

<https://doi.org/10.15388/vu.thesis.1010>

<https://orcid.org/0000-0002-2972-1851>

VILNIUS UNIVERSITY

Aurimas Kopūstas

Mechanistic Studies of Bacterial Defense System Proteins' Interactions with Nucleic Acids

DOCTORAL DISSERTATION

Natural Sciences,
Biochemistry (N 004)

VILNIUS 2026

The dissertation was prepared between 2021 and 2025 at Vilnius University.

The research was supported by the Research Council of Lithuania that granted the funding for the doctoral student's research visit to KU Leuven (Leuven, Belgium) for two weeks during the spring semester in 2023.

Academic Supervisors:

Assoc. Prof. Dr. Marijonas Tutkus (Vilnius University, Natural Sciences, Physics – N 002; Supervision time frame: 1st of October 2024 – 30th of September 2025);

Dr. Mindaugas Zaremba (Vilnius University, Natural Sciences, Biochemistry – N 004; Supervision time frame: 1st of October 2021 – 30th of September 2024). **Academic Consultants:**

Dr. Mindaugas Zaremba (Vilnius University, Natural Sciences, Biochemistry – N 004; Consulting time frame: 1st of October 2024 – 30th of September 2025);

Assoc. Prof. Dr. Marijonas Tutkus (Vilnius University, Natural Sciences, Physics – N 002; Consulting time frame: 1st of October 2021 – 30th of September 2024).

This doctoral dissertation will be defended in a public meeting of the Dissertation Defence Panel:

Chairman – Assoc. Prof. Dr. Rima Budvytytė (Vilnius University Life Sciences Center, Natural Sciences, Biochemistry – N 004). **Members:**

Dr. Rūta Gerasimaitė (Max Planck Institute, Germany, Natural Sciences, Biochemistry – N 004),

Prof. Dr. Johannes Hohlbein (Brandenburg University of Technology, Germany, Natural Sciences, Biophysics – N 011),

Prof. Dr. Saulius Serva (Vilnius University, Natural Sciences, Biochemistry – N 004),

Dr. Justas Svirelis (Aligned Bio, Sweden, Natural Sciences, Biochemistry – N 004).

The dissertation shall be defended at a public meeting of the Dissertation Defence Panel at 1 PM on the 27th of November 2026 in Room R401 of the Life Sciences Center (Vilnius University). The dissertation will be defended in English language.

Address: Saulėtekio Ave. 7, R401, Vilnius, Lithuania

Tel. +370 5 223 4419; e-mail: info@gmc.vu.lt

<https://doi.org/10.15388/vu.thesis.1010>

<https://orcid.org/0000-0002-2972-1851>

VILNIAUS UNIVERSITETAS

Aurimas Kopūstas

Bakterinių apsaugos sistemų baltymų sąveikos su nukleorūgštimis mechanistiniai tyrimai

DAKTARO DISERTACIJA

Gamtos mokslai,
Biochemija (N 004)

VILNIUS 2026

Disertacija rengta 2021–2025 metais Vilniaus universitete. Mokslinius tyrimus rėmė Lietuvos mokslo taryba paskirdama finansavimą doktoranto akademinei išvykai į Leveno katalikiškąjį universitetą (Levenas, Belgija) dviem savaitėms 2023 m. pavasario semestrą.

Moksliniai vadovai:

doc. dr. Marijonas Tutkus (Vilniaus universitetas, gamtos mokslai, fizika – N 002; vadovavimo laikotarpis: 2024 m. spalio mėn. 1 d. – 2025 m. rugsėjo mėn. 30 d.);

dr. Mindaugas Zaremba (Vilniaus universitetas, gamtos mokslai, biochemija – N 004; vadovavimo laikotarpis: 2021 m. spalio mėn. 1 d. – 2024 m. rugsėjo mėn. 30 d.).

Moksliniai konsultantai:

dr. Mindaugas Zaremba (Vilniaus universitetas, gamtos mokslai, biochemija – N 004; konsultavimo laikotarpis: 2024 m. spalio mėn. 1 d. – 2025 m. rugsėjo mėn. 30 d.);

doc. dr. Marijonas Tutkus (Vilniaus universitetas, gamtos mokslai, fizika – N 002; konsultavimo laikotarpis: 2021 m. spalio mėn. 1 d. – 2024 m. rugsėjo

mėn. 30 d.).

Gynimo taryba:

Pirmininkė – doc. dr. Rima Budvytytė (Vilniaus universitetas Gyvybės mokslų centras, gamtos mokslai, biochemija – N 004).

Nariai:

dr. Rūta Gerasimaitė (Makso Planko institutas, Vokietija, gamtos mokslai, biochemija – N 004),

prof. dr. Johannes Hohlbein (Brandenburgo technologijos universitetas, Vokietija, gamtos mokslai, biofizika – N 011),

prof. dr. Saulius Serva (Vilniaus universitetas, gamtos mokslai, biochemija – N 004),

dr. Justas Svirelis (Aligned Bio, Švedija, gamtos mokslai, biochemija – N 004).

Disertacija ginama viešame Gynimo tarybos posėdyje 2026 m. lapkričio mėn. 27 d. 13:00 val. Vilniaus universiteto Gyvybės mokslų centro R401 auditorijoje. Disertacija bus ginama anglų kalba. Adresas: Saulėtekio al. 7, R401, Vilnius, Lietuva, tel. +370 5 223 4419; el. paštas info@gmc.vu.lt

CONTENTS

LIST OF ABBREVIATIONS	7
REPRODUCTION LICENSES	9
INTRODUCTION	13
SCIENTIFIC NOVELTY	18
STATEMENTS OF THE THESIS	20
1. LITERATURE REVIEW.....	21
1.1. Defense systems – why do bacteria need them?	21
1.2. Prevalence and diversity of bacterial defense systems.....	22
1.3. Mechanistic variety of bacterial defense systems.....	24
1.4. Restriction-modification systems.....	28
1.5. Structure and mechanisms of BfiI restriction endonuclease	38
1.6. CRISPR-Cas systems.....	42
1.7. Structure and mechanisms of Type I CRISPR-Cas surveillance complexes.....	47
1.8. Prokaryotic Argonaute proteins	51
1.9. Short prokaryotic Argonaute protein-mediated immunity in bacteria	55
1.10. Structure and biological role of α -synuclein and S100A9 proteins ...	58
1.11. Open-microscopy frameworks for super-resolution imaging	61

1.12. Single-molecule methods for protein-DNA interaction studies.....	64
2. RESULTS.....	69
2.1. Development of the optical fluorescence microscopy system miEye	69
2.2. Characterization of droplets formed by S100A9 and α -synuclein proteins.....	71
2.3. Mechanisms of BfiI restriction endonuclease interaction with DNA	72
2.4. Target search mechanisms of Type I-F Cascade complex in the context of anti-CRISPR proteins.....	75
2.5. SPARDA defense system proteins' mechanisms of action.....	76
3. DISCUSSION.....	79
CONCLUSIONS.....	85
REFERENCES	86

5

SANTRAUKA.....	127
ACKNOWLEDGEMENTS.....	140
LIST OF PUBLICATIONS.....	141
LIST OF CONFERENCES	143
COPIES OF PUBLICATIONS	144
CURRICULUM VITAE.....	149
NOTES	153

LIST OF ABBREVIATIONS

2D two-dimensional
 3D three-dimensional
 Aa *Aggregatibacter actinomycetemcomitans* D7S-1 Abi abortive infection
 Acr anti-CRISPR
 ADPR adenosine 5'-diphosphate-ribose
 AFM atomic force microscopy
 Agos Argonaute proteins
 APAZ analog of PAZ (PIWI-Argonaute-Zwille) ATP
 adenosine triphosphate
 bp base pair
 Cas CRISPR-associated
 CMOS complementary metal-oxide semiconductor CRISPR clustered regularly interspaced short palindromic repeats crRNA CRISPR RNA
 cryo-EM cryo-electron microscopy
 DBD DNA-binding domain
 dCTP deoxycytidine triphosphate
 DREN DNA and RNA effector nuclease
 dsDNA double-stranded DNA
 dSTORM direct stochastic optical reconstruction microscopy
 EaeSPARDA SPARDA complex from *Enhydrobacter aerosaccus*
 eAgos eukaryotic Argonaute proteins
 eGFP enhanced green fluorescent protein
 FM fluorescence microscopy
 FRET Förster resonance energy transfer
 gRNA guide RNA
 HILO highly inclined and laminated optical sheet IR infrared
 LLPS liquid-liquid phase separation
 MGEs mobile genetic elements
 MMF multi-mode fiber
 MTase methyltransferase
 NA nucleic acids
 NAD⁺ nicotinamide adenine dinucleotide (oxidized form) NbaSPARDA
 SPARDA complex from *Novosphingopyxis baekryungensis*
 NPC nuclear pore complex

nt nucleotides
 pAgos prokaryotic Argonaute proteins
 PAM protospacer adjacent motif

PEG polyethylene glycol
 PLD phospholipase D
 pre-crRNA pre-CRISPR RNA
 PSF point spread function
 REase restriction endonuclease
 RM restriction-modification
 RNase ribonuclease
 RRM RNA recognition motif
 SAM S-adenosyl-L-methionine
 SiAgo short prokaryotic Argonaute protein from *Sulfolobus islandicus*
 SIM structured illumination microscopy
 SIR2 silent information regulator 2
 SM single-molecule
 SMF single-mode fiber
 SMFM single-molecule fluorescence microscopy smFRET
 single-molecule FRET
 SMLM single-molecule localization microscopy SPARDA short
 prokaryotic Argonaute, DNase associated SPARSA short
 prokaryotic Argonaute SIR2-APAZ SPARTA short prokaryotic
 Argonaute TIR-APAZ SR super-resolution
 SSB single-stranded DNA-binding
 ssDNA single-stranded DNA
 TA toxin-antitoxin
 TIR Toll-interleukin receptor
 TIRF total internal reflection fluorescence
 TmuSPARDA SPARDA complex from *Thermococcus municipalis*
 tracrRNA *trans*-activating CRISPR RNA
 tRNA transfer RNA
 WT wild-type
 XauSPARDA SPARDA complex from *Xanthobacter autotrophicus* Py2
 α -syn α -synuclein

Figure 3A and Figure 3C reproduced from Sasnauskas, G.; Zakrys, L.; Zaremba, M.; Cosstick, R.; Gaynor, J. W.; Halford, S. E.; Siksnys, V. A Novel Mechanism for the Scission of Double-Stranded DNA: BfiI Cuts Both 3'–5' and 5'–3' Strands by Rotating a Single Active Site. *Nucleic Acids Res.* **2010**, *38* (7), 2399–2410. <https://doi.org/10.1093/nar/gkp1194> under Creative Commons Attribution-NonCommercial 4.0 International License (CC BY NC 4.0).

Figure 3B reproduced from Grazulis, S.; Manakova, E.; Roessle, M.; Bochtler, M.; Tamulaitiene, G.; Huber, R.; Siksnys, V. Structure of the Metal Independent Restriction Enzyme BfiI Reveals Fusion of a Specific DNA Binding Domain with a Nonspecific Nuclease. *Proc. Natl. Acad. Sci.* **2005**, *102* (44), 15797–15802. <https://doi.org/10.1073/pnas.0507949102> under free online availability through the PNAS open access option.

Figure 4 reproduced from Newsom, S.; Parameshwaran, H. P.; Martin, L.; Rajan, R. The CRISPR-Cas Mechanism for Adaptive Immunity and Alternate Bacterial Functions Fuels Diverse Biotechnologies. *Front. Cell. Infect. Microbiol.* **2021**, *10*. <https://doi.org/10.3389/fcimb.2020.619763> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 5 reproduced from Wang, F.; Xu, H.; Zhang, C.; Xue, J.; Li, Z. Target DNA-Induced Filament Formation and Nuclease Activation of SPARDA Complex. *Cell Res.* **2025**, *35* (7), 510–519. <https://doi.org/10.1038/s41422-025-01100-z> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 6A reproduced from Martens, K. J. A.; van Beljouw, S. P. B.; van der Els, S.; Vink, J. N. A.; Baas, S.; Vogelaar, G. A.; Brouns, S. J. J.; van Baarlen, P.; Kleerebezem, M.; Hohlbein, J. Visualisation of dCas9 Target Search in Vivo Using an Open-Microscopy Framework. *Nat. Commun.* **2019**, *10* (1), 3552. <https://doi.org/10.1038/s41467-019-11514-0> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

9

Figure 6B reproduced from Zehrer, A. C.; Martin-Villalba, A.; Diederich, B.; Ewers, H. An Open-Source, High-Resolution, Automated Fluorescence Microscope. *eLife* **2024**, *12*, RP89826. <https://doi.org/10.7554/eLife.89826> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 6C reproduced from Moya Muñoz, G. G.; Brix, O.; Klocke, P.; Harris, P. D.; Luna Piedra, J. R.; Wendler, N. D.; Lerner, E.; Zijlstra, N.; Cordes, T. Single-Molecule Detection and Super-Resolution Imaging with a Portable and

Adaptable 3D-Printed Microscopy Platform (Brick-MIC). *Sci. Adv.* **2024**, *10* (39), eado3427. <https://doi.org/10.1126/sciadv.ado3427> under Creative Commons Attribution-NonCommercial 4.0 International License (CC BY NC 4.0).

Figure 6D reproduced from Valli, J.; Garcia-Burgos, A.; Rooney, L. M.; Vale De Melo E Oliveira, B.; Duncan, R. R.; Rickman, C. Seeing beyond the Limit: A Guide to Choosing the Right Super-Resolution Microscopy Technique. *J. Biol. Chem.* **2021**, *297* (1), 100791. <https://doi.org/10.1016/j.jbc.2021.100791> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 7 reproduced from Kopūstas, A.; Zaremba, M.; Tutkus, M. DNA Flow Stretch Assays for Studies of Protein-DNA Interactions at the Single Molecule Level. *Appl. Nano* **2022**, *3* (1), 16–41. <https://doi.org/10.3390/applnano3010002> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 9 reproduced from Veiveris, D.; Kopustas, A.; Sulskis, D.; Mikalauskaitė, K.; Alsamsam, M. N.; Tutkus, M.; Smirnovas, V.; Ziaunys, M. Heterotypic Droplet Formation by Pro-Inflammatory S100A9 and Neurodegenerative Disease-Related α -Synuclein. *Biomacromolecules* **2025**, *26* (6), 3525–3537. <https://doi.org/10.1021/acs.biomac.5c00130> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 10 reproduced from Ivanovaitė, Š.; Paksaitė, J.; Kopūstas, A.; Karzaitė, G.; Rutkauskas, D.; Silanskas, A.; Sasnauskas, G.; Zaremba, M.; Jones, S. K. Jr.; Tutkus, M. smFRET Detection of Cis and Trans DNA Interactions by the BfiI Restriction Endonuclease. *J. Phys. Chem. B* **2023**, *127* (29), 6470–6478. <https://doi.org/10.1021/acs.jpcc.3c03269> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

10

Figure 11 reproduced from Kupcinskaite, E.; Tutkus, M.; Kopūstas, A.; Ašmontas, S.; Jankunec, M.; Zaremba, M.; Tamulaitiene, G.; Sinkunas, T. Disarming of Type I-F CRISPR-Cas Surveillance Complex by Anti-CRISPR Proteins AcrIF6 and AcrIF9. *Sci. Rep.* **2022**, *12* (1), 15548. <https://doi.org/10.1038/s41598-022-19797-y> under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Figure 12 reproduced from Jurgelaitis, E.; Zagorskaitė, E.; Kopūstas, A.; Asmontas, S.; Manakova, E.; Dalgėdienė, I.; Tylenyte, U.; Silanskas, A.; Toliūsis, P.; Grybauskas, A.; Tutkus, M.; Venclovas, Č.; Zaremba, M.

Activation of the SPARDA Defense System by Filament Assembly Using a Beta-Relay Signaling Mechanism Widespread in Prokaryotic Argonautes. *Cell Res.* **2025**, 35 (12), 1056–1078. <https://doi.org/10.1038/s41422-025-01198-1> under License No. 6364380788194.

Full article Alsamsam, M. N.; Kopūstas, A.; Jurevičiūtė, M.; Tutkus, M. The miEye: Bench-Top Super-Resolution Microscope with Cost-Effective Equipment. *HardwareX* **2022**, 12, e00368. <https://doi.org/10.1016/j.ohx.2022.e00368> reproduced under Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0).

Full article Veiveris, D.; Kopustas, A.; Sulskis, D.; Mikalauskaitė, K.; Alsamsam, M. N.; Tutkus, M.; Smirnovas, V.; Ziaunys, M. Heterotypic Droplet Formation by Pro-Inflammatory S100A9 and Neurodegenerative Disease-Related α -Synuclein. *Biomacromolecules* **2025**, 26 (6), 3525–3537. <https://doi.org/10.1021/acs.biomac.5c00130> reproduced under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Full article Ivanovaitė, Š.; Paksaitė, J.; Kopūstas, A.; Karzaitė, G.; Rutkauskas, D.; Silanskas, A.; Sasnauskas, G.; Zaremba, M.; Jones, S. K. Jr.; Tutkus, M. smFRET Detection of Cis and Trans DNA Interactions by the BfiI Restriction Endonuclease. *J. Phys. Chem. B* **2023**, 127 (29), 6470–6478. <https://doi.org/10.1021/acs.jpcc.3c03269> reproduced under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Full article Kupcinskaite, E.; Tutkus, M.; Kopūstas, A.; Ašmontas, S.; Jankunec, M.; Zaremba, M.; Tamulaitiene, G.; Sinkunas, T. Disarming of Type I-F CRISPR-Cas Surveillance Complex by Anti-CRISPR Proteins AcrIF6 and AcrIF9. *Sci. Rep.* **2022**, 12 (1), 15548.

<https://doi.org/10.1038/s41598-022-19797-y> reproduced under Creative Commons Attribution 4.0 International License (CC BY 4.0).

Full article Jurgelaitis, E.; Zagorskaitė, E.; Kopūstas, A.; Asmontas, S.; Manakova, E.; Dalgėdienė, I.; Tylenyte, U.; Silanskas, A.; Toliūsis, P.; Grybauskas, A.; Tutkus, M.; Venclovas, Č.; Zaremba, M. Activation of the SPARDA Defense System by Filament Assembly Using a Beta-Relay Signaling Mechanism Widespread in Prokaryotic Argonautes. *Cell Res.* **2025**, 35 (12), 1056–1078. <https://doi.org/10.1038/s41422-025-01198-1> reproduced under License No. 6363670353492.

INTRODUCTION

Over billions of years of evolution, bacteria have developed a multitude of systems responsible for cell integrity maintenance. Such systems protect these unicellular organisms from diverse exogenous nucleic acids (NA) – bacteriophage genomes, plasmids and other mobile genetic elements (MGEs) – that typically act as intracellular parasites exploiting the molecular machinery and vital resources of the living cell^{1,2}. When battling these invaders, bacteria rely on innate and adaptive immunities, both involving a plethora of different proteins³. While auxiliary proteins are just as important for carrying out numerous additional functions, which span from regulation to modification, effector proteins or complexes are the ones that gather all the spotlight when either of the bacterial immune pathways is activated. The effectors usually perform their defensive role by directly interacting with NA, which have breached into the cell, and cleaving them². Deeply understanding the molecular mechanisms of such protein-NA interactions is valuable not only from a fundamental standpoint. As previously demonstrated by numerous examples, different proteins, especially nucleases, found in various bacterial defense systems were successfully applied in molecular biology, genome editing, biotechnology, NA diagnostics, biomedicine, and other fields⁴. Therefore, novel mechanistic studies of less-characterized proteins could unravel new candidates with a high intrinsic potential to become even more effective tools in a wide range of applications.

The rapid advance of bioinformatics⁵, sequencing⁶ and protein structure prediction⁷ tools over the past years have radically changed the general understanding of the defensive arsenal possessed by bacterial cells. Just over a decade ago only several bacterial defense systems were known, which, according to the beliefs of the time, seemed to define almost the entire defensive potential of bacteria. At the present time, it more seems that the actual limits of this potential are far from clear, as the number of newly identified candidates with their plausible roles in bacterial immunity has now surpassed two hundreds⁸. Also, the fact that in the bacterial genomes, the genes, which encode proteins involved in bacterial defense systems, tend to form separate clusters, the so-called “defense islands”, further increases the overall complexity of organization inherent to these systems⁹.

Within this rich and yet to be explored context, the research focus is unavoidably shifting to the completely unknown candidates identified as potential role-players in bacterial protection. For instance, the number of studies on restriction-modification (RM) systems, which nowadays can be considered as a classical example of bacterial defense, is slowly decreasing

each year. It is likely that soon the same tendency will also be visible in the study number of newer but already rather well-characterized bacterial defense systems, such as clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) systems or the ones based on prokaryotic Argonaute proteins (pAgos). Despite this, given the enormous contribution of these systems to the various fields of science, medicine and industry over the years, the research of their key components remains relevant to this day. Since the diversity of RM, CRISPR-Cas and pAgo-based systems is extremely broad, all of them still encompass proteins that are incompletely characterized and possibly could be of a particular use in certain practical applications¹⁰.

When characterizing NA-binding proteins, the most difficult challenge usually is to determine the precise molecular mechanisms governing the dynamic interactions between the protein of interest and NA. This typically requires rather powerful methods distinguished by such features, as high spatiotemporal resolution, high sensitivity, precision at the molecular level or high-throughput, all of them being indispensable in this kind of research¹¹.

Mechanistic studies of protein-NA interactions commonly rely on biochemical (e.g., chromatin immunoprecipitation followed by sequencing (ChIP-seq)¹² or electrophoretic mobility shift assay (EMSA)¹³), functional (e.g., site-directed mutagenesis¹⁴) and structural (e.g., cryo-electron microscopy (cryo-EM) and its derivatives¹⁵) techniques¹⁶. Despite being so widespread, these methods fall short in the following aspects – they either measure only the average values of the chosen molecule-ensemble properties or capture only the snapshots of biomolecules at fixed time points¹⁷. Because of these two major drawbacks, the valuable information about the possible heterogeneity of molecules and their dynamic interplay in the sample of interest is concealed or even lost completely¹⁸.

In contrast to structural or bulk measurements, single-molecule (SM) techniques offer several main advantages: (1) they enable direct visualization of individual protein-NA complexes, (2) allow one to continuously monitor the interaction dynamics between these biomolecules in real time, (3) are able to detect intermediate, rare or short-lived unique states characteristic of these complexes and, thus, (4) help one to reveal the heterogeneity of the observed molecule populations present in the sample of interest¹⁹. There are multiple methods dedicated for studying protein-NA interactions at the SM level: magnetic tweezers, optical tweezers, tethered particle motion, atomic force microscopy (AFM), SM Förster resonance energy transfer (smFRET), different types of DNA flow-stretch assays²⁰. Some of these techniques, like

AFM, are entirely standalone and can operate autonomously, while the others,

for instance, smFRET or DNA flow-stretch assays, usually work in tandem with the already existing base methods.

One of such powerful methods is SM fluorescence microscopy (SMFM)²¹. Over the years, it has been successfully applied in numerous protein-NA interaction-related SM-level studies, and by now can be considered as a golden standard in this research topic²². Nevertheless, some bacterial defensomes can contain cell protection systems which are so intricate that interdisciplinary studies are often required in order to unravel the full mechanistic picture of these systems' functioning pathways²³. To execute their role efficiently, ensure strict regulation and minimize the risk of evasion, bacterial defense systems tend to be multicomponent and typically involve one effector protein or multisubunit effector complex and several other auxiliary proteins²⁴. Multicolor fluorescence imaging is then needed to visualize and observe more than one fluorescently labeled distinct biomolecule of interest, e.g., effector, regulator and NA, during the same experiment²⁵. SMFM-based studies of protein-NA interactions performed *in vitro* involve purified biomolecules, which are placed in a predefined environment of limited complexity. Compared to such an artificial system, intracellular environment of bacterium is far more complex and dynamic. Here, the proteins of interest are prone to participate in transient non-specific interactions with a bacterial chromosome or other large molecules, and they are also exposed to various molecular effects, like crowding, sieving or oligomerization²⁶. It is highly likely that conducting SM studies in living cells rather than *in vitro* reflects a more accurate picture of mechanisms that bacterial defense system proteins use to interact with cognate NA.

Since SMFM is a diffraction-limited technique, it alone is often insufficient to effectively investigate the interplay between the selected protein and NA *in vivo*. Two main reasons for this are: (1) many protein-DNA complexes and their formed structures of a higher order are of sub-diffractive dimensions, and (2) there are usually plenty of copies of the same fluorescently labeled ever-moving protein of interest in a relatively small space. To resolve such supramolecular assemblies or track the diffusion of individual proteins in real time, advanced live-cell fluorescence imaging techniques are needed. In many cases, diverse super-resolution (SR) microscopy approaches, such as SM localization microscopy (SMLM)²⁷, MINFLUX²⁸ or super-resolution optical fluctuation imaging (SOFI)²⁹, perfectly fulfill this requirement. Finally, any taken cell resembles a confined volumetric container, in which the molecules can move and reside along all three axes in space. This calls for three-dimensional (3D) SM fluorescence imaging, as any two-dimensional (2D) measurements greatly compromise the

underlying molecular mechanisms of protein-NA interactions effectively, one often needs to use or even combine several different SM methods and implement various sophisticated experimental-design schemes. To meet the above listed criteria, a versatile high-resolution widefield fluorescence microscope is necessary for maximizing the information output of SM-level mechanistic studies on protein-NA interactions. Such optical microscopy system should exhibit high sensitivity to grant SM detection and extreme stability to ensure proper measurement quality over an extended duration. More importantly, it should be modular enough to fulfill the complex design schemes of the planned experiments, as well as to be able to image a wide variety of sample types. The modularity aspect should be present in all major parts of the microscope – laser sources, excitation and emission paths – and implemented via interchangeability, adaptability and expandability of these fundamental units. Ideally, this microscopy framework should incorporate the following capabilities: (1) multicolor imaging or at least two color simultaneous imaging, (2) imaging in several different modes of excitation, (3) live-cell imaging, (4) SR imaging, (5) 3D imaging modality. The appeal of this bespoke system is further increased if such microscope is also cost effective and provides a robust and user-centric software package for controlling the hardware and both acquiring and analyzing the collected data in one unified environment. However, neither the commercial market selling various optical fluorescence imaging equipment, nor the scientific community specializing in developing distinct home-built open-source fluorescence microscopes can currently offer such a multipurpose high-resolution fluorescence microscopy (FM) platform³¹. The lack of such needed cutting edge technology halts the potential breakthrough in both novel and classical bacterial defense systems' research, since there is still a considerable knowledge gap in the complete understanding of distinct effector protein-NA interaction mechanisms on a molecular basis.

The aim of this dissertation is to, by means of SMFM *in vitro*, elucidate the molecular mechanisms governing the protein-NA interactions in fundamentally different bacterial innate and adaptive immunity systems. To achieve this aim, **four main objectives were set:**

1. Build, characterize and, using different types of biological samples, validate high-resolution widefield fluorescence optical microscopy system miEye.
2. Employing smFRET technique, determine the mechanism by which catalytically inactive K107A mutant of *Bacillus firmus* Type IIS

3. Based on Soft DNA Curtains platform, assess the anti-CRISPR (Acr) protein AcrIF9-induced changes in target search mechanism of Type I-F CRISPR-Cas system Cascade complex from *Aggregatibacter actinomycetemcomitans* D7S-1.
4. Using DNA flow-stretch assay, reveal the mechanism of action of short pAgo SPARDA found in Py2 strain of *Xanthobacter autotrophicus*.

improved iteration of the open-source miCube microscope³². The miEye microscopy system combines fully modular, stable, and cost-effective hardware with Python-based open-source software entitled microEye, which we have written in parallel with the development of our microscopy system. The result of merging these separate components into a single, flexible, yet robust framework – versatile SR microscope miEye. It offers SM fluorescence imaging and SMLM capabilities for diverse sample types, and ensures a convenient user experience by integrating microscope control, image acquisition and data analysis modules within a unified environment of microEye³³. Compared to similar SMFM and SMLM systems developed in house by other research groups to this date^{34–36}, miEye is among the first ones to be an entirely customizable open-source standalone SR imaging platform with an affordable equipment, dedicated software package and in-depth technical documentation.

Using the newly developed miEye microscope for simultaneous two color fluorescence imaging, we show that neurodegenerative disease associated α -synuclein³⁷ (α -syn) and the pro-inflammatory S100A9³⁸ are able to form heterotypic droplet-like condensates under conditions that promote liquid-liquid phase separation (LLPS). This is a new finding in the context of earlier research conducted on these proteins. It is also the first demonstration that our custom-built optical fluorescence microscope miEye can be effectively applied for the qualitative imaging of biological samples exhibiting a high level of heterogeneity, as well as for performing quantitative two-color measurements *in vitro*.

In this dissertation, we employed few different SM-level techniques to reveal the so far unknown mechanisms by which three separate effector proteins, each belonging to a bacterial defense system of entirely different nature, interact with NA. In the case of RM systems, the biological object of study was Type IIS homodimeric REase BfiI from *Bacillus firmus*. Previous biochemical ensemble-level studies provided detailed insights into a unique double-stranded DNA (dsDNA) cleavage mechanism used by this restriction enzyme^{39–41}. The presence of two DNA-binding domains (DBD) allows BfiI to bind to its two DNA targets either *in cis* or *in trans* manner. To date, the direct demonstration of BfiI-DNA *in cis* and *in trans* interactions is still lacking. Here we report SM-level characterization of these two distinct mechanistic modes exploited by BfiI during its interplay with DNA. We also provide the first detailed characterization of conformational dynamics of the

nucleolytically inactive BfiI K107A mutant during its interaction with the cognate DNA, as well as new insights into target search mechanisms characteristic to this REase. Furthermore, in this study, we established a novel assay based on the smFRET technique. The main advantage of this new

approach is that it is universal and can be used to investigate *trans* interactions of other similar proteins, which are capable of simultaneously binding two targets, with DNA.

The representative for studying CRISPR-Cas systems' protein interactions with NA was Type I-F Cascade surveillance complex from *Aggregatibacter actinomycetemcomitans* D7S-1 (Aa). The mechanisms of unperturbed target search carried out by some other Cascade complexes of Type I CRISPR-Cas systems have been already explored at the SM level previously⁴²⁻⁴⁵. However, there are currently no *in vitro* mechanistic SM studies in which the target search of Type I CRISPR-Cas surveillance complexes has been examined in the context of the corresponding class Acr proteins. For the first time, here we show that counter-defense protein AcrIF9 inhibits the target search characteristic of Aa-Cascade complex.

The SPARDA (short prokaryotic Argonaute, DNase associated) complex, found in the mesophilic *Xanthobacter autotrophicus* strain Py2 (XauSPARDA), stood as a member of bacterial defense systems based on short pAgos. Here, we report a full molecular mechanism of action of XauSPARDA. It is in a good agreement with the mechanistic model of action of another SPARDA complex from *Novosphingopyxis baekryungensis* (NbaSPARDA)¹⁶. In the case of our study, for the first time, we performed a direct real-time visualization of the continuous process of activated XauSPARDA complexes' filamentation on individual long DNA molecules. In addition, we qualitatively and quantitatively assessed the distinct phenotypic changes experienced by DNA interacting with the activated XauSPARDA complexes. Along with XauSPARDA, we also characterized three other homologous SPARDA proteins in a similar manner. All of this brings novel mechanistic insights into the extensive topic of pAgo-based immune response of bacteria.

STATEMENTS OF THE THESIS

1. Open-source modular widefield FM system miEye is suitable for performing two-color fluorescence imaging, SMFM and SR imaging of different biological sample types *in vitro*. (**Publications 1 and 2**)

2. Instantaneous closed-to-open state transition is a characteristic *Bacillus firmus* REase BfiI conformational behavior occurring upon its cognate DNA binding both *in cis* and *in trans*. (**Publication 3**)
3. Acr protein AcrIF9 hinders the effective target search of *Aggregatibacter actinomycetemcomitans* Type I-F CRISPR-Cas surveillance complex. (**Publication 4**)
4. Short pAgo SPARDA from *Xanthobacter autotrophicus* operates via filamentation on long dsDNA molecules by effectively compacting and cleaving them afterwards. (**Publication 5**)

1. LITERATURE REVIEW

1.1. Defense systems – why do bacteria need them?

Bacteria are among the oldest organisms on Earth⁴⁶. In virtually any

ecosystem, they face various biotic threats, such as other competing bacteria, bacterial and eukaryotic predators, as well as bacteriophages, which all have shaped the physiology and evolution of bacteria for billions of years¹. Bacteriophages are considered to be the most abundant biological entities on Earth – it is estimated that the number of their particles present in the biosphere reaches 10^{31} ⁴⁷. These bacteria-infecting viruses hijack the host cell's machinery for self replication and further propagation. During adsorption, phages specifically attach to the bacterial host and inject their genome into it⁴⁸. Once inside the cell, phages can undergo two main types of lifecycle, depending on various factors⁴⁹. The lysogenic cycle involves integration of the phage genome into the bacterial host genome⁵⁰. Here, a prophage forms and spreads across all future generations of such bacterium by means of inheritance. In contrast, the lytic cycle involves three major stages: (1) replication and expression of the phage genome, (2) assembly of the new virions and (3) release of the phage progeny typically via endolysin and holin-induced lysis of the host cell^{51,52}. Besides temperate (lysogenic) phages and lytic phages, there are also other classes of phages that use lifecycles different from these common two⁴⁹. For example, chronic (filamentous) phages undergo a slow replicative cycle during which new phages are continuously released from the host usually without killing it, hereby allowing the bacterial cells to survive the infection⁵³.

Bacteriophages have evolved alongside bacteria for nearly 4 billion years⁵¹. The everlasting coexistence of these two ultimate enemies decidesthat phage-resistant hosts protect the entire generations of bacteria, while phages tirelessly continue to pose a threat to such bacterial strains. This warfare is mostly driven by phages constantly developing novel strategies to evade the numerous and diverse antiviral defense systems operating in bacteria⁵⁴. Phages play a crucial role in regulating bacterial populations across different environments on Earth⁵⁵. Moreover, phages contribute to 20–40% of daily mortality of bacteria⁵⁶ and have a significant impact on biogeochemical cycles⁵⁷.

Besides bacteriophages, MGEs also include plasmids, transposons, integrative conjugative elements and genetic material that migrates by means of transformation, transduction and conjugation². Interactions between these major MGEs and their hosts can be either mutualistic or parasitic. In exchange

for maintaining a particular MGE, the host cell receives various benefits, such as antibiotic resistance and virulence^{58,59}, genomic rearrangement enhancement, which stimulates evolutionary innovation⁶⁰, and facilitation of the horizontal gene transfer, which is responsible for the inexhaustible plasticity of prokaryotic genomes by constantly shuffling the gene pool for selection^{61,62}. On the contrary, the host cell is exploited for propagation⁶³,

conjugation⁶⁴ and selfish maintenance through post-segregational killing⁶⁵ by, for example, plasmids or temperate phages, which overall is costly in cellular resources². In addition, MGEs carrying potentially advantageous traits can be outcompeted by the “initial” MGEs that were already present in the cell⁶⁶. Phage “superinfection exclusion”⁶⁷ and other incompatibility or competition mechanisms^{68,69} usually govern such a parasitic behavior of certain MGEs.

All the threats described above primarily target the essential cellular components, fundamental processes and vital functions of bacteria. This, for instance, can include: (1) damage to the genomic DNA, cell membrane or peptidoglycan wall, (2) arrest of DNA gyrase, RNA synthetase or ribosome activities and (3) disruption of transcription or NA synthesis¹. Being constantly exposed to different kinds of attackers lurking in their natural habitats, bacteria, over time, have developed a rich arsenal of self-defense mechanisms, which can operate on various scales^{1,10}. Defense at the molecular level encompass modification of the vulnerable target or destruction of the harmful agent⁷⁰. Cellular-level protection involve bacterial cell motility, efflux pumps, secreted membrane vesicles and diverse macromolecular barriers, such as the cell membrane, S-layer or lipopolysaccharides⁵⁵. At the multicellular level, dense cell clusters, resistant subpopulations and various collective barriers, e.g., biofilms or extracellular polymeric substances, keep bacteria safe from the predatory environmental hazards¹.

1.2.Prevalence and diversity of bacterial defense systems

Due to the ubiquitous abundance of phages and other lethal threats, bacteria are always under evolutionary pressure to acquire novel defense systems capable of rapidly halting the spread of virus infection, as well as effectively neutralizing any other type of intrusion⁷¹. Recent advancements in genome sequencing, bioinformatics and experimental analysis have led to the discovery of more than 200 new defense systems, most of which confer protection against bacteriophages^{72–75}. Multiple bacterial defense systems of distinct nature tend to colocalize in special genomic regions known as defense islands^{9,72,74}. Yet, the exact reasons for this clustering remain unclear. In many cases, defense islands associate with phages, phage satellites, integrative

conjugative elements, plasmids and transposons⁹. This suggests that genomic clustering may facilitate the horizontal transfer of defense systems. Synergy between different defense systems or the need for coordinated regulation could also be the driving forces for the origination of defensive islands^{2,8,9}. Defense islands are enriched with putative operons and contain abundant families of widespread genes. Detailed analysis of the sequences of proteins encoded by these genes has shown that most of them are diverged variants of

components present in the already known defense systems. The remaining genes exhibit features like characteristic operon organization, which may be an indication of entirely new means of bacterial protection⁹.

Depending on the system type, there is a large variation in the abundance of defense systems found in bacteria. RefSeq database analysis based on DefenseFinder⁸, a tool designed to systematically detect antiviral defense systems in prokaryotic genomes, has been performed recently^{3,8}. The results revealed that 84% and 40% of bacterial genomes contain RM and CRISPR Cas systems, respectively, whereas more than half of the other detected systems are encoded in <3% of bacterial genomes. Several systems can cooperate to establish a single defensive arsenal that is characteristic of a specific bacterial strain. Of the 21,364 fully sequenced bacterial genomes, 78% encode more than two defense systems (Fig. 1A) with the apparent important variations between distinct strains^{3,8}. For instance, some bacterial species encode over 50 different defense systems (e.g., 57 systems are found in the genome of *Desulfonema limicola*), while the ones belonging to the Chlamydiae phylum carry few-to-none defense systems^{8,74} (Fig. 1B). It is more than likely that antiviral immunity is highly conserved across both prokaryotic kingdoms, as a substantial amount of bacteria-encoded defense systems can also be found in archaea³. After detecting viral infection, bacterial defense systems engage in the phage reproduction cycle termination by mostly targeting NA, single nucleotides (nt), membrane of the cell and other susceptible elements³. In bacterial genomes, NA-degrading effectors are extremely widespread, whereas other types of antiphage protection systems are less common⁸(Fig. 1C).

In respect to its nature, prokaryotic immunity can be innate, which is relatively non-specific, or adaptive, exhibiting a high specificity⁸⁰. Despite their extreme variety, RM modules are currently one of the best-characterized prokaryotic innate immune systems^{81,82}. Other examples of innate immunity in bacteria and archaea include bacteriophage exclusion^{83,84} and pAgos^{85,86}. In contrast, rich and diverse CRISPR-Cas systems perfectly illustrate the concept of adaptive immunity in prokaryotes^{87,87,88}. The final category includes such defense systems as the numerous and highly diverse toxin-antitoxin (TA) systems^{70,89,90} and functionally similar abortive infection (Abi) systems⁹¹.

Another way of categorizing a plethora of mechanistically diverse antiphage protection-ensuring strategies can be based on the specific viral infection cues that are sensed by the different bacterial defense systems. Certain defense systems detect invading NA, when a bacteriophage infects the host cell³(Fig. 2A). In the majority RM and RM-like systems, specific host DNA motifs are methylated⁹² or modified with phosphorothioate^{93–95} to prevent host genome cleavage catalyzed by the NA-degrading effectors, which, in turn, destroy the unmodified foreign DNA. The CRISPR-Cas machinery processes the invading NA and inserts their fragments into the CRISPR array as a part of the host genome⁹⁶. After CRISPR array is expressed and processed, mature guide RNA (gRNA) is produced. The CRISPR-Cas machinery then uses these gRNAs for the recognition of exogenous NA, which, if successful, triggers an immune response. In Type I, II and V CRISPR-Cas systems, the respective effectors are activated directly⁹⁷, whereas in Type III systems, signaling cyclic oligoadenylate molecules, which are produced during the target recognition, act as mediators essential for the effector activation⁷¹. NA fragments serve as guides for pAgos in search of complementary NA. Target recognition and binding by pAgos result in the activation of various effector domains or proteins that lead to cell death due to their cytotoxicity^{98,99}.

Alternatively, some bacterial defense systems directly sense the host cell machinery-produced phage proteins³(Fig. 2B). Before the infecting phage completes its cycle, in some cases, premature cell lysis can occur because of AbiZ protein acting in tandem with the phage-encoded lysins or holins¹⁰⁰. The RexAB system is comprised of two main components – RexA protein responsible for sensing the phage infection and the toxic effector RexB. The latter is an ion channel which can be activated by the RexA protein upon bacteriophage entering the cell, thus triggering cell death¹⁰¹. Serine-threonine kinase Stk2 can be activated by the phage protein PacK to initiate a phosphorylation cascade leading to cell death¹⁰². In the CapRel TA system,

the toxin domain persists in the inhibitory state until the antitoxin domain binds to the capsid protein of the attacking phage. Upon activation, the toxin

domain phosphorylates transfer RNA and promotes translational arrest within the cell¹⁰³. DSR2 proteins possess the nicotinamide adenine dinucleotide (NAD⁺)-degrading activity which is triggered by the phage tail tube protein. The resultant NAD⁺ depletion eventually leads the cell to Abi¹⁰⁴. The phage encoded protein Ocr acts as an inhibitor of RM and bacteriophage exclusion defense systems. Nonetheless, phage anti-restriction-induced system (PARIS) is capable of sensing this Ocr protein and mediates the progression of Abi in such case¹⁰⁵. Sensor domain of Avs systems enables the recognition of phage proteins' conserved structural features through a direct binding. Then, the activation of various types of effector domains occurs¹⁰⁶.

There are quite a few of defense systems that monitor the integrity of the bacterial cell machinery and detect the changes in the host cell, when its machinery is hijacked by the invading phage for the selfish production of various viral components³(Fig. 2C). In the unaffected host cell, PrrC protein is bound to a particular restriction enzyme and, therefore, is unable to perform its function. During phage infection, PrrC is released from this inhibitory state after the phage-encoded anti-restriction protein Stp binds to the same restriction enzyme. The activated PrrC cleaves the cellular tRNA, thus disrupting both the protein synthesis and the phage infection cycle^{107,108}. The reserves of the labile RNA antitoxin of Type III TA system ToxIN are depleted upon phage-induced transcriptional repression in the bacterial host. Such system's toxic effector then becomes functional and non-specifically cleaves the cellular RNA, hereby facilitating Abi of the compromised host¹⁰⁹. Type II RnlAB system consists of the toxin RnlA and unstable antitoxin RnlB. Phage infection causes degradation of RnlB and, in turn, the system's non specific RNase, RnlA, is activated¹¹⁰. Retron Ec48 can be activated due to inhibition of the bacterial RecBCD complex in the phage-infected cells¹¹¹. Such sensing of phage infection results in the recruitment of toxic effectors that mediate the Abi process. The same downstream cellular response occurs when the retrotransposon Sen2 is activated during phage invasion. In this case, a modification of the synthesis of Sen2 multicopy single-stranded DNA (ssDNA) is responsible for the activation of this retrotransposon¹¹². Under equilibrium conditions, RNA antitoxin inhibits the deoxycytidine triphosphate (dCTP) deaminase which is the toxic effector operating in the AvcID TA system. The ongoing phage infection terminates the synthesis of the labile RNA antitoxin, re-enabling the activity of dCTP deaminase. While active, this enzymatic toxin converts dCTP into deoxyuridine monophosphate and, over time, depletes the nucleotide reserves of the host cell¹¹³. Phage-triggered

triphosphate to deoxyguanosine. This, as well, results in the total depletion of nucleotide reserves of bacterium¹¹⁴.

The remaining group of bacterial defense systems sense viral infection through other mechanisms that considerably differ from each other (Fig. 2D). In the pyrimidine cyclase system for antiphage resistance (PYCSAR), oligonucleotide cyclases start to produce signaling molecules when the host cell faces phage infection. Various immune effectors are then activated by these molecules, thus facilitating the progression of Abi response¹¹⁵. Cyclic oligonucleotide-based antiphage signaling system (CBASS) operates in an analogous manner when dealing with bacteria-infecting viruses¹¹⁶. Two proteins, ThsA with a characteristic NAD⁺-degrading activity and Toll interleukin receptor (TIR) domain-containing ThsB, make up the defensive core of Thoeris system. The sensing of the phage infection is performed by ThsB, which initiates the production of signaling molecules if the host cell is under attack. Signal-activated ThsA depletes the cellular NAD⁺ reserves and forces the bacterium to undergo Abi¹¹⁷. In the DarTG TA system, the phage activated toxin DarT inhibits viral replication by ribosylating adenosine diphosphate of the phage DNA¹¹⁸. During phage infection, the associated caspase-like proteases cleave the inhibitory C-terminal domain of bacterial gasdermin (bGSDM) system proteins, thus releasing them from the auto inhibition state. This induces the formation of large pores in the cell membrane and ultimately results in cell death¹¹⁹. Prokaryotic Viperin proteins convert nucleotide triphosphates into their 3'-deoxy-3',4'-didehydro analogs, which inhibit virus transcription in presence of phage invasion. It is likely that, when incorporated into the nascent chains of viral RNA, these modified nucleotides terminate further chain synthesis¹²⁰. Members of *Streptomyces* genus produce secondary metabolites, such as DNA intercalators doxorubicin and daunorubicin¹²¹, as well as certain antibiotics of aminoglycoside class¹²², which exhibit antiphage activity.

cell genome, whereas the methylation pattern-lacking foreign DNA undergoes endonucleolytic cleavage². RM systems exhibit large diversity¹²³ and are found in more than 95% of sequenced bacterial and archaeal genomes¹²⁴. Due to high biotechnological value of restriction enzymes altogether, most RM systems have been already studied and fully characterized over the years⁹².

MTases catalyze the transfer of methyl group from S-adenosyl-L methionine (SAM) molecule to the eligible heterocyclic base in DNA¹²⁵. Adenine and cytosine are the two targets for such chemical modification to occur, converting these bases into 6-methyladenosine, 4-methylcytosine or 5-methylcytosine, respectively¹²⁶. Out of these three, 6-methyladenosine is the most common methylation modification appearing in 75% of all identified methylation cases in prokaryotes¹²⁷. DNA methylation, in general, is a widespread feature of prokaryotic genomes – it is detected in more than 90% of studied species. Besides methylation, in some prokaryotic cells, the defense against exogenous NA can be realized via other DNA modifications². For example, 7-deazapurine-introducing modification in the host chromosome prevents transformation with unmodified plasmids¹²⁸. Alternatively, DNA phosphoribose backbone can be site-specifically modified with phosphorothioate. In this case, the non-bridging oxygen is replaced with sulfur¹²⁹. To incorporate sulfur into the DNA molecule, these unique systems, such as DNA degradation (Dnd) system or ssDNA phosphorothioate (Ssp) modification system, typically use different pathways^{95,130}.

RM systems are classified into four types (I-IV) depending on subunit composition and such biochemical characteristics as protein structure, recognition of restriction sites, cofactor requirements and substrate specificity¹³¹ (Table 1). Based on the evolutionary relatedness of system's components, Type I systems are further divided into five families: A, B, C, D and E¹³². The basis for finer classification of Type II systems is different and relies on the particular enzymatic behavior of REases. Currently, there are eleven subtypes of Type II RM systems: A, B, C, E, F, G, H, M, P, S and T¹³³. This nomenclature is not absolute, since a single enzyme may exhibit several defining properties characteristic of distinct groups. One of the most extreme examples is BcgI protein that belongs to at least four different subtypes of Type II REases (A, B, C and H)¹³⁴.

Over time, phages developed various anti-restriction strategies to bypass the defensive barriers of orthodox RM systems⁵⁴. For instance, instead of a standard cytosine base, the genome of bacteriophage T4 contains α and β glucosylated 5-hydroxymethylcytosines, which confer resistance to any

restriction enzyme responsible for non-methylated or even methylated DNA degradation in the host cell¹³⁵. To ensure the host cell protection against this

modified DNA of foreign nature, some bacterial RM systems have evolved into such unique types as Type IIM and Type IV, which specifically target and cleave DNA molecules bearing methyl group only or methyl group additionally modified with other moieties, respectively¹³⁶.

Type	Characteristic features*	Examp
Type I	• Pentameric complex (2R2MS) acts as REase. • Mg^{2+} ions, SAM and adenosine triphosphate (ATP) are required for cleavage of non-methylated DNA. • Recognition sequence is bipartite and asymmetric. • Cleavage is variable and typically occurs far from the recognition site. • “DEAD-box“ REase – DNA cleavage is preceded by ATP-dependent DNA translocation. • Functional MTase can be of heterodimeric (MS) or heterotrimeric (2MS) form.	EcoKI, CfrAI, StySPI
Type IIP (Palindromic recognition sequence)	• REase is a homodimer (2R). • Only Mg^{2+} ions are required for non-methylated DNA cleavage. • <u>Recognition sequence is palindromic.</u> • Cleavage is defined and occurs within or adjacent to the recognition site. • The cut DNA product may have 3' overhangs, 5' overhangs or blunt ends. • MTase is a monomer (M).	EcoRI, BamHI

31

Type	Characteristic features*	Examp
------	--------------------------	-------

Type IIA (Asymmetric recognition sequence)	• <u>Recognition sequence is asymmetric.</u> • RM systems of this subtype typically consist of one REase and two MTases. • Two methylation enzymes are needed to modify each strand of the asymmetric recognition site.	FokI, AclI, HphI
Type IIB (Both side cleavage)	• <u>Cleavage occurs on both sides of the recognition sequence.</u> • The two sides of target are cleaved on both DNA strands.	BcgI, Bst44I, BsaXI
Type IIC (Combined REase and MTase)	• Recognition sequence can be either symmetric or asymmetric. • <u>Enzymes of this subtype exhibit a hybrid structure – a single polypeptide is responsible both for REase and MTase activities.</u>	GsuI, HaeIV, BcgI
Type IIE (allosteric Effector domains)	• REase can be of monomeric (R) or homodimeric (2R) form. • Mg^{2+} ions and two targets are required for cleavage of non-methylated DNA. • Recognition sequence can be either palindromic or non palindromic.	NaeI, AtuBI, Sau3AI

32

Type	Characteristic features*	Example
	• Cleavage is defined and occurs within the recognition sequence or at a short distance from it. • <u>The second recognition site serves as an allosteric effector by binding to REase <i>in cis</i> or <i>in trans</i>.</u>	

	• MTase is a monomer (M).	
Type IIF (Four DNA strand hydrolysis)	• REase is a homotetramer (4R). • Mg^{2+} ions and two recognition sites are required for non methylated DNA cleavage. • Recognition sequence can be either full or interrupted palindrome. • <u>Cleavage occurs in a concerted reaction within both targets</u>, is defined and may result in 3' or 5' overhangs. • MTase is a monomer (M).	NgoMIV SfiI, Aa
Type IIG (Gamma class DNA methylati on domain)	• <u>DNA-cleavage domain (REase) and gamma-class DNA-methylation domain (MTase) are located on the same polypeptide chain (RM).</u> • Mg^{2+} ions and SAM cofactor are required for cleavage of non-methylated DNA. • Recognition sequence can be either symmetric or asymmetric.	Eco57I, MmeI, B

33

Type	Characteristic features*	Examp
	• Cleavage is defined and occurs at a short distance from the recognition site. • Systems of this subtype also contain a separate monomeric MTase, since the coupled RM polypeptide only methylates one DNA strand.	

Type IIH (Hybrid)	• Recognition sequence can be either symmetric or asymmetric. • <u>Overall gene structure, as well as MTase organization, are similar to Type I RM systems.</u> • MTase subunits M and S are separate and usually act as a tetrameric complex (2M2S).	AhdI, B PshAI
Type IIM (Methylated recognition sequence)	• REase is a homodimer (2R). • Only Mg²⁺ ions are required as a cofactor for DNA cleavage. • <u>For DNA cleavage to occur, recognition site must be methylated.</u> • Recognition sequence can be either palindromic or non palindromic. • MTase is absent in this subtype of Type II RM systems.	DpnI, G

34

Type	Characteristic features*	Examp
Type IIS (Shifted cleavage)	• REase may be of monomeric (R) or homodimeric (2R) form. • Only Mg²⁺ ions are required for the cleavage of non methylated DNA. • In some exceptions, DNA cleavage is completely metal ion independent. • Recognition sequence is asymmetric. • <u>Cleavage occurs at fixed positions usually outside of the recognition site and is shifted to one side of the sequence.</u>	BfiI, Mr SfaNI

	• RM systems of this subtype can contain one or two separate MTases, all of them as monomers (M).	
Type IIT (heterodimer composed of Two different subunits)	• <u>REase can be of heterodimeric ($\alpha\beta$) or heterotetrameric ($2\alpha 2\beta$) form.</u> • Only Mg^{2+} ions are required for non-methylated DNA cleavage. • Recognition sequence may be either symmetric or asymmetric. • Cleavage is defined, occurs within the recognition site or at a short distance from it and results in 3' overhangs.	Bpu10I,

35

Type	Characteristic features*	Examp
	• RM systems of this subtype can contain one or two separate MTases, all of them as monomers (M).	

Type III	• Combined complex (R2M or 2R2M) acts as REase. • Mg^{2+} ions, ATP and two targets located in an inverse orientation are required for the cleavage of non methylated DNA. • Recognition sequence is non-palindromic. • Cleavage occurs at a fixed position – approximately 25 bp away from one of the two recognition sites. • “DEAD-box“ REase – DNA cleavage is preceded by ATP-dependent DNA translocation. • MTase can function independently of REase (2M or 2M2R), only methylates adenines on one strand of DNA, i.e., only performs hemimethylation.	EcoP15I EcoP1II, StyLTI
Type IV	• REase is dependent on DNA methylation and cuts only modified DNA, containing methylated, hydroxymethylated or glucosyl-hydroxymethylated bases. • In some cases, e.g., in <i>E. coli</i> McrBC restriction system, guanosine triphosphate is required for DNA cleavage.	McrBC, M

36

Type	Characteristic features*	Examp
	• Cleavage is variable and occurs outside of the recognition sequence. • These RM systems exhibit a rather large variety in different types and proteins, which are usually unpaired and unrelated.	

*For each different subtype of Type II restriction-modification systems, the group defining feature is underlined and marked in bold. Notation meaning: R – restriction subunit, M – modification subunit, S – specificity subunit.

**The up and down arrows indicate the exact DNA cleavage positions in the given example recognition sequences. Notation meaning: R – purine, Y – pyrimidine.

1.5. Structure and mechanisms of BfiI restriction endonuclease

BfiI protein, isolated from *Bacillus firmus* S8120 strain, belongs to Type IIS REases and binds asymmetric hexanucleotide recognition sequence 5'-ACTGGG-3'. This restriction enzyme catalyzes DNA cleavage 5 (on the top strand) and 4 (on the bottom strand) nucleotides downstream of the recognition site, thus producing a single base 3' overhang¹⁴⁰. Type II REases generally require Mg^{2+} ions as an essential cofactor for cutting the bound DNA molecule. Other divalent cations, such as Mn^{2+} , Co^{2+} , Zn^{2+} , Ni^{2+} or Cu^{2+} , can serve as a replacement for Mg^{2+} ions in the active sites of some REases¹⁴¹. Certain restriction enzymes, like PacI or NotI, are fully operational only in the presence of Zn^{2+} or Fe^{2+} ions, respectively^{142,143}. In most cases, DNA hydrolysis is completely inhibited by Ca^{2+} ions¹⁴⁴. Given the wide variety of divalent cations that different Type II REases use for catalysis, BfiI is a clear exception amongst the other members of this family. To perform DNA cleavage, neither Mg^{2+} ions, nor any other divalent cations are necessary to BfiI REase¹⁴⁵. Such a complete independence from metal ions as core catalytic components is characteristic of only few other restriction enzymes, e.g., BmrI¹⁴⁶ and PabI¹⁴⁷.

The N-terminal domain of BfiI shows a high degree of structural similarity to the *Salmonella typhimurium* non-specific endonuclease Nuc, which belongs to the PLD superfamily¹⁴⁸. All enzymes of the PLD superfamily contain two copies of HXX, a conserved sequence motif, and use a two-step metal-independent mechanism to hydrolyze the phosphodiester bonds in the bound DNA molecule^{149–151}. The His residues of the two HXX motifs are crucial for the catalytic activity of PLD enzymes^{152,153}. Both

monomers of BfiI REase dimer possess a single HXX motif. These two copies of HXX motif contribute to the single active site of BfiI situated at the subunit interface. There are two identical 2-fold symmetry-related histidines, one from each subunit, present in the catalytic center of this restriction enzyme^{39,148}. In the proposed mechanism for BfiI-catalyzed DNA cleavage³⁹ (Fig. 3A), the first reaction step involves attack on the scissile phosphodiester bond by the His105 residue from subunit A. This generates a covalent intermediate with the 5'-terminal phosphate. In the meantime, protonation of the 3'-leaving group hydroxyl occurs, with the main role in this stabilization process belonging to the B subunit-situated H105 residue. The second reaction step involves hydrolysis of the covalent phosphohistidine intermediate performed by a water molecule. The H105 from subunit B can subtract a proton from the water molecule, thus facilitating the reaction. Resolving of the covalent

38

intermediate ultimately results in the release of His105 of subunit A, as well as the cleaved DNA bearing a 5'-phosphate³⁹. In contrast, the active site of Type II metal-dependent REases usually contains one or two divalent cations. One of them always occupies the same position in the catalytic center of the restriction enzyme and is likely responsible for the transition state stabilization achieved by neutralizing the accumulation of negative charge on the phosphorus. If present, the second divalent cation shows rather variable positioning, although it tends to reside near the 3'-leaving group¹³³. For example, molecular dynamics simulation-based studies of EcoRV REase suggest that the second Mg²⁺ ion plays a key role in the formation of a tightly bound protein-DNA complex and aids in stabilizing a distorted DNA conformation¹⁵⁴. The proposed one-metal-ion mechanism of catalytic action of EcoRV also indicates the activity-regulating role of the second metal ion, whereas the first metal ion essentially activates a water molecule, involved in the leaving group protonation, and thus directly contributes to the DNA cleavage reaction¹⁵⁵⁻¹⁵⁷.

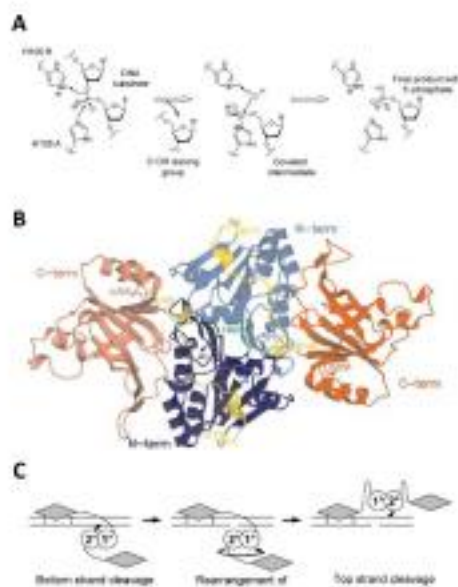


Figure 3. Molecular and structural mechanisms of BfiI restriction endonuclease (REase). **(A)** Hypothetical reaction mechanism of BfiI-catalyzed DNA cleavage. H105:A and H105:B stand for histidine residue at position 105 from subunits A and B, respectively. Adapted from Sasnauskas *et al.*³⁹. **(B)** Crystal structure of BfiI REase homodimer. The two N-terminal domains are colored in light and dark blue, C-terminal DNA-binding domains in light and dark red, N- and C-terminal domains connecting linkers in yellow and active site residues of the dimeric catalytic core in green. Adapted from Grazulis *et al.*¹⁴⁸. **(C)** A model explaining the orientation-related

39

changes of the single BfiI active site engaged in the cleavage of both bottom and top strands of DNA. The primary and the secondary subunits of the BfiI homodimer are indicated as 1° and 2°, respectively. Adapted from Sasnauskas *et al.*³⁹.

In solution, BfiI REase forms a dimer that is capable of binding two copies of its recognition sequence. BfiI interaction with DNA can occur *in cis* if both protein recognition sites are present on the same DNA molecule, or *in trans*, when BfiI targets are located on the two separate molecules of DNA⁴¹. A single polypeptide chain of BfiI contains N-terminal phospholipase D (PLD)-like catalytic domain and C-terminal DBD (Fig. 3B). Two isolated N-terminal domains of BfiI act as non-specific nuclease¹⁴⁸. The C-terminal domain of this restriction enzyme is structurally highly similar to the N-terminal effector domain of Mg²⁺-dependent Type IIE REase EcoRII and B3 DBD of plant transcription factors. Since they all share a common structural fold, which is classified as the DNA-binding pseudobarrel, BfiI protein C-terminal domain is attributed to the B3-like domain family¹⁵⁸. The functionally active BfiI homodimer is formed by the two directly contacting N-terminal domains, which make a catalytic core flanked by C-terminal

domains. The DBDs do not contact each other, as they are situated on the opposite sides of the dimeric core. However, the same single C-terminal domain makes contacts with both N-terminal domains of BfiI, thereby contributing to the protein's interdomain interface. In the monomer of BfiI REase, N-terminal and C terminal domains are connected via a peptide linker, which mostly adopts an extended conformation. This interdomain linker contains 28 amino acid residues, of which the negatively charged Asp and Glu residues are abundant, and runs across the putative DNA-binding surface at the N-terminal catalytic domain, thus serving as an important component for sterical surface occlusion¹⁴⁸.

Experiments performed with the engineered N9C/S333C double mutant of BfiI REase and a small artificial phosphodiester substrate bis(*p* nitrophenyl) phosphate have provided valuable mechanistic insights into BfiI catalytic activity control¹⁵⁹. In the established model, the active site of BfiI is inaccessible to DNA due to the physical blockage created by the interdomain peptide linker. When the cognate DNA is missing, this restriction enzyme resides in the closed state. DNA cleavage is unable to occur in such state, but hydrolysis of bis(*p*-nitrophenyl) phosphate can still take place, indicating that the active site is functionally predisposed for cleavage. Cognate DNA binding to the C-terminal domain of BfiI dislodges the interdomain linker and the entrance to the active site positioned at the two-contacting N-terminal domains becomes available. This means that, upon binding to its recognition sequence,

40

BfiI REase undergoes a substantial conformational change and transitions from the closed to the open state. The latter is catalytically competent and engages in a site-specific cleavage of DNA. In comparison, the fully cross linked BfiI variant, in which the N-terminal and C-terminal domains of both protein subunits are covalently tied to each other via a disulfide bridge, remains “locked” in the closed, DNA hydrolysis-incapable state even after binding to cognate DNA. Therefore, the intrasteric, active site-directed inhibition is the underlying mechanism of the catalytic activity regulation of BfiI restriction enzyme¹⁵⁹.

BfiI catalyzed cleavage of one recognition site-containing plasmid DNA is slower than the cutting of two-sites plasmid, indicating that single recognition sequence-bound BfiI displays low enzymatic activity. The DNA cleavage activity is maximized after this REase concurrently binds to two copies of its recognition sequence⁴¹. The active site is situated in the intersubunit cleft of BfiI and there is only one active site present per protein homodimer¹⁴⁸. The cleavage of the two DNA strands occurs sequentially rather than both at the same time – BfiI REase initially cuts the bottom strand

of DNA and then proceeds with the cleavage of the top DNA strand (Fig. 3C). Low pH conditions selectively block BfiI from cutting the top strand and force it into a nicking enzyme “state”, which is capable of cleaving the bottom strand only⁴⁰. Truncated BfiI heterodimers, lacking the DBD in one of the subunits, were generated and used to elucidate the molecular mechanism for BfiI-catalyzed cleavage of dsDNA³⁹. The target phosphodiester bonds of the bottom and top strands are attacked one by one by the H105 residue located on the DNA-free subunit of BfiI. In both hydrolysis reactions, the proton donor/acceptor role is adopted likely by the equivalent His from the subunit that is bound to the recognition sequence. In such a scenario, a 180° rotation of the protein’s active site must take place in between the two successive reactions to match the anti-parallel polarity of both DNA strands. It is possible that, when a wild-type (WT) BfiI homodimer is bound to two copies of its recognition sequence, these two catalytic-center H105 residues from individual subunits switch roles during the cleavage of both DNA target sites. However, it is certain that BfiI REase always uses histidine of one specific subunit to perform scission of the bottom and top DNA strands at each recognition site³⁹. In contrast, most dsDNA-cleaving nucleases are comprised of two identical subunits related by rotational symmetry. In this case, the two active sites of such nuclease homodimer are arranged so that one from the first subunit attacks the particular DNA strand, while the other catalytic center of the second subunit, which is oriented oppositely, acts on the remaining strand of DNA¹⁶⁰.

41

Although a complete biochemical and structural characterization of BfiI has been already done, till now there has been published only one SM-level study, in which this REase was used solely for the validation of a novel DNA Curtains platform¹⁶¹. Therefore, the dynamic information about the interplay between BfiI and DNA is still lacking. Addressing this knowledge gap would provide better mechanistic insights into the unique properties characteristic to BfiI, as well as broaden the fundamental understanding about proteins with two DNA-binding domains in general. It has been shown that such proteins’ transient *in cis* and *in trans* interactions with two target sites often play a considerable role in target search^{162–164}. Also, these interactions are a key factor in many cellular processes, such as transcriptional regulation¹⁶⁵ and DNA transposition¹⁶⁶.

1.6.CRISPR-Cas systems

CRISPR-Cas systems belong to the adaptive immunity of prokaryotes and are capable of retaining memory of past infections. Repeatedly exposed to the same infectious agent, these defense systems deploy RNA-guided

nucleases that neutralize phages, plasmids and other MGEs in a sequence specific manner⁹⁶. CRISPR-Cas modules are encoded in the genomes of bacteria and archaea. This so-called CRISPR locus consists of CRISPR array and various *cas* genes located next to this array. The CRISPR array itself is composed of short direct repeats separated by short variable DNA sequences called spacers²⁴. CRISPR loci are found in 42% of bacterial and 85% of archaeal genomes, as well as in some bacteriophages and other MGEs⁹⁷. Three main stages – adaptation, expression and interference – are the cornerstones of CRISPR-Cas-based cellular protection. During the adaptation stage, a dedicated Cas protein complex usually at first recognizes a specific short motif called the protospacer-adjacent motif (PAM). The adaptation complex then binds to the target DNA, cuts a certain fragment out of this DNA – the so-called protospacer – and inserts it into the CRISPR array, where the protospacer becomes a spacer. For spacer integration to occur, the repeat at the 5' end of the CRISPR array must be already duplicated^{167,168}. In all adaptation mechanisms known to date, Cas1 and Cas2 proteins are required for capturing the protospacer and catalytically inserting it into the CRISPR array¹⁶⁹. Uniquely, in Type III CRISPR-Cas systems, RNA-derived spacer integration can also occur, as these systems contain reverse transcriptase-Cas1 fusion proteins¹⁷⁰.

During the expression stage, CRISPR array is transcribed into a single long transcript termed pre-CRISPR RNA (pre-crRNA). Transcription

42

typically occurs from a promoter located in the leader region upstream of the CRISPR array. The synthesized pre-crRNA is then processed into mature CRISPR RNA (crRNA) molecules¹⁶⁷. In general, two fundamental features are characteristic of any mature crRNA molecule: (1) a constant palindromic sequence derived from the repeats of the CRISPR array and (2) a variable sequence derived from the spacers of this array. The latter is complementary to the foreign NA invading the cell and is crucial for the downstream CRISPR Cas immune response^{171–173}. Largely the Cas5d or Cas6 endoribonuclease performs the processing of pre-crRNA in Type I and III CRISPR-Cas systems. In most Type I systems, the resultant crRNA molecules are safe from the cellular ribonucleases (RNases) because, after processing, Cas6 remains bound to the crRNA. Such Cas6:crRNA complex is afterwards incorporated into the surveillance complex, a large multiprotein formation. In Type III systems, the leftover hairpin of the crRNA is removed during the second cleavage step, after which the Cas6 protein releases this fully matured crRNA^{97,174}.

In Type II, V and VI CRISPR-Cas systems, the same cellular machinery that is responsible for defensive functions also participates in crRNA processing, since all CRISPR-Cas systems of Class 2 lack factors dedicated for

the maturation of crRNA. The pre-crRNA can have certain features, like 5'-overhangs or pseudoknots, which are recognized and specifically cleaved by V-A subtype and Type VI systems. Type II and V-B subtype systems, in contrast, employ a second RNA molecule called *trans*-activating CRISPR RNA (tracrRNA). By annealing to the fixed region of the pre-crRNA, it acts as an intermediate between the nascent crRNA and the Cas9 or Cas12b effector. The CRISPR-Cas9 system of *Streptococcus pyogenes* additionally relies on the host cell RNases, namely RNase III, to process the formed pre crRNA:tracrRNA complex administered by the Cas9 protein^{97,169,175}. The interference stage begins from the formation of a ribonucleoprotein complex, after the matured crRNA is loaded onto Cas proteins. Guided by the crRNA, this complex scans all available NA for sequence complementarity in the host cell cytosol. When the invading NA is encountered, its base pairing with the interrogative crRNA spacer sequence takes place and, if successful, the recognized target NA is cleaved. In Type II, III, V and VI CRISPR-Cas systems, the crRNA molecule is loaded onto the Cas nuclease, which is itself capable of cleaving within the target sequence or near it. The cleavage mechanism in Type I systems differs from the rest, as here the crRNA molecule is included into the nuclease activity-deficient Cascade complex. Once this crRNA-guided complex locates the target, it recruits the helicase/nuclease Cas3 for DNA cleavage in the vicinity of the target

43

sequence¹⁶⁸. Most CRISPR-Cas systems are being repurposed as genome editing tools^{4,176} due to a high precision and programmable sequence-specific cleavage of target NA, the two unique attributes characteristic of CRISPR-Cas adaptive immunity^{177,178}.

There is a variation in the main substrates of different Cas nucleases. In Type I, II and V systems, the corresponding Cas proteins recognize dsDNA through the formation of an R-loop – a triple-stranded NA structure composed of a target DNA:crRNA hybrid and displaced non-complementary DNA strand⁹⁷. The prokaryotic cell invaders typically produce RNA transcripts in the host before propagating further. These RNA molecules are detected by Type III and VI CRISPR-Cas systems via the complementary base pairing of crRNA with the target transcript⁹⁷. This, in turn, activates diverse non-specific nucleases, e.g., deoxyribonuclease activity of the Cas10 subunit characteristic of Type III CRISPR-Cas effector complex. In addition to the non-specific degradation of the template DNA performed by the activated Cas10 protein, ATP conversion into cyclic oligoadenylates also begins. This is catalyzed by the activated Palm domain of the Cas10 subunit. Cyclic oligoadenylates act as secondary signaling molecules that bind to the non-specific RNases Csm6/Csx1. Upon activation, these enzymes induce host growth arrest, thus helping the CRISPR-Cas defense system to battle the threat⁷¹. Recognition of

the target RNA molecule by the ribonucleoprotein complex of Type VI systems activates the non-specific RNase activity of Cas13 protein. Similarly, this as well drives the host cell into a quiescent state that prevents the further spreading of the infection¹⁷⁹. To prevent the host cell of possible self destruction, these both CRISPR-Cas system types ensure that the target RNA would be cleaved shortly after its recognition, thereby shutting off the activity of non-specific nucleases¹⁶⁸.

To avoid self-targeting of the host cell, adaptive immunity-defining CRISPR-Cas systems must distinguish between self and non-self NA¹⁶⁹. Different types of CRISPR-Cas systems exhibit distinct key features ensuring that only the intruder would be targeted and destroyed. Type I, II and V systems make use of the PAM sequence – a DNA motif that flanks the complementary RNA-DNA region. Alternatively, Type VI systems benefit from an RNA motif called the protospacer flanking sequence that is present in the target RNA molecule. In Type III systems, the complementarity between the 5'-end of crRNA and the 3'-end of the target RNA is lacking¹⁸⁰. Currently, 2 classes, 7 types and 46 subtypes of CRISPR-Cas systems are known¹⁸¹. CRISPR-Cas systems are divided into two classes based on architecture of the effector modules involved in the crRNA biogenesis and interference steps¹⁶⁷. Multiple different proteins make up the effector modules

44

of Class I systems. The formed multisubunit crRNA-binding complexes, such as the Cascade complex in Type I systems, perform the processing of pre crRNA and execute the invading-target interference with the assistance of ancillary Cas proteins. In contrast, Class 2 systems solely rely on a single, multidomain crRNA-binding protein, e.g., Cas9 in Type II systems, for fulfilling these interference and, in most cases, pre-crRNA processing tasks¹⁸². The classification of CRISPR-Cas systems is also related to the classification of *cas* genes, with more than 13 major families of these genes known to date¹⁸¹. Differences in the composition of *cas* genes, as well as in the sequences and structures of Cas proteins, determine the further classification of CRISPR-Cas systems into the corresponding types and subtypes. Alternatively, CRISPR-Cas systems can be grouped based on the distinct structural and sequence peculiarities of the repeats themselves^{183,184}.

Types I, III, IV and VII belong to Class I CRISPR-Cas systems¹⁸¹. These systems account for ~90% of all CRISPR loci found in bacterial and archaeal genomes²⁴. The effector complexes of Class 1 systems share a common organization, suggesting their collective origin. Numerous paralogous proteins from the Cas5 and Cas7 families form the backbone of Class 1 effector complexes. Derived variants of the RNA recognition motif domain are characteristic of these two families. The third family of such domain

containing Cas proteins, namely Cas6, includes separate RNases responsible for pre-crRNA processing in most Class 1 systems. Additionally, two homologous proteins – the large (Cas8 and Cas10 in Type I and Type III, respectively) and small (Cas11 in both types) subunits – constitute the effector complexes of Class 1¹⁸⁵. Type I effector modules are additionally reinforced by another, usually separate component – Cas3 protein (Fig. 4A). It is crucial for target DNA cleavage, since the fused helicase and HD nuclease domains are often situated within this large enzyme. In Type III systems, the HD nuclease is fused to the effector complex's large subunit Cas10 and catalyzes the transcription-dependent cleavage of DNA¹⁶⁷ (Fig. 4B). Similar interference-ensuring nucleases are absent in Type IV CRISPR-Cas systems, which effector complex is typically composed of Cas5, Cas6 and Cas8 subunits (one copy of each) and multiple copies of Cas7 protein¹⁸⁶ (Fig. 4C). Type IV systems, however, feature a signature CRISPR-associated protein DinG, which is an ATP-dependent 5'-3' dsDNA and RNA/DNA hybrid helicase involved in transcriptional interference¹⁸⁷.

Class 2 CRISPR-Cas systems include Types II, V and VI, and constitute nearly 10% of all CRISPR-Cas systems identified in bacteria¹⁸¹. In archaea, however, the Class 2 systems are extremely rare¹⁸⁵. Out of these three types, Type II systems are predominant in bacteria¹⁶⁷. The effector protein of Type

45

II systems is Cas9 endonuclease¹⁸⁵ (Fig. 4D). This enzyme is composed of the recognition (REC) lobe, which participates in target binding, and the nuclease (NUC) lobe, which contains two nuclease domains – HNH and RuvC¹⁸⁸. When Cas9 engages in target DNA cleavage, target and non-target strands of the bound DNA molecule are cleaved by the HNH and RuvC nuclease domains, respectively¹⁷⁷. Type V systems encode the large effector protein Cas12¹⁸⁵ (Fig. 4E). In contrast to Cas9-mediated cleavage, Cas12 cleaves both strands of the target DNA in quick succession, since this enzyme only has a single conserved RuvC nuclease domain, while the HNH domain is lacking¹⁸⁹. Successful target DNA cleavage by Cas9 and Cas12 proteins generally results in double-stranded breaks with blunt and staggered ends, respectively¹⁹⁰.

Additionally, Cas12 exhibits a *trans*-cleavage activity that causes non-specific degradation of the surrounding ssDNA molecules¹⁹¹. Cas13 protein is the effector of Type VI systems which exclusively target RNA¹⁸⁵ (Fig. 4F). Being unrelated to the effectors of Type II and Type V CRISPR-Cas systems, Cas13 functions via its two HEPN RNase domains¹⁹². Transcripts of the invading DNA genomes, as well as RNA phages, are the primary targets of Cas13-mediated immunity¹⁹³. Upon target RNA recognition, collateral non-specific RNase activity of Cas13 protein is activated, thus providing an additional layer of cellular protection due to induced dormancy in virus-infected bacteria¹⁷⁹.

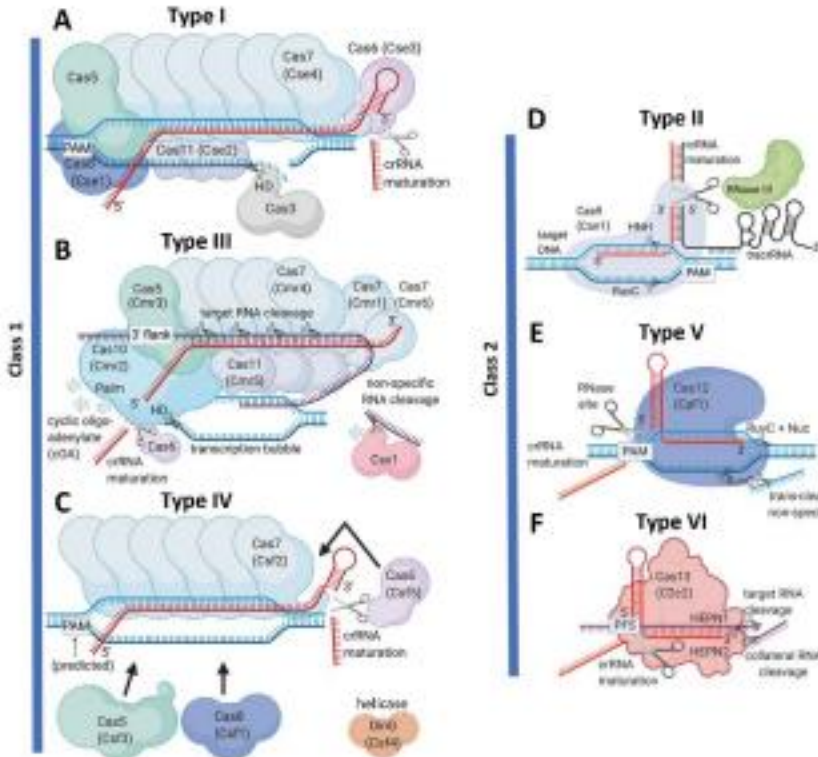


Figure 4. Classification of CRISPR-Cas systems and schematic structural composition of each system's signature interference complex with its primary and secondary functions. DNA is depicted in blue, CRISPR RNA in red, trans-activating CRISPR RNA in black and target or non-specifically cleaved RNA in purple. The legacy names of most proteins are indicated in parentheses. PAM – protospacer adjacent motif, PFS – protospacer flanking site. (A) Type I-E Cascade complex from *Escherichia coli*. (B) Type III-B interference complex from *Pyrococcus furiosus*. (C) Type IV-A1 CRISPR-Cas effector complex from *Aromatoleum aromaticum*. (D) Cas9 is the single large effector protein of Type II CRISPR-Cas systems. (E) In Type V systems, the large single effector enzyme is Cas12. (F) A single large protein Cas13 represents the effectors of Type VI CRISPR-Cas systems. Reproduced and modified from Newsom *et al.*¹⁶⁹.

1.7. Structure and mechanisms of Type I CRISPR-Cas surveillance complexes

Out of four types (Type I, III, IV and VII) that belong to Class 1 CRISPR Cas systems, Type I systems exhibit the largest diversity and prevalence¹⁸¹. Currently, 8 different subtypes of Type I CRISPR-Cas systems are known: A, B, C, D, E, F1, F2 and G¹⁸¹. In all Type I systems, targeting of invading DNA is performed by a multi-subunit crRNA-guided surveillance complex, which is termed Cascade (CRISPR-associated complex for antiviral defense), and

the subsequent degradation of the recognized foreign DNA is catalyzed by the recruited *trans*-acting nuclease/helicase Cas3, a signature protein of such type

systems^{194,195}. The Type I-F CRISPR-Cas surveillance complex also has an alternative legacy name – Csy (CRISPR subtype Ypest (*Yersinia pestis*)) complex, originally given after the discovered *csy* genes^{196,197}. Almost all Type I systems possess two separate proteins, Cas2 and Cas3, which are involved in the adaptation and interference stages of CRISPR-Cas immunity, respectively¹⁹⁴. The distinguishing feature of Type I-F systems is the fusion of Cas2 and Cas3 into a single polypeptide – Cas2/3¹⁹⁸. In the Type I-F1 CRISPR-Cas system from *Pectobacterium atrosepticum*, Cas1 and Cas2/3 proteins assemble into Cas14:Cas2/3₂ complex, in which the endonuclease activity of Cas2/3 is suppressed by the Cas1 proteins¹⁹⁹. Such complex is then only capable of facilitating the spacer integration into CRISPR array^{199,200}. For DNA cleavage to occur, the target-bound Type I-F1 surveillance complex must activate the Cas2/3 fusion protein²⁰¹.

The structural architecture of the typical Type I Cascade complex resembles a seahorse²⁰². The “foot” is formed by the large subunit Cas8 and Cas5 subunit which is positioned at the 5’ end of the crRNA. The crRNA spacer region is surrounded by Cas7 subunits that assemble into a helical filament, the so-called “backbone”. It is capped by the crRNA-bound (at the 3’ end) Cas6 subunit that forms the “head” of the Cascade complex, while Cas11 subunits make up the “belly”^{195,203}. The Type I-F Csy complex from *Pseudomonas aeruginosa* shows a similar 3D structure. This complex is composed of a single 60 nt-long crRNA molecule bound by four different Cas proteins in an unequal stoichiometry manner (Cas8f₁:Cas5f₁:Cas7f₆:Cas6f₁:crRNA₁) and adopts an architecture of an asymmetric spiral^{204,205}. A crRNA-processing enzyme Cas6f is stably associated with the 3’ end of the matured crRNA and caps one end of the spiral, thus forming the so-called “head”. A heterodimer comprised of Cas5f protein and N-terminal domain of Cas8f caps the other end of the spiral, i.e., the “tail”, which accommodates the 5’ end of the processed crRNA. The “backbone” of the Csy complex is made up out of six Cas7f subunits that oligomerize along the crRNA molecule^{204,206}. However, the Cascade complex’s small subunit, Cas11, is absent in the Type I-F surveillance complex¹⁹⁸. Cas11 is known to be involved in R-loop locking, where it stabilizes the non-target strand of DNA²⁰⁷. In Type I-F CRISPR-Cas systems, this function of displaced DNA strand stabilization is taken over by a domain within the Cas7f subunits^{198,205}.

Based on comprehensive biochemical analysis and multiple experiments *in vivo*, the DNA interference mechanism of Type I-F1 Cascade complex from

target located next to the PAM sequence exhibiting promiscuity. Dinucleotides 5'-NC-3' and 5'-CT-3' can serve as the PAMs in this particular Aa-CRISPR-Cas system. In Type I-E and Type I-F CRISPR-Cas systems, PAM recognition is performed by the large subunit Cas8^{209,210}. After binding to the potential target DNA, the Aa-Cascade complex initiates the formation of an R-loop. Here, the basepairing between the spacer of the crRNA and the target strand of the protospacer takes place, whereas the non-target strand of the protospacer is displaced as a ssDNA. A DNA interference checkpoint then occurs, during which the Cascade complex interrogates the length of the R loop. In the case of 32 bp-long (WT length) or longer R-loop, DNA interference is switched on, and the subsequent DNA cleavage is initiated.

This is executed by the helicase/nuclease Cas2/3 which docks to the non-target strand of the R-loop and catalyzes a unidirectional degradation of DNA, starting from the PAM-distal end of the R-loop and proceeding towards the PAM sequence itself and beyond it. The DNA stretching downstream from the protospacer is left intact during this directional degradation process. On the contrary, if the length of the formed R-loop is 26 bp or shorter, DNA interference remains inactive, thus preventing the following cleavage initiation step. Such scenario can possibly happen when Aa-Cascade complex encounters mismatches at the PAM-distal end of the potential target or this complex is of truncated nature²⁰⁸.

During the target search conducted by the Type I-E Cascade complex from *Escherichia coli*, Cascade tends to instantly dissociate from the PAM lacking DNA sites sampled through non-specific interactions. If the PAM sequence is present, Cascade complex can initiate dsDNA unwinding for the subsequent invasion of the crRNA strand. At the potential target with a fully non-matching sequence, rapid dissociation of the bound Cascade occurs. At the potential target with a partially matching sequence, the formation of an R loop can commence in a directional manner, away from the PAM. When a mismatch is encountered, the R-loop can undergo one of the three following fates: (1) stall, (2) continue to form further or (3) promote the dissociation of Cascade. In this case, the mismatch location and thermodynamic barrier determine the exact outcome. At the potential target with a perfectly matching sequence, the R-loop is able to form fully, which results in locking of the Cascade complex via a highly stable interaction¹⁹⁸.

To date, there is a rather limited number of studies that explore the target search mechanisms of Type I CRISPR-Cas surveillance complexes at the SM level. Using smFRET, real-time visualization of DNA search by the Type I-E

Cascade complex was performed. It was shown that a random sampling of dsDNA conducted by Cascade is driven by the short-lived non-specific electrostatic contacts. They form between the dsDNA phosphate backbone and

two lysine-rich loops of the corresponding Cas7 subunits of the Cascade complex. This study also revealed that the observed PAM searching-related binding kinetics of Cascade fall into three main categories: (1) ~100 ms dwell time, when the PAM sequence is absent, (2) ~1 s dwell time, when the PAM sequences are present and their density is low and (3) >1 s dwell time, when there is a high density of PAMs⁴². In another study, SMFM, namely total internal reflection fluorescence (TIRF) imaging, was combined with magnetic tweezers-based measurements to elucidate the target search mechanism of the Cascade complex of *Streptococcus thermophilus* Type I-E CRISPR-Cas system. It was demonstrated that Cascade uses 3D diffusion for its non specific binding to DNA and limited 1D diffusion in a random walk fashion for PAM scanning. The Cascade complex interrogates each encountered PAM sequence for ~0.5 ms. On average, it takes 80 ms until a full R-loop is formed in the case of a productive 1D search event. In contrast, the average duration of non-productive search events is ~150 ms, after which Cascade dissociates from the DNA and continues the target search process via a 3D diffusion⁴³. Furthermore, the target search of Type I-E Cascade complex was directly visualized in a native CRISPR-Cas system *in vivo*. Here, the abundance and mobility of the Cascade complexes were monitored in living *E. coli* cells by the means of a single-particle tracking photoactivated localization microscopy. The findings of this study indicated that Cascade scans the entire nucleoid region of a bacterial cell by interacting with DNA through PAM dependent and PAM-independent strategies. Such search process is rapid (~30 ms dwell time), time-consuming (takes half of Cascade's search time) and driven by the complex's PAM-interacting large subunit Cas8e. Upon target identification, two temporal parameters – search time and critical time – ultimately determine the outcome of CRISPR-Cas-mediated protection of the host cell, which can result in either successful neutralization or further propagation of the invader⁴⁵. The two other studies employed DNA Curtains, a defined high-throughput variation of the standard DNA flow-stretch assays²¹¹, to perform SM-level characterization of the Type I CRISPR-Cas surveillance complex⁴⁴, as well as primed acquisition complex²³. The first study reported that the Cascade complex from *E. coli* is able to use two distinct pathways – a highly efficient PAM-dependent pathway and less efficient PAM-independent pathway – for its target search⁴⁴. The second study concluded that *Thermobifida fusca* Cascade utilizes facilitated 1D diffusion to rapidly scan the non-specific foreign DNA²³. Besides the target search process

characteristic of Type I CRISPR-Cas surveillance complexes, other molecular mechanisms of the RNA-guided Cascade complex and its functional co partner Cas3 helicase/nuclease, such as R-loop formation²¹² and DNA unwinding²¹³, were also examined at the SM level by employing

magnetic

tweezers-based twisting and smFRET assays, respectively. Acr proteins act against CRISPR-Cas systems and inhibit the activity of their effector complexes²¹⁴. AcrIF family proteins can perturb the interference stage of CRISPR-Cas defense in many different ways: bind to the Cascade complex and thus block target DNA binding sites^{206,215,216}, enzymatically inactivate the Cascade complex²¹⁷, allosterically inhibit the cleavage initiation²¹⁶ or DNA hydrolysis^{218–220}. However, it is yet unknown whether such a diverse activity of AcrIF family proteins affects the target search mechanisms of Type I-F CRISPR-Cas surveillance complex. Interestingly, due their inhibiting properties, Acr proteins could be used as effective negative regulators of CRISPR-Cas effector complexes since, while being out of control, these complexes can cause unwanted genetic changes, which generally limits their use in genome editing applications²²¹. There are multiple studies demonstrating the possible application of Type I CRISPR-Cas systems as molecular tools for eukaryotic genome editing^{222–225}. To date, there are still no *in vitro* SMFM-based studies of Type I Cascade that characterize the target search process of this complex, which is interacting with the corresponding Acr proteins. Precise kinetic information that is typically obtained from such sort of studies could aid in optimizing the time related parameters important in CRISPR-Cas-mediated genome engineering⁴.

1.8.Prokaryotic Argonaute proteins

Across all life domains, oligonucleotides, which usually are 15-30 nt in length, serve as guides for Argonaute proteins (Agos), enabling their binding to complementary target NA²²⁶. Agos were originally discovered in eukaryotes²²⁷. Eukaryotic Agos (eAgos) are essentially responsible for RNA interference process – they form RNA-induced silencing complexes and use small single-stranded RNA molecules to either directly cleave the target RNA or recruit additional silencing proteins²²⁸. Therefore, in the cell, the functional roles of eAgos are rather versatile, since these proteins are capable of: (1) post transcriptionally regulating the gene expression, (2) protecting the host cell from the invading RNA viruses and (3) reducing the transposon mobility, thus maintaining the integrity of the host cell's genome²²⁹. In terms of structural organization, all eAgos contain N-terminal, L1 (Linker 1), PAZ (PIWI

Argonaute-Zwille), L2 (Linker 2), MID (Middle) and PIWI (P-element 51

Induced Wimpy Testis) domains^{230,231}. Based on sequence and structural features, two main clades of eAgos – eAGO and ePIWI – are identified currently^{226,230}.

Compared to eAgos, the homologous pAgos show much more diversity in sequence and domain composition²³². Early studies reported that Agos are

found in ~30% and ~10% of all sequenced archaeal and bacterial genomes, respectively^{86,233}. The genes of pAgo often reside in defense islands, where they associate with the other prokaryotic defense-related genes of the host cell^{9,234}. Phylogeny-based classification of pAgo includes two major clades of these proteins – long and short pAgo, which mostly differ in their structural architecture²³². Long pAgo are further divided into the long-A and long-B clades, whereas four phylogenetic clades of short pAgo – S1A, S1B, S2A and S2B – are known to date. Further subdivision of clade S2B short pAgo into 9 subclades (S2B-1 to S2B-9) is possible. Since not all short pAgo are phylogenetically related, some other truncated pAgo are referred to as pseudo-short pAgo. For example, the short pAgo from *Sulfolobus islandicus* (SiAgo) and its homologs form a separate clade named the SiAgo-like clade²²⁶. Even more intriguing example is pAgo encoded by a hyperthermophilic archaeon *Archaeoglobus fulgidus*²³⁵. It is phylogenetically classified as a long-B pAgo, but its structural composition perfectly resembles the one of short pAgo with only MID and catalytically inactive PIWI domains present in the protein. Furthermore, recently it has been shown that this pAgo forms a heterodimeric complex with a protein structurally equivalent to the N L1-L2 domains of long pAgo, thus being highly reminiscent of a long PAZ less pAgo²³⁶.

The six structural domains of long pAgo – N-terminal, L1, PAZ, L2, MID and PIWI – are highly reminiscent of eAgo's domains. Both eAgo and long pAgo exhibit a bilobal structure which is composed of N-PAZ and MID PIWI lobes. The guide and target NA are typically situated between these two lobes²³⁰. Being the least conserved, the N-terminal domain acts as a wedge and participates in guide loading and target cleavage. During guide loading, it facilitates the removal of the passenger strand. After the target undergoes cleavage, this domain facilitates the release of target NA from the protein^{237,238}. The PAZ domain interacts with the 3' end of the guide usually in a sequence-specific manner²³⁹. Thus, it both protects the guide from degradation and holds the designated end of the guide during the target search carried out by the Ago. When the protein encounters a complementary target sequence, the 3' end of the guide is released from the PAZ domain²⁴⁰. This enables a complete guide-target base-pairing, following which the Ago undergoes conformational changes that trigger its endonuclease activity,

ultimately resulting in target cleavage^{226,230,241}. The MID domain forms a charged binding pocket to anchor the 5'-phosphorylated guide^{226,230}. Specifically, in most pAgo, four conserved residues (Y/R, K, Q and K) and a coordinated divalent cation are involved in the anchoring of the guide's 5'-phosphate group^{232,242}. A hydrophobic binding pocket (I, V/I, P, L and Y) and the lack of a divalent cation allow certain pAgo to bind guides bearing a

5'-hydroxyl group. For instance, MpAgo and MhAgo preferentially bind 5'-hydroxylated RNA guides^{243,244}, but there are some other pAgos that accept both 5'-phosphorylated and 5'-hydroxylated guides without any clear preference for a particular moiety²⁴⁵. The MID domain pocket also binds the 5'-nucleotide of the guide, thereby making it unavailable during the base pairing of the emerging guide-target duplex²⁴⁶. The main role of the PIWI domain in the operation of Agos is twofold^{226,230}. This domain promotes the base-pairing between the guide and the target by arranging the guide's 2-7 or 2-8 nt (the so-called "seed" sequence) in a helical conformation. The entropic costs of the guide-target duplex formation are reduced due to the solvent exposed bases of the guide molecule²⁴⁷. More importantly, the PIWI domain is crucial for target NA cleavage²⁴⁸⁻²⁵⁰. The endonuclease activity characteristic of Agos is determined by the catalytic tetrad of amino acids, termed the DEDX (where X is D, H or K) motif, which is found in the PIWI domain and is dependent on another structural feature known as the glutamate finger²⁵¹. The DEDX motif is capable of coordinating two divalent cations and engaging in a nucleophilic attack of the bound substrate. The DEDX tetrad mediated cleavage of the target strand typically occurs between the 10th and 11th nt and strictly relies on the coordinated metal ions that catalyze the hydrolysis of the phosphodiester backbone²⁵². Interestingly, RNase H, which concretely cuts the RNA strand in the RNA-DNA hybrid substrates, is homologous to the PIWI domain²⁵³. Based on the current knowledge, the L1 and L2 domains are not involved in the functional activity of pAgos and serve only as linker sequences^{230,231,254}.

Catalytically active PIWI domain and full-sized PAZ domain are characteristic of most long-A pAgos²³². The majority of characterized long-A pAgos use DNA guides to target and cleave DNA molecules²⁵⁵⁻²⁵⁷. Nonetheless, in some other long-A pAgos, the nature of suitable guides and targets may differ: (1) DNA guides are used for RNA targeting²⁵⁸, (2) RNA guides are used for DNA targeting²⁵⁹ and (3) no clear preference for DNA/RNA acting as the guides/targets is observed^{244,260}. The primary role of long-A pAgos is protection of the host cell against exogenous NA, such as foreign plasmids and invading viruses^{256,261,262}. Besides taking a part in the innate immunity of prokaryotes, long-A pAgos are also capable of promoting

homologous recombination²⁶³ and supporting DNA replication²⁶⁴. In contrast, long-B pAgos are far less studied than the members of the long-A clade²²⁶. The signature structural feature of all long-B pAgos is the inactive PIWI* domain, since the absence of the DEDX motif renders this domain incapable of any catalytic activity (therefore, denoted as PIWI*)²³². Instead of a regular PAZ domain, long-B pAgos often contain the PAZ* domain that is partially truncated (therefore, denoted as PAZ*)^{226,232}. Because of the shortened

sequence, some structural elements of the PAZ* domain's pocket involved in the guide 3' end binding are missing, which could possibly enable more extensive formation of the guide-target duplex^{265,266}. The pAgo from *Rhodobacter sphaeroides* is a rare well-characterized example of canonical long-B pAgos. It has been demonstrated in *E. coli* that, by performing RNA guided binding of target DNA, this particular protein is able to lower the expression level of plasmid-encoded genes and promote degradation of plasmids, even being totally devoid of nuclease activity²⁶⁷.

Short pAgos are the most abundant and comprise 59% of all proteins that are classified as pAgos²³². Short pAgos are predominantly found in bacteria (94%), while archaea encode only a small portion of these proteins (6%)²²⁶. Compared to the general architecture of long pAgos, MID and PIWI* domains are the only two structural elements that are retained by short pAgos²³⁰. Lacking the DEDX catalytic tetrad, all short pAgos themselves are incapable of target cleavage. Instead, in most cases, short pAgos genetically associate with the analog of PAZ (APAZ) domain, which N-terminus, in turn, is typically fused to a variable catalytic domain. This is generally achieved via a particular short pAgo being co-encoded in an operon or fused to another protein that contains the APAZ domain^{86,230,232}. Strikingly, the APAZ domain is not a functional analog of the “original” PAZ domain, but it is rather a homolog of the N-terminal domain of long pAgos^{248,268}. Depending on the specific clade, APAZ domain can be fused to: (1) Silent information regulator 2 (SIR2) domain (in clades S1A and S1B), (2) TIR domain (in clade S2A), (3) Schlafen/Alba, Mrr-like and other domains (in the different subclades of clade S2B)²²⁶. In prokaryotes, SIR2 domains, for example, are NAD⁺-dependent deacetylases that are involved in stress resistance²⁶⁹, immune response modulation²⁷⁰ and Abi systems reliant on NAD⁺ depletion¹⁰⁴. Alternatively, TIR domains possess NADase activity²⁷¹ and play important roles in prokaryotic Abi response by depleting the cellular NAD⁺ resources^{271,272} and in downstream signaling by producing various signaling molecules, such as cyclic adenosine 5'-diphosphate-ribose (ADPR) and its derivatives¹¹⁷. Similar genetic colocalization with other proteins is also characteristic of pseudo-short pAgos²²⁶. Therefore, acting in tandem with their associated proteins, short

54

pAgos ensure the host cell's immunity to diverse MGEs, mainly viruses and plasmids^{98,99}. Unlike long pAgos, which are often capable of directly degrading the invading DNA, defense systems based on short pAgos mostly mediate the Abi response to eliminate the compromised cells and, thus, prevent the further spreading of the infection in the whole population²²⁶.

1.9. Short prokaryotic Argonaute protein-mediated immunity in bacteria

Bacterial defense systems SPARTA (short prokaryotic Argonaute TIR APAZ) and SPARSA (short prokaryotic Argonaute SIR2-APAZ) rely on short pAgos associated with other protein effectors, TIR-APAZ and SIR2-APAZ, respectively²²⁶. In both systems, the functional unit typically attains the form of a heterodimeric complex composed of short pAgo and corresponding effector protein²⁷³. SPARTA is involved solely in anti-plasmid defense, as robust SPARTA-mediated immunity against phages has not been demonstrated yet²⁷⁴. On the contrary, SPARSA confers cellular protection against both these types of foreign DNA. For instance, certain SPARSA systems are able to defend the host cell from dsDNA bacteriophages, while other SPARSAs ensure protection against plasmid DNA that contains a CloDF13 origin of replication^{98,104,275}. The general mechanistic principles of SPARTA and SPARSA-mediated bacterial immunity are rather similar. Transcription of the invading dsDNA generates a sufficient amount of attacker's messenger RNA molecules in the cell. These transcripts are exploited by SPARTA and SPARSA proteins, which then use such acquired small gRNAs to recognize the unwelcome target DNA in a sequence-specific manner²²⁶. In their PIWI* domain, short pAgos of both systems have a conserved sensor loop, which detects the formation of the RNA guide-target ssDNA duplex²⁷³. The resultant conformational changes activate the TIR and SIR2 effector domains of SPARTA and SPARSA systems, respectively. After being turned on, the NADase activity catalyzes the conversion of NAD⁺ into nicotinamide and ADPR²⁷⁴⁻²⁷⁶. Thus, both systems operate by depleting the cellular resources of the central metabolite NAD⁺, which, in turn, triggers host cell death and limits the infection spread within the bacterial population²²⁶. In this whole process, the underlying difference between SPARTA and SPARSA defense systems is the specific requirement for catalytic activation of the corresponding effector domains²⁷³. In order to form the active pocket, TIR domains must undergo oligomerization. This criterion is usually fulfilled through tetramerization of four heterodimeric SPARTAs, each of them being in a complex with the guide-target duplex^{277,278}. The domain activation

requirement, however, is less stringent for SPARSA-mediated immunity, since here, the flexibility increase caused by the target DNA binding is sufficient to fully activate the SIR2 domain²⁷⁵.

Another novel bacterial defense system named SPARDA, which involves short pAgos and protein effectors DREN-APAZ associated with them, has been discovered recently^{279,280}. The very first study focused on short pAgo from *Thermocrisipum municipal* and demonstrated that it forms a heterodimeric complex with the associated protein, back then called

DUF4365-APAZ, where DUF stands for domain of unknown function²⁷⁹. In this *in vitro* study, it was revealed that this complex can bind RNA guides, led by which it recognizes a complementary target ssDNA in a sequence-specific manner. Consequently, the short pAgo-DUF4365-APAZ complex undergoes activation and engages in non-specific cleavage of both ssDNA and dsDNA. These main findings suggested an antiphage-defense role of this endonuclease activity-exhibiting DUF4365-APAZ protein-associated short pAgo system, which could mediate the Abi process²⁷⁹. Such a protective role characteristic of SPARDA system (Fig. 5A) was shortly confirmed by another work, in which short pAgos from *Novosphingopyxis baekryungensis* and *Cupriavidus metallidurans* associated with a putative effector nuclease were studied²⁸⁰. In this study, the DUF4365 effector was designated as DNA and RNA effector nuclease (DREN) and the term SPARDA was coined as a result, thus referring to the short pAgo systems that contain this DREN domain (Fig. 5B). In corroboration with the initial work, it was shown that both short pAgos of interest and the co-encoded protein effector DREN-APAZ form corresponding heterodimeric complexes, which unleash their inherent collateral nuclease activity upon recognizing target DNA in an RNA-guided manner. Additionally, it was demonstrated that activated SPARDA complexes indiscriminately cleave not only ssDNA and dsDNA, but also single-stranded RNA and DNA-RNA hybrid substrates *in vitro*. Experiments *in vivo* unveiled non-specific degradation of cellular DNA by the activated SPARDA. It was concluded that the innate immunity system SPARDA acts against phages and plasmids and ensures protection of the vulnerable bacterial population, supposedly through an Abi-like mechanism, although cell lysis induction does not seem to be a characteristic trait of SPARDA functioning²⁸⁰.

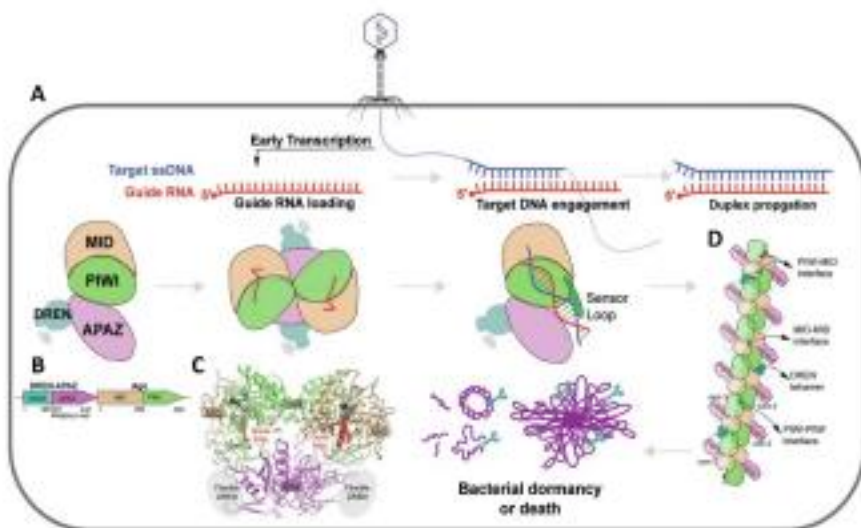


Figure 5. Short prokaryotic Argonaute, DNase associated (SPARDA)-mediated immunity in bacteria. **(A)** A proposed model of cellular protection conferred by the SPARDA complex from *Novosphingopyxis baekryungensis* (NbaSPARDA). The signature feature of NbaSPARDA-mediated defense against foreign DNA is the “monomer-dimer-monomer-filament” transition of this complex, which triggers non specific degradation of collateral nucleic acids and ultimately results in dormancy or death of the host cell. **(B)** Genome organization of the NbaSPARDA system. The four colored domains constitute a heterodimeric NbaSPARDA complex (the so-called “monomer”). **(C)** Cryo-electron microscopy structure of the guide RNA-bound NbaSPARDA complex (the so-called “dimer”). **(D)** Cartoon representation of the topological filament structure. The first, second, third and fourth activated NbaSPARDA complexes are labeled as com.1, com.2, com.3 and com.4, respectively. All panels of this figure were adapted from Wang *et al.*¹⁶.

Cryo-EM-based structural characterization of NbaSPARDA dictated a comprehensive mechanistic model of SPARDA-mediated immunity in bacteria¹⁶. In its apo state, NbaSPARDA exists as a heterodimeric complex (referred as monomer) composed of short prokaryotic Argonaute and effector DREN-APAZ proteins. Once loaded with a 5'-phosphorylated 5'-AU sequence motif-containing gRNA, such NbaSPARDA complex undergoes dimerization (referred as dimer) (Fig. 5C). In the following step, target DNA engagement takes place and the formed NbaSPARDA complex dimers are broken into monomers again. Due to the further propagation of the guide target duplex, filament assembly occurs. This process is controlled by an intrinsic checkpoint mechanism – for filaments to form, the guide-target duplex must be of sufficient length (at least 15 bp, ideally 19-21 bp) and exhibit a perfect complementarity at the central region. The sensor loop of this short pAgo undergoes substantial conformational change exerted by the

propagation of the guide-target heteroduplex. Being pushed outward in the active complex, the sensor loop enables the formation of the PIWI-PIWI interface, a prerequisite for filament assembly. In the NbaSPARDA filament,

interlocking short pAgos form the backbone of the filament, whereas DREN effector domains make up the filament's inner layer (Fig. 5D). The formation of the filament forces monomeric DREN domain to tetramerize. As a result, the initially flexible DREN domain is stabilized and becomes activated. With its collateral nuclease activity unleashed, all the surrounding NA are targeted for non-specific cleavage. Hence, the SPARDA defense system found in *N. baekryungensis* operates via the “monomer-dimer-monomer-filament” structural transition of NbaSPARDA complex and utilizes the multi-turnover collateral nuclease activity of the activated DREN effector domain to indiscriminately degrade both genomic DNA of the host and invading DNA of foreign plasmids or infecting bacteriophages, which ultimately leads to cell death or dormancy¹⁶.

A complete molecular mechanism of SPARDA action, however, remains unclear in the sense of the exact changes that long DNA molecules, such as plasmids or genomic DNA, actually undergo, when they engage in a sequence-*in trans* interactions with the activated SPARDA complexes. A direct real-time visualization of this dynamic multi-step process at the SM level could provide unique insights into the mechanistic features of proteins involved in filamentation-based SPARDA defense system. In terms of its practical applications, the activated SPARDA complex, similarly to CRISPR-Cas12¹⁹¹ and TnpB²⁸¹ proteins, exhibits a collateral nuclease activity, which makes it a potential biomolecular tool for NA diagnostics^{279,280}.

1.10. Structure and biological role of α -synuclein and S100A9 proteins

Neurons of the central and peripheral nervous system, blood cells and other tissues express α -syn – a small (140 amino acids, 14 kDa) acidic protein^{282,283}. Cytosol of synapses is the main location of α -syn, while nuclear and synaptic partitions sometimes can also contain this protein²⁸⁴. Three major structural domains comprise the α -syn protein: (1) an amphipathic feature containing N-terminal domain (1-60), (2) a highly hydrophobic central domain (61-95) and (3) a hydrophilic C-terminal domain (96-140)²⁸⁵. The N terminal domain contains a multi-repeated consensus sequence (KTKEGV), has an alpha-helical structure and is crucial for binding to the cellular membranes. The central domain, also known as the non-amyloid-beta component, is usually folded as an alpha helix and participates in the aggregation of α -syn, when the protein adopts a beta-sheet structure. The C

terminal domain provides flexibility to the polypeptide due to numerous negative charged and proline residues and is involved in binding of the Ca^{2+} ions, as well as in chaperone-like activity^{282,284,285}. Generally referred as a “natively unfolded” monomer²⁸⁶, endogenous α -syn largely acquires a

tetrameric structure²⁸⁷. Being mostly localized at the presynaptic sites, α -syn interacts with proteins that are responsible for the release and reuptake of neurotransmitters^{282,288}. For example, human α -syn can regulate dopamine neurotransmission by managing both the dopamine storage in synaptic vesicles and the expression and activation of vesicular monoamine transporter 2 in nigral neurons^{284,289}. Control of the synaptic vesicle pool and trafficking²⁹⁰, as well as sensing and stabilization of the curved membranes²⁹¹ are another important cellular tasks undertaken by α -syn. Acting as a molecular chaperone, α -syn can promote assembly of the soluble N ethylmaleimide attachment protein receptor complexes, which are involved in the mediation of different exocytosis process' phases, at the synapse^{282,292}. It has been shown that α -syn also performs maintenance of the mitochondrial membrane proteins, such as the electron transport chain complex, by binding to the mitochondrial membrane²⁹³. Abnormal aggregations of α -syn are closely associated with the so-called α -synucleinopathies – sporadic neurodegenerative diseases which include idiopathic Parkinson's disease, dementia with Lewy bodies, multiple systems atrophy, pure autonomic failure and REM sleep behavior disorder^{282,284,285}. Aggregates of α -syn include various formations like cytotoxic oligomers and amyloid fibrils²⁸⁵. Different morphologies, e.g., spherical, elliptical, circular, tubular and flat-shaped, are characteristic of α -syn oligomers that can be composed of more than a hundred polypeptide chains²⁹⁴. Amyloid fibrils are highly compact and beta-sheet structure-enriched aggregates formed during the self-assembly of α -syn²⁹⁵. Various point mutations (A30P, H50Q, A53E and others) of α -syn are known to modulate its aggregation and are associated with the familial forms of α synucleinopathies^{296,297}.

S100A9, also known as MRP14 or calgranulin B, is a Ca^{2+} -binding protein that belongs to the S100 family²⁹⁸. Members of this family are approximately 10 kDa in size and were named S100 due to their solubility in 100% saturated ammonium sulfate²⁹⁹. Full human S100A9 protein consists of 113 amino acid residues, whereas its truncated isoform has 110 amino acids³⁰⁰. S100A9 is found in myeloid cells, cancer cells and in tumor stroma, with the localization of this protein being important for its physiologic functions³⁰¹. For instance, in pancreatic cell lines, S100A9 resides in the cytoplasm and plasma membrane³⁰², but it also can be located within the nuclei, like in the case of esophageal mucosa³⁰³. In normal myeloid cells, such as neutrophils and

monocytes, S100A9 is expressed at a high level and localizes in the cytoplasm of the cell³⁰¹. S100A9 protein contains two Ca^{2+} -binding sites: a high-affinity one at the C-terminus and a low-affinity one at the N-terminus. The canonical high-affinity Ca^{2+} -binding site is composed of 12 amino acids and resembles the shape of a human hand and is, therefore, called an EF-hand motif

characteristic of most S100 proteins. Helix 3 (E), loop 2 and helix 4 (F) are the three major components of this site, while helix 1 (E), loop 1 and helix 2 (F) are the main structural elements of the non-canonical low-affinity Ca^{2+} -binding site, which has a total of 14 amino acids. Hinge region connects the helices 2 and 3³⁰¹. A conformational change, namely the rotation of the helix 3, is triggered upon Ca^{2+} binding, which, in turn, exposes the hydrophobic cleft of the protein. This then serves as an anchoring point for a large number of macromolecular interactions with various target proteins, spanning from receptors to other members of S100 family³⁰⁴. For example, it was demonstrated that Ca^{2+} -bound S100A9 protein readily interacts with the receptor for advanced glycation endproducts, major fatty-acid transporter CD36 and fibronectin³⁰¹. S100A9 can attain several different oligomeric states: (1) it can exist as a homodimer, (2) it can form a non-covalently bonded heterodimer S100A8/A9, also known as calprotectin, with the S100A8 protein, and (3), in the presence of Ca^{2+} , it can also appear as a heterotetramer composed of calprotectin and S100A9/A9^{298,301}. The S100A8/A9 heterodimer is predominant in most biological interactions, but the natural state of S100A9 can as well be different depending on the local environment in which this protein resides³⁰¹. Primary role of S100A9 is directly related to inflammatory processes³⁰⁰. Preferentially secreted during inflammation, S100A9 protein promotes inflammation by boosting phagocyte migration and stimulating the secretion of pro-inflammatory cytokines, reactive oxygen species and enzymes responsible for tissue degradation²⁹⁸. Transendothelial migration enhancement is a direct consequence of S100A9 action, as it causes neutrophil adhesion to the extracellular matrix proteins, as well as prolongs the activation of CD11b³⁰⁵. S100A9 protein stimulates the production of nitric oxide and reactive oxygen species at the inflammatory site³⁰⁶, where it also induces degranulation of neutrophils assisted by JNK and p38³⁰⁷. Neutrophil phagocytosis and phosphorylation of syk, Erk1/2 and Akt, which results in the killing of bacterial cells, are highly stimulated by S100A9³⁰⁷. This protein also activates NF- κ B translocation to the nucleus, thus triggering the release of TNF, IL-1 β and IL-6³⁰⁸. In addition, S100A9 plays a role in thrombosis, since activated platelets were shown to secrete this protein³⁰⁹. Calprotectin, a S100A8/A9 heterodimer, is involved in cytoskeleton rearrangement and arachidonic acid metabolism³⁰⁰. Numerous cancer types, e.g., breast cancer,

gastric cancer, colorectal cancer, lung cancer and others, were shown to have an up-regulated expression of S100A9³⁰¹.

A substantial cross-interaction between α -syn and S100A9 proteins has been revealed recently³¹⁰. Having amyloidogenic properties on its own, S100A9 is now considered to play a possible role in the onset of certain neurodegenerative disorders. Some of these disorders, like Alzheimer's or

Parkinson's disease, together with various cancer types have been linked with the aberrant LLPS process^{311,312}. It has been demonstrated that high molecular crowding conditions *in vitro* provoke α -syn to readily form protein droplets via LLPS³¹³. Namely, the LLPS of α -syn precedes its aggregation – a subsequent liquid-to-solid transition drives the formation of oligomer and fibrillar species-containing amyloid hydrogel from the initial α -syn liquid-like droplets. In general, LLPS is a reversible physicochemical process that involves a dilute phase and a dense phase, which consists of aggregates of large molecular components³¹⁴. During this spontaneous demixing of biomolecules, various membraneless organelles, also known as biomolecular condensates or droplets, form within cells at the nanoscale³¹⁵. The key components and mediators of the LLPS process are proteins, DNA and RNA. Intrinsically disordered regions, enriched in aromatic, charged and hydrophilic residues, and low-complexity regions, which adopt flexible conformations and are devoid of stable tertiary structures, are often found in proteins that exhibit LLPS behavior. Multiple weak interactions, such as π stacking, cation- π interactions, electrostatic forces and hydrophobic contacts, occur between protein amino acid residues and other macromolecules, thus providing a molecular basis for LLPS to begin³¹⁶. LLPS can be also induced by scaffold proteins, which engage in multivalent specific interaction networks³¹⁷, and RNA-containing repetitive sequences³¹⁸. Various environmental parameters, e.g., component concentration, system temperature, salinity and pH, affect the course of LLPS³¹⁹. The overall role of LLPS in the cellular context is extremely multifaceted, as this biomolecular condensation process participates in: transcriptional regulation, genome organization and remapping, immune response (autophagy), neuronal synaptic signaling and the formation of DNA damage repair foci, proteasomes and stress granules^{314–316,320}.

1.11. Open-microscopy frameworks for super-resolution imaging

Interaction between protein and DNA is a dynamic process that typically involves three fundamental stages carried out by the protein: (1) target search,

61

(2) target recognition and (3) functional activity within the formed protein DNA complex³²¹. SM-level real-time direct visualization of such complexes and their precise kinetic information extraction require a fluorescence microscope with a high sensitivity, stability, spatio-temporal resolution and semi-to-full modularity³². Additionally, mechanistic studies of protein-DNA interactions often benefit from SR imaging, which becomes even more relevant when such studies are conducted in living cells³²². Most commercial

FM optical systems exhibiting such specifications are very expensive (hundreds of thousands of Euros) and, in most cases, are capable of imaging a rather limited variety and complexity of biological samples due to their lack of modularity. Various open-microscopy frameworks developed by many independent research groups effectively address these downsides characteristic of commercial microscopes³¹.

The research laboratory led by Johannes Hohlbein developed an open source, modular microscope, the miCube, offering SMFM and SR imaging possibilities (Fig 6A). The accessibility and versatility of this microscopy system come from combining 3D-printed parts with the commercially available optical elements, lasers and other components, which all together constitute a highly modular hardware of the miCube microscope³². Diederich and co-authors presented a fully modular and easily customizable microscopy platform named UC2 (Fig. 6B). This open-hardware microscope is based on 3D-printed cubes and inserts that serve as individual building blocks for users to flexibly assemble a desired custom optical configuration in a Lego-style manner. The UC2 system offers a wide range of different imaging modes, spanning from the conventional brightfield and widefield fluorescence microscopies to SR structured illumination microscopy (SIM), image scanning microscopy and intensity diffraction tomography³²³. SMLM is also possible with the expanded version of UC2 microscopy framework³²⁴. The research group led by Jonas Ries comprehensively described a custom-built open-source microscope which is optimized for SMLM, ultra-stable 3D imaging and multicolor functionalities³⁴. Danial and colleagues developed a cost-efficient and easy-to-assemble SR microscope called NanoPro. Similarly to other open-microscopy solutions, this instrument enables high-quality SMLM in a high-throughput manner³⁵. Moya Munoz *et al.* introduced the Brick-MIC, an affordable open-source portable 3D-printed microscopy platform designed both for SM detection and spectroscopy measurements (Fig. 6C). Its versatility and adaptability allow one to perform SMFM and SR imaging along with fluorescence correlation spectroscopy and time-correlated single-photon counting³²⁵. Niederauer *et al.* introduced an open-microscopy framework, the K2, dedicated for SMFM, live-cell imaging and fluorescence

62

recovery after photobleaching (FRAP) experiments. It is a TIRF microscope which features a focus lock and overall modular design with an extended detection pathway for the specific applications requiring a simultaneous triple color imaging³⁶. Huo and co-authors presented a low-cost compact open hardware TIRF microscope named Open Blink. It features a powerful homogeneous quad-line laser illumination and active focus-lock stabilization, which both, in combination with the dual-channel detection, allow for quantitative multicolor SR imaging using this instrument³²⁶. Besides these

specific aforementioned open-microscopy systems, there are multiple other alternatives that also specialize in SR microscopy, e.g., easySTORM³²⁷, fiberSIM³²⁸, liteTIRF³²⁹, openSIMMO³³⁰, openSIM³³¹ and cellSTORM³³². To date, various open-source software solutions were introduced to enable both the control of such fluorescence microscopes and acquisition of the images and other data, and/or to facilitate the data analysis process: ImSwitch³³³, μ Manager³³⁴, Pycro-Manager³³⁵, Microscope-Cockpit³³⁶, Fiji³³⁷, SMAP³³⁸ and napari³³⁹.

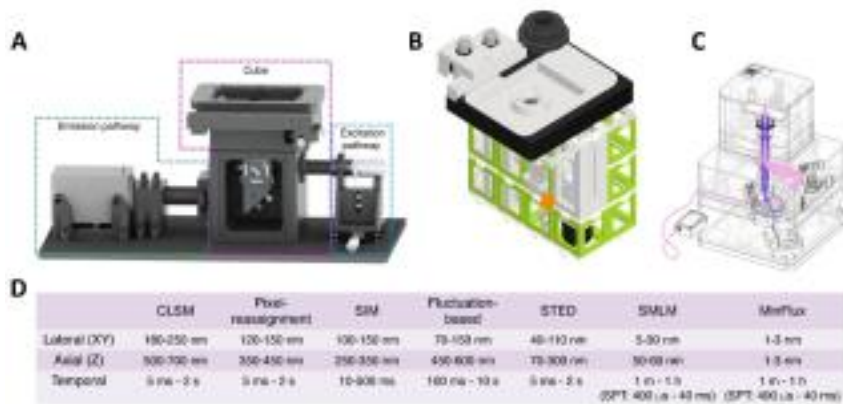


Figure 6. Various openly available super-resolution (SR) microscopy systems and comparison of maximum spatio-temporal resolution achievable with a standard confocal laser scanning microscopy (CLSM) and commonly used different SR imaging techniques. **(A)** Open-source modular SR microscope miCube. Blue, magenta and green dashed lines denote the components of excitation path, main cube and detection pathway, respectively. Adapted from Martens *et al.*³². **(B)** Open hardware fully modular high-resolution widefield fluorescence microscope UC2. Adapted from Zehrer *et al.*³²⁴. **(C)** Portable and adaptable open-microscopy platform Brick-MIC capable of SR imaging, namely, single-molecule localization microscopy. Excitation light path is colored pink, and emission light pathway is marked in purple. Adapted from Moya Munoz *et al.*³²⁵. **(D)** Lateral, axial and temporal resolutions' typical range characteristic of CLSM and various SR imaging techniques. SIM – structured illumination microscopy, STED – stimulated emission depletion, SMLM –

63

single-molecule localization microscopy, SPT – single-particle tracking, MinFlux – minimal photon fluxes. Adapted from Valli *et al.*³⁴⁰.

The highest resolution achievable with various traditional FM methods, such as widefield epi-FM, typically reaches ~200–300 nm in the lateral plane and ~500–700 nm in the axial plane^{341,342}. This maximum limit in standard FM exists because of the wave-particle duality of the light. Specifically, due to its wave nature, the light is subject to diffraction³⁴³. SR microscopy, also known as high-resolution microscopy, fluorescence nanoscopy and diffraction unlimited imaging, is capable of effectively “breaching” the classical barrier

of light diffraction inherent in optical resolution^{341,344} (Fig. 6D). To date, numerous SR microscopy techniques have been developed, of which the following are among the most widely used: stimulated emission depletion (STED) microscopy³⁴⁵, SIM³⁴⁶, photoactivated localization microscopy (PALM)³⁴⁷, direct stochastic optical reconstruction microscopy (dSTORM)³⁴⁸ and point accumulation for imaging in nanoscale topography (PAINT)^{349,350}. The last three methods constitute a separate branch of SR imaging termed SMLM²⁷. Some of these methods serve as the basis for other frequently used SR microscopy approaches, such as single-particle tracking (SPT), also known as SM tracking (SMT)³⁰, and minimal fluorescence photon fluxes (MINFLUX)²⁸. In addition to all these widely used SR microscopy techniques, there are other high-resolution microscopy methods that offer certain advantages: reversible saturable optical fluorescence transition (RESOLFT) microscopy³⁵¹, Airyscan³⁵², 4Pi microscopy³⁵³, fluorescence imaging with one nanometer accuracy (FIONA)³⁵⁴, SR optical fluctuation imaging (SOFI)²⁹, SR radial fluctuations (SRRF) microscopy³⁵⁵ and expansion microscopy³⁵⁶.

1.12. Single-molecule methods for protein-DNA interaction studies

FRET is a non-radiative and distance-dependent energy transfer from a donor to an acceptor, the two fluorescent dye molecules that are generally termed as the “FRET pair”³⁵⁷. Dipole-dipole interaction between the donor’s and acceptor’s electronic states is the main driving force of this energy transfer process³⁵⁸. Namely, an optically generated electronic coherence of the donor exhibits oscillations, which in the case of their resonance with the acceptor’s electronic energy gap, result in the occurrence of FRET³⁵⁹. Collectively, FRET can only occur when all the following three criteria are fully met: (1) there is an overlap between the emission spectrum of a donor and the absorption spectrum of an acceptor, (2) the distance between the donor and the acceptor is no more than 10 nm, and (3) the donor’s and acceptor’s transition dipole moments are relatively oriented so that the energy transfer is allowed to

64

happen^{360,361}. The efficiency of energy transfer is formulated as $E_{\text{FRET}} = (1 + (r/R_0)^6)^{-1}$, where r is the distance between the donor and the acceptor, and R_0 is the Förster radius – the characteristic distance between a donor and an acceptor at which 50% of energy is transferred^{357,358,360}. Changes in the distance from ~1 to 10 nm between the donor and acceptor fluorophores can be detected by measuring the FRET efficiency^{362,363}. This makes ensemble FRET a powerful tool for investigating structural features and dynamics of various biological systems^{364,365}.

Ensemble-based FRET methods are, however, highly susceptible to any dynamic or static heterogeneity that may be present in the sample³⁶⁶. A dedicated FRET technique – smFRET – effectively resolves this sample

heterogeneity-related issue by enabling the observation of FRET at the SM level^{357,360} (Fig. 7A). In smFRET, rather than on multiple different molecules at a time, the same pair of donor and acceptor fluorophores is excited and detected on an individual molecule^{367,368}. Measurements of smFRET are usually performed using either a confocal microscopy system or a TIRF microscopy setup capable of widefield imaging^{358,368}. Ratiometric methods are typically employed to experimentally deduce the efficiency of energy transfer in smFRET. In the intensity-based smFRET, the apparent FRET efficiency is calculated as $E^* = I_A / (I_A + I_D)$, where I_A is the fluorescence intensity of an acceptor and I_D is the fluorescence intensity of a donor. In the lifetime-based smFRET, the FRET efficiency is given as $E_{\text{FRET}} = 1 - \tau_{\text{DA}} / \tau_{\text{D}}$, where τ_{DA} and τ_{D} are the donor's fluorescence lifetime in the presence and absence of the acceptor, respectively^{357,358,361}. Being comparable to the protein size or lipid membrane thickness, the favorable distance (1-10 nm) for the energy transfer to occur is the key advantage of smFRET technique, often specified as a "spectroscopic ruler"^{357,358}. Application wise, smFRET enables measurement and quantification of the dynamic interactions between the neighboring fluorescently labeled biomolecules, for example, protein-protein³⁶⁹ and protein-DNA³⁷⁰ interactions³⁷¹. Conformational changes and dynamics of single DNA³⁷², RNA³⁷³ and protein³⁷⁴ molecules, as well as individual protein folding³⁷⁵, also can be observed by the means of smFRET, when the labeling with the donor and acceptor fluorophores is performed at the two selected different sites on the same single molecule³⁷¹.

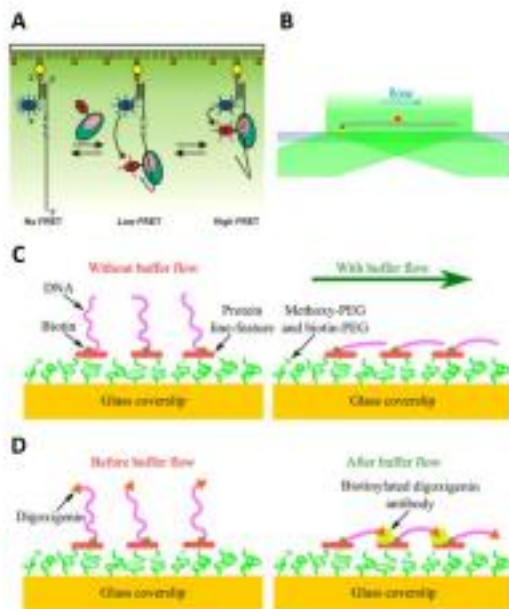


Figure 7. Several examples of the major techniques used for studying protein-DNA interactions at the single-molecule level. **(A)** Single-molecule Förster resonance energy transfer. The donor (D) fluorophore is depicted in blue, acceptor (A) fluorophore in red. **(B)** DNA flow-stretch assay. The protein of interest is shown in red, generated evanescent field in green. **(C)** Single-tethered Soft DNA Curtains. PEG – polyethylene glycol. **(D)** Double-tethered Soft DNA Curtains. Adapted from Kopustas *et al.*³⁷⁶.

In a typical DNA flow-stretch assay, a single end of the individual DNA molecules is randomly immobilized on the glass coverslip surface^{377,378} (Fig. 7B). Streptavidin-biotin interaction, a non-covalent, specific, high affinity ($K_D \approx 10^{-15}$ M) interaction, is a long-standing gold standard for tethering single DNA molecules to the surface^{379,380}. To achieve a single-end anchoring of DNA, the glass coverslip surface is usually chemically modified with polyethylene glycol (PEG) molecules that have additional moieties, such as the methoxy group or biotin. This ensures both passivation and functionalization of the glass surface, which is then coated with streptavidin, thus enabling a non-predefined attachment of the biotinylated DNA molecules^{381,382}. Visualization of DNA is performed by first staining the surface-immobilized DNA molecules with fluorescent intercalating dyes, like SYTOX Green or YOYO-1. In the next step, a continuous buffer flow is applied and the exerted hydrodynamic force causes the stretching of single tethered DNA molecules along the surface. Experiments with the surface attached λ DNA revealed that a shear flow of 0.5-2.0 mL/min can induce

forces up to 4 pN, meaning that with such fluid flow rates, in principle, it is possible to achieve ~90% of the full contour length extension for various

single-tethered DNA molecules³⁸³. In addition, the lateral movement of the immobilized DNA is reduced in the presence of a buffer flow³⁸⁴. Since the flow-stretched DNA molecules reside close to the glass coverslip surface, their imaging is usually performed using a TIRF microscope, which allows for the selective excitation of only a very thin (~100 nm) section above the glass surface^{385,386}. With the background fluorescence removed, the fluorescently labeled contour of the flow-extended individual DNA molecules can be effectively visualized at a high signal-to-noise ratio^{385,387}. By specifically anchoring the second end of single-tethered DNA molecules to the surface, such DNAs can then remain in a stretched conformation even in the absence of a buffer flow³⁸⁸. DNA flow-stretch assays can accommodate both short and long DNA substrates, of which the latter suits the best for uncovering the true potential of this SM technique. Being 48 502 bp in size, bacteriophage λ genomic DNA exhibits a B-form contour length of ~16.5 μm and is the most commonly used substrate in this methodological field^{389,390}. The main application of DNA flow-stretch assays involves protein-DNA interaction studies at the SM level³⁷⁶. By fluorescently labeling the proteins of interest, these can then be injected into a flow cell, which has the DNA flow stretch assay already prepared, to visualize and observe individual protein-DNA interaction dynamics in real time³⁹¹. Namely, this includes characterization of protein-binding events, e.g., position on the DNA and dwell time^{392,393}, tracking of the protein movement and diffusion along the extended DNA molecule^{394,395}, investigation of the DNA unwinding³⁹⁶ and repair³⁹⁷ processes, and fundamental studies of various DNA-binding protein target search mechanisms^{398,399}.

There are several different adaptations of the standard DNA flow-stretch assay, for instance, DNA tightropes⁴⁰⁰ and DNA flow-stretch approach based on the use of a magnetic bead⁴⁰¹. Another huge branch of the standard DNA flow-stretch assays is the so-called DNA Curtains platform that enables imaging of hundreds of surface-immobilized individual DNA molecules, which are all simultaneously visible in a single field of view of the microscope⁴⁰². The superiority of DNA Curtains over a typical DNA flow stretch assay comes from a high-throughput nature of this experimental platform, as large amounts of statistical data can be collected only in a single experiment performed using this platform⁴⁰³. Supported lipid bilayers and lipid lateral diffusion barriers, which are made out of metal (usually chromium) or negative resist (mainly hydrogen silsesquioxane), compose the core technological design of traditional DNA Curtains^{402,404}. There are

numerous different types of these conventional DNA Curtains: (1) single tethered DNA Curtains^{402,405,406}, (2) double-tethered DNA Curtains^{211,407}, (3) single-stranded DNA Curtains⁴⁰⁸, and (4) others^{409–411}. However, all of them

share several fundamental drawbacks related to the preparation of lipid bilayer-based systems (poses system stability and defect management risks) and, most importantly, to the nanofabrication of physical lipid diffusion barriers (requires expensive equipment, specialized knowledge and lots of time)³⁷⁶. In contrast, Soft DNA Curtains combine scanning probe microscopy and soft lithography techniques to offer a simple and cost-effective platform for high-throughput SM-level studies of protein-DNA interactions^{161,412}. Fabrication of Soft DNA Curtains involves AFM, which is used for the inscription of nanometer-sized lines on the surface of Si master template, and protein lift-off microcontact printing, which enables precise nanopatterning of PEGylated glass coverslip surface with the protein (streptavidin or traptavidin) line-features (Fig. 7C). Biotinylated DNA molecules can then be immobilized on these stable anchor points with a predefined alignment and organized in a massively parallel manner by applying a buffer flow¹⁶¹. Like in any other DNA flow-stretch assay, the main imaging modality for visualizing these flow-extended single DNA molecules is TIRF microscopy. In double tethered Soft DNA Curtains, individual DNA molecules are immobilized on the surface via both of their ends in a defined orientation⁴¹² (Fig. 7D). This is achieved by modifying a second end of the biotinylated DNA molecule with digoxigenin and flow-attaching this digoxigenin-labeled end of a single tethered DNA molecule to the adjacent protein line-feature, which was coated with the biotinylated digoxigenin antibodies in one step before. Besides traditional DNA Curtains and Soft DNA Curtains, few other technological variations of this high-throughput DNA flow-stretch assays' SM method were also demonstrated – Suspended DNA Curtains⁴¹³ and DNA Skybridge⁴¹⁴.

2. RESULTS

2.1. Development of the optical fluorescence microscopy system miEye

In this publication, we described in detail the design, building, characterization and validation of optical FM system miEye. Our widefield fluorescence microscope miEye offers SMFM and SR imaging capabilities. The miEye is compact, fully modular and cost-effective – all its equipment can be purchased for approximately 50 thousand Euro. Although the design of our microscope's main body is based on the open-source miCube microscopy platform³², the entirety of optomechanical components, optical elements and paths is completely modified and expanded in the case of miEye, thus significantly improving its flexibility and stability. The major differences between these two fluorescence imaging systems are the layout of the microscope body (two layers in miCube, three layers in miEye), the origin of the laser engine (commercial in miCube, custom design developed by us in miEye), the number of possible sample excitation schemes (only one based on single-mode fiber (SMF) in miCube, two different ones based either on SMF or on multi-mode fiber (MMF) in miEye) and the presence of sample autofocus system (absent in miCube, present in miEye). The miEye

microscopy system is comprised of five main parts: • **Laser engine.** A pivotal component of the laser engine is a 4- wavelength (405, 488, 520 and 638 nm) laser diode combiner. After exiting the combiner, these free-space lasers are coupled into a SMF or MMF (Fig. 8A).

• **Excitation path.** In the excitation path, the laser beam exiting from one of these fibers is subsequently directed into the same single optical path. This enables a flexible possibility to image even the same sample in two different ways, i.e., using a SMF based excitation scheme (Fig. 8B) or a MMF-based excitation scheme (Fig. 8C). In the first approach, the sample can be illuminated in three different modes of excitation – TIRF, HILO or *Epi*-mode. Alternatively, the MMF output is imaged onto the sample plane of the objective. This constricts the excitation along the optical axis and enables imaging modality similar to the optical sectioning of the sample. In this case, a home-built MMF agitation module is used to produce a flat homogeneous illumination profile of the sample.

• **Microscope body.** The three-layer microscope body is manufactured by first CNC milling an aluminum block and then

- **Autofocusing path based on infrared (IR) laser.** The IR laser based autofocusing system is responsible for compensating for the drift of the sample along the Z-axis, thereby stabilizing the sample in focus. This active drift-suppression system uses a closed-feedback loop between the IR laser position, which is monitored in real time by a detection camera, and the objective piezo stage (Fig. 8E).
- **Emission path.** The emission path is arranged in a 4f optical configuration, supports two-color fluorescence imaging, and features a cost-effective industrial-grade CMOS camera offering a sufficient quantum efficiency across the entire visible light spectrum together with a high frame rate (Fig. 8F).

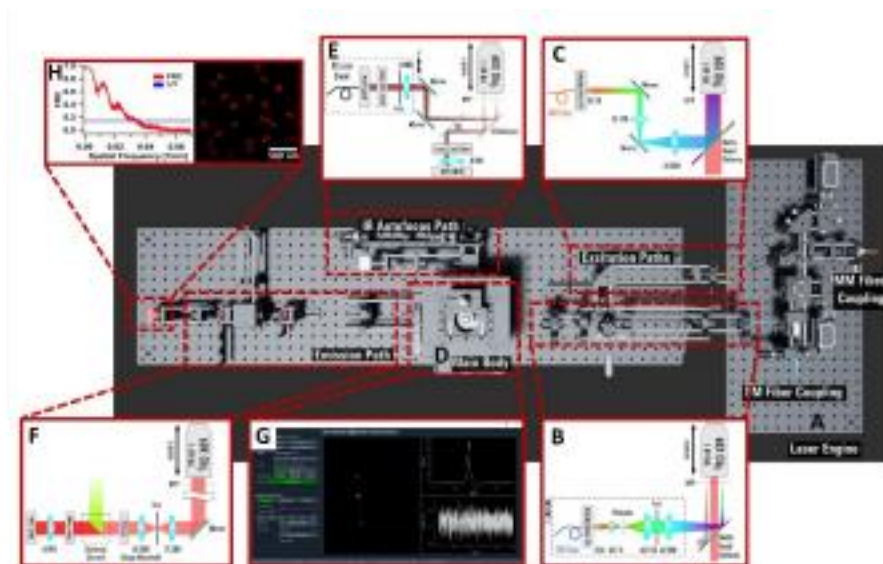


Figure 8. Overview of the main parts and key features of the open-source super resolution microscope miEye. (A) Laser-engine panel. (B) Single-mode fiber-based excitation pathway. (C) Multi-mode fiber-based excitation pathway. (D) Main body of the microscope. (E) Autofocusing pathway based on the infrared laser. (F) Fluorescence emission pathway. (G) Microscope Control module available in the microEye software. (H) Direct stochastic optical reconstruction microscopy (dSTORM) imaging of fluorescently labeled nuclear pore complexes in fixed U2OS Nup96-SNAP cells (on the left – Fourier ring correlation (FRC) curve, on the right – reconstructed super-resolution image). SM – single-mode, MM – multi-mode, AS – aspheric lens, AC – achromat lens, BFP – back-focal plane, TL – tube lens, B – B coated N-BK7 optics, BS – beamsplitter.

The hardware of miEye microscope is controlled by our open-source software, microEye, written in Python³³ (Fig. 8G). The microEye package is dedicated not only for reliable control of the microscope (Control module), but also for the streamlined data acquisition (Acquisition module) and quick

analysis (Tiff viewer module).

To characterize the stability of our microscopy system along the XYZ, under various conditions we performed imaging of 200 nm-diameter red fluorescent polystyrene microspheres immobilized on the glass coverslip surface. We determined that the sample experiences a drift of <7 nm/min in both the lateral and axial planes without the constant active automatic upkeep of the optimal distance between the sample and the objective, i.e., without autofocus. With the autofocus system active, we observed a stabilization of the sample's drift along the Z-axis, and the drift itself

fluctuated within 10 nm around the same stable value throughout the measurement. The sample drift along the XY, in this case, was ~ 3 nm/min. To demonstrate that miEye microscope can be used for nanoscopy, specifically SMLM, we employed a MMF-based excitation scheme and performed dSTORM of fluorescently labeled NPCs in fixed U2OS cells. After analyzing the data, we obtained the precise localizations of fluorescent spots and generated a digital reconstructed SR image in which individual NPCs and their characteristic ring-like structure are clearly visible (Fig. 8H). Using the information of the extracted localizations for Fourier ring correlation method based estimation of resolution, we determined that with this particular sample the resolution of our microscopy system reached 34.7 nm. Finally, to validate that miEye can also achieve SR when the microscope is set to a SMF-based excitation scheme, we performed imaging of commercially available DNA PAINT sample. After reconstructing the SR image and correcting for the lateral sample drift, we were able to resolve individual DNA nanostructures with their characteristic pattern of three spots spaced 80 nm apart from each other. In this case, Fourier ring correlation-based quantitative assessment of resolution achieved by our microscopy system indicated a value of ~ 14 nm.

2.2.Characterization of droplets formed by S100A9 and α -synuclein proteins

In the study described above, we have successfully demonstrated that our newly developed microscope miEye exhibits a high performance in single color fluorescence imaging. Still, it remains unclear whether miEye maintains such an effective performance in simultaneous two-color fluorescence imaging *in vitro*. This could be even more complicated when the sample

exhibits a high level of heterogeneity in terms of the morphological and fluorescence properties of the objects of interest. Protein LLPS is a process during which biomolecules condense into high-concentration membraneless particles^{320,415}. Protein droplet-like molecular condensates formed as a final

product of LLPS typically are highly heterogeneous in terms of the aforementioned aspects³¹³, therefore they are well suited for demonstrating the extended capabilities of our new imaging system.

Here, we investigated the cross interaction between neurodegenerative disease-related α -syn³⁷ and pro-inflammatory S100A9³⁸, mainly focusing on the formation of protein condensates. Using two-color FM, we performed imaging of samples containing unlabeled α -syn mixed with enhanced green fluorescent protein (eGFP)-labeled α -syn and S100A9-mCherry and quantified their colocalization (Fig. 9). We saw that most small droplets consisted primarily of S100A9-mCherry, whereas larger droplets contained either α -syn-eGFP alone or both fluorescently labeled proteins. After performing a statistical analysis of the particles ($n = 2581$), we determined that $44.7 \pm 3.2\%$ of the droplets were heterotypic. Out of the remaining condensates, $43.1 \pm 2.5\%$ of the droplets contained only S100A9-mCherry, while $12.2 \pm 2.5\%$ of the droplets consisted solely of α -syn-eGFP. To conclude, in this study, we demonstrated that α -syn and S100A9 proteins can form both homotypic and heterotypic droplets under high molecular crowding conditions.

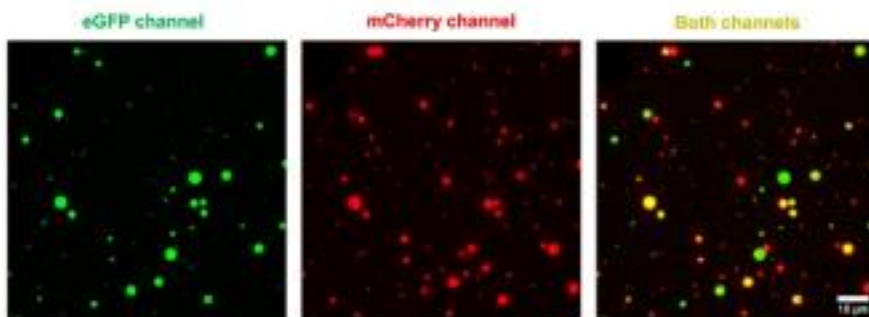


Figure 9. Total internal reflection fluorescence microscopy images of α -synuclein and α -synuclein-eGFP or S100A9-mCherry droplets deposited on the glass coverslip surface. Reproduced and modified from Veiveris *et al.*⁴¹⁶.

2.3. Mechanisms of BfiI restriction endonuclease interaction with DNA

In this work, we developed a novel smFRET-based assay which we then used to investigate *in cis* and *in trans* interactions of the active site mutant K107A of BfiI with the DNA. During *in cis* interaction, when both BfiI targets

are positioned on the same DNA molecule (Fig. 10A), the FRET efficiency distribution was comprised of two populations that showed a good fit when using the double-Gaussian function (Fig. 10B). The first population was distinguished by a peak located at ~ 0.1 FRET efficiency, which we attributed to the non-looped DNA state. The center of the second population was shifted

only by ~ 0.02 units to the higher end of FRET efficiency, and we associated this population with the looped DNA state, in which the DNA should adopt a “U”-shaped conformation.

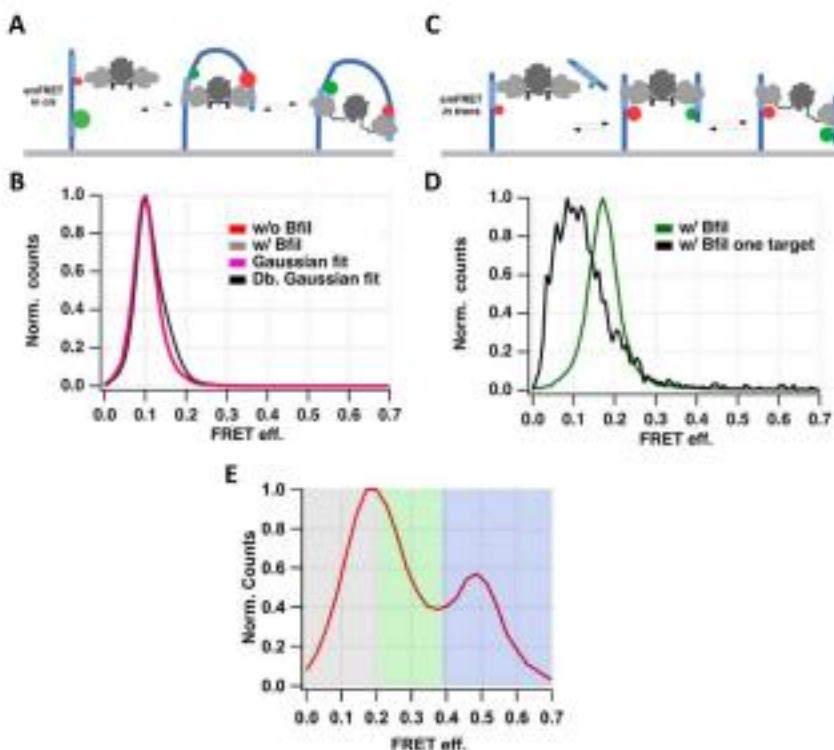


Figure 10. Single-molecule FRET (smFRET) assays for BfiI-DNA *in cis* and *in trans* interactions’ studies. **(A)** Schematic illustration of the smFRET-based DNA looping assay in which BfiI interacts with its two target sites *in cis*. Donor fluorophore is depicted in green, acceptor fluorophore in red, BfiI protein in light and dark grey, double-stranded DNA molecule in dark blue, BfiI target sites in light blue. **(B)** Histograms of FRET efficiencies of selected single-molecule signals with and without BfiI protein in the smFRET *in cis* assay. **(C)** Schematic illustration of the smFRET assay for probing BfiI-DNA interactions *in trans*. **(D)** Histograms of FRET efficiencies of the detected binding events with both BfiI targets present and with only one BfiI target site in the smFRET *in trans* assay. **(E)** Distribution of FRET efficiencies measured in the smFRET *in trans* assay with the cross-linked catalytically inactive BfiI variant. Grey – non-cross-linked population of BfiI, green – partially cross-linked population of BfiI, blue – fully cross-linked population of BfiI. Adapted from Ivanovaite *et al.*⁴¹⁷.

Next, we altered the arrangement of BfiI recognition sites – we kept one BfiI target on the surface-immobilized DNA molecule, whereas we moved the other target to the separate untethered short dsDNA oligonucleotide (Fig. 10C). Because of this rational design, we were able to isolate and observe the

interaction *in trans* over an extended period since the possibility for DNA looping to occur was completely abolished. After visually analyzing and manually selecting the proper SM signals, we determined that approximately 20% of all acquired signals represented BfiI binding events. The graph depicting the distribution of FRET efficiency exhibited a distinct peak centered at a FRET efficiency value of ~ 0.18 (Fig. 10D). We concluded that this conformational state characteristic to BfiI is most likely the same as in the assay of *in cis* interaction, where it showed a lower FRET efficiency (~ 0.12) of the formed BfiI-DNA complex, possibly due to DNA looping effect on the DNA bending at the fluorophore attachment sites. In both interaction cases – *in cis* and *in trans* – the measured FRET efficiencies were lower than the theoretical value (~ 0.4), which we expected to obtain when the DNA-bound BfiI exists in the closed state. Such result suggested that, upon double on target binding, BfiI REase could be changing its conformation from the closed to the open state so quickly that we were not able to detect this higher FRET efficiency-exhibiting closed state due to the limited temporal resolution of the immobilization-based smFRET technique paired with TIRF microscopy⁴¹⁸.

To eliminate this conformational shift intrinsic to BfiI in our experiments, we decided to repeat the smFRET measurements this time using a catalytically inactive cross-linked variant of BfiI, named BfiI-SS¹⁵⁹. In this BfiI version, both DNA-binding domains are coupled to the cleavage domain via a short linker, thus “locking” the protein in a closed state. Such cross-linking previously performed on BfiI, however, was not fully effective and resulted in $\sim 70\%$ of completely cross-linked REase, $\sim 15\%$ of partially cross-linked BfiI (here only one random DNA-binding domain is linked to the cleavage domain) and $\sim 15\%$ of non-cross-linked enzyme. When probing DNA-BfiI-SS interaction *in trans*, we observed stable binding events with three different levels of FRET efficiency (Fig. 10E): ~ 0.2 (characteristic to non-cross-linked BfiI), ~ 0.35 (characteristic to partially cross-linked protein) and ~ 0.5 (characteristic to fully cross-linked REase). Based on this experimental approach, we demonstrated that, as soon as the non-cross-linked version of BfiI binds both of its targets during *in cis* or *in trans* interaction with DNA, it rapidly transitions into an open state. In addition, we characterized BfiI-DNA

interaction dynamics by correlating the average FRET efficiency (i.e., protein conformation) with the duration of binding events (i.e., stability of the complex). We found that BfiI exhibited short (< 0.5 s) off-target binding events, while this restriction enzyme remained bound to both of its cognate DNA sequences for longer (4.9 s).

2.4. Target search mechanisms of Type I-F Cascade complex in the

context of anti-CRISPR proteins

In this work, we studied the potential of ten AcrIF families to inhibit the Type I-F CRISPR-Cas system from *Aggregatibacter actinomycetemcomitans* D7S-1. We found that the activity of Aa-CRISPR-Cas immunity *in vivo* was blocked by only two AcrIF family proteins – AcrIF6 and AcrIF9. After performing the biochemical analysis, we demonstrated that both AcrIF6 and AcrIF9 directly interact with the Aa-Cascade complex and prevent it from recognizing its target DNA. Additionally, AcrIF9 exhibited a unique feature that promotes the non-specific binding of Aa-Cascade to DNA.

To study the target search dynamics of Aa-Cascade and Aa-Cascade AcrIF9 at the SM level in real time, we employed a double-tethered Soft DNA Curtains platform, in which both ends of the flow-stretched individual bacteriophage λ DNA molecules (48.5 kbp) are immobilized on the surface printed traptavidin lines^{161,412}. While probing the interactions between Aa Cascade and DNA, we observed two populations of binding events: short lasting (<20 s) events, which were distributed over the entire length of the λ DNA molecule, and long-lasting (>20 s) events, which were mainly located at the target site (Fig. 11A). Aa-Cascade in complex with the AcrIF9 protein was predominantly binding to DNA in a sequence-non-specific manner (Fig. 11B). In this case, the long binding events became spread across the entire length of the λ DNA molecule, and their population percentage surged from ~11 to ~30%. With the AcrIF9 protein bound to the Aa-Cascade, the frequency of on-target binding decreased from ~18 to ~8%, while the amount of off-target binding events increased from ~31 to 42%. Lastly, upon adding AcrIF9, the mean dwell-time of Aa-Cascade-AcrIF9 complex on the off-target DNA sites increased from ~7.8 to 18.6 s.

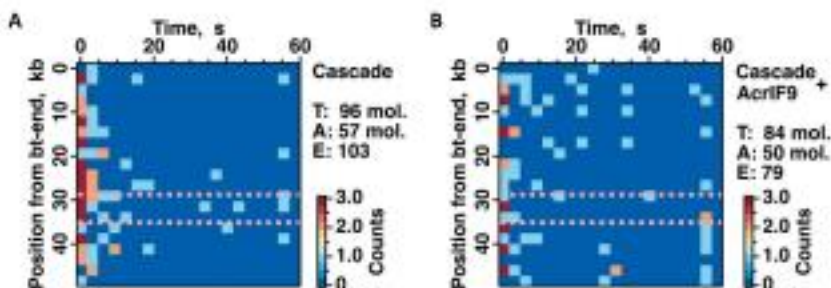


Figure 11. Binding position and dwell time of Type I-F Cascade surveillance complex from *Aggregatibacter actinomycetemcomitans* D7S-1 (Aa) and Aa-Cascade-AcrIF9 measured using double-tethered Soft DNA Curtains. **(A)** 2D histogram plot showing the binding location vs. binding duration for the Aa-Cascade complex. **(B)** 2D histogram plot indicating the binding position vs. dwell time for the pre-assembled Aa-Cascade-AcrIF9 complex. T – total number of examined DNA molecules, A – DNA molecules with at least one respective Cascade complex binding event, E – total number of individual binding events of the respective Cascade complex, bt-end – biotinylated end of λ DNA. Reproduced and modified from Kupcinskaite *et al.*⁴¹⁹.

These results suggest that the prolonged dwell-time of Aa-Cascade AcrIF9 at the off-target DNA sites is the most probable factor driving the non specific binding of Aa-Cascade-AcrIF9 complex to DNA, while the increased off-target binding frequency is of a rather small contribution. Initial biochemical experiments in bulk and complementary AFM measurements also supported this hypothesis. Based on all the collected results, we concluded that AcrIF9 protein inhibits the target-specific recognition of Aa Cascade and promotes its non-specific interactions with the DNA, forcing this Type I-F CRISPR-Cas effector complex to spend more time at the off-target DNA sites.

2.5.SPARDA defense system proteins' mechanisms of action

In this paper, we conducted functional, biochemical, structural and SM level studies of two so far uncharacterized homologous SPARDA defense systems – XauSPARDA and SPARDA complex from the thermophilic *Enhydrobacter aerosaccus* (EaeSPARDA). Here, we proposed a comprehensive molecular mechanism of action of XauSPARDA. We showed that *in vivo* both XauSPARDA and EaeSPARDA grant protection against foreign plasmids by catalyzing the non-specific degradation of DNA and thereby inducing cell death. Our results revealed that the activation of SPARDA proteins occurs through their filament assembly.

By combining SMFM with *in vitro* DNA flow-stretch assay, we visualized SPARDA filamentation on the long λ DNA molecules immobilized

on the surface by their single end (Fig. 12A). Using a relatively slow continuous buffer flow, we demonstrated that in a Ca^{2+} -enriched environment the activated WT XauSPARDA protein, i.e., in complex with gRNA and a fully complementary (17 bp) ssDNA target, was able to compact individual λ DNA molecules into separate bundles (Fig. 12B). Compaction of dsDNA was reversible – this required a higher ionic strength buffer to be injected in the flow cell and incubated there for an extended period. Upon replacing the Ca^{2+} ions with the Mg^{2+} ones in the flow cell, we observed close to a complete cleavage of λ DNA compacted by the activated XauSPARDA. On the contrary, we saw no effect on the single end-immobilized λ DNA molecules when using the assembled XauSPARDA complex comprised of gRNA and a low-complementarity (8 bp) ssDNA target.

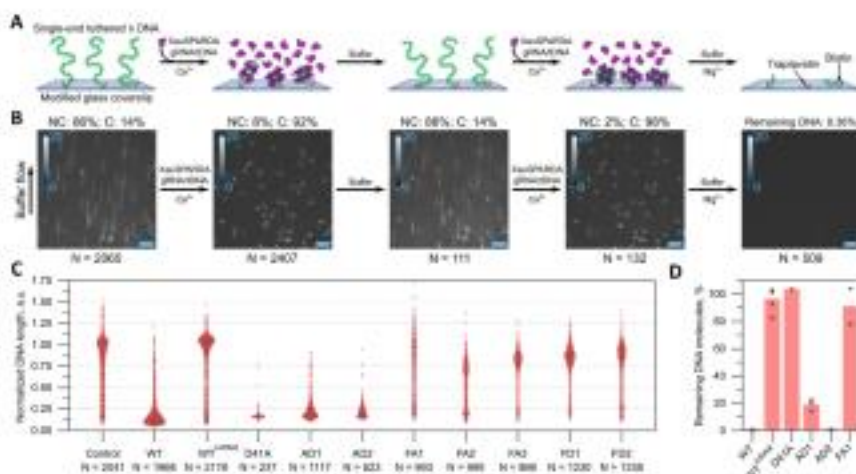


Figure 12. SPARDA mechanism of action revealed by a direct real-time single molecule total internal reflection fluorescence (TIRF) microscopy. **(A)** Schematic illustration of DNA flow-stretch assay designed for probing SPARDA and double stranded DNA individual interaction dynamics. **(B)** TIRF microscopy images showing the reversible single end-immobilized λ DNA compaction process characteristic of the wild-type XauSPARDA-guide RNA (gRNA)/target DNA (tDNA) complex in the presence of Ca^{2+} and the subsequent cleavage of λ DNA in the presence of Mg^{2+} . N – total number of analyzed λ DNA molecules, NC – mean percent value of non-compacted DNA, C – mean percent value of compacted DNA. Scale bar (light blue) is 10 μm . Calibration bar (upper left) indicates minimum and maximum grey scale values. **(C)** Individual λ DNA molecule length distribution obtained after incubation with the complexes of different XauSPARDA protein variants and the complementary gRNA/tDNA duplex in the presence of Ca^{2+} . Control – before incubation, N – total number of λ DNA molecules analyzed, λ ctDNA – low complementarity DNA target. **(D)** Cleavage activity of wild-type XauSPARDA and

various XauSPARDA mutants in the presence of Mg^{2+} . Bars indicate the average percentage, and dots represent the separate values of each imaged surface position. Reproduced and modified from Jurgelaitis *et al.*⁴²⁰.

Here we also examined various XauSPARDA mutants (Fig. 12C) at the SM level and found that: (1) the active site mutant XauD41A specifically

compacted the surface-immobilized λ DNA but did not cleave it (Fig. 12D); (2) apo dimerization mutants AD1 and AD2 exhibited properties analogous to those of the WT XauSPARDA; (3) filamentation by Ago mutants FA1, FA2, FA3 and filamentation by DREN mutants FD1 and FD2 had no effect on λ DNA even when their respective complexes contained a fully complementary ssDNA target. Finally, we demonstrated that the filament formation is also characteristic of other homologous SPARDA systems – EaeSPARDA, NbaSPARDA and SPARDA complex from *Thermocrispum municipale* (TmuSPARDA) – whose proteins, when activated with a fully complementary ssDNA target, were also able to efficiently compact and then cleave the individual λ DNA molecules immobilized on the surface by their single end.

3. DISCUSSION

The studies presented in this dissertation can be divided into three parts: (1) development and validation of a fluorescence imaging system miEye, (2) demonstration of the expanded capabilities of miEye and (3) elucidation of the mechanisms by which proteins from three divergent bacterial defense systems interact with NA. Firstly, we addressed an issue related to the current lack of versatile, easily customizable and at the same time user-friendly high resolution widefield fluorescence microscopes both on the commercial market and in the scientific community³¹. For this purpose, we have developed an integrated open-source solution named miEye. Compared to the existing similar microscopy platforms, the fundamental advantage of the miEye microscope is its full modularity. Our microscopy system is equipped with a custom-made laser engine (405, 488, 520 and 638 nm wavelength laser diodes), has two easily interchangeable schemes of excitation (SMF- or MMF based), as well as modifiable and expandable emission path. All of this enables a flexible imaging experience for the user with the possibility to tailor the microscopy setup according to one's needs. Thus, our bench-top optical microscopy framework can be used for standard widefield fluorescence imaging, SMFM and SR-requiring measurements. High stability is the other strength of miEye, as the microscope has an IR laser-based autofocus feature dedicated for actively maintaining the stabilization of the specimen movement along the Z-axis during imaging. The miEye experiences lateral sample drift of less than 7 nm/min, which is another indication of the system robustness. The total cost of hardware needed to build such system reaches at around 50 thousand Euro, which is way below the typical price range (>100 thousand Euro) of other available open-source microscopes with a similar set of capabilities^{32,34,36}.

Along with the miEye's hardware, we also developed a Python-based open-source software called microEye³³. It is a user-centric solution dedicated for controlling the microscope, acquiring images and analyzing the collected data simultaneously during the same experiment. These three functional modules – Control module, Acquisition module and Tiff viewer module – are integrated into a single environment of microEye, hereby ensuring its stable and reliable performance, which can be sometimes missing in other stand alone software packages of a similar purpose, e.g., µManager³³⁴. To determine the resolution of miEye microscope, we performed dSTORM-based imaging of NPCs, which were previously reported to be a well suited reference standard for quantitative SR microscopy⁴²¹, in fixed U2OS cells, and as well imaged a commercial DNA-PAINT sample *in vitro* for additional validation.

validated the SR capabilities of miEye, demonstrating that this microscopy platform can be efficiently used to study various biomolecules and their interactions at the nanoscale.

In this study, we did not explore live-cell imaging possibilities of our microscopy system. Living cells are far more complex biological system than virtually any artificial *in vitro* sample²⁶ and, for example, mammalian cells usually require at least a proper ambient temperature during the longer measurements in order to maintain their viability⁴²². Thus, it remains to be tested whether miEye can be effectively used for *in vivo* FM applications. In addition, here we performed only 2D imaging of the selected specimens. Since the emission path of our microscope is fully adaptable, the miEye, in principle, should be also suitable for super-resolution imaging in 3D, which is yet to be demonstrated. This imaging modality is typically implemented through introducing astigmatism into the emission optical path⁴²³ or performing more advanced insertable phase mask-based PSF engineering that converts the regular PSF shape into the form resembling, e.g., a double helix⁴²⁴ or tetrapod⁴²⁵.

The second part of this dissertation was dedicated to demonstrating the additional functionalities of the miEye microscopy framework developed here. We aimed to perform simultaneous two-color fluorescence imaging of an *in vitro* sample that differs from most of our regular specimens, containing protein-DNA complexes or being involved in SMLM measurements, in its properties and complexity. We anticipated that the membraneless droplet-like formations originating during the LLPS process should be well-suited for this purpose, where they could serve as a model object exhibiting high heterogeneity³¹³. We employed the miEye microscope for studying the cross interaction between the pro-inflammatory S100A9³⁸ and the neurodegenerative disease-associated α -syn³⁷, as well as their ability to form heterotypic droplets under LLPS-inducing conditions. We found that fluorescently labeled α -syn and S100A9 proteins formed heterotypic droplets to a fairly high degree. We determined that the group of homotypic droplets was dominated by S100A9-mCherry particles, which were characterized by small size, whereas there were few, mainly large condensates comprised solely from α -syn-eGFP. Along with such obtained results dictated by the cross-interaction taking place between α -syn and S100A9 proteins, here we also validated the two-color imaging capabilities of our home-built widefield fluorescence optical microscopy system by effectively visualizing and quantitatively characterizing the highly heterogeneous biomacromolecular products of LLPS.