



Skolkovo Institute of Science and Technology

RECOGNITION STRATEGIES OF TYPE I AND TYPE V BREX SYSTEMS  
*Doctoral Thesis*

by

JULIA GORDEEVA

DOCTORAL PROGRAM IN LIFE SCIENCES

Supervisor  
Professor Konstantin Severinov

Moscow - 2022

© Julia Gordeeva 2022

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Julia Gordeeva)

Supervisor (Prof. Konstantin Severinov)

## Abstract

A strong pressure of viruses on prokaryotes leads to the development of adaptations such as, for example, the emergence of defense systems. Bacteria and archaea encode a variety of immune mechanisms and use this arsenal to combat viruses at each stage of infection. Most defense systems are clustered to special genome regions forming ‘defense islands’. In addition to well-known restriction-modification, toxin-antitoxin, and CRISPR-Cas immunity systems, many other, poorly characterized, systems were found in the defense islands.

One such family, denoted BREX, for *bacteriophage exclusion*, comprises systems with two core genes encoding a putative alkaline phosphatase PglZ/BrxZ and a large protein PglY/BrxC with an ATP-binding domain. Depending on the composition of other *brx* genes, BREX family was divided into six types. The most abundant and studied BREX systems are BREX systems Type I.

This thesis is devoted to an investigation and comparison of non-self DNA recognition strategies by two BREX system types from different organisms: a Type I BREX system from a natural isolate *Escherichia coli* HS (BREX<sup>Ec</sup>) and a Type V BREX system from an archaeon *Haloarcula hispanica* (BREX<sup>HAR</sup>).

In addition to core genes, BREX<sup>Ec</sup> encodes a methyltransferase PglX/BrxX, a predicted Lon-like protease BrxL and two proteins BrxA and BrxB with unknown functions. We showed that BREX<sup>Ec</sup>, when expressed from plasmids, makes *E. coli* cells resistant to phage  $\lambda$  infection. Induction of  $\lambda$  prophages from cells with BREX<sup>Ec</sup> leads to production of viral particles that are no longer subject to exclusion by BREX due to methylation of an adenine residue in an asymmetric 6-nt motif. The same modification is also found in the genome of BREX<sup>Ec</sup> cells. The results establish, for the first time, that immunity to BREX system defense is provided by epigenetic modification. While the nature of BREX defense remains to be established, our results suggest that BREX<sup>Ec</sup> system works at the earliest stages of infection by targeting unmodified foreign DNA (1).

Judging by live fluorescent microscopy and sequencing of total DNA extracted from infected cultures, unmodified phage DNA is very rapidly degraded in BREX<sup>Ec</sup> cells.

Unlike BREX<sup>Ec</sup>, BREX<sup>HAR</sup> encodes two methyltransferases BrxX1 and BrxX2, two ATPases BrxC1 and BrxC2, and a putative helicase BrxHI instead of a protease BrxL. Our work revealed that BREX<sup>HAR</sup> provides protection against archaeal viruses. The BREX<sup>HAR</sup> mechanism of self/non-self discrimination is mediated by methylation of a cytosine residues in two non-palindromic motifs. We suggest that each Brx methyltransferase modifies its own *brx* site: BrxX1 methylates GTAYCCG while BrxX2 methylates GACCC. Similar to BREX<sup>Ec</sup>, BREX<sup>HAR</sup> inhibits production of viral progeny in infected cells.

## Publications

1. **Gordeeva J**, Morozova N, Sierro N, Isaev A, Sinkunas T, Tsvetkova K, Matlashov M, Truncaite L, Morgan RD, Ivanov NV, Siksnys V, Zeng L, Severinov K. BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Research*. 2019 Jan 10;47(1):253-265.
2. Isaev A, Drobiazko A, Sierro N, **Gordeeva J**, Yosef I, Qimron U, Ivanov NV, Severinov K. Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. *Nucleic Acids Research*. 2020 Jun 4;48(10):5397-5406.
3. Fraser A, Sokolova M, Drobysheva A, **Gordeeva J**, Borukhov S, Jumper J, Severinov K, Leiman P. Structural basis of template strand deoxyuridine promoter recognition by a viral RNA polymerase. *Nature communications*, 2022 Jun 20;13(1):3526.

## Conferences

1. **Gordeeva J**, Morozova N, Sierro N, Isaev A, Ivanov N, Zeng L, Severinov K. BREX is a new prokaryotic defense system. *Global Young Scientists Summit 2020*, Singapore, 2020 - poster presentation
2. **Gordeeva J**, Fraser A, Sokolova M, Drobysheva A, Borukhov S, AlphaFold team, Severinov K, Leiman P. Promoter recognition by multisubunit RNA polymerase of Jumbo phage AR9. *33<sup>d</sup> RNA Polymerase Workshop*, Newcastle, 2022 – oral presentation

## Acknowledgements

I would like to thank my supervisor Professor Konstantin Severinov who gave me opportunities to work on an interesting project in a high-level laboratory with top students and to travel worldwide to learn new techniques. He was very supportive and showed an image of a scientist to strive for. Besides participation in my scientific career, Konstantin Viktorovich also helped to meet my husband by encouraging me to give an interview for a journal Schrödinger's cat.

Furthermore, I would like to acknowledge everyone from the laboratory of molecular genetics of microorganisms at Institute of Gene Biology, Moscow, for fruitful seminars and kind atmosphere in the lab. Especially, I wish to thank Svetlana Dubiley, who looks after all lab members, Dmitry Sutormin for his support with data analysis and interesting discussions, and our technician Valentina, who performed very important work in the lab. I thank my colleagues Artem Isaev and Alena Drodiazko who also work on BREX and contributed to the current understanding of this system. I would like to thank Evgeny Klimuk for his invaluable advice and help with protein purification. I thank Natalia Morozova from Peter the Great St. Petersburg State Polytechnic University who spent a lot of time monitoring cells fate under microscope.

I wish to thank my Individual Committee Members Professor Sergei Borukhov and Professor Konstantin Lukyanov for their interest to my work and useful recommendations.

I need to thank Skoltech for the possibility to participate in international conferences and for supporting long-term mobility. I thank Professor Lanying Zeng for hosting me at TAMU (College Station, USA) and giving me opportunity to perform live microscopy experiments. I wish to thank Philip Morris International company (Neuchâtel, Switzerland) and Dr. Nicolas Sierro and Dr. Nikolai Ivanov for PacBio sequencing of BREX<sup>Ec</sup> samples. I thank Professor Hua Xiang for hosting me at Microbiology Institute, Chinese Academy of Science (Beijing, China) and teaching me methods how to

work with archaea. I would like to thank Dr. Tim Blower, Durham University (Durham, UK) for very warm welcome at his lab during difficult time in February-April 2022.

I would like to express my gratitude and appreciation for Professor Petr Leiman and my friend Maria Sokolova for an opportunity to participate in another interesting project about the structure of an unusual phage RNA polymerase. Years of Petr and Masha's hard work resulted in an excellent Nature Communication paper.

I thank Skoltech Genomics Core facility and Maria Logacheva for sequencing of my samples that became a part of this thesis.

Although my PhD was interrupted by COVID-19 lockdown and affected by Russian special military operation in Ukraine, I was supported by my family, friends and colleagues. I thank all other people who somehow participated in this work.

## Table of Contents

Abstract.....	3
Publications.....	5
Conferences.....	5
Acknowledgements.....	6
Table of Contents.....	8
List of Symbols, Abbreviations .....	10
List of Figures .....	12
List of Tables .....	13
Introduction.....	14
<b>Chapter 1. Review of the Literature .....</b>	<b>15</b>
1.1 Receptor recognition avoidance and inhibition of phage DNA injection.....	15
1.2 DNA modification systems.....	17
1.2.1 Classical Restriction-modification systems .....	17
1.2.2 DISARM .....	20
1.2.3 DPD systems.....	21
1.2.4 Phosphorothioation-based systems .....	21
1.2.5 Pgl system .....	22
1.2.6 BREX systems .....	25
1.3 CRISPR-Cas systems.....	30
1.4 Abortive infection and toxin-antitoxin systems .....	33
1.4.1 Abortive infection systems .....	33
1.4.2 TA systems.....	35
1.4.3 Retrons .....	37
1.4.4 CBASS.....	37
1.5 Novel defense systems.....	38
<b>Chapter 2. Materials and methods .....</b>	<b>41</b>
2.1 Strains and plasmids .....	41
2.2 Media and buffers .....	42
2.3 Cloning of BREX constructs .....	43
2.4 Plasmid transformation of <i>E. coli</i> cells.....	45
2.5 Plasmid transformation of <i>H. hispanica</i> cells.....	45
2.6 Knock-out of <i>brx</i> genes in <i>H. hispanica</i> .....	46
2.7 DNA purification .....	46
2.8 Efficiency of plaquing (EOP) assay.....	47
2.9 Phage and virus propagation.....	48
2.10 Growth curves of <i>E. coli</i> infected cultures .....	48
2.11 Adsorption assay for <i>E. coli</i> cells .....	48
2.12 Lysogenization assay for <i>E. coli</i> cells.....	49
2.13 Fluorescence microscopy for <i>E. coli</i> cells .....	49
2.14 Pacific Bioscience sequencing.....	50
2.15 Phage lambda DNA degradation experiment .....	51
2.16. Restriction endonuclease activity assay.....	51

<b>Chapter 3. Results</b> .....	52
3.1 The <i>Escherichia coli</i> HS BREX system provides defense against phage $\lambda$ .....	52
3.2 BREX <sup>Ec</sup> has no effect on lysogenic induction .....	56
3.3 Phage that overcomes BREX <sup>Ec</sup> protection contains modified DNA .....	58
3.4 The role of individual <i>brx</i> genes in BREX <sup>Ec</sup> protection and DNA modification ...	61
3.5 Evidence of phage DNA degradation in BREX <sup>Ec+</sup> cells .....	65
3.6 BREX <sup>Ec</sup> protection is circumvented by glycosylation of phage DNA .....	67
3.7 BREX <sup>HAR</sup> system Type V from <i>Haloarcula hispanica</i> .....	69
3.8 The role of individual <i>brx</i> genes on modification and recognition module.....	72
Conclusions.....	75
Discussion.....	76
Bibliography .....	80
Appendix A.....	87
Appendix B.....	92

## List of Symbols, Abbreviations

A – adenine

Abi – Abortive infection

ATP – adenosine triphosphate

ATPase - Adenosylpyrophosphatase

BREX – Bacteriophage Exclusion system

bp – base pairs

C – cytosine

CBASS – Cyclic-Oligonucleotide-Based Anti-Phage Signaling System

CFU - Colony Forming Unit

CRISPR - Clustered Regularly Interspaced Short Palindromic Repeats

DISARM – Defence Island System Associated with Restriction-Modification

DNA – deoxyribonucleic acid

dsDNA – double-stranded DNA

dNTPs – deoxynucleotides

*E. coli* – *Escherichia coli*

EOP – Efficiency of Plaquing

5'-FOA - 5'-Fluoroorotic acid

G – guanine

*H. hispanica* – *Haloarcula hispanica*

kDa – kiloDalton

LB - Luria-Bertani broth

LPS - Lipopolysacharide

MOI – Multiplicity of infection

Ocr – Overcoming classical restriction protein

OD – optical density

PAGE - polyacrylamide gel electrophoresis

pAgo – prokaryotic Argonaute

PAM - Protospacer Adjacent Motif  
PCR – polymerase chain reaction  
Pgl – Phage growth limitation  
PEG – polyethylene glycol  
PFU – Plaque Forming Unit  
R-M – Restriction – Modification system  
RT – room temperature  
SAM - S-adenosyl methionine  
SDS – sodium dodecyl sulfate  
Sie – Superinfection exclusion  
ssDNA – single-stranded DNA  
T – thymine  
TA – Toxin – Antitoxin  
w/v – weight/volume

## List of Figures

Figure 1. Prokaryotic adsorption-blocking mechanisms .....	16
Figure 2. Mechanisms for modification and cleavage by four groups of R-M systems .....	18
Figure 3. Schematics of the Pgl phenotype .....	23
Figure 4. Schematics of the Pgl mechanism .....	24
Figure 5. PglZ phylogeny and classification of BREX subtypes .....	26
Figure 6. The BREX cassette type I from <i>Bacillus cereus</i> .....	26
Figure 7. Schematics of pEFER-mediated defense against diverse phages .....	29
Figure 8. The mechanism for recognition and elimination of the target by CRISPR-Cas ...	31
Figure 9. Mechanisms of Abi and TA systems .....	34
Figure 10. Current division of TA systems into types .....	36
Figure 11. BREX <sup>Ec</sup> protects cells from phage $\lambda$ infection .....	53
Figure 12. BREX <sup>Ec</sup> effects on lysogenization, adsorption, and DNA injection by phage $\lambda$ .	55
Figure 13. Induction of $\lambda$ prophage from BREX+ lysogens .....	56
Figure 14. Phage $\lambda$ induced from BREX+ lysogens is not subject to inhibitory action of BREX <sup>Ec</sup> .....	57
Figure 15. BREX system epigenetically modifies phage $\lambda$ .....	59
Figure 16. Modified phage appears in the course of infection of BREX+ cells and overcomes protection .....	60
Figure 17. Deletions of individual <i>brx</i> genes on pBREXAL plasmid .....	62
Figure 18. The state of methylation of host DNA and ability to withstand phage infection by cells carrying BREX plasmids lacking indicated genes .....	63
Figure 19. BrxL inhibits FtsZ-ring formation and induces SOS-response .....	65
Figure 20. BREX <sup>Ec</sup> system degrades phage DNA <i>in vivo</i> .....	66
Figure 21. Ability of BREX <sup>Ec</sup> to protect cells from infections other than $\lambda$ .....	68
Figure 22. Comparison of <i>brx</i> gene compositions from BREX <sup>HAR</sup> system Type V from <i>Haloarcula hispanica</i> and BREX <sup>Ec</sup> system Type I from <i>Escherichia coli</i> HS .....	69
Figure 23. BREX <sup>HAR</sup> defense activity .....	70
Figure 24. Dynamics of <i>H. hispanica</i> carrying BREX <sup>HAR</sup> infected with HHPV3 .....	71
Figure 25. Schematics of the BREX <sup>Ec</sup> function .....	77
Figure 26. Schematics of the BREX <sup>HAR</sup> function .....	79

## List of Tables

Table 1. Main characteristics of different R-M systems.....	19
Table 2. BREX protection against a wide range of Bacillus phages .....	27
Table 3. Results of PacBio sequencing of <i>Haloarcula hispanica</i> strains.....	72
Table 4. Defense and methylation activities of <i>H. hispanica</i> stains with individual deletions of <i>brx</i> genes. ....	73
Table 5. The number of <i>brx</i> sites in <i>H. hispanica</i> viruses .....	74
Table 6. BREX <sup>Ec</sup> and BREX <sup>HAR</sup> comparison .....	75

## Introduction

*The Introduction section is modified from Master thesis prepared by the author: Julia Gordeeva (2018). Mechanism of defense function of BREX system from Escherichia coli. Master thesis, Skoltech.*

Bacterial viruses – bacteriophages or phages for short – are thought to outnumber their bacterial hosts and thus present a source of constant and immediate danger that can wipe out bacterial populations. Indeed, analysis of oceanic communities demonstrates that massive phage-driven sweeps in bacterial communities occur at large scale and force important environmental processes (2, 3). The host-virus antagonism mediated by cellular defense mechanisms that evolved to prevent total annihilation of the host by viruses is discussed in the Chapter 1 of the thesis. Known mechanisms include innate immunity conferred by restriction-modification systems (4, 5), adaptive immunity mediated by CRISPR-Cas systems (6–8), and an assortment of poorly studied suicidal abortive infection mechanisms that often involve toxin-antitoxin interactions and limit phage propagation through the population while killing infected cells (9–11). Considering the variety of virul attack strategies and the fact that defense islands may constitute up to 10% of prokaryotic genomes (12), there is little doubt that many host resistance systems remain to be uncovered and recent bioinformatics efforts support this view (13). Sequence analysis of defense islands present in many prokaryotic genomes revealed overrepresentation of genes encoding proteins homologous to a predicted alkaline phosphatase PglZ (14). A census done in 2014 led the authors to propose a separate BREX (for BacteRiophage EXclusion) family for clusters containing *pglZ*-like genes and several other conserved neighboring genes (15). Based on the composition of *brx* genes, BREX family was divided into six types.

We tried to shed light on the defense mechanisms of a Type I BREX system from a bacterium *Escherichia coli* HS (BREX<sup>Ec</sup>) and a Type V BREX system from an archaeon *Haloarcula hispanica* (BREX<sup>HAR</sup>). We deciphered how these two systems distinguish between host and viral DNA and determined the *brx*-mediated modifications. Moreover, we studied effects of *brx* genes deletions on the function of methylation and

defensive modules of each BREX system studied. For BREX<sup>Ec</sup>, we managed to establish the stage of phage infection that is blocked by the defense.

## Chapter 1. Review of the Literature

*Review of the Literature section is based on Master thesis prepared by the author extended by a review of new systems discovered during 2018-2022 years:*

*Julia Gordeeva (2018). Mechanism of defense function of BREX system from Escherichia coli. Master thesis, Skoltech.*

A strong selective pressure on prokaryotes leads to the development of adaptations such as the emergence of antiviral systems. In turn, viruses counter-adapt and evolve strategies that by-pass the host defense (16). Studies of new prokaryotic defense systems are highly significant and over the years provided us with tools for molecular cloning and genomic engineering that underlie most of the recent progress in biotechnology and biomedicine.

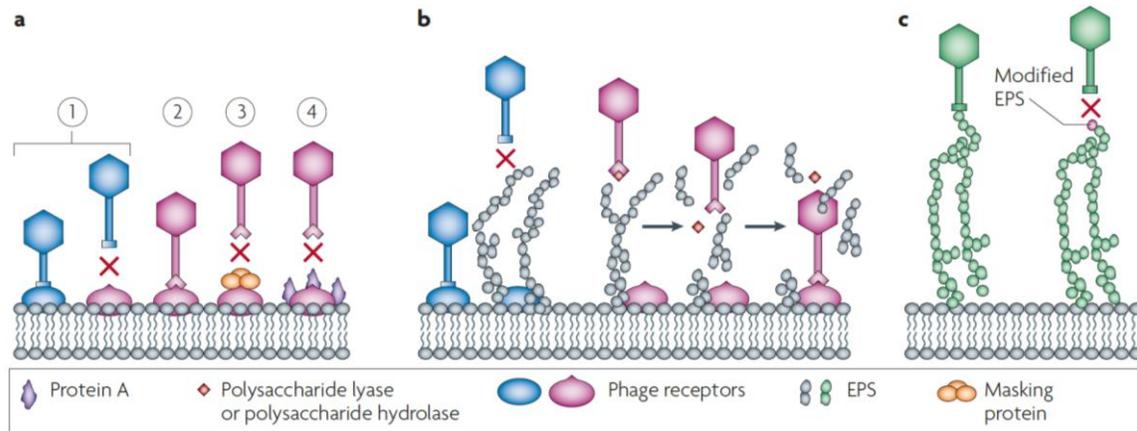
The literature review is focused on the main defense mechanisms of prokaryotes such as Restriction-Modification, CRISPR-Cas, various Abi systems, as well as newly discovered antiviral mechanisms like CBASS, DISARM, etc.

### 1.1 Receptor recognition avoidance and inhibition of phage DNA injection

To initiate the infection process, phages recognize specific receptors on bacterial cell surface, such as polysaccharides, lipopolysaccharides, and surface proteins. Bacteria can prevent phage adsorption via blocking of phage receptors by mutations, masking, or modifications, production of extracellular matrix or competitive inhibitors (Fig. 1) (17).

Since some bacterial receptors are nonessential, mutations in these genes do not affect bacterial fitness but prevent phage attachment. For instance, mutations in the outer-membrane protein A (OmpA), which serves as the receptor for T-even phages entry, help bacteria survive the infection (18). In their turn, phages adapted and developed different strategies to overcome adsorption inhibition by recognition of new or altered receptors.

Studies of phage tail revealed that a combination of four mutations in receptor-binding protein J makes phage  $\lambda$  capable to infect the host by targeting a new receptor, OmpF, when expression of the cognate receptor LamB is reduced (19).



**Figure 1. Prokaryotic adsorption-blocking mechanisms.**

**a.** Bacteria modify their receptors to avoid phage infection (1). In turn, phages adapt to an altered receptor (2). Bacteria also use their proteins to mask phage receptors (3). The example of receptor masking by expression of protein A by *Staphylococcus aureus*.

**b.** Production of exopolysaccharide (EPS) by bacteria inhibits phage attachment, however, viruses circumvent this strategy using a lyase or a hydrolase to cleavage EPS.

**c.** Phages have evolved to recognize extracellular polymers such as O antigens and K antigens. The figure is taken from (17) with permission.

Production of extracellular polymers may protect bacteria against harsh environmental conditions and mask phage receptors. For instance, bacteria carrying the F plasmid can avoid phage adsorption due to interaction of the outer-membrane protein TraT, encoded by the F plasmid, with OmpA (20). In turn, phages use hydrolyzing enzymes that degrade the bacterial cell wall and hence enable phage attachment (21, 22).

Some bacteria produce molecules that bind to a receptor and render it unavailable for phage adsorption. Such masking strategy is associated with an iron transporter FhuA from *E. coli* which also serves as the port of entry for phages T1, T5, and  $\Phi$ 80. An antimicrobial peptide microcin J25 is produced by some *E. coli* strains under nutrient depletion conditions and transported into sensitive cells via FhuA (23). Thus, microcin J25 competes with phages and prevents phage attachment.

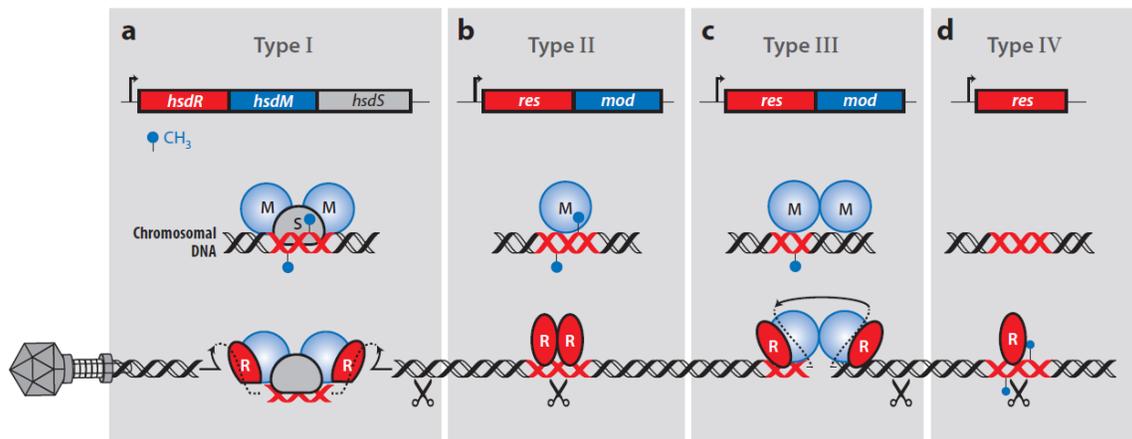
Prevention of adsorption in archaea is almost unexplored (24). However, mutations in two *Sulfolobus* gene clusters encoding cell surface and Type IV secretion proteins leads to inhibition of attachment by an archaeal virus SIRV2 (25).

One of the best-studied systems which block phage DNA entry is Sie (Superinfection Exclusion). This mechanism is encoded by phages or prophages and prevents reinfection of bacteria already infected by other viruses. For example, the Sie system of coliphage T4 consists of two proteins Imm and Sp that stop translocation of DNA by other superinfecting T-even phages (26). Imm blocks phage DNA penetration, while Sp inhibits phage-encoded lysozymes, blocking DNA translocation.

## 1.2 DNA modification systems

### 1.2.1 Classical Restriction-modification systems

If a phage managed to overcome defense systems operational at the earliest stages of infection, the invading phage DNA may be degraded later by restriction-modification systems (R-M). R-M systems are considered to be the most well-characterized (27) defense mechanisms. The REBASE database lists thousands of validated and predicted R-M components (28). R-M systems act by recognizing and attacking foreign DNA, whereas self (host) DNA is protected by modifications. Every restriction-modification system displays two activities. Methyltransferase modifies a specific sequence of own DNA by transferring a methyl group from the S-adenosyl methionine (SAM) to cytosine or adenine, while restriction endonuclease recognizes the same motif and cleaves it, if unmethylated. In addition, some R-M systems possess a separate target recognition domain that defines the specificity of MTase and REase. R-M systems are divided into four major groups by their subunit composition, recognition site, co-factor requirement and cleavage position (Fig. 2, Table 1) (27).



**Figure 2. Mechanisms for modification and cleavage by four groups of R-M systems.**

**a.** Type I forms two complexes:  $M_2S_1$  methylates asymmetric motifs in DNA;  $R_2M_2S_1$  recognizes the unmodified DNA fragment, moves along the DNA, and cleaves it far from the site upon the collision of two complexes.

**b.** Methyltransferase of Type II modifies palindromic sequences. Restriction endonuclease binds to unmodified sites and cuts them.

**c.** A Type III system is composed of two components: Methyltransferase modifies one of the two DNA chains, the  $M_2R_2$  complex recognizes unmodified motifs and cleaves it at a fixed location (>20 bp away from the recognition sites) following a collision of the two complexes.

**d.** Restriction endonucleases of Type IV recognize methylated DNA (or other modifications) and cut it.

Abbreviations: M – methyltransferase, R – restriction endonuclease, S – specificity subunit.

The figure is taken from (29) with permission.

The Type II of R-M systems form the most characterized group due to their usage for DNA analysis and molecular cloning. Typically, REase and MTase are separate enzymes and recognize a 4-8 bp palindromic DNA motif. MTase modifies unmethylated or hemi-methylated host DNA after the replication process, while cleavage of unmethylated phage DNA occurs at a fixed position near or within the recognition sequence and requires  $Mg^{2+}$  ions (30). However, there are plenty of subtypes that are exceptions of these simple rules, like IIB, IIG, and IIH where MTase and REase are fused into a single polypeptide chain (31).

Type I R-M systems encode an additional subunit S responsible for specificity.  $M_2S_1$  methylates asymmetric motifs in DNA, whereas  $R_2M_2S_1$  recognizes the unmodified DNA fragment and cleaves it at random positions far from the recognized site after a collision of the two complexes that processively move along the DNA via ATP-

dependent DNA translocation (32). Type I systems recognize asymmetric sites that contain 3-4 bp and 4-5 bp motifs separated by a non-specific 6-8 bp sequence. For example, the recognition site of EcoKI is  $\underline{A}AC(N)_6GTGC$ , where adenines on opposite DNA strands are methylated (33).

**Table 1. Main characteristics of different R-M systems (27).**

	Type I	Type II	Type III	Type IV
Example R-M system	EcoKI	EcoRI	EcoP1I	EcoMcrBC
Genes	<i>hsdR, hsdM, hsdS</i>	<i>ecorIR, ecorIM</i>	<i>mod, res</i>	<i>mcrB, mcrC</i>
Subunits	Three different subunits (R, M and S) combine to form $R_2M_2S_1$ and $M_2S_1$	Two different subunits (R and M) combine to form $R_2$ or $M_1$	Two different subunits (mod and res) combine to form $mod_2res_2$	Two different subunits are present, McrB and McrC
Enzyme activities	REase, MTase and ATPase	REase or MTase	REase, MTase and ATPase	REase and GTPase
Co-factors required for DNA cleavage	ATP, SAM, $Mg^{2+}$	$Mg^{2+}$	ATP, $Mg^{2+}$ (SAM)	GTP, $Mg^{2+}$
Co-factors required for methylation	SAM	SAM	SAM	No methylation
Recognition sequence	Asymmetric and bipartite, e.g. EcoKI, 5' $\underline{A}AC(N)_6GTGC$	Mostly symmetric, e.g. EcoRI, 5'GAATTC	Asymmetric, e.g. EcoP1I, 5'AGACC	Bipartite and methylated, e.g. EcoMcrBC, 5'RmC( $N_{30-4000}$ )RmC
Cleavage site	Variable locations 1000 bp from recognition site	Fixed location at or near the recognition site	Fixed location 25–27 bp from recognition site	Between methylated bases at multiple sites
DNA translocation	Yes	No	Yes	Yes

Similar to Type I, Type III systems recognize non-palindromic DNA sites and use  $Mg^{2+}$  and ATP-hydrolysis for DNA cleavage. Type III R-M systems possess hetero-oligomeric *res* and *mod* assemblies and methylate only one DNA strand (34). Recognition of two unmethylated sites in an inverted repeat orientation is required for cleavage that occurs at fixed positions - 25-27 bp far from the site (35).

In contrast to other R-M types, Type IV systems eliminate modified DNA and serve as protection from phages with methylated, hydroxymethylated or glucosyl-hydroxymethylated bases in their genomes while the host DNA remains unmodified and is not subject to restriction. One such example is the EcoMcrBC system that consists of an REase only and recognizes two copies of a two-base sequence – a purine followed by methylated cytosine separated by 40 – 3000 bp. Cleavage occurs 30 bp away from any of the two sites (27).

Phages have developed several antirestriction strategies to escape the R-M systems action. Phages with fewer restriction sites in their genome are less prone to DNA cleavage by restriction endonucleases (36). Thus, one phage strategy to avoid R-M systems is elimination of recognition sites in viral genome by mutations or incorporation of modified nucleotides. For example, substitution of cytosine for hydroxymethylcytosine in T4 genome protects the phage against R-M systems whose recognition sites contain cytosine (17). Some phages have adapted to R-M systems defense by coding their own methyltransferase that protects the phage from restriction endonucleases of the host (37). Another approach to resist R-M systems is their inhibition by phage-encoded proteins. For instance, T7 phage expresses Ocr, a protein that mimics  $\beta$ -form of DNA, binds to a restriction endonuclease and thereby inactivates it (38).

### 1.2.2 DISARM

New mechanisms similar to R-M systems were recently discovered. One of such systems named DISARM (Defense Island System Associated with Restriction–Modification) was divided into two classes (39). DISARM from *Bacillus paralicheniformis* is related to class II and consists of five components, including a 5-cytosine DNA methyltransferase DrmMII, a putative helicase DrmA, a protein with a phospholipase D domain DrmC, a protein with the DUF1998 domain, and one protein with an unknown function DrmE. The modification module of DISARM includes a methyltransferase similar to the classical R-M systems. However, the restriction module is thought to contain multiple subunits. Although the exact mechanism of DISARM still remains to be characterized, this system does not affect phage DNA injection, but causes intracellular phage DNA decay (39).

### 1.2.3 DPD systems

Another modification coupled with prokaryotic defense was found encoded in some genomic defense islands. 7-deazapurine-based modifications were thought to be a part of tRNA posttranslational modification pathway (40). However, gene clusters responsible for incorporation of 7-deazaguanine in DNA were also found (41). Besides 7-deazaguanine biosynthesis enzymes, these DPD systems (from 7-deazapurine in DNA) encode various helicases and nucleases. Although their antiphage activity was not tested, transformation efficiency of *Salmonella enterica* serovar Montevideo carrying a DPD system was low compared to cells lacking the system (41).

### 1.2.4 Phosphorothioation-based systems

The DNA sugar-phosphate backbone can serve as a target for modifications. Several prokaryotic defense systems with phosphorothioate (PT) modification, where a non-bridging oxygen is replaced by a sulphur atom, were discovered.

A PT-based system named Dnd (stands for DNA Degradation) was found in diverse prokaryotic genera (42). The modification module consists of a *dndABCDE* cluster, whose product incorporates the modification into host DNA, while DndFGH recognizes non-PT-modified DNA and eliminates it. DndA is a homolog of a pyridoxal 5'-phosphate (PLP)-dependent cysteine desulphurase; DndC has sequence homology to a 3'-phosphoadenosine-5'-phosphosulphate (PAPS) reductase; DndD is an ATPase that provides the system with energy for DNA nicking during sulphur incorporation; DndE is a small protein that binds non-specifically to nicked dsDNA; DndB regulates transcription of the *dndBCDE* cluster (42). Degradation of genomic DNA by DndFGH cluster lacking DndABCDE was shown (43). Interestingly, complete PT-modification of host genome is not necessary for the system. The relative geometry of sites is thought to dictate the modification state (44).

Similar PT-modification systems were found in archaea. In all analyzed archaea the genomic neighborhood of *dndCDEA* contains a *pbeABCD* cluster (for phosphorothioate-blocked DNA exclusion) sharing no sequence homology with *dndFGH* (45). PbeA comprises three domains: a phospholipase D (PLD)-like domain, a *res* subunit of Type III REase and a DEAD-family helicase C-terminal domain. PbeC contains an N-terminal AAA domain. PbeA and PbeD are proteins with domains of unknown function. Instead of DNA degradation showed for bacterial Dnd systems, archaeal DndCDEA-PbeABCD system uses a novel mechanism of virus elimination by inhibition of DNA replication (45).

A new PT-modification-based system distinct from the Dnd systems was discovered. This SspABCD–SspE system encodes a cysteine desulfurase SspA, a DNA nickase SspB, an ATPase SspC, an ATP pyrophosphatase SspD and a nicking endonuclease SspE with NTPase activity triggered by PT-modified sites. The distinctive feature of the Ssp system is modification of a non-palindromic motif in only one DNA strand (46).

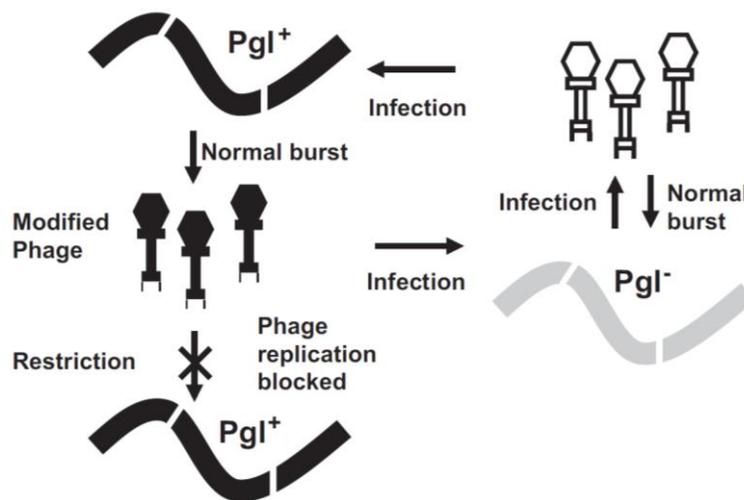
### 1.2.5 Pgl system

Bioinformatics analysis of sequenced prokaryotic genomes revealed that genes involved in defense systems such as R-M and TA systems are clustered in specific compact defense islands (14). One of the genes found enriched within these islands encodes a putative alkaline phosphatase PglZ and is present in 10% of sequenced genomes. Earlier, this gene was reported as a component of defense system Pgl (Phage growth limitation) in *Streptomyces coelicolor* A3(2) (47).

*S. coelicolor* strains carrying the Pgl system are resistant to the temperate phage  $\phi$ C31 and its homoimmune relatives (47). One-step growth experiments revealed no difference in burst size and lysis times of  $\phi$ C31-infected Pgl<sup>+</sup> and Pgl<sup>-</sup> cells. However, the progeny in Pgl<sup>+</sup> cells was attenuated for growth in the second round of infection. The result seemed to suggest that phages released after the first round of infection of Pgl<sup>+</sup>

cells become modified and lose the ability for further infection of these cells but not the Pgl<sup>-</sup> cells (Fig. 3) (48).

The Pgl mechanism still remains unclear, but two functional modules - the phage alteration/protective module and the defensive module that blocks the development of modified phage – are required for the Pgl function. The proposed mechanism of action of the Pgl, if true, would differ from all known methyl-specific restriction systems (in fact, the suggested marking of phage DNA and flagging it for restriction in later rounds of infection can be considered as the opposite of the standard R-M mechanism (48)).



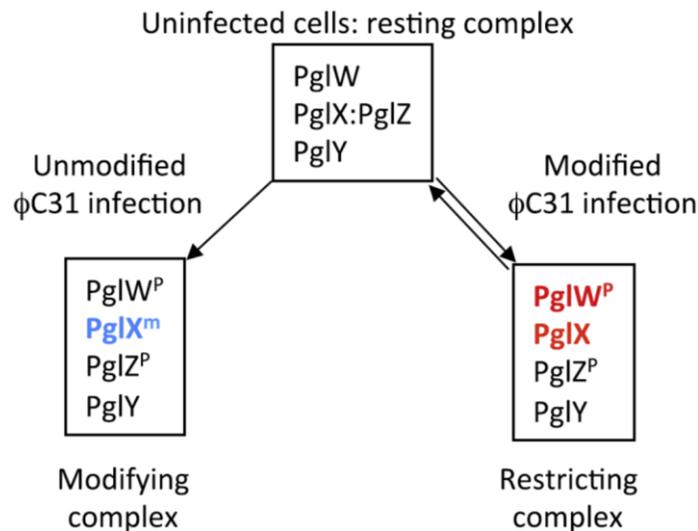
**Figure 3. Schematics of the Pgl phenotype.**

Pgl<sup>+</sup> cells produce modified progeny whose development is blocked in the second cycle of infection. However, in Pgl<sup>-</sup> cells the phage ability to infect the culture is not changed. The figure is taken from (48) with permission.

The genetic determinants of the Pgl system were uncovered (47–49). The complete system was shown to consist of four genes *pglWXYZ*, organized in two (*pglWX* and *pglYZ*) operons. Bioinformatics analysis suggested that PglW may be a kinase, PglX – a SAM dependent methyltransferase, PglY – an ATPase with Walker motifs, while PglZ carries domain with alkaline phosphatase-like fold. For PglW, PglX and PglY these predictions were confirmed *in vitro*. All four Pgl proteins are large (from 974 aminoacid-

long PglZ to 1557-aminoacid PglW), so in addition to activities predicted based on the presence of short sequence motifs they may contain other functions (48).

Knock-out or null mutations showed that all *pgl* genes are necessary for functional defense (48). Moreover, a deletion of alkaline phosphatase gene *pglZ* led to accumulation of mutations in the methyltransferase *pglX* gene. The Pgl system is thus likely to encode a TA module in which antitoxin PglZ neutralizes the PglX toxin. The current model of Pgl function posits that the systems switches between three functional states: resting, modifying, and restricting (Fig. 4). In uninfected cells Pgl proteins assemble into a “resting” complex in which toxic activity of PglX is repressed by PglZ. During infection the resting complex transforms into either the modifying or restricting complex, depending on the modification state of the incoming phage. The modifying complex is likely to have N<sub>6</sub>-adenine activity due to PglX. The mechanism of restriction is not clear but it is suggested to be activated by modified phage and mediated by the nuclease domain of PglZ, host methyl-specific restriction endonucleases, or the PglX toxin. However, no DNA modifications were observed in infected or uninfected cell or viral genomes (48).



**Figure 4. Schematics of the proposed mechanism of Pgl defence.**

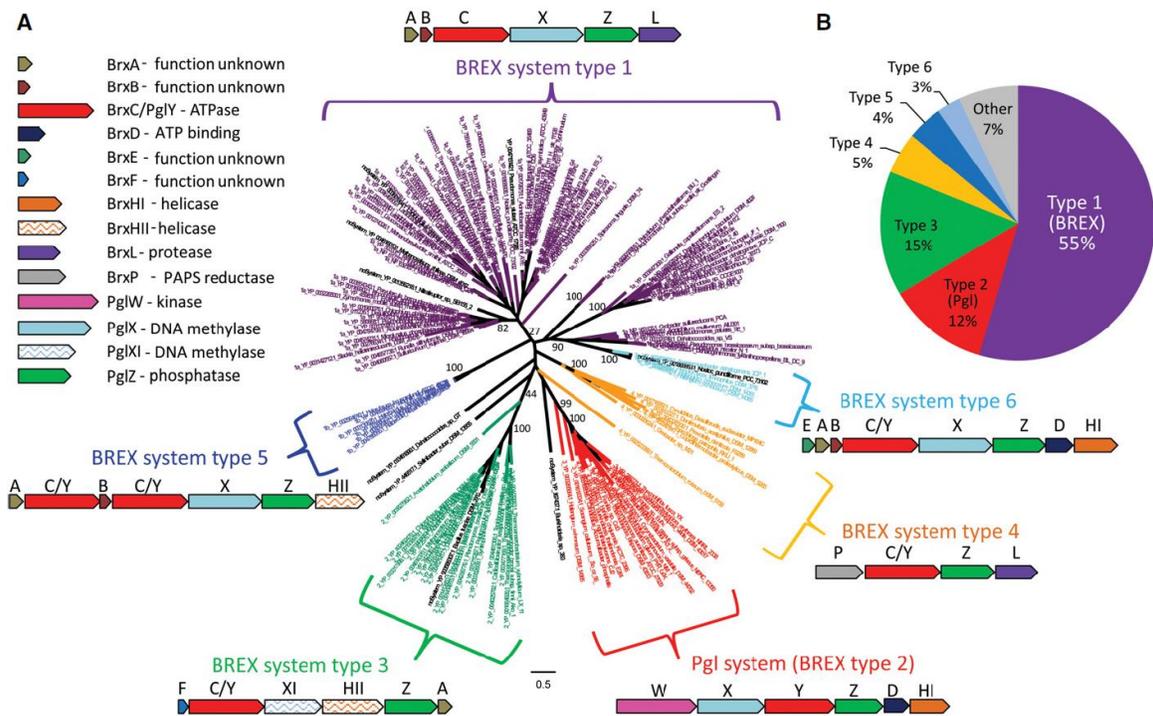
The resting complex assembles in uninfected Pgl<sup>+</sup> cells in which a toxicity of PglX is inhibited by PglZ. Depending on the modification state of the incoming phage, the resting complex is transformed into a modifying or restricting complex. The modifying complex has the N6-adenine-methylase activity of PglX (blue color). Candidates that may be involved in the restriction activity are colored in red. The cleavage mechanism is likely to be mediated by a nuclease domain of PglZ, bacterial methyl-specific restriction endonucleases, or a toxic PglX. PglW<sup>P</sup> and PglZ<sup>P</sup> are probably phosphorylated by the PglW kinase, which becomes activated during the infection.

The figure is taken from (48) with permission.

### 1.2.6 BREX systems

As it was mentioned above, sequence analysis of defense islands present in many prokaryotic genomes revealed overrepresentation of genes coding for proteins homologous to *pglZ* (14). A census done in 2014 led the authors to propose a separate BREX family for clusters containing *pglZ*-like genes and several other conserved neighboring genes (15). The family was divided into 6 types according to genes content and locations (Fig. 5). Previously described Pgl system was classified as a Type II BREX system. Curiously, in almost 90% cases, *pgl* genes were accompanied by two other genes - *brxD* and *brxHI* - that encode a small protein with ATP-binding motif and a predicted helicase, respectively. The limited defense activity of Pgl against the  $\phi$ C31 phage only was suggested to be explained by incomplete gene composition of the system. On the other hand, these additional genes probably may take part in processes other than defense *per se*, as *cas1* and *cas2* genes do in the case of CRISPR-Cas systems.

The most abundant BREX system type is Type I. It consists of six genes: two genes *brxX* and *brxZ* are homologous to genes *pglX* and *pglZ*, the remaining genes encode a predicted Lon-like protease BrxL, a protein with a RNA-binding motif BrxA, a large protein with an ATPase domain BrxC and a protein with an unknown function BrxB. To verify if BREX systems provide protection against phages, the whole BREX cluster of *Bacillus cereus* was cloned into *B. subtilis* genome lacking an endogenous BREX system (Fig. 6) (15).

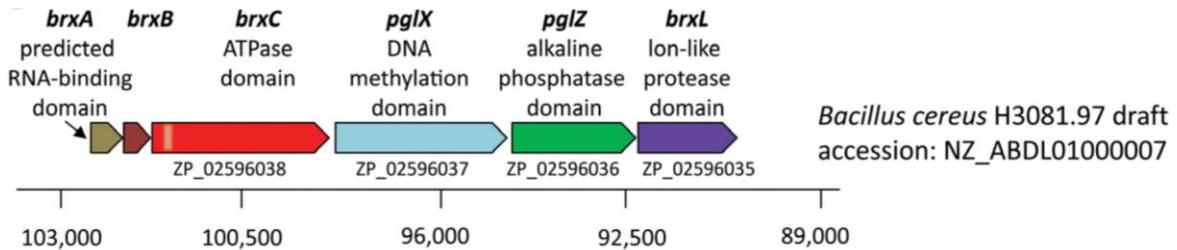


**Figure 5. PglZ phylogeny and classification of BREX subtypes.**

a. A phylogenetic tree of the PglZ protein. BREX family is divided on six types depending on gene composition.

b. Distribution of BREX system types.

The figure is taken from (15) with permission.



**Figure 6. The BREX cassette Type I from *Bacillus cereus*.**

The figure is taken from (15) with permission.

Ten phages including Myoviridae (SPO1 and SP82G), lambda-like Siphoviridae (*Φ105*, *rho10*, *rho14*, and SPO2), and SPβ-like Siphoviridae (*SPβ*, *Φ3T*, SP16, and *Zeta*) were selected to determine BREX defense activity (Table 2). Efficiency of protection was calculated as the ratio between the number of plaques formed on the BREX- strain divided by the number of plaques formed on the BREX+ strain.

**Table 2. BREX protection against a wide range of *Bacillus* phages (15).**

Phage	Genus	Family	Life cycle	Infection blocked by BREX?	Efficiency of BREX protection
<i>SPβ</i>	SPβ-like	Siphoviridae	Temperate	Yes	$> 10^5$
SP16	SPβ-like	Siphoviridae	Temperate	Yes	$> 10^5$
<i>Zeta</i>	SPβ-like	Siphoviridae	Temperate	Yes	$> 10^5$
<i>Φ3T</i>	SPβ-like	Siphoviridae	Temperate	Yes	$> 10^5$
SPO2	Lambda-like	Siphoviridae	Temperate	Yes	$> 10^5$
<i>Φ105</i>	Lambda-like	Siphoviridae	Temperate	No	1
<i>rho10</i>	Lambda-like	Siphoviridae	Temperate	No	1
<i>rho14</i>	Lambda-like	Siphoviridae	Temperate	No	1
SPO1	SPO1-like	Myoviridae	Obligatory lytic	Yes	$8 \times 10^2 \pm 0.02$
SP82G	SPO1-like	Myoviridae	Obligatory lytic	Yes	$1.8 \times 10^1 \pm 0.08$

BREX protected bacteria against 7 of 10 phages tested, including both lytic and temperate viruses. The inability of the BREX system to resist against *Φ105*, *rho10* and *rho14* phages can be explained by either the presence of strategies that circumvent the BREX defense or the lack of BREX targets in viral genomes (15)

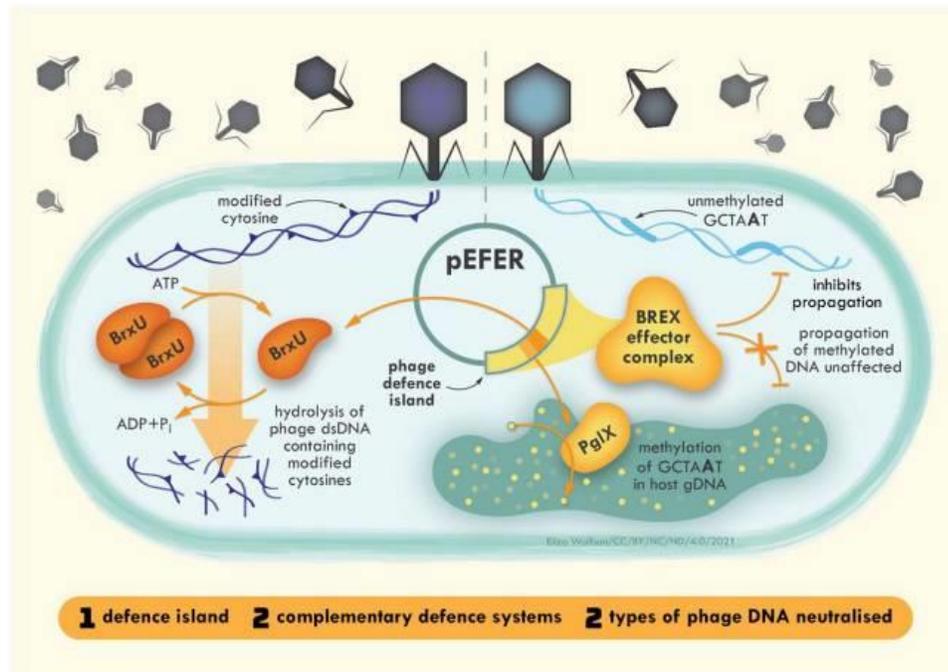
To compare the Pgl and BREX systems, one-step growth experiments with phage *φ3T* were carried out in BREX+ and BREX- cultures. No phage progeny was found in BREX+ cell even after 2 hours of infection, likely due to a block of phage multiplication before the end of the first cycle. In contrast to Pgl, the BREX system inhibited lysogenization. The results established that BREX does not display the Pgl phenotype despite homology of *pglX* with *brxX* and *pglZ* with *brxZ* genes (15).

To determine which stage of phage infection is blocked in BREX+ cells, phage adsorption and replication of its DNA were studied. No effects on phage adsorption were revealed, while phage DNA replication inside BREX+ cells was blocked (15).

Due to the presence of a predicted m6A DNA adenine methylase PglX/BrxX in the BREX system, host DNA and total cellular DNA (including chromosomal DNA and intracellular phage DNA) extracted 10 and 15 minutes after infection were sequenced using PacBio platform. The analysis detected m6A methylation on the 5<sup>th</sup> position of the non-palindromic hexamer TAGGAG in the bacterial genome, whereas no modifications in invading phage genome was observed (15). The result seemed to suggest that the BREX system acts in a way akin to restriction-modification systems. However, deletion of methyltransferase PglX/BrxX did not lead to cell death as in R-M systems. Moreover, no phage DNA degradation was detected in the BREX+ culture (15).

Another Type I BREX system from *Lactobacillus casei* Zhang strain was studied (50). *L. casei* Brx proteins are homologs of *B. cereus* Brx proteins. The BREX system methylates an adenine residue at the fifth position of an ACRCAG motif. The transformation efficiency of plasmids containing ACRCAG sites was lower in BREX+ cells comparing to  $\Delta$ *pglX* strain. In contrast, no differences were found for transformation of BREX+/BREX- cultures with plasmids lacking *brx* sites (50).

Studies of BREX systems Type I was expanded by investigation of a multidrug resistance plasmid pEFER from *Escherichia fergusonii* (51). In addition to canonical *brx* genes, the plasmid carries genes *brxR*, *brxS*, and *brxT* upstream of the cluster and *brxU* between *brxC* and *pglX/brxX*. BrxR is a WYL-domain transcriptional regulator, BrxS is a putative IS3 family transposase, BrxT is a protein of unknown function, BrxU is a homolog of Type IV restriction enzyme GmrSD. Deletion of BrxU abolishes protection against a set of phages but does not affect methylation of adenine in the GCTAAT *brx* motif of this system. *In vitro*, BrxU shows an ATP-dependent endonuclease activity for DNA containing modified cytosines. Thus, unmethylated phages are thought to be subject of BREX defense while phages with modified cytosines can avoid the antiphage activity of BREX but not of restriction endonuclease BrxU (51) (Fig. 7).



**Figure 7. Schematics of pEFER-mediated defense against diverse phages.**

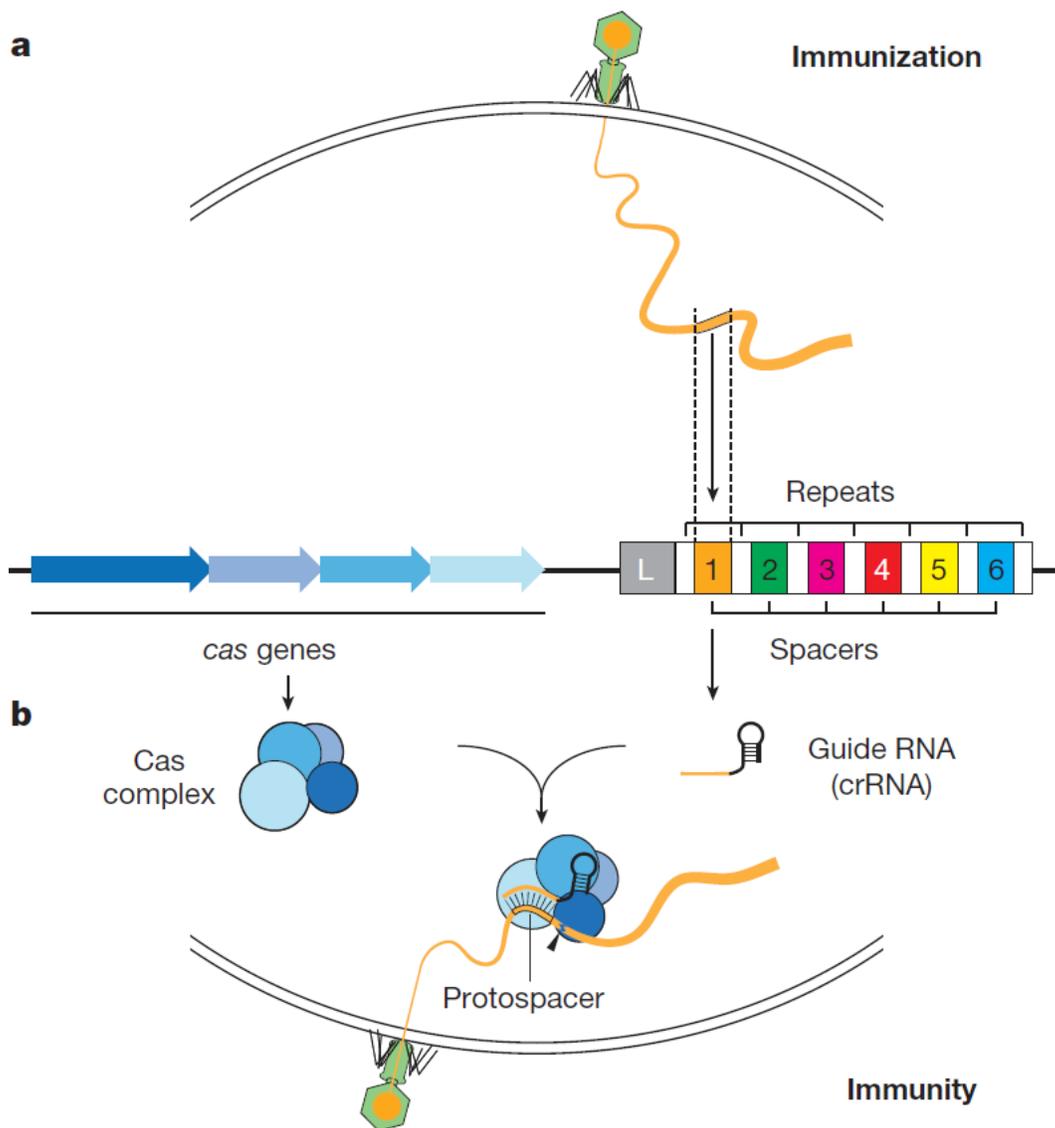
The figure is taken from (51) with permission.

The WYL-domain transcription regulator BrxR was studied by two scientific groups independently in pathogenic *Escherichia fergusonii* (52) and *Acinetobacter baumannii* (53). BrxR negatively regulates the expression of *brxA* and its own transcription by binding to inverted DNA repeats (52). BREX system Type I from *A. baumannii* comprises a canonical set of *brx* genes and *brxR*. The system leads to methylation of the fifth position of GTAGAT sites. BrxR and BrxL do not participate into the modification but all Brx proteins are necessary for defense. Deletions of *brxR*, *brxB* and *pglZ/brxZ* lead to a decrease in cell growth. Unlike BREX from *Bacillus cereus*, this system is transcribed as a single mRNA (53).

### 1.3 CRISPR-Cas systems

Clustered, regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated (Cas) proteins mediate an adaptive immune system, which is widespread in bacteria and archaea (54). Each CRISPR locus consists of an array of short repeats (20-50 bp long) separated by unique sequences (spacers) of viral or plasmid origin (55). CRISPR defense mechanism protects cells from invaders when a spacer sequence is complementary to foreign DNA.

Acquisition of new spacers occurs during infection through the action of universal for all CRISPR systems Cas1 and Cas2 proteins in a process called “adaptation” (Fig. 8a). The variety of spacers in CRISPR locus correlates with the resistance spectrum of the cell and is inherited by daughter cells. The main problem during the adaptation stage is to prevent autoimmunity. Selected for integration sequences of foreign DNA are called protospacers and are usually adjacent to a 2-3 bp short motif known as PAM (from Protospacer Adjacent Motif). Since some CRISPR systems use PAMs for distinguishing between self and non-self (56) and mutations in the PAM region result in avoidance of CRISPR defense (57). Mutations in the protospacer that affect complementarity with the spacer are also important for CRISPR defense. Phages with protospacers introducing even single mismatches close to PAM circumvent CRISPR, but those with mismatches (even multiple) at PAM-distal protospacer positions do not.



**Figure 8. The mechanism for recognition and elimination of target nucleic acids by CRISPR-Cas.**

**a.** In the adaptation/immunization stage, new sequences of invading phage DNA (protospacers) are integrated into the first position of the CRISPR locus by the Cas protein adaptation complex.  
**b.** In the immunity stage, the CRISPR array is transcribed into pre-CRISPR RNA (pre-crRNA) that is processed by Cas proteins into mature CRISPR RNAs (crRNAs). These small RNAs act as antisense guides for Cas RNA-guided nucleases, which recognize the target and cleaves it.  
 The figure is taken from (7) with permission.

Transcription of CRISPR locus activates the immune system (Fig. 8b). CRISPR loci are transcribed into long precursor CRISPR RNA (pre-crRNA) that is processed by Cas protein complexes into mature crRNA. All CRISPR systems are divided into two classes that are subdivided into 6 types and 33 subtypes according to the Cas proteins participating the effector modules (Cas proteins participating in the interference stage), crRNA biogenesis, as well as target recognition and cleavage (58). Class 1 CRISPR Cas systems with multisubunit effector complexes are the most widely distributed in both bacteria and archaea (90% of identified CRISPR systems). The remaining 10% of CRISPR systems belong to Class 2 CRISPR-Cas systems with monomeric effector proteins and previously were thought to be present only in bacteria (59). However, recent studies identified the first Class 2 systems in the archaeal domain (60). Due to their simplicity, the Class 2 CRISPR-Cas systems are widely used as genome editing tool (61).

The final stage of CRISPR defense is interference: a complex of Cas proteins uses crRNA as a guide to recognize and cut the target. Partially matching invader DNA and crRNA provokes incorporation of new spacers in a process known as priming or primed adaptation (62). Although CRISPR systems usually cleave DNA, systems targeting RNA were also found (63).

In addition to mutations in the PAM or in the protospacer, phages escape from CRISPR defense by coding anti-CRISPR proteins. These proteins abolish interference by preventing the formation or blocking the action of effector complexes (64). Moreover, some phages encode their own CRISPR systems. For example, *Vibrio cholerae* serogroup O1 phages contain a CRISPR-Cas system with a spacer targeting host antiphage system PLE (from Phage-inducible island-Like Element). Upon infection PLE system exploits molecular machinery and structural components of phage ICP1 for production of virions carrying PLE DNA. Therefore, phages hijack bacterial CRISPR-Cas systems to overcome host defense (65).

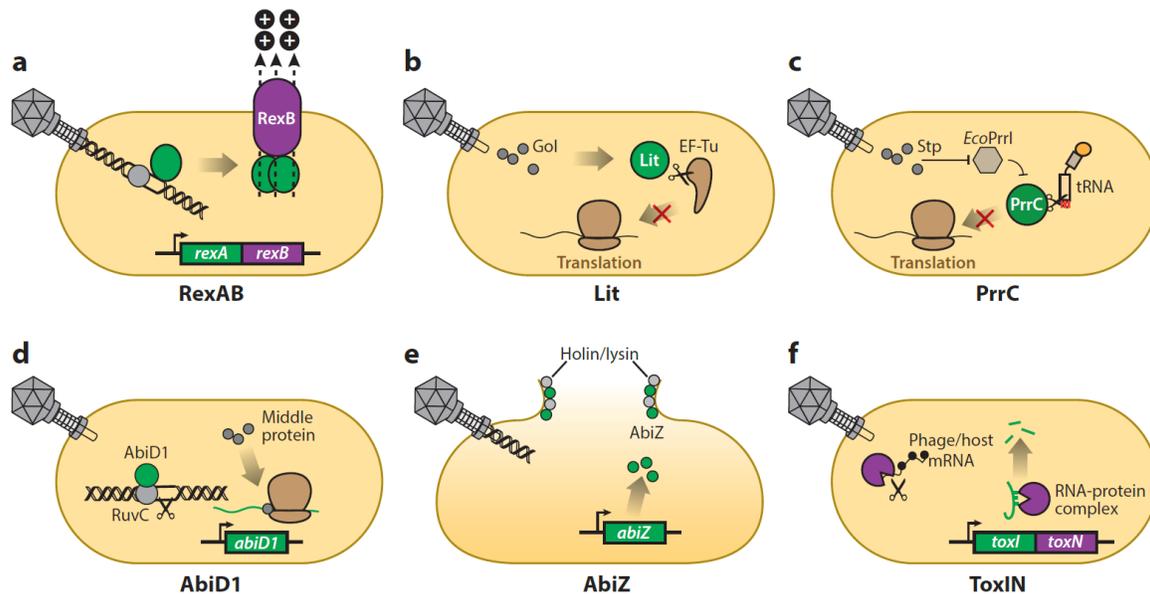
## 1.4 Abortive infection and toxin-antitoxin systems

Infected bacteria are able to inhibit their own metabolic processes and, as a result, commit suicide in order to protect the clonal bacterial population against the spread of viral infection (29). This principle is typical for two defense mechanisms: abortive infection (Abi) and toxin–antitoxin (TA) systems. These systems are very diverse and can interrupt various stages of phage propagation (Fig. 9).

### 1.4.1 Abortive infection systems

Abi systems are typically encoded by prophages or plasmids and become activated once bacteria are infected. The first characterized Abi system was *rexAB* encoded by  $\lambda$ -lysogenized *E. coli* strains. The *rexAB* locus encodes two proteins: RexA binds to phage DNA-protein complex assembled during replication and activates an ion channel formed by RexB in the cell wall (Fig. 9a). Depolarized bacterial membrane leads to a decrease in intracellular ATP and therefore causes death of infected cell and prevents the completion of phage infection cycle (66). However, phages can escape from RexAB by mutations in a transcriptional co-activator that redirects host RNA polymerase and switches phage transcription from early to middle genes (67). Since membrane depolarization is inhibited, phage is able to complete its life cycle (66). Similar to RexAB, a PifA system encoded by the F plasmid aborts T7 phage infection. Phage Gp1.2 and Gp10 proteins activate an ion channel formed by the PifA system (68, 69).

Until recently serine/threonine kinases were thought to be components of antiviral response in eukaryotes only but an Abi system with a serine/threonine kinase was discovered in *Staphylococcus epidermidis* (70). A viral protein PacA of a temperate phage CNP $\times$  activates a Stk2 kinase resulting in phosphorylation of essential metabolic proteins that causes cell death.



**Figure 9. Mechanisms of Abi and TA systems.**

- a.** RexAB is activated by a phage DNA-protein complex. RexB forms an ion channel and, once activated by RexA, depolarizes the membrane, leading to cell death.
- b.** The Lit system inhibits translation by cleavage of the elongation factor EF-Tu in response of phage infection, stopping phage maturation and bacterial metabolic processes.
- c.** PrrC also affects translation. The PrrC activity is blocked by Type I R-M system *EcoPrrI*. In the presence of phage protein Stp, *EcoPrrI* is inactivated, following the release of the PrrC endonuclease that cuts tRNA<sup>Lys</sup>.
- d.** The *abiD1* mRNA is unstable and is not translated in the absence of infection. Phage proteins stabilize the mRNA and enable the AbiD1 protein synthesis. AbiD1 blocks expression of the phage protein ORF3, which resolves branched phage DNA.
- e.** Products of AbiZ bind to holin and lysin of phage  $\phi$ 31, causing premature bacterial lysis.
- f.** ToxIN encodes a toxic endonuclease ToxN, which is neutralized by non-coding ToxI RNA. ToxN cleaves the host and phage RNA, resulting in infection abortion.
- The figure is taken from (30) with permission.

Abi systems Lit and PrrC inhibit translation. A capsid protein Gp23 (or Gol) of T4 phage binds to the ribosomal elongation factor Tu (EF-Tu), which activates a Lit protease encoded by defective e14 prophage of *E. coli* (Fig. 9b). The protease cleaves EF-Tu and stops bacterial protein synthesis (71, 72). Nevertheless, *gp23*-mutated phages are able to replicate in Lit-carrying strains (73). Likewise, PrrC system from a defective prophage expresses an endonuclease that cleaves tRNA<sup>Lys</sup> (74, 75). The PrrC system is inhibited by Type I R-M system *EcoPrrI*, that, in turn, can be inactivated by a T4 phage peptide Stp

(76, 77). Activated by Stp, endonuclease PrrC stops phage and bacterial translation (Fig. 9c).

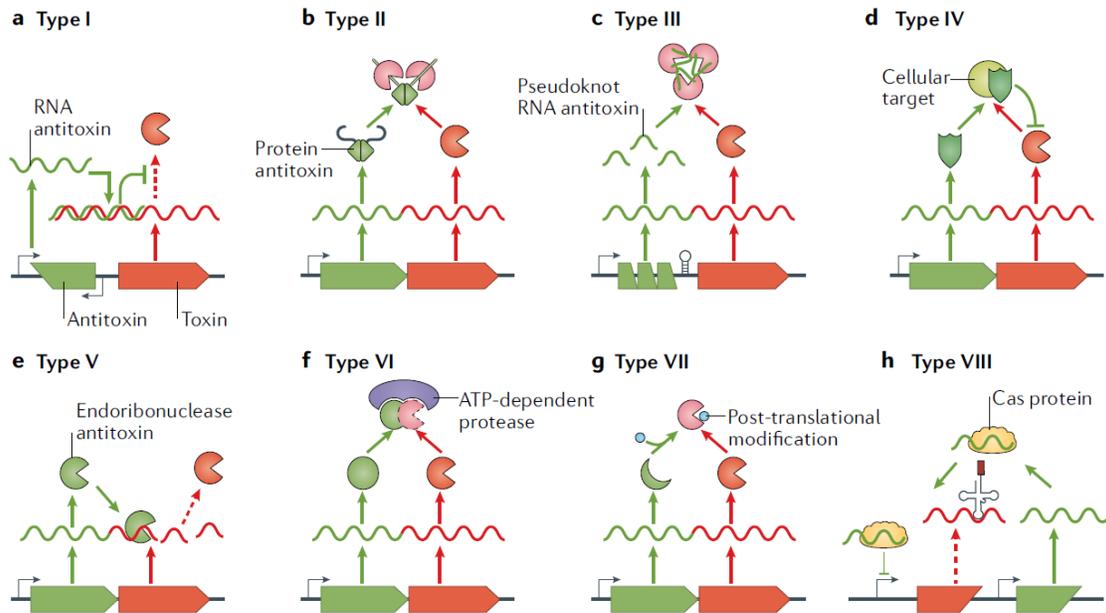
The largest number of Abi systems were found in *Lactococcus lactis*. To date, there are 23 classified Abi mechanisms encoded by plasmids and transferred by conjugation (29). One of the best-characterized Abi systems of *L. lactis* is AbiD1. A protein ORF1, derived from phage bIL66, stabilizes *abiD1* mRNA and enables protein synthesis (Fig. 9d). AbiD1 blocks expression of a phage protein ORF3 which resolves branched DNA structures during phage replication and thereby interferes with phage propagation (78). Spontaneous mutations in *orf1* render phage insensitive to this defense mechanism (79). Phage-encoded holin and lysin proteins can be bound by AbiZ components, which induces premature bacterial lysis (Fig. 9e) (80).

Some Abi systems are toxic for the host and stop cell growth. For example, AbiK, AbiN, and AbiO, when cloned into high-copy-number plasmids, cannot be maintained in bacteria (81–83). Toxic Abi systems require tight control at transcription, translation and post-translation stages to prevent cell death in the absence of phage infection (29).

#### 1.4.2 TA systems

Toxin–antitoxin (TA) systems typically encode a stable toxin whose action is inhibited by an unstable antitoxin. Toxins interfere with essential processes as translation (ToxN, MazF, RelE, HipA), replication (CcdB, ParE) and cytoskeletal/cell wall formation (CbtA, PezT) (29). All TA systems are classified into 8 different types according to regulation of a toxin by its antitoxin (Fig. 10). Antitoxin mRNA of Type I is antisense to toxin mRNA. Types II and III encode translated and untranslated RNA antitoxins, correspondingly, that interact with their cognate toxins directly and inhibit its activity. For Type IV, a protein-antitoxin protects a toxin target from the toxin action. Type V systems encode nucleases that cleave toxin transcripts. Antitoxins of Type VI serve as an adaptor for host proteases that cleave the toxin. Type VII toxins are neutralized by post-translational modification of antitoxins. The last type is TA systems

that are based on transcription repression of toxins by antitoxins (84). One such system, CreA-CreT, uses a Type I CRISPR-Cas system Cascade effector (Effector module of Type I CRISPR-Cas systems) to regulate toxin transcription (85).



**Figure 10. Current division of TA systems into types.**

- a. Type I: Antitoxin mRNA is antisense to toxin mRNA.
  - b. Type II: Protein antitoxin sequesters the protein toxin.
  - c. Type III: Antitoxin RNA interacts with a protein toxin and inhibits it.
  - d. Type IV: Antitoxin hides a toxin target from toxin activity.
  - e. Type V: Antitoxin is an endonuclease that degrades toxin mRNA.
  - f. Type VI: Antitoxin serves as an adaptor for host proteases that cleave the toxin.
  - g. Type VII: Toxins are neutralized by post-translational modification of antitoxins.
  - h. Type VIII: Antitoxin in complex with other proteins regulates transcription of the toxin gene.
- The figure is taken from (84) with permission.

The best characterized TA system is ToxIN of *Pectobacterium atrosepticum* (Fig. 9f). The *toxIN* locus encodes an endoribonuclease ToxN whose toxicity is neutralized by noncoding *toxI* RNA through protein-RNA complex formation. Once activated, ToxN cuts host and phage RNA and therefore causes cell death (86).

### 1.4.3 Retrons

Up until recently, the function of retons was unknown. Taking into account that some retons are localized in defense islands, antiviral activity of retons was predicted and tested (87). Retrons consist of a reverse transcriptase (RT) and non-coding RNA (ncRNA) that is used by RT as a template for synthesis of a covalently linked RNA/DNA hybrid. Examination of the genomic environment of retons showed that the vast majority of them contained an additional protein-coding gene encoding a ribosyltransferase fused to a DNA-binding domain, two predicted transmembrane helices, or endonuclease domain. Overall, six of ten types of effector proteins were tested for antiphage activity. In general, retons defend cells from viruses via an abortive infection strategy. The abortive mechanism of one retron was investigated. Inhibition of the RecBCD system by phage-encoded proteins from phage  $\lambda$  and T7 activates a retron-based Ec48 system and causes cell death (87).

### 1.4.4 CBASS

The cyclic GMP–AMP synthase (cGAS) protein is a sensor of cytosolic viral DNA that produces a cyclic GMP–AMP (cGAMP) signaling molecule and thus activates the immune response (88). CBASS (from Cyclic-oligonucleotide-Based Antiphage Signaling System) consists of a four-gene cluster encoding a cGAS, two proteins with eukaryotic-like domains E1, E2 and JAB, and an effector protein such as a phospholipase, an endonuclease, proteins with TIR or transmembrane domains. More than 10% of sequenced prokaryotic genomes contain the system. An *E. coli*-derived CBASS system with a phospholipase as an effector protein triggers, upon phage infection, the production of cGAMP by cGAS. The signaling nucleotides activate cell membrane degradation by a phospholipase CapV. Thus, CBASS is an Abi system that causes cell death and arrests phage propagation (88).

## 1.5 Novel defense systems

Variety of new prokaryotic defense systems was predicted by bioinformatics tools and validated in model prokaryotes for antiviral activity (13, 89).

One approach is based on a search of genes enriched in defense islands and clustered next to known antiphage systems (13). Several novel systems were found by this method.

The first system named Zorya (after a deity from Slavic mythology) was discovered by the enrichment of a gene with a domain of unknown function pfam15611 denoted as *zorC*. The system consists of a four-gene cluster *zorABCD* and showed protection against tested phages. ZorA and ZorB share homology with flagella motor proteins MotA and MotB, respectively, while ZorD has predicted helicase and nuclease domains. Another type of Zorya systems – Type II - was found by searching for *zorA* and *zorB* homologs. The ZorABE cluster protects the host from T7 and ssDNA phage SECphi17 infection. ZorE contains a HNH-endonuclease domain. No phage progeny was found in Zorya-carrying cells suggesting that the system works via abortive infection mechanism. ZorA and ZorB are hypothesized to depolarize membrane potential upon phage infection. Zorya system was not identified in archaea (13).

Another defense system named Thoeris (Egyptian protective deity of childbirth and fertility) was found by a Toll-Interleukin Receptor (TIR) pfam08937 domain. Curiously, TIR domain is known to be a part of eukaryotic innate immunity. The system encodes a sirtuin (SIR2) domain protein ThsA with a catalytic NADase activity and a TIR domain protein ThsB. The mechanism of phage infection recognition is unknown but ThsB activates upon infection and triggers production of signaling cyclic ADP-ribose molecules that in turn activate NADase activity of ThsA which lethal for cells (13, 90).

The Druantia systems (named after a deity from Gallic mythology) is divided in two types with a common gene coding for a large protein with DUF1998 and a helicase domain. The DUF1998 gene was found to be a component of a previously described

DISARM and Dpd systems. Cloned Type I system showed antiviral activity against 4 out of 6 tested phages (13).

No antiphage activity was identified for a system denoted Wadjet (god protector of ancient Egypt). Wadjet contains a highly enriched *jetABCD* cluster with homology to housekeeping condensin system proteins MukF, MukE and MukB. The Wadjet system protects a host from invasion by plasmids (13).

Viperins are thought to be a part of immune system in eukaryotes (91). However, recent studies showed that homologs of viperins are localized near known defense systems genes and participate in antiviral activity of prokaryotes (92). In addition to ddhCTP produced by eukaryotic viperins, prokaryotic viperins (pVips) produce other chain-terminating ribonucleotides such as ddhGTP and ddhUTP. pVips protect *E. coli* cells from T7 phage infection by inhibition of viral polymerase-dependent transcription. However, though pVips do not affect bacterial transcription, they protect the host from phage  $\lambda$  and P1 that use host transcription machinery only (92).

Another method for defense system identification, which is based on localization in defense islands but irrespective of domain composition, was developed. New antiphage systems with genes previously not implicated in defense were discovered using this method.

The first system found by this approach is a defense system named RADAR (phage Restriction by an Adenosine Deaminase Acting on RNA). The system consists of an ATPase RdrA and an adenosine deaminase RdrB that can be accompanied by additional genes encoding a small membrane protein like a SLATT domain or a Type VI-B CRISPR ancillary protein Csx27. Whole-transcriptome sequencing revealed an enrichment of A to G substitutions at specific sites in sequenced reads upon phage infection. The RADAR system is thought to work via an Abi mechanism (89).

Another family of defense systems with reverse transcriptase (RT) was discovered. Six members of uncharacterized RT family without mobility hallmarks protect cells from phage infection. Such defense systems were named DRTs (from Defense-associated RTs). DRTs systems can consist of a single RT or have additional

genes. For example, DRT Type I contains a large RT protein fused to a nitrilase domain that suppresses phage late genes expression but not the expression of middle and early genes (89).

NTPases of the STAND (signal transduction ATPases with numerous associated domains) superfamily were thought to participate in eukaryotic programmed cell death and their function in prokaryotes was unclear. Antiviral ATPases/NTPases of the STAND superfamily (AVAST) contains a STAND NTPases and an effector protein with a putative nuclease, SIR2, a trypsin-like serine protease or an uncharacterized helical domain (89). Defense activity of AVAST was not experimentally validated.

## Chapter 2. Materials and methods

### 2.1 Strains and plasmids

<i>E. coli</i> strain	Comments	Source
HS	Natural isolate with BREX <sup>Ec</sup> , <i>Str</i> <sup>R</sup>	Paul Cohen
HS ΔBREX <sup>Ec</sup>	Natural isolate with deleted BREX <sup>Ec</sup> , <i>Str</i> <sup>R</sup> , <i>Cm</i> <sup>R</sup>	Lab stock
BW25113	Strain for routine work	Lab stock
JM110	Dam- stain for PacBio sequencing	Lab stock
MC4100 lysogen	Source of λ <i>cl</i> <sub>857</sub> <i>bor</i> :: <i>Cm</i> <sup>R</sup> , <i>Cm</i> <sup>R</sup>	Ryland Young
MG1655 seqA-ECFP Dam- LE392	SeqA-experiment <i>sup</i> <sup>F</sup> and <i>sup</i> <sup>F</sup> host	Lanying Zeng
LE392(λ <sub>LZ1</sub> )	Prophage induction experiment, <i>Kan</i> <sup>R</sup>	Lanying Zeng
MG1655 (λD-EYFP <i>cl</i> <sub>857</sub> <i>bor</i> :: <i>Kan</i> <sup>R</sup> ) [pBR322-Plate*D]	Source of λD-EYFP <i>cl</i> <sub>857</sub> <i>bor</i> :: <i>Kan</i> <sup>R</sup> , <i>Kan</i> <sup>R</sup>	Lanying Zeng
CSH50 λ <i>sfiA</i> :: <i>lacZ</i>	SOS-response reporter	Lab stock
DH5α	Strain for cloning	Lab stock
<b><i>E. coli</i> plasmids</b>		
pBTB-2	ara-promoter vector	Lab stock
pBREXAL	BREX <sup>Ec</sup> system under a natural promoter, <i>Kan</i> <sup>R</sup>	Lab stock
pBREXALΔN, N=A, B, C, X, Z, L	pBREXAL with deleted <i>brxN</i> gene, <i>Kan</i> <sup>R</sup>	This work
pTG-BREX	BREX <sup>Ec</sup> system under a natural promoter, <i>Cm</i> <sup>R</sup>	This work
pTG	Plasmid with the p15A origin, ara promoter and <i>Cm</i> <sup>R</sup>	This work
pBREX1	ara-inducible plasmid with <i>brxABC</i> , <i>Kan</i> <sup>R</sup>	Lab stock
pBREX2	ara-inducible plasmid with <i>brxXZL</i> , <i>Amp</i> <sup>R</sup>	Lab stock
pBREX1ΔN, N=A, B, C	pBREX1 plasmid with deleted N gene	This work
pBREX2ΔN, N=X, Z, L	pBREX2 plasmid with deleted N gene	This work
pBAD- <i>brxN</i> , N=A, B, C, X, Z, L	ara-inducible plasmid with <i>brx</i> gene, <i>Amp</i> <sup>R</sup>	Lab stock
<b><i>Haloarcula hispanica</i> strain</b>		
DF60 Δ <i>cas</i>	Natural isolate with BREX <sup>HAR</sup> system with deleted <i>pyrF</i> and <i>cas1-cas8</i> genes	Hua Xiang
ΔBREX	DF60 Δ <i>cas</i> strain with deleted BREX <sup>HAR</sup> system	This work
ΔX1-HI	DF60 Δ <i>cas</i> strain with deleted X1-Z-X2-HI genes	This work
BREXΔN,N=A,B,C1,C2,X1,X2,Z,HI,NucS	DF60 Δ <i>cas</i> strain with deleted <i>brx</i> gene	This work
<b><i>Haloarcula hispanica</i> viruses</b>		
His2	dsDNA spindle-shaped halovirus	Dennis Bamford
HHPV1	dsDNA	Dennis Bamford
HHPV2	ssDNA	Hua Xiang
HHPV3		Dennis Bamford
HHPV4		Dennis Bamford
SH1	dsDNA icosahedral lipid-containing halovirus	Dennis Bamford
HHIV-2		Dennis Bamford
HHTV-1		Dennis Bamford
<b><i>Haloarcula hispanica</i> plasmids</b>		
pHAR	<i>pyrF</i> -based suicide vector for genome editing in <i>H. hispanica</i> , <i>pyrF</i> , <i>Amp</i> <sup>R</sup>	Hua Xiang
pHAR-ΔBREX	Plasmid for BREX <sup>HAR</sup> system deletion, <i>pyrF</i> , <i>Amp</i> <sup>R</sup>	This work
pHAR-Δ X1-HI	Plasmid for X1-Z-X2-HI genes deletion, <i>pyrF</i> , <i>Amp</i> <sup>R</sup>	This work
pHAR-ΔN, N=A,B,C1,C2,X1,X2,Z,HI,NucS	Plasmid for deletion of individual <i>brx</i> genes, <i>pyrF</i> , <i>Amp</i> <sup>R</sup>	This work

## 2.2 Media and buffers

### **AS168:**

20% NaCl  
2% MgSO<sub>4</sub>·7H<sub>2</sub>O  
0.2% KCl  
0.3% trisodium citrate  
0.1% sodium glutamate  
0.005% FeSO<sub>4</sub>·7H<sub>2</sub>O  
MnCl<sub>2</sub>·4H<sub>2</sub>O (traces)  
0.5% casamino acids  
0.5% yeast extract  
pH 7.2

### **AS168YE-:**

AS168 without addition of yeast extract

### **LB:**

1% tryptone  
0.5% yeast extract  
1% NaCl

### **LB agar:**

LB  
1.5% agar

### **LB soft agar:**

LB  
0.7% agar

### **SW:**

24% NaCl  
3% MgCl<sub>2</sub>·6H<sub>2</sub>O  
3.5% MgSO<sub>4</sub>·7H<sub>2</sub>O  
0.7% KCl  
5% 1M CaCl<sub>2</sub>·2H<sub>2</sub>O  
0.5% 1M Tris-HCl pH 7.5

### **23% MGM:**

76.7% SW  
0.5% peptone (Oxoid)  
0.1% yeast extract (Oxoid)  
0.5% 1M Tris-HCl pH 7.5

### **20% MGM agar:**

66.7% SW  
0.5% peptone (Oxoid)  
0.1% yeast extract (Oxoid)  
1.5% agar  
0.5% 1M Tris-HCl pH 7.5

### **18% MGM top agar:**

60% SW  
0.5% peptone (Oxoid)  
0.1% yeast extract (Oxoid)  
0.4% agar  
0.5% 1M Tris-HCl pH 7.5

### **M9:**

11.3 g/L M9 salts  
1 mM MgSO<sub>4</sub>  
0.5 µg/mL thiamin HCl  
0.1% casamino acids  
100 µM CaCl<sub>2</sub>  
0.4% maltose

### **Revive MGM:**

23% MGM  
15% sucrose

### **SOB:**

2% tryptone  
0.5% yeast extract  
10 mM NaCl  
2.5 mM KCl  
10 mM MgCl<sub>2</sub>  
10 mM MgSO<sub>4</sub>

### **BSS-LS:**

5.85% NaCl  
0.2% KCl  
15% sucrose  
5% 1M Tris-HCl pH 8.2

**BSS-LS + glycerol:**

BSS-LS  
15% glycerol

**UBSS-LS:**

5.85% NaCl  
0.2% KCl  
15% sucrose  
pH 7.5

**SM:**

100 mM NaCl  
10 mM MgSO<sub>4</sub>  
0.01% gelatin  
50 mM Tris-HCl pH 7.5

**Stop solution:**

2% phenol  
75% ethanol  
8 mM EDTA  
20 mM sodium acetate pH 5.2

**STM:**

100 mM NaCl  
10 mM MgSO<sub>4</sub>  
50 mM Tris-HCl pH 7.5

**TE:**

1 mM EDTA  
10 mM Tris-HCl pH 8

**Inoue buffer:**

55 mM MnCl<sub>2</sub>  
15 mM CaCl<sub>2</sub>  
250 mM KCl  
10 mM PIPES pH 6.7

### 2.3 Cloning of BREX constructs

The table of all primers used in this work are listed in Appendix A Supplementary Table S1. PCR products were amplified with Phusion polymerase according to the manufacture's protocols. GeneJET PCR Purification Kit was used for purification of the PCR products from gel. The fragments digested with appropriate enzymes and ligated with a vector using T4 DNA ligase according to the manufacture's protocols. Chemocompetent DH5 $\alpha$  cells were transformed with a ligase mixture after 12 – hour incubation.

For experiments with fluorescent phage lambda (KanR) induction, a 14 kbp fragment of *Escherichia coli* HS genome (positions 340 559–354 275, NC 009800.1) was ligated into a pTG plasmid. This vector was constructed by ligation of a fragment of the p15A origin and chloramphenicol resistance gene amplified from pACYC184 (using primers

pACYC184\_F and pACYC184\_R) and a DNA fragment containing the arabinose promoter amplified from the pBAD30 plasmid (using primers pBAD30\_F and pBAD30\_R). The BREX<sup>Ec</sup> fragment used to construct pBREXAL was ligated with pTG fragment amplified using pTG-BREX\_F and pTG-BREX\_R primers.

Deletions of *brxA* and *brxC* on pBREX1 were created using outside PCR amplification of pBREX1 by delA\_BglII\_F and delA\_BglII\_R, delC\_BglII\_F and delC\_BglII\_R primer pairs, respectively. Internal fragments of *brxX* and *brxL* genes were deleted from pBREX2 by, correspondingly, SpeI/SdaI, and BglII restriction endonuclease digestion and religation. Deletions of *brxB* on pBREX1 and *brxZ* on pBREX2 were constructed by overlap extension PCR (93) with delB\_F1 and delB\_R1, delB\_F2 and delB\_R2, delZ\_F1 and delZ\_R1, delZ\_F2 and delZ\_R2 primers.

Deletion of *brxA* on pBREXAL plasmid was created by overlap extension PCR with primers Brx\_pro\_SacI\_F and DelA\_R, DelA\_F and DelB\_XbaI\_R, treated with SacI and XbaI and ligated with pBREXAL digested with the same restriction endonucleases. Deletion of *brxB* was constructed similarly with primers Brx\_pro\_SacI\_F and DelB\_R, DelB\_F and DelB\_XbaI\_R. pBREXALdelC was created by ligation of the PCR fragment of *brxA* and *brxB*, obtained by primers Brx\_pro\_SacI\_F and DelB\_XbaI\_R, and pBREXAL restricted by SacI and XbaI. Deletions of *brxX*, *brxZ*, *brxL* were created by restriction and religation of pBREXAL with, correspondingly, NdeI, MluI, BglII and NotI.

For deletion of the whole BREX<sup>HAR</sup> system, a 600 bp upstream region of *brxA* and a downstream sequence of *brxHI* were amplified with primers US\_ΔBREX\_F and US\_ΔBREX\_R, DS\_ΔBREX\_F and DS\_ΔBREX\_R, respectively. Two PCR fragments were merged by overlap extension PCR, treated with KpnI and BamHI, and ligated with pHAR plasmid digested with these restriction endonucleases. Ligation mixtures were transformed into *E. coli* DH5α. For construction of a partial deletion of BREX<sup>HAR</sup> (fragment *brxX1-brxZ-brxX2-brxHI*), the same procedure was repeated with primers US\_ΔX1-HI\_F and US\_ΔX1-HI\_R, DS\_ΔX1-HI\_F and DS\_ΔX1-HI\_R. For deletion of individual *brx* genes, pHAR-ΔN, where N = A, B, C1, C2, X1, X2, Z, HI, NucS, was

created with primers US\_ΔN\_F and US\_ΔN\_R, DS\_ΔN\_F and DS\_ΔN\_R. All plasmids were verified by DNA sequencing

#### 2.4 Plasmid transformation of *E. coli* cells

Competent *E. coli* BW25113 cells were prepared using the standard protocol (27). Briefly, overnight culture of *E. coli* was diluted 1:100 in SOB media and incubated at 18°C until OD<sub>600</sub> = 0.5. Cells were harvested by centrifugation at 3 000g for 10 minutes at 4°C. Supernatant was decanted and cells were resuspended in 50 ml of Inoue buffer. Washing step was repeated one more time, resuspended in 1 ml of Inoue buffer with addition of 7% DMSO, frozen in liquid nitrogen and stored at -80°C. Frozen chemocompetent cells were thawed on ice for 10 minutes and mixed with a ligation mixture or 25 ng of plasmid. After cultivation on ice for 25-30 minutes, cells were replaced to 42°C for 90 seconds and returned back on ice for 2-3 minutes. Cells were resuspended in 1 ml of LB media and incubated at 37°C for 1-1.5 hours. After incubation, cells were harvested by centrifugation, plated with appropriate antibiotic and the plate was stored at 37°C for 24 hours.

For plasmid transformation assay, competent *E. coli* BW25113 cells were transformed with 25 ng of pBTB-2 or pBREXAL plasmids. After 1.5 hours of incubation at 37°C, the mixture was serially diluted and plated on LB (to measure the number of living cells) and LB + kanamycin (to measure the amount of transformed cells) media. Plates were incubated at 37°C overnight. For each plasmid, transformation was repeated three times. Transformation efficiency was calculated as a ratio of antibiotic-resistant transformant colonies to the total number of colony forming units formed on LB plates.

#### 2.5 Plasmid transformation of *H. hispanica* cells

Overnight culture of *H. hispanica* was diluted 1:20 in 23% MGM supplemented with uracil (final concentration 50 μg/ml) and incubated at 37°C for 24 hours until OD<sub>600</sub> =

0.6-0.8. Cells (1-1.5 ml) were harvested by centrifugation at 6000 rpm for 3 minutes at room temperature (RT) and diluted in 200  $\mu$ l BSS-LS. Cells were centrifuged again and the pellet was resuspended into 100  $\mu$ l BSS-LS + glycerol. 10  $\mu$ l of EDTA was added to the mixture and incubated for 10 minutes at RT. Plasmid (300-500 ng) was added to cells and left for 5 minutes. 120  $\mu$ l of PEG<sub>600</sub>, diluted with UBSS-LS to final concentration of 60%, was mixed by tapping the tube until a homogeneous opalescent suspension is obtained. After 20-30 minutes of incubation at RT, 1 ml of revive MGM was added to remove PEG. Cells were centrifuged, the pellet was resuspended into 600  $\mu$ l of revive MGM and placed for 2-12 hours at 37°C. After recovering cells were plated on yeast extract-subtracted AS168 (AS168YE-) by spreading on plates and left at 42°C until colonies is obtained.

## 2.6 Knock-out of *brx* genes in *H. hispanica*

Deletions of *brx* genes were cloned by a pop-in-pop-out method (94). Briefly, *H. hispanica* cells were transformed with pHAR- $\Delta$ N and plated on AS168YE- to positively select the single-crossover recombinants (pop-in). Transformants were inoculated into 5 ml of AS168 medium supplemented with uracil (50  $\mu$ g/ml) and 5'-FOA (150  $\mu$ g/ml) to excise the plasmid (pop-out). When cultures reached OD<sub>600</sub>=0.6, cells were plated on AS168 plates with uracil and 5'-FOA. The mutants were screened by PCR analysis to confirm the deletions.

## 2.7 DNA purification

Plasmids were purified with Plasmid Miniprep Kit according to the protocols provided by Thermo Scientific. Archaeal and bacterial genomic DNA was extracted by GeneJet Genomic DNA Purification kit. Viral DNA was purified using Phage DNA Isolation Kit, Norgen, or by phenol-chloroform extraction. Before purification, archaeal virus lysate was diluted to obtain NaCl concentration up to 2%. The lysate was treated

with DNase I and RNase A at 37°C for 30 minutes. After addition of NaCl to the final concentration of 1 M and PEG<sub>6000</sub> to 10%, the mixture was kept at 4°C on rotating shaker for 1 hour or overnight. Viral particles were centrifuged 4000g for 1 hour at 4°C, then the pellet was resuspended in 400 µl of STM buffer, mixed with same volume of CH<sub>3</sub>Cl and vortexed for 1-2 minutes to elute PEG. After 5-minute centrifugation at 2000g 4°C, supernatant was collected and incubated with EDTA to the final concentration of 20 mM, Proteinase K to 100-200 µg/ml, SDS to 0.5% at 55°C for 1 hour. After proteolysis, 1V of phenol pH 7.8 was added and centrifuged at 10000g at 4°C for 10 minutes. Water phase was collected and mixed with 1V of phenol + CH<sub>3</sub>Cl and centrifuged again. Water phase was mixed with 1V of phenol + CH<sub>3</sub>Cl + Isoamyl alcohol (25:24:1). After addition of sodium acetate to the 300 mM and 2-3V of ethanol, DNA was left precipitated from water phase at -70°C for 30 minutes. Precipitate was sedimented at 10000g for 10 minutes at 4°C and washed with 70% ethanol twice and after one additional wash in 96% ethanol genomic DNA was dried and resuspended in 50 µl of TE buffer.

## 2.8 Efficiency of plaquing (EOP) assay

*E.coli* cultures were grown until OD<sub>600</sub> = 0.6 in LBMM medium (LB supplemented with 10 mM MgSO<sub>4</sub> and 0.2% maltose) with the addition of 50 µg/ml kanamycin, mixed with soft (0.7%) LBMM agar and poured on the surface of precast LBMM 1.5% agar plates. 10 µl aliquots of phage lysates or their serial (10<sup>-1</sup> -10<sup>-8</sup>) dilutions were deposited in drops on the surface of freshly poured lawns. After 18 hours of incubation at 37°C, EOP was determined as a ratio of phage titers on BREX<sup>+</sup> to BREX<sup>-</sup> lawns. The plaques were counted in each drop of phage dilution.

Late exponential phase culture of *H. hispanica* (150 µl) was mixed with 3 ml of 18% MGM top-agar supplemented with uracil (for  $\Delta$ *pyrF* strains) and poured on the surface of precast 20% MGM agar plates. Virus lysates were diluted and deposited in drops on the surface of lawns. After 72 hours of incubation at 30 – 37°C, EOP was determined as a ratio of phage titers on BREX<sup>+</sup> to  $\Delta$ BREX lawns.

## 2.9 Phage and virus propagation

Lysogenic *E. coli* cultures were grown at 30°C until OD<sub>600</sub> = 0.4 in LB with the addition of appropriate antibiotics, placed at 42°C for 15 minutes to trigger prophage induction and put at 37°C for phage propagation. The lysate was mixed with chloroform (1:100) to lyse any remaining cells, incubated at 37°C for 1 hour and centrifuged at 10000g for 15 minutes. The titer of phage  $\lambda$  in the supernatant was measured.

Single plaque of a virus was picked up and put into 10 ml of culture of *H. hispanica* at OD<sub>600</sub> = 0.6-0.8. Infected culture was incubated at 30-37°C depending on virus preference for 12-24 hours. Virus was separated from cells by 15-minute centrifugation 10000g, then the titer was measured.

## 2.10 Growth curves of *E. coli* infected cultures

Overnight cultures were diluted 1:100 into LBMM supplemented with appropriate antibiotic and grown at 37°C until OD<sub>600</sub> = 0.6. Phage  $\lambda$  *cI857 bor::Cm* was added to reach appropriate MOI and growth was monitored using EnSpire Multimode Plate Reader (PerkinElmer). At various times post-infection (0, 80 and 180 minutes), aliquots from infected cultures were taken to determine phage titer (PFU) and the number of living cells (CFU).

## 2.11 Adsorption assay for *E. coli* cells

Overnight cultures of BREX+ and BREX- cells were diluted 1:100 in LBMM with kanamycin. Cultures were grown until OD<sub>600</sub> = 0.6, mixed with phage  $\lambda$  *cI857 bor::Cm* at MOI=0.002 and placed in a rotary shaker at 37°C. 100  $\mu$ l culture aliquots were withdrawn at various times post-infection (0, 1, 3, 7, 15, 25 minutes), cells were pelleted by centrifugation at 10000g for 3 minutes and the titer of unabsorbed phage in the

supernatant was determined on BREX- cells. Percentage of unadsorbed phages was next calculated assuming the initial titer of phage (in the absence of cells) as 100%.

### 2.12 Lysogenization assay for *E. coli* cells

1 ml of overnight cultures of BREX+ and BREX- cells were diluted 1:100 in LBMM media with kanamycin, cultivated for 4 hours at 37°C, mixed with phage  $\lambda$  *cI857 bor::Cm* at MOI of 1 to 5 and placed at 30°C. After one-hour incubation, the bacteria-phage mixtures were serially diluted and plated on LB plates supplemented with kanamycin and kanamycin + chloramphenicol followed by overnight growth at 30°C. Lysogenization frequency was calculated as a ratio of the number of colonies grown on kanamycin + chloramphenicol plates to the number of colonies formed on plates with kanamycin only.

### 2.13 Fluorescence microscopy for *E. coli* cells

*Visualization of phage lambda induction.* Overnight cultures of LE392( $\lambda$ LZ1) lysogens (95) transformed with pTG or pTG-BREX plasmids were diluted 1:100 and cultivated in LB with chloramphenicol at 30°C. When OD<sub>600</sub> reached 0.4, cultures were transferred to 42°C for 15 minutes to trigger phage induction. 1-ml culture aliquots were centrifuged for 3 minutes at 4300g, cells were diluted in 300  $\mu$ l LB and 1  $\mu$ l was placed on an LB + 1.5% agarose slab (~1 mm thick) resting on a large 24  $\times$  50 mm coverslip (Fisher Scientific). After 1-minute drying the slab was covered by a small 18  $\times$  18 mm coverslip (Fisher Scientific).

*Visualization of injected phage DNA.* The procedure is described in detail elsewhere (29). Briefly, overnight cultures of LZ204 (96) transformed with pBTB-2 or pBREXAL plasmids were diluted 1:100 in M9 + 0.4% maltose (M9M) with kanamycin and allowed to grow at 37°C until OD<sub>600</sub> reached 0.4. 1-ml culture aliquots were centrifuged at 6000 rpm for 3 minutes and cells were resuspended in 150  $\mu$ l of cold M9M. For experiments

with propidium iodide staining, bacterial culture was mixed with propidium iodide to the final concentration of 20 $\mu$ M. 10  $\mu$ l of 10<sup>9</sup> fluorescent phage was mixed with the same volume of cells and incubated on ice for 30 minutes. The phage-cells mixture was diluted 1:3 in cold M9M and placed to 35°C for 5 minutes to trigger phage DNA injection. 1  $\mu$ l of the mixture was placed on 1.5% agarose M9M/kanamycin slab as described above.

Fluorescent phages were prepared based on standard protocols (96). Lysogenic culture of cells carrying the  $\lambda$ D-eyfp *cI857 bor::KanR* prophage and a plasmid expressing the wild-type  $\lambda$  gpD capsid protein (to avoid capsid instability) was transformed with pTG or pTG-BREX plasmids. BREX+ and BREX- phages were produced by heat induction from lysogenic cultures and purified as described previously (96) using ultracentrifugation in CsCl gradient.

Imaging was performed on a Nikon Eclipse Ti inverted epifluorescence microscope. Either 8 or 16 stages were used for time-lapse movies. In the first frame of the movie, phage was visualized via z-stacks ( $\pm$ 1.2  $\mu$ m, 0.3  $\mu$ m each step) with EYFP filter (200 ms exposure). During the movie, the sample was imaged in phase contrast for cells detection (100 ms), EYFP for phage (100 ms) and ECFP for SeqA (30 ms) channels.

#### 2.14 Pacific Bioscience sequencing

Extracted genomic DNA was sheared to a mean size of 500 bp using an ultrasonicator (Covaris) and purified with AMPure PB beads (Pacific Biosciences). PacBio sequencing libraries were prepared using the SMRTbell Template Prep kit 1.0 (Pacific Biosciences). Protocols for polymerase binding were generated by the Pacific Biosciences Binding Calculator. Sequencing was performed on the PacBio RS II (Pacific Biosciences). The PacBio SMRT Analysis software was used for reads align and modification-motif searches.

## 2.15 Phage lambda DNA degradation experiment

Overnight cultures of BREX+ and BREX- cells were diluted 1:100 in LBMM with kanamycin. Cultures were grown until  $OD_{600} = 0.6$ , mixed with phage  $\lambda$  *cI857 bor::Cm* at MOI=1 and placed in a rotary shaker at 37°C. At different times post infection (0, 2, 3, 5 and 15 minutes) 10 ml of infected cultures was mixed with ice-cold stop solution, cells were pelleted by centrifugation at 10000g for 5 minutes and total DNA was extracted by GeneJet Genomic DNA Purification kit. Library preparation and sequencing of DNA were performed by Skoltech Genomics Core Facility. DNA was sequenced on the NextSeq platform with 150+150 paired-end, 8 million reads for each sample. Reads quality was estimated with QC software and trimming was carried with the Trimmomatic program. Reads were mapped on bacterial and phage genomes, the ratio between them was calculated.

## 2.16. Restriction endonuclease activity assay

BREX+ or BREX- cells were incubated overnight without shaking at 37°C. *Escherichia coli* K12 BW25113 with the EcoRV R-M system components was used as a positive control. To prepare crude lysates 1 ml of overnight culture was spun down by centrifugation, resuspended in 1 ml of buffer (40 mM Tris-HCl pH 7.5, 150mM NaCl, 1mM ethylenediaminetetraacetic acid, 7 mM  $\beta$ -mercaptoethanol) and disrupted by sonication with a brief (5–10 s) impulse. Reactions were carried out in 20  $\mu$ l volume with 2  $\mu$ l of crude cell extracts and 200 ng of phage  $\lambda$  DNA for 30 min at 37°C using the following buffer: 10 mM Tris-HCl pH 7.5, 50 mM NaCl, 10 mM MgCl<sub>2</sub>, 0.1 mg/ml bovine serum albumin with optional addition of ATP to 1 mM.

## Chapter 3. Results

*The results section about BREX<sup>Ec</sup> is based on the paper published by the author (1) extended by new unpublished experiments.*

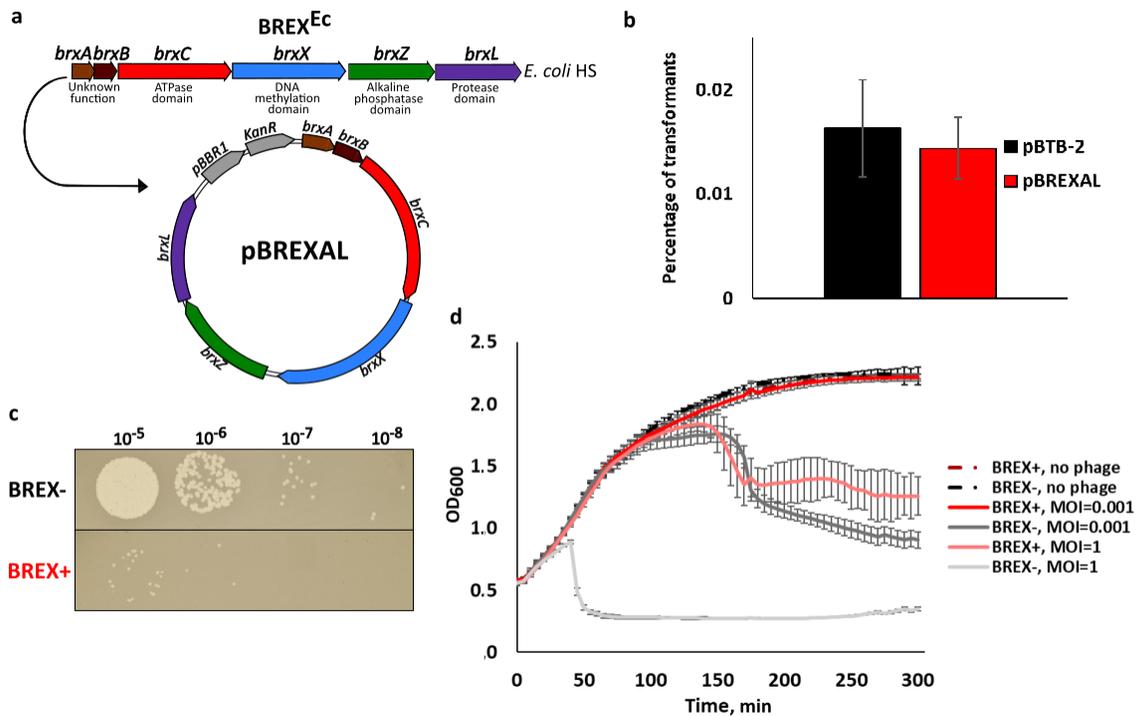
### 3.1 The *Escherichia coli* HS BREX system provides defense against phage $\lambda$ infection

Bioinformatics analysis reveals the presence of closely related six-gene *brxABCXZL* clusters in multiple *Escherichia coli* isolates (15). Previously, in our laboratory, *E. coli* HS natural isolate, which contains such BREX locus, was tested for resistance to several phages ( $\lambda$ , T4, T5, and T7) along with an isogenic strain that lacked it. All strains were found to be fully resistant to these phages. Thus, no conclusions about the contribution of the *brx* cluster to phage resistance could be made. The entire *E. coli* HS *brx* cluster was therefore cloned, together with upstream sequences, on a low-copy *E. coli* pBTB-2 plasmid. The resulting plasmid was named pBREXAL (Fig. 11a). A laboratory *E. coli* K12 strain BW25113, lacking endogenous *brx* genes, was transformed with pBTB-2 and pBREXAL plasmids with equal efficiency (Fig. 11b).

Below, we refer to *E. coli* HS cluster as BREX<sup>Ec</sup>; BW25113 cells carrying plasmid-borne BREX<sup>Ec</sup> are referred to as BREX<sup>+</sup>; control cells carrying the pBTB-2 vector are referred to as BREX<sup>-</sup>.

To test whether the BREX<sup>Ec</sup> system provides protection against  $\lambda$  phage infection, a modified phage carrying the *cl*<sub>857</sub> mutation in the *cl* gene coding for phage repressor, and marked with a chloramphenicol resistance gene that replaced the *bor* gene (encoding phage outer membrane lipoprotein) was obtained by thermal induction from lysogenic *E. coli* MC4100. The ability of phage  $\lambda$  to form plaques on BREX<sup>+</sup> and BREX<sup>-</sup> lawns or infect liquid cultures was tested. The efficiency of plaquing (EOP) was 100-fold reduced on BREX<sup>Ec</sup>-containing cells compared to the control strain transformed with the empty plasmid (Fig. 11c). No change in growth dynamics of uninfected BREX<sup>+</sup> and BREX<sup>-</sup> bacteria in liquid cultures was observed (Fig. 11d). At multiplicity of infection (MOI) of 0.001, the BREX<sup>-</sup> culture collapsed after 180 minutes of infection while the BREX<sup>+</sup> culture continued growing as the uninfected culture (Fig. 11d). At MOI of 1, BREX<sup>+</sup> cultures grew at the rate of uninfected control for ~100 minutes but at later times optical

density of the culture slowly declined (Fig. 11d). Thus, BREX<sup>Ec</sup> is functional in phage defense similarly to the BREX system from *B. cereus*.



**Figure 11. BREX<sup>Ec</sup> protects cells from phage  $\lambda$  infection.**

**a.** The BREX gene cluster from *Escherichia coli* HS is schematically shown; bioinformatically predicted putative functions of *brx* gene products are listed. The entire BREX<sup>Ec</sup> cluster was ligated into the pBTB-2 vector to yield the pBREXAL plasmid.

**b.** Efficiency of transformation of empty pBTB-2 vector and the pBREXAL plasmid into laboratory BW25113 *E. coli*. Obtained transformants are referred to, correspondingly, as BREX<sup>-</sup> and BREX<sup>+</sup> cells.

**c.** Lawns formed by BREX<sup>+</sup> and BREX<sup>-</sup> cells were spotted with indicated dilutions of  $\lambda$  phage lysate. Results of overnight growth at 37°C are shown.

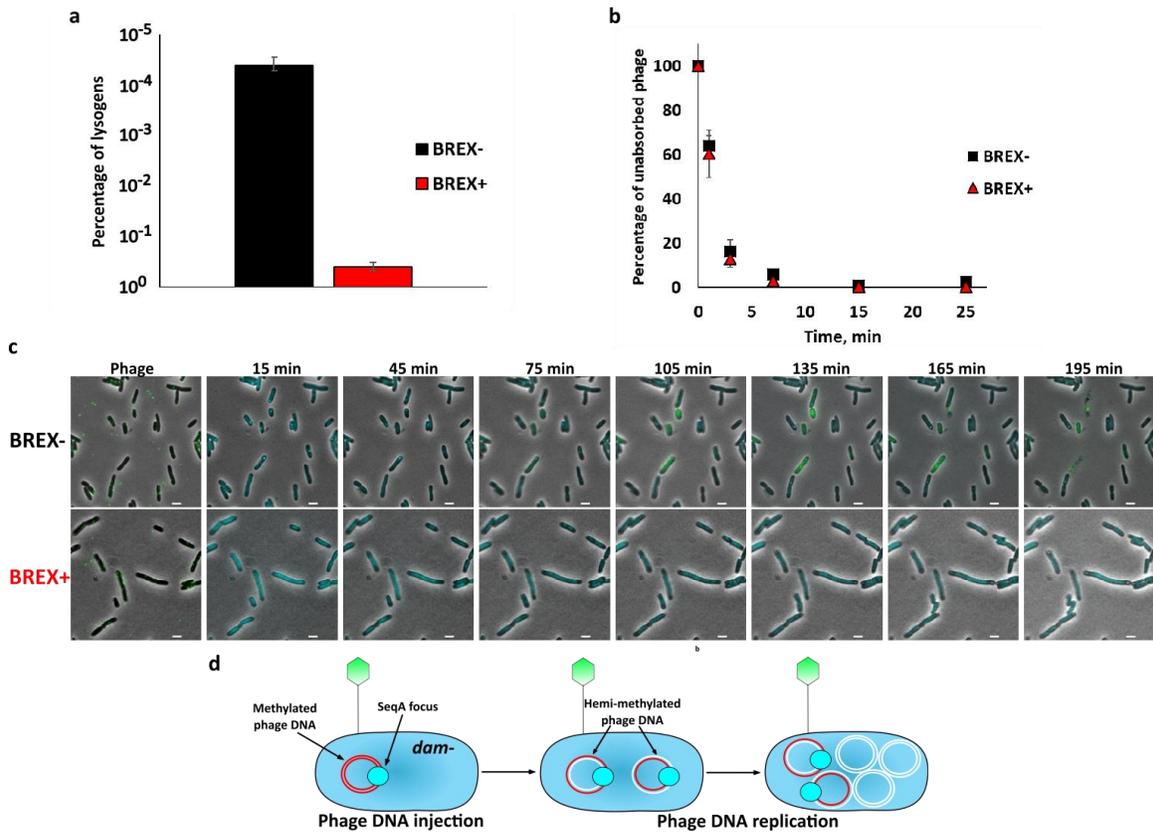
**d.** Growth curves of BREX<sup>+</sup> and BREX<sup>-</sup> cultures in the absence of infection, and during infection with  $\lambda$  phage at MOI of 0.001 and 1. Phage was added at t=0. Each growth curve shows mean optical density values and standard deviations obtained from three independent experiments. The lines for BREX<sup>+</sup>/<sup>-</sup> with no phage addition are masked by the line BREX<sup>+</sup> MOI=0.001.

The figure is taken from (1).

Previous experiments with the *B. cereus* BREX system using a PCR assay demonstrated that BREX abolished lysogenization (15). To test if BREX<sup>Ec</sup> also affects lysogenization, BREX<sup>+</sup> and BREX<sup>-</sup> cultures were infected with phage  $\lambda$  at MOI ~ 5 and after 1-hour incubation cells were plated on selective media containing chloramphenicol,

where only lysogens could form colonies. Unlike the situation reported for the BREX system from *B. cereus*, a few chloramphenicol-resistant colonies were observed in BREX+ culture, though there was 10<sup>4</sup>-fold reduction in colonies numbers (Fig. 12a). Since lysogenization does not require phage DNA replication (97) but only successful adsorption and DNA injection, the result indicates that the defensive action of BREX<sup>Ec</sup> manifests itself either at the stage of phage adsorption or during injection of phage DNA. However, as can be seen from Fig. 12b, the BREX<sup>Ec</sup> system does not affect phage lambda adsorption in the BREX+ culture. A similar observation was made when phage adsorption to *B. subtilis* cells with or without BREX system was monitored (15).

Using live fluorescence microscopy, the process of phage infection was monitored using phage  $\lambda$  with Dam-methylated DNA and capsid-decoration protein fused to EYFP, while *dam*- *E. coli* MG1655 strain encoded SeqA fused to ECFP. The system allows to monitor phage DNA injection and replication processes due to specific binding SeqA to methylated or hemimethylated *dam* sites in invading DNA (96, 98). Bound SeqA forms distinct fluorescent foci on injected viral DNA: one or two dots correspond to methylated (injection) or hemimethylated (after replication) DNA, respectively (Fig.12d). Under our conditions (MOI = 1 - 5), approximately 53% of BREX- cells in the culture were infected as judged by the appearance of at least one SeqA focus (Fig.12c). In some infected cells two SeqA foci appeared, followed by accumulation of yellow fluorescence due to the synthesis of capsid-decoration protein::EYFP and further lysis. BREX- cells where only one SeqA focus was observed survived and likely underwent lysogenic conversion. In contrast, in the BREX+ culture no ECFP foci or EYFP fluorescence accumulation was observed. The infected cells stopped dividing and ~18% of 911 cells examined became mildly elongated. Judging by the lack of propidium iodide staining, the membrane of infected BREX+ cells remained intact (data not shown) and no drop in CFU was observed when infected cells were deposited on agar plates. Thus, the BREX<sup>Ec</sup> system prevents SeqA accumulation which may indicate a rapid phage DNA degradation or inhibition of invading DNA injection.



**Figure 12. BREX<sup>Ec</sup> effects on lysogenization, adsorption, and DNA injection by bacteriophage  $\lambda$ .**

**a.** The bars show the numbers of chloramphenicol-resistant lysogenic colonies formed after high-MOI infection of BREX+ and BREX- cells with  $\lambda$  phage marked with a chloramphenicol resistance cassette.

**b.** The experiment was conducted as in **a** but at an MOI of 0.01. At indicated times, cells were removed by centrifugation and unadsorbed infectious phage particles in the supernatant were counted on BREX- cells lawns.

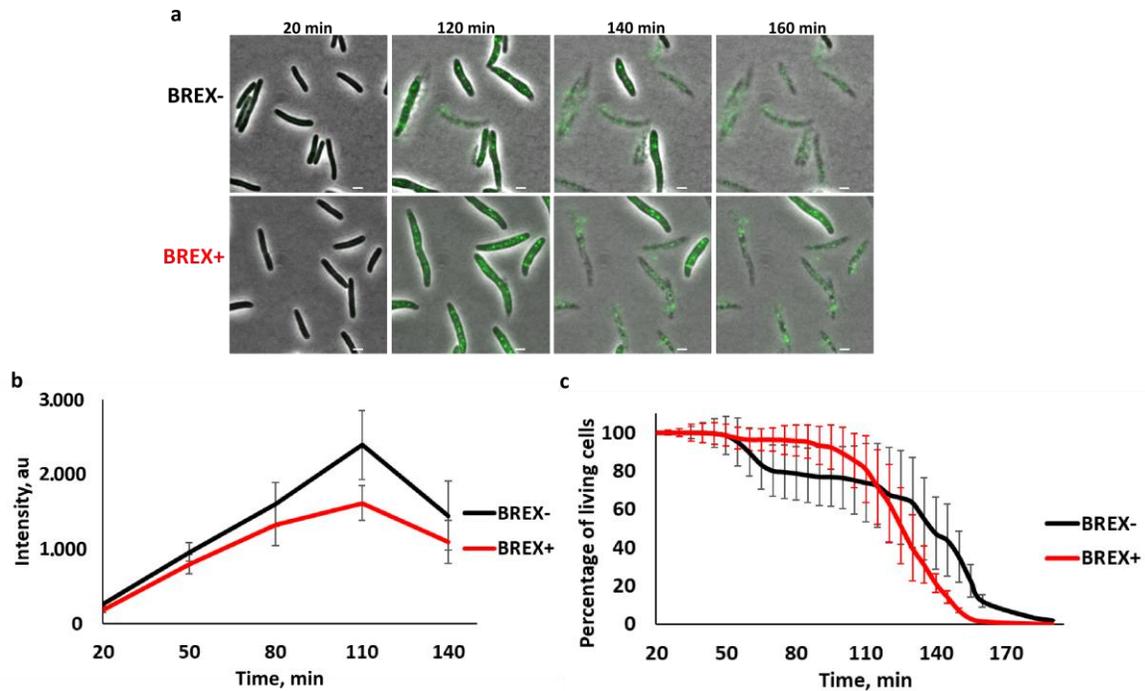
**c.** Live microscopy of BREX- and BREX+ cells infected with phage  $\lambda$ . Images from a time-lapse movie show phage DNA injection. In the first picture, the fluorescent phage appears as a green dot on the cell surface. At 15 min, the SeqA-ECFP foci accumulate as one or two cyan dots, representing the injected and replicated phage DNA, respectively. Scale bar, 2  $\mu$ m.

**d.** Schematics of the experiment whose results are shown in **c**. Phage  $\lambda$  with capsid-decoration protein fused to EYFP injects Dam-methylated DNA that attracts host SeqA-ECFP and leads to formation of a focus. Upon replication of methylated DNA, two hemi-methylated DNAs are produced and thus two fluorescent foci appear.

The figure is modified from (1).

### 3.2 BREX<sup>Ec</sup> has no effect on lysogenic induction

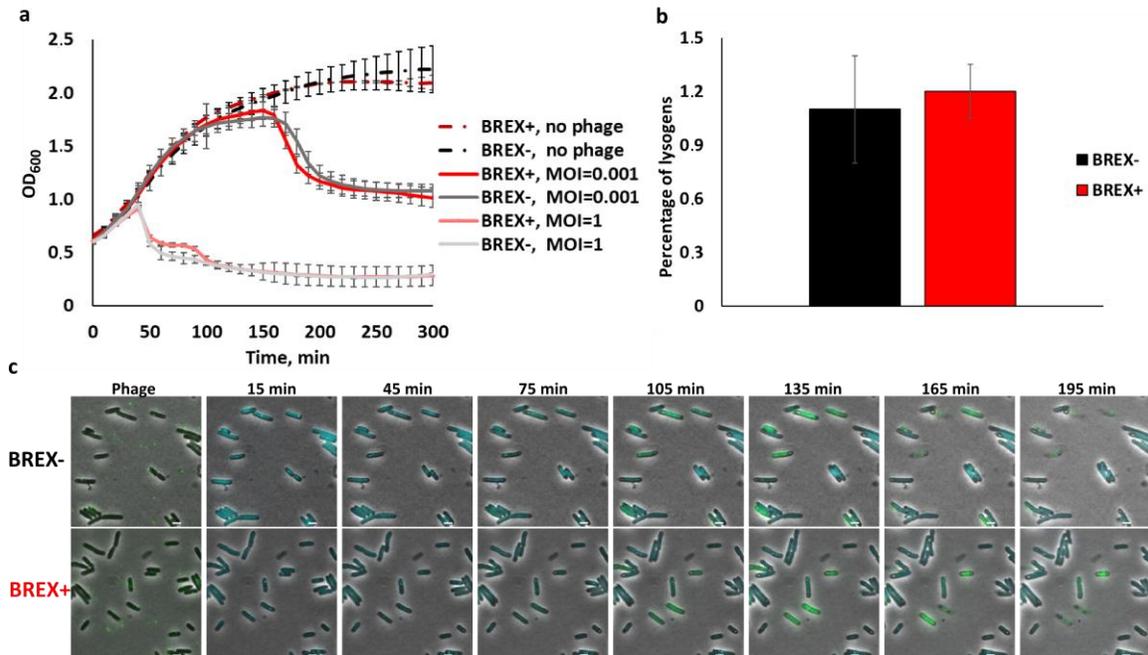
To examine if the BREX system affects phage induction, lysogens carrying fluorescent-labeled phage lambda  $\lambda D\text{-}EYFP\text{ }cI_{857}\text{ }bor::Kan^R$  were transformed with either an empty vector or a plasmid with the BREX<sup>Ec</sup> system. Grown to OD<sub>600</sub>=0.4, BREX+ and BREX- lysogenic cultures were shifted to 42°C for 20 minutes to trigger prophage induction. The process of phage development was monitored using live fluorescent microscopy (Fig. 13a).



**Figure 13. Induction of  $\lambda$  prophage from BREX+ lysogens.**

- a.** Images of cells from BREX- and BREX+ lysogenic cultures taken at indicated times after thermal induction. The induced phage encodes capsid decoration protein fused to EYFP. Scale bar, 2  $\mu$ m.
- b.** Changes in EYFP fluorescence in induced cells are presented at various times post-induction. Mean numbers were calculated from fluorescence intensities obtained with ca. 300 cells. Standard deviations of mean numbers obtained in three independent experiments are shown.
- c.** Quantification of a representative kinetic series showing decrease in live cells during microscopic observation of induced lysogenic cultures.
- The figure is taken from (1).

Shortly after the induction, both BREX<sup>+</sup> and BREX<sup>-</sup> cells became fluorescent, indicating accumulation of the capsid-decoration protein fused to EYFP. At later times, fluorescent speckles corresponding to assembled phage capsids appeared inside the cells. The intensity of diffuse cytoplasmic fluorescence and the number of dots were nearly the same in BREX<sup>+</sup> and BREX<sup>-</sup> cells (Fig. 13b). Most cells in both induced cultures lysed ~3 hours post-induction and the kinetics of cell lysis was the same in BREX<sup>+</sup> and BREX<sup>-</sup> cultures (Fig. 13c). Therefore, we concluded that BREX<sup>Ec</sup> has no effect on lysogenic induction.



**Figure 14. Phage  $\lambda$  induced from BREX<sup>+</sup> lysogens is not subject to inhibitory action of BREX<sup>Ec</sup>.**

**a.** Growth curves of liquid cultures of BREX<sup>+</sup> and BREX<sup>-</sup> cells in the absence of infection, and during infection with  $\lambda$  phage obtained after induction of BREX<sup>+</sup> lysogens. Results of infections at MOI 0.001 and 1 are presented. In infected cultures, phage was added at  $t=0$ .

**b.** The bars show the number of chloramphenicol-resistant lysogenic colonies formed after infection (MOI=5) of BREX<sup>+</sup> and BREX<sup>-</sup> cells with phage  $\lambda$  phage obtained after induction of BREX<sup>+</sup> lysogens (the phage is marked with a chloramphenicol resistance cassette).

**c.** Images of BREX<sup>+</sup> and BREX<sup>-</sup> cultures infected (MOI=1) with phage  $\lambda$  induced from BREX<sup>+</sup> lysogens. Scale bar, 2  $\mu$ m.

The figure is taken from (1).

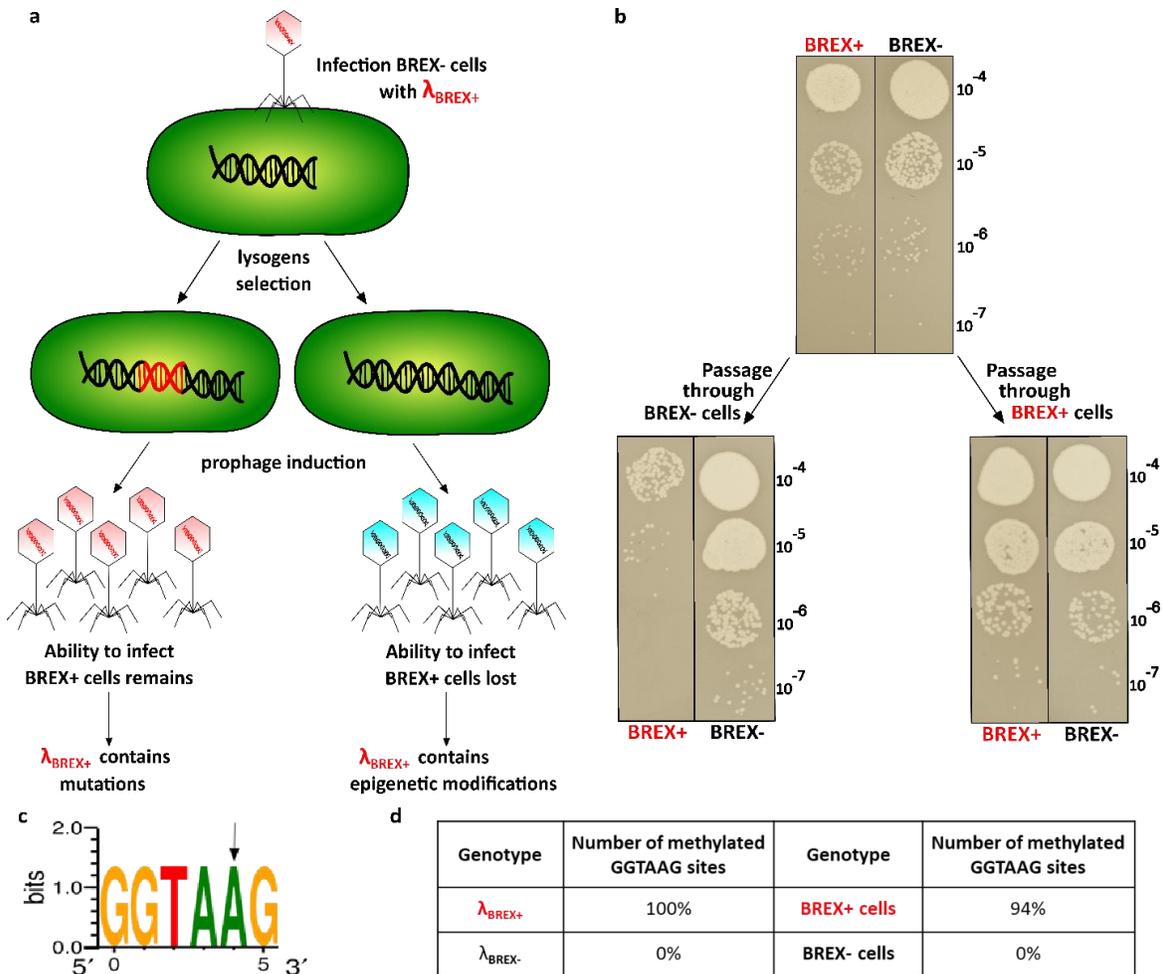
Phages collected after the induction of BREX<sup>+</sup> lysogens were used for infection of non-lysogenic BREX<sup>+</sup> and BREX<sup>-</sup> cultures. This phage efficiently infected both cultures (Fig. 14a). Moreover, no difference in the lysogenization level of BREX<sup>+</sup> and BREX<sup>-</sup> cultures with the phage was observed (Fig. 14b). Live fluorescence microscopy experiments using SeqA-ECFP cells also demonstrated that phage induced from the BREX<sup>+</sup> lysogens infected both BREX<sup>+</sup> and BREX<sup>-</sup> cells normally (Fig. 14c). Thus, the BREX<sup>+</sup> system is unable to protect cells from phages that were generated by the BREX<sup>+</sup> host.

### 3.3 Phage that overcomes BREX<sup>Ec</sup> protection contains modified DNA

The BREX<sup>Ec</sup> system either mutated or modified phage DNA during the induction. To gain further insight into the mechanism, phage progeny formed after infection of BREX<sup>-</sup> cells with phage induced from BREX<sup>+</sup> lysogens (Fig. 15a) was collected and spotted on BREX<sup>+</sup> and BREX<sup>-</sup> lawns. As can be seen from Fig. 15b, after a passage through the BREX<sup>-</sup> culture, phage was no longer able to infect BREX<sup>+</sup> cells efficiently (Fig. 15b).

In contrast, progeny phage produced after passage through BREX<sup>+</sup> cells continued to infect BREX<sup>+</sup> cells well (Fig. 15b). The result suggests that phages induced from BREX<sup>+</sup> lysogens contain an epigenetic modification.

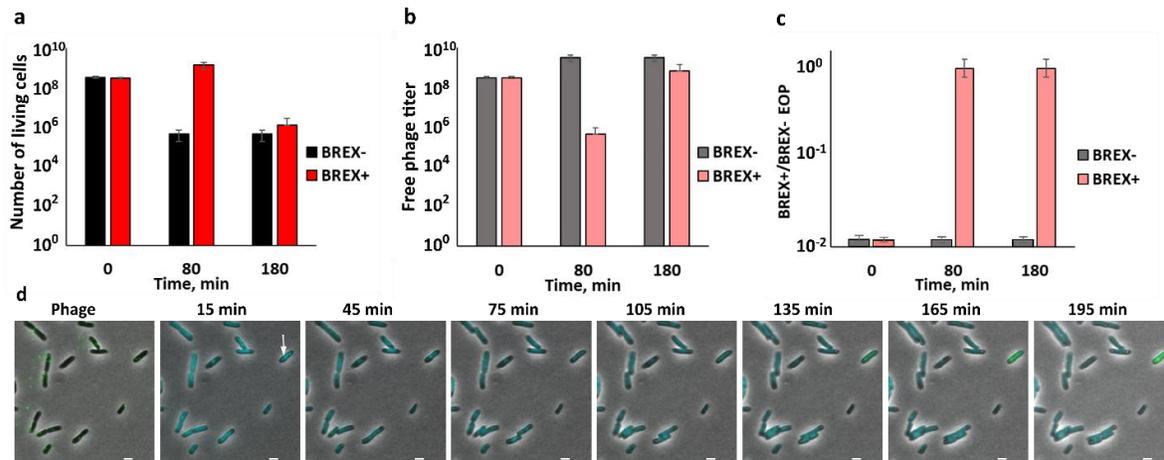
DNA prepared from phage virions induced from BREX<sup>+</sup> lysogens was sequenced on a PacBio platform that detects a range of base modifications (99). The results showed that all 18 GGTAAG sites present in the phage  $\lambda$  genome were methylated at the fifth adenosine residue (Fig. 15c, d). No such modification was present in DNA of phages induced from BREX<sup>-</sup> cells. Sequencing of BREX<sup>+</sup> cells DNA showed that 94% of the 1708 genomic GGTAAG sites were also modified. No other BREX<sup>+</sup> specific modifications were detected. (Fig. 15d). Moreover, no *brx*-mediated modification was detected in BREX<sup>-</sup> DNA. Therefore, BREX<sup>Ec</sup> methylates GGTAAG sites and BREX<sup>Ec-</sup> modified phages can productively infect BREX<sup>+</sup> cells.



**Figure 15. BREX system epigenetically modifies phage  $\lambda$ .**

- a.** The scheme shows to possible ways of appearance of phages that overcome BREX action.
- b.** At the top, lawns of BREX+ and BREX- cells were spotted with indicated dilutions of  $\lambda$  phage lysate obtained after induction of BREX+ lysogens. Results of overnight growth at 37°C are shown. Phages from a plaque obtained on BREX+ cells were used to re-infect BREX+ or BREX- cells. The results of spotting phage progeny from these infections on BREX- and BREX+ cell lawns are shown below.
- c.** The site modified in genomic DNA of *E. coli* BW25113 containing either the pBTB-2 or pBREXAL plasmid, and in the genome of phage  $\lambda$  induced from BREX+ lysogens. The arrow shows the site of BREX-dependent methylation.
- d.** Statistics of BREX site modification in BREX+ cells genome and in phage induced from these cells. The figure is taken from (1).

Based on the epigenetic mechanism of overcoming the BREX protection, we reexamined infections of BREX+ cells with unmodified phage from Fig. 11d. At MOI=1, the number of surviving cells in BREX- culture was 1000-times less than in BREX+ cultures after 80 minutes of infection (Fig. 16a). Nevertheless, at a later time point, the number of surviving cells became equally low in both cultures.



**Figure 16. Modified phage appears in the course of infection of BREX+ cells and overcomes protection.**

- a.** The amount of colony forming units (CFU) in BREX+ and BREX- cultures infected at the MOI of 1 and  $t=0$  is shown 80 and 180 minutes post-infection.  
**b.** The amount of plaque forming units (PFUs) in BREX+ and BREX- cultures from panel **a**.  
**c.** EOP of phages collected from infected cultures from panel **b**. EOP was determined by calculating the ratio of phage titers on BREX+ and BREX- cell lawns.  
**d.** Live microscopy observation of BREX+ cells infected (MOI=1) with phage  $\lambda$ . An arrow shows a productively infected cell. Scale bar, 2  $\mu\text{m}$ .  
The figure is taken from (1).

Phage titer in infected cultures was tested on BREX- and BREX+ lawns. The ratio of these two titers, EOP, allows one to determine the state of phage DNA modification. Ca. 1% of initial phage used for infection was able to form plaques on BREX+ cells (see also Fig. 11c). During infection of BREX- cells, the overall phage titer increased but the proportion of phages able to form plaques of BREX+ lawns remained the same (Fig. 16b, c). In infected BREX+ cultures, the overall phage titer decreased

dramatically (10,000 times) 80 minutes post-infection but all phages were able to infect BREX+ cells. At later times, the phage titer grew, and the ability to infect BREX+ cells with an EOP of 1 was retained. We conclude that during the infection of BREX+ culture with unmodified phage, modified phages appear in low frequency and then proceed to overtake the population. It is the appearance of these modified phages that must be responsible for the drop in optical density of BREX+ cultures shown in Fig. 11d. Live fluorescence microscopy analysis is consistent with this interpretation. While most BREX+ cells were not host to productive infection, CFP foci appeared in rare (~1%) cells followed by subsequent accumulation of EYFP fluorescence and cell lysis, indicating productive infection (an example of such behavior can be seen in Fig. 16d).

### 3.4 The role of individual *brx* genes in BREX<sup>Ec</sup> protection and DNA modification

In contrast to the situation in R-M systems, deletion of predicted SAM-methyltransferase BrxX from the pBREGAL plasmid was not toxic for cells. Moreover, *brxX*-lacking culture was sensitive to phage  $\lambda$  and did not contain any DNA modifications (Fig. 17). This effect was complemented by the introduction of a compatible *brxX* expression plasmid.

The role of other *brx* genes on phage protection/DNA modification was also investigated. Deletion of a predicted Lon-like protease BrxL from pBREGAL led to unintended complete deletions of *brxX* and *brxZ* and partial deletion in *brxC* (Fig.17a). Likewise, deletion of a putative alkaline phosphatase BrxZ resulted in 17-bp deletion in *brxX*. The only genes that we successfully deleted without mutations were *brxB* and *brxX*.

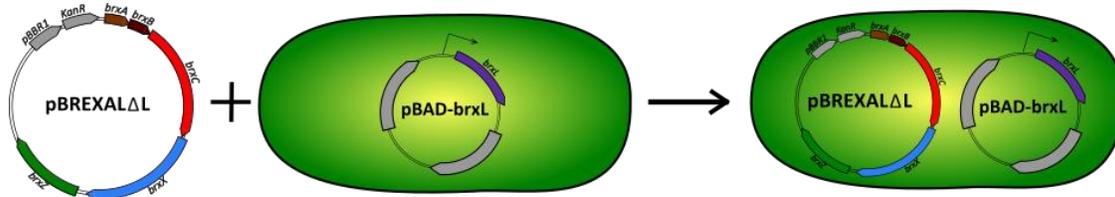
Due to our inability to clone plasmids with *brx* deletions, *E.coli* BW25113 carrying the deleted gene on a separate inducible complementation plasmid was transformed with a ligation mixture containing a plasmid with a deleted gene to complete the system and avoid toxicity (Fig.17b). For *brxZ* and *brxL*, the complementation system was obtained. Curiously, pBREGAL plasmid with deleted *brxA* cannot be complemented

with a plasmid carrying a single *brxA* gene but only with a *brxA-brxB* pair. It may happen due to problems with transcription or translation of *brxB* as the stop codon of BrxA overlaps with the start codon of BrxB. Such constructions are transcribed as a single mRNA and, thus, deletion of *brxA* destroys transcription and translation of downstream gene *brxB*. No colonies with deleted *brxC* gene were found in the complemented system. The failure to complete the system could be explained by *brx* gene regulation interruptions due to the usage of an inducible promoter for *brxC*.

**a**

<i>brx</i> gene	pBEXALdelN construction mutations	pBEXALdelN + pBAD- <i>brxN</i> complementation
<i>brxA</i>	frameshift in <i>brxZ</i>	Only with pBAD- <i>brxAB</i> but not with pBAD- <i>brxA</i>
<i>brxB</i>	-	+
<i>brxC</i>	transposase in <i>brxX</i>	-
<i>brxX</i>	-	+
<i>brxZ</i>	frameshift in <i>brxX</i>	+
<i>brxL</i>	partial deletion of <i>brxC</i> and complete deletions of <i>brxX-brxZ</i>	+

**b**



**Figure 17. Deletions of individual *brx* genes on pBEXAL plasmid.**

**a.** Each *brx* gene was deleted from the pBEXAL plasmid. Only *brxB* and *brxX* can be deleted without accumulation of additional mutations. To prove the toxicity of constructions with deletions without secondary mutations, *E.coli* BW25113 carrying the deleted gene was transformed with ligated mixtures of plasmids with deleted genes to complement the system.

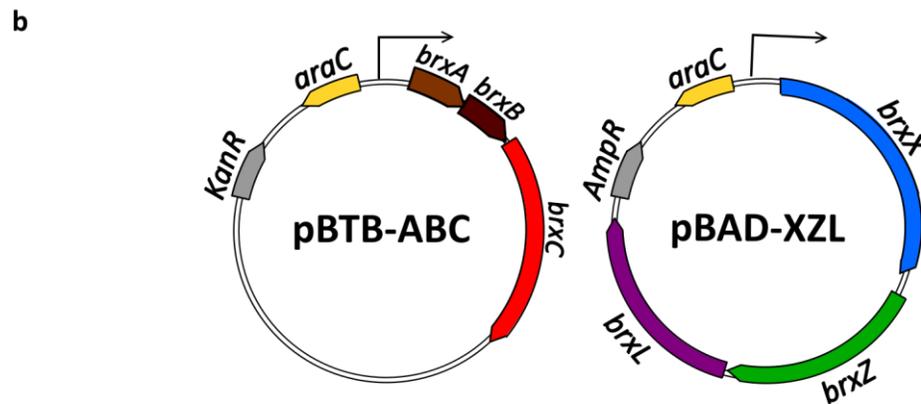
**b.** Schematics of cloning of toxic pBEXAL plasmid derivatives with deletions via complementation.

To tightly control the expression of *brx* genes, we created two compatible plasmids separately expressing *brxABC* and *brxXZL* from arabinose inducible promoters.

This two-plasmid system allowed, in the presence of arabinose, the same protection of cells from  $\lambda$  infection as pBEXAL. The genome of phage  $\lambda$  induced from lysogenic cells containing the full complement of *brx* genes on two plasmids and grown in the presence of arabinose was modified at GGTAAG sites (Fig. 18). As expected, deletion of the *brxX* gene in the context of the two-plasmid system led to the absence of protection and modification. Deletion of *brxA* affected neither protection from phage infection nor modification of phage/host DNA. Deletions of *brxB*, *brxC*, and *brxZ* abolished both protection and modification.

**a**

Genotype	BREX methylation	BREX defense
pBEXAL $\Delta$ X	-	-
BrxX only	-	-
$\Delta$ A	+	+
$\Delta$ B	-	-
$\Delta$ C	-	-
$\Delta$ X	-	-
$\Delta$ Z	-	-
$\Delta$ L	+	-
BREX+ (2 plasmids)	+	+



**Figure 18.** The state of methylation of host DNA and ability to withstand phage infection by cells carrying BREX plasmids lacking indicated genes.

**a.** Cultures carrying the *brxX* gene, deletion of *brxX* on pBREXAL, or indicated deletions in the two-plasmid BREX system were tested for the defense activity by EOP and for BREX modification by PacBio sequencing of host genomes.

**b.** Schematics of the two-plasmid BREX system.

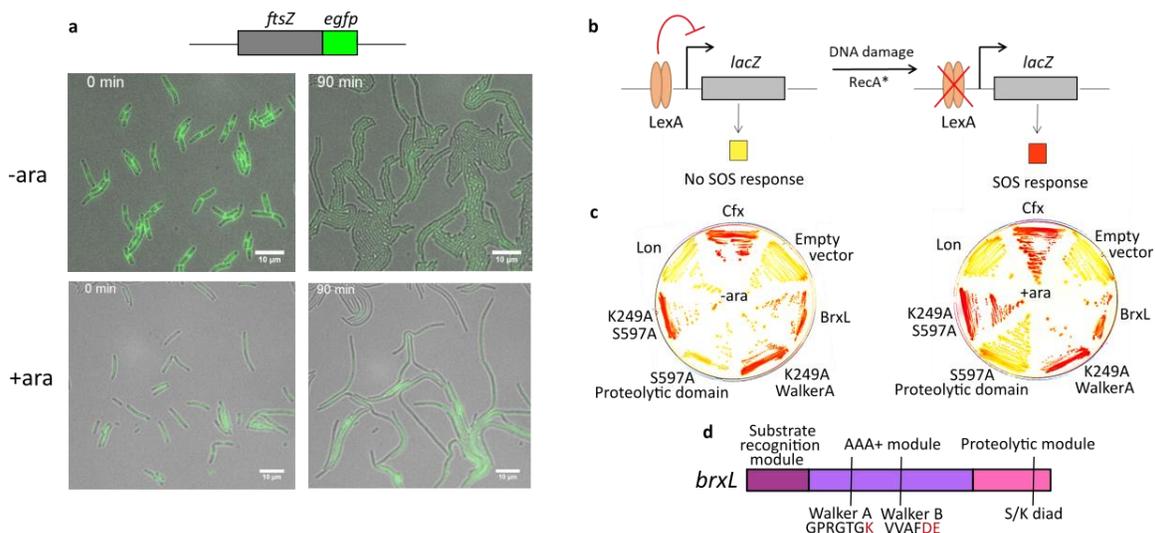
The figure is modified from (1).

The most mysterious was the deletion of *brxL* from the inducible BREX system. Interestingly, introduction of the *brxABCXZ* cluster resulted in the appearance of two types of colonies. Cells of the first type formed normal *E. coli* colonies, however, these colonies could not grow in liquid culture. Colonies of the other type were small and transparent and grew in liquid culture. Real-time PCR analysis revealed that expression of the *brxABCXZ* cluster caused a decrease in the copy number of the *brxXZ*-carrying plasmid (data not shown). Moreover, phage progeny induced from these cells was not completely methylated (17 rather than 18 GGTAAG sites modified in phage genome). Such behavior can be explained by exclusion of the BREX methyltransferase-containing plasmid by cells. In addition, this deletion abolished protection from infection.

Individual expression of *brxA*, *brxB*, *brxC*, and *brxZ* had no effect on cell growth and did not lead to protection from phage infection. Expression of *brxX* mildly inhibited cell growth and had no effect on phage infection. DNA prepared from these cells was not modified. Thus, BrxX alone is not sufficient for DNA modification by the BREX system.

At the same time, expression of *brxL* was highly toxic: cell growth ceased after induction and the culture density gradually declined with time. The experiment with fluorescent labeled FtsZ cells (100) showed that the present of BrxL prevented the formation of FtsZ-ring and caused filamentation (Fig. 19a). To check if induction of BrxL leads to SOS-response, a strain with *lacZ* gene under SOS-response promoter was used (Fig. 19b). Unlike the Lon protease, the Lon-like BrxL induced SOS-response even when arabinose is not added. A leak of arabinose promoter in the absence of the inducer leads to a low level expression of the protein. Substitution of a serine S597 in the serine-lysine proteolytic diad of BrxL abolished toxicity (Fig. 19b, c). Such modification of BrxL on pBREXAL background had no effect on either methylation or on defense activity of the BREX system (data not shown). Mutation in ATP-binding Walker A motif did not affect SOS-response and a double mutation K249A S597A in BrxL also did not

change the phenotype. Moreover, purified BrxL protein did not show any proteolytic activity with fluorescein isothiocyanate (FITC)-labeled casein as a substrate (Abcam's Protease Activity Assay Kit) but possesses an ATP-binding activity (data not shown). The absence of cleavage for BrxL may be explained by non-optimal distance between serine and lysine in the proteolytic domain.



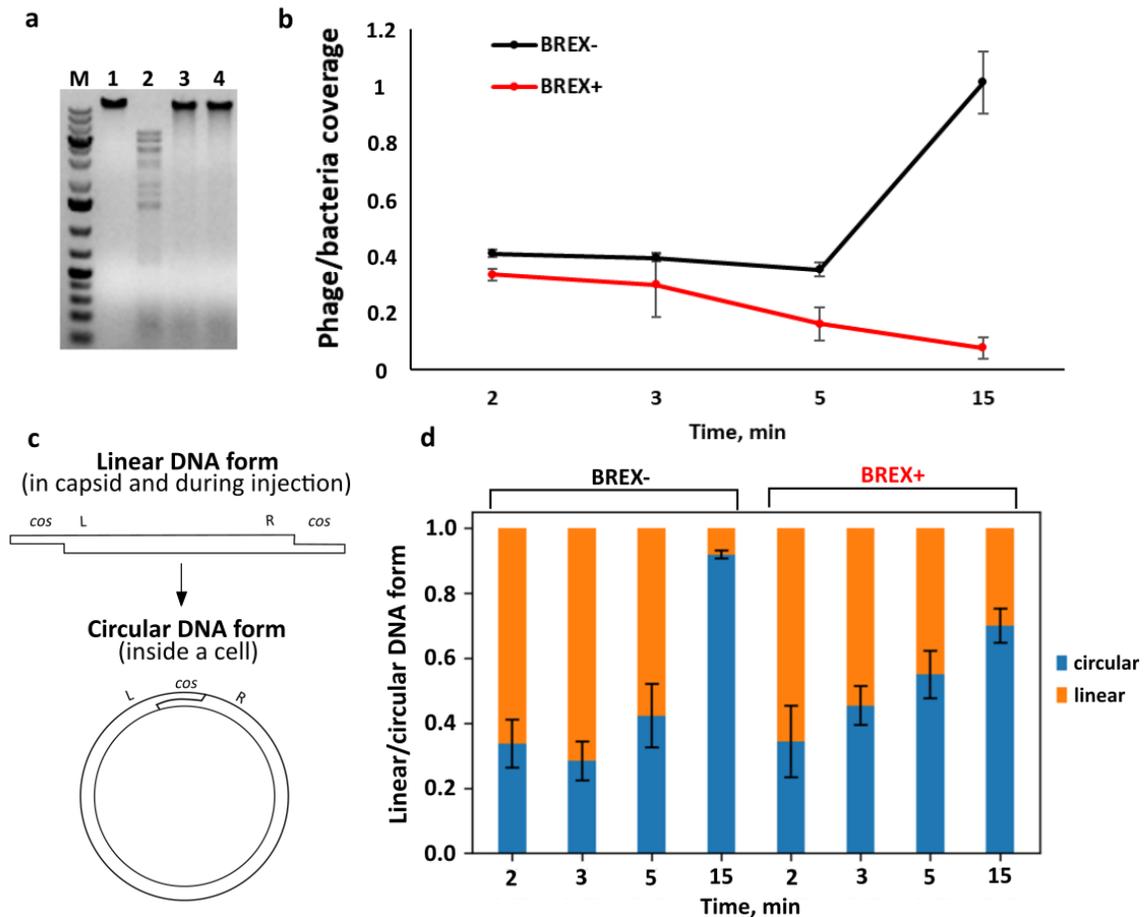
**Figure 19. BrxL inhibits FtsZ-ring formation and induces SOS-response.**

- a.** Live microscopy observation of FtsZ-EGFP *E. coli* strain with the BrxL expression plasmid. Scale bar, 10  $\mu$ m
- b.** Schematics of the SOS-response experiment. The *lacZ* gene is under the *sulA* promoter control and is repressed by LexA. After DNA damage, activated RecA facilitates LexA self-cleavage and thus *lacZ* is transcribed. Cells expressing LacZ hydrolyze lactose that leads to changes in pH and in colony colors.
- c.** Plates with and without the addition of arabinose after SOS-response experiment depicted in **b**. Cells with empty vector serve as a negative control. Cells with empty vector covered with ciprofloxacin (Cfx) are used as a positive control (SOS-response expected). Other sectors in clockwise order are cells expressing BrxL; BrxL with K249A mutation in Walker A motif; BrxL with S597A substitution in the proteolytic domain; BrxL with double K249A, S597A mutation; and Lon protease.
- d.** Domains organization of BrxL.

### 3.5 Evidence of phage DNA degradation in BREX<sup>Ec</sup>+ cells

The data presented so far are consistent with a mechanism of BREX action that is similar to the classical R-M system action, though none of the *brx* genes products have a

predicted nuclease function. When extracts of BREX<sup>+</sup> cells were combined with unmodified  $\lambda$  DNA, no cleavage was observed, compared to BREX<sup>-</sup> cells lysates (Fig. 20a). In contrast, extract of cells containing a plasmid with the EcoRV Type II R-M system (21 recognition of sites in  $\lambda$  genome) readily cleaved  $\lambda$  DNA at these conditions (these experiments were done by Artem Isaev).



**Figure 20. BREX<sup>Ec</sup> system degrades phage DNA *in vivo*.**

- a.** Cell extracts prepared from *E. coli* cells harboring a plasmid containing EcoRV restriction-modification system genes (lane 2), pBREXAL plasmid (lane 3), or pBTB2 control vector (lane 4) were combined with unmodified  $\lambda$  phage DNA, incubated and reaction products were resolved by agarose gel electrophoresis. Lane 1 is a control lane, phage genome incubated with buffer. M is a molecular weight marker. The gel is provided by Artem Isaev
- b.** Dynamics of phage DNA change in infected BREX<sup>+</sup> cells. Phage DNA coverage normalized to the coverage of bacterial DNA.
- c.** Schematics of phage  $\lambda$  DNA transformation after injection.
- d.** Fractions of circular and linear forms of phage  $\lambda$  DNA in BREX<sup>+</sup> and BREX<sup>-</sup> cultures during infection.

To track phage DNA dynamics in infected cells, total DNA was extracted from BREX+ and BREX- cultures at different time points post infection and sequenced on the Illumina platform (Fig. 20b). The experiment started 2 minutes after infection as complete injection time of phage  $\lambda$  DNA is about 2 minutes (101). The amount of phage DNA in BREX+ culture decreases and almost disappears in contrast to clear phage DNA replication in BREX- cells. Moreover, we calculated the fractions of linear and circular forms to track a translocation process because phage lambda DNA is circularized upon the injection (Fig. 20c) (102). We calculated the number of reads covering *cos* sites and 150 bp from the left and the right from *cos* sites to the number of reads ending on *cos* sites. A proportion of the circular form increases in BREX+ cells proving that the injection stage proceeded successfully (Fig. 20d).

### 3.6 BREX<sup>Ec</sup> protection is circumvented by glycosylation of phage DNA

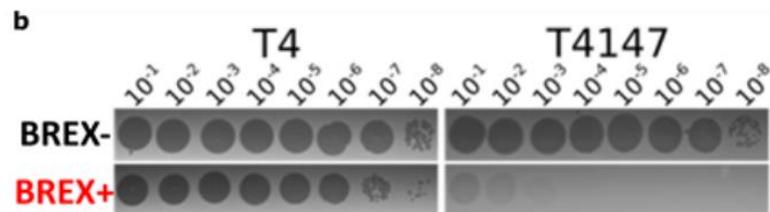
To determine the generality of the protective effect of BREX<sup>Ec</sup>, BREX+ cells were infected with a set of different phages (Fig. 21a). No protection from infection with Q $\beta$ , an RNA phage, was observed. Likewise, there was no protection against infection with M13, a single-stranded DNA phage with a double-stranded replicative intracellular form. In PhD thesis of Dr. Artem Isaev it was shown that accumulation of dsDNA forms of M13 phage is inhibited in BREX+ cells. The rest of the phages tested have double-stranded DNA genomes. The extent of BREX protection from these phages varied significantly without apparent dependence of the number of GGTAAG BREX sites in their genomes. The apparent lack of dependence of restriction on the number of GGTAAG methylation sites is striking, considering the site number dependence in tested restriction-modification systems (103). The results can be explained by anti-BREX strategies used by phages.

Similar to phage  $\lambda$ , ~100-fold protection from T5 and T7 infection in plaque-forming assay was observed. In stark contrast, no protection was observed from T4 infection (Fig. 21a). The DNA of T4 is hydroxymethylated at cytosines and is

additionally glycosylated (104). To test if glycosylation protects from BREX system action, we assayed T4<sub>147</sub>, a mutant of T4 bearing unglycosylated hydroxymethylcytosines in its genome (23). The extent of BREX protection from T4<sub>147</sub> phage was high (the experiment was done by colleagues in the laboratory of Prof. Siksnyš, Fig. 21b). We conclude that phages can overcome BREX system in at least two ways: (i) GGTAAG sites methylation, which requires *brxX* and *brxBCZ*, or (ii) glycosylation of their DNA, most probably at cytosines, in the opposite strand of the GGTAAG sites.

**a**

Phage	Genome	Number of GGTAAG sites	BREX <sup>Ec</sup> protection*
Q $\beta$	ssRNA	3**	10 <sup>0</sup>
M13	ssDNA	3***	10 <sup>0</sup>
T4	dsDNA	40	10 <sup>0</sup>
T4147	dsDNA	39	>10 <sup>5</sup>
T5	dsDNA	65	10 <sup>2</sup>
T7	dsDNA	44	10 <sup>2</sup>
$\lambda$	dsDNA	18	10 <sup>2</sup>



**Figure 21. Ability of BREX<sup>Ec</sup> to protect cells from infections other than  $\lambda$ .**

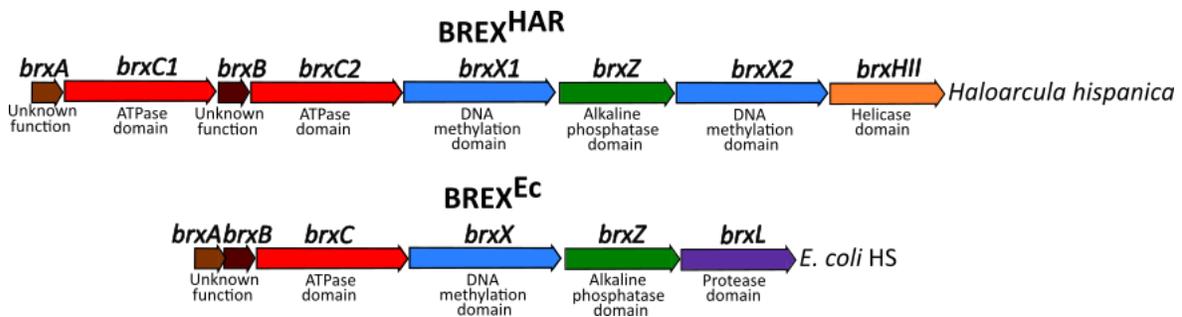
**a.** EOP of different phages for BREX+ cells. \* The ratio of the number of phage plaques on the BREX- lawn to the number of plaques on the BREX+ lawn. \*\* GGTAAG sites. \*\*\* Positive strand

**b.** An effect of cytosines glycosylation on BREX protection (provided by V. Siksnyš)

The figure is taken from (1).

### 3.7 BREX<sup>HAR</sup> system Type V from *Haloarcula hispanica*

A natural isolate of *Haloarcula hispanica* contains an eight-gene cluster related to BREX systems Type V (BREX<sup>HAR</sup>). The distinctive feature of the cluster is a set of genes coding for two ATPases BrxC1 and BrxC2, two methyltransferases BrxX1 and BrxX2, and a putative helicase BrxHII instead of a Lon-like protease BrxL (Fig. 22). Type V BREX systems were found only in archaea (15).



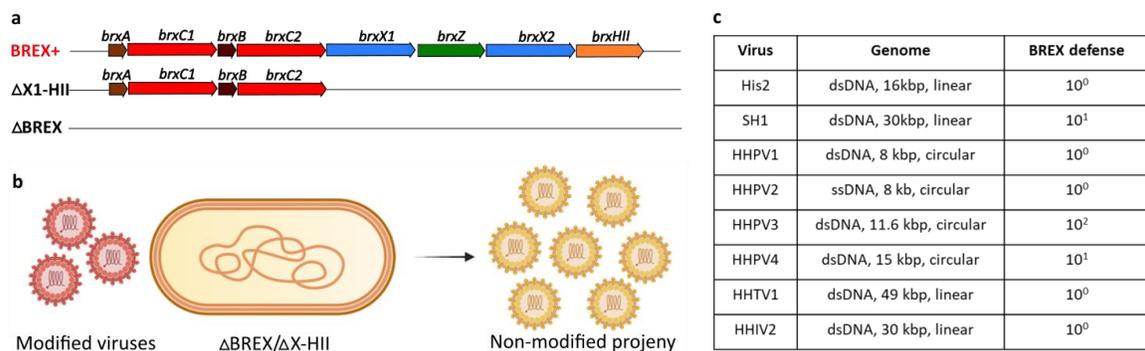
**Figure 22. Comparison of gene compositions for Type V BREX<sup>HAR</sup> system from *Haloarcula hispanica* and Type I BREX<sup>Ec</sup> system from *Escherichia coli* HS.**

The defense activity of CRISPR-Cas Type I-B system from *Haloarcula hispanica* was studied using an uracil auxotrophic (*pyrF*-deleted) strain DF60 (94, 105). The strain can be easily genetically modified by a gene knockout system named ‘pop-in-pop-out’ (94). To disrupt an effect of a CRISPR-Cas system encoded in the *H. hispanica* genome, a DF60  $\Delta casI-8$  strain (85) was used for further experiments.

We hypothesized that Type V BREX systems work by modification of host DNA and elimination of non-modified invader’s DNA like it was shown for *E. coli* Type I BREX system. All *H. hispanica* viruses were isolated and produced on a wild type ATCC 33960 strain carrying the BREX<sup>HAR</sup> system. Thus, all known *H. hispanica* viruses should

contain the proposed modification in their genomes and should not be subject to BREX<sup>HAR</sup> exclusion.

Strains lacking the entire 25-kbp BREX cassette ( $\Delta$ BREX) or with a partial deletion of *brx* genes ( $\Delta$ X1-  $\Delta$ HII) were created (Fig. 23a). A collection of *H. hispanica* viruses was propagated on  $\Delta$ BREX or  $\Delta$ X1- $\Delta$ HII cultures to get rid of possible *brx* modifications (Fig. 23b). Such presumably ‘unmodified’ viruses were used for the efficiency of plaque formation (EOP) assay on BREX+,  $\Delta$ BREX, and  $\Delta$ X1-  $\Delta$ HII lawns to evaluate BREX defense activity. As can be seen from Fig. 23c, upon the infection, *H. hispanica* strain containing the BREX system showed resistance to HHPV3 virus that we chose as a model for further experiments.



**Figure 23. BREX<sup>HAR</sup> defense activity.**

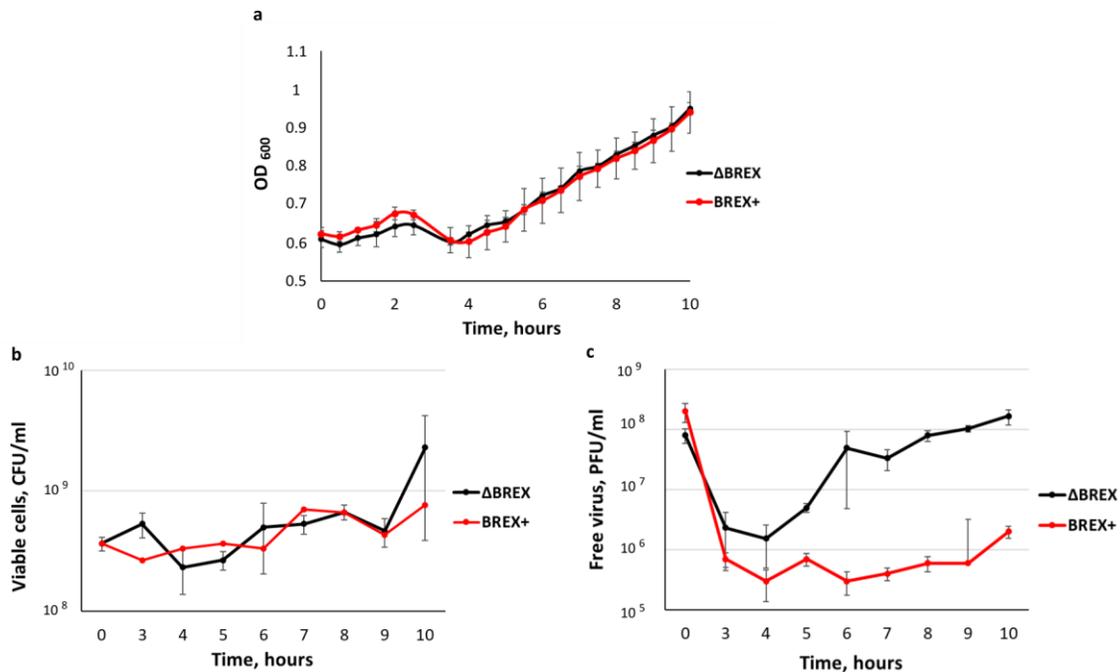
**a.** Gene composition of  $\Delta$ BREX and  $\Delta$ X1-  $\Delta$ HII strains.

**b.** Schematics of procedure used for non-modified viruses production.

**c.** Ability of BREX<sup>HAR</sup> protects cells from viruses. BREX defense is calculated as a ratio of the number of plaques on the  $\Delta$ BREX lawn to the number of plaques on the BREX+ lawn.

The growth rates of infected  $\Delta$ BREX and BREX+ cells were monitored in liquid culture. At the  $t = 0$  cultures were infected with HHPV3 virus, at  $t = 3$  hours post-infection unabsorbed viruses were washed away. As HHPV3 virus undergoes non-lytic lifestyle without cell lysis (106), no decline in cell forming units (CFU) or in optical density (OD) was observed (Fig. 24ab). The HHPV3 virus release was detected starting from  $\sim 6$ -hours post-infection in cells lacking the BREX<sup>HAR</sup> system (Fig. 24c). Virus titer in BREX+ culture was around 10<sup>6</sup> PFU/ml after wash and did not change until the end of

the experiment. Therefore, similar to BREX<sup>Ec</sup>, cells with BREX<sup>HAR</sup> system do not produce progeny of sensitive viruses.



**Figure 24. Dynamics of infection of *H. hispanica* carrying BREX<sup>HAR</sup> and infected with HHPV3.**

**a.** Culture dynamics of virus-infected ΔBREX and BREX+ strains of *H. hispanica*. Cultures were infected at the MOI of 0.2 and time = 0 h, at 3h cultures were washed from unabsorbed viruses.  
**b.** The amount of colony forming units (CFU) in BREX+ and ΔBREX cultures from panel **a**.  
**c.** The amount of plaque forming units (PFU) in BREX+ and ΔBREX cultures from panel **a**.  
 Each experiment was performed three times with three technical triplicates for each biological replicate. Error bars represent SEM

We proved our hypothesis that Type V BREX systems act as Type I systems by modification of self DNA and elimination of non-methylated DNA. To find *brx*-mediated modifications, genomic DNA extracted from *H. hispanica* of ΔBREX, ΔX1-HII and BREX+ strains were sequenced by Single Molecule Real-Time (SMRT) sequencing from Pacific Biosciences. All genomes contain methylated cytosines (m4C) at CTAG sites. In BREX+ cells two more modifications were found (Table 3). One *brx*-mediated modification is a methylation of cytosine (m4C) at the fifth position of GACCCC sites.

The other modification is a methylation of cytosine (m4C) at the sixth position of GTAYCCG.

**Table 3. Results of PacBio sequencing of different *Haloarcula hispanica* strains.**

<i>H. hispanica</i> genome	Percentage of methylated <u>CTAG</u>	Percentage of methylated <u>GACCCC</u>	Percentage of methylated <u>GTAYCCG</u>
<b>ΔBREX</b>	94.9%	0%	0%
<b>ΔX1-HI</b>	95.5%	0%	0%
<b>BREX+</b>	95.2%	95.4%	83%

### 3.8 The role of individual *brx* genes on modification and recognition module

Given the presence of two methyltransferases BrxX1 and BrxX2 in BREX<sup>HAR</sup>, it is likely that each methyltransferase installs a methyl group into its own *brx* site. To check this idea, we created strains with deletions of *brxX1* and *brxX2* genes separately (Table 4). The HHPV3<sub>ΔBREX</sub> virus was titrated on ΔX1 and ΔX2 lawns to test defense. As can be seen from Table 4 deletion of *brxX2* abolished protection. The titer of the virus on lawns with *brxX1* deletion was between titers observed on ΔBREX and BREX+ lawns, indicating partial defense. The methylation status was checked by production of HHPV3 on ΔX1 or ΔX2 cultures and further determination of EOP on ΔBREX and BREX+ lawns. The assay showed that HHPV3<sub>ΔX1</sub> virus had an equal titer on both lawns, indicating complete modification of viral DNA. EOP for HHPV3<sub>ΔX2</sub> virus was 10<sup>1</sup>, indicating partial modification of the viral genome.

These data are consistent with a view that each Brx methyltransferase modifies its own *brx* site, where BrxX1 methylates GTAYCCG and BrxX2 methylates GACCCC. HHPV3 virus contains five GACCCC motifs and two GTAYCCG motifs. Deletion of *brxX1* leads to recognition of GACCCC sites by a complex with BrxX2

methyltransferase, deletion of *brxX2* – GTAYCCG sites by BrxX1. As the number of BrxX2 motifs in HHPV3 is larger than the number of BrxX1 motifs, the defense activity of the system with BrxX2 methyltransferase only is higher than that with BrxX1 only ( $10^1$  for  $\Delta X1$  and  $10^0$  for  $\Delta X2$ ). In the same manner, HHPV3 $_{\Delta X1}$  produced on the  $\Delta X1$  lawn has five non-methylated GACCCC and therefore EOP for HHPV3 $_{\Delta X1}$  is higher than for HHPV3 $_{\Delta X2}$  with only two non-methylated GTAYCCG sites. This explanation will be checked by PacBio sequencing.

To evaluate the role of individual *brx* genes in modification and defense, strains with deletions of individual *brx* genes were created (Table 4). Similar to BREX<sup>Ec</sup> system, *brxA* and *brxHII* (*brxL* for BREX<sup>Ec</sup>) are not essential for DNA methylation. Moreover, deletion of *H. hispanica brxB* also did not affect the methylation status. Proteins BrxB, BrxC1, BrxC2, BrxZ, BrxHII together with BrxX1 and Brx2 participate in defense.

**Table 4. Defense and methylation activities of *H. hispanica* stains with individual deletions of *brx* genes.**

<i>Strain</i>	<i>Defense</i>	<i>Methylation</i>
$\Delta A$	+	+
$\Delta B$	-	+
$\Delta C1$	-	-
$\Delta C2$	-	-
$\Delta X1$	partial	+
$\Delta X2$	-	Partial
$\Delta Z$	-	-
$\Delta HII$	-	+
$\Delta NucS$	+	+

Inside the *brx* cassette a gene encoding a predicted endonuclease NucS is found (between the *brxC2* and *brxX1* genes). We deleted this gene from *H. hispanica* to evaluate its role in BREX defense against HHPV3 and in DNA methylation. Table 4 shows that deletion of *nucS* has no effect on either BREX-mediated defense or on DNA modification. A similar insertion of an endonuclease BrxU was discovered in Type I BREX system from *Escherichia fergusonii* (51). NucS and BrxU do not share sequence homology.

All *H. hispanica* viruses were analyzed for the presence of GACCCC and GTAYCCG motifs. As can be seen from Table 5, the number of *brx*-recognized sites in closely related viruses SH1 and HHIV2 is higher than in HHPV3 even if the size of viral DNA is considered, but the level of BREX protection against these viruses is lower. It can be explained by additional modification of viral genome like hydroxymethylation and glycosylation of cytosines in phage T4 or by anti-BREX proteins as was observed in the case of T7 phage (107). Bioinformatics analysis showed that none of these viruses encode any modification enzymes. The search of anti-BREX proteins in SH1 and HHIV-2 genomes is our next task to decipher the mechanism of BREX Type V systems.

**Table 5. The number of *brx* sites in *H. hispanica* viruses.**

<i>Virus</i>	<i>DNA</i>	<i>GACCCC sites</i>	<i>GTAYCCG sites</i>	<i>Defense</i>
<b>His2</b>	ds, 16 kbp, linear	0	2	10 <sup>0</sup>
<b>SH1</b>	ds, 30 kbp, linear	42	4	10 <sup>1</sup>
<b>HHPV1</b>	ds, 8 kbp, circular	2	2	10 <sup>0</sup>
<b>HHPV2</b>	ss, 8 kb, circular	0	1	10 <sup>0</sup>
<b>HHPV3</b>	ds, 11.6 kbp, circular	5	2	10 <sup>2</sup>
<b>HHPV4</b>	ds, 15 kbp, circular	7	3	10 <sup>1</sup>
<b>HHTV1</b>	ds, 49 kbp, linear	0	0	10 <sup>0</sup>
<b>HHIV2</b>	ds, 30 kbp, linear	34	6	10 <sup>0</sup>

## Conclusions

In this study we

- checked the BREX<sup>Ec</sup>/BREX<sup>HAR</sup> defense activity;
- established BREX<sup>Ec</sup>/BREX<sup>HAR</sup> gene composition in defensive and methylation complexes (Table 6);
- identified *brx*-mediated modifications;
- found BREX<sup>Ec</sup> degradation activity evidence;
- explained a decline in OD for BREX<sup>Ec</sup>-carrying cells at high MOI infection;
- proposed recognition strategies for BREX<sup>Ec</sup> and BREX<sup>HAR</sup>;
- explored toxicity of BREX<sup>Ec</sup> gene deletions and effects of individual *brx* genes on cell growth.

**Table 6. BREX<sup>Ec</sup> and BREX<sup>HAR</sup> comparison**

	BREX Type I from <i>E. coli</i> (BREX <sup>Ec</sup> )	BREX Type V from <i>H. hispanica</i> (BREX <sup>HAR</sup> )
BREX localization	plasmid	genome
Gene composition	A, B, C, X, Z, L	A, B, C1, C2, X1, X2, Z, HII
Methylation complex	B, C, X, Z	C1, C2, X1, X2, Z
Defense complex	B, C, X, Z, L	B, C1, C2, X1, X2, Z, HII
Brx-mediated modification	GGTA <u>A</u> G	GAC <u>CC</u> C, GTAY <u>CC</u> G
Recognition strategy	Methylation of self DNA and elimination of invader's DNA	
Effect on virus production	No virus production	
Defense strategy	Phage DNA degradation	Unknown
Anti-BREX proteins	Ocr (Isaev <i>et al.</i> , 2020)	Unknown

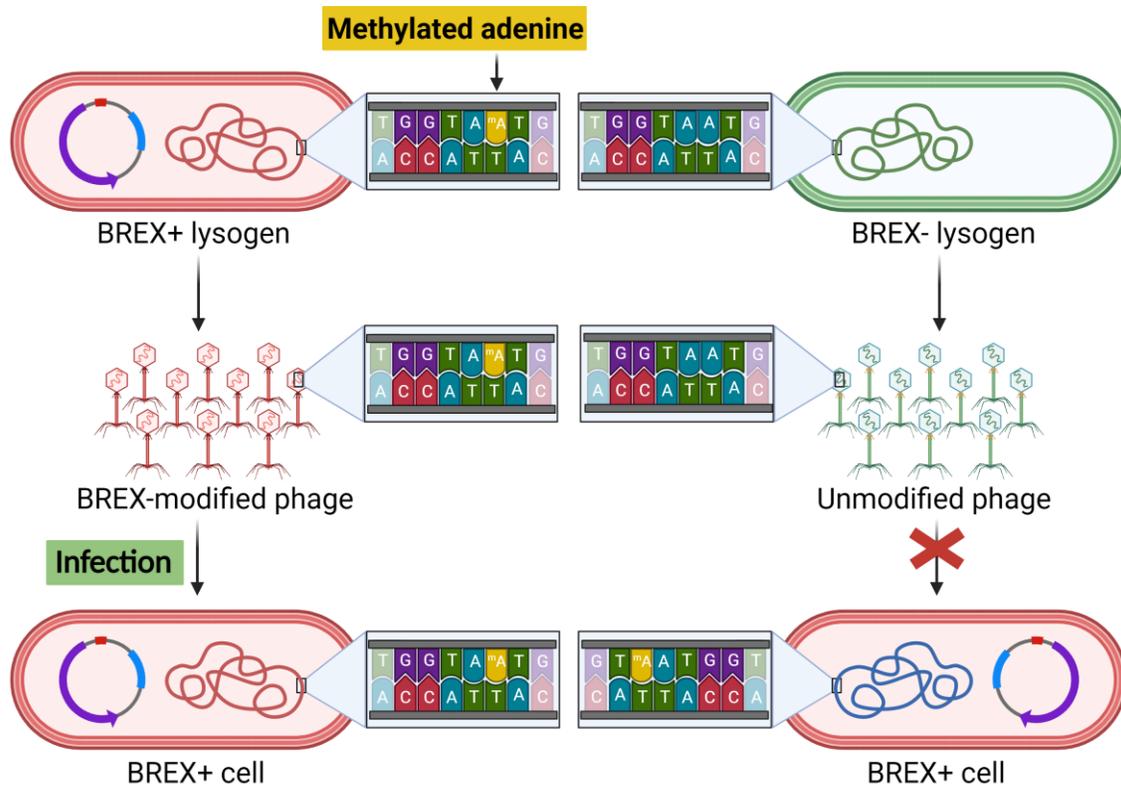
## Discussion

In this work, we demonstrate that the *E. coli* BREX cluster protects cells from infection by diverse dsDNA phages. At least one of BREX genes, *brxA*, which codes for a protein of unknown function, is not essential for phage defense at our conditions. However, we could not delete this gene from pBREXAL plasmid with BREX<sup>Ec</sup> system under natural promoters without additional mutations. We suggest that *brxA* can function as a transcription regulator of *brx* genes and its deletion affects BREX function and leads to toxicity. The use of inducible arabinose promoter allows to circumvent the putative regulatory role of *brxA*.

We showed that the modification module of *E. coli* BREX system functions by methylating a specific asymmetric site in phage DNA (Fig. 25). The enzymatic activity of BrxX in the presence of BrxB, BrxC, and BrxZ is likely responsible for this modification. Site-specific methylation of phage DNA or global glycosylation of cytosine residues allows a phage to bypass the BREX defense. Thus, in this regard the *E. coli* BREX system functions similarly to R-M systems and is distinct from the Pgl system. However, unlike R-M systems, pBREXAL plasmid expressing the BREX system and an empty pBTB-2 vector is transformed with equal efficiency. The observation may be explained by more sophisticated mechanism of gene regulation. The absence of the restriction activity while DNA-methyltransferase is deleted supports this view.

Earlier work with BREX from *B. cereus* failed to observe modification of DNA of phage progeny collected after BREX<sup>+</sup> infection (15). The result is probably due to a very low level of phage progeny production in these experiments, which caused the authors to analyze unabsorbed/non-infecting phage particles, whose genomes naturally remained unmodified. While the sites of methylation in *E. coli*, *E. fergusonii*, and *B. cereus* BREX systems have unrelated sequences (GGTAAG, GCTAAT, and TAGGAG, respectively), there are also important commonalities – they are asymmetric, and it is an adenine in the fifth position of the recognition site that is being methylated. Phylogenetic analysis indicates that BREX systems can be divided into six types based on the presence

of characteristic genes. The *E. coli* and *B. cereus* BREX systems belong to the same Type I (15) and may share a common mechanism of self-versus-non-self differentiation. The Pgl system belongs to a different type, Type II, which may help explain the apparent differences in reported behavior of Pgl and the studied Type I systems.



**Figure 25. Schematics of the BREX<sup>Ec</sup> function.**

Our results establish the mechanism responsible for self-protection of BREX carrying cells and means by which phages can overcome the BREX defense. The logic appears to be similar to that typical for R-M systems, where rare modified genetic invaders eventually multiply and take over the population of initially protected cells. It is interesting that BREX systems may undergo phase variation due to a homopolymeric tract in the *brxX* coding sequence (15) and this strategy may allow the cells to alternate the proportion of BREX+ and BREX- cells in the population which may help withstand phage predation with BREX modified DNA.

For most-studied Type II R-M systems, the palindromic nature of the recognition site ensures that epigenetic protective modification is heritable. The BREX site is asymmetric and the maintenance of epigenetic marks likely requires interactions between different sites, as observed for R-M Type I and Type III systems (35, 108, 109).

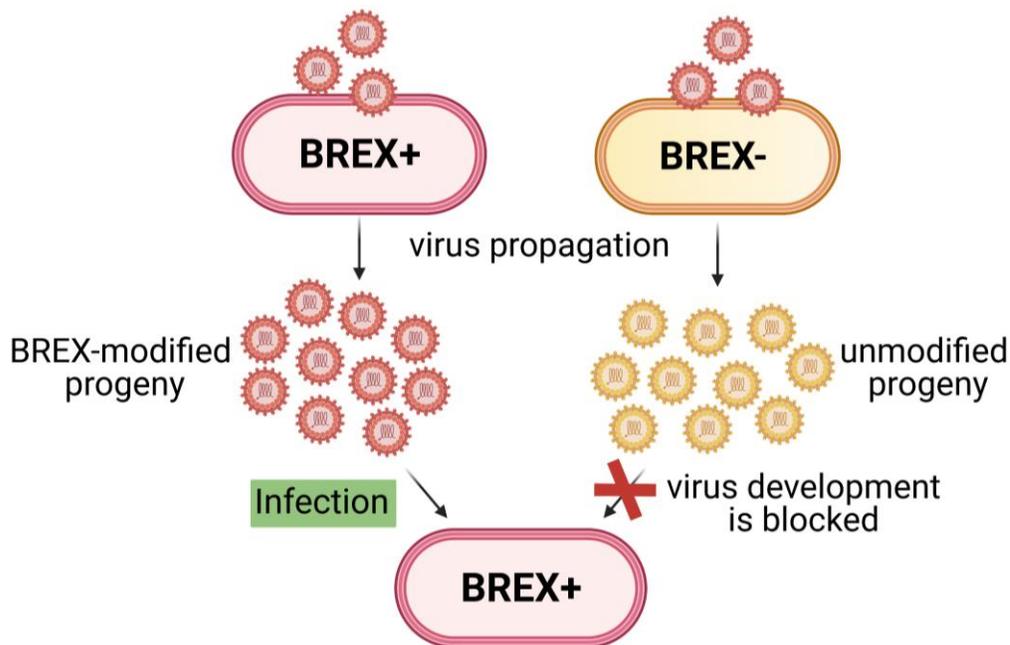
The nature of defensive action of BREX remains elusive. The  $\lambda$  infection must be blocked at a very early stage after the absorption of the phage. Moreover, we observed an accumulation of the circular DNA forms of phage  $\lambda$  that indicates successful DNA injection. In principle, rapid degradation of injected DNA prior to its replication is consistent with our lysogenization assay, prophage induction experiment, live microscopy data and sequencing of total DNA extracted from infected BREX+ cells. Identification of the defense mechanism by BREX systems will require establishment of functional interactions between essential system components, since our mutagenesis results as well as earlier data collected with the Pgl system indicate that there must be several toxin-antitoxin type interactions that allow stable maintenance of the system. Interestingly, BrxL, a putative protease and a likely toxin is absent in Pgl, suggesting that BREX systems may employ multiple strategies to limit viral infections. We speculate that BrxL, which is necessary for the defensive function but not for methylation, may be activated by phage infection and its toxicity is responsible for invader's elimination. The protein could inhibit host mechanisms while the system eliminates a phage. However, we postulate that BREX system does not work as Abi systems.

A recent study from our laboratory showed that BREX defensive action can be eliminated by phage T7 protein Ocr, a DNA mimic protein that inhibits Type I R-M systems (107). Although Ocr forms a complex with methyltransferase BrxX, it does not abolish BREX methylation completely. The presence of large numbers of non-methylated *brx* sites in bacterial genomes raises a question of how self-destruction of BREX+ cells is avoided.

Experiments aimed at characterizing partial biochemical activities of *E. coli* Brx proteins failed except for the observed ATPase activity of BrxL and BrxC. The crystal structure of Brx protein complexes or complexes of anti-BREX protein with Brx

components will shed light on the interactions between Brx proteins and their functions. These experiments are currently ongoing.

The Type V BREX system found only in archaea was studied here and shown to work akin to Type I BREX systems (Fig. 26). BREX<sup>HAR</sup> encodes two active methyltransferases that modify DNA independently and, thus, in genomes of cells with deletions of individual methyltransferases only one *brx* motifs is methylated. In the absence of *brxXI* that incorporates methyl groups in GACCCC sites, BREX system targets only the GTAYCCG motifs modified by BrxX2. Possibility to create such deletions can be explained by participation of methyltransferases in the restriction process: deletion of a gene eliminates both the respective modification activity but also the restriction activity.



**Figure 26. Schematics of the BREX<sup>HAR</sup> function.**

Further study of BREX<sup>HAR</sup> includes identification evidence for the possible presence of anti-BREX proteins encoded by the SH1 and HHIV-2 viruses.

## Bibliography

1. Gordeeva,J., Morozova,N., Sierro,N., Isaev,A., Sinkunas,T., Tsvetkova,K., Matlashov,M., Truncaite,L., Morgan,R.D., Ivanov,N. V., et al. (2019) BREX system of *Escherichia coli* distinguishes self from non-self by methylation of a specific DNA site. *Nucleic Acids Research*, 10.1093/nar/gky1125.
2. Suttle,C. (2007) Marine viruses--major players in the global ecosystem. *Nature Reviews Microbiology*, 5, 801–812.
3. Dammeyer,T., Bagby,S.C., Sullivan,M.B., Chisholm,S.W. and Frankenberg-Dinkel,N. (2008) Efficient phage-mediated pigment biosynthesis in oceanic cyanobacteria. *Current Biology*, 18, 442–448.
4. Williams,R.J. (2003) Restriction endonucleases: classification, properties, and applications. *Molecular. Biotechnology*, 23, 225–243.
5. Loenen,W.A.M., Dryden,D.T.F., Raleigh,E.A., Wilson,G.G. and Murray,N.E. (2014) Highlights of the DNA cutters: a short history of the restriction enzymes. *Nucleic Acids Research*, 42, 3–19.
6. Koonin,E. V and Makarova,K.S. (2009) CRISPR-Cas: an adaptive immunity system in prokaryotes. *RNA Biology*, 10, 679–686.
7. Marraffini,L.A. (2015) CRISPR-Cas immunity in prokaryotes. *Nature*, 526, 55–61.
8. Shmakov,S., Smargon,A., Scott,D., Cox,D., Pyzocha,N., Yan,W., Abudayyeh,O.O., Gootenberg,J.S., Makarova,K.S., Wolf,Y.I., et al. (2017) Diversity and evolution of class 2 CRISPR–Cas systems. *Nature Reviews Microbiology*, 15, 169–182.
9. Chopin,M.C., Chopin,A. and Bidnenko,E. (2005) Phage abortive infection in lactococci: variations on a theme. *Current Opinion Microbiology*, 8, 473–479.
10. Blower,T.R., Pei,X.Y., Short,F.L., Fineran,P.C., Humphreys,D.P., Luisi,B.F. and Salmond,G.P.C. (2011) A processed noncoding RNA regulates an altruistic bacterial antiviral system. *Nature Structural&Molecular Biology*, 18, 185–190.
11. Masuda,H. and Inouye,M. (2017) Toxins of prokaryotic toxin-antitoxin systems with sequence-specific endoribonuclease activity. *Toxins (Basel)*, 9.
12. Koonin,E. V., Makarova,K.S. and Wolf,Y.I. (2017) Evolutionary genomics of defense systems in archaea and bacteria. *Annual Review Microbiology*, 71, 233–261.
13. Doron,S., Melamed,S., Ofir,G., Leavitt,A., Lopatina,A., Keren,M., Amitai,G. and Sorek,R. (2018) Systematic discovery of antiphage defense systems in the microbial pangenome. *Science* 359(6379).
14. Makarova,K.S., Wolf,Y.I., Snir,S. and Koonin,E. V. (2011) Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of Bacteriology*, 193, 6039–6056.
15. Goldfarb,T., Sberro,H., Weinstock,E., Cohen,O., Doron,S., Charpak-Amikam,Y., Afik,S., Ofir,G., Sorek,R., Bikard,D., et al. (2015) BREX is a novel phage resistance system widespread in microbial genomes. *EMBO Journal*, 34, 169–183.
16. Stern,A. and Sorek,R. (2012) The phage-host arms-race: Shaping the evolution of microbes. *Bioessays*, 33, 43–51.
17. Labrie,S.J., Samson,J.E. and Moineau,S. (2010) Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*, 8, 317–327.

18. Koebnik,R. (1999) Structural and functional roles of the surface-exposed loops of the beta-barrel membrane protein OmpA from *Escherichia coli*. *Journal of Bacteriology*, 181, 3688–3694.
19. Meyer,J.R., Dobias,D.T., Weitz,J.S., Barrick,J.E., Quick,R.T. and Lenski,R.E. (2012) Repeatability and contingency in the evolution of a key innovation in phage lambda. *Science* 335(6067):428-32.
20. Achtman,M., Kennedy,N. and Skurray,R. (1977) Cell-cell interactions in conjugating *Escherichia coli*: role of traT protein in surface exclusion. *PNAS*, 74, 5104–5108.
21. Moak,M. and Molineux,I.J. (2004) Peptidoglycan hydrolytic activities associated with bacteriophage virions. *Molecular Microbiology*, 51, 1169–1183.
22. Mosig,G., Lin,G.W., Franklin,J. and Fan,W.H. (1989) Functional relationships and structural determinants of two bacteriophage T4 lysozymes: a soluble (gene e) and a baseplate-associated (gene 5) protein. *New Biology*, 1, 171–180.
23. Destoumieux-Garzón,D., Duquesne,S., Peduzzi,J., Goulard,C., Desmadril,M., Letellier,L., Rebuffat,S. and Boulanger,P. (2005) The iron–siderophore transporter FhuA is the receptor for the antimicrobial peptide microcin J25: role of the microcin Val 11 Pro 16  $\beta$ -hairpin region in the recognition mechanism. *Biochemistry Journal*, 389, 869–876.
24. Quemain,E.R.J. and Quax,T.E.F. (2015) Archaeal viruses at the cell envelope: entry and egress. *Frontiers in Microbiology*, 6, 552.
25. Deng,L., He,F., Bhoobalan-Chitty,Y., Martinez-Alvarez,L., Guo,Y. and Peng,X. (2014) Unveiling Cell Surface and Type IV Secretion Proteins Responsible for Archaeal Rudivirus Entry. *Journal of Virology*, 88, 10264.
26. Lu,M.J. and Henning,U. (1994) Superinfection exclusion by T-even-type coliphages. *Trends in Microbiology*, 2, 137–139.
27. Tock,M.R. and Dryden,D.T.F. (2005) The biology of restriction and anti-restriction. *Current Opinion in Microbiology*, 8, 466–472.
28. Roberts,R.J., Vincze,T., Posfai,J. and Macelis,D. (2015) REBASE-a database for DNA restriction and modification: Enzymes, genes and genomes. *Nucleic Acids Research*, 43, D298–D299.
29. Dy,R.L., Richter,C., Salmond,G.P.C. and Fineran,P.C. (2014) Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annual Review of Virology*, 1, 307–331.
30. Pingoud,A., Wilson,G.G. and Wende,W. (2014) Type II restriction endonucleases - a historical perspective and more. *Nucleic Acids Research*, 42, 7489.
31. Roberts,R.J., Belfort,M., Bestor,T., Bhagwat,A.S., Bickle,T.A., Bitinaite,J., Blumenthal,R.M., Degtyarev,S.K., Dryden,D.T.F., Dybvig,K., et al. (2003) SURVEY AND SUMMARY: A nomenclature for restriction enzymes, DNA methyltransferases, homing endonucleases and their genes. *Nucleic Acids Research*, 31, 1805.
32. Loenen,W.A.M., Dryden,D.T.F., Raleigh,E.A. and Wilson,G.G. (2014) Type I restriction enzymes and their relatives. *Nucleic Acids Research*, 10.1093/nar/gkt847.
33. Neaves,K.J., Cooper,L.P., White,J.H., Carnally,S.M., Dryden,D.T.F., Edwardson,M.J. and Henderson,R.M. (2009) Atomic force microscopy of the EcoKI Type I DNA restriction enzyme bound to DNA shows enzyme dimerization and DNA looping. *Nucleic Acids Research*, 37, 2053.

34. Butterer,A., Pernstich,C., Smith,R.M., Sobott,F., Szczelkun,M.D. and Tóth,J. (2014) Type III restriction endonucleases are heterotrimeric: comprising one helicase–nuclease subunit and a dimeric methyltransferase that binds only one specific DNA. *Nucleic Acids Research*, 42, 5139.
35. Van Aelst,K., Tóth,J., Ramanathan,S.P., Schwarz,F.W., Seidel,R. and Szczelkun,M.D. (2010) Type III restriction enzymes cleave DNA by long-range interaction between sites in both head-to-head and tail-to-tail inverted repeat. *PNAS*, 107, 9123–9128.
36. Wilson,G. (1991) Restriction And Modification Systems. *Annual Reviews in Genetics*, 25, 585–627.
37. Zabeau,M., Friedman,S., Van Montagu,M. and Schell,J. (1980) The *ral* gene of phage lambda. Identification of a non-essential gene that modulates restriction and modification in *E. coli*. *Molecular Genetics and Genomics*, 179, 63–73.
38. Walkinshaw,M.D., Taylor,P., Sturrock,S.S., Atanasiu,C., Berge,T., Henderson,R.M., Edwardson,J.M. and Dryden,D.T.F. (2002) Structure of Ocr from bacteriophage T7, a protein that mimics b-form DNA. *Molecular Cell*, 9, 187–194.
39. Ofir,G., Melamed,S., Sberro,H., Mukamel,Z., Silverman,S., Yaakov,G., Doron,S. and Sorek,R. (2018) DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nature Microbiology*, 3, 90–98.
40. El Yacoubi,B., Bailly,M. and De Crécy-Lagard,V. (2012) Biosynthesis and function of posttranscriptional modifications of transfer RNAs. *Annual Reviews in Genetics*, 46, 69–95.
41. Thiaville,J.J., Kellner,S.M., Yuan,Y., Hutinet,G., Thiaville,P.C., Jumpathong,W., Mohapatra,S., Brochier-Armanet,C., Letarov,A. V., Hillebrand,R., et al. (2016) Novel genomic island modifies DNA with 7-deazaguanine derivatives. *PNAS*, 113, E1452–E1459.
42. Wang,L., Jiang,S., Deng,Z., Dedon,P.C. and Chen,S. (2019) DNA phosphorothioate modification - a new multi-functional epigenetic system in bacteria. *FEMS Microbiology Reviews*, 43, 109.
43. Xu,T., Yao,F., Zhou,X., Deng,Z. and You,D. (2010) A novel host-specific restriction system associated with DNA backbone S-modification in *Salmonella*. *Nucleic Acids Research*, 38, 7133.
44. Wu,X., Wu,X., Wu,X., Cao,B., Cao,B., Cao,B., Aquino,P., Chiu,T.P., Chiu,T.P., Chiu,T.P., et al. (2020) Epigenetic competition reveals density-dependent regulation and target site plasticity of phosphorothioate epigenetics in bacteria. *PNAS*, 117, 14322–14330.
45. Xiong,L., Liu,S., Chen,S., Xiao,Y., Zhu,B., Gao,Y., Zhang,Y., Chen,B., Luo,J., Deng,Z., et al. (2019) A new type of DNA phosphorothioation-based antiviral system in archaea. *Nature Communications*, 10.
46. Xiong,X., Wu,G., Wei,Y., Liu,L., Zhang,Y., Su,R., Jiang,X., Li,M., Gao,H., Tian,X., et al. (2020) SspABCD–SspE is a phosphorothioation-sensing bacterial defence system with broad anti-phage activities. *Nature Microbiology*. 57, 5, 917–928.

47. Chinenova, T.A., Mkrtumian, N.M. and Lomovskaia, N.D. (1982) Genetic characteristics of a new phage resistance trait in *Streptomyces coelicolor* A3(2). *Genetika*, 18, 1945–52.
48. Hoskisson, P.A., Sumbly, P. and Smith, M.C.M. (2015) The phage growth limitation system in *Streptomyces coelicolor* A(3)2 is a toxin/antitoxin system, comprising enzymes with DNA methyltransferase, protein kinase and ATPase activity. *Virology*, 477, 100–109.
49. Sumbly, P. and Smith, M.C.M. (2002) Genetics of the phage growth limitation (Pgl) system of *Streptomyces coelicolor* A3(2). *Molecular Microbiology*, 44, 489–500.
50. Hui, W., Zhang, W., Kwok, L.Y., Zhang, H., Kong, J. and Suna, T. (2019) A Novel Bacteriophage Exclusion (BREX) System Encoded by the pglX Gene in *Lactobacillus casei* Zhang. *Applied Environment Microbiology*, 85.
51. Picton, D.M., Luyten, Y.A., Morgan, R.D., Nelson, A., Smith, D.L., Dryden, D.T.F., Hinton, J.C.D. and Blower, T.R. (2021) The phage defence island of a multidrug resistant plasmid uses both BREX and type IV restriction for complementary protection from viruses. *Nucleic Acids Research*, 49, 11257.
52. Picton, D.M., Harling-Lee, J.D., Duffner, S.J., Went, S.C., Morgan, R.D., Hinton, J.C.D. and Blower, T.R. (2022) A widespread family of WYL-domain transcriptional regulators co-localizes with diverse phage defence systems and islands. *Nucleic Acids Research*, 50.
53. Luyten, Y.A., Hausman, D.E., Young, J.C., Doyle, L.A., Higashi, K.M., Ubilla-Rodriguez, N.C., Lambert, A.R., Arroyo, C.S., Forsberg, K.J., Morgan, R.D., et al. (2022) Identification and characterization of the WYL BrxR protein and its gene as separable regulatory elements of a BREX phage restriction system. *Nucleic Acids Research*, 50.
54. Marraffini, L.A. (2015) CRISPR-Cas immunity in prokaryotes. *Nature*, 526, 55–61.
55. Sorek, R., Lawrence, C.M. and Wiedenheft, B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annual Reviews in Biochemistry*, 82, 237–266.
56. Bolotin, A., Quinquis, B., Sorokin, A. and Dusko Ehrlich, S. (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, 151, 2551–2561.
57. Samson, J.E., Magadán, A.H., Sabri, M. and Moineau, S. (2013) Revenge of the phages: defeating bacterial defences. *Nature Reviews Microbiology*, 11, 675–87.
58. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., et al. (2020) Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature Reviews Microbiology*, 18, 67–83.
59. Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A., Saunders, S.J., Barrangou, R., Brouns, S.J.J., Charpentier, E., Haft, D.H., et al. (2015) An updated evolutionary classification of CRISPR-Cas systems. *Nature Reviews Microbiology*, 13, 722–736.
60. Burstein, D., Harrington, L.B., Strutt, S.C., Probst, A.J., Anantharaman, K., Thomas, B.C., Doudna, J.A. and Banfield, J.F. (2017) New CRISPR-Cas systems from uncultivated microbes. *Nature*, 542, 237–241.
61. Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157, 1262–1278.

62. Datsenko, K. a., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K. and Semenova, E. (2012) Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nature Communications*, 3, 945.
63. Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M. and Terns, M.P. (2009) RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*, 139, 945–956.
64. Bondy-Denomy, J., Pawluk, A., Maxwell, K.L. and Davidson, A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, 493, 429–32.
65. Seed, K.D., Lazinski, D.W., Calderwood, S.B. and Camilli, A. (2013) A bacteriophage encodes its own CRISPR/Cas adaptive response to evade host innate immunity. *Nature*, 494, 489–91.
66. Parma, D.H., Snyder, M., Sobolevski, S., Nawroz, M., Brody, E. and Gold, L. (1992) The Rex System of Bacteriophage Lambda: Tolerance and Altruistic Cell Death. *Genes and Development*, 6, 497–510.
67. Hinton, D.M. (2010) Transcriptional control in the prereplicative phase of T4 development. *Virology Journal*, 7.
68. Schmitt, C.K. and Molineux, I.J. (1991) Expression of gene 1.2 and gene 10 of bacteriophage T7 is lethal to F plasmid-containing *Escherichia coli*. *Journal of Bacteriology*, 173, 1536–1543.
69. Cheng, X., Wang, W. and Molineux, I.J. (2004) F exclusion of bacteriophage T7 occurs at the cell membrane. *Virology*, 326, 340–352.
70. Depardieu, F., Didier, J.P., Bernheim, A., Sherlock, A., Molina, H., Duclos, B. and Bikard, D. (2016) A Eukaryotic-like Serine/Threonine Kinase Protects *Staphylococci* against Phages. *Cell Host Microbe*, 20, 471–481.
71. Georgiou, T., Yu, Y.N., Ekunwe, S., Buttner, M.J., Zuurmond, a, Kraal, B., Kleanthous, C. and Snyder, L. (1998) Specific peptide-activated proteolytic cleavage of *Escherichia coli* elongation factor Tu. *PNAS*, 95, 2891–5.
72. Kao, C. and Snyder, L. (1988) The lit gene product which blocks bacteriophage T4 late gene expression is a membrane protein encoded by a cryptic DNA element, e14. *Journal of Bacteriology*, 170, 2056–2062.
73. Champness, W.C. and Snyder, L. (1982) The gol site: A Cis-acting bacteriophage T4 regulatory region that can affect expression of all the T4 late genes. *Jornal of Molecular Biology*, 155, 395–407.
74. Levitz, R., Chapman, D., Amitsur, M., Green, R., Snyder, L. and Kaufmann, G. (1990) The optional *E. coli prr* locus encodes a latent form of phage T4-induced anticodon nuclease. *EMBO Journal*, 9, 1383–1389.
75. Kaufmann, G., David, M., Borasio, G.D., Teichmann, A., Paz, A., Amitsur, M., Green, R. and Snyder, L. (1986) Phage and host genetic determinants of the specific anticodon loop cleavages in bacteriophage T4-infected *Escherichia coli* CTr5X. *Jornal of Molecular Biology*, 188, 15–22.
76. Chapman, D., Morad, I., Kaufmann, G., Gait, M.J., Jorissen, L. and Snyder, L. (1988) Nucleotide and deduced amino acid sequence of stp: The bacteriophage T4 anticodon nuclease gene. *Jornal of Molecular Biology*, 199, 373–377.

77. Amitsur, M., Morad, I., Chapman-Shimshoni, D. and Kaufmann, G. (1992) HSD restriction-modification proteins partake in latent anticodon nuclease. *EMBO Journal*, 11, 3129–3134.
78. Bidnenko, E., Chopin, M.C., Ehrlich, S.D. and Anba, J. (2002) *Lactococcus lactis* AbiD1 abortive infection efficiency is drastically increased by a phage protein. *FEMS Microbiology Letters*, 214, 283–287.
79. Bidnenko, E., Ehrlich, D. and Chopin, M.C. (1995) Phage operon involved in sensitivity to the *Lactococcus lactis* abortive infection mechanism AbiD1. *Journal of Bacteriology*, 177, 3824–3829.
80. Durmaz, E. and Klaenhammer, T.R. (2007) Abortive phage resistance mechanism AbiZ speeds the lysis clock to cause premature lysis of phage-infected *Lactococcus lactis*. In *Journal of Bacteriology*. Vol. 189, pp. 1417–1425.
81. Emond, E., Holler, B.J., Boucher, I., Vandenberg, P.A., Vedamuthu, E.R., Kondo, J.K. and Moineau, S. (1997) Phenotypic and genetic characterization of the bacteriophage abortive infection mechanism AbiK from *Lactococcus lactis*. *Applied and Environmental Microbiology*, 63, 1274–1283.
82. Prevots, F., Tolou, S., Delpech, B., Kaghad, M. and Daloyau, M. (1998) Nucleotide sequence and analysis of the new chromosomal abortive infection gene abiN of *Lactococcus lactis* subsp. *cremoris* S114. *FEMS Microbiology Letters*, 159, 331–336.
83. Prevots, F. and Ritzenthaler, P. (1998) Complete sequence of the new lactococcal abortive phage resistance gene abiO. *Journal of Dairy Science*, 81, 1483–5.
84. Jurėnas, D., Fraikin, N., Goormaghtigh, F. and Van Melderen, L. (2022) Biology and evolution of bacterial toxin-antitoxin systems. *Nature Reviews Microbiology*, 20.
85. Li, M., Gong, L., Cheng, F., Yu, H., Zhao, D., Wang, R., Wang, T., Zhang, S., Zhou, J., Shmakov, S.A., et al. (2021) Toxin-antitoxin RNA pairs safeguard CRISPR-Cas systems. *Science* 372(6541)
86. Fineran, P.C., Blower, T.R., Foulds, I.J., Humphreys, D.P., Lilley, K.S. and Salmond, G.P.C. (2009) The phage abortive infection system, ToxIN, functions as a protein-RNA toxin-antitoxin pair. *PNAS*, 106, 894–9.
87. Millman, A., Bernheim, A., Stokar-Avihail, A., Fedorenko, T., Voichek, M., Leavitt, A., Oppenheimer-Shaanan, Y. and Sorek, R. (2020) Bacterial Retrons Function In Anti-Phage Defense. *Cell*, 183, 1551-1561.e12.
88. Cohen, D., Melamed, S., Millman, A., Shulman, G., Oppenheimer-Shaanan, Y., Kacen, A., Doron, S., Amitai, G. and Sorek, R. (2019) Cyclic GMP–AMP signalling protects bacteria against viral infection. *Nature*, 5747780, 574, 691–695.
89. Gao, L., Altae-Tran, H., Böhning, F., Makarova, K.S., Segel, M., Schmid-Burgk, J.L., Koob, J., Wolf, Y.I., Koonin, E. V. and Zhang, F. (2020) Diverse Enzymatic Activities Mediate Antiviral Immunity in Prokaryotes. *Science*, 369, 1077.
90. Ofir, G., Herbst, E., Baroz, M., Cohen, D., Millman, A., Doron, S., Tal, N., Malheiro, D.B.A., Malitsky, S., Amitai, G., et al. (2021) Antiviral activity of bacterial TIR domains via immune signalling molecules. *Nature*, 6007887, 600, 116–120.
91. Mattijssen, S. and Pruijn, G.J.M. (2012) Viperin, a key player in the antiviral response. *Microbes Infections*, 14, 419–426.

92. Bernheim,A., Millman,A., Ofir,G., Meitav,G., Avraham,C., Shomar,H., Rosenberg,M.M., Tal,N., Melamed,S., Amitai,G., et al. (2021) Prokaryotic viperins produce diverse antiviral molecules. *Nature*, 589, 120.
93. Horton,R.M., Ho,S.N., Pullen,J.K., Hunt,H.D., Cai,Z. and Pease,L.R. (1993) Gene splicing by overlap extension. *Methods Enzymology*, 217, 270–9.
94. Liu,H., Han,J., Liu,X., Zhou,J. and Xiang,H. (2011) Development of *pyrF*-based gene knockout systems for genome-wide manipulation of the archaea *Haloferax mediterranei* and *Haloarcula hispanica*. *Journal of Genetics and Genomics*, 38, 261–269.
95. Zeng,L., Skinner,S.O., Sippy,J., Feiss,M. and Golding,I. (2010) Decision making at a subcellular level determines the outcome of bacteriophage infection. *Cell*, 141, 682–91.
96. Shao,Q., Hawkins,A. and Zeng,L. (2015) Phage DNA dynamics in cells with different fates. *Biophysical Journal*, 108, 2048–2060.
97. Oppenheim,A.B., Kobilier,O., Stavans,J., Court,D.L. and Adhya,S. (2005) Switches in Bacteriophage Lambda Development. *Annual Reviews in Genetics*, 39, 409–429.
98. Babic,A., Lindner,A.B., Vulic,M., Stewart,E.J. and Radman,M. (2008) Direct visualization of horizontal gene transfer. *Science*, 319, 1533–6.
99. Rhoads,A. and Au,K.F. (2015) PacBio Sequencing and Its Applications. *Genomics, Proteomics, Bioinformatics*, 13, 278–289.
100. Moore,D.A., Whatley,Z.N., Joshi,C.P., Osawa,M. and Erickson,H.P. (2017) Probing for Binding Regions of the FtsZ Protein Surface through Site-Directed Insertions: Discovery of Fully Functional FtsZ-Fluorescent Proteins. *Journal of Bacteriology*, 199.
101. Garcia,L.R. and Molineux,I.J. (1995) Rate of translocation of bacteriophage T7 DNA across the membranes of *Escherichia coli*. *Journal of Bacteriology*, 177, 4066.
102. Clark,D.P., Pazdernik,N.J. and McGehee,M.R. (2018) Molecular biology. 10.1016/C2015-0-06229-3.
103. Pleška,M. and Guet,C.C. (2017) Effects of mutations in phage restriction sites during escape from restriction–modification. *Biology Letters*, 13.
104. Lehman,I.R. and Pratt,E.A. (1960) On the structure of the glucosylated hydroxymethylcytosine nucleotides of coliphages T2, T4, and T6. *Journal of Biological Chemistry*, 235, 3254–3259.
105. Li,M., Wang,R., Zhao,D. and Xiang,H. (2014) Adaptation of the *Haloarcula hispanica* CRISPR-Cas system to a purified virus strictly requires a priming process. *Nucleic Acids Research*, 42, 2483–2492.
106. Demina,T.A., Atanasova,N.S., Pietilä,M.K., Oksanen,H.M. and Bamford,D.H. (2016) Vesicle-like virion of *Haloarcula hispanica* pleomorphic virus 3 preserves high infectivity in saturated salt. *Virology*, 499, 40–51.
107. Isaev,A., Drobiazko,A., Sierro,N., Gordeeva,J., Yosef,I., Qimron,U., Ivanov,N. V and Severinov,K. (2020) Phage T7 DNA mimic protein Ocr is a potent inhibitor of BREX defence. *Nucleic Acids Research*, 10.1093/nar/gkaa290.
108. Dreier,J., MacWilliams,M.P. and Bickle,T.A. (1996) DNA cleavage by the type IC restriction-modification enzyme EcoR124II. *Journal of Molecular Biology*, 264, 722–33.
109. Meisel,A., Bickle,T.A., Krieger,D.H. and Schroeder,C. (1992) Type III restriction enzymes need two inversely oriented recognition sites for DNA cleavage. *Nature*, 355, 467–9.

## Appendix A

Supplementary Table S1. Primers used for cloning in this work (5'-3' orientation).

<b>Name</b>	<b>Sequence</b>
<b>pACYC184_F</b>	<i>tatatctcgagccaatcaattcttgcgg</i>
<b>pACYC184_R</b>	<i>Atataaggcctccggcggtgctttgccg</i>
<b>pBAD30_F</b>	<i>tatataggcctgctagcggagtgtatactgg</i>
<b>pBAD30_R</b>	<i>atatactcgagtgagcggatacatatttg</i>
<b>pTG-BREX_F</b>	<i>tatatgcatgcaagcttggct</i>
<b>pTG-BREX_R</b>	<i>atatagagctcttgtaacgaatcagacaattg</i>
<b>delA_BglII_F</b>	<i>atattagatcttctateccatgccttgc</i>
<b>delA_BglII_R</b>	<i>atgattaaaaatgacaaggcaagatctaatat</i>
<b>delC_BglII_F</b>	<i>atataagatctattaaagcaggcgatcg</i>
<b>delC_BglII_R</b>	<i>tatatagatcttagaggtttttaaaaatctg</i>
<b>delB_F1</b>	<i>atatactgcaggcagtttacctttaccgg</i>
<b>delB_R1</b>	<i>gattcaatgttgctgcgggtcccgtagacaggcgatattcaagaacggga</i>
<b>delB_F2</b>	<i>tcccgttcttgaatatcgctgtctacgggacccgcagcaacattgaatc</i>
<b>delB_R2</b>	<i>atagtctgcagcttcagcatcattt</i>
<b>delZ_F1</b>	<i>atatagacgtcacctctcgcgtcaaagag</i>
<b>delZ_R1</b>	<i>aatcatcctggaacgccagatcgattttaaggccagcaataaagtctga</i>
<b>delZ_F2</b>	<i>tcaggactttattgctggccttaaatcgaatctggcgttcaggatgatt</i>
<b>delZ_R2</b>	<i>gaaatg gcgccagcaaatggctg</i>
<b>Brx_pro_SacI_F</b>	<i>ttatagagctcctaaagcatgtggttctttatttg</i>
<b>DelA_R</b>	<i>aatagatatcaagtcctatccatgccttgcattttta</i>
<b>DelA_F</b>	<i>aaggcatggataggagacttgatatctattctggagggaaaaac</i>
<b>DelB_XbaI_R</b>	<i>gtcgtttctagaacttcaccacc</i>
<b>DelB_F</b>	<i>cccgttcttgaatatgcagcaacattgaatcctcaatg</i>
<b>DelB_R</b>	<i>attcaatgttgctgcatattcaagaacgggatctatcacc</i>

<b>DelC_XbaI_R</b>	<i>tctagactgttcaatattcattgttatgctc</i>
<b>US_ΔA_XbaI_F</b>	<i>tatattctagaagtcctcgaagaaggcaacgt</i>
<b>US_ΔA_R</b>	<i>actcgacgcagccggtcgatgacacg</i>
<b>DS_ΔA_F</b>	<i>ccggctgcgtcgagtcgagcgacgg</i>
<b>DS_ΔA_KpnI_R</b>	<i>tatatggtacctactcggagccgtgtcaacat</i>
<b>US_ΔB_XbaI_F</b>	<i>tatattctagagagtatcgacggttacgaggg</i>
<b>US_ΔB_R</b>	<i>ttggagatgaagtcgccgaaggacgt</i>
<b>DS_ΔB_F</b>	<i>ggacttcatctccaatgagcgcacac</i>
<b>DS_ΔB_KpnI_R</b>	<i>tatatggtaccacacatectcgaacgtctggc</i>
<b>US_ΔC1_BamHI_F</b>	<i>tatatggatcctccctctattgtccagctccc</i>
<b>US_ΔC1_R</b>	<i>taataaccgtcggcagcgaaccagcc</i>
<b>DS_ΔC1_F</b>	<i>tgccgacggttattatccgatgacgacaaaaacg</i>
<b>DS_ΔC1_KpnI_R</b>	<i>tatatggtaccggccttcgatctggtgggc</i>
<b>US_ΔC2_BamHI_F</b>	<i>tatatggatccggacttcacggatcggatcg</i>
<b>US_ΔC2_R</b>	<i>atcacagagaagatgtcgtgaatgtgtgtgc</i>
<b>DS_ΔC2_F</b>	<i>catcttctctgtgatggtgatcgagtcgc</i>
<b>DS_ΔC2_KpnI_R</b>	<i>tatatggtaccgttcgagaactcgtcaagcacc</i>
<b>US_ΔX1_BamHI_F</b>	<i>tatatggatcctgacgcgcttgaatctcagg</i>
<b>US_ΔX1_R</b>	<i>ctccggcttgcgggggtgggtaga</i>
<b>DS_ΔX1_F</b>	<i>cccgaagccggaagtcgttgaggacaag</i>
<b>DS_ΔX1_KpnI_R</b>	<i>tatatggtacccttgttctagaagtttctctc</i>
<b>US_ΔX2_BamHI_F</b>	<i>tatatggatccgaaccggcggaaactcgt</i>
<b>US_ΔX2_R</b>	<i>ctccggatgagagtcagtgatatttcccag</i>
<b>DS_ΔX2_F</b>	<i>actctcatccggaagtgggtggcggac</i>
<b>DS_ΔX2_KpnI_R</b>	<i>tatatggtaccgcgacgaacaggacgcgg</i>
<b>US_ΔZ_BamHI_F</b>	<i>tatatggatcctcaactcaaagagaccactatcatgacc</i>
<b>US_ΔZ_R</b>	<i>aaatcgagggcttcgggaagggtctgtttc</i>
<b>DS_ΔZ_F</b>	<i>cgaagccctcgatttcatcagcatcacgcag</i>

<b>DS_ΔZ_KpnI_R</b>	<i>tatatggtaccaggctgtaggcggtcga</i>
<b>US_ΔHI_XbaI_F</b>	<i>tatattctagaagacggtacaggagaacaaagacg</i>
<b>US_ΔHI_R</b>	<i>tagttcttggcccggacgagaatgaagaatc</i>
<b>DS_ΔHI_F</b>	<i>cgggcaagagaactactgtttgactctccatt</i>
<b>DS_ΔHI_KpnI_R</b>	<i>tatatggtaccgctcgcgagacaagtcattcc</i>
<b>US_ΔNucS_BamHI_F</b>	<i>tatatggatccgaggcctcgtctggtcg</i>
<b>US_ΔNucS_R</b>	<i>gagtatctccgactctaagcatgatgtcgaga</i>
<b>DS_ΔNucS_F</b>	<i>gagtcggagatactcggcgagactaccg</i>
<b>DS_ΔNucS_KpnI_R</b>	<i>tatatggtaccgttcttcttctgtcgagctggg</i>
<b>US_ΔBREX_BamHI_F</b>	<i>atataggatccgacatcgtgttgacgacgctgc</i>
<b>US_ΔBREX_R</b>	<i>gccgaagacgtgcctgctaaatcggtttctgaatgagactacctc</i>
<b>DS_ΔBREX_F</b>	<i>gaggtagtctcattcagaaaccgatttagcagggcacgtcttcggc</i>
<b>DS_ΔBREX_KpnI_R</b>	<i>atataggtaccttttacgagtgcctggatcgacg</i>
<b>US_ΔX1-HI_BamHI_F</b>	<i>atataggatcccacctgtaaccgcaatcaggaga</i>
<b>US_ΔX1-HI_R</b>	<i>gaggtagtctcattcagaaaccgatacagaaatccgtgatgtttcaag</i>
<b>DS_ΔX1-HI_F</b>	<i>cttgaaacatcacggattctcgtatcggtttctgaatgagactacctc</i>
<b>DS_ΔX1-HI_KpnI_R</b>	<i>atataggtaccttttacgagtgcctggatcgacg</i>

Supplementary Table S2. Primers used for BREX<sup>Ec</sup> sequencing (5'-3' orientation).

<b>Brx_F1</b>	<i>gcttgctggtccgtaaactc</i>	<b>Brx_F10</b>	<i>gatgatectcaagagcctgg</i>
<b>Brx_F2</b>	<i>aatgggtggaacaactcgac</i>	<b>Brx_F11</b>	<i>gcagaaaaaacagccgcag</i>
<b>Brx_F3</b>	<i>cgatttctctcgatcaaca</i>	<b>Brx_F12</b>	<i>cgtgatcagtgatcgatgt</i>
<b>Brx_F4</b>	<i>cgctggttgaacatatctgt</i>	<b>Brx_F13</b>	<i>gactgccggtaatggttaagc</i>
<b>Brx_F5</b>	<i>aaaaatatgccccacaggc</i>	<b>pBTB_F</b>	<i>gcacggcgtcacactttgc</i>
<b>Brx_F6</b>	<i>ggttaccgtaccctgacat</i>	<b>pBAD_F</b>	<i>atgccatagcattttatcc</i>
<b>Brx_F7</b>	<i>ctcttggaactcgtgctgtg</i>	<b>Brx_R1</b>	<i>gtactaatggcagcgcttg</i>
<b>Brx_F8</b>	<i>acgagtggttgaggacgac</i>	<b>Brx_R2</b>	<i>actgaccgacttcacc</i>
<b>Brx_F9</b>	<i>cgcctcatgaacaagactgg</i>	<b>Brx_R3</b>	<i>agtcttgcttgtagccgga</i>

<b>Brx_R4</b>	<i>gttttccccggattgttc</i>
<b>Brx_R5</b>	<i>ttatctgcagcgatcccaag</i>
<b>Brx_R6</b>	<i>agcatcagcatcgcaaatc</i>
<b>Brx_R7</b>	<i>aatttatgctgactccggc</i>
<b>Brx_R8</b>	<i>cttcggcgataaattcgag</i>
<b>Brx_R9</b>	<i>cagcgtgttcaggggatg</i>

<b>Brx_R10</b>	<i>cttcgttccagttacggctt</i>
<b>Brx_R11</b>	<i>cgctcgcaacaggctcact</i>
<b>Brx_R12</b>	<i>ctccacgaacgggatcatac</i>
<b>Brx_R13</b>	<i>gcttaccgattcaaattgc</i>
<b>pBTB_R</b>	<i>cgaagcgagccagccgg</i>
<b>pBAD_R</b>	<i>gatttaatctgtatcagg</i>

Supplementary Table S3. Primers used for BREX<sup>HAR</sup> sequencing (5'-3' orientation).

<b>HAR_F1</b>	<i>ctcgtgaggtacacggtacat</i>
<b>HAR_F2</b>	<i>cattccggctcgggtaac</i>
<b>HAR_F3</b>	<i>cgatccggcaaagtcgatt</i>
<b>HAR_F4</b>	<i>ggcattcgacgtcaagga</i>
<b>HAR_F5</b>	<i>cagcagtcgtacgactgg</i>
<b>HAR_F6</b>	<i>gtagacgaggtgttcgtcg</i>
<b>HAR_F7</b>	<i>ctacagcaagtcgagcac</i>
<b>HAR_F8</b>	<i>cgctgtacctgctgaatca</i>
<b>HAR_F9</b>	<i>attcctcgacgagtacgc</i>
<b>HAR_F10</b>	<i>aattcgacggcgtaaccat</i>
<b>HAR_F11</b>	<i>atcgaactcaaacgagggcg</i>
<b>HAR_F12</b>	<i>tagaatccgctatttcggc</i>
<b>HAR_F13</b>	<i>cctacagcctgatcgacc</i>
<b>HAR_F14</b>	<i>tcacagtcttcgagatgcag</i>
<b>HAR_F15</b>	<i>agagacgggtagacggtt</i>
<b>HAR_F16</b>	<i>catatggaagctcagtacgga</i>
<b>HAR_F17</b>	<i>cagcgagtgaagatgaagga</i>
<b>HAR_F18</b>	<i>gagatcgcgcgattctc</i>
<b>HAR_F19</b>	<i>gaggaccactcctatcagttc</i>
<b>HAR_F20</b>	<i>ccgaactctgtactacct</i>

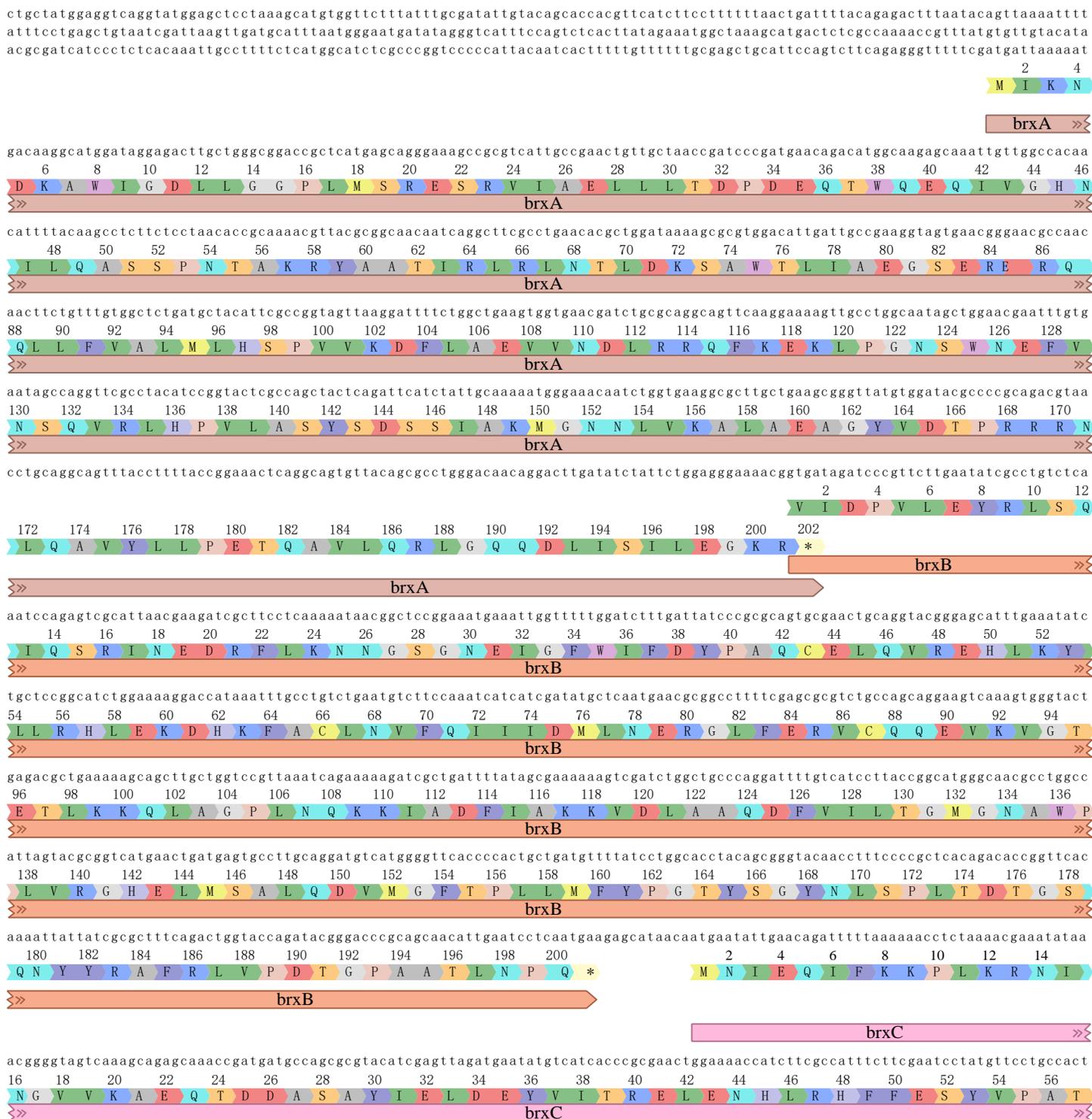
<b>HAR_F21</b>	<i>gcgttcgacttctctcag</i>
<b>HAR_F22</b>	<i>cagtacgatcatccacacc</i>
<b>HAR_F23</b>	<i>aacgacgaatcgctctcg</i>
<b>HAR_F24</b>	<i>gtacatcgatccgttctc</i>
<b>HAR_F25</b>	<i>cttggcctctcgacgac</i>
<b>pHAR_F</b>	<i>gaaacagctatgaccatgattacg</i>
<b>HAR_R1</b>	<i>ttcagcatagtcaaactcgctc</i>
<b>HAR_R2</b>	<i>gaagaaggcaacgtcgagtt</i>
<b>HAR_R3</b>	<i>ggctcgatgaggtcgtaga</i>
<b>HAR_R4</b>	<i>tcctgtaagtgttcgtcgc</i>
<b>HAR_R5</b>	<i>gtttggcgctcgttgaac</i>
<b>HAR_R6</b>	<i>gtagacgaggtgttcgtcg</i>
<b>HAR_R7</b>	<i>cgtegatcgactgcttga</i>
<b>HAR_R8</b>	<i>cttgaggaggcagacag</i>
<b>HAR_R9</b>	<i>cgtagcgagcgtgatcag</i>
<b>HAR_R10</b>	<i>ccggatgttctgaggagc</i>
<b>HAR_R11</b>	<i>ccacgacgtggacttgaa</i>
<b>HAR_R12</b>	<i>gggatcgagtttctgaacc</i>
<b>HAR_R13</b>	<i>gcaggtcatcgaggagtt</i>
<b>HAR_R14</b>	<i>ctgttccgcgatccgtag</i>

<b>HAR_R15</b>	<i>agagagtggtacaagtcgc</i>
<b>HAR_R16</b>	<i>gcgaagttcagctttccg</i>
<b>HAR_R17</b>	<i>ggctttgctcgatcaaca</i>
<b>HAR_R18</b>	<i>ggggctgtcgtcgatatac</i>
<b>HAR_R19</b>	<i>tagtcgcgcaactcgtc</i>
<b>HAR_R20</b>	<i>cagtacgtcgaacgcata</i>

<b>HAR_R21</b>	<i>tctcaccctttgcaacatac</i>
<b>HAR_R22</b>	<i>gagggtttgctcgactcg</i>
<b>HAR_R23</b>	<i>cgttcctgattcacgctg</i>
<b>HAR_R24</b>	<i>cgataccgagttgacctc</i>
<b>HAR_R25</b>	<i>tctaccactcctcaagc</i>
<b>pHAR_R</b>	<i>gcctcgtcaactcctcg</i>

## Appendix B

### pBREXAL plasmid sequence map:



ggccccgaaagatccgctatggaacaagaatcgccgctatgggttcaggcttcttcggttcaggtaaatcgcactttattaagattctttcttattctttatctaacccgaaagtacacataa  
 58 60 62 64 66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98  
 G P E R I R M E N K I G V W V S G F F G S G K S H F I K I L S Y L L S N R K V T H N  
 >> brxC >>

cggtagcgaacgtaaatgcttactccttcttgaagagaaaatcaagatgcattatccttgcgatattaacaaagggtgcattaccgactgaagtcattctgttcaatattgattcgctg  
 100 102 104 106 108 110 112 114 116 118 120 122 124 126 128 130 132 134 136 138 140  
 G T E R N A Y S F F E E K I K D A L F L A D I N K A V H Y P T E V I L F N I D S R  
 >> brxC >>

ccaacgtagatgacaaagaagatgccattcttaaagtcttctgaaagtttcaacgaacgcattggatactgcgctgattttccgcatattgcccatcttgagcgcgagctggataaacgcggt  
 142 144 146 148 150 152 154 156 158 160 162 164 166 168 170 172 174 176 178 180 182  
 A N V D D K E D A I L K V F L K V F N E R I G Y C A D F P H I A H L E R E L D K R G  
 >> brxC >>

cagtatgaaacctttaaagccgcttggcagatcaatggctcgcgctgggaagacgagcgcgacgcttactacttcatcagcgatgacatggcacaagcattaagccaggccacgcagcagag  
 184 186 188 190 192 194 196 198 200 202 204 206 208 210 212 214 216 218 220 222 224  
 Q Y E T F K A A F A D I N G S R W E D E R D A Y Y F I S D D M A Q A L S Q A T Q Q S  
 >> brxC >>

tcttgaagcctccgccaatgggtggaacaactcgacaaaacttcccgtggatatacaataattttgcccagtggtgtaaaagagtggtggatgacaatgtaagaacatcctctttaggtgg  
 226 228 230 232 234 236 238 240 242 244 246 248 250 252 254 256 258 260 262 264  
 L E A S R Q W V E Q L D K N F P L D I N N F C Q W V K E W L D D N G K N I L F M V  
 >> brxC >>

atgaagtcggtcagttcattggcaaaaatcgcaaatgatgctgaagctgcagactattactgaaaaccttgggtaatttgcggtggccgcgcatgggttatcgtgacttcgaggccgatc  
 266 268 270 272 274 276 278 280 282 284 286 288 290 292 294 296 298 300 302 304 306  
 D E V G Q F I G K N T Q M M L K L Q T I T E N L G V I C G G R A W V I V T S Q A D I  
 >> brxC >>

aacgcggcaatcggtggtatgagcagtcgcgacggacagacttctccaagatccaggggcgttctctacacgctgcaactttccagctctaacacatcagaagttatccagaaacgtttgtt  
 308 310 312 314 316 318 320 322 324 326 328 330 332 334 336 338 340 342 344 346 348  
 N A A I G G M S S R D G Q D F S K I Q G R F S T R L Q L S S S N T S E V I Q K R L L  
 >> brxC >>

ggtaaaagactgacgaagcaaaagcggcgtggcaaaaagtgggcaagagaaaggcgatactctcgtaaacagctggctttgacactacaacaactactgcactacgctctttaccagcgaag  
 350 352 354 356 358 360 362 364 366 368 370 372 374 376 378 380 382 384 386 388 390  
 V K T D E A K A A L A K V W Q E K G D I L R N Q L A F D T T T T T A L R P F T S E  
 >> brxC >>

aagagtctgtgacaactaccgcttctccgctggcactatcagattctgcaaaaagtgttgaatctatcggacgaaaggtgcagcgggtaacaatggccatgggtgagcgttctcagctg  
 392 394 396 398 400 402 404 406 408 410 412 414 416 418 420 422 424 426 428 430 432  
 E E F V D N Y P F V P W H Y Q I L Q K V F E S I R T K G A A G K Q L A M G E R S Q L  
 >> brxC >>

gaggcattccagacggcggcgcagcaaatctcagcgcagggtggattctctggtgccttcttgccgcttctatgcccattgagagcttcttggaaacctgcccgttagccgaccatca  
 434 436 438 440 442 444 446 448 450 452 454 456 458 460 462 464 466 468 470 472 474  
 E A F Q T A A Q Q I S A Q G L D S L V P F W R F Y A A I E S F L E P A V S R T I T Q  
 >> brxC >>

ggcttgcagaatggcattcttgatgagttcgaatggcaacctgcttaaacgctgttcttgatccgctatgtgaaacgctgaaagcacctggataacctggtcacattgtctatcgatagga  
 476 478 480 482 484 486 488 490 492 494 496 498 500 502 504 506 508 510 512 514  
 A C Q N G I L D E F D G N L L K T L F L I R Y V E T L K S T L D N L V T L S I D R  
 >> brxC >>

tcgatccgataaagttgagttgcccggcgggtcgaaaaagtctcaacacgcttgaacccgctgatgctcattgcccgcggttgaagataaataatgtgttcttgaccaacgaagaaaaagagatc  
 516 518 520 522 524 526 528 530 532 534 536 538 540 542 544 546 548 550 552 554 556  
 I D A D K V E L R R R V E K S L N T L E R L M L I A R V E D K Y V F L T N E E K E I  
 >> brxC >>

gaaaacgagatccgtaacgttgatgctgatttctctgcgatcaacaaaaactggcatcgatcatcttggatgacattctgaaaagccgtaaatatcgttatccggctaacaaagcaagactttga  
 558 560 562 564 566 568 570 572 574 576 578 580 582 584 586 588 590 592 594 596 598  
 E N E I R N V D V D F S A I N K K L A S I I F D D I L K S R K Y R Y P A N K Q D F D  
 >> brxC >>

tatcagccgcttctgaaacggcattagacggcagctgcttaacgatctggtggtgaaagatcctgaccctaaagatccgacttattcgttctataacagcgatgcgacctgctgcctt  
 600 602 604 606 608 610 612 614 616 618 620 622 624 626 628 630 632 634 636 638 640  
 I S R F L N G H P L D G A V L N D L V V K I L T P K D P T Y S F Y N S D A T C R P  
 >> brxC >>

atacgtcagaagcgcagcgtgattttgattcgtctgcccgaagaggccgcttactggagcgatattgatttagctgctcagactgaaaagttcctcaagataacgccgggcaacgtccggaa  
 642 644 646 648 650 652 654 656 658 660 662 664 666 668 670 672 674 676 678 680 682  
 Y T S E G D G C I L I R L P E E G R T W S D I D L V V Q T E K F L K D N A G Q R P E  
 >> brxC >>

caggctaccctgctctcagaaaaagcggctgaaacacgaaccgggaaaaattactccggttgcagttggaatcactactgcagaagcagatgtctggcgattggcgagcgttaccgaaaaa  
 684 686 688 690 692 694 696 698 700 702 704 706 708 710 712 714 716 718 720 722 724  
 Q A T L L S E K A R E N S N R E K L L R V Q L E S L L A E A D V W A I G E R L P K K  
 >> brxC >>

atcctccacgccatcgaaacattgctgatgaagcctgcccgttacgtgattgaaaacaccttcggcaagctgaagatgctgccccttttaacggcgacatctcccgtgaaattcatgcattactga  
 726 728 730 732 734 736 738 740 742 744 746 748 750 752 754 756 758 760 762 764  
 S S T P S N I V D E A C R Y V I E N T F G K L K M L R P F N G D I S R E I H A L L  
 >> brxC >>

cggttgagaacgacaccgaactggatctcggtaacctcgaagagtcacccccgacgccatgcgcgaggtagaacctggatcagatgaatatcgaatacaataaacccgggtatttacggat  
 766 768 770 772 774 776 778 780 782 784 786 788 790 792 794 796 798 800 802 804 806  
 T V E N D T E L D L G N L E E S N P D A M R E V E T W I S M N I E Y N K P V Y L R D  
 >> brxC >>

attctgaaccatttcgcgctgccccttatggctggcccgaagacgaagtgaacctgctggtagcccctgctgacctgcaaagtaattcagcttcagccagcaaaacaacaacgctgcaacgaaa  
 808 810 812 814 816 818 820 822 824 826 828 830 832 834 836 838 840 842 844 846 848  
 I L N H F A R R P Y G W P E D E V K L L V A R L A C K G K F S F S Q Q N N N V E R K  
 >> brxC >>

acaggcgtgggaattathtaataacagcccgcccatagcgaattgctctgcataaagttcgccgctatgatgaagcgcaggctgctaaagccgcaaacatggctgacatcgtcagcagc  
 850 852 854 856 858 860 862 864 866 868 870 872 874 876 878 880 882 884 886 888 890  
 Q A W E L F N N S R R R H S E L R L H K V R R H D E A Q V R K A A Q T M A D I A Q Q  
 >> brxC >>

cgtttaacgaacgggaagagccggcgtggttgaacatattcgtcaggtttttgaaagtggaagcaagagctgaacctattccgcgccaaagcagagggcggaacaatccggggaaaaacgag  
 892 894 896 898 900 902 904 906 908 910 912 914 916 918 920 922 924 926 928 930 932  
 P F N E R E E P A L V E H I R Q V F E E W K Q E L N V F R A K A E G G N N P G K N E  
 >> brxC >>

atgtaaccggctgcccctgcttaatgccattcttaatgagaaagaagatttggccctgatgaaaagctcctatcctgctgaaagatgaactcttgatttcagcgaagaccgtgaagatttgg  
 934 936 938 940 942 944 946 948 950 952 954 956 958 960 962 964 966 968 970 972 974  
 I E S G L R L L N A I L N E K E D F A L I E K V S S L K D E L L D F S E D R E D L V  
 >> brxC >>

cgacttctaccgtaagcaattcggccactggcaaaaactgggtgctgcccctgaattggcagctttaaacttaaccgcagcgcctggaaaaagacggcgagcggttaaagcgtggcgagctgg  
 976 978 980 982 984 986 988 990 992 994 996 998 1000 1002 1004 1006 1008 1010 1012 1014  
 D F Y R K Q F A T W Q K L G A A L N G S F K S N R S A L E K D A A A V K A L G E L  
 >> brxC >>

aaagcatctggcaaatgcccgaaccttataagcatctcaatgcatacagctgttgattgaacaggtccagaacgtcaaccatcagtttagtcgaacagcatcggcagcagccctcgaaacgatt  
 1016 1018 1020 1022 1024 1026 1028 1030 1032 1034 1036 1038 1040 1042 1044 1046 1048 1050 1052 1054 1056  
 E S I W Q M P E P Y K H L N R I T L L I E Q V Q N V N H Q L V E Q H R Q H A L E R I  
 >> brxC >>

gacgcccgcacgaggaagccgtcaacgctgctcgaagcgcacgccacgtcagaactgcaaaacagcttctgctgcccgatgcaaaaagccagaaaaagcgcctgaggtcagccagcttattcc  
 1058 1060 1062 1064 1066 1068 1070 1072 1074 1076 1078 1080 1082 1084 1086 1088 1090 1092 1094 1096 1098  
 D A R I E E S R Q R L L E A H A T S E L Q N S V L L P M Q K A R K R A E V S Q S I P  
 >> brxC >>

ggaaatttggcggaacagcaagagacaaaagcgtgcaaatggatcgagataaaaagattaaacctggtgatcgacagctgcgtaaaaagcaagaagcacaactcgggcagcaaatgaagcca  
 1100 1102 1104 1106 1108 1110 1112 1114 1116 1118 1120 1122 1124 1126 1128 1130 1132 1134 1136 1138 1140  
 E I L A E Q C E T K A L Q M D A D K K I N L W I D E L R K K Q E A Q L R A A N E A  
 >> brxC >>

aacgcgctcggagtcagaacagacttatgttgggtggaaaaaccgcttatccaaccggtaccgaaaaaacgcacatctggtgaatgctgccagtgagatgcgtaatgccaccggtggtgaagtt  
 1142 1144 1146 1148 1150 1152 1154 1156 1158 1160 1162 1164 1166 1168 1170 1172 1174 1176 1178 1180 1182  
 K R A A E S E Q T Y V V V E K P V I Q P V P K K T H L V N V A S E M R N A T G G E V  
 >> brxC >>

ctgaaacgaccgaacaggtggaaaaggcctcgcacagcttacgcacaacgctgctggccgcccattaaagcagggcagatcgattcgccttcagtaactcccatttcagggcagcactctgctgcc  
 1184 1186 1188 1190 1192 1194 1196 1198 1200 1202 1204 1206 1208 1210 1212 1214  
 L E T T E Q V E K A L D T L R T T L L A A I K A G D R I R L Q \*  
 >> brxC >>

ctttcaggttttctatgaataccaataacattaaaaaatatgccccacaggcccgttaacgacttccgcatgagggtgatccagaagctaacgacgcttgggatcgtgcagataaaaaaggca  
 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30 32 34 36  
 M N T N N I K K Y A P Q A R N D F R D A V I Q K L T T L G I A A D K K G  
 >> brxX >>

atttcagattcggcggcgaaccattggcgagaccgtcggttacggctcagtttgattaccgcttaccgacctccccgccggaacggctggtaaacgcgcccgtgagcagggttttag  
 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68 70 72 74 76 78  
 N L Q I A E A E T I G E T V R Y G Q F D Y P L S T L P R R E R L V K R A R E Q G F E  
 >> brxX >>

gtgctggttgagcactgcgctacacctggttaaccgcttatgtgcaattcgctatatggagctgcacggttatcttgatcacggcttccgtatggtgtcaccgggagactccgaccgctt  
 80 82 84 86 88 90 92 94 96 98 100 102 104 106 108 110 112 114 116 118 120  
 V L V E H C A Y T W F N R L C A I R Y M E L H G Y L D H G F R M L S H P E T P T A F  
 >> brxX >>

tgaggctgctggatcatgtgcccgaagtggaagctctgctaccagaaagtaagcgcagttggttgaatgaagctttccgtaatcaggacgaagctctgtaccgtgaaactgctgctggggc  
 122 124 126 128 130 132 134 136 138 140 142 144 146 148 150 152 154 156 158 160  
 E V L D H V P E V A E A L L P E S K A Q L V E M K L S G N Q D E A L Y R E L L L G  
 >> brxX >>

agtgccacgcccctgaccacgcgatgcccgttccctgtttgaaagcggtagatgacgaagcggaaactgctgctgccggataaacctgaccgctactgactccattctacgtgggctgggtgatgatatt  
 162 164 166 168 170 172 174 176 178 180 182 184 186 188 190 192 194 196 198 200 202  
 Q C H A L H H A M P F L F E A V D D E A E L L L P D N L T R T D S I L R G L V D D I  
 >> brxX >>

ccggaagaagactgggagcaggtagaggttatcggctggcigtatcagttctatatttcggaagaaagatgccgtgattggcaaagtgggaaagcgaagatattcctgccgccaccagct  
 204 206 208 210 212 214 216 218 220 222 224 226 228 230 232 234 236 238 240 242 244  
 P E E D W E Q V E V I G W L Y Q F Y I S E K K D A V I G K V V K S E D I P A A T Q L  
 >> brxX >>

gtttacgccgaactggattgtgcagtatctgggtacagaactccgttggccgccagtggttgcagacctaccggactcaccgctgaaagacaaaatggagtactacattgagcctgccggagcaaa  
 246 248 250 252 254 256 258 260 262 264 266 268 270 272 274 276 278 280 282 284 286  
 F T P N W I V Q Y L V Q N S V G R Q W L Q T Y P D S P L K D K M E Y Y I E P A E Q  
 >> brxX >>

gcgggaagtgcaggcgcagctggcggcgattaccggccagcattgagcctgaaagcattaaagtctggatccggcctcggttccgggcatatcctg Gcagaagcctataacgtgctgaag  
 288 290 292 294 296 298 300 302 304 306 308 310 312 314 316 318 320 322 324 326 328  
 T P E V Q A Q L A A I T P A S I E P E S I K V L D P A C G S G H I L A E A Y N V L K  
 >> brxX >>

gccatttacgaagagcgggttaccgtaccgctgacattccgcagctgattctggaaaaacaatatttcggcctggacatcgtatgctgctggcgcagcttccggatttgcgatgctgatgct  
 330 332 334 336 338 340 342 344 346 348 350 352 354 356 358 360 362 364 366 368 370  
 A I Y E E R G Y R T R D I P Q L I L E N N I F G L D I D D R A A Q L S G F A M L M L  
 >> brxX >>

ggcacgccaggtgaccgctgattctgggcccgttgggtgtagggctgaatattgtctccctacaggaagtaagctggatattgtctgaggttggaccaagctgaacttccatcagcatatgcaac  
 372 374 376 378 380 382 384 386 388 390 392 394 396 398 400 402 404 406 408 410  
 A R Q D D R R I L G R G V R L N I V S L Q E S K L D I A E V W T K L N F H Q H M Q  
 >> brxX >>

gcggcagcatgggagatattttaccgaagggagcggcactggccaataccgacagcgggaatataagctactgatcgtactctggcattgttccactagtgcaaaaacgctgggatcactgatt  
 412 414 416 418 420 422 424 426 428 430 432 434 436 438 440 442 444 446 448 450 452  
 R G S M G D M F T Q G T A L A N T D S A E Y K L L M R T L A L F T S A K T L G S L I  
 >> brxX >>

caggtaccgcaggaagacgaagcggcgttgaagcgtttctggagagattgtatcgtctggcgttgaaggtgatattcagcagaagaggctgcggcagagctgatcccatattcagcaggc  
 454 456 458 460 462 464 466 468 470 472 474 476 478 480 482 484 486 488 490 492 494  
 Q V P Q E D E A A L K A F L E R L Y R L A V E G D I Q Q K E A A A E L I P Y I Q Q A  
 >> brxX >>

gtggatactggcgcaacgttatgatgctgtggctcggaacccgcgctatattgggggggaaagggatgaatgggtgatcgaagagtttgcataaaaaacaattccggacagcaagctggatttt  
 496 498 500 502 504 506 508 510 512 514 516 518 520 522 524 526 528 530 532 534 536  
 W I L A Q R Y D A V V A N P P Y M G G K G M N G D L K E F A K K Q F P D S K S D L  
 >> brxX >>

ttgcgatgtttatgcaacatgcattttcttactcaagaaaacggttttaatgcgcaagtgaatgacagctatgagtgcttctgctagctatgaagcgttctggtgatgggtgctggataat  
 538 540 542 544 546 548 550 552 554 556 558 560 562 564 566 568 570 572 574 576 578  
 F A M F M Q H A F S L L K E N G F N A Q V N M Q S W M F L S S Y E A L R G W L L D N  
 >> brxX >>

aaaacctttattacgatggcacatttgggggctcgagcatttgggtcaaatctctggagaagtgttccaactaccgcatgggtgattaaaaataaccattcaggattttataaacctgtattttt  
 580 582 584 586 588 590 592 594 596 598 600 602 604 606 608 610 612 614 616 618 620  
 K T F I T M A H L G A R A F G Q I S G E V V Q T T A W V I K N N H S G F Y K P V F F  
 >> brxX >>

ccgttagttgacgataatgaagaacataaaaaaaaaacaatcttgaatcggatgaattgcttataaacactctgcagaatgactttataaaaaatcctggttaccaccattgcttactgggcaa  
 622 624 626 628 630 632 634 636 638 640 642 644 646 648 650 652 654 656 658 660  
 R L V D D N E E H K K N N L L N R M N C C F K N T L Q N D F K K I P G S P I A Y W A  
 >> brxX >>

cattggcgtttatataatctcttcttaaatgacctgctcttggaaactgctgctgtaaaggactagatacaaacgggtccattgatgtattttgctgagatggccggaagtgcagcataatcc  
 662 664 666 668 670 672 674 676 678 680 682 684 686 688 690 692 694 696 698 700 702  
 T L A F I N S F L K L P A L G T R A V K G L D T N G S I D V F L R R W P E V S I N S  
 >> brxX >>

ttgatgcattagggaaagtaactcaaaatggtttccgattgcaaaaaggggtgagttgaggaagtggttcggaaatcatgagtataatcaaaactatgaaatgatggcattgaattaagaaa  
 704 706 708 710 712 714 716 718 720 722 724 726 728 730 732 734 736 738 740 742 744  
 F D A L G K G N S K W F P I A K G G E L R K W F G N H E Y I I N Y E N D G I E L R K  
 >> brxX >>

aaacaaagcaaaccttgagaataaggatattgactttcagaggggggaacctggactgttgtatcaaccaccggttctcaatgagatatatgccaaaaggattccttttgaccaagcggtt  
 746 748 750 752 754 756 758 760 762 764 766 768 770 772 774 776 778 780 782 784 786  
 N K A N L R N K D M Y F Q E G G T W T V V S T T G F S M R Y M P K G F L F D Q G G  
 >> brxX >>

ctgctgtttttgtgaaataatgatgagctatcgatataataacatcctggcctgcatgaactctaataatatacaactactctgcaagtttaattgtcctacgcttaactttacaaccgggat  
788 790 792 794 796 798 800 802 804 806 808 810 812 814 816 818 820 822 824 826 828  
S A V F C E N N D E L S I Y N I L A C M N S K Y I N Y S A S L I C P T L N F T T G D  
»»  
brxX

gttaggaatccctgttataaaaaacaatcaccttgaagatttagcaaaaaagcaattgagatcaaaaagcagactggaaccaattcgagacatcttgggagttcagtaaaaaaagttat  
830 832 836 838 840 842 844 846 848 850 852 854 856 858 860 862 864 866 868 870  
V R K F P V I K N N H L E D L A K K A I E I S K A D W N Q F E T S W E F S K N K L I  
»»

»»  
brxX  
tgaacacaaggaacgttgccttattcgtatgcttagctattgtaactttcaagataaactgtatgaacagctagtttaattgaaaaaataattaataacataatcgaagaataactgggttta  
872 874 876 878 880 882 884 886 888 890 892 894 896 898 900 902 904 906 908 910  
E H K G N V A Y S Y A S Y C N F Q D K L Y E Q L V N I E K N I N N I I E E I L G F  
»»

»»  
brxX

aaatagaacaacagagaatagtgatttaattacataaactcgaacaaaatatatcgttatgggcaaatgaaaccaatgatacatcctaacaggcatcgagtgaccatttcggaactc  
912 914 916 918 920 922 924 926 928 930 932 934 936 938 940 942 944 946 948 950 952  
K I E T T E N S E L I T L N S N K I Y R Y G Q S E T N D T F L N R H R S D T I S E L  
»»

»»  
brxX

atctcatattcagtcggcctgccaatgggacgctattccctcgatcgcaagcctcgtctacgccatgaaggcaacaaagctttgcccgaacttgcgcctgaagtgctacaaaaccttccc  
954 956 958 960 962 964 966 968 970 972 974 976 978 980 982 984 986 988 990 992 994  
I S Y S V G C Q M G R Y S L D R E G L V Y A H E G N K G F A E L A A E G A Y K T F P  
»»

ggcagacaatgacggcatcctgcccgtgatggatgacgagtggttgaggacgactcaccctcgcgtcaaagagtttgcgcaccgtctggggcgaagagcacctgcagaaaacctcgaat  
996 998 1000 1002 1004 1006 1008 1010 1012 1014 1016 1018 1020 1022 1024 1026 1028 1030 1032 1034 1036  
A D N D G I L P L M D D E W F E D D V T S R V K E F V R T V W G E E H L Q E N L E  
»»

ttatcgccgaaagtctctgtttatacgcgatcaagccgaaaaaggcgaatctgcgctggagaccattcgtcgtatctttccacacagttctggaaagatcatatgaagatgtataaaaagcgc  
1038 1040 1042 1044 1046 1048 1050 1052 1054 1056 1058 1060 1062 1064 1066 1068 1070 1072 1074 1076 1078  
F I A E S L C L Y A I K P K K G E S A L E T I R R Y L S T Q F W K D H M K M Y K K R  
»»

ccaatctactggttattcagctccggtaaaagagaagccttcgagtgctggtgtatctgcacgctataacgacgccacgctgtcgagaatgctaccgaatattgtggtgccgctgctggcagc  
1080 1082 1084 1086 1088 1090 1092 1094 1096 1098 1100 1102 1104 1106 1108 1110 1112 1114 1116 1118 1120  
P I Y W L F S S G K E K A F E C L V Y L H R Y N D A T L S R M R T E Y V V P L L A R  
»»

ttatcaggccaatcgatecgcctgaacgatcaacttgatgaagctctgcccgggtaagccacacgcctgaaacgcgaacgcgacagcctgatcaaaaaattagcgaactgcgtagctatgac  
1122 1124 1126 1128 1130 1132 1134 1136 1138 1140 1142 1144 1146 1148 1150 1152 1154 1156 1158 1160  
Y Q A N I D R L N D Q L D E A S G G E A T R L K R E R D S L I K K F S E L R S Y D  
»»

accgctcgcgtcactatgctgatatgagaatcagttatgatctcgacgatggcgttaaggttaactacggcaagtttggcgatctgctggcagatgtcaaaagccatcaccggcaatgccccagag  
1162 1164 1166 1168 1170 1172 1174 1176 1178 1180 1182 1184 1186 1188 1190 1192 1194 1196 1198 1200 1202  
D R L R H Y A D M R I S I D L D D G V K V N Y G K F G D L L A D V K A I T G N A P E  
»»

gggatcctaaccagacggcgcttcccctgctgcccgggttctgcccgggtggcaaataccaccgggaaacgcgcgctgctggcatttctccacctcacttcatgataaaatgcgccaccgtgca  
1204 1206  
A I \*  
»»

aaatcctcttttcgcttttggcgcttcttattcatcgtaacaacatgggatttgaacttgcaaaatcaggactttattgctggccttaaagctaaatttggcgaacatcgatcgttttctg  
2 4 6 8 10 12 14 16 18 20 22 24  
V N L Q N Q D F I A G L K A K F A E H R I V F W  
»»  
brxZ

gcacgateccgataaaagcttttattgaggaactggaacagctcaagcttgaaaagcgtcagcctaatcaacatgacccacgcgtcacagctggcggtgaaaaaacgcatcgagatggatgagccag  
26 28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64  
H D P D K R F I E E L E Q L K L E N V T L I N M T H A S Q L A V K K R I E M D E P  
»»

aacagcagttcctgctgtggttcccctatgatgcgccctcatgaacaagactggctgctggatattccgctttacagcagcgaattccatgccgattttgccgccatcaccctgaacagctg  
66 68 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 100 102 104 106  
E Q Q F I L W F P H D A P P H E Q D I L D I R L Y S S E F H A D F A A I T I N T L  
»»

»»  
brxZ

ggcattcccagcttggcctgctgagcatattcagcagcgaaggccttcttcagcactaaactacgcagcgtgaaaaacctggtgacggaacaggaagatgaaacctcgtggaatgaa  
108 110 112 114 116 118 120 122 124 126 128 130 132 134 136 138 140 142 144 146 148  
G I P Q L G L R E H I Q R R K A F F S T K R T Q A L K N L V T E Q E D E T S L D K K  
»»

aatgattgCGGtGatGctgGcgcaaaagaccgCGaaaaccgaagacattttgtttaacctgattaccagTactgTtaaccaacaacagaagacGacagcGaaactggaaaacacgcagGcgatgt  
150 152 154 156 158 160 162 164 166 168 170 172 174 176 178 180 182 184 186 188 190  
M I A V I A G A K T A K T E D I L F N L I T Q Y V N Q Q T E D D S E L E N T Q A M  
»» brxZ »»

tgaaacgccacggtctggaccgggtattgtgggaaatgctcaaccacgaaatgggctaccagGagggacccatcgctggaaaacctgctcctgaaactgtttgtaccgatctctctgccag  
192 194 196 198 200 202 204 206 208 210 212 214 216 218 220 222 224 226 228 230 232  
L K R H G L D P V L W E M L N H E M G Y Q A E E P S L E N L L L K L F C T D L S A Q  
»» brxZ »»

gccgacagacagcagcgcgctggctggaaaaaatgctcctgctgacgccatccggcagagcatccgccctggcatttatggTgacctggcgtgccgatcgctcgtataaagaggcttatgacta  
234 236 238 240 242 244 246 248 250 252 254 256 258 260 262 264 266 268 270 272 274  
A D R Q Q R A W L E K N V L L T P S G R A S A L A F M V T W R A D R R Y K E A Y D Y  
»» brxZ »»

ctgcgctcagcaaatgcagGcgccctgCGccccggaagatcattaccgactcagttcgccgTatgatTTgcacgaatgGaaaaccacctcagcatGaaacaaccattattcatcgctggtaa  
276 278 280 282 284 286 288 290 292 294 296 298 300 302 304 306 308 310 312 314  
C A Q Q M Q A A L R P E D H Y R L S S P Y D L H E C E T T L S I E Q T I I H A L V  
»» brxZ »»

cgcaactgctggaagagacaccacgctcgatcgGgaagcctttaaaaaactgctcctgaaGccagagcaaaatctggTgccagacacaaccagatattacgccatctatgacgGctgGc  
316 318 320 322 324 326 328 330 332 334 336 338 340 342 344 346 348 350 352 354 356  
T Q L L E E S T T L D R E A F K K L L S E R Q S K Y W C Q T Q P E Y Y A I Y D A L R  
»» brxZ »»

caggctgagcgTttgctgaatcTgCGcaatcgccacatcgatggTttccactaccagGacagcgccaccttctggaaaGctactgGaaagaactgttccgcttcgaccagGcttatcgccTgtt  
358 360 362 364 366 368 370 372 374 376 378 380 382 384 386 388 390 392 394 396 398  
Q A E R L L N L R N R H I D G F H Y Q D S A T F W K A Y C E E L F R F D Q A Y R L F  
»» brxZ »»

taatgaatatgccctgctgTttcacagcaaaagggcgatgatcctcaagacgctggatgattatctgagGcgtctacagcaactggTatctggcagaattagccgtaactggaacgaagTgc  
400 402 404 406 408 410 412 414 416 418 420 422 424 426 428 430 432 434 436 438 440  
N E Y A L L V H S K G A M I L K S L D D Y I E A L Y S N W Y L A E L S R N W N E V  
»» brxZ »»

tggaagcgGaaaatcgTatgcagGcgTggcaaatccccgGcgtgCCgcgtcagaagaacttttcaatgagGtgGtgaagccgagTtccaaaatccGcaaatcaaacgGgttctgtgataatt  
442 444 446 448 450 452 454 456 458 460 462 464 466 468 470 472 474 476 478 480 482  
tccgatccctcgcttatgaagTggcgGagagctggGaatcaaatcaatccGagaacgctttaccGagaactcgctcGcagctcggGgtgctccccagctacaccaactggGaatggc  
484 486 488 490 492 494 496 498 500 502 504 506 508 510 512 514 516 518 520 522 524  
S D A L R Y E V A E E L G N Q I N T E K R F T A E L R S Q L G V L P S Y T Q L G M A  
»» brxZ »»

ggcattgctGccccatgaacaacttTgctatcaaccaagtaacggcgacatcgTttatgctgatggGctgTgacctcgGgtatttccTaccGgataaccattctgaagaactataagGgaatgg  
526 528 530 532 534 536 538 540 542 544 546 548 550 552 554 556 558 560 562 564  
A L L P H E Q L C Y Q P S N G D I V Y A D G L S T S G I P N R D T I L K N Y K G M  
»» brxZ »»

cgataaaatcgaaggaccttctggagTtaaaaaatcaggaagggcgGaccttattcgGattacgaagTgTtatatctggcataacacgattgatgccaccggtgacacggcatccacggaa  
566 568 570 572 574 576 578 580 582 584 586 588 590 592 594 596 598 600 602 604 606  
A I K S K D L L E L K N Q E G R D L I R D Y E V V Y I W H N T I D A T G D T A S T E  
»» brxZ »»

L E A E N R M Q A W Q I P G V P R Q K N F F N E V V K P Q F Q N P Q I K R V F V I I  
gataaaaccttgaagcgTgcccacggcgGactgaaactgaaagatttagTcaccagGtaataaccgTctccacggcacacgcatTTTgttacggcgGatcacggttctcgttccagca  
608 610 612 614 616 618 620 622 624 626 628 630 632 634 636 638 640 642 644 646 648  
D K T F E A C R T A V T E L K D L V T K V I N R L H G T R I F V T A D H G F L F Q Q  
»» brxZ »»

acagGcgcttctggttcagGataaaaccactctgcaaatTaagccgGaaaacaccatcaagaaccacaaacgctttattatcgGccatcagcttcccccgatgattttgctgGaaagGaaag  
650 652 654 656 658 660 662 664 666 668 670 672 674 676 678 680 682 684 686 688 690  
Q A L S V Q D K T T L Q I K P E N T I K N H K R F I I G H Q L P A D D F C W K G K  
»» brxZ »»

tggcgGataaccgagGcgtgagcgacaacagcgagTtcttattccGaaagggatccagcgcttccatttctgGcgGcgGcgttctgTtcatggcgGaccatgTtgcagGagTtgcgtt  
692 694 696 698 700 702 704 706 708 710 712 714 716 718 720 722 724 726 728 730 732  
V A D T A G V S D N S E F L I P K G I Q R F H F S G G A R F V H G G T M L Q E V C V  
»» brxZ »»

ccggtattgcagataaaagccctgcaaaaaaccgGcagaaaaaacgGcagcgGccccggTggatattgTcgettaccatcccatgattaaagctggTgaaacaatatgataaaagtGagcct  
734 736 738 740 742 744 746 748 750 752 754 756 758 760 762 764 766 768 770 772 774  
P V L Q I K A L Q K T A A E K Q P Q R R P V D I V A Y H P M I K L V N N I D K V S L  
»» brxZ »»

gttgcagacgcatccgggtggggaactttatgaaccgctatcctgaacatttacattgtcgcataacgccaacaatgtggctcgggcaagagcgcatcagctttgacagcgataacagcacca  
 776 778 780 782 784 786 788 790 792 794 796 798 800 802 804 806 808 810 812 814  
 L Q T H P V G E L Y E P R I L N I Y I V D N A N N V V S G K E R I S F D S D N S T  
 >> brxZ >>

tggaaaaacggtacgcgaagttagcgtgaagctgattggcgctaacttcaaccgtcgcaatgagctactgggtgatactgggaagacgcacaacggaaacggggtatcagaagtaccgggtcatt  
 816 818 820 822 824 826 828 830 832 834 836 838 840 842 844 846 848 850 852 854 856  
 M E K R V R E V T L K L I G A N F N R R N E Y W L I L E D A Q T E T G Y Q K Y P V I  
 >> brxZ >>

atcgatctggcgttccaggatgatttcttctaagtgaggcgatgcaaacccatcatgatttacctgtttcaggcgatcccgagggaattgccctccgagggttacgatctggacgacctgc  
 858 860 862 864 866 868 2 4 6 8 10 12 14 16 18 20 22 24 26  
 I D L A F Q D D F F \* M Q T H H D L P V S G V S A G E I A S E G Y D L D A L  
 >> brxZ >> brxL >>

tgaaccagcattttgctggctggtgctgtaagatctccaagcaactcaaggaaggggcaaacgtcccgggtgatgtgctggagatctgctcggcatgtactgcccctctgacgatgac  
 gacgtggctgagcaaggggttcaaaacggttaagcgtattctggctgataaactatgtgcgccggatgaagcagagaaagtgaagctgctgatcccgagcgtgggttcgtacaaaatcatcgataa  
 70 72 74 76 78 80 82 84 86 88 90 92 94 96 98 100 102 104 106 108 110  
 D V V E Q G L Q N V K R I L A D N Y V R P D E A E K V K S L I R E R G S Y K I I D K  
 >> brxL >>

agtgtcggtagaagctaaaccagaaaaaacgctttacgaagcccagcttttcaacctcggcatcaaacgagcgcctgggtgccatcgagatgggttaaagacaacgagaagctactaacggggcgta  
 112 114 116 118 120 122 124 126 128 130 132 134 136 138 140 142 144 146 148 150 152  
 V S V K L N Q K K D V Y E A Q L S N L G I K D A L V P S Q M V K D N E K L L T G G  
 >> brxL >>

tctggctgatgattaccgtcaactatttcttgaagaagggcagaagacttcgcccttctcattgatgacgcttaagcctatccagatgccgaatggtgatggaagaagtgttcgatgcccgt  
 154 156 158 160 162 164 166 168 170 172 174 176 178 180 182 184 186 188 190 192 194  
 I W C M I T V N Y F F E E G Q K T S P F S L M T L K P I Q M P N M D M E E V F D A R  
 >> brxL >>

28 30 32 34 36 38 40 42 44 46 48 50 52 54 56 58 60 62 64 66 68  
 L N Q H F A G R V V R K D L T K Q L K E G A N V P V Y V L E Y L L G M Y C A S D D D  
 >> brxL >>

aaacactttaaccgtgatcagtgagctgatgtgctgctgcctcagtggtgatggagcccgccaatattgagcaacgcaccaaatggcaccattatcacccgtatgatcccggttcgtggagaacaa  
 196 198 200 202 204 206 208 210 212 214 216 218 220 222 224 226 228 230 232 234 236  
 K H F N R D Q W I D V L L R S V G M E P A N I E Q R T K W H L I T R M I P F V E N N  
 >> brxL >>

ctataacggttcgagctggggccgctggcaccggtaaaagccatggtataaagagtggttctcctaactcctggttagtttctggcgggcaaacgaccgttgccaactgttctacaacatgg  
 238 240 242 244 246 248 250 252 254 256 258 260 262 264 266 268 270 272 274 276  
 Y N V C E L G P R G T G K S H V Y K E C S P N S L L V S G G Q T T V A N L F Y N M  
 >> brxL >>

ccagtcgccagatcgccctggttggcatgtgggatgtgtagcgttcgacgaagtcgccccgatcactttcaagataaagacggcgtgcaaatcatgaaagattacatggcgtcaggatctttc  
 278 280 282 284 286 288 290 292 294 296 298 300 302 304 306 308 310 312 314 316 318  
 A S R Q I G L V G M W D V V A F D E V A G I T F K D K D G V Q I M K D Y M A S G S F  
 >> brxL >>

tctcggcgagagattcgattgaaggtaaagcgtcgatggttttctcgtggcaacatcaatcaaaagctagagactctcgttaaaccagccatttgcggcgcatttccggctgcatgattga  
 320 322 324 326 328 330 332 334 336 338 340 342 344 346 348 350 352 354 356 358 360  
 S R G R D S I E G K A S M V F V G N I N Q S V E T L V K T S H L L A P F P A A M I D  
 >> brxL >>

tactgcatttttcgaccgtttcatgcttatattcccggttgggaaatccccaaaatgcccggaaatttttaccaaccgttacggctgattacggattatctcgtggaatataatgcggaaa  
 362 364 366 368 370 372 374 376 378 380 382 384 386 388 390 392 394 396 398 400 402  
 T A F F D R F H A Y I P G W E I P K M R P E F F T N R Y G L I T D Y L A E Y M R E  
 >> brxL >>

tgcgcaaacgagtttctcgtgatgcatgataaattctttaagctgggtaacaacctcaaccagctgacgttatttgcgcttcgacgtaccgctcggggttgttaaactcatgcatcccgat  
 404 406 408 410 412 414 416 418 420 422 424 426 428 430 432 434 436 438 440 442 444  
 M R K R S F S D A I D K F F K L G N N L N Q R D V I A V R R T V S G L L K L M H P D  
 >> brxL >>

ggcgctacagcaagaagatgtgcgagctgcctgacctatgcatggaagtctcgtcccgcgtgaaagagcaactaaaaaacggggcgtcggagttcttcgatgtgaacttagctacat  
 446 448 450 452 454 456 458 460 462 464 466 468 470 472 474 476 478 480 482 484 486  
 G A Y S K E D V R V C L T Y A M E V R R V K E Q L K K L G G L E F F D V N F S Y I  
 >> brxL >>

cgacaacgaaacgtggaagagtttttgtgagcgtaccggaacagggcggcagcgaacttatctcgcggaatgccaaagccgggtgttgtgcatctggctcactcaggcagaaaagcggcatga  
488 490 492 494 496 498 500 502 504 506 508 510 512 514 516 518 520 522 524 526  
D N E T L E E F F V S V P E Q G G S E L I P A G M P K P G V V H L V T Q A E S G M  
»» brxL »»

ccggcctgtatcgttttgaaacacagatgactgccggtaatggtaagcatagtgatcgggctcgggttcaaatacctccgcgaaagaagctatccgcctcggtttcgattacttcaaaggcaat  
528 530 532 534 536 538 540 542 544 546 548 550 552 554 556 558 560 562 564 566 568  
T G L Y R F E T Q M T A G N G K H S V S G L G S N T S A K E A I R V G F D Y F K G N  
»» brxL »»

ttgaatcgggtaagcgcggccgcgaaattctccgatcatgaatatcaccttcatgtcgttgaactgcataatactggcccaagcaccgcaaccagtccttgcctcgttatcgtttatgttcgat  
570 572 574 576 578 580 582 584 586 588 590 592 594 596 598 600 602 604 606 608 610  
L N R V S A A A K F S D H E Y H L H V V E L H N T G P S T A T S L A A L I A L C S I  
»» brxL »»

attgctggcaaaacgggtgcaggaacagatgggtggttgggagcagtgacgcttgggtgggtaataaccgggtgcagatcttgcgcagcttacagctcgccttcgacagcgggtgcaaaac  
612 614 616 618 620 622 624 626 628 630 632 634 636 638 640 642 644 646 648 650 652  
L L A K P V Q E Q M V V L G S M T L G G V I N P V Q D L A A S L Q L A F D S G A K  
»» brxL »»

gggttctgttgccgatgctcctcggctatggatattccaacgggtccggcagagttattaccaagtttcaggtagtttttactcagaccgggtgatgctgtttataaggcctgggtgtgaat  
654 656 658 660 662 664 666 668 670 672 674 676 678 680 682 684 686 688 690 692 694  
R V L L P M S S A M D I P T V P A E L F T K F Q V S F Y S D P V D A V Y K A L G V N  
»» brxL »»

taagcatgcaacgaattcaagcttgatgatcattcaggacgagcctcagactccagcgtactggactgaaacaaactaaagcccttgtggcgttttagttttgttccgcggccaccggctgg  
\*  
»»

ctcgttccgtcggccccgtggacaaccctgtcggacaagctgatggacagcctgcccctgccacagccttgaccacagggattgcccaccggctaccagccttcgaccacataccaccggct  
ccaactgcgcggcctcggccttgccecatcaatttttttaattttctctggggaaaagcctccggcctgcggcctgcgcgcttcgcttgcgggttgacaccaagtggaaggcgggtcaagcct  
cgcgcagcagaccgcgcagcggcttggccttgacgcgcctggaacgacccaagcctatgcgagtgggggcagtcgaaagggcgaagcccccccgcctgcccccgagcctcagggcggcagtgccg  
gggttccaagggggcagcgcacccttgggcaagggcgaagggcgcgagtcgatcaacaagccccggagggggccactttttgcccggagggggagccgcgccaagggcgtgggggaacccccgacgg  
gggtgcccttctttggccacaaagaactagatataggggcaaatgcgaagacttaaaaatcaacaacttaaaaaaggggggtacgcaacagctcattgcggaacccccgcaatagctcattgc  
gtaggttaagaaaatctgtaattgactgccacttttacgcaacgcataattgttgcgcgctgccgaaaagttgcagctgattgcatggtgcccgaaccgtgcggcaacccctacgcgatgga  
gataagcatggccacgcagtcagagaaaatggcattcaagccaagaacaagccccggtcactgggtgcaaacggaaacgaaagcgcagtagggcgtgggcccgggcttattgagagaaacccacgg  
cggcaatgctgctgcatcacctcgtggcgcagatgggcccaccagaaccccgtgggtggtcagccagaagacactttccaagctcactcggaagttctttgcgagcgttcccaatcgcagtcaggac  
ttggtggccgagcgtgatctccgtcgtgaagctcaacggccccggcaccgctgctggcctacgtggtcaatgaccgctggcgtggggccagccccgcgaccagttgcgctgctggtgttcag  
tgcccctggtggttgatcagcagcaccaggacgaatcgtgttggggcattggcgacctgcgcgcacatccgacctgtatccgggcgagcagcaactaccgaccggccccggcagggagccgc  
ccagccagccccgcatccgggcatggaaccagacctgccagccttgaccgaaacggaggaaatgggaacgacgcgggcagcagcgcctgccaatgccgatgagccgtgttttctggacgatggc  
gagccgttggagccgcgacagggctacgctgcgcgcccgtacgtaagaggttccaactttcaccataatgaaataagatcactaccggcgtatttttgagttatcgagatttccaggagct  
aaggaagctaaaatgagccatattcaacgggaaacgtcttctcagagccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataatgctgggcaatcaggtgc  
gacaatctatcgattgtatgggaagcccgatgcgccagagttgttctgaaacatggcaaaagtagcgttgccaatgatgttacagatgagatggtcagcgtaaactggctgacggaatttatgc  
ctctccgaccatcaagcattttatecgtactcctgatgatgcatggttactcaccactgcgacccagggaaacagcattccaggtattagaagaatctcgtattcaggtgaaaatattgtt  
gatgcgctggcagtgctcctcgcgggttgcatcagctcctgctttgtaattgtccttttaacggcagtcgcttattcgtctcgtcagcgcgcaateacgaatgaataacgggttgggtggtgc  
gagtgatttggatgacgagcgaatggcctggtgaaacagcttgaaagaaatgcataagcttttgccattctcaccggattcagctgctcactcatggtgatttctcacttgataacctta  
ttttgacagggggaaatgaataggttggatgattggtgacagagtcggaatcgcagaccgataaccggatcttgccatcctatggaactgcctcggtagttttctccttcattacgaaacgg  
ctttttcaaaaatattggtattgataatcctgatataaataattgcagtttcaacttgatgctcagtagtttttctgagggcggatccccctcaagctcaaaagcctccggctcggaggtttttgac  
ttt