CAS12A IS GOVERNED IN ITS ACTIVITY AND CONFORMATIONAL TRANSITIONS BY THE TWO INTERPLAYING KEY ELEMENTS HELIX 1 AND BRIDGE HELIX

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)
der Fakultät für Biologie und Vorklinische Medizin
der Universität Regensburg

vorgelegt von

Elisabeth Johanna Viera Wörle

aus
München

Regensburg im Jahr 2022

Das Promotionsgesuch wurde eingereicht am: 10.06.2022

Die Arbeit wurde angeleitet von: Prof. Dr. Dina Grohmann

Unterschrift: ....

Elisabeth Wörle

The presented thesis was conducted between November 2018 and June 2022 at the chair for Microbiology at the Institute for Biochemistry, Genetics, and Microbiology at the Faculty of Biology and Preclinical Medicine and was instructed by Prof. Dr. Dina Grohmann.

Parts of this thesis have been published in the following articles:

Wörle E., Jakob L., Schmidbauer A., Zinner G., Grohmann D. (2021) ‘Decoupling the bridge helix of Cas12a results in a reduced trimming activity, increased mismatch sensitivity and impaired conformational transitions’, *Nucleic Acids Research*, 49(9), pp. 5278–5293. doi: 10.1093/nar/gkab286

Wörle E., Newman A., Burgio G., Grohmann D. (2022) ‘Allosteric activation of CRISPR-Cas12a requires the concerted movement of the bridge helix and helix 1 of the RuvC II domain’, *bioRxiv.* doi https://doi.org/10.1101/2022.03.15.484427
Currently under revision at *Nucleic Acids Research*.

TABLE OF CONTENTS

1. **Abstract** **5**
  1.1 Abstract (English) 5
  1.2 Zusammenfassung (deutsch) 6

2. **Introduction** **7**
  2.1 Innate prokaryotic antiviral defense systems 8
    2.1.1 Inhibition of adsorption 8
    2.1.2 Blockage of DNA injection 8
    2.1.3 Restriction modification 9
    2.1.4 Abortive infection 10
    2.1.5 Prokaryotic Argonaute proteins 10
    2.1.6 Chemical defense 11
  2.2 CRISPR-Cas – the prokaryotic adaptive immune system 11
    2.2.1 The three stages of immune response 13
      2.2.1.1 Adaptation 13
      2.2.1.2 Expression and maturation 14
      2.2.1.3 Interference 14
    2.2.2 Classification 16
    2.2.3 Anti-CRISPR systems 18
    2.2.4 Applications of CRISPR 18
      2.2.4.1 Genome editing 19
      2.2.4.2 Plant biotechnology 20
      2.2.4.3 Therapeutics and disease management 20
      2.2.4.4 Disease detection 21
    2.2.5 Cas12a 22
  2.3 Objective 26

3. **Materials** **27**
  3.1 Chemicals 27
  3.2 Equipment 29
  3.3 Consumables 31
  3.4 Software 32
  3.5 Buffers and Solutions 33
    3.5.1 Buffers 33
    3.5.2 Solutions 34
    3.5.3 Culture media 35
    3.5.4 Polyacrylamide gel electrophoresis 35

3.6 Molecular biology kit systems 38
3.7 Electrophoresis ladders 38
3.8 Enzymes 39
3.9 Bacterial strains 39
3.10 Vectors and plasmids 40
3.11 Oligonucleotides 40
3.11.1 Primers for cloning reactions 40
3.11.2 Primers for site directed mutagenesis 41
3.11.3 Oligonucleotides for analysis 43
3.12 Fluorescent Dyes 44

4. Methods 45
4.1 Microbiological methods 45
4.1.1 Preparation of equipment and solutions 45
4.1.2 Cell culture and storage of *E. coli* strains 45
4.1.3 Preparation of chemically competent *E. coli* cells 45
4.1.4 Chemical transformation of *E. coli* cells 46
4.1.5 Disposal of microorganisms 46
4.2 Molecular biological methods 46
4.2.1 Nucleic acid concentration determination by absorption spectroscopy 46
4.2.2 Polymerase chain reaction 47
4.2.3 DNA cloning 48
4.2.3.1 DNA restriction 49
4.2.3.2 DNA ligation 49
4.2.4 PCR-mediated site-specific mutagenesis 50
4.2.5 Isolation and purification of plasmid DNA from *E. coli* 51
4.2.6 Agarose gel electrophoresis 52
4.2.7 DNA sequencing 52
4.2.8 RNA *in vitro* transcription 53
4.3 Protein biochemical methods 54
4.3.1 Recombinant expression of FnCas12a variants in *E. coli* 54
4.3.2 Recombinant expression of AzF-modified FnCas12a variants in *E. coli* 54
4.3.3 Cell lysis and protein extraction 55
4.3.4 Glutathione affinity chromatography 56
4.3.5 Heparin affinity chromatography 57
4.3.6 Dialysis of protein solutions 58
4.3.7 Concentrating of protein solutions 58
4.3.8 SDS polyacrylamide gel electrophoresis 58
4.3.9 Labeling of AzF-modified FnCas12a variants with fluorescent dyes 59

4.4 Analytical methods 60
4.4.1 Determination of protein concentration 60
4.4.2 Circular dichroism spectroscopy 61
4.4.3 Protein Thermal Shift melting curves 61
4.4.4 Preparation of double-stranded target oligonucleotides 62
4.4.5 Electrophoretic mobility shift assay 63
4.4.5.1 FnCas12a binding to RNA 63
4.4.5.2 FnCas12a binding to DNA 63
4.4.6 DNA *cis*-cleavage assays 64
4.4.6.1 Plasmid cleavage assay 64
4.4.6.2 Short DNA target cleavage assay 64
4.4.7 *Trans*-cleavage assay using single-stranded DNA....................65
4.4.8 Mass photometry 66
4.4.9 Structure prediction with AlphaFold2 67
4.5 Fluorescence-based single-molecule spectroscopy 68
4.5.1 Förster resonance energy transfer....................68
4.5.2 Confocal single-molecule fluorescence microscopy 71
4.5.3 Confocal single-molecule FRET measurements 73
4.5.3.1 Sample preparation and measurement 73
4.5.3.2 Data Analysis of confocal single-molecule FRET experiments....................74
4.5.3.3 Determination of correction factors 75
4.5.3.4 Calculation of FRET efficiency histograms....................76
4.5.3.5 Burst variance analysis 76

5. Results 77
5.1 Characterization of WT FnCas12a....................77
5.1.1 Biochemical characterization of WT FnCas12a....................77
5.1.2 Conformational transitions of an RNA/DNA hybrid 84
5.1.3 Conformational transitions of FnCas12a 84
5.2 Collaborational analysis of variants to alter the target strand affinity 92
5.3 Characterization of FnCas12a bridge helix variants 95
5.3.1 Biochemical characterization of FnCas12a bridge helix variants....................96
5.3.2 Confocal single-molecule FRET analysis of FnCas12a bridge helix variants 105
5.4 Characterization of FnCas12a helix 1 variants....................108
5.4.1 Biochemical characterization of FnCas12a helix 1 variants....................109
5.4.2 Comparison of cleavage efficiencies of all FnCas12a variants....................119
5.4.3 Confocal single-molecule FRET analysis of FnCas12a helix 1 variants 120

6. Discussion 122
  6.1 Conformational changes of Cas12a throughout its activity cycle 122
  6.2 Bridge helix decoupling reduces trimming activity, increases mismatch sensitivity, and impairs conformational transitions 125
  6.3 Helix 1 is the primary stabilizing and connecting structural element in Cas12a .. 131
  6.4 Conclusion and outlook 134

7. References 135

List of figures 149

List of tables 152

List of abbreviations 154

Appendix 158
  I. Vector and plasmid maps 158
  II. Protein sequences 160
    II.i. Nucleotide sequence 160
    II.ii. Amino acid sequence 162
    II.iii. Multiple sequence alignment of Cas12a variants 163
  III. Extended data 166
    III.i. Protein purification 166
    III.ii. Single-molecule FRET measurements on a doubly labeled RNA/DNA hybrid 167
    III.iii. Single-molecule FRET measurements on doubly labeled WT FnCas12a 168
    III.iv. Single-molecule FRET measurements on doubly labeled FnCas12a HKK variants 172
    III.v. Single-molecule FRET measurements on doubly labeled dCas12a 174
    III.vi. Single-molecule FRET measurements on doubly labeled FnCas12a variants to alter the target strand affinity 176
    III.vii. Single-molecule FRET measurements on doubly labeled FnCas12a bridge helix variants 181
    III.viii. Single-molecule FRET measurements on doubly labeled FnCas12a helix 1 variants 188

Acknowledgement 193

1. ABSTRACT

1.1 Abstract (English)

The CRISPR-Cas system (clustered regularly interspaced short palindromic repeats and CRISPR associated proteins) is a widespread and versatile prokaryotic immune system, that evolved to sequence-dependently recognize and cleave foreign nucleic acids. The single-effector nuclease of the type V-A CRISPR-Cas system, Cas12a, binds and cleaves double-stranded target DNA with the aid of a CRISPR-RNA (crRNA). Similar to Cas9, Cas12a was increasingly used in genome editing applications during the past years. Cas12a comprises a bilobal structure with the REC and the Nuc lobe connected by the bridge helix (BH) and helix 1 of the REC2 domain. These helices perform a tandem movement upon the transition from the binary to the ternary complex.

In this thesis, Cas12a from *Francisella novicida* (Fn) was analyzed utilizing several biochemical and smFRET (single-molecule Förster Resonance Energy Transfer) assays to elucidate its conformational flexibility and transitions when binding to crRNA and target DNA. The smFRET measurements revealed the hitherto unknown open conformational state of the apo form of Cas12a that is in equilibrium with a closed state of the protein. The binding of the crRNA leads to a shift to the closed conformation. Binding of target DNA requires the re-opening of Cas12a, which consequently adopts a semi-closed conformation in the ternary complex.

In a mutational analysis, the roles of the BH and helix 1 on the enzymatic activity and conformational flexibility of FnCas12a were investigated. Thereby, the importance of the BH for the trimming activity of Cas12a at the non-target strand of the DNA as well as its influence on the mismatch sensitivity was revealed. This resulted in two variants with improved mismatch recognition and cleavage accuracy in *in vitro* assays. smFRET measurements showed that BH variants preferentially adopt the open conformation in the apo state and that the transition to the closed conformation is inefficient for the binary complex. The semi-closed conformation of the ternary complex, however, is readily adopted even if the BH is deleted in its entirety. In summary, it was shown that the BH impacts the catalytic activity and conformational transitions of FnCas12a.

Analyses on helix 1 variants revealed that this is the predominant structural element linking the REC and the Nuc lobe conferring stability to the whole protein. Deletion of helix 1 leads to a conformational collapse of FnCas12a and the closure of the protein is no longer possible. Disturbance of the concerted tandem movement of BH and helix 1 drastically decreases the cleavage activity of FnCas12a but does not affect the conformational transitions. The structural remodeling of the helices upon DNA binding is linked to the closed-to-open movement of the ‘lid’, a loop that closes the active site in the binary complex and opens if DNA is bound. Therefore, the tandem movement of BH and helix 1 is involved in the allosteric activation of the active site.

1.2 Zusammenfassung (deutsch)

Das CRISPR-Cas-System (clustered regularly interspaced short palindromic repeats und CRISPR associated proteins) ist ein vielseitiges und weit verbreitetes prokaryotisches Abwehrsystem, das fremde Nukleinsäuren sequenzabhängig erkennt und spaltet. Die eigenständige Effektornuklease des CRISPR-Cas-Systems V-A, Cas12a, bindet und spaltet doppelsträngige Ziel-DNA mit Hilfe einer CRISPR-RNA (crRNA). Ähnlich wie Cas9 wurde auch Cas12a in den letzten Jahren zunehmend in Genom-Editierungsanwendungen eingesetzt. Cas12a weist eine zweiflüglige Struktur auf, wobei der REC- und der Nuc-Flügel durch die Bridge Helix (BH) und Helix 1 der REC2-Domäne verbunden sind. Diese Helices führen beim Übergang vom binären zum ternären Komplex eine Tandembewegung aus.

In dieser Arbeit wurde Cas12a von *Francisella novicida* (Fn) mit Hilfe verschiedenster biochemischer und biophysikalischer Experimente charakterisiert. Dies schloss Einzelmolekül-FRET (Förster Resonanz Energie Transfer) Messungen ein. Die FRET-Messungen deckten den bisher unbekannten offenen Zustand von Cas12a in seiner Apo-Form auf, der im Equilibrium mit einer geschlossenen Konformation des Proteins steht. Die Bindung der crRNA führt zu einem Übergang in die geschlossene Konformation. Bindung einer Ziel-DNA zusätzlich zur crRNA führt zu einer Öffnung von Cas12a, welches dadurch eine halbgeschlossene Konformation einnimmt.

In einer Mutationsanalyse wurden BH und Helix 1 auf ihre Rolle in der enzymatischen Aktivität und Konformationsflexibilität von FnCas12a untersucht. Die BH beeinflusst die Trimmaktivität von Cas12a am Nicht-Zielstrang der DNA und hat Auswirkungen auf die *Mismatch*-Empfindlichkeit des Enzyms. Durch die Mutationsstudie wurden zwei Cas12a Varianten mit verbesserter *Mismatch*-Erkennung und Spaltgenauigkeit identifiziert. Einzelmolekül-FRET Messungen zeigten, dass die BH-Varianten im Apo-Zustand die offene Konformation bevorzugen und dass hier der Übergang in die geschlossene Konformation des binären Komplexes ineffizient ist. Die halbgeschlossene Konformation kann jedoch selbst bei vollständiger Deletion der BH angenommen werden. Zusammenfassend wurde gezeigt, dass die BH die katalytische Aktivität und die konformationellen Zustände von Cas12a beeinflusst. Analysen von Helix 1-Varianten zeigten, dass diese Helix das strukturelle Element ist, das den REC- und den Nuc-Flügel verbindet und dem gesamten Protein Stabilität verleiht. Die Deletion von Helix 1 führt zu einem Zusammenbruch der konformationellen Flexibilität von FnCas12a und das Schließen des Proteins ist nicht mehr möglich. Eine Störung der konzertierten Tandembewegung von BH und Helix 1 verringert die Spaltungsaktivität von FnCas12a drastisch, ohne jedoch die einzelnen Konformationszustände zu beeinflussen. Es deutet darauf hin, dass die strukturellen Neuordnungen der Helices bei DNA-Bindung mit der geschlossen/offen-Bewegung des ‚Deckels‘ verbunden sind, einer Schleife, die das aktive Zentrum im binären Komplex verschließt und dieses öffnet, sobald DNA gebunden wird. Die Tandembewegung von BH und Helix 1 ist daher an der allosterischen Aktivierung des aktiven Zentrums beteiligt.

2. INTRODUCTION

The majority of the living world cannot be seen by our naked eye. Thousands of years passed until Antony van Leeuwenhoek developed a microscope with which he was able to see bacteria (Greek βακτήριον (baktérion), meaning 'staff, cane') for the first time in 1676 [1]. Almost another 300 years passed until Frederick Twort (1915) and Félix d'Hérelle (1917) discovered viruses (Latin vīrus, meaning 'poison') even though these are the most abundant biological entities in the world [2–4]. Viruses that infect and harm bacteria are known as 'bacteriophages' or short 'phages' (Greek φαγεῖν (phageín), meaning 'to devour'). With 10^{31} phages in the biosphere, they outnumber bacteria by a factor of 10 [5–7].

According to their infection strategy, phages can be classified into four categories [8]: (i) Lytic and non-temperate phages also known as virulent pages. After infection, they do not integrate into the host genome but lead to the production of new virus particles and lysis of the bacterial cell with subsequent release of daughter virions [8,9]. (ii) Chronic and non-temperate phages that do not integrate into the host genome lead to the production of new virus particles, and a continuous release of daughter virions over a long period without lysis of the host cell [8]. (iii) Lytic and temperate phages, these are either able to integrate into the bacterial chromosome, in this case called prophages, or are maintained extra-chromosomally in the bacterial cell. Upon the production of new viral particles and the release of daughter virions, the host cells are lysed [5,8]. (iv) Chronic and temperate phages are either able to integrate into the bacterial chromosome or are maintained extra-chromosomally in the bacterial cell. With the production of new virus particles, daughter virions are continuously released without lysis of the host cell [8,10].

To protect themselves from viral attacks, Bacteria and Archaea developed different defense systems against various phages that target different steps of the life cycle of a phage, which led to an evolutionary arms race between phages and their prokaryotic hosts [11–13]. Bacteria and Archaea evolved a variety of innate defense systems, many of them are still not discovered yet. Additionally, with the CRISPR-Cas system, they evolved an adaptive immune system with a memory that continuously grows upon the infection with a new phage.

2.1 Innate prokaryotic antiviral defense systems

2.1.1 Inhibition of adsorption

As the first step of the infection, phages adsorb to the bacterial cell surface by binding to receptors followed by the injection of their genome [13]. Attachment points are the lipopolysaccharide (LPS) layer, membrane proteins, pili, or flagella [11]. Prokaryotes can prevent adsorption by modifying the entry points through mutations or masking of susceptible membrane proteins with other proteins. The latter prevents recognition, but still retains the function of the protein [14,15]. Another possibility is the positive selection of prokaryotes that lose the genes for specific proteins that serve as entry points for phages from their genome or that evolved strategies to downregulate them [12,16,17]. Entry points can also be blocked and occluded by the overproduction of an extracellular matrix, like capsular polysaccharides, proteins, lipids, or extracellular DNA [18–21]. Specific binding to phage entry points can be achieved by the production of competitive inhibitors that outcompete the phages [22]. A strategy that reduces productive infections is decoying phages by the production of outer membrane vesicles (OMV) with the respective phage attachment proteins. The phages bind and infect the OMV without harming the bacterial cell [23,24]. By phase variation, the availability of entry points can be decreased through reversible changes in the regulation of gene and protein expression. Thereby, the prokaryotic cell surface can be altered [25–27].

In response to these prokaryotic strategies to overcome infection by inhibiting adsorption, phages themselves evolved counterstrategies like modifying their tail fibers to recognize new or altered proteins and attachment possibilities [28].

2.1.2 Blockage of DNA injection

Adsorption of the phage to the bacterial cell is followed by the injection of the viral genome into the host cell. Superinfection exclusion systems (Sie) prevent the entry or replication of phage DNA and are commonly phage or prophage encoded. By this, they protect bacteria from a secondary infection with closely related phages. Sie proteins are predicted to be either membrane-anchored or associated with other membrane components, but the underlying mechanisms of action are yet not well understood [29–31].

2.1.3 Restriction modification

Upon successful injection of the phage genome into a prokaryotic cell, the restriction modification system (RM) is a possibility to cleave invading DNA and to avoid damages caused by the phage. The RM system, which is considered the prokaryotic innate immune system, is the best characterized phage-resistance mechanism and is widely exploited in biotechnology to site-specifically cleave DNA [13]. It is a highly diverse and ubiquitous system that is present in 75% of all prokaryotes and about 90% of the Bacteria [32]. The two-component system is composed of a methyltransferase (MT) and a restriction endonuclease (RE). The MT methylates endogenous DNA on both strands at specific palindromic sequences of four, six, or eight base pairs, mostly at adenine or cytosine bases, and has a higher specificity to hemimethylated than to unmethylated DNA. The RE recognizes the same sequence motives as the respective MT, and cleaves unmethylated DNA, thereby assuring discrimination between self and non-self DNA [33–35]. Consequently, unmethylated foreign DNA will be cleaved by the RE, whereas the genomic DNA stays intact.

Phages developed different strategies to circumvent RM systems. For example, phages express their own MT upon host cell entry or activate a host MT leading to methylation of the phage DNA, and disguising of their genome from endogenous REs [35–39]. Another strategy is the introduction of point mutations at RM sites, leading to underrepresentation or absence of recognition sites in the phage genome which makes them less vulnerable for RM systems [35,36,40]. Some phages can inactivate REs by expressing Ocr proteins (overcome classical restriction) directly after infection. Ocr proteins mimic DNA and tightly bind REs with a 50-fold higher affinity than DNA [41–43]. To avoid recognition by an RE some phages use modified bases like uracil, hydroxymethyluracil, or hydroxymethylcytosine. As a response to this, type IV REs evolved that only cleave modified DNA [44–46]. Generally, MTs are more conserved than REs as the latter have to constantly adapt to mutations in the phage genome and undergo rapid evolution [47].

In recent years, many RM-like systems have been discovered by analyzing gene clusters around known defense genes, called defense islands. One of these systems is the BREX (bacteriophage exclusion) system, which consists of a gene cassette composed of six genes and includes among others a phosphatase (PglZ) and a methyltransferase (PglX). After DNA injection, BREX impedes phage replication without cleavage of the DNA. The exact mechanism of action still needs to be elucidated. Discrimination between self and non-self DNA is again achieved by methylation of the host DNA [48].

Another RM-like system associated with defense islands is DISARM (defense island system associated with restriction-modification). The system encodes for five genes including among others a protein with a helicase-related domain, a methyltransferase, and a phospholipase/nuclease D domain-containing protein. Although the mechanism of DISARM is not yet completely understood, it is known that phage DNA circularization and thereby replication is prevented, and that methylation of the host DNA is used to discriminate between self and non-self DNA [49].

2.1.4 Abortive infection

Abortive infection systems (Abi) are considered altruistic because this defense system does not protect an individual single cell, but instead a bacterial population [12,50]. In this post-infection defense mechanism, the host cell dies and hence only releases a few or even no phages at all [50]. Dormant proteins that are activated upon infection are responsible for the disruption of essential phage processes, such as replication, transcription, translation, or DNA packaging [11,18].

A subset of Abi systems are the toxin-antitoxin mechanisms (TA) which can include regulatory interactions of different molecules like protein-protein, protein-RNA, or RNA-RNA interactions [18]. Toxins are molecules targeting essential cellular processes, like replication, translation, or cell wall formation, and lead to bacterial dormancy or cell death [51]. In an equilibrium, antitoxins regulate the activity of toxins. After phage infection, the antitoxin promoter is repressed, transcription is terminated, or the antitoxin is degraded leading to the release of the toxin and dormancy or cell death [52,53]. A possible response of phages is to encode antitoxins themselves to circumvent the effect of toxins [54].

Other Abi strategies involve dCTP deaminase and dGTPase proteins that degrade the deoxynucleotides dCTP and dGTP. This, on the one hand, stops phage replication, but on the other hand, also leads to the death of the host cell [55,56].

Some bacterial genomes encode a phage-inducible chromosomal island (PICI) that can interfere with the reproduction of certain phages. After phage infection, the PICI is excised, circularized, and replicated. During phage assembly, mature phage particles are not loaded with phage DNA but instead loaded with PICI DNA. The host cell dies upon lysis by the phage, but the released phages are not able to infect further cells [57,58].

2.1.5 Prokaryotic Argonaute proteins

Argonaute (Ago) proteins are present in all domains of life and have a key function in RNA interference (RNAi) in eukaryotes [59]. In prokaryotes, they are often encoded in defense islands, in direct proximity to other defense systems, supporting their role of defending the cell against foreign nucleic acids and phages [60]. Many Ago proteins are nucleases mostly guided by 5’-phosphorylated nucleic acids, termed guides [61]. The guide can either be single-stranded (ss) DNA or RNA and helps the protein to identify complementary viral sequences that are subsequently cleaved by Ago [62–64]. Additionally, many prokaryotic Ago proteins do not show catalytic activity by themselves but are co-encoded with nucleases or other proteins that they form complexes with [60,65]. Target binding by the Ago protein subsequently unleashes the enzymatic activity of its interaction partner.

2.1.6 Chemical defense

Not all defense strategies of prokaryotes involve proteins. Chemical defense strategies employ bioactive secondary metabolites that target invading phages. These small molecules block phage genome replication, transcription, and circularization by intercalating into the DNA. So far, it is not clear how host DNA is discriminated from phage DNA. The small-sized metabolites can diffuse out of the cell and hamper phages before an infection took place and without the need for a cellular environment. Consequently, protection of a whole prokaryotic community can be achieved. Therefore, chemical defense has been proposed to represent an innate defense strategy [66].

With further research, especially focusing on defense islands, constantly new defense systems are discovered and described, but many still remain unknown [67].

2.2 CRISPR-Cas – the prokaryotic adaptive immune system

Besides the various defense systems that provide the same defense response every time they encounter a phage, there is also a defense system that is adaptive. This means that also prokaryotic organisms can build an immunological memory for efficient defense against reoccurring phage attacks. For a long time, it was not known that prokaryotes have a defense strategy that acts like an adaptive immune system. This defense is now known as CRISPR-Cas (clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins).
CRISPR sequences were first mentioned in 1987 by Ishino *et al.* as ‘unusual structures’ in the genome of *E. coli* K17 after *in silico* analysis. They found a set of five homologous repetitive sequences with 29 nt of length which were separated by spacers of 32 nt [68]. In 2002, these structures of repeats and spacers were named CRISPR after a joint decision by Mojica *et al.* and Jansen *et al.* [69,70]. In the same year, Tang *et al.* discovered that the CRISPR loci are transcribed into small RNAs [71] and Jansen *et al*. found that the *cas* genes are a gene family associated with the CRISPR loci [70]. In 2005, for the first time, three groups

almost simultaneously connected spacer sequences with sequences derived from phages or mobile genetic elements (MGE) [72–74]. One year later, in 2006, a computational analysis of CRISPR loci and their *cas* genes predicted a function similar to eukaryotic RNAi, involving Cas proteins equipped with small CRISPR RNAs (crRNA) to detect complementary sequences. This finding led to the assumption that CRISPR might be involved in prokaryotic defense strategies [75]. The demonstration of this hypothesis was published in 2007 by Barrangou *et al.* even surpassing it by showing that immunity is adaptive and new spacers are acquired during infection, building a memory of the invaders and protecting the host from subsequent attacks by the same phage [76]. In the years after, it was shown that immunity crucially relies on the crRNAs [77] and that the targeted molecule is DNA rather than RNA as previously assumed for a mechanism similar to RNAi [78].

Until today, CRISPR sequences were detected in 42.3% of sequenced bacterial genomes and 85.2% of archaeal genomes. In hyperthermophilic archaea, even 96.7% of all sequenced organisms reveal a CRISPR locus [79]. A CRISPR locus consists of a CRISPR array, containing a variety of spacers separated by short direct repeats with identical sequences with an average length of 32 bp (21–47 bp, **Figure 1**). The spacers in one array have a fixed length with averages between 21 and 72 bp and are highly variable in their sequence as they originate from viral sequences [80]. Adjacent to the CRISPR array is the leader sequence, an often AT-rich sequence of up to 500 bp in length, that is involved in the incorporation of new spacers [70,81,82]. The CRISPR-associated genes (*cas*) are a cluster of genes coding for different Cas proteins that are responsible for the acquisition of new spacers and the interference with foreign nucleic acids. Because of their binding affinity to nucleic acids, it was formerly thought that these proteins could be involved in DNA repair [83]. The *cas* genes in a CRISPR locus vary in their orientation and order. Based on sequence similarities, the 93 identified genes are grouped into 35 families which are part of different types and subtypes of the CRISPR system (2.2.2) [84]. The number of CRISPR loci in one genome can vary from one to twenty containing a few to hundreds of repeat-spacer sequences [85].

2.2.1 The three stages of immune response

The adaptive immunity provided by the CRISPR-Cas system can be divided into three mechanistic steps: (i) adaptation and spacer acquisition, (ii) expression and maturation, and (iii) the interference with foreign invading nucleic acids (**Figure 1**).

2.2.1.1 Adaptation

During the adaptation phase, protospacers acquired from invading nucleic acid sequences are integrated as new spacers into the CRISPR locus of the host genome. This process is mediated by the adaptation module containing the heterohexameric Cas1-Cas2 complex as a core component. Cas1 and Cas2 are usually encoded in the same operon together with other Cas proteins and are universally conserved in all CRISPR subtypes [86–89]. Cas1 is an integrase and endonuclease displaying a unique α-helical fold [90–92]. Cas2 shows various cleavage activities against RNA [93] and DNA [94] substrates [95]. Studies on the adaptation module of *Campylobacter jejuni* suggest an involvement of the Cas4 protein in the adaptation process [96]. Additionally, other Cas and non-Cas proteins can aid in the process like the integration host factor (IHF) [97]. In the heterohexameric $(Cas1_2\text{-}Cas2)_2$ complex, the two Cas1 proteins of the homodimers adopt different conformations [86,87,90–92,98]. This suggests that the two Cas1 protomers fulfill different biological functions. Cas1a cleaves the invading DNA, called protospacer, thereby generating a 3'-OH group, Cas1b is responsible for the interaction with Cas2 and the complex formation [90]. A Cas2 homodimer connects two Cas1 dimers building a binding surface for the protospacer DNA (**Figure 1A**). Thereby, Cas2 functions as a space holder and its main purpose is to measure the protospacer's length between the two Cas1 dimers [90]. Consequently, neither the intrinsic activity of Cas2 nor its catalytic residues are crucial for spacer acquisition [87]. After the recognition of new potential spacer sequences, and the acquisition by the Cas1-Cas2 complex through the initial cleavage steps by Cas1, the spacers are integrated into the host genome [99]. The integration of new spacers into the CRISPR locus proceeds at the junction between the leader sequence and the first repeat of the locus [100,101]. Through two nucleophilic attacks at the complementary strands, both on the 5'-end of the first repeat, the protospacer is inserted and becomes an additional spacer. The gapped duplex at the repeat sequences flanking the new spacer is repaired by a non-Cas polymerase and covalently rejoined by a non-Cas ligase, thereby leading to a duplication of the repeat [99,100]. In this manner, the CRISPR locus is ordered chronologically with the spacer most recently acquired in direct proximity to the leader [99].

Still, the adaptation phase remains the least investigated and least understood process of CRISPR interference.

2.2.1.2 Expression and maturation

During the second phase of CRISPR immunity, crRNA biogenesis, processing, and maturation take place. The CRISPR array is transcribed into a long precursor crRNA (pre-crRNA) containing all spacers and repeats (**Figure 1B**). The pre-crRNA is processed by cleavage within the repeat sequences [102] to yield the single mature crRNAs containing a single spacer flanked by parts of the repeat sequence [103]. Depending on the subtype of the CRISPR system, different proteins are involved in the maturation process. The cleavage step is either performed (i) by a subunit of a multi effector complex with endonuclease activity, typically Cas6 [104], (ii) by the bacterial RNase III with the aid of an additional tracrRNA (*trans*-activating CRISPR RNA) [105], or (iii) by large single-effector Cas proteins that contain a distinct catalytic site for crRNA processing [79].

2.2.1.3 Interference

The ultimate step, interference (**Figure 1C**), starts with the formation of the effector complex. Depending on the CRISPR subtype, this can either be a multiprotein complex or a single protein, both equipped with crRNA. The ribonucleoprotein complex recognizes invading sequences based on their complementarity to the crRNA through base-pairing with the spacer sequence. After recognition, a nuclease subunit cleaves and inactivates the foreign target. The discrimination between self and non-self sequences occurs via the detection of a protospacer adjacent motif (PAM), a 2–5 nt sequence upstream or downstream of the protospacer that is not present in the bacterial genome [74,89,106–109]. The complete mechanism of self/non-self discrimination is not yet entirely understood. Short PAM sequences of 2–3 nt are not specific enough for discrimination and can be found in almost every prokaryotic genome.

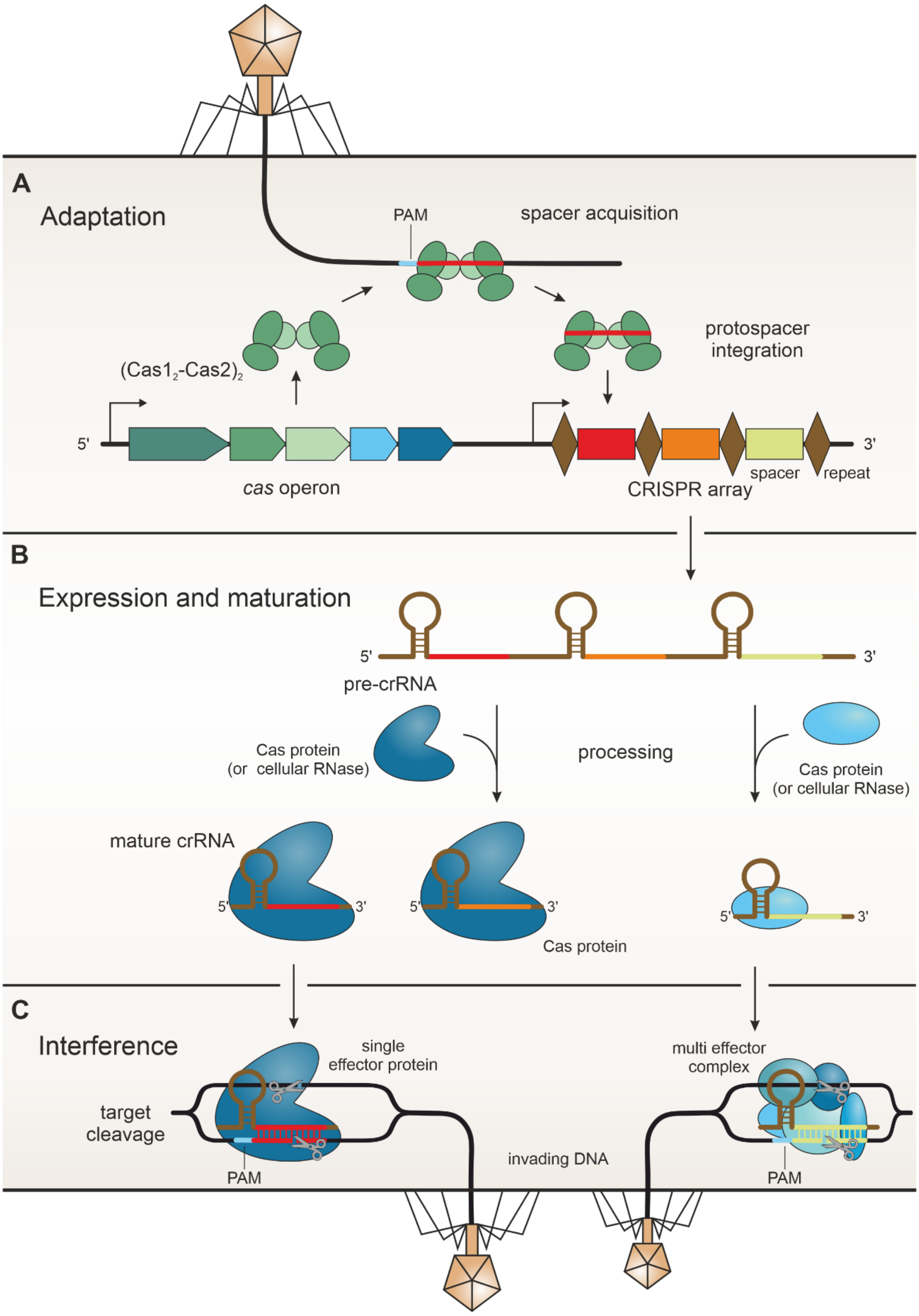


Figure 1: The three stages of CRISPR-Cas immune response.
(**A**) In the adaptation phase, the Cas1-Cas2 complex acquires new spacers from invading nucleic acids and integrates them into the CRISPR array (brown: repeats; red, orange, yellow: spacers). (**B**) During expression and maturation, the CRISPR array is transcribed into the pre-crRNA. This long RNA is processed by Cas proteins or cellular nucleases to yield mature crRNAs. (**C**) Together with the effector complex, the crRNAs recognize newly invading foreign DNA. Upon binding of the crRNA to the complementary sequences, the effector complex cleaves the foreign target. Left: schematic representation of a single-effector complex of class II, right: schematic representation of a multi-effector complex of class I. (adapted from Hille *et al.* 2018 [110])

2.2.2 Classification

With the discovery of more and more CRISPR operons, the diversity of Cas proteins, and their different mechanisms of action, it became clear that the various types required a systematized classification. Makarova *et al.* developed a classification system and constantly update it following a 'polyethic' approach taking into account the differences in Cas protein composition, phylogenetic, genomic, and structural studies [79,84,111]. The current classification divides the CRISPR-Cas system into two classes, six types, 33 subtypes, and several variants [79].

Class 1 comprises types I, III, and IV, these systems possess multiprotein effector complexes interfering with the foreign nucleic acids [79,84]. The effector complexes of these types consist of a backbone formed by Cas5 and multiple Cas7 proteins, both proteins of the RAMP family (repeat-associated mysterious protein) [79]. A third RAMP protein, Cas6, featuring RNase activity, is responsible for pre-crRNA processing in most of these systems (**Figure 2**) [79].

Type I, containing seven subtypes, is characterized by the signature protein Cas3 [84]. Cas3 is a nuclease and helicase that unwinds double-stranded DNA (dsDNA) and RNA-DNA heteroduplexes [112–114]. After recognition of the PAM and local unwinding of the target DNA by Cas8, subsequent conformational changes lead to the recruitment of Cas3 as part of the CRISPR-associated complex for antiviral defense (Cascade), which then cleaves the foreign DNA [79,84,110].

The type III CRISPR system contains six subtypes and is distinguished by the signature protein Cas10 [84]. Most operons of type III subtypes lack the *cas6* gene. Cas6 is responsible for pre-crRNA processing in other systems. Consequently, type III CRISPR systems depend on a Cas6 protein provided in *trans* by another locus of the same genome [115]. Some subtypes additionally lack the *cas1* and *cas2* genes, making them dependent on adaptation proteins encoded in other loci [84]. The effector complex of the type III system, namely Csm or Cmr (depending on the subtype) co-transcriptionally degrades single-stranded RNA (ssRNA) employing the nuclease subunit Cas7 and cleaves the DNA via Cas10 [79,84,110]. Type III systems deploy a signaling pathway, involving the synthesis of cyclic oligoadenylate (coA) by Cas10. These second messengers bind the CARF (CRISPR-associated Rossmann fold) domain of Csm6 or Csx1 and thereby activate the indiscriminate RNase activity of their HEPN (higher eukaryotes and prokaryotes nucleotide-binding) domain leading to promiscuous RNA cleavage [116,117].

The type IV system consists of three subtypes and is almost exclusively localized on plasmids [79,118]. As these subtypes lack the proteins responsible for adaptation and target cleavage, they rely on the *cas* genes of other loci in the genome and are proposed to perform other yet unknown functions besides the adaptive immunity like the maintenance or mobility of plasmids [79,119]. Recent studies showed that the Cas5 and Cas7 components that are part of the effector complex share sequence homologies with the respective proteins in type III systems, suggesting that type IV could be a derivative of type III systems [79,120,121].

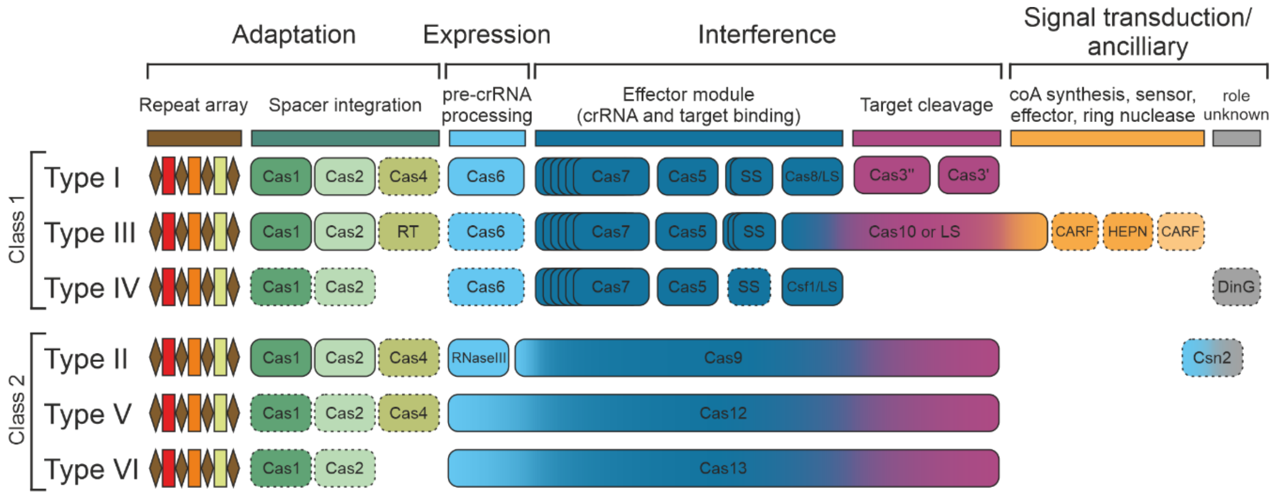


Figure 2: The modular organization of CRISPR-Cas systems.
Class 1 CRISPR systems utilize an effector module composed of multiple Cas proteins. These form a complex to bind the crRNA and target DNA and subsequently sequence-specifically cleave the DNA. Class 2 systems have one single multidomain protein as effector module that accomplishes all tasks. Shown are the relationships between the organizations of the six types of CRISPR-Cas systems. Dashed outlines indicate dispensable or (in some subtypes) missing components. LS, large subunit; SS small subunit. (adapted from Makarova *et al.* [79])

Class 2 CRISPR systems utilize single, large, multidomain Cas proteins complexed with crRNA as effector complexes. Part of this class are the types II, V, and VI (**Figure 2**) [79].
The type II CRISPR system comprises three subtypes and is characterized by the effector protein Cas9, which is not only responsible for target DNA cleavage but also involved in spacer acquisition [79,122]. Additionally, the Cas9 protein is required for pre-crRNA processing in a joint action with RNase III, a non-Cas protein from the bacterial host, and a *trans*-activating crRNA (tracrRNA) [115]. The tracrRNA forms a duplex structure with the pre-crRNA via a complementarity in the repeat-derived part of the crRNA. After processing by RNase III, the RNA duplex remains stably bound to Cas9 and is part of the effector complex [123–126]. Cas9 comprises a bilobal structure with two nuclease sites for the cleavage of both strands of a dsDNA target. Particularly, the HNH (histidine-asparagine-histidine) domain cleaves the target strand (TS) and the RuvC domain subsequently cleaves the non-target strand (NTS) [127–130]. The enzyme RNase III is absent in Archaea, consequently, the type II CRISPR systems are exclusively found in Bacteria [84,115,131].
Type V contains ten subtypes, all characterized by the single-effector protein Cas12 forming a complex with a single crRNA [79]. Cas12 proteins of subtype V-A (Cas12a) are able to self-process their pre-crRNA, whereas other type V subtypes harness the activity of RNase III

to produce mature crRNAs [115,132]. Cleavage of the target DNA is performed by a single active site in the RuvC domain that cleaves both DNA strands in a sequential manner [133]. CasΦ, also named Cas12j, a hypercompact Cas endonuclease encoded exclusively by huge bacteriophages is also classified as a type V protein sharing 7% sequence identity with other type V proteins [134,135]. It is supposed that these phages employ their own CRISPR endonuclease to compete against other phages that invade the same prokaryotic cell.

The type VI CRISPR system encompasses four subtypes. Their single-effector protein Cas13 contains two HEPN domains featuring RNase activity. One active site is responsible for processing the pre-crRNA, the other for sequence-specific cleavage of foreign RNA [79,115,136,137].

2.2.3 Anti-CRISPR systems

The arms race between phages and their hosts, which has been observed in many other prokaryotic defense strategies (0), is also visible in the CRISPR-Cas system. Phages can overcome the system by point mutations or deletions in important parts of the DNA sequence, like the PAM or the seed region (initial base-pairing between crRNA and target nucleic acid) [138]. A direct counterstrategy is the inhibition of the host's CRISPR system by anti-CRISPR (Acr) proteins with a variety of mechanisms [139–141]. 92 families of Acr proteins have been identified so far that inhibit types I, II, III, V, and VI CRISPR systems [142]. The different mechanisms of action reach from preventing the Cas proteins from binding or cleaving target DNA by directly binding to the effector protein [140,141], over preventing crRNA loading [143], or triggering unspecific DNA binding [144,145], to inducing the degradation of Cas proteins [146] or necessary second messenger molecules [147].

The arms race between bacteria and phages did not stop with the evolution of Acr proteins by phages but continues with the development of anti-anti-CRISPR systems in prokaryotes. For example, prokaryotic genomes can encode repressors that inactivate the Acr expression in phages [148].

2.2.4 Applications of CRISPR

In 2012, Jennifer Doudna and Emmanuelle Charpentier anticipated the potential of the CRISPR-Cas9 system for the development of a programmable DNA nuclease that would find use in diverse technologies like genome editing [149]. With the CRISPR-Cas system, the intention of targeting a specific sequence or site in a genome could be realized. With the possibility of changing the crRNA sequence to any desired sequence and hence, programming the Cas protein, opportunities opened that would have never been reached with the nuclease arsenal known so far. Eight years later, in 2020, Doudna and Charpentier got awarded the Nobel prize in Chemistry by the Royal Swedish Academy of Science 'for the development of a method for genome editing'. Currently, especially class 2 CRISPR-Cas proteins are harnessed for different applications as their handling as single-effector nucleases is much easier than for class 1 protein complexes. Additionally, the fusion of tracrRNA and crRNA to a single guide RNA (sgRNA) made the previously dual-guided Cas9 protein even more suitable for applications [149].

Nevertheless, Jennifer Doudna and Emmanuelle Charpentier were not the only ones seeing the great potential of CRISPR-Cas. They have an ongoing patent dispute with Feng Zhang from the Massachusetts Institute of Technology (MIT). Doudna and Charpentier applied for a patent right after their publication in 2012, already at the end of 2012 also Feng Zhang filed a patent application regarding the first utilization of CRISPR-Cas in human and mouse cells [150,151]. With that, the patent for the application in higher eukaryotic cells was first rewarded to Feng Zhang, although Doudna and Charpentier hold the patent for the general method. Despite many patent lawsuits, the dispute is not over yet, even though Feng Zhang was confirmed multiple times, most recently in February 2022. Until now many hundreds of CRISPR patents have been applied for, mostly with small variations to the original method, like the exchange of Cas9 by another Cas nuclease.

2.2.4.1 Genome editing

The resulting double-strand break (DSB) in a genome after DNA cleavage by a Cas protein will be repaired by the endogenous repair systems, namely non-homologous end joining (NHEJ) or homology-directed repair (HDR) [152]. NHEJ is an error-prone pathway that often results in insertions or deletions (indel) at the DSB site through the removal of some nucleotides prior to the direct conjunction of the DNA ends [153,154]. By contrast, HDR is a high-precision repair mechanism but requires a homologous DNA template. This requirement allows for specific insertions or deletions with the aid of a specifically designed HDR template DNA strand and therefore the insertion of new genes or the knock-out of a desired gene [152].

Recruitment of native Cas9 and Cas12a proteins is based on the presence of a PAM [107], limiting their area of application to these sequences. To broaden their operating range, Cas9 and Cas12a variants with altered or less strict PAM specificities were generated by rational design or random mutagenesis [155–159]. A reoccurring problem of CRISPR nucleases are off-target effects, meaning that sequences that only show partial complementarity to the crRNA get cleaved and could harm the host [160–162]. Several groups tried to overcome this issue by producing Cas9 variants with increased target specificity [163–166].

To directly exchange single bases and introduce precise point mutations without the need for a DSB, catalytically inactive Cas proteins are modified and fused to deaminases. The deaminase only acts on ssDNA, which is provided by the Cas protein by binding to DNA and unwinding the target site. This way, various base conversions can be introduced site-specifically, e.g. the conversion of a C-G pair into a T-A pair (via cytosine base editors – CBEs) or an A-T base pair into a G-C pair (via adenine base editors – ABEs) [167,168].

2.2.4.2 Plant biotechnology

The great potential of the CRISPR-Cas technology was also noticed in agriculture, aiming to increase productivity by generating genome-edited crops, rice, and corn with increased grain size, weight, and quality [169–171]. Furthermore, the composition of ingredients was enhanced, like decreasing the amount of gluten, unsaturated fatty acids, or the levels of toxic steroidal glycoalkaloids or increasing amylose, starch, or oleic acid levels [172–175]. CRISPR-Cas systems are also employed to incorporate resistances into plants. A resistance against plant targeting RNA viruses can be gained through Cas13a [176,177], whereas CBEs are used to introduce herbicide resistances into rice, Arabidopsis, or watermelon [178–180]. A hope associated with the use of CRISPR in plant biotechnology is to increase food production to feed the ever-growing population on planet earth [103].

2.2.4.3 Therapeutics and disease management

The application of the CRISPR-Cas system in genome editing also opened up the possibility to cure human genetic diseases. Potential diseases to be targeted are monogenetic diseases like Duchenne muscular dystrophy [181], α-1 antitrypsin deficiency (AATD) [182], hemophilia [183], hematopoietic diseases [184,185], or hearing loss [186].

Blood disorders, like transfusion-dependent β-thalassemia (TDT) and sickle cell disease (SCD) are severe and potentially life-threatening diseases. They are caused by wrong or insufficient production of hemoglobin β-chains, leading to a decrease in oxygen-carrying capacities. A strategy to cure these diseases is the upregulation of fetal γ-hemoglobin expression, which is normally repressed by the transcription factor BCL11A in non-fetal cells. Blood stem cells are continuously built new, rendering them a good target for CRISPR applications without the need to alter embryonic or germline cells. To this end, hematopoietic stem cells from healthy donors are treated by CRISPR-Cas9, introducing a deletion in the BCL11A gene [184,185]. In a corporate study of researchers and companies under the lead of Selim Corbacioglu (University Hospital, Regensburg), two patients received Cas9-edited cells with great success, showing a pan-cellular distribution of fetal hemoglobin and resulting in the independence of blood transfusions [185].

The human immunodeficiency virus (HIV) is a retrovirus that integrates into the human genome [187]. Current therapies stop viral replication but are not able to target latent infections. Therefore, the CRISPR-Cas technology has been implemented to fight HIV by destroying its prophages [188].

In November 2018, He Jiankui attracted attention by reporting the birth of two genome-edited babies at the Second International Summit on Human Genome Editing in Hong Kong [189,190]. Jiankui deleted 32 amino acids (aa) of the chemokine receptor type 5 (CCR5) gene, one of two coreceptors necessary for HIV to enter a host cell [189,191]. With this, he aimed to provide resistance against the virus infection for the babies. After their birth, Lulu's and Nana's genome was sequenced with DNA samples from their placenta, umbilical cord, and cord blood revealing that not both alleles of the CCR5 gene were targeted resulting in mosaicism in the girls as well as off-target effects in both twins [103].

The use of CRISPR-Cas in human gene editing is a highly discussed topic in ethics and law. Questioned are for example the still high risk of off-target effects and the resulting hazard of uncontrollable and random mutations in the genome as well as the responsibility towards future generations if the germline is edited. The economic side is also taken into account as a society divided into the rich and healthy, and the poor and sick is apprehended [103,192–194].

2.2.4.4 Disease detection

The CRISPR technology has also been applied to detect diseases like cancer. The SHERLOCK (specific high-sensitivity enzymatic reporter unlocking) platform uses Cas13a to rapidly detect specific sequences that are amplified via T7 RNA transcription with high sensitivity [195]. During the recent COVID-19 (coronavirus disease 2019) pandemic caused by SARS-CoV-2 (severe acute respiratory syndrome Coronavirus 2), a multitude of detection assays based on Cas12a or Cas13a activity were developed [196–213].

2.2.5 Cas12a

In this thesis, Cas12a, the signature and single-effector protein of the type V-A CRISPR-Cas system from *Francisella novicida* (Fn) was characterized. Cas12a was formerly named CRISPR-associated endonuclease in *Prevotella* and *Francisella* 1 (Cpf1) [214,215]. Its similarity to TnpB nucleases that are encoded by transposons and its homology to genes found outside the CRISPR-Cas context suggest that Cas12a is recruited from transposable elements, similar to other type V nucleases [84,118,215,216].

Cas12a is a 150 kDa protein with 1300 aa that adopts a bilobal structure with the nuclease (Nuc) lobe, composed of the domains RuvC I, RuvC II, RuvC III, and BH (bridge helix), and the recognition (REC) lobe, encompassing the REC1 and REC2 domains [217–221]. The wedge (WED) domain connects both lobes and acts as a hinge module (**Figure 3A–C**).

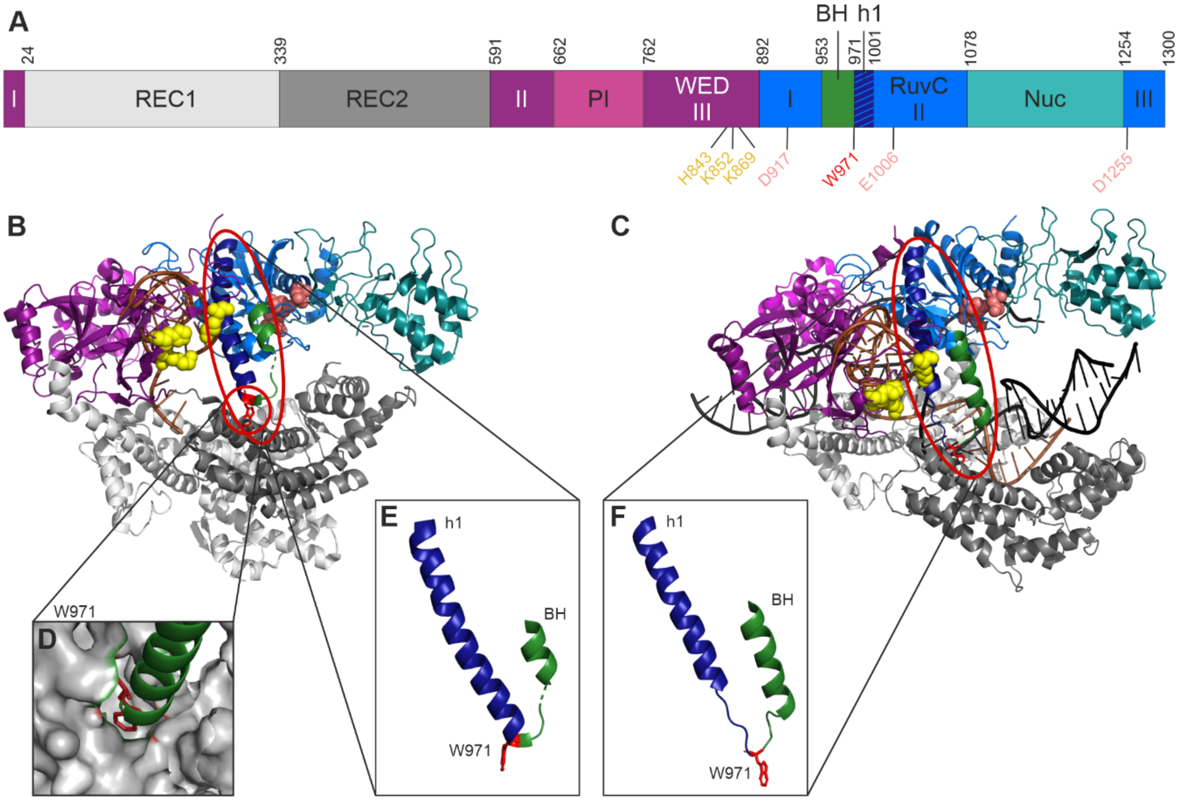


Figure 3: Structural organization of FnCas12a.
(**A**) Domain organization of FnCas12a with residues of the RNase catalytic site indicated in yellow, residues of the DNase catalytic site indicated in salmon, and W971 indicated in red. (**B**) Structure of the binary complex of FnCas12a with crRNA (brown) (PDB: 5NG6). (**C**) Structure of the ternary complex of FnCas12a together with crRNA (brown) and target DNA (black) (PDB: 6I1K). (**B**) and (**C**) Protein domains are colored according to the domain organization in (**A**). Helix 1 (h1), part of the RuvC II domain, is colored in dark blue, the bridge helix (BH) in green, residues of the RNase catalytic site are shown in yellow, residues of the DNase catalytic site in salmon, and W971 in red. (**D**) Close-up of the anchoring of W971 at the C-terminal end of the bridge helix in a hydrophobic pocket formed by the REC2 domain. (**E**) and (**F**) Close-up of helix 1 (blue) and BH (green) in the binary (**E**) and ternary (**F**) state, showing the change of helical lengths upon the rearrangement from binary to ternary complex. W971 is shown in red.

With a ribonuclease-active site in the WED domain (formed by residues H843, K852, and K869 in FnCas12a), Cas12a processes its own pre-crRNA via an acid-base catalytic mechanism (**Figure 4**) [132,220].

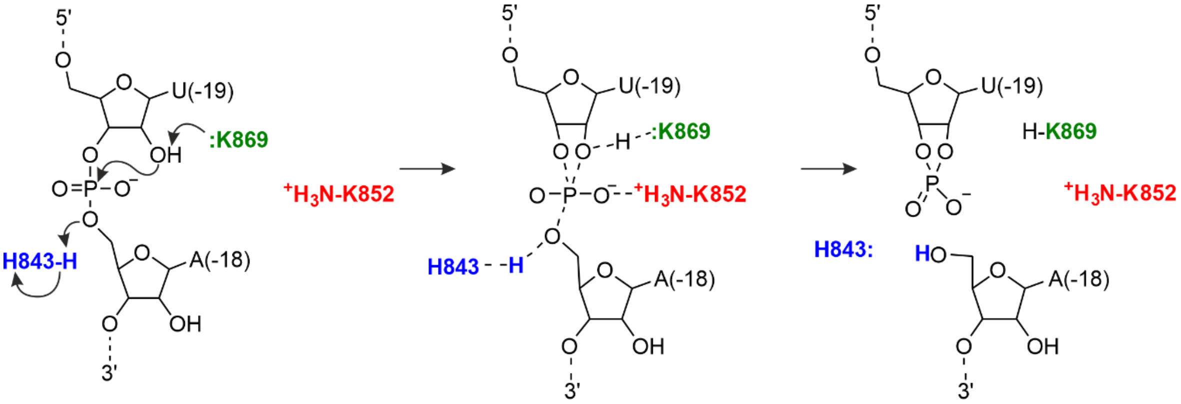


Figure 4: Acid-base reaction mechanism of pre-crRNA cleavage of FnCas12a.
Reaction mechanisms of pre-crRNA cleavage by FnCas12a, involving the catalytic residues H843, K852, and K869. (adapted from Swarts *et al.* [220])

The mature crRNA remains bound to Cas12a forming a characteristic 5’-pseudoknot structure with its repeat-derived part [217,218]. A pseudoknot is an RNA-secondary structure formed by two integrated hairpins of the 5’-handle (**Figure 5**). The spacer-derived part of the crRNA interacts with the protein, pre-ordering and exposing the nucleotides to form base pairs with the TS of the target DNA [220]. Upon infection, the ribonucleoprotein (RNP) composed of Cas12a and its crRNA (binary complex) scans the DNA via irregular binding and dissociation for the PAM sequence 5’-TTTV-3’ (V is any base, but not T) and for complementarity with the crRNA [222,223]. Binding the PAM via interactions with a flexible pocket formed by the WED, REC1, and PI domains (**Figure 3B, C**), results in ATP-independent local melting of the double-stranded target DNA at the seed region (5–6 nt), directly downstream of the PAM [218,220,221,224–227]. The PAM is recognized by a shape-specific recognition of a narrowed minor groove of the DNA, combined with a base-specific recognition by K671 and K613 in FnCas12a [220]. A third lysine, K667, inserts between the DNA strands directly downstream of the PAM and separates the DNA strands [220]. Base-pairing through the complementarity of crRNA and TS further melts the DNA and leads to the formation of an RNA-DNA heteroduplex and an R-loop structure within the REC lobe [219,221,222]. When the RNP is stably bound to the target DNA, it is called ternary complex, consisting of Cas12a, crRNA, and target DNA. The seed region of the R-loop is especially sensitive to mismatches between crRNA and target DNA. At least 14 nt of complementarity are necessary for stable formation of the heteroduplex, 16 nt to detect cleavage, and 18 nt to achieve efficient DNA cleavage [214,228,229]. Especially compared to Cas9, Cas12a is more sensitive to mismatches, particularly at the PAM-distal part of the spacer region [229–231]. As long as target sequences contain complementarity to the crRNA, Cas12a cleaves dsDNA but also ssDNA [214].

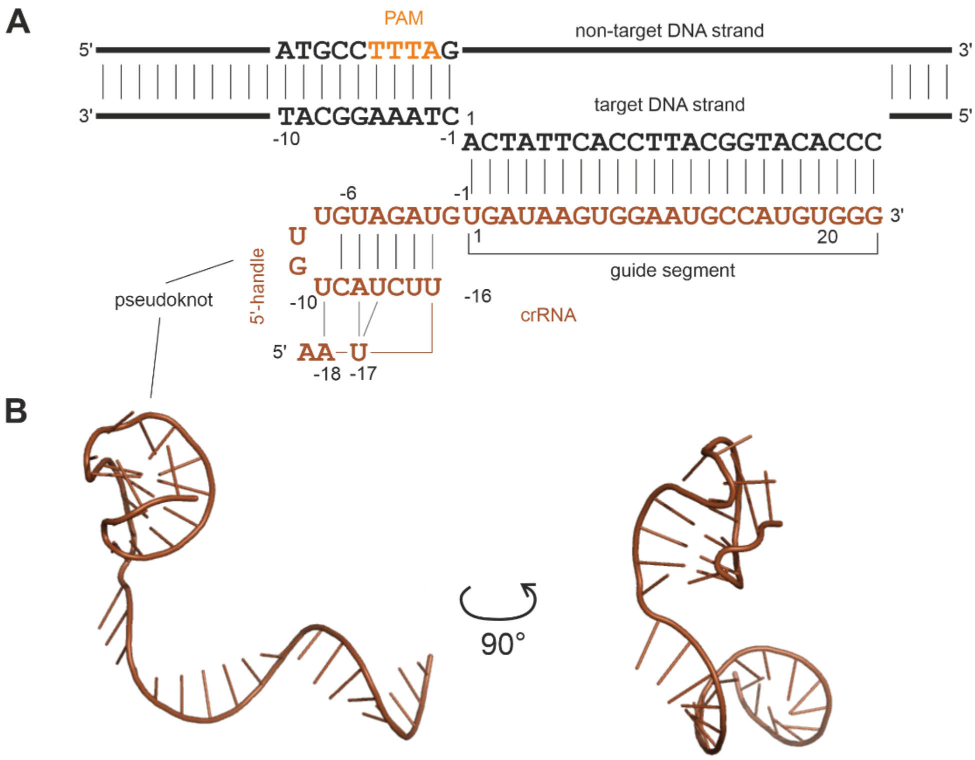


Figure 5: Pseudoknot formation in the crRNA of Cas12a.
(**A**) The 5'-handle of the crRNA (brown) forms a pseudoknot structure. The 3' guide segment base pairs with the complementary sequence on the target strand of the DNA double strand (black). During this heteroduplex formation, the non-target strand is displaced. (**B**) Structure of the crRNA forming the pseudoknot at the 5'-end (PDB: 6I1K). (adapted from Yamano *et al.* [218])

To activate target DNA cleavage, different checkpoints have to be passed. The 'REC linker', a loop connecting REC1 and REC2 domains, undergoes conformational changes to form interactions with the crRNA [226]. The 'lid' is a flexible and unstructured loop that contains one of the catalytic residues for DNA cleavage, E1006. Additionally, it harbors the residue K1013 which is in polar interactions with D917, another catalytic residue. In the binary complex, the lid closes the catalytic pocket. Upon transitioning to the ternary complex, the lid forms a new α-helix opening the catalytic pocket for the NTS to enter (**Figure 6**) [226]. Additionally, the 'finger' is a helix in the REC1 domain that undergoes a 30 Å movement to form interactions with the crRNA [226]. This movement is part of the closed-to-open conformational change of Cas12a upon the transition to the ternary complex to accommodate the target DNA [219,220].

Cleavage of both target DNA strands, called *cis*-cleavage, occurs sequentially in the RuvC active site formed by the catalytic residues D917, E1006, and D1255 in FnCas12a [132] (**Figure 3A–C**). The displaced NTS docks into the active site and is cleaved between the 16th and 17th nucleotides downstream of the PAM, resulting in the nicking of the DNA [222,231]. The strand is released but can re-enter the active site leading to trimming of the NTS up to four nucleotides [231,232]. As the active site is initially incorrectly orientated for TS cleavage, the Nuc domain, which is inactive in catalysis, helps to orient the target strand for the cleavage between the 22nd, 23rd, or 24th nt in the RuvC active site [220,233–235]. Because of large re-orientations in the protein necessary to load the TS, cleavage of the NTS is 20 times faster ($1.2 \cdot 10^{-2}$ s^{-1}) than cleavage of the TS [226,231]. Cleavage of the TS was accelerated in Cas12a from *Lachnospiraceae bacterium* (LbCas12a) by the removal of W355, an aromatic gate that terminates the R-loop via stacking interactions. The removal increases downstream DNA breathing and results in the activation of the second cleavage event [235]. The sequential cleavage of both strands results in a DSB with staggered ends and a 5'-overhang of 5–8 nt [110,133,214,220].

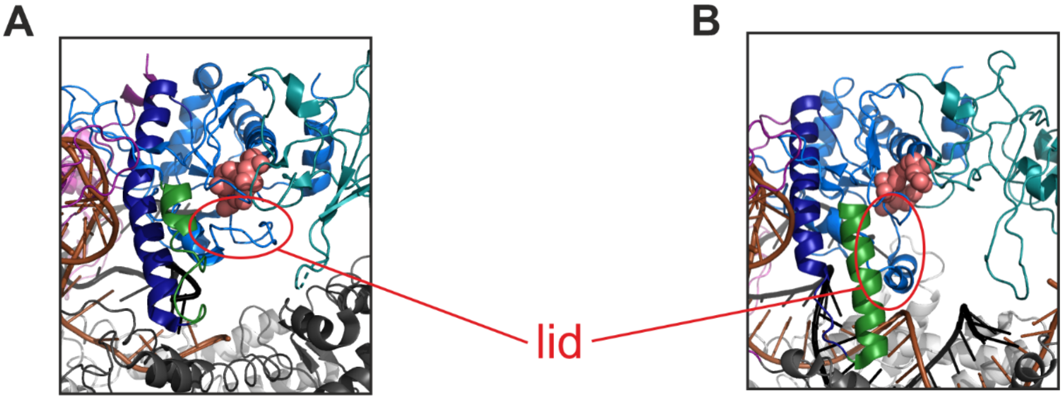


Figure 6: Conformational transitions of the structural element 'lid' in Cas12a.
(**A**) In the binary complex and at the beginning of the conformational transitions of Cas12a to the ternary complex after DNA binding, the flexible structural element of the 'lid' is an unstructured loop making the active site (salmon) inaccessible for the target DNA (black). PDB: 6GTC. (**B**) During the transition to the ternary complex conformation, the lid forms a new α-helix and opens the active site for the target DNA. PDB: 6GTG. [226] Color code according to **Figure 3**, dark blue: helix 1, green: BH.

After cleavage, Cas12a directly releases the PAM-distal cleavage product but remains bound to the PAM-proximal part. The very slow dissociation of the PAM-proximal part on the one hand effectively defines Cas12a as a single-turnover enzyme [161,221,222,229,230]. On the other hand, it activates Cas12a for indiscriminate cleavage of ssDNA, ssRNA, and dsDNA in *trans*, as Cas12a remains in an active and cleavage-competent state [236,237]. This has been widely exploited for the development of quantitative platforms to detect nucleic acids even though the exact mechanism of *trans*-cleavage is not completely understood yet [238–241].

Several Acr proteins have been identified that inhibit Cas12a. Amongst them, AcrVA1 which mimics the DNA structure Cas12a binds to. Additionally, AcrVA1 is capable to cleave the crRNA when it is bound to Cas12a [242,243]. The protein AcrVA4 binds to the REC2, WED, and BH domains of Cas12a, induces dimerization, and hinders the formation of the full R-loop and consequently, the activation of Cas12a [242,244,245]. AcrVA5 is an acetyltransferase that modifies a lysine residue of Cas12a required for PAM recognition. The modified Cas12a variant completely loses its ability to bind dsDNA and is not able to cleave the target DNA anymore [246].

2.3 Objective

The CRISPR-Cas technology harbors great potential for a wide range of applications. But until now, the risks often prevail in ethical considerations about utilizing the technology in human therapy. To obtain a better understanding of the existing risks and to elude and dispel them, the molecular mechanisms of CRISPR-Cas enzymes utilized for genome editing have to be completely understood. To this end, it is important to constantly extend the current knowledge about CRISPR-Cas in general and its core proteins, the nucleases, in particular.

As single-effector protein, Cas12a comprises at least the same technological potential as Cas9 but important mechanistic details are still not understood. Therefore, this thesis aimed to elucidate conformational transitions as well as the function of distinct structural parts of Cas12a.

The Nuc and the REC lobe of Cas12a are connected by the WED domain that acts as a hinge between the two lobes. The two lobes are additionally linked by the BH domain, a central, α-helical structure that also interacts with the sugar-phosphate backbone of the TS. Structural studies of *Acidaminococcus sp.* Cas12a (AsCas12a) revealed furthermore that residues of the BH are involved in the recognition of the crRNA-DNA heteroduplex [218]. At the C-terminal end, the BH harbors a tryptophan residue (W971 in FnCas12a) that is accommodated in a hydrophobic pocket formed by the REC2 domain. These hydrophobic interactions stabilize the closed and active conformation of Cas12a [218,221] (**Figure 3D**). Crystal structures show that the BH is further connected to a helix of the RuvC domain (termed helix 1) [133,217–221,226]. These two adjacent helices act in concert upon the transition from the binary to the ternary complex: in the binary complex, the BH is short and helix 1 is long, in the ternary complex, by contrast, the BH is elongated, whereas helix 1 is shortened by six amino acids (**Figure 3E, F**). The tandem movement of BH and helix 1 is part of significant conformational rearrangements in the REC lobe upon binding of the target DNA and the transition from the binary to the ternary complex and ensures that the tryptophan (W971) stays in its hydrophobic pocket in the REC2 domain. In FnCas12a, contacts between the residues lysine 978, isoleucine 977, and asparagine 976 of helix 1 and the TS were described [221].

The objective of this thesis was to conduct single-molecule Förster resonance energy transfer (smFRET) measurements on diffusing molecules to follow the conformational transitions of FnCas12a through its activity cycle. Additionally, the influence of the BH and helix 1 on the catalytic activity, structural integrity, and conformational transitions should be assessed via mutational analysis and the generation of FnCas12a BH and helix 1 variants.

3. MATERIALS

3.1 Chemicals

The chemicals used had the highest possible purity grade (listed in alphabetical order).

Table 1. List of chemicals and reagents.

Reagent	Supplier
40% Acrylamide/Bisacrylamide solution (19:1)	Roth-Chemie, Karlsruhe (DE)
30% Acrylamide/Bisacrylamide solution (37.5:1)	Roth-Chemie, Karlsruhe (DE)
Acetic acid, p.a. ≥ 99.8%	Merck, Darmstadt (DE)
Agar	VWR International, Radnor (PA, USA)
Agarose LE	Biozyme, St. Joseph (MO, USA)
Ammonium persulfate (APS)	Roth-Chemie, Karlsruhe (DE)
Ampicillin sodium salt	AppliChem GmbH, Darmstadt (DE)
L-Arabinose	Merck, Darmstadt (DE)
Boric acid	Merck, Darmstadt (DE)
Bovine serum albumin (BSA)	Sigma-Aldrich, St. Louis (MO, USA)
Bromophenol blue	AppliChem, Karlsruhe (DE)
Calcium chloride dihydrate	Merck, Darmstadt (DE)
Chloramphenicol	Roth-Chemie, Karlsruhe (DE)
cOmplete EDTA-free Protease Inhibitor Cocktail Tablets	Roche Diagnostics International AG, Rotkreuz (CH)
Coomassie Brilliant Blue G250	Serva Electrophoresis GmbH, Heidelberg
Deoxynucleotide (dNTP) solution Mix (10 mM)	New England Biolabs, Ipswich (MA, USA)
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich, St. Louis (MO, USA)
Dithiothreitol (DTT)	Roth-Chemie, Karlsruhe (DE)
Ethanol, p.a. ≥ 99.8%	Sigma-Aldrich, St. Louis (MO, USA)
Ethanol, technical 99%	J.T. Baker, Deventer (NL)
Ethylenediaminetetraacetic acid (EDTA, Titriplex III)	Merck, Darmstadt (DE)
Ficoll 400	Sigma-Aldrich, St. Louis (MO, USA)
Formamide	Merck, Darmstadt (DE)
L-Glutathione reduced	Roth-Chemie, Karlsruhe (DE)
Glycerol 99%, Ph.Eur.	Roth-Chemie, Karlsruhe (DE)
Glycine	Sigma-Aldrich, St. Louis (MO, USA)
Hellmanex	Hellma, Müllheim (DE)

Heparin sodium salt	Sigma-Aldrich, St. Louis (MO, USA)
Hydrochloric acid 37%	Merck, Darmstadt (DE)
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Roth-Chemie, Karlsruhe (DE)
Isopropyl-β-D-thiogalactopyranoside (IPTG)	Roth-Chemie, Karlsruhe (DE)
Kanamycin	Sigma-Aldrich, St. Louis (MO, USA)
Magnesium acetate	Merck, Darmstadt (DE)
Magnesium chloride hexahydrate	Merck, Darmstadt (DE)
Manganese chloride tetrahydrate	Merck, Darmstadt (DE)
β-Mercaptoethanol	Roth-Chemie, Karlsruhe
MilliQ water	University of Regensburg (DE)
Nuclease-free water	Cell Signaling Technology, Cambridge (UK)
para-azido-L-Phenylalanine (AzF)	Chem-Impex Internatonal Inc., Wood Dale (IL, USA)
Phenol-chloroform-isoamyl alcohol (25:24:1)	VWR International, Radnor (PA, USA)
Polyoxoethylenesorbitan Monolaurate (Tween20)	Sigma-Aldrich, St. Louis (MO, USA)
Potassium acetate	Merck, Darmstadt (DE)
Potassium chloride	Merck, Darmstadt (DE)
Potassium dihydrogen phosphate	Merck, Darmstadt (DE)
Sigmacote	Sigma-Aldrich, St. Louis (MO, USA)
Sodium chloride	neoFroxx, Einhausen (DE)
Sodium dihydrogen phosphate	Merck, Darmstadt (DE)
Sodium dodecyl sulfate (SDS)	Merck, Darmstadt (DE)
Sodium hydroxide	Fisher Scientific, Loughborough (UK)
SYBR Gold nucleic acid gel stain	ThermoFisher Scientific, Waltham (MA, USA)
SYBR Safe DNA Gel stain	ThermoFisher Scientific, Waltham (MA, USA)
N,N,N',N'-Tetramethylethylenediamine (TEMED)	Roth-Chemie, Karlsruhe (DE)
Tris(hydroxymethyl)aminomethane (Tris)	Sigma-Aldrich, St. Louis (MO, USA)
Tryptone	Difco, Lawrence (KS, USA)
Urea	Sigma-Aldrich, St. Louis (MO, USA)
Xylene cyanol	Merck, Darmstadt (DE)
Yeast extract	Voigt Global Distribution Inc., Lawrence (KS, USA)

3.2 Equipment

Table 2. List of Equipment.

Equipment	Producer
Analytical Balance Sartorius CP2202S	Sartorius, Göttingen (DE)
Autoclav Sanoclav	Maschinenbau Wolf GmbH, Bad Überkingen (DE)
Affinity column HiTrap Heparin High Performance	GE Healthcare, Chicago (IL, USA)
Affinity column GSTrap Fast Flow	GE Healthcare, Chicago (IL, USA)
Agarose gel system Mini Sub cell GT	Bio-Rad Laboratories Inc., Hercules (CA, USA)
ÄKTA purifier	GE Healthcare Life Science, Freiburg (DE)
Centrifuge Beckman Avanti® J-25 centrifuge	Beckman Coulter, Brea (CA, USA)
Centrifuge 5424 R	Eppendorf, Hamburg (DE)
Centrifuge 5425	Eppendorf, Hamburg (DE)
Centrifuge 5810 R	Eppendorf, Hamburg (DE)
Centrifuge CT15RE	VWR International, Radnor (PA, USA)
Centrifuge Galaxy MiniStar	VWR International, Radnor (PA, USA)
Centrifuge Sorvall RC 5C Plus	ThermoFisher Scientific, Waltham (MA, USA)
Centrifuge Mastercycler personal	Eppendorf, Hamburg (DE)
Centrifuge mini MCF-2360	LMS Co., LTP, Tokyo (JP)
ChemiDoc MP Imaging System	Bio-Rad Laboratories Inc., Hercules (CA, USA)
Centrifuge rotor JA.10	Beckman Coulter, Brea (CA, USA)
Centrifuge rotor JA25.50	Beckman Coulter, Brea (CA, USA)
Centrifuge rotor GS3	ThermoFisher Scientific, Waltham (MA, USA)
Centrifuge rotor SS34	ThermoFisher Scientific, Waltham (MA, USA)
Glassware	VWR International (Schott), Radnor (PA, USA)
Heater and magnetic stirrer SU1350	Sunlab Sustainable Lab Instruments Labdiscount GmbH, Mannheim (DE)
Ice flaker Scotsman MF26	Scotsman ice Systems Vernon Hills (Il, USA)
Incubation Shaker INFORS HT Multitron Pro	Infors AG, Bottmingen (CH)
Filtration Unit, Nalgene	ThermoFisher Scientific, Waltham (MA, USA)
Freezer Liebherr Comfort	Liebherr, Bulle (CH)
Freezer New Brunswick Innova U725-G	Eppendorf, Hamburg (DE)
Fridge Liebherr UK1720	Liebherr, Bulle (CH)
Fridge Sanyo Labcool MPR-720	Ewald Innovationstechnik, Bad Nenndorf (DE)

Mass photometer TwoMP	Refeyn, Oxford (GB)
Microwave Europa style	Tchibo GmbH, Hamburg (DE)
Mini-PROTEAN-Tetra System	Bio-Rad Laboratories Inc., Hercules (CA, USA)
NanoDrop One^{C}	ThermoFisher Scientific, Waltham (MA, USA)
pH-meter pHM82 Standard	Radiometer A/S, Kopenhagen (DK)
Photometer WPA CO 8000 biowave Cell Density Meter	Biochrom Ltd., Cambridge (GB)
Pipetting aid Pipetus	Hirschmann Laborgeräte GmbH & Co. KG, Eberstadt (DE)
Plate reader Victor Nivo Multimode	PerkinElmer, Waltham (MA, USA)
PowerPac Basic	Bio-Rad Laboratories Inc., Hercules (CA, USA)
PowerPac HV	Bio-Rad Laboratories Inc., Hercules (CA, USA)
Precision balance Acculab Atilon ATL-224-I	Sartorius, Göttingen (DE)
Quartz Crystal Cuvette, 0.2 mm	Hellma Analytics, Müllheim (DE)
Rotation shaker Roto-Shake Genie	Scientific industries Inc., Bohemia (NY, USA)
Rotor-Gene Q	Qiagen, Venlo (NL)
Sequi-Gen GT sequencing Cell	Bio-Rad Laboratories Inc., Hercules (CA, USA)
Sonotrode SONOPULS H2070	BANDELIN electronic, Berlin (DE)
Spectropolarimeter J-815	Jasco Corporation, Tokyo (JP)
Thermocycler MJ Mini	Bio-Rad Laboratories Inc., Hercules (CA, USA)
Thermocycler C1000 Touch Thermal Cycler	Bio-Rad Laboratories Inc., Hercules (CA, USA)
Thermocycler Primus25 advanced	Peqlab Biotechnologie GmbH, Erlangen (DE)
ThermoMixer C	Eppendorf, Hamburg (DE)
Transferpette (1000, 200, 20, 10, 1 µl)	Sigma-Aldrich, St. Louis (MS, USA)
Ultrasonic bath Elmasonic S 30 H	Elma Schmidbauer GmbH, Singen (DE)
Vacuum pump Laboport N 86 KN.18 KNF	Neuberger, Freiburg im Breisgau (DE)
Vortex-Genie 2	Scientific Industries Inc., Bohemia (NY, USA)

Table 3. List of optical components of the MicroTime 200 confocal microscope (PicoQuant, DE).

Component	Producer
Bandpass filter (emission) 582/64 nm BrightLine	Semrock, Rochester (NY, USA)
Bandpass filter (emission) H690/70	Chroma Technology, Bellow Falls (VT, USA)
Dichroic mirror (emission) T 635LPX	Chroma Technology, Bellow Falls (VT, USA)
Dichroic mirror (excitation) ZT 532/640RPC	AHF Analysentechnik, Tübingen (DE)
Narrow band cleanup filter ZET 635/10	AHF Analysentechnik, Tübingen (DE)
Objective UplanSApo ×60/1.20 W	Olympus, Tokyo (JP)
Olympus IX72 (microscope body)	Olympus, Tokyo (JP)
Optical workstation Vision IsoStation	Newport Corporation, Irvine (CA, USA)
Pulsed laser diode head LDH-P-FA-530B	PicoQuant, Berlin (DE)
Pulsed laser diode head LDH-D-C-640	PicoQuant, Berlin (DE)
Single photon detector SPCM-AQRH-14-TR	Excelitas Technologies, Dumbery (CAN)
TCSPC module HydraHarp 400	PicoQuant, Berlin (DE)

3.3 Consumables

Table 4. List of Consumables.

Consumable	Producer
Amicon Ultra 0.5 ml Centrifugal Filters 100 K	Merck, Darmstadt (DE)
CELLview slides (50 x 26 mm)	Greiner Bio-One Inc., Monroe (NC, USA)
CultureWell reusable gaskets	Grace Bio-Labs, Bend (OR, USA)
Falcon Tubes (15 ml, 50 ml)	Sarstedt, Nümbrecht (DE)
Filter tips (10 µl, 100 µl, 200 µl, 1000 µl)	Sarstedt, Nümbrecht (DE)
Lens cleaning tissue	Thorlabs, Lübeck (DE)
Membrane filters Whatman 0.2 µm	GE Healthcare, Chicago (IL, USA)
Membrane filters Whatman 0.45 µm	GE Healthcare, Chicago (IL, USA)
Microscope cover glasses (high precision) 24 x 60 x 0.17 mm	Roth-Chemie, Karlsruhe (DE)
Nick Columns Sephadex G-50 DNA Grade	GE Healthcare, Chicago (IL, USA)
Nitrile Gloves size S, M	Roth-Chemie, Karlsruhe (DE)
Parafilm „M“ Laboratory Film	Bemis NA, Neenah (WI, USA)
PCR tubes	VWR International, Radnor (PA, USA)
PD SpinTrap G-25 columns	Merck, Darmstadt (DE)

Pipette tips (10 µl, 200 µl, 100 µl)	Sarstedt, Nümbrecht (DE) STARLAB Group, Hamburg (DE)
Reaction tubes (0.2, 0.5, 1.5, 2 ml)	Sarstedt, Nümbrecht (DE) STARLAB Group, Hamburg (DE)
Slide-A-Lyzer Dialysis Cassette 0.1–0.5 ml 100.000 MWCO	ThermoFisher Scientific, Waltham (MA, USA)
Syringe filter Rotilab PES sterile (0.22 µm)	Roth-Chemie, Karlsruhe (DE)
Syringe filter Rotilab CA sterile (0.45 µm)	Roth-Chemie, Karlsruhe (DE)
Syringe (3 ml, 10 ml, 20 ml)	Henke-Sass, Wolff, Tuttlingen (DE)
96-well plate (black with clear bottom)	Invitrogen, Carlsbad (CA, USA)

3.4 Software

Table 5. List of Software.

Software	**Producer**
AcquireMP 2022 R1	Refeyn, Oxford (GB)
ÄKTA Unicorn Version 5.01	GE Healthcare, Chicago (IL, USA)
ChemDraw Professional 20.0.0.41	PerkinElmer, Waltham (MA, USA)
CorelDRAW 2020 22.1.1.523	COREL Corporation, Ottawa (CA)
DiscoverMP 2022 R1	Refeyn, Oxford (GB)
ImageLab 6.0	Bio-Rad Laboratories, Hercules (CA, USA)
Mendeley Desktop 1.19.8	Mendeley, London (UK)
MS Office 365	Microsoft Corporation, Redmond (WA, USA)
NanoDrop 1000 3.8.0	ThermoFisher Scientific, Waltham (MA, USA)
Origin 2021	OriginLab Corporation, Northampton (MA, USA)
PIE Analysis in MATLAB (PAM) 1.22	Laboratory of Prof. Dr. Don C. Lamb, LMU Munich (DE) [247]
PyMOL 2.3.1	Schrödinger LLC, München (DE)
Serial Cloner 2.6.1	SerialBasics (FR)
SymPhoTime 64 2.5	PicoQuant, Berlin (DE)
Spectra Analysis 2.09.06	Jasco Corporation, Tokyo (JP)
Spectra Manager	Jasco Corporation, Tokyo (JP)
Rotor-Gene ScreenClust HRM	Qiagen, Venlo (NL)

3.5 Buffers and Solutions

All solutions and buffers were stored at room temperature unless stated otherwise. Buffers used on chromatography units were sterile-filtered prior to use (Whatman filter Ø 33 mm; 0.22 µm pore size; Roth-Chemie).

3.5.1 Buffers

Table 6. List of buffers.

Buffer	Composition	Concentration
5x Annealing buffer (pH adjusted with acetic acid)	HEPES pH 7.4 Potassium acetate Magnesium acetate ddH_2O	15 mM 50 mM 1 mM
Äkta Cas A buffer	Tris/HCl pH 7.5 NaCl $MgCl_2$ Glycerol ddH_2O	20 mM 50 mM 5 mM 5% (v/v)
Äkta Cas B buffer	Tris/HCl pH 7.5 NaCl $MgCl_2$ Glycerol Glutathione ddH_2O	20 mM 50 mM 5 mM 5% (v/v) 10 mM
Äkta Cas high salt buffer	Tris/HCl pH 7.5 NaCl $MgCl_2$ Glycerol ddH_2O	20 mM 1 M 5 mM 5% (v/v)
Cas batch buffer	Tris/HCl pH 7.5 NaCl $MgCl_2$ Glycerol Tween20 ddH_2O	50 mM 200 mM 5 mM 2% (v/v) 0.1% (v/v)
10x Cas buffer	Tris/HCl pH 7.5 NaCl Glycerol DTT $MgCl_2$ Tween20 ddH_2O	200 mM 1 M 20% (v/v) 10 mM 100 mM 0.5% (v/v)

CCMB80 buffer (4 °C)	KOAc pH 7.0 $CaCl_2$ $MnCl_2$ $MgCl_2$ Glycerol ddH_2O (adjust pH before adding $MnCl_2$, otherwise manganese dioxide is precipitating)	10 mM 80 mM 20 mM 10 mM 10%
MP buffer	Tris/HCl pH 7.5 NaCl $MgCl_2$ ddH_2O	20 mM 50 mM 5 mM
trans-cleavage buffer	Tris/HCl pH 7.9 NaCl $MgCl_2$ BSA ddH_2O	10 mM 100 mM 10 mM 50 µg/ml

3.5.2 Solutions

Table 7. List of solutions.

Solution	**Composition**	**Concentration**
Ampicillin stock solution (1000x, −20 °C)	Ampicillin (sodium salt) ddH_2O	150 mg/ml
L-Arabinose stock solution (−20 °C)	L-Arabinose ddH_2O	20% (w/v)
Chloramphenicol stock solution (1000x, −20 °C)	Chloramphenicol 70% Ethanol	30 mg/ml
IPTG stock solution (−20 °C)	IPTG ddH_2O	1 M

3.5.3 Culture media

All culture media were autoclaved for 20 min at 121 °C. Heat labile additions, like antibiotics, were added from sterile filtered (pore size 0.22 µm) stock solutions to the cooled media.

Table 8. List of culture media.

Culture medium	Composition	Concentration
LB (Lysogeny Broth) full medium	Yeast extract NaCl Tryptone ddH_2O	0.5% (w/v) 1.0% (w/v) 1.0% (w/v)
LB (Lysogeny Broth) agar plates	Yeast extract NaCl Tryptone Agar ddH_2O	0.5% (w/v) 1.0% (w/v) 1.0% (w/v) 1.5% (w/v)
SOB (Super Optimal Broth) full medium	Yeast extract Tryptone NaCl KCl $MgCl_2$ $MgSO_4$ Glucose (after autoclaving) ddH_2O	0.5% (w/v) 2% (w/v) 10 mM 2.5 mM 10 mM 10 mM 20 mM
SOC (Super Optimal Catabolite) full medium	Yeast extract Tryptone NaCl KCl ddH_2O	0.5% (w/v) 2% (w/v) 10 mM 2.5 mM

3.5.4 Polyacrylamide gel electrophoresis

Table 9. List of buffers and solutions for polyacrylamide gel electrophoresis.

Buffer	Composition	Concentration
10% ammonium persulfate (APS, −20 °C)	APS ddH_2O	10% (w/v)
Coomassie destaining solution	Ethanol Acetic acid ddH_2O	20% (v/v) 10% (v/v)

Coomassie staining solution	Coomassie brilliant blue G250	0.1% (w/v)
	Acetic acid	10% (v/v)
	Ethanol	40% (v/v)
	ddH_2O	
6x DNA loading buffer	Tris/HCl pH 7.5	10 mM
	Bromophenol blue	0.03% (w/w)
	Glycerol	60%
	EDTA	6 mM
	ddH_2O	
10x Laemmli running buffer	Tris/HCl pH 8.3	240 mM
	Glycine	1.92 M
	SDS	35 mM
	ddH_2O	
4x Laemmli sample buffer	Tris/HCl pH 7.5	250 mM
	Glycerol	40% (v/v)
	SDS	8% (w/v)
	β-Mercaptoethanol	8% (v/v)
	Bromophenol blue	0.04% (w/v)
	ddH_2O	
Native Laemmli running buffer	Tris/HCl pH 8.3	240 mM
	Glycine	1.92 M
	ddH_2O	
2x Native loading buffer	Tris/HCl pH 6.8	125 mM
	Ficoll 400	5% (w/v)
	ddH_2O	
15% PAA, TBE	Polyacrylamide (19:1)	15% (v/v)
	TBE-buffer	1x
15% PAA, 6 M Urea, TBE	Polyacrylamide (19:1)	15% (v/v)
	Urea	6 M
	TBE-buffer	1x
PBS (phosphate buffered saline)	NaCl	137 mM
	KCl	2.7 mM
	NaH_2PO_4	10 mM
	KH_2PO_4	2 mM
	ddH_2O	
2x RNA loading buffer	Formamide	95% (v/v)
	SDS	0.02% (w/v)
	Bromophenol blue	0.02% (w/v)
	Xylene cyanol	0.01% (w/v)
	EDTA	1 mM
	ddH_2O	

2x RNA loading buffer (colorless)	Formamide SDS EDTA ddH_2O	95% (v/v) 0.02% (w/v) 1 mM
10% SDS	SDS ddH_2O	10% (w/v)
50x TAE (Tris-Acetate-EDTA)	Acetate EDTA Tris-Base pH 8.0 ddH_2O	1 M 50 mM 2 M
10x TBE (Tris-Borate-EDTA)	Tris Borate EDTA ddH_2O	893 mM 890 mM 27 mM

Table 10. List of polyacrylamide gel compositions.

Gel	Composition	Volume
6% PAA native gel	Polyacrylamide (37.5:1) Tris/HCl pH 8.8 (1.5 M) 10% APS TEMED ddH_2O	1.4 ml 1.82 ml 70 µl 7 µl 3.78 ml
15% PAA native TBE gel	15% PAA, TBE 10% APS TEMED	7 ml 70 µl 7 µl
15% PAA, 6 M Urea, TBE gel	15% PAA, 6 M Urea, TBE 10% APS TEMED	7 ml 70 µl 7 µl
15% PAA, 6 M Urea, TBE gel for high-resolution sequencing gel	15% PAA, 6 M Urea, TBE 10% APS TEMED	50 ml 500 µl 50 µl
10% SDS resolving gel	Polyacrylamide (37.5:1) Tris/HCl pH 8.8 (1.5 M) 10% SDS 10% APS TEMED ddH_2O	1.67 ml 1.3 ml 50 µl 50 µl 5 µl 1.93 ml
5% SDS stacking gel	Polyacrylamide (37.5:1) Tris/HCl pH 6.8 (1 M) 10% SDS 10% APS TEMED ddH_2O	0.33 ml 0.25 ml 20 µl 20 µl 2 µl 1.4 ml

3.6 Molecular biology kit systems

All kit systems were used according to the manufacturers manual.

Table 11. List of commercially available laboratory kit systems.

Kit	Producer
NucleoSpin Gel and PCR clean-up Kit	Macherey-Nagel, Düren (DE)
NucleoSpin Plasmid	Macherey-Nagel, Düren (DE)
Protein Thermal Shift	ThermoFisher Scientific, Waltham (MA, USA)
T7 RiboMAX Express Large Scale RNA Production System	Promega Corporation, Madison (WI, USA)

3.7 Electrophoresis ladders

To estimate the size of gel bands of proteins and DNA, ladders were used. DNA ladders were prepared according to manufacturer's instructions. Per gel slot 5–8 µl of ladder were applied.

Table 12. List of electrophoresis ladders.

Ladder	Producer
PageRuler prestained protein ladder	ThermoFisher Scientific, Waltham (MA, USA)
PageRuler unstained protein ladder	ThermoFisher Scientific, Waltham (MA, USA)
GeneRuler 1 kb Plus DNA Ladder	ThermoFisher Scientific, Waltham (MA, USA)

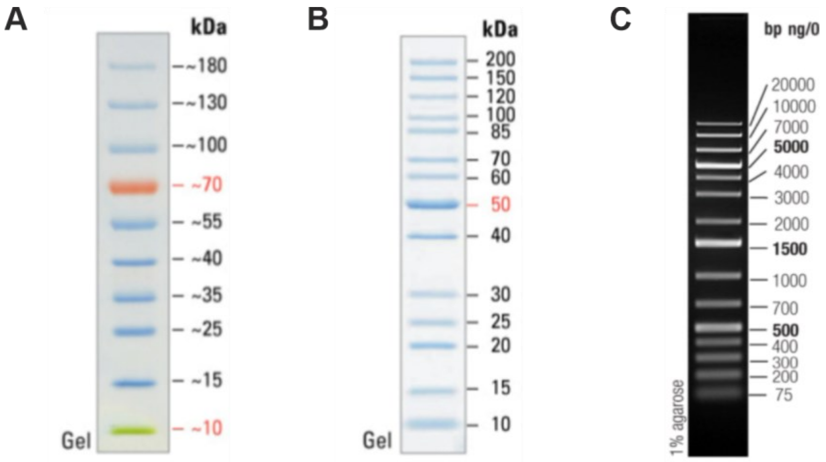


Figure 7: Electrophoresis ladders.
(**A**) PageRuler prestained protein ladder, (**B**) PageRuler unstained protein ladder, and (**C**) GeneRuler 1 kb Plus DNA Ladder. (Pictures derived from www.thermofisher.com)

3.8 Enzymes

Table 13. List of utilized enzymes.

Enzyme	**Producer**
T4 DNA ligase (400 U/µl)	New England Biolabs, Ipswich (MA, USA)
Fast AP (alkaline Phosphatase, 1 U/µl))	ThermoFisher Scientific, Waltham (MA, USA)
FastDigest BamHI (10 U/µl)	ThermoFisher Scientific, Waltham (MA, USA)
FastDigest DpnI (10 U/µl)	ThermoFisher Scientific, Waltham (MA, USA)
FastDigest EcoRI (10 U/µl)	ThermoFisher Scientific, Waltham (MA, USA)
Pierce Nuclease Mix (250 U/µl)	ThermoFisher Scientific, Waltham (MA, USA)
Phusion high fidelity DNA polymerase (2 U/µl)	ThermoFisher Scientific, Waltham (MA, USA)
Proteinase K (>0.6 U/µl)	ThermoFisher Scientific, Waltham (MA, USA)
SmaI (20 U/µl)	New England Biolabs, Ipswich (MA, USA)
Thrombin Protease (500 U)	Sigma-Aldrich, St. Louis (MO, USA)

Table 14. List of utilized enzyme reaction buffers.

Enzyme reaction buffer	**Producer**
rCutSmart (10x)	New England Biolabs, Ipswich (MA, USA)
FastDigest buffer (10x)	ThermoFisher Scientific, Waltham (MA, USA)
FastDigest Green buffer (10x)	ThermoFisher Scientific, Waltham (MA, USA)
Phusion HF Puffer (5x)	ThermoFisher Scientific, Waltham (MA, USA)
T4 ligase buffer (10x)	New England Biolabs, Ipswich (MA, USA)

3.9 Bacterial strains

Table 15. List of bacterial strains.

Bacterial strain	**Genotype**	**Comment**	**Supplier**
E. coli BL21(DE3) (chemically competent)	fhuA2 ompT gal (λ DE3) ΔhsdS λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5:: T7gene1) i21 Δnin5	Expression strain, copy of a T7 RNA-polymerase gene, controlled by lacUV5 promotor, expression inducible by IPTG	Merck, Darmstadt (DE)
E. coli BL21(DE3)-pEVOL-pAzF (chemically competent)	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pEVOL-pAzF (CamR)	Expression strain, increased transformation efficiency, expression inducible by IPTG, orthogonal tRNA synthetase and tRNA for AzF	Peter G. Schultz, Scripps Research Institute, La Jolla (CA, USA) [248,249]

E. coli DH5α (chemically competent)	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17	plasmid amplification strain, efficient transformation of unmethylated DNA	New England Biolabs, Ipswich (MA, USA)

3.10 Vectors and plasmids

Table 16. List of vectors and plasmids.

Vector	Resistance	Length	Supplier
pEVOL-pAzF	Chloramphenicol	6125 bp	Peter G. Schultz, Scripps Research Institute (La Jolla, CA, USA)
pGEX-2TK	Ampicillin	4969 bp	GE Healthcare, Chicago (IL, USA)
pY002 (pFnCpf1_min) #69975	Chloramphenicol	8410 bp	Addgene, Watertown (MA, USA)
EGFP-hAgo2 #21981	Kanamycin	7280 bp	Addgene, Watertown (MA, USA)

Vector maps are displayed in Appendix I.

3.11 Oligonucleotides

3.11.1 Primers for cloning reactions

The following oligodeoxyribonucleotide primers were obtained from Eurofins Genomics, Luxemburg (LUX).

Table 17. List of primers used for cloning.

Name	Sequence (5' → 3')	Specifics
FnCpf1_fw BamHI	`ATATATGGATCCATGTCAATTTATCAAGAATTTG`	Introduction of a BamHI cleavage site
FnCpf1_rev EcoRI	`ATATATGAATTCTTAGTTATTCCTATTCTGCAC`	Introduction of a EcoRI cleavage site

3.11.2 Primers for site directed mutagenesis

The following oligodeoxyribonucleotide primers were obtained from (a) Eurofins Genomics, Luxemburg (LUX) or (b) Microsynth, Balgach (CH).

Table 18. List of primers for site directed mutagenesis.
Mutated nucleotides are underlined, and positions of deletions are indicated by a vertical line.

Name	Mutation	Sequence (5' → 3')	
M9f	E1006A	CTATTGTGGTTTTTGCGGATTTAAATTTTG	(a)
M9r		CAAAATTTAAATCCGCAAAAACCACAATAG	(a)
M11f	K647stop	GATGATAAAGCTATCTAGGAAAATAAAGGCGAG	(a)
M11r		CTCGCCTTTATTTTCCTAGATAGCTTTATCATC	(a)
M23f	T1222stop	CAAATGCGTAACTCAAAATAGGGTACTGAGTTAGATTATC	(a)
M23r		GATAATCTAACTCAGTACCCTATTTTGAGTTACGCATTTG	(a)
M24f	D470stop	GCATAGAGATATATAGAAACAGTGTAGG	(a)
M24r		CCTACACTGTTTCTATATATCTCTATGC	(a)
M28f	I960P	CATGATAAGCTTGCTGCACCAGAGAAAGATAGGGATTC	(a)
M28r		GAATCCCTATCTTTCTCTGGTGCAGCAAGCTTATCATG	(a)
M29f	W971D	GGATTCAGCTAGGAAAGACGACAAAAAGATAAATAACATC	(a)
M29r		GATGTTATTTATCTTTTTGTCGTCTTTCCTAGCTGAATCC	(a)
M30f	W971K	GATTCAGCTAGGAAAGACAAGAAAAAGATAAATAACATC	(a)
M30r		GATGTTATTTATCTTTTTCTTGTCTTTCCTAGCTGAATC	(a)
M31f	W971A	GATTCAGCTAGGAAAGACGCCAAAAAGATAAATAACATC	(a)
M31r		GATGTTATTTATCTTTTTGGCGTCTTTCCTAGCTGAATC	(a)
M32f	W971F	GATTCAGCTAGGAAAGACTTTAAAAAGATAAATAACATC	(a)
M32r		GATGTTATTTATCTTTTTAAAGTCTTTCCTAGCTGAATC	(a)
M33f	ΔBH (Y953–K969)	GGTAATGATAGAATGAAAAC | AAACGACTGGAAAAAGATAAATAAC	(a)
M33r		GTTATTTATCTTTTTCCAGTCGTTTG | TTTTCATTCTATCATTACC	(a)
M34f	ΔΔBH (Y953–W971)	GGTAATGATAGAATGAAAAC | AAACAAAAAGATAAATAACATCAAAGAG	(b)
M34r		CTCTTTGATGTTATTTATCTTTTTG | TTTGTTTTCATTCTATCATTACC	(b)
M35f	K823A	CTTCAAGATGTGGTTTATGCTCTAAATGGTGAGGCAGAG	(b)
M35r		CTCTGCCTCACCATTTAGAGCATAAACCACATCTTGAAG	(b)
M36f	K1066A	CCTTTTGAGACTTTTAAAGCTATGGGTAAACAAACAGGT	(b)
M36r		ACCTGTTTGTTTACCCATAGCTTTAAAAGTCTCAAAAGG	(b)
M37f	I195R	GTTTATAGTAGCAATGATAGACCTACATCTATTATTTATAGG	(b)
M37r		CCTATAAATAATAGATGTAGGTCTATCATTGCTACTATAAAC	(b)
M38f	D194R	GTTTATAGTAGCAATGATATTAGAACATCTATTATTTATAGG	(b)
M38r		CCTATAAATAATAGATGTTCTAATATCATTGCTACTATAAAC	(b)

M39f	N188R	GGTTTTCATGAAAATAGAAAAAGAGTTTATAGTAGCAATGATATTCC	(b)
M39r		GGAATATCATTGCTACTATAAACTCTTTTTCTATTTTCATGAAAACC	(b)
M40f	K978A	GGAAAAAGATAAATAACATCGCTGAGATGAAAGAGGG	(b)
M40r		CCCTCTTTCATCTCAGCGATGTTATTTATCTTTTTCC	(b)
M41f	K981P	GATAAATAACATCAAAGAGATGCCTGAGGGCTATCTATCTCAGG	(b)
M41r		CCTGAGATAGATAGCCCTCAGGCATCTCTTTGATGTTATTTATC	(b)
M42f	ΔBH/Δh1 (Y953–N1000)	GGTAATGATAGAATGAAAACAAAC | GCTATTGTGGTTTTTGAGGATTTAAATTTTGG	(b)
M42r		CCAAAATTTAAATCCTCAAAAACCACAATAG | CGTTTGTTTTCATTCTATCATTACC	(b)
M43f	Δh1 (K972–N1000)	GGGATTCAGCTAGGAAAGACTGG | GCTATTGTGGTTTTTGAGGATTTAAATTTTGG	(b)
M43r		CCAAAATTTAAATCCTCAAAAACCACAATAGC | CCAGTCTTTCCTAGCTGAATCCC	(b)
M45f	H843A K852A	CCTAAAAAAATCACTGCTCCAGCTAAAGAGGCAATAGCTAATGCTAACAAAGATAATCC	(b)
M45r		GGATTATCTTTGTTAGCATTAGCTATTGCCTCTTTAGCTGGAGCAGTGATTTTTTTAGG	(b)
M46f	K869A	GTGTTTTTGAATATGATTTAATCGCTGATAAACGCTTTACTGAAGATAAG	(b)
M46r		CTTATCTTCAGTAAAGCGTTTATCAGCGATTAAATCATATTCAAAAACAC	(b)

Table 19. List of primers used for commercial Sanger sequencing.

Name	Sequence (5' → 3')	Binding site in *FnCas12a*	
Seq6	AAAGACAAAGCTCCAGAAGC	nt 664–683 forward	(a)
Seq7	AAGCTCAAAAACTTGATTTGAGTAAA	nt 1144–1170 forward	(a)
Seq8	CTCAAAAGCCATATAGTGATGAGAAA	nt 1759–1785 forward	(a)
Seq9	GGTTTATAAGCTAAATGGTGAGGC	nt 2459–2483 forward	(a)
Seq10	GAGAAGCAGGTCTATCAAAAGTT	nt 3057–3080 forward	(a)

3.11.3 Oligonucleotides for analysis

The following DNA and RNA oligonucleotides were obtained from (a) Eurofins Genomics, Luxemburg (LUX), (b) Microsynth, Balgach (CH), (c) Biomers, Ulm (DE), (d) IBA Lifesciences, Göttingen (DE), or (e) Integrated DNA Technologies (IDT), Coralville (IA, USA). The oligonucleotides were ordered at smallest production scales possible. Non-modified DNA constructs were purified salt-free, modified oligos via HPLC or PAGE purification. Upon arrival, the concentration was checked via absorption spectroscopy and the oligonucleotides stored at −20 °C.

Table 20. List of DNA and RNA oligonucleotides.
PAM sequences are underlined and bold, mismatches in MM DNA strands are colored in pink, fluorescent dyes are indicated.

Name	Sequence (5' → 3')	
KW3single (TS)	GGTCGACTCTAGAGGATCCCCGGGTACCGAGCTC**GAA**TTCG**TAA**TCATGG T= coupled to Atto532	(d)
KW5ts (TS)	ACTCAATTTTGACAGCCCACATGGCATTCCACTTATCAC**TAAA**GGCATCCTTCCACGT [133]	(a)
KW5ts-Cy5 5' (TS)	Cy5-ACTCAATTTTGACAGCCCACATGGCATTCCACTTATCAC**TAAA**GGCATCCTTCCACGT	(a)
KW5ts-Cy5 3' (TS)	ACTCAATTTTGACAGCCCACATGGCATTCCACTTATCAC**TAAA**GGCATCCTTCCACGT-Cy5	(a)
KW5tsMM2-Cy5 5' (TS)	Cy5-ACTCAATTTTGACAGCCCACATGGCATTCCACTTATCTC**TAAA**GGCATCCTTCCACGT	(a)
KW5tsMM10-Cy5 5' (TS)	Cy5-ACTCAATTTTGACAGCCCACATGGCATTCGACTTATCAC**TAAA**GGCATCCTTCCACGT	(a)
KW5tsMM20-Cy5 5' (TS)	Cy5-ACTCAATTTTGACAGCCCAGATGGCATTCCACTTATCAC**TAAA**GGCATCCTTCCACGT	(a)
KW5nts (NTS)	ACGTGGAAGGATGCC**TTTA**GTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGT [133]	(a)
KW5nts-Cy3 5' (NTS)	Cy3-ACGTGGAAGGATGCC**TTTA**GTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGT	(a)
KW5nts-Cy3 3' (NTS)	ACGTGGAAGGATGCC**TTTA**GTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGT-Cy3	(a)
KW5ntsMM2-Cy3 5' (NTS)	Cy3-ACGTGGAAGGATGCC**TTTA**GAGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGT	(a)
KW5ntsMM10-Cy3 5' (NTS)	Cy3-ACGTGGAAGGATGCC**TTTA**GTGATAAGTCGAATGCCATGTGGGCTGTCAAAATTGAGT	(a)
KW5ntsMM20-Cy3 5' (NTS)	Cy3-ACGTGGAAGGATGCC**TTTA**GTGATAAGTGGAATGCCATCTGGGCTGTCAAAATTGAGT	(a)
KWcpf1#2	UAAUUUCUACUGdTUGUAGAUCGAAUUCGAGCUCGGUACCCGGGG dT = coupled to Atto647N	(d)
KW5RNA-Atto532 5'	Atto532-UAAUUUCUACUGUUGUAGAUGUGAUAAGUGGAAUGCCAUGUGGG	(c)

KW5RNA-Cy5 3'	UAAUUUCUACUGUUGUAGAUGUGAUAAGUGGAAUGCCAUGUGGG-Cy5	(b)
KW5RNA-Cy3 5'	Cy3-AAUUUCUACUGUUGUAGAUGUGAUAAGUGGAAUGCCAUGUGGG	(b)
FQ substrate	FAM-TTTTTTTTTT-IABk	(e)

Table 21. List of DNA oligonucleotides used for T7 *in vitro* transcription to synthesize RNA strands.

Name	Sequence (5' → 3')	
T7 promotor	GGGATATCACTCAGCATAATAGCTAG	(a)
T7crGFP	GGTCGAGCTGGACGGCGACGATCTACAACAGTAGAAATTACCCTATAGTGAGTCGTATTATCGATC	(a)
T7crKW5	CCCACATGGCATTCCACTTATCACATCTACAACAGTAGAAATTACCCTATAGTGAGTCGTATTATCGATC	(a)
T7pre-crRNA KW5	CCCACATGGCATTCCACTTATCACATCTACAACAGTAGAAATTATTTAAAGTTCTTAGACCCGTTTTTGCCTAAATCAGCCCCTATAGTGAGTCGTATTATCGATC	(a)

Table 22. List of T7-transcribed RNA.

Name	Sequence (5' → 3')
crGFP	UAAUUUCUACUGUUGUAGAUCGUCGCCGUCCAGCUCGACC
crKW5	UAAUUUCUACUGUUGUAGAUGUGAUAAGUGGAAUGCCAUGUGGG
pre-crRNA KW5	GCUGAUUUAGGCAAAAACGGGUCUAAGAACUUUAAAUAAUUUCUACUGUUGUAGAUGUGAUAAGUGGAAUGCCAUGUGGG

3.12 Fluorescent Dyes

Table 23. List of fluorescent dyes.
All dyes were solved in water-free DMSO to a final concentration of 10 mM.

Name	Producer
DyLight 550 Phosphine (DL550)	ThermoFisher Scientific, Waltham (MA, USA)
DyLight 650 Phosphine (DL650)	ThermoFisher Scientific, Waltham (MA, USA)

4. METHODS

4.1 Microbiological methods

4.1.1 Preparation of equipment and solutions

Glassware and heat-stable consumables like pipette tips and tubes were sterilized by autoclaving for 40 min at 121 °C and 2 bar and subsequently dried at 60 °C in a drying cabinet. Thermostable solutions such as media were autoclaved for 20 min at 121 °C and 2 bar prior to use. Heat-labile solutions were sterile filtered with a syringe filter (0.22 µm, Roth-Chemie, Karlsruhe (DE)). Solutions for chromatographic systems were filtered sterile (Whatman filter Ø 33 mm, 0.22 µm pore size, Roth-Chemie; Filtration Unit, Nalgene) and degassed for at least 30 min in an ultrasonic bath (Elmasonic, Elma Schmidbauer GmbH, Singen (DE)) prior to use.

4.1.2 Cell culture and storage of *E. coli* strains

Escherichia coli (E. coli) strains were grown at 37 °C in liquid LB media unless stated otherwise. The culture vessels were shaken at 160 rpm. For strains with selection markers, the respective antibiotic was added to the media (from a 1000-fold concentrated stock solution). The end concentration for ampicillin was 150 µg/ml and for chloramphenicol 30 µg/ml. The growth of the cells was tracked by photometric measurement of the optical density at 600 nm (OD_{600}).

For the growth of single colonies, the cell suspension was plated on selective agar plates and incubated at 37 °C. For temporary storage, plates and suspensions were sealed and stored at 4 °C. For permanent storage of *E. coli* strains, a 1 ml aliquot of an overnight culture was mixed in a sterile tube with 0.3 ml of 75% glycerol (v/v), flash-frozen in liquid nitrogen, and stored at −80 °C.

4.1.3 Preparation of chemically competent *E. coli* cells

For the preparation of chemically competent cells, 250 ml SOB medium was inoculated with 1 ml over-night preculture and incubated at 20 °C and 160 rpm. At a maximal OD_{600} of 0.3, the cells were harvested through centrifugation (JA.10 rotor, 800 g, 4 °C, 10 min) and the supernatant was discarded. The pellet was gently resuspended on ice in 80 ml ice-cold CCMB80 buffer and incubated on ice for 20 min. The cells were centrifuged a second time (1000 g, 4 °C, 10 min). The supernatant was discarded, and the pellet was resuspended in

10 ml ice-cold CCMB80 buffer. The OD_{600} was determined for a 1:5 dilution with SOB medium and cold CCBM80 buffer was added to yield a final OD_{600} of 5.0. After incubation on ice for 20 min, the cells were aliquoted in 100 µl, flash-frozen in liquid nitrogen, and stored at −80 °C. [250]

4.1.4 Chemical transformation of *E. coli* cells

50 µl of chemically competent *E. coli* DH5α (for plasmid amplification), *E. coli* BL21(DE3) (for protein expression), or *E. coli* BL21(DE3) pEVOL-pAzF (for expression of modified proteins based on the Amber-suppression strategy [248]) cells were thawed on ice for approximately 10 min. 100–200 ng of plasmid were added and incubated on ice for 20 min. After a heat shock (45 s at 42 °C), the cells were incubated on ice for another 5 min and mixed with 250 µl of warm (37 °C) SOC medium. After curing for 1 h at 37 °C and 160 rpm, 100–200 µl of the transformation mix were plated on LB plates supplemented with the relevant antibiotics (150 µg/ml ampicillin and/or 30 µg/ml chloramphenicol) and incubated for 12–16 h at 37 °C.

4.1.5 Disposal of microorganisms

Bacterial cultures and contaminated equipment were autoclaved at 121 °C for 40 min and 2 bar prior to disposal.

4.2 Molecular biological methods

4.2.1 Nucleic acid concentration determination by absorption spectroscopy

Concentrations of DNA in aqueous solutions were measured spectroscopically using a Nanodrop spectrometer (NanoDrop One^{C}, ThermoFisher Scientific) at a wavelength of 260 nm. The DNA concentrations were calculated using the Lambert-Beer law:

$$E = log\frac{I_0}{I} = \varepsilon_\lambda \cdot c \cdot d \qquad \textbf{Equation 1}$$

E	absorbance
I_0	intensity of the incident light
I	intensity of the transmitting light
ε_λ	extinction coefficient ($l\ mol^{-1}\ cm^{-1}$) at a given wavelength λ
c	concentration (mol/l)
d	length of the light path (cm)

Nucleic acid concentrations are calculated based on the absorbance value at 260 nm, with ε depending on the nucleic acid sample (dsDNA: $\varepsilon = 50\ (\mu g/ml)^{-1}\ cm^{-1}$; ssDNA: $\varepsilon = 33\ (\mu g/ml)^{-1}\ cm^{-1}$; ssRNA: $\varepsilon = 40\ (\mu g/ml)^{-1}\ cm^{-1}$).

Before each measurement, a blank (buffer-only) spectrum was measured and subtracted from the subsequently recorded spectra. For DNA and RNA, 1 µl of the sample or the respective reference was used. Each sample was measured in triplicates and average values were used for the final concentration calculation. A pure DNA solution has no measurable absorption above 300 nm, and an A_{260}/A_{280} ratio of at least 1.8.

4.2.2 Polymerase chain reaction

The polymerase chain reaction (PCR) is used to amplify defined DNA fragments. Two oligonucleotide primers are used to limit the fragment on each end. In some instances in this thesis, primers carried a sequence to introduce restriction endonuclease recognition sites. The template DNA, a thermostable DNA polymerase, and deoxyribonucleotides (dNTPs) are used in a buffer containing magnesium ions to stabilize the phosphates. The PCR is a cyclic reaction in three steps: (i) denaturation of the double-stranded template DNA, (ii) primer hybridization to the ssDNA, and (iii) elongation of the DNA by a polymerase starting at the 3'-end of the primer.

Here, the FnCas12a gene was amplified from the plasmid pY002 (pFnCpf1_min, Addgene #69975) using the primers shown in **Table 17**.

The PCR reaction mixture included the following components:

Table 24. PCR reaction mix.

5x HF buffer	4 µl
DMSO	0.6 µl
dNTP-mix (10 mM)	0.4 µl
forward primer (10 µM)	1.25 µl
reverse primer (10 µM)	1.25 µl
DNA template	≈ 20 ng
Phusion polymerase (2 U/µl)	1 µl
nuclease-free H_2O	ad 20 µl

The reaction was carried out under hot start conditions in a preheated thermocycler (Eppendorf) with a lid temperature of 105 °C. The PCR program was carried out as listed in **Table 25**:

Table 25. PCR program.

1. Initial denaturation	98 °C	30 s
2. Denaturation	98 °C	10 s
3. Annealing	55 °C	20 s
4. Extension	72 °C	2 min
5. Final extension	72 °C	10 min
6. Storage	16 °C	∞

Steps 2 to 4 were repeated 30 times. The extension time was calculated considering the polymerization speed of the DNA polymerase used (approximately 40 s/kb). The optimum melting temperature T_m (°C) and annealing temperature T_A (°C) of the primers were calculated according to the following equations:

$$T_m = 69.3 + 0.41 \cdot (\%GC) - \frac{650}{N} \qquad \textbf{Equation 2}$$

$$T_A = \left(\frac{T_{m1}+T_{m2}}{2}\right) - 3\ °C \qquad \textbf{Equation 3}$$

%GC GC-content of primers in %
N number of nucleotides
T_{m1} melting temperature of forward primer
T_{m2} melting temperature of reverse primer

To remove interfering PCR components, the amplified DNA was purified using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel) and eluted in a volume of 43 µl elution buffer.

4.2.3 DNA cloning

DNA cloning describes the insertion of a linear DNA fragment into a circular plasmid (vector) via restriction endonuclease (RE) digestion. The insert and the vector are digested with the same RE, creating the same overhangs at the DNA ends. In the next step, these ends are covalently linked by a ligase. To ensure the right orientation of the insert, both ends are cleaved with different REs. Additionally, the vector is treated with alkaline phosphatase to prevent re-ligation by removing the 5'-phosphate.

Here, the FnCas12a gene was cloned into the pGEX-2TK vector (GE Healthcare). This resulted in an expression plasmid that encoded the wildtype FnCas12a sequence with an N-terminal GST-tag and a Thrombin cleavage site in between the FnCas12a sequence and the sequence for the GST-tag.

4.2.3.1 DNA restriction

REs recognize specific palindromic motifs and cleave dsDNA resulting in sticky or blunt ends. For restriction digestion, mainly REs that produce sticky ends are used, facilitating the oriented ligation with another DNA fragment [251].

The DNA restriction mixture included the following components:

Table 26. Reaction mix for DNA restriction.

10x FastDigest green buffer	5 µl
FastDigest BamHI (10 U/µl)	1 µl
FastDigest EcoRI (10 U/µl)	1 µl
DNA template (purified in 4.2.2, ≈ 200 ng/µl)	43 µl

The restriction mix was incubated at 37 °C for 1–2 h. 1 µl of alkaline phosphatase was additionally added to the vector restriction mix and incubated for 1 h at 37 °C. The mix was completely loaded onto a preparative 1% agarose gel. After electrophoresis (4.2.6), the band of the DNA fragment with the correct length was excised from the gel and purified with the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel).

4.2.3.2 DNA ligation

During DNA ligation 3'- and 5'-ends of DNA fragments are covalently linked by a ligase via the formation of a phosphate diester bond. Hybridization of ssDNA overhangs at the ends, created by sticky-end-cleavage of REs, facilitates the ligation reaction. The DNA ligation mixture included the following components:

Table 27. Reaction mix for DNA ligation.

restricted vector	100 ng
restricted insert	threefold molar excess
10x T4 DNA ligase buffer	2 µl
T4 DNA ligase (400 U/µl)	1 µl
nuclease-free H_2O	ad 20 µl

The volume of the insert was calculated as follows:

$$\frac{V(insert)}{V(vector)} = 3 \cdot \frac{c(vector) \cdot N_{nt}(insert)}{N_{nt}(vector) \cdot c(insert)} \qquad \textbf{Equation 4}$$

V	volume
N_{nt}	number of nucleotides (length)

The ligation mix was incubated at room temperature for 20 min, prior to incubation at 65 °C for 10 min to inactivate the DNA ligase. The complete ligation mix was transformed into *E. coli* DH5α cells (4.1.4) for plasmid preparation.

4.2.4 PCR-mediated site-specific mutagenesis

For the introduction of a site-specific mutation into a gene encoded on a plasmid, a PCR with specifically designed mutagenesis primers was utilized. The primers contain a mutation compared to the wildtype sequence, coding for a different amino acid [252]. To compensate for the introduced mismatches in the primers, the primer sequence is extended to ensure stable primer hybridization by more Watson-Crick base pairings despite a mismatched region. The PCR reaction mixture included the following components:

Table 28. Reaction mix for site-specific mutagenesis.

5x HF buffer	10 µl
dNTP-mix (10 mM)	1 µl
forward primer (10 µM)	1 µl
reverse primer (10 µM)	1 µl
DNA template (plasmid with the gene)	≈ 50 ng
Phusion polymerase (2 U/µl)	1 µl
nuclease-free H_2O	ad 50 µl

Sequences of the primers used are shown in **Table 18**. The reaction was carried out under hot start conditions in a preheated thermocycler (Eppendorf) with a lid temperature of 105 °C. The PCR program was composed as follows:

Table 29. PCR program for site-specific mutagenesis.

1. Initial denaturation	98 °C	30 s
2. Denaturation	98 °C	10 s
3. Annealing	55 °C	20 s
4. Extension	72 °C	5 min
5. Final extension	72 °C	10 min
6. Storage	16 °C	∞

Steps 2 to 4 were repeated 20 times.

5 µl of the PCR reaction were separated on a 0.9% agarose gel (4.2.6) to monitor the efficiency of the mutagenesis PCR. To remove interfering PCR components, the rest of the amplified DNA was purified using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel) and eluted in a volume of 43 µl.

To digest the template DNA, the purified PCR product was incubated with DpnI for 1–2 h at 37 °C. The restriction enzyme DpnI (recognition sequence GATC) is specific for methylated and hemi-methylated DNA, only leading to the degradation of template DNA that was purified from *E. coli* cells but not the PCR-amplified DNA.

Table 30. Reaction mix for DpnI digestion.

purified PCR product	43 µl
10x FastDigest buffer	5 µl
FastDigest DpnI (10 U/µl)	2 µl

For plasmid propagation, 100–200 ng of the mutated plasmid were transformed into *E. coli* DH5α (4.1.4) and isolated according to 4.2.5. The successful introduction of the mutations was confirmed by commercial Sanger sequencing (4.2.7).

4.2.5 Isolation and purification of plasmid DNA from *E. coli*

Purification of plasmid DNA was performed according to the principle of alkaline cell lysis [253]. Bacterial cells are lysed by the addition of SDS and sodium hydroxide. After neutralization with ammonium acetate, proteins and genomic DNA remain precipitated, whereas the covalently closed circular plasmid DNA is restored quickly and gets redissolved in the solution.

The isolation and purification of plasmid DNA from *E. coli* cells was performed with the NucleoSpin Plasmid kit (Macherey-Nagel). For this purpose, cells of a 5-ml overnight culture were grown in a 15-ml tube. The cells were harvested by centrifugation (10 min, 4000 rpm, 4 °C, EPPENDORF Centrifuge 5810R). The isolation of the DNA from the cell pellet was carried out according to the manufacturer's instructions. The DNA was eluted in 50 µl nuclease-free H_2O and the DNA concentration was determined as described in 4.2.1. The DNA was stored at −20 °C.

4.2.6 Agarose gel electrophoresis

Agarose gel electrophoresis was used for the analysis of DNA fragments after PCR and cleavage assays. In an electrophoresis, charged molecules are separated by migration in an electric field. Nucleic acids have a strong negative charge at the phosphate backbone, therefore, in an electric field, they move from the cathode (negative pole) to the anode (positive pole) [254].

Agarose is a polysaccharide that forms a network with pores that are smaller at higher agarose concentrations. DNA fragments move through the gel along the electric field and are separated according to their length, with large nucleic acids migrating slower through the pores, whereas small fragments can pass the pores easily. In the presence of an intercalating dye like SYBR-Safe, DNA becomes visible under UV light.

For the gel, the respective amount of agarose (0.8%, 0.9%, 1%, or 1.5% (w/v)) was dissolved by boiling in 1x TAE buffer. After cooling to approximately 40 °C, 3.2 µl SYBR-Safe (per 50 ml agarose solution, ThermoFisher Scientific) was added. The agarose solution was poured into the agarose gel system Mini Sub cell GT (BioRad) with a suitable comb to form the sample slots of the gel. After complete solidification, the agarose gel was covered with 1x TAE buffer. The DNA sample was mixed with 6x DNA loading dye and the sample was loaded onto the gel. The electrophoresis was carried out at 100 V for 30 min. The length of the DNA fragments can be estimated with the aid of a DNA length standard (GeneRuler 1 kb Plus DNA Ladder, ThermoFisher Scientific). For documentation, the gels were visualized using a ChemiDoc Imaging System (590/110 nm standard filter, BioRad).

4.2.7 DNA sequencing

Sequencing service for plasmids containing FnCas12a genes was provided by Microsynth, Balgach (CH) or Macrogen, Amsterdam (NL). The primers for sequencing were either chosen from the standard primer list of the company (pGEX fw, GEX3 rev, GEX5 for) or Cas12a-specific sequencing primers were supplied with the DNA sample to be sequenced (**Table 19**). Sequencing results were analyzed using the program Serial Cloner (SerialBasics) to verify that the point mutations and deletions were inserted at the correct site in the *FnCas12a* gene.

4.2.8 RNA *in vitro* transcription

For RNA *in vitro* transcription, a single-stranded DNA template with a short double-stranded promotor region is used to produce single-stranded RNA by a T7 RNA polymerase.

Using the T7 RiboMax Express large-scale RNA Production System kit (Promega), RNA was transcribed from DNA templates (**Table 21**) that are reverse complementary to the desired RNA sequence, additionally containing a 26 nt T7 promotor sequence at the 3'-end. First, the DNA template was annealed to a short DNA oligonucleotide complementary to the T7 promoter sequence.

Table 31. T7 annealing reaction.

5x Annealing buffer	5 µl
T7 promotor DNA (80 µM)	3.13 µl
DNA template (100 µM)	2.5 µl
nuclease-free H_2O	ad 25 µl

The annealing reaction was heated to 95 °C for 3 min and passively cooled to room temperature to allow annealing of the DNA template and the T7 promotor sequence.

According to the manufacturer's instructions of the kit, RNA transcription was performed with the annealed template for 30 min at 37 °C. After digestion with RQ1 DNase for 15 min at 37 °C, the reaction mix was cleaned using PD SpinTrap G-25 columns. The columns were prepared by five washing steps with 400 µl nuclease-free H_2O and centrifugation at 800 g, 4 °C, for 1 min. The T7 reaction was loaded onto the column and eluted by centrifugation at 800 g, 4 °C, for 2 min.

The RNA was purified by phenol-chloroform extraction. To this end, 400 µl phenol-chloroform-isoamyl alcohol was added to the eluted RNA, mixed thoroughly, and centrifuged at 8000 rpm, 4 °C, for 15 min. The aqueous phase containing the RNA was transferred to a new reaction cup and mixed with 40 µl sodium acetate and 1 ml ice-cold 100% ethanol to precipitate the RNA. The mix was incubated at −20 °C for at least 1 h. The RNA was pelleted by centrifugation at 15000 rpm, 4 °C, for 30 min. The supernatant was discarded, and the pellet was washed with 1 ml ice-cold 80% ethanol and centrifuged at 15000 rpm, 4 °C, for 10 min. After quantitatively removing the supernatant, the pelleted RNA was air-dried and resuspended in 50–200 µl nuclease-free H_2O. RNA concentrations were determined as described in 4.2.1. The RNA was stored in aliquots at −20 °C.

4.3 Protein biochemical methods

4.3.1 Recombinant expression of FnCas12a variants in *E. coli*

Protein expression of FnCas12a WT and FnCas12a variants was performed in *E. coli* BL21(DE3) cells. 1 l of LB medium supplemented with 150 µg/ml ampicillin was inoculated with a preculture of *E. coli* BL21(DE3) cells transformed with the pGEX-2TK vector containing the FnCas12a gene and grown at 37 °C and 160 rpm. At an OD_{600} of 0.4, the cells were transferred to 18 °C. Protein expression was induced with IPTG (final concentration: 1 mg/l) at an $OD_{600} = 0.6$. The cells were further incubated at 18 °C and 160 rpm overnight to allow protein production. The cells were harvested by centrifugation at 6000 g, 4 °C, for 10 min (Beckmann, JA.10). The cell pellet was either lysed and the overexpressed FnCas12a variant was subsequently purified or stored at −80 °C for later use.

4.3.2 Recombinant expression of AzF-modified FnCas12a variants in *E. coli*

To produce fluorescently labeled protein variants of FnCas12a for single-molecule FRET measurements, the Amber-suppressor strategy was employed to site-specifically incorporate the unnatural amino acid *para*-azido-L-phenylalanine (AzF) into the FnCas12a protein [248,249,255]. The Amber-stop-codon (TAG) was introduced into the wildtype *FnCas12a* gene or the *FnCas12a* gene variants that already carried mutations in the bridge helix or helix 1. Amber-stop-codons were introduced via site-specific mutagenesis at gene positions that encode for D470, K647, and/or T1222 in the wildtype (**Figure 22**). The pGEX-2TK-FnCas12a plasmids carrying the corresponding Amber-codons were expressed in BL21(DE3) cells, which additionally harbored the pEVOL-pAzF plasmid (Addgene #31186 [248,249]). This plasmid encodes an additional arabinose-inducible promotor that governs the expression of an Amber-suppressor tRNA ($tRNA^{CUA}$) and a bioorthogonal tRNA synthetase, both originated from *Methanocaldococcus jannaschii*. In that way, the Amber stop codon in *E. coli* is reprogrammed and binds the modified tRNA, which is loaded with the desired unnatural amino acid, instead of the release factor 1 (RF1) in unmodified cells (**Figure 8**).

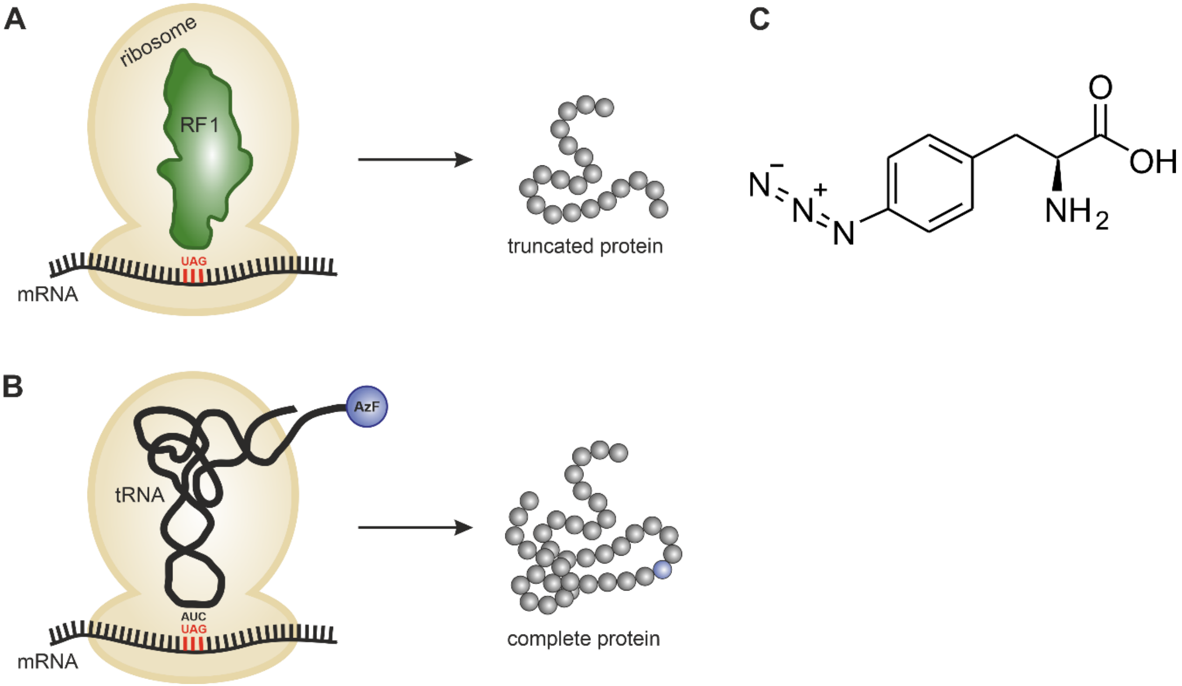


Figure 8: Incorporation of an unnatural amino acid in an *E. coli* expression system.
An orthogonal tRNA synthetase couples an unnatural amino acid (here AzF) to an orthogonal tRNA. During translation, the ribosome recognizes the loaded tRNA that matches the base triplet and adds the next amino acid to the growing polypeptide chain. At the Amber stop codon, the ribosome usually recruits the release factor RF1 (**A**) and releases the polypeptide chain. In this case, it also recognizes the orthogonal tRNA (**B**) which is complementary to the Amber codon and the loaded unnatural amino acid is incorporated into the protein, leading to the full-length protein. (**C**) Structure of *para*-azido-L-phenylalanine (AzF).

1 l of LB medium supplemented with 150 µg/ml ampicillin and 30 µg/ml chloramphenicol were inoculated with a preculture of *E. coli* BL21(DE3) cells transformed with the pGEX-2TK vector containing the FnCas12a gene and grown at 37 °C and 160 rpm. At an OD_{600} of 0.3, 200 mg/ml AzF was added, at an OD_{600} of 0.4, L-Arabinose at a final concentration of 20 mg/l was added and the cells were transferred to 18 °C. Protein expression was induced with IPTG (1 mg/l) at an OD_{600} = 0.6. The cells were further incubated at 18 °C and 160 rpm overnight to allow protein production. The cells were harvested by centrifugation at 6000 g, 4 °C, for 10 min (Beckmann, JA.10). The cell pellet was either lysed and the overexpressed FnCas12a variant was subsequently purified or stored at −80 °C for later use.

4.3.3 Cell lysis and protein extraction

The cell pellet from 4.3.1 or 4.3.2 was resuspended on ice in 20 ml Cas A buffer per liter expression volume. The suspension was mixed with 1 µl nuclease mix (Pierce Nuclease Mix, 250 U/µl, ThermoFisher Scientific) to degrade host nucleic acids and one tablet protease inhibitor (25x cOmplete proteinase inhibitor, Roche Diagnostics) to prevent unspecific protein degradation. Cells were disrupted on ice via sonication with a flat-head nozzle (80% power, 3x 3 min, Cycle 5, 1 min pause, SONOPULS H2070, BANDELIN electronic) and the cell debris was removed via centrifugation (1 h, 25000 rpm, 4 °C, Beckmann, JA25.50). Subsequently, the crude extract was filtered (pore size 0.45 µm). The target protein was purified from the soluble fraction via column chromatography (Äkta purifier, GE Healthcare). FnCas12a and its variants were expressed as fusion proteins with an N-terminal GST-tag and purified via glutathione affinity chromatography (4.3.4) and heparin affinity chromatography (4.3.5).

4.3.4 Glutathione affinity chromatography

The purification of proteins via glutathione affinity chromatography was performed using the chromatography unit Äkta purifier (GE Healthcare) with a GSTrap FF column (GE Healthcare, CV: 5 ml). The column consists of a glutathione Sepharose (FastFlow) matrix. Proteins that are genetically fused to a GST-tag (glutathione-S-transferase) bind specifically to glutathione. A high-salt washing step is included to liberate protein-bound nucleic acids. Elution of the bound protein is achieved by the addition of glutathione to the buffer, which competes with the glutathione fused to the column matrix and liberates the GST-tagged target protein.

The purification of FnCas12a and its variants was performed according to the following protocol.

Table 32. Protein purification via glutathione affinity chromatography.

Flow rate	1 ml/min
Equilibration	3 CV Cas A buffer
Sample application	25 ml protein solution in Cas A buffer (flow rate 0.5 ml/min)
Washing	2 CV Cas A buffer
Washing high salt	10 CV Cas high salt buffer
Washing	2 CV Cas A buffer
Elution	7 CV Cas B buffer
Flushing	5 CV H_2O
Storage of the column	in 20% EtOH

For buffer composition see 3.5.1.

Elution of protein was monitored via absorption measurements at 280 nm. Fractions of 1 ml volume were collected. To check the purity of eluted fractions, 15 µl of protein solution were mixed with 5 µl 4x Laemmli sample buffer and further analyzed by SDS-PAGE (4.3.8). Fractions containing the protein of interest were pooled and incubated overnight with

Thrombin protease (100 U/ml, 50:1 (v:v)) at 4 °C to allow the cleavage of the GST-tag. To remove the GST-tag and the Thrombin protease, further purification was done via heparin affinity chromatography (4.3.5).

4.3.5 Heparin affinity chromatography

For heparin affinity chromatography a HiTrap Heparin HP column (GE Healthcare, CV: 5 ml) was used that was connected to the chromatography unit Äkta purifier (GE Healthcare). The column contains a heparin Sepharose (high performance) matrix for the purification of proteins with affinity to heparin, such as DNA binding proteins. In addition to the intrinsic affinity of proteins to heparin, heparin acts as an efficient cation exchanger due to its anionic sulfate groups. Elution of the target protein is achieved via outcompeting the ionic interactions by a salt gradient.

The purification of FnCas12a and its variants was performed according to the following protocol.

Table 33. Protein purification via heparin affinity chromatography.

Flow rate	1 ml/min
Equilibration	3 CV Cas A buffer
Sample application	10 ml of protein solution (flow rate 0.5 ml/min)
Washing	2 CV Cas A buffer
Elution	8 CV gradient Cas A to Cas high salt buffer
Final wash-out	3 CV Cas high salt buffer
Flushing	5 CV H_2O
Storage of the column	in 20% EtOH

For buffer composition see 3.5.1.

The elution of the protein was monitored via absorption measurements at 280 nm. Fractions of 0.5 ml volume were collected. To check the purity of the eluted fractions, 15 µl of protein solution were mixed with 5 µl 4x Laemmli sample buffer and further analyzed by SDS-PAGE (4.3.8). Aliquots of the purified protein were flash-frozen in liquid nitrogen and stored at −80 °C.

4.3.6 Dialysis of protein solutions

To exchange buffers, protein solutions were dialyzed against at least a 100-fold volume of the respective dialysis buffer. The samples were dialyzed overnight at 4 °C in a dialysis cassette (Slide-A-Lyzer Dialysis Cassette, ThermoFisher Scientific) with a molecular exclusion limit of 100 kDa. The dialysis cassette was equilibrated with buffer solution prior to usage.

4.3.7 Concentrating of protein solutions

If necessary, the purified proteins were concentrated using Amicon centrifugal filter devices (Amicon Ultra 0.5 ml Centrifugal Filters, Merck) with a molecular weight cut-off of 100 kDa via centrifugation (1500 g, 4 °C, VWR Centrifuge CT15RE). The length of centrifugation depended on the desired end volume and concentration of the protein solution.

4.3.8 SDS polyacrylamide gel electrophoresis

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) is a technique to separate proteins according to their molecular mass. By the addition of SDS, β-Mercaptoethanol to break disulfide bonds, and heating the sample to 95 °C, proteins are denatured and dissociate into their single polypeptide chains. SDS, an anionic detergent, attaches to the protein backbone (in a ratio of 1.4 g SDS/1 g polypeptide) resulting in a negative overall charge proportional to the size of the protein. Through voltage application, the proteins are separated according to their size and not by their shape or intrinsic charge [256].

The resolution of an SDS-PAGE is increased by a discontinuous buffer system. The buffer reservoir contains a Tris-Glycine buffer with pH 8.3. The low percent acrylamide stacking gel as well as the applied sample contain Tris-HCl buffers with pH 6.8, whereas the high percent acrylamide resolving gel contains a Tris-HCl buffer with pH 8.8 [257,258]. At the lower pH in the stacking gel, glycine is only weakly ionized and has low mobility in the electric field. The small chloride ions are fast and function as leading ions in front of the proteins that have an intermediate velocity. At the transition from stacking gel to the resolving gel with smaller pore size, the proteins are slowed down, whereas Glycine as the previously slowest fraction and therefore trailing ions becomes negatively charged at the higher pH in the resolving gel and speeds up. Consequently, the glycine ions pass the proteins leading to stacking and concentration of the proteins. In the following resolving gel, they are separated according to their size.

For casting and gel run, the Mini-PROTEAN-Tetra System was used. Protein solutions containing FnCas12a were separated with 10% polyacrylamide resolving and 5% stacking gels. Gel compositions are listed in **Table 10**.

Protein samples were prepared by mixing them at a 1:4 volume ratio with 4x Laemmli sample buffer and denatured for 5 min at 95 °C. The polymerized gel was placed in the electrophoresis chamber and filled with Laemmli running buffer. After short centrifugation (30 s, 21000 rpm), to remove precipitates, 10–15 µl of the sample or 5 µl of the marker were loaded into the gel slots. The electrophoresis was performed at 300 V for 30 min.

For visualization, unspecific staining of proteins was done in Coomassie Brilliant Blue R-250 for 1 h while shaking and subsequent destaining in Coomassie destaining solution. Stained gels were documented with a ChemiDoc Imaging System (590/110 standard filter, BioRad).

Gels containing protein samples labeled with fluorescent dyes DyLight 550 and DyLight 650 were directly visualized employing a ChemiDoc Imaging System (DyLight 550: 577–613 nm filter, DyLight 650: 675–725 nm filter, BioRad).

4.3.9 Labeling of AzF-modified FnCas12a variants with fluorescent dyes

For single-molecule FRET measurements, FnCas12a variants were site-specifically labeled with two fluorescent dyes. To this end, the unnatural amino acid *para*-Azido-L-phenylalanine (**Figure 8C**) was incorporated (4.2.4 and 4.3.2). The azide group of AzF is a versatile bioorthogonal chemical reporter that can react in different coupling reactions [259,260]. One of these reactions is the Staudinger-Bertozzi ligation (**Figure 9**), a selective reaction between triphenylphosphines (soft nucleophiles) and azides to aza-ylide intermediates [261]. The subsequent processing with water leads to hydrolysis and the formation of an amide bond.

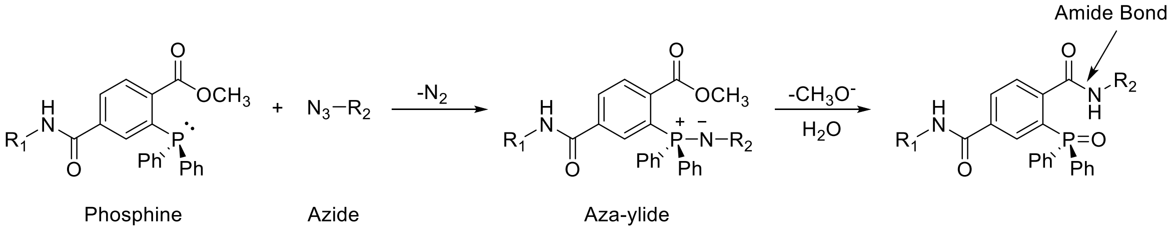


Figure 9: Staudinger-Bertozzi ligation of azides.
Staudinger-Bertozzi ligation with a triphenylphosphine results in the formation of an amide bond. R_1 represents the labeling agent (fluorophore), R_2 represents the protein with built-in AzF.

The incorporated unnatural amino acid AzF at positions D470, K647, and/or T1222 was coupled in a stochastic Staudinger reaction with DyLight 650 and DyLight 550. During the whole procedure, it was taken care that the dye solutions were protected from light.

The coupling reaction was performed by mixing a protein sample of 100 µl (1–3 µM) with 0.5 µl of 10 mM DyLight 650 Phosphine and 0.5 µl of 10 mM DyLight 550 Phosphine stock solutions (10-fold molar excess, dissolved in DMSO) and incubated for 90 min at room temperature in darkness. Subsequently, the labeling reaction was centrifuged for 10 min at 21000 g. The supernatant was purified via size exclusion chromatography using a Sephadex G-50 illustra NICK column (GE Healthcare) and Cas batch buffer to remove the excess of dye. The labeled protein was directly used for single-molecule measurements. Labeled protein that was stored at −20 °C lost quality and showed reduced molecule counts in the measurements compared to fresh protein.

4.4 Analytical methods

4.4.1 Determination of protein concentration

The aromatic amino acids tryptophan, tyrosine, and phenylalanine, as well as disulfide bonds (cystines), absorb UV-light between 250 and 300 nm. Absorbance at a wavelength of 280 nm is therefore used to calculate protein concentrations. Absorption spectroscopy was conducted employing a NanoDrop One^{C} spectrophotometer (ThermoFisher Scientific) and a sample volume of 2 µl.

The concentrations were calculated according to the Lambert-Beer law (**Equation 1**). The molar extinction coefficient at 280 nm (ε_{280}, M^{-1} cm^{-1}) can be calculated from the amino acid composition as follows [262].

$$\varepsilon_{280} = \sum Trp \cdot 5500 + \sum Tyr \cdot 1490 + \sum Cystine \cdot 125 \qquad \textbf{Equation 5}$$

The molar extinction coefficient of FnCas12a was calculated using the ProtParam tool (https://web.expasy.org/protparam/) as 143 830 M^{-1} cm^{-1}.

4.4.2 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy in the far-UV range is used to analyze the secondary structure of proteins. Right- and left-circular polarized light is absorbed to a different extent when passing through a solution of chiral molecules. As a result, the CD signal over the corresponding wavelengths can adopt positive and negative values. Different secondary structure elements like α-helix, β-sheet, and random coil have characteristic CD spectra in the range of 170–250 nm. In this work, CD spectra were recorded from 190 to 260 nm with a quartz crystal cuvette (0.2 mm, Hellma) in a spectropolarimeter J-815 (JASCO). The data were normalized to obtain the molar ellipticity per amino acid according to the following equation:

$$\theta_{MRW} = \frac{\theta_{obs} \cdot 100 \cdot MRW}{\beta \cdot d} = \frac{\theta_{obs} \cdot 100 \cdot MW}{\beta \cdot d \cdot N_{aa}} = \frac{\theta_{obs} \cdot 10^5}{c \cdot d \cdot N_{aa}} \qquad \textbf{Equation 6}$$

θ_{MRW}	average ellipticity per amino acid (deg cm^2 $dmol^{-1}$)
θ_{obs}	measured ellipticity (mdeg)
MRW	mean residue weight (kDa)
β	protein concentration (mg/ml)
c	protein concentration (µmol/l)
d	thickness of cuvette (cm)
MW	molecular weight of the protein (kDa)
N_{aa}	number of amino acids in the protein

4.4.3 Protein Thermal Shift melting curves

Protein thermal shift assays monitor the stability of a protein at increasing temperatures. Melting curves of FnCas12a variants were generated using the Protein Thermal Shift™ kit (ThermoFisher Scientific) and the Rotor-Gene Q qPCR cycler from Qiagen and the HRM (High-Resolution Melt) software. The Protein Thermal Shift dye binds to hydrophobic parts of the protein. Its fluorescence depends on its chemical surroundings. With the heating process, the protein unfolds and more of the proteins' hydrophobic residues get accessible for the dye, leading to an increase in fluorescence. When the protein aggregates, the dye gets excluded and the fluorescence decreases again.

The proteins were diluted in water and measurements were performed in quadruplicates. The data were analyzed using the Origin 2019b software (OriginLab).

4.4.4 Preparation of double-stranded target oligonucleotides

Double-stranded DNA targets or RNA/DNA hybrids, for binding and cleavage studies of FnCas12a, were prepared by hybridizing two complementary single-stranded oligonucleotide strands. RNA and DNA oligonucleotides used in this thesis are listed in **Table 20**.

For hybridization, both strands were used at a concentration of 3 µM in an annealing reaction as follows:

Table 34. Oligonucleotide annealing mix.

5x Annealing buffer	10 µl
ss oligonucleotide 1 (100 µM)	1.5 µl
ss oligonucleotide 2 (100 µM)	1.5 µl
nuclease-free H_2O	ad 50 µl

The annealing mix was heated to 95 °C for 3 min and passively cooled to room temperature to allow annealing. Hybridized oligonucleotides were stored at −20 °C.

The EGFP-hAgo2 (Addgene #21981) plasmid that was further used as target DNA in plasmid cleavage assays with FnCas12a (4.4.6.1) was linearized with SmaI. SmaI cleaves the DNA in the middle of the recognition sequence 5’-CCCGGG-3’ with blunt ends.

The DNA restriction mixture included the following components:

Table 35. Reaction mix for DNA linearization with SmaI.

10x FastDigest buffer	5 µl
SmaI (20 U/µl)	1 µl
plasmid DNA (≈ 1 µM)	4 µl
nuclease-free H_2O	40 µl

The restriction mix was incubated at 37 °C for 1–2 h or at room temperature overnight. The mix was purified with the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel) and the linearization success was verified on a 0.8% agarose gel (4.2.6). The DNA was stored at −20 °C.

4.4.5 Electrophoretic mobility shift assay

The electrophoretic mobility shift assay (EMSA) was used to study the binding of RNA and DNA oligonucleotides to FnCas12a and its variants and to monitor the formation of binary and ternary complexes. It is based on the retardation of nucleic acids during electrophoresis when being bound to a protein [263]. To be able to unravel the size difference due to binding, native PAGEs like native Tris-Glycine PAGE or native TBE PAGE (**Table 10**) are used that do not interfere with the interactions between proteins and nucleic acids. For casting and gel run, the Mini-PROTEAN-Tetra System (Bio-Rad) was used.

4.4.5.1 FnCas12a binding to RNA

To monitor the binding of FnCas12a to the crRNA, crRNA-Cas12a complexes were formed in 1x Cas buffer with a two-fold molar excess of protein (20 nM) over fluorescently labeled crRNA (10 nM). The complexes were incubated in a total volume of 10 µl at room temperature for 10 min. 2x Native loading buffer was added and the samples were separated on a non-denaturing TBE gel (15% PAA, 300 V, 10 min). In-gel fluorescence of the labeled crRNA was visualized using a ChemiDoc Imaging System (Atto532, Cy3: 577–613 nm filter, Cy5: 675–725 nm filter, Bio-Rad).

4.4.5.2 FnCas12a binding to DNA

To monitor the binding of FnCas12a to the target DNA, the binary complex composed of 200 nM FnCas12a protein and 200 nM crRNA was pre-formed in 1x Cas buffer for 10 min at room temperature. 75 nM of the binary complex was added to 10 nM of fluorescently labeled short target DNA in a total volume of 10 µl and incubated at 37 °C for 1 h. To avoid unspecific interactions of Cas12a with nucleic acids, Heparin at a final concentration of 32 µg/ml was added and the reaction was incubated for an additional 10 min at room temperature. 2x Native loading buffer was added and the samples were separated by non-denaturing Tris-Glycine gel electrophoresis (6% PAA, 230 V, 30 min) with Native Laemmli running buffer. The gel was visualized using a ChemiDoc Imaging System (Cy3: 577–613 nm filter, Cy5: 675–725 nm filter, Bio-Rad).

4.4.6 DNA *cis*-cleavage assays

Cis-cleavage assays were performed to monitor the cleavage activity and efficiency of FnCas12a variants on different DNA targets. Evaluation of gel pictures and the quantification of fluorescent signals was done with the software ImageLab (Bio-Rad).

4.4.6.1 Plasmid cleavage assay

Plasmid cleavage assays were performed using the EGFP-hAgo2 plasmid (Addgene #21981) as target DNA. It was used either in its supercoiled form or after linearization with SmaI (0). The binary complex, composed of FnCas12a and crRNA, was pre-formed by incubating 100 nM of protein and 100 nM of crRNA for 10 min at room temperature in 1x Cas buffer. 37.5 nM of preincubated binary complexes were added to 5 nM of target DNA. The reaction was incubated in 1x Cas buffer in a total volume of 15 µl for 1 h at 37 °C. The reaction was stopped with 0.36 U (0.6 µl) Proteinase K (ThermoFisher Scientific). After incubation for 30 min at 55 °C, the reaction was mixed with 6x DNA loading buffer. The reaction with the linearized plasmid was analyzed on a 0.8% agarose 1x TAE gel (100 V, 30 min), the reaction using supercoiled DNA on a 1.5% agarose 1x TAE gel (80 V, 60 min) and visualized using SYBR-Safe (ThermoFisher Scientific) (4.2.6).

In order to analyze the time-dependent cleavage of FnCas12a variants, cleavage kinetics with linearized and supercoiled plasmid DNA were conducted. To this end, 37.5 nM of the binary complex were added to 5 nM of target DNA and incubated at 37 °C. After 30 s, 1 min, 5 min, 10 min, 20 min, 40 min, or 1 h 10 µl-aliquots of the reaction were stopped with EDTA (final concentration 83.3 mM) and 0.36 U Proteinase K. After incubation for 30 min at 55 °C, the reaction was mixed with 6x DNA loading buffer, and loaded onto a 0.8% agarose 1x TAE gel (100 V, 30 min) for the linearized plasmid and on a 1.5% agarose 1x TAE gel (80 V, 60 min) for the supercoiled plasmid DNA. The DNA was visualized using SYBR-Safe (ThermoFisher Scientific) (4.2.6).

4.4.6.2 Short DNA target cleavage assay

To analyze the *cis*-cleavage by FnCas12a and the resulting reaction products, cleavage assays with short DNA targets were conducted. The binary complex of 200 nM FnCas12a protein and 200 nM crRNA was pre-formed in 1x Cas buffer for 10 min at room temperature. In a final concentration of 75 nM, the binary complex was added to 10 nM of fluorescently labeled double-stranded short target DNA in a total volume of 10 µl and incubated at 37 °C for 1 h. The reaction was stopped by addition of 0.36 U (0.6 µl) Proteinase K and incubated at 55 °C for 30 min. 2x RNA loading buffer (with or without dye) was added and the samples were heated to 95 °C for 5 min. The samples were separated on a pre-heated 15% PAA, 6 M Urea, 1x TBE gel (300 V, 30 min). For casting and gel run, the Mini-PROTEAN-Tetra System (Bio-Rad) was used. Fluorescent signals were visualized using a ChemiDoc Imaging System (Cy3: 577–613 nm filter, Cy5: 675–725 nm filter, Bio-Rad).

To follow cleavage kinetics, reactions were stopped after 1 s, 30 s, 1 min, 3 min, 5 min, 10 min, 20 min, 30 min, 45 min, and 60 min with EDTA (final concentration 125 mM) and 0.36 U Proteinase K. Further steps were conducted according to the cleavage assay as described above.

Sequencing gels were used to resolve the cleavage reaction products at the nucleotide level. They enable a resolution down to one nucleotide and have a thickness of 0.4 mm and a length of 50 cm. The cleavage reactions for the high-resolution sequencing gel (15% PAA, 6 M Urea, 1x TBE) were carried out with 375 or 750 nM of the binary complex and 50 or 100 nM of fluorescently labeled target DNA, respectively. After incubation for 1 h at 37 °C, the reaction was stopped with the addition of 0.36 U Proteinase K and incubated at 55 °C for 30 min. 2x RNA loading buffer was added and the samples were heated to 95 °C for 5 min. The samples were separated on a pre-heated sequencing gel (45 °C, 65 W, 1 h). For casting and gel run, the Sequi-Gen GT sequencing Cell (Bio-Rad) system was used. Fluorescent signals were visualized using a ChemiDoc Imaging System (Cy3: 577–613 nm filter, Cy5: 675–725 nm filter, Bio-Rad).

4.4.7 *Trans*-cleavage assay using single-stranded DNA

To test for indiscriminate *trans*-cleavage activity of FnCas12a Gaetan Burgio and Anthony Newman from the Australian National University (Canberra, AUS) performed assays with different FnCas12a variants. The ribonucleoprotein complexes were pre-formed that prime FnCas12a for *cis*-cleavage: 1 µM FnCas12a and 1 µM crRNA were complexed at 25 °C for 10 min in *trans*-cleavage buffer. Afterwards, 50 nM of a single-stranded DNA oligo ‘activator’ was added, and the sample was incubated at 30 °C for 20 min in *trans*-cleavage buffer. To detect *trans*-cleavage activity, the *cis*-cleavage reaction was diluted 10x in buffer to a total volume of 150 µl, and three technical replicates of 50 µl for each condition were loaded onto a 96-well plate (black with clear bottom, Invitrogen). 50 µl of 200 nM fluorescence quencher (FQ) substrate (diluted in 1x buffer) were added to each well, the plate was immediately transferred to a plate reader (Victor Nivo Multimode, PerkinElmer), shaken at 300 rpm (orbital) for 30 s, and the fluorescence intensity was measured over 90 min (excitation filter 480/30 nm, emission filter 530/30 nm). The final concentrations were as follows: 100 nM FnCas12a RNP, 5 nM ssDNA, 100 nM FQ reporter, and an upper limit of 5 nM enzyme. Using WT FnCas12a and variants at variable concentrations of FQ reporter, a calibration curve was constructed to convert changes in fluorescence intensity to *trans*-cleavage rates and concentration of cleaved FQ reporter over time. The slope of change in fluorescence intensity was extracted from the first 600 seconds of the reaction, and relative rates of *trans*-cleavage were estimated from a calibration curve.

4.4.8 Mass photometry

Mass photometry (MP) combines the principles of interference microscopy [264] and interferometric microscopy [265]. Molecules hitting a surface scatter the illumination light in a defined way depending on the particle's volume and refractive index. The interference between the scattered light of a particle and the reflected light of the surface is measured (mass photometry contrast). The signal is directly proportional to the molecular mass of the molecule as the refractive indices of biomolecules only vary by a few percent and can be neglected [266,267]. Consequently, it is possible to weigh single molecules or complexes with light (**Figure 10**).

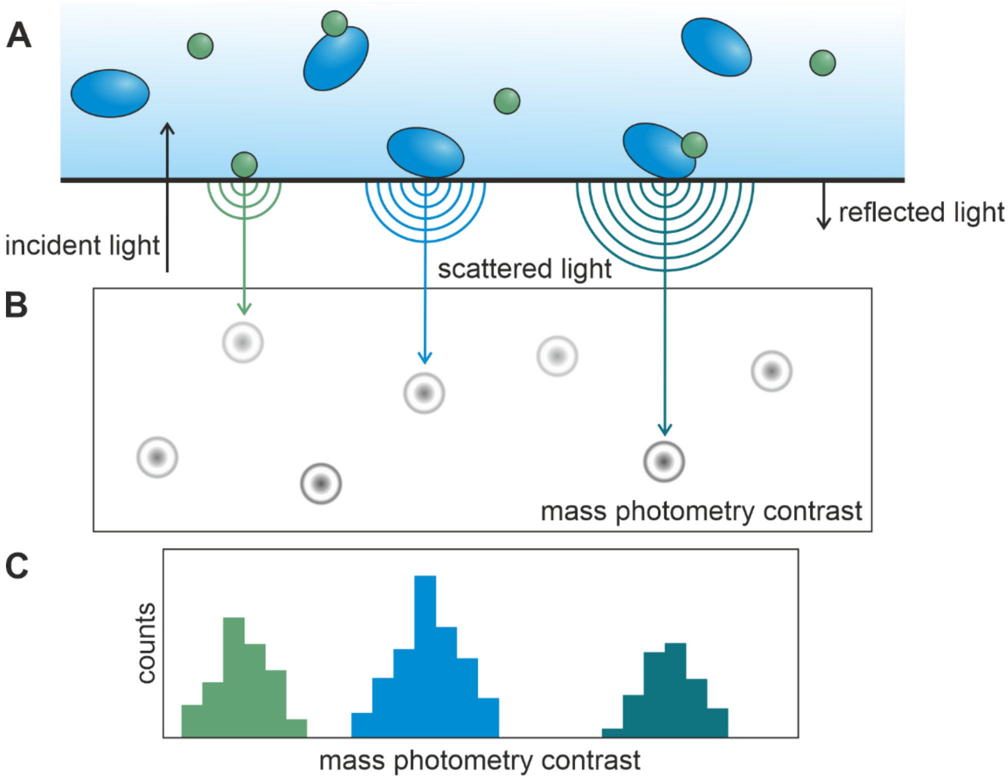


Figure 10: The principle of mass photometry.
(**A**) A sample of molecules is illuminated by light. Molecules at the surface scatter the incident light depending on their volume. (**B**) The mass photometry contrast is derived from the interference between the scattered light of a particle and the reflected light of the surface. (**C**) To create the histogram, mass photometry contrast counts are summarized. (figure adapted from Young *et al.* [267])

MP experiments were conducted on a TwoMP Instrument (Refeyn) at room temperature. Data were acquired using the AcquireMP (Refeyn) software. MP movies were recorded at 500 Hz with an exposure time of 1.95 ms. Small wells were created by placing a gasket foil (Grace Bio-Labs) onto a microscope cover slip (50 x 26 mm, Greiner Bio-One). The gasket foils and cover slips were cleaned with 5 min in 50% (v/v) isopropanol and pure Milli-Q H_2O. 15–19 µl of MP buffer were placed onto a well and the focus was adjusted and locked via ‘droplet dilution’ mode. 1–5 µl of the sample were added to the buffer droplet to adjust the volume to 20 µl and the measurement was started.

Prior to the sample measurement, a calibration with the standard proteins IgG (150 kDa), TG (660 kDa), and BSA (66 kDa) was performed in MP buffer. Apo FnCas12a samples were diluted to 200 nM, binary complex samples were pre-incubated at room temperature for 10 min at an equimolar concentration of protein and crRNA of 2 µM and prior to the measurement, diluted to 200 nM. The samples were measured in triplicates and analyzed using the DiscoverMP software.

4.4.9 Structure prediction with AlphaFold2

Structural prediction was performed using the Google Colab version of Alphafold2 [268] (available at https://colab.research.google.com/github/sokrypton/ColabFold/blob/v1.1-premultimer/AlphaFold2.ipynb) with the default settings and no custom multiple sequence alignment (MSA) template as a default option. Approximately 300 amino acids around BH and helix 1 were used as input sequence (V777–G1072), obtained from FnCas12a crystal structure bound to the crRNA and target DNA (PDB: 6I1K) and modified with the respective mutations. The accuracy of the structural model was tested using a per-residue measure of local confidence (predicted Local Distance Difference Test, pLDDT) and indicated a high confidence in accurately predicting the residues and domains. Structures were analyzed with PyMol version 2.3.1 (Schrodinger LLC).

4.5 Fluorescence-based single-molecule spectroscopy

4.5.1 Förster resonance energy transfer

Förster resonance energy transfer (FRET) is the physical process of the radiationless transfer of energy from an excited fluorophore (donor) in the first excited state (S_1) to a second fluorophore (acceptor) in direct proximity (2–10 nm) without the emission of a photon (**Figure 11**). Upon this transfer, the donor returns to the ground state S_0 while the acceptor is excited from the S_0 to the S_1 state. The acceptor can decay to S_0 by the emission of a photon. The energy transfer between the fluorophores is non-radiative but mediated by dipole-dipole interactions [269,270].

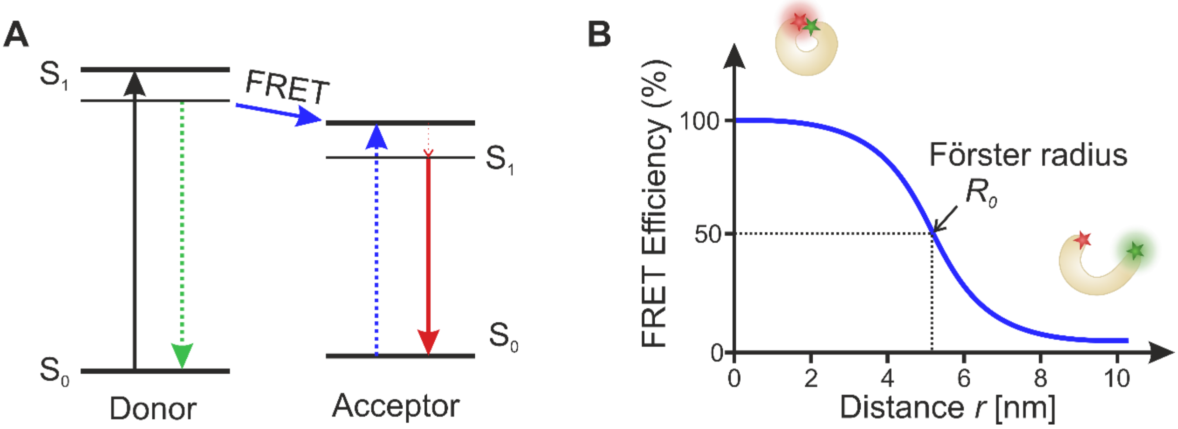


Figure 11: The principle of FRET and the Förster radius.
(**A**) Simplified Jablonski Diagram of the Förster resonance energy transfer (FRET) between two fluorophores. An electron of the donor molecule is excited from the ground state (S_0) to the first excited state (S_1, black arrow). The return to the ground state can either be achieved via direct fluorescence and emission of a photon of slightly lower energy (longer wavelength, green dotted arrow) or, in the presence of a suitable acceptor, via nonradiative transfer of energy like FRET (blue) to another fluorophore. Thereby, an electron of the acceptor molecule is excited from S_0 to S_1. Via emission of a photon, the acceptor returns to S_0 (red arrow). (**B**) Exponential dependency of the transfer efficiency and the distance r separating the donor (green) and the acceptor (red). The distance at 50% FRET efficiency is called Förster radius R_0. Shown for a theoretical pair of fluorophores with a Förster distance of 5 nm.

FRET only occurs if special requirements are fulfilled. (i) The emission spectra of the donor and the absorption spectra of the acceptor have to overlap (**Figure 12**). The higher the overlap, the more efficient is the energy transfer from donor to acceptor.

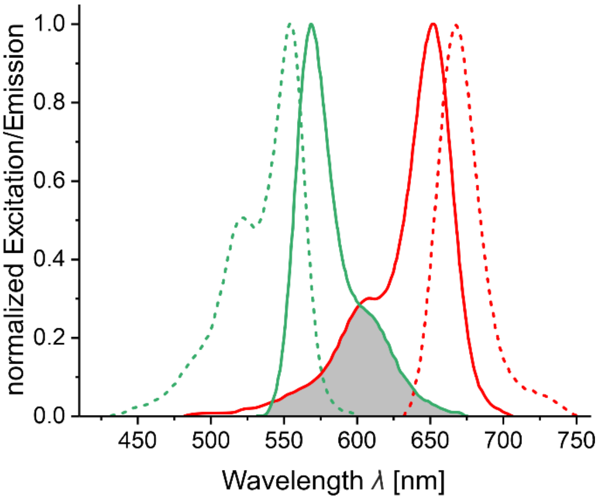


Figure 12: Spectral overlap of donor emission and acceptor excitation.
DyLight 550 (donor) excitation (green, dotted line) and emission (green, solid line) spectra and DyLight 650 (acceptor) excitation (red, solid line) and emission (red, dotted line) spectra. The spectral overlap (grey area) of the donor emission and the acceptor excitation allows the energy transfer via FRET. (spectral information derived from www.chroma.com/spectra-viewer)

(ii) The dipole moments of donor and acceptor fluorophore have to be in a suitable alignment with each other. The orientation factor κ^2 defines the angular relation of both dipoles. κ^2 is maximal with a value of 4 when the orientation of the dipole-dipole transition moments of the molecules is head to tail, and minimal (0), when the orientation is orthogonal [271]. At room temperature freely rotating fluorophores have an approximate κ^2 of $^2/_3$.

(iii) The transfer of energy occurs via a nonradiative dipole-dipole resonance interaction between donor and acceptor and is limited to small distances between 2 and 10 nm. The efficiency of the energy transfer E_{FRET} is dependent on the distance r between the centers of donor and acceptor and is proportional to its inverse sixth power [270]. The efficiency is given by the Förster equation:

$$E_{\mathrm{FRET}} = \frac{1}{1+\left(\frac{r}{R_0}\right)^6} \qquad \textbf{Equation 7}$$

R_0 is the Förster radius, the specific distance for each donor and acceptor pair, where the FRET efficiency is 50%. This means every second time the donor gets excited the energy is transferred to the acceptor (**Figure 11A**). R_0 is defined as:

$$R_0 = \sqrt[6]{\frac{9(\ln 10)\cdot\kappa^2\cdot\Phi_D\cdot J}{128\pi^5\cdot N_A\cdot n^4}}\ \text{Å} \qquad \textbf{Equation 8}$$

J	spectral overlap integral between emission spectra of donor and absorption spectra of acceptor
κ^2	relative orientation of donor and acceptor fluorophore dipoles
n	refractive index of the surrounding media
Φ_D	quantum yield of the donor fluorophore
N_A	Avogadro constant

The Förster radius for typically used dye pairs is in a range between 3 and 8 nm. The Förster radius of the DyLight 550-DyLight 650 pair, used in this thesis, is 5.2 nm [272].

By transferring the energy from the excited state of the donor to the acceptor fluorophore via FRET, the fluorescence of the donor is partly quenched. With that, the energy transfer is in direct competition with the radiative decay of the donor to S_0 via the emission of a photon. Therefore, E_{FRET} also depends on the ratio of photons emitted by the donor in the presence and absence of the acceptor molecule [273,274]:

$$E_{\mathrm{FRET}} = 1 - \frac{I_{D(A)}}{I_{D(0)}} \qquad \textbf{Equation 9}$$

$I_{D(A)}$ intensity of donor emission in the presence of acceptor
$I_{D(0)}$ intensity of donor emission in the absence of acceptor

According to **Equation 7** and **Equation 9**, measurements of the FRET efficiency can be used to determine the distance between two fluorophores (inter- or intramolecular). For this, it is required that the actual distance between the two fluorophores is in the same range as the fluorophores' Förster radius R_0 [270]. This makes FRET a useful tool for the characterization of biomolecules as their dimensions are often in the accessible range between 2 and 10 nm [274].

The first FRET measurements were performed in ensemble approaches [275–277], but biological interactions and dynamics are often inhomogeneous. Conformational subpopulations or intermediates could not be assessed with classical FRET methods due to averaging effects. FRET spectroscopy gained importance with the development of single-molecule methods [278] and has been applied to analyze structure [279], function, folding, and dynamics [280] of diverse biomolecules and molecular machines [272,281,282].

In this thesis, single-molecule FRET (smFRET) measurements were applied to analyze intramolecular distances and conformational transitions in fluorescently labeled variants of FnCas12a in the absence and presence of crRNA and target DNA. Additionally, conformational changes of a doubly labeled RNA/DNA hybrid upon binding by FnCas12a were monitored.

4.5.2 Confocal single-molecule fluorescence microscopy

Single-molecule FRET measurements can be conducted in solution at a confocal fluorescence microscope (**Figure 13**) if the fluorescently labeled sample is used in a pico- to nanomolar concentration range to assure that only one labeled molecule at a time diffuses through the detection volume. In consequence, intermolecular interactions of two labeled molecules can only be monitored if the dissociation constant K_D does not exceed 50 nM [283]. If interactions between a doubly labeled molecule and an unlabeled interaction partner are monitored, the concentration of the unlabeled partner may exceed 50 nM. To additionally reduce the signal-to-noise ratio (SNR) only a small, femtoliter-sized volume of the sample (confocal volume) is illuminated by an excitation laser beam focused by a dichroic mirror through the high numerical aperture (NA) objective of the microscope. These two requirements assure that only one molecule at a time gets excited by the laser beam, thereby achieving single-molecule resolution. All molecules diffusing through the confocal volume over time are monitored and the emitting light is collected through the same objective (epifluorescence). It passes through the dichroic mirror and is focused onto a micrometer-sized pinhole to eliminate background signals by excluding any light out-of-focus emission. The emission signal is spectrally separated to allow the separate detection of donor and acceptor photons and ultimately recorded by an efficient single-photon avalanche diode (SPAD) detector, with single-photon sensitivity [284,285]. The constant excitation of molecules residing approximately 0.5–5 ms in the confocal volume, depending on the diffusion speed of the sample, leads to the emission of hundreds to thousands of photons that are for each molecule condensed to a single “burst”.

In this thesis, single-molecule FRET measurements on freely diffusing molecules in solution were conducted using a commercial MicroTime 200 confocal fluorescence microscope (Picoquant).

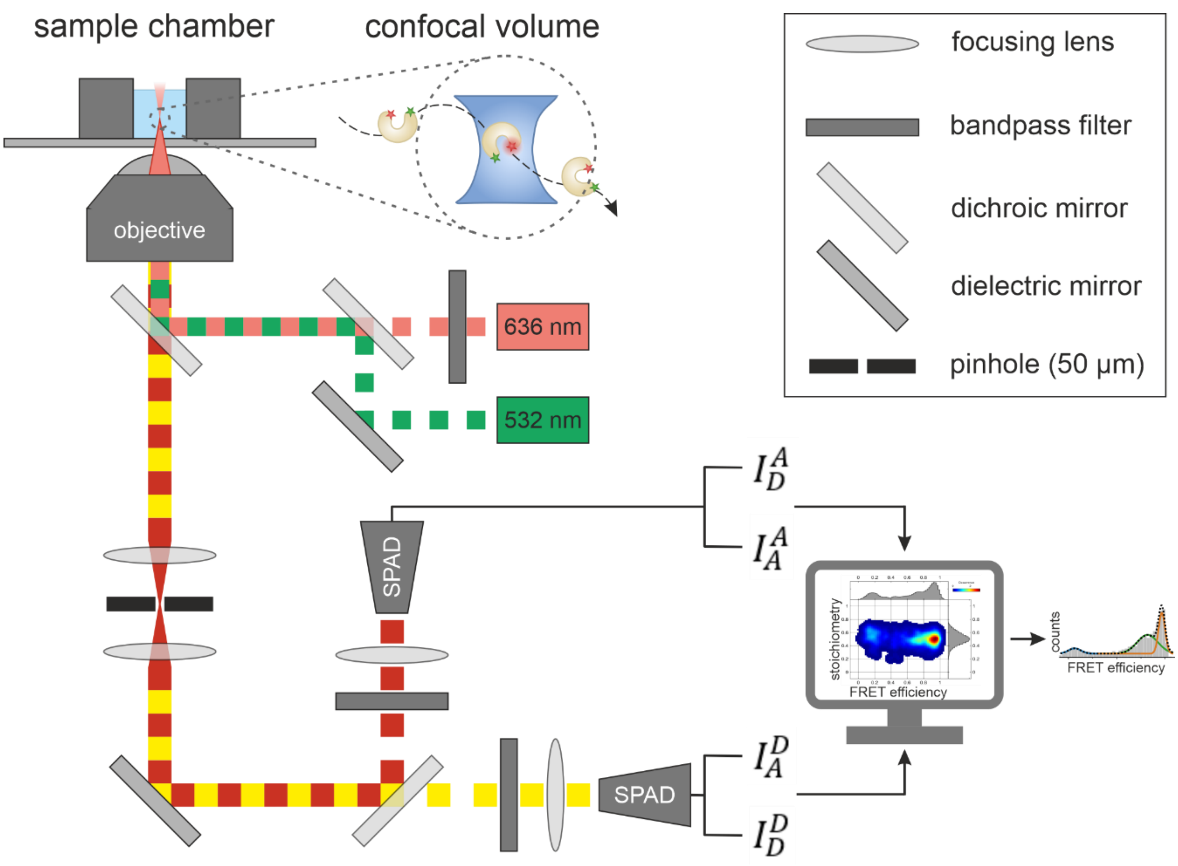


Figure 13: Confocal microscope setup.
Schematic overview of the PicoQuant MicroTime 200 confocal fluorescence microscope setup. Pulsed laser diodes (532 nm (LDH-P-FA-530B, green), 636 nm (LDH-D-C-640, red)) are combined via redirection by dielectric and dichroic filters. The 636 nm beam is cleaned by a bandpass filter (ZET 635; Chroma). The excitation beams are focused through a 1.2 NA, ×60 water-immersion microscope objective (UplanSApo ×60/1.2 W; Olympus) into the sample chamber creating the confocal volume. Emitted fluorescence of excited molecules (yellow, dark red) is collected by the same objective and focused with lenses and a 50 µm confocal pinhole. Donor and acceptor fluorescence are separated by a dichroic mirror (T635lpxr; Chroma) and further filtered with bandpass filters (donor: ff01-582/64; Chroma; acceptor: H690/70; Chroma). Both photon streams are detected by an avalanche photodiode detector (SPAD, SPCM-AQRH-14-TR, Excelitas Technologies) and subsequently analyzed by a TCSPC-capable PicoQuant HydraHarp 400 module. The obtained data is processed and analyzed using the SymphoTime64 and PAM software. (adapted from Kramm, 2020 [286])

To be able to correct spectral crosstalk caused by broad donor emission spectra, donor and acceptor fluorophores were excited with an alternating excitation scheme called pulsed interleaved excitation (PIE) [287]. Two synchronized lasers alternate on a nanosecond timescale, separated by a 9.2 ns time gap to allow complete decay of fluorescence before the next pulse. This leads to a sequential excitation of donor and acceptor fluorophores passing through the confocal volume at 532 nm and 636 nm. The direct excitation of the acceptor enables the FRET independent detection of acceptor fluorescence and allows for spectral sorting. This is important to calculate the donor and acceptor stoichiometry and to confirm that a detected particle is coupled to both fluorophores (**Equation 11** and **Equation 13**). Using PIE, results in four different emission signals: donor emission upon donor excitation, acceptor emission upon donor excitation, donor emission upon acceptor excitation, and acceptor emission upon acceptor excitation. This information is subsequently used to calculate correction factors like crosstalk effects.

4.5.3 Confocal single-molecule FRET measurements

4.5.3.1 Sample preparation and measurement

Sample chambers for confocal measurements (Cellview slide, Greiner Bio-One) were passivated with a passivation solution (2 mg/ml BSA in PBS) for 10 min to prevent unspecific adhesion of biomolecules to the surfaces. The measurement chamber was subsequently washed with 1x Cas buffer.

Labeled FnCas12a apo protein was diluted to picomolar concentrations in 1x Cas buffer and transferred into the sample chamber, yielding a burst rate of approximately 10–20 counts/s. The binary complex composed of FnCas12a protein with crRNA (final concentration 0.5 nM, 1 nM, 2 nM, or 4 nM) was pre-incubated for 30 min at room temperature in a small volume of 20 µl and prior to the measurement diluted by a factor of 10 with 1x Cas buffer to yield a final volume of 200 µl. For the ternary complexes, the binary complex composed of FnCas12a with crRNA (final concentration 1 nM) was pre-incubated for 10 min prior to the addition of double-stranded or single-stranded target DNA (final concentration 1 nM, 2 nM, or 4 nM) and incubated in a volume of 20 µl for 20 min. Prior to the measurements, the sample was diluted by a factor of 10 with 1x Cas buffer to yield the final volume of 200 µl.

For smFRET measurements with doubly labeled crRNA/DNA hybrid, a labeled RNA and a labeled DNA were annealed (0) and diluted to 10 nM. 100 pM of the hybrid were transferred into the sample chamber and 0, 25, 50, or 100 pM of WT FnCas12a were added and incubated in a volume of 20 µl for 10 min. Prior to the measurement, the sample was diluted by a factor of 10 with 1x Cas buffer to yield the final volume of 200 µl.

Excitation of the fluorophores at 532 nm and 636 nm was achieved by an LDH-P-FA-530B laser at 20 µW and an LDH-D-C-640 laser (clean-up filter: ZET 635; Chroma) at 16 µW, respectively. The PIE pulse frequency was 40 MHz with a 9.2 ns delay for the donor excitation. SmFRET measurements of freely diffusing molecules were carried out for 30 min.

4.5.3.2 Data Analysis of confocal single-molecule FRET experiments

Confocal smFRET data of diffusing molecules were analyzed with the software package PAM (PIE Analysis with MATLAB) [247]. The program sorts the photons of the detected signals (donor/acceptor) into three PIE channels: (i) donor detection upon donor excitation (DD, donor signal), (ii) acceptor detection upon donor excitation, (DA, FRET, and crosstalk), and (iii) acceptor detection upon acceptor excitation (AA, acceptor signal). Photon bursts were extracted by photon burst search algorithms. In PAM, two variants are implemented: (i) The all-photon burst search (APBS) algorithm, searching for all detected photons together, including signals from molecules only labeled with one of both fluorophores, and (ii) the dual-channel burst search (DCBS) algorithm that searches independently for photons resulting from the donor (DD+DA) or acceptor excitation (AA) and only keeps bursts detected in both searches, resulting in signals exclusively from doubly labeled molecules. Bursts were registered if at least *L* photons with at least *M* neighboring photons within a time interval *T* were detected (APBS: $L = 100$, $M = 15$, $T = 500$ µs, DCBS: $L = 100$, $DD{+}DA = 15$, $AA = 15$, $T = 500$ µs) [288].

The FRET efficiency as proximity ratio E_{PR} and the raw stoichiometry factor S_{raw} for an APBS were calculated as follows:

$$E_{PR} = \frac{N_{DA}}{N_{DD}+N_{DA}} \qquad \textbf{Equation 10}$$

$$S_{raw} = \frac{N_{DD}+N_{DA}}{N_{DD}+N_{DA}+N_{AA}} \qquad \textbf{Equation 11}$$

N_{DD}	number of photons detected in the DD PIE channel
N_{DA}	number of photons detected in the DA PIE channel
N_{AA}	number of photons detected in the AA PIE channel

E_{PR} and S_{raw} were subsequently used for calculation of the correction factors for donor leakage c_{leak} and direct excitation c_{dir}.

For each DCBS burst, the FRET efficiency *E* and the stoichiometry factor *S* were calculated as follows:

$$E = \frac{N_{DA}-(c_{leak}\cdot N_{DD}+c_{dir}\cdot N_{AA})}{\gamma\cdot N_{DD}+N_{DA}-(c_{leak}\cdot N_{DD}+c_{dir}\cdot N_{AA})} \qquad \textbf{Equation 12}$$

$$S = \frac{\gamma\cdot N_{DD}+N_{DA}-(c_{leak}\cdot N_{DD}+c_{dir}\cdot N_{AA})}{\gamma\cdot N_{DD}+N_{DA}+\beta\cdot N_{AA}-(c_{leak}\cdot N_{DD}+c_{dir}\cdot N_{AA})} \qquad \textbf{Equation 13}$$

N_{DD}	number of photons detected in the DD PIE channel
N_{DA}	number of photons detected in the DA PIE channel
N_{AA}	number of photons detected in the AA PIE channel
c_{leak}	correction factor for donor leakage
c_{dir}	correction factor for direct excitation
γ	correction factor for detection
β	correction factor for excitation

4.5.3.3 Determination of correction factors

To determine correct FRET efficiencies, the implemented correction tool in PAM was used. The obtained burst data were corrected for donor leakage and acceptor direct excitation, summarized as crosstalk effects, the influence of the different quantum yields of the dyes, the detection efficiencies of the donor and acceptor dyes (γ-factor), and the different excitation efficiencies of the dyes (β-factor, **Figure 14**).

When spectral separation is incomplete and therefore donor emission is detected in the acceptor channel, donor leakage occurs. From the proximity ratio of the donor only population $E_{PR}(DD)$ (with $0.8 < S_{raw} < 1.1$) of the APBS data, the correction factor c_{leak} was calculated (**Figure 14A**):

$$c_{leak} = \frac{E_{PR}(DD)}{1-E_{PR}(DD)} \qquad \textbf{Equation 14}$$

During excitation of the donor fluorophore, direct excitation of some acceptor fluorophores can occur. For correction, the factor c_{dir} was determined from the raw stoichiometry factor of the acceptor only population $S_{raw}(AA)$ ($E > 0.1$, $-0.2 < S_{raw} < 0.2$) in an analogous way to the c_{leak} calculation (**Figure 14A**):

$$c_{dir} = \frac{S_{raw}(AA)}{1-S_{raw}(AA)} \qquad \textbf{Equation 15}$$

After applying the c_{leak} and c_{dir} correction to the DCBS histograms, γ and β factors were determined by applying an internal fit on distinct FRET populations in the ES-histograms. The populations were fitted with one or more Gaussian distributions to determine their mean E and S positions (**Figure 14B**). These were plotted as $1/S$ against E and linearly fitted. The slope m and the y-axis intercept n were used to calculate the γ and β factors as follows:

$$\gamma = \frac{n-1}{n+m-1} \qquad \textbf{Equation 16}$$

$$\beta = n + m - 1 \qquad \textbf{Equation 17}$$

The set of correction factors was determined for all data acquired on the same day. Further analysis was conducted with fully corrected datasets (**Figure 14C**).

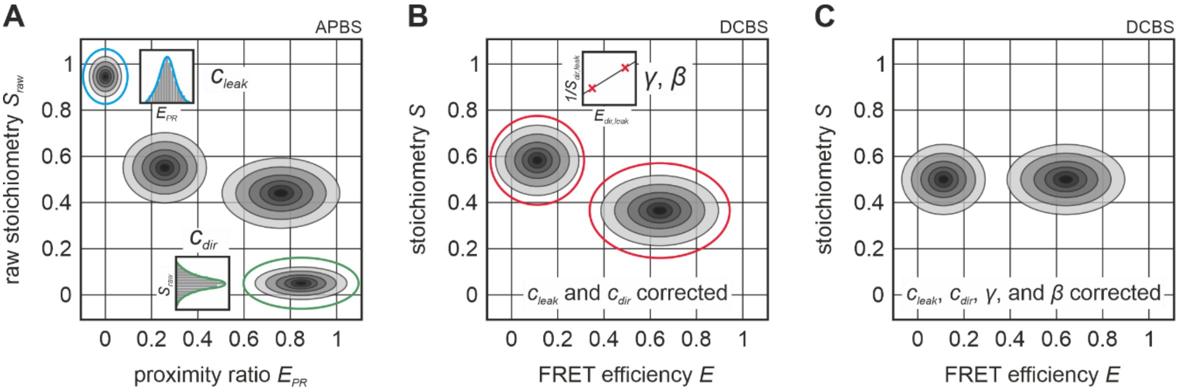


Figure 14: Correction factor determination for confocal single-molecule FRET data.
(**A**) From the E_{PR}/S_{raw} histograms derived from the all-photon burst search (APBS) the correction factors for donor leakage c_{leak} and direct excitation of the acceptor c_{dir} are determined. Based on the mean FRET efficiency of the donor-only population (blue circle) for c_{leak}, and the mean stoichiometry of the acceptor-only population (green circle) for c_{dir}, Gaussian fits (inserts) are calculated. (**B**) With the E/S histogram based on a dual channel burst search (DCBS) corrected by c_{leak} and c_{dir}, the excitation and detection correction factors β and γ are calculated. For at least two E/S separated populations, the mean E and S values are determined via Gaussian fits (red circles). The values are plotted as $1/S$ versus E (insert) and with the linear fit the y-axis intercept n and the slope m are determined. (**C**) The histogram corrected for c_{leak}, c_{dir}, γ, and β has a mean stoichiometry of $S = 0.5$. (adapted from Kramm, 2020 [286])

4.5.3.4 Calculation of FRET efficiency histograms

The corrected DCBS data were eventually analyzed using the Origin 2019b software (OriginLab). The FRET efficiency histograms were calculated from all bursts of a measurement using a binning range of −0.1–1.1 with a bin size of 0.032 (37 bins). After normalization to the total number of counts, the mean value of triplicate measurements was plotted as an *E*-histogram. The histograms were fitted with either one, two, or three Gaussian fits to determine the mean FRET efficiency of each population.

4.5.3.5 Burst variance analysis

To determine if heterogeneity of a sample results from dynamic processes or from the coexistence of multiple static conformations, a burst variance analysis (BVA) was performed [289]. The method compares the standard deviation of the FRET efficiency from individual molecules over time to the theoretically expected standard deviation of static molecules. If a molecule is dynamic and shows fluctuations in its FRET efficiency, the standard deviation is increased.

The BVA analysis tool implemented in PAM was employed, with the settings of 50 bursts per bin and a bin number of 20. On the x-axis, the proximity ratio was applied.

5. RESULTS

In this thesis, the CRISPR-Cas single-effector nuclease FnCas12a was deeply characterized via a variety of biochemical assays and single-molecule Förster resonance energy transfer (smFRET) measurements on diffusing molecules. The smFRET measurements were conducted to follow the conformational transitions of the FnCas12a protein through its activity cycle. Additionally, the influence of two central helices in the structure of Cas12a, namely the BH and helix 1, on the catalytic activity, structural integrity, and conformational transitions was investigated via a mutational analysis and the generation of FnCas12a BH and helix 1 variants.

The Cas12a gene from *Francisella novicida* was amplified from the plasmid pFnCpf1_min (Addgene #69975, **Table 16**) and successfully cloned into the pGEX-2TK vector via BamHI and EcoRI restriction sites (**Table 17**). This resulted in an expression plasmid that encoded the wildtype FnCas12a sequence with an N-terminal GST-tag and a Thrombin cleavage site (**Figure 15A**).

5.1 Characterization of WT FnCas12a

5.1.1 Biochemical characterization of WT FnCas12a

In this section, WT FnCas12a was characterized biochemically. As negative control and comparison, a deactivated FnCas12a (dCas12a) was generated by mutating E1006, one of the catalytic residues (D917, E1006, and D1255, **Figure 3**), to alanine via site-directed mutagenesis. Both proteins could be purified to homogeneity (**Figure 15B**, Appendix III.i, **Figure 61**). WT FnCas12a has a molecular weight of 151.9 kDa. In SDS-PAGEs WT FnCas12a and all subsequently purified variants migrated faster and were detected below the 150 kDa band of the protein ladder. The activity of the FnCas12a proteins was analyzed in a plasmid cleavage assay utilizing an EGFP-hAgo2 plasmid linearized with SmaI as target DNA (7300 bp, **Table 16**). The addition of WT FnCas12a together with the crRNA resulted in the expected formation of two cleavage products (4400 and 2900 bp), whereas dCas12a did not show any cleavage activity, neither did the crRNA alone nor WT FnCas12a alone (**Figure 15C**).

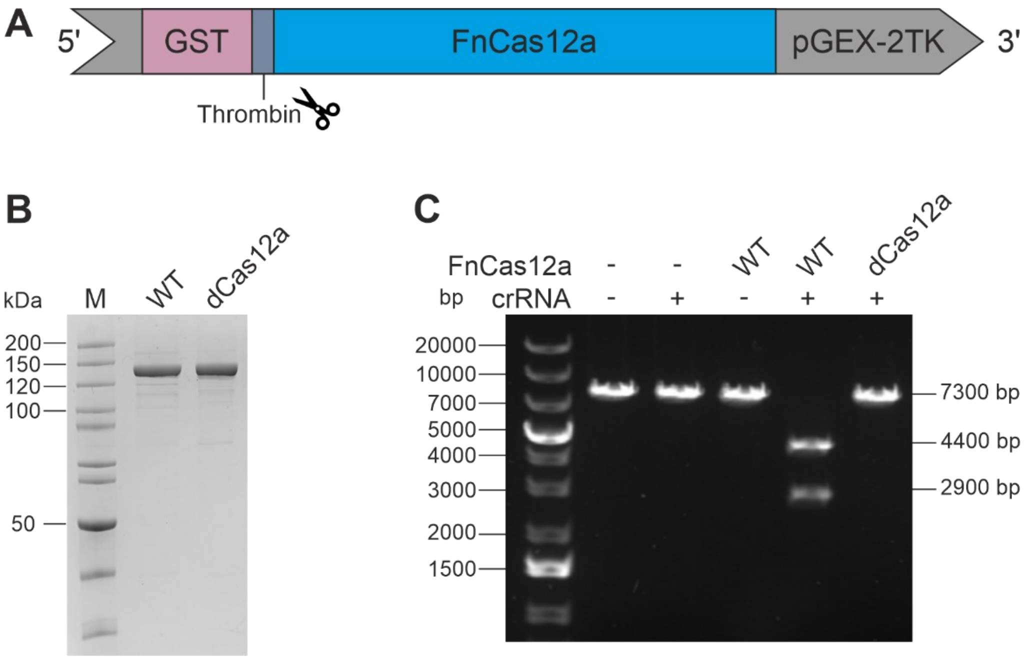


Figure 15: Purification and plasmid cleavage assay of FnCas12a.
(**A**) Schematic representation of the FnCas12a expression construct on the pGEX-2TK vector (grey). FnCas12a (blue) was expressed with an N-terminal GST-tag (pink), separated by a thrombin cleavage site (purple). (**B**) SDS-PAGE of purified WT FnCas12a and dCas12a (1.7 µg each, molecular weight of WT FnCas12a: 151.9 kDa). M: PageRuler unstained (ThermoFisher Scientific). (**C**) Plasmid cleavage assay using WT FnCas12a and dCas12a (37.5 nM) and linearized target DNA at 5 nM. The reaction was incubated at 37 °C for 1 h. The linearized target DNA plasmid (7300 bp) is cleaved into two products (4400 and 2900 bp). 1 kb plus DNA ladder was used as standard and the reaction separated on a 0.8% agarose 1x TAE gel.

In order to better evaluate the cleavage pattern of FnCas12a, a cleavage assay established by Swarts *et al.* based on fluorescently labeled short DNA target strands (KW5 (**Table 20**), 58 nt, **Figure 16A**) was used [133]. To follow the cleavage reaction on both strands simultaneously, a Cy3 and a Cy5 label were attached to the 5'-end of the non-target strand (NTS) and the target strand (TS), respectively. To first assess the binding behavior of FnCas12a to the short target DNA, an electrophoretic mobility shift assay (EMSA) of WT FnCas12a and dCas12a using increasing protein concentrations was conducted (**Figure 16B**). Both proteins bind the target DNA, visible in a shift of protein-bound DNA. The bands for WT FnCas12a are mainly green, as the protein is active and cleaves the target DNA but remains bound to the PAM-proximal part of the DNA (Cy3-labeled) after cleavage. dCas12a does not cleave the DNA, consequently, the Cy5-labeled DNA is still bound to the protein, resulting in a yellow band.

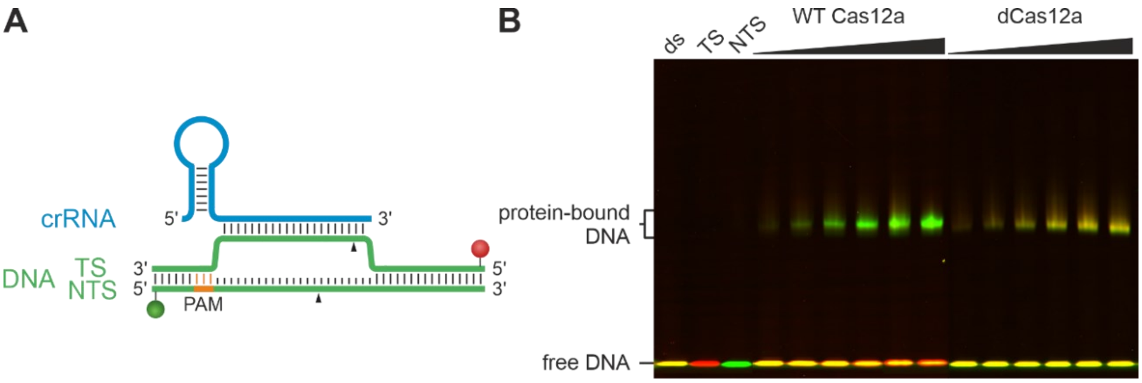


Figure 16: EMSA of FnCas12a-crRNA-target DNA complexes.
(**A**) Schematic model of the short target DNA (KW5, 58 nt) doubly labeled with a Cy3 label (green sphere) at the 5' end of the non-target strand (NTS) and a Cy5 label (red sphere) at the 5'-end of the target strand (TS). (**B**) WT FnCas12a and dCas12a were used in increasing equimolar concentrations of protein and crRNA (10, 20, 50, 75, 100, or 150 nM) over target DNA (10 nM). After incubation for 1 h at 37 °C, the products were analyzed on a non-denaturing Tris-Glycine gel (6% PAA). Free DNA was detected at the bottom of the gel, protein-bound DNA was retarded and detected as a slowly migrating complex.

Next, the cleavage activity of FnCas12a was tested on the short doubly labeled target DNA (**Figure 17**). With increasing concentrations of WT FnCas12a, more target DNA was cleaved. dCas12a did not show any cleavage activity.

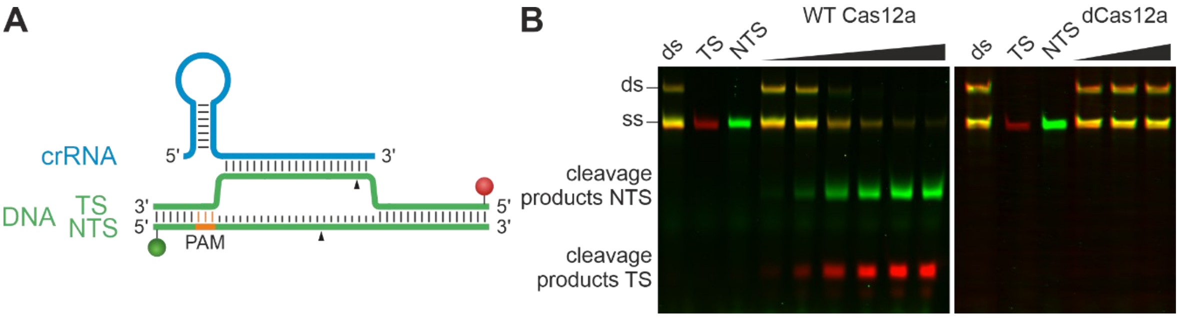


Figure 17: Cleavage assay of WT FnCas12a and dCas12a on a short target DNA.
(**A**) Schematic model of the short target DNA (KW5, 58 nt) doubly labeled with a Cy3 label (green sphere) at the 5'-end of the non-target strand (NTS) and a Cy5 label (red sphere) at the 5'-end of the target strand (TS). Cleavage positions are indicated with black arrowheads. (**B**) WT FnCas12a and dCas12a were used in increasing equimolar concentrations of protein and crRNA (WT: 10, 20, 50, 75, 100, or 150 nM; dCas12a: 10, 75, or 150 nM) over target DNA (10 nM). After incubation for 1 h at 37 °C, the cleavage products were analyzed on a 15% denaturing PAA gel.

To further analyze the cleavage activity of WT FnCas12a, cleavage kinetics in a time range of 1 s to 1 h were conducted with the short doubly labeled target DNA as substrate (**Figure 18A, C**). The observed cleavage pattern detected for reaction times up to three minutes suggested that an initial cleavage product is produced that is subsequently shortened. Therefore, to analyze cleavage products with single-nucleotide resolution, the cleavage products were resolved on a denaturing sequencing gel (**Figure 18E**). This confirmed that at the start of the reaction, the initial cleavage product of FnCas12a is longer and gets shortened over time, to yield the main product. This is the result of a trimming activity that shortens the NTS by up to 4 nt [231]. After cleavage of the target DNA, the PAM-distal part of the DNA is released but Cas12a remains bound to the PAM-proximal part [229] (here labeled with Cy3 at the NTS). While still bound, the NTS can re-enter the active site and Cas12a further shortens it via its intrinsic trimming activity.

To analyze the cleavage pattern on the PAM-proximal part, the same target DNA construct carrying the fluorescent labels at the 3'-ends was used (**Figure 18B, D, F**). For the NTS, two cleavage products were detected, appearing simultaneously after a very short cleavage time. This suggests an initial imprecise cleavage at two different positions, before the trimming of the PAM-proximal part takes place, and is in agreement with data by Swarts *et al.* [220]. The cleavage pattern for the TS shows one main cleavage product and two additional products 1 nt longer and 1 nt shorter than the main product, observed for the 3'- as well as for the 5'-labeled TS (**Figure 18E, F**). The kinetic analysis showed that no trimming takes place at the PAM-distal part of the TS. This data is in agreement with previous studies that reported the imprecise cleavage of the TS by AsCas12a and FnCas12a [226,231].

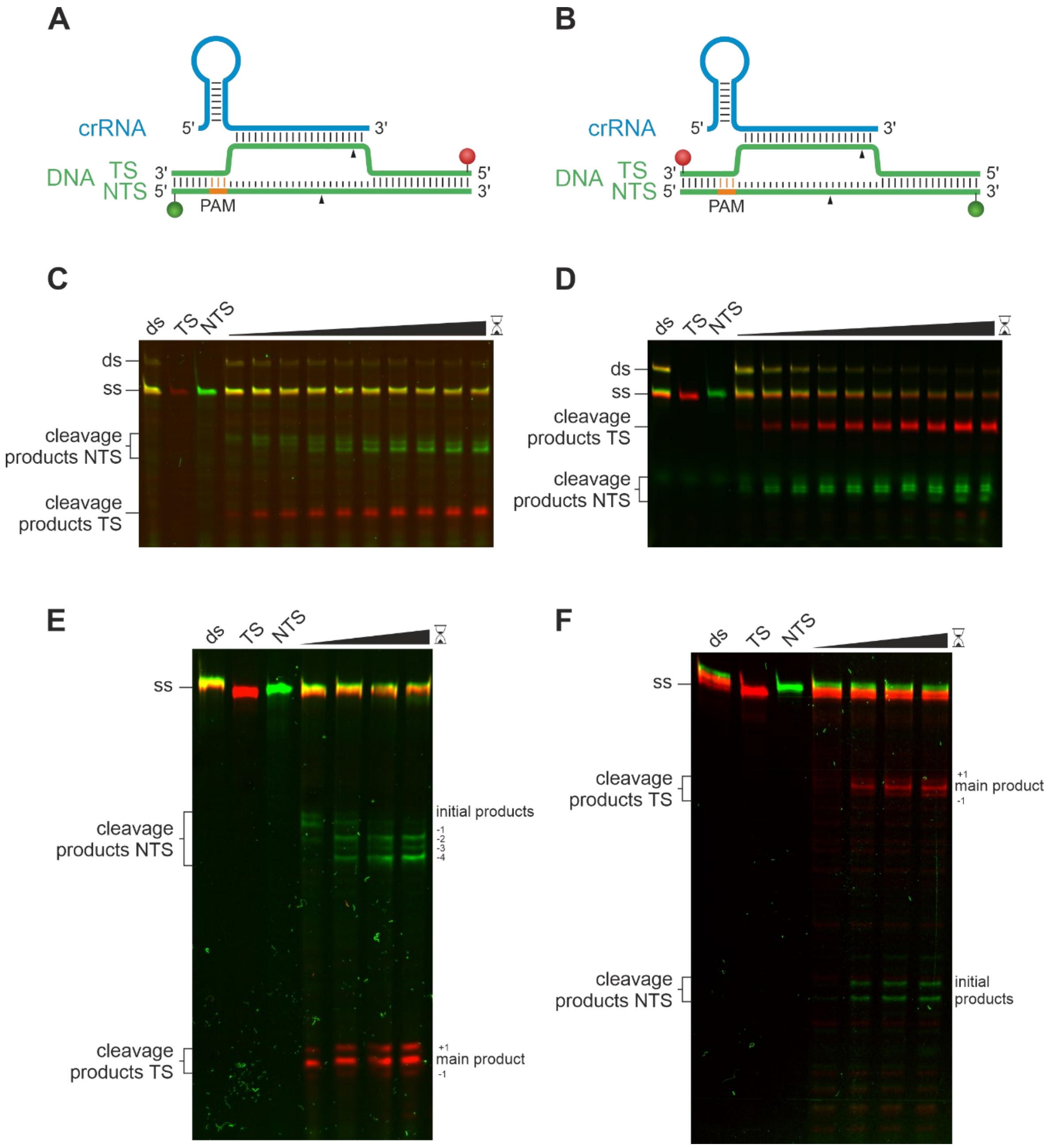


Figure 18: Cleavage kinetics of WT FnCas12a on a short doubly labeled target DNA.
Reactions contained WT FnCas12a and crRNA (C, D 75 nM; E, F 750 nM) in 7.5-fold molar excess over the target DNA (58 nt; C, D 10 nM; E, F 100 nM). Reactions were stopped with 125 mM EDTA after different time intervals. (**A**) and (**B**) Schematic representation of the labeled DNA construct with crRNA. (**A**) Cy3 (green sphere) and Cy5 (red sphere) labels are coupled to the 5'-end of the DNA strands. (**B**) Cy3 (green sphere) and Cy5 (red sphere) labels are attached to the 3'-end of the DNA strands. (**C**) and (**D**) Cleavage kinetics: reactions were stopped after 1 s, 30 s, 1 min, 3 min, 5 min, 10 min, 20 min, 30 min, 45 min, or 1 h incubation at 37 °C. Samples were analyzed on a 15% denaturing PAA gel. (**E**) and (**F**) Cleavage kinetics analyzed on a high-resolution gel. Reactions were stopped after 1 s, 5 min, 30 min, or 1 h incubation at 37 °C. (**C**) and (**E**) 5'-labeled target DNA; (**D**) and (**F**) 3'-labeled target DNA.
(adapted from Wörle *et al.* [290])

Next, FnCas12a cleavage on the TS only was assayed (**Figure 19**), showing that WT FnCas12a together with its crRNA can process the TS even in the absence of the NTS. Cleavage of TS only did not produce one main product and two side products, but three products with comparable signal intensity, indicating that without the NTS, the accommodation of the TS in Cas12a is more flexible and leads to an equal distribution of cleavage events on the different possible cleavage sites.

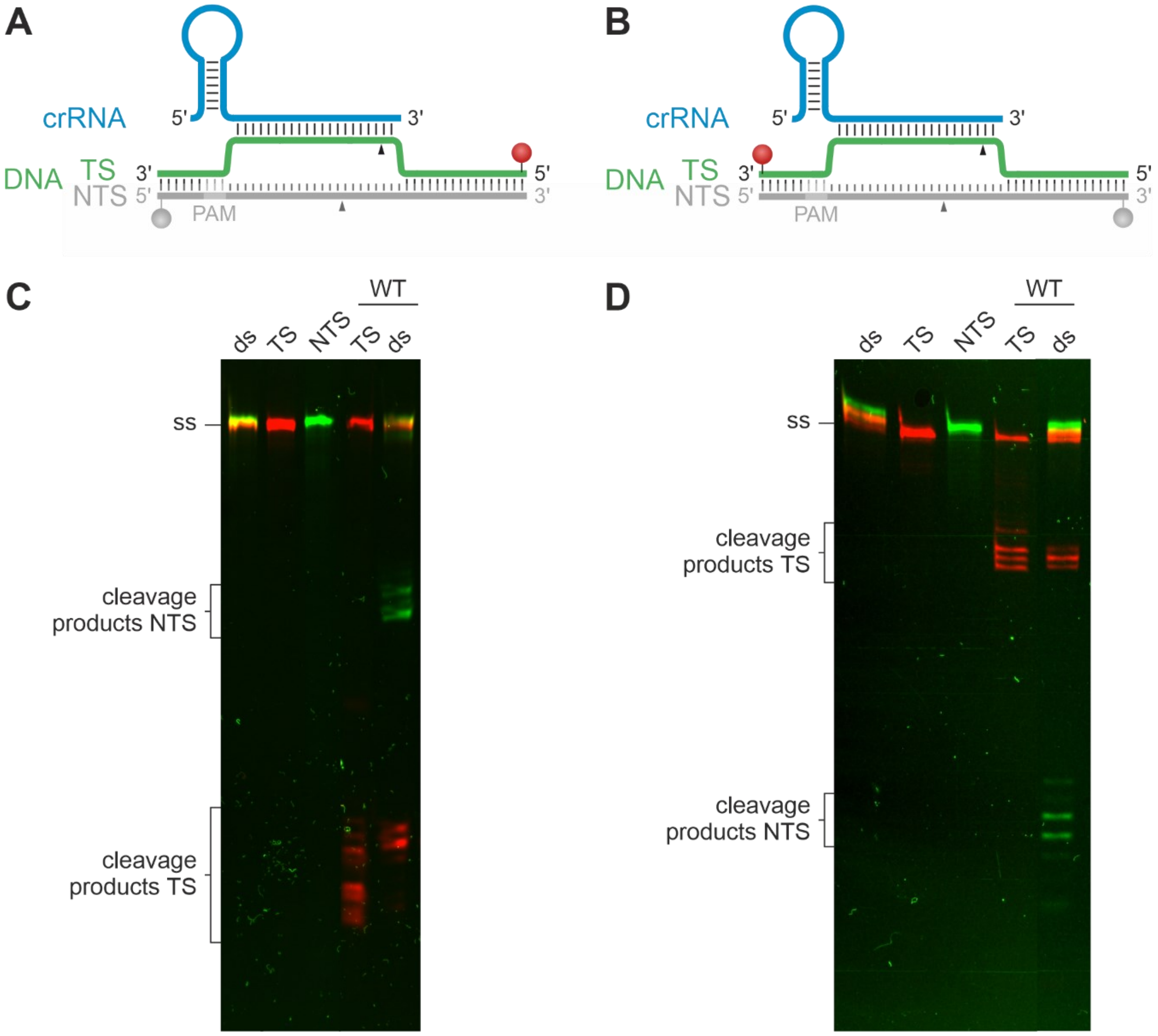


Figure 19: Cleavage assay of WT FnCas12a using a single-stranded target strand.
The cleavage reactions contained WT FnCas12a and crRNA (75 nM) in 7.5-fold molar excess over the target strand (TS) or ds target DNA (58 nt; 10 nM). After incubation of the reaction at 37 °C for 1 h, samples were analyzed on a high-resolution denaturing 15% PAA gel. (**A**) and (**B**) Schematic representation of the labeled DNA construct with crRNA. (**A**) Cy5 (red sphere) was coupled to the 5'-end of the DNA strand. (**B**) Cy5 (red sphere) was attached to the 3'-end of the DNA strand. The grey-colored NTS was excluded in the reaction with TS only and carries a Cy3 label in the reaction with dsDNA. (**C**) Cleavage assay with 5'- labeled and (**D**) with 3'-labeled DNA strands. (adapted from Wörle *et al.* [290])

In addition to its ability to cleave target DNA with the DNase active site in the RuvC domain, Cas12a is also able to self-process its own pre-crRNA with an RNase active site in the WED domain [132]. To compare the activity of WT FnCas12a to an inactive protein, the FnCas12a HKK variant (H843A/K852A/K869A) was generated in which the RNase active site residues were mutated to alanine (**Figure 3**). To monitor the pre-crRNA processing activity, a pre-crRNA was incubated with FnCas12a (**Figure 20C**). This experiment showed that the pre-crRNA is processed by WT FnCas12a and revealed that even the crRNA used is further processed. Interestingly, the crRNA and the pre-crRNA substrates were not processed to the same length, with the pre-crRNA yielding a second product slightly shorter than the product of the processed crRNA. The catalytically inactive FnCas12a HKK variant did not process RNA. An additional assay with a 5'-labeled pre-crRNA (**Figure 20D**) showed the pre-crRNA processing activity of WT FnCas12a and dCas12a by the cleavage of the 5'-label which was detected at the bottom of the gel. FnCas12a HKK was not able to process the labeled pre-crRNA.

Both, WT FnCas12a and FnCas12a HKK cleaved the doubly labeled short target DNA with pre-crRNA, albeit with a slightly reduced cleavage efficiency compared to the mature crRNA (**Figure 20E**).

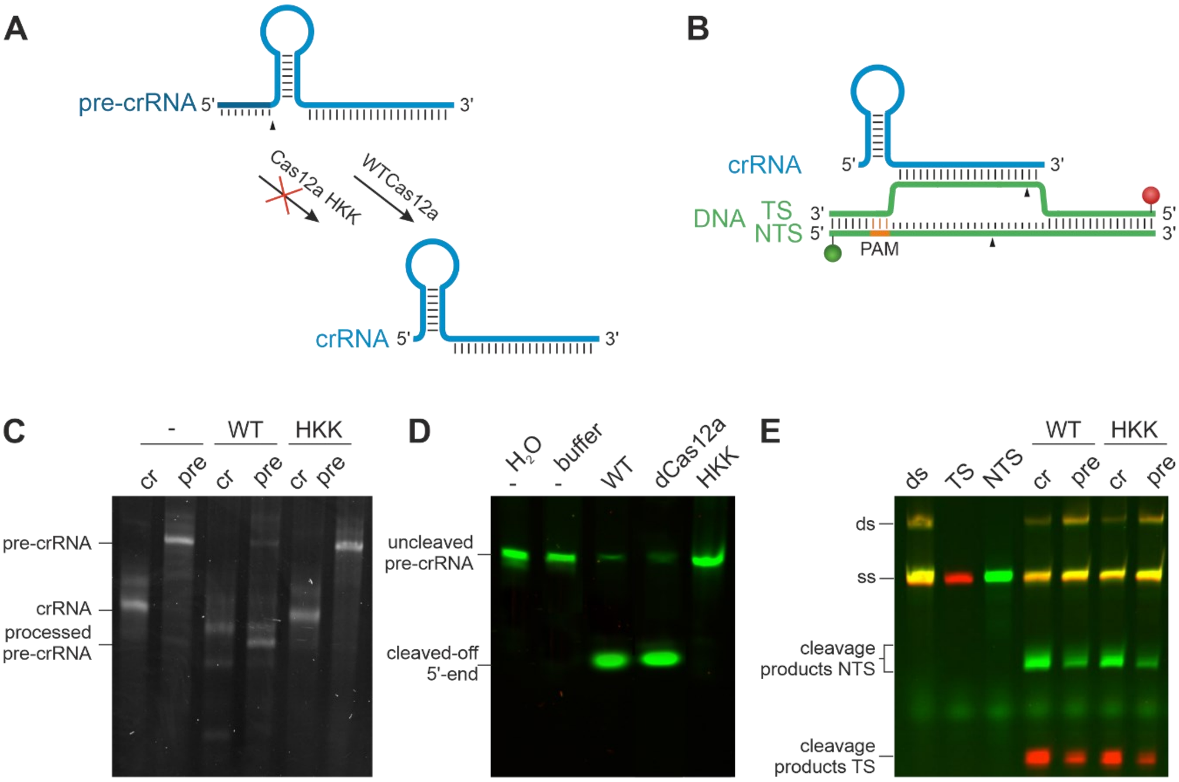


Figure 20: Cleavage assay of WT FnCas12a and the HKK variant using pre-crRNA as substrate. (**A**) Schematic model of a pre-crRNA processed by WT FnCas12a to the mature crRNA, but not by FnCas12a HKK. (**B**) Schematic model of the crRNA and the doubly labeled target DNA (Cy3 (green sphere) at the 5'-end of the NTS and Cy5 (red sphere) at the 5'-end of the TS). (**C**) Cleavage assay of WT FnCas12a and FnCas12a HKK (150 nM) with mature crRNA (cr) or pre-crRNA (pre) (150 nM). Reaction products were analyzed on a 15% denaturing PAA gel and stained with SYBR Gold for 10 min. (**D**) WT FnCas12a, dCas12a, and FnCas12a HKK were used in a two-fold molar excess of protein (20 nM) over Cy3 5'-labeled pre-crRNA (10 nM). Pre-crRNA processing could be followed by the cleavage of the 5'-end. (**E**) Cleavage assay of WT FnCas12a and FnCas12a HKK (75 nM) with pre-crRNA or mature crRNA (75 nM) on 5'-labeled DNA (10 nM). Reaction products were analyzed on a 15% denaturing PAA gel. The cleavage efficiency of FnCas12a loaded with pre-crRNA or mature crRNA is comparable but slightly reduced with pre-crRNA.

5.1.2 Conformational transitions of an RNA/DNA hybrid

To observe conformational changes in a preformed crRNA/target DNA hybrid upon FnCas12a binding, the crRNA strand (KWcpf1#2) labeled with Atto647N as acceptor and the TS of the target DNA (KW3single) labeled with Atto532 as donor were hybridized (**Figure 21A**). The constructs were analyzed in confocal smFRET measurements using 100 pM of the crRNA/DNA hybrid. In the FRET efficiency histogram, a single population with a high FRET efficiency of $E = 0.78$ was found for the RNA/DNA hybrid, whereas the addition of WT FnCas12a led to a significant shift to lower FRET efficiencies ($E = 0.20$). The drastic conformational change is a result of RNA/DNA splaying and the incorporation of the nucleic acids into the protein upon association of Cas12a.

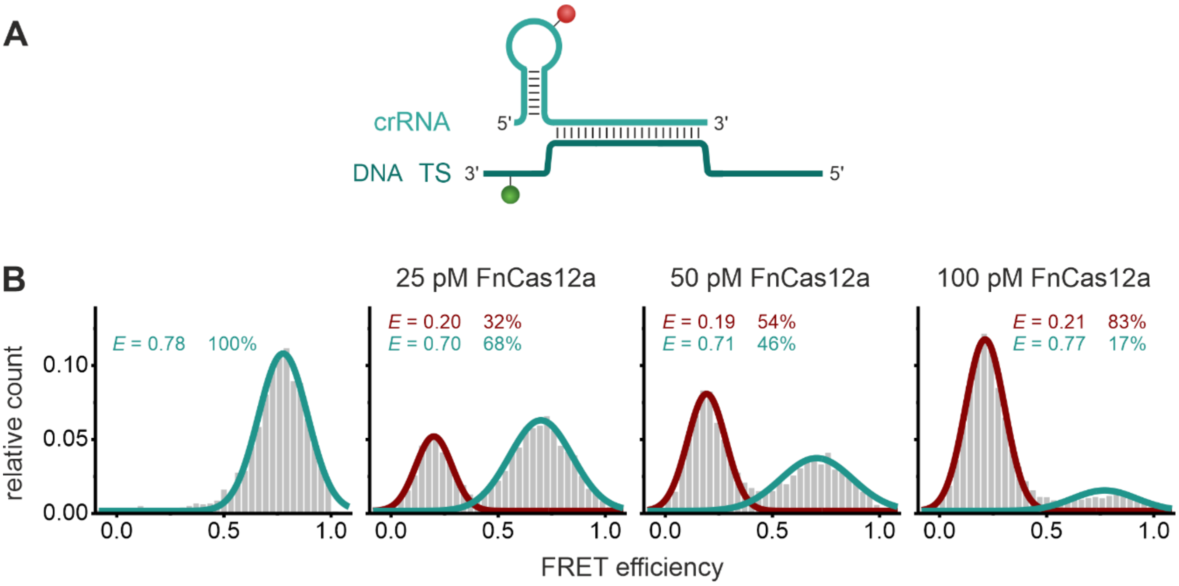
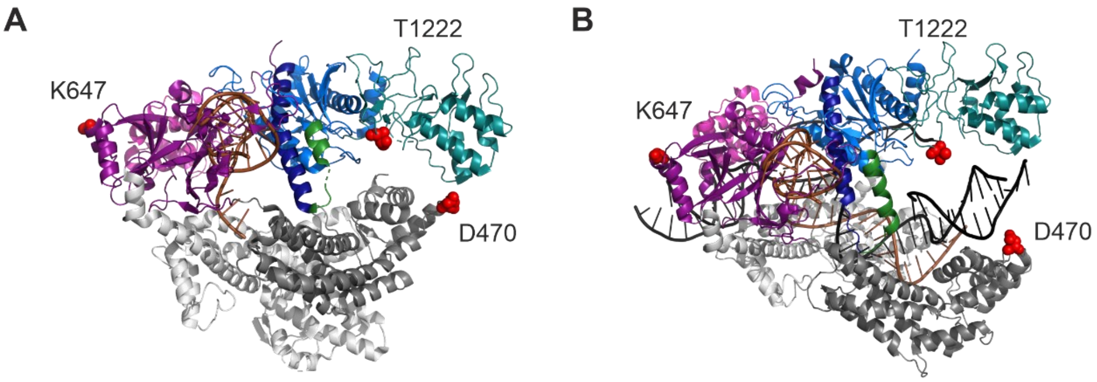


Figure 21: Single-molecule FRET measurements of an RNA/DNA hybrid with WT FnCas12a. (**A**) Schematic model of the crRNA/DNA hybrid. The crRNA KWcpf1#2 was labeled with Atto647N (red sphere), the DNA strand KW3single (TS) was labeled with Atto532 (green sphere). (**B**) FRET efficiency histograms of the measurements with 100 pM of doubly labeled RNA/DNA hybrid and the addition of 25 pM, 50 pM, or 100 pM of WT FnCas12a protein. The mean FRET efficiencies E and the percentage distributions of the fitted populations are given. Measurements were performed in triplicates. Additional fitting parameters and molecule counts can be found in Appendix III.ii.

5.1.3 Conformational transitions of FnCas12a

Several X-ray and cryo-EM structures are available for Cas12a that show the architecture of the binary (mature crRNA + Cas12a) and the ternary (crRNA + Cas12a + target DNA) complexes [133,217–221,224,226,242,244–246,291] revealing that the transition from the binary to the ternary complex is accompanied by a structural rearrangement from a closed to a more open state of the enzyme. Nevertheless, structural information for the apo Cas12a enzyme and the pre-crRNA-Cas12a complex in the absence of a DNA target is not available. smFRET measurements on diffusing molecules were performed to follow the conformational transitions of the protein and to determine relative intramolecular distances in WT FnCas12a. FnCas12a harbors nine cysteine residues, which makes site-specific labeling via a cysteine-labeling strategy unpractical. Therefore, the unnatural amino acid *para*-azido-L-phenylalanine (AzF) was site-specifically incorporated via the Amber-suppressor strategy [249,255]. The label was either placed in the WED domain (K647AzF), the Nuc lobe (T1222AzF), or the REC lobe (D470AzF) yielding the following variants: FnCas12a$^{K647AzF/T1222AzF}$ (termed FnCas12a$^{WED\text{-}Nuc}$), FnCas12a$^{D470AzF/K647AzF}$ (termed FnCas12a$^{WED\text{-}REC}$), and FnCas12a$^{D470AzF/T1222AzF}$ (termed FnCas12a$^{REC\text{-}Nuc}$) (**Figure 22**). According to crystal structures, the distances between the chosen residues in the binary and the ternary complex are: positions K647 and T1222: 78.1 and 79.8 Å, D470 and K647: 98.0 and 104.4 Å, and D470 and T1222: 30.8 and 36.0 Å, respectively [133,220] (**Figure 22C**).



C

	binary complex	ternary complex	ternary complex −NTS
K647-T1222	78.1 Å	79.8 Å	77.3 Å
D470-K647	98.0 Å	104.4 Å	105.7 Å
D470-T1222	36.0 Å	30.8 Å	33.2 Å

Figure 22: Labeling positions in FnCas12a.
The positions D470 (REC lobe), K647 (WED domain), and T1222 (Nuc lobe) in FnCas12a for the incorporation of the unnatural amino acid AzF and subsequent labeling are shown as red spheres. Color coding of the FnCas12a domains is according to **Figure 3**. (**A**) Binary complex (PDB: 5NG6) and (**B**) ternary complex (PDB: 6I1K) of FnCas12a with labeling positions. (**C**) Distances between the Cα-atoms of the labeling positions in the binary complex, ternary complex, and ternary complex without the NTS according to the structural information [133,220]. (adapted from Wörle *et al.* [290])

To rule out any influences of the incorporated unnatural amino acid on the nucleic acid binding capacity or activity of FnCas12a, EMSAs and activity assays were performed (**Figure 23**). These demonstrated that all variants were able to bind nucleic acids and cleave the target DNA.

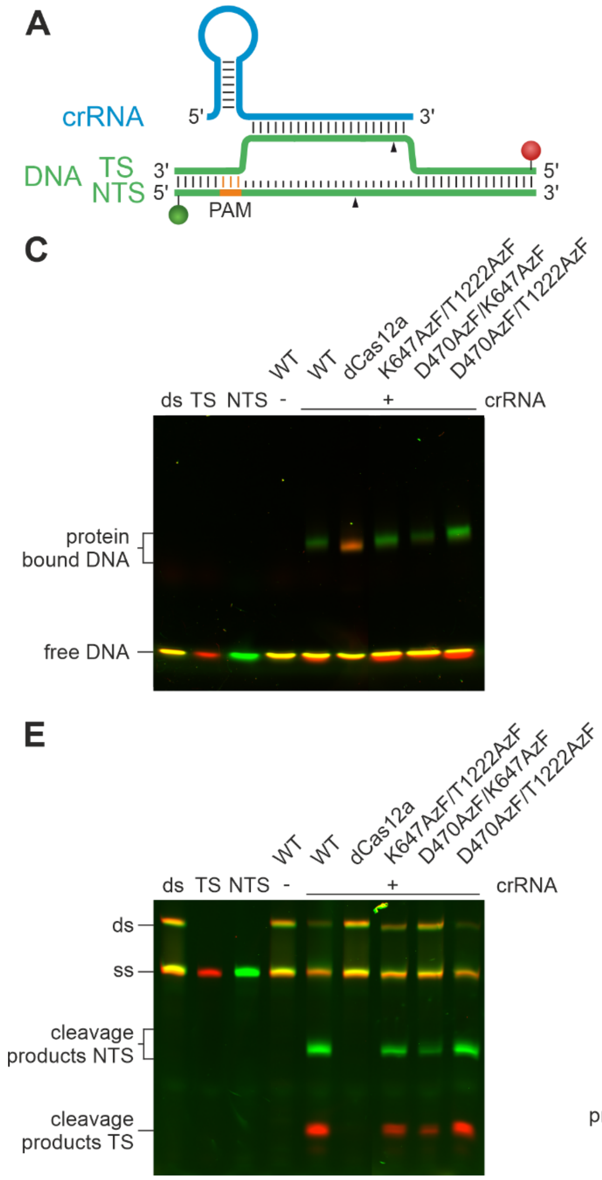


Figure 23: Activity assays of FnCas12a variants carrying two unnatural amino acid residues. WT FnCas12a, dCas12a, and FnCas12a variants with the site-specifically engineered unnatural amino acid AzF were used in a 7.5-fold molar excess of protein and crRNA (75 nM) over target DNA (10 nM). Reactions were incubated for 1 h at 37 °C. The short target DNA (58 nt) was doubly labeled with a Cy3 label (green) at the non-target strand (NTS) and a Cy5 label (red) at the target strand (TS). (**A**) Cy3 (green sphere) and Cy5 (red sphere) labels were coupled to the 5'-end of the DNA strands, respectively. (**B**) Cy3 (green sphere) and Cy5 (red sphere) labels were attached to the 3'-end of the DNA strands, respectively. (**C**) EMSA of FnCas12a variants with 5'-labeled DNA. (**D**) EMSA of FnCas12a variants with 3'-labeled DNA. (**E**) Cleavage assay of FnCas12a variants with 5'-labeled DNA. (**F**) Cleavage assay of FnCas12a variants with 3'-labeled DNA. Due to low protein expression yield, the FnCas12a$^{\text{D470AzF/K647Azf}}$ variant could only be used at a reduced concentration (63 nM) resulting in a reduced fluorescence signal for the DNA-protein complex and the cleaved DNA.

The fluorescent dyes DyLight 550 and DyLight 650 were coupled to the azide group of AzF via the Staudinger-Bertozzi ligation [261]. The stochastic labeling approach used here, yields proteins carrying a donor-acceptor dye pair, or proteins that only carry the donor or only the acceptor fluorophore. During single-molecule measurements, FnCas12a molecules that carry the donor-acceptor pair can be spectroscopically sorted via separate detection channels for donor and acceptor emission based on the PIE excitation scheme [287]. After each labeling procedure, a validating PAGE was performed to verify that the labeling reaction was successful (**Figure 24**). Efficient labeling was visible in clear and equal bands in the fluorescent scans.

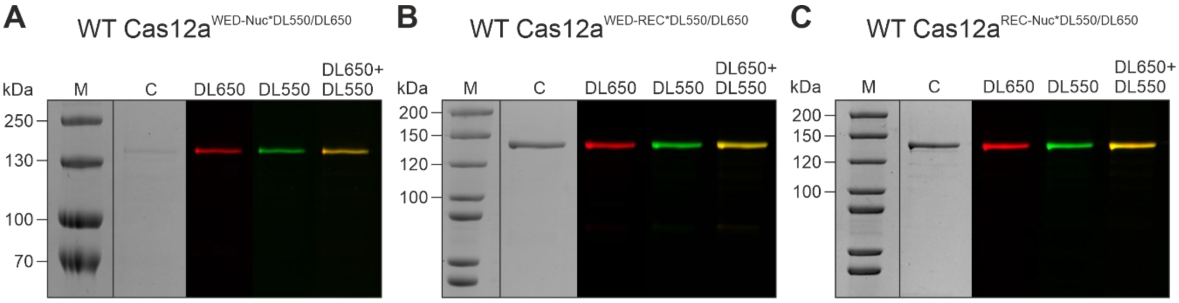


Figure 24: SDS-PAGE analysis of doubly labeled FnCas12a variants.
(**A**)–(**C**) Validating SDS-PAGE of doubly labeled FnCas12a variants (fluorophores are DyLight 550 (DL550) and DyLight 650 (DL650)). M: PageRuler prestained (**A**) / PageRuler unstained (**B**, **C**); C: Coomassie-stained gel; DL650: fluorescence scan at 625–650 nm excitation for DyLight 650; DL550: fluorescence scan at 520–545 nm excitation for DyLight 550; DL650+DL550: overlay of fluorescence scans DL650 and DL550. (adapted from Wörle *et al.* [290])

To follow conformational changes of FnCas12a along the WED-Nuc axis and the WED-REC axis, smFRET measurements with the FnCas12a$^{WED-Nuc}$ and FnCas12a$^{WED-REC}$ variants, respectively, were performed. The following samples were measured (i) the apo enzyme, (ii) the binary complex composed of FnCas12a with pre-crRNA, (iii) the binary complex composed of FnCas12a with mature crRNA, (iv) the ternary complex (FnCas12a + mature crRNA + dsDNA target), and (v) a ternary complex loaded with a single-stranded target DNA (ternary complex −NTS). Unlabeled KW5 nucleic acid constructs were used for the smFRET measurements (**Table 20**). Both, the FnCas12a$^{WED-Nuc}$ as well as the FnCas12a$^{WED-REC}$ variant, showed a single low FRET population at 0.15 ± 0.01 and 0.13 ± 0.02 FRET efficiency, respectively, for all tested conditions (**Figure 25B**). In agreement with structural studies (**Figure 22**) [133,220], this shows that the axial distances between the WED domain and the Nuc lobe as well as between the WED domain and the REC lobe do not significantly change when FnCas12a progresses through its activity cycle. Probing the FnCas12a$^{Nuc-REC}$ variant, carrying the fluorescent labels in the two lobes, a low FRET population with a mean FRET efficiency of 0.12 (± 0.04) and a high FRET population (E = 0.97 ± 0.01) was detected for the apo state of FnCas12a (**Figure 25B**). This indicates that the nucleic acid-free FnCas12a can adopt a closed and an open conformation with the majority of molecules (68 ± 7%) in the closed state (high FRET population). The addition of crRNA to the protein led to a shift to the high FRET population (84 ± 2%), both for the addition of mature crRNA as well as upon the addition of pre-crRNA, indicating a higher percentage of closed proteins. Increasing the crRNA concentration led to an increase in the high FRET population (Appendix III.iii, **Figure 63**), indicating that the formation of the

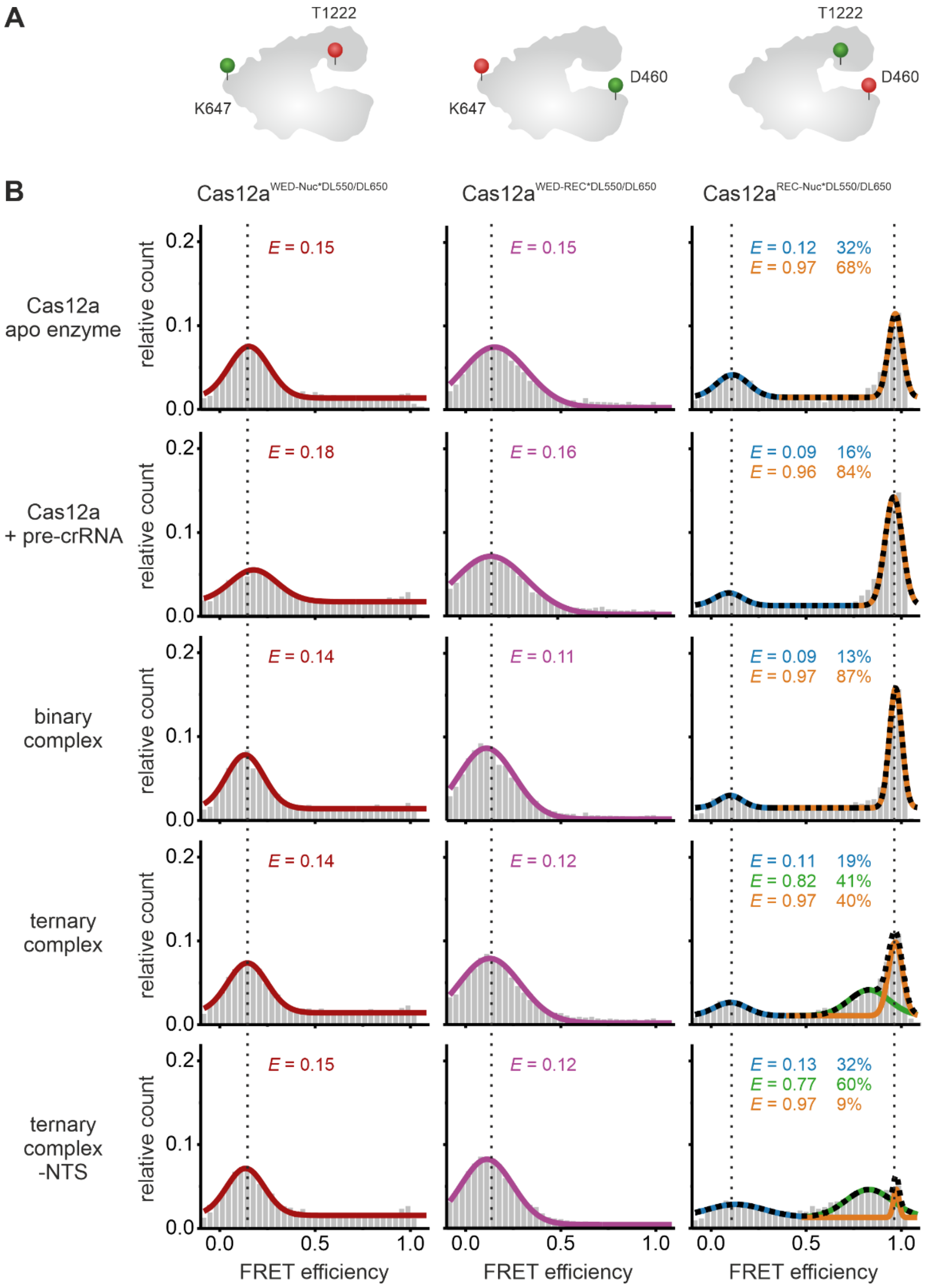


Figure 25: Confocal single-molecule FRET measurements on doubly labeled WT FnCas12a. (**A**) Schematic model of FnCas12a showing the labeling positions D470 (REC domain), K647 (WED domain), and T1222 (Nuc domain). (**B**) FRET efficiency distributions for the FnCas12a apo enzyme, FnCas12a with pre-crRNA (1 nM pre-crRNA), the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM target DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS DNA). FnCas12a$^{\text{WED-Nuc*DL550/DL650}}$ and FnCas12a$^{\text{WED-REC*DL550/DL650}}$ displayed one single FRET population for the apo, RNA-, and RNA-and-DNA-bound state. FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ showed two FRET populations for the apo enzyme and the RNA-bound state, with the high FRET population increasing with the addition of RNA. In the ternary complex, FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ showed three FRET populations with an additional medium FRET population. Histograms show the average data of three independent measurements. The histograms were fitted with a single, double, or triple Gaussian function and the overall fit is shown in black (dotted line). The mean FRET efficiencies E and the percentage distributions of the fitted populations are given in the histograms. Additional information regarding fit parameters and molecule counts can be found in Appendix III.iii. (adapted from Wörle *et al.* [290])

binary complex is dependent on the crRNA concentration. This validates that the high FRET population represents the binary complex and that the binding of the crRNA induces the closure of the protein. Upon loading of the DNA, a third population with a FRET efficiency of 0.82 (± 0.02) arose. Hence, the protein re-opens to provide enough space to bind the target DNA. With increasing target DNA concentrations, the relative number of molecules in the medium FRET population increased, whereas the number of molecules in the high FRET population decreased (Appendix III.iii, **Figure 63**, **Table 39**). This indicates that the loading of the target DNA is responsible for the formation of the medium FRET population. Even at elevated target DNA concentrations, loading was only moderately efficient and approximately 30–50% of the proteins loaded with a crRNA remained in the closed state. Additionally, a ternary complex lacking the NTS was assayed corresponding to the cleavage assay with TS only (**Figure 19**). Here, the high FRET population dropped to 9% (in the presence of the NTS: 40%), and the medium FRET population increased to 60%. This indicates that loading of the sterically less demanding single-stranded TS is very efficient and allows FnCas12a to successfully transition from the binary to the ternary complex.

Burst variance analysis (BVA) [289] revealed that FnCas12a showed some dynamic switching between the open and closed conformation in the apo state of the protein (**Figure 26**). The binary complex of FnCas12a predominantly adopted the closed conformation with a high FRET efficiency, but little switching occurred between low and high FRET efficiencies. In the ternary complex, dynamics increased, especially around the medium FRET population but also at low FRET efficiencies.

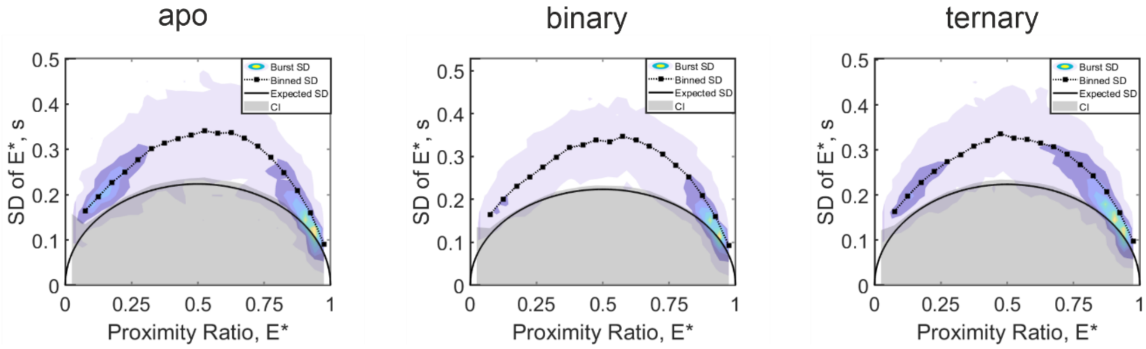


Figure 26: Burst variance analysis of WT FnCas12a$^{REC-Nuc*DL550/DL650}$.
The burst variance analysis of the FRET measurements of WT FnCas12a$^{REC-Nuc*DL550/DL650}$ showed some dynamic switching between low and high FRET efficiencies, especially for the apo protein. Dynamics were less prominent in the binary complex but increased for the ternary complex. No dynamic would be indicated if the black dotted line (binned standard deviation (SD) of the measurement) was identical to the black solid line (expected SD). The colored area shows the standard deviation of every single burst (burst SD) and the grey area shows the confidence interval (CI).

To rule out any effects that might have been missed because of the possibly fast processing of the pre-crRNA by WT FnCas12a during the incubation and single-molecule measurement time, the RNase-inactive variant FnCas12a HKK was employed to verify results with the RNase-deficient FnCas12a variant (**Figure 27**). The FnCas12a$^{WED-Nuc*DL550/DL650}$ HKK and

FnCas12a$^{WED-REC*DL550/DL650}$ HKK variants also displayed a low FRET population for all tested conditions including the binary complexes with pre-crRNA and mature crRNA. Like WT FnCas12a$^{REC-Nuc*DL550/DL650}$ also FnCas12a$^{REC-Nuc*DL550/DL650}$ HKK showed two FRET populations for the apo enzyme and the RNA-bound states. The high FRET population increased with the addition of RNA. Hence, the pre-crRNA/FnCas12a complex did not give rise to an additional FRET population, therefore, it can be assumed that the pre-crRNA/FnCas12a complex does not adopt a distinct conformation.

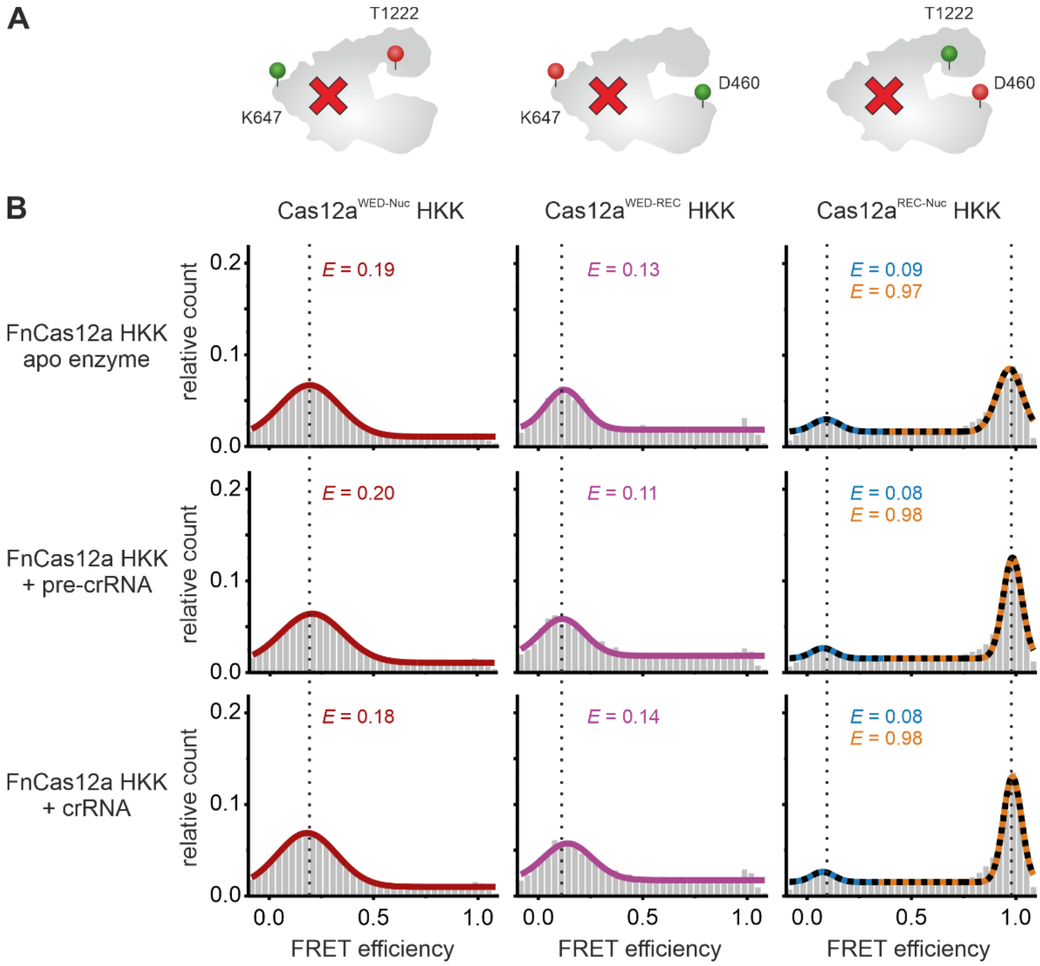


Figure 27: Confocal single-molecule FRET measurements using doubly labeled FnCas12a RNase-inactive variants.
(**A**) Schematic model of FnCas12a HKK showing the labeling positions D470 (REC domain), K647 (WED domain), and T1222 (Nuc domain). (**B**) FRET efficiency distributions for the FnCas12a HKK apo enzyme, FnCas12a HKK with pre-crRNA (1 nM pre-crRNA), and the binary complex (1 nM crRNA). FnCas12a$^{WED-Nuc*DL550/DL650}$ HKK and FnCas12a$^{WED-REC*DL550/DL650}$ HKK displayed a single FRET population for the apo and RNA-bound states. FnCas12a$^{REC-Nuc*DL550/DL650}$ HKK showed two FRET populations for the apo enzyme and the RNA-bound states, with the high FRET population increasing with the addition of RNA. Histograms show the average data of three independent measurements. The histograms were fitted with a single or double Gaussian function and the overall fit is shown in black (dotted line). The mean FRET efficiencies *E* of the fitted populations are given in the histograms. Additional information regarding fit parameters and molecule counts can be found in Appendix III.iv.

The measurements with REC-Nuc labeling positions were also performed with the DNase-inactive variant of FnCas12a (dCas12a). Highly comparable data to the WT measurement were obtained (**Figure 28**) that demonstrate that the medium FRET population corresponds to the ternary complex with bound target DNA and that the conformations of the ternary complex before cleavage (dCas12a) and after cleavage (WT FnCas12a) of the target DNA do not significantly differ.

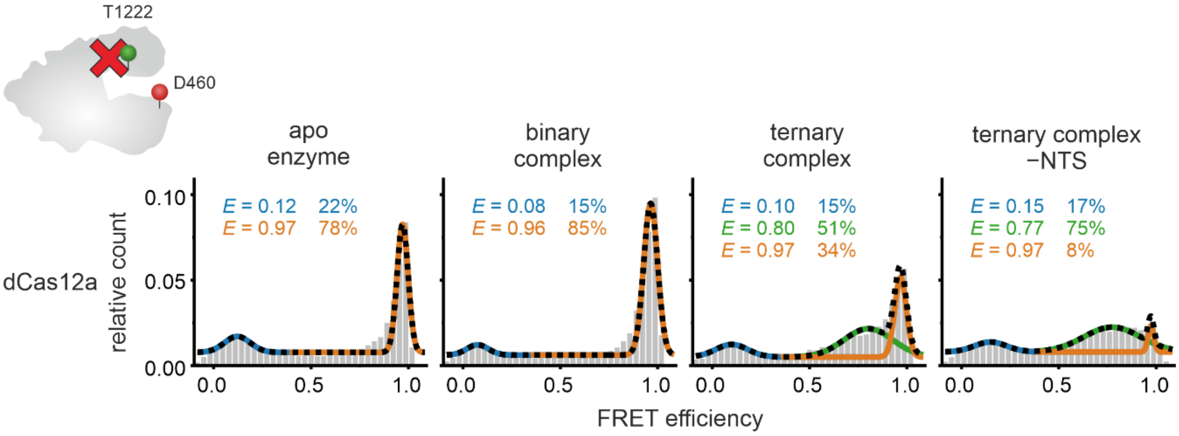


Figure 28: Confocal single-molecule FRET measurements using a doubly labeled FnCas12a DNase-inactive variant.
smFRET measurements were conducted with the dCas12a$^{\text{REC-Nuc*DL550/DL650}}$ apo enzyme, the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM target DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS). Histograms show the average data of three independent measurements. The histograms were fitted with a double or triple Gaussian function, the overall fit is shown in black (dotted line). The mean FRET efficiencies E and the percentage distribution of the populations are given in the histograms. Additional information regarding fit parameters and molecule counts can be found in Appendix 0. (adapted from Wörle *et al.* [290])

Taken together, the smFRET measurements recapitulate the available structural data of the binary complex (closed state) and the opening movement of FnCas12a upon target DNA binding (semi-closed state). A single-stranded TS is sufficient and even more efficient to induce the transition to the semi-closed state. Furthermore, the hitherto unknown structural state of the apo enzyme could be assessed with the smFRET measurements, which showed that apo FnCas12a can adopt an open and a closed conformation.

5.2 Collaborational analysis of variants to alter the target strand affinity

In this part, data are presented that were collected in a collaborational project with colleagues from the Australian National University in Canberra (AUS). Anthony Newman and Gaetan Burgio designed FnCas12a variants with the intention to change the affinity of Cas12a to the target strand of the DNA. This is one of the major interactions in the surveillance complex of Cas12a searching for a matching target sequence. The rationale was that by changing the target strand binding strength, the *cis*- or *trans*-cleavage activity might increase. The designed mutations in FnCas12a were K823A, K1066A, I195R, D194R, and N188R (**Figure 29**). The lysine to alanine mutations were designed with the intention to reduce the affinity of the protein to the TS, the arginine substitutions aimed to increase this affinity.

In collaboration, these variants were (i) cloned and purified to conduct cleavage assays, and (ii) cloned and purified with the unnatural amino acid AzF for subsequent labeling at positions D470 and T1222 to perform confocal smFRET measurements.

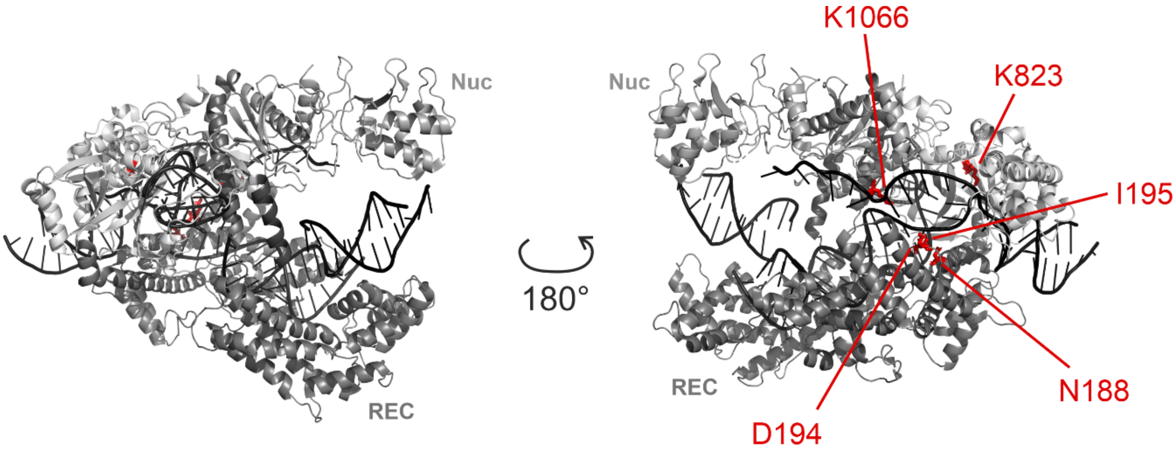


Figure 29: Positions of mutations to alter the target strand affinity.
In the ternary complex structure of FnCas12a (PDB: 6I1K) the positions for mutations to alter the TS affinity (N188, D194, I195, K823, and K1066) are highlighted in red.

All variants could be purified to homogeneity (**Figure 30A**) and cleavage assays with a short doubly labeled target DNA (KW5, **Table 20**) were performed (**Figure 30B, C**), revealing that all variants were active and cleaved the target DNA with comparable efficiency as WT FnCas12a with the exception of the FnCas12a N188R variant that showed a reduced cleavage efficiency. Additionally, FnCas12 K823A showed a slightly reduced cleavage efficiency.

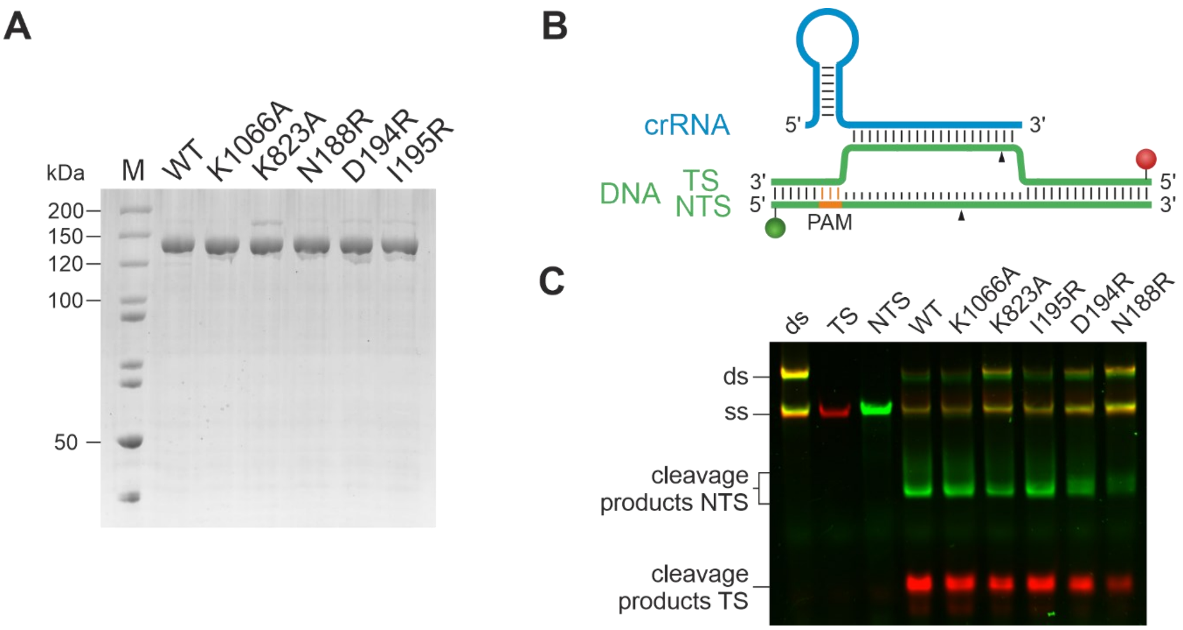


Figure 30: Purification and cleavage assay of FnCas12a variants to alter the target strand affinity. (**A**) SDS-PAGE of purified WT FnCas12a and FnCas12a variants to alter the TS affinity (1.7 µg each, molecular weight of WT FnCas12a: 151.9 kDa). M: PageRuler unstained (ThermoFisher Scientific). (**B**) Schematic model of the short target DNA (KW5, 58 nt) doubly labeled with a Cy3 label (green sphere) at the 5'-end of the non-target strand (NTS) and a Cy5 label (red sphere) at the 5'-end of the target strand (TS). (**B**) FnCas12a WT and FnCas12a variants to alter the TS affinity were used in 7.5-fold molar excess of protein and crRNA (75 nM) over target DNA (10 nM). After incubation for 1 h at 37 °C, the reaction products were analyzed on a 15% denaturing PAA gel.

The FnCas12a variant N188R, in which also AzF should be incorporated at positions D470 and T1222, failed to be expressed in sufficient quantities. Hence, smFRET measurements were not conducted with this variant.

All variants tested in the smFRET measurements showed highly comparable data to WT FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ (**Figure 31**). For the apo enzyme, all variants exhibited either an open (mean FRET efficiency of 0.11 ± 0.01) or a closed (mean FRET efficiency of 0.97 ± 0.004) conformation with 79% of molecules in the closed conformation. The addition of the crRNA efficiently shifted the equilibrium further to the closed conformation (92%). The addition of target DNA induced the formation of the medium FRET population ($E = 0.83 \pm 0.03$) for all variants. For FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ K823A the loading with target DNA was the least efficient with 53% of molecules still in the high FRET population (compared to ≈ 42% in the other variants), this is congruent with the slightly reduced cleavage efficiency in the cleavage assay. The measurements without the NTS showed efficient loading of the TS into the FnCas12a variants (65% in the medium FRET population) as seen for WT FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$. All variants exhibited a broad distribution for the medium FRET population lacking the NTS. This indicates that the altered TS affinity intended by the mutations, in fact, lowered both, the affinity for and the exact positioning of the TS for all variants.

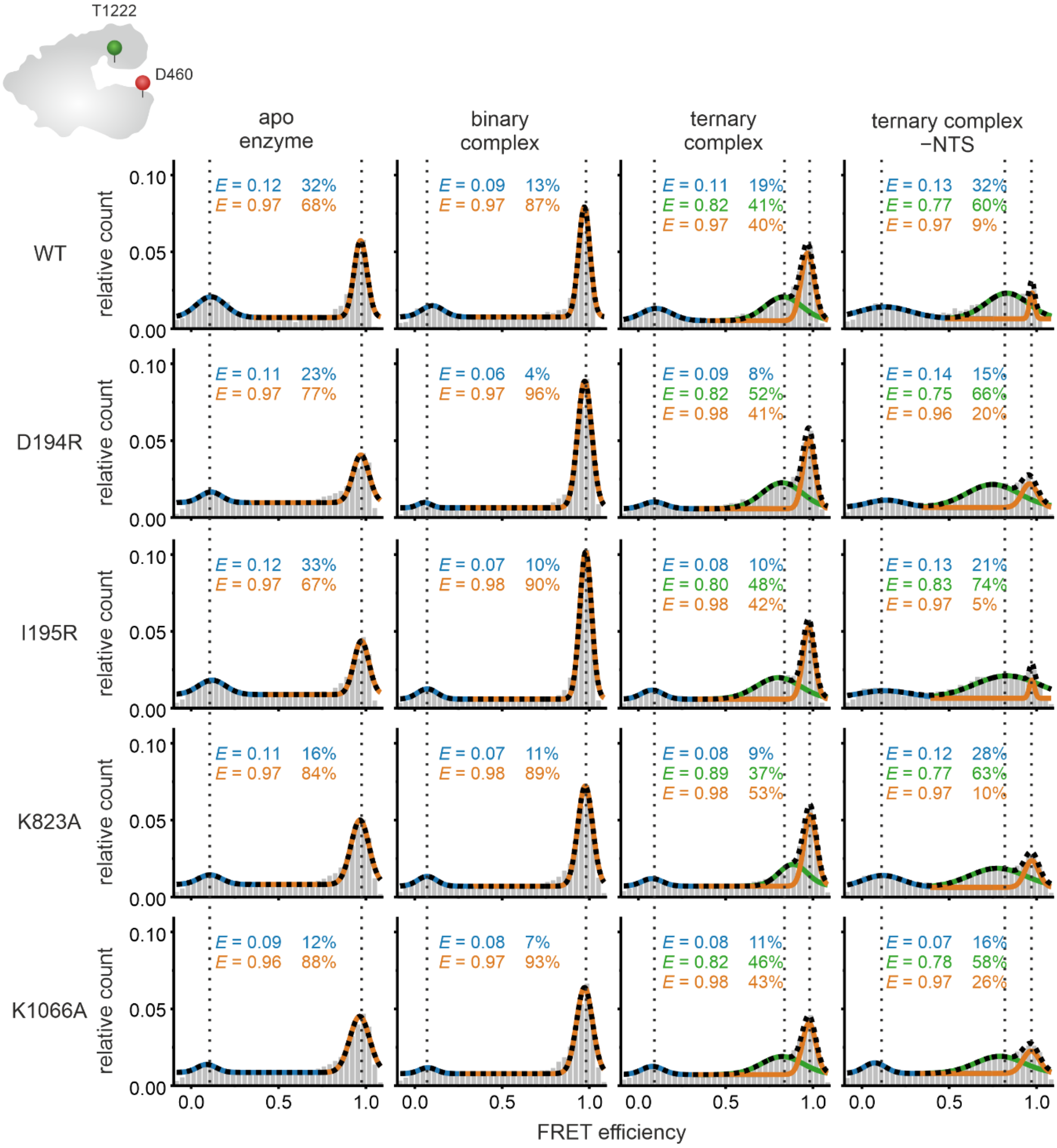


Figure 31: Confocal single-molecule FRET measurements using doubly labeled FnCas12a variants to alter the target strand affinity.

smFRET measurements were conducted with the apo enzymes of FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ variants with altered TS affinity, the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM target DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS). Histograms show the average data of three independent measurements. The histograms were fitted with a double or triple Gaussian function, the overall fit is shown in black (dotted line). The mean FRET efficiencies *E* and the percentage distribution of the populations are given in the histograms. Additional information regarding fit parameters and molecule counts can be found in Appendix 0.

In *cis*- and *trans*-cleavage assays performed by Anthony Newman, none of the variants showed an increase in cleavage activities. On the contrary, most of the variants exhibited a decreased catalytic activity. Hence, these introduced mutations were not further investigated.

5.3 Characterization of FnCas12a bridge helix variants

The bridge helix (BH) is a central helix in Cas12a and a connecting element that bridges the REC and the Nuc lobe. The BH was the main focus of a mutational study summarized in Wörle *et al.* (2021) [290].

In order to shed light on the function and importance of the BH for the activity and conformation of FnCas12a, a series of mutations was introduced via site-directed mutagenesis. FnCas12a I960P was generated to disrupt the α-helical nature of the BH by the introduction of a proline (**Figure 32**). The tryptophan at position 971 in FnCas12a is conserved amongst Cas12a proteins of various species (Appendix II.iii, **Figure 60**). It is accommodated in a hydrophobic pocket generated by the REC2 domain and composed of residues like valine, isoleucine, lysine, leucine, and tyrosine (**Figure 3**, **Figure 32B**) [218,221]. W971 was mutated to a range of amino acids with different chemical properties, differing in size and charge (W971A, W971D, W971K, W971F). By replacing the tryptophan, the impact of the bridge helix anchoring in the REC lobe on the structural integrity and catalytic function was investigated. Additionally, variants with a deleted bridge helix, including or not including the tryptophan 971 in the deletion, were created (ΔBH: ΔY953–K969 → deletion of 17 amino acids; ΔΔBH: ΔY953–W971 → deletion of 19 amino acids).

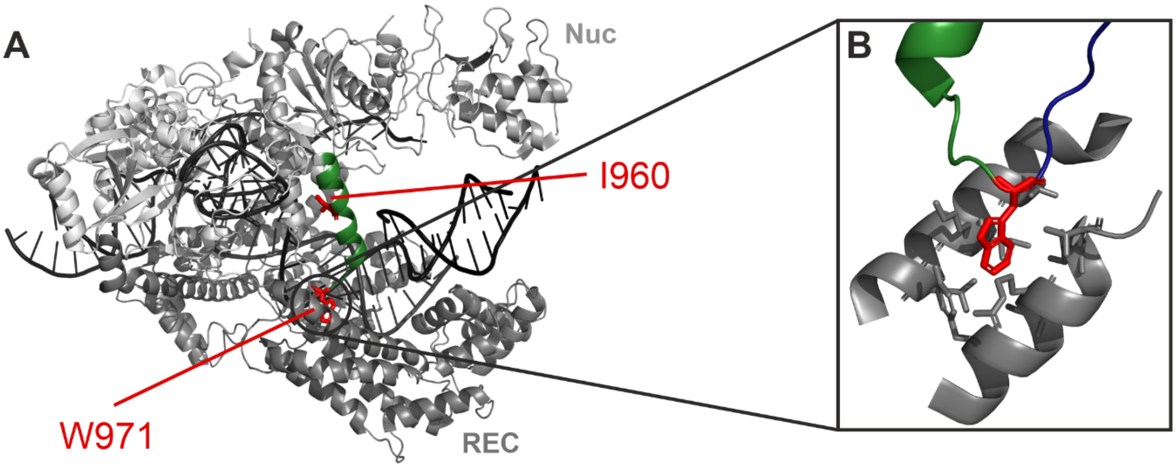


Figure 32: Positions of introduced mutations in the bridge helix of FnCas12a.
(**A**) In the ternary complex structure of FnCas12a (PDB: 6I1K) the BH is colored in green and the positions for the BH mutations (I960 and W971) are highlighted in red. (**B**) Hydrophobic pocket and anchoring point of W971 (red) in the REC2 domain.

5.3.1 Biochemical characterization of FnCas12a bridge helix variants

All FnCas12a BH variants could be cloned and purified to homogeneity (**Figure 33A**). They exhibited comparable thermal stability to the WT enzyme with an average melting temperature of 48.2 °C as assessed by a thermal shift assay (**Figure 33B**).

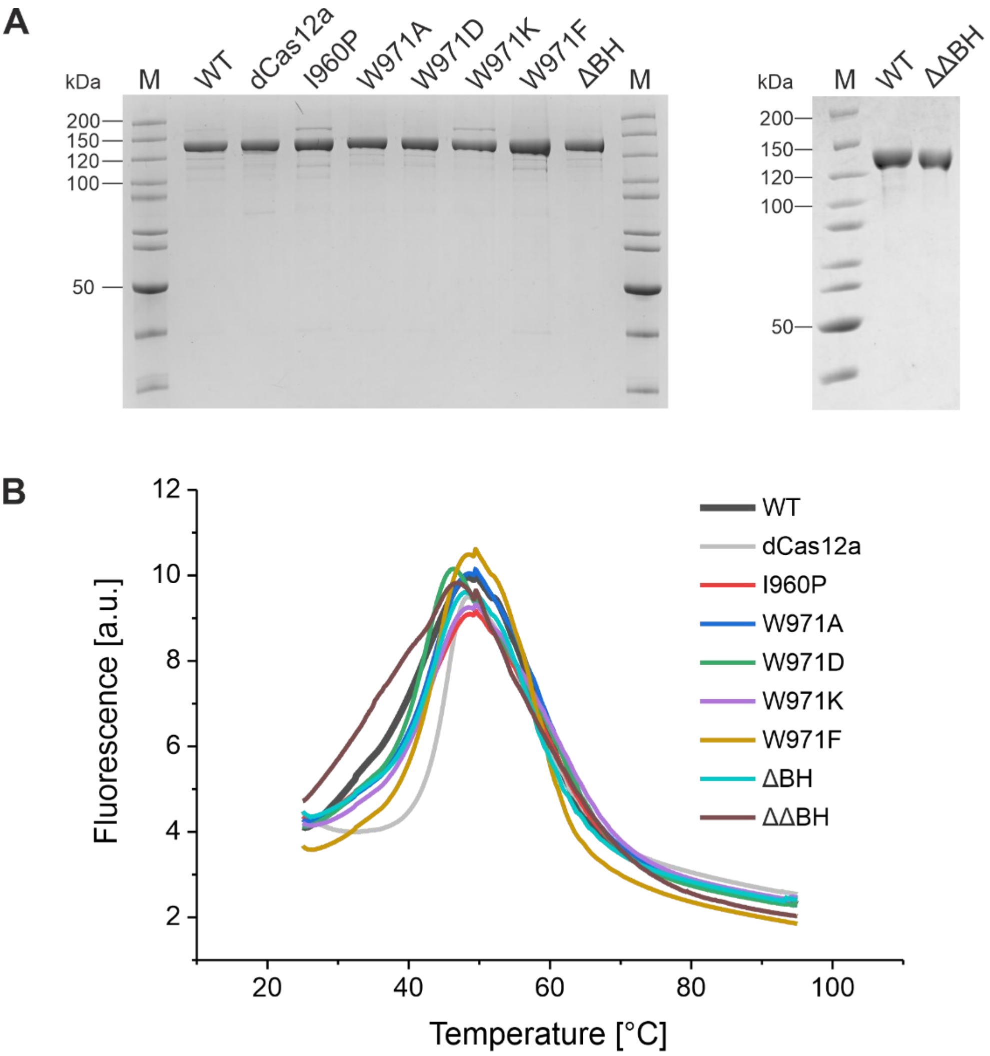


Figure 33: Purification and melting curves of FnCas12a bridge helix variants.
(**A**) SDS-PAGE of purified FnCas12a BH variants (1.7 µg each, molecular weight of WT FnCas12a: 151.9 kDa). M: PageRuler unstained (ThermoFisher Scientific). (**B**) Protein Thermal Shift™ (ThermoFisher Scientific) melting curves of FnCas12a BH variants (2 µg) recorded in a temperature range from 25 to 95 °C, with an average T_m = 48.2 ± 0.7 °C. The average of four replicates is shown.
(adapted from Wörle *et al.* [290])

The activity of the FnCas12a BH variants was analyzed in a plasmid cleavage assay utilizing either plasmid DNA (EGFP-hAgo2, **Table 16**) linearized with SmaI or supercoiled plasmid DNA as target DNA (**Figure 34**). All BH variants were catalytically active but showed clear differences in their cleavage efficiency. All variants produced the two expected fragments of 4400 bp and 2900 bp in length for the linearized plasmid (**Figure 34A**). Only FnCas12a ΔBH additionally produced smeared fragments longer than 4400 bp, resulting in a relatively high total DNA turn-over but a rather low production of effective products. The FnCas12a

ΔΔBH variant was the least active mutant with only 14.5% of target DNA being cleaved as compared to 99.5% cleaved by the WT enzyme after 60 min. FnCas12a I960P was reduced to 50% of the WT activity, whereas the mutation of W971 (W971A, W971D, W971K, W971F) did not or only marginally affect the plasmid DNA cleavage activity.

Utilizing supercoiled plasmid DNA as the target, linearization by FnCas12a could be monitored (**Figure 34C**). WT FnCas12a, as well as FnCas12a W971A and W971F completely linearized the target DNA. Additionally, FnCas12a W971D and W971K almost completely linearized the substrate and FnCas12a I960P only showed a slightly reduced cleavage efficiency compared to the WT enzyme. FnCas12a ΔBH and ΔΔBH showed the least linearization activity. FnCas12a ΔBH again showed an additional product smear above the linearized DNA band, whereas FnCas12a ΔΔBH accumulated nicked DNA.

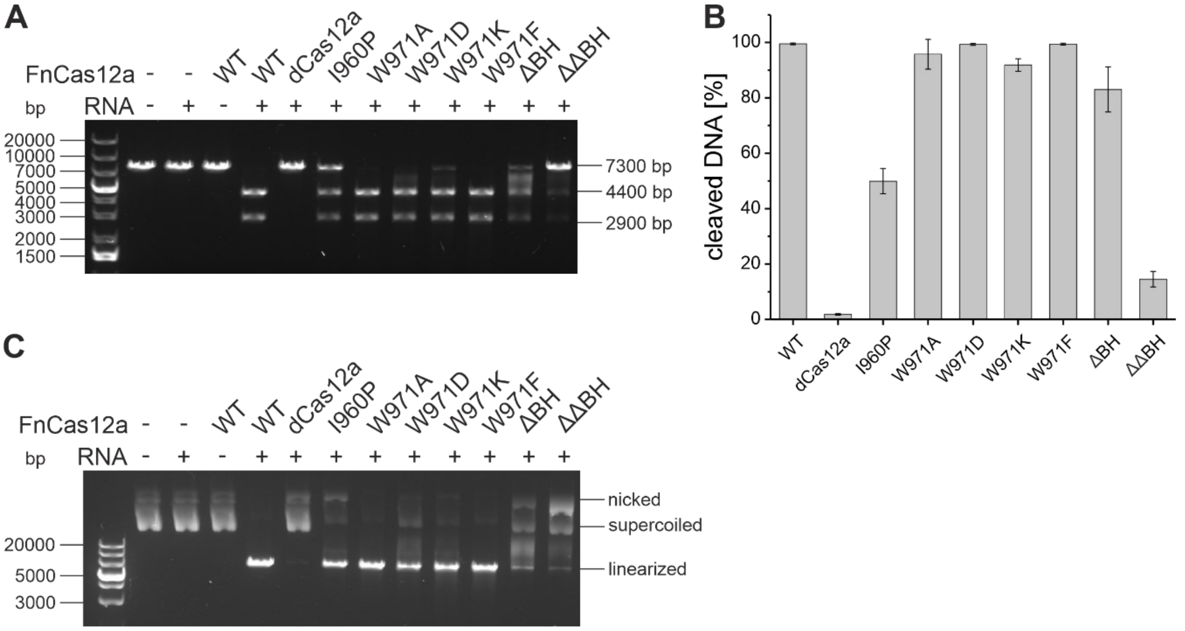


Figure 34: Plasmid cleavage assay of FnCas12a bridge helix variants.
The binary complex (37.5 nM FnCas12a + crRNA) was incubated with 5 nM of target DNA at 37 °C for 1 h. (**A**) The linearized target DNA plasmid (7300 bp) was cleaved into two products (4400 and 2900 bp) by FnCas12a. Products were analyzed on a 0.8% agarose 1x TAE gel. (**B**) The supercoiled plasmid DNA was nicked or linearized by FnCas12a variants. Products were analyzed on a 1.5% agarose 1x TAE gel. 1 kb plus DNA ladder was used as standard. (**C**) Signal quantification of the cleavage efficiency for the FnCas12a BH variants with the linearized plasmid DNA. Shown is the average, error bars show the standard deviation of three independent assays.

The affinity of FnCas12a BH variants to the crRNA was assessed with electrophoretic mobility shift assays using a fluorescently labeled crRNA (KW5, **Table 20**, **Figure 35A, B**), revealing that all variants bind the crRNA similarly to the WT FnCas12a protein. Notably, the RNA-FnCas12a complex stayed in the pockets of the 15% PAA gel. WT FnCas12a has a theoretical isoelectric point of 8.57. In the used TBE buffer system with a pH of 8.0 the protein is almost uncharged, in combination with the high-percentage PAA gel and the large protein, the FnCas12a-crRNA complex was not mobile in the gel and remained in the pockets.

In an EMSA experiment using a labeled pre-crRNA substrate, the pre-crRNA processing activity of the FnCas12a BH variants was investigated (**Figure 35C, D**). The RNase-inactive variant FnCas12a HKK was employed as a negative control. All BH variants showed highly comparable pre-crRNA processing efficiencies to the WT protein.

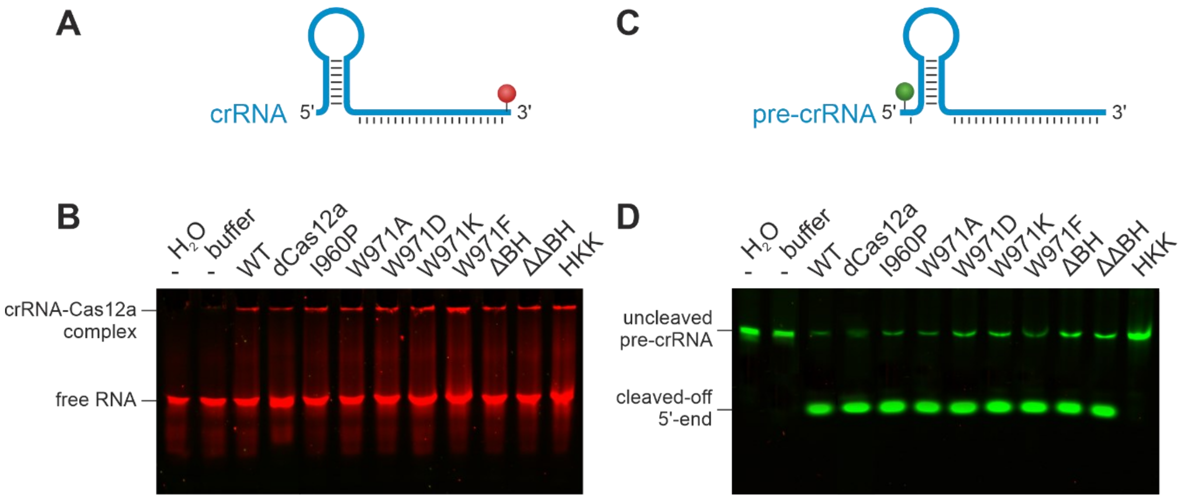


Figure 35: Binding behavior to crRNA and pre-crRNA processing of FnCas12a bridge helix variants.
(**A**) Schematic representation of crRNA labeled with Cy5 at the 3'-end (red sphere). (**B**) EMSA of FnCas12a-crRNA complexes. WT FnCas12a and FnCas12a BH variants were used in a two-fold molar excess of protein (20 nM) over crRNA (10 nM, 43 nt). Free RNA migrated into the gel and was separated according to its size, whereas RNA bound to protein stayed in the gel pockets. (**C**) Schematic representation of pre-crRNA with an additional nucleotide at the 5'-end (44 nt) labeled with Cy3 at the 5'-end (green sphere). (**D**) Investigation of pre-crRNA processing by FnCas12a BH variants. WT FnCas12a and FnCas12a BH variants were used in a two-fold molar excess of protein (20 nM) over crRNA (10 nM). Pre-crRNA processing could be followed by the cleavage of the single labeled 5'-nucleotide.

To verify the binding of FnCas12a to the crRNA, mass photometry measurements were carried out (**Figure 36**). WT FnCas12a showed a mass of 155 kDa (95% of the counts, blue) which is congruent with its actual molecular weight of 151.9 kDa. 5% of the counts were detected at a mass of 324 kDa, which would match a dimeric Cas12a protein. Initially, it was assumed that Cas12a appears as a dimer [214], but subsequent studies showed that it acts in a monomeric fashion [132,217,218]. The mass photometry data showed that a small fraction of the protein indeed dimerizes, but the majority of molecules remain monomeric. The addition of a 43 nt crRNA (approximately 15 kDa) led to a shift of 21 kDa (green), to a detected molecular weight of 176 kDa, showing effective binding of the crRNA to the protein and the formation of the binary complex. The mass photometry data additionally showed that even though the crRNA-FnCas12a complex in the EMSA experiments remained in the gel pockets (**Figure 35B**), the signal did not originate from aggregate formation.

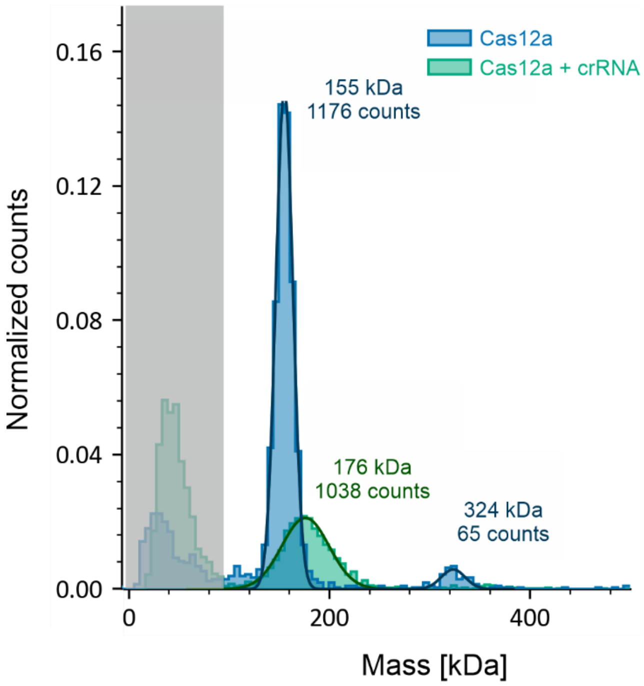


Figure 36: Mass photometry of FnCas12a WT and FnCas12a WT with crRNA.
Mass histogram showing the mass photometry data of WT FnCas12a with a mass of 155 kDa (blue, expected molecular weight of 151.9 kDa). The addition of crRNA (43 nt, ≈ 15 kDa) led to a shift to higher masses (176 kDa). The data underneath the greyed area originates from background molecules in the buffer solution.

The binding behavior of FnCas12a BH variants to the target DNA was assayed in EMSAs using the short doubly labeled target DNA substrates (KW5, **Table 20**, **Figure 37A–D**). These assays revealed that all BH variants bind the DNA efficiently. Consequently, reduced cleavage activity observed for some variants is exclusively caused by a catalytic defect of the variants, not by an impairment of binding to the target DNA.

The cleavage activity of FnCas12a variants was further analyzed on the short target DNA (**Figure 37E, F**). All variants showed cleavage activity with the FnCas12a ΔBH being the least active and only showing faint bands for the cleaved products. Because of the low cleavage activity of FnCas12a ΔBH, the shifted band of the FnCas12a-crRNA-DNA complex in the EMSA (**Figure 37C, D**) is differently colored than the bands of the completely cleavage-competent variants, as both fluorophores remained part of the complex in this case.

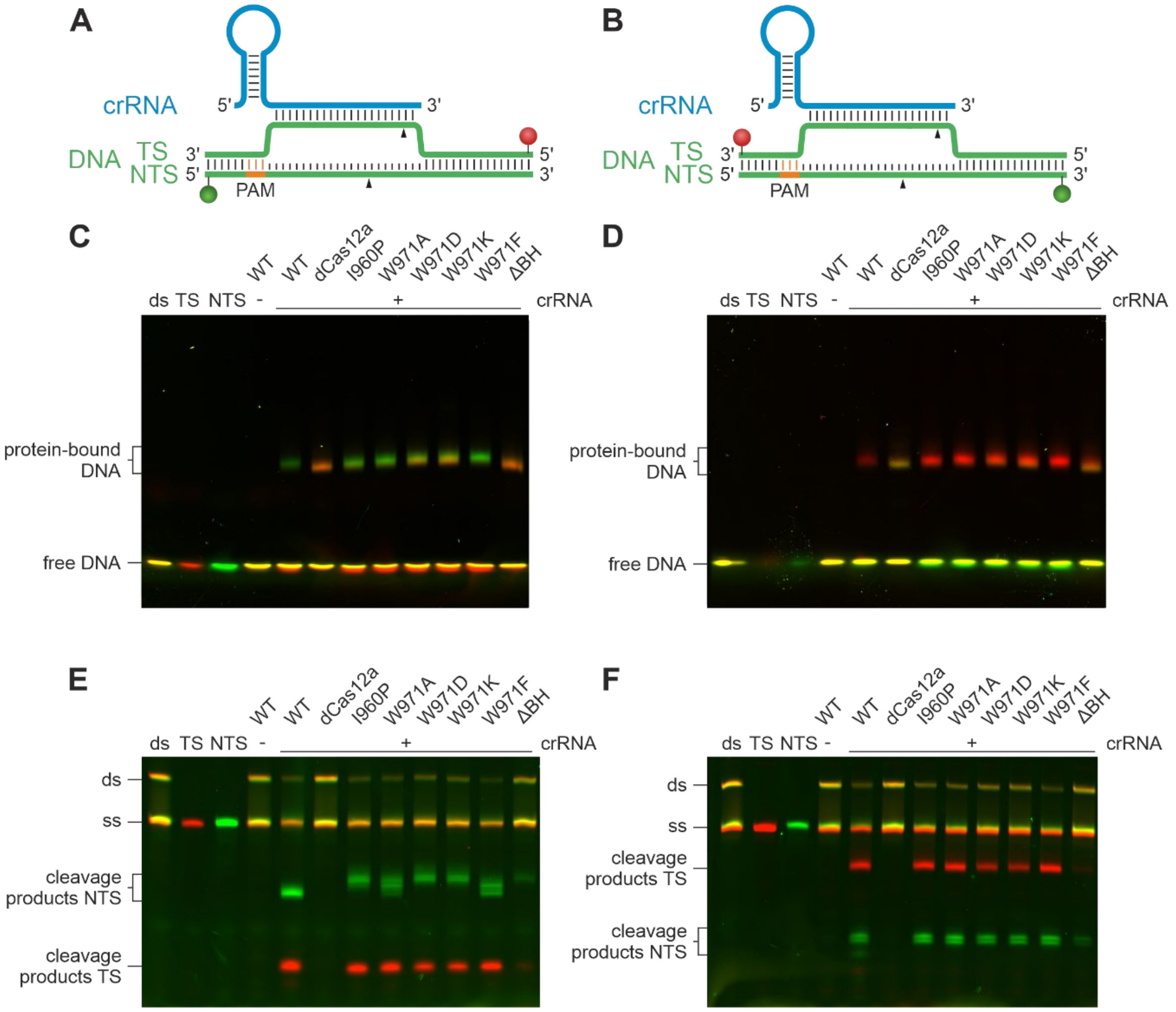


Figure 37: Activity assays of FnCas12a bridge helix variants using short double-stranded DNA as target.
WT FnCas12a and FnCas12a BH variants were used in a 7.5-fold molar excess of protein and crRNA (75 nM) over target DNA (10 nM). Reactions were incubated for 1 h at 37 °C. The short target DNA (58 nt) was doubly labeled with a Cy3 label (green) at the non-target strand (NTS) and a Cy5 label (red) at the target strand (TS). (**A**) Cy3 (green sphere) and Cy5 (red sphere) labels were coupled to the 5'-end of the DNA strands, respectively. (**B**) Cy3 (green sphere) and Cy5 (red sphere) labels were attached to the 3'-end of the DNA strands, respectively. (**C**) EMSA of FnCas12a BH variants with 5'-labeled DNA. (**D**) EMSA of FnCas12a BH variants with 3'-labeled DNA. (**E**) Cleavage assay of FnCas12a BH variants with 5'-labeled DNA. (**F**) Cleavage assay of FnCas12a BH variants with 3'-labeled DNA.

In cleavage assays with 5'-labeled DNA constructs (**Figure 37E**), the cleavage product bands for the NTS seemed to run at different heights. In consequence, to analyze the cleavage products with a single-nucleotide resolution, the cleavage reactions were resolved on a denaturing sequencing gel (**Figure 38**). These analyses revealed that the cleavage pattern of the FnCas12a BH variants for the NTS indeed differed from that of the WT enzyme (**Figure 38B**). For WT FnCas12a a predominant cleavage product was visible and two minor products, one and two nucleotides longer, were detected. This is a result of the trimming activity of Cas12a (**Figure 18E**). In contrast, the BH variants showed mainly longer cleavage products, being up to 5 nt longer than the main product of the WT. Only the W971A and W971F variants generated a product as short as the WT via the trimming activity. Noteworthy, this is not the main product of these variants, showing that in the mutated enzymes, the NTS trimming activity is less efficient. Even after prolonged reaction times (up to 5 h, **Figure 38E**), full trimming to the WT product could not be completed. In contrast, all other variants never yielded the final cleavage product of the WT enzyme. This suggests that the variants FnCas12a I960P, W971D, W971K, and ΔBH lack the trimming activity altogether. The cleavage pattern for the TS was the same for all FnCas12a BH variants and showed one main cleavage product and two additional products as previously detected for WT FnCas12a (**Figure 18**).

The high-resolution cleavage assay with target DNA constructs carrying the fluorescent labels at the 3'-end (**Figure 38D**) revealed that the PAM-proximal part of the TS, that remains bound to the protein, showed the same cleavage pattern for all FnCas12a BH variants. The 3'-labeled TS was processed yielding three products as already observed for the 5'-labeled DNA. For the NTS, two initial cleavage sites were detected for all FnCas12a BH variants.

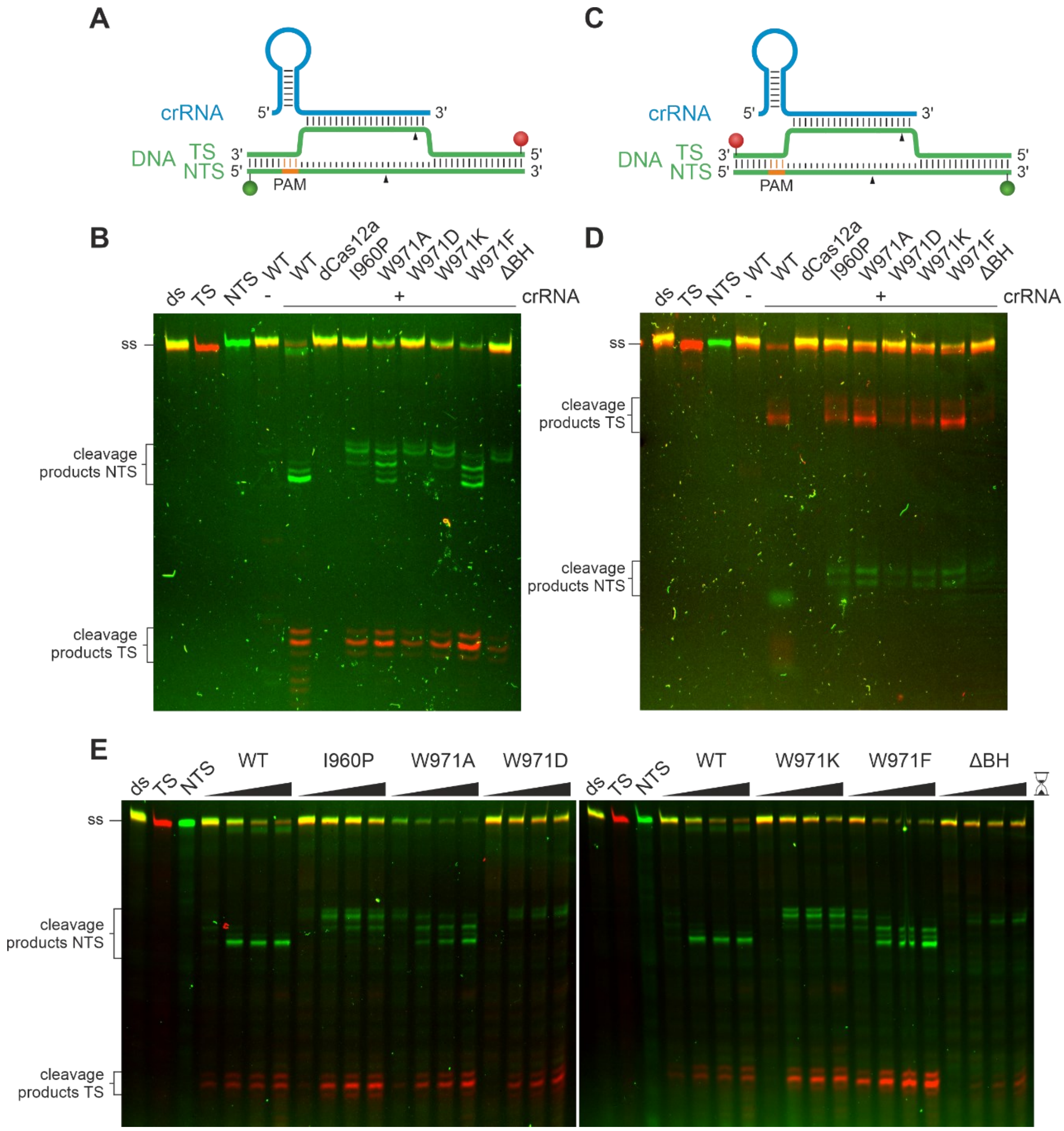


Figure 38: High-resolution target DNA cleavage assays of FnCas12a bridge helix variants. Cleavage assay of FnCas12a BH variants using a 7.5-fold molar excess of protein and crRNA (375 nM) relative to the target DNA (50 nM). The reaction was incubated at 37 °C for 1 h. The short target DNA (58 nt) was doubly labeled with a Cy3 label (green) at the non-target strand (NTS) and a Cy5 label (red) at the target strand (TS). (**A**) Cy3 (green sphere) and Cy5 (red sphere) labels were attached to the 5'-end of the DNA strands. (**B**) The cleaved NTS showed a different cleavage pattern for the various FnCas12a BH variants. The TS cleavage pattern was the same for all FnCas12a BH variants. (**C**) Cy3 (green sphere) and Cy5 (red sphere) labels were positioned at the 3'-end of the DNA strands. (**D**) The cleavage patterns of target and non-target strands were the same for all FnCas12a BH variants. (**E**) Cleavage kinetics of FnCas12a BH variants using a short 5' doubly labeled DNA target (**A**). The reactions contained FnCas12a and crRNA (375 nM) in 7.5-fold molar excess over target DNA (58 nt; 50 nM). Reactions were stopped with 125 mM EDTA after 1 min, 1 h, 2.5 h, and 5 h at 37 °C. Samples were analyzed on a high-resolution 15% denaturing PAA gel. (adapted from Wörle *et al.* [290])

It has been reported that BH variants of Cas9 from *Streptococcus pyogenes* are altered in their sensitivity to mismatches between crRNA and target DNA [292,293]. In this work, the influence of mismatches on the cleavage specificity of WT FnCas12a and the BH variants was tested in an additional set of cleavage assays. To this end, substrates carrying a single mismatch at position 2 (MM 2, mismatch in the seed region), 10 (MM 10), or 20 (MM 20) of the R-loop (**Figure 39A, Table 20**) were tested. WT FnCas12a showed moderate sensitivity to substrates with a mismatch at position 2 or 10, but the MM 20 substrate was cleaved to the same extent as the fully complementary substrate (MM 0). A mismatch at position 2 decreased the cleavage efficiency for the wildtype and to a similar extent in all BH variants (**Figure 39B, C**). FnCas12a W971F is an exception as this variant was significantly less active when using the MM 2 substrate as compared to the wildtype. Notably, FnCas12a W971F was equally active as the WT when using a fully complementary substrate. All BH variants showed pronounced sensitivity to a mismatch at position 10. Additionally, the BH variants FnCas12a W971D and W971K exhibited major discrimination against a substrate with a mismatch at position 20. For the FnCas12a BH variants, the relative reduction in cleavage efficiencies between fully complementary substrates and mismatched substrates was more pronounced than for WT FnCas12a. In the case of FnCas12a W971D, the cleavage efficiency was reduced to 50% of the WT level for the fully complementary target (MM 0), for MM 10 only 3.6% of the wildtype cleavage efficiency was reached. The BH deletion variants did not show enough activity for quantification. Taken together, these data show that the target selectivity of FnCas12a is modulated by the BH.

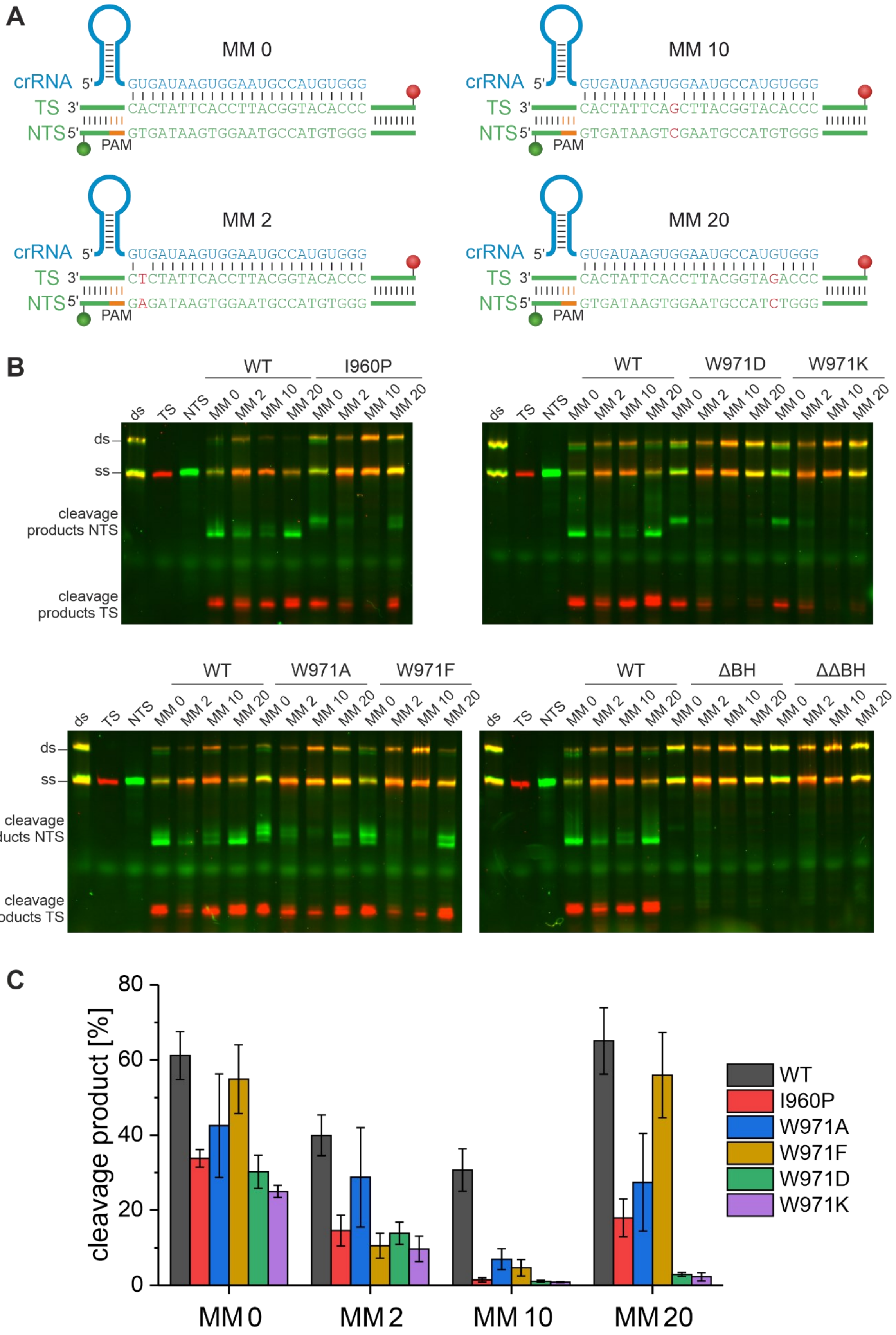


Figure 39: DNA cleavage specificity of FnCas12a bridge helix variants at mismatched substrates. (**A**) The short target DNA (58 nt) was doubly labeled with a Cy3 label (green sphere) at the 5'-end of the non-target strand (NTS) and a Cy5 label (red sphere) at the 5'-end of the target strand (TS). In addition to the fully complementary substrate (MM 0), substrates that carry a single mismatch at position 2 (MM 2), 10 (MM 10), or 20 (MM 20) were used. (**B**) Cleavage assays of FnCas12a BH variants using a 7.5-fold molar excess of protein and crRNA (75 nM) relative to the target DNA (10 nM). The reactions were

incubated at 37 °C for 1 h. (**C**) Quantification of the fluorescence signal for the cleavage efficiency of the NTS for the WT and FnCas12a BH variants using fully complementary and mismatched substrates. Shown is the average, error bars show the standard deviation of three independent assays. While a substrate mismatched at position 10 was still cleaved by WT FnCas12a, the BH variants did not cleave this substrate. BH variants W971D and W971K additionally did not cleave a substrate with a mismatch at position 20. (adapted from Wörle *et al.* [290])

5.3.2 Confocal single-molecule FRET analysis of FnCas12a bridge helix variants

For a better understanding of the function of the bridge helix and its involvement in the stabilization of the conformational states of Cas12a, the FnCas12a BH variants were analyzed employing single-molecule FRET measurements using a confocal setup. Donor and acceptor fluorophores were engineered at positions D470 and T1222. For all FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants, low, medium, and high FRET populations were detected (**Figure 40**). However, differences in the relative distribution of these states were observed. In the apo state, all FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ W971 variants showed higher populated low FRET populations. This suggests that the destabilization or altered anchoring of the BH in the REC lobe increases the probability to adopt the open state. In the case of the BH deletion variants FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ ΔBH and ΔΔBH, the distribution and FRET efficiencies of the respective populations did not change, and high and low FRET states were almost equally populated.

Upon loading of the crRNA into the enzyme, all variants adopted the closed conformation. However, titration experiments with increasing concentrations of crRNA revealed that for the variants FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ I960P, W971A, W971D, and W971K the formation of the closed state (high FRET population) was less efficient (Appendix 0, **Figure 70**). The variants FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ W971F and ΔBH, by contrast, readily formed the closed conformation even at low crRNA concentrations (Appendix 0, **Figure 70**). Taken together, for variants that preferentially adopted the open state in the apo form, the loading of crRNA was inefficient, resulting in reduced amounts of the binary complex (**Figure 41A**). This suggests that for some FnCas12a BH variants the binding of crRNA or the conformational transition from the open to the closed state is impaired.

Target DNA addition efficiently led to the formation of the ternary complex in all FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants (**Figure 40** and **Figure 70**, Appendix 0). Hence, the implemented BH mutations did not affect target DNA loading. Nevertheless, all variants displayed a relatively broadly distributed medium FRET population (full width half maximum value (FWHM) higher than for WT FnCas12a, **Figure 41B**), suggesting that this conformation is more flexible than the high FRET conformation.

Loading of FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants with the TS was very effective and resulted in an almost complete loss of the high FRET population, implying that most of the proteins loaded with crRNA also bind to the TS. This was especially pronounced in the

FnCas12a$^{REC\text{-}Nuc^{*}DL550/DL650}$ ΔBH variant, indicating that also the unloading of the protein is less efficient without the bridge helix.

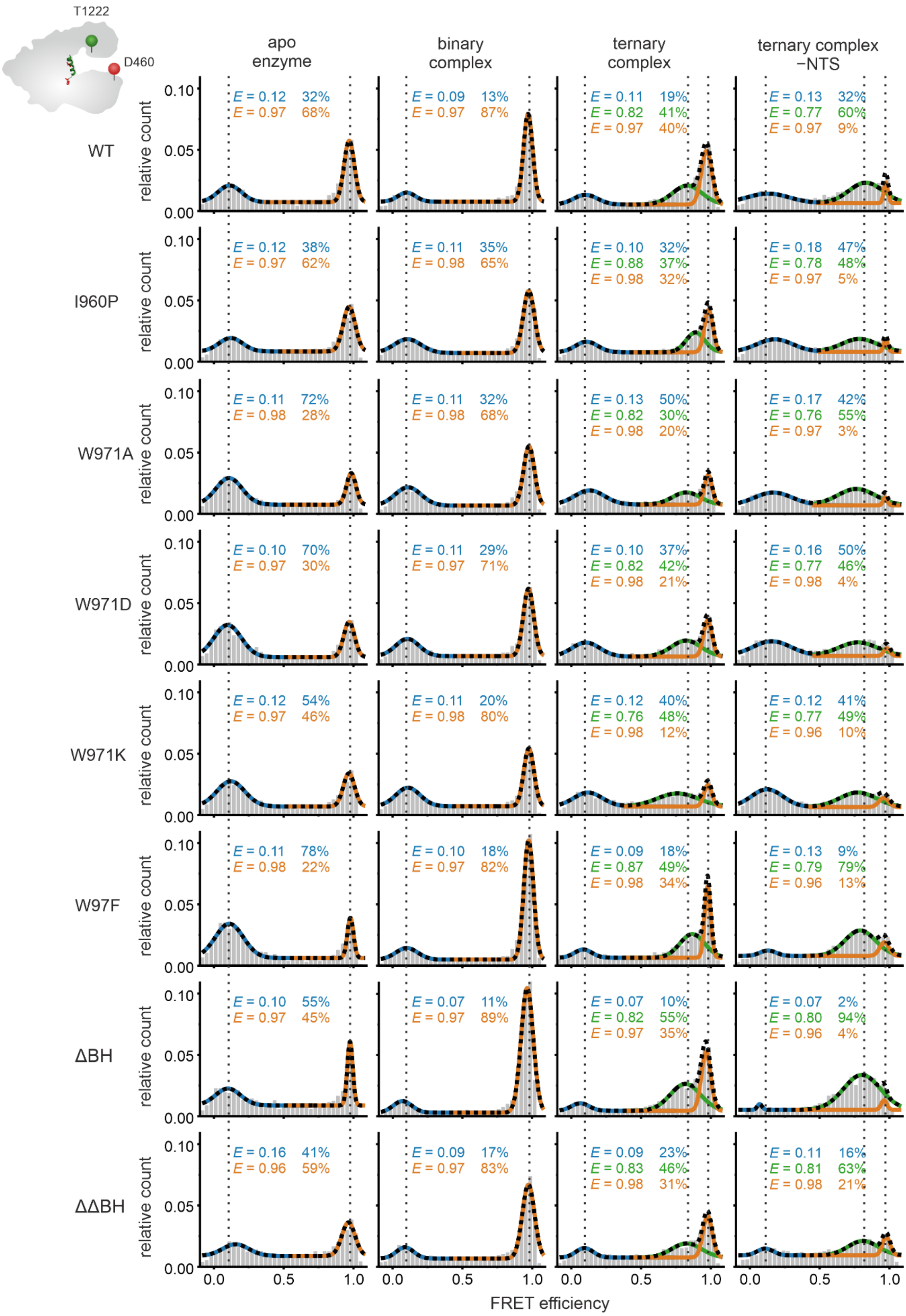


Figure 40: Confocal single-molecule FRET measurements using doubly labeled FnCas12a bridge helix variants.
smFRET measurements were conducted with the apo enzymes of FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants, the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM target DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS). Histograms show the average data of three independent measurements. The histograms were fitted with a double or triple Gaussian function, the overall fit is shown in black (dotted line). The mean FRET efficiencies E and the percentage distribution of the populations are given in the histograms. Additional information regarding fit parameters and molecule counts can be found in Appendix 0. (adapted from Wörle *et al.* [290])

Taken together, the smFRET data indicate that part of the FnCas12a BH variants are not only impaired in their cleavage activity but that perturbing the helical structure of the bridge helix by the introduction of a proline residue or loosening the anchoring of the BH by replacing the tryptophan with charged amino acids also disturb efficient binary complex formation. While a shift towards the low FRET population was observed for the W971 variants in the apo state and BH variants do not adopt the high FRET conformation as efficiently as the wildtype enzyme, the defined structural states represented by the low and high FRET population can still be adopted by the BH deletion variants. Therefore, the BH is presumably not the main structural element that defines the open and closed state of Cas12a.

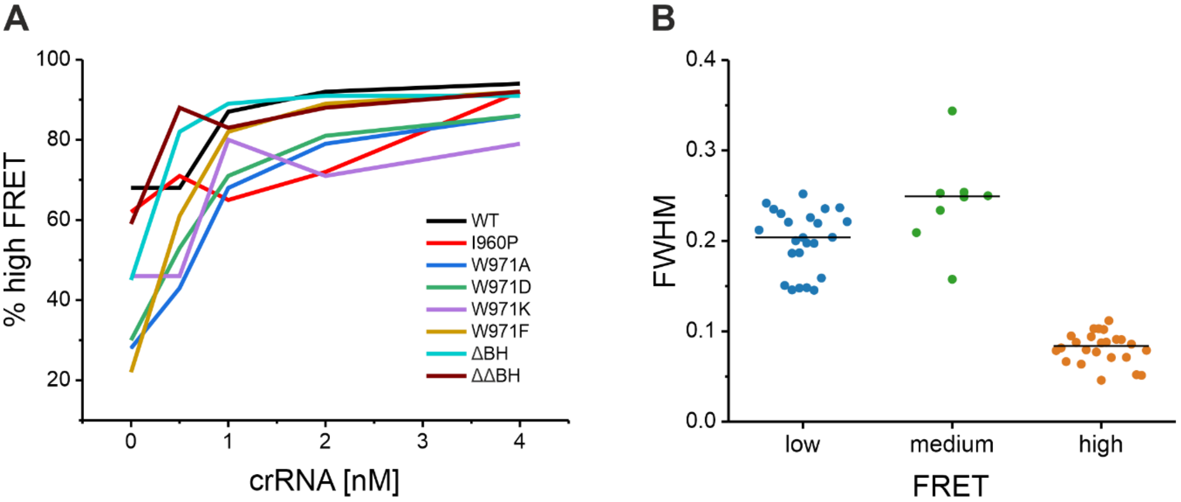


Figure 41: Quantification of single-molecule FRET measurements of FnCas12a bridge helix variants.
(**A**) Quantification of the relative number of molecules in the high FRET population in the binary complex of FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants compared to WT FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ as a function of crRNA concentration. The relative amount of the high FRET population increases upon the addition of increasing concentrations of crRNA (0 nM, 0.5 nM, 1 nM, 2 nM, 4 nM, FRET efficiency histograms in **Figure 70**, Appendix 0). (**B**) Quantification of FWHM values of WT FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ and FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants. Quantified are the fits displayed in **Figure 40** for the low, medium, and high FRET populations of the apo, binary, and ternary complex.
(adapted from Wörle *et al.* [290])

5.4 Characterization of FnCas12a helix 1 variants

The biochemical and biophysical characterization of FnCas12a variants that carry mutations in or deletions of the BH revealed that the BH is indeed an important structural element for the trimming activity and cleavage efficiency, but not for the structural integrity of Cas12a. Investigation of the available structures revealed that the BH is connected to the adjacent helix 1 which is the focus of this chapter. Results presented in this chapter are summarized in Wörle *et al.* (2022) [294].

Structural studies show that the BH and helix 1 undergo a tandem movement upon transition from the binary to the ternary state [133,217–221,226]. During this transition, helix 1 is shortened, whereas the BH is elongated (**Figure 3E, F**). To gain more insights into the dependency of BH and helix 1 and the role and importance of helix 1 on the stability and activity of FnCas12a, a mutational study on this helix was performed and the resulting FnCas12a helix 1 variants were analyzed biochemically and in smFRET measurements (**Figure 42**). To this end, a proline was introduced into helix 1 to disrupt the α-helical structure of this structural element. Here, the single proline variant FnCas12a K981P, as well as the double proline variant FnCas12a I960P/K981P that additionally carries a proline substitution in the BH were generated. Structural predictions of this variant using the AlphaFold2 algorithm [268] show the kinks introduced by the prolines into the helices (**Figure 42B**). The positively charged and conserved lysine at position 978 interacts with the backbone of the DNA TS [221] (Appendix II.iii). To disrupt this interaction, the mutation K978A was introduced. Moreover, deletion variants that result in FnCas12a variants that lack helix 1 or both helices (BH and helix 1) were generated (Δh1: ΔK972–N1000 → deletion of 29 amino acids; ΔBH/Δh1: ΔY953–N1000 → deletion of 48 amino acids).

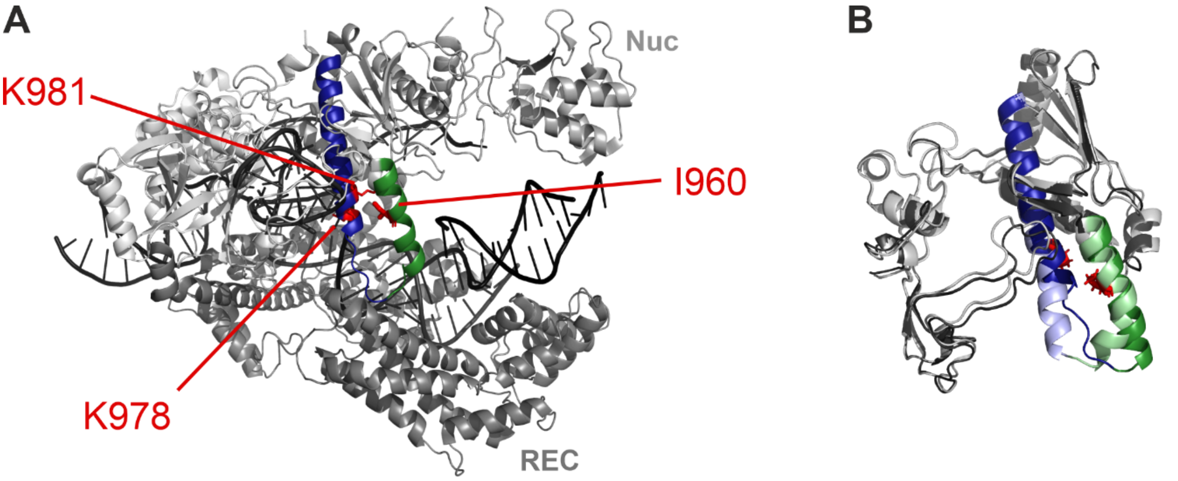


Figure 42: Positions of mutations introduced into helix 1 of FnCas12a.
(**A**) Ternary complex structure of FnCas12a (PDB: 6I1K) with helix 1 colored in blue, BH in green, and positions of helix 1 mutations (K981 and K987 plus I960 in the BH) highlighted in red. (**B**) Structural prediction generated with AlphaFold2 [268] of the 300 aa surrounding helix 1 and the BH of FnCas12a I960P/K981P (brighter colors) in comparison to the prediction of WT FnCas12a (dark color) showing the kinks introduced by the two proline residues. The mutations I960P and K981P are colored in red.

5.4.1 Biochemical characterization of FnCas12a helix 1 variants

All FnCas12a helix 1 variants were cloned successfully. The variants FnCas12a K978A, FnCas12a K981P, and FnCas12a I960P/K981P were purified to homogeneity and exhibited comparable thermal stability with an average melting temperature of 47.9 ± 0.2 °C as assessed by a thermal shift assay (**Figure 43A, B**). The deletion variants FnCas12a Δh1 and FnCas12a ΔBH/Δh1 could be purified, but due to altered elution behavior during the heparin affinity chromatography, they could not be purified to full homogeneity (**Figure 43A**). Protein thermal shift assays revealed an additional thermal melting point at 35.9 °C for the variants FnCas12a Δh1 and ΔBH/Δh1 (**Figure 43B**). This probably results from the byproducts during purification but could also indicate a partial destabilization of the protein caused by the deletion of helix 1.

Circular dichroism experiments revealed that all helix 1 variants exhibit two minima at 208 nm and 222 nm as well as a maximum at 192–198 nm in the far-UV spectrum, typically for proteins with a predominantly α-helical structure (**Figure 43C**). This corresponds well to the structure of FnCas12a which is almost exclusively composed of alpha helices. The FnCas12a helix 1 variants appear not to deviate from this overall structural organization.

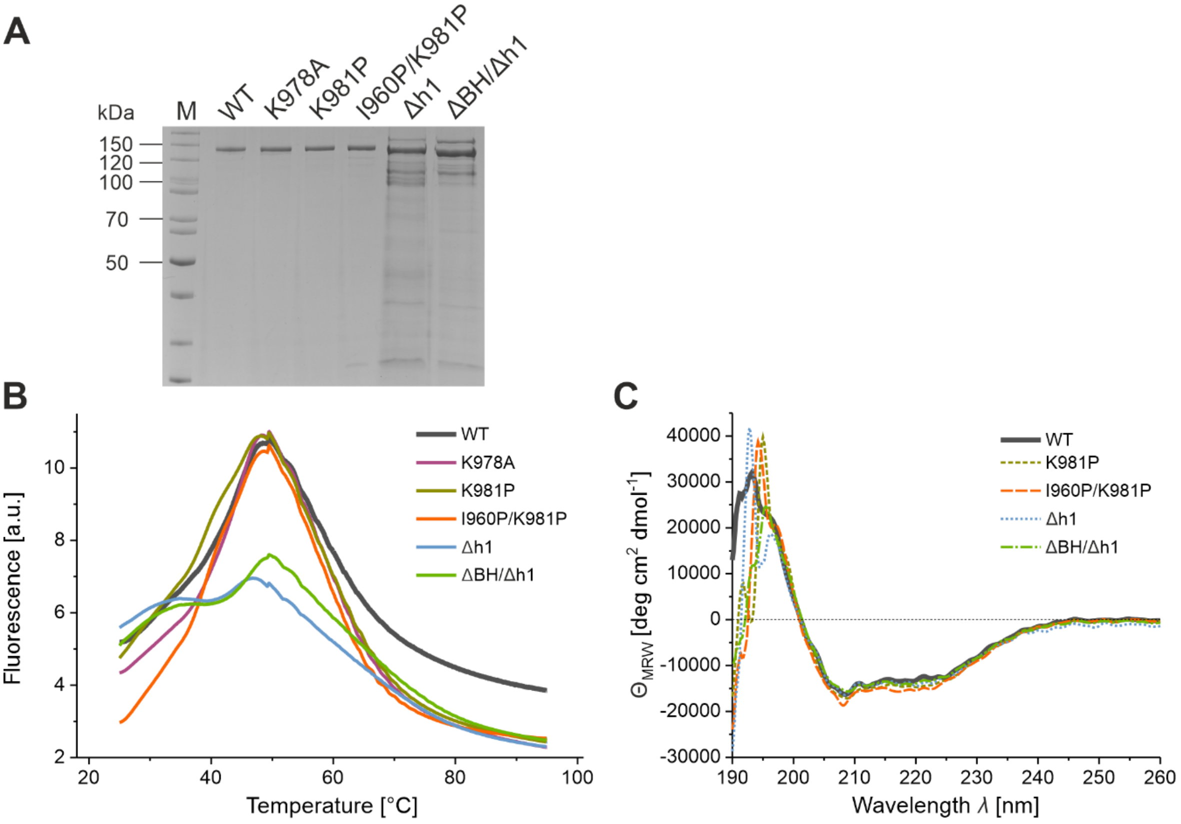


Figure 43: Purification, melting curves, and CD spectra of FnCas12a helix 1 variants.
(**A**) SDS-PAGE of purified FnCas12a helix 1 variants (1.7 µg each, molecular weight of WT FnCas12a: 151.9 kDa). M: PageRuler unstained (ThermoFisher Scientific). (**B**) Protein Thermal Shift™ (ThermoFisher Scientific) melting curves of FnCas12a helix 1 variants (2 µg) recorded in a temperature range from 25 to 95 °C. The average of four replicates is displayed. (**C**) Far-UV CD spectra of WT FnCas12a and FnCas12a helix 1 variants (2 µM) at 25 °C. The average of five replicates is displayed.
(adapted from Wörle *et al.* [294])

The activity of the FnCas12a helix 1 variants was analyzed in a plasmid cleavage assay utilizing either plasmid DNA (EGFP-hAgo2, **Table 16**) linearized with SmaI or the supercoiled plasmid as target DNA (**Figure 44**). For direct comparison, relevant BH variants (FnCas12a I960P, ΔBH, and ΔΔBH) were analyzed on the same gels. Using the linearized plasmid DNA, FnCas12a K978A displays the same cleavage efficiency of 100% as WT FnCas12a, whereas the other FnCas12a helix 1 variants were less efficient (**Figure 44A, B**). FnCas12a K981P, carrying the proline in helix 1, cleaved 83% of the linearized target DNA but similarly to the previously analyzed FnCas12a ΔBH variant it additionally produced smeared fragments above the 4400 bp fragment, resulting in a higher percentage of cleaved DNA (reduction of target DNA) than effective product formation. The previously analyzed FnCas12a I960P, with the introduced proline in the BH, cleaved 50% of the target DNA. Product formation of both proline-substituted variants was similar to each other. Combining both proline mutations in the FnCas12a I960P/K981P variant resulted in a complete loss of cleavage activity using the linearized plasmid DNA. The same was observed for the helix 1 deletion variants FnCas12a Δh1 and ΔBH/Δh1 – both did not cleave the linearized plasmid DNA.

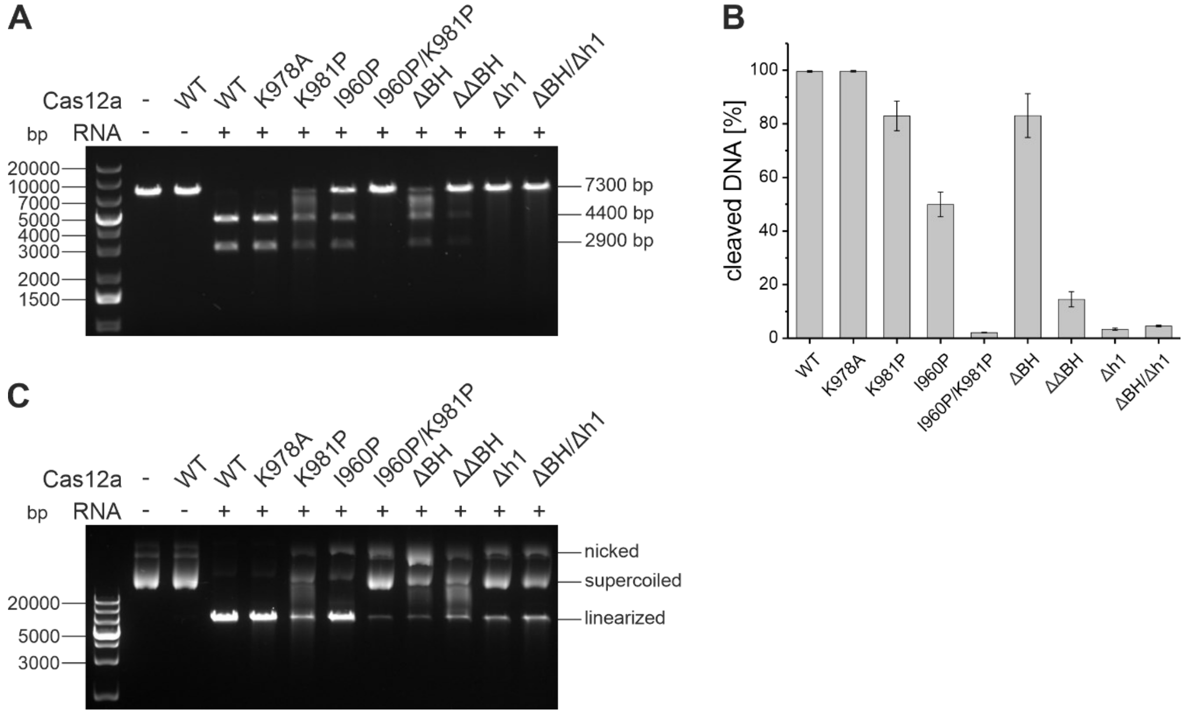


Figure 44: Plasmid cleavage assay of FnCas12a helix 1 variants.
The binary complex (37.5 nM FnCas12a + crRNA) was incubated with 5 nM of target DNA at 37 °C for 1 h. (**A**) The linearized target DNA plasmid (7300 bp) was cleaved into two products (4400 and 2900 bp) and analyzed on a 0.8% agarose 1x TAE gel. (**B**) Signal quantification of the cleavage efficiency for the FnCas12a helix 1 variants with the linearized plasmid DNA. Shown is the average of three independent assays, error bars show the standard deviation. (**C**) The supercoiled plasmid DNA is nicked or linearized by FnCas12a variants and analyzed on a 1.5% agarose 1x TAE gel. 1 kb plus DNA ladder was used as standard. (adapted from Wörle *et al.* [294])

Analysis of FnCas12a-mediated cleavage on supercoiled plasmid DNA (**Figure 44C**) revealed that some cleavage can also be observed for the variants FnCas12a I960P/K981P, Δh1, and ΔBH/Δh1. As observed in the cleavage assay with linearized target DNA, FnCas12a K978A showed the same cleavage efficiency as the WT enzyme, and FnCas12a K981P, ΔBH, and ΔΔBH had a reduced cleavage efficiency but produced a smear between the supercoiled and the linearized cleavage product, indicating that these variants cleave the plasmid DNA in an imprecise fashion.

In time-resolved plasmid cleavage assays (**Figure 45**), the cleavage behavior of selected variants was further analyzed. Already after a few seconds, WT FnCas12a completely cleaved the target DNA. For both DNA substrates, linearized and supercoiled plasmid, FnCas12a K981P was faster in the initial cleavage than FnCas12a ΔBH, but FnCas12a ΔBH showed more effective cleavage after approximately 40–50 min. The variants FnCas12a ΔΔBH, I960P/K981P, and Δh1 were only analyzed on the supercoiled plasmid DNA as they did not show any cleavage when using the linearized plasmid as target DNA. These variants only showed slow linearization of the supercoiled plasmid DNA. Interestingly, in the time-resolved assays neither FnCas12a ΔBH nor FnCas12a K981P showed smeared product bands.

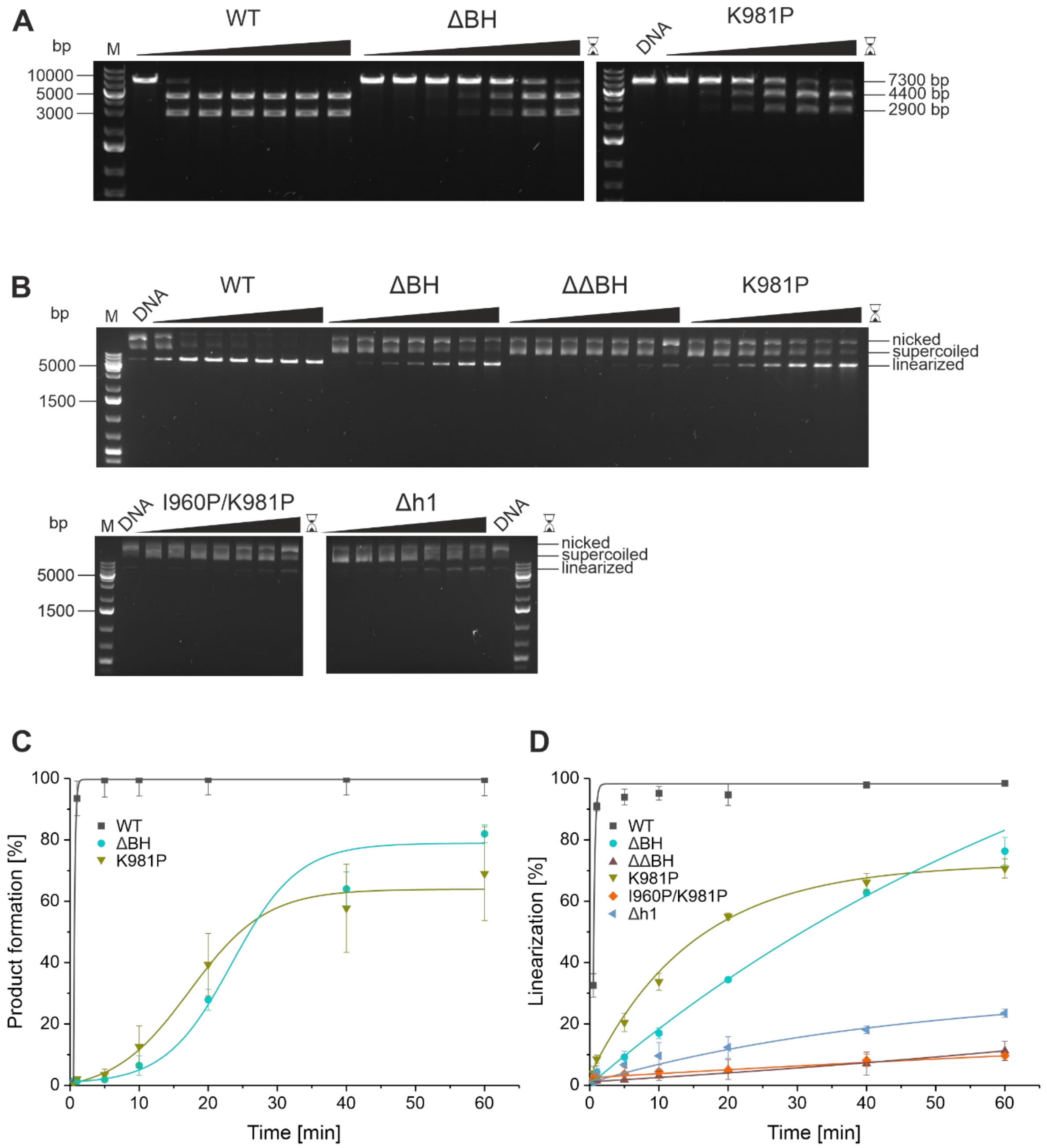


Figure 45: Time-resolved cleavage assay of selected FnCas12a bridge helix and helix 1 variants. The binary complex (37.5 nM FnCas12a + crRNA) was incubated with 5 nM of target DNA at 37 °C. The reaction was stopped with EDTA and Proteinase K after 30 s, 1 min, 5 min, 10 min, 20 min, 40 min, or 1 h. (**A**) Time-resolved cleavage activity of WT FnCas12a, FnCas12a ΔBH, and FnCas12a K981P on linearized plasmid DNA. (**B**) Time-resolved cleavage activity of WT FnCas12a, FnCas12a ΔBH, FnCas12a ΔΔBH, FnCas12a K981P, FnCas12a I960P/K981P, and FnCas12a Δh1 using supercoiled plasmid DNA as target. (**C**) Signal quantification of three replicates of cleavage assays using linearized target DNA is shown in (**A**). (**D**) Signal quantification of three replicates of cleavage assays using supercoiled target DNA as target is shown in (**B**). Error bars show the standard deviation of three independent assays. The data points were fitted with an exponential equation (ExpDec1) using the origin software. The data points for FnCas12a variants ΔBH and K981P using linearized target DNA (**C**) were fitted with a Boltzmann equation.

In electrophoretic mobility shift assays using a fluorescently labeled crRNA (KW5, **Table 20**), the binding efficiency of the FnCas12a helix 1 variants to crRNA was assessed (**Figure 46A–C**). FnCas12a K978A bound crRNA as effectively as the WT protein, the other variants carrying a single or double mutation demonstrated slightly reduced levels of crRNA binding compared to WT FnCas12a. The helix 1 deletion variants FnCas12a Δh1 and ΔBH/Δh1 were impaired in the binding of the crRNA (Δh1 10.7%, ΔBH/Δh1: 22.9% binding efficiency relative to the wildtype). Whereas the FnCas12a ΔBH and ΔΔBH variants still bound the crRNA, albeit, with a reduction of 50% compared to the wildtype, the FnCas12a Δh1 variant showed a more drastic reduction, indicating that helix 1 is important for the crRNA binding capacity.

The pre-crRNA processing activity of the FnCas12a helix 1 variants was investigated in an EMSA with a labeled pre-crRNA substrate (**Figure 46D–F**). The RNase-inactive variant FnCas12a HKK was employed as a negative control. The FnCas12a Δh1 and ΔBH/Δh1 variants cleaved the pre-crRNA very inefficiently, which is congruent with their reduced RNA binding capacity. All other FnCas12a helix 1 variants were able to process the pre-crRNA substrate with the same or only slightly changed processing efficiency as the WT protein.

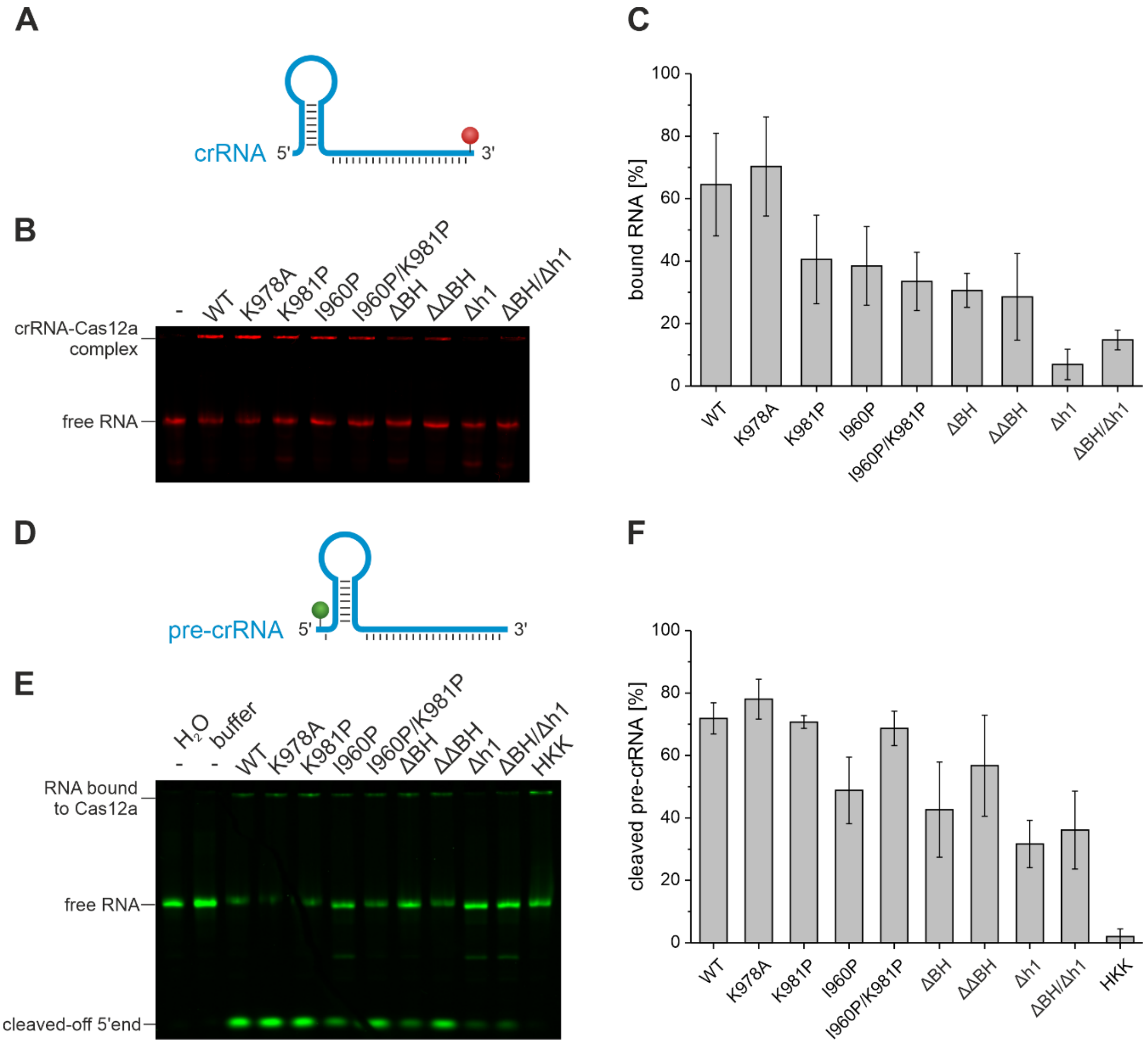


Figure 46: Complex formation of FnCas12a helix 1 variants and crRNA and pre-crRNA processing of FnCas12a helix 1 variants.
(**A**) Schematic representation of crRNA labeled with Cy5 at the 3'-end (red sphere). (**B**) Electrophoretic mobility shift assay of FnCas12a-crRNA complexes. WT FnCas12a, FnCas12a helix 1 variants, and

relevant FnCas12a BH variants for comparison were used in a two-fold molar excess of protein (20 nM) over crRNA (10 nM, 43 nt). Free RNA migrated into the gel and was separated according to its size, whereas RNA bound to protein stayed in the gel pockets. (**C**) Signal quantification of fluorescent bands that correspond to the crRNA-Cas12a complexes. Data are shown from three independent assays. (**D**) Schematic representation of pre-crRNA with an additional nucleotide at the 5'-end (44 nt) labeled with Cy3 at the 5'-end (green sphere). (**E**) Investigation of pre-crRNA processing by FnCas12a helix 1 variants. WT FnCas12a, FnCas12a helix 1 variants, and relevant FnCas12a BH variants for comparison were used in a two-fold molar excess of protein (20 nM) over crRNA (10 nM). Pre-crRNA processing could be followed by the cleavage of the single labeled 5'-nucleotide. (**F**) Signal quantification of pre-crRNA-Cas12a cleavage from three independent assays. (adapted from Wörle *et al.* [294])

After probing the binding behavior of FnCas12a helix 1 variants to the crRNA, their binding behavior to the target DNA was assayed in EMSAs using short doubly labeled target DNA substrates (KW5, **Table 20**, **Figure 47A–D**). These assays showed that FnCas12a K978A binds the DNA with a higher affinity than WT FnCas12a, whereas all other helix 1 variants exhibited a reduced binding capacity. Furthermore, the EMSAs revealed that the helix 1 deletion variants lost their ability to bind short target DNA.

The short target DNA was additionally used to further analyze the cleavage activity of FnCas12a helix 1 variants (**Figure 47E, F**). Congruent with the missing ability to bind the target DNA of FnCas12a Δh1 and ΔBH/Δh1, these variants showed no cleavage activity. Additionally, the FnCas12a I960P/K981P variant was not able to cleave the short DNA substrate, consistent with the observations in the plasmid cleavage assay with linearized DNA. FnCas12a variants that only carried one of the introduced prolines, FnCas12a I960P, and FnCas12a K981P, demonstrated cleavage activity, albeit, with reduced efficiency. For FnCas12a K981P a shift in NTS cleavage products as previously seen for the FnCas12a BH variants was detected. Similar observations were made for the FnCas12a I960P variant. The variant carrying the alanine mutation at the TS-contacting site K978 cleaved the target DNA with the same cleavage efficiency as WT FnCas12a.

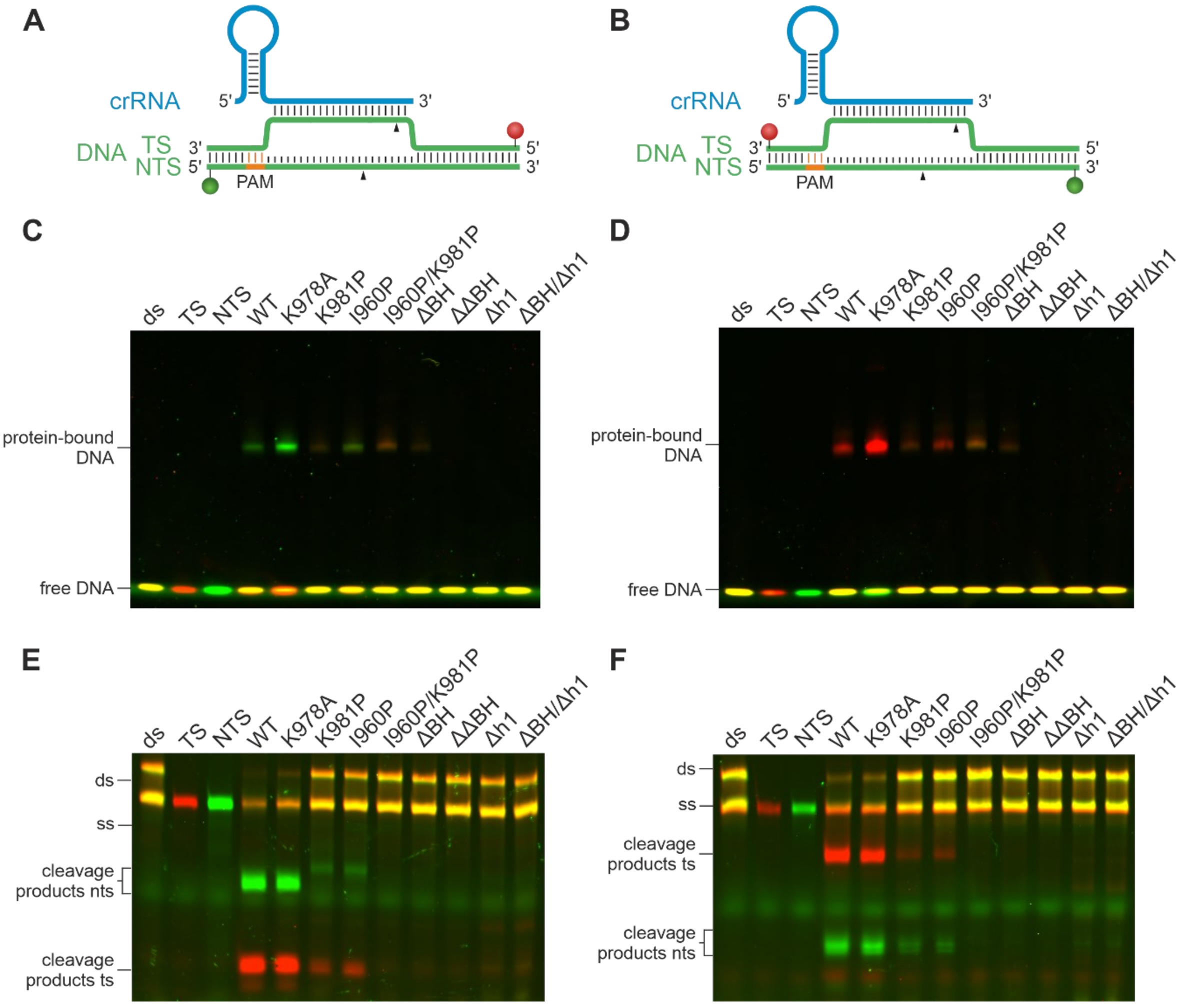


Figure 47: Activity assays of FnCas12a helix 1 variants using short double-stranded target DNA. WT FnCas12a, FnCas12a helix 1 variants, and for comparison relevant FnCas12a BH variants were used in a 7.5-fold molar excess of protein and crRNA (75 nM) over target DNA (10 nM). Reactions were incubated for 1 h at 37 °C. The short target DNA (58 nt) was doubly labeled with a Cy3 label (green) at the non-target strand (NTS) and a Cy5 label (red) at the target strand (TS). (**A**) Cy3 (green sphere) and Cy5 (red sphere) labels were coupled to the 5'-end of the DNA strands, respectively. (**B**) Cy3 (green sphere) and Cy5 (red sphere) labels were attached to the 3'-end of the DNA strands, respectively. (**C**) EMSA of FnCas12a helix 1 variants with 5'-labeled DNA. (**D**) EMSA of FnCas12a helix 1 variants with 3'-labeled DNA. (**E**) Cleavage assay of FnCas12a helix 1 variants with 5'-labeled DNA. (**F**) Cleavage assay of FnCas12a helix 1 variants with 3'-labeled DNA.

To thoroughly analyze the cleavage behavior of the FnCas12a helix 1 variants on the short target DNA, the cleavage products were separated on a high-resolution sequencing gel (**Figure 48**). Here, the FnCas12a K978A variant revealed the same cleavage pattern resulting from the trimming activity as WT FnCas12a, indicating that the disruption of the interaction between K978 and the TS does neither influence cleavage efficiency, nor cleavage accuracy. As indicated by the low-resolution cleavage assay (**Figure 46E**), FnCas12a K981P cleaved the target DNA inefficiently with a cleavage pattern similar to FnCas12a I960P and the other BH variants. This indicates that the disruption of the helical nature of either the BH or helix 1 results in a strongly reduced NTS trimming activity. Even though binding of the target DNA could not be detected for the FnCas12a Δh1 and ΔBH/Δh1 variants (**Figure 47C, D**), faint cleavage products could be detected on the high-resolution sequencing gels (**Figure 48C, D**). Notably, the cleavage position at the TS was significantly shifted, however, precise at a defined position. This indicates a cleavage position shifted towards the 3'-end of the TS. In contrast, the cleavage position at the NTS was not altered, but trimming activity could again not be detected.

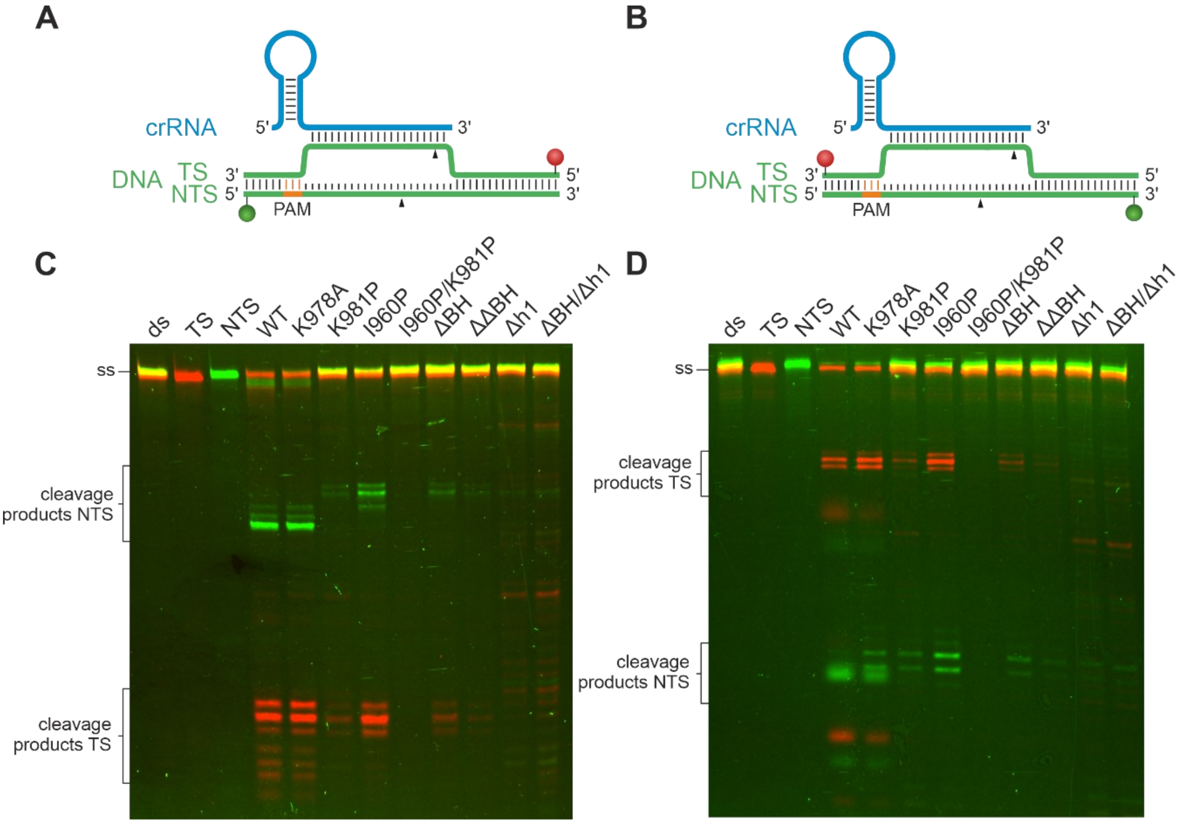


Figure 48: High-resolution cleavage assays of FnCas12a helix 1 variants.
Cleavage assay of FnCas12a helix 1 variants using a 7.5-fold molar excess of protein and crRNA (750 nM) relative to the target DNA (100 nM). The reaction was incubated at 37 °C for 1 h. The short target DNA (58 nt) was doubly labeled with a Cy3 label (green) at the non-target strand (NTS) and a Cy5 label (red) at the target strand (TS). (**A**) Cy3 (green sphere) and Cy5 (red sphere) labels were attached to the 5'-end of the DNA strands. (**B**) Cy3 (green sphere) and Cy5 (red sphere) labels were positioned at the 3'-end of the DNA strands. (**C**) High-resolution cleavage assay of FnCas12a helix 1 variants with 5'-labeled DNA. (**D**) High-resolution cleavage assay of FnCas12a helix 1 variants with 3'-labeled DNA.

(adapted from Wörle *et al.* [294])

Since the FnCas12a variants mutated in the BH were more sensitive to DNA mismatches in the target DNA (**Figure 39**), the cleavage reactions with FnCas12a helix 1 variants and mismatched DNA strands were performed as well (**Table 20**, **Figure 49**). Whereas FnCas12a K978A performed equally well as the WT protein in cleaving mismatched DNAs, the FnCas12a K981P, I960P/K981P, Δh1, and ΔBH/Δh1 variants did not cleave any of the mismatched DNAs.

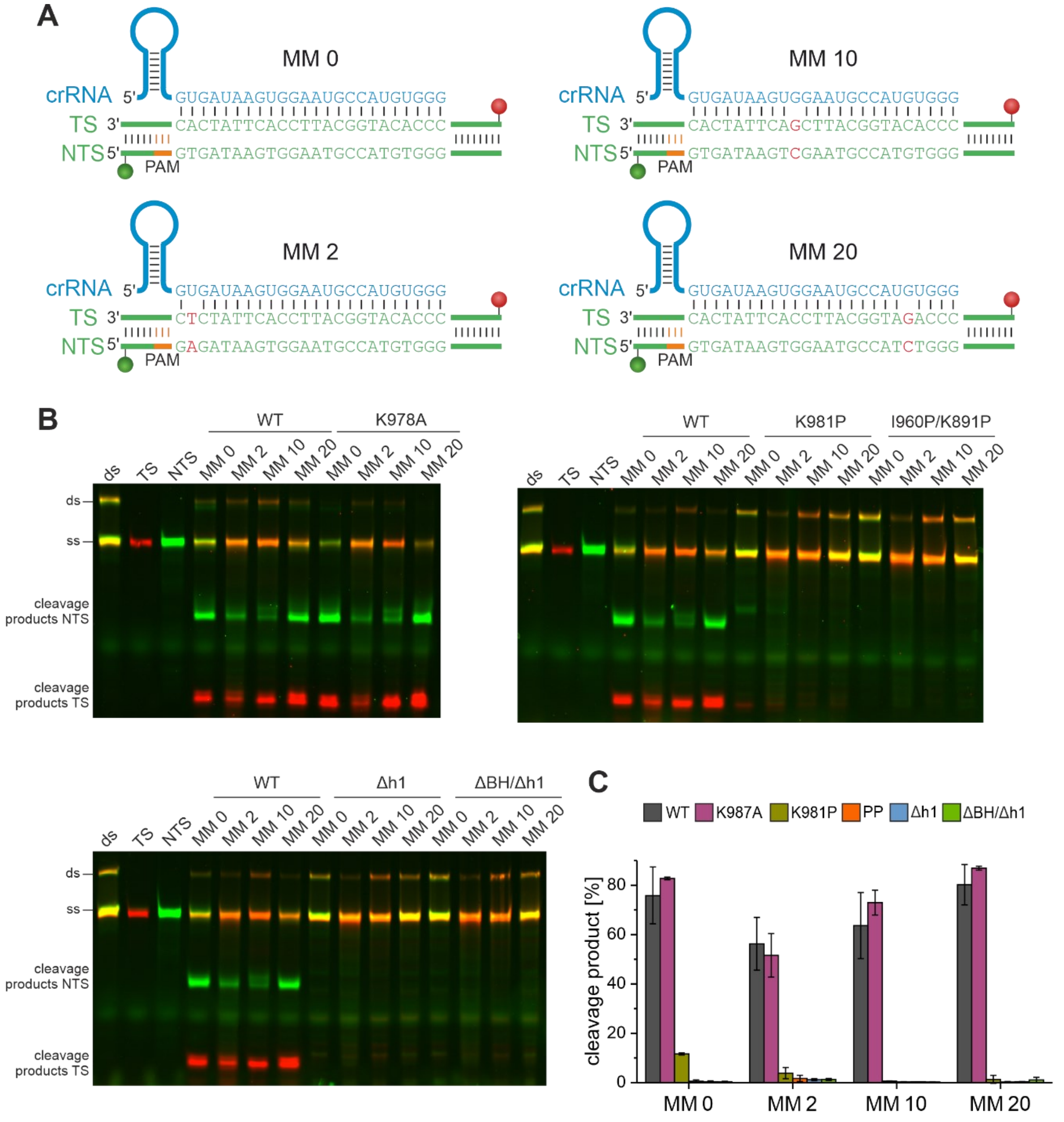


Figure 49: DNA cleavage specificity of FnCas12a helix 1 variants at mismatched substrates. (**A**) The short target DNA (58 nt) was doubly labeled with a Cy3 label (green sphere) at the 5'end of the NTS and a Cy5 label (red sphere) at the 5'-end of the TS. In addition to the fully complementary substrate (MM 0), substrates that carry a single mismatch at position 2 (MM 2), 10 (MM 10), or 20 (MM 20) were used. (**B**) Cleavage assay of FnCas12a helix 1 variants using a 7.5-fold molar excess of protein and crRNA (75 nM) relative to the target DNA (10 nM). The reaction was incubated at 37 °C for 1 h. (**C**) Signal quantification of the cleavage efficiency of the NTS for the WT and FnCas12a helix 1 variants using fully complementary and mismatched DNA substrates. Shown is the average of three independent assays, error bars show the standard deviation. PP: I960P/K981P. (adapted from Wörle *et al.* [294])

Gaetan Burgio and Anthony Newman, collaboration partners from the Australian National University (Canberra, AUS), performed *trans*-cleavage assays with some of the FnCas12a helix 1 variants as well as selected BH variants for comparison (**Figure 50**). After cleavage of ds or ssDNA in *cis*, Cas12a is primed for indiscriminate cleavage of nucleic acids, mainly ssDNA, in *trans* [236,237].

A short FAM-labeled ssDNA target additionally carrying a fluorescence quencher on the other end was used as a reporter substrate. The fluorescence quencher prevents the emission of fluorescence light resulting in low fluorescence of the intact DNA strand. *Trans*-cleavage leads to the separation of quencher and fluorophore, hence, the fluorescence increases, and the extent of *trans*-cleavage activity can be detected by the increase in fluorescence (**Figure 50A**).

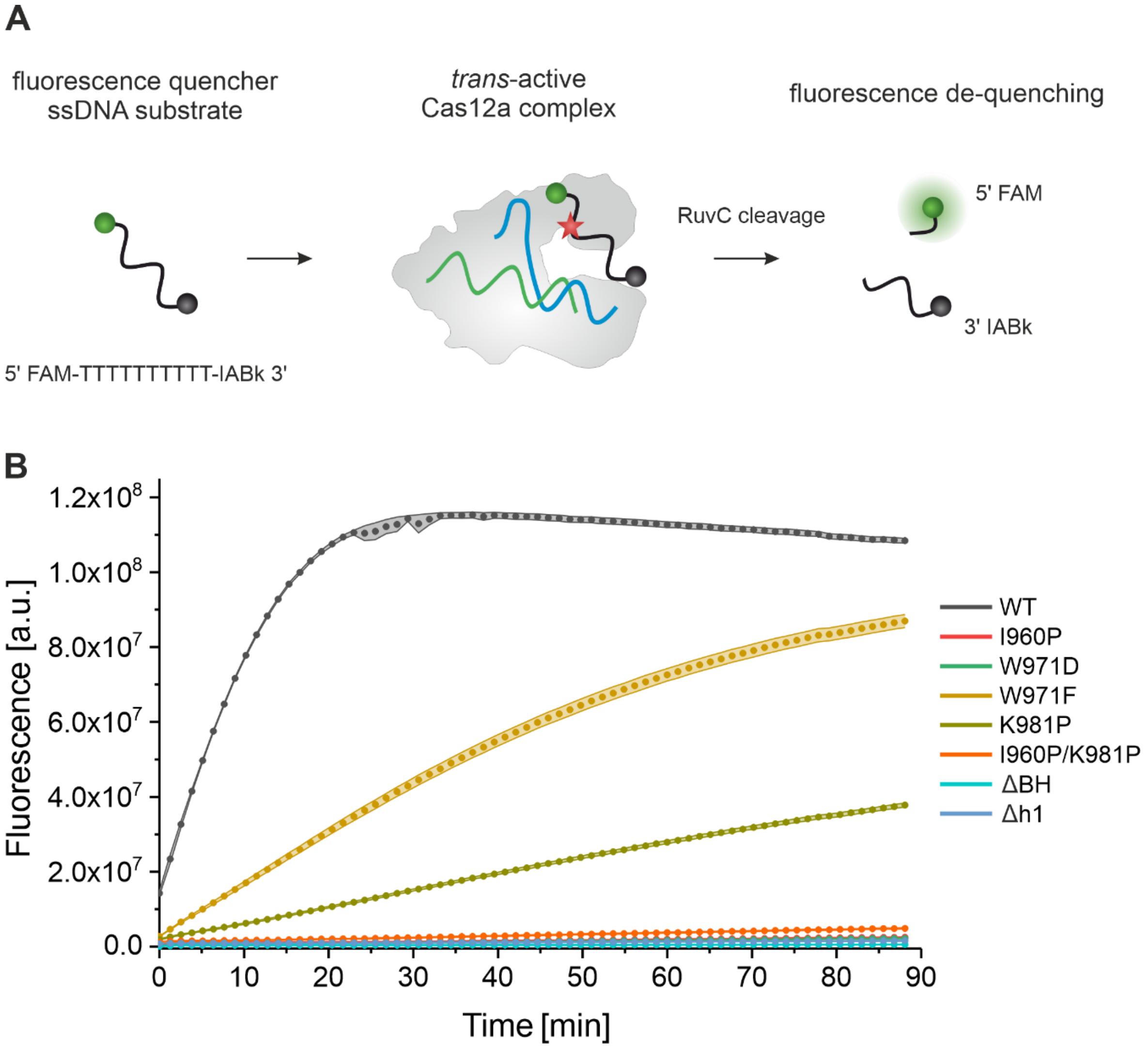


Figure 50: *Trans*-cleavage of FnCas12a helix 1 variants in comparison to FnCas12a bridge helix variants.
(**A**) The binary complex (100 nM of premixed 1 μM FnCas12a and 1 μM crRNA) was incubated with an ssDNA activator (5 nM ssDNA) and 100 nM Fluorescence-Quencher (FQ) reporter substrate in a *trans*-cleavage buffer for 20 min. (**B**) The fluorescence intensity of the cleaved reporter strand was measured on a plate reader over time for 90 min. Data points represent three replicates from independent experiments with the standard deviation indicated as colored areas. (adapted from Wörle *et al.* [294])

For WT FnCas12a and the variants W971F and K981P, *trans*-cleavage activity could be detected, although the activity of the variants was drastically reduced (**Figure 50B, Table 36**). All other mutants tested in the *trans*-cleavage assay (FnCas12a I960P, W971D, I960P/K981P, ΔBH, and Δh1) did not show any activity. This leads to the conclusion that the BH and helix 1 are highly important for *trans*-cleavage activity.

Table 36: Kinetic parameters for FnCas12a *trans*-cleavage.

	WT	**W971F**	**K981P**
v_{max} [M/s]	$5.07 \cdot 10^{-10}$	$3.14 \cdot 10^{-9}$	$4.17 \cdot 10^{-11}$
K_M [M]	$9.93 \cdot 10^{-7}$	$3.85 \cdot 10^{-6}$	$3.76 \cdot 10^{-7}$
k_{cat} [1/s]	$1.69 \cdot 10^{-3}$	$1.05 \cdot 10^{-2}$	$1.39 \cdot 10^{-4}$
k_{cat}/K_M [M^{-1} s^{-1}]	$1.70 \cdot 10^{3}$	$2.72 \cdot 10^{3}$	$3.70 \cdot 10^{1}$

5.4.2 Comparison of cleavage efficiencies of all FnCas12a variants

The following table shows a summary of cleavage activities and efficiencies of all FnCas12a variants produces during this thesis (**Table 37**).

Table 37: Comparison of cleavage efficiencies of all FnCas12a variants generated in this thesis. The efficiency of cleavage or activity is indicated with: +++ very good cleavage/activity, ++ good cleavage/activity, + faint cleavage/activity, − no cleavage/activity, n.a. not assessed.

FnCas12a	**pre-crRNA processing**	**linearized plasmid**	**supercoiled plasmid**	**short dsDNA**	**trimming activity**	***trans* cleavage**
WT	+++	+++	+++	+++	+++	+++
dCas12a	+++	−	−	−	−	n.a.
HKK	−	n.a.	n.a.	+++	n.a.	n.a.
K1066A	n.a.	n.a.	n.a.	+++	n.a.	n.a.
K823A	n.a.	n.a.	n.a.	++	n.a.	n.a.
I195R	n.a.	n.a.	n.a.	+++	n.a.	n.a.
D194R	n.a.	n.a.	n.a.	+++	n.a.	n.a.
N188R	n.a.	n.a.	n.a.	+	n.a.	n.a.
I960P	++	++	++	++	+	−
W971A	+++	+++	+++	+++	++	n.a.
W971D	+++	+++	++	++	−	−
W971K	+++	+++	++	++	−	n.a.
W971F	+++	+++	+++	+++	++	++
ΔBH	++	+	+	+	−	−
ΔΔBH	++	+	+	+	−	n.a.
K978A	+++	+++	+++	+++	+++	n.a.
K981P	+++	++	++	+	−	+
I960P/K981P	+++	−	+	−	−	−
Δh1	++	−	+	+	−	−
ΔBH/Δh1	++	−	+	+	−	n.a.

5.4.3 Confocal single-molecule FRET analysis of FnCas12a helix 1 variants

To analyze the conformational states of the FnCas12a helix 1 variants confocal single-molecule FRET measurements were conducted with the variants FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ I960P/K981P, FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ Δh1, and FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ ΔBH/Δh1 with fluorescent labels engineered in the REC and the Nuc lobe at positions D470 and T1222 (**Figure 51**).

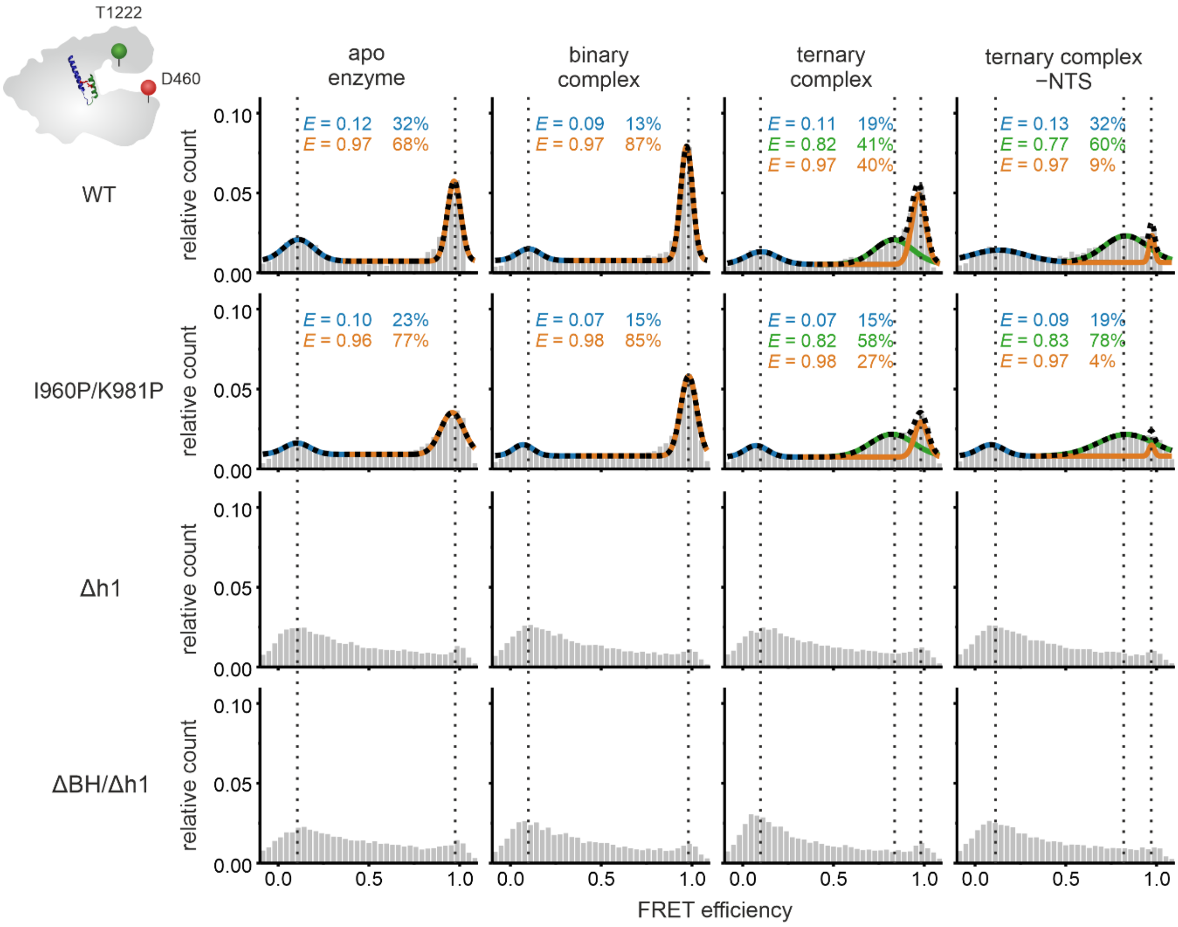


Figure 51: Confocal single-molecule FRET measurements using doubly labeled FnCas12a helix 1 variants.
smFRET measurements were conducted with the apo enzymes of FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ helix 1 variants, the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM target DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS). Histograms show the average data of three independent measurements. The histograms were fitted with a double or triple Gaussian function, the overall fit is shown in black (dotted line). The mean FRET efficiencies *E* and the percentage distribution of the populations are given in the histograms. Additional information regarding fit parameters and molecule counts can be found in Appendix III.viii. (adapted from Wörle *et al.* [294])

Although two prolines were introduced into the BH and helix 1, FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ I960P/K981P adopted the low and the high FRET populations for the closed and open conformations of the apo and the binary complex. Furthermore, the additional medium FRET population after the addition of target DNA is detected, indicating the formation of the ternary complex and the re-opening of the protein. This suggests that FnCas12a$^{\text{REC-}}$

$^{\mathrm{Nuc*DL550/DL650}}$ I960P/K981P efficiently undergoes the transitions from the apo state to the binary complex (open to closed) and from the binary to the ternary complex (closed to semi-closed). By contrast, the FnCas12a$^{\mathrm{REC\text{-}Nuc*DL550/DL650}}$ Δh1 and ΔBH/Δh1 variants almost completely lost the high FRET population indicating that in these FnCas12a helix 1 variants the closure of the protein cannot commence. Additionally, neither the addition of crRNA, nor the addition of crRNA and target DNA induced any changes in the FRET distribution. Most likely this is a result of the defect in crRNA and DNA binding of these FnCas12a variants lacking helix 1 that was already observed in EMSA experiments. Hence, the smFRET data could not be reliably fitted with Gaussian distributions.

Overall, structural disruptions of helix 1 in FnCas12a reveal major impairments in the binding of nucleic acids, the catalytic activity, and the conformational transitions of the protein. This leads to the assumption that helix 1 is the main structural element for the coordination of the REC and the Nuc lobes as well as the correct positioning of the nucleic acids in the protein.

6. DISCUSSION

A step to a broader application of the great potential of the CRISPR-Cas gene-editing technology in selective human therapy of non-stem cells is the full understanding of the Cas nuclease used for gene editing and the processes leading to DNA cleavage, repair, and the resulting DNA modification. To get there, the knowledge about structural conformations and transitions in the nucleases upon nucleic acid binding and cleavage as well as their cleavage behavior is of crucial importance. With this, Cas nucleases can be engineered to accomplish different tasks and to enhance their activity and/or specificity. In Cas9, for example, the engineering of the sgRNA and its 3'-end extension facilitate the usage of Cas9 but preserve its activity [129,295]. The finding of two independent active sites for NTS and TS cleavage led to the effective conversion of Cas9 into a DNA nickase by inactivating one of the two active sites [149]. To reduce or even suppress off-target effects, high-fidelity Cas9 variants were generated that minimize non-specific DNA interactions by reducing the protein-DNA contacting sites [164] or by destabilizing the NTS-binding groove in Cas9 to encourage rehybridization of DNA strands that are not fully complementary to the sgRNA [163]. A broader spectrum of possible target sequences can be achieved by altering the PAM contacting residues to recognize different PAM sequences [296].

To contribute to that knowledge, in this study, mechanistic details of the FnCas12a activity cycle were unraveled via single-molecule FRET measurements. Additionally, the function and importance of two central structural elements, namely the bridge helix and helix 1 of the RuvC II domain, were deciphered. These analyses revealed that both helices are important for the cleavage activity but that helix 1 is the key element for the structural integrity of FnCas12a.

6.1 Conformational changes of Cas12a throughout its activity cycle

Crystal and cryo-electron microscopy (EM) structures [133,217–221,224,226,242,244–246,291] as well as smFRET studies [222,226,228] showed the conformational flexibility of Cas12a and the conformational changes upon the transition from binary Cas12a-crRNA complexes to ternary Cas12a-crRNA-target DNA complexes. Bound to the crRNA, Cas12a adopts a compact and closed conformation. Binding to the target DNA induces an inter-lobal conformational change and leads to a re-opening of the enzyme. Nevertheless, the conformation of the apo state of Cas12a was not elucidated by these studies. Crystal and EM structures only provide static snapshots of a potentially highly dynamic enzyme and even less information is available on the apo enzyme. Based on cryo-EM structure analysis of the binary Cas12a complex of *Lachnospiraceae bacterium* (Lb) and negative staining experiments of apo LbCas12a, Dong *et al*. proposed the existence of a conformationally flexible open conformation for the apo state [217]. Formation of binary complexes with crRNA results in at least one defined structural state that they analyzed employing cryo-EM. Applying smFRET studies with immobilized proteins, Stella *et al*. found additional indications for the existence of an open conformation [226]. For the doubly labeled proteins carrying the labels in the REC and the Nuc lobe (positions C473 and C1190, respectively), they found broadly distributed FRET efficiency histograms for the apo enzyme as well as for nucleic acid bound states. This indicated that Cas12a in its apo form adopts a wide range of conformations that could not be identified as discrete states. Exclusively in the apo state, a relatively high fraction of molecules adopted a low FRET state, supporting their suggestion of an open conformation. Nonetheless, the mechanistic details of the activity cycle of Cas12a proteins remained enigmatic.

In this thesis, an in-depth smFRET study on diffusing molecules with site-specifically attached fluorescence labels was conducted. Intramolecular smFRET measurements enabled the observation of movements of the REC and the Nuc lobe of Cas12a of *Francisella novicida* (Fn) in its different nucleic acid bound states. In this study, a variety of positions for fluorescent labels were chosen to perform measurements on the WED-REC or the WED-Nuc axis to observe potential intra-lobal movements. This unraveled only small conformational transitions along both of the intra-lobal axes with the chosen labeling positions which are in line with structural studies of Cas12a [133,217–221,224,226,242,244–246,291]. Instead, the measurements of a variant with labels at the REC and the Nuc lobe revealed that the major flexibility and majority of conformational transitions occur inter-lobal and the individual lobes are relatively static. Especially in the case of the apo enzyme, conformational heterogeneity was detected, showing two discrete conformational states with a low FRET and a high FRET population. In line with the negative staining results of Dong *et al*. and the smFRET results of Stella *et al*. [217,226], this study clearly reveals a mixture of open and closed complexes for apo FnCas12a. In contrast to Stella *et al.* the measurements presented here, however, reveal that the apo enzyme adopts two discrete states and shows dynamic switching between these states.

Intramolecular smFRET measurements of the binary and ternary complex conducted in this study are in agreement with structural data of FnCas12a and AsCas12a (*Acidaminococcus sp.*) – both enzymes were found to adopt the closed conformation in the binary complex and re-open upon target DNA binding [133,217–221,224,226,242,244–246,291]. The binary complex adopts a closed and defined conformational state, reflected by the high FRET population. The data presented in this study, show that the binding of Cas12a to the pre-crRNA already induces the closure of the protein and that processing of the crRNA does not lead to the formation of a different distinct conformation detectable with the chosen labeling positions.

Up to now, the structure of Cas12a solely bound to a pre-crRNA was not assessed (a crystal structure of an FnCas12a ternary complex bound to pre-crRNA and target DNA is available, PDB: 5NFV [220]). The closed conformation is stabilized by intermolecular interactions of crRNA and Cas12a protein. Residues from the WED, RuvC, and REC2 domain coordinate the pseudoknot of the RNA, whereas other residues of the WED domain and additional amino acids from the REC1 domain form interactions with the sugar-phosphate backbone of the crRNA and orient the seed region for base-pairing with the target DNA [220]. Via 1D diffusion on DNA, Cas12a scans the DNA for the PAM and target sequence complementary to the crRNA [222]. In the ternary complex, crRNA and the target strand of the DNA form a heteroduplex spanning 20 nucleotides. This heteroduplex is in extensive interactions with the REC1 and REC2 domains, whereas the displaced non-target strand forms contacts to the PI domain [220]. The DNA duplex upstream of the crRNA-target strand heteroduplex contains the PAM sequence and interacts with residues of the PI, WED, and REC domains [220]. The finding of a more broadly distributed medium FRET population in this study, which arises with the addition of target DNA and the consequently more flexible ternary complex is in line with molecular dynamics simulations that show increased flexibility of the REC and Nuc lobe in the ternary complex [225]. The data presented in this study show for the first time that loading of the target DNA is more efficient for a ssDNA target, suggesting that the formation of the Cas12a-crRNA-dsDNA complex is energetically demanding. Interestingly, the FRET data suggest that the expansion of the REC and the Nuc lobe upon transition to the ternary complex is not depending on the single-stranded or double-stranded nature of the DNA, as loading of the TS is sufficient to induce the re-opening in Cas12a resulting in an identical FRET efficiency for both, ssDNA and dsDNA. This is in agreement with crystal structures by Swarts *et al.* [133].

These findings can be summarized in the following model: FnCas12a is a structurally flexible protein that can either adopt an open or a closed conformation in its apo state (**Figure 52**). The binding of (pre-)crRNA induces the closure of the protein and stabilizes the closed conformation. In this complex, the crRNA is stably bound to the protein priming the complex for target search and DNA scanning [222]. Cas12a re-opens to a semi-closed conformation to bind and accommodate a target DNA and to proceed within its activity cycle with the sequential cleavage of both DNA strands.

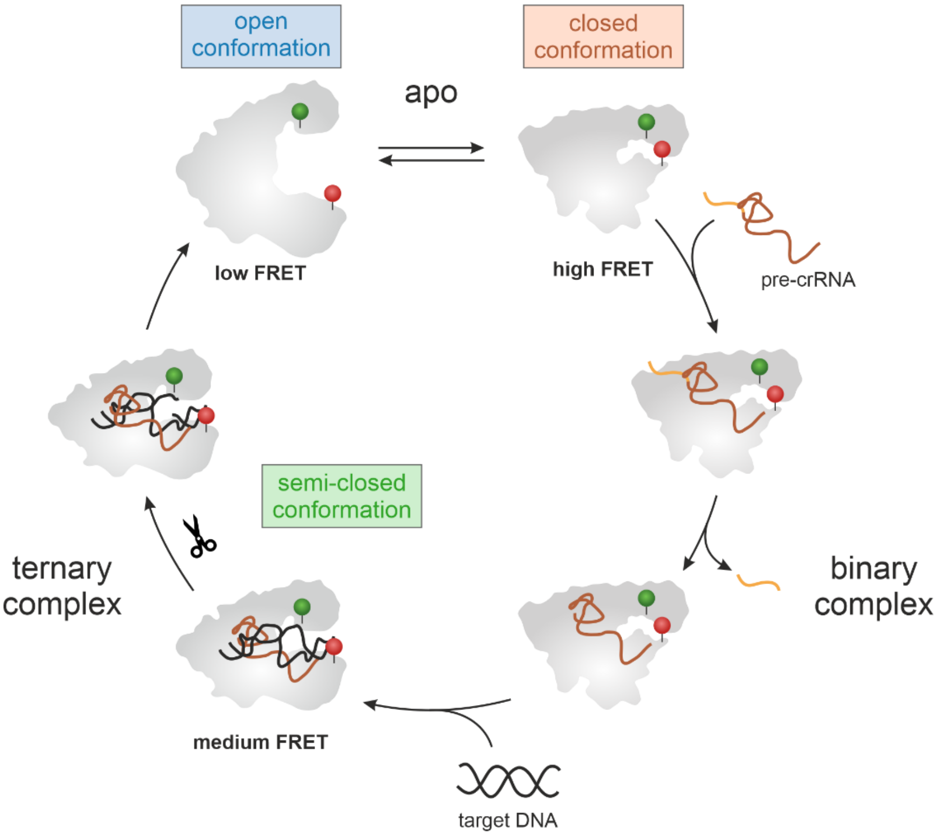


Figure 52: The conformational transitions of FnCas12a during its activity cycle.
In the apo state, FnCas12a can adopt two conformations: a closed (detected as a high FRET population in smFRET measurements) and an open state (detected as a low FRET population in smFRET measurements). The binding of (pre-)crRNA induces the closure of the enzyme. In order to bind a single-stranded or double-stranded target DNA, Cas12a partly re-opens again, giving rise to a medium FRET population in smFRET measurements. (adapted from Wörle *et al.* [290])

6.2 Bridge helix decoupling reduces trimming activity, increases mismatch sensitivity, and impairs conformational transitions

The inter-lobal movements of Cas12a between the Nuc and the REC lobe are the defining conformational changes upon the transition from the apo to the binary and finally to the ternary complex. Apart from the connection via the WED domain, both lobes are additionally connected by the structural element of the bridge helix (BH). An element that occupies such a central position in a protein prompts questions about the functional role of the element.

Cas9, the single-effector nuclease of CRISPR type II systems, has a bilobal structure homologous to Cas12a (**Figure 53A**). The BH of Cas9 also connects the REC and Nuc lobes and reaches deep into the REC lobe, where it, directly and indirectly, contacts the target DNA, crRNA, and tracrRNA [127–130,155,297]. With an arginine-rich site, the BH of Cas9 coordinates the crRNA seed nucleotides and exposes them to the solvent for base-pairing with the target DNA. In the case of *Streptococcus pyogenes* Cas9 (SpyCas9), mutations in the BH resulted in altered RNA and DNA binding, enhanced mismatch sensing, and reduced RNA-loop stability [129,292,293,298]. This raises the question if mutations in the BH of Cas12a result in similar findings as observed for Cas9.

In Cas12a, the REC and Nuc lobe-connecting BH is anchored in the REC lobe as well (**Figure 53B**). But the interaction of the tryptophan residue (FnCas12a: W971) bound in a hydrophobic pocket of REC2 is not as deep as the insertion of the Cas9 BH into the REC lobe of Cas9 [133,220]. The BH of Cas12a also contacts the RNA-DNA-heteroduplex at the sugar-phosphate backbone of the crRNA and at the DNA (shown for AsCas12a), but these interactions are not as extensive as in the case of Cas9 [218].

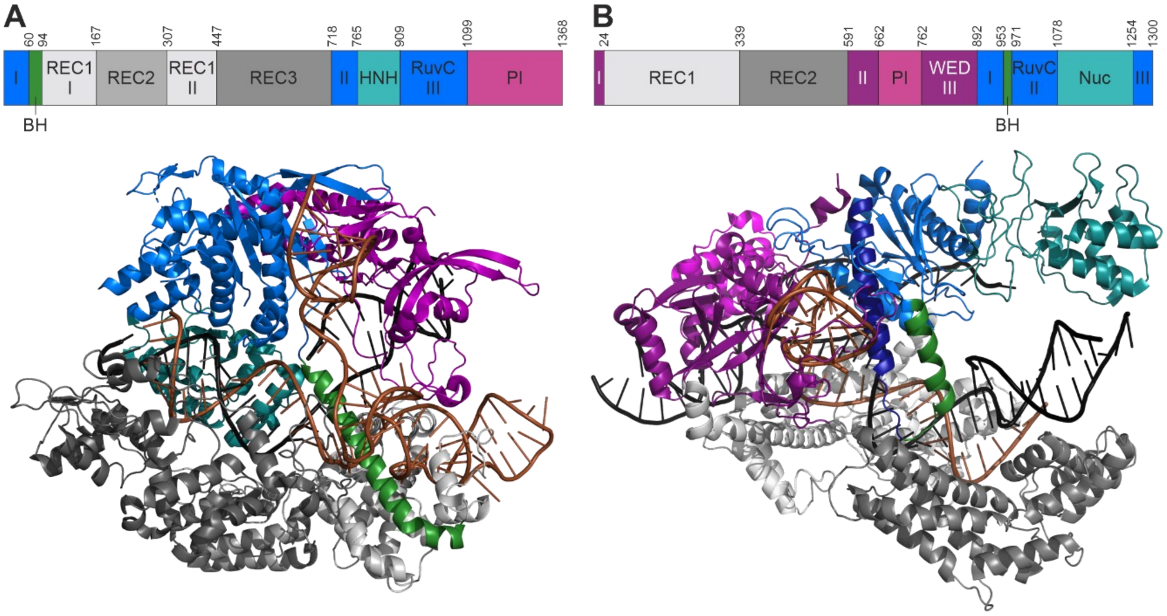


Figure 53: Comparison of Cas9 and Cas12a crystal structures.
(**A**) Crystal structure of SpyCas9 in complex with an sgRNA (brown) and a target DNA (black) (PDB: 4UN3). The bridge helix (BH, green) connects the REC lobe (grey) and the Nuc lobe (containing RuvC (blue) and HNH (cyan) domains). It reaches deep into the REC lobe and forms extensive contacts with RNA and DNA. (**B**) Crystal structure of FnCas12a in complex with crRNA (brown) and target DNA (black) (PDB: 6I1K). The bridge helix (BH, green) connects the REC lobe (grey) and the Nuc lobe (containing the RuvC (blue) and Nuc (cyan) domains). It is anchored in the REC lobe and forms some contacts with the crRNA and the target DNA strand.

In this thesis, the mutational study on the BH of FnCas12a showed that the BH influences the catalytic activity and specificity as the deletion of the BH results in a severely reduced DNA cleavage activity. It turned out that the BH is a major determinant of precise cleavage and the trimming activity of Cas12a on the NTS. Comparably, the deletion of the BH in SpyCas9 yielded an inactive variant [299]. Measurements of FnCas12a with a proline substitution in the BH additionally revealed that the full helical nature of the BH is required for effective cleavage and the NTS trimming activity. This is congruent with the findings of Parameshwaran *et al.* [300] who tested a double proline variant of the BH of FnCas12a. The authors hypothesized that the BH plays a role in the coordination of NTS and TS into the active site for cleavage. Also, the disruption of the BH anchoring in the REC domain influences the overall cleavage and trimming activity on the NTS. Whereas the replacement of W971 with phenylalanine or alanine did not affect the cleavage activity and only slightly reduced the trimming activity, the substitution to charged amino acids resulted in a reduction of the DNA cleavage rate, especially with a short target DNA, and the loss of the trimming activity. Biochemically, the hydrophobic phenylalanine most likely can functionally replace the likewise hydrophobic tryptophan residue, similarly, alanine as a small and uncharged amino acid, probably still fits into the pocket in the REC2 domain. By contrast, the charged amino acids aspartate and lysine provoke a repulsion of the BH from the hydrophobic pocket and therefore exhibit stronger effects.

Reduced trimming activity was also observed by Swarts *et al.* with different FnCas12a variants (FnCas12a D917A, E920A, F1012A, S1083A, R1218A, D1255N, D1227A, and D1227N) [220]. However, these mutated residues are close to the active site or in contact with the catalytic amino acids, which is not the case for the BH variants that, nevertheless, show a similar phenotype.

In a next step of this study, the effect of mismatches between the crRNA and the target sequence on the FnCas12a cleavage activity was tested. All FnCas12a BH variants show higher sensitivity to mismatches than WT FnCas12a. For example, a mismatch at position 10 after the PAM was not tolerated by any of the variants, and FnCas12a W971K and W971D additionally were sensitive to mismatches at position 20.

Parameshwaran *et al.* also tested their BH variant FnCas12a K969P/D970P in cleavage assays using supercoiled mismatched target DNAs and observed selective nicking of the DNA, which was highly distinct between positions 12 and 17 (counted from the PAM) [300]. Similar to the FnCas12a I960P variant used in this thesis, FnCas12a K969P/D970P did not show discrimination against mismatches at position 19 and thereafter.

Swarts *et al.* tested the mismatch tolerance of the WT FnCas12a cleavage activity using a linearized plasmid DNA [220]. They observed a complete loss of activity with a mismatch at position 2 after the PAM, in contrast to the data collected in this thesis. Here only an activity decrease of 35% compared to the fully matched substrate was found. Parameshwaran *et al.* [300] even reported almost no difference if a target DNA with a mismatch at position 2 was used. In the study of Swarts *et al.* mismatches at positions 10 or 20 resulted in approximately 50% and 100% cleavage activity, respectively [220], which is in agreement with the data presented in this thesis. For both mismatch positions, Parameshwaran *et al.* did not observe a decrease in cleavage efficiency for the wildtype enzyme [300]. This discrepancy might be explained by the substrates used in the various studies. In this thesis as well as in the other two studies (Swarts *et al.* [220], Parameshwaran *et al.* [300]) different types of substrates (supercoiled plasmid, linearized plasmid, and short dsDNA) were utilized.

Another study on the mismatch tolerance of a Cas12a variant (AsCas12a) employed a short target DNA. In that study, Strohkendl *et al.* showed that mismatches between crRNA and target DNA in the R-loop region slow down the binding of the RNP to the DNA and reduce their affinity to each other [231]. With the R-loop formation being the rate-limiting step in the cleavage reaction [231], the reduced affinity directly decreases the target cleavage rate. Consequently, Strohkendl *et al.* presented drastically decreased cleavage rates for substrates with mismatches at positions 2 or 4 downstream of the PAM and only a minor impact on the cleavage rate if mismatches are at positions 10, 18, or 21, which are not involved in R-loop formation.

In SpyCas9, the introduction of two adjacent proline residues in the BH increases the selectivity, pointing to the stabilizing role of Cas9's BH in R-loop formation [293]. The interactions of the BH in Cas12a are not as tight as in Cas9. In the case of Cas12a, the BH is only contacting the phosphate backbone of crRNA and DNA and stabilizes the R-loop in a rather indirect way [218,220]. Hence, as the FnCas12a variants with destabilized BH anchoring, FnCas12a W971D and W971K, showed reduced cleavage and trimming activities, but at the same time, an increase in mismatch sensitivity, the anchoring of the BH seems to be essential for the stable formation of the R-loop in FnCas12a.

In contrast, as the FnCas12a W971D and W971K variants are reduced in their trimming activity, they produce a more precise cleavage pattern than the wildtype, with cleavage occurring mainly at two distinct sites. For comparison, WT FnCas12a produces at least four intermediate trimming products on the way to the two final products. Together with the higher specificity, resulting from better discrimination against mismatched DNA substrates, these variants might be better suited for gene editing applications than WT FnCas12a. In nature, the lower specificity of WT Cas12a might be an advantage for a prokaryotic defense system. A more flexible mechanism might be better to recognize a higher variety of closely related and rapidly evolving invading phages. In contrast, for genome editing applications a highly specific and precise cleavage is favored to not encounter undesired off-target effects. Hence, these variants would be promising for further testing and engineering to yield specific and precise genome editing enzymes with higher cleavage efficiency.

To further elucidate functional and mechanistic details of the conformational transitions of the FnCas12a BH variants, smFRET measurements were conducted. It seemed possible that the deletion of the central BH leads to a collapse of Cas12a, however, this was not observed. In fact, all BH variants displayed the same two or three FRET efficiency populations as the wildtype protein for the apo and binary or the ternary complex, respectively. This indicates that the structural integrity remains preserved and loading with crRNA and target DNA is efficient for all BH variants. Notably, whereas for the W971 variants the open state is favored, in the BH deletion variants the open state is less populated, and the closed state is more favored, indicating that the BH influences the equilibrium between open and closed state.

Inspection of all available Cas12a structures [133,217–222,224,226,242,244–246,291] revealed that the BH expands upon the transition to the ternary complex. This structural change occurs in tandem with a shortening of the adjacent helix 1 of the RuvC II domain (**Figure 54**). The tryptophan at position 971 is part of helix 1 in the binary state. Upon the transitions in the REC lobe, leading to the formation of the ternary complex, the tryptophan at position 971 shifts by 8.7 Å to remain docked in its binding pocket in the REC2 domain. During this re-positioning, W971 is released from helix 1 into a flexible loop, connecting both helices. Helix 1 is thereafter shortened by six amino acids, whereas the BH is extended by two residues in FnCas12a and four residues in LbCas12a. A conserved lysine (K978 in FnCas12a, Appendix II.iii, **Figure 60**) is positioned in the middle of helix 1 in the binary complex [220]. Upon the formation of the ternary complex, helix 1 is shifted and orients K978 ideally to form interactions with the sugar-phosphate backbone of the TS. In the re-ordered helix 1, K978 is positioned at the N-terminal end of the helix but not released into the flexible loop between helix 1 and the BH, which suggests that the positioning of K978 in the helix is important for the formation of its DNA contact. The tandem helices and their rearrangement are conserved in all Cas12a variants studied so far [133,217,218,224,242,246,291].

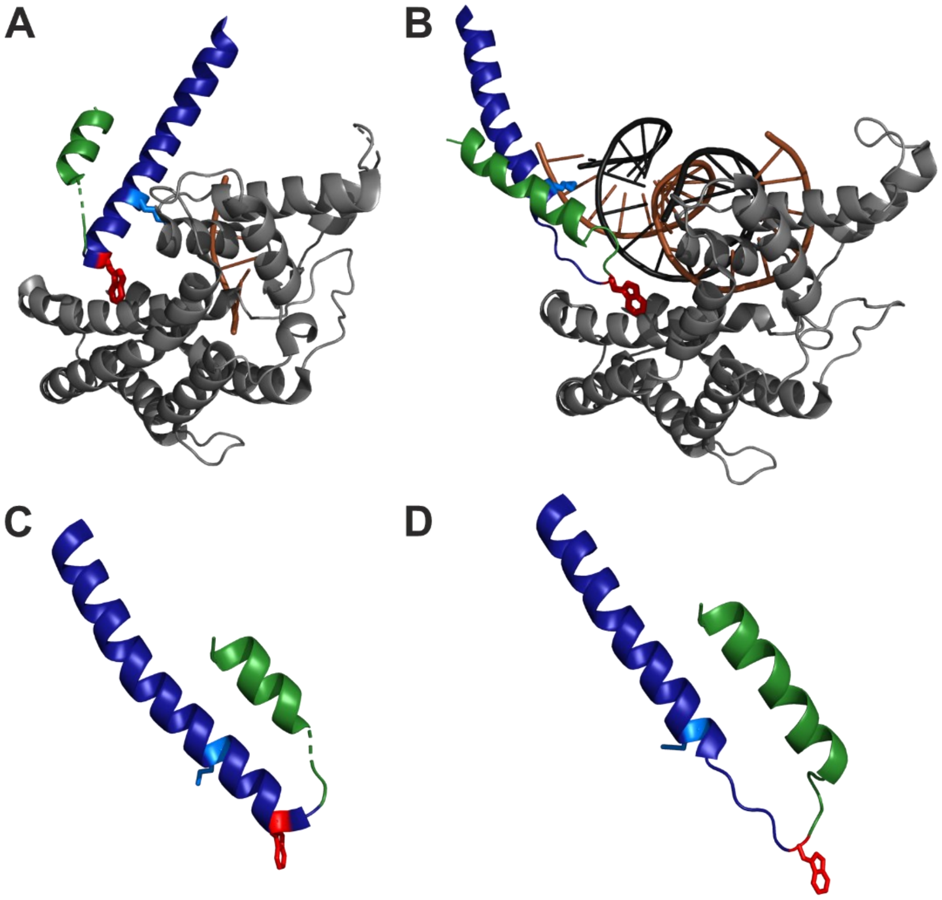


Figure 54: Tandem movement of bridge helix and helix 1.
(**A**) and (**B**) Comparison of the structural rearrangement of the BH and helix 1 relative to the REC domain upon transition from the binary (PDB: 5NG6) to ternary (PDB: 6I1K) complex. The structures were aligned relative to the REC domain as the reference point. For clarity, only the bridge helix (BH, green), helix 1 (blue), and the REC domain (grey) are shown (RNA: brown, DNA: black). Tryptophan 971 is highlighted in red and lysine 978 in light blue. (**C**) and (**D**) Comparison of the length of BH, helix 1, and the connecting linker in the (**C**) binary (PDB: 5NG6) and (**D**) ternary complex (PDB: 6I1K). Tryptophan 971 is highlighted in red and lysine 978 in light blue. (adapted from Wörle *et al.* [290])

Hence, it is likely that some of the mutations introduced into the BH influence the tandem movement of BH and helix 1, possibly leading to a slightly altered positioning of the REC domains and the DNA contacts, especially at the R-loop. This might result in a negative influence on R-loop formation, which, in turn, increases the sensitivity of FnCas12a to mismatches between crRNA and target DNA. Additionally, the observed trimming defect is probably caused by these disruptions. For example, the deleted BH completely disrupts the helical tandem movement. Similarly, the proline introduced instead of I960, in the middle of the BH, might prevent the formation of the extended helix in the ternary complex. Likewise, the exchange of W971 by charged amino acids (W971D and W971K) and the resulting repulsion from the hydrophobic pocket might have an impact on the tandem movement.

6.3 Helix 1 is the primary stabilizing and connecting structural element in Cas12a

Considering the concerted tandem movement of the tightly connected BH and helix 1, it seemed reasonable that helix 1 could be the main structural element that is responsible for coordinating the distance between the REC and the Nuc lobe thereby mediating the opening and closing of the enzyme to allow the loading of nucleic acids in the protein. Other elements like the BH or the PI and WED domains might aid in the coordination of these rearrangements but, like the BH, only influence the equilibrium between closed and open states and not the specific coordination of the distance between the lobes.

To elucidate this mechanistic detail, a mutational study on helix 1 was conducted. It showed that the lysine at position 978 is not important for the catalytic activity, even though it is conserved (Appendix II.iii, **Figure 60**) and contacts the backbone of the TS DNA [220]. K978 was mutated to alanine which did not result in any loss of cleavage activity or efficiency. The structural importance of helix 1 was assessed via deletion variants (Δh1 and ΔBH/Δh1) and the introduction of prolines to break the helical structure (K981P and I960P/K981P).

The CD-spectra predicted a mainly α-helical structure and an overall intact folding for all tested helix 1 variants and the melting curve analysis exhibited similar melting temperatures for all variants. For the helix 1 deletion variants (Δh1 and ΔBH/Δh1) an additional melting point was observed that probably originates from remaining impurities from the purification. It could also indicate that the deletion of helix 1 in these variants partially destabilizes the protein and leads to a shifted melting behavior.

Plasmid cleavage assays revealed that the variants FnCas12a Δh1 and ΔBH/Δh1 are not competent to cleave linear target DNA in *cis* nor ssDNA in *trans* but show minimal *cis*-cleavage activity on a supercoiled target DNA, the same holds true for FnCas12a I960P/K981P. As shown previously, cleavage efficiencies are higher on supercoiled DNA compared to linear DNA [227,301], originating from relaxation and breathing events in the supercoiled DNA that expose ssDNA and thereby, facilitate R-loop formation and cleavage. Additionally, it is not known if the nicking and linearization shown by variants FnCas12a Δh1, ΔBH/Δh1, and I960P/K981P is sequence-specific. As the DNase active site itself is intact, the double-strand break could be executed at a transiently bound, by DNA breathing randomly melted site of the supercoiled DNA without proper positioning of the DNA in the R-loop and the protein. The FnCas12a variant with a helix disruption in helix 1 (K981P) is more reduced in its cleavage activity than FnCas12a I960P, the variant with proline substitution in the BH. This highlights the importance of helix 1 for the function of Cas12a.

Although they display reduced or only weak binding to the crRNA, the pre-crRNA is processed by all FnCas12a helix 1 variants, albeit very inefficiently by the variants with a deleted helix 1. Probably the pre-crRNA only binds transiently to the proteins, leading to processing in the RNase active site in the WED domain that is structurally not directly connected to helix 1 or the BH and therefore not influenced in its activity but not leading to a stable binding that could be monitored in EMSAs. Additionally, considering the smFRET measurements that indicate the destabilization of the helix 1 deletion variants as they mainly adopt the open conformation, they also do not show the shift to high FRET efficiencies as typically observed after the addition of RNA and the stable formation of the binary complex. In the smFRET measurements for FnCas12a Δh1 and ΔBH/Δh1, also the conformational changes required for target DNA binding and ternary complex formation are prevented. In conclusion, as observed in EMSAs, these variants cannot bind the DNA anymore. Showing that the closed conformation and the re-opening of the enzyme are required to proceed in the catalytic cycle. Hence, the structural function of helix 1 in the stabilization and overall structure of Cas12a cannot be compensated by the BH or other structural elements of the protein. Without helix 1, FnCas12a is not able to adopt its closed conformation, whereas the deletion of the BH only leads to minor structural and functional impairments. This suggests that helix 1 is the functional coupling of the Nuc and the REC lobe.

The pre-crRNA processing activity is also not significantly impacted if the helical structure of helix 1 and/or the BH is disrupted (K981P, I960P/K981P). Interestingly, FnCas12a I960P/K981P can efficiently bind to RNA and DNA and go through all conformational transitions as observed in EMSAs and smFRET measurements but does not cleave linear target DNA in *cis* nor ssDNA in *trans*. As the major conformational transitions leading to the binary and ternary complex formation are still possible in this variant and because neither helix 1 nor the BH considerably interact with the crRNA or the target DNA, it seems plausible that the allosteric activation of the DNase active site in the RuvC domain is prevented. The smFRET data show that FnCas12a I960P/K981P can still undergo the large rotational movement of the REC lobe. Nevertheless, structural studies showed large but also small structural rearrangements leading to the activation of the active site after R-loop formation [133,217,218,220,226]. Especially all the small conformational changes cannot be addressed by smFRET measurements with the here applied labeling positions. Smaller conformational rearrangements occur for example in the ‘REC linker’ that connects the domains REC1 and REC2, the ‘lid’ which is part of the RuvC domain, and the ‘finger’ that is located in the REC1 domain [226]. The lid is closed in the binary complex and prevents access to the catalytic site formed by D917, E1006, and D1255. Upon the transition to the ternary complex, the lid opens and is restructured. A small α-helix is formed within the loop, which is then part of polar interactions (R918 with E1006 and K1013 with D917) stabilizing the active site. This way, the active site in the RuvC domain is activated for cleavage. The lid (residues F1005–K1021 in FnCas12a) is only located four amino acids away from the C-terminal end of helix 1 (K972–N1000) and is therefore under structural influence from helix 1 and its conformational changes. The introduction of prolines into helix 1 and/or the BH probably inhibits the tandem movement of these helices (**Figure 55**). Without the tandem movement, most likely the closed-to-open transition of the lid cannot be completely accomplished and therefore, the catalytic residues, especially E1006, as part of the lid, are not correctly oriented for cleavage. This results in a cleavage-incompetent and catalytically inactive FnCas12a protein (**Figure 55B**).

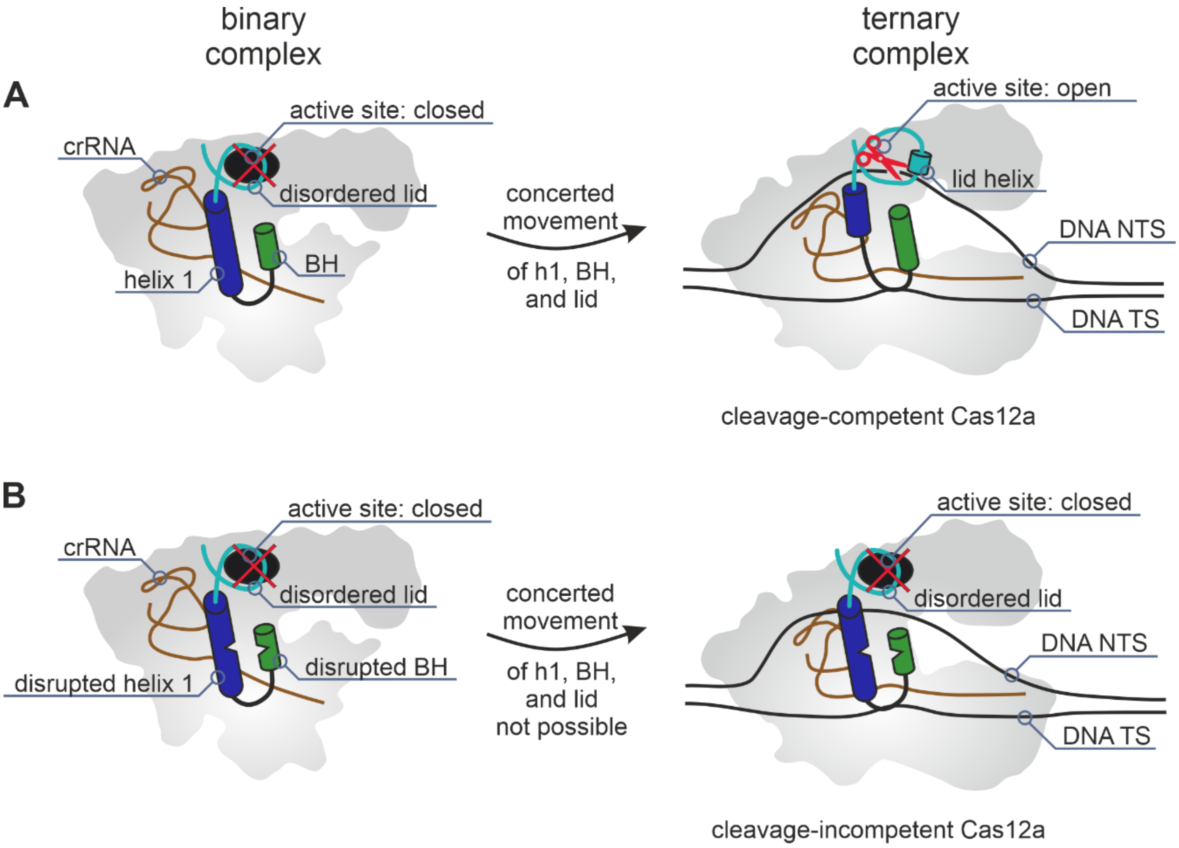


Figure 55: Mechanistic model summarizing the impact of helix 1 on the activity of FnCas12a. (**A**) Loading of target DNA leads to the formation of the ternary complex. This transition is accompanied by an opening of the enzyme and a tandem movement of helix 1 and the BH. Helix 1 (blue cylinder) is shortened and the BH (green cylinder) is extended. Additionally, the 'lid' (cyan) is restructured inducing the formation of a short helix and ultimately the opening and the activation of the active site (red star) in the RuvC II domain. (**B**) The introduction of proline residues in helix 1 and the BH (kink in the cylinders), respectively, prevents the concerted tandem movement of both helices. The impaired movement of helix 1 is translated to the lid domain that cannot reorder, ultimately preventing the opening of the lid. Consequently, the active site is not activated (red cross), and DNA cleavage activity cannot be observed. (adapted from Wörle *et al.* [294])

Taken together, the study reveals that the structural collapse initially anticipated as a consequence of BH deletion was observed for the deletion of helix 1, leading to the conclusion that not the BH but helix 1 is the central helix in Cas12a that is inevitable for the structure and the function of the enzyme. Furthermore, it is hypothesized that the interplay of BH and helix 1 fine-tunes the lid as a structural element and that distortions of this interaction directly affect the active site via the opening or closure of the lid.

6.4 Conclusion and outlook

In this work, smFRET measurements were conducted to elucidate the conformational landscape of the apo state of the bilobal FnCas12a enzyme. Additionally, the conformational transitions of FnCas12a were followed that take place upon the binding of the crRNA and the formation of the binary complex and further the binding of the target DNA and the rearrangements to the cleavage-competent ternary complex. These data showed that FnCas12a adopts an open and a closed conformation in the apo state, closes upon crRNA binding, and re-opens to a semi-closed conformation to accommodate the target DNA in the ternary complex.

Mutational studies on two central helices of FnCas12a, the bridge helix and helix 1, revealed that both helices are crucial for the trimming activity of FnCas12a on the NTS and that helix 1 is an essential structural element for nucleic acid binding and the cleavage of the target DNA.

Fixing the lid in its closed conformation and thereby blocking the active site for DNA cleavage, like with the introduction of prolines into helix 1 and bridge helix, gives another strategy of producing an inactive Cas12a variant. Inactive Cas nucleases can be used to effectively regulate transcription. By sequence-specific binding to DNA without subsequent cleavage of the DNA strands, they can block DNA binding sites for transcription-associated proteins [302]. As fusion proteins, inactive Cas nucleases can direct transcription modulators, but also methyltransferases, deaminases, or fluorescent proteins to the desired site of action [167,168,303–305].

With the BH variants FnCas12a W971D and FnCas12a W971K, two proteins with enhanced cleavage accuracy and mismatch sensitivity were generated. Additionally, the indiscriminate *trans*-cleavage activity that could cause problems in gene editing applications is suppressed (shown for FnCas12a W971D). Together with mutations or modifications for an enhanced cleavage efficiency like the covalent conjunction of Cas12a and crRNA or the broadening of the recognized PAM spectrum [306–308], these variants could be interesting for further investigations and *in vivo* applications in genome editing.

7. REFERENCES

1. Porter, J. R. Antony van Leeuwenhoek. Tercentenary of his discovery of bacteria. *Bacteriol. Rev.* **40**, 260–269 (1976).
2. Twort, F. W. An investigation on the nature of ultra-microscopic viruses. *Lancet* **186**, 1241–1243 (1915).
3. Summers, W. C. In the beginning... *Bacteriophage* **1**, 50–51 (2011).
4. D'Hérell, F. Sur un microbe invisible antagoniste des bacilles dysentériques. *Comptes rendus Acad Sci Paris* **165**, 373–375 (1917).
5. Fortier, L.-C. & Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **4**, 354–365 (2013).
6. Wommack, K. E. & Colwell, R. R. Virioplankton: Viruses in aquatic ecosystems. *Microbiol. Mol. Biol. Rev.* **64**, 69–114 (2000).
7. Brüssow, H. & Hendrix, R. W. Phage genomics: small is beautiful. *Cell* **108**, 13–16 (2002).
8. Hobbs, Z. & Abedon, S. T. Diversity of phage infection types and associated terminology: the problem with 'Lytic or lysogenic'. *FEMS Microbiol. Lett.* **363**, 1–8 (2016).
9. Ackermann, H.-W. Tailed bacteriophages: the order Caudovirales. *Adv. Virus Res.* **51**, 135–201 (1998).
10. Clokie, M. R. J., Millard, A. D., Letarov, A. V. & Heaphy, S. Phages in nature. *Bacteriophage* **1**, 31–45 (2011).
11. Dy, R. L., Richter, C., Salmond, G. P. C. & Fineran, P. C. Remarkable mechanisms in microbes to resist phage infections. *Annu. Rev. Virol.* **1**, 307–331 (2014).
12. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary ecology of prokaryotic immune mechanisms. *Microbiol. Mol. Biol. Rev.* **80**, 745–763 (2016).
13. Hampton, H. G., Watson, B. N. J. & Fineran, P. C. The arms race between bacteria and their phage foes. *Nature* **577**, 327–336 (2020).
14. Hyman, P. & Abedon, S. T. Bacteriophage host range and bacterial resistance. *Adv. Appl. Microbiol.* **70**, 217–248 (2010).
15. Ohshima, Y., Schumacher-Perdreau, F., Peters, G. & Pulverer, G. The role of capsule as a barrier to bacteriophage adsorption in an encapsulated *Staphylococcus simulans* strain. *Med. Microbiol. Immunol.* **177**, 229–233 (1988).
16. Castillo, D., Christiansen, R. H., Dalsgaard, I., Madsen, L. & Middelboe, M. Bacteriophage resistance mechanisms in the fish pathogen *Flavobacterium psychrophilum*: Linking genomic mutations to changes in bacterial virulence factors. *Appl. Environ. Microbiol.* **81**, 1157–1167 (2015).
17. Azam, A. H., Hoshiga, F., Takeuchi, I., Miyanaga, K. & Tanji, Y. Analysis of phage resistance in *Staphylococcus aureus* SA003 reveals different binding mechanisms for the closely related Twort-like phages ϕSA012 and ϕSA039. *Appl. Microbiol. Biotechnol.* **102**, 8963–8977 (2018).
18. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nat. Rev. Microbiol.* **8**, 317–327 (2010).
19. Hammad, A. M. M. Evaluation of alginate-encapsulated *Azotobacter chroococcum* as a phage-resistant and an effective inoculum. *J. Basic Microbiol.* **38**, 9–16 (1998).
20. Hanlon, G. W., Denyer, S. P., Olliff, C. J. & Ibrahim, L. J. Reduction in Exopolysaccharide Viscosity as an Aid to Bacteriophage Penetration through *Pseudomonas aeruginosa* Biofilms. *Appl. Environ. Microbiol.* **67**, 2746–2753 (2001).
21. Kjems, E. Studies on streptococcal bacteriophages. I. Technique of isolating phage-producing strains. *Acta Pathol. Microbiol. Scand.* **36**, 433—440 (1955).

22. Destoumieux-Garzón, D. *et al.* The iron-siderophore transporter FhuA is the receptor for the antimicrobial peptide microcin J25: Role of the microcin Val11-Pro16 β-hairpin region in the recognition mechanism. *Biochem. J.* **389**, 869–876 (2005).

23. Reyes-Robles, T. *et al. Vibrio cholerae* outer membrane vesicles inhibit bacteriophage infection. *J. Bacteriol.* **200**, 1–9 (2018).

24. Manning, A. J. & Kuehn, M. J. Contribution of bacterial outer membrane vesicles to innate bacterial defense. *BMC Microbiol.* **11**, (2011).

25. Uhl, M. A. & Miller, J. F. Integration of multiple domains in a two-component sensor protein: The *Bordetella pertussis* BvgAS phosphorelay. *EMBO J.* **15**, 1028–1036 (1996).

26. Seed, K. D. *et al.* Phase Variable O Antigen Biosynthetic Genes Control Expression of the Major Protective Antigen and Bacteriophage Receptor in *Vibrio cholerae* O1. *PLoS Pathog.* **8**, (2012).

27. Gencay, Y. E., Sørensen, M. C. H., Wenzel, C. Q., Szymanski, C. M. & Brøndsted, L. Phase variable expression of a single phage receptor in *Campylobacter jejuni* NCTC12662 influences sensitivity toward several diverse CPS-dependent phages. *Front. Microbiol.* **9**, (2018).

28. Meyer, J. R. *et al.* Repeatability and contingency in the evolution of a key innovation in phage λ. *Bone* **335**, 528–432 (2012).

29. Lu, M. & Henning, U. Superinfection exclusion by T-even-type coliphages. *Trends Microbiol* **2**, 137–139 (1994).

30. Mahony, J., McGrath, S., Fitzgerald, G. F. & Van Sinderen, D. Identification and characterization of lactococcal-prophage-carried superinfection exclusion genes. *Appl. Environ. Microbiol.* **74**, 6206–6215 (2008).

31. Sun, X., Göhler, A., Heller, K. J. & Neve, H. The ltp gene of temperate *Streptococcus thermophilus* phage TP-J34 confers superinfection exclusion to *Streptococcus thermophilus* and *Lactococcus lactis*. *Virology* **350**, 146–157 (2006).

32. Oliveira, P. H., Touchon, M. & Rocha, E. P. C. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Res.* **42**, 10618–10631 (2014).

33. Loenen, W. A. M., Dryden, D. T. F., Raleigh, E. A., Wilson, G. G. & Murrayy, N. E. Highlights of the DNA cutters: A short history of the restriction enzymes. *Nucleic Acids Res.* **42**, 3–19 (2014).

34. Tock, M. R. & Dryden, D. T. F. The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* **8**, 466–472 (2005).

35. Krüger, D. H. & Bickle, T. A. Bacteriophage survival: Multiple mechanisms for avoiding the deoxyribonucleic acid restriction systems of their hosts. *Microbiol. Rev.* **47**, 345–360 (1983).

36. Samson, J. E., Magadán, A. H., Sabri, M. & Moineau, S. Revenge of the phages: Defeating bacterial defences. *Nat. Rev. Microbiol.* **11**, 675–687 (2013).

37. Loenen, W. A. M. & Murray, N. E. Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage λ. *J. Mol. Biol.* **190**, 11–22 (1986).

38. Semerjian, A. V., Malloy, D. C. & Poteete, A. R. Genetic structure of the bacteriophage P22 PL operon. *J. Mol. Biol.* **207**, 1–13 (1989).

39. Günthert, U. & Reiners, L. *Bacillus subtilis* phage SPR codes for a DNA methyltransferase with triple sequence specificity. *Nucleic Acids Res.* **15**, 3689–3702 (1987).

40. Pleška, M. & Guet, C. C. Effects of mutations in phage restriction sites during escape from restriction-modification. *Biol. Lett.* **13**, 10–13 (2017).

41. Walkinshaw, M. D. *et al.* Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell* **9**, 187–194 (2002).

42. Bandyopadhyay, P. K., Studier, F. W., Hamilton, D. L. & Yuan, R. Inhibition of the type I restriction-modification enzymes EcoB and EcoK by the gene 0.3 protein of bacteriophage T7. *J. Mol. Biol.* **182**, 567–578 (1985).

43. Atanasiu, C., Su, T., Sturrock, S. S. & Dryden, D. T. F. Interaction of the ocr gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res.* **30**, 3936–3944 (2002).

44. Stewart, F. J., Panne, D., Bickle, T. A. & Raleigh, E. A. Methyl-specific DNA binding by McrBC, a modification-dependent restriction enzyme. *J. Mol. Biol.* **298**, 611–622 (2000).

45. Sutherland, E., Coe, L. & Raleigh, E. A. McrBC: a multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol.* **225**, 327–348 (1992).

46. Janosi, L., Yonemitsu, H., Hong, H. & Kaji, A. Molecular cloning and expression of a novel hydroxymethylcytosine-specific restriction enzyme (PvuRts1I) modulated by glucosylation of DNA. *J. Mol. Biol.* **242**, 45–61 (1994).

47. Gupta, R., Capalash, N. & Sharma, P. Restriction endonucleases: Natural and directed evolution. *Appl. Microbiol. Biotechnol.* **94**, 583–599 (2012).

48. Goldfarb, T. *et al.* BREX is a novel phage resistance system widespread in microbial genomes . *EMBO J.* **34**, 169–183 (2015).

49. Ofir, G., Melamed, S., Sberro, H., Mukamel, Z. & Silverman, S. DISARM is a widespread bacterial defence system with broad anti-phage activities. *Nat. Microbiol.* **3**, 90–98 (2018).

50. Chopin, M. C., Chopin, A. & Bidnenko, E. Phage abortive infection in lactococci: Variations on a theme. *Curr. Opin. Microbiol.* **8**, 473–479 (2005).

51. Page, R. & Peti, W. Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nat. Chem. Biol.* **12**, 208–214 (2016).

52. Dy, R. L., Przybilski, R., Semeijn, K., Salmond, G. P. C. & Fineran, P. C. A widespread bacteriophage abortive infection system functions through a Type IV toxin-antitoxin mechanism. *Nucleic Acids Res.* **42**, 4590–4605 (2014).

53. Amitai, S., Yassin, Y. & Engelberg-Kulka, H. MazF-mediated cell death in *Escherichia coli*: A point of no return. *J. Bacteriol.* **186**, 8295–8300 (2004).

54. Otsuka, Y. & Yonesaki, T. Dmd of bacteriophage T4 functions as an antitoxin against *Escherichia coli* LsoA and RnlA toxins. *Mol. Microbiol.* **83**, 669–681 (2012).

55. Severin, G. B. *et al.* A broadly conserved deoxycytidine deaminase protects bacteria from phage infection. *bioRxiv* (2021) doi:10.1101/2021.03.31.437871.

56. Tal, N. *et al.* Antiviral defense via nucleotide depletion in bacteria. *bioRxiv* 2021.04.26.441389 (2021).

57. Carpena, N., Manning, K. A., Dokland, T., Marina, A. & Penadés, J. R. Convergent evolution of pathogenicity islands in helper cos phage interference. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, (2016).

58. Martínez-Rubio, R. *et al.* Phage-inducible islands in the Gram-positive cocci. *ISME J.* **11**, 1029–1042 (2017).

59. Ketting, R. F. The many faces of RNAi. *Dev. Cell* **20**, 148–161 (2011).

60. Willkomm, S., Makarova, K. S. & Grohmann, D. DNA silencing by prokaryotic Argonaute proteins adds a new layer of defense against invading nucleic acids. *FEMS Microbiol. Rev.* **42**, 376–387 (2018).

61. Hegge, J. W., Swarts, D. C. & Van Der Oost, J. Prokaryotic argonaute proteins: Novel genome-editing tools? *Nat. Rev. Microbiol.* **16**, 5–11 (2018).

62. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507**, 258–261 (2014).

63. Zander, A. *et al.* Guide-independent DNA cleavage by archaeal Argonaute from *Methanocaldococcus jannaschii*. *Nat. Microbiol.* **2**, 17034 (2017).

64. Olovnikov, I., Chan, K., Sachidanandam, R., Newman, D. K. & Aravin, A. A. Bacterial Argonaute samples the transcriptome to identify foreign DNA. *Mol. Cell* **51**, 594–605 (2013).

65. Koopal, B. *et al.* Short prokaryotic Argonaute systems trigger cell death upon detection of invading DNA. *Cell* **185**, 1471-1486.e19 (2022).

66. Kronheim, S. *et al.* A chemical defence against phage infection. *Nature* **564**, 283–286 (2018).

67. Tal, N. & Sorek, R. SnapShot: Bacterial immunity. *Cell* **185**, 578-578.e1 (2022).

68. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakatura, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in *Escherichia coli*, and identification of the gene product. *J. Bacteriol.* **169**, 5429–5433 (1987).

69. Mojica, F. J. M., Díez-Villaseñor, C., Soria, E. & Juez, G. Biological significance of a family of regularly spaced repeats in the genomes of *Archaea*, *Bacteria* and mitochondria. *Mol. Microbiol.* **36**, 244–246 (2000).

70. Jansen, R., Van Embden, J. D. A., Gaastra, W. & Schouls, L. M. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol. Microbiol.* **43**, 1565–1575 (2002).

71. Tang, T.-H. *et al.* Identification of 86 candidates for small non-messenger RNAs from the archaeon *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci.* **99**, 7536–7541 (2002).

72. Mojica, F. J. M., Díez-Villasenõr, C., García-Martínez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**, 174–182 (2005).

73. Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* **151**, 653–663 (2005).

74. Bolotin, A., Quinquis, B., Sorokin, A. & Dusko Ehrlich, S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551–2561 (2005).

75. Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **1**, 1–26 (2006).

76. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science.* **315**, 1709–1712 (2007).

77. Brouns, S. J. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science.* **321**, 960–964 (2008).

78. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science.* **322**, 1843–1845 (2008).

79. Makarova, K. S. *et al.* Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).

80. Karginov, F. V & Hannon, G. J. The CRISPR system: small RNA-guided defense in *Bacteria* and *Archaea*. *Mol Cell* **37**, 1–7 (2010).

81. Alkhnbashi, O. S. *et al.* Characterizing leader sequences of CRISPR loci. *Bioinformatics* **32**, i576–i585 (2016).

82. Wei, Y., Chesne, M. T., Terns, R. M. & Terns, M. P. Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in *Streptococcus thermophilus*. *Nucleic Acids Res.* **43**, 1749–1758 (2015).

83. Makarova, K. S., Aravind, L., Grishin, N. V., Rogozin, I. B. & Koonin, E. V. A DNA repair system specific for thermophilic *Archaea* and *Bacteria* predicted by genomic context analysis. *Nucleic Acids Res.* **30**, 482–496 (2002).

84. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat. Rev. Microbiol.* **13**, 722–736 (2015).

85. Jore, M. M., Brouns, S. J. J. & van der Oost, J. RNA in defense: CRISPRs protect prokaryotes against mobile genetic elements. *Cold Spring Harb. Perspect. Biol.* **4**, 1–12 (2012).

86. McGinn, J. & Marraffini, L. A. Molecular mechanisms of CRISPR-Cas spacer acquisition. *Nat. Rev. Microbiol.* **17**, 7–12 (2019).

87. Nuñez, J. K. *et al.* Cas1-Cas2 complex formation mediates spacer acquisition during CRISPR-Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **21**, 528–534 (2014).

88. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comparative genomics of defense systems in *Archaea* and *Bacteria*. *Nucleic Acids Res.* **41**, 4360–4377 (2013).

89. Deveau, H. *et al.* Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.* **190**, 1390–1400 (2008).

90. Wang, J. *et al.* Structural and mechanistic basis of PAM-dependent spacer acquisition in CRISPR-Cas systems. *Cell* **163**, 840–853 (2015).

91. Babu, M. *et al.* A dual function of the CRISPR-Cas system in bacterial antivirus immunity and DNA repair. *Mol. Microbiol.* **79**, 484–502 (2011).

92. Kim, T. Y., Shin, M., Huynh Thi Yen, L. & Kim, J. S. Crystal structure of Cas1 from *Archaeoglobus fulgidus* and characterization of its nucleolytic activity. *Biochem. Biophys. Res. Commun.* **441**, 720–725 (2013).

93. Beloglazova, N. *et al.* A novel family of sequence-specific endoribonucleases associated with the clustered regularly interspaced short palindromic repeats. *J. Biol. Chem.* **283**, 20361–20371 (2008).

94. Nam, K. H. *et al.* Double-stranded endonuclease activity in *Bacillus halodurans* clustered regularly interspaced short palindromic repeats (CRISPR)-associated Cas2 protein. *J. Biol. Chem.* **287**, 35943–35952 (2012).

95. Samai, P., Smith, P. & Shuman, S. Structure of a CRISPR-associated protein Cas2 from *Desulfovibrio vulgaris*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **66**, 1552–1556 (2010).

96. Hooton, S. P. T. & Connerton, I. F. *Campylobacter jejuni* acquire new host-derived CRISPR spacers when in association with bacteriophages harboring a CRISPR-like Cas4 protein. *Front. Microbiol.* **6**, 1–9 (2015).

97. Nuñez, J. K., Bai, L., Harrington, L. B., Hinder, T. L. & Doudna, J. A. CRISPR Immunological Memory Requires a Host Factor for Specificity. *Mol. Cell* **62**, 824–833 (2016).

98. Wiedenheft, B. *et al.* Structural basis for DNase activity of a conserved protein implicated in CRISPR-mediated genome defense. *Structure* **17**, 904–912 (2009).

99. Amitai, G. & Sorek, R. CRISPR-Cas adaptation: Insights into the mechanism of action. *Nat. Rev. Microbiol.* **14**, 67–76 (2016).

100. Nuñez, J. K., Lee, A. S. Y., Engelman, A. & Doudna, J. A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193–198 (2015).

101. Rollie, C., Schneider, S., Brinkmann, A. S., Bolt, E. L. & White, M. F. Intrinsic sequence specificity of the Cas1 integrase directs new spacer acquisition. *Elife* **4**, e08716 (2015).

102. Hochstrasser, M. L. & Doudna, J. A. Cutting it close: CRISPR-associated endoribonuclease structure and function. *Trends Biochem. Sci.* **40**, 58–66 (2015).

103. Nidhi, S. *et al.* Novel CRISPR-Cas systems: An updated review of the current achievements, applications, and future research perspectives. *Int. J. Mol. Sci.* **22**, 1–42 (2021).

104. Wang, R., Preamplume, G., Terns, M. P., Terns, R. M. & Li, H. Interaction of the Cas6 riboendonuclease with CRISPR RNAs: recognition and cleavage. *Structure* **19**, 257–264 (2011).

105. Deltcheva, E. *et al.* CRISPR RNA maturation by *trans*-encoded small RNA and host factor RNase III. *Nature* **471**, 602–607 (2011).

106. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733–740 (2009).

107. Shah, S. A., Erdmann, S., Mojica, F. J. M. & Garrett, R. A. Protospacer recognition motifs: Mixed identities and functional diversity. *RNA Biol.* **10**, 891–899 (2013).

108. Heler, R., Marraffini, L. A. & Bikard, D. Adapting to new threats: The generation of memory by CRISPR-Cas immune systems. *Mol. Microbiol.* **93**, 1–9 (2014).

109. Gleditzsch, D. *et al.* PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA Biol.* **16**, 504–517 (2019).

110. Hille, F. *et al.* The biology of CRISPR-Cas: backward and forward. *Cell* **172**, 1239–1259 (2018).

111. Makarova, K. S. *et al.* Evolution and classification of the CRISPR-Cas systems. *Nat. Rev. Microbiol.* **9**, 467–477 (2011).

112. Sinkunas, T. *et al.* Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. *EMBO J.* **30**, 1335–1342 (2011).

113. Gong, B. *et al.* Molecular insights into DNA interference by CRISPR-associated nuclease-helicase Cas3. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 16359–16364 (2014).

114. Huo, Y. *et al.* Structures of CRISPR Cas3 offer mechanistic insights into Cascade-activated DNA unwinding and degradation. *Nat. Struct. Mol. Biol.* **21**, 771–777 (2014).

115. Charpentier, E., Richter, H., van der Oost, J. & White, M. F. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* **39**, 428–441 (2015).

116. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science.* **357**, 605–609 (2017).

117. Niewoehner, O. *et al.* Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548**, 543–548 (2017).

118. Faure, G. *et al.* CRISPR-Cas in mobile genetic elements: counter-defence and beyond. *Nat. Rev. Microbiol.* (2019) doi:10.1038/s41579-019-0204-7.

119. Faure, G., Makarova, K. S. & Koonin, E. V. CRISPR-Cas: Complex functional networks and multiple roles beyond adaptive immunity. *J. Mol. Biol.* **431**, 3–20 (2019).

120. Özcan, A. *et al.* Type IV CRISPR RNA processing and effector complex formation in *Aromatoleum aromaticum*. *Nat. Microbiol.* **4**, 89–96 (2019).

121. Zhou, Y. *et al.* Structure of a type IV CRISPR-Cas ribonucleoprotein complex. *iScience* **24**, 102201 (2021).

122. Heler, R. *et al.* Cas9 specifies functional viral targets during CRISPR-Cas adaptation. *Nature* **519**, 199–202 (2015).

123. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6105 (2014).

124. Briner, A. E. & Barrangou, R. Guide RNAs: A glimpse at the sequences that drive CRISPR-Cas systems. *Cold Spring Harb. Protoc.* **2016**, 594–600 (2016).

125. Faure, G. *et al.* Comparative genomics and evolution of *trans*-activating RNAs in Class 2 CRISPR-Cas systems. *RNA Biol.* **16**, 435–448 (2019).

126. Chyou, T. yuan & Brown, C. M. Prediction and diversity of tracrRNAs from type II CRISPR-Cas systems. *RNA Biol.* **16**, 423–434 (2019).

127. Jiang, F. *et al.* Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science.* **351**, 867–871 (2016).

128. Jinek, M. *et al.* Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science.* **343**, (2014).

129. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).

130. Nishimasu, H. *et al.* Crystal structure of *Staphylococcus aureus* Cas9. *Cell* **162**, 1113–1126 (2015).

131. Garrett, R. A., Vestergaard, G. & Shah, S. A. Archaeal CRISPR-based immune systems: Exchangeable functional modules. *Trends Microbiol.* **19**, 549–556 (2011).

132. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517–521 (2016).

133. Swarts, D. & Jinek, M. Mechanistic insights into the *cis*- and *trans*-acting DNase activities of Cas12a. *Mol. Cell* **73**, 589-600.e4 (2018).

134. Pausch, P. *et al.* CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science.* **369**, 333–337 (2020).

135. Pausch, P., Soczek, K. M., Herbst, D. A. & Al-shayeb, B. DNA interference states of the hypercompact CRISPR-CasΦ effector. *Nat. Struct. Mol. Biol.* **28**, 1–40 (2021).

136. East-Seletsky, A. *et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide RNA processing and RNA detection. *Nature* **538**, 270–273 (2016).

137. Liu, L. *et al.* Two distant catalytic sites are responsible for C2c2 RNase activities. *Cell* **168**, 121–134 (2017).

138. Deveau, H., Garneau, J. E. & Moineau, S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu. Rev. Microbiol.* **64**, 475–493 (2010).

139. Bondy-Denomy, J., Pawluk, A., Maxwell, K. L. & Davidson, A. R. Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**, 429–432 (2013).

140. Bondy-Denomy, J. *et al.* Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. *Nature* **526**, 136–139 (2015).

141. Trasanidou, D. *et al.* Keeping CRISPR in check: Diverse mechanisms of phage-encoded anti-CRISPRs. *FEMS Microbiol. Lett.* **366**, 1–14 (2019).

142. Mohanraju, P. *et al.* Alternative functions of CRISPR–Cas systems in the evolutionary arms race. *Nat. Rev. Microbiol.* (2022) doi:10.1038/s41579-021-00663-z.

143. Thavalingam, A. *et al.* Inhibition of CRISPR-Cas9 ribonucleoprotein complex assembly by anti-CRISPR AcrIIC2. *Nat. Commun.* **10**, 1–11 (2019).

144. Hirschi, M. *et al.* AcrIF9 tethers non-sequence specific dsDNA to the CRISPR RNA-guided surveillance complex. *Nat. Commun.* **11**, 1–6 (2020).

145. Lu, W. T., Trost, C. N., Müller-Esparza, H., Randau, L. & Davidson, A. R. Anti-CRISPR AcrIF9 functions by inducing the CRISPR-Cas complex to bind DNA non-specifically. *Nucleic Acids Res.* **49**, 3381–3393 (2021).

146. Osuna, B. A. *et al.* Listeria phages induce Cas9 degradation to protect lysogenic genomes. *Cell Host Microbe* **28**, 31-40.e9 (2020).

147. Athukoralage, J. S. *et al.* An anti-CRISPR viral ring nuclease subverts type III CRISPR immunity. vol. 577 (2020).

148. Osuna, B. A. *et al.* Critical anti-CRISPR locus repression by a bi-functional Cas9 inhibitor. *Physiol. Behav.* **176**, 139–148 (2017).

149. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science.* **337**, 816–821 (2012).

150. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).

151. Mali, P. *et al.* RNA-Guided human genome engineering via Cas9. 823–827 (2013).

152. Scully, R., Panday, A., Elango, R. & Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **20**, 698–714 (2019).

153. Li, Y., Glass, Z., Huang, M., Chen, Z.-Y. & Xu, Q. *Ex vivo* cell-based CRISPR/Cas9 genome editing for therapeutic applications. *Biomaterials.* **234**, 119711 (2020).

154. Ceccaldi, R., Rondinelli, B. & D'Andrea, A. D. Repair pathway choices and consequences at the double-strand break. *Trends Cell Biol.* **26**, 52–64 (2016).

155. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).

156. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).

157. Kleinstiver, B. P. *et al.* Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. *Nat. Biotechnol.* **33**, 1293–1298 (2015).

158. Gao, L. *et al.* Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.* **35**, 789–792 (2017).

159. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).

160. Fu, Y. *et al.* High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.* **31**, 822–826 (2013).

161. Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. & Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62–67 (2014).

162. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res.* **24**, 132–141 (2014).

163. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science.* **351**, 84–88 (2016).

164. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

165. Casini, A. *et al.* A highly specific SpCas9 variant is identified by *in vivo* screening in yeast. *Nat. Biotechnol.* **36**, 265–271 (2018).

166. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407–410 (2017).

167. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).

168. Kim, Y. B. *et al.* Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).

169. Miao, J. *et al.* Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* **23**, 1233–1236 (2013).

170. Shan, Q., Wang, Y., Li, J. & Gao, C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* **9**, 2395–2410 (2014).

171. Ma, X. *et al.* A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Mol. Plant* **8**, 1274–1284 (2015).

172. Jiang, W. Z. *et al.* Significant enhancement of fatty acid composition in seeds of the allohexaploid, *Camelina sativa*, using CRISPR/Cas9 gene editing. *Plant Biotechnol. J.* **15**, 648–657 (2017).

173. Sánchez-León, S. *et al.* Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant Biotechnol. J.* **16**, 902–910 (2018).

174. Aznar-Moreno, J. A. & Durrett, T. P. Simultaneous targeting of multiple gene homeologs to alter seed oil production in *Camelina sativa*. *Plant Cell Physiol.* **58**, 1260–1267 (2017).

175. Sun, Y. *et al.* Generation of high-amylose rice through CRISPR/Cas9-mediated targeted mutagenesis of starch branching enzymes. *Front. Plant Sci.* **8**, 1–15 (2017).

176. Aman, R. *et al.* RNA virus interference via CRISPR/Cas13a system in plants. *Genome Biol.* **19**, (2018).

177. Khan, M. Z., Amin, I., Hameed, A. & Mansoor, S. CRISPR–Cas13a: Prospects for plant virus resistance. *Trends Biotechnol.* **36**, 1207–1210 (2018).

178. Shimatani, Z. *et al.* Targeted base editing in rice and tomato using a CRISPR-Cas9 cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 441–443 (2017).

179. Chen, Y. *et al.* CRISPR/Cas9-mediated base-editing system efficiently generates gain-of-function mutations in Arabidopsis. *Sci. China Life Sci.* **60**, 520–523 (2017).

180. Tian, S. *et al.* Engineering herbicide-resistant watermelon variety through CRISPR/Cas9-mediated base-editing. *Plant Cell Rep.* **37**, 1353–1356 (2018).

181. Min, Y. L. *et al.* CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. *Sci. Adv.* **5**, 1–13 (2019).

182. Bjursell, M. *et al.* Therapeutic genome editing with CRISPR/Cas9 in a humanized mouse model ameliorates α1-antitrypsin deficiency phenotype. *EBioMedicine* **29**, 104–111 (2018).

183. Ohmori, T., Mizukami, H., Ozawa, K., Sakata, Y. & Nishimura, S. New approaches to gene and cell therapy for hemophilia. *J. Thromb. Haemost.* **13**, S133–S142 (2015).

184. Khosravi, M. A. *et al.* Targeted deletion of BCL11A gene by CRISPR-Cas9 system for fetal hemoglobin reactivation: A promising approach for gene therapy of beta thalassemia disease. *Eur. J. Pharmacol.* **854**, 398–405 (2019).

185. Frangoul, H. *et al.* CRISPR-Cas9 gene editing for sickle cell disease and β-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).

186. György, B. *et al.* Allele-specific gene editing prevents deafness in a model of dominant progressive hearing loss. *Nat Med* **25**, 1123–1130 (2019).

187. Craigie, R. & Bushman, F. D. HIV DNA integration. *Cold Spring Harb. Perspect. Med.* **2**, a006890 (2012).

188. Hu, W. *et al.* RNA-directed gene editing specifically eradicates latent and prevents new HIV-1 infection. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11461–11466 (2014).

189. Rusconi, S. & Giacomelli, A. CRISPR in HIV: Dangers of CCR5 deletion. *Future Virol.* **15**, 207–209 (2020).

190. Cyranoski, D. The CRISPR-baby scandal: what's next for human gene-editing. *Nature* **566**, 440–442 (2019).

191. Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722–725 (1996).

192. Locke, L. G. The promise of CRISPR for human germline editing and the perils of 'playing God'. *Cris. J.* **3**, 27–31 (2020).

193. Schleidgen, S. *et al.* Human germline editing in the era of CRISPR-Cas: Risk and uncertainty, inter-generational responsibility, therapeutic legitimacy. *BMC Med. Ethics* **21**, 1–12 (2020).

194. Furtado, R. N. Gene editing: the risks and benefits of modifying human DNA. *Rev. Bioética* **27**, 223–233 (2019).

195. Huang, C. H., Lee, K. C. & Doudna, J. A. Applications of CRISPR-Cas enzymes in cancer therapeutics and detection. *Trends in Cancer* **4**, 499–512 (2018).

196. Ding, X. *et al.* Ultrasensitive and visual detection of SARS-CoV-2 using all-in-one dual CRISPR-Cas12a assay. *Nat. Commun.* **11**, 1–10 (2020).

197. Hou, T. *et al.* Development and evaluation of a rapid CRISPR-based diagnostic for COVID-19. *PLoS Pathog.* **16**, 1–12 (2020).

198. Jiang, Y. *et al.* Detection of SARS-CoV-2 by CRISPR/Cas12a-enhanced colorimetry. *ACS Sensors* **6**, 1086–1093 (2021).

199. Broughton, J. P. *et al.* Rapid detection of 2019 novel coronavirus SARS-CoV-2 using a CRISPR-based DETECTR lateral flow assay. *medRxiv* (2020) doi:10.1101/2020.03.06.20032334.

200. Li, S. *et al.* A one-step, one-pot CRISPR nucleic acid detection platform (CRISPR-top): Application for the diagnosis of COVID-19. *Talanta* **233**, 122591 (2021).

201. Yang, Y., Liu, J. & Zhou, X. A CRISPR-based and post-amplification coupled SARS-CoV-2 detection with a portable evanescent wave biosensor. *Biosens. Bioelectron.* **190**, 113418 (2021).

202. Aquino-Jarquin, G. Recent progress on rapid SARS-CoV-2/COVID-19 detection by CRISPR-Cas13-based platforms. *Drug Discov. Today* **26**, 2025–2035 (2021).

203. Patchsung, M. *et al.* Clinical validation of a Cas13-based assay for the detection of SARS-CoV-2 RNA. *Nat. Biomed. Eng.* **4**, 1140–1149 (2020).

204. Arizti-Sanz, J. *et al.* Streamlined inactivation, amplification, and Cas13-based detection of SARS-CoV-2. *Nat. Commun.* **11**, (2020).

205. Wang, Y. *et al.* Detection of SARS-CoV-2 and its mutated variants via CRISPR-Cas13-based transcription amplification. *Anal. Chem.* **93**, 3393–3402 (2021).

206. Wang, R. *et al.* opvCRISPR: One-pot visual RT-LAMP-CRISPR platform for SARS-cov-2 detection. *Biosens. Bioelectron.* **172**, 112766 (2021).

207. Broughton, J. P. *et al.* CRISPR-Cas12-based detection of SARS-CoV-2. *Nat. Biotechnol.* **38**, 870–874 (2020).

208. Nouri, R. *et al.* CRISPR-based detection of SARS-CoV-2: A review from sample to result. *Biosens. Bioelectron.* **178**, (2021).

209. Wang, X. *et al.* Rapid and sensitive detection of COVID-19 using CRISPR/Cas12a-based detection with naked eye readout, CRISPR/Cas12a-NER. *Sci. Bull.* **65**, 1436–1439 (2020).

210. Huang, Z. *et al.* Ultra-sensitive and high-throughput CRISPR-powered COVID-19 diagnosis. *Biosens. Bioelectron.* **164**, 112316 (2020).

211. Lucia, C., Alejandara, G. C. & Federico, P.-B. An ultrasensitive, rapid, and portable coronavirus SARS-CoV-2 sequence detection method based on CRISPR-Cas12. *bioRxiv* (2020).

212. Xiang, X. *et al.* CRISPR-Cas systems based molecular diagnostic tool for infectious diseases and emerging 2019 novel coronavirus (COVID-19) pneumonia. *J. Drug Target.* **28**, 727–731 (2020).

213. Liang, Y. *et al.* CRISPR-Cas12a-based detection for the major SARS-CoV-2 variants of concern. *Microbiol. Spectr.* **9**, (2021).

214. Zetsche, B. *et al.* Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).

215. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR-Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).

216. Koonin, E. V. & Makarova, K. S. Mobile genetic elements and evolution of CRISPR-Cas systems: All the way there and back. *Genome Biol. Evol.* **9**, 2812–2825 (2017).

217. Dong, D. *et al.* The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532**, 522–526 (2016).

218. Yamano, T. *et al.* Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell* **165**, 949–962 (2016).

219. Yamano, T. *et al.* Structural basis for the canonical and non-canonical PAM recognition by CRISPR-Cpf1. *Mol. Cell* **67**, 633-645.e3 (2017).

220. Swarts, D. C., van der Oost, J. & Jinek, M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. *Mol. Cell* **66**, 221-233.e4 (2017).

221. Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546**, 559–563 (2017).

222. Jeon, Y. *et al.* Direct observation of DNA target searching and cleavage by CRISPR-Cas12a. *Nat. Commun.* **9**, (2018).

223. Specht, D. A., Xu, Y. & Lambert, G. Massively parallel CRISPRi assays reveal concealed thermodynamic determinants of dCas12a binding. *PNAS* **117**, 11274–11282 (2020).

224. Gao, P., Yang, H., Rajashankar, K. R., Huang, Z. & Patel, D. J. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.* **26**, 901–913 (2016).

225. Saha, A. *et al.* DNA-induced dynamic switch triggers activation of CRISPR-Cas12a. *J. Chem. Inf. Model.* **60**, 6427–6437 (2020).

226. Stella, S. *et al.* Conformational activation promotes CRISPR-Cas12a catalysis and resetting of the endonuclease activity. *Cell* **175**, 1856–1871 (2018).

227. van Aelst, K., Martínez-Santiago, C., Cross, S. & Szczelkun, M. The effect of DNA topology on observed rates of R-loop formation and DNA strand cleavage by CRISPR Cas12a. *Genes (Basel).* **10**, 169 (2019).

228. Zhang, L. *et al.* Conformational dynamics and cleavage sites of Cas12a are modulated by complementarity between crRNA and DNA. *iScience* **19**, 492–503 (2019).

229. Singh, D. *et al.* Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 (Cas12a). *Proc. Natl. Acad. Sci.* **115**, 5444–5449 (2018).

230. Singh, D., Sternberg, S. H., Fei, J., Doudna, J. A. & Ha, T. Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. *Nat. Commun.* **7**, 1–8 (2016).

231. Strohkendl, I., Saifuddin, F. A., Rybarski, J. R., Finkelstein, I. J. & Russell, R. Kinetic basis for DNA target specificity of CRISPR-Cas12a. *Mol. Cell* **71**, 816-824.e3 (2018).

232. Cofsky, J. C. *et al.* CRISPR-Cas12a exploits R-loop asymmetry to form double strand breaks. *Elife* **9**, 2020.02.10.937540 (2020).

233. Swartjes, T., Staals, R. H. J. & Van Der Oost, J. Editor's cut: DNA cleavage by CRISPR RNA-guided nucleases Cas9 and Cas12a. *Biochem. Soc. Trans.* **48**, 207–219 (2020).

234. Swarts, D. C. & Jinek, M. Cas9 versus Cas12a/Cpf1: Structure-function comparisons and implications for genome editing. *Wiley Interdiscip. Rev. RNA* **9**, 1–19 (2018).

235. Naqvi, M. M., Lee, L., Montaguth, O. E. T. & Szczelkun, M. D. A gate and clamp regulate sequential DNA strand cleavage by CRISPR-Cas12a. *bioRxiv Biochem.* (2021).

236. Chen, J. S. *et al.* CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science.* **360**, 436–439 (2018).

237. Fuchs, R. T., Curcuru, J., Mabuchi, M., Yourik, P. & Robb, G. B. Cas12a *trans*-cleavage can be modulated *in vitro* and is active on ssDNA, dsDNA, and RNA. *bioRxiv* (2019) doi:10.1101/600890.

238. Li, S. Y. *et al.* CRISPR-Cas12a has both *cis*- and *trans*-cleavage activities on single-stranded DNA. *Cell Res.* **28**, 491–493 (2018).

239. Li, S. Y. *et al.* CRISPR-Cas12a-assisted nucleic acid detection. *Cell Discov.* **4**, 18–21 (2018).

240. Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science.* **6387**, 439–444 (2018).

241. Varble, A. & Marraffini, L. A. Three new C's for CRISPR: Collateral, Communicate, Cooperate. *Trends Genet.* **35**, 446–456 (2019).

242. Zhang, H. *et al.* Structural basis for the inhibition of CRISPR-Cas12a by anti-CRISPR proteins. *Cell Host Microbe* **25**, 815-826.e4 (2019).

243. Knott, G. J. *et al.* Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nat. Struct. Mol. Biol.* **26**, 315–321 (2019).

244. Knott, G. J. *et al.* Structural basis for AcrVA4 inhibition of specific CRISPR-Cas12a. *Elife* **8**, 1–25 (2019).

245. Peng, R. *et al.* Structural insight into multistage inhibition of CRISPR-Cas12a by AcrVA4. *Proc. Natl. Acad. Sci.* **116**, 18928–18936 (2019).

246. Dong, L. *et al.* An anti-CRISPR protein disables type V Cas12a by acetylation. *Nat. Struct. Mol. Biol.* **26**, 308–314 (2019).

247. Schrimpf, W., Barth, A., Hendrix, J. & Lamb, D. C. PAM: A framework for integrated analysis of imaging, single-molecule, and ensemble fluorescence data. *Biophys. J.* **114**, 1518–1528 (2018).

248. Young, T. S., Ahmad, I., Yin, J. A. & Schultz, P. G. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. *J. Mol. Biol.* **395**, 361–374 (2010).

249. Chin, J. W. *et al.* Addition of *p*-azido-L-phenylalanine to the genetic code of *Escherichia coli*. *J. Am. Chem. Soc.* **124**, 9026–9027 (2002).

250. Hanahan, D., Jessee, J. & Bloom, F. R. Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol.* **204**, 63–113 (1991).

251. Meselson, M., Yuan, R. & Heywood, J. Restriction and modification of DNA. *Annu. Rev. Biochem.* **41**, 447–466 (1972).

252. Bachman, J. Site-directed mutagenesis. Methods in Enzymology. vol. 529 (Elsevier Inc., 2013).

253. Le Gouill, C., Parent, J. L., Rola-Pleszczynski, M. & Stankova, J. Analysis of recombinant plasmids by a modified alkaline lysis method. *Anal. Biochem.* **219**, 164 (1994).

254. Berg, J. M., Tymoczko, J. L., Gatto jr., G. J. & Stryer, L. *Stryer Biochemie*. *Springer Spektrum* vol. 8 (2018).

255. Grohmann, D., Werner, F. & Tinnefeld, P. Making connections – strategies for single-molecule fluorescence biophysics. *Curr. Opin. Chem. Biol.* **17**, 691–698 (2013).

256. Laemmli, U. K. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).

257. Ornstein, L. Disc Electrophoreses I – Background and theory. *Ann. New York Acad. Sci.* **121**, 321–349 (1962).

258. Davis, B. J. Disc Electrophoresis II – Method and application to human serum proteins. *Clin. Appl.* 404–427 (1962).

259. Prescher, J. A. & Bertozzi, C. R. Chemistry in living systems. *Nat. Chem. Biol.* **1**, 13–21 (2005).

260. Agard, N. J., Baskin, J. M., Prescher, J. A., Lo, A. & Bertozzi, C. R. A comparative study of bioorthogonal reactions with azides. *ACS Chem. Biol.* **1**, 644–648 (2006).

261. Saxon, E. & Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science.* **287**, 2007–2010 (2000).

262. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. & Gray, T. How to measure and predict the molar absorption-coefficient of a protein. *Protein Sci.* **4**, 2411–2423 (1995).

263. Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2**, 1849–1861 (2007).

264. Verschueren, H. Interference reflection microscopy in cell biology: methodology and applications. *J. Cell Sci.* **75**, 279–301 (1985).

265. Ortega-Arroyo, J. & Kukura, P. Interferometric scattering microscopy (iSCAT): New frontiers in ultrafast and ultrasensitive optical microscopy. *Phys. Chem. Chem. Phys.* **14**, 15625–15636 (2012).

266. Cole, D., Young, G., Weigel, A., Sebesta, A. & Kukura, P. Label-free single-molecule imaging with numerical-aperture-shaped interferometric scattering microscopy. *ACS Photonics* **4**, 211–216 (2017).

267. Young, G. *et al.* Quantitative mass imaging of single biological macromolecules. *Science.* **360**, 423–427 (2018).

268. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

269. Förster, T. Energiewanderung und Fluoreszenz. *Naturwissenschaften* **33**, 166–175 (1946).

270. Stryer, L. & Haugland, R. P. Energy transfer: a spectroscopic ruler. *Proc. Natl. Acad. Sci. U. S. A.* **58**, 719–726 (1967).

271. Gust, A. *et al.* A starting point for fluorescence-based single-molecule measurements in biomolecular research. *Molecules* **19**, 15824–15865 (2014).

272. Schulz, S. *et al.* TFE and Spt4/5 open and close the RNA polymerase clamp during the transcription cycle. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E1816–E1825 (2016).

273. Lottspeich, F. & Engels, J. W. *Bioanalytik*. (Springer Spektrum, Mannheim, Germany, 2012).

274. Stryer, L. Fluorescence energy transfer as a spectroscopic ruler. *Annu. Rev. Biochem.* **47**, 819–846 (1978).

275. Bunting, J. R. & Cathou, R. E. Energy transfer distance measurements in immunoglobulins. *J. Mol. Biol.* **77**, 223–235 (1973).

276. Matsumoto, S. & Hammes, G. G. Fluorescence energy transfer between ligand binding sites on aspartate transcarbamylase. *Biochemistry* **14**, 214–224 (1975).

277. Katchalski-Katzir, E., Haast, E. & Steinberg, I. Z. Study of conformation and intermolecular motility of polypeptides in solution by a novel fluorescence method. *Ann. New York Acad. Sci.* **366**, 44–61 (1981).

278. Ha, T. *et al.* Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor. *Proc. Natl. Acad. Sci.* **93**, 6264–6268 (1996).

279. Muschielok, A. *et al.* A nano-positioning system for macromolecular structural analysis. *Nat. Methods* **5**, 965–971 (2008).

280. Schuler, B. & Hofmann, H. Single-molecule spectroscopy of protein folding dynamics – expanding scope and timescales. *Curr. Opin. Struct. Biol.* **23**, 36–47 (2013).

281. Abbondanzieri, E. A., Greenleaf, W. J., Shaevitz, J. W., Landick, R. & Block, S. M. Direct observation of base-pair stepping by RNA polymerase. *Nature* **438**, 460–465 (2005).

282. Seidel, R. & Dekker, C. Single-molecule studies of nucleic acid motors. *Curr. Opin. Struct. Biol.* **17**, 80–86 (2007).

283. Holzmeister, P., Acuna, G. P., Grohmann, D. & Tinnefeld, P. Breaking the concentration limit of optical single-molecule detection. *Chem. Soc. Rev.* **43**, 1014–1028 (2014).

284. Nie, S., Chiu, D. T. & Zare, R. N. Probing individual molecules with confocal fluorescence microscopy. *Science.* **266**, 1018–1021 (1994).

285. Deniz, A. A. *et al.* Single-pair fluorescence resonance energy transfer on freely diffusing molecules: Observation of Förster distance dependence and subpopulations. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 3670–3675 (1999).

286. Kramm, K. O. Single-molecule analysis of transcription initiation in archaea and eukaryotes. *Dissertation*, University Regensburg (2020).

287. Müller, B. K., Zaychikov, E., Bräuchle, C. & Lamb, D. C. Pulsed interleaved excitation. *Biophys. J.* **89**, 3508–3522 (2005).

288. Nir, E. *et al.* Shot-Noise Limited Single-Molecule FRET Histograms: Comparison between Theory and Experiments. *J. Phys. Chem. B* **110**, 22103–22124 (2006).

289. Torella, J. P., Holden, S. J., Santoso, Y., Hohlbein, J. & Kapanidis, A. N. Identifying molecular dynamics in single-molecule FRET experiments with burst variance analysis. *Biophys. J.* **100**, 1568–1577 (2011).

290. Wörle, E., Jakob, L., Schmidbauer, A., Zinner, G. & Grohmann, D. Decoupling the bridge helix of Cas12a results in a reduced trimming activity, increased mismatch sensitivity and impaired conformational transitions. *Nucleic Acids Res.* **49**, 5278–5293 (2021).

291. Nishimasu, H. *et al.* Structural basis for the altered PAM recognition by engineered CRISPR-Cpf1. *Mol. Cell* **67**, 139-147.e2 (2017).

292. Bratovič, M. *et al.* Bridge helix arginines play a critical role in Cas9 sensitivity to mismatches. *Nat. Chem. Biol.* **16**, 587–595 (2020).

293. Babu, K. *et al.* Bridge helix of Cas9 modulates target DNA cleavage and mismatch tolerance. *Biochemistry* **58**, 1905–1917 (2019).

294. Wörle, E., Newman, A., Burgio, G. & Grohmann, D. Allosteric activation of CRISPR-Cas12a requires the concerted movement of the bridge helix and helix 1 of the RuvC II domain. *bioRxiv* (2022) doi:https://doi.org/10.1101/2022.03.15.484427.

295. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Physiol. Behav.* **63**, 1–18 (2014).

296. Hirano, H. *et al.* Structure and Engineering of *Francisella novicida* Cas9. *Cell* **164**, 950–961 (2016).

297. Yamada, M. *et al.* Crystal structure of the minimal Cas9 from *Campylobacter jejuni* reveals the molecular diversity in the CRISPR-Cas9 systems. *Mol. Cell* **65**, 1109-1121.e3 (2017).

298. Zeng, Y. *et al.* The initiation, propagation and dynamics of CRISPR-SpyCas9 R-loop complex. *Nucleic Acids Res.* **46**, 350–361 (2018).

299. Shams, A. *et al.* Comprehensive deletion landscape of CRISPR-Cas9 identifies minimal RNA-guided DNA-binding modules. *Nat. Commun.* **12**, (2021).

300. Parameshwaran, H. P. *et al.* The bridge helix of Cas12a imparts selectivity in *cis*-DNA cleavage and regulates *trans*-DNA cleavage. *FEBS Lett.* **595**, 892–912 (2021).

301. Murugan, K., Seetharam, A. S., Severin, A. J. & Sashital, D. G. CRISPR-Cas12a has widespread off-target and dsDNA-nicking effects. *J. Biol. Chem.* **295**, 5538–5553 (2020).

302. Huang, C. J. *et al.* A naturally DNase-free CRISPR-Cas12c enzyme silences gene expression. *Mol. Cell* **82**, 2148-2160.e4 (2022).

303. Xu, X., Hulshoff, M. S., Tan, X., Zeisberg, M. & Zeisberg, E. M. Crispr-Cas derivatives as novel gene modulating tools: Possibilities and *in vivo* applications. *Int. J. Mol. Sci.* **21**, (2020).

304. Gardiner, J., Ghoshal, B., Wang, M. & Jacobsen, S. E. CRISPR-Cas-mediated transcriptional control and epi-mutagenesis. *Plant Physiol.* **188**, 1811–1824 (2022).

305. Adli, M. The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, (2018).

306. Ling, X. *et al.* Improving the efficiency of CRISPR-Cas12a-based genome editing with site-specific covalent Cas12a-crRNA conjugates. *Mol. Cell* 1–10 (2021) doi:10.1016/j.molcel.2021.09.021.

307. Liu, P. *et al.* Enhanced Cas12a editing in mammalian cells and zebrafish. *Nucleic Acids Res.* **47**, 4169–4180 (2019).

308. Kleinstiver, B. P. *et al.* Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).

309. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, (2011).

LIST OF FIGURES

Figure 1: The three stages of CRISPR-Cas immune response. 15
Figure 2: The modular organization of CRISPR-Cas systems......................................17
Figure 3: Structural organization of FnCas12a. 22
Figure 4: Acid-base reaction mechanism of pre-crRNA cleavage of FnCas12a. 23
Figure 5: Pseudoknot formation in the crRNA of Cas12a. 24
Figure 6: Conformational transitions of the structural element 'lid' in Cas12a. 25
Figure 7: Electrophoresis ladders. 38
Figure 8: Incorporation of an unnatural amino acid in an *E. coli* expression system......55
Figure 9: Staudinger-Bertozzi ligation of azides. 59
Figure 10: The principle of mass photometry. 66
Figure 11: The principle of FRET and the Förster radius. 68
Figure 12: Spectral overlap of donor emission and acceptor excitation..........................69
Figure 13: Confocal microscope setup..72
Figure 14: Correction factor determination for confocal single-molecule FRET data. 76
Figure 15: Purification and plasmid cleavage assay of FnCas12a. 78
Figure 16: EMSA of FnCas12a-crRNA-target DNA complexes. 79
Figure 17: Cleavage assay of WT FnCas12a and dCas12a on a short target DNA. 79
Figure 18: Cleavage kinetics of WT FnCas12a on a short doubly labeled target DNA. ...81
Figure 19: Cleavage assay of WT FnCas12a using a single-stranded target strand. 82
Figure 20: Cleavage assay of WT FnCas12a and the HKK variant using pre-crRNA as substrate...83
Figure 21: Single-molecule FRET measurements of an RNA/DNA hybrid with WT FnCas12a. 84
Figure 22: Labeling positions in FnCas12a. 85
Figure 23: Activity assays of FnCas12a variants carrying two unnatural amino acid residues. 86
Figure 24: SDS-PAGE analysis of doubly labeled FnCas12a variants. 87
Figure 25: Confocal single-molecule FRET measurements on doubly labeled WT FnCas12a. 88
Figure 26: Burst variance analysis of WT FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$. 89
Figure 27: Confocal single-molecule FRET measurements using doubly labeled FnCas12a RNase-inactive variants...90
Figure 28: Confocal single-molecule FRET measurements using a doubly labeled FnCas12a DNase-inactive variant. 91
Figure 29: Positions of mutations to alter the target strand affinity. 92

Figure 30: Purification and cleavage assay of FnCas12a variants to alter the target strand affinity. 93
Figure 31: Confocal single-molecule FRET measurements using doubly labeled FnCas12a variants to alter the target strand affinity. 94
Figure 32: Positions of introduced mutations in the bridge helix of FnCas12a. 95
Figure 33: Purification and melting curves of FnCas12a bridge helix variants. 96
Figure 34: Plasmid cleavage assay of FnCas12a bridge helix variants. 97
Figure 35: Binding behavior to crRNA and pre-crRNA processing of FnCas12a bridge helix variants. 98
Figure 36: Mass photometry of FnCas12a WT and FnCas12a WT with crRNA. 99
Figure 37: Activity assays of FnCas12a bridge helix variants using short double-stranded DNA as target. 100
Figure 38: High-resolution target DNA cleavage assays of FnCas12a bridge helix variants. 102
Figure 39: DNA cleavage specificity of FnCas12a bridge helix variants at mismatched substrates. 104
Figure 40: Confocal single-molecule FRET measurements using doubly labeled FnCas12a bridge helix variants. 107
Figure 41: Quantification of single-molecule FRET measurements of FnCas12a bridge helix variants. 107
Figure 42: Positions of mutations introduced into helix 1 of FnCas12a. 108
Figure 43: Purification, melting curves, and CD spectra of FnCas12a helix 1 variants. 109
Figure 44: Plasmid cleavage assay of FnCas12a helix 1 variants. 110
Figure 45: Time-resolved cleavage assay of selected FnCas12a bridge helix and helix 1 variants. 112
Figure 46: Complex formation of FnCas12a helix 1 variants and crRNA and pre-crRNA processing of FnCas12a helix 1 variants. 113
Figure 47: Activity assays of FnCas12a helix 1 variants using short double-stranded target DNA. 115
Figure 48: High-resolution cleavage assays of FnCas12a helix 1 variants. 116
Figure 49: DNA cleavage specificity of FnCas12a helix 1 variants at mismatched substrates. 117
Figure 50: *Trans*-cleavage of FnCas12a helix 1 variants in comparison to FnCas12a bridge helix variants. 118
Figure 51: Confocal single-molecule FRET measurements using doubly labeled FnCas12a helix 1 variants. 120
Figure 52: The conformational transitions of FnCas12a during its activity cycle. 125

Figure 53: Comparison of Cas9 and Cas12a crystal structures. 126
Figure 54: Tandem movement of bridge helix and helix 1. 130
Figure 55: Mechanistic model summarizing the impact of helix 1 on the activity of FnCas12a. 133
Figure 56: Vector pEVOL-pAzF. 158
Figure 57: Vector pGEX-2TK. 158
Figure 58: Plasmid pY002 (pFnCpf1_min). 159
Figure 59: Plasmid EGFP-hAgo2. 159
Figure 60: Multiple sequence alignments of Cas12a variants from different organisms. 165
Figure 61: Example for the purification of FnCas12a. 166
Figure 62: FRET efficiency histograms with standard deviation for measurements of the RNA/DNA hybrid with WT FnCas12a. 167
Figure 63: Confocal single-molecule FRET measurements using doubly labeled WT FnCas12a with different RNA/DNA concentrations. 169
Figure 64: FRET efficiency histograms with standard deviation for measurements of doubly labeled WT FnCas12a. 171
Figure 65: FRET efficiency histograms with standard deviation for measurements of doubly labeled FnCas12a HKK variants. 173
Figure 66: Confocal single-molecule FRET measurements using doubly labeled dCas12a with different RNA/DNA concentrations. 174
Figure 67: FRET efficiency histograms with standard deviation for measurements of doubly labeled dCas12a. 175
Figure 68: Confocal single-molecule FRET measurements using doubly labeled FnCas12a variants to alter the target strand affinity with different RNA/DNA concentrations. 177
Figure 69: FRET efficiency histograms with standard deviation for measurements of doubly labeled FnCas12a variants with altered TS affinity. 180
Figure 70: Confocal single-molecule FRET measurements using doubly labeled FnCas12a bridge helix variants with different RNA/DNA concentrations. 182
Figure 71: FRET efficiency histograms with standard deviation for measurements of doubly labeled FnCas12a BH variants. 187
Figure 72: Confocal single-molecule FRET measurements using doubly labeled FnCas12a helix 1 variants with different RNA/DNA concentrations. 189
Figure 73: FRET efficiency histograms with standard deviation for measurements of doubly labeled FnCas12a helix 1 variants. 191

LIST OF TABLES

Table 1. List of chemicals and reagents. 27
Table 2. List of Equipment. 29
Table 3. List of optical components of the MicroTime 200 confocal microscope (PicoQuant, DE). 31
Table 4. List of Consumables. 31
Table 5. List of Software. 32
Table 6. List of buffers. 33
Table 7. List of solutions. 34
Table 8. List of culture media. 35
Table 9. List of buffers and solutions for polyacrylamide gel electrophoresis. 35
Table 10. List of polyacrylamide gel compositions. 37
Table 11. List of commercially available laboratory kit systems. 38
Table 12. List of electrophoresis ladders. 38
Table 13. List of utilized enzymes. 39
Table 14. List of utilized enzyme reaction buffers. 39
Table 15. List of bacterial strains. 39
Table 16. List of vectors and plasmids. 40
Table 17. List of primers used for cloning. 40
Table 18. List of primers for site directed mutagenesis. 41
Table 19. List of primers used for commercial Sanger sequencing. 42
Table 20. List of DNA and RNA oligonucleotides. 43
Table 21. List of DNA oligonucleotides used for T7 *in vitro* transcription to synthesize RNA strands. 44
Table 22. List of T7-transcribed RNA. 44
Table 23. List of fluorescent dyes. 44
Table 24. PCR reaction mix. 47
Table 25. PCR program. 48
Table 26. Reaction mix for DNA restriction. 49
Table 27. Reaction mix for DNA ligation. 49
Table 28. Reaction mix for site-specific mutagenesis. 50
Table 29. PCR program for site-specific mutagenesis. 51
Table 30. Reaction mix for DpnI digestion. 51
Table 31. T7 annealing reaction. 53
Table 32. Protein purification via glutathione affinity chromatography. 56
Table 33. Protein purification via heparin affinity chromatography. 57

Table 34. Oligonucleotide annealing mix. 62
Table 35. Reaction mix for DNA linearization with SmaI. 62
Table 36: Kinetic parameters for FnCas12a *trans*-cleavage. 119
Table 37: Comparison of cleavage efficiencies of all FnCas12a variants generated in this thesis. 119
Table 38. Molecule counts and fitting parameters for single-molecule FRET measurements with an RNA/DNA hybrid. 167
Table 39. Molecule counts and fitting parameters for FRET measurements with doubly labeled WT FnCas12a. 169
Table 40. Molecule counts and fitting parameters for FRET measurements with doubly labeled FnCas12a HKK variants. 172
Table 41. Molecule counts and fitting parameters for FRET measurements with doubly labeled dCas12a. 174
Table 42. Molecule counts and fitting parameters for FRET measurements with doubly labeled FnCas12a variants to alter the target strand affinity.........................177
Table 43. Molecule counts and fitting parameters for FRET measurements with doubly labeled FnCas12a BH variants. 183
Table 44. Molecule counts and fitting parameters for FRET measurements with doubly labeled FnCas12a helix 1 variants. 189

LIST OF ABBREVIATIONS

°C	degree Celsius
aa	amino acid(s)
AATD	α-1 antitrypsin deficiency
ABE	adenine base editor
Abi	abortive infection
Acr	anti-CRISPR
Ago	Argonaute
Amp	ampicillin
APBS	all-photon burst search
APS	ammonium peroxydisulfate
Ara	arabinose
As	*Acidaminococcus sp.*
a.u.	arbitrary units
AzF	*para*-Azido-L-phenylalanine
BH	bridge helix
bin	binary complex
bp	base pair(s)
BREX	bacteriophage exclusion system
BVA	burst variance analysis
c	concentration
Cam	chloramphenicol
CARF	CRISPR-associated Rossmann fold domain
Cas	CRISPR-associated
Cascade	CRISPR-associated complex for antiviral defense
CBE	cytosine base editor
CCR5	chemokine receptor type 5
CD	circular dichroism
coA	cyclic oligoadenylate
COVID-19	coronavirus disease 2019
Cpf1	CRISPR-associated endonuclease in *Prevotella* and *Francisella* 1
CRISPR	clustered regularly interspaced palindromic repeats
crRNA	CRISPR RNA
CV	column volume
dCas12a	deactivated FnCas12a
ddH_2O	double distilled water

dNTPs	deoxyribonucleotides
DCBS	dual-channel burst search
deg	degree
DISARM	defense island system associated with restriction-modification
DNA	desoxyribonucleic acid
ds	double-stranded
DSB	double-strand break
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EM	electron microscopy
EMCCD	electron multiplying charge coupled device
EMSA	electrophoretic mobility shift assay
EtOH	ethanol
Fn	*Francisella novicida*
FQ	Fluorescence quencher
FRET	Förster resonance energy transfer
FWHM	full with half maximum
g	gravity of Earth, 9.81 m s^{-2}
GST	glutathione S-transferase
h	hour(s)
h1	helix 1
HDR	homology-directed repair
HEPN	higher eukaryotes and prokaryotes nucleotide-binding domain
HIV	human immunodeficiency virus
HPLC	high-performance liquid chromatography
indel	insertion/deletion
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilo Dalton
Lb	*Lachnospiraceae bacterium*
LB	lysogenic broth, full media for *E. coli*
LPS	lipopolysaccharide
MGE	mobile genetic element
min	minute(s)
MM	mismatch
MP	mass photometry
MSA	multiple sequence alignment

MT	methyltransferase
n	refractive index
NA	numerical aperture
N_A	Avogadro constant, $6.022 \cdot 10^{23}$ mol^{-1}
n.a.	not assessed
NHEJ	non-homologous end joining
nt	nucleotides
Nuc	nuclease
NTS	non-target strand
Ocr	overcome classical restriction
OD_x	optical density at x nm
OMV	outer membrane vesicle
PAA	polyacrylic acid
PAGE	polyacrylamide gel electrophoresis
PAM	PIE Analysis with MATLAB
PAM	protospacer adjacent motif
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDB	protein data bank
PEG	polyethylene glycol
pH	potential of hydrogen
pI	isoelectric point
PICI	phage-inducible chromosomal island
PIE	pulsed interleaved excitation
pre-crRNA	precursor crRNA
RE	restriction endonuclease
REC	recognition
RAMP	repeat-associated mysterious protein
RM	restriction modification
RNA	ribonucleic acid
RNAi	RNA interference
RNP	ribonucleoprotein
RT	room temperature
rpm	rotations per minute
SARS-CoV-2	severe acute respiratory syndrome Coronavirus 2
SCD	sickle cell disease
SD	standard deviation

SDS	sodium dodecyl sulfate
s.e.	standard error
sgRNA	single guide RNA
SHERLOCK	specific high-sensitivity enzymatic reporter unlocking
Sie	superinfection exclusion systems
sm	single-molecule
SNR	signal-to-noise ratio
sp.	species
Spy	*Streptococcus pyogenes*
SPAD	single-photon avalanche diode
ss	single-stranded
TA	toxin-antitoxin
TAE	Tris-Acetate-EDTA
TBE	Tris-Borate-EDTA
TDT	transfusion-dependent β-thalassemia
TEMED	N, N, N', N'-Tetramethylethylenediamine
ter	ternary complex
TIRF	total internal reflection fluorescence
tracrRNA	trans-activating CRISPR RNA
TS	target strand
UV	ultraviolet
vol	volume
v/v	volume per volume (*volume/volume*)
w/v	weight per volume (*weight/volume*)
ε_x	molar extinction coefficient at x nm
λ	wavelength

Amino acids are abbreviated by the 1- or 3-letter code.

Nucleotides are abbreviated by the 1-letter code.

APPENDIX

I. Vector and plasmid maps

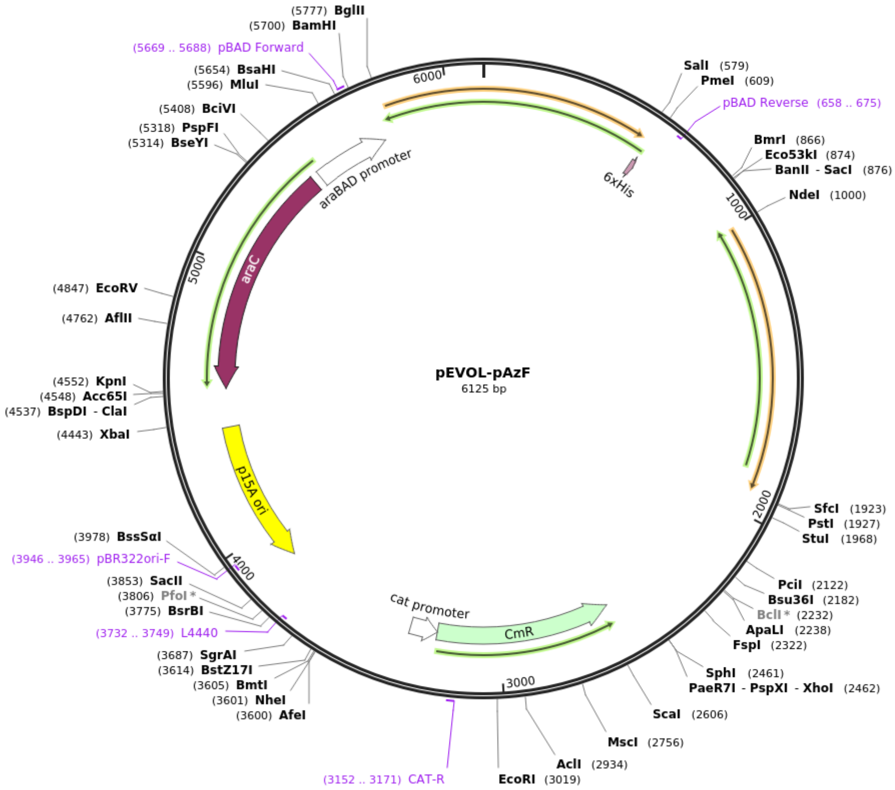


Figure 56: Vector pEVOL-pAzF.
Vector map created with SnapGene.

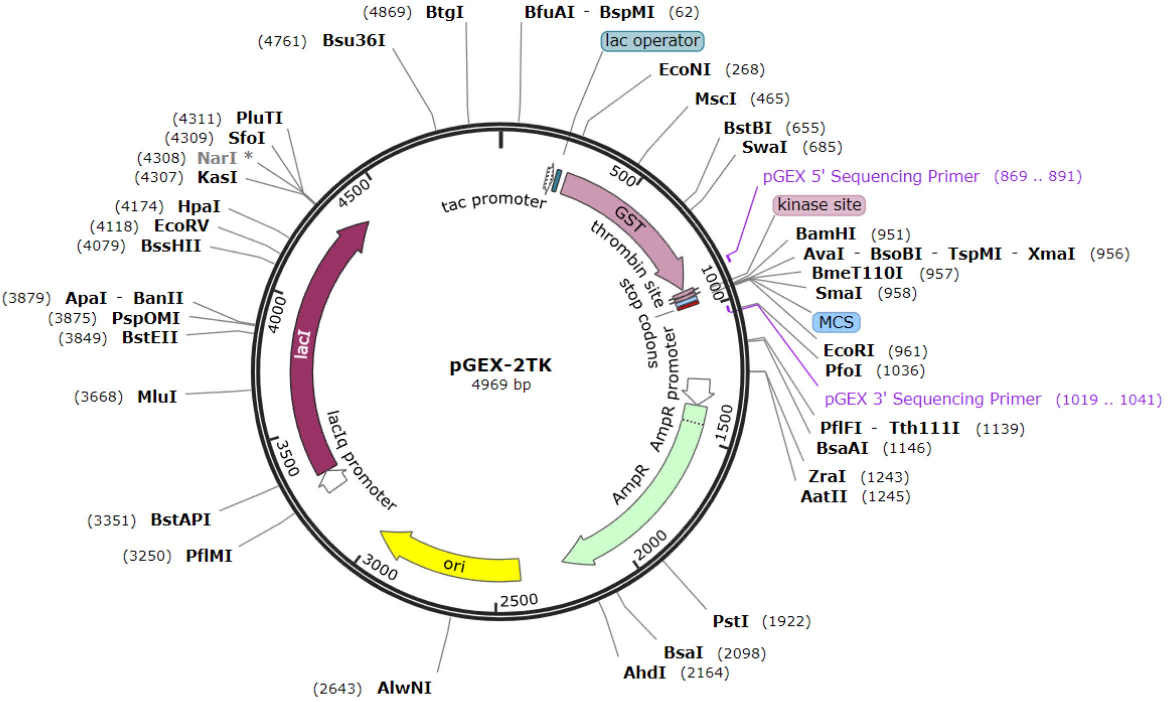


Figure 57: Vector pGEX-2TK.
Vector map created with SnapGene.

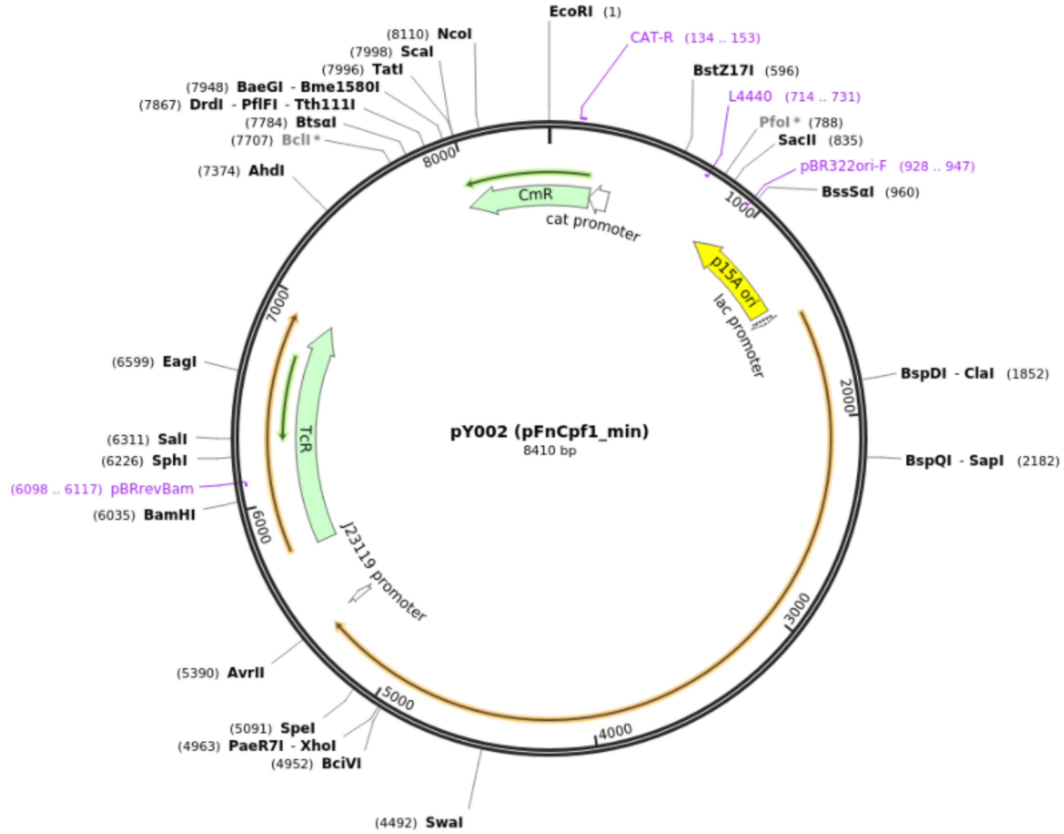


Figure 58: Plasmid pY002 (pFnCpf1_min).
Plasmid map created with SnapGene.

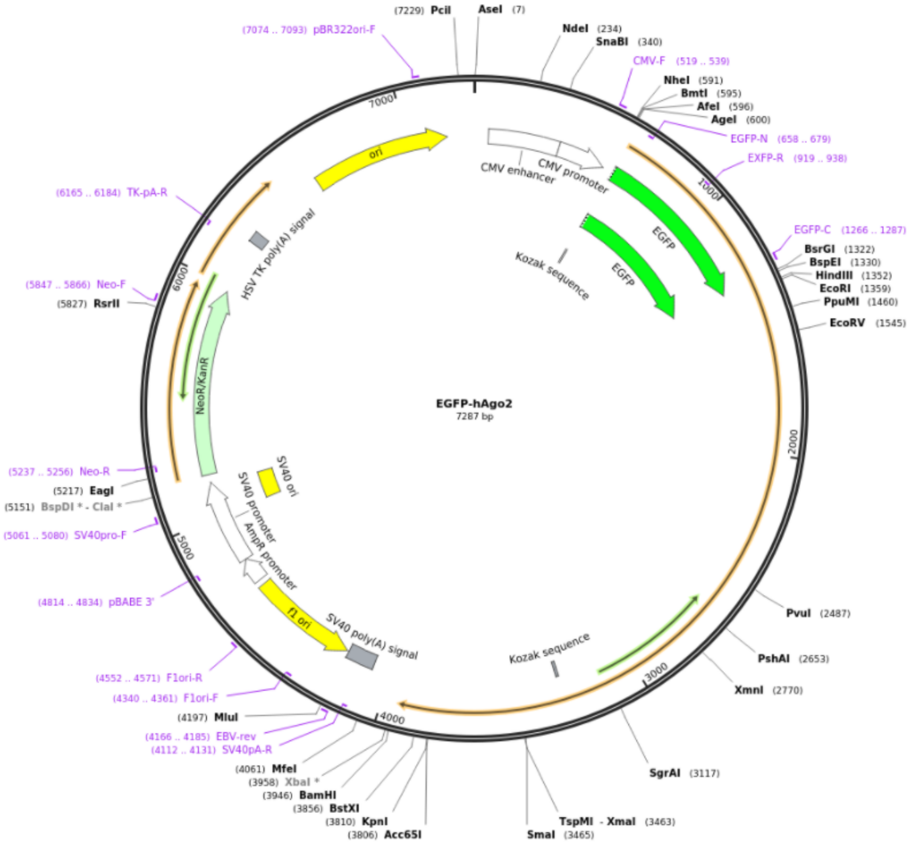


Figure 59: Plasmid EGFP-hAgo2.
Plasmid map created with SnapGene.

II. Protein sequences

II.i. Nucleotide sequence

FnCas12a WT

```
ATG TCA ATT TAT CAA GAA TTT GTT AAT AAA TAT AGT TTA AGT AAA ACT CTA
AGA TTT GAG TTA ATC CCA CAG GGT AAA ACA CTT GAA AAC ATA AAA GCA AGA
GGT TTG ATT TTA GAT GAT GAG AAA AGA GCT AAA GAC TAC AAA AAG GCT AAA
CAA ATA ATT GAT AAA TAT CAT CAG TTT TTT ATA GAG GAG ATA TTA AGT TCG
GTT TGT ATT AGC GAA GAT TTA TTA CAA AAC TAT TCT GAT GTT TAT TTT AAA
CTT AAA AAG AGT GAT GAT GAT AAT CTA CAA AAA GAT TTT AAA AGT GCA AAA
GAT ACG ATA AAG AAA CAA ATA TCT GAA TAT ATA AAG GAC TCA GAG AAA TTT
AAG AAT TTG TTT AAT CAA AAC CTT ATC GAT GCT AAA AAA GGG CAA GAG TCA
GAT TTA ATT CTA TGG CTA AAG CAA TCT AAG GAT AAT GGT ATA GAA CTA TTT
AAA GCC AAT AGT GAT ATC ACA GAT ATA GAT GAG GCG TTA GAA ATA ATC AAA
TCT TTT AAA GGT TGG ACA ACT TAT TTT AAG GGT TTT CAT GAA AAT AGA AAA
AAT GTT TAT AGT AGC AAT GAT ATT CCT ACA TCT ATT ATT TAT AGG ATA GTA
GAT GAT AAT TTG CCT AAA TTT CTA GAA AAT AAA GCT AAG TAT GAG AGT TTA
AAA GAC AAA GCT CCA GAA GCT ATA AAC TAT GAA CAA ATT AAA AAA GAT TTG
GCA GAA GAG CTA ACC TTT GAT ATT GAC TAC AAA ACA TCT GAA GTT AAT CAA
AGA GTT TTT TCA CTT GAT GAA GTT TTT GAG ATA GCA AAC TTT AAT AAT TAT
CTA AAT CAA AGT GGT ATT ACT AAA TTT AAT ACT ATT ATT GGT GGT AAA TTT
GTA AAT GGT GAA AAT ACA AAG AGA AAA GGT ATA AAT GAA TAT ATA AAT CTA
TAC TCA CAG CAA ATA AAT GAT AAA ACA CTC AAA AAA TAT AAA ATG AGT GTT
TTA TTT AAG CAA ATT TTA AGT GAT ACA GAA TCT AAA TCT TTT GTA ATT GAT
AAG TTA GAA GAT GAT AGT GAT GTA GTT ACA ACG ATG CAA AGT TTT TAT GAG
CAA ATA GCA GCT TTT AAA ACA GTA GAA GAA AAA TCT ATT AAA GAA ACA CTA
TCT TTA TTA TTT GAT GAT TTA AAA GCT CAA AAA CTT GAT TTG AGT AAA ATT
TAT TTT AAA AAT GAT AAA TCT CTT ACT GAT CTA TCA CAA CAA GTT TTT GAT
GAT TAT AGT GTT ATT GGT ACA GCG GTA CTA GAA TAT ATA ACT CAA CAA ATA
GCA CCT AAA AAT CTT GAT AAC CCT AGT AAG AAA GAG CAA GAA TTA ATA GCC
AAA AAA ACT GAA AAA GCA AAA TAC TTA TCT CTA GAA ACT ATA AAG CTT GCC
TTA GAA GAA TTT AAT AAG CAT AGA GAT ATA GAT AAA CAG TGT AGG TTT GAA
GAA ATA CTT GCA AAC TTT GCG GCT ATT CCG ATG ATA TTT GAT GAA ATA GCT
CAA AAC AAA GAC AAT TTG GCA CAG ATA TCT ATC AAA TAT CAA AAT CAA GGT
AAA AAA GAC CTA CTT CAA GCT AGT GCG GAA GAT GAT GTT AAA GCT ATC AAG
GAT CTT TTA GAT CAA ACT AAT AAT CTC TTA CAT AAA CTA AAA ATA TTT CAT
ATT AGT CAG TCA GAA GAT AAG GCA AAT ATT TTA GAC AAG GAT GAG CAT TTT
TAT CTA GTA TTT GAG GAG TGC TAC TTT GAG CTA GCG AAT ATA GTG CCT CTT
TAT AAC AAA ATT AGA AAC TAT ATA ACT CAA AAG CCA TAT AGT GAT GAG AAA
TTT AAG CTC AAT TTT GAG AAC TCG ACT TTG GCT AAT GGT TGG GAT AAA AAT
AAA GAG CCT GAC AAT ACG GCA ATT TTA TTT ATC AAA GAT GAT AAA TAT TAT
CTG GGT GTG ATG AAT AAG AAA AAT AAC AAA ATA TTT GAT GAT AAA GCT ATC
AAA GAA AAT AAA GGC GAG GGT TAT AAA AAA ATT GTT TAT AAA CTT TTA CCT
GGC GCA AAT AAA ATG TTA CCT AAG GTT TTC TTT TCT GCT AAA TCT ATA AAA
TTT TAT AAT CCT AGT GAA GAT ATA CTT AGA ATA AGA AAT CAT TCC ACA CAT
ACA AAA AAT GGT AGT CCT CAA AAA GGA TAT GAA AAA TTT GAG TTT AAT ATT
GAA GAT TGC CGA AAA TTT ATA GAT TTT TAT AAA CAG TCT ATA AGT AAG CAT
```

```
CCG GAG TGG AAA GAT TTT GGA TTT AGA TTT TCT GAT ACT CAA AGA TAT AAT
TCT ATA GAT GAA TTT TAT AGA GAA GTT GAA AAT CAA GGC TAC AAA CTA ACT
TTT GAA AAT ATA TCA GAG AGC TAT ATT GAT AGC GTA GTT AAT CAG GGT AAA
TTG TAC CTA TTC CAA ATC TAT AAT AAA GAT TTT TCA GCT TAT AGC AAA GGG
CGA CCA AAT CTA CAT ACT TTA TAT TGG AAA GCG CTG TTT GAT GAG AGA AAT
CTT CAA GAT GTG GTT TAT AAG CTA AAT GGT GAG GCA GAG CTT TTT TAT CGT
AAA CAA TCA ATA CCT AAA AAA ATC ACT CAC CCA GCT AAA GAG GCA ATA GCT
AAT AAA AAC AAA GAT AAT CCT AAA AAA GAG AGT GTT TTT GAA TAT GAT TTA
ATC AAA GAT AAA CGC TTT ACT GAA GAT AAG TTT TTC TTT CAC TGT CCT ATT
ACA ATC AAT TTT AAA TCT AGT GGA GCT AAT AAG TTT AAT GAT GAA ATC AAT
TTA TTG CTA AAA GAA AAA GCA AAT GAT GTT CAT ATA TTA AGT ATA GAT AGA
GGT GAA AGA CAT TTA GCT TAC TAT ACT TTG GTA GAT GGT AAA GGC AAT ATC
ATC AAA CAA GAT ACT TTC AAC ATC ATT GGT AAT GAT AGA ATG AAA ACA AAC
TAC CAT GAT AAG CTT GCT GCA ATA GAG AAA GAT AGG GAT TCA GCT AGG AAA
GAC TGG AAA AAG ATA AAT AAC ATC AAA GAG ATG AAA GAG GGC TAT CTA TCT
CAG GTA GTT CAT GAA ATA GCT AAG CTA GTT ATA GAG TAT AAT GCT ATT GTG
GTT TTT GAG GAT TTA AAT TTT GGA TTT AAA AGA GGG CGT TTC AAG GTA GAG
AAG CAG GTC TAT CAA AAG TTA GAA AAA ATG CTA ATT GAG AAA CTA AAC TAT
CTA GTT TTC AAA GAT AAT GAG TTT GAT AAA ACT GGG GGA GTG CTT AGA GCT
TAT CAG CTA ACA GCA CCT TTT GAG ACT TTT AAA AAG ATG GGT AAA CAA ACA
GGT ATT ATC TAC TAT GTA CCA GCT GGT TTT ACT TCA AAA ATT TGT CCT GTA
ACT GGT TTT GTA AAT CAG TTA TAT CCT AAG TAT GAA AGT GTC AGC AAA TCT
CAA GAG TTC TTT AGT AAG TTT GAC AAG ATT TGT TAT AAC CTT GAT AAG GGC
TAT TTT GAG TTT AGT TTT GAT TAT AAA AAC TTT GGT GAC AAG GCT GCC AAA
GGC AAG TGG ACT ATA GCT AGC TTT GGG AGT AGA TTG ATT AAC TTT AGA AAT
TCA GAT AAA AAT CAT AAT TGG GAT ACT CGA GAA GTT TAT CCA ACT AAA GAG
TTG GAG AAA TTG CTA AAA GAT TAT TCT ATC GAA TAT GGG CAT GGC GAA TGT
ATC AAA GCA GCT ATT TGC GGT GAG AGC GAC AAA AAG TTT TTT GCT AAG CTA
ACT AGT GTC CTA AAT ACT ATC TTA CAA ATG CGT AAC TCA AAA ACA GGT ACT
GAG TTA GAT TAT CTA ATT TCA CCA GTA GCA GAT GTA AAT GGC AAT TTC TTT
GAT TCG CGA CAG GCG CCA AAA AAT ATG CCT CAA GAT GCT GAT GCC AAT GGT
GCT TAT CAT ATT GGG CTA AAA GGT CTG ATG CTA CTA GGT AGG ATC AAA AAT
AAT CAA GAG GGC AAA AAA CTC AAT TTG GTT ATC AAA AAT GAA GAG TAT TTT
GAG TTC GTG CAG AAT AGG AAT AAC TAA
```

Length:	3909 bp
GC-content:	29%

calculated with http://biotools.nubic.northwestern.edu/OligoCalc2.0.html

II.ii. Amino acid sequence

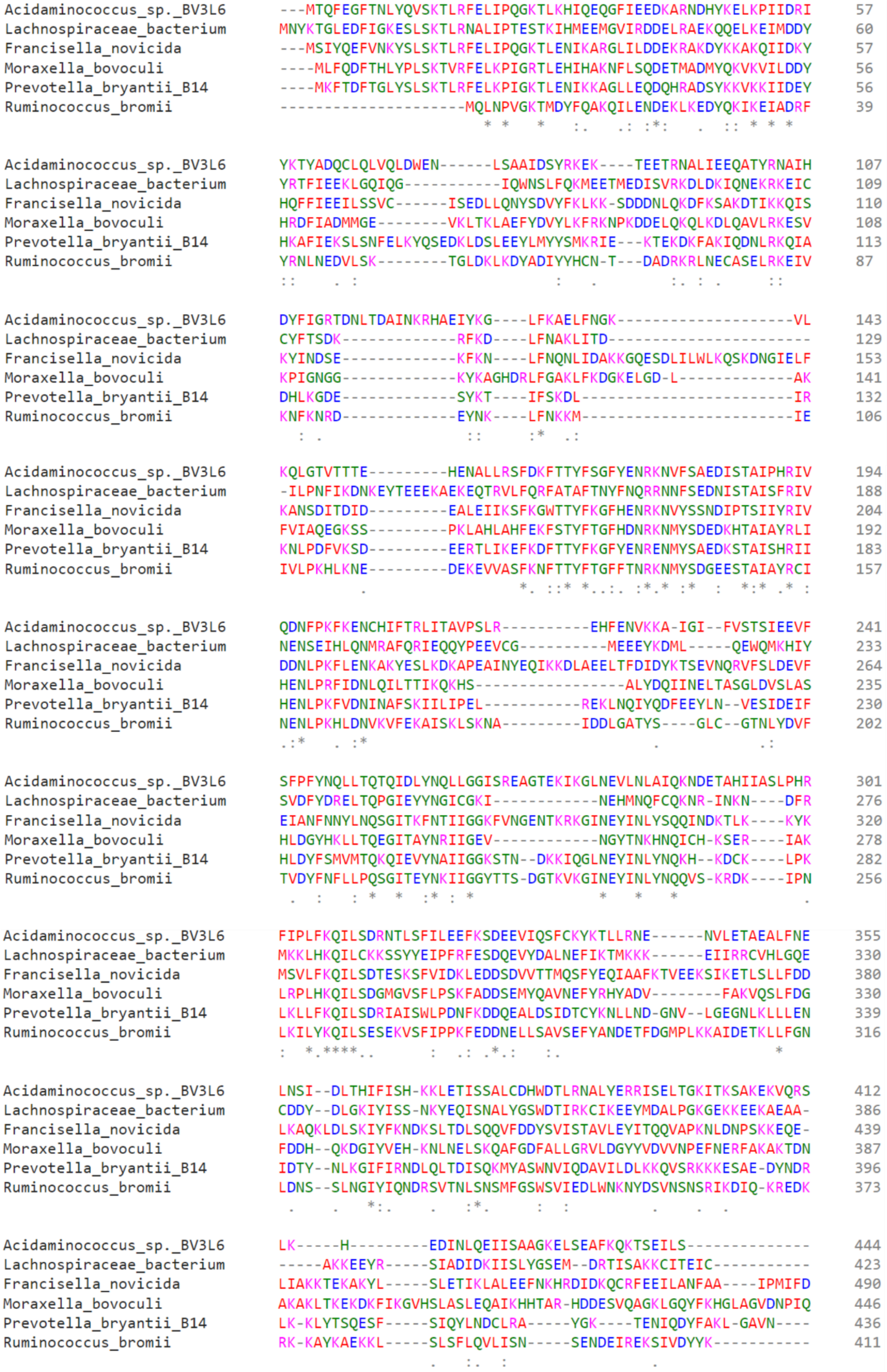
FnCas12a WT

```
MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKYHQFFIEEIL
SSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQISEYIKDSEKFKNLFNQNLIDAKK
GQESDLILWLKQSKDNGIELFKANSDITDIDEALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTS
IIYRIVDDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVF
EIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILS
DTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDDLKAQKLDLSKIYFKND
KSLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLETIKLALEE
FNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKKDLLQASAEDDVKAIKD
LLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDE
KFKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYK
LLPGANKMLPKVFFSAKSIKFYNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQ
SISKHPEWKDFGFRFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKD
FSAYSKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANKNKDNPK
KESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKANDVHILSIDRGERHLA
YYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVH
EIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQ
LTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKG
YFEFSFDYKNFGDKAAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHG
ECIKAAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKNMPQDA
DANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEYFEFVQNRNN
```

Length:	1300 aa
Molecular weight:	151.915 kDa
Extinction coefficient:	143 830 M^{-1} cm^{-1}
Theoretical pI:	8.57

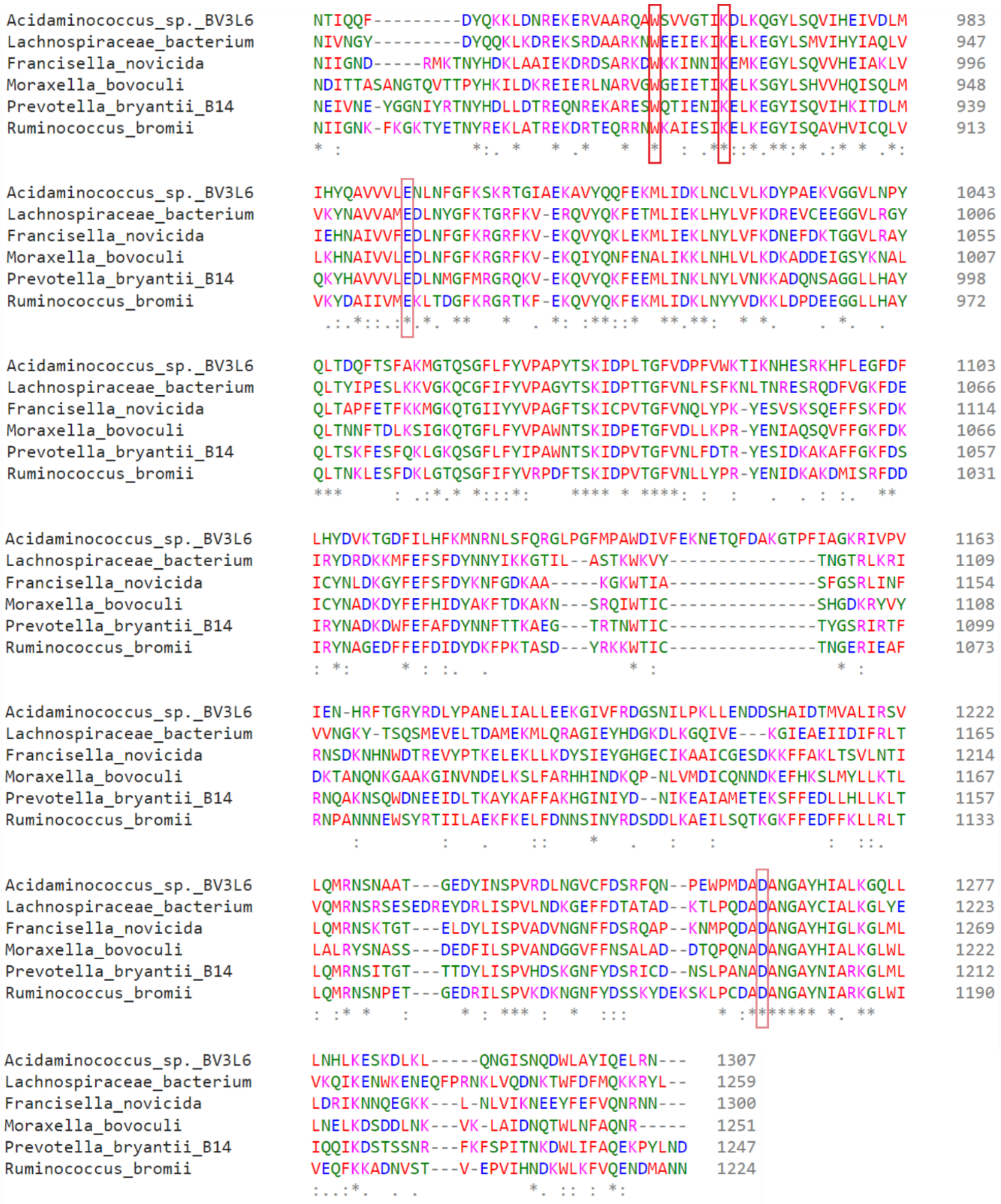
calculated with https://web.expasy.org/protparam/

II.iii.Multiple sequence alignment of Cas12a variants



```
Acidaminococcus_sp._BV3L6      ---MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKELKPIIDRI      57
Lachnospiraceae_bacterium      MNYKTGLEDFIGKESLSKTLRNALIPTESTKIHMEEMGVIRDDELRAEKQQELKEIMDDY      60
Francisella_novicida           ---MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKAKQIIDKY      57
Moraxella_bovoculi             ----MLFQDFTHLYPLSKTVRFELKPIGRTLEHIHAKNFLSQDETMADMYQKVKVILDDY      56
Prevotella_bryantii_B14        ----MKFTDFTGLYSLSKTLRFELKPIGKTLENIKKAGLLEQDQHRADSYKKVKKIIDEY      56
Ruminococcus_bromii            ---------------------MQLNPVGKTMDYFQAKQILENDEKLKEDYQKIKEIADRF      39
                                                      * *   *   :.   .: :*:   .  :: * * *

Acidaminococcus_sp._BV3L6      YKTYADQCLQLVQLDWEN------LSAAIDSYRKEK----TEETRNALIEEQATYRNAIH      107
Lachnospiraceae_bacterium      YRTFIEEKLGQIQG-----------IQWNSLFQKMEETMEDISVRKDLDKIQNEKRKEIC      109
Francisella_novicida           HQFFIEEILSSVC------ISEDLLQNYSDVYFKLKK-SDDDNLQKDFKSAKDTIKKQIS      110
Moraxella_bovoculi             HRDFIADMMGE--------VKLTKLAEFYDVYLKFRKNPKDDELQKQLKDLQAVLRKESV      108
Prevotella_bryantii_B14        HKAFIEKSLSNFELKYQSEDKLDSLEEYLMYYSMKRIE---KTEKDKFAKIQDNLRKQIA      113
Ruminococcus_bromii            YRNLNEDVLSK--------TGLDKLKDYADIYYHCN-T---DADRKRLNECASELRKEIV      87
                               ::    . :                      :   .        :. : .     ::

Acidaminococcus_sp._BV3L6      DYFIGRTDNLTDAINKRHAEIYKG----LFKAELFNGK--------------------VL      143
Lachnospiraceae_bacterium      CYFTSDK-------------RFKD----LFNAKLITD-----------------------      129
Francisella_novicida           KYINDSE-------------KFKN----LFNQNLIDAKKGQESDLILWLKQSKDNGIELF      153
Moraxella_bovoculi             KPIGNGG-------------KYKAGHDRLFGAKLFKDGKELGD-L-------------AK      141
Prevotella_bryantii_B14        DHLKGDE-------------SYKT----IFSKDL------------------------IR      132
Ruminococcus_bromii            KNFKNRD-------------EYNK----LFNKKM------------------------IE      106
                                 : .                ::     :*  .:

Acidaminococcus_sp._BV3L6      KQLGTVTTTE---------HENALLRSFDKFTTYFSGFYENRKNVFSAEDISTAIPHRIV      194
Lachnospiraceae_bacterium      -ILPNFIKDNKEYTEEEKAEKEQTRVLFQRFATAFTNYFNQRRNNFSEDNISTAISFRIV      188
Francisella_novicida           KANSDITDID---------EALEIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIV      204
Moraxella_bovoculi             FVIAQEGKSS---------PKLAHLAHFEKFSTYFTGFHDNRKNMYSDEDKHTAIAYRLI      192
Prevotella_bryantii_B14        KNLPDFVKSD---------EERTLIKEFKDFTTYFKGFYENRENMYSAEDKSTAISHRII      183
Ruminococcus_bromii            IVLPKHLKNE---------DEKEVVASFKNFTTYFTGFFTNRKNMYSDGEESTAIAYRCI      157
                                        .                 *. ::* *..:. :*.* :*  :  *:* .* :

Acidaminococcus_sp._BV3L6      QDNFPKFKENCHIFTRLITAVPSLR----------EHFENVKKA-IGI--FVSTSIEEVF      241
Lachnospiraceae_bacterium      NENSEIHLQNMRAFQRIEQQYPEEVCG----------MEEEYKDML-----QEWQMKHIY      233
Francisella_novicida           DDNLPKFLENKAKYESLKDKAPEAINYEQIKKDLAEELTFDIDYKTSEVNQRVFSLDEVF      264
Moraxella_bovoculi             HENLPRFIDNLQILTTIKQKHS-----------------ALYDQIINELTASGLDVSLAS      235
Prevotella_bryantii_B14        HENLPKFVDNINAFSKIILIPEL-----------REKLNQIYQDFEEYLN--VESIDEIF      230
Ruminococcus_bromii            NENLPKHLDNVKVFEKAISKLSKNA---------IDDLGATYS----GLC--GTNLYDVF      202
                               .:*   . :*                                .           .:

Acidaminococcus_sp._BV3L6      SFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVLNLAIQKNDETAHIIASLPHR      301
Lachnospiraceae_bacterium      SVDFYDRELTQPGIEYYNGICGKI------------NEHMNQFCQKNR-INKN----DFR      276
Francisella_novicida           EIANFNNYLNQSGITKFNTIIGGKFVNGENTKRKGINEYINLYSQQINDKTLK----KYK      320
Moraxella_bovoculi             HLDGYHKLLTQEGITAYNRIIGEV------------NGYTNKHNQICH-KSER----IAK      278
Prevotella_bryantii_B14        HLDYFSMVMTQKQIEVYNAIIGGKSTN--DKKIQGLNEYINLYNQKH--KDCK----LPK      282
Ruminococcus_bromii            TVDYFNFLLPQSGITEYNKIIGGYTTS-DGTKVKGINEYINLYNQQVS-KRDK----IPN      256
                                .  :   : *  *  :* : *              *   *   *              .

Acidaminococcus_sp._BV3L6      FIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLRNE------NVLETAEALFNE      355
Lachnospiraceae_bacterium      MKKLHKQILCKKSSYYEIPFRFESDQEVYDALNEFIKTMKKK------EIIRRCVHLGQE      330
Francisella_novicida           MSVLFKQILSDTESKSFVIDKLEDDSDVVTTMQSFYEQIAAFKTVEEKSIKETLSLLFDD      380
Moraxella_bovoculi             LRPLHKQILSDGMGVSFLPSKFADDSEMYQAVNEFYRHYADV--------FAKVQSLFDG      330
Prevotella_bryantii_B14        LKLLFKQILSDRIAISWLPDNFKDDQEALDSIDTCYKNLLND-GNV--LGEGNLKLLLEN      339
Ruminococcus_bromii            LKILYKQILSESEKVSFIPPKFEDDNELLSAVSEFYANDETFDGMPLKKAIDETKLLFGN      316
                               :  *.****..      :  .: .*.:   :.                        *

Acidaminococcus_sp._BV3L6      LNSI--DLTHIFISH-KKLETISSALCDHWDTLRNALYERRISELTGKITKSAKEKVQRS      412
Lachnospiraceae_bacterium      CDDY--DLGKIYISS-NKYEQISNALYGSWDTIRKCIKEEYMDALPGKGEKKEEKAEAA-      386
Francisella_novicida           LKAQKLDLSKIYFKNDKSLTDLSQQVFDDYSVISTAVLEYITQQVAPKNLDNPSKKEQE-      439
Moraxella_bovoculi             FDDH--QKDGIYVEH-KNLNELSKQAFGDFALLGRVLDGYYVDVVNPEFNERFAKAKTDN      387
Prevotella_bryantii_B14        IDTY--NLKGIFIRNDLQLTDISQKMYASWNVIQDAVILDLKKQVSRKKKESAE-DYNDR      396
Ruminococcus_bromii            LDNS--SLNGIYIQNDRSVTNLSNSMFGSWSVIEDLWNKNYDSVNSNSRIKDIQ-KREDK      373
                                .    .   *:.    .   :*.     :  :         .    .  .

Acidaminococcus_sp._BV3L6      LK-----H---------EDINLQEIISAAGKELSEAFKQKTSEILS--------------      444
Lachnospiraceae_bacterium      -----AKKEEYR-----SIADIDKIISLYGSEM--DRTISAKKCITEIC-----------      423
Francisella_novicida           LIAKKTEKAKYL-----SLETIKLALEEFNKHRDIDKQCRFEEILANFAA----IPMIFD      490
Moraxella_bovoculi             AKAKLTKEKDKFIKGVHSLASLEQAIKHHTAR-HDDESVQAGKLGQYFKHGLAGVDNPIQ      446
Prevotella_bryantii_B14        LK-KLYTSQESF-----SIQYLNDCLRA-----YGK----TENIQDYFAKL-GAVN----      436
Ruminococcus_bromii            RK-KAYKAEKKL-----SLSFLQVLISN-----SENDEIREKSIVDYYK-----------      411
                                                .   :.  :                .
```

```
Acidaminococcus_sp._BV3L6          ---------------------HAHAALDQPLPTT-------LKKQEEKEILKSQLDSLLG     476
Lachnospiraceae_bacterium          ---------------------DMAGQISID-PLVCNSDIKLLQNKEKTTEIKTILDSFLH     461
Francisella_novicida               EIAQNK---------------DNLAQISIKYQNQGKKDLLQASAEDDVKAIKDLLDQTNN     535
Moraxella_bovoculi                 KIHNNHSTIKGFLERERPAGERALPKI------KSG----KNPEMTQLRQLKELLDNALN     496
Prevotella_bryantii_B14            ---NEHEQTINLFAQVRNAYTSVQAILTTPYPENAN----LAQDKETVALIKNLLDSLKR     489
Ruminococcus_bromii                ------TSLMQLTDNLSDKYKEAAPLFSENYDNEKG----LKNDDKSISLIKNFLDAIKE     461
                                                             :                       :*  **

Acidaminococcus_sp._BV3L6          LYHLLDWFA------VDESNEVDPEFSARLTGIKLEMEPSLSFYNKARNYATKKPYSVEK     530
Lachnospiraceae_bacterium          VYQWGQ------TFIVSDIIEKDSYFYSELEDVLEDFEGITTLYNHVRSYVTQKPYSTVK     515
Francisella_novicida               LLHRLKIFHISQSEDKANILDKDEHFYLVFEECYFELANIVPLYNKIRNYITQKPYSDEK     595
Moraxella_bovoculi                 VAHFAKLL-----MTKTTLDNQDGNFYGEFGVLYDELAKIPTLYNKVRDYLSQKPFSTEK     551
Prevotella_bryantii_B14            LQRFIKPL-----LGKGDESDKDERFYGDFTPLWETLNQITPLYNMVRNYMTRKPYSQEK     544
Ruminococcus_bromii                IEKFIKPL-----SETNITGEKNDLFYSQFTPLLDNISRIDILYDKVRNYVTQKPFSTDK     516
                                   : :  .              : :  *   :      :     :*:  *.* ::**:*  *

Acidaminococcus_sp._BV3L6          FKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGIMPKQKGRYKALSFEPTEKTSEGF     590
Lachnospiraceae_bacterium          FKLHFGSPTLANGWSQSKEYDNNAILLMRDQKFYLGIFNVRNKPDKQIIKGHEKEEKGDY     575
Francisella_novicida               FKLNFENSTLANGWDKNKEPDNTAILFIKDDKYYLGVMNKKNNKIFDDKAIKE-NKGEGY     654
Moraxella_bovoculi                 YKLNFGNPTLLNGWDLNKEKDNFGVILQKDGCYYLALLDKAHKKVFDNAPN---TGKNVY     608
Prevotella_bryantii_B14            IKLNFENSTLLGGWDLNKEHDNTAIILRKNGLYYLAIMKKSANKIFDKDKLD--NSGDCY     602
Ruminococcus_bromii                IKLNFGNSQLLNGWDKDKEREYGAVLLCKDEKYYLAIIDKSNNSILENIDFQDCNESDYY     576
                                    **:*    * .**. .** :  .::: ::  :**.::                     :

Acidaminococcus_sp._BV3L6          DKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHTTPILLSNNFIEPLEITKEIYDLNNPEK     650
Lachnospiraceae_bacterium          KKMIYNLLPGPSKMLPKVFITSRSGQ--------------ETYKPSKHILDGYNEKRH--     619
Francisella_novicida               KKIVYKLLPGANKMLPKVFFSA-KSI--------------KFYNPSEDILRIRNHSTH--     697
Moraxella_bovoculi                 QKMIYKLLPGPNKMLPKVFFAK-SNL--------------GYYNPSAELLDKYAQGTH--     651
Prevotella_bryantii_B14            EKMVYKLLPGANKMLPKVFFSK-SRI--------------DEFKPSENIIENYKKGTH--     645
Ruminococcus_bromii                EKIVYKLLTKINGNLPRVFFSE-KRK--------------KLLSPSDEILKIYKSGTF--     619
                                   .*: *. :      :*:                          .*

Acidaminococcus_sp._BV3L6          EPKKFQTAYAKKTGDQKGYRE------ALCKWIDFTRDFLSKYTKTTSIDLSSLRPSSQY     704
Lachnospiraceae_bacterium          ---------IKSS--------PKFDLGYCWDLIDYYKECIHKHPDWKNYDF-HFSDTKDY     661
Francisella_novicida               ---------TKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGF-RFSDTQRY     747
Moraxella_bovoculi                 ---------KKG---------DNFNLKDCHALIDFFKAGINKHPEWQHFGF-KFSPTSSY     692
Prevotella_bryantii_B14            ---------KKG---------ANFNLADCHNLIDFFKSSISKHEDWSKFNF-HFSDTSSY     686
Ruminococcus_bromii                ---------KKG---------DKFSLDDCHKLIDFYKESFKKYPNWLIYNF-KFKNTNEY     660
                                             *                     **: :  : *: .    .:  :  :. *

Acidaminococcus_sp._BV3L6          KDLGEYYAELNPLLYHISFQRIAEKEIMDAVETGKLYLFQIYNKDFAKGHHGKPNLHTLY     764
Lachnospiraceae_bacterium          EDISGFYREVEMQGYQIKWTYISADEIQKLDEKGQIFLFQIYNKDFSVHSTGKDNLHTMY     721
Francisella_novicida               NSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAYSKGRPNLHTLY     807
Moraxella_bovoculi                 RDLSDFYREVEPQGYQVKFVDINADYIDELVEQGQLYLFQIYNKDFSPKAHGKPNLHTLY     752
Prevotella_bryantii_B14            EDLSDFYREVEQQGYSISFCDVSVEYINKMVEKGDLYLFQIYNKDFSEFSKGTPNMHTLY     746
Ruminococcus_bromii                NDISEFYNDVASQGYNISKMKIPTTFIDKLVDEGKIYLFQLYNKDFSPHSKGTPNLHTLY     720
                                   ..:. :* ::    * :.   :    * .  : *.::***:*****:    *  *:**:*

Acidaminococcus_sp._BV3L6          WTGLFSPENLAKTSIKLNGQAELFYRPKSRM-KRMAHRLGEKMLNKKL-------KDQKT     816
Lachnospiraceae_bacterium          LKNLFSEENLKDIVLKLNGEAELFFRKASIKTP--IVHKKGSVLVNRSYTQTVGNKEIRV     779
Francisella_novicida               WKALFDERNLQDVVYKLNGEAELFYRKQSIP-KKITHPAKETIANKNK------------     854
Moraxella_bovoculi                 FKALFSEDNLADPIYKLNGEAQIFYRKASLDMNETTIHRAGEVLENKN------------     800
Prevotella_bryantii_B14            WNSLFSKENLNNIIYKLNGQAEIFFRKKSLNYKRPTHPAHQAIKNKNK------------     794
Ruminococcus_bromii                FKMLFDERNLEDVVYKLNGEAEMFYRPASIKYDKPTHPKNTPIKNKNT------------     768
                                    .  **.  **  .   ****:*::*:*  *             :  :.

Acidaminococcus_sp._BV3L6          PIPDTLYQELYDYVNHRLSHDLSDEARALL--PNVITKEVSHEIIKDRRFTSDKFFFHVP     874
Lachnospiraceae_bacterium          SIPEEYYTEIYNYLNHIGKGKLSSEAQRYLDEGKIKSFTATKDIVKNYRYCCDHYFLHLP     839
Francisella_novicida               -----------------------------D--NPKKESVFEYDLIKDKRFTEDKFFFHCP     883
Moraxella_bovoculi                 -----------------------------P--DNPKKRQFVYDIIKDKRYTQDKFMLHVP     829
Prevotella_bryantii_B14            -----------------------------C--NEKKESIFDYDLVKDKRYTVDKFQFHVP     823
Ruminococcus_bromii                -----------------------------L--NDKKASTFPYDLIKDKRYTKWQFSLHFP     797
                                                                             :::*: *:   :: :* *

Acidaminococcus_sp._BV3L6          ITLNYQAANSPS-KFNQRVNAYLK-EHPETPIIGIDRGERNLIYITVIDSTGKILEQRSL     932
Lachnospiraceae_bacterium          ITINFKAKSDVA--VNERTLAYIA-KKEDIHIIGIDRGERNLLYISVVDVHGNIREQRSF     896
Francisella_novicida               ITINFKSSGANK--FNDEINLLLKEKANDVHILSIDRGERHLAYYTLVDGKGNIIKQDTF     941
Moraxella_bovoculi                 ITMNFGVQGMTIKEFNKKVNQSIQ-QYDEVNVIGIDRGERHLLYLTVINSKGEILEQRSL     888
Prevotella_bryantii_B14            ITMNFKSTGNTN--INQQVIDYLR-TEDDTHIIGIDRGERHLLYLVVIDSHGKIVEQFTL     880
Ruminococcus_bromii                ITMNFKAPDKAM--INDDVRNLLK-SCNNNFIIGIDRGERNLLYVSVIDSNGAIIYQHSL     854
                                   **:*:   .     .*.     :     :  ::.******:* *  :::  * *  * ::
```



```
Acidaminococcus_sp._BV3L6           NTIQQF---------DYQKKLDNREKERVAARQAWSVVGTIKDLKQGYLSQVIHEIVDLM      983
Lachnospiraceae_bacterium           NIVNGY---------DYQQKLKDREKSRDAARKNWEEIEKIKELKEGYLSMVIHYIAQLV      947
Francisella_novicida                NIIGND-----RMKTNYHDKLAAIEKDRDSARKDWKKINNIKEMKEGYLSQVVHEIAKLV      996
Moraxella_bovoculi                  NDITTASANGTQVTTPYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVVHQISQLM      948
Prevotella_bryantii_B14             NEIVNE-YGGNIYRTNYHDLLDTREQNREKARESWQTIENIKELKEGYISQVIHKITDLM      939
Ruminococcus_bromii                 NIIGNK-FKGKTYETNYREKLATREKDRTEQRRNWKAIESIKELKEGYISQAVHVICQLV      913
                                    * :             *:. *   * .*   *  *  : .**::*.**:* .:* * .*:

Acidaminococcus_sp._BV3L6           IHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDKLNCLVLKDYPAEKVGGVLNPY      1043
Lachnospiraceae_bacterium           VKYNAVVAMEDLNYGFKTGRFKV-ERQVYQKFETMLIEKLHYLVFKDREVCEEGGVLRGY      1006
Francisella_novicida                IEHNAIVVFEDLNFGFKRGRFKV-EKQVYQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAY      1055
Moraxella_bovoculi                  LKHNAIVVLEDLNFGFKRGRFKV-EKQIYQNFENALIKKLNHLVLKDKADDEIGSYKNAL      1007
Prevotella_bryantii_B14             QKYHAVVVLEDLNMGFMRGRQKV-EKQVYQKFEEMLINKLNYLVNKKADQNSAGGLLHAY      998
Ruminococcus_bromii                 VKYDAIIVMEKLTDGFKRGRTKF-EKQVYQKFEKMLIDKLNYYVDKKLDPDEEGGLLHAY      972
                                     .:.*::.:*.*. **   *  . *: :**::*  **.**:  * *.    . *.  .

Acidaminococcus_sp._BV3L6           QLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTGFVDPFVWKTIKNHESRKHFLEGFDF      1103
Lachnospiraceae_bacterium           QLTYIPESLKKVGKQCGFIFYVPAGYTSKIDPTTGFVNLFSFKNLTNRESRQDFVGKFDE      1066
Francisella_novicida                QLTAPFETFKKMGKQTGIIYYVPAGFTSKICPVTGFVNQLYPK-YESVSKSQEFFSKFDK      1114
Moraxella_bovoculi                  QLTNNFTDLKSIGKQTGFLFYVPAWNTSKIDPETGFVDLLKPR-YENIAQSQVFFGKFDK      1066
Prevotella_bryantii_B14             QLTSKFESFQKLGKQSGFLFYIPAWNTSKIDPVTGFVNLFDTR-YESIDKAKAFFGKFDS      1057
Ruminococcus_bromii                 QLTNKLESFDKLGTQSGFIFYVRPDFTSKIDPVTGFVNLLYPR-YENIDKAKDMISRFDD      1031
                                    ***     : .:*.* *:::*:    **** * ****: :  :   .  . : :.  **

Acidaminococcus_sp._BV3L6           LHYDVKTGDFILHFKMNRNLSFQRGLPGFMPAWDIVFEKNETQFDAKGTPFIAGKRIVPV      1163
Lachnospiraceae_bacterium           IRYDRDKKMFEFSFDYNNYIKKGTIL--ASTKWKVY---------------TNGTRLKRI      1109
Francisella_novicida                ICYNLDKGYFEFSFDYKNFGDKAA-----KGKWTIA---------------SFGSRLINF      1154
Moraxella_bovoculi                  ICYNADKDYFEFHIDYAKFTDKAKN---SRQIWTIC---------------SHGDKRYVY      1108
Prevotella_bryantii_B14             IRYNADKDWFEFAFDYNNFTTKAEG---TRTNWTIC---------------TYGSRIRTF      1099
Ruminococcus_bromii                 IRYNAGEDFFEFDIDYDKFPKTASD---YRKKWTIC---------------TNGERIEAF      1073
                                    : *:     * : :.  .              *  :                  * :

Acidaminococcus_sp._BV3L6           IEN-HRFTGRYRDLYPANELIALLEEKGIVFRDGSNILPKLLENDDSHAIDTMVALIRSV      1222
Lachnospiraceae_bacterium           VVNGKY-TSQSMEVELTDAMEKMLQRAGIEYHDGKDLKGQIVE---KGIEAEIIDIFRLT      1165
Francisella_novicida                RNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIKAAICGESDKKFFAKLTSVLNTI      1214
Moraxella_bovoculi                  DKTANQNKGAAKGINVNDELKSLFARHHINDKQP-NLVMDICQNNDKEFHKSLMYLLKTL      1167
Prevotella_bryantii_B14             RNQAKNSQWDNEEIDLTKAYKAFFAKHGINIYD--NIKEAIAMETEKSFFEDLLHLLKLT      1157
Ruminococcus_bromii                 RNPANNNEWSYRTIILAEKFKELFDNNSINYRDSDDLKAEILSQTKGKFFEDFFKLLRLT      1133
                                        :        :   .    ::    *   .   :   :           :  ::.

Acidaminococcus_sp._BV3L6           LQMRNSNAAT---GEDYINSPVRDLNGVCFDSRFQN--PEWPMDADANGAYHIALKGQLL      1277
Lachnospiraceae_bacterium           VQMRNSRSESEDREYDRLISPVLNDKGEFFDTATAD--KTLPQDADANGAYCIALKGLYE      1223
Francisella_novicida                LQMRNSKTGT---ELDYLISPVADVNGNFFDSRQAP--KNMPQDADANGAYHIGLKGLML      1269
Moraxella_bovoculi                  LALRYSNASS---DEDFILSPVANDGGVFFNSALAD--DTQPQNADANGAYHIALKGLWL      1222
Prevotella_bryantii_B14             LQMRNSITGT---TTDYLISPVHDSKGNFYDSRICD--NSLPANADANGAYNIARKGLML      1212
Ruminococcus_bromii                 LQMRNSNPET---GEDRILSPVKDKNGNFYDSSKYDEKSKLPCDADANGAYNIARKGLWI      1190
                                    : :* *   :     * : *** :  *  :::         * :******* *. **

Acidaminococcus_sp._BV3L6           LNHLKESKDLKL-----QNGISNQDWLAYIQELRN---      1307
Lachnospiraceae_bacterium           VKQIKENWKENEQFPRNKLVQDNKTWFDFMQKKRYL--      1259
Francisella_novicida                LDRIKNNQEGKK---L-NLVIKNEEYFEFVQNRNN---      1300
Moraxella_bovoculi                  LNELKDSDDLNK---VK-LAIDNQTWLNFAQNR-----      1251
Prevotella_bryantii_B14             IQQIKDSTSSNR---FKFSPITNKDWLIFAQEKPYLND      1247
Ruminococcus_bromii                 VEQFKKADNVST---V-EPVIHNDKWLKFVQENDMANN      1224
                                    :..:*.  . .           *. :: : *:
```

Figure 60: Multiple sequence alignments of Cas12a variants from different organisms.
Multiple sequence alignments of Cas12a variants from different organisms were generated using the Clustal Omega tool [309] (https://www.ebi.ac.uk/Tools/msa/clustalo/, version 1.2.4). The aligned Cas12a sequences originate from *Acidaminococcus sp.* BV3L6 (AsCas12a), *Lachnospiracaea bacterium* (LbCas12a), *Francisella novicida* (FnCas12a), *Moraxella bovoculi* (MbCas12a), *Prevotella bryantii* B14 (PbCas12a), and *Ruminococcus bromii* (RbCas12a). The conserved amino acids of the DNase active center are framed in salmon and the conserved amino acids of the RNase active center are framed in yellow. W971 and K978 in the FnCas12a sequence and the corresponding conserved amino acids in the other sequences are framed in red. The residues of the protein sequences are color-coded according to their chemical properties. Red: small/hydrophobic (A, V, F, P, M, I, L, W), blue: acidic (D, E), pink: basic (R, K), green: hydroxyl/sulfhydryl/amine (S, T, Y, H, C, N, G, Q). * Residues conserved throughout all sequences, : highly conserved residues with groups of similar properties, . conserved residues with groups of weakly similar properties.

III. Extended data

In this section, additional data and data corresponding to figures in the results section are shown.

III.i. Protein purification

WT FnCas12a and FnCas12a variants were expressed in *E. coli* BL21(DE3) or *E. coli* BL21(DE3)-pEVOL-pAzF cells and purified in two steps via affinity chromatography. First, the N-terminally GST-tagged protein was purified via glutathione affinity chromatography. The GST-tag was thereafter removed by thrombin protease cleavage. Eventually, the protein was purified via heparin affinity chromatography to dispose the cleaved-off GST-tag and to reduce nucleic acid levels (**Figure 61**).

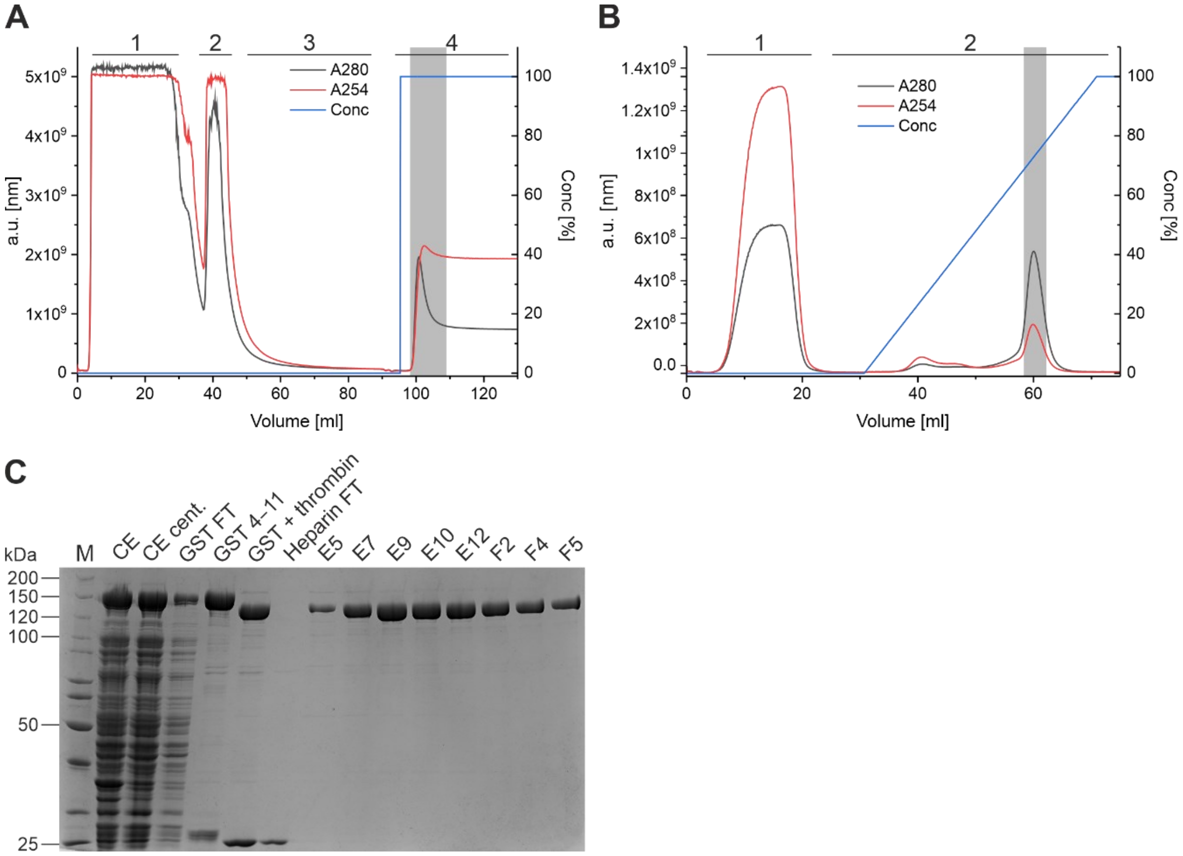


Figure 61: Example for the purification of FnCas12a.
(**A**) Chromatogram of the WT FnCas12a purification via glutathione affinity chromatography. The absorption measurement at 280 nm is colored in black and the absorption measurement at 254 nm in red, both corresponding to the left y-axis. The concentration of elution buffer is displayed in blue corresponding to the right y-axis. 1: sample application, 2: washing step with high salt buffer, 3: washing step with CasA buffer, 4: elution with CasB buffer. The greyed area marks the eluted fractions for further purification. (**B**) Chromatogram of the WT FnCas12a purification via heparin affinity chromatography. The absorption measurement at 280 nm is colored in black and the absorption measurement at 254 nm in red, both corresponding to the left y-axis. The concentration of elution buffer is displayed in blue corresponding to the right y-axis. 1: sample application, 2: elution with high salt buffer. The greyed area marks the eluted fractions that were stored at −80 °C and used for protein characterization. (**C**) 10% SDS

page of the WT FnCas12a purification. CE: crude extract, CE cent.: centrifuged crude extract, GST FT: flowthrough during glutathione affinity chromatography sample application, GST 4–11: pooled fraction 4–11 of the glutathione affinity chromatography for further purification (corresponding to fractions of the greyed area in **A**), GST + thrombin: pooled fraction 4–11 of the glutathione affinity chromatography after digestion with thrombin protease, FnCas12a runs at approximately 150 kDa, the cleaved-off GST-tag at 25 kDa, Heparin FT: flowthrough during heparin affinity chromatography sample application, E5–F5: elution fractions containing purified WT FnCas12a (corresponding to fractions of the greyed area in **B**), M: PageRuler unstained (ThermoFisher Scientific).

III.ii. Single-molecule FRET measurements on a doubly labeled RNA/DNA hybrid

Table 38. Molecule counts and fitting parameters for single-molecule FRET measurements with an RNA/DNA hybrid.
The FRET efficiency histograms (corresponding to **Figure 21**) were fitted with one or two Gaussian distributions. The number of molecules N_M used to calculate each histogram, the coefficient of determination R^2, the mean FRET efficiency E (± s.e.), and the peak area A (± s.e.) of each Gaussian are listed. * Indicates fixed values for the fit.

Sample	N_M	E_1 E_2	± s.e. E_1 ± s.e. E_2	A_1 A_2	± s.e. A_1 ± s.e. A_2	R^2
hybrid	4863	0.776	0.003	0.031	0.0007	0.9824
+25 pM Cas12a	8375	0.200 0.700	0.003 0.003	0.011 0.022	0.0004 0.0005	0.9824
+50 pM Cas12a	9203	0.193 0.706	0.003 0.008	0.018 0.015	0.0005 0.0007	0.9751
+100 pM Cas12a	10557	0.212 0.77*	0.002 –	0.027 0.005	0.0005 0.0005	0.9877

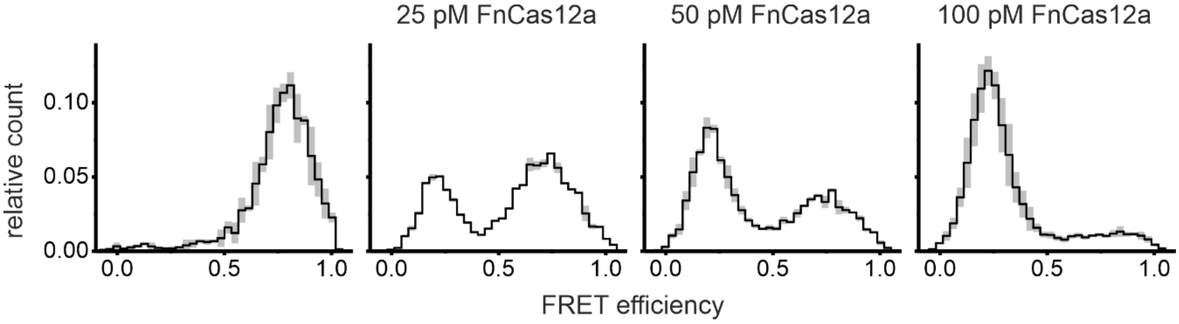


Figure 62: FRET efficiency histograms with standard deviation for measurements of the RNA/DNA hybrid with WT FnCas12a.
FRET efficiency histograms of the measurements with a doubly labeled RNA/DNA hybrid and the addition of 25 pM, 50 pM, and 100 pM of WT FnCas12a protein (corresponding to **Figure 21**). Histograms show the mean FRET efficiency (solid line) and the standard deviation (gray area) of three replicates.

III.iii. Single-molecule FRET measurements on doubly labeled WT FnCas12a

The smFRET measurements shown in **Figure 25** were additionally conducted with different concentrations of RNA and DNA. FnCas12a$^{\text{WED-Nuc*DL550/DL650}}$ and FnCas12a$^{\text{WED-REC*DL550/DL650}}$ display one single FRET population for all measured conditions in the apo, RNA-, and RNA-and-DNA-bound state that do not drastically change with the increase of RNA or DNA concentrations (**Figure 63**). FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ shows an increase in the high FRET population with the increase of crRNA, whereas increasing the target DNA concentrations leads to a decrease in the high FRET population and an increase in the medium FRET population (green).

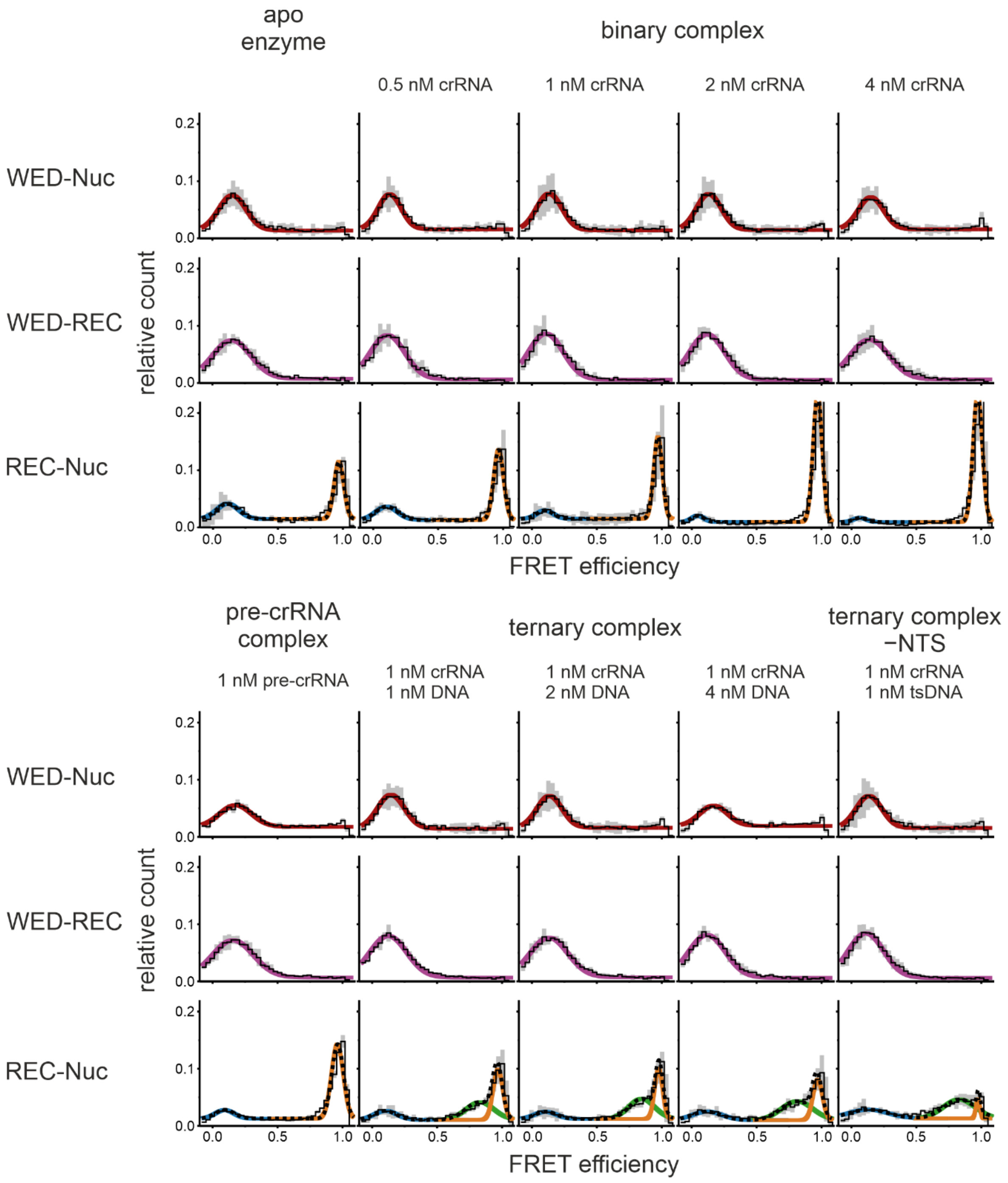


Figure 63: Confocal single-molecule FRET measurements using doubly labeled WT FnCas12a with different RNA/DNA concentrations.

FRET efficiency distributions for the FnCas12a apo enzyme, FnCas12a with pre-crRNA, the binary complex, the ternary complex, and the ternary complex without NTS. FnCas12a$^{\text{WED-Nuc*DL550/DL650}}$ and FnCas12a$^{\text{WED-REC*DL550/DL650}}$ display one single FRET population for all tested conditions. FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ shows two FRET populations for the apo enzyme and the RNA-bound state, with the high FRET population increasing with the addition of RNA and further increasing with higher crRNA concentrations. In the ternary complex, FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ shows three FRET populations with an additional medium FRET population that increases with higher target DNA concentrations, whereas the high FRET population decreases. Histograms show the average data of three independent measurements. The histograms were fitted with a single, double, or triple Gaussian function. The solid black line shows the mean FRET efficiency, colored lines show the Gaussian fit, the black dotted line the overall fit, and grey bars the standard deviation. Additional information regarding fit parameters and molecule counts can be found in **Table 39**.

Table 39. Molecule counts and fitting parameters for FRET measurements with doubly labeled WT FnCas12a.

The FRET efficiency histograms (corresponding to **Figure 25** and **Figure 63**) were fitted with one, two, or three Gaussian distributions. The number of molecules N_M used to calculate each histogram, the coefficient of determination R^2, the mean FRET efficiency E (± s.e.), and the peak area A (± s.e.) of each Gaussian are listed. Indicated are concentrations of RNA, and DNA, respectively, 'apo' stands for apo FnCas12a, 'preRNA' for FnCas12a +pre-crRNA, 'bin' for the binary complex, 'ter' for the ternary complex, and '−NTS' for the ternary complex without NTS. * Indicates fixed values for the fit.

Sample	N_M	E_1 E_2 E_3	± s.e. E_1 ± s.e. E_2 ± s.e. E_3	A_1 A_2 A_3	± s.e. A_1 ± s.e. A_2 ± s.e. A_3	R^2
WED-Nuc apo	7035	0.153	0.005	0.016	0.0008	0.9556
preRNA 1 nM	9007	0.177	0.009	0.012	0.0011	0.8737
bin 0.5 nM	6172	0.134	0.005	0.014	0.0008	0.9357
bin 1 nM	5435	0.133	0.005	0.016	0.0009	0.9492
bin 2 nM	5871	0.132	0.005	0.015	0.0009	0.9402
bin 4 nM	5787	0.150	0.007	0.014	0.0010	0.9051
ter 1 nM, 1 nM	4541	0.145	0.005	0.015	0.0008	0.9493
ter 1 nM, 2 nM	4647	0.135	0.006	0.013	0.0009	0.9235
ter 1 nM, 4 nM	7020	0.169	0.012	0.009	0.0012	0.7710
−NTS 1 nM, 1 nM	5153	0.135	0.006	0.014	0.0009	0.9217
WED-REC apo	4567	0.148	0.004	0.025	0.0008	0.9855
preRNA 1 nM	7600	0.156	0.003	0.026	0.0007	0.9894
bin 0.5 nM	4787	0.118	0.004	0.026	0.0009	0.9800
bin 1 nM	5083	0.107	0.003	0.027	0.0009	0.9837
bin 2 nM	5861	0.119	0.003	0.027	0.0007	0.9899
bin 4 nM	8143	0.139	0.004	0.026	0.0008	0.9850
ter 1 nM, 1 nM	6587	0.124	0.003	0.026	0.0007	0.9885
ter 1 nM, 2 nM	6412	0.131	0.004	0.025	0.0008	0.9828
ter 1 nM, 4 nM	6345	0.118	0.004	0.027	0.0008	0.9855
−NTS 1 nM, 1 nM	6599	0.112	0.003	0.027	0.0007	0.9898

REC-Nuc	apo	4439	0.113	0.163	0.006	0.0012	0.9038
			0.969	0.003	0.009	0.0007	
	preRNA 1 nM	15610	0.093	0.036	0.003	0.0015	0.9110
			0.960	0.003	0.015	0.0011	
	bin 0.5 nM	5541	0.103	0.019	0.005	0.0012	0.9285
			0.965	0.003	0.012	0.0008	
	bin 1 nM	7646	0.101	0.027	0.002	0.0010	0.9409
			0.970	0.002	0.012	0.0007	
	bin 2 nM	10195	0.046	0.038	0.001	0.0010	0.9636
			0.971	0.002	0.019	0.0009	
	bin 4 nM	10868	0.064	0.062	0.001	0.0012	0.9587
			0.973	0.002	0.020	0.0008	
	ter 1 nM, 1 nM	7856	0.103	0.205	0.004	0.0012	0.9556
			0.831	0.033	0.008	0.0028	
			0.969	0.003	0.008	0.0018	
	ter 1 nM, 2 nM	7137	0.104	0.026	0.003	0.0011	0.9504
			0.845	0.022	0.009	0.0022	
			0.975	0.002	0.007	0.0013	
	ter 1 nM, 4 nM	8114	0.120	0.026	0.004	0.0008	0.9324
			0.812	0.026	0.010	0.0018	
			0.970	0.003	0.006	0.0013	
	−NTS 1 nM, 1 nM	7271	0.13*	−	0.006	0.0009	0.7289
			0.83*	−	0.010	0.0011	
			0.974	0.004	0.001	0.0004	

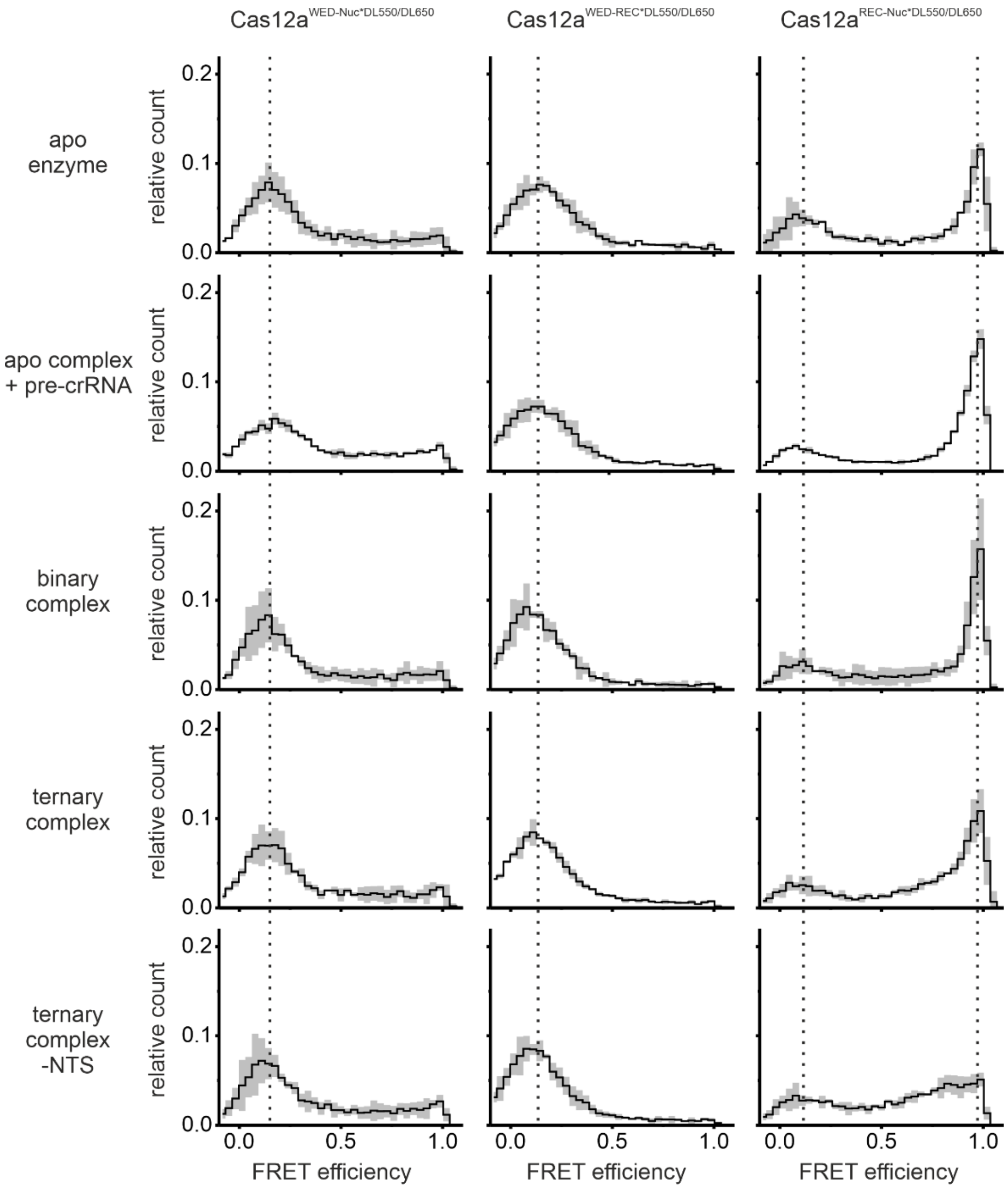


Figure 64: FRET efficiency histograms with standard deviation for measurements of doubly labeled WT FnCas12a.

FRET efficiency histograms of the measurements with doubly labeled FnCas12a$^{WED-Nuc*DL550/DL650}$, FnCas12a$^{WED-REC*DL550/DL650}$, and FnCas12a$^{REC-Nuc*DL550/DL650}$ for the apo enzyme, with pre-crRNA (1 nM pre-crRNA), the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM target DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS DNA) (corresponding to **Figure 25**). Histograms show the mean FRET efficiency (solid line) and the standard deviation (gray area) of three replicates. (adapted from Wörle *et al.* [290])

III.iv. Single-molecule FRET measurements on doubly labeled FnCas12a HKK variants

Table 40. Molecule counts and fitting parameters for FRET measurements with doubly labeled FnCas12a HKK variants.
The FRET efficiency histograms (corresponding to **Figure 27**) were fitted with one or two Gaussian distributions. The number of molecules N_M used to calculate each histogram, the coefficient of determination R^2, the mean FRET efficiency E (± s.e.), and the peak area A (± s.e.) of each Gaussian are listed. ‘apo’ stands for apo FnCas12a, ‘pre-crRNA’ for FnCas12a +pre-crRNA (1 nM), and ‘crRNA’ for the binary complex of FnCas12a +crRNA (1 nM).

Sample		N_M	E_1 E_2	± s.e. E_1 ± s.e. E_2	A_1 A_2	± s.e. A_1 ± s.e. A_2	R^2
WED-Nuc	apo	22133	0.194	0.003	0.0201	0.0005	0.9883
HKK	pre-crRNA	25775	0.203	0.003	0.0201	0.0006	0.9880
	crRNA	19547	0.181	0.003	0.0283	0.0006	0.9894
WED-REC	apo	5468	0.125	0.006	0.0100	0.0007	0.9105
HKK	pre-crRNA	6593	0.141	0.008	0.0108	0.0009	0.8819
	crRNA	5459	0.141	0.008	0.0117	0.0009	0.8979
REC-Nuc HKK	apo	29951	0.091 0.969	0.023 0.004	0.0021 0.0105	0.0009 0.0008	0.9070
	pre-crRNA	31284	0.077 0.983	0.025 0.002	0.0016 0.0124	0.0007 0.0006	0.9567
	crRNA	30146	0.075 0.983	0.027 0.002	0.0015 0.0125	0.0008 0.0007	0.9512

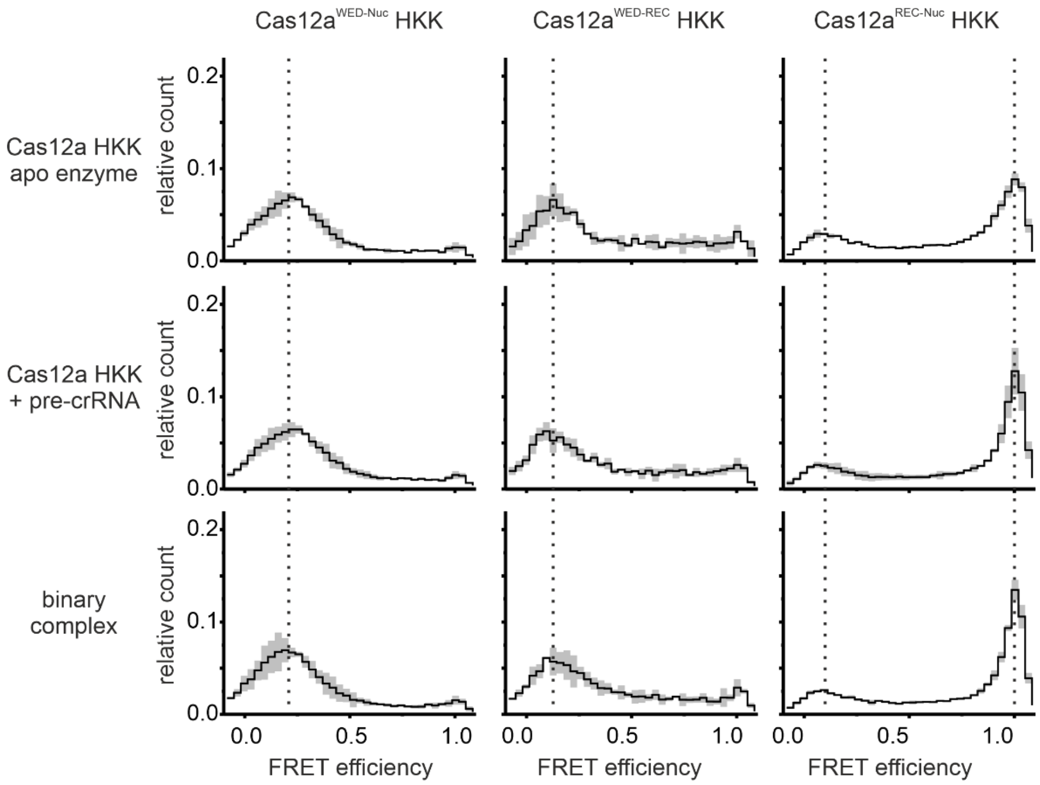


Figure 65: FRET efficiency histograms with standard deviation for measurements of doubly labeled FnCas12a HKK variants.

FRET efficiency histograms of the measurements with doubly labeled FnCas12a$^{WED-Nuc*DL550/DL650}$ HKK, FnCas12a$^{WED-REC*DL550/DL650}$ HKK, and FnCas12a$^{REC-Nuc*DL550/DL650}$ HKK for the apo enzyme, with pre-crRNA (1 nM pre-crRNA), and the binary complex (1 nM crRNA) (corresponding to **Figure 27**). Histograms show the mean FRET efficiency (solid line) and the standard deviation (gray area) of three replicates.

III.v. Single-molecule FRET measurements on doubly labeled dCas12a

The smFRET measurements shown in **Figure 28** were additionally conducted with different concentrations of crRNA and DNA. dCas12a$^{\text{REC-Nuc*DL550/DL650}}$ shows the same populations and RNA/DNA dependency as WT FnCas12a does (**Figure 66**).

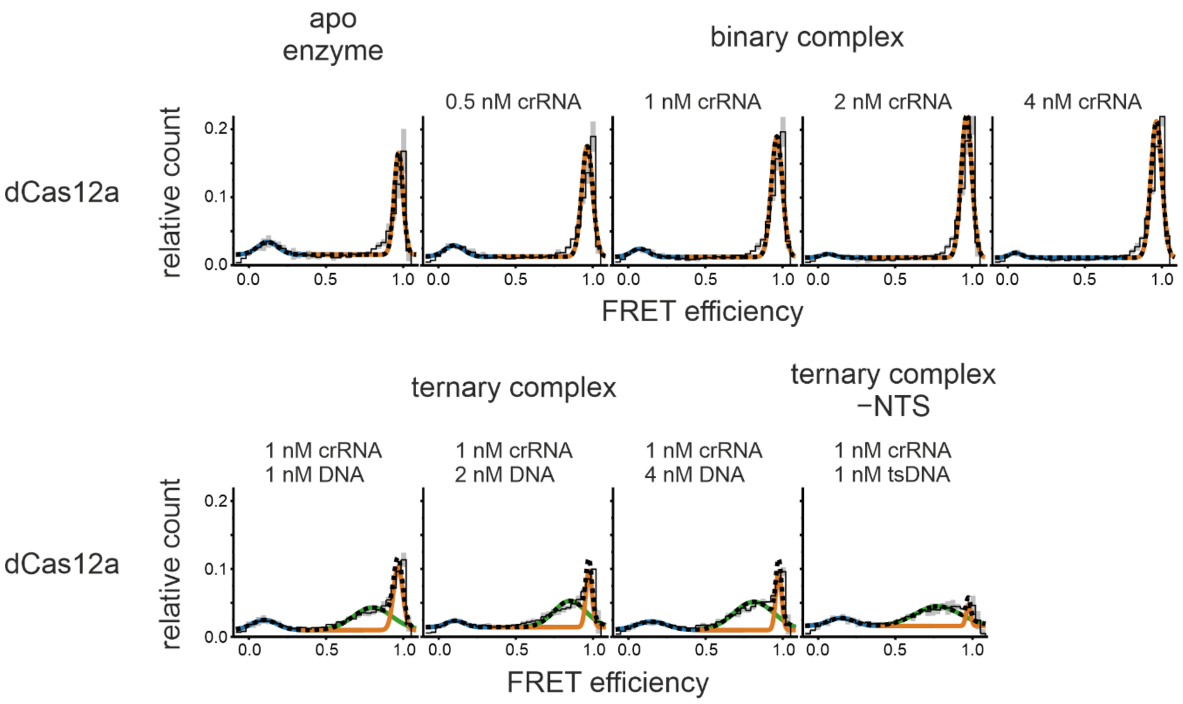


Figure 66: Confocal single-molecule FRET measurements using doubly labeled dCas12a with different RNA/DNA concentrations.
FRET efficiency distributions for the dCas12a apo enzyme, the binary complex, the ternary complex, and the ternary complex without NTS. dCas12a$^{\text{REC-Nuc*DL550/DL650}}$ shows two FRET populations for the apo enzyme and the RNA-bound state, with the high FRET population increasing with the addition of crRNA and further increasing with higher crRNA concentrations. In the ternary complex, dCas12a$^{\text{REC-Nuc*DL550/DL650}}$ shows three FRET populations with an additional medium FRET population that increases with higher target DNA concentrations, whereas the high FRET population decreases. Histograms show the average data of three independent measurements. The histograms were fitted with a double or triple Gaussian function. The solid black line shows the mean FRET efficiency, colored lines show the Gaussian fit, the black dotted line the overall fit, and grey bars the standard deviation. Additional information regarding fit parameters and molecule counts can be found in **Table 41**.

Table 41. Molecule counts and fitting parameters for FRET measurements with doubly labeled dCas12a.
The FRET efficiency histograms (corresponding to **Figure 28** and **Figure 67**) were fitted with two or three Gaussian distributions. The number of molecules N_M used to calculate each histogram, the coefficient of determination R^2, the mean FRET efficiency E (± s.e.), and the peak area A (± s.e.) of each Gaussian are listed. Indicated are concentrations of crRNA, and DNA, respectively, 'apo' stands for apo dCas12a, 'bin' for the binary complex, 'ter' for the ternary complex, and '−NTS' for the ternary complex without NTS. * Indicates fixed values for the fit.

Sample		N_M	E_1 E_2 E_3	± s.e. E_1 ± s.e. E_2 ± s.e. E_3	A_1 A_2 A_3	± s.e. A_1 ± s.e. A_2 ± s.e. A_3	R^2
dCas12a	apo	6783	0.121 0.970	0.003 0.003	0.0031 0.0107	0.0017 0.0010	0.8601

bin 0.5 nM	8967	0.102	0.043	0.0026	0.0019	0.8717
		0.964	0.003	0.0144	0.0013	
bin 1 nM	9838	0.075	0.052	0.0017	0.0016	0.9019
		0.964	0.003	0.018	0.0012	
bin 2 nM	11403	0.065	0.115	0.0010	0.0010	0.9147
		0.964	0.003	0.0181	0.0013	
bin 4 nM	11621	0.049	0.081	0.0007	0.0014	0.9125
		0.965	0.003	0.0181	0.0013	
ter 1 nM, 1 nM	11171	0.100	0.310	0.003	0.0010	0.9013
		0.801	0.029	0.010	0.0021	
		0.967	0.003	0.007	0.0013	
ter 1 nM, 2 nM	11861	0.107	0.030	0.0014	0.0011	0.8908
		0.85*	−	0.0099	0.0019	
		0.974	0.003	0.0041	0.0006	
ter 1 nM, 4 nM	11500	0.149	0.035	0.0031	0.0008	0.9067
		0.813	0.017	0.0134	0.0015	
		0.973	0.002	0.0041	0.0007	
−NTS 1 nM, 1 nM	12518	0.149	0.027	0.0024	0.0007	0.8059
		0.774	0.016	0.0106	0.0011	
		0.973	0.0036	0.0011	0.0003	

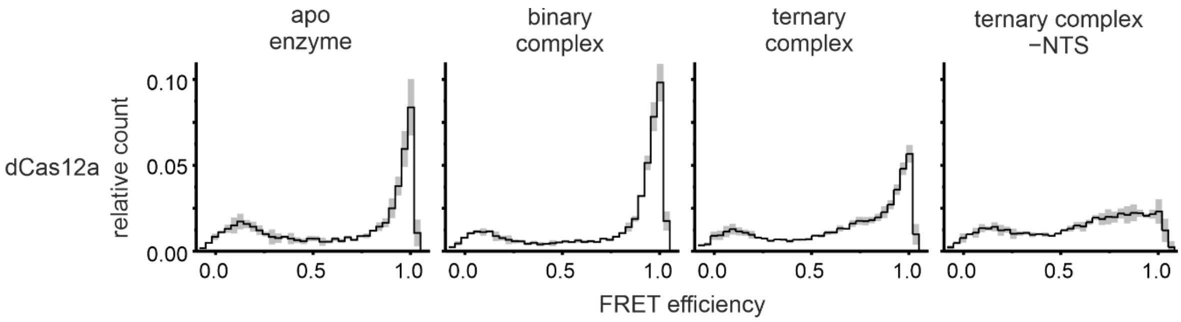
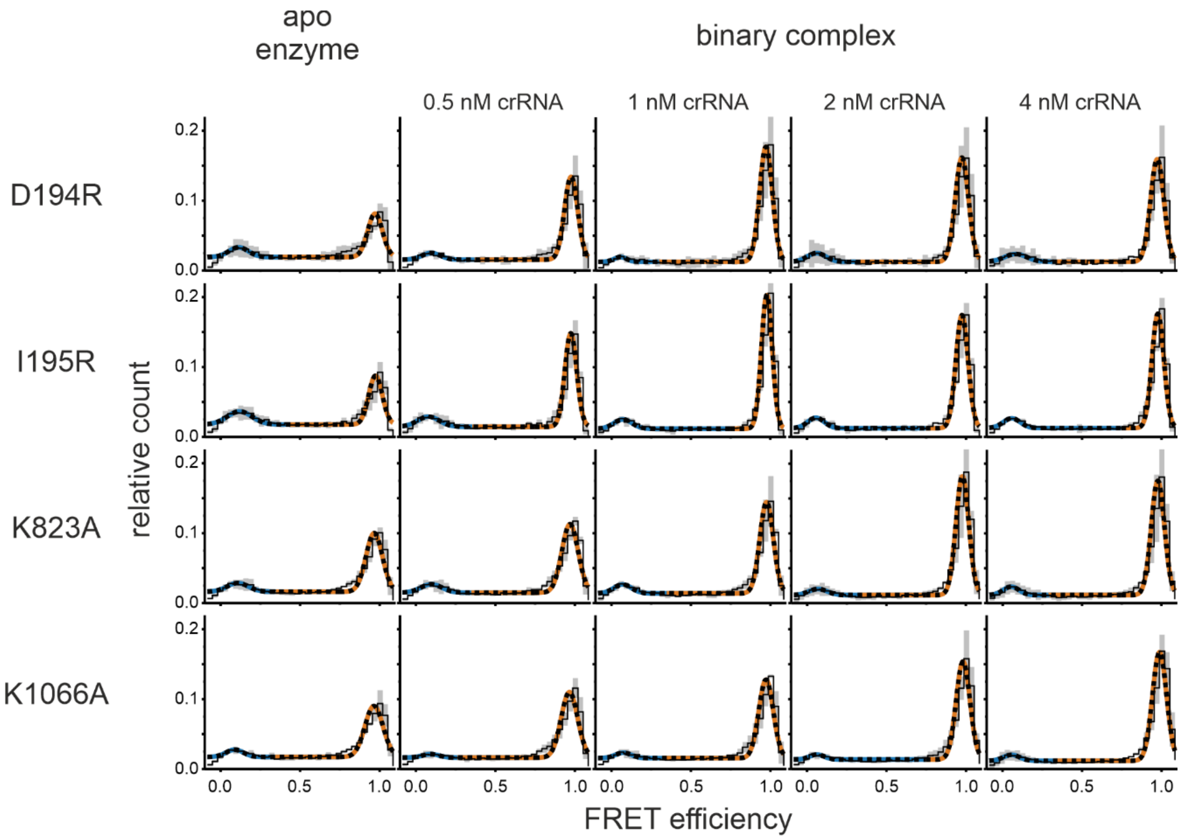


Figure 67: FRET efficiency histograms with standard deviation for measurements of doubly labeled dCas12a.

FRET efficiency histograms of the measurements with doubly labeled dCas12a$^{\text{REC-Nuc*DL550/DL650}}$ for the apo enzyme, the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS DNA) (corresponding to **Figure 28**). Histograms show the mean FRET efficiency (solid line) and the standard deviation (gray area) of three replicates.

III.vi. Single-molecule FRET measurements on doubly labeled FnCas12a variants to alter the target strand affinity

The smFRET measurements shown in **Figure 31** were additionally conducted with different concentrations of crRNA and DNA. All FnCas12a$^{REC\text{-}Nuc*DL550/DL650}$ variants with altered TS affinity show an increase in the high FRET population with the increase of crRNA, whereas increasing target DNA concentrations lead to a decrease in the high FRET population and an increase in the medium FRET population (**Figure 68**).



The figure continues on the next page.

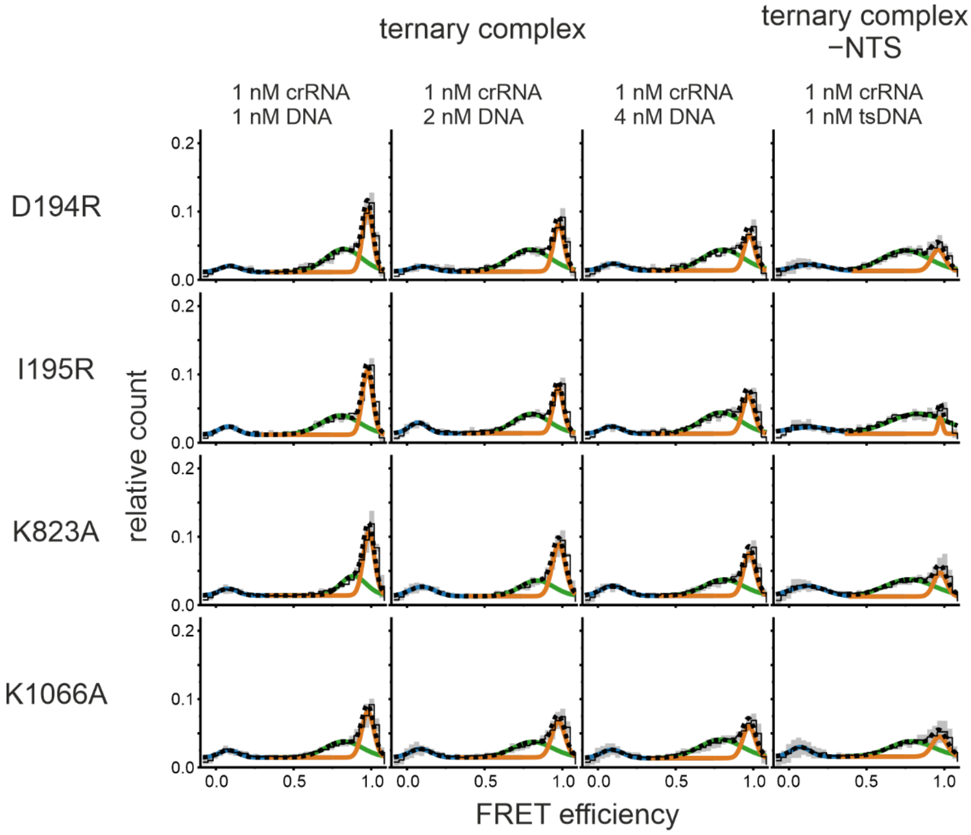


Figure 68: Confocal single-molecule FRET measurements using doubly labeled FnCas12a variants to alter the target strand affinity with different RNA/DNA concentrations.
FRET efficiency distributions for the apo enzyme of FnCas12a variants with altered TS affinity, the binary complex, the ternary complex, and the ternary complex without NTS. FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ variants with altered TS affinity show two FRET populations for the apo enzyme and the RNA-bound state, with the high FRET population increasing with the addition of crRNA and further increasing with higher crRNA concentrations. In the ternary complex, FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ variants with altered TS affinity show three FRET populations with an additional medium FRET population that increases with higher target DNA concentrations, whereas the high FRET population decreases. Histograms show the average data of three independent measurements. The histograms were fitted with a double or triple Gaussian function. The solid black line shows the mean FRET efficiency, colored lines show the Gaussian fit, the black dotted line the overall fit, and grey bars the standard deviation. Additional information regarding fit parameters and molecule counts can be found in **Table 42**.

Table 42. Molecule counts and fitting parameters for FRET measurements with doubly labeled FnCas12a variants to alter the target strand affinity.
The FRET efficiency histograms (corresponding to **Figure 31** and **Figure 68**) were fitted with two or three Gaussian distributions. The number of molecules N_M used to calculate each histogram, the coefficient of determination R^2, the mean FRET efficiency E (± s.e.), and the peak area A (± s.e.) of each Gaussian are listed. Indicated are concentrations of crRNA, and DNA, respectively, 'apo' stands for apo FnCas12a, 'bin' for the binary complex, 'ter' for the ternary complex, and '−NTS' for the ternary complex without NTS. * Indicates fixed values for the fit.

Sample		**N_M**	**E_1** **E_2** **E_3**	**± s.e. E_1** **± s.e. E_2** **± s.e. E_3**	**A_1** **A_2** **A_3**	**± s.e. A_1** **± s.e. A_2** **± s.e. A_3**	**R^2**
D194R	apo	10810	0.113 0.971	0.029 0.006	0.0021 0.0072	0.0011 0.0009	0.7874
	bin 0.5 nM	10524	0.096 0.976	0.029 0.006	0.0014 0.0123	0.0009 0.0007	0.7874

	bin 1 nM	7859	0.055 0.973	0.043 0.002	0.0007 0.0167	0.0007 0.0007	0.9654
	bin 2 nM	7378	0.060 0.974	0.027 0.002	0.0018 0.0157	0.0009 0.0007	0.9655
	bin 4 nM	7714	0.083 0.973	0.034 0.002	0.0019 0.0154	0.0010 0.0007	0.9621
	ter 1 nM, 1 nM	9342	0.091 0.823 0.977	0.029 0.021 0.002	0.0014 0.0098 0.0076	0.0008 0.0021 0.0011	0.9643
	ter 1 nM, 2 nM	9271	0.1* 0.792 0.977	− 0.002 0.003	0.0013 0.0166 0.0059	0.0010 0.0022 0.0009	0.9479
	ter 1 nM, 4 nM	7936	0.1* 0.804 0.975	− 0.027 0.004	0.0018 0.0100 0.0046	0.0011 0.0026 0.0012	0.9020
	−NTS 1 nM, 1 nM	10103	0.14* 0.749 0.960	− 0.019 0.005	0.0026 0.0114 0.0034	0.0005 0.0012 0.0009	0.9303
I195R	apo	10056	0.119 0.974	0.023 0.005	0.0037 0.0075	0.0012 0.0008	0.829
	bin 0.5 nM	8082	0.079 0.974	0.024 0.002	0.0025 0.0124	0.0009 0.0006	0.954
	bin 1 nM	8610	0.071 0.978	0.026 0.001	0.0019 0.0163	0.0009 0.0007	0.9717
	bin 2 nM	6153	0.056 0.975	0.023 0.002	0.0019 0.0153	0.0009 0.0007	0.9617
	bin 4 nM	8617	0.058 0.977	0.027 0.002	0.0019 0.0153	0.0010 0.0008	0.9561
	ter 1 nM, 1 nM	7778	0.078 0.802 0.975	0.021 0.024 0.002	0.0018 0.0088 0.0078	0.0008 0.0021 0.0010	0.9599
	ter 1 nM, 2 nM	7880	0.072 0.080 0.973	0.015 0.019 0.003	0.0022 0.0082 0.0055	0.0007 0.0016 0.0009	0.9443
	ter 1 nM, 4 nM	7101	0.088 0.790 0.967	0.024 0.024 0.004	0.0017 0.0099 0.0052	0.0009 0.0023 0.0012	0.9303
	−NTS 1 nM, 1 nM	8471	0.13* 0.83* 0.973	− − 0.005	0.0037 0.0138 0.0008	0.0008 0.0012 0.0003	0.7730
K823A	apo	19992	0..08 0.966	0.029 0.004	0.0021 0.0108	0.0010 0.0008	0.9112

	bin 0.5 nM	22762	0.097	0.033	0.0020	0.0011	0.9171
			0.969	0.004	0.0126	0.0009	
	bin 1 nM	20991	0.071	0.026	0.0018	0.0008	0.9555
			0.976	0.002	0.0139	0.0007	
	bin 2 nM	20634	0.069	0.040	0.0013	0.0009	0.9706
			0.978	0.002	0.0173	0.0007	
	bin 4 nM	18530	0.052	0.025	0.0015	0.0007	0.9731
			0.978	0.002	0.0170	0.0007	
	ter 1 nM, 1 nM	24269	0.079	0.021	0.0015	0.0006	0.9778
			0.885	0.039	0.0058	0.0026	
			0.984	0.002	0.0083	0.0022	
	ter 1 nM, 2 nM	19778	0.097	0.017	0.0028	0.0008	0.9665
			0.837	0.043	0.0062	0.0024	
			0.977	0.003	0.0080	0.0017	
	ter 1 nM, 4 nM	15288	0.096	0.017	0.0028	0.0008	0.9462
			0.809	0.030	0.0075	0.0022	
			0.974	0.003	0.0059	0.0011	
	−NTS 1 nM, 1 nM	15151	0.120	0.018	0.0044	0.0006	0.9001
			0.775	0.026	0.0100	0.0014	
			0.972	0.005	0.0015	0.0009	
K1066A	apo	19760	0.089	0.032	0.0014	0.0009	0.8894
			0.964	0.005	0.0103	0.0009	
	bin 0.5 nM	23674	0.092	0.075	0.0006	0.0010	0.9066
			0.966	0.004	0.0121	0.0009	
	bin 1 nM	24743	0.075	0.043	0.0010	0.0009	0.9341
			0.972	0.003	0.0129	0.0008	
	bin 2 nM	25637	0.062	0.041	0.0008	0.0007	0.9639
			0.981	0.002	0.0152	0.0007	
	bin 4 nM	24320	0.056	0.035	0.0011	0.0008	0.9722
			0.987	0.002	0.0172	0.0007	
	ter 1 nM, 1 nM	27742	0.081	0.021	0.0017	0.0070	0.9530
			0.824	0.034	0.0070	0.0022	
			0.978	0.003	0.0065	0.0013	
	ter 1 nM, 2 nM	24652	0.087	0.020	0.0021	0.0009	0.9268
			0.810	0.041	0.0072	0.0026	
			0.975	0.004	0.0056	0.0016	
	ter 1 nM, 4 nM	20780	0.089	0.019	0.0021	0.0009	0.9287
			0.791	0.028	0.0094	0.0025	
			0.972	0.004	0.0047	0.0012	
	−NTS 1 nM, 1 nM	20789	0.074	0.013	0.0022	0.0006	0.9224
			0.789	0.035	0.0078	0.0022	
			0.006	0.006	0.0035	0.0013	

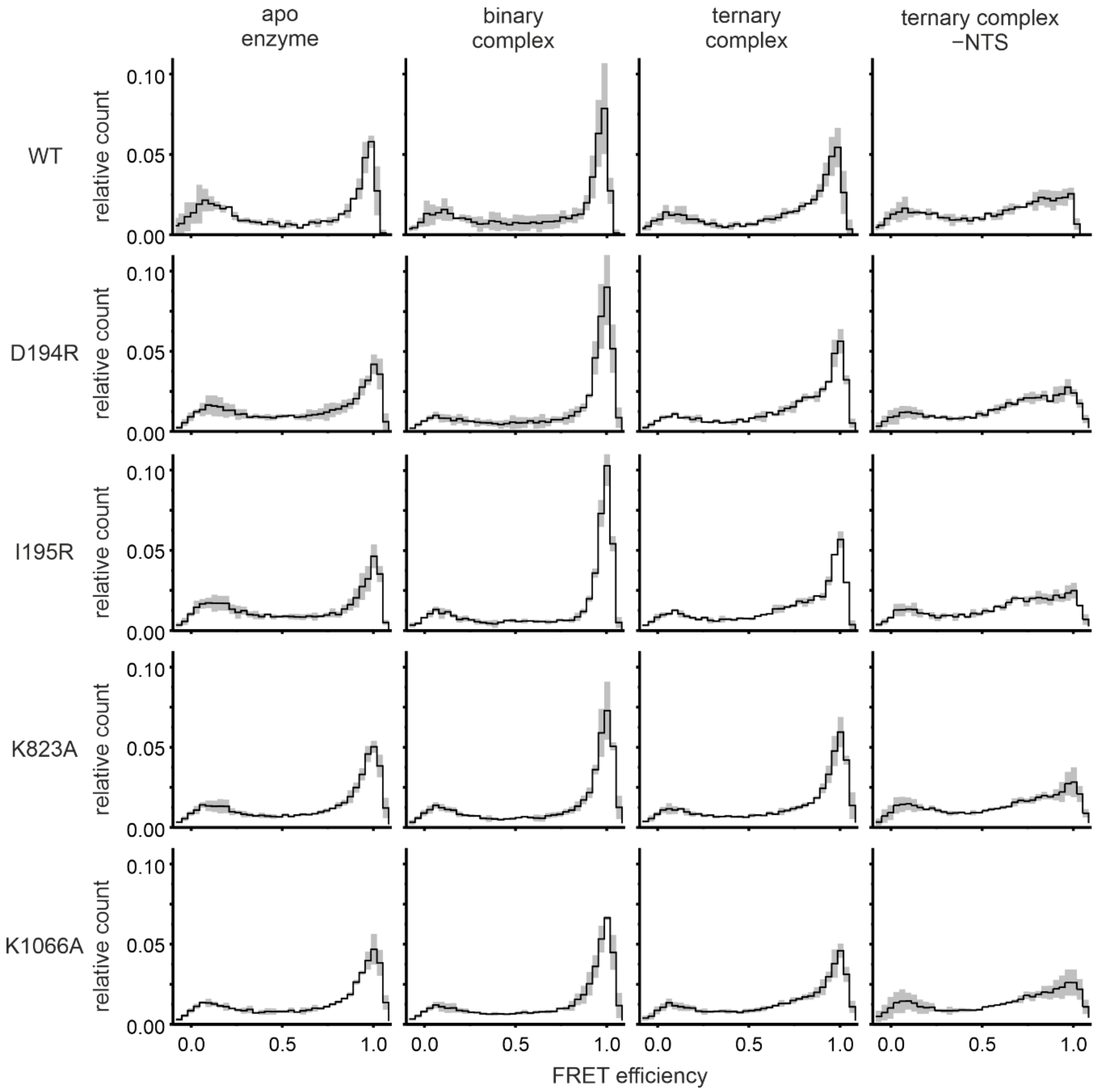
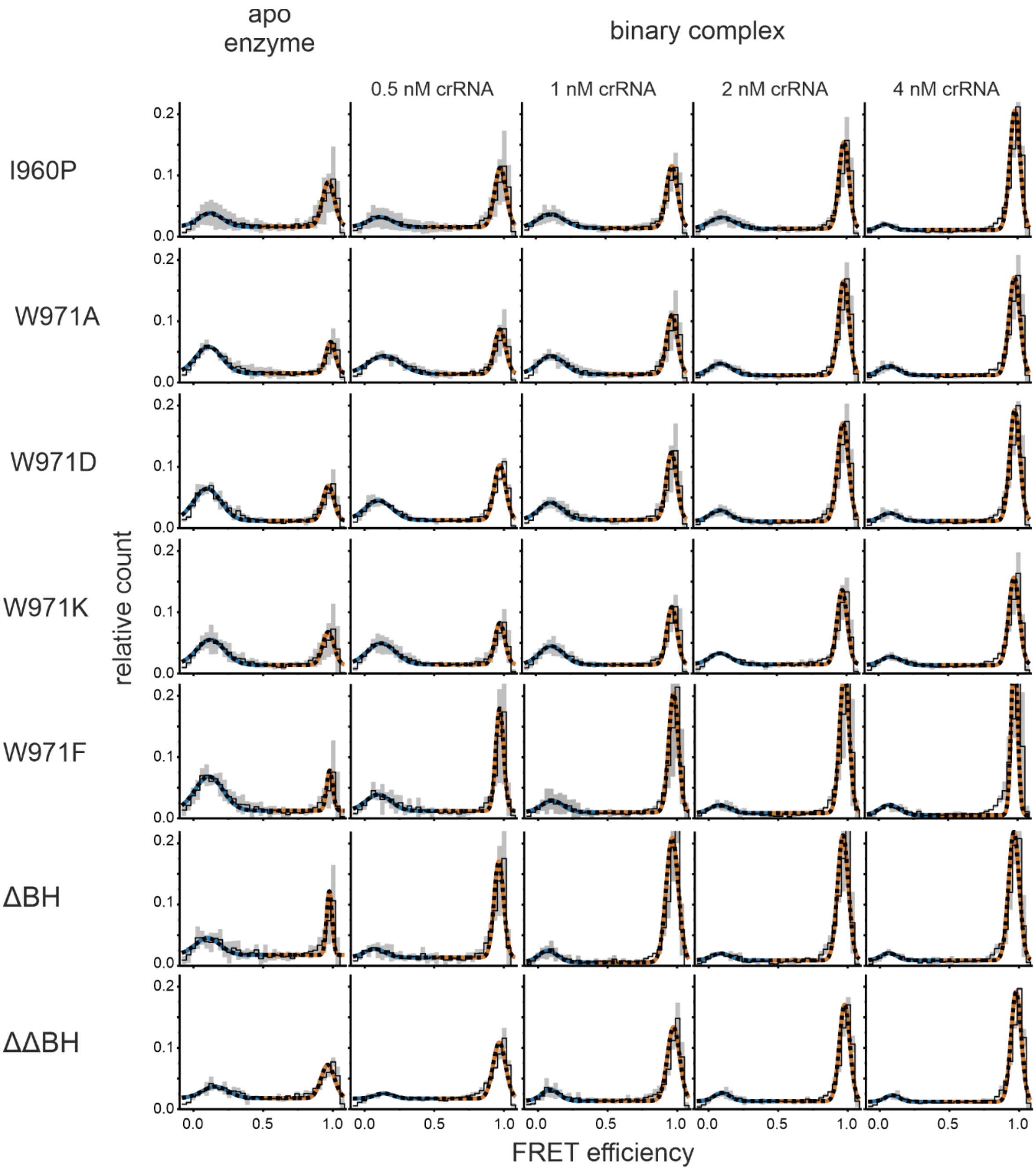


Figure 69: FRET efficiency histograms with standard deviation for measurements of doubly labeled FnCas12a variants with altered TS affinity.
FRET efficiency histograms of the measurements with doubly labeled FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ variants with altered TS affinity for the apo enzyme, the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS DNA) (corresponding to **Figure 31**). Histograms show the mean FRET efficiency (solid line) and the standard deviation (gray area) of three replicates.

III.vii. Single-molecule FRET measurements on doubly labeled FnCas12a bridge helix variants

The smFRET measurements shown in **Figure 40** were additionally conducted with different concentrations of crRNA and DNA. All FnCas12a$^{REC-Nuc*DL550/DL650}$ BH variants show an increase in the high FRET population with the increase of crRNA, whereas increasing target DNA concentrations lead to a decrease in the high FRET population and an increase in the medium FRET population (**Figure 70**).



The figure continues on the next page.

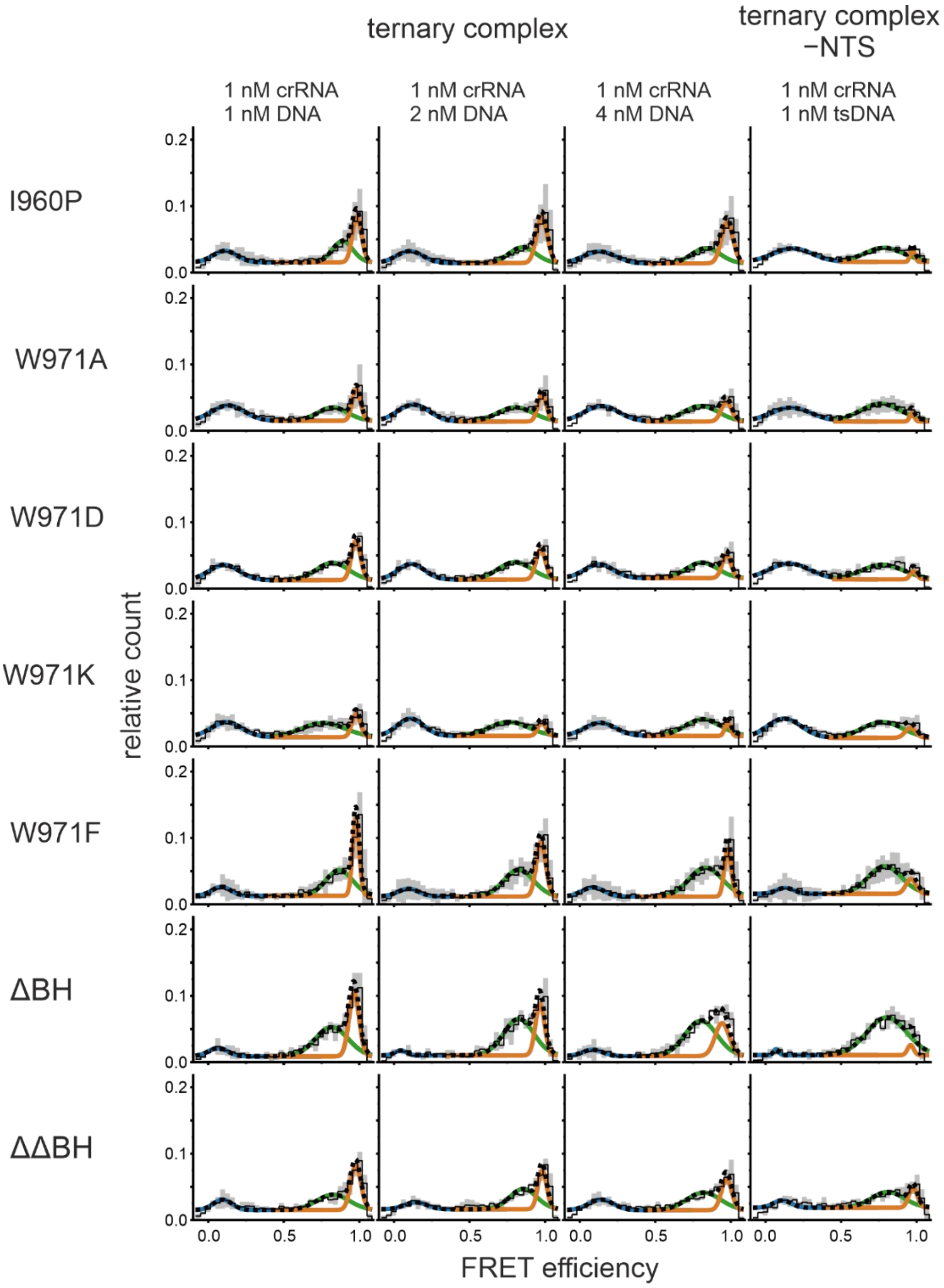


Figure 70: Confocal single-molecule FRET measurements using doubly labeled FnCas12a bridge helix variants with different RNA/DNA concentrations.

FRET efficiency distributions for the apo enzyme of FnCas12a BH variants, the binary complex, the ternary complex, and the ternary complex without NTS. FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants show two FRET populations for the apo enzyme and the RNA-bound state, with the high FRET population increasing with the addition of crRNA and further increasing with higher crRNA concentrations. In the ternary complex, FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants show three FRET populations with an additional medium FRET population that increases with higher target DNA concentrations, whereas the high FRET population decreases. The populations are similar to the WT but differ in their distributions. Histograms show the average data of three independent measurements. The histograms were fitted with a double or triple Gaussian function. The solid black line shows the mean FRET efficiency, colored lines show the Gaussian fit, the black dotted line the overall fit, and grey bars the standard deviation. Additional information regarding fit parameters and molecule counts can be found in **Table 43**.

(adapted from Wörle *et al.* [290])

Table 43. Molecule counts and fitting parameters for FRET measurements with doubly labeled FnCas12a BH variants.

The FRET efficiency histograms (corresponding to **Figure 40** and **Figure 70**) were fitted with two or three Gaussian distributions. The number of molecules N_M used to calculate each histogram, the coefficient of determination R^2, the mean FRET efficiency E (± s.e.), and the peak area A (± s.e.) of each Gaussian are listed. Indicated are concentrations of crRNA, and DNA, respectively, 'apo' stands for apo FnCas12a, 'bin' for the binary complex, 'ter' for the ternary complex, and '−NTS' for the ternary complex without NTS. * Indicates fixed values for the fit.

Sample		N_M	E_1 E_2 E_3	± s.e. E_1 ± s.e. E_2 ± s.e. E_3	A_1 A_2 A_3	± s.e. A_1 ± s.e. A_2 ± s.e. A_3	R^2
I960P	apo	9702	0.121 0.969	0.020 0.004	0.0047 0.0079	0.0013 0.0008	0.8516
	bin 0.5 nM	11100	0.112 0.974	0.029 0.003	0.0035 0.0100	0.0014 0.0008	0.8949
	bin 1 nM	11027	0.107 0.976	0.020 0.003	0.0054 0.0101	0.0014 0.0008	0.9143
	bin 2 nM	11858	0.106 0.981	0.028 0.002	0.0041 0.0127	0.0015 0.0008	0.9328
	bin 4 nM	10842	0.045 0.980	0.046 0.002	0.0012 0.0180	0.0011 0.0009	0.9576
	ter 1 nM, 1 nM	10601	0.103 0.88* 0.978	0.022 − 0.004	0.0039 0.0052 0.0062	0.0013 0.0016 0.0007	0.8959
	ter 1 nM, 2 nM	9879	0.103 0.85* 0.978	0.215 − 0.004	0.0039 0.0052 0.0062	0.0013 0.0016 0.0007	0.8959
	ter 1 nM, 4 nM	10387	0.131 0.84* 0.976	0.024 − 0.004	0.0048 0.0052 0.0059	0.0018 0.0018 0.0008	0.8726
	−NTS 1 nM, 1 nM	16204	0.179 0.784 0.972	0.020 0.021 0.005	0.0066 0.0068 0.0007	0.0009 0.0010 0.0003	0.6882
W971A	apo	9025	0.108 0.983	0.007 0.004	0.0102 0.0043	0.0009 0.0005	0.9092
	bin 0.5 nM	10942	0.142 0.972	0.015 0.004	0.0087 0.0072	0.0015 0.0007	0.8854
	bin 1 nM	12732	0.111 0.977	0.015 0.003	0.0070 0.0094	0.0013 0.0007	0.9119
	bin 2 nM	12262	0.095 0.979	0.022 0.002	0.0037 0.0146	0.0012 0.0008	0.9542
	bin 4 nM	12121	0.076 0.976	0.029 0.002	0.0025 0.0157	0.0012 0.0008	0.9523

	ter 1 nM, 1 nM	11618	0.131 0.824 0.981	0.016 0.028 0.004	0.0062 0.0050 0.0033	0.0016 0.0019 0.0008	0.8275
	ter 1 nM, 2 nM	11438	0.128 0.803 0.980	0.014 0.025 0.004	0.0067 0.0062 0.0029	0.0008 0.0012 0.0007	0.8227
	ter 1 nM, 4 nM	11770	0.143 0.802 0.970	0.017 0.032 0.008	0.0067 0.0067 0.0025	0.0009 0.0016 0.0012	0.7679
	−NTS 1 nM, 1 nM	9242	0.168 0.763 0.969	0.019 0.016 0.008	0.0070 0.0091 0.0005	0.0009 0.0010 0.0003	0.7510
W971D	apo	4576	0.098 0.969	0.008 0.005	0.0128 0.0056	0.0013 0.0007	0.8947
	bin 0.5 nM	8589	0.100 0.971	0.015 0.003	0.0078 0.0093	0.0014 0.0008	0.8958
	bin 1 nM	10440	0.106 0.973	0.014 0.002	0.0059 0.0103	0.0011 0.0007	0.9319
	bin 2 nM	10419	0.091 0.975	0.023 0.002	0.0033 0.0161	0.0012 0.0008	0.9593
	bin 4 nM	10248	0.080 0.978	0.039 0.002	0.0020 0.0165	0.0013 0.0010	0.9434
	ter 1 nM, 1 nM	8010	0.105 0.822 0.977	0.015 0.025 0.003	0.0057 0.0069 0.0045	0.0014 0.0021 0.0010	0.8963
	ter 1 nM, 2 nM	9173	0.114 0.800 0.971	0.014 0.024 0.005	0.0052 0.0073 0.0036	0.0013 0.0021 0.0010	0.8618
	ter 1 nM, 4 nM	8452	0.127 0.809 0.971	0.018 0.027 0.006	0.0049 0.0060 0.0026	0.0008 0.0014 0.0010	0.7743
	−NTS 1 nM, 1 nM	9449	0.157 0.774 0.98*	0.019 0.025 −	0.0082 0.0077 0.0007	0.0010 0.0013 0.0006	0.6562
W971K	apo	5997	0.121 0.967	0.010 0.005	0.0103 0.0055	0.0013 0.0007	0.8718
	bin 0.5 nM	9429	0.119 0.971	0.011 0.003	0.0087 0.0061	0.0011 0.0006	0.8962
	bin 1 nM	12056	0.110 0.976	0.013 0.003	0.0066 0.0086	0.0011 0.0006	0.9189
	bin 2 nM	8677	0.093 0.968	0.021 0.002	0.0035 0.0110	0.0011 0.0007	0.9398

	bin 4 nM	10153	0.090 0.972	0.034 0.002	0.0024 0.0136	0.0013 0.0009	0.9338
	ter 1 nM, 1 nM	10647	0.120 0.761 0.979	0.017 0.028 0.006	0.0058 0.0078 0.0025	0.0008 0.0013 0.0007	0.7521
	ter 1 nM, 2 nM	11179	0.109 0.761 0.978	0.012 0.023 0.009	0.0057 0.0068 0.0014	0.0007 0.0011 0.0006	0.7501
	ter 1 nM, 4 nM	11107	0.135 0.81* 0.979	0.019 − 0.011	0.0050 0.0073 0.0010	0.0008 0.0010 0.0004	0.7148
	−NTS 1 nM, 1 nM	10775	0.12* 0.77* 0.959	− − 0.013	0.0082 0.0078 0.0013	0.00079 0.0011 0.0006	0.7745
W971F	apo	3172	0.110 0.978	0.008 0.003	0.0141 0.0036	0.0013 0.0005	0.9024
	bin 0.5 nM	2561	0.097 0.971	0.021 0.002	0.0061 0.0119	0.0016 0.0008	0.9318
	bin 1 nM	3567	0.102 0.975	0.028 0.0018	0.0037 0.0169	0.0015 0.0009	0.9563
	bin 2 nM	4085	0.078 0.976	0.040 0.002	0.0023 0.0198	0.0014 0.0009	0.9632
	bin 4 nM	3993	0.077 0.976	0.049 0.022	0.0030 0.0153	0.0018 0.0138	0.3658
	ter 1 nM, 1 nM	3837	0.086 0.867 0.977	0.019 0.014 0.001	0.0021 0.0089 0.0063	0.0007 0.0014 0.0008	0.9684
	ter 1 nM, 2 nM	3886	0.095 0.825 0.972	0.026 0.017 0.003	0.0020 0.0099 0.0064	0.0010 0.0019 0.0012	0.9475
	ter 1 nM, 4 nM	4030	0.087 0.833 0.979	0.021 0.012 0.002	0.0026 0.0120 0.0041	0.0010 0.0017 0.0007	0.9409
	−NTS 1 nM, 1 nM	3827	0.130 0.790 0.963	0.032 0.016 0.010	0.0013 0.0114 0.0019	0.0006 0.0014 0.0010	0.8903
ΔBH	apo	1288	0.101 0.975	0.018 0.002	0.0062 0.0051	0.0014 0.0006	0.8403
	bin 0.5 nM	1779	0.071 0.965	0.048 0.003	0.0026 0.0135	0.0018 0.0011	0.8916
	bin 1 nM	2153	0.070 0.965	0.040 0.003	0.0032 0.0223	0.0020 0.0012	0.9218

	bin 2 nM	22024	0.090 0.966	0.059 0.002	0.0019 0.0198	0.0018 0.0012	0.9362
	bin 4 nM	837	0.064 0.963	0.038 0.002	0.0017 0.0200	0.0012 0.0009	0.9638
	ter 1 nM, 1 nM	2989	0.065 0.817 0.965	0.027 0.021 0.003	0.0020 0.0119 0.0080	0.0010 0.0025 0.0015	0.9542
	ter 1 nM, 2 nM	2895	0.044 0.824 0.966	0.039 0.016 0.003	0.0007 0.0137 0.0058	0.0007 0.0024 0.0015	0.9386
	ter 1 nM, 4 nM	2751	0.085 0.801 0.942	0.033 0.032 0.007	0.0016 0.0141 0.0058	0.0010 0.0042 0.0033	0.9474
	−NTS 1 nM, 1 nM	2713	0.069 0.80* 0.96*	0.037 – –	0.0004 0.0187 0.0008	0.0005 0.0019 0.0005	0.9015
ΔΔBH	apo	8572	0.156 0.964	0.024 0.006	0.0045 0.0065	0.0014 0.0009	0.7833
	bin 0.5 nM	11087	0.138 0.967	0.056 0.004	0.0014 0.0102	0.0013 0.0010	0.8618
	bin 1 nM	9320	0.089 0.983	0.024 0.003	0.0028 0.0132	0.0011 0.0007	0.9208
	bin 2 nM	11341	0.097 0.977	0.026 0.002	0.0019 0.0148	0.0009 0.0007	0.9618
	bin 4 nM	9173	0.089 0.978	0.040 0.002	0.0014 0.0164	0.0011 0.0009	0.9538
	ter 1 nM, 1 nM	7570	0.095 0.826 0.976	0.020 0.038 0.004	0.0024 0.0062 0.0057	0.0010 0.0024 0.0014	0.9004
	ter 1 nM, 2 nM	8464	0.143 0.856 0.981	0.035 0.028 0.004	0.0019 0.0068 0.0042	0.0012 0.0023 0.0014	0.8654
	ter 1 nM, 4 nM	7103	0.140 0.820 0.971	0.024 0.037 0.006	0.0033 0.0074 0.0040	0.0015 0.0030 0.0016	0.8347
	−NTS 1 nM, 1 nM	7024	0.11* 0.810 0.976	– 0.028 0.008	0.0016 0.0064 0.0021	0.0010 0.0020 0.0010	0.7122

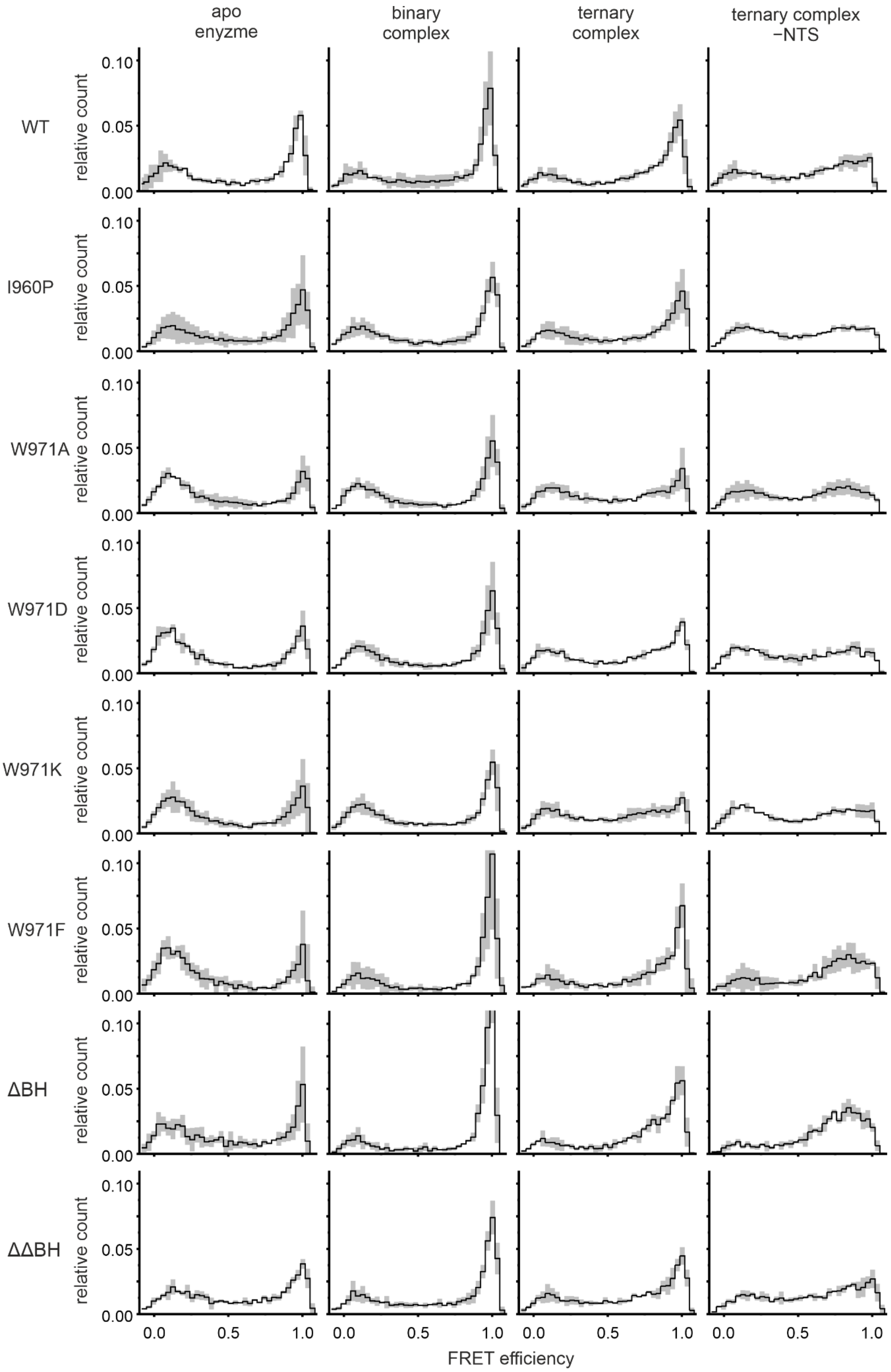
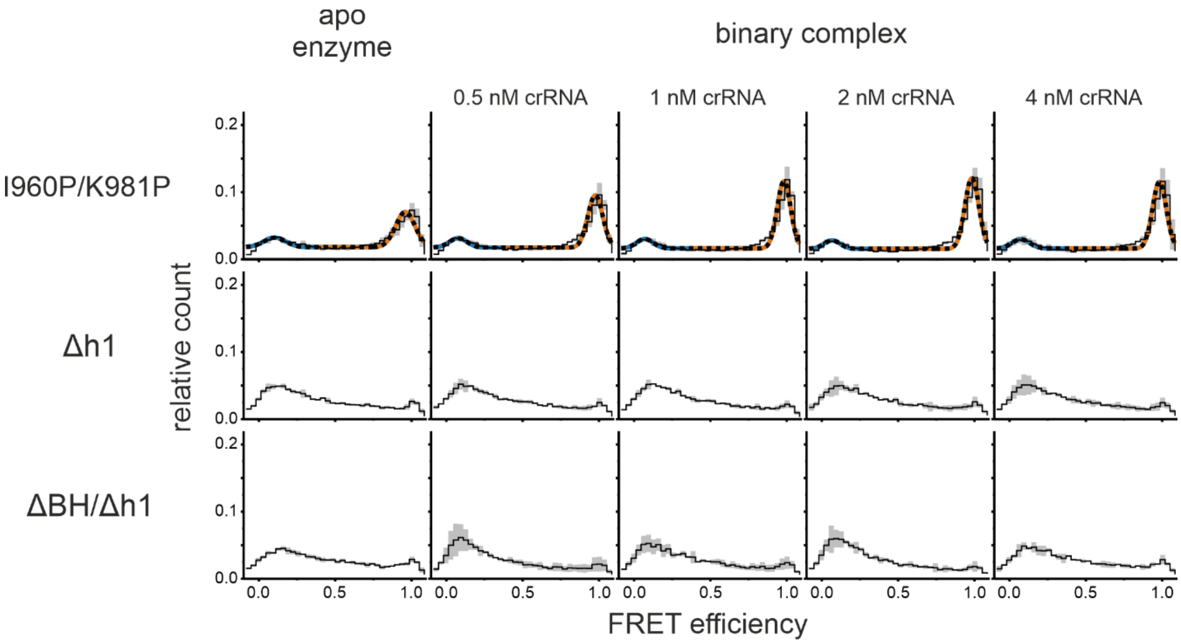


Figure 71: FRET efficiency histograms with standard deviation for measurements of doubly labeled FnCas12a BH variants.

FRET efficiency histograms of the measurements with doubly labeled FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ BH variants for the apo enzyme, the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS DNA) (corresponding to **Figure 40**). Histograms show the mean FRET efficiency (solid line) and the standard deviation (gray area) of three replicates. (adapted from Wörle *et al.* [290])

III.viii. Single-molecule FRET measurements on doubly labeled FnCas12a helix 1 variants

The smFRET measurements shown in **Figure 51** were additionally conducted with different concentrations of crRNA and DNA. FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ I960P/K981P shows an increase in the high FRET population with the increase of crRNA, whereas increasing target DNA concentrations lead to a decrease in the high FRET population and an increase in the medium FRET population, as observed in measurements with WT FnCas12a. Variants with deleted helix 1 did not show defined populations, not even with higher nucleic acids concentrations.



The figure continues on the next page.

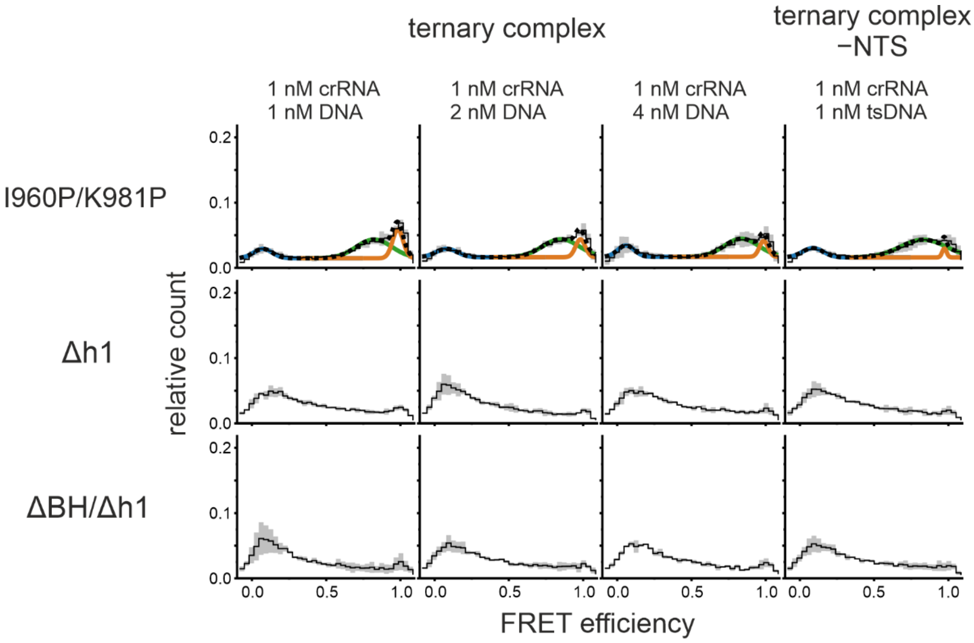


Figure 72: Confocal single-molecule FRET measurements using doubly labeled FnCas12a helix 1 variants with different RNA/DNA concentrations.

FRET efficiency distributions for the apo enzyme of FnCas12a helix 1 variants, the binary complex, the ternary complex, and the ternary complex without NTS. FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ I960P/K981P shows two FRET populations for the apo enzyme and the RNA-bound state, with the high FRET population increasing with the addition of crRNA and further increasing with higher crRNA concentrations. In the ternary complex, FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ I960P/K981P shows three FRET populations with an additional medium FRET population that increases with higher target DNA concentrations, whereas the high FRET population decreases. The populations are similar to the WT but differ in their distributions. The helix 1 deletion variants FnCas12a$^{\text{REC-Nuc*DL550/DL650}}$ Δh1 and ΔBH/Δh1 do not show defined populations. Histograms show the average data of three independent measurements. The histograms were fitted with a double or triple Gaussian function. The solid black line shows the mean FRET efficiency, colored lines show the Gaussian fit, the black dotted line the overall fit, and grey bars the standard deviation. Additional information regarding fit parameters and molecule counts can be found in **Table 44**.

Table 44. Molecule counts and fitting parameters for FRET measurements with doubly labeled FnCas12a helix 1 variants.

The FRET efficiency histograms (corresponding to **Figure 51** and **Figure 73**) were fitted with two or three Gaussian distributions. No fit could be performed for the measurements with helix 1 deletion variants. The number of molecules N_M used to calculate each histogram, the coefficient of determination R^2, the mean FRET efficiency E (± s.e.), and the peak area A (± s.e.) of each Gaussian are listed. Indicated are concentrations of crRNA, and DNA, respectively, 'apo' stands for apo FnCas12a, 'bin' for the binary complex, 'ter' for the ternary complex, and '−NTS' for the ternary complex without NTS. * Indicates fixed values for the fit.

Sample		N_M	E_1 E_2 E_3	± s.e. E_1 ± s.e. E_2 ± s.e. E_3	A_1 A_2 A_3	± s.e. A_1 ± s.e. A_2 ± s.e. A_3	R^2
I960P/ K981P	apo	23380	0.105 0.959	0.023 0.006	0.0025 0.0083	0.0009 0.0008	0.8605
	bin 0.5 nM	25291	0.078 0.976	0.021 0.003	0.0020 0.0095	0.0008 0.0007	0.9162
	bin 1 nM	25924	0.070 0.982	0.020 0.003	0.0019 0.0110	0.0007 0.0006	0.9441

	bin 2 nM	25804	0.067 0.983	0.023 0.003	0.0017 0.0121	0.0007 0.0007	0.9499
	bin 4 nM	24225	0.074 0.982	0.020 0.002	0.0020 0.0112	0.0007 0.0006	0.9521
	ter 1 nM, 1 nM	22137	0.074 0.823 0.982	0.013 0.020 0.003	0.0022 0.0086 0.0039	0.0006 0.0016 0.0009	0.9484
	ter 1 nM, 2 nM	23028	0.081 0.85* 0.982	0.015 - 0.005	0.0023 0.0085 0.0022	0.0007 0.0012 0.0004	0.9225
	ter 1 nM, 4 nM	21777	0.059 0.84* 0.986	0.010 - 0.005	0.0024 0.0081 0.0018	0.0005 0.0010 0.003	0.9145
	−NTS 1 nM, 1 nM	31589	0.088 0.832 0.973	0.012 0.011 0.005	0.0025 0.0105 0.0005	0.0004 0.0006 0.0002	0.9038
Δh1	apo	16177	–	–	–	–	–
	bin 0.5 nM	15888	–	–	–	–	–
	bin 1 nM	15170	–	–	–	–	–
	bin 2 nM	15066	–	–	–	–	–
	bin 4 nM	16658	–	–	–	–	–
	ter 1 nM, 1 nM	13812	–	–	–	–	–
	ter 1 nM, 2 nM	15835	–	–	–	–	–
	ter 1 nM, 4 nM	15846	–	–	–	–	–
	−NTS 1 nM, 1 nM	15458	–	–	–	–	–
ΔBH/Δh1	apo	20091	–	–	–	–	–
	bin 0.5 nM	20680	–	–	–	–	–
	bin 1 nM	20703	–	–	–	–	–
	bin 2 nM	18622	–	–	–	–	–
	bin 4 nM	18372	–	–	–	–	–
	ter 1 nM, 1 nM	19996	–	–	–	–	–
	ter 1 nM, 2 nM	20254	–	–	–	–	–
	ter 1 nM, 4 nM	18569	–	–	–	–	–
	−NTS 1 nM, 1 nM	19119	–	–	–	–	–

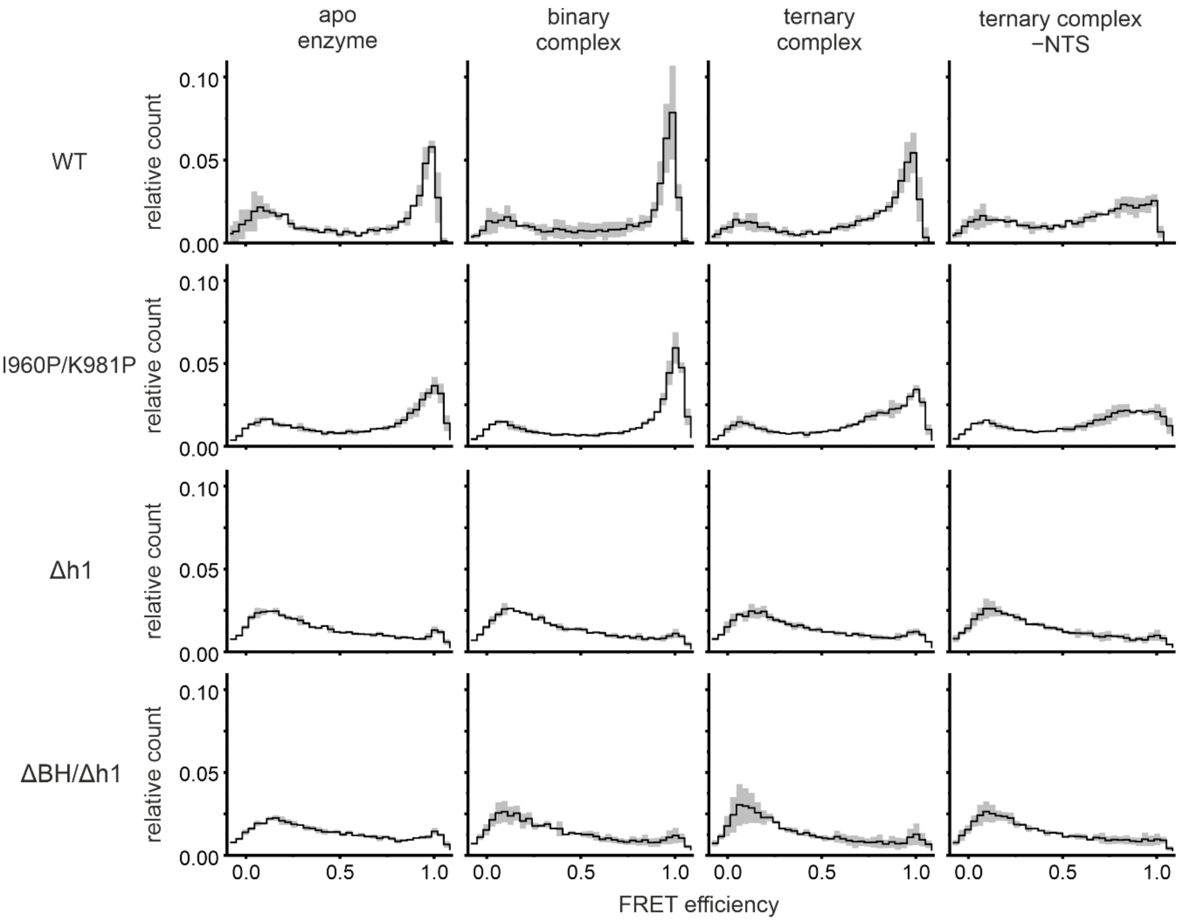


Figure 73: FRET efficiency histograms with standard deviation for measurements of doubly labeled FnCas12a helix 1 variants.

FRET efficiency histograms of the measurements with doubly labeled FnCas12a$^{\mathrm{REC\text{-}Nuc*DL550/DL650}}$ helix 1 variants for the apo enzyme, the binary complex (1 nM crRNA), the ternary complex (1 nM crRNA, 1 nM DNA), and the ternary complex without NTS (1 nM crRNA, 1 nM TS DNA) (corresponding to **Figure 51**). Histograms show the mean FRET efficiency (solid line) and the standard deviation (gray area) of three replicates.

ACKNOWLEDGEMENT

Finally, I want to thank everyone who contributed to the success of my thesis and who supported me during this work.

First, I want to cordially thank Prof. Dr. Dina Grohmann, my main supervisor, for giving me the opportunity to work on this interesting topic and to conduct my thesis in her research group. I am thankful for the great supervision, lively discussions, and her continuous support during the last years.

Also, I would like to thank my mentoring team Prof. Dr. Gunter Meister and Dr. Sabine Schneider for always having an open ear and broad suggestions for the progress of my project. I additionally thank Prof. Dr. Gunter Meister for accepting to be the second Ph.D. assessor.

Special gratitude applies to my cooperation partners Dr. Gaetan Burgio and Anthony Newman from Canberra (Australia) for conducting measurements with my FnCas12a variants and for the willingness and support to host me during our planned exchange. A pity that COVID-19 crossed our plans and we could not meet in person. Further thank applies to Dr. Sabine Schneider for trying to crystallize some of my FnCas12a variants.

A sincere thank you applies to my colleagues Dr. Sarah Willkomm, Benedikt Moissl, Dr. Kevin Kramm, Andreas Schmidbauer, Dr. Felix Grünberger, Henri Michel, and Daniela Tarau for great collaboration and discussions. You motivated and helped me with many questions and always had an open ear and new approaches for me. Furthermore, I thank my former colleague Dr. Leonhard Jakob for the introduction to the project and the coincidence that I was the student he initially started the project with during an internship, not knowing where it once would end.

In addition, I thank all further members of the Microbiology chair for their aid in whatever question I had during the daily lab routine and for the pleasant working atmosphere in the group.

I want to thank my friends Benedikt, Fabian, Michaela, and Thomas for supporting me regardless of the distance. It was always a great time being with you in person again. It was great to have some additional sources for papers if I didn’t have the access myself.

Last but not least, I thank my family and Sonja for their infinite support during the last years, which I can always rely on.