Department of

Protein-Nucleic Acids Interactions

Chief Scientist and Head

Prof. Virginijus Šikšnys, PhD

phone: 370 5 2691884 ; fax: 370 5 2602116

e-mail: siksnys@ibt.lt; virginijus.siksnys@bti.vu.lt

http://www.ibt.lt/en/laboratories/

Scientific staff

Saulius Gražulis, PhD
Giedrius Sasnauskas, PhD
Elena Manakova, PhD
Mindaugas Zaremba, PhD
Giedrė Tamulaitienė, PhD
Gintautas Tamulaitis, PhD
Arūnas Šilanskas, PhD
Dima Golovenko, PhD
(on a post doc leave)
Giedrius Gasiūnas, PhD
Marija Mantvyda Grušytė, M.Sc.

Postdoctoral associates

Lina Jakutytė-Giraitienė, PhD

PhD students

Georgij Kostiuk, M.Sc.
Tautvydas Karvelis, M.Sc.
Tomas Šinkūnas, M.Sc.
Paulius Toliušis, M.Sc.
Evelina Zagorskaitė, M.Sc.
Inga Songailienė, M.Sc.
Marius Rutkauskas,
(PhD student at Munster University since 2014)
Andrius Merkys, M.Sc.

Undegraduate Students

Miglė Kazlauskienė
Algirdas Mikšys
Justas Lavišius
Rokas Grigaitis
Irmantas Mogila
Greta Bigelytė
Kotryna Kauneckaitė
Šarūnas Tumas
Gediminas Drabavičius
Tomas Urbaitis
Antanas Vaitkus

Technician

Ana Tunevič

Research overview

Bacterial viruses (bacteriophages) provide a ubiquitous and deadly threat to bacterial populations. To survive in hostile environments, bacteria have developed a multitude of antiviral defense systems. The overall research theme in our department is the **structural and functional characterization of enzymes and enzyme assemblies that contribute to the bacteria defense systems that target invading nucleic acids.** In particularly, we are interested in the molecular machinery involved in the CRISPR function and structural and molecular mechanisms of restriction enzymes. We are using X-ray crystallography, mutagenesis, and functional biochemical and biophysical assays to gain information on these systems.

Structure and molecular mechanisms of CRISPR-Cas systems

CRISPR-Cas systems in bacteria and archea provide acquired immunity against viruses and plasmids. CRISPR (clustered regularly interspaced short palindromic repeats) locus usually consists of short and highly conserved DNA repeats that are interspaced by variable sequences, called spacers (Figure 1). CRISPR arrays are typically located in the direct vicinity of *cas* (CRISPR-associated) genes. In response to phage infection, some bacteria integrate new spacers that are derived from phage genomic sequences, which contribute to the CRISPR-mediated phage resistance in the subsequent rounds of infection. Many mechanistic steps involved in invasive element recognition, novel repeat manufacturing, and spacer selection and integration into the CRISPR locus remain uncharacterized (see below for the details).

Streptococcus thermophilus DGCC7710 contains four distinct systems: CRISPR1, CRISPR2, CRISPR3 and CRISPR4, which belong to the three distinct Types (Figure 2). We aim to establish molecular mechanisms of CRISPR-immunity provided by the CRISPR systems of *S. thermophilus.*

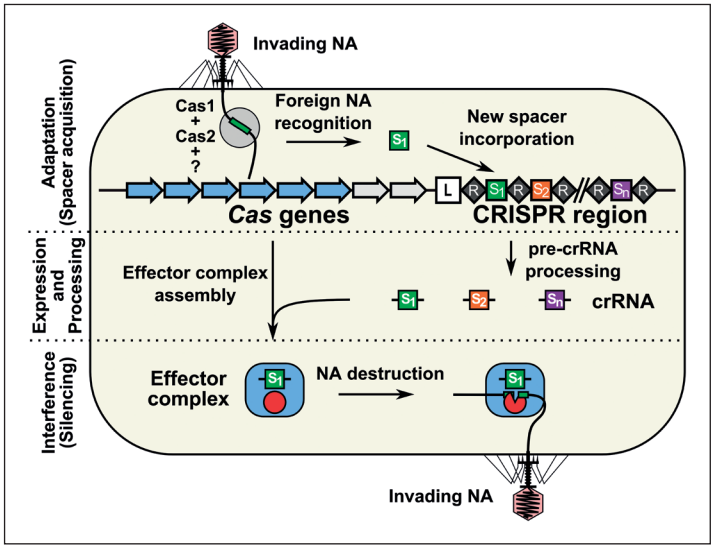

Figure 1. ***Schematic representation of CRISPR-Cas mechanism. The CRISPR–Cas mechanism is arbitrarily divided into three main stages: (1) adaptation or spacer acquisition, (2) expression and processing (crRNA generation), and (3) interference or silencing. During adaptation, Cas proteins recognize invasive nucleic acid (NA) and integrate short pieces of foreign DNA into the CRISPR region as new spacers (S1). Spacers are inserted at the leader (L) proximal end followed by duplication of the repeat (R). In the expression and processing stage, the CRISPR repeat-spacer array is transcribed into a long primary RNA transcript (pre-crRNA) that is further processed into a set of small crRNAs, containing a conserved repeat fragment and a variable spacer sequence (guide) complementary to the invading nucleic acid. crRNAs further combine with Cas proteins into an effector complex. In the interference or silencing stage, the effector complex recognizes the target sequence in the invasive nucleic acid by base pairing and induces sequence-specific cleavage, thereby preventing proliferation and propagation of foreign genetic elements.***

Cas9-dual RNA complex provides DNA silencing in the type II system

Type II CRISPR-Cas systems typically consist of only four Cas genes. We have shown that in the CRISPR3 system of *Streptococcus thermophilus* Cas9 protein associates with a dual crRNA:tracrRNA molecule to form an effector complex which specifically cleaves matching target dsDNA. The ternary Cas9-crRNA-tracrRNA complex of the type II CRISPR-Cas system functions as an Mg^{2+}-dependent RNA-directed DNA endonuclease that locates its DNA target guided by the crRNA and introduces a double-strand break at a specific site in DNA (Figure 2). The simple modular organization of Cas9, where specificity for the DNA target is encoded by a small crRNA and the cleavage reaction is executed by the Cas9 endonuclease, provides a versatile platform for the engineering of universal RNA-directed

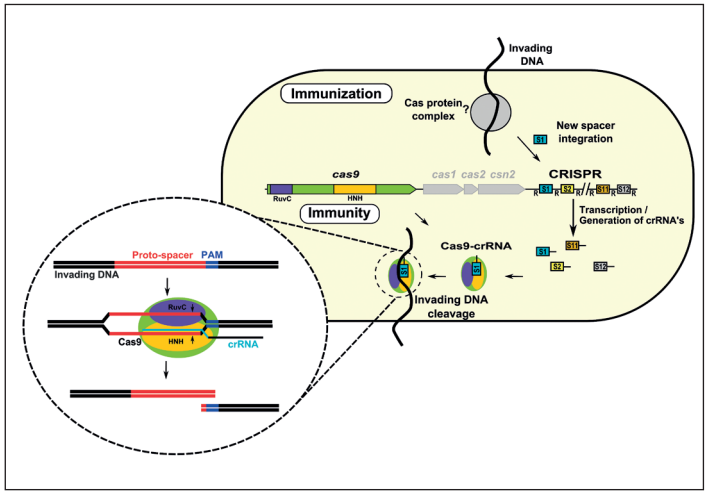

Figure 2. ***The Cas9-crRNA complex functions as an RNA-guided DNA endonuclease. Guided by the crRNA it finds a specific sequence in the target DNA and binds to it forming an R-loop. In the presence of Mg^{2+} ions, the Cas9 protein nicks each DNA strand 3 nt -upstream of the PAM sequence to generate blunt DNA ends, through RuvC- and HNH-like active sites that act on separate DNA strands.***

DNA endonucleases. Using *Streptococcus thermophilus* Cas9 as a model system, we demonstrated the feasibility of Cas9 as a programmable molecular tool for *in vitro* and *in vivo* DNA manipulations.

Cas9 for genome editing applications

Cas9 of *Streptococcus pyogenes* is currently used as a model system for genome editing applications. Typically, the DNA expression cassettes encoding nucleus-targeted codon-optimized Cas9 protein and sgRNAs are transfected into the cells. The efficiency of DNA cleavage by plasmid-delivered Cas9 in eukaryotic cells depends on multiple factors, including expression vector design, transfection efficiency, cell type, recovery yield of functional Cas9 complex, and usually requires optimization of a set of experimental conditions. Cas9 delivery by plasmid transfection is still difficult to achieve for some hard-to-transform cell lines including human primary cells and pluripotent stem cells. Moreover, plasmid transfection occasionally results in undesirable integration of vector plasmid into the genome and is often inefficient and stressful to cells. We developed an alternative way for the Cas9-mediated genome modification in eukaryotic cells by chemical transfection of *in vitro* reconstituted functionally active Cas9-crRNA-tracrRNA complex of *Streptococcus thermophilus* CRISPR3-Cas system.

Molecular basis for CRISPR immunity in type I systems

CRISPR-encoded immunity in type I systems relies on the Cascade ribonucleoprotein complex, which triggers foreign DNA degradation by an accessory Cas3 protein. To establish the mechanism for adaptive immunity provided by the *Streptococcus thermophilus* CRISPR4-Cas system (St-CRISPR4-Cas), we isolated an effector complex (St-Cascade) containing 61-nucleotide CRISPR RNA (crRNA). We show that St-Cascade, guided by crRNA, binds *in vitro* to a matching proto-spacer if a proto-spacer adjacent motif (PAM) is present. Surprisingly, the PAM sequence determined from binding analysis is promiscuous and limited to a single nucleotide (A or T) immediately upstream (-1 position) of the proto-spacer. In the presence of a correct PAM, St-Cascade binding to the target DNA generates an R-loop which serves as a landing site for the Cas3 ATPase/nuclease. We show that Cas3 binding to the displaced strand in the R-loop triggers DNA cleavage, and

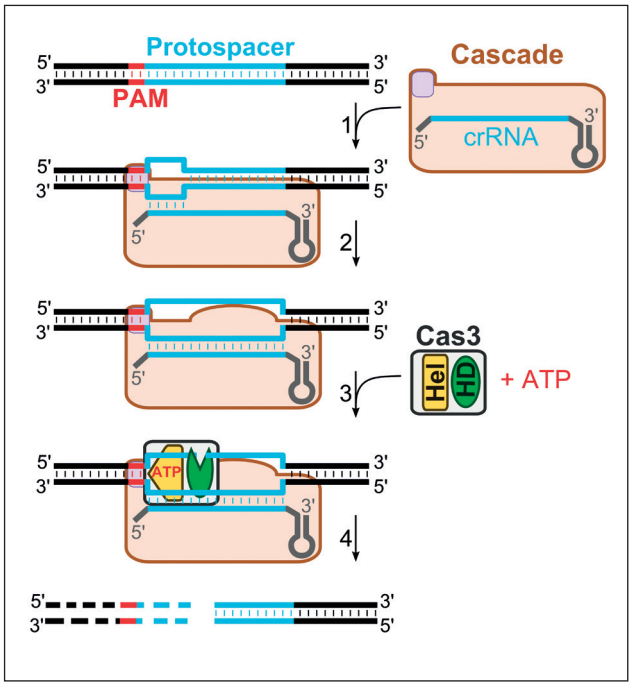

Figure 3. ***DNA-interference in the type I-E CRISPR-Cas systems. Cascade scans DNA for a protospacer sequence and PAM. Once the correct PAM and a short primary hybridization sequence ("seed") is identified (1), the guide of crRNA basepairs with a complementary DNA strand forming an R-loop, which is stabilized (locked) if a PAM-distal end of the protospacer is hybridized with the guide (2). Displaced DNA strand of the R-loop serves as a landing site for the Cas3 (3). Cas3 translocates in a 3'→5' direction powered by a helicase domain (Hel), whereas the HD domain degrades DNA in a unidirectional manner (4).***

if ATP is present, Cas3 further degrades DNA in a unidirectional manner. These findings establish a molecular basis for CRISPR immunity in St-CRISPR4-Cas and other Type I systems (Figure 3).

Single molecule observation of R-loop intermediate in the Cas9 and Cascade effector complex

Central to the defense against invading foreign DNA provided by the CRISPR-Cas systems is a ribonucleoprotein complex that produces RNA-guided cleavage of foreign nucleic acids. In DNA-targeting CRISPR-Cas systems, the RNA component of the complex encodes target recognition by forming a site-specific hybrid (R-loop) with its complement (protospacer) on an invading DNA while displacing the non-complementary strand. Subsequently, the R-loop structure triggers DNA degradation. Although these reactions have been reconstituted, the exact mechanism of R-loop formation has not been fully resolved. We used single-molecule DNA supercoiling to directly observe and quantify the dynamics of torque-dependent R-loop formation and dissociation for both Cascade- and Cas9-based CRISPR-Cas systems. We found that the protospacer adjacent motif (PAM) affects primarily the R-loop association rates, whereas protospacer elements distal to the PAM affect primarily R-loop stability. Furthermore, Cascade has higher torque stability than Cas9 by using a conformational locking step. Our data provide direct evidence for directional R-loop formation, starting from PAM recognition and expanding toward the distal protospacer end.

Type III-A CRISPR2 system of Streptococcus thermophilus targets RNA

Immunity against viruses and plasmids provided by CRISPR-Cas systems relies on a ribonucleoprotein effector complex that triggers the degradation of invasive nucleic acids (NA). Effector complexes of type I (Cascade) and II (Cas9-dual RNA) target foreign DNA. Intriguingly, the genetic evidence suggests that the type III-A Csm complex targets DNA, whereas biochemical data show that the type III-B Cmr complex cleaves RNA. We aimed to investigate NA specificity

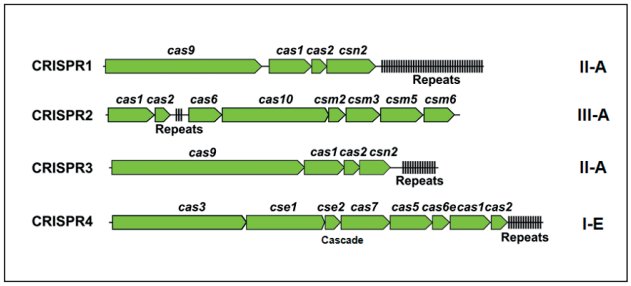

***Figure 4.** CRISPR-Cas systems of S. thermophilus DGCC7710. CRISPR1 and CRISPR3 systems belong to the type II, CRISPR2 to the type III whilst CRISPR4 belongs to the type I (E. coli subtype).*

and mechanism of CRISPR interference for the *Streptococcus thermophilus* Csm (III-A) complex (StCsm). When expressed in *Escherichia coli*, two complexes of different stoichiometry copurified with 40 and 72 nt crRNA species, respectively. Both complexes targeted RNA and generated multiple cuts at 6 nt intervals. The Csm3 protein, present in multiple copies in both Csm complexes, acts as endoribonuclease. In the heterologous *E. coli* host, StCsm restricts MS2 RNA phage in a Csm3 nuclease-dependent manner. Thus, our results demonstrate that the type III-A StCsm complex guided by crRNA targets RNA.

Structure and function of restriction endonucleases

Restriction-modification (RM) systems commonly act as sentries that guard bacterial cells against invasion by bacteriophage. RM systems typically consist of two complementary enzymatic activities, namely restriction endonuclease (REase) and methyltransferase (MTase). In typical RM systems REase cuts foreign DNA but does not act on the host genome because target sites for REase are methylated by accompanying MTase. REases from 4000 bacteria species with nearly 350 distinct specificities have been characterised. REases have now gained widespread application as indispensable tools for the *in vitro* manipulation and cloning of DNA. However, much less is known about how they achieve their function.
In the Department of Protein-Nucleic acids Interactions we focus on the structural and molecular mechanisms of restriction enzymes. Among the questions being asked are: How do the restriction enzymes recognize the particular DNA sequence? What common structural principles exist among restriction enzymes that recognize related nucleotide sequenc-

es? How do the sequence recognition and catalysis are coupled in the function of restriction enzymes? Answers to these questions are being sought using X-ray crystal structure determination of restriction enzyme-DNA complexes, site-directed mutageneses and biochemical studies to relate structure to function (see below for the details).

NTP-dependent restriction enzymes

The stress-sensitive RM system CglI from *Corynebacterium glutamicum* and the homologous NgoAVII RM system from *Neisseria gonorrhoeae* FA1090 are composed of three genes: a DNA methyltransferase (M.CglI and M.NgoAVII), a putative restriction endonuclease (R.CglI and R.NgoAVII, or R-proteins) and a predicted DEAD-family helicase/ATPase (H.CglI and H.NgoAVII or H-proteins). Size-exclusion chromatography and SAXS experiments revealed that the isolated R.CglI, R.NgoAVII and H.CglI proteins form homodimers, while H.NgoAVII is a monomer in solution. Moreover, the R.CglI and H.CglI proteins assembled in a complex with R2H2 stoichiometry. Next, we showed that H-proteins have ATPase activity that is dependent on double-stranded DNA and is stimulated by the R-proteins. Functional ATPase activity and extensive ATP hydrolysis (~170 ATP/s/monomer) are required for site-specific DNA cleavage by R-proteins. We further showed that ATP-dependent DNA cleavage by R-proteins occurs at fixed positions (6-7 nucleotides) downstream of the asymmetric recognition sequence 5'-GCCGC-3'. Despite similarities to both Type I and II restriction endonucleases, the CglI and NgoAVII enzymes may employ a unique catalytic mechanism for DNA cleavage.

To establish the molecular mechanism of the sequence recognition by NTP-dependent restriction enzymes we have solved crystal structures of the R.NgoAVII apo-protein and the R.NgoAVII C-terminal domain bound to a specific DNA (Figure 5). R.NgoAVII is composed of two domains: an N-terminal nucleolytic PLD domain; and a C-terminal B3-like DNA-binding domain identified previously in BfiI and EcoRII REases, and in plant transcription factors.

Structural comparison of the B3-like domains of R.NgoAVII, EcoRII, BfiI and the plant transcription factors revealed a conserved DNA-binding surface comprised of N- and C-arms that together grip the DNA. The C-arms of R.NgoAVII, EcoRII, BfiI and plant B3 domains are similar in size, but

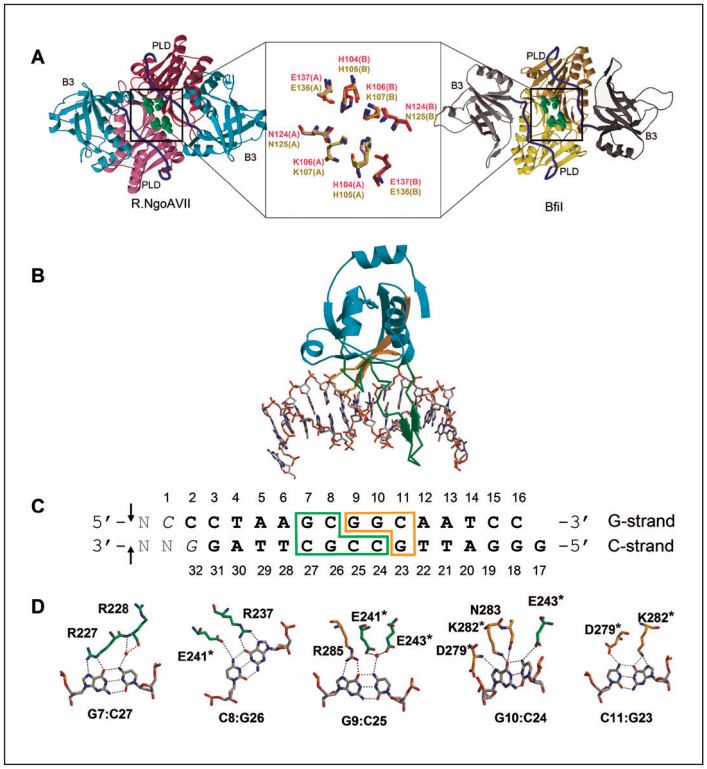

Figure 5. Crystal structure and DNA recognition by R.NgoAVII-B3. (A) Structure of R.NgoAVII. The R.NgoAVII N-terminal PLD domains (pink) form a dimer with a single active site (active site residues H104 and K106 are depicted in green). The C-terminal B3-like domains (cyan) are connected to the catalytic domains by long linkers (blue) and are positioned on both sides of a dimeric core. The R.NgoAVII dimer is similar to BfiI (PDB ID 2C1L). (B) Overall structure of the B3-like domain in DNA-bound form. The N-arm is colored green and the C-arm is colored orange. (C) The cognate oligoduplex used in co-crystallization. (D) Recognition of the individual base pairs by R.NgoAVII-B3. Residues from the N-arm are colored green and residues from the C-arm are colored orange. Residues involved in the recognition of more than one base pair are marked by asterisk.

the R.NgoAVII N-arm which makes the majority of the contacts to the target site is much longer. The overall structures of R.NgoAVII and BfiI are similar; however, whilst BfiI has stand-alone catalytic activity, R.NgoAVII requires an auxiliary cognate N.NgoAVII protein and ATP hydrolysis in order to cleave DNA at the target site. The structures will help formulate future experiments to explore the molecular mechanisms of intersubunit crosstalk that control DNA cleavage by R.NgoAVII and related endonucleases.

B3-like DNA binding domain of BfiI restriction enzyme

The B3 DNA-binding domains (DBDs) of plant transcription factors (TF) and DBDs of EcoRII and BfiI restriction endonucleases (EcoRII-N and BfiI-C) share a common structural fold, classified as the DNA-binding pseudobarrel. The B3 DBDs in the plant TFs recognize a diverse set of target sequences. The only available co-crystal structure of the B3-like DBD is that of EcoRII-N (recognition sequence 5'-CCTGG-3'). In order to understand the structural and molecular mechanisms of specificity of B3 DBDs, we have solved the crystal structure of BfiI-C (recognition sequence 5'-ACTGGG-3') complexed with 12-bp cognate oligoduplex (Figure 6). Structural comparison of BfiI-C-DNA and EcoRII-N-DNA complexes reveals a conserved DNA-binding mode and a conserved pattern of interactions with the phosphodiester backbone. The determinants of the target specificity are located in the loops that emanate from the conserved structural core. The BfiI-C-DNA structure presented here expands a range of templates for modeling of the DNA-bound complexes of the B3 family of plant TFs.

Structure and mechanism of modification-dependent restriction enzymes

Unlike conventional restriction endonucleases (REases) that recognize and cleave unmodified DNA sequences, cytosine modification-dependent REases recognize DNA sites containing 5-methylcytosine (5mC), 5-hydroxymethylcytosine (5hmC) or glucosylated cytosine. To date, several families of cytosine modification-dependent restriction endonucleases (REases) were identified, and some of them were employed as molecular tools for the epigenetic profiling of cytosine modifications in eukaryotic DNA. The MspJI family of modification-dependent REases recognize 5mC and 5hmC in various sequence contexts, and cleave DNA at a fixed distance from the modified cytosine. We aim to understand the structural and molecular mechanisms of MspJI family REases.

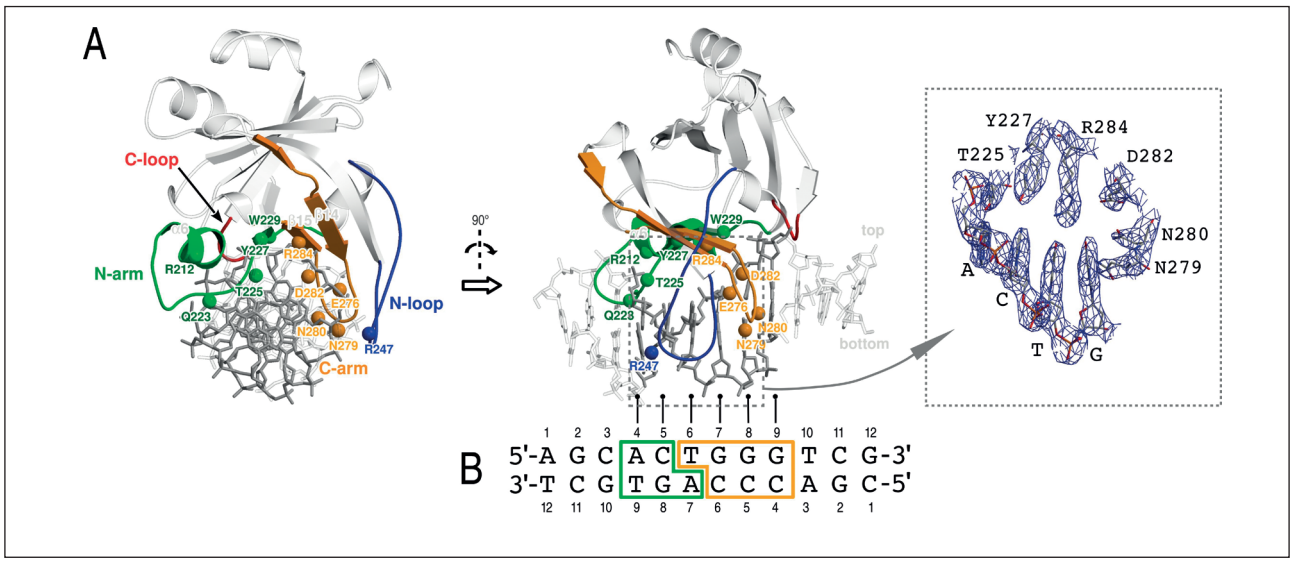

Figure 6. DNA recognition by BfiI-C. (A) The view of the BfiI-C–DNA complex along the long DNA axis (left) and the side view (right). The DNA-recognition site is colored dark grey. Spheres represent the Cα atoms of the DNA-recognition residues from the N- and C-arms. The N-loop is colored blue and the C-loop is red. A region of the top DNA strand (nucleotides A4-G7) and adjacent recognition residues are shown against their mFO-DFC SIGMAA-weighted-electron density contoured at 2.0 σ level. (B) The sequence and numbering of the cognate 12/12 oligoduplex used in this study. DNA bases that interact with the N- and C-arms are boxed in green and orange, respectively.

Collaboration

Dr. Philippe Horvath, DuPont, France
Prof. Dr. Mark Sczcelkun, Bristol University, United Kingdom
Prof. Dr. Ralf Seidel, Technische Universitat Dresden/Munster University
Dr. Mark Dickman, Sheffield University, United Kingdom
Dr. Rodolphe Barrangou, DuPont/NC State University, USA

Funding

European Social Fund under the Global Grant Measure
Research Council of Lithuania
Agency for Science, Innovation and Technology

Publications 2013-2014

1. **Tamulaitis G., Kazlauskiene M., Manakova E., Venclovas Č.**, Nwokeoji A.O, Dickman M.J., Horvath P. and **Siksnys V.** Programmable RNA shredding by the Type III-A CRIPSR-Cas system of Streptococcus thermophilus. Mol Cell 2014, 56(4):506-17.

2. Glemzaite M., Balciunaite E., **Karvelis T., Gasiunas G.,** Alzbutas G., Jurcyte A., Anderson E.M., Maksimova M., SmithA.J., Lubys A., Zaliauskiene L., **Siksnys V.** Targeted gene editing by transfection of in vitro reconstituted Streptococcus thermophilus Cas9 nuclease complex, RNA Biology, 2014, in press

3. Szczelkun M.D., Tikhomirova M.S., **Sinkunas T., Gasiunas G., Karvelis T.,** Pschera P., **Siksnys V.,** and Seidel R. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proc Natl Acad Sci USA 2014, 111(27):9798-803.

4. **Zaremba M., Toliusis P., Grigaitis R., Manakova E., Silanskas A., Tamulaitiene G.**, Szczelkun M.D., **Siksnys V.** DNA cleavage by CgII and NgoAVII requires interaction between N- and R-proteins and extensive nucleotide hydrolysis. Nucleic Acids Res 2014, 42(22):13887-13896.

5. **Tamulaitiene G., Silanskas A., Grazulis S., Zaremba M., Siksnys V.** Crystal structure of the R-protein of the multisubunit ATP-dependent restriction endonuclease NgoAVII. Nucleic Acids Res 2014, 42(22):14022-14030.

6. **Golovenko D., Manakova E., Zakrys L., Zaremba M., Sasnauskas G., Grazulis S., Siksnys V.** Structural insight into the specificity of the B3 DNA-binding domains provided by the co-crystal structure of the C-terminal fragment of BfiI restriction enzyme. Nucleic Acids Res 2014, 42(6):4113-4122.

7. **Zagorskaitė E., Sasnauskas G.** Chemical Display of Pyrimidine Bases Flipped Out by Modification-Dependent Restriction Endonucleases of MspJI and PvuRts1I Families. PLoS One 2014, 9(12):e114580.

8. Rutkauskas D., Petkelyte M., Naujalis P., **Sasnauskas G., Tamulaitis G., Zaremba M., Siksnys V.** Restriction Enzyme Ecl18kI-Induced DNA Looping Dynamics by Single-Molecule FRET. J Phys Chem B 2014, 118(29):8575-82.

9. Horvath P., **Gasiunas G., Siksnys V.** and Barrangou R. Applications of the Versatile CRISPR-Cas Systems //CRISPR-Cas Systems. RNA-mediated adaptive immunity in Bacteria and Archea. R. Editors: Barrangou and J. van der Oost. Springer-Verlag Berlin Heidelberg 2013, 267-286.

10. **Sinkunas T., Gasiunas G.,** Waghmare S.P., Dickman M.J., Barrangou R., Horvath P., **Siksnys V.** In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. EMBO J 2013, 32(3):385-94.

11. **Karvelis T., Gasiunas G., Miksys A.,** Barrangou R., Horvath P., **Siksnys V.** crRNA and tracrRNA guide Cas9-mediated DNA interference in Streptococcus thermophilus. RNA Biol 2013, 10(5):841-51.

12. **Gasiunas G., Sinkunas T., Siksnys V.** Molecular mechanisms of CRISPR-mediated microbial immunity. Cell Mol Life Sci 2013, 71(3):449-65.

13. **Gasiunas G., Siksnys V.** RNA-dependent DNA endonuclease Cas9 of the CRISPR system: Holy Grail of genome editing? Trends Microbiol 2013, 21(11):562-7.

14. **Karvelis T., Gasiunas G., Siksnys V.** Programmable DNA cleavage in vitro by Cas9. Biochem Soc Trans 2013, 41(6):1401-6.

PATENT APPLICATIONS

1. **Siksnys V., Gasiunas G., Karvelis T.,** Lubys A., Zaliauskiene L., Glemzaite M., Smith A. RNA-directed DNA cleavage by the Cas9-crRNA complex from CRISPR3/Cas immune system of Streptococcus thermophilus. WO/2013/142578, PCT/US2013/033106. 2013/03/20

2. **Šikšnys V., Gasiūnas G., Karvelis T.** RNA-directed DNA cleavage by the Cas9-crRNA complex from CRISPR3/Cas immune system of Streptococcus thermophilus. WO/2013/141680, PCT/LT2013/000006. 2013/03/15

3. **Siksnys V., Kazlauskiene M., Tamulaitis G.** Programmable RNA shredding by the Type III-A CRSIPR-Cas system of Streptococcus thermophilus. U.S. Patent Application Serial No. 62/046, 384.