ląsteles.

Intarpinės mutagenezės bei mutantų atrankos sąlygų parinkimui pirminiuose eksperimentuose GAP metodas buvo išbandytas naudojant žinomos struktūros modelinį baltymą - FokI REazę. Nespecifiniu nukleaziniu aktyvumu pasižyminčių mutantų atrankai buvo panaudotos ER1992 kamieno ląstelės su įvestu FokI metiltransferazės genu. Atrinkus nespecifiniu nukleaziniu aktyvumu pasižyminčius intarpinius mutantus paaiškėjo, jog už SOS atsako indukciją yra atsakingas tik C galinis fragmentas. Išanalizavus kasetės įsistatymo sritis aktyviuose mutantuose buvo nustatyta, jog šis polipeptidas atitiko nukleazinį FokI domeną (Li ir kiti, 1992; Wah ir kiti, 1997). Be to, visuose mutantuose FokI C-galinio polipeptido transliacijos iniciacijos vietos yra išsidėsčiusios siaurame regione, kuris puikiai atitinka tarpdomeninę FokI REazės jungtį. Tokiu būdu FokI pavyzdžiu buvo pademonstruota, jog GAP metodas tinkamas funkcionalių domenų atrankai bei tarpdomeninės jungties lokalizacijai.

Sekančiame etape GAP buvo panaudota BfiI REazės domeninės organizacijos tyrimui. Išanalizavus kasetės įsistatymo vietas atrinktuose nespecifiniu nukleaziniu aktyvumu pasižyminčiuose mutantuose paaiškėjo, jog visais atvejais N-galinio polipeptido sintezė baigiasi siaurame regione, kuris apima aminorūgštis nuo Asp185 iki Thr204. Iškritų analizė parodė, jog už SOS atsaką yra atsakingas tik N-galinis BfiI polipeptidas. Tokiu būdu buvo netiesiogiai buvo parodyta, jog BfiI N-galinė dalis atsakinga už nukleazinę BfiI domeno funkciją. C-galinio polipeptido tyrimai parodė jog ši dalis atsakinga už specifinę sąveiką su DNR taikiniu. Išgrynintas polipeptidas, apimantis nuo 193 iki 358 C-galines aminorūgštis rišosi su BfiI taikinį turinčiu dvigrandžiu oligonukleotidu, tuo tarpu sąveika su nespecifiniu oligonukleotidu nebuvo stebima. Skirtingo ilgio N- ir C-galinių polipeptidų analizė leido identifikuoti tarpdomeninės jungties ribas: mažiausi polipeptidai, pasižymintys aktyvumu, buvo BfiIN¹⁻¹⁸⁵ bei BfiIC¹⁹³⁻³⁵⁸, todėl tarpdomeninė jungtis turėtų apimti BfiI baltymo sritį nuo Asp185 iki Leu193. Šiuos tyrimų rezultatus bei iškeltas prielaidas patvirtino nepriklausomai Biotechnologijos instituto Baltymų nukleorūgščių sąveikos tyrimo laboratorijoje atlikti BfiI restrikcijos endonukleazės struktūriniai tyrimai. Dalinės proteolizės eksperimentų metu proteazė termolizinas suskaldė BfiI REazę ties Leu191, tuo būdu atskirdamas katalitinį ir taikinio atpažinimo domenų (Zaremba ir kiti, 2004), o rentgeno-struktūrinė analizė parodė, kad BfiI tarpdomeninė jungtis apima aminorūgštis nuo Thr171 iki Arg198 (Grazulis ir kiti, 2005).

Tiek BfiI, tiek FokI atvejais be nespecifiniu nukleaziniu aktyvumu pasižyminčių mutantų buvo atrinkti mutantai, pasižymintys taikiniui specifiniu nukleaziniu aktyvumu. Atrinktų mutantų iškritų analizė parodė, jog abiem atvejais specifinių taikinių karpymui *in vivo* reikalingi abu polipeptidai, t.y. N- ir C-galiniai fragmentai komplementuoja vienas kitą formuodami funkciškai aktyvų baltymą. Analizuojant padalinimo vietas paaiškėjo, jog dauguma jų yra lokalizuotos baltymo paviršiuje eksponuotose srityse. Šie rezultatai rodo, jog GAP metodas gali būti sėkmingai panaudotas ne tik baltymų domenų bei tarpdomeninių jungčių identifikavimui, bet ir komplementuojančių baltymo fragmentų gavimui.

References

1. Ausubel, F.M., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and Struhl, K. (1992). Short Protocols In Molecular Boilogy (Green Publishing Associates and John Wiley & Sons).
2. Bertolaet, B.L., and Knowles, J.R. (1995). Complementation of fragments of triosephosphate isomerase defined by exon boundaries. *Biochemistry* 34, 5736-5743.
3. Biondi, R.M., Baehler, P.J., Reymond, C.D., and Veron, M. (1998). Random insertion of GFP into the cAMP-dependent protein kinase regulatory subunit from Dictyostelium discoideum. *Nucleic Acids Res* 26, 4946-4952.
4. Bjorklund, A.K., Ekman, D., Light, S., Frey-Skott, J., and Elofsson, A. (2005). Domain rearrangements in protein evolution. *J Mol Biol* 353, 911-923.
5. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L., and Boyer, H.W. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. *Gene* 2, 95-113.
6. Bullock, W.O., Fernandex, J.M., and Short, J.M. (1987). X11-Blue: A High Efficiency Plasmid Transforming recA Escherichia coli Strain With Beta-Galactosidase Selection. *Biotechniques* 5, 376-379.
7. Cabantous, S., Terwilliger, T.C., and Waldo, G.S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nature biotechnology* 23, 102-107.
8. Chan, S.H., Bao, Y., Ciszak, E., Laget, S., and Xu, S.Y. (2007). Catalytic domain of restriction endonuclease BmrI as a cleavage module for engineering endonucleases with novel substrate specificities. *Nucleic Acids Res* 35, 6238-6248.
9. Chang, A.C., and Cohen, S.N. (1978). Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *Journal of bacteriology* 134, 1141-1156.
10. 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.
11. Ghosh, I., Hamilton, A.D., and Regan, L. (2000). Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J Am Chem Soc* 122, 5658-5659.
12. Grazulis, S., Manakova, E., Roessle, M., Bochtler, M., Tamulaitiene, G., Huber, R., and Siksnys, V. (2005). Structure of the metal-independent restriction enzyme BfiI reveals fusion of a specific DNA-binding domain with a nonspecific nuclease. *Proceedings of the National Academy of Sciences of the United States of America* 102, 15797-15802.
13. Gunn, J.S., and Stein, D.C. (1997). The Neisseria gonorrhoeae S.NgoVIII restriction/modification system: a type IIs system homologous to the Haemophilus parahaemolyticus HphI restriction/modification system. *Nucleic Acids Res* 25, 4147-4152.
14. Heitman, J., Zinder, N.D., and Model, P. (1989). Repair of the Escherichia coli chromosome after in vivo scission by the EcoRI endonuclease. *Proceedings of the National Academy of Sciences of the United States of America* 86, 2281-2285.

15. Herrera, J.E., and Chaires, J.B. (1994). Characterization of preferred deoxyribonuclease I cleavage sites. *J Mol Biol* 236, 405-411.
16. Huai, Q., Colandene, J.D., Topal, M.D., and Ke, H. (2001). Structure of NaeI-DNA complex reveals dual-mode DNA recognition and complete dimer rearrangement. *Nature structural biology* 8, 665-669.
17. Hubbard, S.J., Eisenmenger, F., and Thornton, J.M. (1994). Modeling studies of the change in conformation required for cleavage of limited proteolytic sites. *Protein Sci* 3, 757-768.
18. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33, 103-119.
19. Yoshioka, K. (2002). KyPlot – A User-oriented Tool for Statistical Data Analysis and Visualization. . In *CompStat* pp. 425–437.
20. Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America* 93, 1156-1160.
21. Kriukiene, E. (2006). Domain organization and metal ion requirement of the Type IIS restriction endonuclease MnlI. *FEBS Lett*.
22. Ladurner, A.G., Itzhaki, L.S., de Prat Gay, G., and Fersht, A.R. (1997). Complementation of peptide fragments of the single domain protein chymotrypsin inhibitor 2. *J Mol Biol* 273, 317-329.
23. Lagunavicius, A., Sasnauskas, G., Halford, S.E., and Siksnys, V. (2003). The metal-independent type IIs restriction enzyme BfiI is a dimer that binds two DNA sites but has only one catalytic centre. *J Mol Biol* 326, 1051-1064.
24. Li, L., Wu, L.P., and Chandrasegaran, S. (1992). Functional domains in Fok I restriction endonuclease. *Proceedings of the National Academy of Sciences of the United States of America* 89, 4275-4279.
25. Lubys, A., Lubiene, J., Kulakauskas, S., Stankevicius, K., Timinskas, A., and Janulaitis, A. (1996). Cloning and analysis of the genes encoding the type IIS restriction-modification system HphI from *Haemophilus parahaemolyticus*. *Nucleic Acids Res* 24, 2760-2766.
26. Lutz, R., and Bujard, H. (1997). Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* 25, 1203-1210.
27. Manoil, C., and Bailey, J. (1997). A simple screen for permissive sites in proteins: analysis of *Escherichia coli* lac permease. *J Mol Biol* 267, 250-263.
28. Marko, M.A., Chipperfield, R., and Birnboim, H.C. (1982). A procedure for the large-scale isolation of highly purified plasmid DNA using alkaline extraction and binding to glass powder. *Analytical biochemistry* 121, 382-387.
29. Marmur, J. (1961). A procedure for the isolation of deoxyribonucleic acid from microorganisms. *Journal of Molecular Biology* 3, 208-218.
30. Mucke, M., Grelle, G., Behlke, J., Kraft, R., Kruger, D.H., and Reuter, M. (2002). EcoRII: a restriction enzyme evolving recombination functions? *Embo J* 21, 5262-5268.
31. Ostermeier, M., Nixon, A.E., Shim, J.H., and Benkovic, S.J. (1999). Combinatorial protein engineering by incremental truncation. *Proceedings of the National Academy of Sciences of the United States of America* 96, 3562-3567.

32. Ponting, C.P., and Kerr, I.D. (1996). A novel family of phospholipase D homologues that includes phospholipid synthases and putative endonucleases: identification of duplicated repeats and potential active site residues. *Protein Sci* 5, 914-922.
33. Remy, I., and Michnick, S.W. (1999). Clonal selection and in vivo quantitation of protein interactions with protein-fragment complementation assays. *Proceedings of the National Academy of Sciences of the United States of America* 96, 5394-5399.
34. 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.
35. Roberts, R.J., Vincze, T., Posfai, J., and Macelis, D. (2007). REBASE--enzymes and genes for DNA restriction and modification. *Nucleic Acids Res* 35, D269-270.
36. Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*, 2 edn (Cold Spring Harbor Laboratory).
37. Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* 74, 5463-5467.
38. Shiba, K., and Schimmel, P. (1992). Functional assembly of a randomly cleaved protein. *Proceedings of the National Academy of Sciences of the United States of America* 89, 1880-1884.
39. Szomolanyi, E., Kiss, A., and Venetianer, P. (1980). Cloning the modification methylase gene of *Bacillus sphaericus* R in *Escherichia coli*. *Gene* 10, 219-225.
40. Tasayco, M.L., and Carey, J. (1992). Ordered self-assembly of polypeptide fragments to form natively like dimeric trp repressor. *Science* 255, 594-597.
41. Timinskas, A., Butkus, V., and Janulaitis, A. (1995). Sequence motifs characteristic for DNA [cytosine-N4] and DNA [adenine-N6] methyltransferases. Classification of all DNA methyltransferases. *Gene* 157, 3-11.
42. Ullmann, A., Jacob, F., and Monod, J. (1967). Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the beta-galactosidase structural gene of *Escherichia coli*. *J Mol Biol* 24, 339-343.
43. Vilkaitis, G., Lubys, A., Merkiene, E., Timinskas, A., Janulaitis, A., and Klimasauskas, S. (2002). Circular permutation of DNA cytosine-N4 methyltransferases: in vivo coexistence in the BcnI system and in vitro probing by hybrid formation. *Nucleic Acids Res* 30, 1547-1557.
44. Vitkute, J., Maneliene, Z., Petrusyte, M., and Janulaitis, A. (1998). BfiI, a restriction endonuclease from *Bacillus firmus* S8120, which recognizes the novel non-palindromic sequence 5'-ACTGGG(N)5/4-3'. *Nucleic Acids Res* 26, 3348-3349.
45. Wah, D.A., Bitinaite, J., Schildkraut, I., and Aggarwal, A.K. (1998). Structure of FokI has implications for DNA cleavage. *Proceedings of the National Academy of Sciences of the United States of America* 95, 10564-10569.
46. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I., and Aggarwal, A.K. (1997). Structure of the multimodular endonuclease FokI bound to DNA. *Nature* 388, 97-100.
47. Way, J.C., Davis, M.A., Morisato, D., Roberts, D.E., and Kleckner, N. (1984). New Tn10 derivatives for transposon mutagenesis and for construction of lacZ operon fusions by transposition. *Gene* 32, 369-379.

48. Whitehead, P.R., and Brown, N.L. (1985). A simple and rapid method for screening bacteria for type II restriction endonucleases: enzymes in *Aphanothece halophytica*. *Archives of microbiology* *141*, 70-74.
49. Zaremba, M., Urbanke, C., Halford, S.E., and Siksnys, V. (2004). generation of the BfiI restriction endonuclease from the fusion of a DNA recognition domain to a non-specific nuclease from the phospholipase D superfamily. *J Mol Biol* *336*, 81-92.
50. Zhang, P., Bao, Y., Higgins, L., and Xu, S.Y. (2007). Rational design of a chimeric endonuclease targeted to NotI recognition site. *Protein Eng Des Sel* *20*, 497-504.