Investigating the Conformational Landscape of Cas13a

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Maria Francesca Vicino

aus

Köln

Bonn 2024

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät
der Rheinischen Friedrich-Wilhelms-Universität Bonn

Gutachter/Betreuer: Prof. Dr. Olav Schiemann

Gutachter: Prof. Dr. Ulrich Kubitscheck

Tag der Promotion: 23.04.2025

Erscheinungsjahr: 2025

Table of Contents

Abstract

Bacteria and bacteriophages are in a constant arms race to develop strategies to coexist. One adaptive defense system, which bacteria and archaea have developed against phages is the CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats – CRISPR Associated proteins) system. Out of a variety of known CRISPR-Cas types, the type VI CRISPR effector protein Cas13a (Cas-associated 13a), composed of a recognition (REC) lobe and a nuclease (NUC) lobe, has unique properties. In contrast to the most prominent Cas protein representative Cas9, Cas13a binds and cleaves RNA and not DNA. In addition, it catalyzes the maturation of a precursor-CRISPR RNA (pre-crRNA) and the cleavage of the target RNA, which activates a unique sequence non-specific collateral RNA cleavage. This collateral cleavage was harnessed as nucleic acid detection tool, to detect viral RNA, causing human diseases. Further Cas13a has therapeutic applications for example in inhibiting cancer cell growth.

Despite the high medical relevance and application, the underlying coordinating mechanism and its structure-dynamics-function relationship was not investigated. Structures from all functionally relevant complexes are described (*apo*, pre-crRNA bound, crRNA bound, and cr- and target RNA bound), but they origin from proteins of different organisms and contain different extends of truncation. Additionally, it is unknown how the structures relate to functional aspects of this system. Thus, to identify and follow conformational changes on the molecular level, information of all states from one organism are needed.

In this work, mainly Pulsed Electron-Electron Double Resonance (PELDOR) spectroscopy was used to analyze conformational changes during the functional pathway of Cas13a. In the first part of this thesis, basic biochemical techniques were used to optimize the expression and purification approach for Cas13a from *Leptotrichia buccalis* (*Lbu*) and to optimize RNA cleavage assays in our laboratory. The second part of this thesis focuses on the development of active double labelled protein constructs, for PELDOR studies. A mutational analysis was used, to find replacements for three native cysteines, generating a protein construct that retained cleavage activity after spin labelling. The third part of the thesis focuses on validating

the structures of *Lbu*Cas13a and following its conformational changes through the entire functional pathway.

The PELDOR measurements showed a high flexibility in the *apo* state of the REC lobe. This flexibility was not described by any known structure or prediction. In contrast, the *apo* state of the NUC lobe is rigid, with the helical-2 domain being the exception. The REC and NUC lobes are flexible towards each other. The pre-crRNA bound structure is not known. We reveal this structure to match the crRNA bound state. The expected conformational changes from the experimentally known binary crRNA-bound complex to the experimentally known ternary complex were found to occur in frozen solution. Interestingly, by using a combination of PELDOR spectroscopy and AFM measurements, a previously unknown dimeric structure of the protein complex with cr- and target RNA was found. This structure is different than the structure observed in the asymmetric unit of the cr- and target RNA bound *Lbu*Cas13a.

By summing up, for the first time, the entire functional cycle of the Cas13a protein from one organism is structurally and dynamically investigated and new details are added.

Acknowledgments

First of all, I would like to express my sincere gratitude to Prof. Dr. Olav Schiemann, for his immense support throughout my PhD thesis. Thank you for the exciting and constructive discussions, the criticism, the development of new ideas and the readiness to help me in my scientific challenges. Further, I would also like to thank him for challenging me with tasks, that have greatly contributed to my growth as a scientist and as a person and for creating a stimulating working atmosphere in his group.

I would like to thank Prof. Dr. Günter Mayer for being my second mentor, for the very constructive meetings and ideas, and for accepting the role of the second reviewer of this thesis. In this context, I would like to thank Prof. Dr. Ulrich Kubitscheck, and Prof. Dr. Christoph Thiele for their participation in my doctoral committee.

During my PhD thesis I had the great opportunity to work with and to learn from many inspiring scientists. I would like to thank Prof. Benjamin Kaupp for his support, discussions and for ideas he brought to this project. Further I would like to thank PD Dr. Gregor Hagelüken for his biochemical help during the first months of my PhD thesis. I would also like to thank Dr. Dinar Abdullin, Dr. Tobias Hett, and Hamed Alaei for their advice and technical support with the EPR spectrometers.

Further, the Deutsche Forschungsgemeinschaft (DFG, project number 420322655, Koselleck grant to U. B. Kaupp) is greatly acknowledged for funding parts of this research.

I consider myself a very lucky person, since I have met many amazing people during my PhD, who have all contributed to this thesis and supported me in different ways. I would like to thank all current and former members of the Schiemann group and the Kaupp group for creating a wonderful working atmosphere, especially those who shared the office with me, namely Catrin Allar, Olivia Kendall, Dr. Tobias Hett, and Kevin Kopp. I would like to express my special gratitude to Dr. Christine Wübben introducing me to the RNA world, and Catrin Allar, for choosing to work on Cas13a and for her motivated work, especially during her Master's thesis.

I would like to thank some very special colleagues, who I am honored to now call friends. Thank you, Sarah Löffelsender, Dr. Anja Haags, Beatrice Wolff, Alina Geis,

Olivia Kendall, and Catrin Allar, for supporting me and for the great time we spent together.

Thank you, Emily Goehlmann, for your constant encouragements and for accepting my numerous scientific nonsenses. I look forward to our next adventure together.

A very special thank you goes to my parents, for their patience and for supporting and believing in me in every step of my life. Again, I am a very lucky person to have found not only friends but also love in the Clausius Institute. I would like to thank my partner Martin, for grounding me and for being my rock every day, for his immense support and patience, and for letting me focus on the beautiful things in life and not the failed experiments.

Contributions

I am grateful to all Master and Bachelor students, which I supervised and who contributed to the topic of this thesis. This holds especially for Catrin Allar, who decided to continue as a PhD student on this topic. Thus, I would like to acknowledge their work by showing in the following table to which part of my PhD thesis they have contributed.

Lejla Tokic, bachelor student	• Purification of wt *Lbu*Cas13a, Figure 2.1-7

Catrin Allar, master student and PhD student	• Testing cleavage assay conditions with L-arginine and L-glutamic acid
	• Testing trehalose, L-arginine and L-glutamic acid for PELDOR sample preparation
	• Cloning, purification, labelling, and measuring once the following construct sets:
	  ○ 756/926 (Figure 4.2-3)
	  ○ 462/926 (Figure 4.2-4)
	  ○ 190/756 (Figure 4.3-1)
	  ○ 190/926 (Figure 4.3-2)
	• Cloning of constructs $756_{dtarget}$ and $926_{dtarget}$

Yannik Limbach, master student (supervised by Catrin Allar), PhD student	• Cloning, purification, labelling of $190/756_{dtarget}$ (Figure 4.3-1)
Tim Mevs, master student (supervised by Catrin Allar)	• Cloning, purification, labelling of the $138/926_{ative}$ construct (Figure 4.3-3)
Dr. Joshua Lee Wort, postdoctoral fellow	• ITC measurements
Dr. Christian Renzl, former PhD student in the laboratory of Prof. G. Mayer	• Supervision of the first *in-vitro* transcription, performed in their laboratories

List of Abbreviations

Abbreviations	
660/926	cysfree-V with cysteines at positions 660 and 926
aaRS	aminoacyl-tRNA synthetase
Abi	abortive infection
Acr	anti-CRISPR
AF2 / AF3	AlphaFold2 / AlphaFold3
APS	ammonium peroxydisulfate
CBASS	cyclic oligonucleotide-based anti-phage signalling system
CDA	ComparativeDeerAnalyzer
cOA	cyclic oligoadenylates
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CV	column volume
cysfree-V	*Lbu*Cas13a C293A C348V C1141A
d-EG	d_4-ethylene glycol
del	deletion
dNTPs	deoxynucleotide triphosphates
dsDNA	double-stranded DNA
EDTA	ethylenediaminetetraacetic acid
EPR	electron paramagnetic resonance
EU	European Union
FD	fast digest
FRET	Förster resonance energy transfer
HEPN domain	higher eucaryotes and procaryotes nucleotide binding domain
IDP	intrinsically disordered protein
ins	insertion
IPTG	isopropyl-β-D-1-thiogalactopyranoside
IVT	*in vitro* transcription
LB	Lysogeny Broth
Lba	*Lachnospiraceae bacterium*
Lbu	*Leptotrichia buccalis*
Lsh	*Leptotrichia shahii*
Lwa	*Leptotrichia wadei*
MBP	maltose binding protein
MSA	multiple sequence alignment
MTSL	methanethiosulfonate spin label
Ni-AC	nickel affinity chromatography
nt	nucleotide
NTD	N-terminal domain
NUC lobe	nuclease lobe
PAE	predicted aligned error
PAGE	polyacrylamide gel electrophoresis

PAM	protospacer adjacent motif
PCR	polymerase chain reaction
PDS	pulsed dipolar spectroscopy
PELDOR	pulsed electron-electron double resonance
PFS	protospacer flanking site
pLDDT	predicted local distance difference test
PMSF	phenylmethylsulfonylfluorid
pre-crRNA	precursor CRISPR RNA
Rc	*Rhodobacter capsulatus*
REC lobe	recognition lobe
RM	restriction modification
RT	room temperature
SANS	small-angle neutron scattering
SAXS	small-angle X-ray scattering
SDS	sodium dodecylsulfate
SEC	size exclusion chromatography
Sie	superinfection exclusion
SLIM	short linked maleimide
ssDNA	single-stranded DNA
TAE Buffer	tris-acetic acid-EDTA buffer
TBE Buffer	tris-boric acid-EDTA buffer
Tcc	*Thermoclostridium caenicola*
TCEP	tris-(2-carboxyethyl)-phosphine
TEMED	tetramethylethylenediamine
TEV	tobacco etch virus
tracrRNA	trans-activating crRNA
UAA	unnatural amino acid
wt	wildtype
XL-MS	cross-linking mass spectrometry

Chapter 1

Introduction

1.1 Viruses

It is estimated that 10^{31} viruses exist on Earth, making them the most abundant and diverse organisms[1]. Viruses were discovered independently in 1915 and 1917 by Frederick Twort and Félix d'Hérelle, respectively, and it was very soon that d'Hérelle recognized a potential antibacterial property that could be exploited for human health e.g. wound treatment[2]. Viruses that only attack and replicate in bacteria and archaea are called bacteriophages, or just phages. However, the ability of phages to attack bacteria and alter the host genome can pose a serious threat to humans. Several, by nature non-dangerous bacterial strains as *Escherichia*, *Salmonella*, *Staphylococcus*, *Streptococcus,* or *Vibrio* can cause human death by acquiring phage genomes. This acquisition leads to the production of virulence factors such as specific enzymes or exotoxins[3]. In addition, there are many viruses that can directly infect human organisms. Some of the most infamous viruses threatening public health are influenza virus, Zika virus, coronaviruses, hepatitis B virus, hepatitis C virus, Ebola virus, human immunodeficiency virus (HIV), and dengue virus. Already the influenza virus, causing the commonly known seasonal influenza disease, is a respiratory infection that causes between 290,000 to 650,000 deaths per year[4]. The total number of infections with these eight viruses is estimated to be larger than 1.3 billion per year[4–6], which means that about 16% of the world's population is infected each year (approximate world population: 8 billion[7]).

Table 1.1-1: Selection of viruses threatening public health. The genome type is abbreviated with +/- for positive or negative sense strands and ss for single stranded or ds for double stranded genomes. The infection numbers are given in global infections per year, except for the Zika virus, where the infection cases are shown for the European Union[4–6].

Virus	Genome	Infections per year
Influenza virus	(+/-) ssRNA	3-5 million
Zika virus	(+) ssRNA	> 300 cases in EU
Coronavirus	(+) ssRNA	772 million in 2023
Hepatitis B virus	dsDNA/ssDNA	1.5 million
Hepatitis C virus	(+) ssRNA	1.5 million
Ebola virus	(-) ssRNA	170 cases in 2022
HIV	(+) ssRNA	1.3 million
Dengue virus	(+) ssRNA	390 million

Interestingly, all these viruses are RNA viruses, with hepatitis B virus being the only exception (Table 1.1-1)[8]. It is therefore particularly important to find sensitive

detection strategies and effective treatments for viruses (especially RNA viruses). These strategies should additionally be easily adapted to different viral sequences since viruses rapidly change their genome[9]. One of the promising ways to do this is to study and understand how smaller organisms, such as bacteria, defeat viruses. These small organisms are in a constant arms race with phages, trying to evolve and gain an evolutionary advantage by declining the phage's fitness (Red Queen Hypothesis) through the development of different defense systems[10]. Moreover, some prokaryotic defense systems are believed to be the ancestors of human defense systems and there are many parallels between them[11]. Many of the systems found in eukaryotes as cGAS-STING[12], argonautes[13], or viperins[14] have a prokaryotic analogue. Thus, the next Chapter will focus on how bacteria defend themselves against phages.

1.2 Bacterial immune system

The bacterial immune system, like the human immune system, is divided into two parts, the innate and the adaptive immune system. The innate immune system is the rapid, and pattern-based response of a cell against invaders. In contrast, the adaptive immune system is a response to specific pathogens, which takes longer since it needs to recognize the invader as an invader. Further, the adaptive immune system remembers previous infections, so that it can act faster in a second infection[15]. In general, the bacterial immune system varies widely within a cell colony. Each cell can encode multiple defense systems that differ from those of other cells in the same colony, which is the result of the co-evolutionary arms race between phages and bacteria[16]. Bacterial defense systems are found clustered on the genome in so-called defense islands, which are used to discover new anti-phage defense systems[17]. However, additionally to the variability of defense systems in a cell colony, there is also a temporal variability of defense systems. Defense systems can be horizontally transferred and lost in short time scales, making it a very dynamic system[9]. The following paragraphs will give a broad summary of the most studied antiphage defense systems (Figure 1.2-1). New defense systems are constantly added to this list, such as Gabija[18], Shedu[18], Zorya[18], and Lamassu[19]. But since these mechanisms are yet unknown, they will not be covered in this summary.

Antiphage defense systems can act at each step of the infectious, lytic cycle of phages (Figure 1.2-1 a). During the first step of the infection cycle, phages recognize and dock to membrane proteins, lipopolysaccharides, or other cell-surface features. The phage attachment to the cell surface can be inhibited by diverse mechanisms, including phage receptor blocking, extracellular matrix production, production of competitive inhibitors, and modification of cell surface components (Figure 1.2-1 b)[10,20]. An example of phage receptor blocking can be found in *Staphylococcus aureus* that synthesizes the protein immunoglobulin G-binding protein A. This binds to immunoglobulin G on the cell surface, masking this receptor and leading to reduced viral adsorption[21]. Alternatively, some bacteria produce an extracellular matrix that acts as a physical barrier. For example, *Pseudomonas spp.* synthesizing alginates, which block phage adsorption. Interestingly, phages can overcome the alginate coat by synthesizing alginate lyases that degrade the coat[20]. The third strategy, to inhibit phage binding is the production of competitive inhibitors. These are molecules, which are present in the bacterial environment, and which block phage receptors by binding to them. One example is the iron transporter FhuA, present in *E. coli*. This is blocked by the molecule microcin J25, rendering the receptor unavailable for T1 and T5 phage binding[22]. Finally, the phage adsorption can also be reduced by mutating or deleting receptors, but this can have a negative impact on cellular fitness[10].

If these defense strategies can be overcome by the phage and it docks onto the cell surface, the injection and subsequent entering of the genome into the cell can be inhibited (Figure 1.2-1 c). This is done by superinfection exclusion systems (Sie), which are often membrane-anchored proteins[10]. The Sie systems differ depending on whether the bacteria are Gram-negative or Gram-positive, and instead of inhibiting the first phage genome injection, they inhibit subsequent phage infections. This system is unique, since it is controlled by a virus and since it requires a specific viral protein[23]. If the T4 phage, for example, attacks a Gram-negative bacterium, it releases proteins that digest the peptidoglycan layer (Figure 1.2-1 c). Then, the Sie imm (immunity to superinfection) and sp (spackle periplasmic) systems are injected. This prevents the injection by other T-even phages, such as T2, T4, T6[24]. However, the mechanism is not fully understood. What is known is, that imm blocks the phage genome from entering and that sp prevents local peptidoglycan degradation[24].

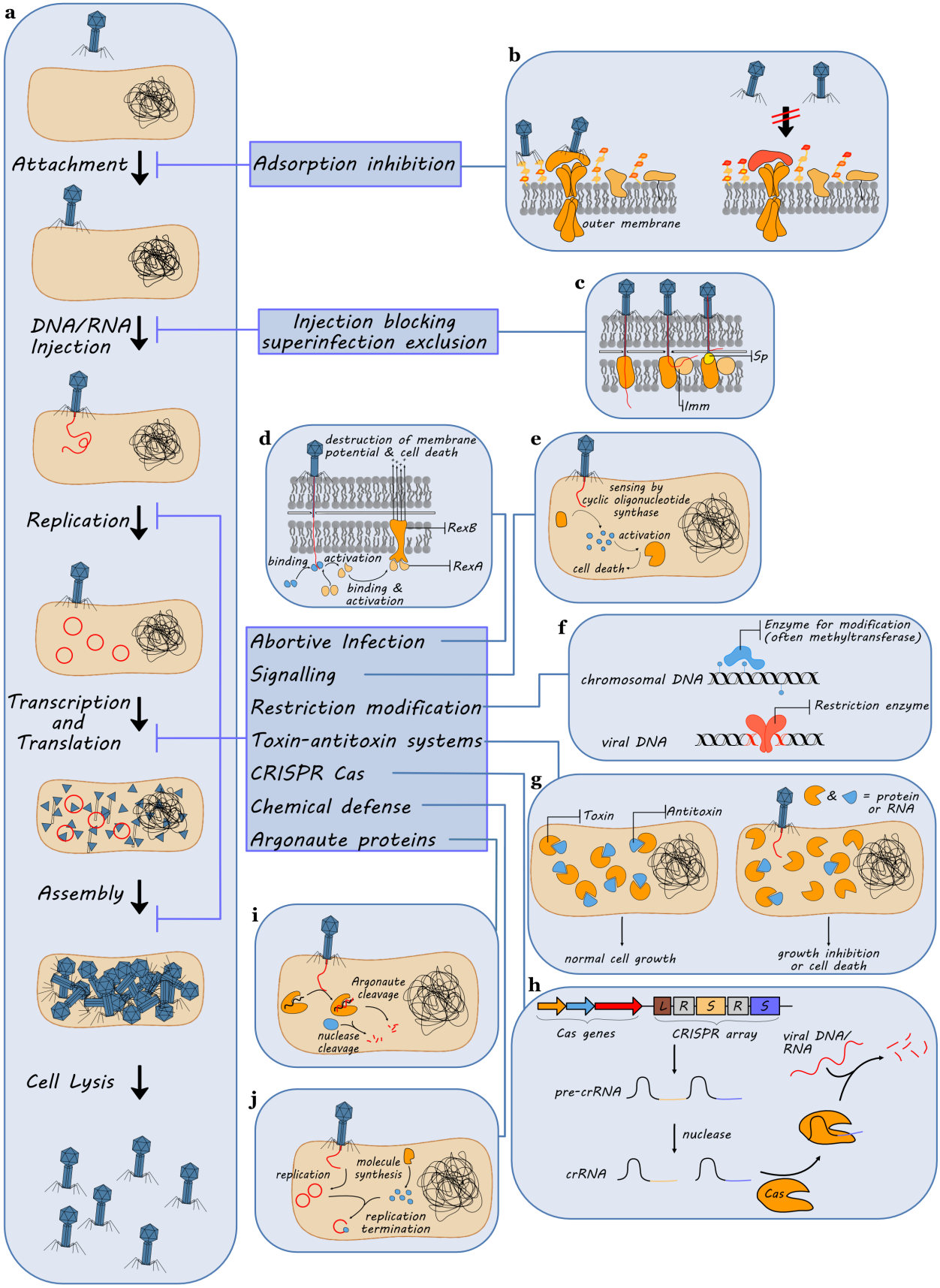


Figure 1.2-1: Overview of the most studied prokaryotic antiphage defense systems. These defense systems are organized according to the phage reproduction stage they act on. **a,** Explains the lytic phage cycle and **b-j,** show schematically how the respective defense systems work, which is explained in detail in the main text.

If the bacterial cell gets infected by a phage and the phage's genome is in the cytosol, a cascade of cellular defense mechanisms can be activated. These cycles of defense mechanisms end with an altruistic cellular behavior culminating in cell death or with the degradation of the phage genome. The systems, that end with cell death are abortive infection (Figure 1.2-1 d), cyclic nucleotide signalling (Figure 1.2-1 e), toxin-antitoxin systems (Figure 1.2-1 g), and some clustered regularly interspaced short palindromic repeats (CRISPR) CRISPR-associated (Cas) systems (Figure 1.2-1 h). Abortive infection (Abi) systems are well-studied defense systems that target specific mechanisms during the phage lytic cycle, such as replication, transcription, and translation[20] (Figure 1.2-1 d). An example of an Abi system is the RexA-RexB system found in λ-lysogenic *E. coli*. This system has been important in understanding the life cycle and, more generally, the biology of the T4 phage. Upon phage infection, the complex between phage DNA and a phage protein activates RexA in the bacterial cell. Two activated RexA bind to the membrane protein RexB, which reduces ATP levels in the cell and depolarizes the bacterial membrane, leading to cell growth inhibition or cell death[25–27]. A special form of Abi systems is the cyclic oligonucleotide-based anti-phage signalling system (CBASS, Figure 1.2-1 e), which is very similar to the cyclic GMP-AMP synthase (cGAS)-STING pathway that is part of the eukaryotic innate immune system. The phage is sensed by a cyclic-oligoadenylate synthase, which produces cyclic oligoadenylates, as cGMP-cAMP. These signalling molecules activate downstream genes such as phospholipases, which destabilize the cell membrane and lead to cell death, before phage replication is terminated[12,28]. A similar system often associated with Abi is the toxin-antitoxin system (Figure 1.2-1 g), which can be based on RNA-RNA (type I systems), protein-RNA (type III systems) or protein-protein (type II systems) interactions[29]. These systems consist of a toxin, which is produced by the cell, and its counterpart the antitoxin, which counteracts the toxin's function. The toxin can disrupt cellular processes such as translation, replication, and cell membrane destabilization. Usually, these systems are tightly regulated by promoter repression or specific transcriptional termination because even a small imbalance will cause the toxin to kill the bacterial cell[20]. Cell toxicity can also result from CRISPR-Cas systems (Figure 1.2-1 h), such as the Type VI system Cas13a, which is the subject of this thesis. Cas13a is an RNA-guided nuclease that can cleave phage RNA in a non-specific manner as a defense response. This means, that both phage and cellular RNA are cleaved, leading to a state of cell dormancy that eventually culminates in cell death[30].

In contrast to the presented defense systems that lead to cell death, there are other defense systems that degrade the phage genome, including for example restriction modification (RM, Figure 1.2-1 f), argonaute proteins (Figure 1.2-1 i), chemical defense (Figure 1.2-1 j), and CRISPR-Cas (Figure 1.2-1 h). The RM systems are part of the innate immune system and recognize foreign DNA (Figure 1.2-1 f). They consist of two classes of enzymes, a restriction endonuclease that cleaves foreign DNA, and often a methyltransferase that methylates cellular DNA. This methylation protects the host genome from degradation by the restriction endonuclease. However, this system is not very specific and sometimes foreign DNA is also methylated and protected from degradation[31]. Argonaute systems differ conceptionally from RM systems in their recognition mechanisms. RM systems directly recognize modifications on the cellular DNA, whereas argonaute systems use RNA or DNA molecules for self, versus non-self-discrimination[32]. Argonaute-RNA or argonaute-DNA complexes bind foreign oligonucleotides and degrade these either themselves or by recruiting other nucleases[13,33] (Figure 1.2-1 i), inhibiting phage replication. As already mentioned for CBASS systems, argonaute systems also have a eukaryotic argonaute analogue, which preferentially targets RNA[34]. Another system, inhibiting phage replication is chemical defense (Figure 1.2-1 j). There, a small molecule is synthesized by an enzymatic pathway that inhibits phage replication. Recently, prokaryotic viperins were discovered, which are predicted to be ancestors of eukaryotic viperins[14]. The prokaryotic protein viperin converts CTP, GTP and UTP to their ddh variants. These lack the hydroxy group at the 3' carbon of the ribose, which inhibits chain elongation once incorporated, therefore leading to the termination of phage RNA replication[14]. Interestingly, we make use of this principle in human treatments, such as treatment of the herpes virus. The most commonly used treatment is acyclovir, which is also a chain terminator, mimicking guanin[35]. In contrast to RM and argonaute systems, CRISPR-Cas systems are part of the adaptive immune system and are RNA guided DNA or RNA nucleases (Figure 1.2-1 h). As this thesis focuses on a CRISPR-Cas system, the following Chapters will explain what CRISPR-Cas systems are, how they work and how they are classified.

1.3 CRISPR-Cas as adaptive immune system

CRISPR-Cas systems are RNA-guided adaptive immune systems that confer phage resistance in bacteria and archaea[36]. These systems are one of the most prevalent antiphage systems and can be found in 50% of bacteria and 87% of archaea[37]. Although these systems differ in the proteins involved, the underlying mechanisms, and the general scheme of how phage immunity is acquired is the same[38]. The acquisition of phage resistance is divided into three main steps, adaptation, crRNA maturation and expression, and interference (Figure 1.3-1). Adaptation begins with the attack of a new RNA or DNA phage, injecting its genome into the cell, or by a mobile genetic element entering the cell. The adaptation machinery is often composed of the proteins Cas1 and Cas2, which incorporate small DNA sequences of the phage genome called protospacers, into the CRISPR array. This CRISPR array is part of the host genome. If the invading oligonucleotide is RNA, this RNA needs to be reverse transcribed prior incorporation into the CRISPR array[39]. In DNA-cleaving CRISPR-Cas systems, Cas1-Cas2 choose specific foreign sequences to be incorporated, which is dependent on the protospacer adjacent motif (PAM). The PAM is a 3-5 bp long sequence that is not integrated into the protospacer, but it is crucial to distinguish foreign sequences from the host sequences[40,41]. This incorporation into the host genome ensures a heritable immunity[42] and the incorporated spacers are typically sequences of 25-50 bp in lengths, but they can reach up to 72 bp[43,44].

The CRISPR array is part of the host genome and is composed of a leader sequence (L, brown in Figure 1.3-1), palindromic repeats (R, gray in Figure 1.3-1), and spacers (S in Figure 1.3-1) in a clustered fashion. The repeats are short sequences that are repeated and interspaced by spacer sequences. These spacers derive from previous infections and act as a cellular memory. Once the protospacer is incorporated into this memory, the CRISPR array is transcribed by cellular machinery and additionally one or more Cas nucleases are transcribed and translated. Generally, genes of Cas nucleases can be found in vicinity of the CRISPR array of the host genome, as defense island[45]. The transcription product of the CRISPR array is the precursor-crRNA (pre-crRNA), which is cleaved by Cas ribonucleases or host RNases into its monomeric units, the crRNAs[46]. These crRNAs are recruited by individual Cas proteins, or Cas protein complexes to form the surveillance complex. This surveillance complex can bind to the viral genome

through sequence complementarity of the crRNA and the viral RNA or DNA, when the phage attacks for a second time. Binding activates the nuclease, resulting in single-strand or double-strand breaks and sometimes even to a collateral, sequence-independent oligonucleotide cleavage. In DNA-cleaving CRISPR-Cas systems the Cas nuclease recognizes the PAM on the foreign sequence prior to cleavage. This prevents binding of the surveillance complex to the CRISPR array and degradation of the host

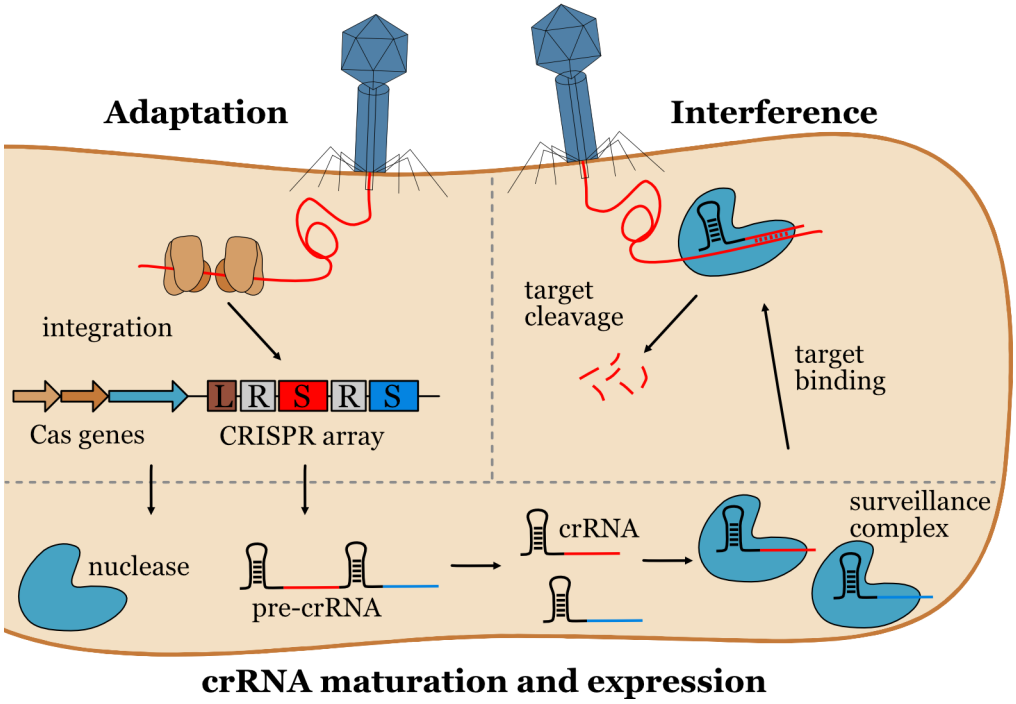


Figure 1.3-1: General CRISPR-Cas immune mechanism. CRISPR-Cas immunity is divided into the adaptation, the crRNA maturation and expression, and the interference step. This mechanism is described in detail in the main text.

genome.[40]

It is worth noting that phages have evolved to counteract CRISPR-Cas systems with small proteins produced by the phage or encoded in the prophage and synthesized by the host cell. These small proteins are called anti-CRISPR (Acr) [47,48]. There are several mechanisms by which Acrs can inhibit CRISPR-Cas systems. Acrs can bind to Cas proteins by mimicking DNA or RNA, such as AcrIIA4[49], or they can bind to the protein region that interacts with PAM, such as AcrVA1[50]. Both mechanisms inhibit oligonucleotide cleavage. Some Acrs can induce dimer formation of Cas nucleases, such as AcrIIC3 and Cas9, which inhibits DNA binding[51]. Acrs have also been shown to have enzymatic activity as for example AcrVA1, which binds to the surveillance complex of Cas12a, altering the nuclease activity such to degrade crRNA[50].

1.4 Class 1 CRISPR-Cas systems

CRISPR Cas systems, are divided into two classes[52,53] (Figure 1.4-1). Both have similar adaptation modules, which often consist of Cas1 and Cas2, but they differ in their effector complexes which drive crRNA maturation and interference. In class 1 systems, interference is carried out by several proteins acting as subunits of one effector complex[54]. In contrast, the effector complex of class 2 systems is composed of one multidomain protein[52]. These two classes are further subdivided into types based on the effector proteins involved. Class 1 consists of type I, III, and IV and represents up to 90% of naturally occurring CRISPR-Cas systems[55]. Class 1 nucleases cleave double-stranded DNA (dsDNA), single-stranded DNA (ssDNA), and RNA. Some type III nucleases produce cyclic oligoadenylates (cOA) during RNA cleavage, which act as messenger molecules that initiate cellular responses[56]. All class 1 effector complexes, called cascades (CRISPR-associated complex for antiviral defense), have a seahorse-like structure that wraps around the crRNA (Figure 1.5-1, left)[55].

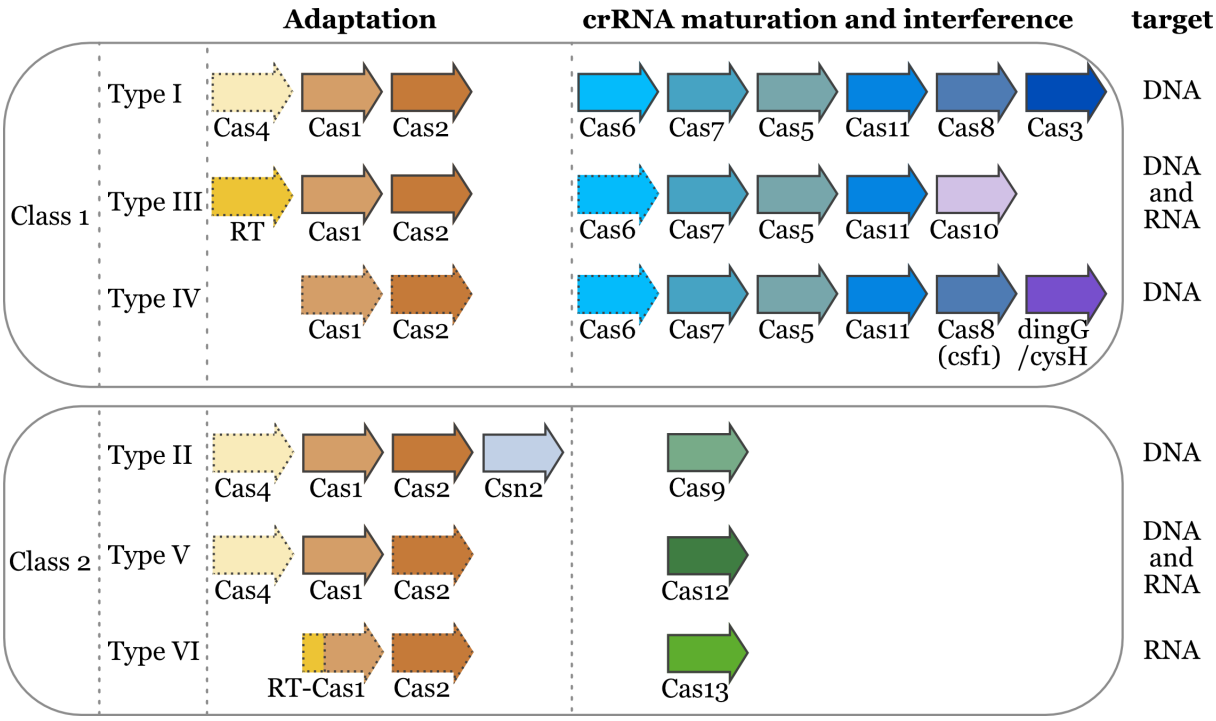


Figure 1.4-1: Organization of CRISPR-Cas systems in 2 classes and 6 types. The genes that make up the adaptation, the crRNA maturation, and interference modules are showed as arrows. The genes that are only present in some subtypes are outlined with dashed lines and in RNA-targeting systems a reverse-transcriptase (RT) is recruited during adaptation. This figure was adapted from the review by Wang et al.[57]

1.5 Class 2 CRISPR-Cas systems

Class 2 CRISPR-Cas systems are divided in type II, V, and VI and it is believed that these evolved later than class 1 systems, since the functions of class 1 Cascade is performed here by one multidomain-protein (Figure 1.5-1)[55]. The class 2 signature proteins of the type II, type V, and type VI CRISPR-Cas systems are Cas9, Cas12, and Cas13, respectively[54] (Figure 1.5-1). These proteins are of particular interest for diagnostic and medical applications, such as genome engineering, as they are easier to deliver compared to multi-protein complexes of five or more proteins. Cas9 is the most studied Cas nuclease and has been extensively used to develop a genome editing tool[58]. DNA cleavage is performed by Cas9 via an HNH domain and a RuvC-like domain, each of which cleave a target DNA strand[52]. In addition to these two nuclease domains, Cas9 also has a recognition (REC) lobe that binds to target DNA.[59] Cas9 uses two RNA strands to recognize and cleave DNA, the crRNA and the trans-activating crRNA (tracrRNA), forming partially a duplex (Figure 1.5-1). For gene editing purposes, a fusion of crRNA and tracrRNA, called single-guide RNA (sgRNA), was developed. This facilitates targeting of specific sequences preceding the PAM 5′-NGG-3′ [52,60].

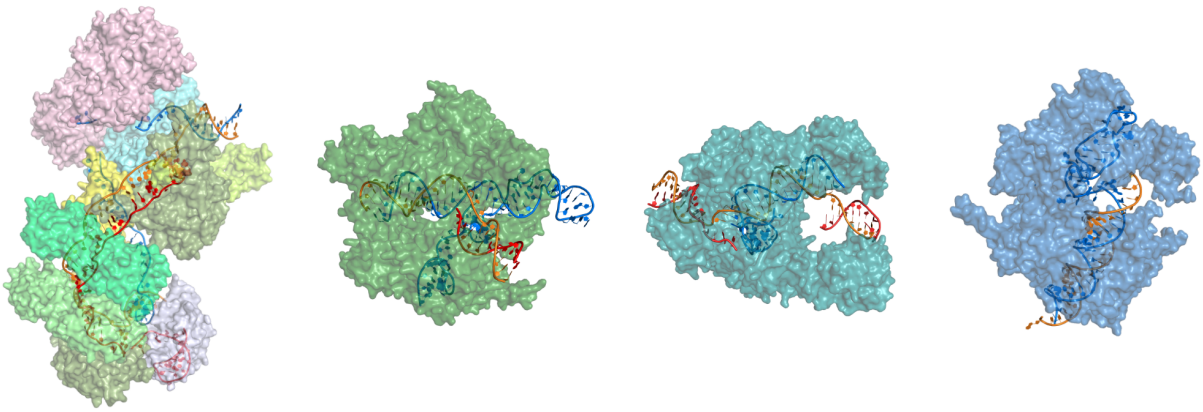


Figure 1.5-1: Structures of different Cas nuclease-complexes with crRNA and targets. From left to right: Example of a class 1 type I cascade complex in comparison to structures of class 2 Cas complexes. Type II and V systems target DNA, the target strand is highlighted in orange and the non-target strand in red. The crRNA is colored in blue. The target of type VI systems is RNA thus the orange oligonucleotide is the target RNA. The PDB IDs from left to right are, 6C66[61], 4UN3[62], 5NFV[63], and 5XWP[64].

Like Cas9, also Cas12 has a RuvC-like nuclease domain. However, the domain architecture and the cleavage mechanism are different. Cas12, which has one nuclease

domain, has to cleave one DNA strand after the other, whereas Cas9 cleaves both DNA strands simultaneously[63]. In general, the Cas12 family is highly diverse, including effector complexes that cleave ssDNA and dsDNA (Cas12f), or even complexes that show non-specific collateral DNA or RNA cleavage (e.g. Cas12g)[65,66]. The maturation of crRNA also varies between Cas12 and Cas9 systems. For example, Cas12a cleaves the pre-crRNA to form the crRNA by itself, while in type II systems, other nucleases are recruited, such as RNaseIII for crRNA maturation[52]. In type VI systems, the pre-crRNA cleavage is performed by a specific active site in Cas13. In general type V and type VI systems are very similar, they exert a collateral cleavage activity. This means, that both cleave RNA and DNA (for type V) or just RNA (type VI) in a sequence unspecific manner[30,54]. Type VI systems target exclusively RNA and cleave via two higher eukaryotes and procaryotes nucleotide (HEPN) binding domains[67]. The advantage of this RNA cleaving enzyme is, that it can provide immunity against RNA phages and against DNA phages, since also DNA phages produce RNA[68–70]. Interestingly, the sequence of the two HEPN domains in Cas13 are similar to the HEPN domains of Abi systems and it is likely, that the Cas13 HEPN domains have evolved from these Abi HEPN domains[53].

1.6 Cas13a as an RNA targeting CRISPR-Cas system

Type VI CRISPR systems are categorized into 4 main subtypes, each with the respective signature protein Cas13a-d[52]. Additionally, the subtypes Cas13X and Cas13Y were recently discovered[71]. Cas13a, Cas13b, and Cas13c are approximately of the same protein size (1228-1118 amino acids), whereas the Cas13d variants are 20-30% smaller (928 amino acids)[52,72]. Independent of this, they all target ssRNA, catalyze pre-crRNA cleavage and target RNA cleavage[52]. In addition, they have collateral cleavage activities. Especially the collateral cleavage makes Cas13a a powerful tool for RNA detection, RNA knockdown, and therapeutics (Chapter 1.6.2). Before explaining how this tool can be exploited, the next Chapter will briefly explain how the immune response with Cas13a works, as Cas13a is the main character of this thesis.

1.6.1 CRISPR Cas13a immunity

In general, the anti-phage response for Cas13a (Figure 1.6-1) is very similar to the CRISPR-Cas immune response shown in Figure 1.3-1. The adaptation of protospacers is done by a Cas1-Cas2 complex. But in type VI-A systems, Cas1 is expressed as a fusion protein with a reverse-transcriptase, that reverse transcribes viral RNA to DNA before incorporation into the CRISPR array[39]. Interestingly, Cas13a has no PAM preference. Some Cas13a variants recognize the protospacer flanking site (PFS) of the target RNA (for example *Leptotrichia shahii (Lsh)* Cas13a), other Cas13a variants don't recognize any PAM or PFS motif (e.g. *Leptotrichia wadei (Lwa)* Cas13a)[73]. PFS and PAM are essentially both small sequences that are recognized by Cas enzymes for self, versus non-self-discrimination. But PAM is found on the target DNA, whereas PFS is found on the target RNA[74].

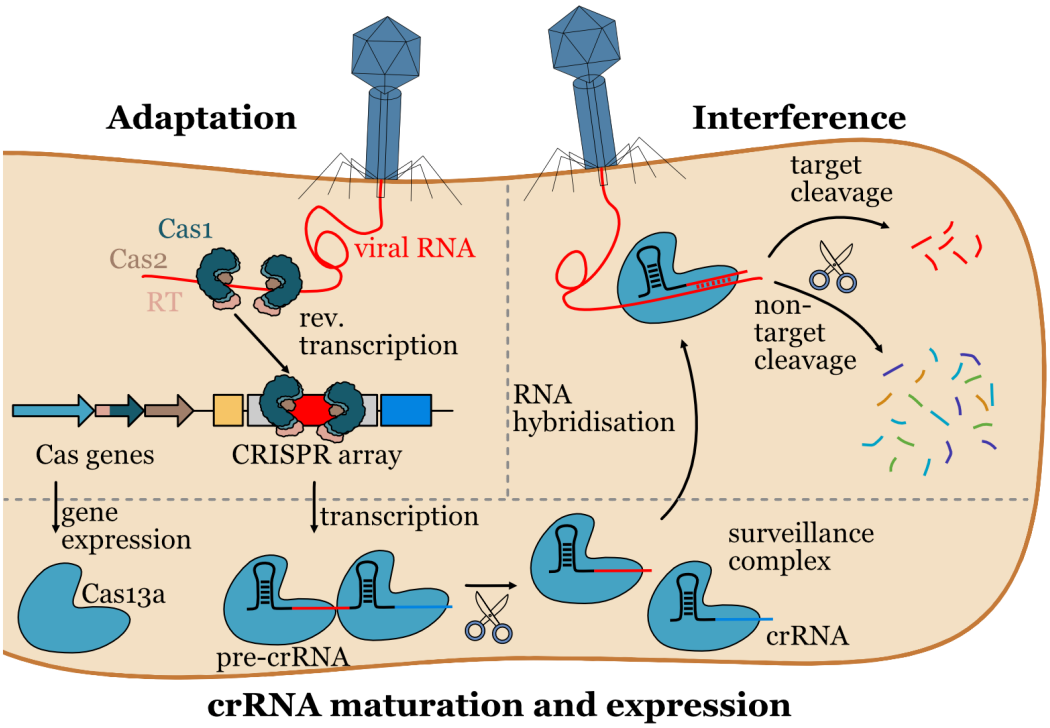


Figure 1.6-1: Schematic overview of the anti-phage response of Cas13a. RT means reverse-transcriptase. The overall mechanism is described in the main text.

After transcription of the CRISPR-array, Cas13a binds to the transcription product, which is the pre-crRNA. The pre-crRNA is cleaved by Cas13a to form the crRNA. This cleavage is done in a divalent ion-independent manner in the whole Cas13a family[72]. Once the crRNA is formed, Cas13a remains bound to this RNA, forming the

surveillance complex[75]. Upon recognition of complementary, invading RNA, the surveillance complex binds and degrades this foreign RNA. Upon binding of the target RNA, a conformational change occurs, which brings both HEPN domains in vicinity, activating the unspecific collateral cleavage of RNA (non-target cleavage, see Chapter 1.6.4)[64]. This collateral cleavage can provide broader immunity against multiple viruses, infecting the cell contemporarily. Even though Cas13a systems show an unspecific cleavage, they do have a nucleotide cleavage preference. *Lbu*Cas13a and *Lsh*Cas13a prefer RNA cleavage at U[76], whereas Cas13a from *Lwa* and *Lba* prefer AU and AC[77].

It was shown that the collateral cleavage activity targets not only foreign, viral RNA, but that Cas13a cleaves also bacterial RNA and has a bias towards cellular tRNA. This tRNA cleavage negatively affects translation of cellular proteins, leading to cell dormancy[30,78].

1.6.2 Applications of Cas13a

The Cas protein that has been exploited the most for genome editing, transcriptional regulation, and medical application is by far Cas9, which is reviewed somewhere else[79]. However, also Cas13a, was applied as a diagnostic and as a therapeutical tool for infamous diseases caused by RNA viruses, which is reviewed in the next paragraphs.

Based on the collateral cleavage activity of Cas13a, a method for nucleic acid detection was developed, called Specific High-Sensitivity Enzymatic Reporter Unlocking (SHERLOCK)[80]. In this paper-based test, the nucleic acids are pre-amplified. An RNA sequence containing a fluorophore and a quencher in vicinity is added to the reaction mixture. If Cas13a with the corresponding crRNA is added and it does not detect the target RNA strand, no fluorescence is seen. If Cas13a with the corresponding crRNA detects the target, it activates the collateral cleavage of the fluorophore-quencher labelled RNA. The fluorophore and quencher are no longer in close proximity and fluorescence can be detected. It was found that RNA and DNA can be detected *in vitro* and *in vivo* with attomolar sensitivity. This platform was shown to work for Zika and Dengue virus detection, on mutated cell-free tumor DNA and on distinguishing pathogenic bacteria from non-pathogenic bacteria. SHERLOCK's sensitivity and

applicability was further improved by simultaneously searching for several pathogenic nucleic acids in one reaction, by combining Cas13 with the Type III ribonuclease Csm6, and by a lateral flow readout, similar to SARS-CoV2 (Severe acute respiratory syndrome coronavirus type 2) tests[77]. In this improved assay, Cas13a cleaves RNA, when the RNA sequence of interest is detected, forming 2',3'-cyclic phosphates. These are recognized by Csm6, leading to an additional non-specific RNA cleavage and resulting in signal amplification[77]. Instead of using fluorophore and quencher, for which specific fluorescence detection equipment is needed, FAM and biotin are used. This enables detection on a lateral-flow readout with antibodies. SHERLOCK was clinically validated by establishing a detection assay of SARS-CoV-2 RNA in throat-swap samples[81] and was also used in Nigeria and Sierra Leone to detect the Ebola virus and Lassa virus on-site[82].

Further, Cas13a was used to detect specific exosomal miRNAs that are biomarkers in breast cancer patients[83]. Apart from the detection of cancer biomarkers, Cas13a was also used to stop cancer growth. It was shown that treatment with Cas13a led to apoptosis and cell growth inhibition in human cancer cells, leaving healthy human cell growth unaffected. Since tumor growth in mice was also inhibited, Cas13a is an interesting system for novel cancer gene therapy approaches[84]. Also glioma cell growth, which are cerebral tumor cells, has been addressed by Cas13a leading to the inhibition of intracranial tumors in mice[85].

Another example of potential Cas13a usage are respiratory diseases. As mentioned above, SARS-CoV-2 was successfully detected by SHERLOCK in throat-swap samples[81], but it can also be used to treat it. It was shown that Cas13a degraded influenza virus in lung tissues of mice and that it reduced SARS-CoV-2 replication and mitigate infections and infection symptoms[86].

In summary, these applications show the massive potential of the CRISPR-Cas13 system for RNA detection and treatment of human diseases caused by RNA. In the next Chapter the molecular and structural basis of Cas13a will be discussed, which is the foundation of the proteins function. Further also structurally and mechanistically unknown features on the molecular level will be targeted.

1.6.3 Structural similarities in the Cas13a protein family

This Chapter deals with the structural features of Cas13a proteins. Special emphasis will be placed on Cas13a from *Leptotrichia buccalis* (*Lbu*Cas13a), as this it is the protein studied in this thesis.

Cas13a proteins have a bilobed architecture consisting of the recognition (REC) lobe and the nuclease (NUC) lobe (Figure 1.6-2).

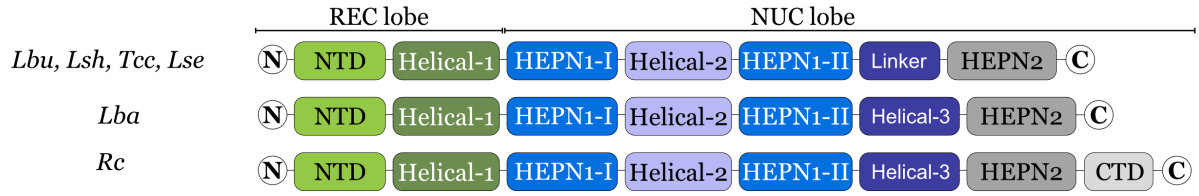


Figure 1.6-2: Domain architecture of different Cas13a proteins. NTD and CTD are the N-terminal and C-terminal domains, respectively[87–93]. The abbreviations for the organisms are: *Leptotrichia buccalis* (*Lbu*), *Leptotrichia shahii* (*Lsh*), *Thermoclostridium caenicola* (*Tcc*), *Listeria seeligeri* (*Lse*), *Rhodobacter capsulatus* (*Rc*), *Lachnospiraceae bacterium* (*Lba*).

Throughout the Cas13a protein family, the REC lobe consists of the same two domains, the N-terminal domain (NTD) and the helical-1 domain. In contrast, the NUC lobe domain architecture slightly varies. The NUC lobe in most organisms shown here (Figure 1.6-2) consists of the HEPN1-I, helical-2, HEPN1-II, Linker, and HEPN2 domains. In *Lba*Cas13a the linker domain is exchanged by a helical-3 domain, while in *Rc*Cas13a the linker domain is exchanged by a helical-3 domain and a C-terminal domain is added. Structures of Cas13a from all organisms are known in *apo* and in different RNA bound states (Figure 1.6-3). But no complete pathway from *apo* to the cr-and target RNA bound complex is structurally described. Beginning with the *apo* state of Cas13a, two structures have been solved. One from *Lsh*Cas13a and the second from *Tcc*Cas13a and both structures are lacking the NTD. For the pre-crRNA bound complex only the structure of *Lba*Cas13a is known, in which the pre-crRNA is 2 nt longer than the crRNA that was resolved in the corresponding crRNA-bound structure[89]. Interestingly, if both known complexes of *Lba*Cas13a are aligned, no structural change is visible in the protein. The Cas13a complex with the most experimentally solved structures is the crRNA bound complex. Only the structure of the crRNA bound TccCas13a is unknown.

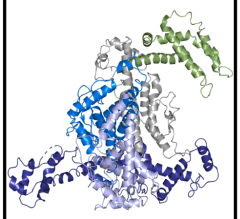
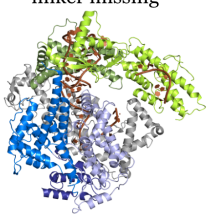
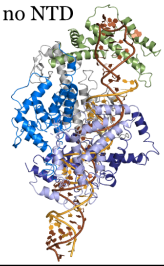
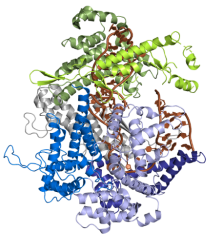
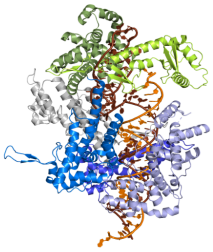
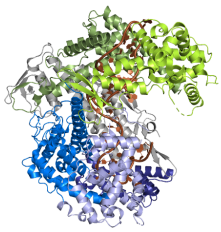
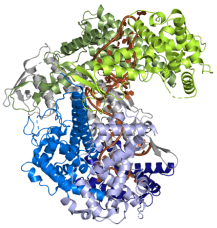
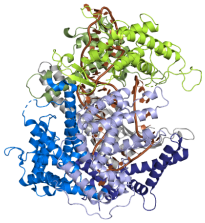
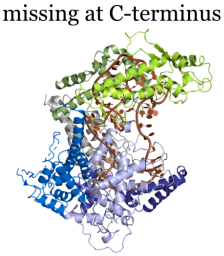
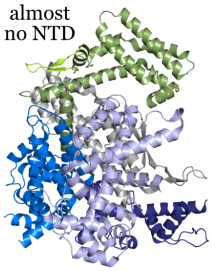
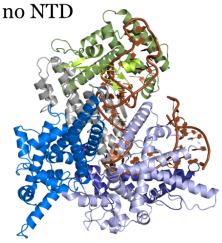
	apo	**pre-crRNA bound**	**crRNA bound**	**cr- & target RNA bound**
Leptotrichia shahii (Lsh)	no NTD 		parts of NTD and linker missing 	no NTD 
Leptotrichia buccalis (Lbu)				
Lachnospiraceae bacterium (Lba)				
Rhodobacter capsulatus (Rc)			>100 amino acids missing at C-terminus 	
Listeria seeligeri (Lse)			>100 amino acids missing at C-terminus 	
Thermoclostridium caenicola (Tcc)	almost no NTD 	almost no NTD 		

Figure 1.6-3: Experimental structures of the Cas13a family. Notes are inserted next to those structures that were truncated or miss a larger number of amino acids[87–93]. The crRNA is colored in brown and the target RNA in orange.

Generally, all crRNA bound structures are similar. The hairpin of the crRNA, the crRNA repeat, is bound in the REC lobe, while the crRNA spacer is bound to the NUC lobe and is more solvent exposed than the crRNA repeat. Interestingly in *Lba*, *Lbu*, *Lsh*, and *Tcc*Cas13a there are certain regions in the crRNA spacer that are not resolved. In *Lbu* and *Lsh*Cas13a these regions comprise 9 and 6 nt, respectively. In *Lbu*Cas13a these are nucleotides 47-52 (Figure 1.6-4 a, shown as dotted line). Generally, this region is called seed region. These nucleotides in *Lbu* and *Lsh*Cas13a are not resolved in the structure of the binary complex but make extensive contacts with the protein in the ternary complex[87,93]. It is supposed that the seed region is the first part of the crRNA that binds the target RNA, upon which the A-form-like duplex (Figure 1.6-4 b) is formed from the seed region on.

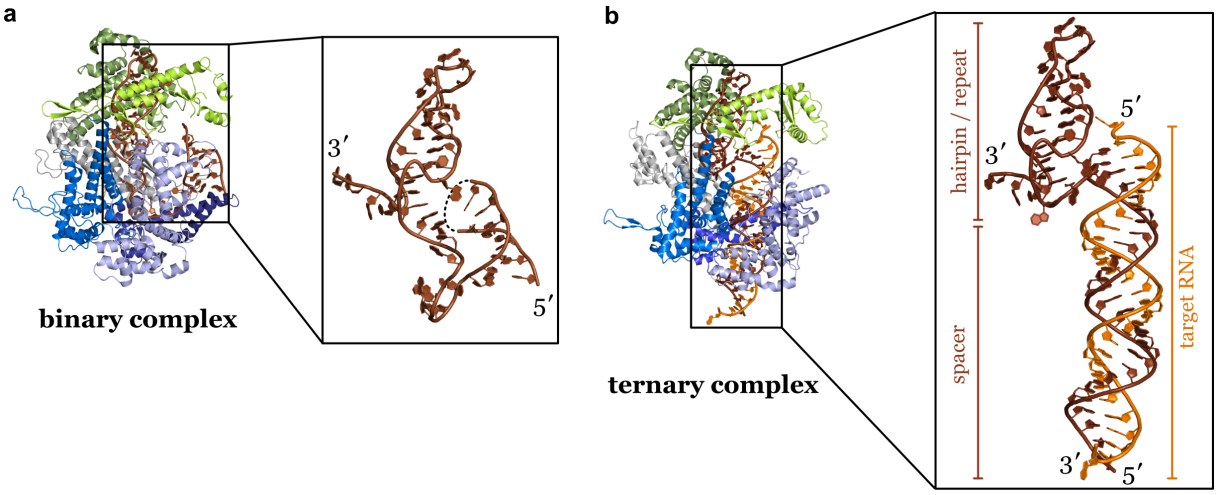


Figure 1.6-4: Structural change of bound crRNA in *Lbu*Cas13a. a, Binary complex (PDB ID: 5XWY) with an insert showing the crRNA structure. The seed region is schematically drawn as dashed line. **b**, Ternary complex (PDB ID: 5XWP) with an insert showing the crRNA-target RNA duplex.

Only two structures are known for the ternary cr- and target RNA bound complex, which are from *Lbu* and *Lsh*. As briefly described before, an A-form-like duplex is formed between the crRNA and the target RNA in both structures (Figure 1.6-4). When focusing on the crRNA in *Lbu*Cas13a (Figure 1.6-4) it can be seen that some parts undergo substantial conformational changes during ternary complex formation. The crRNA hairpin at the 3′ region, which is the CRISPR repeat, shows no conformational change upon target RNA binding. In contrast, the 5′ region of the crRNA, which is the CRISPR spacer, needs to undergo a substantial conformational change to enable duplex formation with the target RNA (Figure 1.6-4). On the protein side, the REC lobe does not undergo a conformational change in *Lbu*Cas13a. But the conformation of the

helical-2 domain (light blue domain in Figure 1.6-4) and, to a lesser extent, the linker domain, change. The helical-2 domain rotates towards the linker domain to allow channel formation, which is needed for the cr- & target RNA duplex. This conformational change is in-line with the conformational change seen for *Lsh*Cas13a from the crRNA bound to the cr-and target RNA bound complex[93]. Additionally, both HEPN domains get closer together upon target RNA binding (see Chapter 1.6.4), to activate the active site in *Lbu*Cas13a for target RNA cleavage.

1.6.4 RNA cleavage in *Lbu*Cas13a

Cas13a catalyzes two cleavage reactions, the pre-crRNA and the target RNA cleavage and the extend of knowledge of both is different. This Chapter summarizes what is mechanistically known about RNA cleavage in Cas13a, especially focusing on *Lbu*Cas13a.

Generally, the active site of pre-crRNA cleavage varies in Cas13 proteins from different organisms[94]. In *Lsh*Cas13a, the pre-crRNA cleavage is catalyzed by the active site consisting of R438 and K441 in the helical-1 domain of the REC lobe[93]. However, in *Lbu*Cas13a it is not exactly known where the pre-crRNA cleavage site is. There it was shown that residues in the helical-1 domain lower pre-crRNA cleavage activity[72,75,76,94]. However, also the NUC lobe modulates pre-crRNA cleavage in *Lbu*Cas13a[76]. There is evidence that R1079 in the HEPN-2 domain is required for cleavage[75]. In previous studies the mutation R1079A abolished pre-crRNA cleavage activity, by retaining binding of pre-crRNA[75]. Furthermore, this mutation does not affect target RNA binding and cleavage. Specifically, the pre-crRNA is cleaved five nucleotides upstream of the hairpin repeat sequence and, interestingly, the cleavage rate can be modulated by the hairpin sequence. Changing the length of the hairpin stem or the hairpin loop attenuates RNA cleavage, while inversion of the hairpin sequence leads to a strong reduction of pre-crRNA cleavage[75]. The mechanism of pre-crRNA cleavage was revealed by the crystal structure of *Lba*Cas13a. There the 5'-end of the pre-crRNA is positioned in such a way that a nucleophilic attack can occur by the 2'-OH group of the ribose on the phosphate, generating a 2'-3'-cyclic phosphate as a classic acid-base reaction product[89]. The cleavage was shown to be metal-ion independent and the postulated mechanism is shown in Figure 1.6-5[75].

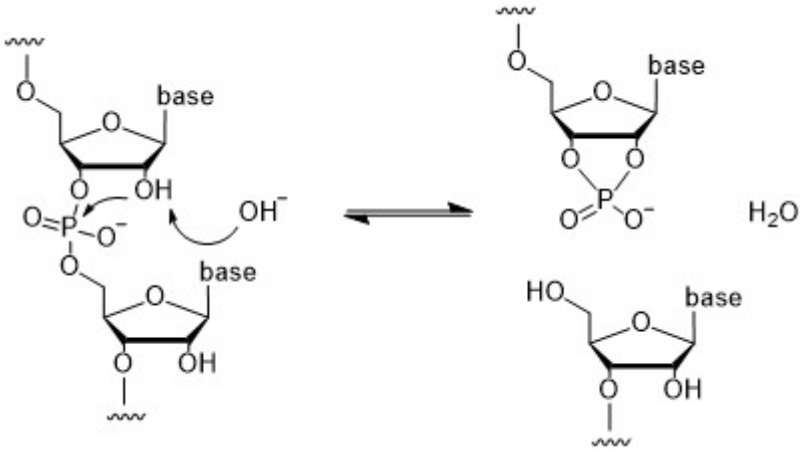


Figure 1.6-5: Postulated mechanism of RNA cleavage[89].

Unlike pre-crRNA cleavage, the target RNA cleavage is divalent metal-ion dependent and it is approximately 80 times faster than pre-crRNA cleavage[75]. The target RNA cleavage site in Cas13a proteins is generally known. Cleavage occurs at both HEPN domains via one R-X_{4-6}-H motif in each domain[94]. In *Lbu*Cas13a this cleavage site is composed of R472 and H477 in the HEPN1 domain and R1048 and H1053 in the HEPN2 domain (Figure 1.6-6 a, b). Mutation of one of the two R-X_4-H motifs to alanine abolishes target RNA cleavage, while retaining pre-crRNA cleavage activity and pre-crRNA, crRNA, and target RNA binding[75].

In *Lbu*Cas13a the target RNA is only cleaved when the complementary crRNA is bound. Once the target RNA binds to the crRNA bound complex, the target RNA is cleaved, and a collateral cleavage is activated, leading to a cleavage of non-target RNA in a sequence-independent manner. It is postulated that the target and non-target RNA cleavages are performed by the same active site since the cleavage rates are very similar.

Knowing the active site of target and non-target RNA cleavage, allows to rationalize why the target RNA needs to bind to the complex to activate cleavage and why collateral cleavage *per se* happens. Upon binding of target RNA, the conformation of the protein changes so that residue H477 is turned around towards the HEPN-2 domain (Figure 1.6-6 a, b). In addition, the two HEPN domains move closer together, resulting in an activated cleavage site. By this, the binding of the target RNA is triggering the conformational change that alters the geometry of both arginine and histidine residues important for cleavage (Figure 1.6-6 a, b). Looking at the overall position of the active site in the protein (Figure 1.6-6 c) it is seen that this part is lying in a cleft between the

HEPN1 and HEPN2 domains. The two domains are shielding this active site to some extent, but it is solvent exposed. This solvent exposure of the cleavage site explains why also non-target RNA strands are degraded once target RNA cleavage is activated.

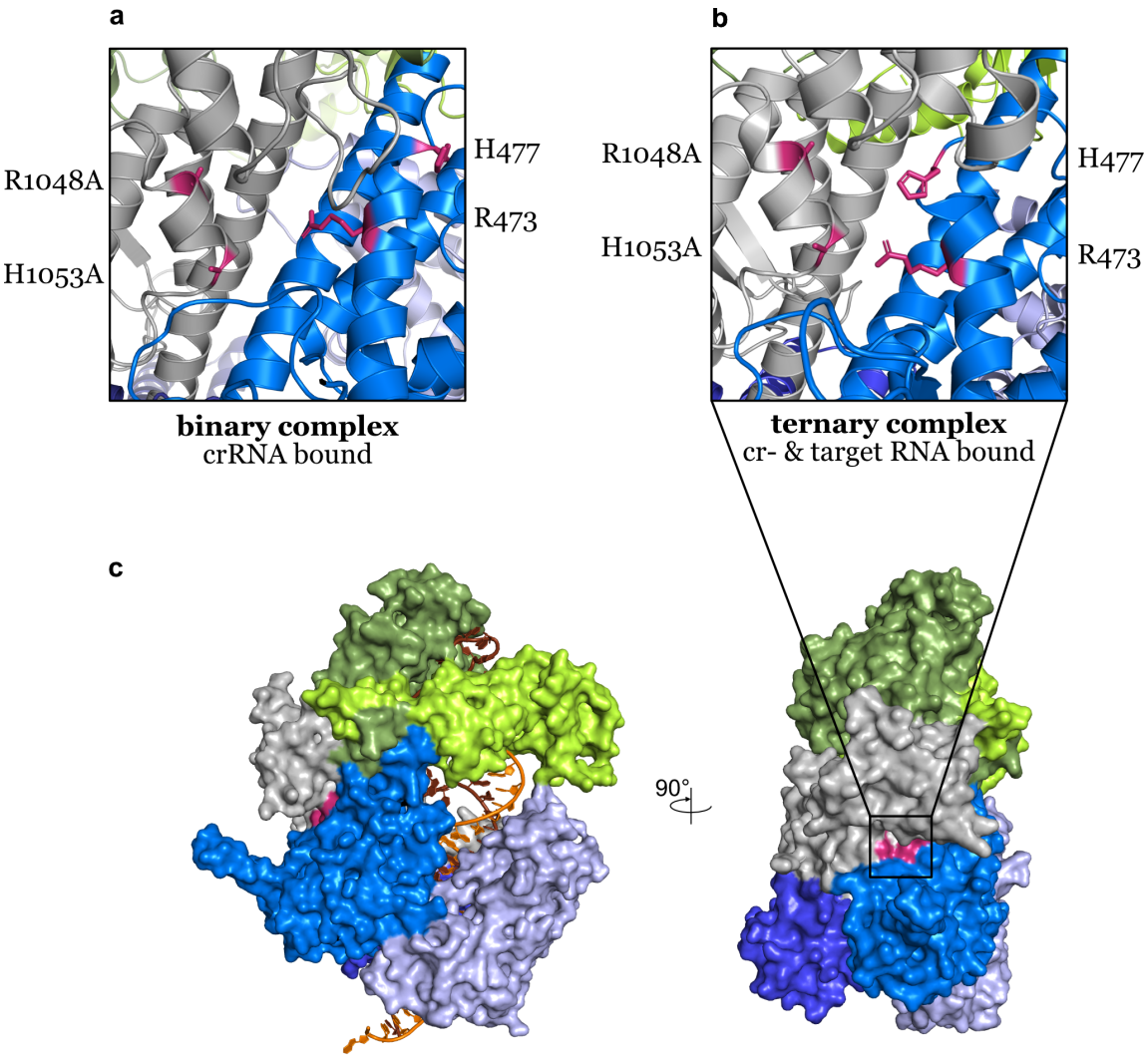


Figure 1.6-6: Comparison of the target RNA cleavage site in the binary and ternary complex. a, Close-up views of the HEPN1 and HEPN2 domains from the cryo-EM structure of the binary complex of *Lbu*Cas13a (PDB ID 5XWY)[87] are shown in blue and grey, respectively. The residues composing the target cleavage site are highlighted in pink and the respective residue is shown at the sides. Two residues were mutated to alanine to inhibit target RNA cleavage. **b**, Same as in a, but for the ternary complex of *Lbu*Cas13a (PDB ID 5XWP)[87]. **c**, Surface representation of the ternary complex, rotated by 90°. The magenta area highlights the active site for target RNA cleavage, which is solvent exposed. The crRNA is colored in brown and the target RNA in orange.

1.6.5 Unknown aspects of *Lbu*Cas13a

As mentioned above, structures of *Lbu*Cas13a in the crRNA bound and in the cr- and target RNA bound states were solved via cryo-EM and X-ray crystallography, respectively. However, important conformations on the functional pathway are still unknown, such as the conformational changes of domains from *apo* to pre-crRNA bound and to crRNA bound complex. This is especially valid for the REC lobe. The only structures of the apo and pre-crRNA bound complexes that exist lack the NTD. This makes it difficult to mechanistically understand the initial RNA binding step. Generally, structures are missing in the functional pathways of all other Cas13a proteins (Figure 1.6-3) as well. This makes it difficult to understand the full process of how Cas13a works. Moreover, more than half of the known structures are truncated, missing over 100 amino acids or even entire domains at the N- or C- terminus. This makes it even more difficult to understand the conformational changes, that underlie the proteins' function. The missing structures can be predicted by AlphaFold2 (Chapter 1.7.3), launched in 2021[95] and AlphaFold3 launched in 2024[96]. However, structures predicted with AF2 and AF3 cannot explain the dynamics of the protein, which will be covered in the next Chapter. Other methods are therefore needed to determine the missing structures and to detect and follow the conformational changes that Cas13a undergoes upon RNA addition, providing information on the dynamics of the protein.

1.7 Methods for structure determination of biomacromolecules

First, a general distinction needs to be made between the structure and the dynamics of biomolecules, e.g. proteins. A protein's structure is only a snapshot of one specific conformation on the conformational landscape[97]. The conformational landscape of a protein consists of many different maxima and minima, each with its free energy. Each point on this energy surface represents a distinct protein conformation. In solution, several points on this surface and several local minima can be populated, since the energy barriers (maxima) between these minima may be small. This leads to proteins switching between these local minima, which means that these are switching between different conformations. This is what dynamics describe. Protein dynamics are time dependent fluctuations of atomic coordinates leading to different conformations of

proteins. This process includes rotation of atomic bonds and rotations of entire domains, but it does not include cleavage and formation of bonds between atoms[98].

Several methods have been developed to determine structures of proteins, oligonucleotides, and their complexes. Some of these methods can only describe one structure, others can also describe to some extend the dynamics of a system, which will be discussed separately for each method.

Focusing on the structure, a distinction is usually made between methods yielding atomic high-resolution structures of biomolecules including X-ray crystallography, cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR), and methods yielding coarse-grained structural information, for example atomic force microscopy (AFM), small-angle X-ray or neutron scattering (SAXS/SANS), Förster resonance energy transfer (FRET), cross-linking mass spectrometry (XL-MS), selective 2' hydroxyl acylation analyzed by primer extension (SHAPE), and electron paramagnetic resonance (EPR) pulsed dipolar spectroscopy (PDS)[99]. Each one of these methods has advantages and limitations. Integrative structural biology is the approach to combine the strengths of multiple experimental and computational methods to develop structural and dynamic models of the biomolecule of interest[99]. The following Chapters will briefly describe the mentioned methods as their 'classical' and most popular experimental setups.

1.7.1 Methods for atomic high-resolution structure determination

The most common techniques used for protein structure determination at atomic-resolution are X-ray crystallography, cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR), making up 98.5% of the experimentally determined structures (Figure 1.7-1). Most of the deposited structures were solved with X-ray crystallography. One advantage of X-ray crystallography is that it has no size limitations regarding the biomolecule and that this technique often provides high resolution structures. However, it is strongly limited in describing protein dynamics, since the proteins are in a crystalline state.

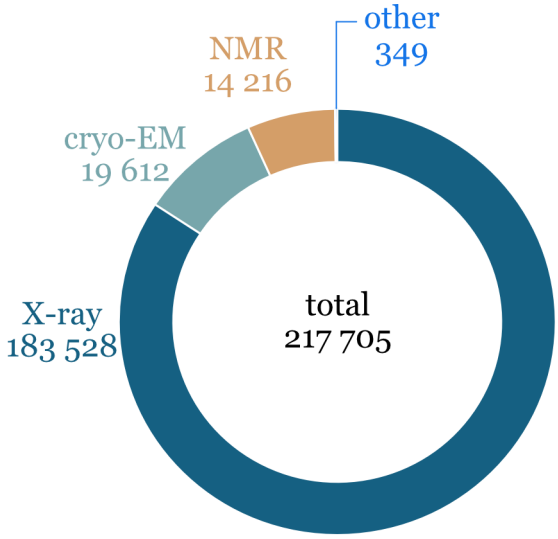


Figure 1.7-1: Diagram showing the composition of known experimental structures in the protein data bank (PDB). 84% of the experimental structures are solved by X-ray crystallography, 9% by cryo-EM, 6.5% by NMR, and 1.5% by other methods. Data source: Research Collaboratory for Structural Bioinformatics Protein Data Bank (PDB), www.rcsb.org, accessed on the 27.03.2024.

Another limiting factor is the crystallization step, which is a trial-and-error process based on large screenings, and which can get very difficult, if not impossible for e.g. membrane proteins, very flexible proteins, and intrinsically disordered proteins (IDP)[100]. Due to this, proteins are often truncated, to delete flexible regions and facilitate crystallization. However, these regions can be important for the proteins function. For nucleic acids, the crystal preparation is also challenging since they have structurally a high electrostatic repulsion that origins from the phosphate backbone. One method to overcome this is co-crystallization with proteins[101]. If the biomolecule successfully crystallizes, the next challenge is to get crystals that are diffracting with high resolution, which can be very tedious. Also, the structures that are solved with X-ray crystallography represent low-lying energy states on the conformational landscape of the protein, leading to crystal packing effects or other artefacts, which are not necessarily features important for the proteins function[100]. Even in the case of collected high-resolution data, X-ray crystallography is faced with the so-called phase problem, to reconstitute an experimental electron density. Each diffraction spot contains three parameters, the wavelength, the amplitude, and the phase. The wavelength is known by the wavelength of X-ray diffraction source, which is not altered by diffraction on a crystal. The amplitude is known by the intensity of the spots in the diffraction image. But a direct determination of the phase is not possible[102]. Methods to overcome this

are molecular replacement, single isomorphous replacement or multiple anomalous dispersion[102,103]. More recently, through the development of AlphaFold2 (AF2, and AF3), also computational structure prediction with high accuracy has become available, facilitating the solution of the phase problem by molecular replacement[95].

Cryo-EM is a complementary technique to X-ray crystallography, which has experienced main developments in the last decades. Initially this technique was used for studying viruses and tissues that were stained with heavy atoms (negative stain cryo-EM). In the 1980s and 1990s, metal grids were developed and a new method of flash-freezing proteins with liquid ethane, being the first part of the so-called resolution revolution[104]. Secondly, in the early 2000s, Henderson, Agard and others led the resolution revolution by working on new digital direct electron detectors that were first available in 2012[104,105]. These new detectors enabled beam-induced motion correction, leading to resolution enhancement[106]. At the same time, the third part of the resolution revolution was happening with the development of software able to model a 3D protein structure out of 2D images. These three key innovations made cryo-EM a method for biomolecular structure determination with atomic resolution, resolving big complexes as the ribosome[107] and transcription assemblies[108]. Unlike X-ray crystallography, cryo-EM takes a snapshot of the protein in solution, at the freezing temperature. This enables to resolve flexibility and multiple conformations of a biomolecule in one sample. Another strength of cryo-EM is the small amount of biomolecule needed, since single particles are resolved[105]. As a disadvantage, this technique has size limitations of the biomolecule, the lower edge being around 40 kDa[109]. Additionally, the sample needs to be vitrified and is measured at cryogenic temperatures[109]. Furthermore, as in X-ray crystallography, very flexible protein regions will not be resolved because the conformations will average out by stacking and summarizing the single particles[109]. One exciting development of cryo-EM is cryo-electron tomography (cryo-ET), enabling 3D reconstitutions of single proteins, but also entire cells[110]. This is interesting, since several conformations of biomolecules can be resolved in their native environment (*in-cellulo)*, such as the ribosome[111].

One of the strengths of in-solution NMR is that it can analyze conformational dynamics at room temperature. The idea behind structure determination via solution NMR is to obtain a set of parameters as dihedral angles and distances between atoms. These are used as restraints to build a 3D model of the biomolecule[112]. For structural modeling, the Nuclear Overhauser Effect (NOE), the residual dipolar couplings (RDC), and the

paramagnetic relaxation enhancement (PRE) are commonly used. NOE is a magnetization transfer process of spins, coupled through space that provides short-range distances (≤6 Å) and J-couplings, which depend on the dihedral angles[112,113]. RDC relies on the dipolar coupling between nuclei and provides information about the relative orientation of the inter-nuclei vectors in biomolecule[114]. In contrast, PRE relies on the dipolar interaction between a paramagnetic center and a nucleus. The classic PRE therefore relies on the insertion of paramagnetic centers, as spin labels and can provide distance constraints between the paramagnetic center and the nuclei of up to 20-30 Å[114]. NMR can further be used to track conformational changes in a time-dependent manner[115]. In addition, solution NMR can be very helpful to determine and confirm secondary structures of RNA sequences. Here, the imino-protons in the base pairs are used, to detect double stranded regions[116,117]. Newer advances show, that through isotope enrichment also protein structures in a cellular environment can be solved[118]. However, even though great improvements have been made, the major disadvantage of solution NMR is the size restriction of biomolecules to approximately <70 kDa[119] for classical NMR. Through advanced NMR techniques as Methyl TROSY, structures of 100 kDa proteins can principally be solved[120]. For larger systems, as protein-protein complexes, solid state NMR (ssNMR) can be used. The advantage of ssNMR is that it has no size limitation on the biomolecule and that methods as isotope labelling[121] can be used to simplify the ssNMR spectrum[122]. Especially for membrane proteins, ssNMR became a powerful tool for conformational characterization, since these are often not soluble in aqueous solutions, difficult to crystallize and have a high molecular weight when considering their membrane mimetic systems[123]. Recently, even a membrane protein structure in *E. coli* cellular inner membrane was solved[124], highlighting the power of ssNMR for structure biology.

1.7.2 Coarse-grained methods for structure determination

For X-ray crystallography and cryo-EM the sample is crystallized or vitrified, respectively, which does not necessarily resemble the native-like liquid state. Additionally, both methods are limited in resolving flexible regions and distinct conformations of a biomolecule. While this is the power of NMR, which can be measured in liquid solutions, solution NMR is restricted to small biomolecules.

Thus, a combination of these methods with low-resolution coarse-grained methods has been shown to be very helpful[125]. Integrative structure biology via coarse-grained methods is the tool of choice and can help to unveil a protein's structure and dynamics. These methods include but are not limited to AFM, SAXS/SANS, XL-MS, SHAPE, FRET, and EPR PDS.

One coarse-grained method is AFM, which measures low-resolution topological images in the submolecular to subcellular range[126]. Typically, the molecule is held in place on a surface during the measurement by electrostatic interactions between the molecule and the surface[127]. The measurement is made by scanning over the surface with a cantilever, equipped with a tip. This tip is in contact with the surface and topological changes e.g. through a biomolecule, result in a height adjustment of the cantilever and tip. This contact mode, can be performed in liquid solution, and can resolve processes in realtime[126]. The lateral resolution of AFM images is about 0.5-1 nm and the vertical resolution is 0.1-0.2 nm[128]. AFM can be used for example to characterize protein structures and shapes, to analyze protein-nucleic acid complexes, observe biomolecular assemblies and aggregation, and analyze ligand-induced changes[129,130]. The main drawbacks are, that the biomolecule must be immobilized on the surface, the error of the biomolecular size depends on the size of tip, and mechanical effects and interactions between the molecule and the tip may occur[131,132].

SAXS and SANS are two complementary techniques that characterize the overall size and shape of a biomolecule through X-ray scattering of electrons or neutron scattering of nuclei[100]. During a SAXS experiment the sample is irradiated by an X-ray beam and the scattered radiation at small angles is detected[133]. SAXS is measured on samples with resolutions between 1 nm and 1000 nm[134]. Main advantages of this technique are that the sample does not need to be vitrified, crystallized or fixed onto a surface. However, an X-ray or neutron radiation source is needed, which is generated for example at synchrotron beamlines. Often SAXS is combined with size exclusion chromatography (SEC-SAXS). This has the advantage to remove aggregates and oligomers prior SAXS, and by this to measure just the scattering contribution of the biomolecule of interest[134]. For integrative structure prediction SAXS helps to restrict the conformational space of the biomolecule of interest and is powerful when combined to other techniques as for example PDS EPR[125].

XL-MS (also called CL-MS) is used to investigate the structure and dynamics of

biomolecules but also to investigate protein-protein interactions[135] and protein-RNA interactions.[136] The idea is to add a reagent to the biomolecules that covalently crosslinks functional groups of the amino acid side chains, followed by enzymatic digestion. Since these crosslinkers have a specific length, this length is used as a constraint together with the position of the amino acid residues, that have been crosslinked. By crosslinking several amino acid residues with different crosslinker lengths, diverse constraints can be measured that enables a 3D structure modeling of the protein or of protein-protein or protein-RNA interfaces. The strength of XL-MS is that also low-affinity interacting partners and transient interaction partners can be trapped and analyzed[135]. However, the analysis of the obtained fragments can be challenging[135].

A method that is used for secondary RNA structure determination at single nucleotide resolution is SHAPE[137]. Unpaired nucleotides adopt more conformations, which increase the nucleophilicity, compared to base-paired nucleotides. This can be exploited, by adding electrophiles, which react faster with the 2'OH of the ribose belonging to an unpaired nucleotide, than with one of a base-paired nucleotide[137]. Importantly, this reaction is mostly independent on the type of RNA base[138]. After this chemical modification of RNA, radiolabeled complementary DNA (cDNA) is annealed to the RNA of interest and a reverse transcription is performed[137]. The reverse transcriptase stops when the RNA modification or a stable RNA structure is encountered and the obtained cDNAs are separated by high-resolution gel electrophoresis. Upon comparison with control reactions, a secondary structure model of an RNA strand can be developed.

FRET measures the non-radiative energy transfer between two fluorophores in dependence of their proximity. In contrast to the other coarse-grained methods for structure determination, classical FRET relies on the introduction of two spectrally different fluorophores, a donor and an acceptor that are attached to the biomolecule via flexible linkers[139]. Both fluorophores are dipolarly coupled and the emission spectrum of the donor and the excitation spectrum of the acceptor overlap[140]. Upon excitation of the donor, a non-radiative energy transfer to the acceptor occurs. From the efficiency of the energy transfer, the FRET efficiency, an interdye distance can be calculated with

$$E_{FRET} = \frac{1}{1 + \left(\frac{r}{R_0}\right)^6} \quad (1)$$

Where r is the interdye distance and R_0 is the distance at which the FRET efficiency equals 50%, also called Förster radius[141]. R_0 is calculated for each fluorophore pair individually and the distance that is measured by FRET is strongly dependent on the Förster radius, thus on the fluorophore pair used. The optimal distance for a FRET measurement is around the Förster radius, because of the high FRET energy changes with respect to the distance in this regime[119]. In addition to the interdye distance, the FRET efficiency also depends on the spectral overlap of the dye pair, and their relative orientation[119,142]:

$$R_0^6 = \frac{0.529\kappa^2\phi_D J(\lambda)}{N_A n^4} \quad (2)$$

Here, κ^2 is the orientation factor, ϕ_D is the quantum yield of the donor, $J(\lambda)$ is the spectral overlap of the donor emission spectrum and acceptor excitation spectrum, N_A is the Avogadro number, and n is the refractive index of the medium[142]. The orientation factor κ^2 is one large uncertainty in FRET efficiencies. This factor describes the orientation of the acceptor and donor dipole to each other. Usually it is set to 2/3, but this is strictly only valid for free-rotating fluorophores in solution[140]. The highest orientation factor value is 4, which is reached by a collinear arrangement[140]. By a parallel orientation of both fluorophores a value of 1 is reached[140]. Since the orientation factor scales with R_0^6, this would lead to a distance change of 26% between a collinear and a parallel label arrangement, which is quite significant[140]. If the dipoles are oriented in a perpendicular fashion, the orientation factor becomes 0[140]. So, the orientation parameter is a significant uncertainty in FRET measurements. Sometimes, the orientation parameter can be estimated through anisotropy measurements, which describe the orientational freedom of fluorophores attached to a molecule in their excited states, but this is not always possible[143]. Besides the orientation factor, the FRET efficiency sometimes needs to be adjusted by setup-dependent parameters[119].

Furthermore, another drawback of FRET is the size of most fluorophores with flexible linkers. The probability of structure perturbation is higher for large labels, compared to PDS EPR spin labels, which are usually smaller (Figure 1.8-4)[144,145]. Further, through their long and flexible linkers, fluorophores can also adopt preferred orientations, which artificially change the distance measured by FRET[119,140]. This means that it is

more difficult to transfer the interdye distance to the distance of the protein regions undergoing conformational changes[119]. Generally, a distance error of ± 5 Å is commonly estimated for FRET[119].

The strengths of classic FRET measurements are that they can be performed in solution with mobile[146] or immobilized[147] molecules, in vitro[147] or in cells[148] with ångström precision[149]. Other strengths of FRET are the sensitivity to distances between 2.5-10 nm, the ability to resolve also dynamic properties of molecules and molecular complexes (as also NMR, PDS EPR and others), and the low sample amount needed[150]. In addition, FRET can also detect transition states and very rare events, which is similar to XL-MS[150]. Further, FRET detects only fluorophores in the biomolecules, hence there is no size restriction of the biomolecule itself, in contrast to solution NMR, and finally one single molecule can be analyzed through smFRET[150].

1.7.3 AlphaFold2 and 3 – A method to predict protein structures

As mentioned before, the release of AF2 in 2021 initiated a revolution in structural biology, allowing to get a predicted structure for every known protein[151]. Because AF2 and AF3 structures were calculated and used in this thesis the following Chapter is dedicated to briefly describe what AF2 is and how it roughly works[95]. Additionally, the differences between AF2 and AF3 are briefly discussed. Even though exclusively AF was used, it is still worth mentioning that also other models for structure prediction have been developed, such as RoseTTaFold[152].

AF2 is an evolutionary-based computational approach that predicts 3D structures of proteins and protein complexes (in its multimer variant[153]) on the basis of the proteins amino acid sequence, with experimental accuracy[95]. It uses a neuronal network that was trained with constraints including physical, evolutionary and geometric constraints from known protein structures[95]. An AF2 calculation starts with the input, being the amino acid sequence of the protein under study. A multiple sequence alignment (MSA) is generated containing various sequences that are similar to the input sequence, but not identical. This is done to detect parts in the protein sequence that are more likely to be mutated during evolution and to detect possible correlations

between these. At the same time, AF2 generates the first structure of the protein of interest, by comparing protein structures in databases as the PDB to the unknown protein. This representation is the so-called pair-representation. Once these two pieces of information are ready, they are used as input for the neuronal network called Evoformer. The Evoformer refines the pair representation based on the MSA and the MSA based on the pair representation, which is done in several cycles. Lastly, the output of the Evoformer, being an improved MSA and an improved pair representation, is put into the structure module. The structure module, also a neuronal network, constructs the final 3D structure[95].

Besides the structure of the protein, AF2 also calculates two scores, the predicted aligned error value (PAE) and the per-residue accuracy/confidence score of the structure, called the predicted local distance difference test (pLDDT). The PAE is a 2D matrix that shows the error of each amino acid position with respect to a second amino acid at a second position. If for example the PAE is very low between two amino acids in two different protein domains, then the domains are predicted to have a defined position and orientation toward each other. If the PAE between two residues is high, the relative position is insufficiently described and should be interpreted with caution[154]. The second value is the pLDDT, which is a common metric used for protein structure prediction[154]. It uses high values of 100-70 for protein regions that are well predicted and low values of 70-0 for regions with low accuracy[154]. It was shown that low pLDDT values can indicate disordered protein regions[155]. In the human proteome it was hypothesized, that 37 – 50% of the proteins include disordered residues and it was shown that these greatly correspond to regions with a low pLDDT[156,157]. However, AF2 overestimates disorder in protein sequences[155]. In addition, low pLDDTs can also origin from protein regions with folded domains, that are badly described by AF2 due to the lack of co-evolutionary information, or can origin from protein hinge regions[158,159]. In contrast, a high pLDDT score can indicate protein regions, that are misfolded in *apo* and folded when bound to a binding partner. It was shown, that AF2 predictions sometimes correspond to a bound structure and not the *apo* structure[160,161]. What this essentially indicates is, that pLDDTs are very difficult to interpret and that other methods, as molecular dynamics (MD) simulations can become a very insightful addition to the AF2 structure prediction. In addition, the AF2 prediction can also be used as an input for MD simulations.

AF2 opened a new path of integrative structure biology, when combined with classic experimental techniques. As briefly mentioned before, AF2 structures are used for molecular replacement in X-ray crystallography, as input for MD simulations and in general as starting point for protein structure determination[162]. Further, AF2 can predict structural effects when mutations are inserted in a protein[163,164] and can generate contact maps between proteins, that can be experimentally used to design mutations, that disrupt these interfaces[159]. However, AF2 results need careful evaluation. What AF2 is unable to predict are ligand-bound states, as RNA bound states and in general binding partners. It also cannot predict conformational ensembles by default, even though there are first attempts to do so[165,166].

In contrast to AF2, AF3[96] can predict structures with binding partners as ions, small molecules and oligonucleotides. Further AF3 can implement several residues with post-translational modifications into the prediction of protein structures. One main difference between AF2 and AF3 is that the code of AF3 is not accessible for research. An AF3 server[167] was launched, where researchers can upload sequences of interest for structure prediction. The prediction is easy to perform and fast. As in the AF2 multimer, oligomeric structures can also be predicted with AF3 but with higher accuracy. This higher accuracy is achieved for example by replacing the Evoformer with a simpler Pairformer module, which reduces the amount of MSA processing. In addition, the structure module is replaced by a diffusion module, which directly predicts atom coordinates, and the training sets were adjusted. Regarding accuracy, protein structure predictions were tested against a test set of proteins and protein complexes. It is worth noting that even though the accuracy of *apo* protein structure prediction was found to be 86.9% for AF3 (85.5% for AF2.3 multimer), the accuracy of complexes can be very low (Table 1.7-1). Especially protein-RNA complexes have a very low mean accuracy of 39.4%. This is still more accurate than RoseTTAFold2NA[168] (19%), but also these structures need careful evaluation. Since AF3 was released during the writing process of this thesis, it is unclear if and to which extend conformational ensembles can be calculated.

Table 1.7-1 Mean accuracy of AlphaFold3[96].

prediction of	mean accuracy [%]
protein monomers	86.9
protein-protein	76.6
protein-dsDNA	64.8
protein-RNA	39.4
protein-small ligands	76.4 - 93.2
RNA	54.5

Despite AF3 revolutionizing protein structure predictions, no information about the dynamic behavior of proteins is given. There, Molecular Dynamics (MD) simulations can be applied as an additional theoretical method. MD simulations generate a time series of protein conformations by solving Newton's equations of motion for every atom in the biomolecule for each time step. These conformations are typically simulated over a time frame of nano- to microseconds[169,170], but also time scales over microseconds have been achieved[171]. A movie is generated from these single conformations called a trajectory. Conveniently, AF3 structure predictions, as well as experimental structures can be used as starting points of these simulations. Despite dynamic information about the protein conformational ensemble, also thermodynamic properties can be obtained, as free energies and entropies[171].

1.8 EPR spectroscopy

In the following Chapters the focus is set on EPR spectroscopy as method for structure-function elucidation in biomolecules. This gives basic knowledge to understand and interpret the measurements performed in this thesis. For in depth understanding of the theory behind EPR spectroscopy and the underlying spin physics, the reader is referred to the following literature[172–178].

Electron paramagnetic resonance (EPR) or electron spin resonance (ESR) spectroscopy is a physical method to detect and characterize paramagnetic species. It is the best resolved and sensitive method used on paramagnetic centers[177]. EPR can be applied on biomolecules *in vitro* in liquid[179] or in frozen solution[125,180], in membranes[181], and in cells[182,183]. Measurements are typically performed at micromolar concentrations[184], but also dilutions up to nanomolar concentrations[182,185,186] have

been reached. Since paramagnetic centers often need to be inserted into a biomolecule and since in EPR electron spins are selectively targeted, EPR has no size restriction regarding the biomolecule[187,188]. Similar to FRET, EPR can resolve conformational changes by distance measurements that can be done via continuous wave (*cw*)-EPR[189] or via pulsed EPR[177]. The distance range that can be measured with *cw*-EPR is up to 2.5 nm[190] and the distance range that can be measured with pulsed EPR is typically 1.5–8 nm[119,184]. But, upon full deuteration of the protein, also longer distances of 16 nm have been resolved with pulsed EPR [191,192]. Further, PDS EPR resolves conformational distributions by performing ensemble measurements with ångström precision. Time resolution in the milli and microsecond regime can be achieved through additional set ups, as freeze-quench techniques[193,194] and microsecond freeze-hyperquenching (MHQ)[195]. But also the nanosecond time regime is accessible through transient EPR techniques[196]. These techniques enable to give time-dependent insights, and to follow conformational changes and ligand binding events.

1.8.1 *cw*-EPR spectroscopy

The object of study in EPR spectroscopy is the electron, which has an angular momentum called spin $\vec{s}$. The length of this vector is

$$|\vec{s}| = \hbar\sqrt{s(s+1)} \tag{3}$$

, where the spin quantum number s is given by s=½ and ħ is the Planck's constant divided by 2π[197]. The spin vector has three components, s_x, s_y, and s_z. The electron is a small magnet and its magnetic moment μ_e is aligned with the spin vector but has the opposite direction due to the negative charge[177].

$$\overrightarrow{\mu_e} = -g_e \frac{e}{2m_e} \vec{s} \tag{4}$$

The magnetic moment of the electron depends on the elementary charge *e,* the resting mass of the electron m_e, and the g-factor of the free electron, being 2.0023[198]. The g-factor is similar to the chemical shift in NMR and defines the position of the absorption signal in EPR, which depends e.g. on the ligands and coordination symmetry[177]:

$$g = \frac{h\upsilon}{\mu_B B_0} = (7.144775 \times 10^{-2}) \frac{\upsilon}{B_0} \; [MHz/mT] \tag{5}$$

If a magnetic field B_0 is externally applied, the spin vector aligns with the external field either in the same direction, parallel, or in the opposite direction, antiparallel to B_0. So, the z-component of μ_e and s can be written as:

$$\mu_{e,z} = -g_e \mu_B m_s \tag{6}$$

$$s_z = m_s \hbar \tag{7}$$

Here, μ_B is the Bohr magneton[198]. This introduces the second quantum number, besides *s*, which is the magnetic quantum number m_s. Since the s_z component is known, s_x and s_y cannot be determined (Heisenberg's uncertainty principle). The magnetic quantum number m_s can take values of +½ and -½ for s=½, or, more generally:

$$m_s = -s, -s+1, \ldots, +s \tag{8}$$

$$M = 2s + 1 \tag{9}$$

M is the multiplicity, which describes the nature of a spin state, e.g. doublet state (M=2) for s=½, triplet state (M=3) for s=1[199].

The energy *E* of the magnetic moment oriented along the z-axis is given by[198]:

$$E = -\mu_{e,z} B_0 = g_e \mu_B m_s B_0 \tag{10}$$

Without an external magnetic field, both electron states (m_s=+½ and m_s= -½) are degenerate. But upon applying an external magnetic field B_0, both states are split into two states with different energies, called Zeemann splitting (Figure 1.8-1 a). The interaction between the electron magnetic moment and B_0 is called Zeemann interaction[200]. Consequently, the energies of the two spin states are:

$$E_{m_s=+\frac{1}{2}} = +\frac{1}{2} g_e \mu_B B_0 \tag{11}$$

$$E_{m_s=-\frac{1}{2}} = -\frac{1}{2} g_e \mu_B B_0 \tag{12}$$

And the energy difference between both states, which is the resonance condition, is:

$$\Delta E = g_e \mu_B B_0 = h\upsilon \tag{13}$$

In general, in an EPR sample, more spins are occupying the lower energy state, being the m_s=-½ state. There, the occupation numbers are given by the Boltzmann distribution:

$$\frac{N_{m_s=+1/2}}{N_{m_s=-1/2}} = e^{\frac{-\Delta E}{kT}} = e^{\frac{-g_e \mu_B B_0}{kT}} \tag{14}$$

N are the occupation numbers of the respective spin state, T denotes the temperature and k is the Boltzmann constant[200]. Interestingly, the ratio of the occupation numbers at T=298 K and B_0=3000 G is 0.9986. As a comparison, in NMR it is 0.99999, which means that the lower state is populated more in EPR than it is in NMR[197].

To acquire a *cw*-EPR spectrum, an external magnetic field is applied to the paramagnetic sample, the energy states are split, and a continuous microwave is additionally applied. In addition to the Zeeman effect for electrons, also the nuclear spins undergo a nuclear-Zeeman splitting. So, in analogy to the spin quantum numbers, the nuclear quantum numbers are $\vec{I}$, the nuclear spin and m_I, the corresponding magnetic quantum number. In EPR a transition occurs when just an electron is flipped, which means that the selection rules can be summed up by[201]:

$$\Delta m_s = \pm 1 \; and \; \Delta m_I = 0 \tag{15}$$

The magnetic moment of the electron spin can interact with the magnetic moment of a nearby nuclear spin I, if it is greater than zero. This interaction is referred to as the hyperfine interaction. The hyperfine coupling leads to the splitting of the EPR signal into

$$M = 2I + 1 \tag{16}$$

lines, where M is the multiplicity. For Nitroxide radicals as MTSL, the naturally occurring isotope ^{14}N has a nuclear spin I =1, consequently a multiplicity of 3 (Figure 1.8-1 d-f). This means that the EPR spectrum of a nitroxide has three lines. Additionally, MTSL has carbon atoms, where the isotope ^{13}C has a nuclear spin I=1/2, leading to an additional doublet signal. However, because of the low natural abundance of ^{13}C, this doublet is often not detected[173]. The extent of this splitting is described by the isotropic coupling constant A_{iso}, which is dependent on the magnetic moment of the nucleus and the spin density at the nucleus, hence the spin density in the s-orbital, since it is the only orbital having spin density at the nucleus. The hyperfine splitting can give information about the environment of the electron and about the different nuclei that are coupled to this electron[201].

For the detection of *cw*-EPR spectra, usually, the magnetic field is swept, and the microwave frequency is kept fixed at the resonator frequency. To increase sensitivity, a second smaller modulated magnetic field is applied to the sample, which enables lock-in detection and the detection of the first derivative of the absorption signal (Figure 1.8-1 c and f).[173]

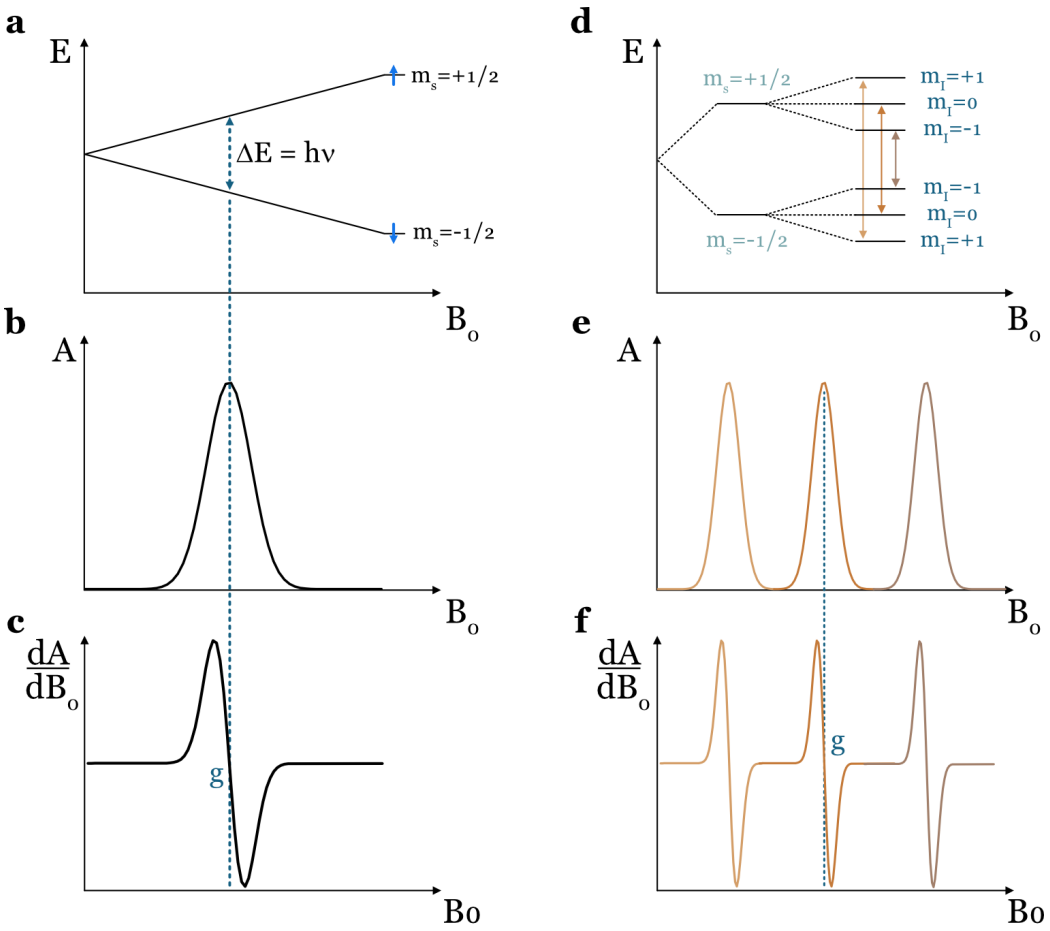


Figure 1.8-1 Schematic energy level diagrams without and with electron-Zeeman interaction and hyperfine interaction. a, For a spin ½ system and an increasing external magnetic field, the two degenerate electron spin states increasingly split into two energetically different states. **b**, When the resonance condition is fulfilled, an absorption line is generated. **c**, The spectrum that is measured is the first derivative of the absorption spectrum in **b**[197]. **d**, Schematic energy diagram of a s=½, I=1 system under consideration of the electron-Zeeman interaction (light blue) and the hyperfine interaction (dark blue)[202]. **e**, The three transitions highlighted in d are shown as an absorption curve. **f**, Analogously to c, the first derivative of e is measured as *cw*-EPR spectrum. The spectrum is centered around the g-factor.

When describing a paramagnetic center in the solid phase, the g anisotropy needs to be considered. For biological samples PELDOR time traces are often acquired in the solid phase, at cryogenic temperatures at e.g. 50 K. In a liquid solution, the radical rotates faster than the EPR scale and it has no preferred orientation to the outer magnetic field[173]. But in a frozen solution, in crystals or solids, the radicals have a fixed orientation in an outer magnetic field, resulting in EPR spectra broadening. Also the increase of viscosity leads to anisotropic behavior, as well as the attachment of a spin label to a biomolecule, due to the decreased rotating motion (Figure 1.8-2)[172,173].

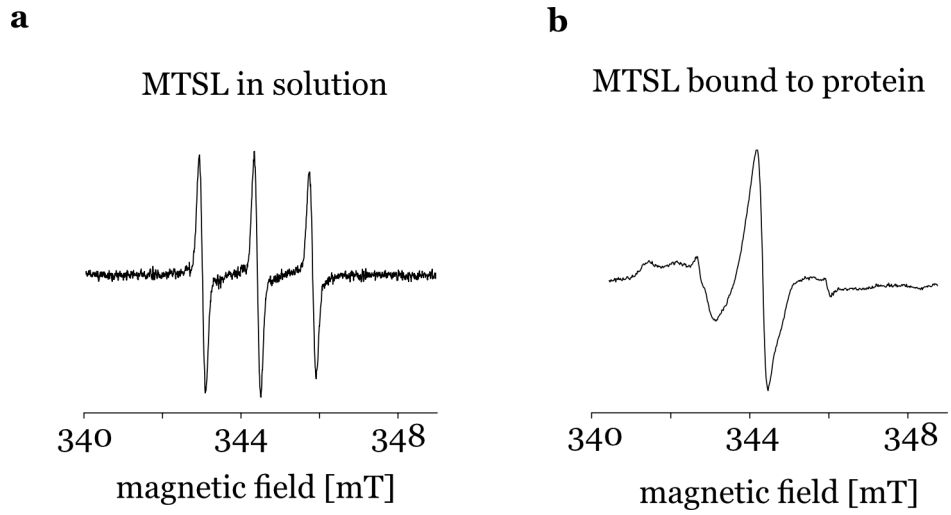


Figure 1.8-2: Exemplary X-band *cw*-EPR spectra of a, the nitroxide spin label MTSL in liquid solution at room temperature, and **b**, MTSL bound to *Lbu*Cas13a at room temperature.

These orientation dependent anisotropic contributions to the EPR spectra can origin from g anisotropy, hyperfine anisotropy, and dipolar and exchange coupling of multiple electrons[173]. If an EPR spectrum has an anisotropic behavior, the g-factor becomes a **g**-tensor with 3x3 values, where the diagonal elements correspond to the g_{xx}, g_{yy}, and g_{zz} factors. Analogously, also A_{iso} becomes an **A**-tensor with A_{xx}, A_{yy}, and A_{zz} values as diagonal[173].

For biochemical purpose, the different shapes of EPR spectra are used to detect and quantify the successful attachment of a spin label to a biomolecule, since the area of the absorption signal is proportional to the number of spins. Further it is used to study the dynamic properties of a biomolecule, e.g. the motion or flexibility changes of a certain protein site upon external changes as ligand binding, and to measure distances up to 2.5 nm[190,172].

1.8.2 Pulsed dipolar spectroscopy

Pulsed EPR techniques can be divided into several groups. The most widely used ones are the hyperfine spectroscopy, and the dipolar spectroscopy. Hyperfine spectroscopy focuses on the interaction between electrons and nuclei, while dipolar spectroscopy focuses on interactions between electrons[203]. The information obtained through hyperfine spectroscopy regards the electronic and spatial structure of a single

paramagnetic center and its environment. The paramagnetic center can be transient or stable e.g. metal centers in proteins. Pulsed dipolar spectroscopy (PDS) techniques enable the measurement of distances between commonly 1.5 nm to 10 nm[191,203]. PDS measures the dipolar coupling ω_{dd} between two paramagnetic centers, which is given by

$$\omega_{dd} = \omega_{dd}^{0}(1 - 3cos^{2}\theta) \quad (17)$$

$$\omega_{dd}^{0} = \frac{\mu_0 \mu_B^2 g_A g_B}{4\pi\hbar r_{AB}^3} \quad (18)$$

, where ω_{dd} is the dipolar splitting, μ_0 is the permeability of the vacuum, $g_A g_B$ are the g-factors of spins A and B, $\hbar$ is the Planck constant divided by 2π, r_{AB} is the interspin distance, and θ is the angle between the magnetic field B_0 and the distance vector r_{AB} (Figure 1.8-3c)[203,204]. Equation (18) shows that the dipolar coupling is inverse proportional to the interspin distance r_{AB}^3. If the dipolar interaction is strong, meaning that the distance between those spins is small, then the EPR line splitting can already be seen in *cw*-EPR[205].

In general, PDS techniques include single frequency and double frequency techniques. In single frequency techniques one microwave frequency is used to excite two different spins, whereas double frequency techniques use two different frequencies to address each spin individually[206]. The most common single frequency techniques are double quantum coherence (DQC)[207], single frequency technique for refocusing dipolar couplings (SIFTER)[208], and relaxation-induced dipolar modulation enhancement (RIDME)[209]. Whereas the most common double frequency technique is pulsed electron-electron double resonance (PELDOR), also known as double electron-electron resonance (DEER)[210].

1.8.2.1 PELDOR

In general, PELDOR is a pulsed double frequency EPR technique, that enables to separate and measure experimentally the dipolar frequency. From the dipolar frequency, the interspin distance is determined according to Equation (18). The initial three pulse PELDOR sequence was developed in the early 1980s[211,212] and was further developed by introducing the dead time-free four pulse PELDOR[210], which is the pulse sequence used in this thesis (Figure 1.8-3, right).

In the four-pulse PELDOR experiment a Hahn echo sequence $\left(\frac{\pi}{2} - \tau_1 - \pi\right)$ is applied at the observer frequency. This generates a Hahn echo of maximal intensity after the time interval of τ_1(Figure 1.8-3, right). If a $\frac{\pi}{2}$ pulse is applied, the magnetization is rotated to the x-y-plane and dephases in the x-y-plane. After the time τ_1, a π pulse is applied, which leads to inversion of the spin packets for 180°. After the same time interval $\tau_1$the spin packets refocus, and an echo is formed.

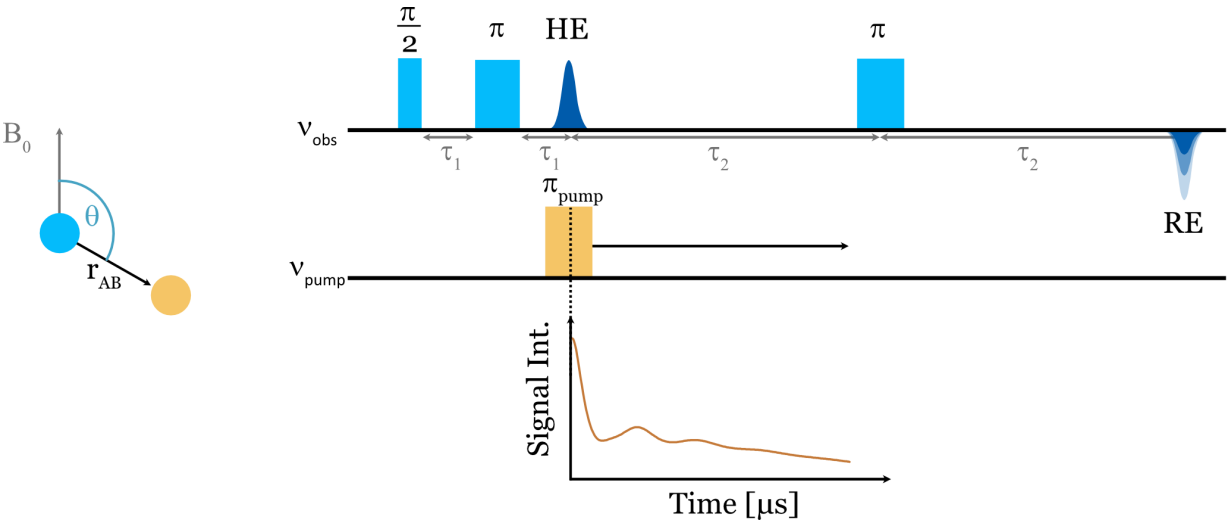


Figure 1.8-3: Four-pulse PELDOR experiment. left, Schematic representation of the distance vector between spin A (blue) and spin B (yellow), the magnetic field B_0, and the angle $\boldsymbol{\theta}$ between both spins[213]. **right,** pulse sequence[214] with the observer frequency (blue) acting on spin A and the pump frequency (yellow) acting on spin B. During the experiment, the area of the refocused echo is measured in dependence of the time between the Hahn echo and the pump pulse, generating the time trace on the bottom. This oscillates with the dipolar frequency. The detailed description of the pulse sequence is in the main text.

In the PELDOR sequence (Figure 1.8-3), after the Hahn echo and after the delay of τ_2, an additional π-Pulse is applied. This generates a refocused echo after the delay time τ_2. Between the second and third observer pulses (both π-pulses) an additional π-pulse is applied to the spin B packet at the pump frequency. This flips spin B for 180°. The integral of the refocused echo is acquired in dependence of the time, between the first Hahn echo and the pump pulse (Figure 1.8-3). This generates the time trace, which oscillates with the dipolar frequency.

Generally, a distinction can be made between physical methods that measure ensembles, e.g. PELDOR, and those measuring single molecules, e.g. FRET. In the context of investigating the dynamics of a system, single molecule measurements describe the dynamics on one single molecule. While ensemble measurements describe the dynamics of a system by looking at many molecules contemporarily, describing the

homogeneity or heterogeneity of a set of molecules. In PELDOR a set of spin pairs in a set of biomolecules are measured at the same time and the PELDOR signal is divided into an intramolecular part, originating from a spin pair in one molecule, and an intermolecular part, originating from spin pairs of two different biomolecules in vicinity. The intramolecular dipolar part of the PELDOR signal, being the echo amplitude, is given by the product of each electron-electron coupling:

$$V_{intra}(t) = \prod_i (1 - \lambda_i(1 - \cos(\omega_{AB,i}t))) \tag{19}$$

with λ_i being the fraction of inverted spins by the pump pulse, and $\omega_{AB,i}$ the dipolar coupling[214]. As a result, PELDOR spectroscopy does not measure a single distance, but an entire distribution of distances in the sample. In a biochemical context, PELDOR spectroscopy enables to measure the most probable distance between two spin labels in a biomolecule and enables to get information about the flexibility of the labelled regions. A narrow distance distribution means that the spin labelled protein regions are rigid towards each other and a wide distance distribution means that the spin labelled protein regions are flexible toward each other. The width of a distance distribution is also dependent on the spin label, since spin labels with a long linker have a higher rotational freedom than spin labels with a short linker. Hence the distance distribution of a spin label with a long linker is wider than a spin label with a short linker (Chapter 1.8.4).

The signal of the time trace does not only include the component from intramolecular spin contributions, but also from intermolecular contributions, meaning, from spins in neighboring biomolecules[214].

$$V(t) = V_{intra}(t)V_{inter}(t) \tag{20}$$

The higher the concentration of the biomolecule gets, the greater is the intermolecular component[214,215].

$$V_{inter}(t) = V_{t=0}\exp\left(-\frac{ct\lambda}{1.0027 mmol L^{-1}\mu s}\right) \tag{21}$$

Here, λ is the fraction of inverted spins, c is the spin label concentration, and the constant $1.0027 mmol L^{-1}\mu s = 9\sqrt{3}\hbar/(2\pi g_A g_B \mu_0 \mu_B^2 N_A)$, where N_A is the Avogradro constant. For samples, in which the distribution of biomolecules is homogeneous, $V_{inter}(t)$ simplifies to the stretched exponential function[216]:

$$V_{inter}(t) = \exp\left(-k_{dec}t^{D/3}\right) \quad (22)$$

D is the dimensionality or fractal dimension, being 3 for a homogeneous sample in a three-dimensional solution and 2 for systems in a two-dimensional space as lipid bilayers and k_{dec} is the decay rate constant. From this equation the local spin concentration can be determined by a semi-logarithmic plot [217]:

$$V_{inter}(t) = \exp\left(-\lambda k_0 c t^{D/3}\right) \quad (23)$$

From the slope:

$$k^{-1} = 1.0027\frac{10^3}{\lambda c} \quad (24)$$

c is the molar concentration and usually the fraction of inverted spins λ at Q-band is 0.2-0.35[217], depending on the pump pulse length. To reliably separate the intramolecular from the intermolecular signal component it is favorable to measure the PELDOR time trace such that the modulation is fully damped, however this is not always experimentally achievable[215].

Since PELDOR and FRET are very similar distance measurements they are briefly compared in the next paragraph.

1.8.3 Comparison of PELDOR and smFRET

Both techniques enable distance measurements. As stated above, PELDOR resolves distance distributions in the regime of 1.5-10 nm and up to 16 nm for deuterated proteins[191]. For smFRET the distance range is around 3-8 nm[119]. A distance error of ± 5 Å is commonly estimated for FRET, which is slightly larger than for PELDOR, being around ± 3.5 Å[119]. A major difference between both techniques is that smFRET is a single molecule measurement and PELDOR an ensemble measurement. As described before, smFRET investigates the dynamics of a system by observing one single molecule in a time-dependent manner. In contrast, PELDOR investigates the dynamics of a system by observing a greater number of molecules at the same time, as a snapshot. Hence, PELDOR is mostly measured in frozen solution at 50 K to get a snapshot of the biomolecules in solution. In principle it is also possible to measure PELDOR in aqueous solution, but this requires specific labels as trityls and large or immobilized biomolecules[218]. A smFRET experiment is usually performed at room temperature to

measure the dynamics in real time. It can be performed at cell-culture conditions and time resolution is easily achieved down to micro- and nanoseconds. For PELDOR, freeze quench techniques exist that have achieved microseconds resolution but often with demanding handling[195]. The amount of the sample needed for a PELDOR measurement is 60 µL with typical concentrations around 10-50 µM (corresponding to 0.6-3 nmol spins), but also measurements up to 10 nM[186] have been done. For smFRET, only 100-400 µL of a 15-100 pM protein solution is needed (corresponding to 0.0015-0.04 pmol)[119]. Regarding measurement times, a PELDOR measurement can take from an hour to up to days, while smFRET measurements on diffusing molecules take on average 30-60 min and on immobilized molecules minutes to hours[119]. smFRET is mostly done with two different labels, while for PELDOR just one type is needed. This makes the sample preparation and data analysis easier.

In PELDOR sometimes different cryoprotectants can lead to different distance distributions and conformations, for example due to interactions between the biomolecule and the cryoprotectant. On the other side, the fluorophore labels in smFRET are larger, compared to spin labels and have long linkers (Figure 1.8-4). These linkers permit the fluorophore to interact with the surface of the biomolecule. Sometimes fluorophores have preferred orientations and stick on the protein surface, which then can compromise the distance measurement[119].

1.8.4 Spin labelling of proteins for EPR studies

The prerequisite for EPR on biomolecules is that they need to contain paramagnetic centers. Some proteins inherently incorporate metal ions as cofactors, such as Fe^{3+} [219] or Cu^{2+}[220]. Some other proteins undergo transient paramagnetic states due to radical transfer, as the tyrosyl radical[221]. But many other proteins are diamagnetic and EPR silent[222]. For those proteins paramagnetic centers can be introduced for example through diamagnetic to paramagnetic metal ion exchange[223,224], where Mg^{2+} is replaced by Mn^{2+}. This technique is also convenient when studying oligonucleotides, in which Mg^{2+} is often bound as a cofactor and is important for folding[225,226]. Since many proteins do not bind metal ions, one of the most common technique for spin label incorporation is site-directed spin labelling (SDSL)[227]. SDSL is a combination of site-specific mutagenesis on the DNA level, in which a specific chemical moiety is

introduced into the protein, and a labelling reaction, in which the label carries a second chemical moiety orthogonal to the one on the protein[228].

In the context of SDSL the paramagnetic labels can consist of metal complexes with e.g. Cu^{2+} [229–231] or Gd^{3+} [232–234] or they can consist of stable radicals as nitroxides[235–237] or trityls[182,238–240] (Figure 1.8-4 b-d).

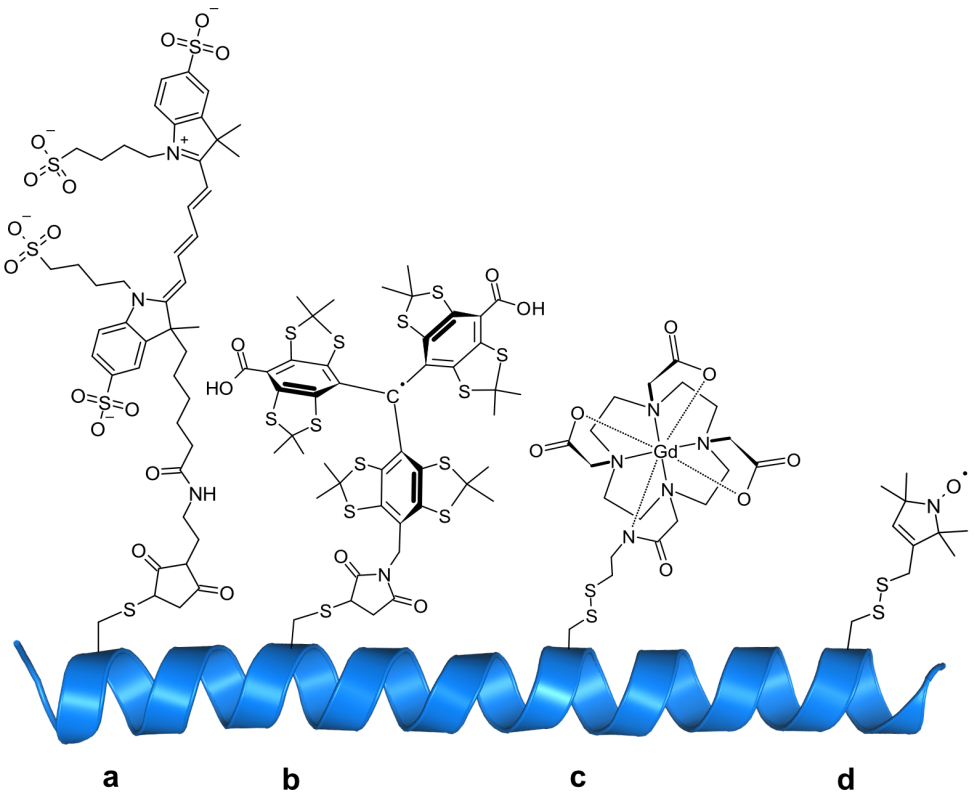


Figure 1.8-4: Chemical structures of selected labels used for smFRET and EPR. a, Alexa Fluor 647 is a common label used for smFRET[241]. **b,** The trityl spin label SLIM[182]. **c,** DOTA-Gd[242]. **d,** The nitroxide spin label MTSL. c, and d form disulfide bonds upon cysteine labelling and all four labels are attached to the protein via cysteines.

Nitroxides are stable radicals, in which the electron spin density is mostly delocalized on the N-O bond, with around 60% at the oxygen and about 40% of spin localization at the nitrogen[243]. The stability of these radical species origins from sterical shielding through e.g. methyl- or ethyl- groups. The larger the shielding, the higher the radical stability in reducing conditions[244]. Even though nitroxides are stable radicals and well suited for *in-vitro* EPR studies, they are not suited for in-cell measurements because they are quickly reduced under cellular conditions. The nitroxide core is attached to a linker and to a bioconjugation group. In the context of SDSL, the bioconjugation

groups on the label and on the protein, and therefore the labelling reaction, can vary widely (Figure 1.8-5).

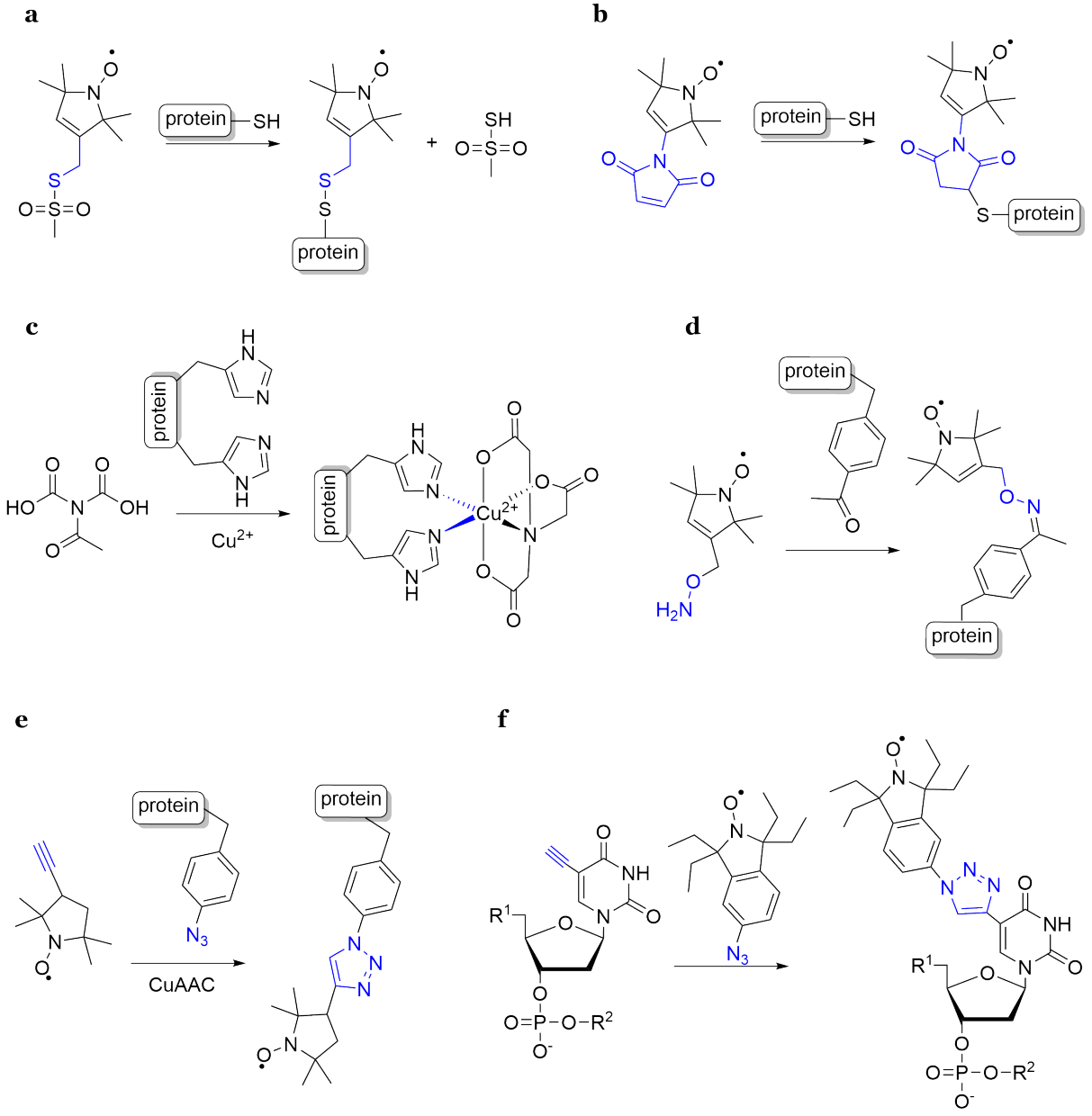


Figure 1.8-5: Selection of labelling reactions for protein and oligonucleotide spin labelling with nitroxides. The linker region between the protein and the label is highlighted in blue. **a**, Labelling reaction of a cysteine residue with MTSL in which a disulfide-bond is created[245]. **b**, Labelling reaction of a cysteine residue with the spin label maleimido-proxyl[246]. **c**, Labelling reaction of copper and nitrilotriacetic acid with the protein containing a double-histidine motif[229].**d**, Labelling reaction of the unnatural amino acid *p*-acetyl-L-phenylalanine with the nitroxide label HO-4120[247]. **e**, Labelling reaction of *p*-azido-L-phenylalanine with alkyne-proxyl[243]. **f**, Labelling reaction of 5-ethynyl-2'-deoxyuridine with a gem-diethyl isoindoline nitroxide bearing an azide moiety. R^1 and R^2 denote the continuing RNA chain[244].

The target moiety at the protein is often a cysteine residue[237,239], but also other targets, such as histidines[231] for copper labelling and unnatural amino acids like *para*-acetyl-

L-phenylalanine[247] or *para*-ethynyl-L-phenylalanine[248] are used (Figure 1.8-5). However, the spin labelling strategy that is used the most in diamagnetic proteins is cysteine labelling with MTSL (Figure 1.8-5 a). The R1 side chain, which is the reaction product, has been shown to be tolerated at various protein sites and various proteins[172]. The labelling reaction itself, as well as the work-up after the reaction are easily performed with high labelling efficiencies, and the reaction is very robust, if no reducing agents are present. Under reducing conditions, such as the presence of TCEP, DTT or reducing conditions in cell, disulfide bonds are reversible and are easily reduced[236]. Nitroxides generally have short lifetimes in the range of minutes to hours, since they are easily reduced to hydroxylamines[236] and in these cases trityls show much longer lifetimes[182]. The linkage of the maleimido-proxyl spin label with a maleimide as bioconjugation group is more stable than the disulfide bond[245] (Figure 1.8-5 b). The labelling reaction is equally easy to perform, compared to MTSL, but the pH of the labelling reaction is crucial for site specific labelling, since maleimides also react with amines. Therefore, the labelling reactions should be performed around pH7, where maleimides react 1000-times faster with sulfhydryls, than amines[249].

Even though cysteines are rare in proteins (0.5-2.26%)[250], the labelling via cysteines requires a protein without native cysteines on the protein's surface, to allow freedom in the choice of labelling sites. This can be difficult to achieve for proteins with many cysteines and for proteins including functionally relevant cysteines. Copper-NTA labelling (Figure 1.8-5 c) and unnatural amino acids (Figure 1.8-5 d, e) overcome this issue. For copper-NTA labelling, two histidines are incorporated into the protein via mutagenesis. These are placed into the amino acid sequence in such a way that they are separated by four amino acids in an α-helix or two amino acids in a β-sheet[229,231]. This ensures that the geometrical requirement for copper-complex formation is fulfilled. Another advantage of this labelling method is the reduced rotational freedom or the radical with respect to the C_α atom of the labelled amino acid residue. This is favorable, since it reduces conformational distributions seen in the PELDOR distance distributions, which could cover distance distributions that origin from protein movements.

The insertion of unnatural amino acids is more difficult from a biochemical perspective. The unnatural amino acid is encoded on the DNA basis via a stop codon. For this, usually the amber codon (UAG) is used, which is rarely present in bacterial

genes[251]. By supplying the bacteria with a plasmid containing a complementary set of tRNA and aminoacyl-tRNA synthetase, the stop codon is recognized as the codon of the unnatural amino acid[252–254]. However, since truncation products are formed, consequently the protein yields are lower, and often the expression of the protein of interest needs to be adapted. Once the protein is isolated, unnatural amino acids containing a unique functional group can be selectively labelled e.g. via copper-I catalyzed azide-alkyne cycloaddition, also known as click-chemistry (Figure 1.8-5 e)[243]. Click chemistry is also used for spin labelling of RNA (Figure 1.8-5 f) on unnatural nucleotides[184,244].

1.9 Aim of this thesis

Cas13a has applications in medical therapies and diagnostics[80,83,85]. Since it is an exclusive RNA nuclease it can also have potential important roles in the future for defeating RNA viruses.

During the functional pathway, Cas13a undergoes a multiple-step RNA-binding and RNA cleaving process with several distinct conformational states. Structures of all complexes on the functional pathway are known, but they origin from Cas13a proteins of different organisms. Further, these available structures are only snapshots of the functional pathway, but no information about the protein's dynamics as a biomolecular machinery is available. To study conformational changes and the dynamics of this protein from one organism, we chose Cas13a from *Leptotrichia buccalis* for two main reasons. Firstly, two structures of this protein have been solved that are not truncated. Secondly, the number of cysteines was important. *Lbu*Cas13a has three native cysteines (C293, C348, and C1141), whereas *Lba*Cas13a, where also two experimental structures were solved, has 12 cysteines. Since spin labelling via cysteines was chosen for this thesis, it is beneficial that the protein has little or no native cysteines.

While for *Lbu*Cas13a experimental structures of the crRNA bound and the cr- and target RNA bound complexes are known, the *apo* state and the pre-crRNA bound state are unknown. Consequently, a whole picture of conformational changes that this protein undergoes is missing and mechanistic insights are not fully understood, such as how the *apo* protein recognizes the pre-crRNA and crRNA. Further, it is unknown if there are potential intermediate states between the *apo*, pre-crRNA bound, crRNA bound, and cr- and target RNA bound states. Intermediate states could also be formed during the cleavage of pre-crRNA and target RNA. Additionally, it is unknown if the ternary complex dimerizes in solution, as seen in the asymmetric unit of the crystal structure of *Lbu*Cas13a, or if this phenomenon is due to crystal packing effects.

This is what is investigated in this thesis, mainly by means of PELDOR spectroscopy. For this, in a first part, the wildtype protein, as well as the RNA sequences need to be produced and labelled in an efficient and reproducible way. So, the first part of the thesis will focus on the development of basic biochemical techniques for our laboratory, such as a robust protein expression and protein purification protocol and

the generation of RNA strands via *in vitro* transcription. To guarantee functional integrity of all subsequently used Cas13a variants, the pre-crRNA and target RNA cleavage assays as functionality assays were optimized. The second part deals with the preparation and optimization of Cas13a for PELDOR spectroscopy measurements. For this, a labelling strategy was developed, which is finally based on a generated cysteine-free Cas13a construct. The last part will focus on conformational changes Cas13a undergoes from the *apo* to the cr- and target RNA bound state and on mechanistic insights obtained mainly by PELDOR spectroscopy.

Summing up, the aim of the thesis is to understand how Cas13a binds, and cleaves RNA, to investigate possible intermediate states, and to resolve conformational changes that enable the proteins' function.

Chapter 2

Biochemical preparation of the *Lbu*Cas13a system

This Chapter describes the first steps to establish the *Lbu*Cas13a system in our laboratory. The protein expression and purification, the *in vitro* transcription of pre-crRNA, and the pre-crRNA and target RNA cleavage assays will be covered here. The expression and purification of *Lbu*Cas13a as well as pre-crRNA and target RNA cleavage assays were done by others before[75,87]. However, the expression and purification of *Lbu*Cas13a, as described there, did not work in our laboratory. Also, the pre-crRNA and target RNA cleavage assays were optimized for our laboratory. Regarding the *in vitro* transcription of RNA, an existing protocol in our laboratory was used and optimized for this system. The optimizations of these procedures will be discussed in this Chapter.

2.1 Expression and purification of wildtype *Lbu*Cas13a

2.1.1 First expression and purification tests using Rosetta (DE3) cells

The following Chapter describes how a robust and reproducible expression and purification protocol for wt *Lbu*Cas13a was established.

The plasmid, containing the *E. coli* codon optimized *Lbu*Cas13a gene was purchased from Addgene (Materials Section 6.5) and contains an ampicillin resistance, a T7 promoter, and the downstream fusion gene of *Lbu*Cas13a. *Lbu*Cas13a is expressed as a fusion protein containing an N-terminal hexa-His tag, followed by a maltose binding protein (MBP), and a tobacco etch virus (TEV) protease cleavage site. The hexa-His tag is used for protein purification via Ni^{2+} affinity chromatography (Ni-AC)[255], the MBP tag is often used to increase protein solubility[256], and the TEV cleavage site[257] is used during protein purification to cleave off the hexa-His and MBP tags, leaving only the protein of interest.

The plasmid was transformed into *E. coli* Rosetta (DE3), since these cells were also used by East-Seletsky et al.[75]. The first goal was to determine the best expression conditions in our laboratory. For this, expression tests were performed in small volumes of 25 mL and the expression temperature, expression time, and the inductor (Isopropyl-β-D-1-thiogalactopyranoside, IPTG) concentration were varied. The cells

were grown in LB medium and SDS-PAGE samples were collected for protein detection through an SDS-PAGE and a western blot, which is shown in Figure 2.1-1.

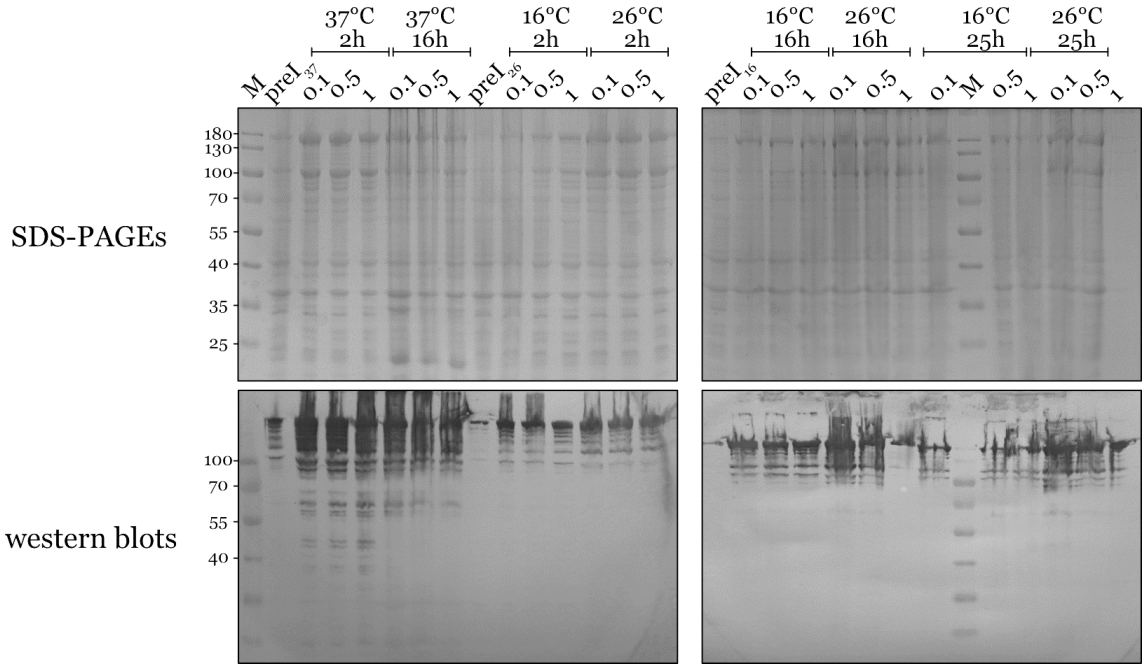


Figure 2.1-1: Expression tests of wt *Lbu*Cas13a in *E. coli Rosetta (DE3)*. Two 10% SDS-PAGEs are shown on top, with the same samples after western blotting at the bottom. Three expression temperatures were tested, 16°C, 26°C, and 37°C. In addition, also three IPTG concentration were tested, 0.1 mM, 0.5 mM, and 1 mM. M denotes the marker, preI is the abbreviation for the samples taken before induction agent was added. Western blots were prepared with an anti-His antibody.

The molecular weight of *Lbu*Cas13a including tags is 183 kDa, which fits to the band at around 180 kDa. This protein seems to be overexpressed, upon addition of inductor and it can be clearly seen in the western blot with an anti-His antibody, that this band is composed of a His-tagged protein. Already the samples taken before induction show a small protein band, indicating a weak leaky expression, which is literature known[258] for T7 promoters. The most prominent target protein bands in the western blots and in the SDS-PAGEs are those, that are induced at 37°C for 2h and for 16h. However, also bands located below the expected protein band in the western blots increase. Since these are His-tagged proteins, it is very likely that these are truncated *Lbu*Cas13a proteins, that could have been degraded by proteases. Further, for the samples taken at 37°C, also the sample loading pockets are stained. This could be an indication that a substantial amount of the protein is insoluble e.g. because of misfolding. The degree of pocket staining is decreased when lowering the temperature to 26°C with incubation time of 16 h and it is further decreased at 16°C and 16 h incubation time. These observations fit to what is reported in literature[259]. By comparing the western blot samples taken at 16°C and 26°C after 16 h and 25 h, the protein amount decreases from

16 h to 25 h. In summary, it seems that an induction at 16°C and for 16 h is the most promising expression condition.

The first test expressions were performed in LB medium. To see if more nutrients in the medium resulted in a higher protein yield and to test if the expression is scalable to 1 L cultures, the expressions were repeated with 1 L LB and 1 L 2YT media. These expressions were done at 37°C and at 16°C to validate the results seen in the SDS-PAGEs and western blots in Figure 2.1-1. SDS-PAGE samples were collected, and the resulting SDS-PAGE is shown in Figure 2.1-2 a.

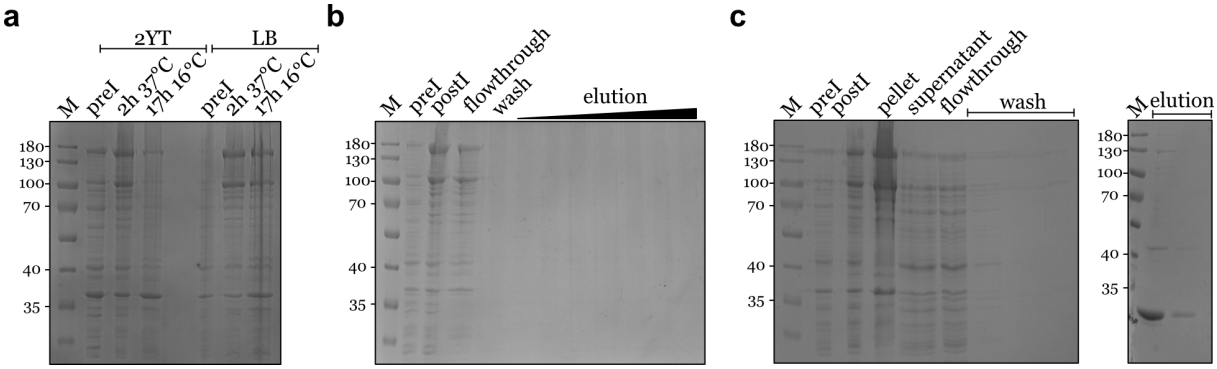


Figure 2.1-2: Expression tests and purification tests of wildtype *Lbu*Cas13a. All SDS-PAGE gels shown are 10% polyacrylamide gels. **a,** Expression tests were performed in *E. coli* Rosetta (DE3) cells with different expression media. M denotes the marker and preI denotes the samples taken before induction. **b,** SDS-PAGE of the purification via benchtop Ni-AC beads. Pre and post induction samples (preI, postI) are showed, as well as the nickel bead flowthrough, the beads wash, and the elution steps. The elution was done three times with 10 mM, two times with 100 mM, once with 150 mM, once with 250 mM and once with 1M imidazole in the binding buffer. **c,** SDS-PAGEs of the purification with a benchtop amylose beads column.

The expression levels of *Lbu*Cas13a in nutrient rich 2YT and LB media at 37°C are very similar. By comparing the expression levels at 16°C, the protein band at about 180 kDa seems to be slightly larger in the sample taken from the expression in LB medium. These tests were repeated several times (data not shown) and pellets from both, 2YT and LB cultures at 16°C were used for first purification tests. To determine if protein can be isolated from one-liter cultures with different purification methods, one LB culture was purified via Ni-AC and one 2YT culture was purified via amylose bead purification. The results of both purifications are shown in Figure 2.1-2 b and c, respectively. The purification via Ni-AC on a benchtop column (Figure 2.1-2 b) shows, that the protein was successfully expressed and that the protein band at 180 kDa is present in the flowthrough but not in the elution. It seems, as if the protein did not bind to the Ni-column. However, the His-tag is present and in principle accessible for binding, which was seen in the western blot showed before (Figure 2.1-1). In addition,

the entire Cas13a gene with fusion tags was sequenced and showed an intact and correct DNA sequence (data not shown). So, one can assume that the protein is intact. Unfortunately, also the purification via a benchtop column containing amylose beads was not successful (Figure 2.1-2 c). Here the protein is visible in the soluble cell lysate (supernatant) as well as in the cell pellet, indicating non-soluble and maybe misfolded protein that could come from inclusion bodies. The protein band in the flowthrough seems to be equally large as the one in the supernatant and no band is visible in the wash and in the elution steps. This means, that also here the protein did not bind to the beads. The band in the elution sample of the amylose purification seen below 35 kDa is the TEV protease (27 kDa[260]), which was used to cleave Cas13a from the tags and from the beads (Methods Section 7.2.4).

Next, a purification with a HisTrap FF 5 mL column on a chromatography system was tested by using a cell pellet from a 3L LB culture. Both, the Ni-beads on a benchtop column and the column material of a HisTrap FF column consist of agarose. So, the material is the same, but the main difference is the pressure and flow rate that is used for sample loading and elution, which is higher on the HisTrap FF column.

The pellet was lysed as described in the Methods Chapter 7.2.4. The cell lysate was directly loaded onto the HisTrap column after centrifugation and washed with binding buffer containing 5 mM imidazole. The elution was done stepwise, by 2 column volumes (CV) of binding buffer containing 10 mM imidazole and 2 CV binding buffer supplemented with 50 mM imidazole. Then, a linear gradient was set from 50 mM to 1 M imidazole with a total volume of 60 mL and SDS-PAGE samples were collected (Figure 2.1-3 a). A band at around 180 kDa is seen in the SDS-PAGE samples from the linear gradient of the elution, which means that the protein did bind to the HisTrap column and that it can be eluted. In addition, the removal of various other proteins is achieved through the first elution step with 50 mM imidazole. Compared to the purifications shown above, the protein has bound to the HisTrap column for the first time and led to a significant improvement in protein isolation. After concentration and TEV cleavage at room temperature for 90 min, a reverse Ni-AC was performed (Figure 2.1-3 b). After sample loading, the column is washed with 1 CV of binding buffer and everything that was bound to the column was eluted with binding buffer supplemented with 1 M imidazole. Since the His- and MBP tag was cleaved, the protein should weight 139 kDa and should not bind to the Ni-column anymore. Thus, it should be present in

the flowthrough, which is what is also seen in the SDS-PAGE gel (Figure 2.1-3 b). Finally, the protein flowthrough was concentrated (Figure 2.1-3 c).

Upscaling of the protein purification to a cell lysate obtained from 6 L LB main culture worked as the purification of 3 L LB main culture. After the reverse HisTrap, the protein containing fractions were pooled, concentrated, and loaded onto a gel filtration (GF, size exclusion chromatography, SEC) column (Superdex 200 10/300). The SDS-PAGE samples of the SEC fractions are shown in Figure 2.1-3 d. The up scaling of the main cultures from 3 L to 6 L led to an increased protein yield, which was expected. However, the GF fractions show that also the impurities increased. Since protein degradation products were seen in western blots of cell culture samples (Figure 2.1-1), protease inhibitor tablets were added in another protein expression and purification trial. The result of this expression and purification is shown in Figure 2.1-3 e.

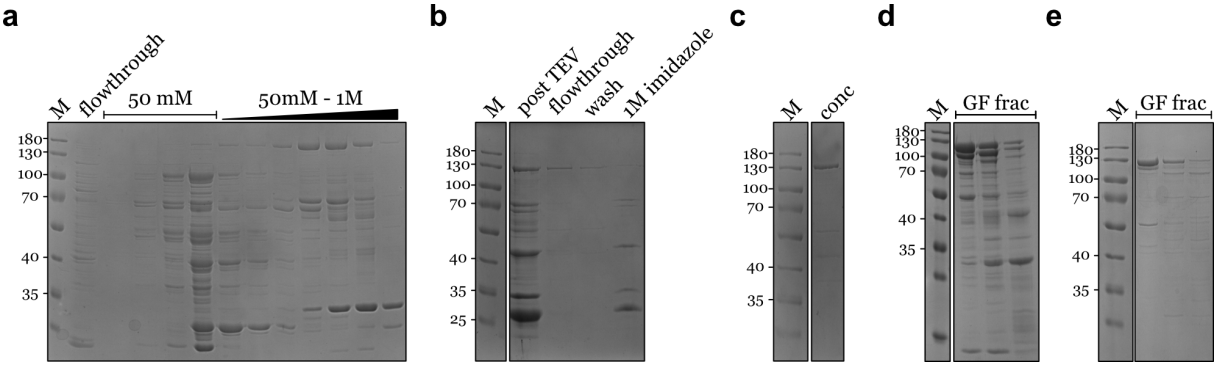


Figure 2.1-3: Protein purification optimizations using the HisTrap column on a chromatography system. 10% SDS-PAGE gels of different purifications are shown. **a**, SDS-PAGE of samples taken from the protein purification with a HisTrap column (3L main culture). **b**, SDS-PAGE of samples taken from the reverse Ni-AC on a chromatography system. This is the same purification as in a. **c**, SDS-PAGE of the concentrated protein is showed, from the same purification as in a and b. **d**, SDS-PAGE of a purification in which cells from 6 1L main cultures are used. The protein as well as the impurities increased, showed by SDS-PAGE samples collected from gel filtration. **e**, SDS-PAGE of a different purification. For this purification, protease inhibitor tablets were used during cell lysis and cells collected from 5 x 1 L main cultures were used.

The expression and purification protocols were not modified, despite only 5 L of main culture was used and that protease inhibitor tablets were added to the cell resuspension, prior to cell lysis. The SDS-PAGE samples of the SEC fractions show that the impurities diminished. It is noted, that since the amount of loaded protein is not the same in Figure 2.1-3 d and e, only the relative intensities of the bands originating from impurities and the ones originating from Cas13a, at around 130 kDa, in one SDS-PAGE gel should be compared. So, the usage of protease inhibitor tablets improved protein isolation quality.

2.1.2 Expression and purification tests using BL21 AI cells

During several protein expression trials, which are not shown here, a non-reproducible cell behavior was noticed. The *E. coli* Rosetta (DE3) cells usually take 2 to 3 h to grow to an optical density (OD_{600}) of 0.8 - 1. However, in these trials, the cells sometimes took 5 to 6 hours to grow at 37°C and 180 rpm. In addition, the expression quality of the wt *Lbu*Cas13a varied. In some cases, no protein was expressed, and in other cases the protein was expressed as shown in the SDS-PAGE gels above. Since *Lbu*Cas13a has a collateral cleavage activity, maybe the leaky expression identified in the western blot caused cell toxicity. Another explanation could be the size of *Lbu*Cas13a. It was shown for the 158 kDa recombinant Cas9 protein that the protein expression levels in *E. coli* Rosetta (DE3) cells are lower than the ones in BL21 cells. This seems to be partly related to the low level of Cas9 mRNA transcription due to the long Cas9 mRNA sequence, which effects transcription rates[261]. This could be also valid for Cas13a, being expressed as a 183 kDa protein. Consequently, expression tests have been performed with various *E. coli* cell types available in our laboratory, including Rosetta (DE3), XL1 Blue, C43 (DE1), MC10 61, and Bl21 AI (Figure 2.1-4 a). BL21 AI cells were recommended from Dr. Christophe Rouillon (Institut Pasteur, Paris, personal communication) since he worked on another Cas13a construct and had positive results. BL21 AI is an L-arabinose induced cell line specifically developed to minimize leaky expression. L-arabinose induces production of T7 RNA polymerase, and its amount can be adjusted to control mRNA transcription. This can be useful when expressing toxic proteins. In addition, some proteases are deleted, compared to BL21 (DE3), which could be beneficial for decreasing the amount of partially degraded Cas13a[262]. The expression tests were done in 50 mL main cultures supplemented with ampicillin and additionally chloramphenicol for Rosetta (DE3). Cells were grown at 37°C and induced during log-phase at 16°C with 0.5 mM IPTG, except for BL21 AI cells that are additionally induced with 0.5% L-arabinose.

The different *E. coli* strains showed a different behavior already before induction of protein expression (Figure 2.1-4 a). The strains MC10 61, C43, and BL21 AI showed the least leaky expression of the protein. XL1 blue expresses the protein to a higher degree and Rosetta 1 and Rosetta 2, and BL21 show even more protein expression. Rosetta 1 and Rosetta 2 are the same Rosetta cells, but from different batches. Rosetta 1 cells

were taken from a glycerol stock from the -80°C freezer. For Rosetta 2, the protein gene containing plasmid was freshly transformed into a new batch of Rosetta cells, plated onto an agar plate and one colony was picked for this test expression. Interestingly, Rosetta 1, Rosetta 2, and XL1 blue reached log-phase after 2.5 h, 3 h and 3 h, respectively, while all other cell strains were 30 min faster. It is noted here, that obviously different cell amounts are pipetted into the same main culture volume, since the precultures had different optical densities. But the preculture of C43, and both Rosetta precultures had the same optical density (2.9) when added to the main culture. So, over all it seems that Rosetta cells need a longer time to reach the log-phase. The SDS-PAGE samples taken after 5 h post-induction (Figure 2.1-4 a, left) already show that Rosetta 1, Rosetta 2, BL21 AI, and BL21 cell strains produce a higher amount of *Lbu*Cas13a than the other cell strains. This stays the same when looking at the SDS-PAGE samples taken after 16 h post induction (Figure 2.1-4 a, right). It seems that the *Lbu*Cas13a overexpression is slightly larger in BL21 AI cells than in all other cell strains.

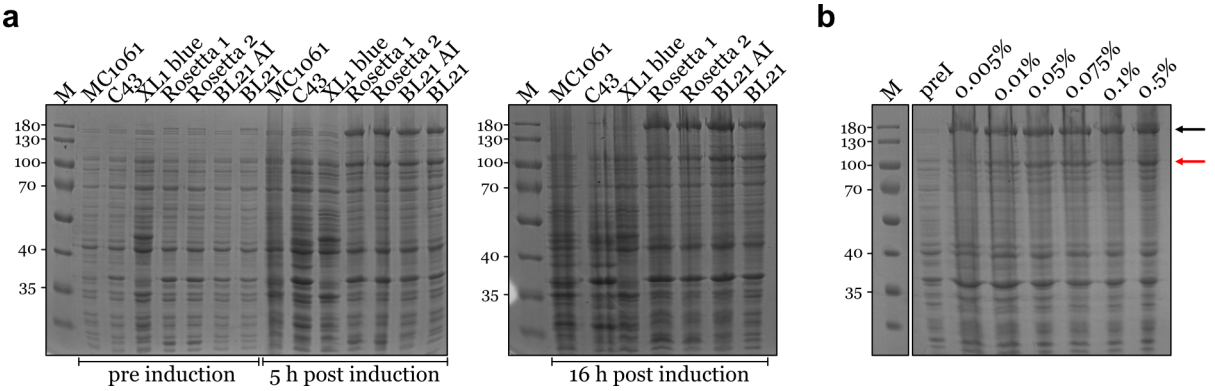


Figure 2.1-4: Test expressions of *Lbu*Cas13a by variation of cell strains and inductor concentration. 10% SDS-PAGE gels are shown in this figure. **a**, SDS-PAGE gels of protein expression with different cell strains are shown before inductor addition, after 5 h and after 16 h that the inductor was added. **b**, SDS-PAGE gel of test expressions performed with BL21 AI cells and with different L-arabinose amounts, depicted in %. The IPTG concentration was held constant at 0.5 mM. The black arrow highlights the protein band of *Lbu*Cas13a, and the red arrow highlights the band that is supposed to be degradation product.

In a separate experiment, the concentration of L-arabinose as inductor for BL21 AI cells was tested and the results are shown in Figure 2.1-4 b. The expression was performed under the same conditions as the one in Figure 2.1-4 a. The L-arabinose concentration was varied by retaining an IPTG concentration of 0.5 mM. Already 0.005% L-arabinose shows successful expression of the protein (Figure 2.1-4 b, black arrow). This band increases when 0.01% L-arabinose was added and increases slightly

more upon addition of 0.05% L-arabinose. From 0.05% to 0.5% L-arabinose the amount of protein is very similar. Contemporarily to the increase of the desired protein band, also another protein band increases in intensity, highlighted by a red arrow. This could be degradation product of the desired protein. Therefore, 0.05% L-arabinose was used for further expressions as it provided a trade-off between the largest amount of *Lbu*Cas13a and the lowest amount of degraded protein.

The protein expression in BL21 AI cells was scaled up to 6 x 1 L main cultures and the protein was purified as described before via Ni-AC, reverse Ni-AC and size exclusion chromatography. The only difference is that in addition to protease inhibitor tablets, also 250 µL of a 100 mM PMSF solution (PMSF dissolved in isopropanol) was added to the cell resuspension, as an additional protease inhibitor. The size exclusion chromatogram and the corresponding SDS-PAGE samples of the size exclusion are shown in Figure 2.1-5. Two monodispersed peaks can be seen in the size exclusion chromatogram at an elution volume of around 50 mL and 72 mL. But absorption of small proteins and degradation products are seen after 72 mL. The corresponding SDS-PAGE gel shows that the yellow region consists of various proteins, the brown region also consists of other proteins but includes *Lbu*Cas13a, the blue region consists of almost only *Lbu*Cas13a, and the light blue region seems to be composed of protein impurities. The main chromatogram peak at around 72 mL was concentrated and stored at -80°C. In general, the protein purification with BL21 AI as expression cells improved the quality of the final protein as well as the yield, being 4 mg.

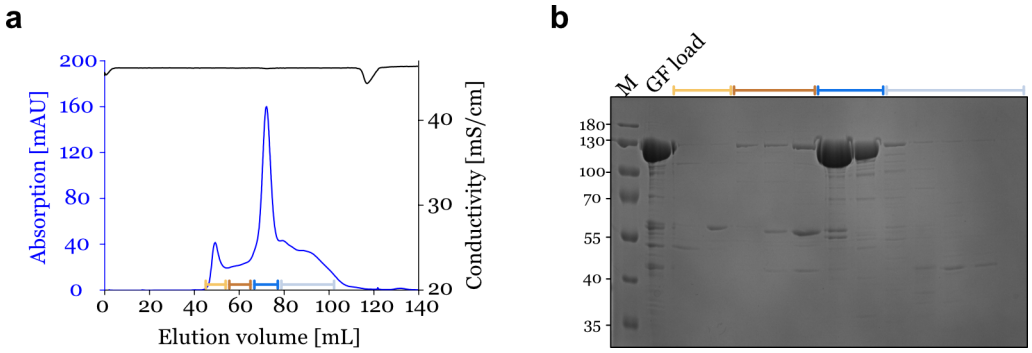


Figure 2.1-5: Purification results of *Lbu*Cas13a in BL21 AI cells, and with PMSF. a, Chromatogram of the size exclusion (HiLoad 200 16/600). The absorption at 280 nm is measured and the colored bars correspond to the bars in b. **b,** SDS-PAGE gel of samples from the size exclusion chromatography. GF load denotes the sample that has been taken from the protein that was loaded onto the size exclusion column.

2.1.3 Protein identification via LC-MS

To prove that the SDS-PAGE band at around 130 kDa is *Lbu*Cas13a, a protein mass identification was done by the Mass-Spectrometry facility of the University of Bonn. For this, an SDS-PAGE was performed of a sample from the purification shown in Figure 2.1-5. The protein band at 130 kDa was cut and given to the facility. The protein is extracted from the gel by the department and after trypsin digestion, the protein fragments are analyzed by liquid chromatography-mass spectrometry (LC-MS). The protein fragments that were identified are highlighted in green with respect to the protein sequence (Appendix, Figure 8.1-1). The result shows a coverage of 44% and an abundancy of 1.9×10^9 for Cas13a and by this, confirming the presence of the protein. A positive feature is that the identified sequence fragments are equally distributed through the protein sequence, which indicates that the purified protein is intact and that it is likely that no protein parts are missing.

2.1.4 Improvement of protein purification using a heparin column

One of the goals is, to purify an intact protein that binds the pre-crRNA and crRNA and cleaves the pre-crRNA and target RNA. To ensure that these processes occur, it is very important that the protein is not bound to cellular RNAs. Therefore, a heparin column was added to the protein purification protocol, after the reverse Ni-AC. Heparin is an RNA competitor[263] on which RNA and DNA binding proteins can bind to. Through heparin binding, proteins should lose cellular oligonucleotides that can be bound to them. The chromatograms and SDS-PAGE gels of the purification through heparin and size exclusion are shown in Figure 2.1-6. During heparin-AC, an unexpected peak is seen at 70 mL elution volume, when the run switches from the column washing step, to the elution step with fractionation. This can happen in some chromatography runs and it is likely to be caused by a failure of the UV-detection system, or by air in the chromatography system (see in Figure 2.1-7 b for an example of a typical chromatogram of a heparin run). Despite the unusual chromatogram shape, a small peak around an elution volume of 85 mL is seen. The SDS-PAGE gel shows that this peak is *Lbu*Cas13a. Most protein impurities remained in the flowthrough and in the

wash step, after sample loading. In contrast to these impurities, *Lbu*Cas13a binds to heparin, hence to the column, and is eluted upon increasing concentration of NaCl in the elution buffer (high salt buffer). Therefore, in addition to removing cellular RNA, the use of a heparin column also greatly improved the purity of the protein solution. At this stage, the Cas13a fraction is almost free of protein impurities and after size exclusion (Figure 2.1-6 b) no proteins are detectable in the SDS-PAGE gel, besides *Lbu*Cas13a. In addition of removing impurities it was tested whether oligonucleotides are bound to *Lbu*Cas13a during expression and purification and whether these oligonucleotides are removed by the heparin column. For this the flowthrough and the wash of the heparin column were both separately concentrated in a vacuum concentrator. Additionally, a proteinase K digestion was done with 85 µg of an *Lbu*Cas13a batch, purified with heparin column, and by incubating 85 µg of an *Lbu*Cas13a batch purified without heparin column. This should digest *Lbu*Cas13a and other proteins and release possible bound oligonucleotides. The samples were then loaded onto an agarose gel, the result is shown in Figure 2.1-6 c.

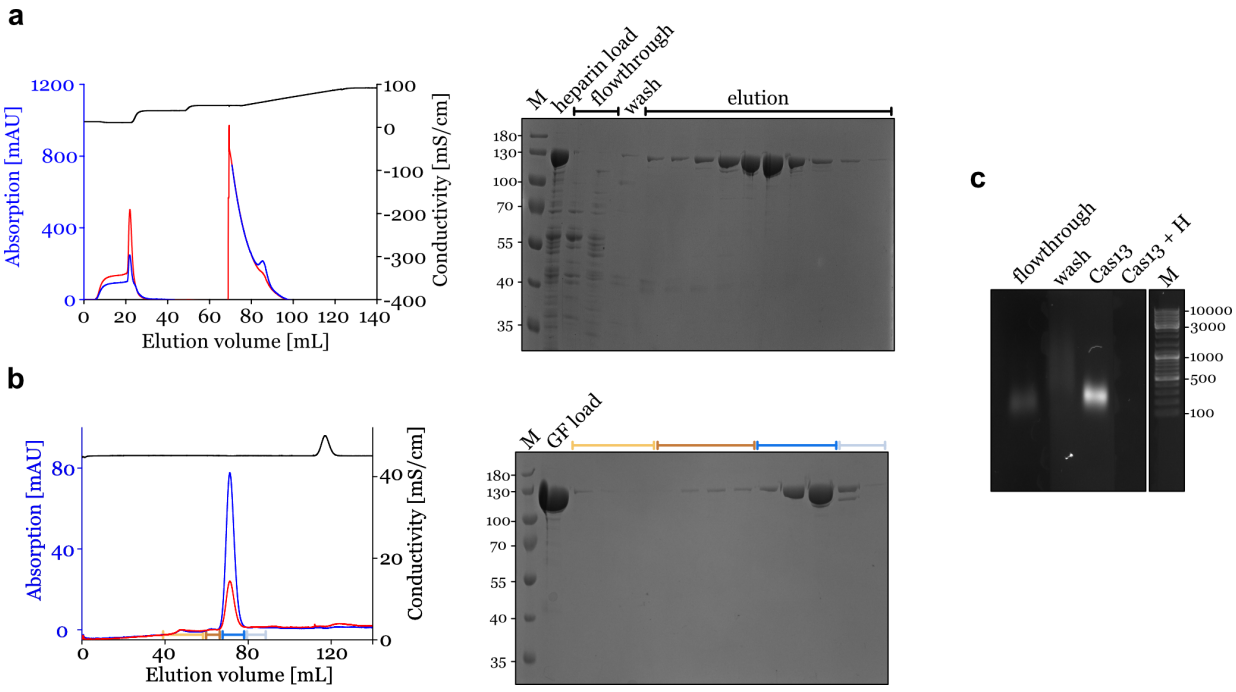


Figure 2.1-6: Protein purification of wt *Lbu*Cas13a with a heparin-AC. a, Chromatogram and SDS-PAGE gel of the heparin column run. The blue curves in the chromatograms show the absorption at 280 nm and the red curves the absorption at 254 nm. **b,** Chromatogram and SDS-PAGE gel of the size exclusion chromatography. The colored bars in the chromatogram highlight the eluted fractions that correspond to the samples that are highlighted by the same-colored bars in the SDS-PAGE gel. **c,** Result of an agarose gel electrophoresis, showing the difference between *Lbu*Cas13a purified with a heparin column (H) and without a heparin column. The DNA ladder in base pairs is showed for reference.

Oligonucleotides of varying sizes were detected as bands in the agarose gel samples of the flowthrough and in the wash of the heparin column run. Interestingly, the sample containing proteinase K digested *Lbu*Cas13a, which was not purified with the heparin column, showed a bright band. This band corresponds to oligonucleotides with similar sizes to the oligonucleotides in the flowthrough of the heparin column. In contrast, no oligonucleotide bands could be detected in the sample containing proteinase K digested *Lbu*Cas13a, previously purified with the heparin column. This indicates that heparin, as RNA competitor, has bound to the RNA binding pocket of *Lbu*Cas13a and that the bound RNA has been removed. The heparin column was therefore added to the protein purification protocol.

To sum this Section up, the results of the final optimized expression and purification procedure of *Lbu*Cas13a will be briefly described.

The chromatograms and the corresponding SDS-PAGE gels of the optimized purification procedure are shown in Figure 2.1-7 and the detailed optimized protocol for protein expression and purification is found in the Methods Sections 7.2.5 and 7.2.6. The expression of *Lbu*Cas13a was done in 1 L main cultures and in BL21 AI cells. After those cells reached the log-phase, they were induced with L-arabinose and IPTG. After incubation at 16°C for 16 h, the cells were harvested by centrifugation and either directly resuspended in cell lysis buffer or stored at -80°C. For protein purification, the cell lysis buffer (binding buffer) was freshly supplemented with PMSF and protease inhibitor tablets. After cell lysis the mixture was centrifugated to remove cell membranes, inclusion bodies, and insoluble cellular components. The SDS-PAGE sample of the pellet (Figure 2.1-7 a) showed a protein band at around 180 kDa, which corresponds to *Lbu*Cas13a. This is an indication that maybe insoluble, aggregated protein or inclusion bodies were still formed and that maybe the expression temperature could have been lowered, to diminish this formation. However, no further expression tests have been done since the yield seemed to be sufficient, which was observed in the SDS-PAGE samples of the Ni-AC elution. There, a prominent band at 180 kDa was observed. The eluted fractions from the Ni-AC were concentrated using a centrifugal concentrator, and TEV cleavage was performed by diluting the sample with low salt buffer and incubating with TEV protease at RT for 2 h. The concentration and dilution steps are particularly important, since imidazole was used in the Ni-AC buffers. The TEV protease precipitates when exposed to high imidazole concentrations,

which could lead to incomplete cleavage of the protein tags. After TEV cleavage, the second Ni-AC was skipped. Instead, the protein mixture was directly loaded onto a heparin-AC column (Figure 2.1-7 b). The reason for skipping the reverse Ni-AC is, that the TEV protease and other proteins are directly removed by the heparin-AC as well, which is nicely seen in the flowthrough fraction in the SDS-PAGE gel. The band at around 40 kDa corresponds to the 46 kDa bis MBP- and hexa-His-tag and the band appearing between 25 and 35 kDa corresponds to the TEV protease.

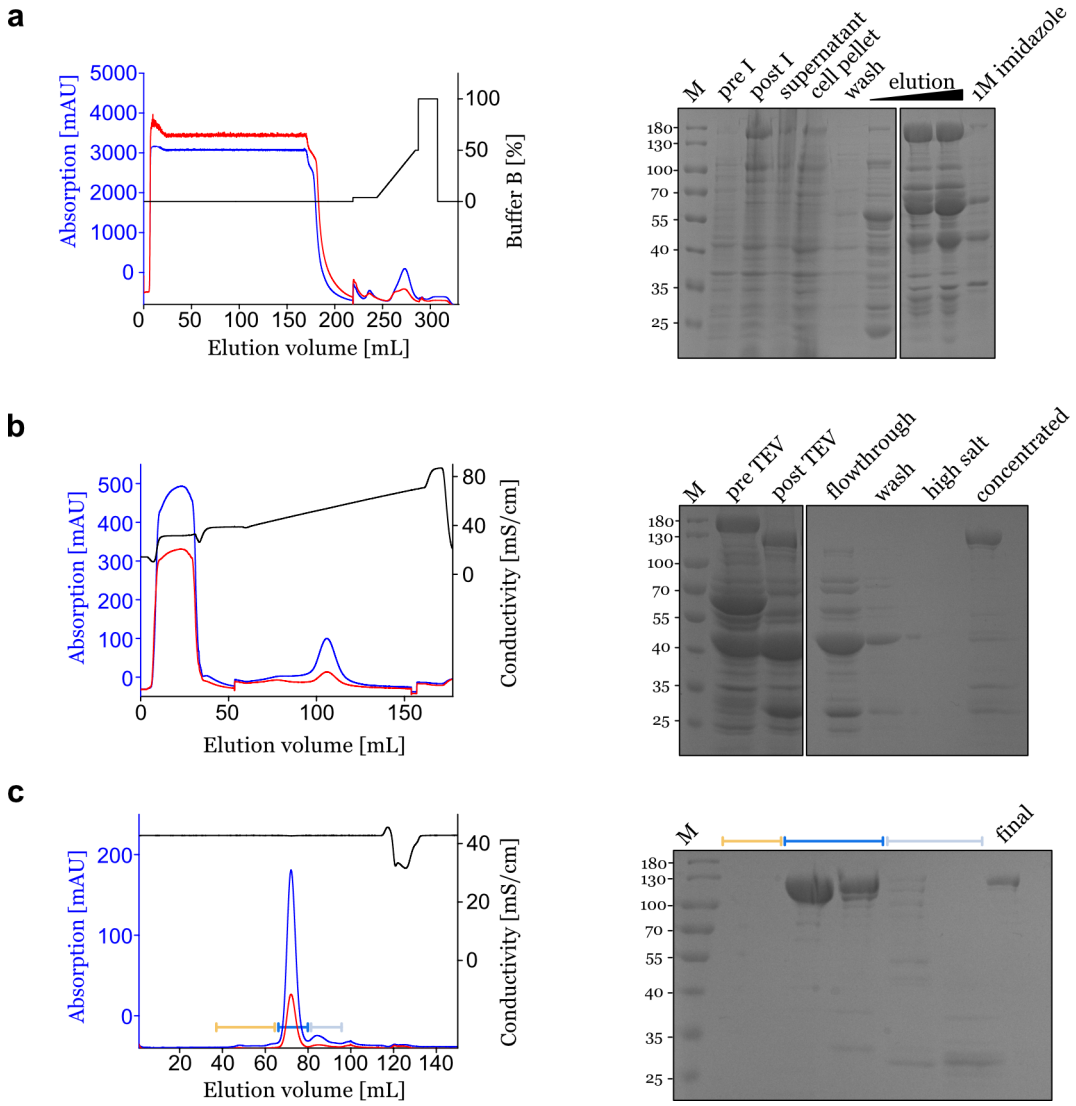


Figure 2.1-7: Optimized expression and purification protocol results for wt *Lbu*Cas13a. a, Chromatogram and SDS-PAGE gel of the Ni-AC. **b,** Chromatogram and SDS-PAGE gel of the heparin-AC. The SDS-PAGE sample denoted as concentrated, is composed of the concentrated fractions of the heparin column elution and is the sample that was loaded onto the size exclusion column. **c,** Chromatogram and SDS-PAGE gel of the size exclusion, with the HiLoad 200 16/600 column. The final concentrated protein is highlighted, and the SDS-PAGE fractions are marked with colored bars that correspond to the colored bars in the chromatogram. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm.

The eluted protein fractions of the heparin-AC that correspond to the elution volumes 70-150 mL were concentrated in a centrifugal concentrator and loaded onto the size exclusion column (HiLoad 200 16/600). The reason for using the HiLoad 200 instead of the Superdex 200, as it was done before, is the enhanced peak separation (Figure 2.1-7 c). The chromatogram peak with the highest intensity corresponds to *Lbu*Cas13a. The fractions of this peak were concentrated with a centrifugal concentrator and as it can be seen from the SDS-PAGE gel of the concentrated protein (Figure 2.1-7 c), a purified *Lbu*Cas13a was obtained. This protein was flash-frozen in liquid nitrogen and stored at -80°C for long-term storage.

2.2 Transcription of pre-crRNA for protein activity assays

Two tests need to be done with purified *Lbu*Cas13a to determine if it is active, a pre-crRNA cleavage assay and a target RNA cleavage assay. For this, the pre-crRNA, the crRNA, and the target RNA is needed. Instead of buying these RNA strands, an *in vitro* transcription procedure was performed. To test, if the yields of an *in vitro* transcription (IVT) are sufficient and if in general this procedure can be done in our laboratories, the pre-crRNA was chosen as a test-product to be transcribed.

First, for IVT, a DNA template is needed. The pre-crRNA sequence was purchased from Eurofins (as a DNA, hence pre-crDNA), embedded in a pEX-A128 vector with ampicillin resistance. This plasmid was constructed such, to have a T7 RNA polymerase promoter at the 5' end and an XbaI restriction site at the 3' end of the pre-crDNA sequence. The XbaI restriction site is not necessarily needed, since large transcription products are removed by urea-PAGE. A transcription terminator sequence was not included, since T7 transcription can also be terminated via polymerase “run off”, where the T7 RNA polymerase falls off when arriving at the end of the DNA sequence[264]. Generally, the IVT procedure is divided into three steps, amplification, transcription, and purification via PAGE. To start with the amplification, a standard PCR was performed by using DNA primers that bind at both ends of the desired pre-crDNA sequence. To remove the polymerase, an ethanol precipitation was performed. In the second step, an IVT was conducted by dissolving the DNA pellet and pipetting the transcription mixture composed of a TRIS buffer, NTPs, $MgCl_2$, DTT, T7 RNA polymerase, RNAsin, and ddH_2O. After incubation of the transcription reaction

overnight, at 37°C the purification step was started. For this, the RNA was precipitated with ethanol and an urea PAGE was run with samples containing the transcription products. The urea gel was covered with transparent foil and the successfully generated transcription product was visualized with an UV-lamp (Figure 2.2-1 a). The bands, containing the transcription product were cut out with a sterile scalpel, which is highlighted as red squares in Figure 2.2-1 a. The prominent bands at the bottom of the picture are the residual NTPs that have not been used by the T7 RNA polymerase. The blue mark on the right is a loading dye, which was used as an indicator of how far the urea gel has run. The purification was continued by electroelution, performed in the Laboratories of Prof. G. Mayer (LIMES Institute, University of Bonn) and terminated by a last ethanol precipitation. The pre-crRNA was dissolved in ddH_2O and stored at -20°C. The amount of RNA that was transcribed from 150 µL PCR reactions was 155.9 pmol.

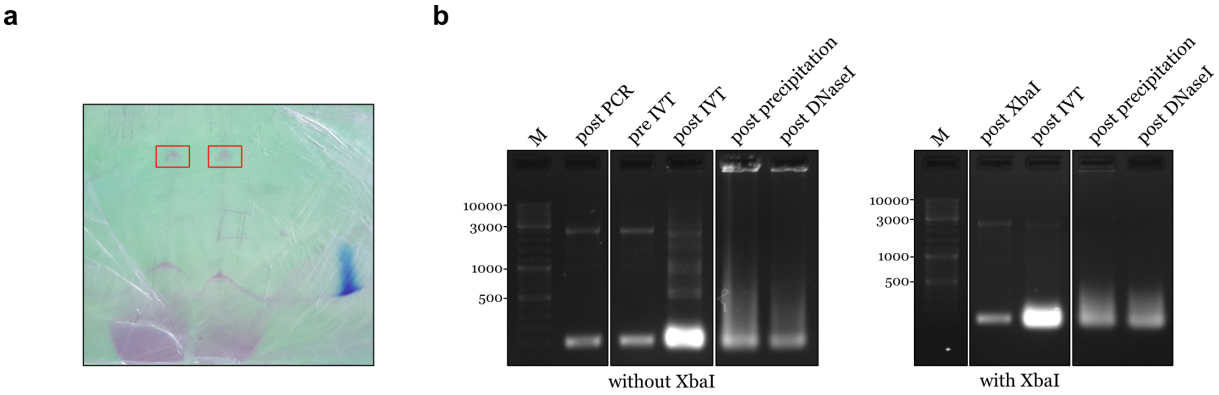


Figure 2.2-1: In vitro transcription of pre-crRNA and analysis of XbaI restriction. a, Image of the urea PAGE after in vitro transcription (IVT). The pre-crRNA bands that were cut out are highlighted with red squares. The prominent bands on the bottom derive from NTPs. **b,** Agarose gels of transcription reactions without (left) and with (right) XbaI restriction.

To test, if the XbaI restriction leads to an increased yield of the transcription product, the transcription procedure was repeated in our laboratory in absence and in presence of XbaI. Three PCR reactions were performed and were split into two reaction tubes. Samples of 5 µL were collected after each step for a visualization in an agarose gel. The result is shown in Figure 2.2-1 b. The sample containing the PCR product shows two bands, one at the top, at around 3000 bp and the other at the bottom, corresponding to a short DNA sequence < 100 bp. The lower band is the short, double stranded amplified PCR product and the band on top can be attributed to the plasmid that is used as the template. The DNA in the sample without XbaI, was precipitated and the

result is seen in the next lane of the same gel (labelled with pre IVT). As expected, the plasmid and the short PCR product have been precipitated. The sample collected after XbaI restriction and DNA precipitation looks the same as the one without XbaI restriction. This makes sense, since the plasmid is linearized through XbaI and the plasmid's size is not altered. The difference between without and with XbaI becomes visible in the samples after IVT. The sample containing the IVT product that was not treated with XbaI shows, additionally to an amplified, bright band on the bottom of the gel, many reaction products with varying lengths. In contrast, the sample which was treated with XbaI has only two bands in the agarose gel, a first, light band coming from the plasmid and a bright band from amplified pre-crRNA. In addition, by comparing the transcription products, the pre-crRNA band from the sample treated with XbaI is slightly larger than the one that was not treated with XbaI. The intensity of the pre-crRNA transcription product decreases after precipitation even though the sample volume is decreased, which is an indication that the precipitation procedure was not quantitative. The sample that was collected after removal of the DNA templates through DNase I is shown in the last lane of the agarose gel. The band of the pre-crRNA transcription product with XbaI restriction seems to contain more homogeneously sized RNA, than the product which was not treated with XbaI. The agarose band of the latter is more blurred towards larger sizes, than the other transcription product. The overall amount of RNA increases slightly with incubation of XbaI. Maybe this could result from the fact that the T7 RNA polymerase takes longer to transcribe long RNA sequences, present in the transcription reaction without XbaI incubation.

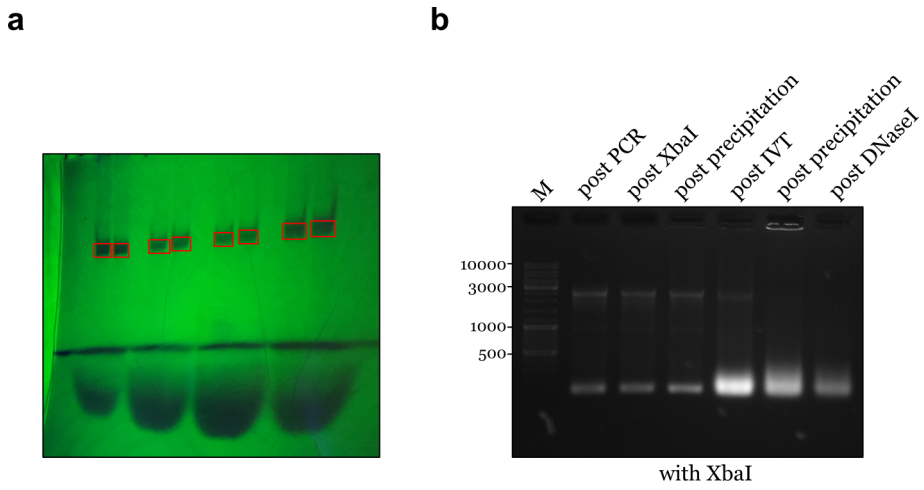


Figure 2.2-2: Results of IVT performed in our laboratory, with XbaI restriction. a, Urea-PAGE imaging with an UV lamp. The bands that are attributed to pre-crRNA are highlighted with red squares. These are cut out for passive elution. **b,** Agarose gel of samples to follow the RNA generation procedure.

Thus, the XbaI incubation was added to the transcription protocol and the transcription was repeated with 4 PCR amplification reactions and with XbaI restriction. The results of the agarose gel and the urea-PAGE of all transcription reactions are shown in Figure 2.2-2 and are very similar to those showed above. The urea-PAGE bands of the transcription products were cut out and, combined with the reactions from Figure 2.2-1 purified via passive elution. Passive elution was used instead of electroelution, because we have no access to an electroelution chamber in our laboratories. But in principle also this method of RNA extraction from a PAGE gel should work. For this, the gel slices were crushed and incubated at 65°C and 1000 rpm in a NaOAc solution. This should enable the RNA to get into the liquid phase. After repeating the elution a second time, the suspensions were filtered, and the RNA precipitated to obtain a pellet. The pellet was resuspended in ddH_2O.

From 7 PCR reactions, 2.37 nmol pre-crRNA was obtained, which is about 338 pmol pre-crRNA per PCR reaction. It should be noted that the yields can vary from different PCR reactions. However, compared to the first attempt, the yield was increased, and the transcription optimized through XbaI restriction.

To sum this Section up, pre-crRNA can be transcribed in our laboratories with sufficient yields for cleavage assays. The passive elution of RNA works, but it is difficult to attribute the increased RNA yield in our laboratory compared to the laboratory of Prof. Mayer to a specific step during the generation of pre-crRNA. Especially the lack of agarose samples collected from the first transcription makes this attribution difficult.

2.3 Establishing pre-crRNA cleavage assays

*Lbu*Cas13a was successfully expressed and purified, and the pre-crRNA was successfully transcribed. The next Chapters deal with the establishment of functional assays for *Lbu*Cas13a.

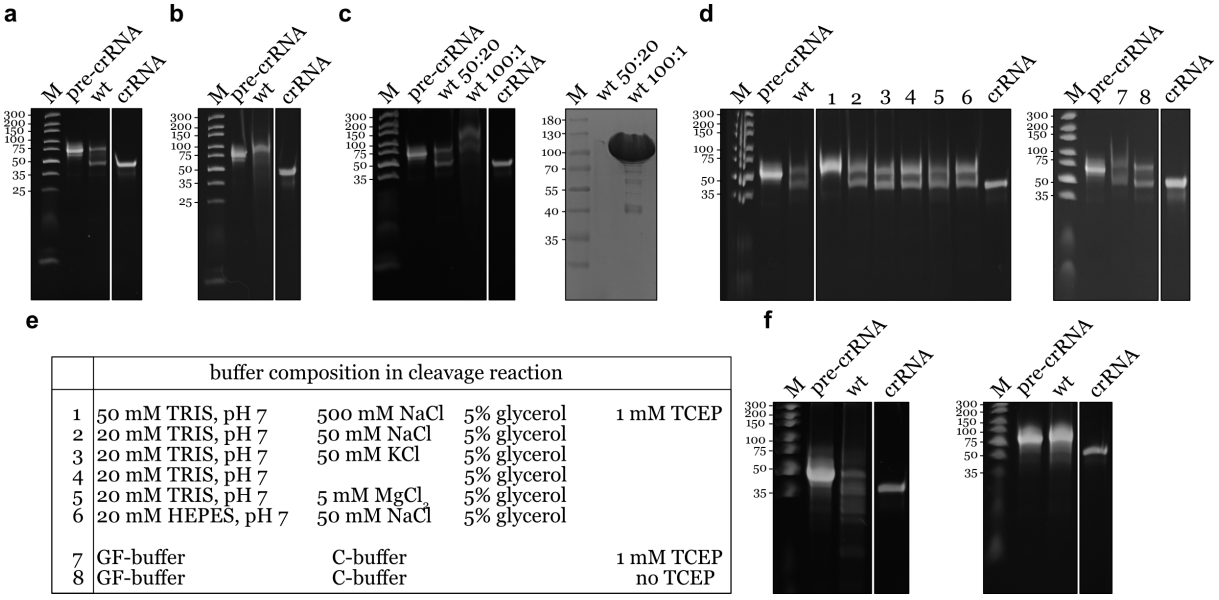
Functionality assays are highly important, especially when the aim is to obtain structural information since the structure and the function of proteins are strongly linked to each other. Cleavage assays were developed to test if *Lbu*Cas13a is active, as reported in literature[75,87].

The general idea of the experiment was to incubate the pre-crRNA with the protein. Subsequently, the protein should be incubated with proteinase K, to quench the reaction and simultaneously release possible bound RNA from the protein. The detection of cleavage products should be done by urea-PAGE, as done for ITV products. When looking into the literature, it seemed that the ratio between the pre-crRNA and the protein would not make a significant difference. In Liu et al.[87] a 1.25 molar excess of pre-crRNA was used, whereas in East-Seletsky et al.[76] a 100 fold molar excess of the protein was used. The approach of East-Seletsky et al. includes a radioactively labelled RNA while the approach used in Liu et al. was more similar to what was aimed for this thesis. Further, the same cleavage buffer was used for this thesis, which has also been used by Liu et al.[87].

The first attempt for a cleavage assay was to use a 2.5-fold excess of protein (50 pmol) to pre-crRNA (20 pmol). The reason of choosing 20 pmol RNA is, that this amount can be easily detected by ethidium bromide staining. The result is shown in Figure 2.3-1 a. The pre-crRNA is a 58 nt sequence, while the crRNA is 51 nt long. However, this does not exactly fit to the bands of the marker (M), being an ultra-low range DNA ladder. RNA can adopt different secondary structures leading to different running behaviors in a native and denaturing PAGE. Since this gel contains urea it should denature secondary and tertiary structures, but this is not always the case, as it was shown by Hegg et al.[265]. Ternary structures have also been detected during target RNA cleavage assays (Figure 2.4-1). Thus, in this case, the DNA ladder is an instrument to compare bands from different gels, rather to attribute absolute sizes to an RNA strand. In Figure 2.3-1 a, the *in-vitro* transcribed pre-crRNA, and crRNA was loaded as a negative control and a positive control, respectively. The pre-crRNA shows two bands at around 75 nt. It remains elusive, if the second, lower band origins from a different secondary structure or from a different sequence. The pre-crRNA that has been incubated with the wildtype is partially cut to form an RNA sequence that matches the running behavior of the crRNA positive control. Hence, the wildtype is active and cleaves pre-crRNA to form the crRNA. This result was promising, however it was not reliably reproducible, as shown in Figure 2.3-1b.

Further experiments were performed to test if this inconsistency could result from an unfavorable protein-RNA-ratio or from wrong buffer conditions. Firstly, the protein ratio was increased to 100:1 protein to pre-crRNA, matching the ratio used by East-

Seletsky et al.[76]. The result is shown in Figure 2.3-1 c. Although the bands for this reaction show cleavage, the bands smear upon increasing the protein amount in the cleavage reactions. Samples of the cleavage reactions were taken to test for protein degradation during cleavage reactions (Figure 2.3-1 c), which could be the reason for inconsistent cleavage efficiencies. Even though the SDS-PAGE was overloaded, most of the protein seems to be intact. Some other protein bands appear at lower molecular weights, but these are neglectable impurities that are visible because of the very high sample concentration.



	buffer composition in cleavage reaction			
1	50 mM TRIS, pH 7	500 mM NaCl	5% glycerol	1 mM TCEP
2	20 mM TRIS, pH 7	50 mM NaCl	5% glycerol	
3	20 mM TRIS, pH 7	50 mM KCl	5% glycerol	
4	20 mM TRIS, pH 7		5% glycerol	
5	20 mM TRIS, pH 7	5 mM $MgCl_2$	5% glycerol	
6	20 mM HEPES, pH 7	50 mM NaCl	5% glycerol	
7	GF-buffer	C-buffer		1 mM TCEP
8	GF-buffer	C-buffer		no TCEP

Figure 2.3-1: Development of pre-crRNA cleavage assays. a, Cleavage assay with wt Cas13a. **b,** Repetition of cleavage assay shown in a. **c, left,** Cleavage assay with wt Cas13a at different protein to pre-crRNA ratios**. right**, 10% SDS-PAGE gel of samples collected after the cleavage assay shown left. **d,** Cleavage assays with wt Cas13a different buffer conditions. **e,** Buffer conditions that were used for the cleavage assays in d. GF-buffer denotes the gel filtration buffer and C-buffer is the cleavage buffer. **f**, Cleavage assays with wt Cas13a without TCEP.

Secondly, different buffer conditions were tested (Figure 2.3-1 d, e), in which salt concentrations and buffer components were varied. For most reactions, buffer components did not change cleavage efficiency. However, the protein seems to be inactive in the cleavage reaction set up containing only GF buffer (Figure 2.3-1 d, lane 4, labelled with 1). This buffer contained a high salt content and TCEP. To test, if TCEP had a negative influence on RNA cleavage, two cleavage reactions were performed with the same buffer components but with and without TCEP (Figure 2.3-1 d, e, second gel in d, lanes labelled with 7 and 8). As it is shown in the cleavage reactions 7 and 8 in

Figure 2.3-1 d, TCEP has no influence on the cleavage rate, but the bands of the reaction with TCEP smear more than the ones without. Thus, other cleavage assays were set up without TCEP (Figure 2.3-1 f left). However, also there was a high inconsistency, with sometimes more or less smeary bands. Additionally, when the cleavage assay was repeated (Figure 2.3-1 f, right), the wt protein was inactive, which is the second inconsistency.

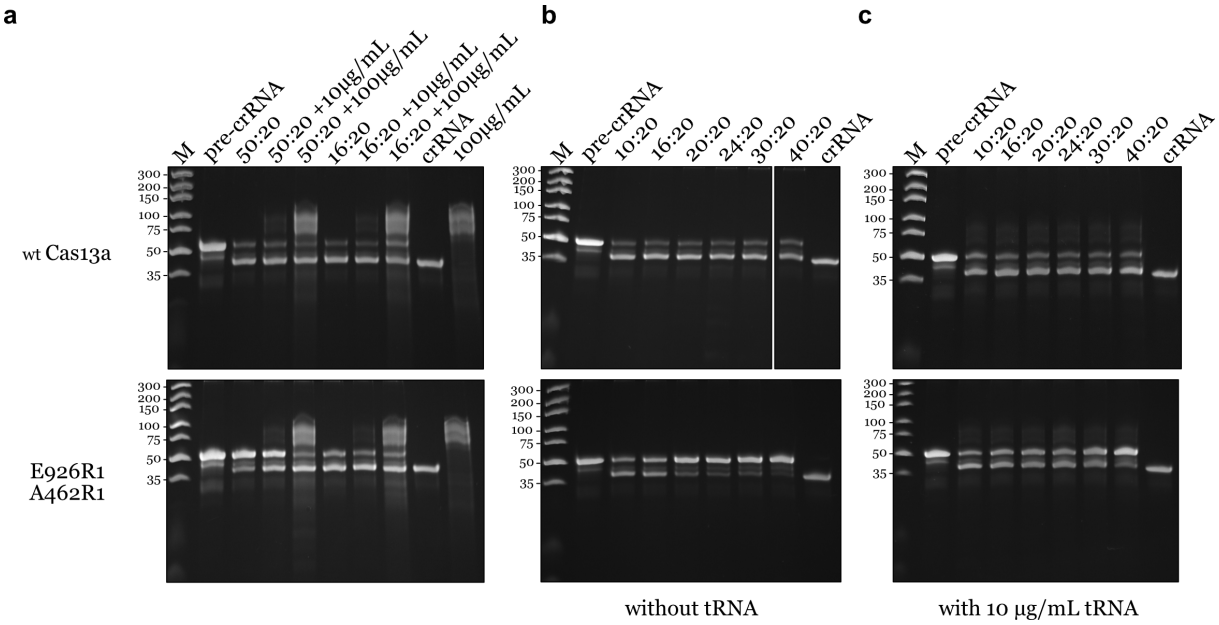


Figure 2.3-2: Influence of tRNA in pre-crRNA cleavage assays. The assays in the top row were performed with wt *Lbu*Cas13a and the ones on the bottom with the E926R1 A462R1 construct, which lacks all three native cysteines. The descriptions on the top are also valid for the assays on the bottom. The protein: pre-crRNA ratio is given in pmol and the tRNA concentration is given in µg/mL, if it was added to the mixture. **a,** First cleavage assays with and without tRNA. The concentration of tRNA is given in the description. Additionally, the protein:pre-crRNA ratio was varied. **b,** Cleavage assays performed without tRNA and by varying the protein:pre-crRNA ratio. **c,** Same cleavage assays as in b, but with 10 µg/mL tRNA in each cleavage reaction.

Among other components, East-Seletsky et al.[75] used total yeast tRNA in the cleavage reactions, maybe to reduce unspecific binding of RNA to the protein, which could cause the inconsistencies, seen in our cleavage assays. Thus, the first component that was analyzed to optimize this assay is the addition of tRNA. For this, 10 and 100 µg/mL tRNA was used in the cleavage reactions. Secondly, pre-crRNA was purchased to investigate its effect on the cleavage behavior, in comparison to self-transcribed RNA. The third variation that has been made was to further adjust the ratio of protein to pre-crRNA. If unspecific RNA binding occurs, this behavior should decrease upon shifting the protein:pre-crRNA ratio from 50:20 to 16:20, instead of 100:1, as done above (Figure 2.3-1 c). In addition to changes in the cleavage reactions, also a technical

change during PAGE preparation was introduced. The urea PAGE was performed with large glass plates (200 x 205 mm) instead of small ones (100 x 100 mm), which should result in an enhanced separation of RNA bands. The results are shown in Figure 2.3-2 a. During the optimization of pre-crRNA cleavage assays, different double-cysteine constructs of *Lbu*Cas13a were generated and labelled for subsequent PELDOR measurements (Chapter 3.4, 3.5, and 4). Thus, to test the influence of tRNA and pre-crRNA ratio on labelled constructs, the same cleavage assay was performed with the E926R1 A462R1 construct, which lacks the three native cysteines C293A C348V C1141A (Chapter 3.2.2, Figure 2.3-2 a, bottom). Interestingly, for this labelled construct, the protein:pre-crRNA ratio and the tRNA has an impact on the cleavage efficiency. The lower the protein amount and the more tRNA is added, the higher is the cleavage product band. To investigate this further, four cleavage assays were performed. Two of them with the wt and the other two with the E926R1 A462R1 construct. One set of cleavage assays was done without and one set with tRNA. Here an amount of 10 µg/mL was chosen, since this tRNA band is not as prominent as in the samples containing 100 µg/mL. The second reason for choosing 10 µg/mL is, that there is not much difference regarding cleavage efficiency between the samples of 16:20 containing 10 µg/mL and 100 µg/mL (Figure 2.3-2 a). The cleavage reactions were set up with varying the protein:pre-crRNA ratio from 10:20, to 40:20, and the results are shown in Figure 2.3-2 b, c.

As it is nicely seen in the cleavage assays of the wt protein (top row in Figure 2.3-2 b and c), the protein:pre-crRNA ratio and tRNA have no impact on the cleavage efficiency. In contrast, by gradually increasing the amount of the E926R1 A462R1 construct in the reaction, a gradual decrease in cleavage is observed. The intensity of the upper band, corresponding to the pre-crRNA, increases while the lower crRNA-band decreases. This trend is also seen in the cleavage assay supplemented with tRNA, but to a lower degree. By comparing each reaction in the construct-based assays in Figure 2.3-2 b and c, it is noted that upon addition of tRNA the protein is more active than without tRNA. These observations indicate that the pre-crRNA non-specifically binds to Cas13a.

Further cleavage assays were performed with 16 pmol protein and 20 pmol pre-crRNA, with 10 µg/mL tRNA. The reasons for choosing this ratio were mainly, practical reasons. By pipetting 16 pmol of protein, usually 0.1-0.3 µL were pipetted from protein

stocks. Upon further decreasing the protein amount, the pipetting error would increase and pipetting errors due to dilutions would occur. The second reason was, that there is not much difference in the cleavage reaction efficiencies of the reactions containing a protein amount of 24 pmol to 10 pmol.

The last change that was made in the pre-crRNA cleavage assay protocol was the addition of 50 mM L-glutamic acid and 50 mM of L-arginine to the cleavage buffer. It was observed during PELDOR sample preparations that the protein precipitates and forms amorphous structures (Chapter 3.6). This precipitation could be decreased and sometimes prevented by the addition of these two amino acids. Therefore, these amino acids have been added to the cleavage buffer to ensure consistency of the functional assays with the PELDOR sample preparation, from which structural information will be derived. The testing of the amino acid combinations was based on literature[266] and was performed by Catrin Allar (PhD student, University of Bonn).

2.4 Development of target RNA cleavage assays

Since *Lbu*Cas13a catalyzes two cleavage reactions, also the target cleavage needed to be tested. The same approach as for the pre-crRNA cleavage assay was used, consisting of a cleavage reaction followed by proteinase K incubation and visualization of RNA via ethidium bromide staining of a large (200 x 205 mm) 20% urea PAGE. Different ratios of target RNA, crRNA and protein were tested, but since this optimization was done by Catrin Allar, it is omitted here. Instead, the final target RNA cleavage assay protocol is discussed.

As described in East-Seletsky et al.[75], after crRNA annealing, a molar ratio of 2:1 protein (40 pmol) to crRNA (20 pmol) was used to form the binary complex (Methods Chapter 7.4.7). Then, 20 pmol target RNA was added and the reaction was performed in the cleavage buffer with L-arginine and L-glutamic acid by omitting tRNA, since East-Seletsky et al.[75] did not use tRNA in their cleavage buffer. The results are shown in Figure 2.4-1. For this assay, wt Cas13a and the double cysteine construct S660C E926C, which lacks all three native cysteines, were used. In addition to this 660/756^{active} construct, also the pre-crRNA cutting deficient 660/756^{dpecr} and target RNA cutting deficient 660/756^{dtarget} constructs were tested regarding their target RNA

cleavage activity. As briefly described in the introduction (Chapter 1.6.5), there are literature known mutations that inhibit pre-crRNA and target RNA cleavage. The pre-crRNA cleavage is inhibited by R1079A, while target RNA cleavage is inhibited by R1048A H1053A. Thus, the dprecr construct should not cleave the pre-crRNA while retaining target RNA cleavage. The dtarget construct should not degrade target RNA, while retaining pre-crRNA cleavage.

Both, the crRNA and the target RNA alone produced sharp bands. Interestingly, when the crRNA and target RNA were mixed and incubated for 1 h at 37°C they formed secondary or tertiary structures which resulted in a blurred band approximately between 50-100 bp. This band is also formed in the other reactions, when crRNA and target RNA are both included. In the reactions, in which the protein is active, no target RNA band was seen, since the target RNA is degraded, and since these small fragments were presumably too short to be visualized.

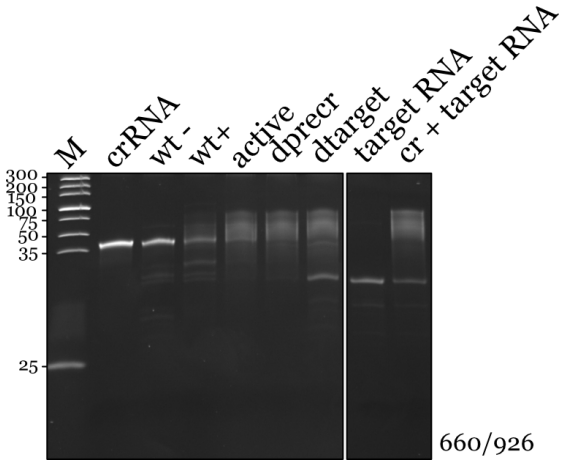


Figure 2.4-1: Target RNA cleavage assay of wt Cas13a and double cysteine construct S660C E926C. The crRNA, target RNA, and the annealed cr- and target RNA together are loaded as negative control. wt- denotes a reaction that was performed with the wt and crRNA, lacking target RNA and wt+ denotes a cleavage reaction containing crRNA, target RNA and the wt as positive control. Active, dprecr, and dtarget shows cleavage reactions that were performed with the active double cysteine construct 660/926, the pre-crRNA cleavage deficient construct containing R1079A, and the target RNA deficient construct containing R1048A H11053A, respectively.

As expected, wt, 660/756^{active}, and 660/756^{dpecr} lack a band that corresponds to the target RNA. Thus, these constructs degrade target RNA and are active. In contrast the sample of 660/756^{dtarget}, which should not cleave target RNA, contains a band that corresponds to the target RNA hence this protein construct is inactive regarding target RNA cleavage. Interestingly, the secondary and ternary structures of cr- and target

RNA were strongly reduced in the reaction containing the wt protein (denoted as wt+). This could be an indication that the wt is more active than the other constructs and that the crRNA and target RNA are degraded to a higher extend.

2.5 Discussion and Conclusion

In this Chapter, biochemical techniques for *Lbu*Cas13a were established, ranging from protein synthesis and protein handling procedures to RNA handling and the establishment of functionality assays.

Different protein expression tests were performed with different cell lines and with variation of temperature and inductor to find the best protein expression conditions. BL21 AI cells were found to yield high protein expression levels, with no leaky expression in a reproducible fashion. In contrast to Rosetta, where cell growth to log-phase took 2 – 6 h, BL21 AI cells showed a consistent growth behavior by reaching log-phase after 2 h. Additionally, as also seen for Cas9 expression rates[261], the Cas13a expression rates in BL21AI were higher than the ones in Rosetta. This is an indication, that maybe the underlying principle is the same, and that Cas13a mRNA transcription rates are lower in Rosetta, leading to lower expression rates. The protein amount and protein quality were significantly increased by expression in BL21 AI. However, one possibility to further optimize protein yields would be to lower the temperature during protein expression, which could diminish inclusion body formation and maximize the amount of protein in the soluble fraction after cell lysis. Since the protein is already expressed at a low temperature of 16°C, also cells specifically developed for low temperature expression, as Arctic Express *E. coli*, could be tried out.

During successful *Lbu*Cas13a purification the heparin column turned out to be the key step for a high-quality *Lbu*Cas13a isolation. Most importantly, oligonucleotides that have been bound to *Lbu*Cas13a during expression and purification were successfully removed by this heparin column. This is of essential importance for functionality assays and for consequent structural analysis. To perform functionality assays as cleavage assays, pre-crRNA was successfully transcribed *in-vitro*. However, the yields were small. It was seen that the ethanol precipitation of RNA during IVT was not quantitative, maybe other precipitation procedures, such as with isopropanol[267], could

have been tested. Another disadvantage of ethanol precipitation is, that despite precipitating small RNA sequences, it may also precipitate proteins[267]. So, an improvement could be to perform a phenol-chlorophorm extraction prior precipitation.

During cleavage assays it was seen, that bought pre-crRNA was cleaved to a higher degree than the IVT pre-crRNA. This could have several reasons, of which one is the 3′-end inhomogeneity of the transcript. Most commonly, transcribed RNA has a single A extension[264],which could have an influence on cleavage behavior. However, it is difficult to tell if this has an influence since the additional A at the 3′-end is 8 nucleotides away from the cleavage site. Since the yields of the pre-crRNA from IVT were small and since the bought pre-crRNA was working well, the focus was set on the preparation of protein constructs that could have been used for PELDOR spectroscopy (see next Chapter) and only the bought pre-crRNA was used for cleavage assays.

Both, the pre-crRNA cleavage, and the target RNA cleavage were successfully established and rely on the visualization with ethidium bromide. The staining works well but is not quantitative and depends on the secondary structure and the length of oligonucleotides in which ethidium bromide intercalates. Thus, only the intensity of bands containing the same RNA sequence can be compared. One additional disadvantage of performing the target cleavage assay as described above is lack of visualization of cleaved target RNA. However, ethidium bromide was the best choice in our laboratory, since we already had a suitable detection setup and since we did not want to characterize the specific, quantitative activity of each protein mutation or RNA mutation. Instead, we sticked to qualitatively determine if a protein construct has a drastically reduced cleavage activity or if this activity is similar to the one of the wt protein.

Chapter 3

Development of active, double labelled *Lbu*Cas13a constructs

3.1 Spin Labelling of wildtype *Lbu*Cas13a

As mentioned above, *Lbu*Cas13a has three native cysteines, 293C, 348C, and 1141C. One common method for protein labelling is via cysteines. This labelling procedure requires a protein without native cysteines on the protein's surface, to allow complete freedom in the choice of labelling sites. By looking at the AF3 structure prediction of *apo Lbu*Cas13a (Figure 3.1-1) it can be hypothesized that only C348 is not solvent accessible. This is supported by the calculation of solvent accessibility through pyMOL, on the basis of the *apo* AF3 prediction, the pre-crRNA bound AF3 prediction, the crRNA bound cryo-EM structure, and the cr- and target RNA bound crystal structure (Table 3.1-1).

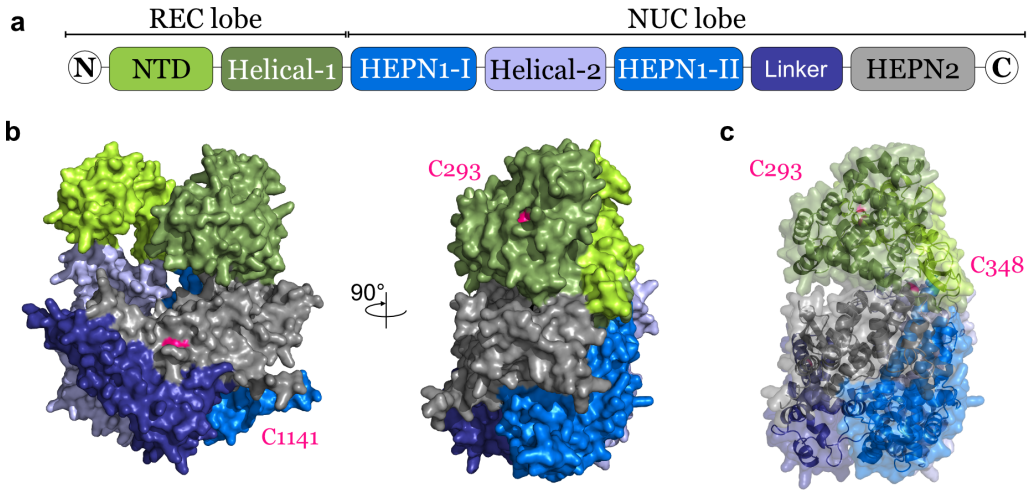


Figure 3.1-1: Surface accessibility of native cysteines. a, Domain architecture of *Lbu*Cas13a. **b,** Surface representations of *Lbu*Cas13a from two perspectives (rotated by 90°). Native cysteines are shown in magenta. **c,** Same representation as in b (right), but with a semi-transparent surface showing the locations of C293 and C348.

Depending on the structure on which the calculation is based, C1141 seems to have the highest solvent accessibility, of 5-12% and it is clearly visible on the protein surface (Figure 3.1-1 b, left). C293 is visible on the protein surface (Figure 3.1-1 b, right) but it is buried in a cleft, which leads to a lowered solvent accessibility of 2-3%. C348 cannot be seen on the surface and was calculated to have a solvent accessibility of 0-3%, meaning that it is very unlikely that C348 is going to be labelled (location of C348 can be seen in Figure 3.1-1 c). Even though a certain degree of solvent accessibility is predicted, generally this degree is very small for each of the three cysteines. So, maybe

none of these native cysteines are labelled, or to a very low extend. This would be favorable, since no, or minimal native cysteine replacements via mutagenesis would be required. However, it is difficult to draw reliable conclusions since each *in-silico* determination is done on a single static protein structure that doesn't account for different protein conformations, local dynamics, and rearrangements of residues or loops. These factors could increase or decrease solvent accessibility and by this, increase or decrease cysteine availability for an unwanted labelling reaction.

Table 3.1-1: Solvent accessibility of native cysteines. The solvent accessibility of native cysteines was determined with pyMOL, based on the AF3 predictions, or on known structures, as indicated in brackets.

Structure	C293	C348	C1141
apo (AF3)	3%	0%	7%
pre-crRNA bound (AF3)	3%	2%	7%
crRNA bound (PDB-ID: 5XWY)	2%	0%	12%
crRNA- & target RNA bound (PDB-ID: 5XWP)	3%	1%	5%

To experimentally determine if one or more native cysteines are solvent exposed and by this, available for the labelling reaction, the *apo* wt *Lbu*Cas13a construct was labelled with cysteine-reactive MTSL (structure and labelling reaction shown in Figure 1.8-5 a). The *cw*-EPR spectrum is shown in Figure 3.1-2 and the labelling procedure is described in Methods Chapter 7.3.1.

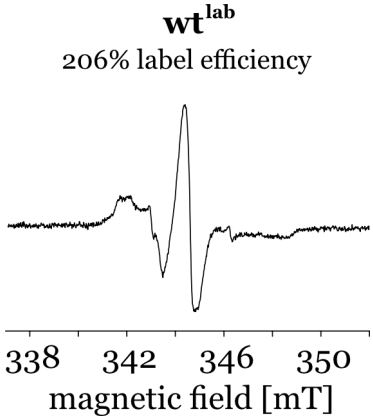


Figure 3.1-2: *cw*-EPR spectrum of the wt *Lbu*Cas13a with MTSL. The protein concentration was 50 µM, spin concentration of 103 µM was determined via spin count against an internal standard, and the spectrum was acquired with 156 scans.

The *cw*-EPR spectrum of *apo* wt *Lbu*Cas13a shows a labelling efficiency of over 200% indicating that at least two native cysteines are labelled. The conclusion that can be drawn from this spectrum is that the wt *Lbu*Cas13a is labelled and that all three native cysteines need to be replaced.

3.2 Searching for replacements of the three native cysteines

3.2.1 Functional importance of native cysteines

It is not straight-forward to decide which amino acid should be used to replace the three native cysteines. This depends on the structure of the protein and the local surroundings of the amino acid positions. A bulky amino acid can perturb the proteins structure because of the enlarged space it requires. A charged amino acid could be the choice for protein regions that are on the protein surface and pointing towards the solvent, but these would be a bad choice for an amino acid buried in the protein, surrounded by hydrophobic amino acid side chains. Since all three cysteines are rather buried, alanine was chosen as replacement, because of its small size. Additionally, serine was chosen as a second option, since serine only differs by the oxygen in the alcohol group that replaces sulfur of the thiol in cysteine (Figure 3.2-1 a). Thus, serine can form hydrogen bonds with surrounding interaction partners, as it is also done by cysteine, lowering structural perturbations.

The purification of C293A C348A C1141A, from this point on called cysfree-A construct, worked well. This is seen in the size exclusion chromatogram (Figure 3.2-1 b) and SDS-PAGE analysis (Figure 3.2-1 c). After spin labelling with MTSL this protein construct was subjected to *cw*-EPR spectroscopy and to a cleavage assay, to test for unspecific labelling and to test for pre-crRNA cleavage activity, respectively. Regarding the *cw*-EPR spectrum, no EPR signal is visible, highlighting that no unspecific spin labelling occurred (Figure 3.2-1 d).

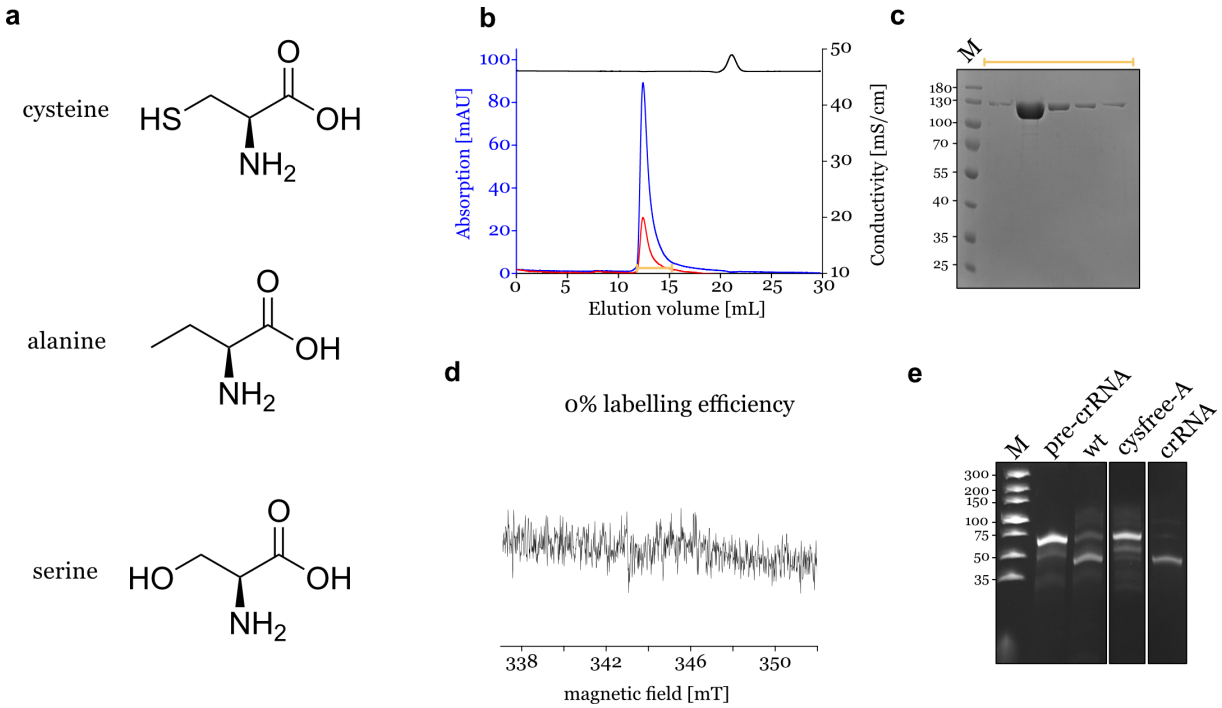


Figure 3.2-1: Analysis of the cysteine-free alanine construct C293A C348A C1141A. a, Structures of cysteine, alanine and serine. **b,** Size exclusion chromatogram of the protein purification with a SD200 10/300 column. The blue curve shows the absorption at 280 nm and the red curve the absorption at 254 nm. The yellow bar indicates the region from which samples were taken for SDS-PAGE analysis shown in c. **c,** SDS-PAGE of the size exclusion shown in b and of the final concentrated protein. **d,** *cw*-EPR spectrum of the protein construct after labelling procedure (100 scans). **e,** pre-crRNA cleavage assay of this protein construct with the wt protein as comparison. This construct was not labelled with MTSL.

By analyzing the pre-crRNA cleavage activity of the unlabelled cysfree-A construct (Figure 3.2-1 c) it becomes evident, that in the lane consisting of a cleavage reaction with cysfree-A protein only the pre-crRNA band is visible. This means that the cleavage activity of this construct is almost completely abolished. In contrast the reaction containing wt *Lbu*Cas13a shows two bands corresponding to the pre-crRNA and the crRNA, the latter being the more intense band. In summary, the cysfree-A construct is not active and cannot be used for structural studies.

As a second attempt, the cysfree-S construct was cloned, containing the mutations C293S, C348S, and C1141S. The results of the purification and the cleavage assay are shown in Figure 3.2-2. The purification yield was 6.2 mg for 6 L LB culture and a HiLoad200 16/600 column was used. The chromatogram shows one distinct peak at an elution volume of 70 mL (Figure 3.2-2 a).

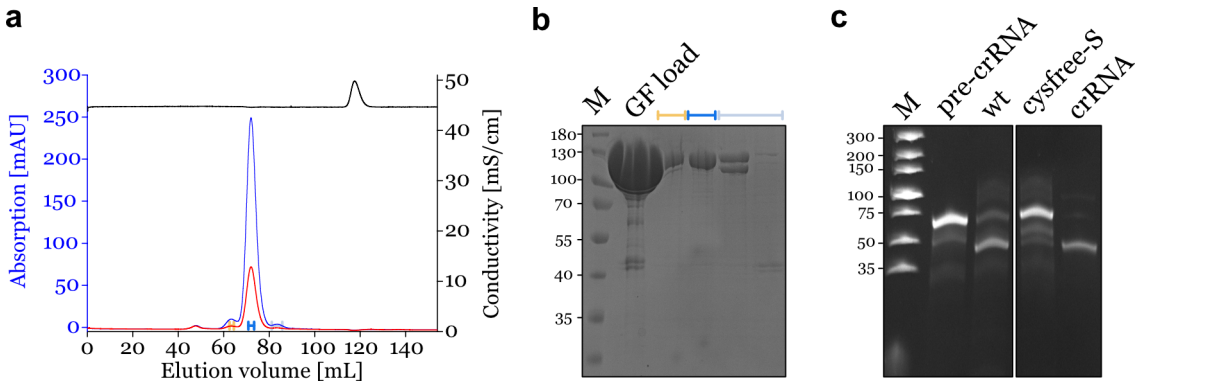


Figure 3.2-2: Analysis of the cysteine free construct C293S C348S C1141S. a, Chromatogram of the size exclusion during protein purification with a HiLoad200 16/600. The blue curve shows the absorption at 280 nm and the red curve the absorption at 254 nm. The colored bars indicate the regions from which samples were taken for SDS-PAGE, shown in b. **b,** SDS-PAGE of the same purification as in a. GF load denotes the sample taken from the protein solution loaded onto the size exclusion column. **c,** pre-crRNA cleavage assay of this protein construct with the wt protein as comparison.

As seen in the SDS-PAGE, the peak consists of the desired protein (Figure 3.2-2 b). However, as the cysfree-A construct, also the cysfree-S construct has a drastically reduced activity regarding pre-crRNA cleavage (Figure 3.2-2 c). Thus, even though the three cysteines were replaced by serine, which should preserve hydrogen bonding, it has a drastic effect on the protein activity.

To determine if one specific cysteine is functionally relevant or all three, each cysteine-to-alanine combination was cloned, expressed, and purified. These six constructs are C293A, C348A, C1141A, C293A C348A, C349A C1141A, and C293A C1141A (Figure 3.2-3). As seen from the SDS-PAGEs in Figure 3.2-3 a-f, each protein construct was purified with a high degree of purity and then subjected to a pre-crRNA cleavage assay (Figure 3.2-3 g). The single cysteine-to-alanine protein constructs C293A and C1141A were similarly active compared to the wt protein. In contrast, C348A presented a reduced cleavage activity. This trend was also visible in the double cysteine-to-alanine constructs. The pre-crRNA cleavage activity of C293A C1141A, in which C348 is still present, was equal to the one of the wt. In contrast, the activity of C348A C1141A was lower than the one of the wt and the activity of C293A C348A was even lower. This indicates that C348 is involved in efficient pre-crRNA cleavage and that the combination of C293A and C348A has an additional negative impact on pre-crRNA cleavage.

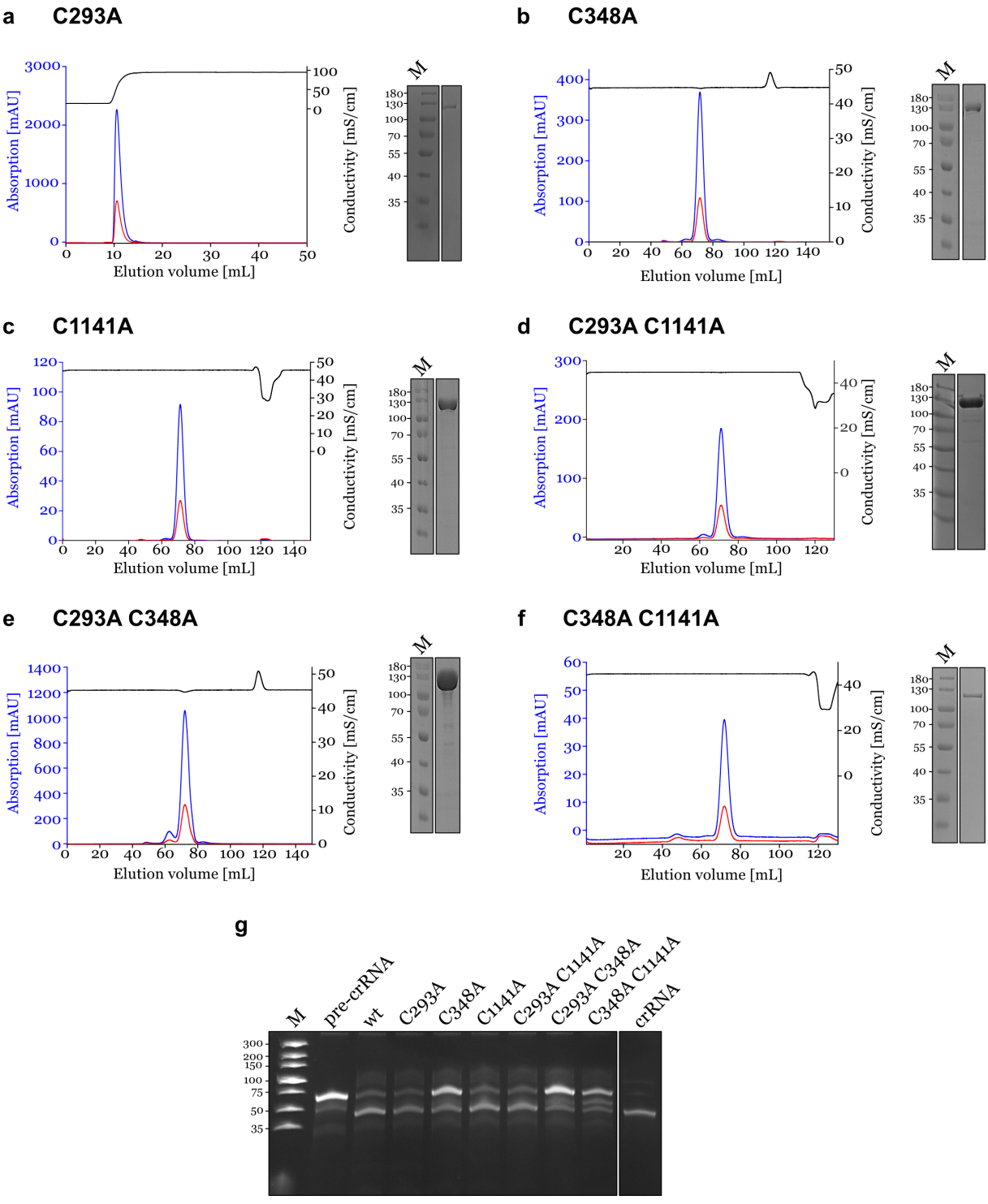


Figure 3.2-3: Purification results of different *Lbu*Cas13a constructs and their pre-crRNA cleavage behavior. The chromatograms of the last purification step with the SDS-PAGE of the eluted protein are shown for each construct. **a,** The chromatogram of the heparin column of C293A and the SDS-PAGE of a sample taken from the main peak is shown. It was not possible to run a size exclusion, due to pressure issues with the size exclusion column. **b-f,** Size exclusion chromatograms and SDS-PAGEs of the main peak in the chromatograms are shown from different protein construct, as denoted on top of each figure section. **g,** pre-crRNA cleavage assay of the protein constructs shown in a-f.

However, it is unknown if C348 and C293 are directly involved in pre-crRNA cleavage, e.g. through direct contacts to the pre-crRNA or through involvement in bond cleavage. Further it is unknown if the exchange to alanine leads to structural changes that could explain this decreased cleavage activity. Structural changes can occur for example due to the loss of interactions between amino acids. So, to investigate if the native cysteines are involved in hydrogen bonds or non-covalent interactions, a Ligplot[268] analysis was performed. Ligplot detects non-covalent interactions in experimental structures from the PDB and generates an interaction map of the residue of interest (Figure 3.2-4 b-d).

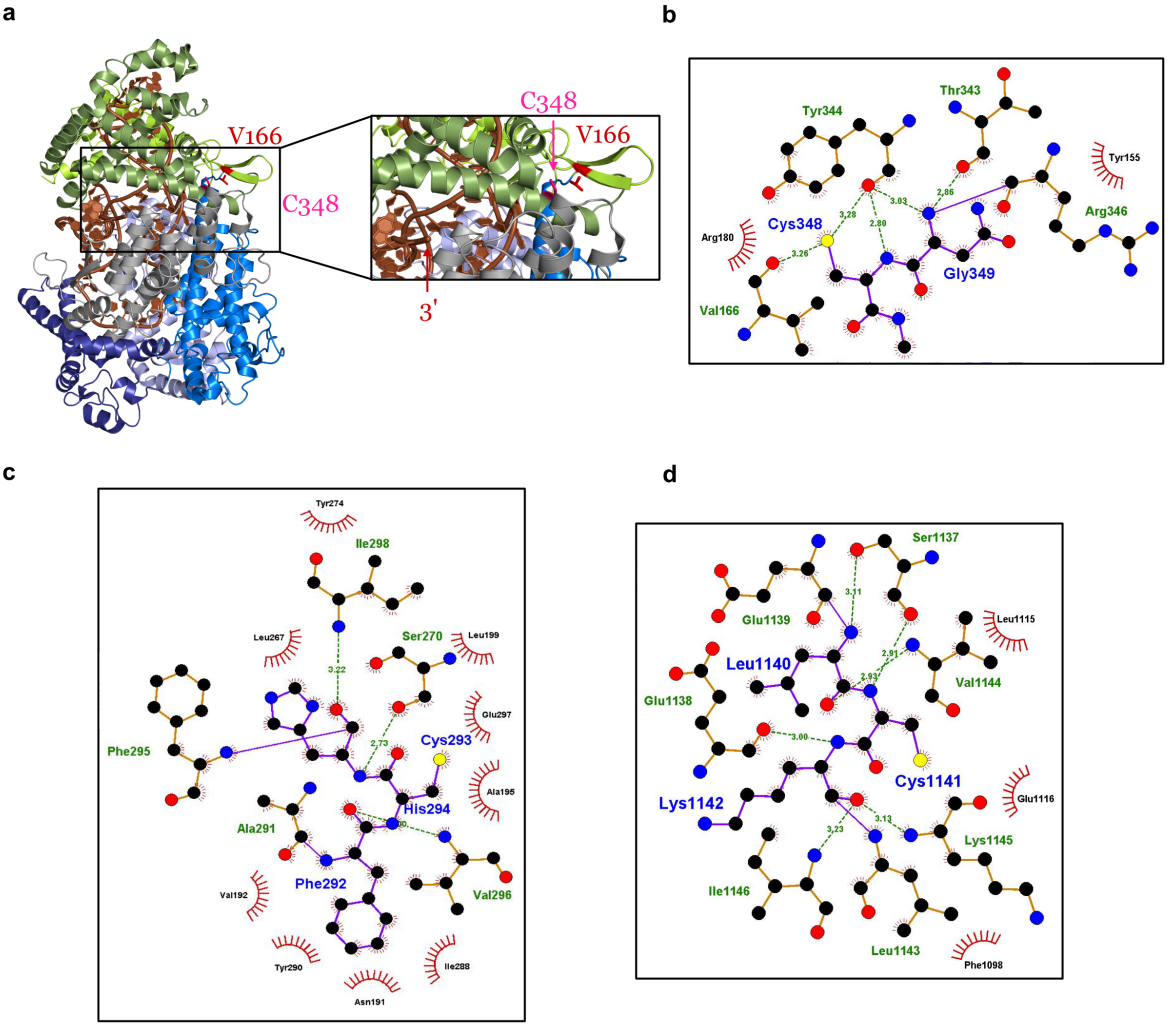


Figure 3.2-4: Structural Ligplot[268] analysis of the three native cysteines. a, cryo-EM structure of the binary *Lbu*Cas13a-crRNA complex. The insert highlights the region in which C348 is located. C348 is colored in pink, while interaction partner V166 is colored red. The 3' end of the crRNA hairpin is highlighted by a red arrow. **b,** Ligplot interaction diagram for C348. Hydrogen bonds are shown with dashed green lines between the atoms. Hydrophobic interactions are shown as red rays. **c,** Ligplot interaction diagram for C293. The color coding is the same as in b. **d,** Ligplot interaction diagram for C1141. The color coding is the same as b.

As it can be seen, C293 and C1141 (Figure 3.2-4 c and d) are not involved in any hydrogen bonds. This fits to their spatial position in the protein, since both are more solvent exposed than C348 (Table 3.1-1, and Figure 3.1-1). In contrast, C348 is involved in hydrogen bonding to the amino acid residues V166 and Y344. This could be an explanation why mutations at this position lead to a drastic reduction of cleavage activity. In addition, C348 is in the helical-1 domain of the REC lobe, directly between the interface of the REC lobe and NUC lobe (HEPN1 and HEPN2 domains), at the closed backbone region of the protein. As described above, the specific active site for pre-crRNA cleavage is unknown. The 3'-end of the crRNA, where cleavage happens, is also located in between the REC and NUC lobe (Figure 3.2-4 a). Thus, the mutation of C348 could lead to a loss of hydrogen bonds and consequently to a perturbed structure in this protein region, which inhibits pre-crRNA cleavage. But since no experimental structure is known, this is only a hypothesis.

In summary, the drastic decrease in activity of the cysfree construct originates mostly from C348A and to a smaller degree from C293. C1141A has little to no influence on the pre-crRNA cleavage.

3.2.2 Development of a cysteine-free active protein construct

On the path of developing a cysteine-free *Lbu*Cas13a construct, it is a valuable information that C348A is mostly responsible for a decreased cleavage activity. By substituting C293A and C1141A the pre-crRNA cleavage efficiency was retained. Thus, the next step was to find an adequate amino acid exchange for C348A. One method to look for an amino acid for C348 exchange, which does not decrease cleavage activity, is to test every natural amino acid. That would include primer design, cloning, expression, purification, and cleavage assays of 19 constructs containing one of the 19 different natural amino acids each (excluding selenocysteine). The other approach is to use a tool called PoPMuSiC[269] embedded in the dezyme server[270]. This tool predicts the thermodynamic stability changes in a protein structure when single site mutations are performed *in-silico*. The thermodynamic function that is used for this is the folding free energy change, $\Delta\Delta G$. A negative sign of $\Delta\Delta G$ corresponds to a stabilization of a protein structure, when a mutation is performed. Consequently, a positive sign indicates a destabilization of the protein's structure upon single point mutation.

$\Delta\Delta G$ was calculated for all 19 natural amino acid exchanges for C348. These calculations were done for each of the two experimentally known structures, being the binary complex and the ternary complex. The results are shown in Table 3.2-1.

Table 3.2-1: Determination of possible stabilizing amino acid mutations for C348. Stabilizing and destabilizing single mutations with their folding free energy change. The amino acids that were used to exchange C348 are highlighted in green.

amino acid	ternary complex ΔΔG of mutation [kcal/mol]	binary complex ΔΔG of mutation [kcal/mol]
ALA	1.24	1.33
VAL	0.61	1.05
LEU	0.78	0.85
ILE	0.65	0.83
PHE	0.81	0.71
TYR	0.74	0.54
TRP	0.72	0.9
MET	1	0.89
SER	1.85	1.69
THR	1.74	1.6
PRO	3.04	2.91
ASN	2.84	2.46
GLN	2.13	2.21
ASP	3.07	2.87
GLU	2.75	2.85
HIS	1.9	1.46
LYS	2.66	2.61
ARG	1.89	1.61
GLY	2.78	2.4

Generally, each mutation of C348 is destabilizing the protein structure. The calculated $\Delta\Delta G$ of a mutation is dependent on the input structure, which is seen by the difference between $\Delta\Delta G$ of a specific amino acid replacement in the binary versus the ternary complex. This difference is mostly 2-28%. The exception is valine, where $\Delta\Delta G$ from the binary complex and ternary complex differs by 72%. Thus, these $\Delta\Delta G$ values should be seen as a general trend, rather than absolute values. Nevertheless, the two amino acids valine and tyrosine were predicted to be less destabilizing than the other amino acids in the ternary and binary complex, respectively. In addition to this $\Delta\Delta G$ analysis, a sequence comparison of the position 348 was made with the two Cas13a proteins that

are phylogenetically most similar to *Lbu*Cas13a, being *Lsh*Cas13a and *Lwa*Cas13a. Surprisingly the sequence alignment showed that C348 corresponds to a tyrosine in *Lwa*Cas13a, which was also suggested by dezyme to be less destabilizing in the binary complex of *Lbu*Cas13a. In *Lsh*Cas13a, leucine corresponds to C348.

Thus, C348 was mutated to tyrosine, valine, and leucine, to test for an active cysteine-free *Lbu*Cas13a construct. The size exclusion chromatograms, the corresponding SDS-PAGE gels, the pre-crRNA cleavage assay results, and the target cleavage assay results are shown in Figure 3.2-5.

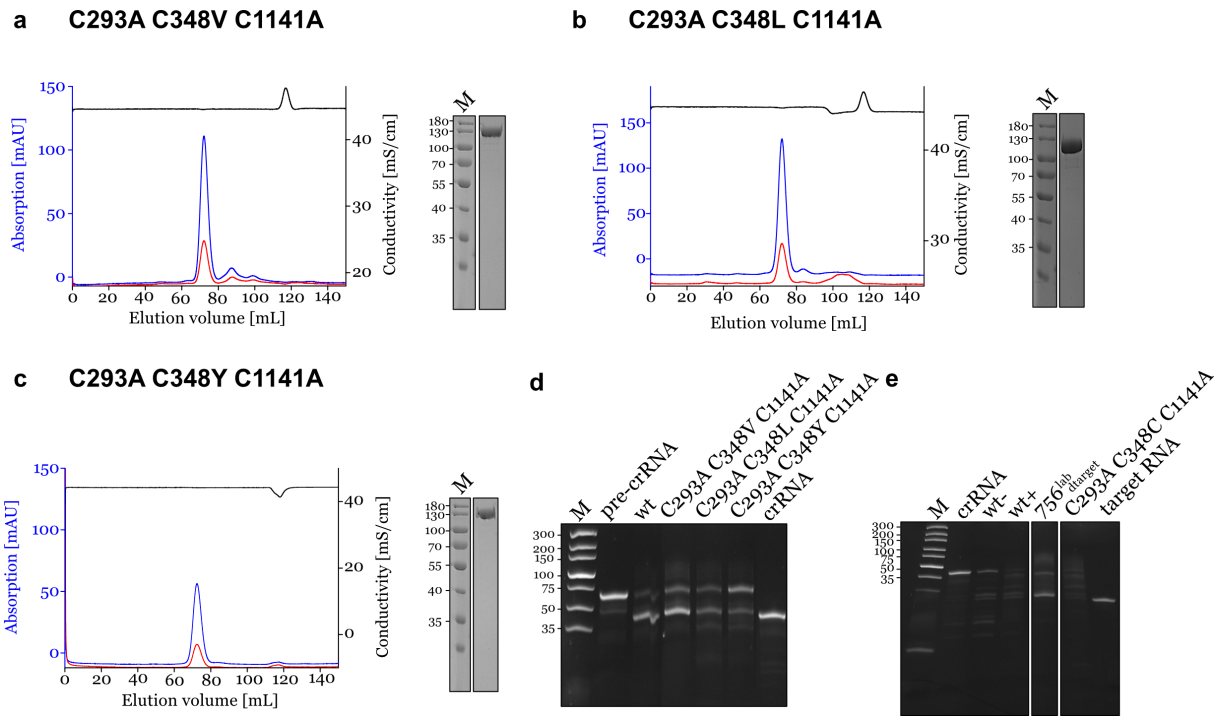


Figure 3.2-5: Purification results and cleavage assays of the three C348X constructs, designed with dezyme[270] and phylogenetic analysis. The chromatograms of the size exclusion run with an SDS-PAGE sample of the main peak are shown from **a,** C293A C348V C1141A, **b,** C293A C348L C1141A, and **c,** C293A C348Y C1141A. The blue curve shows the absorption at 280 nm and the red curve the absorption at 254 nm. **d,** pre-crRNA cleavage assay of the constructs shown in a-c. **e,** Target cleavage assay of C293A C348V C1141A, with a labelled and target cleavage deficient construct (dtarget, 756[lab]) as comparison. wt- denotes a reaction that was performed with the wt and crRNA, lacking target RNA and wt+ denotes a cleavage reaction containing crRNA, target RNA and the wt as positive control.

The protein constructs were successfully isolated (Figure 3.2-5 a-c). All three cysteine-free construct are active regarding pre-crRNA cleavage (Figure 3.2-5 d). However, two constructs, C293A C348L C1141A and C293A C348Y C1141A, showed decreased cleavage, in comparison to the wt Cas13a. In contrast, the construct C293A C348V

C1141A had a similar cleavage activity as the wt protein. In addition to the pre-crRNA cleavage activity, also the target RNA cleavage activity was tested for the C293A C348V C1141A construct, to determine whether the target RNA cleavage is retained (Figure 3.2-5 e). Indeed, also in this assay, the valine construct behaved as the wt Cas13a, with a very similar cleavage pattern. For comparison, an additional construct is shown, which is target RNA cleaving deficient and was spin labelled at the position 756 with MTSL. The target RNA is nicely visible in the cleavage assay of this dtarget construct, in contrast to the wt and to the C293A C348V C1141A construct.

In summary, two of the three native cysteines, C293 and C348, modulate the pre-crRNA cleavage activity of *Lbu*Cas13a. Upon mutation to alanine, it was found that especially C348A drastically reduced pre-crRNA cleavage. Also, the replacement with serine, which should retain possible hydrogen bonding in the protein structure, showed strongly reduced pre-crRNA cleavage. The software PoPMuSiC (dezyme) was used in combination with phylogenetic analysis to determine three possible amino acids that were tested for position 348. Out of these three, one combination was found to yield a cysteine free, active *Lbu*Cas13a, which is C293A C348V C1141A. The activities regarding pre-crRNA and target RNA cleavage were similar to the wt *Lbu*Cas13a. From this point on, this construct is called cysfree-V.

3.3 Development of a labelled *Lbu*Cas13a construct via unnatural amino acids

A second approach was tested, to achieve a labelled protein for PELDOR measurements. This approach consisted of expressing *Lbu*Cas13a with an unnatural amino acid. Since the unnatural amino acid bears a functional group orthogonal to the one at the spin label (Figure 1.8-5 d, e), there is no need for cysteine replacement. The general method to express a protein with an unnatural amino acid (UAA) is to use a stop codon as the codon for the UAA. Commonly, the amber codon (UAC) is used, which is almost absent in bacterial cells and by this has a low influence to expressions of other proteins[271]. This amber codon is recognized by an aminoacyl-tRNA synthetase (aaRS) with the corresponding tRNA. Consequently, the aaRS incorporates the UAA into the amino acid chain. However, if the tRNA is not present in required amounts, the amber codon can be read as a stop codon, leading to a truncated protein construct.

Truncated proteins are mostly inevitable. Thus, it is favourable to have C-terminal affinity tags, which enable to purify only the desired full-length protein. Several cloning steps were made to apply this method to *Lbu*Cas13a. The p2CT-His-MBP plasmid, containing the *Lbu*Cas13a gene had an N-terminal His-MBP-TEV end, followed by the Cas13a gene. The plasmid was modified to have the following scheme: MBP-TEV cleavage site-Cas13a-TEV cleavage site-His, thus a C-terminal TEV cleavage site and His_6-tag. The primers used for cloning are found in the Materials Section, Table 6.5-3. This cloning should ideally lead to total remotion of truncated protein constructs during Ni-AC since truncated products are lacking a His_6-tag.

Once the plasmid was modified, it was co-transformed into BL21 AI cells, with one of three plasmids, being pAcF/pUltra, pAcF/pEVOL, and pAzF/pEVOL (Chapter 7.1.3). These plasmids encode an aminoacyl-tRNA synthetase (aaRS) and the corresponding tRNA for incorporation of acetylphenylalanine (AcF) or azidophenylalanine (AzF). This tRNA-aaRS pair is constructed such that the tRNA is amino-acylated only by the cognate aaRS and not by any other cellular aaRS. This ensures a high incorporation rate.

Expression tests were performed with all three Cas13a-UAA constructs, without amber codon, to test if the UAA in the media influences expression of Cas13a. The procedure is similar to the expression tests without UAA (Chapter 2.1.2 and 7.2.5). The main cultures were set up with the corresponding antibiotics and after 2.5 h at 37°C, different amounts of UAA were added to the main cultures. The temperature was reduced to 16°C prior induction and SDS-PAGE samples of the main cultures were taken before and after 2 h, 6 h, and 16 h post induction. The results are shown in Figure 3.3-1. Generally, all three BL21AI cell cultures overexpress Cas13a, which is the band between 130 kDa and 180k Da. The amount of Cas13a increases with increasing incubation time after induction. When comparing the last three lanes in each of the gels, corresponding to an incubation time of 16 h after induction, it can be seen that the amount of Cas13a increases with increasing UAA concentration for pAcF/pUltra and pAcF/pEVOL. In contrast, the biggest *Lbu*Cas13a band for pAzF/pEVOL was achieved with 0.05% pAzF.

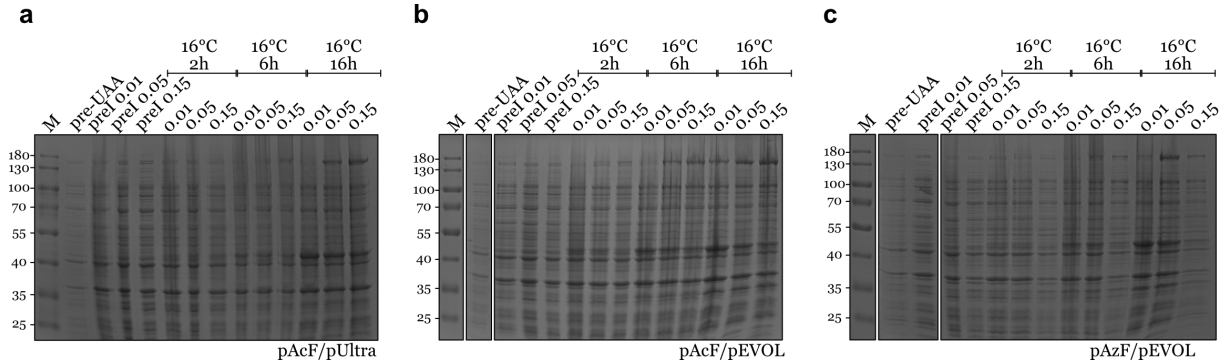


Figure 3.3-1: SDS-PAGE gels of expression tests for UAA incorporation. Samples were taken from the main culture prior addition of the respective UAA (pre-UAA). After dividing the main cultures into three flasks, samples were taken before (preI) and after induction. The numbers 0.01, 0.05, and 0.15 indicate the mass percentage of UAA that was added to each flask. These tests were done with three different aaRS-tRNA containing plasmids, being **a,** pAcF/pUltra **b,** pAcF/pEVOL, and **c,** pAzF/pEVOL.

Another observation, maybe the most important one is, that the greatest protein amount, thus the biggest *Lbu*Cas13a band is obtained with pAcF/pEVOL, an incubation time of 16 h and a pAcF amount of 0.15%. However, the expression tests were performed on a construct without the amber codon. Thus, an amber codon was cloned into the Cas13a gene at amino acid position E32. This position was chosen with a difference distance map as described in Chapter 3.4. Since E32 is close to the N-terminus of *Lbu*Cas13a, it is a suitable candidate to test the purification of the protein and to see whether truncation products are removed during HisTrap affinity chromatography or whether a size exclusion is needed to eliminate the truncated proteins. The expression and purification of *Lbu*Cas13a-E32Amb was performed with pAcF/pEVOL as described in Methods Chapter 7.2.8. The results of this expression are shown in Figure 3.3-2. The induction of protein expression worked well, as a band at around 180 kDa appeared. However, this band is less intense, when compared to the one in the test expressions (Figure 3.3-1 b). The protein amount of full length *Lbu*Cas13a in the elution of the HisTrap affinity chromatography is also small. Additionally, to the band at around 180 kDa, a very prominent band between 40kDa and 55 kDa appeared (marked with a black arrow, Figure 3.3-1 a). This fits to the truncation product, consisting of MBP, TEV cleavage site, and the first 31 amino acids of Cas13a. If this is indeed the truncation product, this should decrease in size upon incubation with TEV, which is observed in Figure 3.3-2 b.

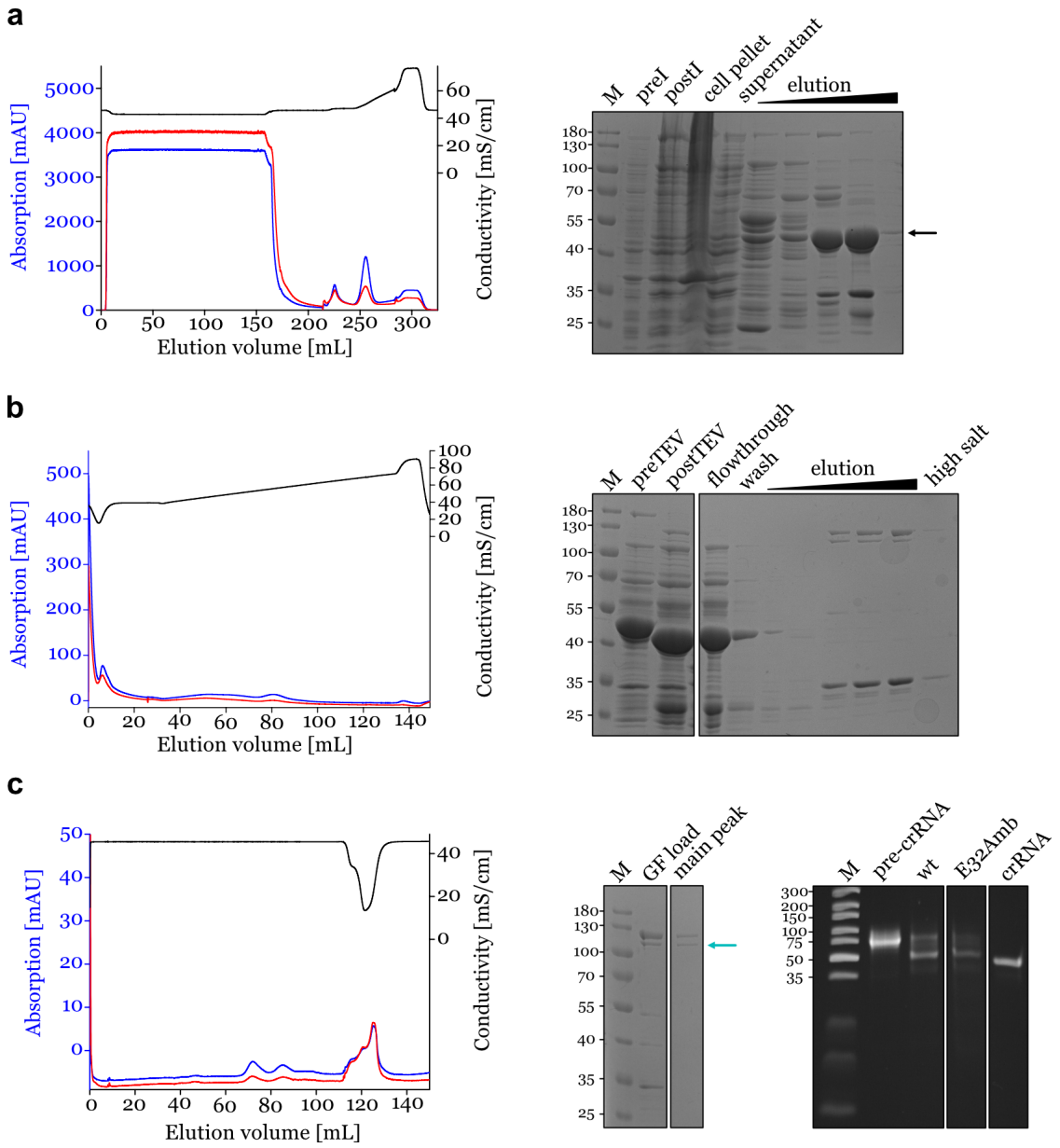


Figure 3.3-2: Purification and cleavage assay of LbuCas13a E32Amb. a, Chromatogram of the HisTrap affinity chromatography with the respective SDS-PAGE gel. preI and postI denote the samples that were collected from the main cultures prior and after induction, respectively. The black arrow highlights the protein band consisting of truncated MBP-Cas13a. **b,** Protein elution part of the chromatogram of the heparin affinity chromatography with the corresponding SDS-PAGE gel. **c,** Chromatogram of the size exclusion chromatography with the corresponding SDS-PAGE gel and an image of the pre-crRNA cleavage assay. The blue arrow highlights a protein band in the SDS-PAGE gel that supposedly origins from degradation of *Lbu*Cas13a E32Amb. The pre-crRNA cleavage assay was performed with the wt Cas13a and the E32Amb construct. The blue curves in the chromatograms show the absorption at 280 nm and the red curves the absorption at 254 nm.

During the heparin affinity chromatography, the full-length protein was eluted with impurities from small size (25-35 kDa) proteins. However, these were fully removed by size exclusion (Figure 3.3-2 c). But the protein yield was so low that almost no peak was seen in the heparin affinity chromatography and in the size exclusion

chromatography (Figure 3.3-2 c). The SDS-PAGE after size exclusion shows one protein band, directly below the one of the desired protein, being copurified with Cas13a (marked with a blue arrow in Figure 3.3-2 c). This band is absent for the first part of the purification and appears after TEV cleavage. The TEV cleavage of this construct was performed over night at 6°C and this band could be a degradation product. Thus, it can be assumed that maybe the insertion of an UAA at this specific position destabilizes the protein and makes it more prone to degradation. Nevertheless, a pre-crRNA cleavage assay was performed with this protein construct (Figure 3.3-2 c). It is important to note here, that this cleavage assay was performed before all assay optimization steps. The assay was done with a small urea gel, without L-arginine and L-glutamic acid, without tRNA, and with self-transcribed pre-crRNA. However, the wt, as well as the E32Amb constructs are both active. Because of the non-optimal experimental setup, it is difficult to determine if E32Amb is more or less active than the wt.

During the establishment of the expression and purification with UAA, a cysteine-free and active construct was developed (Chapter 3.2.2). This last method is the preferred one since the protein yield is much higher than the one with UAA.

3.4 Selection of suitable labelling sites

As mentioned above (Chapter 3.3), the position E32 was chosen as a labelling site. This position was not arbitrarily chosen, and this Chapter will describe how suitable labelling positions were specifically chosen for *Lbu*Cas13a.

But first, general considerations should be made about what a suitable labelling position is. First, a suitable labelling position is solvent exposed, since it yields higher labelling efficiencies. A buried labelling site might be not accessible for the label. Additionally, a labelling at a buried site can perturb the local protein structure through clashes with the protein. Second, a suitable labelling position is located in a rigid secondary structure element as an α-helix or a β-sheet. This is favorable since the distance distributions between two spin labels are narrower when the spin labelling sites are rigid. Third, the spin label position should be in a protein region that does not abolish protein function or folding. Four, a spin label pair should yield a measurable

distance that resides in the PELDOR regime of commonly 1.5 nm to 8 nm[191,203]. The distance distributions between two spin labels, acquired without and with ligand, should ideally not overlap and enable to track conformational changes.

For *Lbu*Cas13a, the last consideration was tackled first. Since two experimental structures were known (at that time AlphaFold was not available), the mtsslSuite server[272,273] was used to calculate a so-called difference distance map (DDM, Figure 3.4-1 a). This map is a three-dimensional map in which the x- and y- axis consist of the amino acids of two conformations of the same protein. For Cas13a, the x-axis consists of the number of amino acid residues in the binary complex, bound to crRNA, and the y-axis consists of the amino acid residues in the ternary complex, bound to cr- and target RNA. The third dimension is colored and describes the difference of the inter-spin distance between two spin label positions in two different protein structures. The lighter the color, the larger is the distance difference between two residues in two different structures. The gray areas in a DDM originate from residues that are not resolved in one specific structure, so their coordinates are unknown. The advantage of this representation is, that all distances, that are not in the PELDOR regime can be sorted out. Additionally, those residues that reside in a flexible loop, can also be sorted out manually, leaving residues in rigid secondary structure elements. The resulting modified DDM is shown in Figure 3.4-1 b. From this DDM, protein regions can be picked that present a big difference distance change (light colors) and that origin from residues in rigid protein regions. Once these are picked, the location on the protein is checked manually, regarding solvent accessibility. Lastly, the spin label location is checked for RNA binding contacts and for active sites, in which the RNA is cleaved. If one of these two features are valid, the labelling position is excluded.

Since one goal of this thesis is to get an overview of the domain movements that *Lbu*Cas13a undergoes, it is favorable to look into the domain movements of the two known structures (Figure 3.4-2 a). Already from the DDM of *Lbu*Cas13a it can be noted that the first 360-380 amino acids apparently do not change their position, since the difference in interspin distance is low (blue box in Figure 3.4-1 a). This region coincides with the REC lobe. Further, the REC lobe, seems to have a fixed position with respect to the last 250-300 amino acids of the protein, which corresponds to the linker and HEPN2 domains (orange boxes in Figure 3.4-1 a).

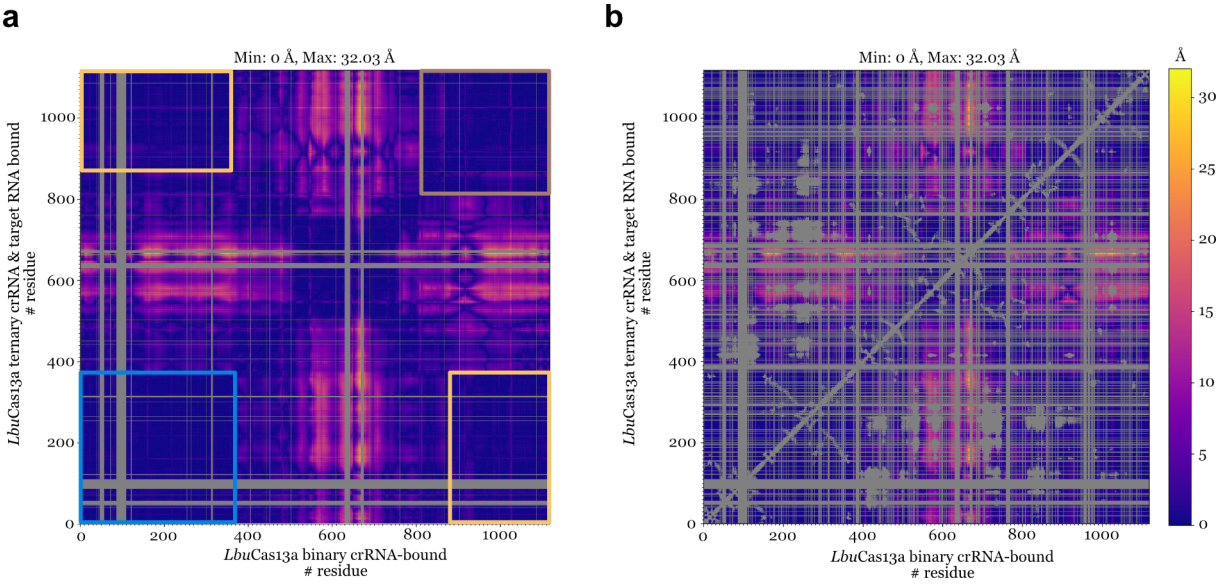


Figure 3.4-1: Difference distance maps (DDM) of *Lbu*Cas13a. a, DDM. The gray areas are residues that are not resolved in one of the two experimental structures. The blue, yellow and brown rectangles highlight protein regions in which the interspin distance changes only marginally in both structures. **b,** Same DDM as in a. Additionally, all flexible loop areas were deleted manually, resulting in more gray areas. All regions that are left consist of rigid secondary structure elements, such as α-helices or a β-sheets. The scale bar in b is also valid for a.[1]

Lastly, a third region with no distance changes can be determined that consists of the linker and HEPN2 domains, which seem to undergo no conformational change between the two known structures (brown box in Figure 3.4-1 a). This indicates that these regions could move in a rigid body fashion. To investigate the conformational change of the binary and ternary complex in more detail, a histogram representation was done (Figure 3.4-2 b). Here, the structures of the binary and ternary complex were aligned with pymol and the C_α displacements of each residue are plotted against the number of the respective residue. The C_α displacements are determined by calculating the vector length of the C_α atom of a specific amino acid in the binary structure and the same C_α atom of the amino acid in the ternary structure. This distance difference is plotted in (Figure 3.4-2 b).

[1] The DDM was only calculated to visualize the conformational change of the experimental binary and ternary complex of *Lbu*Cas13a, since AlphaFold2 (and AF3) were not available at that time. Further it is important to point out, that the DDMs were calculated with the distance distributions, hence with the *in-silico* MTSL-labelled protein, and not with the C_α-displacements.

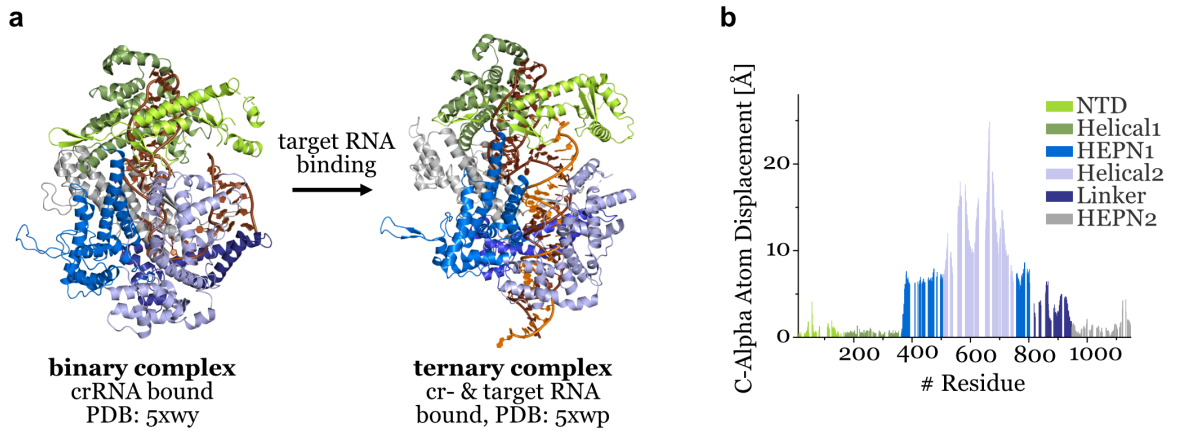


Figure 3.4-2: C_α displacements between the binary and ternary complex. a, cryo-EM and X-ray crystallographic structures of the binary and ternary complex, respectively. **b,** Displacement representation. Displacements were calculated after alignment of both structures. The color coding of the domains is the same as used before.

In this representation it becomes even more visible that the REC lobe, consisting of the NTD and helical-1 domain, as well as the HEPN2 domain seem not to change their position during target RNA addition. The movements that occur, happen mostly in the HEPN1, helical-2, and to a lower extend in the linker domain. This makes sense, since these are the regions that need to move to enable formation of the helix between the crRNA and the target RNA. A second observation is, that the extend of the intrinsic C_α atom displacement in the NTD, helical-1, HEPN1, Linker or HEPN2 domain is very similar. Thus, the path length that the C_α atoms undergo in one domain is very similar. Even though the angle and direction of each C_α atom is omitted in this analysis, this is a hint, that the movements that occur could origin from translational rigid-body movements of entire domains, since a rotation of single domains would lead to different C_α atom path lengths. This also means, that maybe the helical-2 domain undergoes a rotational movement, since there the path lengths of the C_α atoms are very different.

Since most of the domains seem to move as whole and each to a different extend, spin labels should be placed into each domain. This ensures, that each part of the protein will be sufficiently described by distance measurements. The helical-2 domain should be labelled at least twice, since its movement seems to be more complicated. Additionally, the HEPN2 domain is the only domain that will not be labelled. The reasons for this are first, the movements in this domain seem to be small (Figure 3.4-2). Second, and most importantly, there is evidence that this domain is important for pre-

crRNA cleavage and for target RNA degradation[75]. Hence, this domain is functionally very important and should not be labelled to avoid perturbations.

For *Lbu*Cas13a more than 50 combinations of labelling positions yielded distances in the PELDOR regime. These were restricted to 20 based on their solvent accessibility. Out of these 20, 7 combinations were picked, depending on the overlap of the distance distributions of the binary and ternary complex. These 7 were 32/624 measuring the distance between the REC and NUC lobe, 660/756, 756/926, 660/926, 462/660, 462/756 and 462/926, measuring distances within the NUC lobe (Figure 3.4-3). Additionally, four combinations of labelling positions were chosen to investigate the movements in the REC lobe (Figure 3.4-3 green distance measurements), which were 138/222, 63/138, 138/190, and 190/222. Further, three combinations were chosen to investigate the movements of the REC and NUC lobe to each other (Figure 3.4-3 gray distance measurements), 190/756, 190/926, and 138/926. In total, from these 14 different spin labelling site combinations, 11 constructs were obtained that were functionally active. These were used for PELDOR distance measurements (Chapter 4).

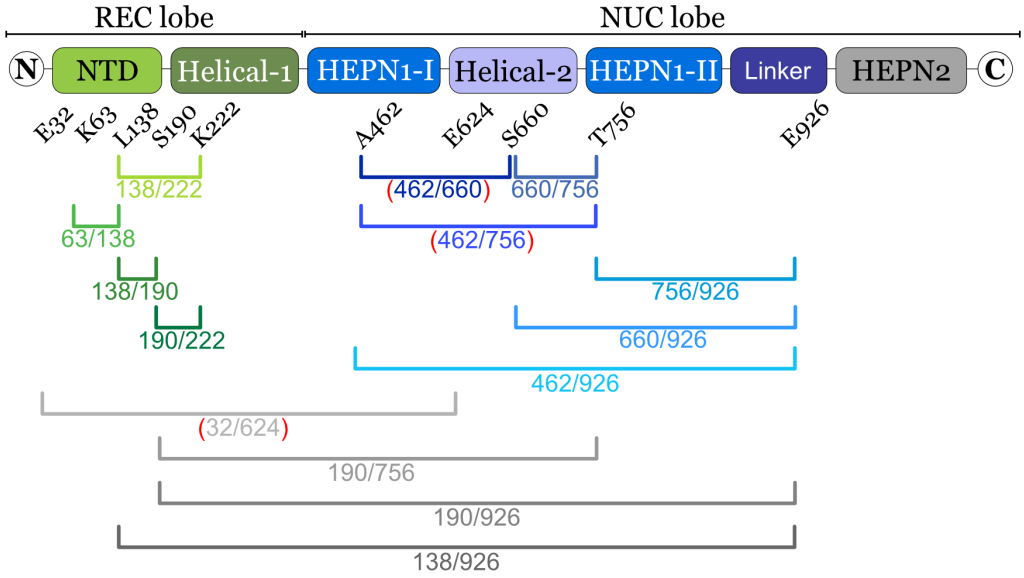


Figure 3.4-3: Spin label positions chosen for PELDOR measurements. The domain architecture of *Lbu*Cas13a is shown on top with selected labelling positions. The spin label combinations that are tested for distance measurements are shown below, as bars. The green label combinations lead to distance measurements within the REC lobe, the blue combinations to distance measurements within the NUC lobe, and the grey combinations to measurements between the REC and NUC lobe. The spin label combination 32/624 resulted in an instable construct and was discarded (Chapter 3.5). Also 462/660 and 462/756 were discarded since these resulted to be inactive regarding RNA cleavage (Chapter 3.5). All discarded constructs are highlighted by red brackets.

3.5 Preparation of active double labelled protein constructs

As described above, several suitable labelling positions were found. The first construct that was cloned, expressed and purified contained E32 and E624 as labelling positions. However, no protein could be purified. As seen in the SDS-PAGE gel sample after cell induction, in Figure 3.5-1, the protein was overexpressed. But in comparison to other purifications (e.g. Figure 2.1-7 a), the expression rate seems to be lower. Additionally, it seems that the protein is either not stable, or that it is expressed in inclusion bodies. This is difficult to distinguish, since the SDS-PAGE sample of the cell pellet is very blurred. What can be seen is, that almost no protein is present in the supernatant, in the elution of the nickel affinity chromatography and in the elution of the heparin affinity chromatography (Figure 3.5-1).

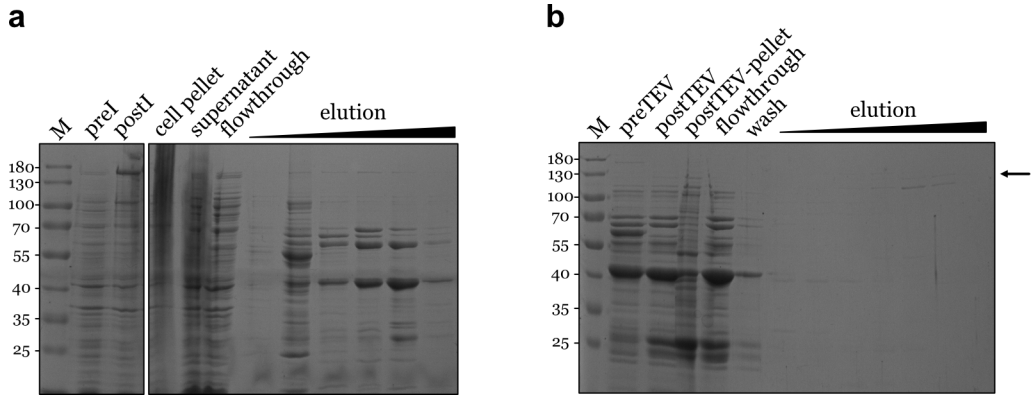


Figure 3.5-1: Purification of the Cas13a construct E32C E624C. a, SDS-PAGE gel with samples from the nickel affinity chromatography. **b,** SDS-PAGE gel with samples from the heparin affinity chromatography. The arrow on the right highlights the protein band, which can be barely seen.

After this result, the focus was set on other spin labelling positions to test whether the low protein stability was caused by this specific protein construct. Therefore, three cysfree-V constructs were cloned, being A462C S660C (462/660), S660C T756C (660/756), and S660C E926C (660/926). These were expressed and purified, and the results of the size exclusion and the corresponding SDS-PAGE gels are shown in Figure 3.5-2. As seen in Figure 3.5-2 a-c, all three purifications were successful. So, the reduced protein stability of 32/624 (Figure 3.5-1) was probably caused by the mutations at positions 32 and 624. Additionally, the three new protein constructs show no other substantial impurities in their respective SDS-PAGE samples. The protein

construct with the highest yield was chosen for MTSL labelling tests, which was the 462/660 construct.

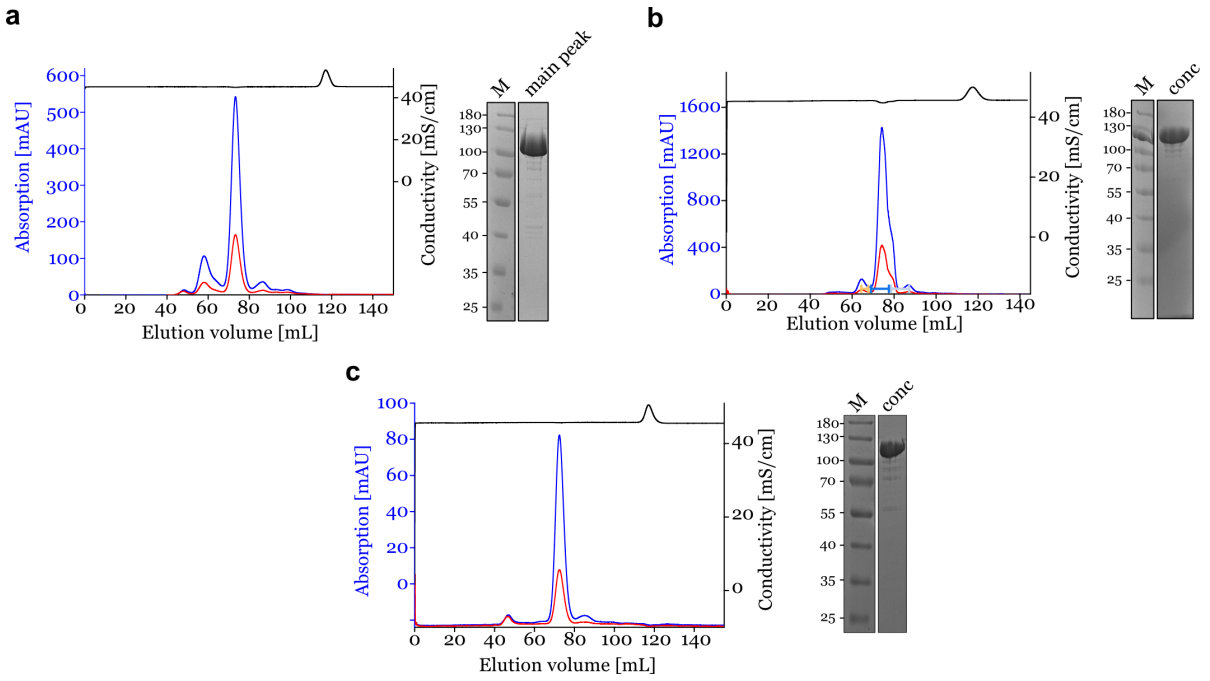


Figure 3.5-2: Purification results of the protein constructs 462/600, 550/756, and 660/926. The chromatogram on the left belongs to the size exclusion run with a HiLoad 200 16/600 column. The corresponding SDS-PAGE gel of the concentrated protein (conc) or of the main peak is shown on the right. The constructs shown are in **a,** 462/660, in **b,** 660/756, and in **c,** 660/926. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm.

Different labelling procedures were tested, in which the temperature and the length of protein-MTSL incubation was varied. Generally, the spin labelling procedure consisted of a 30 min incubation of the protein with TCEP, to reduce existing disulfide bridges between cysteines. TCEP was removed by a PD10-gravity column and the protein was eluted into a falcon containing a 20-fold molar excess of MTSL per cysteine. The incubation was done for 2 h at RT, for 16 h at RT, and for 16 h at 4°C. After incubation, excess MTSL was removed via a PD-10 gravity column, followed by protein concentration through a centrifugal concentrator with an MWCO of 100.000 kDa. The results of all three labelling tests are shown in Figure 3.5-3. In general, all three *cw*-EPR spectra show immobile spin label and the labelling efficiencies, calculated from a spin count, are very similar. Maybe small amounts of free spin label are present in the samples incubated for 2 h at RT and for 16 h at 4°C. This results in small indentations in the spectrum, highlighted with arrows in Figure 3.5-3 a.

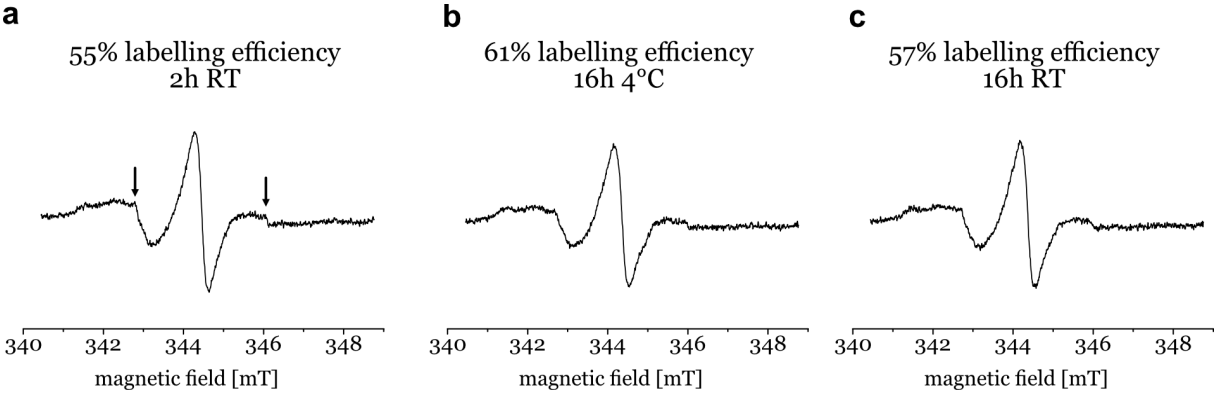


Figure 3.5-3: *cw*-EPR spectra of different labelling trials with the 462/660 construct. Labelling is performed with an incubation of **a,** 2 h at RT. **b,** 16 h at 4°C. **c,** 16 h at RT. The labelling tests were performed without L-arginine and L-glutamic acid.

But the amount of free spin label is so small that it can be neglected. The labelling test that yielded the highest spin labelling efficiency is the one performed over night for 16 h at 4°C. This method was used for further *cw*-EPR measurements. Additionally, to fit the background, the sweep width of the *cw*-EPR spectrum was increased. The addition of the amino acids L-arginine and L-glutamic acid (see Chapter 3.6) during labelling, in combination with the new labelling procedure resulted in a labelling efficiency of 90% (Figure 3.5-4). The constructs 462/660, 660/756, and 660/926 were labelled with the optimized protocol, using an incubation of 16 h and 4°C. As seen in Figure 3.5-4, the labelling efficiency was 90-100%. Additionally, the labelled protein constructs were subjected to a pre-crRNA and a target RNA cleavage assay. The labelled 462/660 construct is active regarding the pre-crRNA cleavage, even though the cleavage seems to be reduced, compared to the wt. Additionally, the target cleavage seems to be drastically reduced (Figure 3.5-4 a). Thus, this construct was excluded from further experiments.

In the *cw*-EPR spectrum of the labelled 660/756 construct, free spin label is still present. Labelling trials were performed in which the protein solution was purified twice via PD10, but the amount of free spin label could not be decreased (data not shown). However, the amount of free spin label did not inhibit pre-crRNA and target RNA cleavage (Figure 3.5-4 b). But during PELDOR sample preparations it was found that this construct fully precipitated, as soon as pre-crRNA or crRNA was added, indicating a reduced protein stability *in-vitro*. It was decided not to use this construct for PELDOR measurements with RNA.

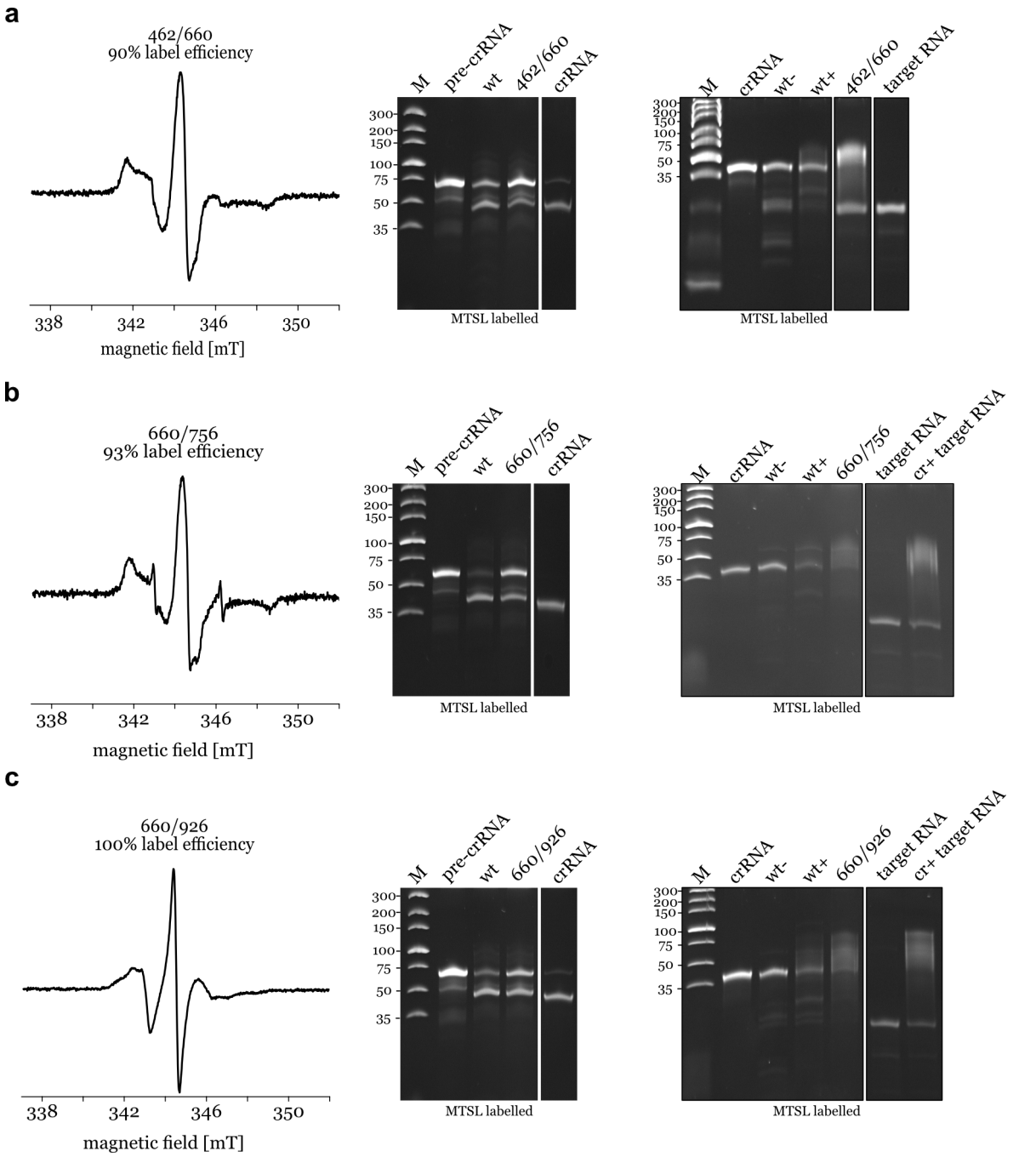


Figure 3.5-4: *cw*-EPR spectra, pre-crRNA, and target RNA cleavage assays of MTSL-labelled constructs. The spectra and the cleavage assays were performed with L-arginine and L-glutamic acid. The spin labelling efficiency is given for each spectrum individually, on top. This data is shown for the protein constructs labelled at positions **a,** 462/600, **b,** 600/756, and **c,** 600/926.

The third labelled construct, finally, seemed to almost behave as the wt. No free spin label can be seen in the *cw*-EPR spectrum, the pre-crRNA cleavage activity is only slightly reduced compared to the wt, and the target RNA seems to be fully cleaved (Figure 3.5-4 c).

In summary, a labelling procedure was established, and the final protocol is described in the Methods Chapter 7.3.1. This protocol was applied to all protein constructs, showed in Figure 3.4-3. Each of these constructs will be introduced and explained in Chapter 4.

As seen in this Chapter, double cysteine mutations or labelling can lead to inactive protein constructs regarding pre-crRNA or target RNA cleavage. Other constructs can become instable, resulting in protein precipitation. The topic of the last part of this Chapter will be how to minimize precipitation.

3.6 Minimizing amorphous structure formation during RNA addition

All PELDOR measurements of the *apo* protein constructs were set up with a concentration of 50 µM protein. But during the preparation of the first PELDOR samples with crRNA it was noticed that upon addition of RNA, the solution became very viscous and turbid. It was hypothesized that the protein or/and the RNA precipitated. Upon lowering the protein concentration to 15 µM this phenomenon still occurred, but to a lower extend, depending on the double cysteine construct that was used. In the context of Catrin Allar's master thesis, it should be tested what this precipitant could be and PELDOR was measured of a protein construct that showed a more pronounced precipitation, which was the double labelled 462/756 construct (Figure 3.6-1 a, b). In later studies this construct resulted to be inactive. However, it was shown that the precipitation could be mostly attributed to labelled protein. Since the distance distribution of the *apo* and the crRNA bound complex are very similar, it was difficult to determine which of both states predominantly precipitated (Figure 3.6-1 a). To avoid the formation of a turbid solution, several additives were tested, these were tRNA, trehalose and amino acids (data in Figure 3.6-1 and Figure 3.6-2 are recorded by Catrin, data shown in Figure 3.6-3 were recorded by me). First, total yeast tRNA was added. Since tRNA decreased unspecific RNA binding in cleavage assays (Chapter 2.3), this may help to reduce macroscopic aggregation during PELDOR sample preparation. But the opposite happened (Figure 3.6-1 b).

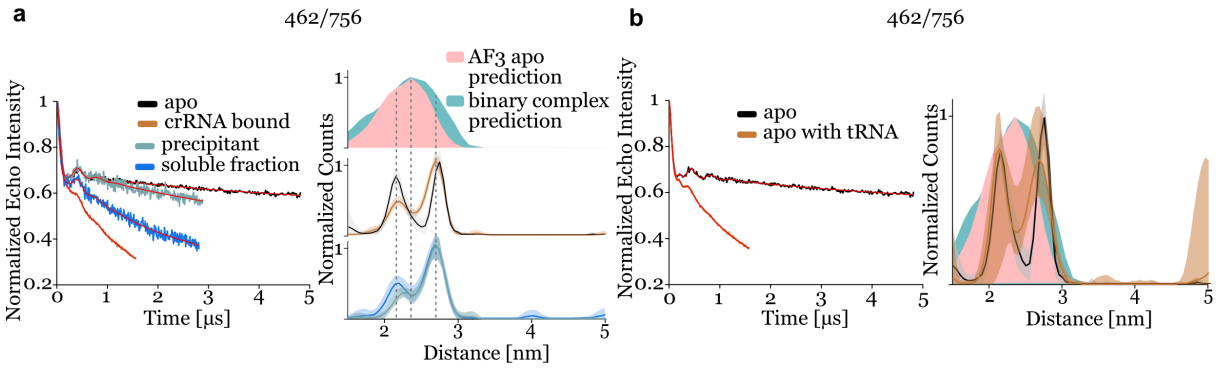


Figure 3.6-1: Investigation of PELDOR sample precipitation. PELDOR time traces and distance distributions of **a,** labelled 462/756 construct. Precipitation occurred during sample preparation, the precipitant and the soluble fraction were measured separately. The predictions of distance distributions on the basis of the AF3 prediction and on the experimental structure of the binary crRNA-bound complex are shown as colored areas. The 95% confidence intervals are drawn as shaded areas in the same color of the distance distributions that it belongs to. **b,** *apo* 462/756 with and without tRNA. The time trace analysis with backgrounds is shown in the Appendix in Figure 8.3-1. All distance distributions were calculated using CDA.

Addition of tRNA increased turbidity, leading to a shorter phase memory time T_m and consequently to a shorter measurable PELDOR time trace. The reason for this is that the environment of the spin label in aggregated proteins is not deuterated, compared to the non-aggregated protein in deuterated solvent, leading to a shorter phase memory time T_m[192]. However, the distance distribution for the *apo* 462/756 construct with and without tRNA are very similar and the impact of this turbidity on the distance distributions is low. This is not the case for the crRNA bound PELDOR measurements with and without tRNA of the 190/756 construct (Figure 3.6-2 a).

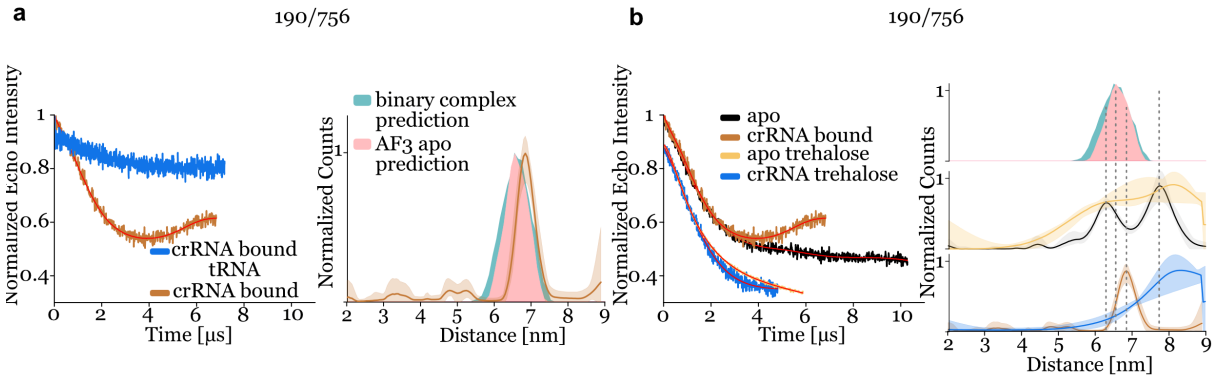


Figure 3.6-2: Investigation of tRNA and trehalose to reduce precipitation. PELDOR time traces and distance distributions of **a,** labelled 190/756 construct with crRNA. As in Figure 3.6-1, the influence of tRNA was tested. The signal to noise ratio of time trace of the sample with tRNA was so low that no distance distribution could be calculated. **b,** Labelled 190/756 construct in its *apo* and crRNA bound states, with and without trehalose. The time traces of the samples with trehalose are shifted on the y-axis by 0.1. The predictions of distance distributions on the basis of the AF3 prediction and on the experimental structure of the binary crRNA-bound complex are shown as colored areas. The 95% confidence intervals are drawn as shaded areas in the same color of the distance distributions that it belongs to. The time trace analysis with backgrounds is shown in the Appendix in Figure 8.3-1. All distance distributions were calculated using CDA.

The distance distribution belonging to the sample without tRNA presents a most probable distance, which fits to the predicted distribution from the binary complex (blue shaded area in Figure 3.6-2 a). The signal to noise ratio of time trace belonging the sample with tRNA is too low for the analysis with ComparativeDeerAnalyzer (CDA) and the time trace shows no oscillations. This could be an indication that the protein almost entirely precipitated and that generally the proteins' structure is perturbed.

The second additive that has been tested was trehalose. Trehalose is produced in unicellular organisms when cellular stress levels rise. It is used by cells to prevent protein degradation, it stabilizes proteins preventing aggregation, and it was used for cryopreservation[274]. Thus, it was hoped that this would help, especially in combination with d_8-glycerol as an additional cryoprotectant. Two samples of the 190/756 construct containing trehalose were prepared with and without crRNA (Figure 3.6-2 b). By comparing the *apo* samples with and without trehalose first, it can be noted that the most probable distances in the distributions are similar. The two peaks are more defined in the sample without trehalose and broader in the sample with trehalose. This could be an indication that the protein adopts more and different conformations with trehalose than without trehalose. However, the time trace with trehalose is significantly shorter than without trehalose, which has an effect on the reliability of the distance distribution. The crRNA bound sample without trehalose shows one distinct peak, with a most probable distance at 6.9 nm (brown trace, Figure 3.6-2 b). This distance fits to the predicted distance on the basis of the experimental structure of the binary complex. As for the samples without RNA, the time trace length of the construct with crRNA and trehalose is 2 µs shorter than the one without trehalose. The time trace of the crRNA bound construct with trehalose is too short to observe any possible oscillations and the resulting distance distribution with a peak at around 8 nm should not be interpreted. What can be said at this point is, that trehalose did not increase protein solubility and that it has an impact on protein conformation.

The third test was to use a different cryoprotectant. It was shown that cryoprotectants can have an impact on the protein's conformation[119] and that different cryoprotectants should be tested since these can have an impact on weak protein-protein interactions and on the proteins state during the freezing process[214]. For this thesis, d_8-glycerol was initially used during sample preparation since it was also present in the buffers for protein purification. But d_4-ethylene glycol (d-EG) is also a common cryoprotectant.

The comparison of using different cryo-protectants is shown in Figure 3.6-3 a.

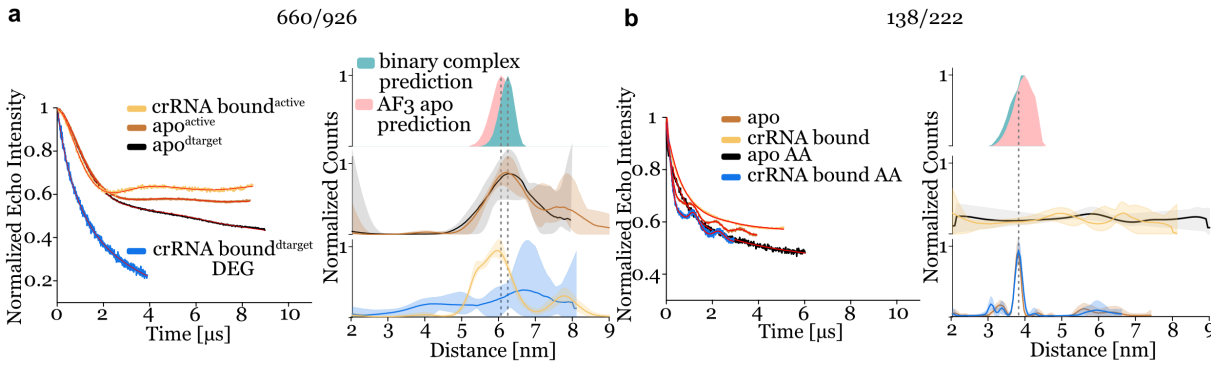


Figure 3.6-3: Influence of different cryoprotectants and amino acids on the formation of precipitants during PELDOR sample preparation. PELDOR time traces and distance distributions of **a,** the labelled 660/926 construct in its active and dtarget form, with and without crRNA. The influence of d_4-ethylene glycol (DEG) is analyzed. **b,** The labelled, cleavage active 138/222 construct with and without the amino acids L-glutamic acid and L-arginine (AA). The time traces of the samples with AA are shifted on the y-axis by 0.1. The predictions of distance distributions on the basis of the AF3 prediction and on the experimental structure of the binary crRNA-bound complex are shown as colored areas. The 95% confidence intervals are drawn as shaded areas in the same color of the distance distributions that it belongs to. The time trace analysis with backgrounds is shown in the Appendix in Figure 8.3-1. All distance distributions were calculated using CDA.

Two different constructs are shown in Figure 3.6-3 a, the active and the target RNA cleavage deficient dtarget form of 660/926. The dtarget modifications (R1048A H1053A) should have no impact on the structure compared to the active form. However, only experimental structures of the dtarget construct are known and no structure of the active form of *Lbu*Cas13a is available. Our data suggest, at least for the region of this double labelled construct, that the structures are very similar (brown and black curves with the blue shaded area in Figure 3.6-3 a). Both show the same most probable distance of 6.1 nm, and the conformational flexibility (the width of the distance distribution) is also very similar. Upon addition of crRNA to the active construct, the most probable distance is slightly shifted to shorter distances and a shoulder appears at 5.4 nm. The sample of the dtarget construct with crRNA and d-EG (blue time trace, in Figure 3.6-3 a) showed a much faster T_m, hence only a shorter time trace could be measured. This is explained by the formation of a significantly more turbid solution, than the crRNA sample without d-EG and with d_8-glycerol. Despite of a deterioration of the turbidity, the distance distribution between both spin labels is different. The most probable distance is shifted to longer distances and the background validation led to distance peaks that almost completely vanished. Additionally, the time trace is similar to an exponential decay. Taken together, this is an indication that the

distance distribution should rather be interpreted as a broad distribution, that no clear distance between spin labels can be measured, and that the structure of the protein is altered.

The fourth and last additives tested, were the amino acid combination L-glutamic acid and L-arginine (AA), in a 1:1 ratio. Golovanov et al.[266] found out that the addition of 50 mM of both amino acids to the protein buffer can decrease the formation of precipitant and aggregation and increase protein stability. Additionally, the long-term protein stability was increased and they stated that protein-protein and protein-RNA interactions were not adversely affected by the presence of these amino acids[266]. To test, if these additives have a positive impact on Cas13a stability, the double labelled construct used for this was the active 138/222 construct (Figure 3.6-3 b). Both *apo* samples, with and without AA showed a time trace without oscillations and a very broad distance distribution. When the crRNA was added to the samples with and without AA, the sample with AA remained almost transparent, while the sample without AA became turbid. After centrifugation to deposit the precipitation, also the sample with AA contained some precipitant on the bottom of the reaction tube, but to a smaller degree than the sample without AA addition. Therefore, AA addition drastically increased solubility. The distance distributions of both samples containing crRNA show the same peak at 3.8 nm. In contrast to the other parameters tested before, the *apo* and the crRNA bound samples with and without AA lead to the same result. This was also the case for other double labelled protein constructs (data not shown).

After this positive result, both amino acids were added to the cleavage buffer for cleavage assays, to the PELDOR buffer for PELDOR sample preparation, and to the *cw*-EPR buffer, for protein labelling. It is noted that no protein precipitation was observed during cleavage assays, during AFM measurements (Chapter 4.4.2) and during any other experiments in which pre-crRNA, crRNA and/or target RNA was added to the protein. The only exception was PELDOR sample preparation, in which sometimes minor precipitations occurred. This is presumably due to the micromolar protein and RNA concentration. Since these were minor precipitations, they were tolerated and the same buffer composition was used for all experiments, for comparable experimental conditions.

3.7 Discussion and Conclusion

This Chapter describes how double labelled *Lbu*Cas13a constructs were obtained that retained pre-crRNA and target RNA cleavage activity and that could be used for structural and dynamic studies via PELDOR spectroscopy.

The wt *Lbu*Cas13a was labelled with MTSL, and it was seen that the native cysteines were labelled, despite their low solvent accessibility. The cysteines were exchanged with alanine and serine, but the cleavage of both protein constructs was almost completely abolished. Upon mutational analysis it was found that C348 was particularly important for pre-crRNA cleavage. The mutation C348A in combination with C293A further reduced pre-crRNA cleavage. So far it is unclear why and how these amino acids are important for pre-crRNA cleavage. Through an interaction diagram analysis with Ligplot it was seen that C348 forms hydrogen bonds with other residues. This could be a first indication, that this cysteine stabilizes the proteins' active structure. It would be interesting to determine the structure of this construct, to potentially determine where the pre-crRNA cleavage site exactly is and to develop a protein construct that even enhances the pre-crRNA cleavage activity.

To develop a cysteine-free protein construct that remains active, C348 was mutated to valine, tyrosine, and lysine. Out of these, the valine modified construct retained pre-crRNA cleavage activity. The program used for this is dezyme, which provides information on protein stability based on the proteins' structure. Even though this approach worked, the ΔΔG values that are given should be carefully used. They provide rough indications of the influence of a specific amino acid at a specific position. But dezyme does not account for the flexibility of the proteins' structure, for the different buffer conditions (pH, viscosity etc.), and it does not include homology comparisons and evolutionary information. Further, dezyme does not include information on an active site or a ligand binding site. Taking e.g. an active site, often the amino acids that make up this site are not the amino acids that are the most stable at these positions[269]. Thus, dezyme is a useful tool but the results should be evaluated with care and need to be experimentally tested.

During the preparation of the active cysteine-free construct C293A C348V C1141A, another approach was tested, which was the insertion of unnatural amino acids. The E32Amb construct was successfully expressed and purified, but with small protein

yields compared to the protein without UAA. A pre-crRNA cleavage assay confirmed this construct to be active. Thus, the expression and purification of *Lbu*Cas13a with UAA was successfully established as a second option for generating a protein construct that can be used for PELDOR measurements. Since an active cysteine-free construct was successfully found, which did not require amber codon suppression and which was purified in higher yields, the labelling via UAA was not further pursued.

Different labelling positions were chosen for double labelled constructs. However, the insertion of two new cysteines needs careful experimental evaluation. Some double cysteine constructs showed a decrease in stability (E32C E624C), which led to full protein degradation during purification. Other double cysteine constructs were stable but became inactive regarding pre-crRNA or target RNA cleavage. This shows how important activity tests are and that the integrity of the proteins needs to be validated. Out of the double cysteine constructs tested, there were several that were stable and active. These were successfully used for structural studies with PELDOR spectroscopy, discussed in the following Chapter.

During PELDOR sample preparation with crRNA it was seen that the labelled protein precipitated, probably mainly because of the high protein and RNA concentrations. To overcome this issue, several additives were tested. These were total yeast tRNA, trehalose, D-EG as cryoprotectant, and L-arginine and L-glutamic acid. It was found that the first three additives did not reduce precipitation. But addition of L-arginine and L-glutamic acid substantially reduced protein precipitation and, depending on the protein construct, completely inhibited precipitation. Furthermore, the PELDOR data suggested that the addition of these amino acids did not change the proteins' structure or the distribution of protein conformations.

Chapter 4

Conformational studies of *Lbu*Cas13a

The previous Chapters lay the biochemical foundation for the subsequent measurements. As mentioned above, it is important to analyze each protein construct separately regarding its purification, its spin labelling, and its pre-crRNA and target RNA cleavage activity. For each protein construct that will be shown in the following Chapter, the last step of the purification, the SDS-PAGE sample of the concentrated protein construct, the *cw*-EPR spectrum, and one of the three replicates of the pre-crRNA and target RNA cleavage assays can be found in the Appendix. All PELDOR samples that are used from this point on were prepared with buffers containing L-arginine and L-glutamic acid.

4.1 Recognition lobe

4.1.1 The REC lobe changes its conformation and rigidifies upon RNA binding

*Lbu*Cas13a is composed of the recognition lobe (REC lobe) and the nuclease lobe (NUC lobe). Only two experimental structures of Cas13a are available, the crRNA bound structure and the cr- and target RNA bound structure. Additionally, to these structures, the AF3 predictions of the *apo* and the pre-crRNA bound complex were calculated and the results are shown in Figure 4.1-1.

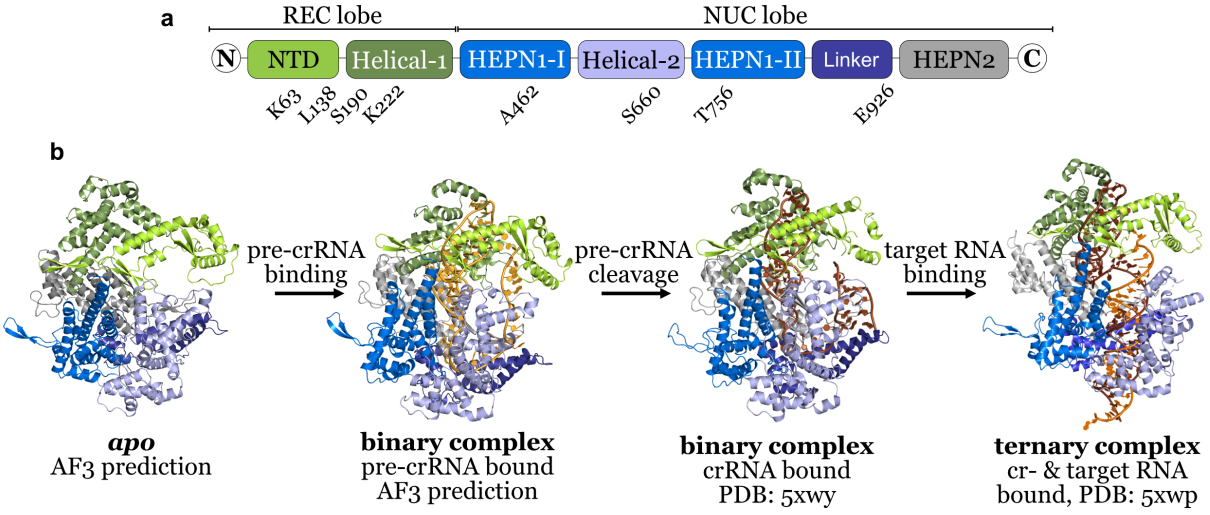


Figure 4.1-1: Molecular architecture of *Lbu*Cas13a conformations. a, Domain architecture of *Lbu*Cas13a with labelling sites highlighted at the bottom. **b,** Conformational pathway with predicted and experimental structures[87].

When looking at experimental and at predicted structures of the various states of *Lbu*Cas13a, the REC lobe does not change its conformation from the *apo*, via the pre-crRNA-bound, crRNA-bound, and the cr- and target RNA bound complex, suggesting a rigid REC lobe. In contrast, from the known structures of other Cas13a proteins (Figure 1.6-3), it can be noted that in four structures, parts or the complete NTD in the REC lobe are missing. These construct modifications can be an indication of an increased flexibility in the REC lobe, since both used methods for structure determination, X-ray crystallography and cryo-EM, work best with rigid proteins. Thus, the questions arise, if the REC lobe in solution is rigid or flexible, how the conformation of the REC lobe changes throughout the functional pathway, and how structurally RNA binding works.

To answer the first two questions, the REC lobe structure was investigated with PELDOR spectroscopy in each state, the *apo*, the pre-crRNA, the crRNA and the cr- and target RNA bound states. To inhibit cleavage of the pre-crRNA and to trap the protein in a pre-crRNA bound state during data acquisition, a double labelled dprecr construct containing R1079A was cloned. To inhibit cleavage of the target RNA and to trap the cr- and target RNA bound complex, a double labelled dtarget construct containing R2048A H1053A was cloned (see Chapter 1.6.5). Biochemical validation data is found in the Appendix 8.2.1.

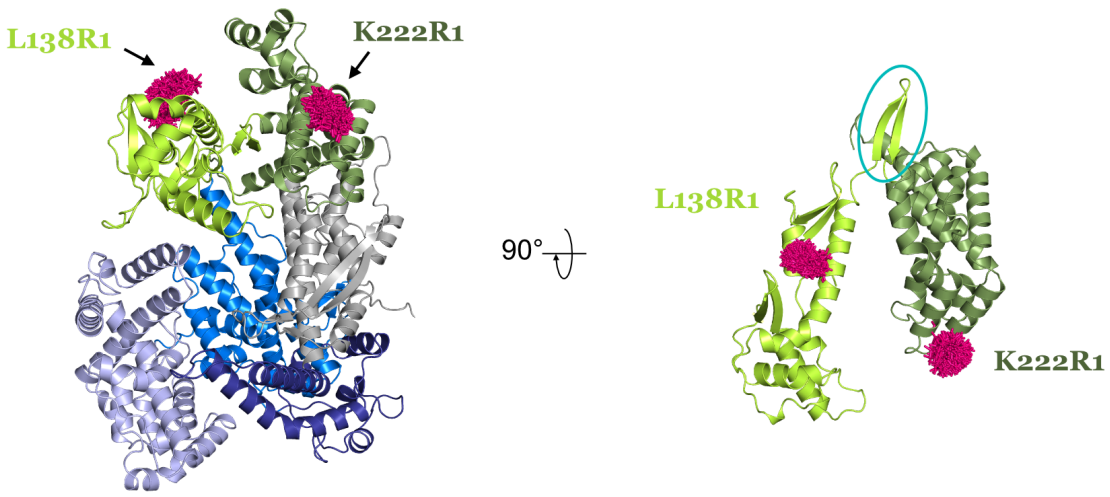


Figure 4.1-2: AF3 structure of *Lbu*Cas13a with focus on the REC lobe. The full AF3 structure of *apo Lbu*Cas13a with label rotamers in magenta at positions 138 and 222 are shown on the left. The REC lobe consisting of NTD (light green) and helical-1 domain (olive green) is shown from the top view. The label rotamers are highlighted in magenta and the β-sheets that could act as a hinge region between both domains are highlighted by a blue oval.

By looking at the structural connection between the NTD and helical-1 domain (Figure 4.1-2) it was noted that two β-sheets in between of both domains could act as a hinge region, leading to a seashell-like opening and closing motion. Thus, a construct was designed to detect if a closed, an open or even multiple REC lobe states can be observed. The labelling positions that were chosen for this were L138 and K222, which are located in the NTD and in the helical-1 domain, respectively (Figure 4.1-2).

The PELDOR results of $138/222^{\text{apo}}_{\text{active}}$ (Figure 4.1-3, top), show a very broad distance distribution without a defined distance between both spin labels. This indicates that the protein region studied here seems to be very flexible and that the protein does not adopt a preferred conformation in this protein region. In fact, it adopts a variety of conformations in frozen solution, that are more or less equally probable. AF3 predicts a structure with an interspin distance of around 3.9 nm (Figure 4.1-3, top, magenta distribution). Together with many other conformations, this conformation is also present in the experimental distribution. Generally, an enhanced flexibility could explain why the NTD was partially or completely truncated in the *apo* structures that were experimentally solved from homolog Cas13a proteins, presumably to facilitate crystallization. Additionally, this flexibility fits to what is known from other Cas nucleases, as Cas12a, where a high extend of conformational flexibility between the helical domain and the RuvC domain was seen in cryo-EM[275]. Through crRNA addition, a conformational change is driven in Cas12a, leading to a more compact structure[275]. This is also the case for Cas13a, the protein dynamics drastically change upon addition of the pre-crRNA to the dprecr construct. In contrast to the time trace of $138/222^{\text{apo}}_{\text{active}}$, the one of $138/222^{\text{pre-crRNA}}_{\text{dprecr}}$ shows well defined oscillations, an indication of a narrow distance distribution. The distance distribution of $138/222^{\text{pre-crRNA}}_{\text{dprecr}}$ has a single, most probable distance at 3.8 nm, which fits well to the AF3 prediction of the pre-crRNA bound complex. Interestingly, this conformation seems to be similar to the one of the *apo* AF3 prediction. As discussed in Chapter 1.7.3, AF2 predictions sometimes correspond to the bound structure and not to the *apo* structure, which is a known AF2 feature[160,161]. Our results indicate that this could also be valid for AF3.

As a next step, we added the crRNA to the *apo* protein and measured PELDOR, shown in the third measurement. The most probable distance of $138/222^{\text{crRNA}}_{\text{active}}$ remains at

3.8 nm, indicating no structural rearrangements. Also, upon addition of target RNA, the most probable distance of $138/222^{\text{cr-target RNA}}_{\text{dtarget}}$ does not change. Both PELDOR distance distributions of $138/222^{\text{crRNA}}_{\text{active}}$ and $138/222^{\text{cr-target RNA}}_{\text{dtarget}}$ fit to the predicted ones based on the experimental structures. It is worth mentioning that the peaks at 1-3 nm and 5 nm in the distance distributions analyzed with Tkh, vanish when analyzed with CDA. But the background of the time traces fitted by CDA is different (see Appendix, Figure 8.3-2). This suggests that these smaller peaks could origin from the background.

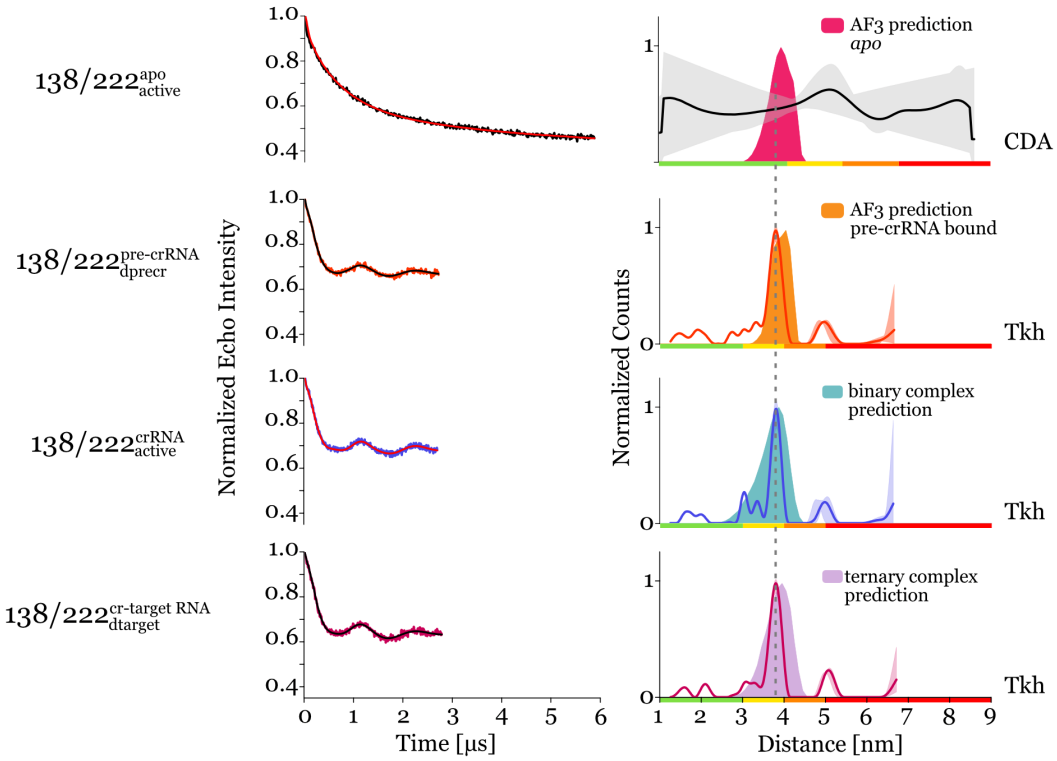


Figure 4.1-3: PELDOR results for the protein constructs labelled at 138/222. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through Tikhonov regularization (Tkh) or through the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models, the binary cryo-EM structure, and the ternary X-ray crystallographic structure are shown as shaded distributions.

Taken together, the PELDOR data show that the REC lobe is very flexible, when no RNA is bound. Once the pre-crRNA binds to *Lbu*Cas13a, the conformation of this protein region strongly rigidifies into one conformation. This conformation does not change anymore through the entire functional pathway. This fits to the known crRNA

bound cryo-EM and to the cr- and target RNA bound crystal structure, as seen from the predicted distributions (Figure 4.1-3, blue and purple shaded distributions).

4.1.2 REC lobe flexibility

From the PELDOR measurement of $138/222^{apo}_{active}$, the question arises, where the enhanced flexibility of the REC lobe origins. One possible reason for the flexibility of $138/222^{apo}_{active}$ could be a rigid body movement of the NTD and the helical-1 domain towards each other. Another possibility could be an enhanced flexibility due to intrinsically disordered protein (IDP) domains or regions, that become ordered once the pre-crRNA binds.

To investigate the origin of the flexibility, three additional double labelled REC lobe constructs were designed (Figure 4.1-4). Two of them include two labelling positions in one domain, to address the intrinsic domain flexibility. These are 63/138, where both positions are located in the NTD and 190/222, where both positions are located in the helical-1 domain. The third construct was labelled at the positions 138/190. These two positions are located as near as possible to the two β-sheets, which potentially act as a hinge region (Figure 4.1-4). This strategic placement of the spin labels enables to examine the flexibility of this hinge region part only, minimizing the influence of both domains. The biochemical validation of these constructs is shown in the Appendix in Figure 8.2-2. In contrast to the flexibility between the NTD and helical-1 domain shown above (Figure 4.1-3 and Figure 4.1-4 b), the NTD only distance distribution with spin labels at positions 63 and 138 is very narrow. The experimental distance of 3.8 nm fits to the AF3 prediction with a slightly longer distance of 4.1 nm.

In contrast to the rigid NTD, the distance distribution of the $190/222^{apo}_{active}$ construct, in which both spin labels are located in the helical-1 domain, is broad. However, the broadness of this distribution is not as prominent as in the $138/222^{apo}_{active}$ construct. The most probable distance of the distribution based on the AF3 prediction fits to the most probable distance of the experiment. Structurally, the dynamic behavior of the helical-1 domain could origin from the kink in this domain (Figure 4.1-5 a, red dashed line). But the dynamics of just that kink alone (Figure 4.1-4 d) are not explaining the enhanced flexibility seen in $138/222^{apo}_{active}$ (Figure 4.1-4 b).

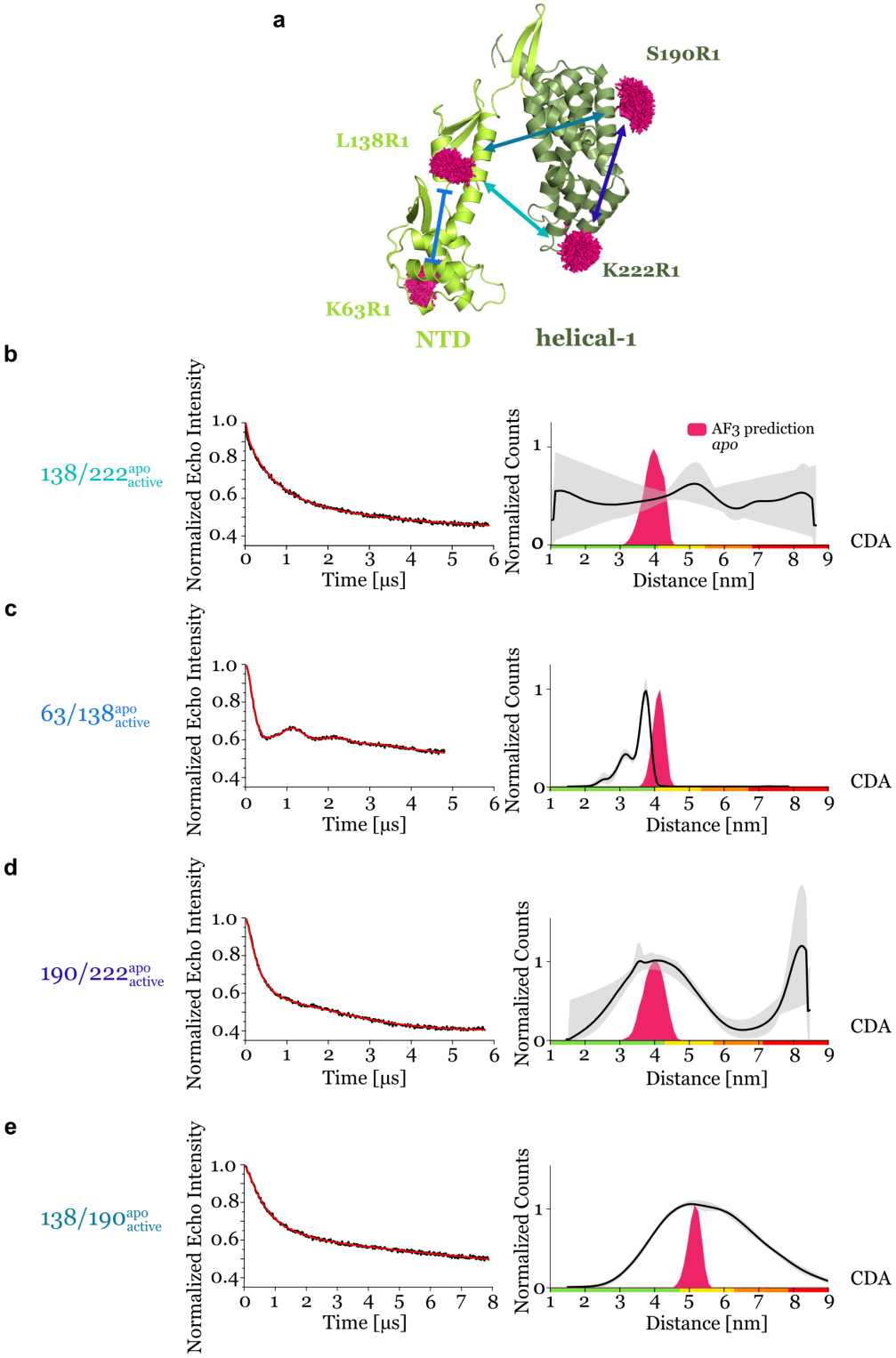


Figure 4.1-4: Investigation of the domain flexibility in the REC lobe. a, REC lobe of the *apo* AF3 prediction, with MTSL rotamer clouds in magenta. The light green domain is the NTD and the olive-green domain is the helical-1 domain. The arrows indicate flexibility between two protein regions and a bar denotes rigidity between two protein regions. **b-e,** Time traces and distance distributions of different double labelled protein constructs. All were analyzed with CDA. The distance distributions on the right contain the 95% confidence interval as shaded area in gray and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 model are shown as shaded distributions.

Consequently, to investigate if the initially observed REC lobe flexibility between 138/222 additionally originates from the possible hinge region between the NTD and helical-1 domain a fourth construct was measured, $138/190^{apo}_{active}$. The PELDOR measurement of this construct shows a broad distance distribution. The most probable distance of the experimental data fits to the most probable distance of the AF3 prediction. This measurement confirms a second degree of flexibility, that probably comes from the hinge region between both REC lobe domains, suggesting an opening and closing motion (Figure 4.1-5 b).

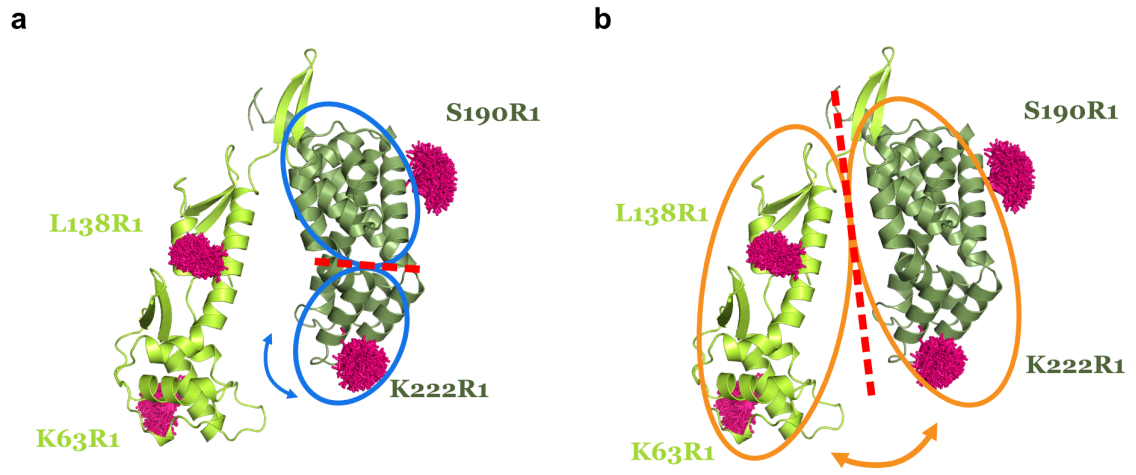


Figure 4.1-5: Possible domain movements in the REC lobe. The shown structure is the REC lobe of the AF3 structure of *apo Lbu*Cas13a. The spin labels are shown as magenta rotamer clouds. **a,** Possible kink region in the NTD is highlighted as a red dashed line. The two regions that may act as rigid bodies are marked with blue circles. **b,** Possible hinge region in the REC lobe, between the NTD and helical-1 domain, marked with a red, dashed line. The two protein regions, which could act as rigid bodies are highlighted with orange circles.

Both identified dynamic behaviors of first, the helical-1 domain through the $190/222^{apo}_{active}$ construct, and second between helical-1 and NTD through the $138/190^{apo}_{active}$ construct together explain the enhanced flexibility seen in the measurements of $138/222^{apo}_{active}$. The labelling positions 138 and 222 are connected via both hinge and kink regions, resulting in a very broad distance distribution, with no preferred distance.

4.1.3 Summary REC lobe

In contrast to the suggestion from the predicted and experimental structures, the PELDOR measurements indicate the NTD to be rigid and the helical-1 domain to be flexible. As shown by different spin label combinations, this flexibility presumably origins from a hinge and a kink region inside the REC lobe. The PELDOR measurements also suggest that it is unlikely that the protein has intrinsically disordered regions in the REC lobe. The PELDOR measurement in the NTD ($63/138^{apo}_{active}$) yielded a well-defined distance distribution, indicating a rigid and defined structure. The measurements performed at the interface of NTD and helical-1 ($138/190^{apo}_{active}$) and in the helical-1 domain ($190/222^{apo}_{active}$) both show a broad distance distribution, but the most probable distance fits to the AF3 predictions. This wouldn't be expected from intrinsically disordered proteins and suggests a flexible but structured protein region. Moreover, the per residue confidence scores (pLDDT) of the AF3 prediction of the REC lobe are high (mostly over 70). A low pLDDT can suggest a high structural variability and can be related to IDPs. However, at least AF2 overestimates disorder[159,277], which is an additional indication for the unlikeliness of intrinsically disordered regions in this protein. Since AF3 was published during the writing process of this thesis it is unknown, if also AF3 overestimates disorder.

The *apo* state, being very flexible and adopting a great variety of conformations, fits to what is known from other Cas-nucleases as Cas12a[49]. Upon addition of pre-crRNA to *Lbu*Cas13a, the REC lobe rigidifies and locks into one conformation. This conformation stays the same throughout the functional pathway up to the ternary complex.

4.2 Nuclease lobe

To test whether the NUC lobe also presents a high degree of flexibility in the structurally unknown *apo* state and whether it undergoes conformational changes, three double cysteine constructs were generated. These constructs contain labelling positions in different domains of the NUC lobe. The first one is 756/926, where the labelling positions are located in the HEPN1-II and linker domains, respectively. The second construct is 462/926, investigating the movements between the HEPN1-I and

the linker domains. The last construct is 660/926, investigating the helical-2 and the linker domains. In doing so, each domain in the NUC lobe is represented, except of the HEPN2 domain. This domain is essential for the proteins' function, thus perturbations that could origin from spin labelling are avoided.

4.2.1 Nuclease lobe is already preorganized in *apo*

For the NUC lobe, three double labelled constructs were designed such that the distance distribution of the crRNA and cr-and target RNA bound structures should change, based on the known structures. The constructs are 756/926 (spin labelled in the HEPN1-II and linker domains), 462/926 (spin labelled in the HEPN1-I and linker domains), and 660/926 (spin labelled in the helical-2 and linker domains). The PELDOR data of the *apo* state of these constructs are shown in Figure 4.2-1.

Beginning with the labelled $756/926^{apo}_{active}$construct (Figure 4.2-1), the distance distribution shows a defined peak with a most probable distance of 6.9 nm, which is a significant difference compared to the flexibility between the REC lobe domains. The AF3 prediction with a most probable distance of 7.8 nm does not fit perfectly.

The second construct, $462/926^{apo}_{active}$, in which the spin labels are located in the HEPN1-I and linker domains, adopts one most probable conformation with an interspin distance of 7.4 nm (Figure 4.2-1, middle). This is an indication that this specific protein region is preorganized already without RNA, as also the HEPN1-II and linker domains shown in Figure 4.2-1, top. The AF3 prediction with a most probable distance of 7.8 nm fits better to the experimental data, then the one of $756/926^{apo}_{active}$. Interestingly, the measurement of the $660/926^{apo}_{active}$construct reveals a bimodal distance distribution with two most probable distances at 5.7 nm and 6.7 nm (Figure 4.2-1, bottom).

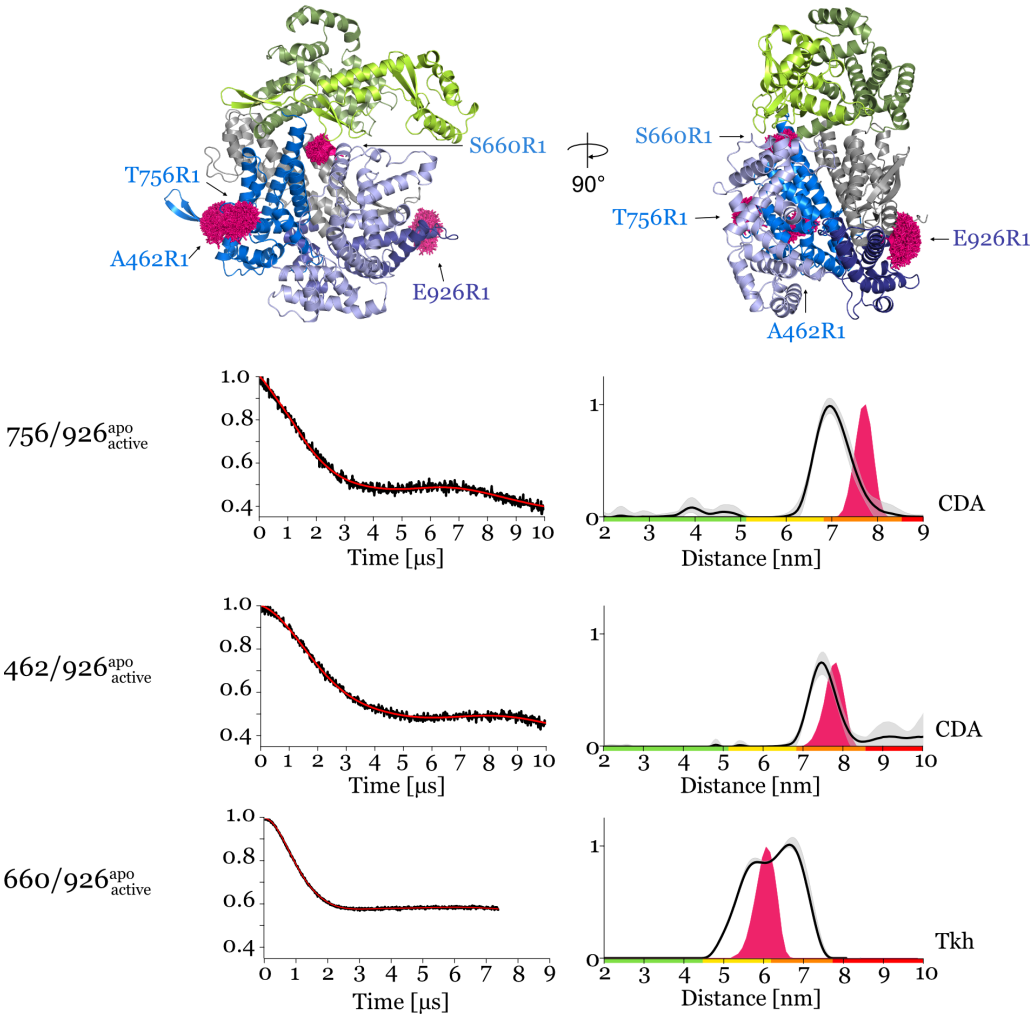


Figure 4.2-1: PELDOR results for the *apo* protein constructs labelled in the NUC lobe. The AF3 structure of *apo* *Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds. From top to bottom measurements are shown from the *apo* state of the active double labelled protein constructs 756/926, 462/926, and 660/926. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through Tikhonov regularization (Tkh) or the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on *apo* AF3 model, are shown as magenta shaded distributions.

All three NUC lobe constructs presented, $462/926^{apo}_{active}$, $756/926^{apo}_{active}$, and $660/926^{apo}_{active}$ share the labelling position 926. But a bimodal distribution in the *apo* state is only observed for the construct $660/926^{apo}_{active}$, indicating that the bimodality most probably origins from the position 660 in the helical-2 domain. To validate this, we additionally created the construct $660/756^{apo}_{active}$, where the labels are located in the helical-2 and HEPN1-II domains (Figure 4.2-2).

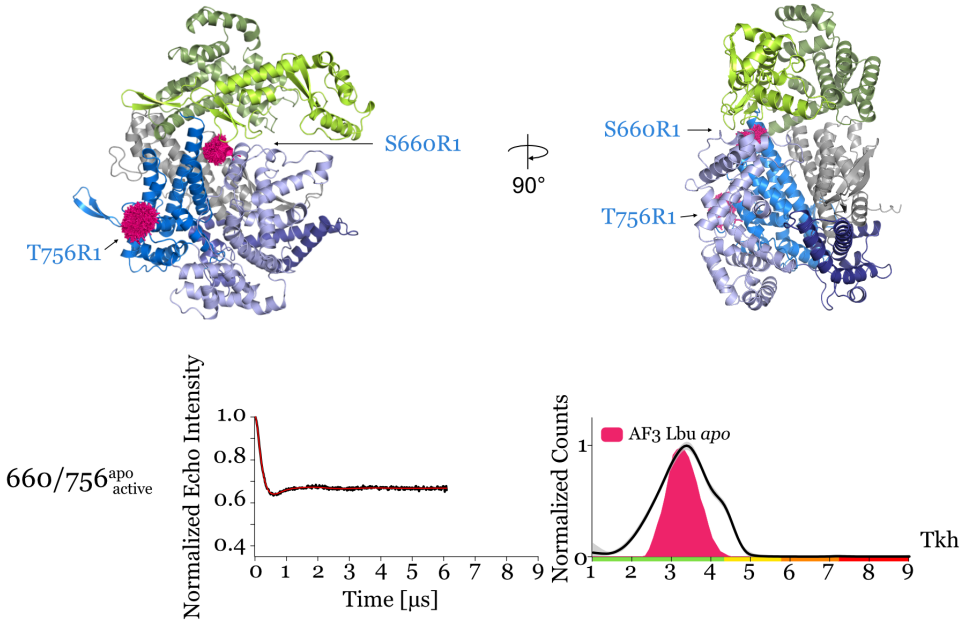


Figure 4.2-2 PELDOR result for the protein construct labelled at 660/756. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 660 and 756. The measurement shown is from the *apo* state of the active double labelled protein construct. The time trace includes the fits that were done with the software DeerAnalysis[276], through Tikhonov regularization. The distance distribution on the right contains the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The prediction of the distance distributions based on AF3 model of *Lbu*Cas13a is shown as shaded distributions.

It can be seen, that also the $660/756^{apo}_{active}$ construct shows a broad distance distribution with a bimodality (Figure 4.2-2). This bimodality is not as prominent as for $660/926^{apo}_{active}$ (Figure 4.2-1, bottom), but a clear most probable distance at 3.4 nm and a shoulder at 4.3 nm are visible. This bimodality can arise from two different protein conformations or from label conformers. Generally, it is difficult to distinguish this, and a detailed discussion is done in Chapter 4.2.3.

Focusing on the distance distributions based on the AF3 prediction of the $660/926^{apo}_{active}$ construct (magenta shaded area, Figure 4.2-1, bottom) and of the $660/756^{apo}_{active}$ construct, no bimodality can be seen. In addition, the most probable distance of this AF3 prediction is laying in between of both most probable distances measured by PELDOR spectroscopy. This indicates that the Cas13a protein is more dynamic than predicted.

4.2.2 NUC lobe undergoes conformational changes upon RNA addition

The three constructs shown above (Chapter 4.2.1) were used to determine if conformational changes occur upon RNA addition, as suggested by the known cryo-EM and crystal structures. Beginning again with the 756/926$^{apo}_{active}$construct, it was already seen above (Figure 4.2-1,), that the AF3 prediction does not fit perfectly.

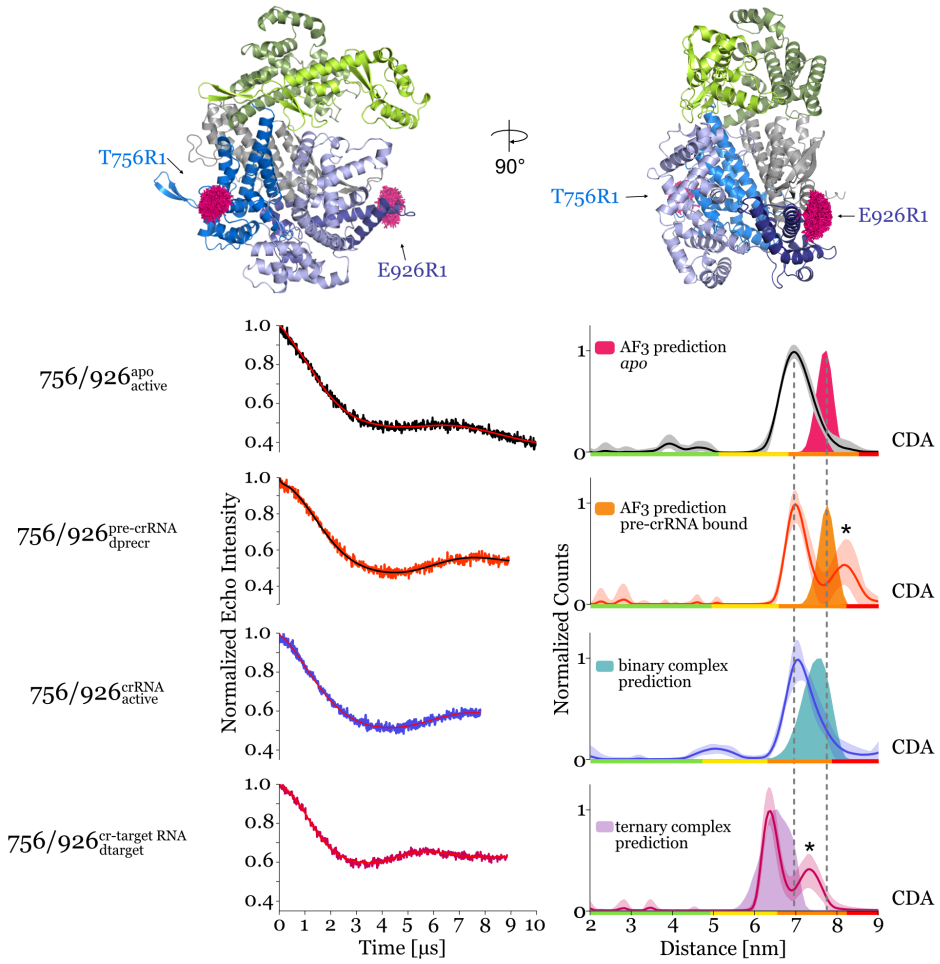


Figure 4.2-3: PELDOR results for the protein constructs labelled at 756/926. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 756 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models, the binary cryo-EM structure, and the ternary X-ray crystallographic structure are shown as shaded distributions. The asterisks in the distance distributions highlight the peak that could origin from dimerization, investigated in Chapter 4.4.

But interestingly this predicted distribution fits to the closed, RNA bound state, seen in the experimental cryo-EM structure of the binary complex with crRNA (blue shaded area in Figure 4.2-3). When the pre-crRNA is added to the dprecr construct, $756/926^{\text{pre-crRNA}}_{\text{dprecr}}$, the most probable distance does not change (orange distribution in Figure 4.2-3). But the distance distribution slightly narrows, meaning that the protein structure rigidifies. The most probable distance does also not change upon crRNA addition to the active construct (blue distribution in Figure 4.2-3). Interestingly, also the AF3 predictions of the *apo*, and the pre-crRNA bound structures, as well as their distance prediction based on the experimental crRNA bound structure are very similar, with mean distances at 6.8 - 6.9 nm. In contrast to these three measurements, the addition of target RNA to the dtarget construct leads to a significant structural change. The most probable distance is shifted to shorter distances of 0.7 nm, which fits to the predictions based on the experimental X-ray and cryo-EM structures (blue and purple shaded area in Figure 4.2-3). Target RNA binding leads to an opening of the structure to accommodate the crRNA-target RNA helix, which brings both labelling position closer to each other.

Importantly, a second peak is seen at larger distances in the distance distributions of the pre-crRNA bound complex $756/926^{\text{pre-crRNA}}_{\text{dprecr}}$ and the cr- and target RNA bound complex $756/926^{\text{cr-target RNA}}_{\text{dtarget}}$ (marked with an asterisk in Figure 4.2-3). This peak is investigated further below in Chapter 4.4.

The second labelling position combination that was investigated was 462/926. This label combination should yield information about the flexibility and conformation of the HEPN1-I and linker domains. The label combinations 756/926 and 462/926 are likely to yield similar distance distributions, since both labelling positions, 462 and 756, are spatially in the same region.

As described above, also this region of the protein, between HEPN1-I and linker domains is preorganized, since the distance distribution is narrow and monomodal with a most probable distance of 7.4 nm (black distribution in Figure 4.2-4). This is also the case for the HEPN1-II and linker domains shown in Figure 4.2-3 (black distribution).

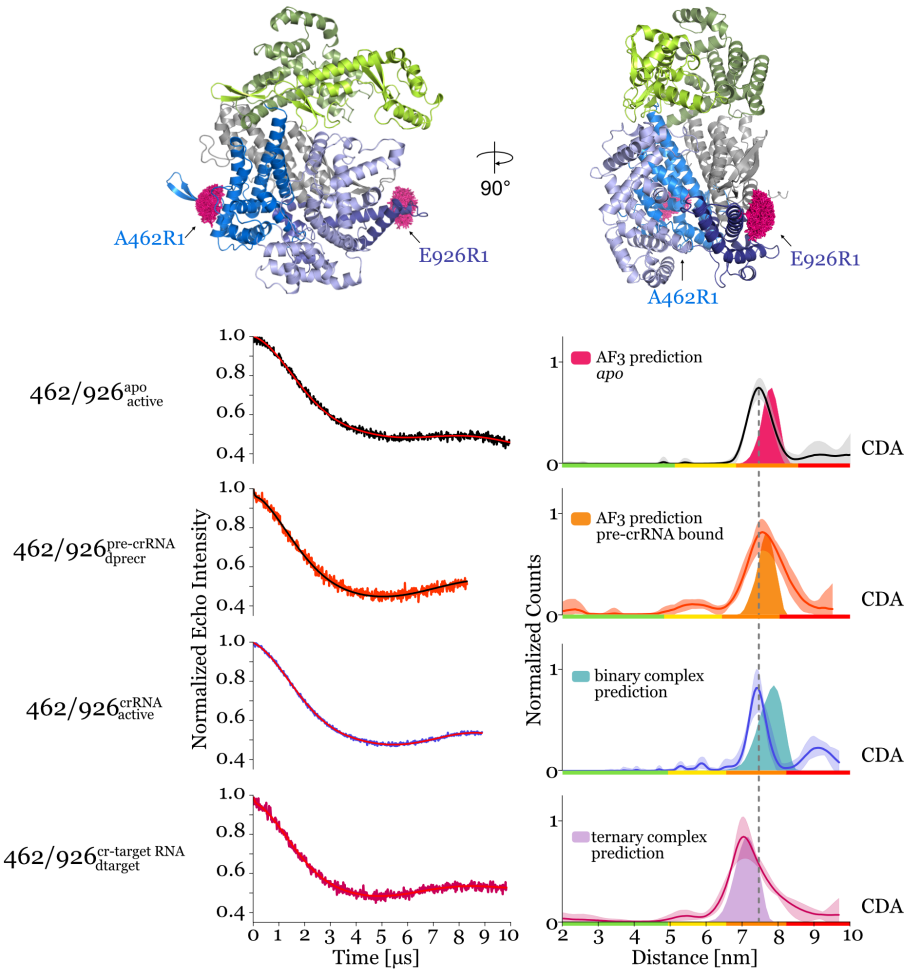


Figure 4.2-4: PELDOR results for the protein constructs labelled at 462/926. The AF3 structure of *apo LbuCas13a* is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 462 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models, the binary cryo-EM structure, and the ternary X-ray crystallographic structure are shown as shaded distributions.

When the pre-crRNA was added to the dprecr construct the most probable distance stayed at 7.4 nm and no structural rearrangement is detected (orange distribution in Figure 4.2-4). This is also valid for the $462/926^{\text{crRNA}}_{\text{active}}$ measurement (blue distribution in Figure 4.2-4). It is difficult to determine if the crRNA has bound to the protein, since no distance change from the *apo* to the crRNA bound complex is observed. However, isothermal titration calorimetry (ITC) was performed with various constructs,

including 756/926. These measurements confirmed crRNA binding, and the K_D did not change upon mutations and spin labelling compared to the wt *Lbu*Cas13a. These measurements were performed by Joshua Lee Wort (data not shown). The predicted distance distribution based on the AF3 prediction of the pre-crRNA bound complex fits nicely to the PELDOR distance distribution, while the prediction based on the crRNA bound cryo-EM structure is slightly off (blue shaded distribution Figure 4.2-4). The predicted distributions for both, the $756/926^{crRNA}_{active}$ and the $462/926^{crRNA}_{active}$constructs, predict the most probable distance to be slightly larger. One reason for this could be label conformers or the different RNA sequence used for both experimental structures. This is discussed in Chapter 4.2.4. From the crRNA to the cr- and target RNA bound state, the distance shifts to shorter distances, as also predicted by the experimentally known structure.

4.2.3 Conformational selection of the helical-2 domain

The third measurement set that was performed with labelling positions exclusively in the NUC lobe is with the 660/926 construct. The labelling position 660 is in the helical-2 domain and the labelling position 926 in the linker domain. The helical-2 domain is of special interest, since it undergoes the largest conformational change when going from the binary crRNA bound to the ternary cr- and target RNA bound complex, as seen in the displacement plot in Figure 3.4-2. Also, in the predicted AF3 structures from the *apo* to the pre-crRNA bound state (Figure 4.1-1), the helical-2 domain is predicted to undergo a conformational change that rotates this domain towards the HEPN1 domain. But from pre-crRNA bound to the crRNA bound complex this domain seems to stay in the same conformation.

As described above, the *apo* measurement of the 660/926 construct reveals a bimodal distance distribution with two most probable distances at 5.7 nm and 6.7 nm (black distribution in Figure 4.2-5). This bimodality origins from the position 660 in the helical-2 domain (Chapter 4.2.1). Focusing on the AF3 prediction of the $660/926^{apo}_{active}$construct (magenta shaded area in Figure 4.2-5), no bimodality can be seen. The most probable distance of this AF3 prediction is laying in between of both most probable distances measured by PELDOR spectroscopy, as if AF3 has predicted

an intermediate-like state. The origin of the bimodality seen in the PELDOR experiments is unknown. It can come from two different protein conformations or from label conformers, which is difficult to distinguish. Upon addition of pre-crRNA, the conformational equilibrium is shifted to the first peak at 5.7 nm (orange distribution in Figure 4.2-5). Also, a technical replicate of the $660/926^{\text{pre-crRNA}}_{\text{dprecr}}$ measurement showed the same shift (Appendix, Figure 8.3-7). The AF3 prediction for $660/926^{\text{pre-crRNA}}_{\text{dprecr}}$ fits to our experimental data, but again, AF3 cannot reproduce the dynamics of the protein. Additionally, AF3 can also not reproduce the conformational change between *apo,* and pre-crRNA bound states, that are visible in the PELDOR measurements. The conformational shift seen through the addition of pre-crRNA continues upon crRNA addition. A monomodal distribution is obtained for $660/926^{\text{crRNA}}_{\text{active}}$, with a most probable distance at 5.7 nm (blue distribution in Figure 4.2-5), which is slightly off with respect to the prediction based on the cryo-EM structure (blue shaded distribution). Thus, both conformations underlying both distance peaks, seen in the *apo* distance distribution can be converted into each other through the addition of RNA. However, upon target RNA addition, the distribution of $660/926^{\text{cr-target RNA}}_{\text{dtarget}}$ is shifted to shorter distances, in line with the prediction of the ternary complex (purple shared distribution).

Regarding the first peak at 5.7 nm, it can be hypothesized that this protein conformation under *apo* and pre-crRNA bound conditions resembles the one from the binary, crRNA bound complex, since the most probable distance is the same in both. However, the second peak can't be explained by comparison of the AF3 prediction, and the main question here is if this conformation origins from the protein or from label conformers. Known experimental structures as well as previous studies give no indication that other conformations could be adopted. Since no experimental *apo* structure of *Lbu*Cas13a exists, it should be tested, if the structure of the phylogenetically similar *apo Lsh*Cas13a would help to understand this bimodal structure-feature.

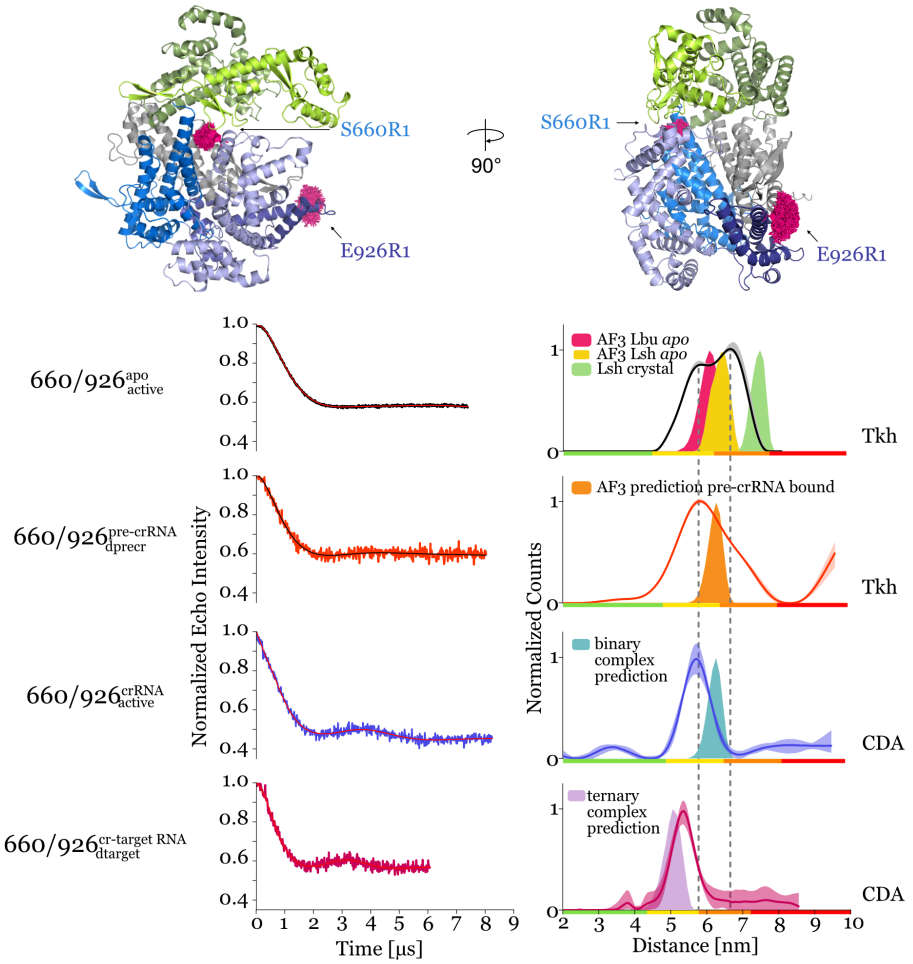


Figure 4.2-5: PELDOR results for the protein constructs labelled at 660/926. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 660 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through Tikhonov regularization and the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models of *Lbu*Cas13a and *Lsh*Cas13a, the X-ray crystallographic structure from *apo Lsh*Cas13a (PDB: 5WTJ)[88], the binary cryo-EM structure of *Lbu*Cas13a, and the ternary X-ray crystallographic structure of *Lbu*Cas13a are shown as shaded distributions.

Careful structural evaluation was done to *in-silico* spin label *Lsh*Cas13a at comparable positions, as in *Lbu*Cas13a. The resulting labelling positions for S660 and T756 in *Lbu*Cas13a are K756 and R1158 in LshCas13a, respectively. The predicted distance distributions from the AF3 prediction and from the *apo Lsh*Cas13a crystal structure of 756/1158 are shown as yellow and green shaded areas in Figure 4.2-5, respectively. The *apo Lsh*Cas13a distribution based on the AF3 prediction yields a similar distribution

compared to the one from *apo Lbu*Cas13a. In contrast, the distribution based on the crystal structure of *Lsh*Cas13a predicts a much longer distance.

To determine the differences between the AF3 prediction of *Lbu*Cas13a and the *Lsh*Cas13a crystal structure, both structures were aligned (with an RMSD of 16.9, Figure 4.2-6 a). The main structural differences between the AF3 prediction of *Lbu*Cas13a and the *Lsh*Cas13a crystal structure are in the helical-2 domain (blue domains in Figure 4.2-6 a). It can be noted, as simplified by the ellipsoids in Figure 4.2-6 a, that the helical-2 domains of both complexes are rotated. Even though the sequence identity between both is only 23% (calculated with Expasy[278]), the secondary structure elements are remarkably similar. As shown in the insert in Figure 4.2-6 a, the helices in the first structure can easily be attributed to the helices in the second structure. Interestingly, also the helical-2 domains in the *Lsh*Cas13a crystal structure and in the *apo* AF3 *Lsh*Cas13a structure are rotated (Figure 4.2-6 b). Here the sequences are identical, and the rotated helical-2 conformations can't result from different amino acid sequences.

Taken together, it seems that the helical-2 domain is rotated in a rigid-body-fashion, when the AF3 prediction of *Lbu*Cas13a and the *Lsh*Cas13a crystal structure are compared. Thus, it is hypothesized that it is this rotation that is seen in *apo Lbu*Cas13a, which leads to two different, distinct conformations, detected by PELDOR spectroscopy. The movement suggested for the *apo* state would also fit to the movement observed in the helical-2 domain between the crRNA bound and the cr- and target RNA bound conformations (Figure 4.2-6 c and d). There the helical-2 domain undergoes the same rigid-body movement. In contrast, the linker domain (Figure 4.2-6 d) in which the labelling position 926 is located, undergoes only small conformational changes.

When looking at the PELDOR data for the whole functional pathway, the most probable interspin distance decreases from a large distance at 6.7 nm and a smaller distance at 5.8 nm of the *apo* protein, via the 5.8 nm distance in the pre-crRNA and crRNA bound complexes, to finally the ternary complex with the shortest distance at 5.3 nm (Figure 4.2-5). The domain movement on which the distance change is based on is nicely seen, when comparing the elliptic shapes from the *apo Lsh*Cas13a crystal structure in Figure 4.2-5 a (light blue domain), via the crRNA bound *Lbu*Cas13a in Figure 4.2-5 c (light purple domain), to the cr- and target RNA bound *Lbu*Cas13a in

Figure 4.2-5 c (light blue domain). There, structurally the domain seems to gradually rotate towards the linker domain.

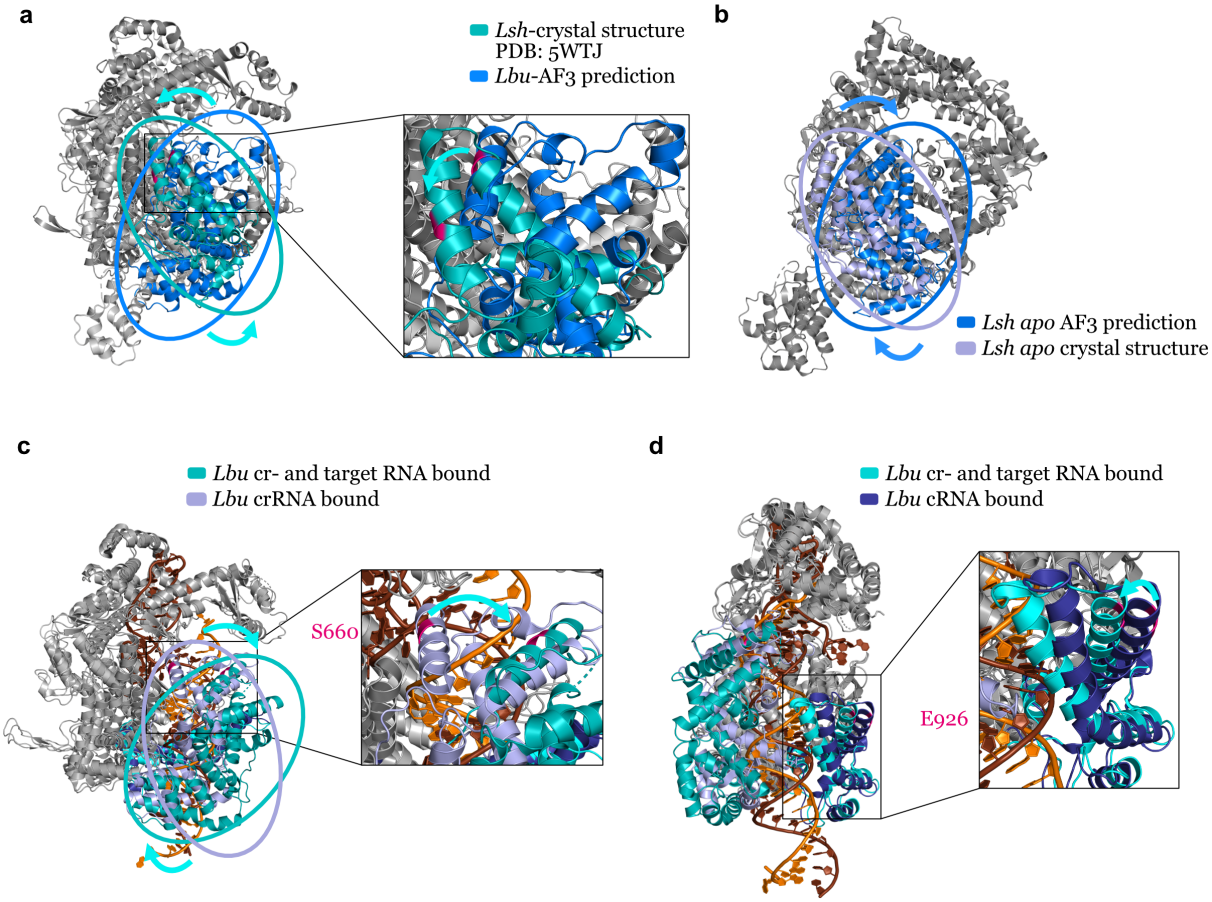


Figure 4.2-6: Helical-2 domain conformations. a, The *Lsh apo* crystal structure (PDB:5WTJ) is aligned with the *apo Lbu*Cas13a AF3 prediction. The helical-2 domains are highlighted in different shades of blue. The domain positions are schematically highlighted with ellipsoids. The insert on the right shows the protein region in the helical-2 domain in which the labelling position 660 (in *Lbu*, and 756 in *Lsh*) resides. Both label positions are highlighted in magenta and the rotation of this helix in both structures is shown with an arrow. **b,** The *Lsh*Cas13a AF3 *apo* structure is aligned with the *Lsh*Cas13a *apo* crystal structure (PDB: 5WTJ). The helical-2 domains are highlighted in different shades of blue. The domain positions are schematically highlighted with ellipsoids. Same insert as in a. **c,** The crRNA bound (PDB: 5XWY) and the cr- and target RNA bound (PDB: 5XWP) structures of *Lbu*Cas13a are shown aligned. The helical-2 domains are colored in different shades of blue and the domain positions and movements from the binary to the ternary structure are schematically drawn as ellipsoids and marked with arrows. The insert shows the location of the labelling position S660. **d,** Same structures and alignment is shown as in c, but rotated by 90°, towards the left. Now, the linker domain is shown in the insert, focusing on the labelling position E926. The movement of the helix, in which the labelling position E926 is located is depicted by a blue arrow.

This indicates, again, that the bimodality seen in our PELDOR data (Figure 4.2-5) origins mainly from a rigid-body movement of the helical-2 domain adopting two preferred orientations in *apo* in frozen solution. The first conformation resembles the

closed crRNA bound state and the second conformation, with a larger interspin-distance, shows a more open conformation in which the helical-2 domain is rotated towards the linker and HEPN-1 domain.

4.2.4 Conformational changes with different RNA sequences

All three measured distributions shown above, in which the crRNA was added (blue distributions for 756/926 in Figure 4.2-3, 462/926 in Figure 4.2-4, and 660/926 in Figure 4.2-5) show a slight discrepancy, when compared to the predicted distance distribution of the complex (blue shaded distributions). This prediction was based on the cryo-EM structure obtained by Liu et al.[87] in which a different crRNA sequence was used. Additionally, the crRNA and target RNA bound complex obtained by X-ray crystallography was also solved by Liu et al.[87], with different cr- and target RNA sequences than those used in the measurements showed in this thesis. All RNA containing measurements that were done for this thesis contained RNA sequences adapted from East-Seletsky et al.[75] Thus, PELDOR measurements were repeated with the same RNA sequences as used by Liu et al., to determine whether the discrepancies between the PELDOR measurements and the experimental structure originate from different crRNA sequences. Even though the PELDOR measurements of the ternary cr- and target RNA complexes are in very nice agreement with the predictions from the structures, also the conformation of the ternary complex was tested to determine whether it is changed upon addition of different cr- and target RNA sequences. The different crRNA and target RNA sequences are shown in Figure 4.2-7.

The repeat of both crRNA sequences is identical, which makes sense, since this is the crRNA region that is recognized by the REC lobe of Cas13a. But the spacer of both crRNA sequences is different (blue sequence in Figure 4.2-7), with a longer sequence used by Liu et al.[87]. Consequently, the target RNA sequences (orange sequences in Figure 4.2-7, bottom sequence) are also different and the one used by Liu et al. is longer. The longer sequences lead to a larger hybridization area between the target RNA and the crRNA spacer.

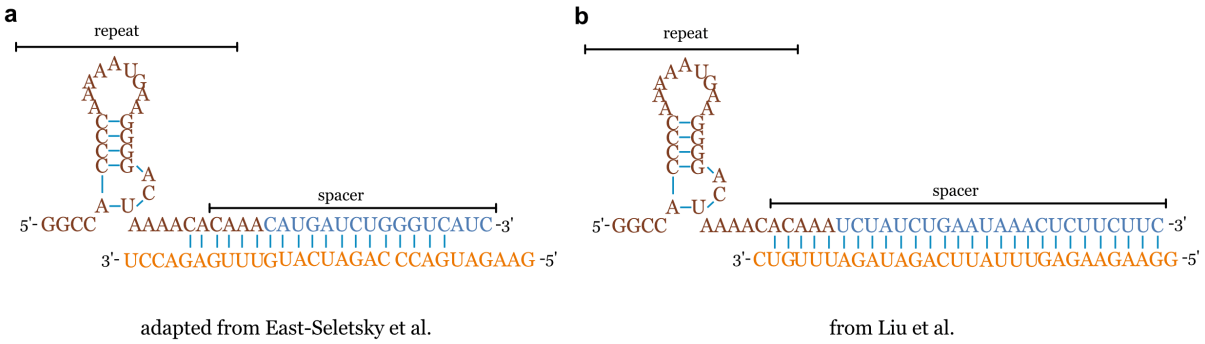


Figure 4.2-7: Comparison of RNA sequences. a, RNA sequences adapted from East-Seletsky et al.[75] and used for PELDOR studies. The crRNA is shown in brown and blue, on top. Target RNA is shown in orange, on the bottom. The region in which crRNA and target RNA hybridize is highlighted by blue bonds. **b,** RNA sequences used for crystallization in Liu et al.[87] The color code and representation is the same as in a. The crRNA region that differs between a and b is colored in blue.

In addition to the measurements shown in Figure 4.2-4 with RNA sequences adapted from East-Seletsky et al., PELDOR measurements with the 462/926 construct containing the RNA sequences from Liu et al. were performed. This specific construct was chosen, since it has a monomodal distance distribution and since a conformational change between the $462/926^{\mathrm{crRNA}}_{\mathrm{active}}$ and $462/926^{\mathrm{cr\text{-}target\ RNA}}_{\mathrm{dtarget}}$ should be observed. A comparison of both used RNA sequences is shown in Figure 4.2-8. Beginning with the binary complex $462/926^{\mathrm{crRNA}}_{\mathrm{active}}$ (time traces and distance distributions on top of Figure 4.2-8), it can be seen that no difference was observed. The most probable distance for both crRNA bound complexes is 7.4 nm and the same distribution width was obtained. The conclusion that can be drawn here is that a different spacer sequence does not seem to alter the proteins conformation, at least in the region in which the spin labels are located. The same is also valid for the comparison between the ternary complexes with different RNA sequences (time traces and distributions on the bottom in Figure 4.2-8). The most probable distance for both, the measurement with RNA sequences from Liu et al. and from East-Seletsky et al., is at 7 nm. This indicates that the different cr- and target RNA sequences tested here do not have an influence on the conformation of the ternary complex.

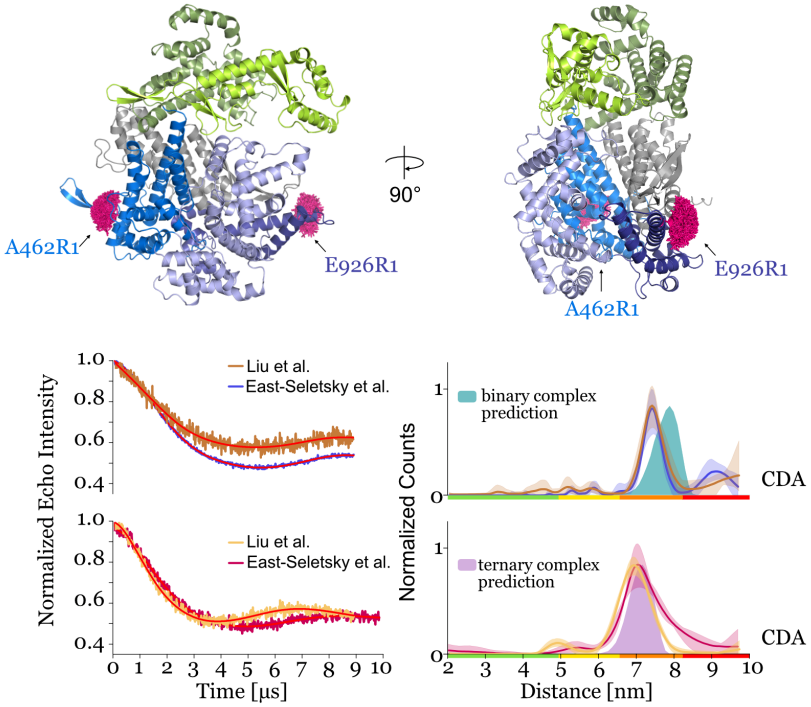


Figure 4.2-8: Influence of different crRNA and target RNA sequences on the protein conformation. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 462 and 926. The 462/926 construct was measured with the RNA sequences shown in Figure 4.2-7. The active protein construct was used for the measurement with crRNA and the dtarget construct for the measurement including cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The reliability bars shown here are the ones for the measurements with the RNA sequences from Liu et al.[87] The reliability bars from the measurements done with the RNA sequences from East-Seletsky et al. can be found in Figure 4.2-4.The predictions of the distance distributions based on the binary cryo-EM structure of *Lbu*Cas13a, and the ternary X-ray crystallographic structure of *Lbu*Cas13a are shown as shaded distributions.

Taken together, the measurements performed with the RNA sequences of Liu et al. show that potential discrepancies between the experimental structures of the binary and ternary complexes and the PELDOR measurements are most likely not caused by a different RNA sequence. Other reasons might be different experimental conditions as buffers and freezing methods, which can result in slightly different arrangements of the helix containing the spin label at position 926. This is discussed in detail in the Discussion in Chapter 4.5.

4.2.5 Summary NUC lobe

In contrast to the REC lobe of the protein, the NUC lobe seems to be preorganized in the *apo* state. Interestingly, the NUC lobe adopts one conformation, which does not change upon pre-crRNA and crRNA addition.

However, there is one exception, which is the helical-2 domain. This domain adopts two conformations in the *apo* state. One of these conformations is identified as the one of the crRNA-bound structure. The other conformation is presumably formed through a rotation of the helical-2 domain, as shown by comparison with homolog Cas13a protein structures. We hypothesize that it is the same rotation also seen in the conformational change between the crRNA bound and the cr- and target RNA bound complex. Upon pre-crRNA and crRNA addition a conformational selection of the first peak is seen, attributed to the crRNA bound state.

For all crRNA-supplemented samples, discrepancies were seen when the experimental distributions were compared to the predicted ones based on the crRNA-bound cryo-EM structure. By repeating the PELDOR experiments with the same crRNA sequence as used for the cryo-EM structure, identical results were obtained. Additionally, also when comparing the ternary complexes formed with different crRNA and target RNA, no difference in protein conformation is observed. This indicates that the different RNA sequences yield the same protein conformation in frozen solution and that the discrepancies most likely come from experimental conditions, as discussed in Chapter 4.5.

4.3 Orientation of REC and NUC lobe towards each other

In the PELDOR measurements shown above, the REC lobe and the NUC lobe were investigated separately regarding their conformations, conformational changes, and dynamics (Chapters 4.1 and 4.2). The next step is to determine if and how the two lobes change their respective orientation towards each other upon RNA binding. As before, the *apo* state as well as the RNA bound complexes were investigated.

4.3.1 pre-crRNA bound complex adopts the same conformation as the crRNA bound complex

To analyze the conformational changes that occur between the two lobes, two double cysteine constructs were designed, with one label in the REC lobe and the other one in the NUC lobe. Additionally, emphasis is placed on the pre-crRNA bound complex. The experimental structure of the pre-crRNA bound complex is unknown. However, the previous PELDOR measurements suggest that the pre-crRNA bound, and the crRNA bound complex adopt the same conformation. The labelling positions that were chosen for REC-NUC investigations are 190/756 (Figure 4.3-1) and 190/926 (Figure 4.3-2). The position 190 is in the helical-1 domain, the position 756 in the HEPN1-II domain, and the position 926 in the linker domain. These two double cysteine constructs enable to measure the distance distributions from the helical-1 domain to one side of the NUC lobe (near to the protein backbone, position 756, see Figure 4.3-1) and from the helical-1 domain to the other side of the NUC lobe (position 926, see Figure 4.3-2). These two label combinations were chosen because of their different structural features. The labelled domains of the construct 190/926 should undergo a conformational change, leading to a shift of the distance distribution of about 0.3 nm from the crRNA-bound to the cr- and target RNA bound state (as predicted based on the experimental structures). In contrast, the distance distribution of the 190/756 construct should not change between the cr-RNA and the cr- and target RNA bound structures.

Generally, 190/756 (Figure 4.3-1) and 190/926 (Figure 4.3-2) show very similar behaviors. The *apo* distance distributions of 190/756 (black distribution, Figure 4.3-1) and 190/926 (black distribution, Figure 4.3-2) are both broad. From the studies shown in Figure 4.1-4, it was seen that under *apo* conditions the distance distribution of 138/190 is broad, and that the distribution of 63/138 is narrow. Consequently, presumably the labelling position 190 is in a flexible region and the distributions of new constructs 190/756 and 190/926 are expected to also show a high flexibility. In contrast, the PELDOR measurement of the 756/926 construct showed that the protein conformation in *apo* between positions 756 and 926 is rigid (Figure 4.2-3). The predicted distance distribution based on the *apo* AF3 structure fits into the middle of the experimental distribution, but obviously it doesn't reflect the flexibility seen in the

PELDOR measurement. This confirms the helical-1 domain being flexible.

Upon pre-crRNA addition, the distributions of $190/756^{\text{pre-crRNA}}_{\text{dprecr}}$ (orange distribution, Figure 4.3-1) and $190/926^{\text{pre-crRNA}}_{\text{dprecr}}$ (orange distribution, Figure 4.3-2) become narrower, hence the protein structure rigidifies. The most probable distance of the 190/756 construct is 6.9 nm (Figure 4.3-1) and the one of the 190/926 construct is 7.2 nm (Figure 4.3-2).

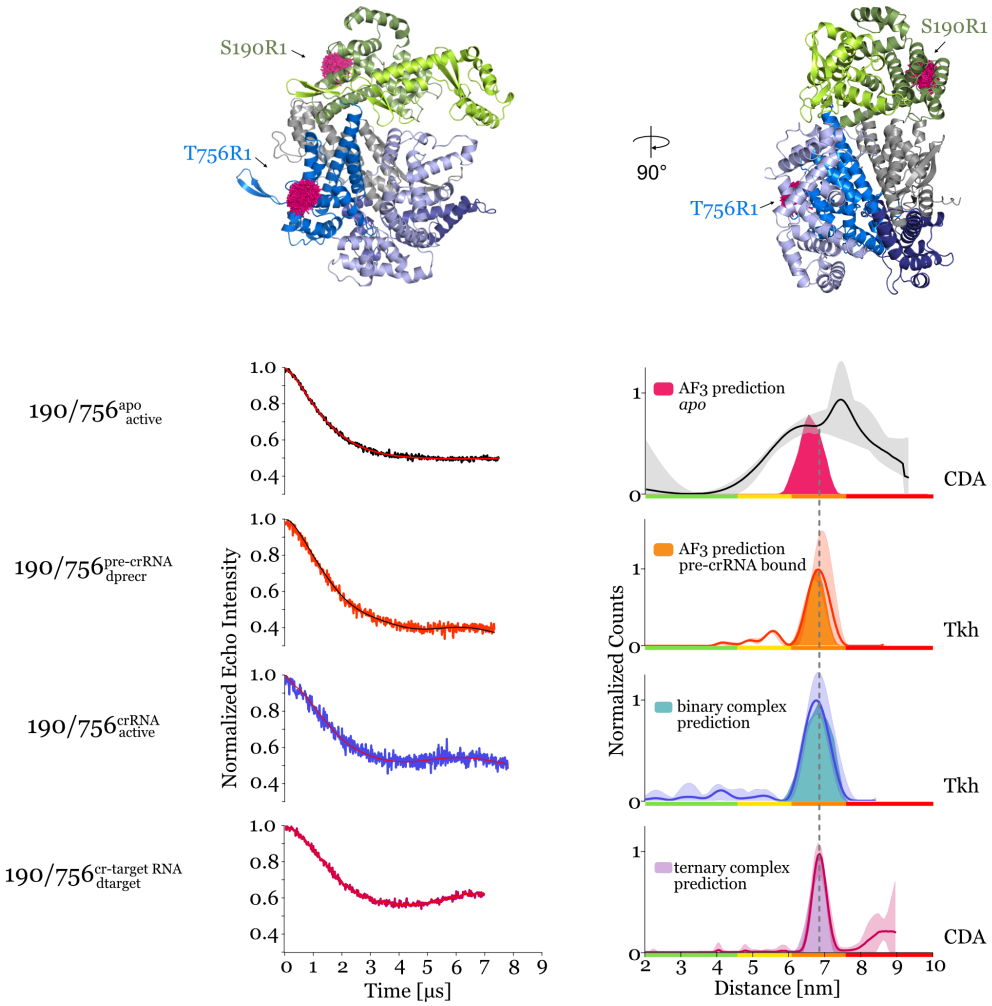


Figure 4.3-1: PELDOR results for the protein constructs labelled at 190/756. The AF3 structure of *apo LbuCas13a* is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 190 and 756. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through Tikhonov regularization and the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models of *Lbu*Cas13a, the binary cryo-EM structure of *Lbu*Cas13a, and the ternary X-ray crystallographic structure of *Lbu*Cas13a are shown as shaded distributions.

The predictions of the distance distributions based on the AF3 pre-crRNA bound complex fit nicely to the experimental data for both constructs, the PELDOR data supports the right conformation predicted by AF3. The distributions of the pre-crRNA bound, and the crRNA bound states of both, 190/756 and 190/926, show no conformational change when the crRNA is added to the active construct (blue distributions). Again, the experimental distributions fit nicely to the prediction based on the known cryo-EM structure (blue shaded distribution). Taken together, neither the PELDOR data nor the distance distributions based on AF3 predictions or experimental structures showed a conformational change upon binding, supporting the hypothesis that the pre-crRNA bound and the crRNA bound complexes adopt the same conformation.

The same is true for the fourth measurement with cr- and target RNA. Upon target RNA addition, no conformational change can be observed for 190/756 (purple distribution, Figure 4.3-1) and 190/926 (purple distribution, Figure 4.3-2) compared to the pre-crRNA and crRNA bound state. Interestingly, the ternary complex seems to be more rigid than the pre-crRNA bound, and the crRNA bound states, as seen by a sharper distance distribution.

Also, for $190/756^{\text{cr-target RNA}}_{\text{dtarget}}$ the experimental distribution fits nicely to the one based on the ternary complex crystal structure (purple shaded distribution). Generally, it seems that the protein conformation in frozen solution changes from the pre-crRNA bound to the cr- and target RNA bound only in the horizontal direction (in the REC and in the NUC lobe, Chapters 4.1, 4.2) and not vertically between the REC and NUC lobe.

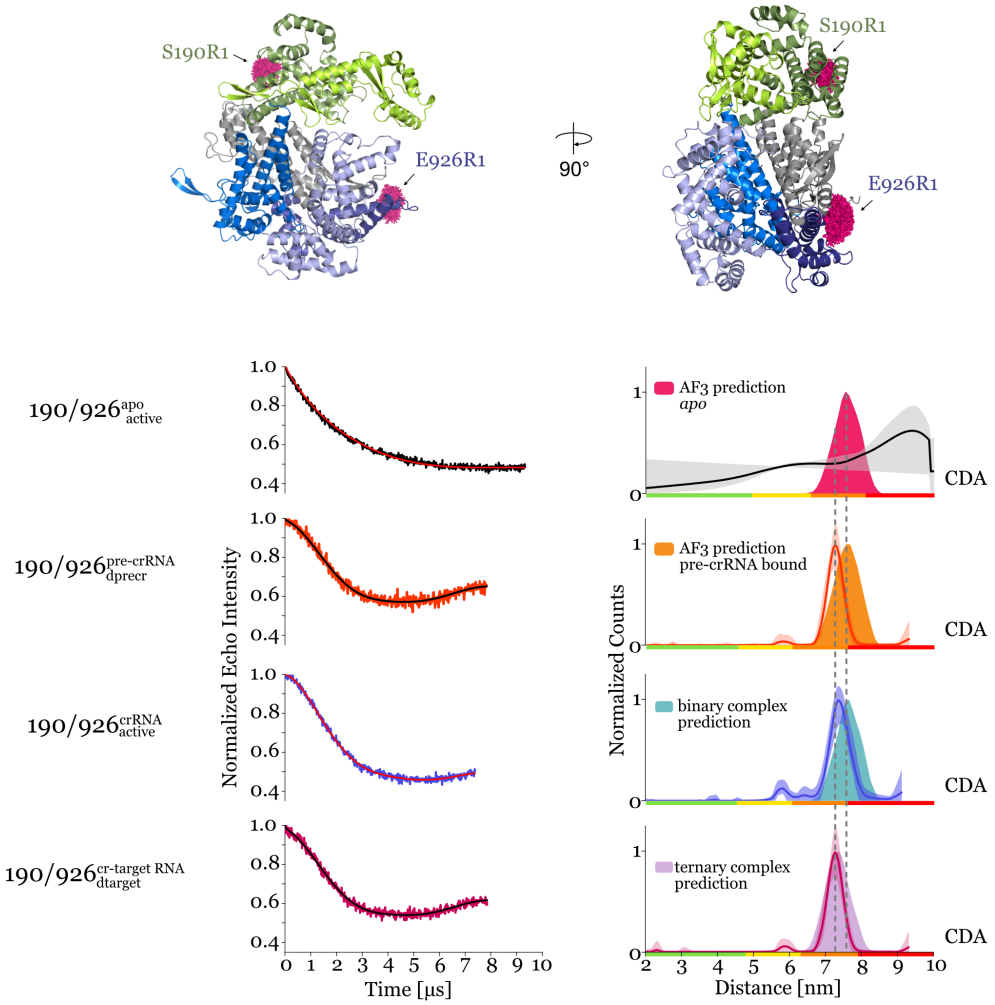


Figure 4.3-2: PELDOR results for the protein constructs labelled at 190/926. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 190 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models of *Lbu*Cas13a, the binary cryo-EM structure of *Lbu*Cas13a, and the ternary X-ray crystallographic structure of *Lbu*Cas13a are shown as shaded distributions.

As described above, the helical-1 domain is flexible towards the NUC lobe. To test, if also the NTD is flexible towards the NUC lobe, the labelling position 138 and 926 were chosen. This construct was created during the master thesis of Tim Mevs (University of Bonn)[279]. In depth functional studies can be found in this master thesis. It can be hypothesized from previous studies in Figure 4.1-4 that the position 138 is principally rigid, since the distance distribution of 63/138 is narrow. Same counts for the

measurement of the 756/926 construct, the protein conformation in *apo* between positions 756 and 926 is rigid (Figure 4.2-3).

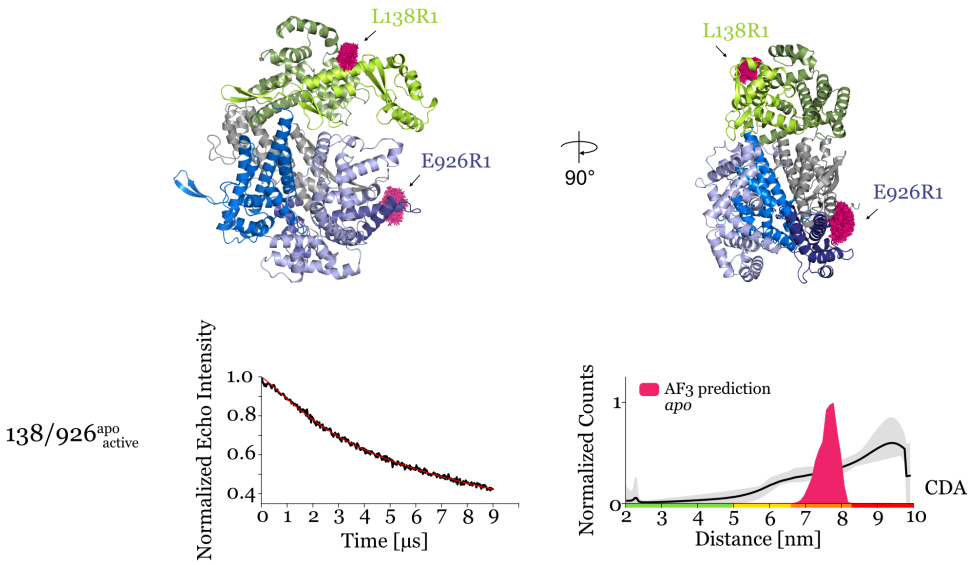


Figure 4.3-3: PELDOR results for the protein constructs labelled at 138/926. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 138 and 926. The PELDOR time trace and distribution are shown from the *apo* state of the active double labelled protein construct. The time traces includes the fit that was done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. The distance distribution on the right contains the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The prediction of the distance distribution based on AF3 model of *Lbu*Cas13a, is shown as magenta shaded distribution. Data recorded by Tim Mevs [279].

However, when looking at the distribution of the 138/926 construct (Figure 4.3-3), a broad distribution is observed. This is a strong indication that the NTD, even though it is intrinsically rigid (see data in Figure 4.1-4), is very flexible with respect to the NUC lobe or that the protein conformation is wide open. This behavior is very similar to the one observed for the helical-1 domain.

4.3.2 REC and NUC lobe summary

Concluding it can be summarized, that in *apo*, the two REC lobe domains are confirmed to be flexible. Additionally, to an intrinsic opening and closing motion with a kinking of the REC lobe (as hypothesized in Chapter 4.1), it is unclear if both REC

lobe domains move in the direction of the NUC lobe. This would mean that a second hinge region is present in the protein backbone, between the REC and NUC lobe, which open and closes both lobes toward each other. This will be further discussed in Chapters 4.4.2 and 4.5.

The pre-crRNA bound complex seems to adopt the same conformation as the crRNA bound complex. The protein conformations from the pre-crRNA bound state, via the crRNA bound state to the cr- and target RNA bound state seem to stay the same between the REC and NUC lobe and no conformational change was detected. The known structures of the binary and ternary complexes fit to the PELDOR data presented.

4.4 Dimerization of *Lbu*Cas13a

In the distance distributions of the NUC lobe construct $756/926^{\text{pre-crRNA}}_{\text{dprecr}}$ (orange distribution, Figure 4.2-3) and $756/926^{\text{cr-target RNA}}_{\text{dtarget}}$ (purple distribution, Figure 4.2-3) a second peak at larger distances was observed. These peaks at longer distances could origin from protein oligomerization, which will be investigated in the following Chapter.

Generally, methods to test for oligomerization are for example analytical size exclusion chromatography or native PAGE. During the purification of proteins, it can happen that peaks at lower elution volumes are detected during SEC runs, which could originate from oligomers. Leila Tokic tried in her bachelor thesis, to establish an analytical size exclusion routine of the wt *Lbu*Cas13a but failed (data shown in [280]). Additionally, the detection of oligomers via native PAGE was not suitable. *Lbu*Cas13a has an isoelectric point of 9.1, which is higher than the pH of the gel, this led to the protein not running through the PAGE gel (data shown in [280]).

Another method that could help to determine the presence of oligomers is to spin label the protein with only one spin label. Since peaks have been identified in the PELDOR measurements of the 756/926 construct, two new constructs were prepared bearing only 756R1 or only 926R1. If the protein oligomerizes and if the distance between the spin labels is in the PELDOR regime, a defined distance should be detected.

Additionally, to these PELDOR experiments, atomic force microscopy (AFM) can yield the desired information, if the protein is big enough that it can be visualized on a surface and if the protein is successfully attached onto the surface. The advantage by using AFM is, that no spin labelled protein is needed and that the wt *Lbu*Cas13a can be used with the respective dprecr and dtarget mutations.

This Chapter will deal with the oligomerization of *Lbu*Cas13a, and how we determined the formation of dimeric structures through PELDOR and AFM. The AFM data were acquired by Daniel Keppner, group of Michael Famulok (University of Bonn).

4.4.1 A defined dimer is formed upon ternary complex formation

The initial hint, that *Lbu*Cas13a possibly forms a dimeric structure came from the crystal structure of the ternary complex, solved by Liu et al.[87]. The asymmetric unit of the measured crystal contained two proteins with a crRNA and a target RNA strand bound to each of them (Figure 4.4-1 a). Interestingly, the 5'-end of the target RNA that is bound to the first protein (Figure 4.4-1 a, orange RNA strand) inserts into a protein cleft of the second protein (Figure 4.4-1 a, insert). This cleft is the target cleavage site and suggests a target RNA-protein interaction driving this dimerization. However, the dimer has not been seen in solution yet and it remained unclear if this is a crystallization artefact. Thus, this dimeric structure was labelled *in-silico* with MTSL at the positions 756 and 926 to predict if the intermolecular distances are in the PELDOR range and if it can be potentially verified with PELDOR measurements (Figure 4.4-1 b).

As seen in Figure 4.4-1 b, most of the intermolecular distance distributions are in the detectable PELDOR regime of up to 10 nm. So, for the experiments, the dprecr and dtarget single cysteine constructs 756 and 926 were mutated, expressed, purified and labelled (Appendix, Figure 8.2-9 for biochemical validation). These dprecr and dtarget single cysteine constructs were measured by PELDOR spectroscopy in their *apo*, pre-crRNA bound, and crRNA bound states.

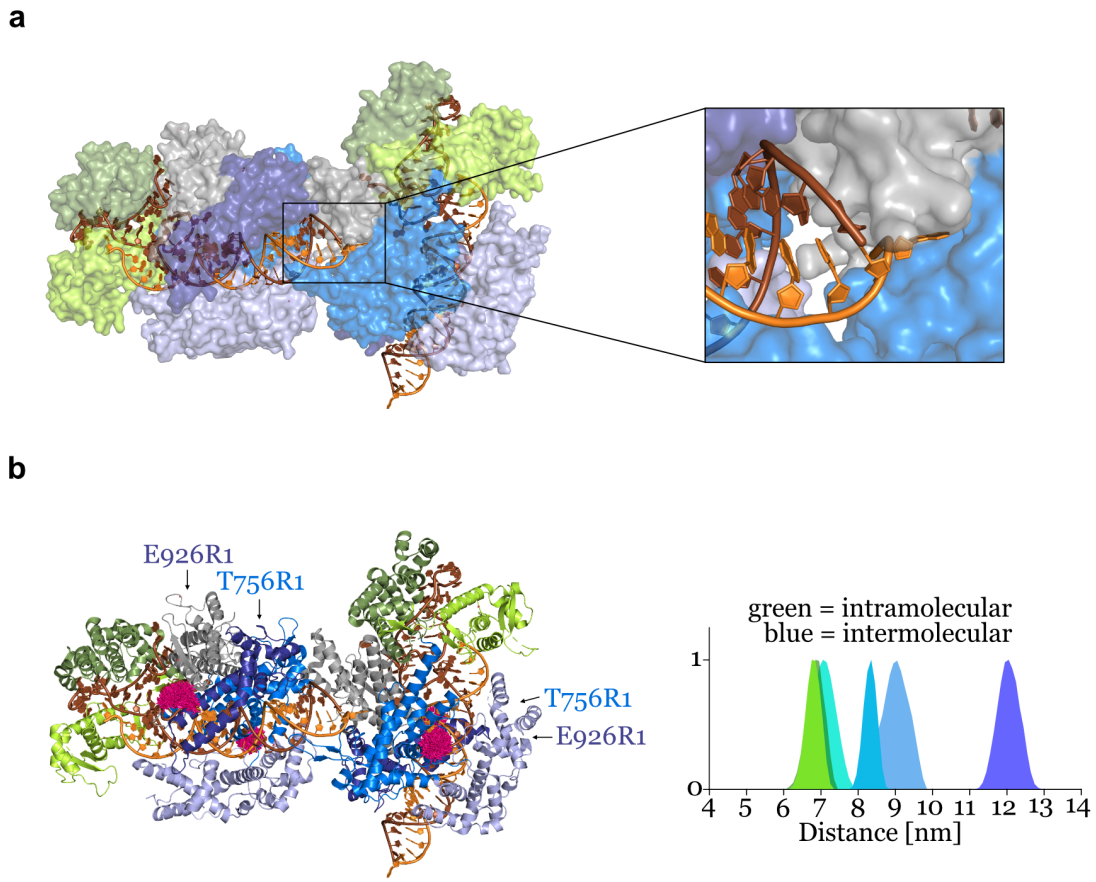


Figure 4.4-1: Dimer formation based on the crystal structure. a, Asymmetric unit of the crystal structure of the ternary *Lbu*Cas13a complex (PDB: 5xwp). The insert highlights the 5'-end of the target RNA, in orange that is inserted and bound in the active site of the second protein complex. The brown RNA strand is the crRNA. **b,** Dimer structure with spin label at positions 756 and 926. The predicted corresponding intra and intermolecular distance distributions are shown on the right.

The *apo* time traces (Figure 4.4-2 a, b) show no modulation depth and no oscillations, these are only straight decaying curves. This indicates that only monomers are present in the solution. A similar result is obtained when pre-crRNA is added to the single labelled dprecr constructs (Figure 4.4-2 c). The 926 construct shows no oscillations and no modulation depth. The 756 construct presents a small modulation depth but without visible oscillations. However, the signal to noise ratio is low and no CDA or Tikhonov analysis could be performed. When the crRNA is added to the dtarget construct, the modulation depth increases to 25% and 20% for the 756 and 926 constructs, respectively.

The maximal modulation depth that can be obtained with our experimental setup in a 4-pulse PELDOR experiment is 35%[184]. By taking a previously determined spin

labelling efficiency of 100% for the 926 construct and 94% for the 756 construct into account (Figure 8.2-9), and by taking a statistic analysis of the sample composition into account, the amount of dimer can be calculated to about 57% for the 926 construct and to about 79% for the 756 construct.

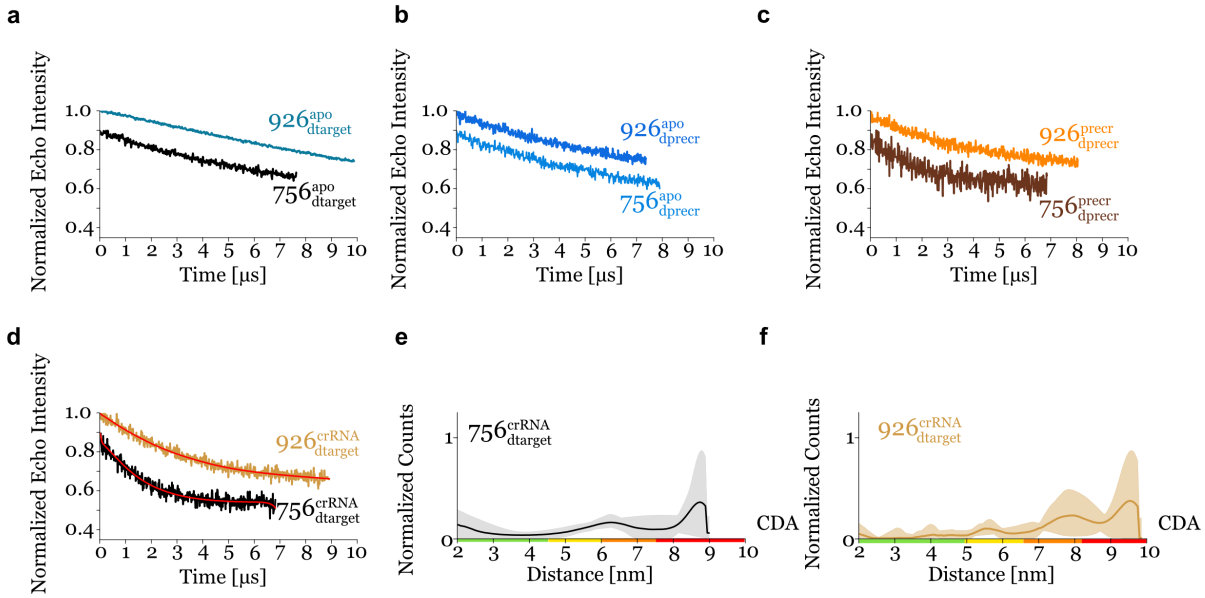


Figure 4.4-2: PELDOR spectroscopic measurements of single labelled 756 and 926 constructs. For visualization all time traces are shown with a y-axis offset of 0.1. **a,** time traces of *apo* dtarget constructs. **b,** Time traces of *apo* dprecr constructs. **c,** Time traces of pre-crRNA bound dprecr constructs. **d,** Time traces of dtarget constructs with crRNA including their fit in red that was done with the software DeerAnalysis[276], through Tikhonov regularization. **e,** Distance distribution of the black time trace shown in d. The distance distribution contains the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). **f,** Distance distribution of the yellow/beige time trace shown in d. The representation is the same as in e.

However, even though the amount of dimer is high, no defined distance is obtained after CDA analysis (Figure 4.4-2 f). Therefore, it is most likely that the dimeric or oligomeric structures formed are based on unspecific interactions forming unspecific oligomers or forming highly flexible oligomers.

When both dtarget single cysteine constructs were used to form their ternary complexes with cr- and target RNA a completely different picture is seen (Figure 4.4-3). The PELDOR measurements of the ternary complexes of the single cysteine constructs 756 and 926 (Figure 4.4-3, first two rows on top), show modulation depths of 24.5% and 25%, respectively. This corresponds to 78% and 71% dimer, by considering the labelling efficiencies (Figure 8.2-9).

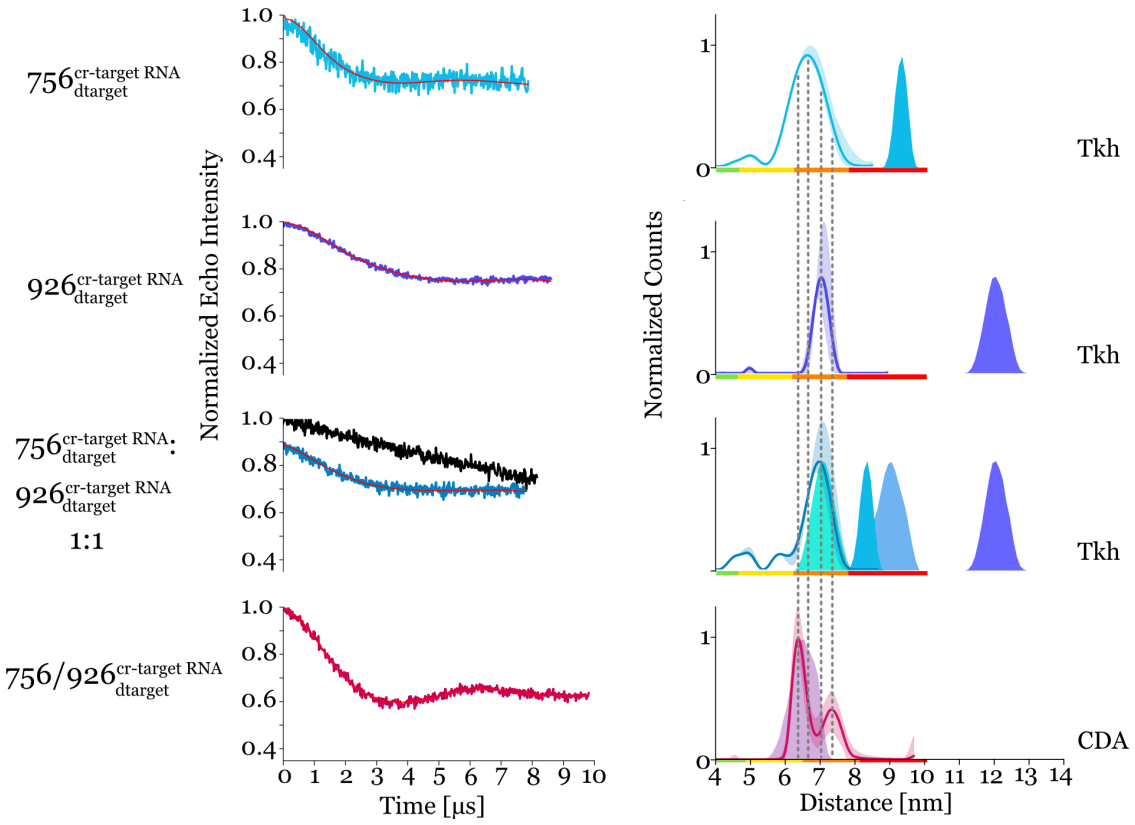


Figure 4.4-3: PELDOR results of the ternary complex with single labelled constructs 756 and 926. Time traces (left) and distance distributions (right) are shown for different constructs. From top to bottom, the dtarget constructs 756 and 926 are shown first, with crRNA and target RNA bound to them. The third row shows the data for a 1:1 mixture of the dtarget 756 and 926 constructs in *apo* (black) and with crRNA and target RNA (blue). The last row shows the dtarget double cysteine construct 756/926, also seen in Figure 4.2-3. The fit of the respective time traces was done with the software DeerAnalysis[276], through Tikhonov regularization (Tkh) or the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predicted distributions are shown as shaded areas and are the same as in Figure 4.4-1 b.

However, the major difference between the PELDOR measurements of the crRNA bound complexes (Figure 4.4-2) and the cr- and target RNA bound complexes are the oscillations seen in the time traces, resulting in defined distances seen in the distributions (Figure 4.4-3). The most probable distances are 6.6 nm and 7 nm for the 756 and 926 constructs, respectively. Interestingly, these distance distributions do not fit to the corresponding predicted distributions based on the dimer seen in the crystal structure (Figure 4.4-3, purple shaded distribution, on the bottom).

With these first two measurements using 756 and 926 constructs separately, only the distances between 756-756 or 926-926 can be measured. Thus, a mixture with a 1:1

ratio of the dtarget 756 construct and the dtarget 926 construct was prepared to measure also hetero-distances between the positions 756 and 926. The result is shown in the third measurement set in Figure 4.4-3.

The *apo* measurement of the 1:1 mixed construct shows no modulation depth (black curve). Hence, under these conditions, no dimer is formed. The experimental distance distribution of the 1:1 mixed sample with cr- and target RNA (Figure 4.4-3, third row) shows one monomodal distance distribution with a most probable distance at 7 nm, which fits to one predicted distance distribution (light blue shaded distribution). The expectations were that both distributions of 756-756 (Figure 4.4-3, top) and 926-926 (Figure 4.4-3, second row) should be seen with additional peaks for the cross distances between 756-926. But since this distribution lies entirely in the orange reliability interval, only the mean distance is reliable. The acquired time trace length seems to be too short to reliably detect several distances and shoulders that are maybe underlying the monomodal distance distribution. The distance difference between the most probable distances of 756-756 and 926-926 may additionally be too small for the time trace length acquired, to separate these when the dtarget constructs 756 and 926 are mixed.

Since the interspin distances of the single cysteine construct didn't match to the predicted distribution based on the crystal structure (Figure 4.4-3, bottom, purple shaded area), this dimeric state from the crystal structure is most likely not present in solution under PELDOR conditions.

Interestingly, the distance distributions of the single cysteine constructs do not match to the second (smaller) peak in the already previously described distance distribution of 756-926 (Figure 4.2-3 purple distribution, bottom, Figure 4.4-3, bottom). Furthermore, a second, smaller peak was also seen in the pre-crRNA bound PELDOR sample of 756/926 (Figure 4.2-3), which cannot be reproduced by the single cysteine PELDOR measurements with pre-crRNA (Figure 4.4-2 c). Thus, the question remains how the structure looks like that caused the second peaks in the 756/926 distance distributions. This will be covered in Chapter 4.4.3.

4.4.2 Validation of dimeric Cas13a with AFM

The PELDOR time traces shown above are acquired from solutions containing a protein concentration of 15 µM. Since this concentration is rather high, there is a high possibility for aggregation artefacts. To exclude that the dimerization seen in our PELDOR data origins from the high concentration, atomic force microscopy was used as a second orthogonal method.

As already described above, one advantage of AFM, compared to PELDOR spectroscopy is that no spin labelled protein is needed. This allows the use of the wt Cas13a, as well as Cas13a constructs bearing just the mutation R1079A, for pre-crRNA cleavage inhibition, and the mutations R1048A H1053A, for target RNA cleavage inhibition. These constructs contained all three native cysteines that were removed for selective spin labelling for PELDOR sample preparation. The second advantage of AFM is that the concentration used is in the picomolar range, which is significantly lower than the concentration in PELDOR samples.

Thus, several AFM measurements without and with RNA were performed and exemplary results are shown in Figure 4.4-4. In total, three AFM measurements were done, which are found in the Appendix in Figure 8.4-1. Generally, the AFM images are dominated by monomers in the samples of *apo* wt *Lbu*Cas13a, pre-crRNA bound dprecr *Lbu*Cas13a, and crRNA bound wt *Lbu*Cas13a, while the sample containing the cr- and target RNA bound dtarget *Lbu*Cas13a is dominated by dimers. The average amount of dimer with the standard deviation and the comparison to PELDOR spectroscopy is given in Table 4.4-1.

Table 4.4-1: Comparison between dimer amount in AFM and PELDOR.

Cas13a-complex	AFM [%]	PELDOR [%]
apo	7.2 ± 4	0
pre-crRNA bound	9.6 ± 0.7	35.3 ± 7.4
crRNA bound	8.7 ± 1	76.7 ±10.7
cr- and target RNA bound	60 ± 13.5	63.8 ± 5.7

Interestingly, the dimer amount in *apo* experiments and in the experiments with the ternary complex with AFM and PELDOR fits quite well. This indicates that indeed

almost no dimer is formed in solution when no RNA is added to *Lbu*Cas13a and that *Lbu*Cas13a predominantly adopts a dimeric structure when the ternary complex with cr- and target RNA is formed. In the PELDOR measurements we additionally see that the dimer has a specific and defined conformation.

However, the AFM and the PELDOR data of the complexes with pre-crRNA and crRNA are not in agreement. The AFM data of the pre-crRNA bound, and crRNA bound complexes show 9.6% and 8.7% dimer, while the PELDOR data suggest 35.3% and 76.7% dimer, respectively. This is an indication that the dimerization seen in PELDOR for the pre-crRNA bound complex and the crRNA bound complex are both concentration-driven and almost not present when analyzed with AFM. Additionally, as a second indication for this dimer to be an artefact or unspecific is that a very broad, undefined distribution is yielded for the crRNA bound complex in the PELDOR measurements (Figure 4.4-2 d-f) and no defined distribution for the pre-crRNA bound complexes could be obtained. So, these dimers are likely formed through unspecific interactions, without defined structure.

A second analysis that can be done with the AFM images is a shape analysis of *Lbu*Cas13a. Even though the resolution is too low to determine single domains in the protein structures, the overall shape of the protein is clearly visible (Figure 4.4-4, enlarged proteins on the left of each AFM images). Interestingly, it seems that the shape of monomeric particles changes upon RNA addition. Beginning with the enlarged images of the *apo* AFM measurement in Figure 4.4-4 a, it seems that these belong to two different monomeric orientations. In the lower red image, two dots can be seen, which could belong to the two protein lobes. This would fit to our PELDOR data, showing that these two lobes are flexible with respect to each other, and it strengthens the hypothesis of an additional hinge region between both lobes that leads to a wide-open structure (Chapter 4.3). The enlarged image on top shows a more compact conformation. Since a more intense signal was recorded (brighter color), this is potentially a different orientation of *Lbu*Cas13a. In AFM the height has a high resolution of 0.1-0.2 nm[128], thus it is likely that the first picture is showing Cas13a from the top (looking directly onto the REC lobe, NUC lobe is underneath) and the second picture is showing Cas13a from a side view, in which the two lobes are shown.

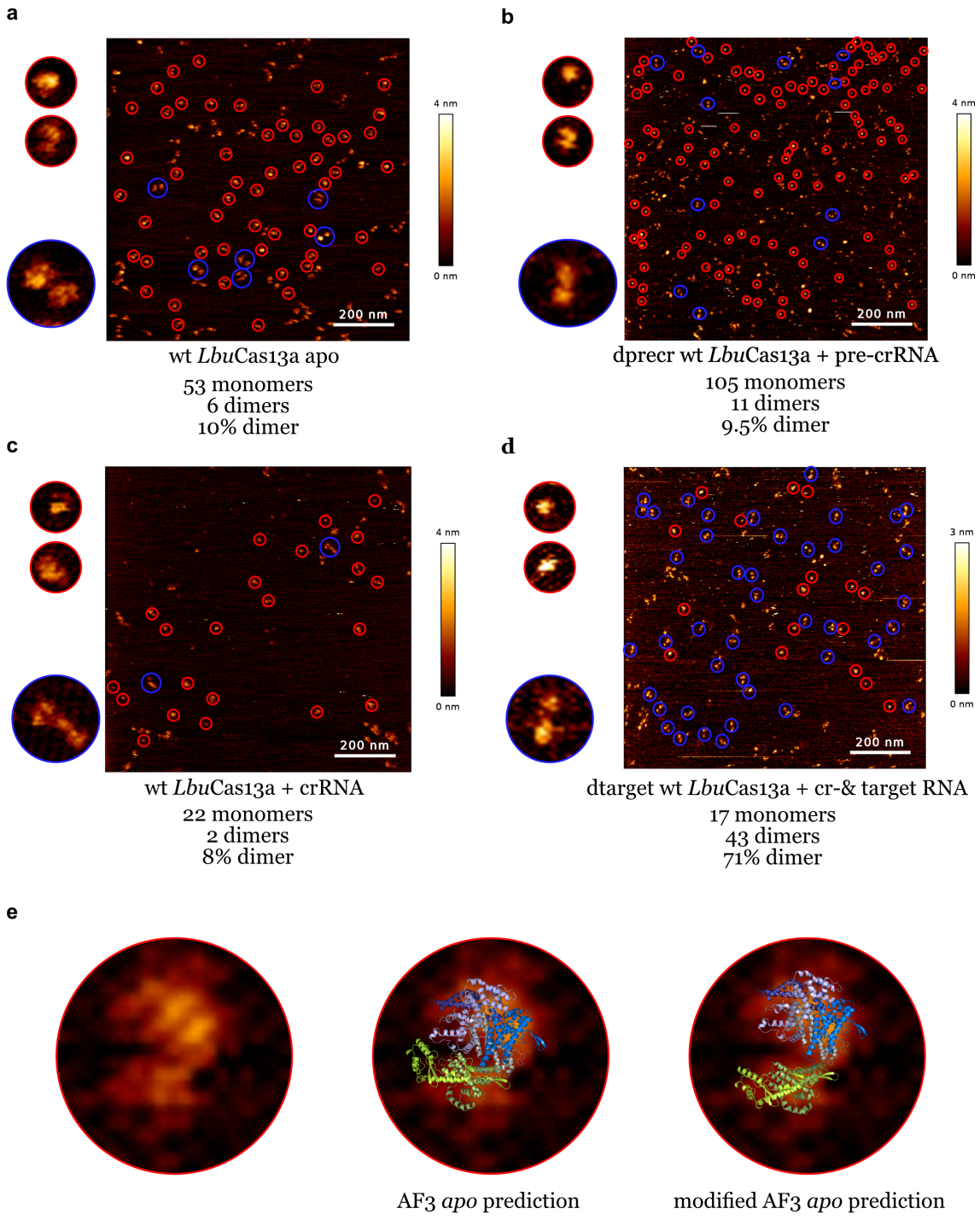


Figure 4.4-4: Exemplary AFM images of each *Lbu*Cas13a state. AFM images acquired from **a,** wt *Lbu*Cas13a without RNA. **b,** dprecr construct of wt *Lbu*Cas13a with pre-crRNA. **c,** wt *Lbu*Cas13a with crRNA. **d,** dtarget construct of wt *Lbu*Cas13a with cr- and target RNA. Blue circles highlight dimers and red circles highlight monomers. The inserts on the left of each AFM image show two monomer structures and one dimer structure in the same color coding. **e**, enlarged monomer particle from a (left), overlay of the particle shown on the left with the *apo* AF3 structure prediction of *Lbu*Cas13a (middle). This AF3 prediction was used to rotate the REC and NUC lobe away from each other, to open a cleft between both. This modified structure is shown on the right, as an overlay to the same recorded particle.

This enlarged image from the side-view of Cas13a, from Figure 4.4-4 a was overlayed with the AF3 predicted structure (Figure 4.4-4 e). Interestingly, when the AF3

structure is just adjusted in size, to fit the intensities seen in AFM, one can already see that the proportions of the particle fit to the AF3 prediction. However, by using the information obtained by PELDOR spectroscopy, the protein structure fits even better (Figure 4.4-4 e, right). To obtain this modified structure, the REC lobes was manually rotated by using PyMOL, since this lobe is flexible in *apo* (see measurements in Figure 4.3-1, Figure 4.3-2, and Figure 4.3-3, black distributions). This modified AF3 prediction was further used as an input structure for MtsslWizard, to test if this conformation is present in our PELDOR measurements. The distance distributions were predicted, based on this modified structure, for the three PELDOR constructs targeting the REC and NUC lobe flexibility. These constructs were $190/756^{apo}_{active}$ (Figure 4.3-1), $138/926^{apo}_{active}$ (Figure 4.3-3), and $190/756^{apo}_{active}$ (Figure 4.3-2), covered in Chapter 4.3. The results are showed in Figure 4.4-5. Interestingly, all distance distributions based on the modified AF3 model are shifted to larger distances, compared to the distributions based on the original AF3 model. Further, these modified distributions fit better to the measured PELDOR distributions, then those based on the original AF3 model. Thus, among a great variety of conformations, that the protein adopts in the *apo* state, the modified AF3 conformation could be one of these conformations and maybe one of the more predominant ones.

For the AFM images of the pre-crRNA and crRNA bound samples (Figure 4.4-4 b and c, red circles) there are proteins that seem to be globular and others that seem to be more open (in both, the more globular protein is shown enlarged on top and the more open is shown below). This could also be an indication, that the protein is present in apo and bound to pre-crRNA under the protein and RNA concentrations used in AFM samples. In contrast, most proteins to which the cr- and target RNA were added show an even more globular form. This is in line with our PELDOR measurements, showing that generally the *apo* protein is the most flexible and that it rigidifies upon RNA addition.

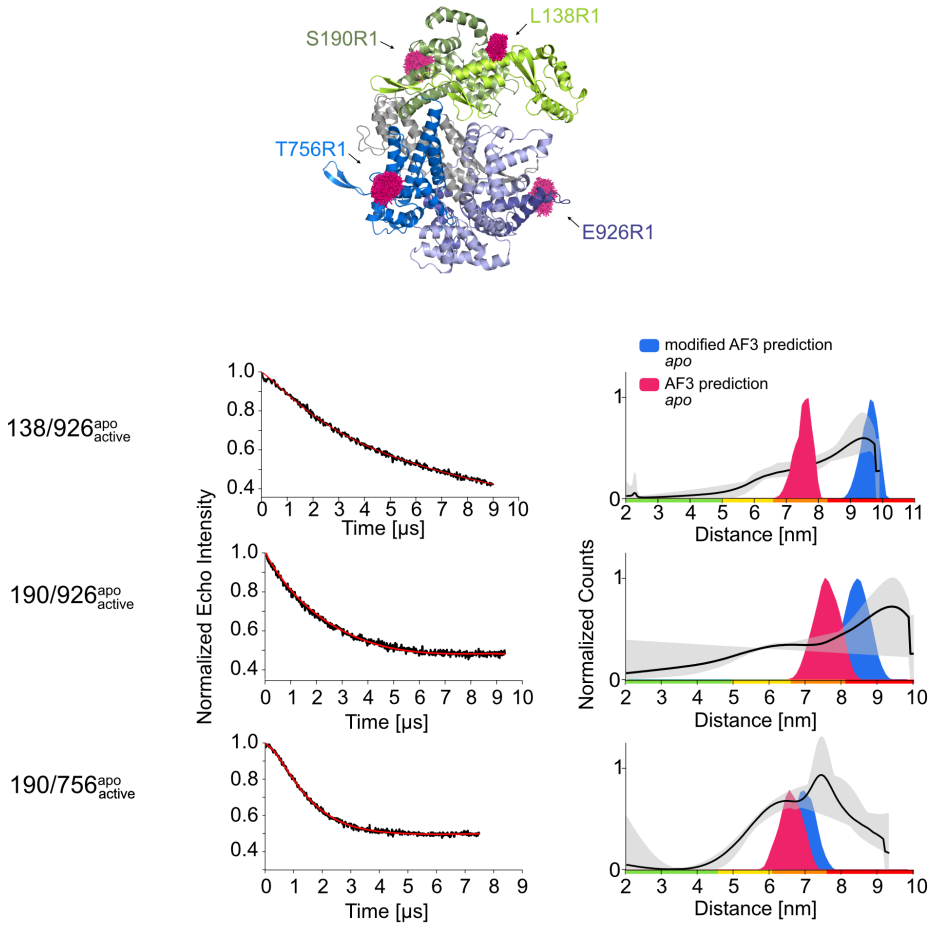


Figure 4.4-5: Comparison of the distance distributions based on the AF3 prediction and the manually modified AF3 prediction. The AF3 structure of *apo Lbu*Cas13a is shown on top, with magenta MTSL rotamer clouds at different labelling positions. The PELDOR time traces and distributions are shown from the *apo* state of the active double labelled protein constructs. The time traces include the fit that was done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. The distance distribution on the right contains the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The prediction of the distance distribution based on AF3 model of *Lbu*Cas13a, is shown as magenta shaded distribution and the prediction based on the modified AF3 model is shown as blue shaded distribution.

4.4.3 Possible dimer conformation

4.4.3.1 Predicting the dimeric structure with mtsslDock

In the Chapters above, PELDOR data indicated a defined oligomeric structure formed by the ternary *Lbu*Cas13a complex with cr- and target RNA. It was hypothesized that this could be a dimer, which was confirmed by AFM. However, the question remains

how this dimeric structure looks like, since the PELDOR distance distributions of the single cysteine constructs do not fit to the dimer seen in the crystal structure (Figure 4.4-3).

To develop a structural hypothesis the mtsslDock implementation in mtsslSuite[273] was used. The idea there is to use PELDOR distance distributions as distance constraints, to build a structural model. The input, used for the docking runs, is the most probable distance of the respective distance distribution and the width of the distribution (the standard deviation). Upon running the docking, one of the two rigid bodies is held fix, here one of the two Cas13a proteins, while the second rigid body is rotated and translated such, to yield the best fit to the PELDOR data. One aspect, that is important to highlight is that the number of constraints, being two, is very small to develop a reliable structure model. Thus, the docking that was done here should be rather considered as a hint of how the dimeric structure could look like.

In the first docking run, the distance constraints of the single cysteine constructs 756 and 926 were used. The output of this run is shown in Figure 4.4-6. The output of the mtsslDock are 20 different structures and interestingly all of them are very similar (Figure 4.4-6 a). The interspin distance results of the docking run, as shown in Figure 4.4-6 b as single lines, fit nicely to the most probable distance of the input distributions (shaded distributions, Figure 4.4-6 b).

Interestingly, when comparing the structure shown in Figure 4.4-6 c, and the dimeric structure from X-ray crystallography, depicted in Figure 4.4-1 a and c, it can be noted that they present substantial differences. The X-ray structure seems to be driven by RNA-protein interactions, while the structure predicted by mtsslDock seems to be driven by protein-protein interactions. Additionally, in contrast to an almost vertical arrangement in the X-ray structure, the proteins in the docked structure are aligned in a parallel fashion.

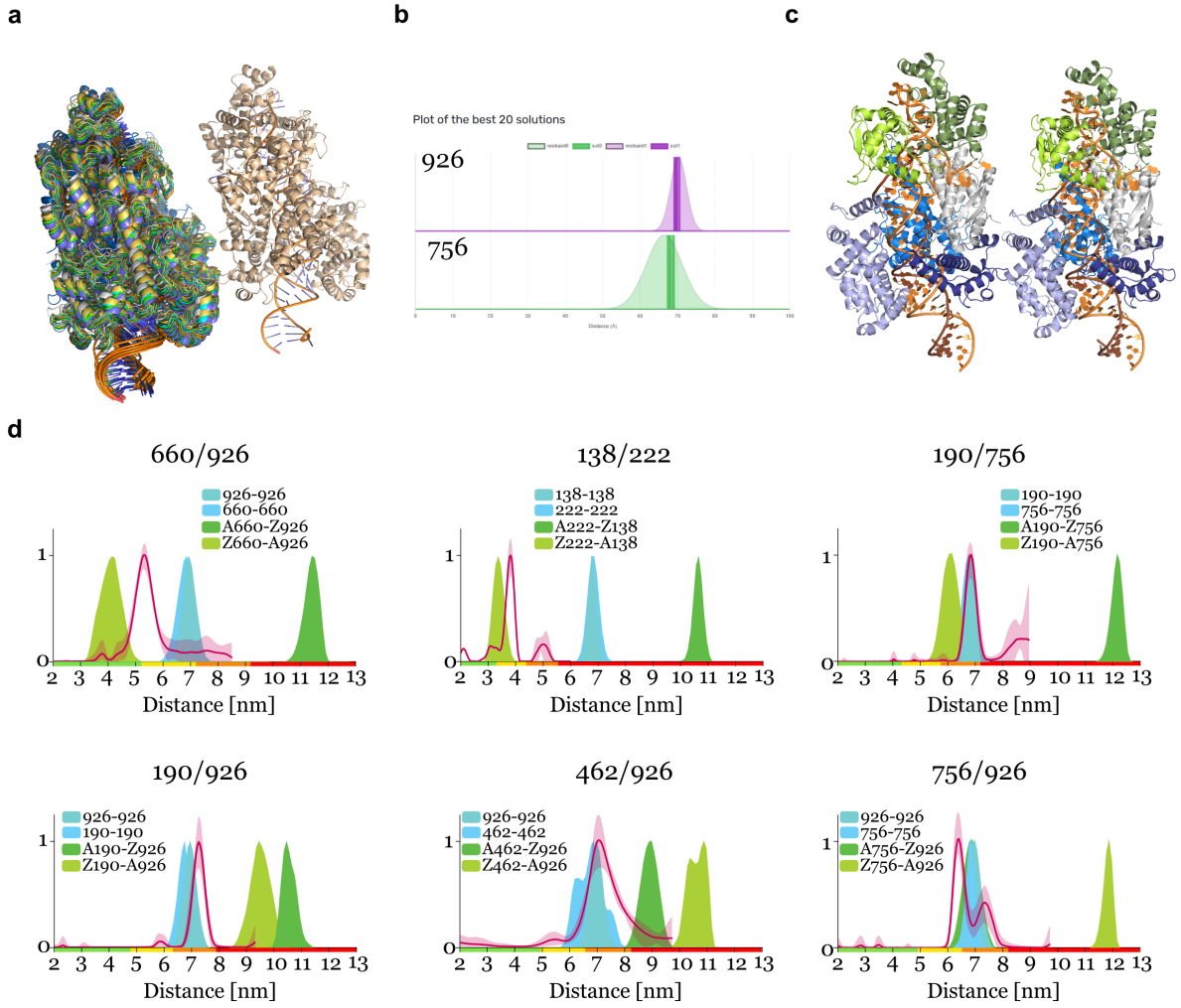


Figure 4.4-6: Docking run for dimer structure prediction. a, One docking run results with 20 different potential dimer structures. These were aligned using pymol. **b,** The distance constraints that were fitted during the docking run, shown as single lines are compared to the input constraints of the intermolecular distance distributions from the labelling positions 756-756 and 926-926. **c,** One of the 20 docking output structure is shown with the same color coding as in Figure 3.4-3. **d,** The dimer structure shown in c was labelled *in-silico* and all intermolecular distance distributions were predicted with the mtsslSuite between the spin labelling positions 756 and 926.

In Chapter 4.4.1, it was described that a second, smaller peak was seen in the crRNA and target RNA bound PELDOR sample of 756-926 (Figure 4.2-3), which cannot be fully reproduced by the single cysteine PELDOR measurements with crRNA and target RNA (Figure 4.4-3). If the structure from mtsslDock is the structure present in solution, it could maybe explain this second peak in our distance distributions. Thus, the dimer was *in-silico* labelled at the positions used for PELDOR spectroscopy and the intermolecular distance distributions were predicted (Figure 4.4-6 d) to understand why a second smaller peak is only visible in the distribution of 756/926. The results show that for some double labelled constructs, the intermolecular

distributions predicted, based on the docked dimer (Figure 4.4-6 b), do not fit to the experimental PELDOR distributions, as for the construct 660/926, or 138/222. The dark green distributions above 10 nm cannot be resolved in our PELDOR measurements. For other constructs, as 190/756, 190/926, and 462/926, one or two out of four distributions are in the region of the most experimental probable distance, hence it is difficult to detect these in PELDOR and to separate the intra- from the intermolecular distributions. Interestingly, in the construct 756/926, three out of four intermolecular distributions yield the same most probable distance in the region of the second smaller peak. It does not fit perfectly, but for example preferred label conformations or slightly different protein conformations in solution could yield a slight shift in the distance distributions. Since three out of four distributions yield the same most probable distance, the probability is higher that a distance distribution is detected in this region.

4.4.3.2 Analyzing the dimerization with AlphaFold 3

Recently, AF3 was published[96,167], which can predict the structure of protein-RNA complexes. Additionally, the algorithm was improved regarding multimeric predictions. Thus, a computation was run with two *Lbu*Cas13a proteins, two crRNA strands and two target RNA strands to determine if the AF3 calculation would predict RNA bound protein dimers. All five outputs are shown in Figure 4.4-7 a. What can be directly seen from the five models is that a dimer is predicted with interaction surfaces formed by protein-protein interactions. The 5 models can be grouped into three categories. The first is a parallel orientation of both proteins (model 4), which is similar to the result of the mtsslDock run but rotated (Figure 4.4-6). The second category shows an anti-parallel conformation of both proteins (model 0 and model 1). The third category predicts the two proteins to be tilted to each other (model 2 and model 3). Additionally, the RNA duplex in these two tilted models is distorted with an incomplete Franklin-Watson-Crick base pairing of the crRNA and target RNA at the 3' end of the target RNA, towards the crRNA hairpin. It is difficult to determine if this "distorted" protein-RNA structure is biologically relevant and if it is present in solution, especially considering that the average accuracy of AF3 for protein-RNA complexes is 39.4% (Table 1.7-1)[96].

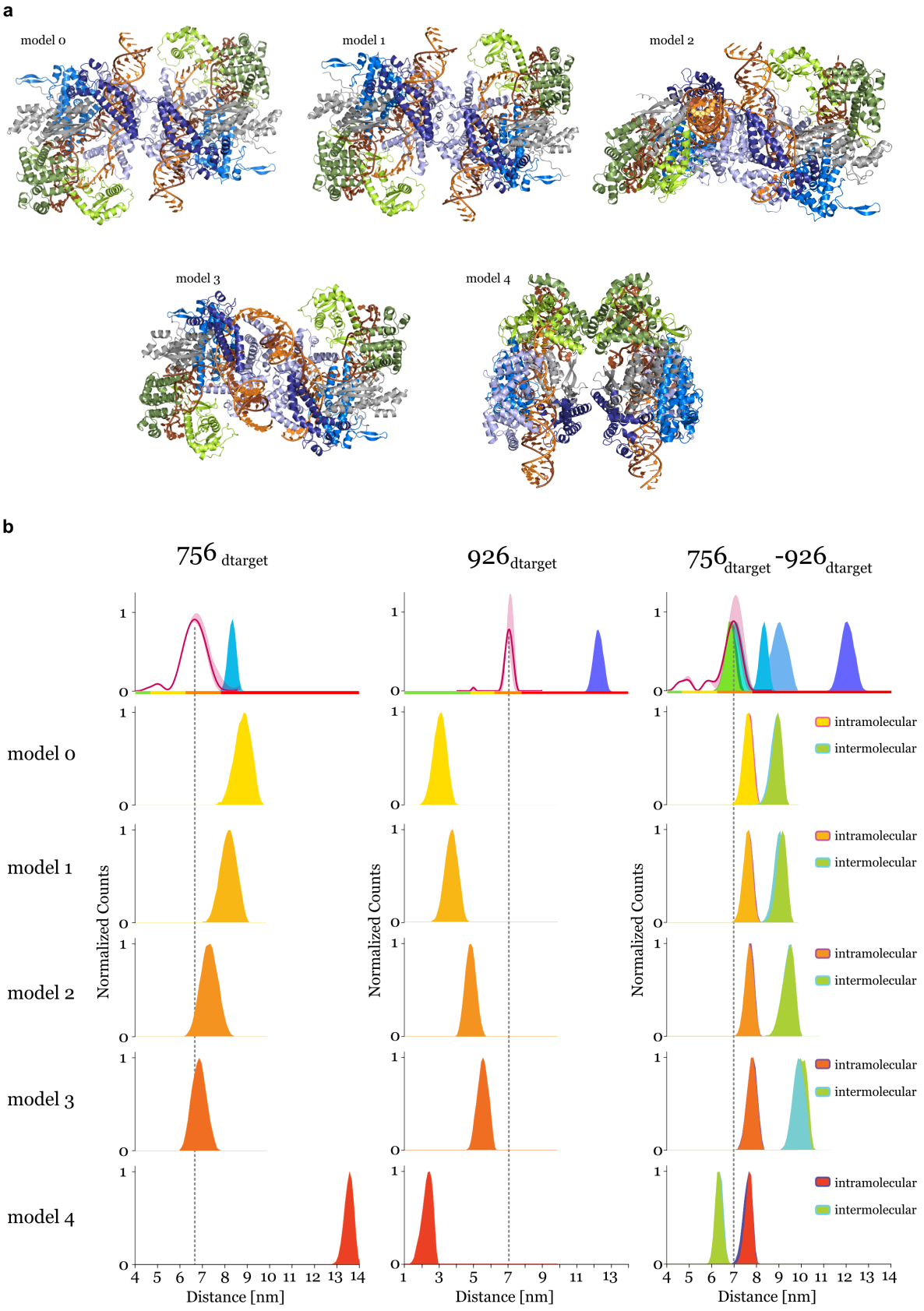


Figure 4.4-7: AF3 run predicting dimeric structures. a, Output structures from AF3 calculations. The nomenclature is the same as used in AF3 with model 0 being the most probable model and model 4 the least probable. **b,** Each protein in each AF3 model was labelled at the positions 756 and 926 and the predicted intra- and intermolecular distance distributions were predicted with the

mtsslWizard (shaded distributions). The experimental distance distribution from PELDOR data is shown on top with the predicted distance distributions from the crystal structure of the ternary complex (PDB: 5xwp). The experimental distance distributions contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable).

All 5 AF3 models were *in-silico* labelled at the positions 756 and 926 to determine, if those structures fit better to the structure seen in the PELDOR measurements, than the structure from mtsslDock. The results are shown in Figure 4.4-7 b. The three experimental distributions are shown on top, with the predictions of the intermolecular distributions as shaded areas, based on the crystal structure (Figure 4.4-1). Upon comparison of the predicted distributions based on AF3 models with the experimental distributions, it becomes clear that models 2 and 3 fit the best. Despite the distorted RNA duplex, apparently those two models present the best fit concerning the protein structure in the dimer structure.

However, when the predicted distributions based on AF3 model 3 (Figure 4.4-7) are compared to the result from the mtsslDock (Figure 4.4-6), the latter method seems to give the best fit.

4.5 Discussion

In this Chapter, PELDOR spectroscopy was used to characterize all *apo* and RNA bound states and the conformational changes, which *Lbu*Cas13a undergoes.

For the first time, our PELDOR data revealed that the REC lobe of *Lbu*Cas13a undergoes substantial changes and that it was found to be very flexible in *apo*. This enhanced flexibility probably explains why no *apo* structure of Cas13a from any organism is known in which parts of the REC lobe are not truncated. Because of the flexibility, also the AF3 model does not represent the soluble state of *Lbu*Cas13a in a sufficient manner, because it can only predict static structures and is unable to predict dynamic behavior of proteins[96].

The REC lobe rigidifies upon binding of the pre-crRNA to one conformation that stays the same upon crRNA, and cr- and target RNA addition (Figure 4.1-3). Interestingly, the predicted distances from 138/222 with RNA, based on the AF3 model, are the same

as for the *apo* protein. It is known for AF2 that it predicts ligand bound structures of *apo* proteins[160,161]. Evidence is shown in this thesis that also AF3 has the same tendency (Figure 4.1-3, Figure 4.3-1, and Figure 4.3-2).

As described above, the flexible REC lobe recognizes RNA in the first steps of the functional pathway of Cas13a. However, it remains elusive what the biological function of this flexibility is. One hypothesis is that the protein tries to compensate for different RNA structures. RNA is a highly dynamic and flexible molecule, which can adopt many conformations. The pre-crRNA and crRNA repeats form a stable hairpin structure, which is recognized and bound by the REC lobe[184,281]. But the spacer should be very flexible. Thus, one hypothesis is that Cas13a searches potential RNA ligands by offering the maximum number of conformations and to account for the variety of RNA conformations that can be bound. Additionally, the PELDOR measurements between the REC and NUC lobe (Figure 4.3-1, Figure 4.3-2, and Figure 4.3-3) showed that these lobes are flexible towards each other. Interestingly, especially in the AFM data of the wt *apo* protein (Figure 4.4-4 a, second red circular insert on the left) many particles are seen that are composed of two densities. These two dots could resemble the REC and NUC lobe of Cas13a being in a wide-open conformation, as seen by the overlap of the protein structure (Figure 4.4-4 e), which would support the idea of an increased accessibility of the RNA binding pocket.

Once the crRNA or pre-crRNA is detected and bound, the REC lobe is pushed into the immobilized RNA-bound conformation. This observation is again supported by the solved experimental structures, in which the main contacts between the REC and NUC lobe are mediated by the RNA sequences and less by direct domain-domain interactions. East-Seletsky et al.[75] showed that the inversion of the pre-crRNA hairpin sequence led to severe reduction of pre-crRNA cleavage. However, the deletion of one base pair in the hairpin led to a lower reduction in pre-crRNA cleavage, which may be an indication that *Lbu*Cas13a can cope with smaller conformational changes in the hairpin and maybe also in the spacer.

Interestingly, for Cas12a, which is the type V effector protein cleaving DNA, a similar behavior was seen. Its REC lobe is flexible in the *apo* state[275,282] and adopts distinct structures with RNA. In addition it was shown by Nguyen et al.[283] in a study of 23 Cas12a orthologs that these exhibit a strong adaptation behavior to different canonical and non-canonical crRNA sequences, which fits to the hypothesis for Cas13a.

A second possible explanation for the flexibility seen in the REC lobe of Cas13a is that this flexibility can enhance spacer acquisition. At least for Cas9 it was shown that its *apo* variant enhances spacer acquisition and that crRNA and tracrRNA act as attenuator for spacer acquisition (unpublished results). Thus, it would be interesting to test if Cas13a interacts with the adaptation machinery Cas1 and Cas2 from *Lbu*.

In contrast to the REC lobe, the NUC lobe is already preorganized without RNA. The conformational changes that were seen from the cryo-EM structure with crRNA to the crystal structure with crRNA and target RNA were also observed in the PELDOR measurements in frozen solution. However, in contrast to the experimental structures with RNA, it was found that the helical-2 domain adopts two different conformations in the *apo* state. The first conformation can be attributed to the pre-crRNA and crRNA bound structure and the second conformation to a rotated and slightly more open conformation, as found in *Lsh*Cas13a. When analyzing all PELDOR data of crRNA and pre-crRNA bound complexes it becomes evident that *Lbu*Cas13a adopts the same conformation when bound to the pre-crRNA and the crRNA. This is in-line with other Cas nucleases, as Cas12a[282], where the same trend was observed. However, since no complete structure of Cas13a bound to pre-crRNA exists, especially the conformation of the pre-crRNA in this complex remained elusive.

It was observed that the distance distributions of the crRNA bound complexes did not perfectly fit to the cryo-EM structure. Since two different crRNA sequences were used, the influence of the crRNA sequence on the structure of *Lbu*Cas13a was tested, which yielded the same *Lbu*Cas13a structure in frozen solution. Also, the ternary complexes with different cr- and target RNA sequences led to the same distance distributions. Thus, it is hypothesized that the structural differences seen between the distributions from PELDOR spectroscopy, and the predicted distributions based on the experimental crRNA bound structure can result from several different experimental conditions. Firstly, the buffers used for cryo-EM and PELDOR were different. The cryo-EM structure was solved with purified protein in 20 mM Tris-HCl, pH 7.5, and 100 mM NaCl[64], while the buffer for PELDOR measurements contained 20 mM HEPES (pH 7.2), 5 mM MgCl, 40 mM KCl, and 40 mM of L-glutamic acid and L-arginine. It may be that the higher salt concentrations and the amino acids in the PELDOR buffer led to the stabilization of different conformations. For RNA it is known that Mg^{2+}-ions are very important for RNAs folding[284,285]. Recently it was found for

Cas12a, for which target cleavage is Mg^{2+} dependent, that it's specificity is highly dependent on the Mg^{2+} concentration[286]. It is unknown if the different specificities at different Mg^{2+} concentrations originate from structural differences of the protein, but the importance of Mg^{2+} becomes evident. The target cleavage of *Lbu*Cas13a is also Mg^{2+} dependent and in contrast to the buffer conditions during structure determination with cryo-EM, Mg^{2+} ions were added to the buffer used for PELDOR spectroscopy.

The second difference between the experimental conditions in cryo-EM and PELDOR spectroscopy is the glassing agent/cryoprotectant and the freezing method. In PELDOR spectroscopy, deuterated glycerol was used as cryoprotectant and it is known that different cryoprotectants can influence protein conformations[119,214]. The cryoprotectant is needed to form a frozen vitrified solution, when the sample is flash-frozen in liquid nitrogen. For cryo-EM the sample sizes are smaller, and the freezing is much faster than for PELDOR, hence, no glassing agent is needed[287]. However, in cryo-EM, proteins at the air-water interface on the cryo-EM grid can cause difficulties. There, the protein can adopt preferred orientations, can undergo conformational changes or can even be structurally damaged, influencing the final protein structure[288]. In addition to experimental conditions, local rearrangements in the experimental structure can be a possible explanation for the observed discrepancy. When certain residues in a protein are more flexible than others, this flexibility could lead to a lower local resolution. In all three measurements (blue distributions for 756/926 in Figure 4.2-3, 462/926 in Figure 4.2-4, and 660/926 in Figure 4.2-5) the position 926 is present. By manually checking the electron density in this region, (EMBD: EMD-6777) it was seen that the labelling position 926 is nicely resolved and the local resolution is sufficient to structurally describe this part of the protein. This means that local rearrangements at this position are unlikely to be the reason for the observed discrepancy. Thus, different buffers and different freezing methods could explain the slight differences in the protein conformations.

During PELDOR measurements of the 756/926 construct (Figure 4.2-3) a second peak was observed for the pre-crRNA bound and the cr-and target RNA bound structure. Initially, these peaks were thought to be formed by a dimer. The measurements of the ternary single labelled protein constructs (Figure 4.4-3) revealed a defined dimeric structure, but this distance distribution does not fit to the distance of the smaller peak seen in 756/926. Thus, it remains elusive where the second peak in the labelled

756/926 construct origins from. One possibility is that the spin labels form interactions with the protein surface and that two preferred orientations are adapted, leading to a bimodal distance distribution. Another possibility could be conformational changes in the protein leading to different label conformations. However, the smaller peak is only seen in those protein complexes, that undergo RNA cleavage. Thus, this peak could also origin from a second protein conformation, important for RNA cleavage. This is additionally likely, since this thesis shows that the conformation of the pre-crRNA bound state and of the crRNA bound states are very similar, if not identical. If this smaller peak origins from spin label conformations, these conformations should also be visible in the cr-RNA bound complex. But to finally attribute this smaller peak, further measurements are needed.

AFM images were acquired to validate the formation of a dimer with crRNA and target RNA. Even though most of the particles could be successfully grouped into monomers or dimers, it is difficult to determine specific conformations of the detected proteins. In Figure 4.4-4 a, as an example, two monomer particles are shown enlarged on the left (contoured in red). The second particle is divided into two parts. These two parts can either be the REC and NUC lobe, or the two REC lobe domains from the top. Since the particle in the first insert is brighter, hence its height is larger, it is likely that the view of the second insert is a side view and that the two dots belong to the REC and NUC lobes. To validate which orientation the protein has adopted on the surface, one method would be to measure the particle size. However, the lateral particle size heavily depends on the tip of the cantilever. During the measurements shown above and in the Appendix (Figure 4.4-4 and Figure 8.4-1) the old tips were sometimes replaced by a new one, which resulted in smaller particle sizes and makes the size determination inaccurate. Interestingly, the two dots in the second enlarged AFM image in Figure 4.4-4 a would fit to the hypothesis of a hinge region, which opens and closes a cleft between the REC and NUC lobes. When looking closely into the AF3 prediction of *apo* *Lbu*Cas13a, both REC lobe domains undergo almost no interaction with the NUC lobe. There are α-helices in the NTD that can interact with the helical-2 domain, but generally no strong interaction can be seen. This interaction can also explain why two conformations of the helical-2 domains are adopted by *Lbu*Cas13a (Chapter 4.2.3, Figure 4.2-5 and Figure 4.2-6). The first one equaling the crRNA bound state, which is the open state, with almost no helical-2 – NTD contact. The second one would be a

state in which the helical-2 domain rotates closer to the NTD to stabilize the closed conformation.

When the number of dimers in AFM images were compared to the amount from PELDOR (Table 4.4-1), a discrepancy is observed for the pre-crRNA bound and the crRNA bound complexes. As for the discrepancies seen in the distance distributions of the cryo-EM and crystal structure, the final buffer used for AFM is different than the one used for PELDOR spectroscopy. AFM images are acquired at room temperature, with immobilized particles on the surface, while PELDOR samples were measured at 50K, in frozen solution. In addition, the concentration of the PELDOR samples, being in the μM regime, is significantly higher than the protein concentration during acquisition of AFM images, being in the pM regime. Interestingly, these differences are valid for all four states, the *apo* state, the pre-crRNA, the crRNA, and the cr- and target RNA bond states. But only the pre-crRNA and crRNA bound states show this discrepancy. Thus, maybe the increased amount of aggregation is a complex-specific artefact that originates from buffer compositions, freezing method and a high concentration.

The dimer identified in AFM images was tried to model *in-silico* by using mtsslDock. For this, the PELDOR constraints of the single labelled protein constructs were used as input. But generally, these two constraints are too few to determine a reliable dimer structure. Another study [272] has shown that the correct protein structure was calculated only when 6 or more constraints were used. Interestingly, the structure predictions calculated by AF3 (Figure 4.4-7) have a worse fit to the PELDOR data. However, these AF3 predictions were calculated in an unbiased manner, in contrast to the mtsslDock prediction that is based on the input constraints. Even though the AF3 predictions show different dimer structures with interaction surfaces, they are calculated with a low average accuracy of 36.4% (Table 1.7-1)[96]. One possibility to work towards a reliable dimer structure would be a combination between adding more distance constraints to the mtsslDock computations and to perform a combination of mutational and conservation analyses on potential interaction surfaces. Most of the interaction surfaces of the AF3 predicted dimer models are composed of both helical-2 domains. Thus, one possibility is to try to disrupt this interaction surface by point mutations on the helical-2 domain surface. Further, information about the overall shape of the dimer could be yielded through SAXS, which could be additionally used as a constraint for mtsslDock[125].

Summary and Conclusion

The first result Chapter of this thesis deals with the establishment of basic biochemical techniques for handling *Lbu*Cas13a. These ranged from protein expression and purification, RNA transcription and the development of reproducible cleavage assays. For protein expression, *E. coli* BL21 AI cells were identified as most suitable because they showed reproducible protein expression with high yields. In the purification of *Lbu*Cas13a, the heparin column was showed to be of essential importance. The heparin column ensured high-quality *Lbu*Cas13a isolation and, more importantly, removed oligonucleotides that were bound to *Lbu*Cas13a during expression and purification. Functionality assays, consisting of pre-crRNA cleavage assays and target RNA cleavage assays were successfully established. Here addition of tRNA in pre-crRNA cleavage assays was found to increase cleavage, presumably by decreasing non-specific interactions between the protein and the pre-crRNA.

The second part of this thesis focused on the development of active, double labelled *Lbu*Cas13a constructs. It was seen that the three native cysteines in wildtype *Lbu*Cas13a were labelled, despite their low solvent accessibility. These cysteines were replaced by alanine and serine, but this resulted in an almost complete loss of the proteins cleavage activity. Through a stepwise mutational analysis, it was found out that the cysteine C348 was particularly important for pre-crRNA cleavage, maybe due to the stabilization of the protein's active structure. A cysteine-free construct, C293A C348V C1141A was found with the software dezyme and with protein homology analysis, which retained pre-crRNA and target RNA cleavage activity. Upon further exchange of two amino acids to cysteines at different domains in *Lbu*Cas13a, several active double labelled *Lbu*Cas13a constructs were obtained for structural investigation via PELDOR spectroscopy.

During PELDOR sample preparations with RNA, protein precipitation was observed. Out of different additives, a combination of 50 mM L-glutamic acid and 50 mM L-arginine was found to minimize or completely inhibit protein and RNA precipitation. These amino acids were used as supplements for all buffers related to experiments as PELDOR measurements and cleavage assays.

The last Chapter of this thesis is focused on structural and dynamic insights of *Lbu*Cas13a. New mechanistic aspects of *Lbu*Cas13a were added to the known mechanistic scheme shown in Figure 4.1-1 and the result is shown in Figure 5-1.

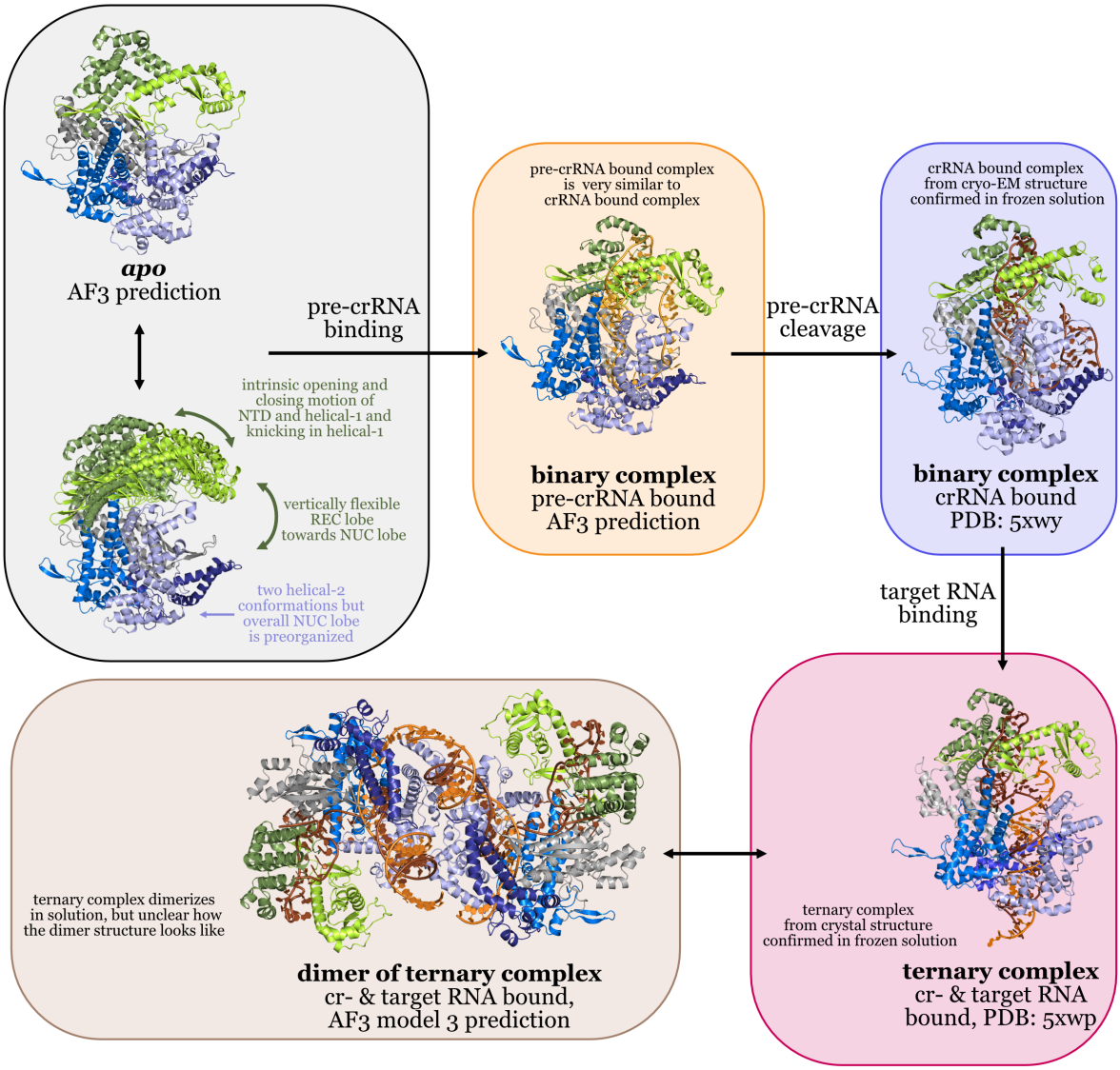


Figure 5-1: Hypothesized mechanism for *Lbu*Cas13a with predicted and experimental structures.

Beginning with the REC lobe of *Lbu*Cas13a, it was found that it is very flexible in its *apo* state. Upon analysis of each REC lobe domain individually it was seen that the NTD is rigid and that the helical-1 domain is intrinsically flexible. The high extend of protein flexibility most likely origins from a hinge region between both REC lobe domains, causing an opening-and-closing motion like a seashell, and a kinking region in the helical-1 domain. Thus, the *apo* REC lobe adopts many conformations and upon

addition of RNA, the two REC lobe domains close at the hinge region and additionally, the kink region closes, locking the RNA into place. Upon pre-crRNA and crRNA binding, the REC lobe drastically rigidifies and adopts one specific conformation that remains the same throughout the conformational pathway.

In contrast, the NUC lobe is preorganized already in *apo* and adopts one conformation, one exception being the helical-2 domain. The helical-2 domain adopts two distinct conformations in *apo*, which could be attributed to the structure of the binary complex and to a structure in which the helical-2 domain is rotated towards the REC lobe. A similar rotation of this domain was observed in the *apo* crystal structure and AF3 prediction of *Lsh*Cas13a. This rotational motion is the same as the one observed in the helical-2 domain between the binary and ternary complexes of *Lbu*Cas13a.

The PELDOR data in which one spin label is located in the REC lobe and the second spin label in the NUC lobe, showed that both lobes are flexible towards each other in *apo*. As also valid for the REC lobe alone, upon RNA addition it was seen that the REC and NUC lobes rigidify.

By considering all PELDOR measurements, it was seen for the unknown pre-crRNA bound complex that it adopts a very similar conformation as the crRNA bound complex in frozen solution. The crRNA bound complex is confirmed to be similar to the experimental cryo-EM structure. The distance changes that are expected from the crRNA bound binary complex to the cr- and target RNA bound ternary complex are also seen in frozen solution. The ternary complex itself was also confirmed.

Interestingly, a dimer of the ternary complex is formed. This dimer was not observed for *apo*, pre-crRNA and crRNA bound complexes and was confirmed to be a dimer through AFM measurements. A structural model was tried to be developed with PELDOR constraints for mtsslDock and with AF3. However, the in-solution structure remains elusive and further experiments need to be done to reliably determine this structure.

Materials and Methods

This Chapter describes all the materials and methods used during the preparation of this thesis. Several methods have been optimized, such as protein expression and purification. For these methods and for the according solution or buffer compositions, only the optimized protocols are described, unless otherwise noted.

Materials

6.1 Consumables

Table 6.1-1: Chemicals.

Chemicals	Manufacturer / Supplier
Acrylamide (Rotiphorese Gel 30)	Carl Roth
Acrylamide with Urea (Rotiphorese Sequenziergelkonzentrat, 25%)	Carl Roth
Agar	AppliChem
Agarose	Carl Roth
Ampicillin, sodium salt	Carl Roth
APS (Ammonium peroxydisulfate)	Carl Roth
Boric Acid	Chemsolute / Th. Geyer
Coomassie Brilliant Blue R-250	Carl Roth
D_2O	Deutero GmbH
dNTPs	New England Biolabs
EDTA	Carl Roth
Ethanol	Fisher Scientific
Ethidiumbromide	AppliChem
Glycerol	VWR Chemicals
Glycerol D_8	Merck KGaA
HEPES	Carl Roth
Imidazole	Carl Roth
IPTG(Isopropyl-β-D-1-thiogalactopyranoside)	Carl Roth
KCl	Carl Roth
L(+)-arabinose	Carl Roth
L-arginine	Carl Roth
L-glutamic acid	Carl Roth
$MgCl_2$	Carl Roth

NaCl	Carl Roth
NaOAc	Carl Roth
PMSF	Carl Roth
SDS, 20% solution	AppliChem
TCEP	Carl Roth
TEMED	Carl Roth
Tris	Carl Roth
Tris-base	Carl Roth
Trypton	Carl Roth
Yeast extract	Carl Roth
β-Mercaptoethanol	AppliChem
TMB solution	Pierce

Table 6.1-2: Labels.

Labels	Manufacturer / Supplier
MTSSL (1-Oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl) methanethiosulfonate)	Sigma-Aldrich
SLIM (short-linked maleimide)	self-made
3-maleimid-proxyl	Sigma-Aldrich

Table 6.1-3: Enzymes and enzyme buffers.

Enzymes and Supplements	Manufacturer / Supplier
Pfu Polymerase	Thermo Fisher Scientific
Rxn buffer (supplied with Pfu)	
Q5 Polymerase	Thermo Fisher Scientific
5x Reaction buffer (buffer supplied with Q5)	
DpnI	New England Biolabs
Tango buffer (supplied with DpnI)	
XbaI	Thermo Fisher Scientific
10x Fast Digest buffer (FD)	Thermo Fisher Scientific
TEV protease	self-made
T7 RNA polymerase	Thermo Fisher Scientific
T7-reaction buffer (supplied with T7)	
Ribolock	Thermo Fisher Scientific
DNAse1	Thermo Fisher Scientific
10x DNAse buffer (supplied with DNAse1)	
Proteinase K	Thermo Fisher Scientific
Yeast total tRNA	Thermo Fisher Scientific

Table 6.1-4: Further material and consumables.

Material	Usage	Manufacturer/Supplier
X-band tube	*cw*-EPR Spectroscopy	Wilmad-LabGlass

X-band capillary, 10 µL	*cw*-EPR Spectroscopy	Hirschmann Laborgeräte
Q-band tube	PELDOR Spectroscopy	Wilmad-LabGlass
Vivaspin Concentrator	Protein purification	Sartorius
Amicon Ultra Concentrator	Protein purification	Merck
Millipore Membrane Filter, 0.22µm pore size	Buffer filtration	Merck
GeneJET Plasmid Miniprep Kit	DNA extraction	Thermo Fisher Scientific
GeneJET PCR Purification Kit	DNA purification	Thermo Fisher Scientific
Mica surface	AFM	75 mm x 25 mm, 15mm V-3
Ultra short cantilevers high-speed AFM	AFM	Nanoworld

6.2 Solutions, buffers, and gels

All solutions, buffers, and gel compositions were prepared by using double deionized water from a Millipore MilliQ-direct water purification system. All buffers are filtered through a 0.22 µm disposable filter or syringe filter (Merck, Table 6.1-4) and degassed. In the following Table only the optimized buffers are included.

Table 6.2-1: Compositions of buffers and solutions.

Buffers and Solutions	Usage	Composition
Binding	Cas13a purification	50 mM Tris-Cl (pH7), 500 mM NaCl, 1 mM TCEP, 10 mM Imidazole, 2.5% Glycerol
Elution	Cas13a purification	50 mM Tris-Cl (pH7), 500 mM NaCl, 1 mM TCEP, 1 M Imidazole, 2.5% Glycerol
Gel Filtration (GF)	Cas13a purification	50 mM Tris-Cl (pH7), 500 mM NaCl, 1 mM TCEP, 2.5% Glycerol
High Salt	Cas13a purification	50 mM Tris-Cl (pH7), 250 mM NaCl, 1 mM TCEP, 2.5% Glycerol
Low Salt	Cas13a purification	50 mM Tris-Cl (pH7), 100 mM NaCl, 1 mM TCEP, 2.5% Glycerol
cw	Cas13a labelling	20 mM HEPES (pH7.2), 5 mM $MgCl_2$, 50 mM KCl, 2.5% Glycerol, 50 mM L-Glutamic Acid, 50 mM L-Arginine
PELDOR	Cas13a PELDOR sample preparation	20 mM HEPES (pH7.2), 5 mM $MgCl_2$, 50 mM KCl, 50 mM L-Glutamic Acid, 50 mM L-Arginine in D_2O
Cleavage Buffer	Cas13a cleavage assays	20 mM HEPES (pH 7), 50 mM KCl, 5 mM MgCl2, 50 mM L-Glutamic Acid, 50 mM L-Arginine

Sample Buffer (4x)	SDS-PAGE	240 mM Tris (pH8.3), 40% Glycerol, 8% SDS, 5% β-Mercaptoethanol, 0.04% Bromphenol blue
SDS Running Buffer	SDS-PAGE	25 mM Tris (pH 8.3), 192 mM Glycine, 2% SDS
Coomassie Staining Solution	SDS-PAGE	2.5g/L Coomassie R250, 450 mL/L Methanol, 100 mL/L Glacial Acetic Acid
TAE Buffer (50x)	Agarose Gel	2M Tris-HCl, 0.5 M EDTA, 5.71% Glacial Acetic Acid
Agarose Gel Sample Buffer (6x)	Agarose Gel	1 M Tris-HCl, 0.5 M EDTA, 60% Glycerol, 0.03% Xylencyanol, 0.03% Bromphenol blue
TBE Buffer (10x)	Urea-PAGE Running Buffer	890 mM Tris-Cl, 890 mM Boric Acid, 20 mM EDTA
Transfer Buffer	Western Blot	25 mM Tris-HCl (pH8.3-8.5), 192 mM Glycin, 0.05% SDS
TBST Buffer	Western Blot	50 mM Tris-HCl (pH7.5), 150 mM NaCl, 0.1% Tween 20
Blocking Solution	Western Blot	5% milk powder in TBST Buffer
AFM Buffer	AFM	10 mM Tris-Cl, 50 mM NaCl, 10 mM $MgCl_2$, pH 7.5

Table 6.2-2: Compositions of various gels for analysis of proteins and oligonucleotides.

Gel	Usage	Composition
Stacking Gel	SDS-PAGE	0.17 mL/mL 30% Acrylamide mix, 0.13mL /mL 1 M Tris (pH 6.8), 10% SDS, 10% Ammonium persulfate, 0.1% TEMED
10% SDS-Gel	SDS-PAGE	0.33 mL/mL 30% Acrylamide mix, 0.25mL /mL 1.5 M Tris (pH 6.8), 10% SDS, 10% Ammonium persulfate, 0.4% TEMED
20% Urea-Gel	RNA Cleavage Assay	0.8mL/mL 25%Acrylamide in 8.3 M Urea, 0.1mL /mL 8.3 M Urea, 0.1 mL /mL 8.3M Urea in 10x TBE Buffer, 10% APS, 0.5% TEMED
1% Agarose Gel	Agarose Gel	1% Agarose in 1x TAE Buffer

6.3 Media

Table 6.3-1: Media composition.

Medium	Composition
LB	10 g/L Tryptone, 5 g/L Yeast extract, 10 g/L NaCl, pH7
2YT	16 g/L Tryptone, 10 g/L Yeast extract, 5 g/L NaCl, pH7

6.4 Bacterial strains

Table 6.4-1: Bacterial strains.

Bacterial Strains	Usage
DH5alpha	Transformation
BL21 AI	Cas13a Expression
BL21 (DE3)	Expression Tests
MC1061	Expression Tests
XL1 Blue	Expression Tests
Rosetta	Expression Tests
C43 (DE3)	Expression Tests

6.5 Plasmids and oligonucleotides

Table 6.5-1: Plasmids and their features.

Plasmids	Features	Origin	Usage
p2CT-His-MBP	N-terminal MBP- and His_6-Tag, TEV cleavage site, Ampicillin Resistance, Lac Promoter	Addgene plasmid #83482, J. Doudna Lab	Cas13a Expression
pEX-A128	Ampicillin Resistance	Eurofins	pre-crRNA transcription
pUltraCNF/pAcF (pEx1)	Spectinomycin Resistance, Lac Promoter	Peter Schulz, Scripps research institute	Cas13a Expression with UAA
pEvolAcF/pAcF (pEx2)	Chloramphenicol Resistance, pBAD Promoter		
pEvolAzF/pAzF (pEx5)	Chloramphenicol Resistance, pBAD Promoter		

The RNA strands were purchased either at Metabion international AG or at Biomers GmbH.

Table 6.5-2: RNA sequences.

Name	RNA Sequence	Origin
pre-crRNA	GGA UUU AGA CCA CCC CAA AAA UGA AGG GGA CUA AAA CAG GGG CAG AGA UGA UGA CCC U	Adapted from East-Seletsky et al. 2016
crRNA	GGC CAC CCC AAA AAU GAA GGG GAC UAA AAC ACA AAC AUG AUC UGG GUC AUC	East-Seletsky et al. 2016
target RNA	GAA GAU GAC CCA GAU CAU GUU UGA GAC CU	Adapted from East-Seletsky et al. 2016
$crRNA_{crystal}$	GGACCACCCCAAAAAUGAAG GGGACUAAAACACAAAUCUAUCUGAAUAAACU CUUCUUC	Liu et al. 2017
target $RNA_{crystal}$	GGAAGA AGAGUUUAUUCAGAUAGAUUUGUC	Liu et al. 2017

The primers used for mutagenesis were designed using the software Geneious and purchased either at Eurofins Scientific, or Microsynth AG.

Table 6.5-3: DNA primer sequences for QC-PCR or standard amplification PCR.

Primer	Sequence
*Lbu*Cas13a C293A	GAAGAGTTGAACGACAAGAACATCAAATACGCGTTTGCACATTTCG CAGCAACTGACTCATTTCGATTTCCACGAAATGTGCAAACGCG
*Lbu*Cas13a C348A	CTGTTAAACAAACTTGACACGTACGTCCGTAATGCAGGAAAGTATA ATT GGCAATTTCGCCGTCTTGCAAATAATAATTATACTTTCCTGCATTAC G
*Lbu*Cas13a C1141A	GAAGTTAATGACTGACCGCAATTCCGAGGAACTTGCAAAATTGGT TTTTTCTCTTCCATTTTGTATTCAAACATAATCTTCACCAATTTTGC AAGTTCCTC
*Lbu*Cas13a C293S	GAAGAGTTGAACGACAAGAACATCAAATACGCGTTTAGCCATTTCG CAGCAACTGACTCATTTCGATTTCCACGAAATGGCTAAACGCG
*Lbu*Cas13a C348S	CTGTTAAACAAACTTGACACGTACGTCCGTAATAGCGGAAAGTATA ATT GGCAATTTCGCCGTCTTGCAAATAATAATTATACTTTCCGCTATTAC G
*Lbu*Cas13a C1141S	GAAGTTAATGACTGACCGCAATTCCGAGGAACTTAGCAAATTGGT TTTTTCTCTTCCATTTTGTATTCAAACATAATCTTCACCAATTTGCT AAGTTCCTC
*Lbu*Cas13a E32Amber	AAGTCAGAATCGGAAGAAAATCGCACAGACTAGCGTCTG CGCATATTAAGCAACGCCGACAGACGCTAGTCTGTG

*Lbu*Cas13a His-del-N-terminus	GAAATAATTTTGTTTAACTTTAAGAAGGAGATATACATATGAAATC TTCTGGTTCTTCTATG CAGATTACCAGTTTACCTTCTTCGATTTTCATAGAAGAACCAGAAG ATTTC
*Lbu*Cas13a TEV deletion	AGATTATGTTTGAATACAAAATGGAAGAGAAAAAGTCTGAAAACTA ATAACATTGG ATCGCGGATCCGTTATCCACTTCCAATGTTATTAGTTTTCAGAC
*Lbu*Cas13a TEV insertion	GGAAGAGAAAAAGTCTGAAAACCTGTACTTCCAATCCCATCACCAT CACCATCACTAATAACATT GGATTGGAAGTACAGGTTTTCAGACTTTTTCTCTTCCATTTTGTATT CAAACATAATCTTCACCAATTTGC
*Lbu*Cas13a His-ins-C-terminus	GGAAGAGAAAAAGTCTGAAAACCATCACCATCACCATCACTAATAA CATTGGAAGTGGATAACGGATC GTGATGGTGATGGTGATGGTTTTCAGACTTTTTCTCTTCCATTTTG TATTCAAACATAATCTTCACCAATTTGC
*Lbu*Cas13a C348Y	CTGTTAAACAAACTTGACACGTACGTCCGTAATTATGGAAAGTATA ATT GCAATTTCGCCGTCTTGCAAATAATAATTATACTTTCCATAATTACG G
*Lbu*Cas13a C348V	TGTTAAACAAACTTGACACGTACGTCCGTAATGTTGGAAAGTATAA TT GCAATTTCGCCGTCTTGCAAATAATAATTATACTTTCCAACATTACG
*Lbu*Cas13a C348L	TGTTAAACAAACTTGACACGTACGTCCGTAATCTGGGAAAGTATAA TT GCAATTTCGCCGTCTTGCAAATAATAATTATACTTTCCCAGATTACG
pre-crRNA amplification	CCGCAAGCGAAATTAATACGACTCACTATAGG AGGGTCATCATCTCTGCCCCTGTTTTAG
*Lbu*Cas13a H1053A	CATCGCCGCATTCAATTATATTCCTCACGCCGAGATCTCACTGC GAATATAATTGAATGCGGCGATGTAATTTCGAATATAAAGGTCCTT CTTTTCTTGCTTC
*Lbu*Cas13a R1048A	CCTTTATATTGCAAATTACATCGCCCACTTCAATTATATTCCTCACG CGATGTAATTTGCAATATAAAGGTCCTTCTTTTCTTGCTTCAACACC
*Lbu*Cas13a S660C-long	GAAGATTCCAAAGGAATACTTGGCGAATATCCAGTGTCTGTAC GTCCTGATTACCGGCATTAATCATGTACAGACACTGGATATTC
*Lbu*Cas13a A462C	CAATATGGACAACAAGAATGAAATCGAAGATTTCTTCTGTAACATC G GGATGGAAGAAATCGCCTCGTCGATGTTACAGAAGAAATC
*Lbu*Cas13a T756C	GTGAGATCAAACTGGGAAACATCCTGAAGTATTGTGAGCG AAGCTTTAAGATAAGGTAGAACATGTTTAAACGCTCACAATACTTC AG
*Lbu*Cas13a E926C	GTAAAGACGAGAAGTTTACAGATGAGGACTATTGTAGTTACAAG ACTCCTCAATATTCTCAATAGCTTGCTTGTAACTACAATAGTCC
*Lbu*Cas13a R472A	CTTCCATCGCACACGGTATTGTCCACTTCAACTTGGAATTAGAAGG CCGTGTGCGATGGAAGAAATCGCCTCGTCGATGTTGGC
*Lbu*Cas13a H477A	GATTTCTTCCATCCGTCACGGTATTGTCGCATTCAACTT AAAGATATCCTTACCTTCTAATTCCAAGTTGAATGCGACAATAC
*Lbu*Cas13a R1079A	GTCCTACGATGCAAAACTGAAAAATGCCGTAATGAAATCAGTAGTT GATATC

	CAGTTTTGCATCGTAGGACAGCAATTTACGCAAATTTTCAAGGACTTCC
*Lbu*Cas13a S190C	CTTATGTAUGCAATGTGAAAGAAGCCTTTGATAAGCTTTACAAGGAA CTTTCACATTGCATACATAAGCGTCACGTTTCGCTGACTC
*Lbu*Cas13a K222C	GAGAAATACUGCATTCGCGAGTTCTACCACGAAATTATTGGACG CGCGAATGCAGTATTTCTCTAACTTCGTAAGGTTCTCAATTTCAAGAA
*Lbu*Cas13a L138C	ATTAAGAAGAAAUGCAACAAAATCAACAGCCTGAAGTACTCATTTGAAAAGAATAA GATTTTGTTGCATTTCTTCTTAATGTCGTTACGAAAAACTTCCAATTCCTC
*Lbu*Cas13a K63C	AACGCATTGGGTGTTTAAAGAAATTCTTCTCAAACAAAATGGTCTATCTTAAAGACAATA TTTCTTTAAACACCCAATGCGTTTTTGATTTTCCTTGGTTTCCGTGCTGC

6.6 Instruments and Columns

Table 6.6-1: Instruments.

Instrument	Name	Manufacturer
Water purification system	Milli-Q® Direct	Merck Millipore
Cell sonicator	SONOPLUS	Bandelin
Cell disruptor		Constant Systems Limit
Centrifuges	5424R	Eppendorf AG
	5810R	Eppendorf AG
	Avanti J-26 XP	Beckmann Coulter
Chromatography systems	Äkta avant	GE Healthcare
	Äkta start	GE Healthcare
Heat block	Thermomixer comfort	Eppendorf AG
Thermocycler	Mastercycler® nexus	Eppendorf AG
Spectrophotometer	Nanodrop 2000	Thermo Fisher Scientific
Water Bath	Wise Bath®	Wise Laboratory Instruments
Incubators	Ecotron	Infors HT
	I26	New Brunswick Scientific
Gel imager		Peqlab
pH meter	pH1000 L	VWR International
Spectrometer	EMX nano	Bruker BioSpin
	ELEXSYS 580	Bruker BioSpin
Resonator	ER 5106QT-II	Bruker BioSpin
Temperature regulator	ITC502	Oxford Instruments
travelling wave tube amplifier	150 W, 187Ka	Applied Systems Engineering

Helium gas-flow cryostat	CF935	Oxford Instruments
AFM	NanoWizard 3 Ultra	JPK

Table 6.6-2: Columns for protein purification, labelling, and buffer exchange.

Column	Manufacturer/Supplier
HisTrap HP column, 5 mL	Cytiva
HiTrap Heparin HP, 5 mL	Cytiva
HiLoad 16/600 Superdex 200pg	GE Healthcare
Superdex 200pg 10/300	GE Healthcare
HiPrep 26/10 Desalting	GE Healthcare
PD-10 column	Cytiva

6.7 Software

Table 6.7-1: Software used for project design, data acquisition, analysis, processing, illustration, and for validation.

Software	Developer/Company
MATLAB R2021b	MathWorks
DeerAnalysis 2022	Jeschke et al., 2006[276]
Xenon nano 1.2a.2	Bruker
Xepr 2.6b.151	Bruker
Geneious 6.1.8	Geneious
OriginPro 8G	OriginLab
PyMOL Version 2.0.7	Schrödinger LLC
Dezyme	Dehouck et al., 2011[270]
ProtParam	Expasy, Swiss Institute of Bioinformatics
mtsslWizard/mtsslsuite	Hagelueken et al., 2015[273]

Methods

7.1 Genetics, cloning, and RNA transcription

7.1.1 Quick change polymerase chain reaction (QC-PCR)

Several protein constructs with point mutations were generated for this thesis. The protocol for the mutation on the plasmid level used here, is a variation of the one described by Liu and Naismith[289]. In this procedure amino acid codes are changed on the DNA level, by incorporating them into the overlapping region of DNA primers. The pipetting scheme and the thermocycler settings are described below.

Table 7.1-1: QC-PCR pipetting scheme.

Component	Stock	Volume/Amount
Reaction buffer	10x	5 µL
dNTP-Mix	10 mM	2 µL
Pfu Polymerase	5 Units/µL	0.5 µL
Template		50- 100 ng
Forward Primer	100 µM	0.5 µL
Reverse Primer	100 µM	0.5 µL
Total Volume		50 µL

Table 7.1-2: QC-PCR Thermocycler settings.

Step	Thermocycler Setting [°C]	Holding Time [s]	Number of Cycles
Initial Denaturation	95	300	1
Denaturation	95	60	
Annealing	$T_{m,overlap}$-5°C	60	3
Elongation	72	900	
Denaturation	95	60	
Annealing	$T_{m,non\ overlap}$-5°C	60	15
Elongation	72	900	
Denaturation	95	60	
Annealing	43	60	2
Elongation	72	900	

7.1.2 Agarose gel-electrophoresis

Agarose gel electrophoresis was performed to image product formation from QC-PCR. For this, 1% agarose gel solved in TBE, was heated in the microwave until transparent and fluid. The gel was supplemented with 1:1000 ethidium bromide, was poured into the gel chamber, and the comb was inserted. After the gel has cooled to RT, the comb was removed, and the gel chamber inserted into the buffer reservoir. After addition of 1x TAE buffer until the gel is completely covered, the gel was loaded with all samples. For the samples, 5 µL of the QC-PCR reaction was mixed with 1 µL 6x Agarose Gel Sample Buffer. The agarose gel run for 30 min at 100 V and 300 mA.

7.1.3 Transformation of DNA plasmids

The transformation of DNA plasmids was done by using a standard lab protocol and was used for plasmid selection via antibiotics, after QC-PCR, or for expression of recombinant protein in *E. coli*. Self-made chemically competent cells of choice (*E. coli*, mostly DH5 alpha or BL21 AI) were thawed on ice for 10 min. About 100 ng of plasmid was added to the competent cells and incubated on ice for 10 min. After incubation, a heat shock is performed by incubation in a pre-heated water bath at 42°C for 45 s. The cells were incubated on ice for 2 minutes and 1 mL LB medium was added. Then these cells are incubated for 1 h, 180 pm, and at 37°C for cell growth and centrifuged at 3800 rpm for 3 min. Almost all growth medium was discarded, and the cells resuspended in the remaining medium. For antibiotic selection, the cells were streaked on an agar plate, which was previously supplemented with the respective antibiotics. This plate was then incubated over night at 37°C.

7.1.4 Isolation of plasmid DNA

Usually, the isolation of plasmid DNA is done after QC-PCR and transformation. The goal is to verify the mutated sequence through Sanger sequencing. For this, one colony is picked with a sterile pipette tip from an agar plate, supplemented with corresponding antibiotics. This tip is put into a falcon with 10 mL LB medium and the respective antibiotic, and shacked at 180 rpm and 37°C, overnight. On the next day, plasmid was isolated with the GeneJet MiniPrep kit, using the manufacturer's protocol. The amount of isolated DNA was determined by measuring the absorption at 260 and 280 nm with the NanoDrop Spectrophotometer and the isolated DNA plasmid was sent to Microsynth AG for sequencing.

7.1.5 pre-crRNA transcription

The pre-crRNA was transcribed to use it for cleavage assays and PELDOR sample preparations. The pre-crDNA sequence was bought in a pEX-A128 vector (Eurofins) with a T7 RNA polymerase promoter at the 5' end, and an XbaI restriction site at the 3' end. Before transcription, the DNA was amplified by standard PCR with the following pipetting scheme and thermocycler settings.

Table 7.1-3: Standard PCR pipetting scheme.

Component	Stock	Volume
Rxn Buffer	10x	5 µL
dNTPs	10 mM each	1 µL
Template		~ 100 ng/µL
Forward primer	100 µM	0.5 µL
Reverse primer	100 µM	0.5 µL
Pfu Polymerase	2.5 U/µL	0.5 µL
Total Volume		50 µL

Table 7.1-4: Standard PCR thermocycler settings.

Step	Thermocycler Setting [°C]	Holding Time [s]	Number of Cycles
Initial Denaturation	95	60	1
Denaturation	95	30	30
Annealing	T_m-5°C	30	
Elongation	72	60	
Elongation	72	300	1

After DNA amplification, the restriction enzyme digestion was done, by mixing 75 µL PCR product, 3.3 µL 10x FD buffer, and 1 µL XbaI and incubating this mixture for 1 h at 37°C. After digestion the DNA template is precipitated by 3 volumes of cold 100% ethanol and 1/10 volume of a 3 M NaOAc solution. After 20 min incubation at -80°C, the DNA is pelleted at 4°C and 20817 g. The pellet was washed with 100 µL 70% ethanol and centrifuged for 15 min. The supernatant was removed with a pipette and the pellet was dried in the thermomixer at 65°C and resuspended in 50 µL water. Then, the in vitro transcription reaction was set up.

Table 7.1-5: Pipetting scheme of in vitro transcription.

Component	Stock	Volume
T7-Rxn Buffer	5x	20 µL
NTPs	25 mM	10 µL
Template		50 µL
DTT	100 mM	5 µL
Ribolock	40 U/µL	1.56 µL
T7 RNA polymerase	20 U/µL	5 µL
Total Volume		100 µL

The transcription reaction was incubated over night at 37°C. The RNA and DNA was precipitated through ethanol with the same protocol as the precipitation after the restriction reaction. The pellet was resuspended in 70 µL water and 8 µL 10x DNAse buffer and 2 µL DNAseI (10 U/µL) were added and incubated for 20 min at 37°C. Then 27 µL of PAA loading dye was added and incubated for 10 min at 65°C. The sample was loaded onto a 10% urea-PAA gel, which ran for 1.5 h at 375 V and 500 mA. The RNA was visualized under an UV-lamp and PAA gel bands were cut out with a sharp scalpel. These gel slices were crushed in a 1.5 mL reaction tube with a pipette tip and 1 mL 0.3 M NaOAc, pH5.4 was added and incubated at 65°C and 1000 rpm. This step was repeated, and the suspensions were filtered through a syringe, which was previously filled with glass wool. The syringe was rinsed with 500 µL 0.3 M NaOAc and the RNA was precipitated for a third time, as described before. The pellet was resuspended in 30 µL deionized water and the yield was determined with the NanoDrop. The RNA was stored in water at -20°C.

7.2 Protein biochemistry

7.2.1 SDS-PAGE

SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis) is a standard procedure in protein biochemistry that is used to separate proteins by their size and to visualize the protein components in a sample. In this thesis the discontinuous electrophoresis is used to monitor the protein composition in each protein purification step and to monitor the purity of the purified protein. The respective compositions of the stacking gel and the separating gel are described in Table 6.2-2. The SDS-PAGES

usually run for 50 min, at 300 mA and 175V. The gels are then stained with the Coomassie solution (Table 6.2-1) for 10 min and destained by placing them into hot water for 10 minutes.

7.2.2 Western blot

The western blot is used to identify *Lbu*Cas13a in a mixture of proteins by electrophoresis and by its tag-antibody binding. For western blotting, two identical SDS-PAGEs are performed, containing the samples that need to be investigated. This protocol is carried out at room temperature. One gel is used as an SDS-PAGE reference and is treated as described above (Chapter 7.2.1). The second gel is equilibrated in the transfer buffer for 15 minutes. Four blotting papers are also equilibrated with the transfer buffer. The polyvinylidene difluoride (PVDF) membrane is washed with methanol, then with deionized water, and finally it is equilibrated in transfer buffer for 5 min. After soaking, everything is transferred into the blotting chamber, by stacking from bottom (+ pole) to top (-pole), 2 blotting papers, the PVDF membrane, the SDS PAGE gel, and the other 2 blotting papers. Care should be taken to remove the air bubbles in the stack, which is done by rolling a falcon from the middle of the stack to the outer edges. After assembling the western blot chamber, the transfer is started by applying 300 mA and 25 V for 90 min. After electrophoresis, the membrane is placed in the blocking solution and incubated on a shaker for 1 h. The membrane is rinsed three times by placing it in TBST buffer for 5 minutes (shaking). The first antibody, the mouse-anti-6xhis-tag, was diluted in 5 mL TBST buffer and the membrane incubated on a shaker in this solution for 90 min. The membrane is washed again three times in TBST buffer, as described before. The second antibody, goat anti mouse was diluted in 10 mL TBST buffer, and the membrane was incubated in this solution for 60 min. The membrane was washed for the third time three times in 10 mL TBST buffer for 5 min each. Lastly, the membrane was incubated with 6 mL of TMB solution for 5-10 min, rinsed with water and imaged.

7.2.3 Test expressions of *Lbu*Cas13a

Test expressions were used to test under which conditions *Lbu*Cas13a is expressed the most. The general workflow corresponds to the workflow for 1 L cultures. A preculture was set up in the afternoon, supplemented with the corresponding antibiotic and by inoculation with the desired cells. The preculture was incubated over night at 37°C and

180 rpm. The next day, the main cultures were set up. These smaller main cultures contained 20-50 mL LB, supplemented with the corresponding antibiotic and 0.2-0.5 mL preculture. Main cultures were grown at 37°C and 180 rpm until an OD600 of around 0.8 was reached. The temperature was reduced to 16°C and cells were induced with 0.5 mM IPTG and 0.05% L-arabinose and incubated for 16 h at 16°C and 130 rpm. Parameters that have been tested and were varied are for example temperature, speed of rotation, concentration of inducing agent, and *E. coli* cell types. SDS-PAGE samples were collected before and after induction, by collecting 1 mL of cell culture, pelleting the cells, discarding the supernatant, adding 100 µL 1x SDS loading dye to each sample and heating the sample up to 95°C for 5 min.

7.2.4 *Lbu*Cas13a purification tests with benchtop columns

The cells were harvested at 4,000 rcf and the pellet was stored at -80°C or resuspended in binding buffer. The suspension was lysed via sonication thrice at 70% amplitude, for 3 min and in intervals of 1 s. The cell debris were removed through centrifugation at 48,500 rcf for 20 min at 10°C. For the purification tests with a benchtop Ni-AC column, the HisPur Ni-NTA resin (Pierce) was equilibrated with two bead volumes of the binding buffer. Then the soluble fraction, the cell lysate was added to the Ni-NTA beads and incubated at room temperature for 1 h by continuous shaking. The resin was transferred to the benchtop column and the flowthrough is collected. The beads were washed once with 50 mL binding buffer and the protein was consecutively eluted with 60 mL of elution buffer containing 10 mM imidazole, 20 mL containing 100 mM imidazole, 10 mL containing 150 mM imidazole, 10 mL containing 250 mM imidazole, and 10 mL of elution buffer containing 1 M imidazole. SDS-PAGE samples were collected from each purification step.

For the purification via a benchtop column containing amylose beads, the cells were harvested and lysed as described for the Ni-AC purified protein. Then the cell lysate was added to the beads and incubated for one hour at room temperature by continuous shaking. The flowthrough was collected, and the beads were washed with 100 mL of binding buffer. The TEV protease (4 mL) was diluted with 16 mL binding buffer and added directly to the beads. The protein was incubated for 3 h at room temperature by continuous shaking and the flowthrough, containing the protein was collected. SDS-PAGE samples were collected from each purification step.

7.2.5 *Lbu*Cas13a Expression

The expression of all *Lbu*Cas13a constructs is done in the same way, unless otherwise noted. Since this protocol was developed during my thesis, only the final optimized protocol is described here. The p2CT-His-MBP vector containing the codon optimized gene of LbuCas13a was transformed into BL21AI cells (Invitrogen) and expressed in LB medium. A preculture was prepared with a single colony, picked from an agar plate or from a glycerol stock, stored at -80°C. Around 75 mL LB medium was supplemented with 0.3 mM ampicillin and incubated over night at 37°C and 180 rpm. The main culture was prepared with 0.3 mM ampicillin and 10 mL preculture per liter of LB medium. Usually, 6 x 1 L cultures were prepared. Main cultures were grown at 37°C and 180 rpm until an OD600 of around 0.8 was reached. The temperature was reduced to 16°C and cells were induced with 0.5 mM IPTG and 0.05% L-arabinose and incubated for 16 h at 16°C and 130 rpm. Cells were harvested by centrifugation at 4000 rcf for 20 min and stored at -80°C or directly resuspended in binding buffer, supplemented with 1 protease inhibitor tablet per 50 mL suspension and 100 µM 250 mM PMSF, for protein purification.

7.2.6 LbuCas13a Purification

The cells were harvested at 4,000 rcf and the pellet was stored at -80°C or resuspended in binding buffer with the addition of 100 µM 250 mM PMSF and 1 protease inhibitor tablet per 50 mL suspension. The suspension was lysed via sonication thrice at 70% amplitude, for 3 min and in intervals of 1 s. The cell debris were removed through centrifugation at 48,500 rcf for 20 min at 10°C. For protein purification a Ni^{2+}-affinity chromatography was performed using the binding and the elution buffer. The affinity chromatography was followed by the concentration of protein containing fractions in a VivaSpin 20 mL 100,000 MWCO to 5 mL. The solution was diluted with low salt buffer to 20 mL and TEV-protease (self-made) cleavage was performed for 1.5-2 h at room temperature. The solution was loaded onto a 5 mL HiTrap Heparin HP column with a linear gradient of high salt buffer and the protein containing fractions were pooled and concentrated in a VivaSpin 20 mL 100,000 MWCO. The protein was loaded onto a HiLoad 200 16/600 SEC column previously equilibrated with the GF buffer and the purified protein was concentrated again, in a VivaSpin 6 mL 100,000 MWCO. The protein was flash-frozen in liquid nitrogen and stored at -80°C.

7.2.7 Molecular cloning for *Lbu*Cas13a with an UAA

The method to express a protein with an unnatural amino acid used here is by amber codon suppression. The amber codon is a stop codon (UAG), which is almost absent in bacterial cells[271]. It is recognized by an aminoacyl-tRNA synthetase (aaRS) with the corresponding tRNA and incorporated into the amino acid chain as an additional amino acid. So, instead of exchanging an amino acid to a cysteine, an amber codon is inserted at the position that needs to be spin labelled. This is done by QC-PCR (see Chapter 7.1.1). The p2CT-His-MBP plasmid, containing the *Lbu*Cas13a gene has an N-terminal His-MBP-TEV end. Several cloning steps were performed, to facilitate the isolation of the full-length *Lbu*Cas13a from the truncated product, which origins from reading the amber codon as a stop codon. The plasmid was modified to have the following scheme: N-terminus-MBP-TEV cleavage site-Cas13a-TEV cleavage site-His. That should ideally lead to the remotion of truncated protein construct already after Nickel affinity chromatography.

7.2.8 *Lbu*Cas13a expression and purification with UAA

To express *Lbu*Cas13a with an UAA, the p2CT-His-TEV cleavage site-Cas13a-TEV cleavage site-MBP plasmid containing the *Lbu*Cas13a gene with an amber codon modification, as well as a second plasmid (pEvolAcF/pAcF), containing the aaRS gene, are co-transformed in *E. coli* BL21 AI cells (see Chapter 7.1.3), with the corresponding antibiotic resistance. The expression was done once, in the same way as for *Lbu*Cas13a without UAA, with two exceptions. First, the cultures were supplemented with two antibiotics (ampicillin and chloramphenicol) and second, all culture flasks were supplemented with 500 mg L-acetyl phenylalanine before induction. The protein was purified in the same way as *Lbu*Cas13a without UAA, for details see Methods Section 7.2.6.

7.2.9 Nickel affinity chromatography

Nickel affinity chromatography (Ni-AC) is a commonly used purification method for multi-His-tagged recombinant proteins. In this thesis, a Hexa-His-tag was co-expressed with a maltose binding protein (MBP)-tag at the N-terminus. The histidines are highly affine to Ni^{2+} Ions, which are bound to the stationary phase of the column. The AC was performed using one or two prepacked Ni^{2+} columns (HisTrap) on the chromatography system Äkta Avant. The AC run consists of loading the centrifuged cell lysate onto the column with a flow rate of 2 mL/min. After washing the column with 10 CV binding buffer at 3 mL/min, the protein is eluted with a linear gradient of

elution buffer from 4% to a final target percentage of the elution buffer of 55%. The gradient volume was set to 7 CV and the flow rate to 2 mL/min. Lastly, the column was washed with 4 CV of 100% elution buffer with a flow rate of 3 mL/min and subsequently equilibrated by 5 CV binding buffer and 3 mL/min. For long-term storage, HisTrap columns are stored in 20% ethanol.

7.2.10 Heparin affinity chromatography

Heparin mimics the polyanionic structure of DNA and RNA, which is the reason why it is used for the purification of DNA and RNA binding proteins. In the purification procedure of Cas13a prepacked heparin columns were used with the low salt and high salt buffers. The protein containing solution was loaded onto the column with a flow rate of 2 mL/min. After washing the column with 30% high salt buffer for 5 CV with 2 mL/min, a linear gradient of the duration of 12 CV is applied from 35% to 65% high salt buffer, with a flow rate of 2 mL/min. Lastly, the column was washed for 2 CV with 100% high salt buffer and re-equilibrated with low salt buffer. For long-term storage, Heparin columns are stored in 20% ethanol.

7.2.11 Size exclusion chromatography

The size exclusion chromatography (SEC) is often the last chromatographic step in a protein purification, where the molecules are separated by their size. The stationary phase of SEC columns can have different pore sizes and the elution volume of a biomolecule is dependent on its hydrodynamic radius. The smaller the biomolecular size, the larger is the diffusion volume in the column for this biomolecule. Once the protein is loaded onto the column, a constant flow rate of 0.5-1 mL/min is applied for 1.2 CV, depending on the amount of glycerol in the GF buffer. For long-term storage, SEC columns are stored in 20% ethanol.

7.3 *Lbu*Cas13a spin labelling

7.3.1 MTSSL spin labelling

The following spin-labelling protocol was developed to label *Lbu*Cas13a for EPR spectroscopic measurements. Up to 50 nmol *Lbu*Cas13a was diluted to 1 mL with *cw*-buffer. The reducing agent TCEP was added to a final concentration of 750 µM and the protein was then incubated for 45 min on ice. TCEP was removed through a disposable

PD10 column, previously equilibrated with *cw*-buffer. Here the manufacturers' protocol was used. Then, a 20-fold excess of MTSL per cysteine was added immediately to the protein after TCEP removal and the solution was incubated over night at 4°C. The next day the spin label was removed in 2 steps with the PD10 column, and the flowthroughs were pooled and concentrated in a VivaSpin 6 mL 100,000 MWCO. The protein was washed in a VivaSpin 6 mL 100,000 MWCO once with *cw*-buffer.

7.4 Analytical Methods

7.4.1 *Lbu*Cas13a *cw*-EPR sample preparation

To determine the labelling efficiency, samples for *cw*-EPR spectroscopy were prepared with labelled protein constructs in *cw*-buffer. The final protein concentration was 50 µM in a total sample volume of 11 µL. These samples were filled in 10 µL capillaries.

7.4.2 *cw*-EPR measurements

The *cw*-EPR spectra were recorded at X-band frequencies on an EMXnano spectrometer from Bruker BioSpin. The measurements were conducted with a microwave power of 10 mW, a modulation frequency of 100 kHz, a modulation amplitude of 1 G, a microwave frequency of 9.6 GHz, a center field of 3346 G, a sweep width of 150 G, a sweep time of 45.13 s, and a time constant of 20.48 ms.

7.4.3 PELDOR sample preparation

For PELDOR measurements, each mutant needs to be rebuffered in deuterated PELDOR buffer. All PELDOR protein samples were prepared from labelled protein constructs in deuterated buffer and d_8-glycerol in a final amount of 30% (v/v) was used as cryoprotectant. The *apo* protein samples had a final protein concentration of 50 µM in a total sample volume of 60 µL, all *holo* protein-RNA samples had a final protein concentration of 15 µM and a final pre-crRNA, crRNA, and target RNA concentration of 18 µM. For PELDOR samples containing pre-crRNA, the right amount of RNA was dried in a vacuum-concentrator, dissolved in deuterated PELDOR buffer, added to the protein, incubated for 5 min at room temperature and gently mixed with a pipette

about 200 times. Finally, the sample was filled into the PELDOR tubes and flash-frozen in liquid nitrogen. For the PELDOR samples with crRNA, the crRNA was dried in a vacuum-concentrator, dissolved in deuterated PELDOR buffer, and annealed by incubating it at 95°C for 5 min and cooling it down to room temperature for 15 min. The protein was then added to the crRNA and incubated for 10 min at 37°C. After gently mixing the solution for at least 200 times with the pipette, it was filled into PELDOR tubes and flash-frozen in liquid nitrogen. The PELDOR samples with cr- and target RNA were prepared, in the first place, as the samples containing crRNA. Additionally, after incubation with the crRNA, target RNA was added, which has been previously dried in a vacuum-concentrator and dissolved in PELDOR buffer. Then the mixture was incubated for 15 min at 37°C, before flash-freezing. All samples were filled into 3 mm quartz tubes and flash-frozen.

7.4.4 PELDOR measurements

The PELDOR experiments were performed as described in the detailed PELDOR protocol in Methods in Molecular Biology[290]. All PELDOR measurements were conducted at Q-band frequencies (~33.7 GHz), and 50 K on an ELEXSYS E580 EPR spectrometer from Bruker BioSpin, equipped with an ER 5106QT-II resonator, a 150 W TWT-amplifier, and a helium gas-flow cryostat in conjunction with an iTC 502 temperature controller. For a standard 2-pulse Electron spin echo envelope modulation experiment (2pESEEM), a standard two-pulse Hahn echo sequences was applied, in which the pulse lengths were set to 12 ns for the $\pi/2$, and 24 ns for the π pulse. The initial τ was set to 200 ns. A standard four-pulse PELDOR experiment was used, in which the pump pulse frequency was set to the maximum of the nitroxide field swept spectrum and the offset between the pump and the observer frequency was 100 MHz. The 4-pulse PELDOR pulse sequence used is $(\pi/2(\upsilon_A) - \tau_1 - \pi(\upsilon_A) - (\tau_1+t) - \pi(\upsilon_B) - (\tau_2-t) - \pi(\upsilon_A) - \tau_2 - \text{echo})$. The $\pi/2$ and π pump pulse length was set to 12 ns and 24 ns, respectively, with an interpulse delay τ_1 of 200 ns. The pump pulse length was set to the length at which the maximum of spins was flipped, which was determined by a 3-pulse ELDOR nutation experiment. The initial τ was set to the maximum of the 2-pulse ESEEM experiment, which was usually 232 ns. Signal averaging was usually done for 8 to 48 h (depending on the sample) to increase the signal-to-noise ratio.

7.4.5 PELDOR data analysis

To plot, analyze, and convert the raw data the software MATLAB was used. Various programs exist to extract the distance distributions from PELDOR time traces. Here two methods were used, first Tikhonov regularization and second, the ComparativeDeerAnalyzer (CDA). Both are implemented in the MATLAB program DeerAnalysis 2022. During analysis of the time traces acquired for this thesis it was seen that the background from the CDA did not fit well for some of the time traces, which show low or no oscillations. So, for those time traces, where the CDA background does not fit, Tikhonov regularization was done.

7.4.6 pre-crRNA cleavage assays

Generally, the aim of pre-crRNA or target RNA cleavage assays is, to detect RNA cleavage by Cas13a and to differentiate between e.g. pre-crRNA and the cleavage product crRNA. For this, pre-crRNA cleavage assays were performed and loaded onto a 20% polyacrylamide gel in standard 1x TBE buffer. For each cleavage reaction, 20 pmol pre-crRNA was incubated at 37°C for 1 h with 16 pmol of Cas13a and 0.9 µL tRNA in cleavage buffer (previously filtered with a 0.22 µm syringe filter) to yield a total reaction volume of 8 µL. The reaction was quenched with 0.3 U of proteinase K. After incubation for 45 min at 37°C, PAA-loading dye was added 1:1 to each reaction, heated to 95°C for 5 min and loaded onto the PAA-gel. The PAA gel pre-ran for 15 min at 150 V, ran for 3 h at 300 V, and was finally stained with ethidium bromide.

7.4.7 Target RNA cleavage assays

As for the pre-crRNA cleavage assays, also the target RNA cleavage assays were performed on a 20% polyacrylamide gel in standard TBE buffer. Before setting up the reactions, 20 pmol crRNA was annealed in 5 µL cleavage buffer by heating it up at 65°C for 5 min and cooling it down at room temperature for 15 min. For binary complex formation, 40 pmol Cas13a (in 3 µL) was added to the crRNA and incubated for 45 min at 37°C. At this point, 40 pmol target RNA was added and incubated for 1 h at 37°C. Next, 0.3 U proteinase K was added to quench the reactions, after 45 min at 37°C PAA-loading dye was added in a 1:1 ratio. The reactions were heated for 5 min to 95°C and loaded onto the PAA gel. As for the pre-crRNA cleavage assay, the PAA-gel pre-ran for

15 min at 150 V and ran, after sample loading, for 3 h at 300 V, and was stained with ethidium bromide.

7.4.8 AFM measurements

AFM measurements were performed on a mica surface on which different protein constructs were loaded. The protein constructs had initial concentrations of 12-22 mg/mL (86 µM - 158 µM). The wt Cas13a was diluted with AFM buffer to 1:18000. For the crRNA bound complex, the crRNA was annealed by heating it up to 65°C for 5 min and cooling down to RT for 15 min. The RNA was mixed with the protein and incubated for 15 min at 37°C, to obtain a 3 µM solution with a protein : crRNA ratio of 1 : 1.2 in a total volume of 20 µL. The pre-crRNA sample was also prepared such to yield 20 µL of a 3 µM protein solution with a 1 : 1.2 protein to RNA ratio. Here the RNA was not annealed, it was directly incubated with the protein for 15 min at 37°C. The cr- and target RNA bound complex was prepared by forming the crRNA-bound complex, as in the crRNA bound sample. The ternary complex was formed by adding the target RNA and incubating the sample for 15 min at 37°C. The final protein : crRNA : target RNA ratio was 1 : 1.2 : 1.2. All RNA containing samples were diluted with AFM buffer to 1:300 – 1:500 and the AFM measurements themselves were done by Daniel Keppner (AG Famulok).

Appendix

8.1 LC-MS identification of *Lbu*Ca13a

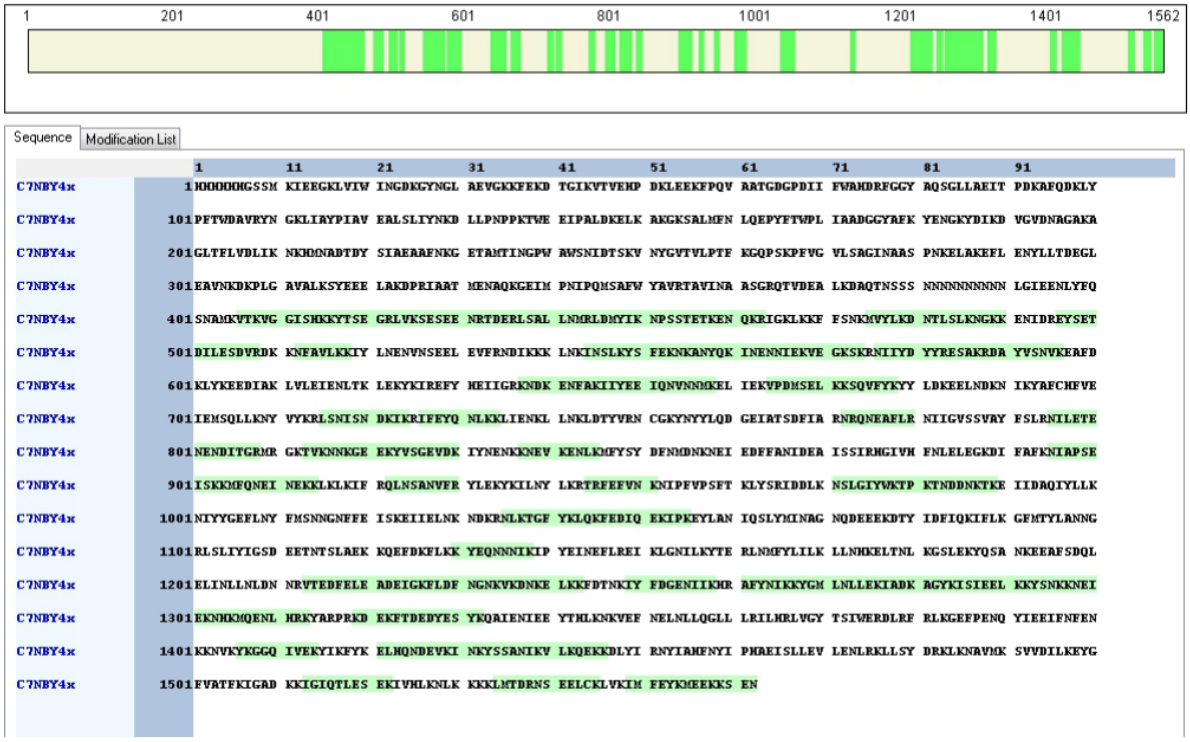


Figure 8.1-1: Result of protein identification through LC-MS. The entire protein sequence with hexa-his tag, MBP tag, TEV cleavage site, and Cas13a is shown. The green regions denote the protein fragments that have been identified. The total coverage is 44%, the last amino acids in the sequence have been identified and the first protein fragment (SNAMK) is too short to be identified with the LC-MS set-up used.

8.2 Biochemical characterization of *Lbu*Cas13a constructs

8.2.1 Construct 138/222

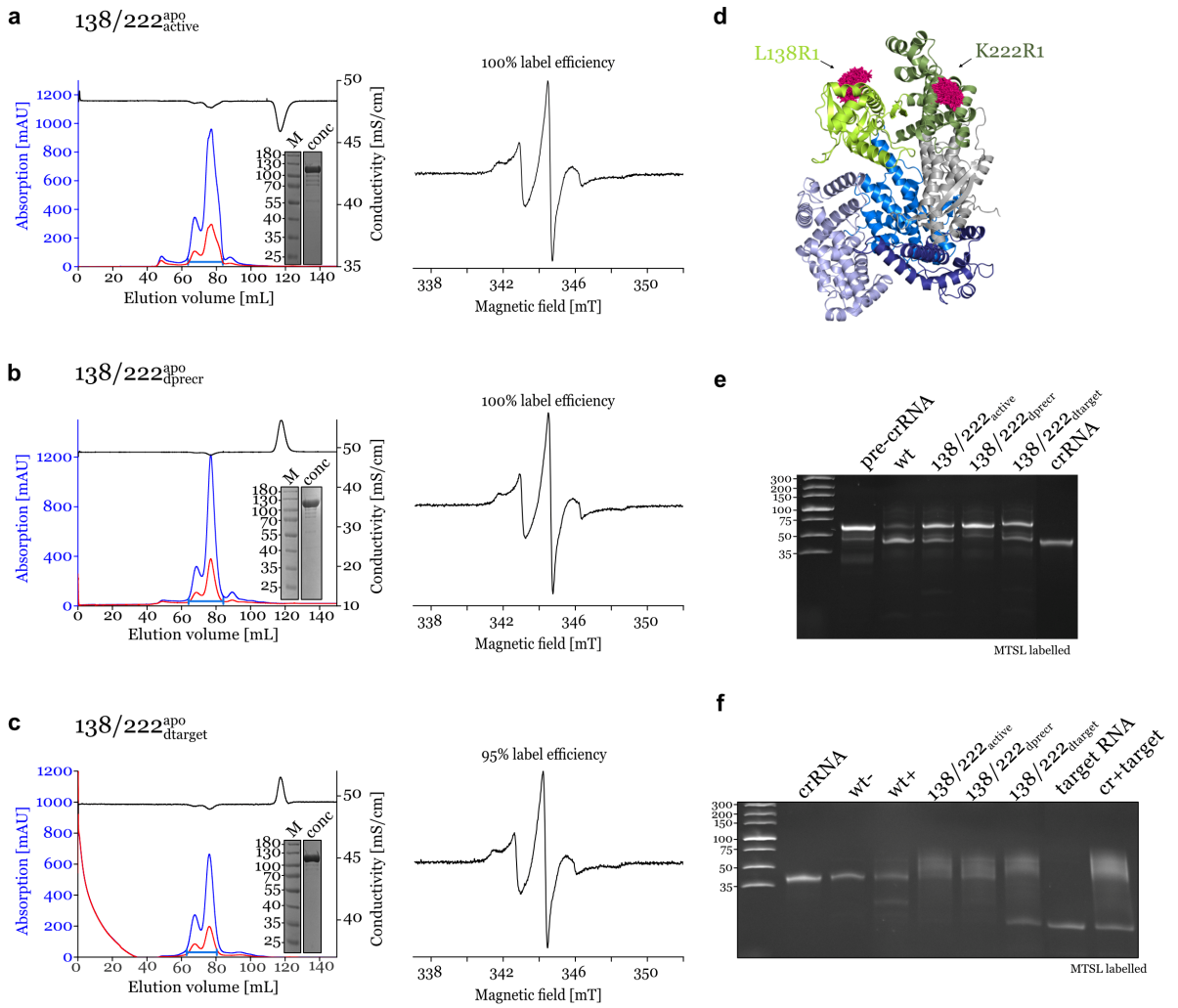


Figure 8.2-1: Biochemical characterization of the different 138C/222C constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the active construct **b,** the pre-crRNA cleavage deficient dprecr construct, and **c,** the target RNA cleavage deficient dtarget construct. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. **d,** AF3 structure prediction of *apo Lbu*Cas13a with MTSL rotamer clouds in magenta, at L138R1 and K222R1. **e,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a-c. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **f,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-c. The crRNA, and the target RNA are loaded as negative controls. Both were mixed and incubated to yield the last lane (cr+target). wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively.

8.2.2 REC lobe constructs for flexibility investigation

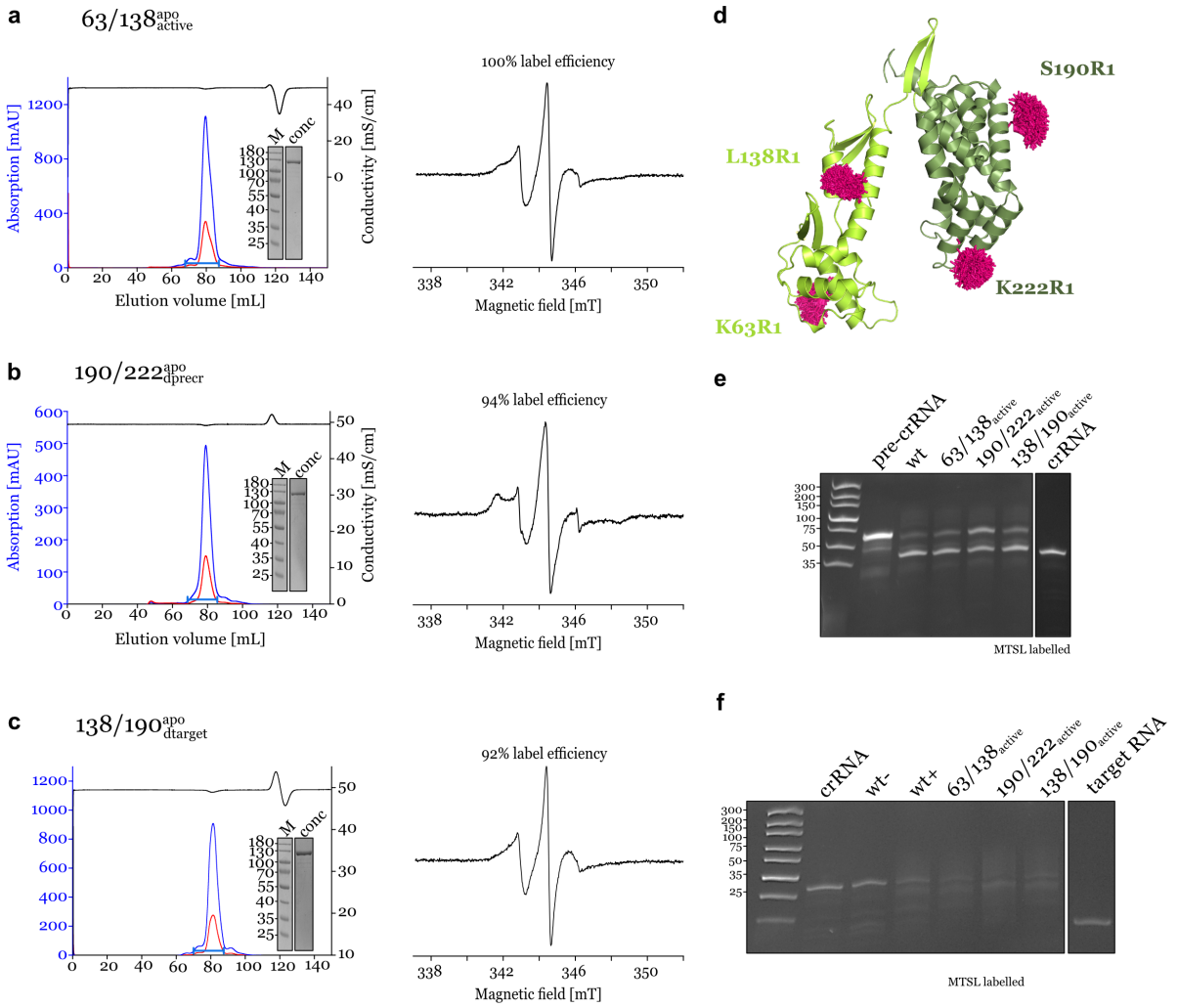


Figure 8.2-2: Biochemical characterization of the different REC-lobe constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the active 63/138 construct **b,** the active 190/222 construct, and **c,** the active 138/190 construct. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. **d,** Structure of the REC lobe predicted by AF3. The MTSL label rotamer clouds are showed in magenta. **e,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a-c. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **f,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-c. The crRNA, and the target RNA are loaded as negative controls. The lanes labelled with wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively.

8.2.3 Construct 756/926

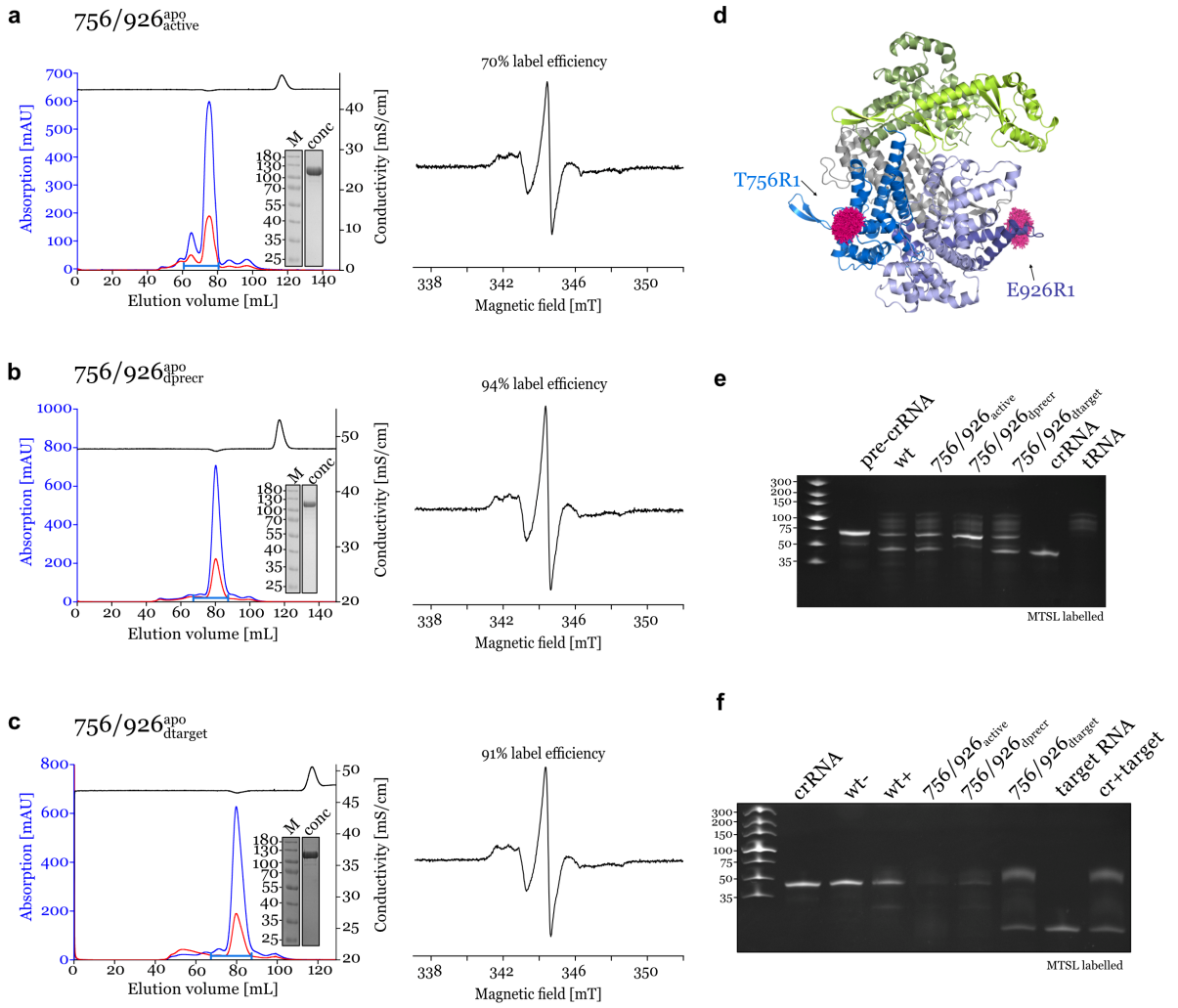


Figure 8.2-3: Biochemical characterization of the different 756C/926C constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the active construct **b,** the pre-crRNA cleavage deficient dprecr construct, and **c,** the target RNA cleavage deficient dtarget construct. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. **d,** AF3 structure prediction of *apo Lbu*Cas13a with MTSL rotamer clouds in magenta, at T756R1 and E926R1. **e,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a-c. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **f,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-c. The crRNA, and the target RNA are loaded as negative controls. Both were mixed and incubated to yield the last lane (cr+target). wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively.

8.2.4 Construct 462/926

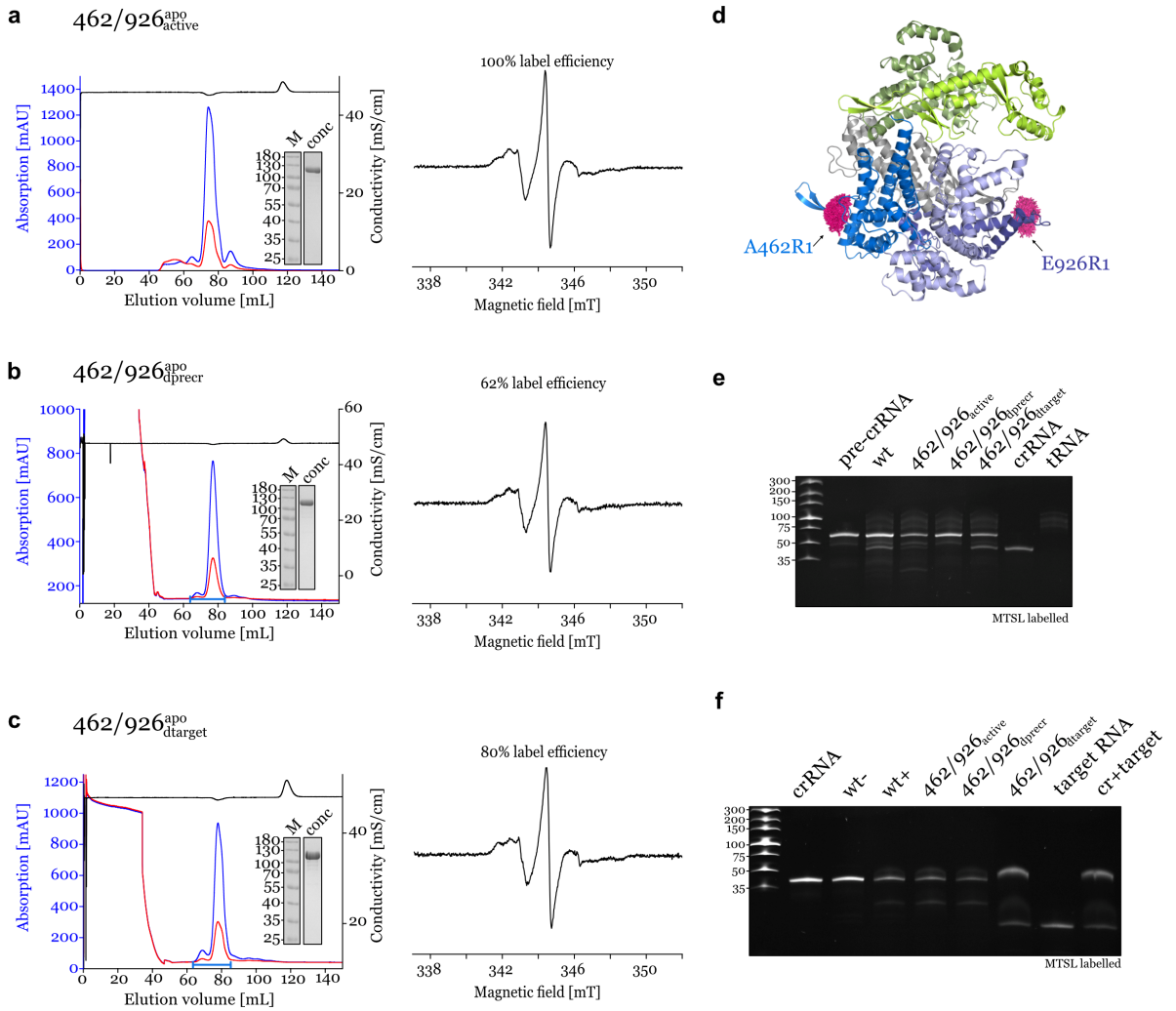


Figure 8.2-4: Biochemical characterization of the different 462C/926C constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the active construct **b,** the pre-crRNA cleavage deficient dprecr construct, and **c,** the target RNA cleavage deficient dtarget construct. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. **d,** AF3 structure prediction of *apo Lbu*Cas13a with MTSL rotamer clouds in magenta, at A462R1 and E926R1. **e,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a-c. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **f,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-c. The crRNA, and the target RNA are loaded as negative controls. Both were mixed and incubated to yield the last lane (cr+target). wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively.

8.2.5 Construct 660/926

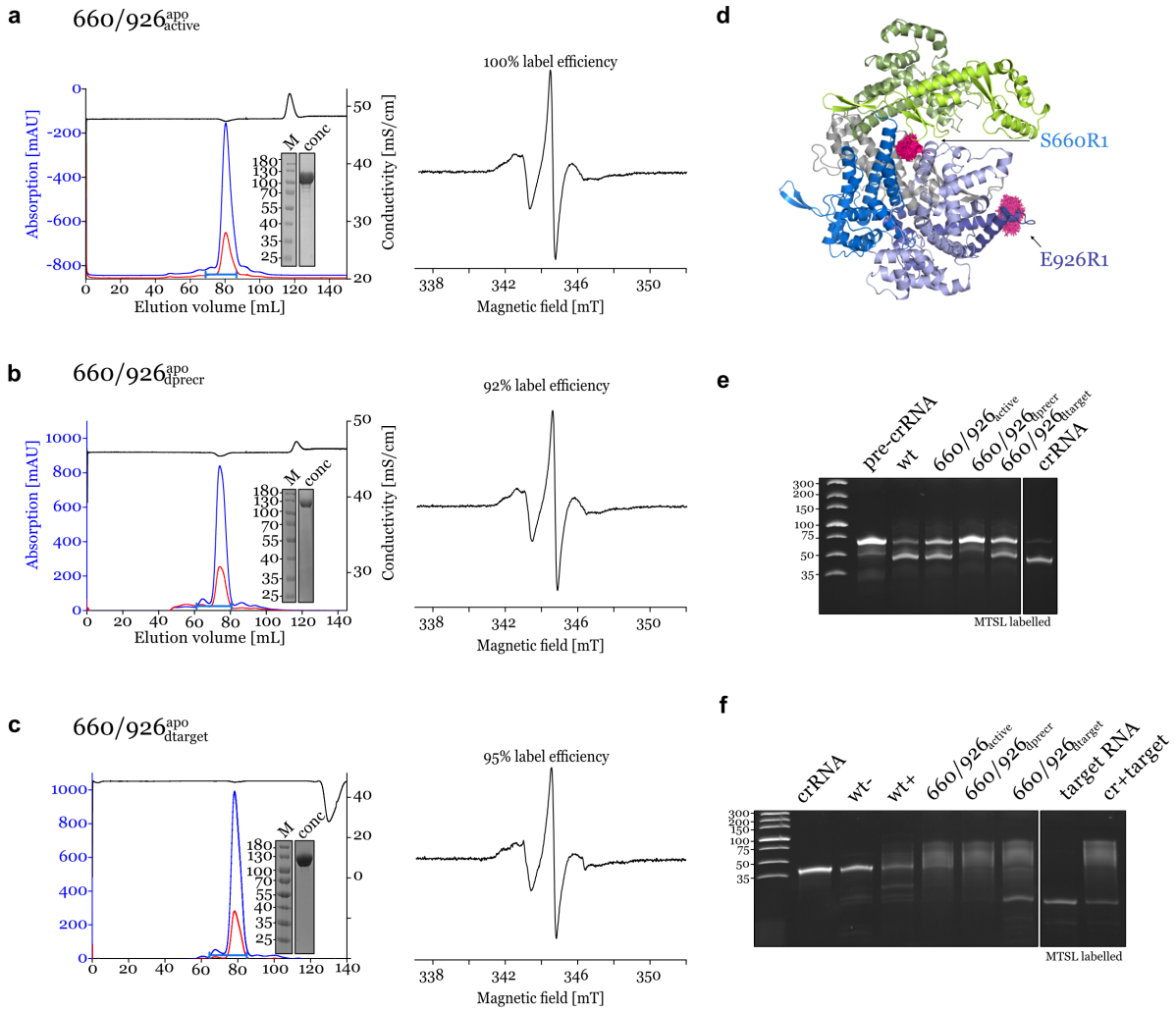


Figure 8.2-5: Biochemical characterization of the different 660C/926C constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the active construct **b,** the pre-crRNA cleavage deficient dprecr construct, and **c,** the target RNA cleavage deficient dtarget construct. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. **d,** AF3 structure prediction of *apo Lbu*Cas13a with MTSL rotamer clouds in magenta, at A660R1 and E926R1. **e,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a-c. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **f,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-c. The crRNA, and the target RNA are loaded as negative controls. Both were mixed and incubated to yield the last lane (cr+target). wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively.

8.2.6 Construct 660/756

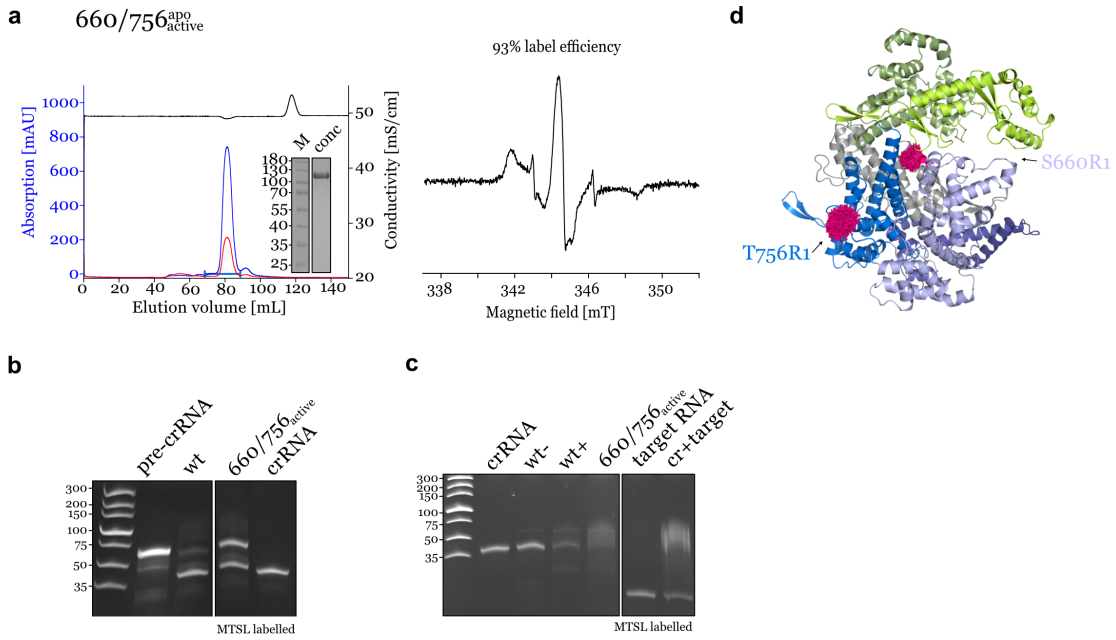


Figure 8.2-6: Biochemical characterization of the different 660C/756C constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the active construct. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. **b,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **c,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-c. The crRNA, and the target RNA are loaded as negative controls. Both were mixed and incubated to yield the last lane (cr+target). wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively. **d,** AF3 structure prediction of *apo Lbu*Cas13a with MTSL rotamer clouds in magenta, at S660R1 and T756R1.

8.2.7 Construct 190/926

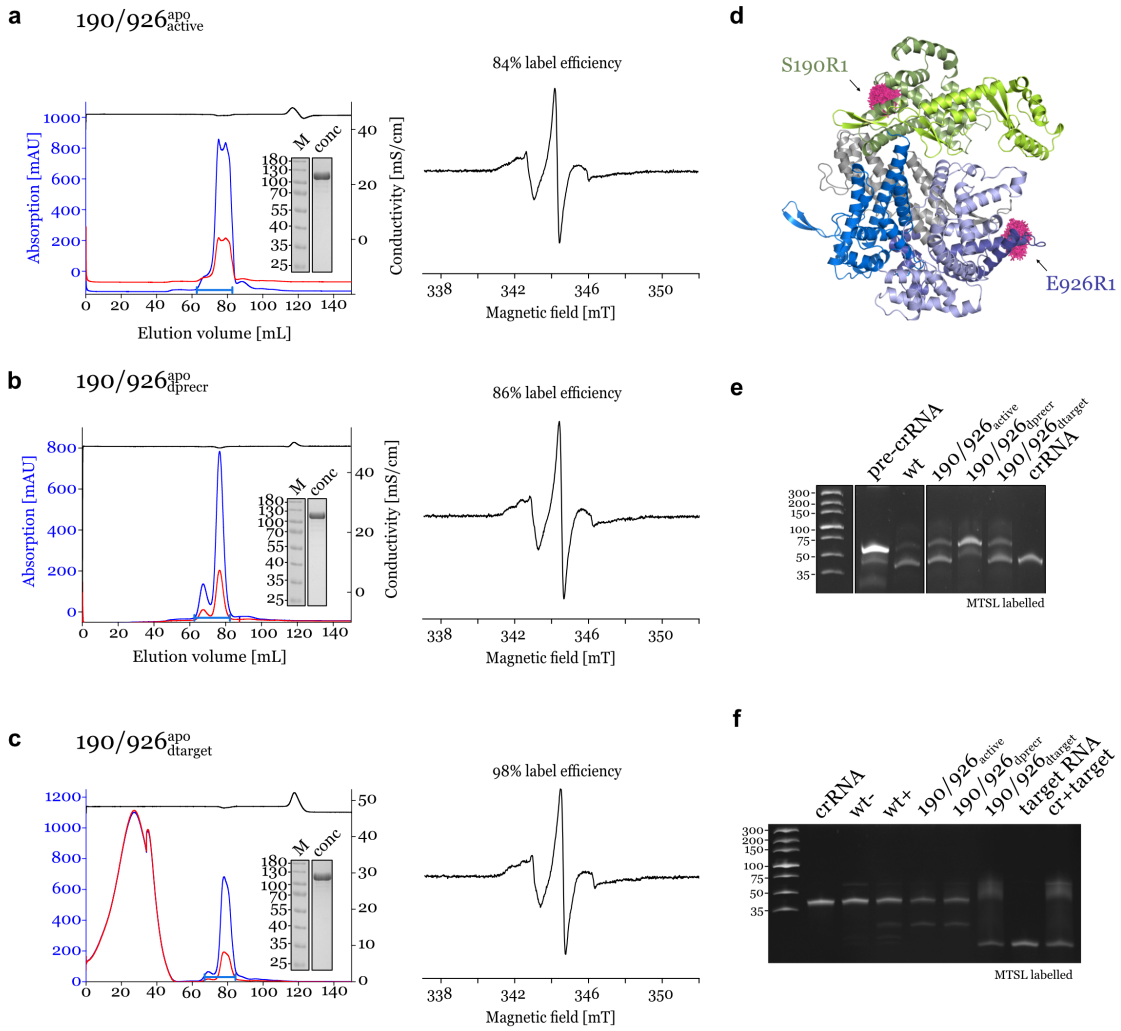


Figure 8.2-7: Biochemical characterization of the different 190C/926C constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the active construct **b,** the pre-crRNA cleavage deficient dprecr construct, and **c,** the target RNA cleavage deficient dtarget construct. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. **d,** AF3 structure prediction of *apo Lbu*Cas13a with MTSL rotamer clouds in magenta, at S190R1 and E926R1. **e,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a-c. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **f,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-c. The crRNA, and the target RNA are loaded as negative controls. Both were mixed and incubated to yield the last lane (cr+target). wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively.

8.2.8 Construct 190/756

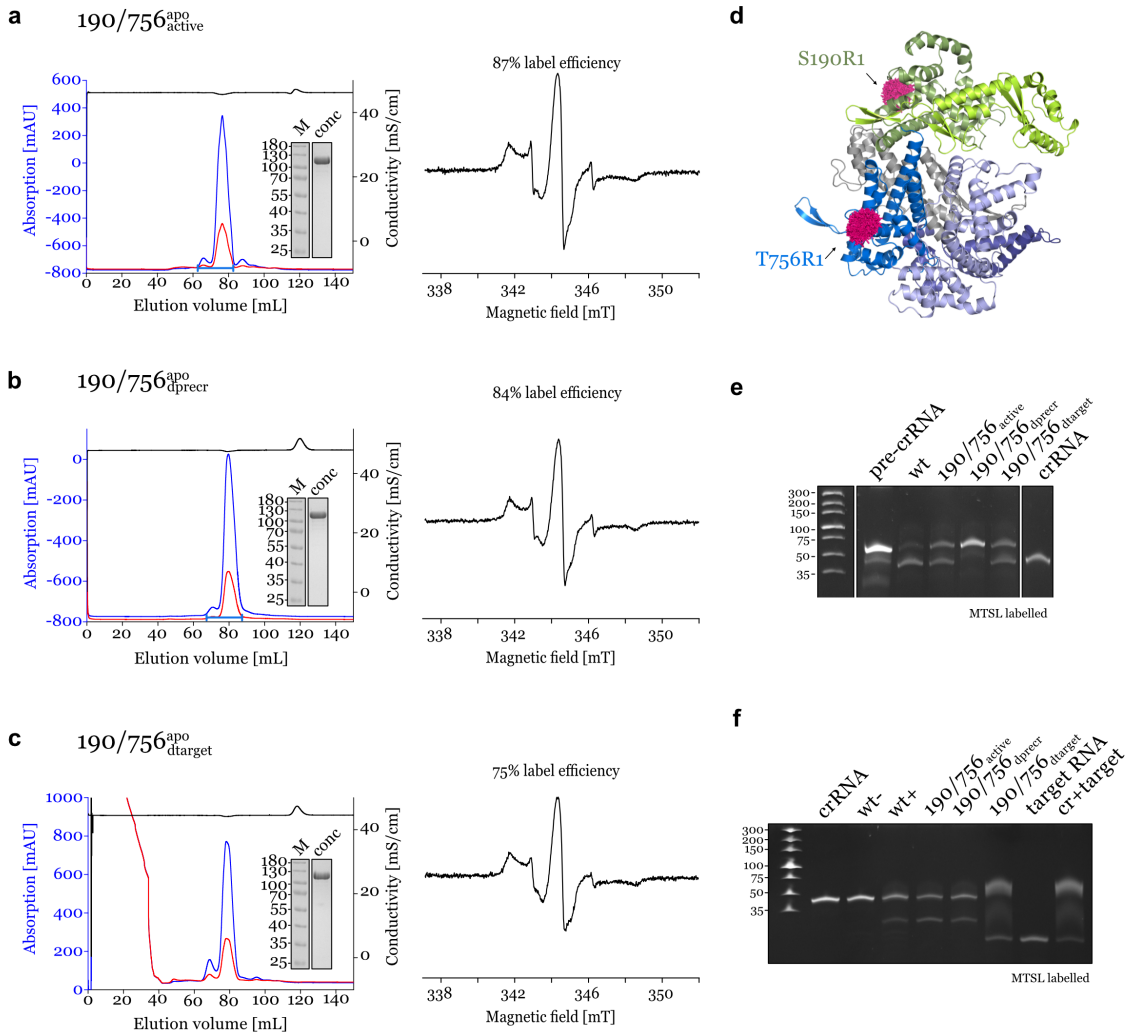


Figure 8.2-8: Biochemical characterization of the different 190C/756C constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the active construct **b,** the pre-crRNA cleavage deficient dprecr construct, and **c,** the target RNA cleavage deficient dtarget construct. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. **d,** AF3 structure prediction of *apo Lbu*Cas13a with MTSL rotamer clouds in magenta, at S190R1 and T756R1. **e,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a-c. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **f,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-c. The crRNA, and the target RNA are loaded as negative controls. Both were mixed and incubated to yield the last lane (cr+target). wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively.

8.2.9 Single labelled constructs 756 and 926

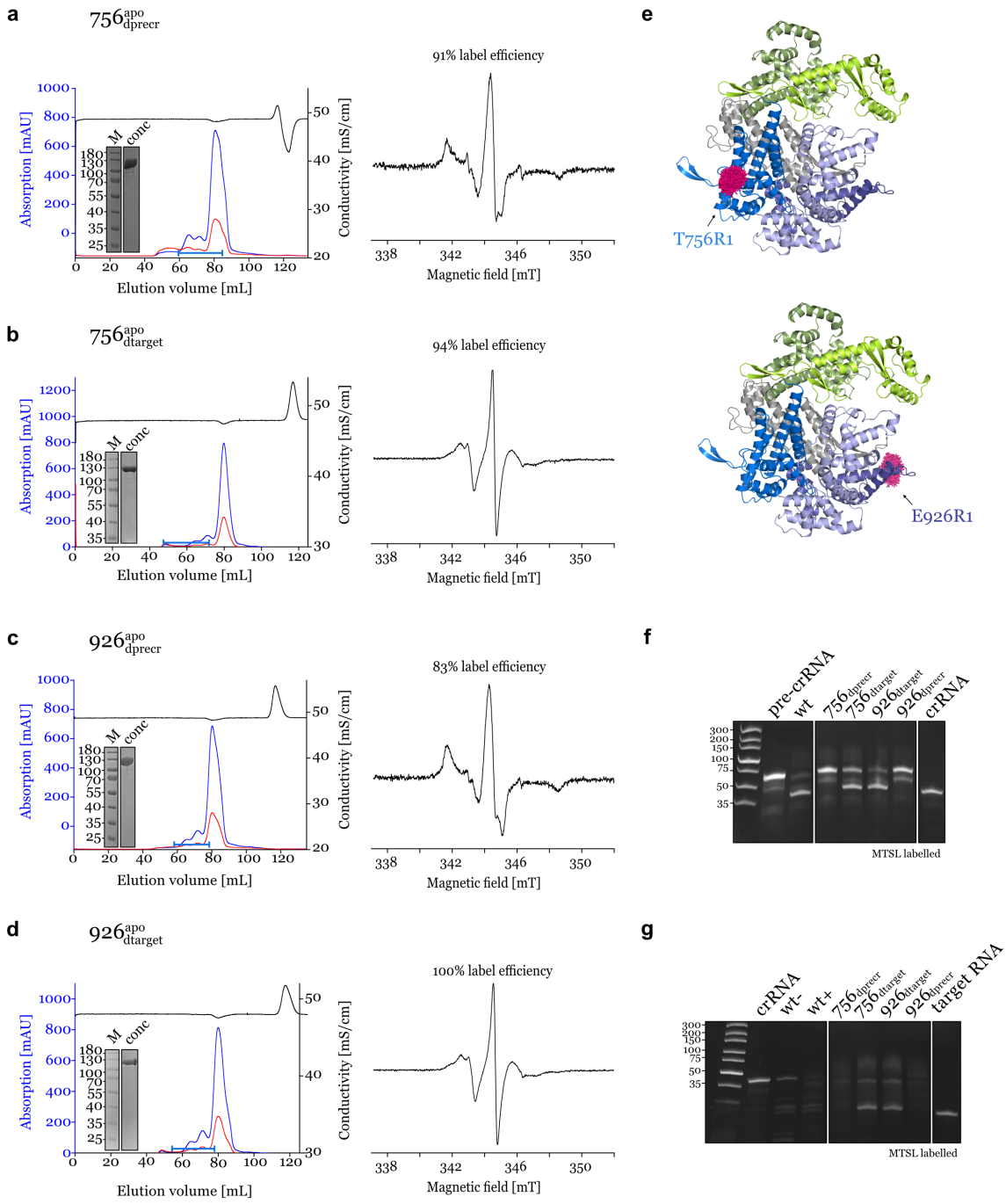


Figure 8.2-9: Biochemical characterization of different single labelled constructs. The size exclusion chromatogram, with an insert of the SDS-PAGE gel of the concentrated protein, and the *cw*-EPR spectrum of the labelled constructs with the corresponding labelling efficiencies are shown from **a,** the 756 dprecr construct **b,** the 756 dtarget construct **c,** the 926 dprecr construct, and **d**, the 926 dtarget construct. The blue curves show the absorption at 280 nm and the red curves the absorption at 254 nm. The blue bar in the chromatogram shows the region, which was concentrated to obtain the final protein. **e,** AF3 structure predictions of *apo Lbu*Cas13a with MTSL rotamer clouds in magenta, at T756R1 and E926R1. **f,** pre-crRNA cleavage assay of the MTSL labelled constructs shown in a-d. The pre-crRNA in the first lane, and the crRNA in the last lane represent the negative and positive control, respectively. **g,** Target RNA cleavage assay performed with the MTSL labelled constructs shown in a-d. The crRNA, and the target RNA are loaded as negative controls. Both were mixed and incubated to yield the last lane (cr+target). wt- and wt+ show reactions in which the wt Cas13a-crRNA complex was incubated without and with target RNA, to get a negative and positive control, respectively.

8.3 Additional Analysis of PELDOR data

8.3.1 Analysis of precipitation during PELDOR sample preparation

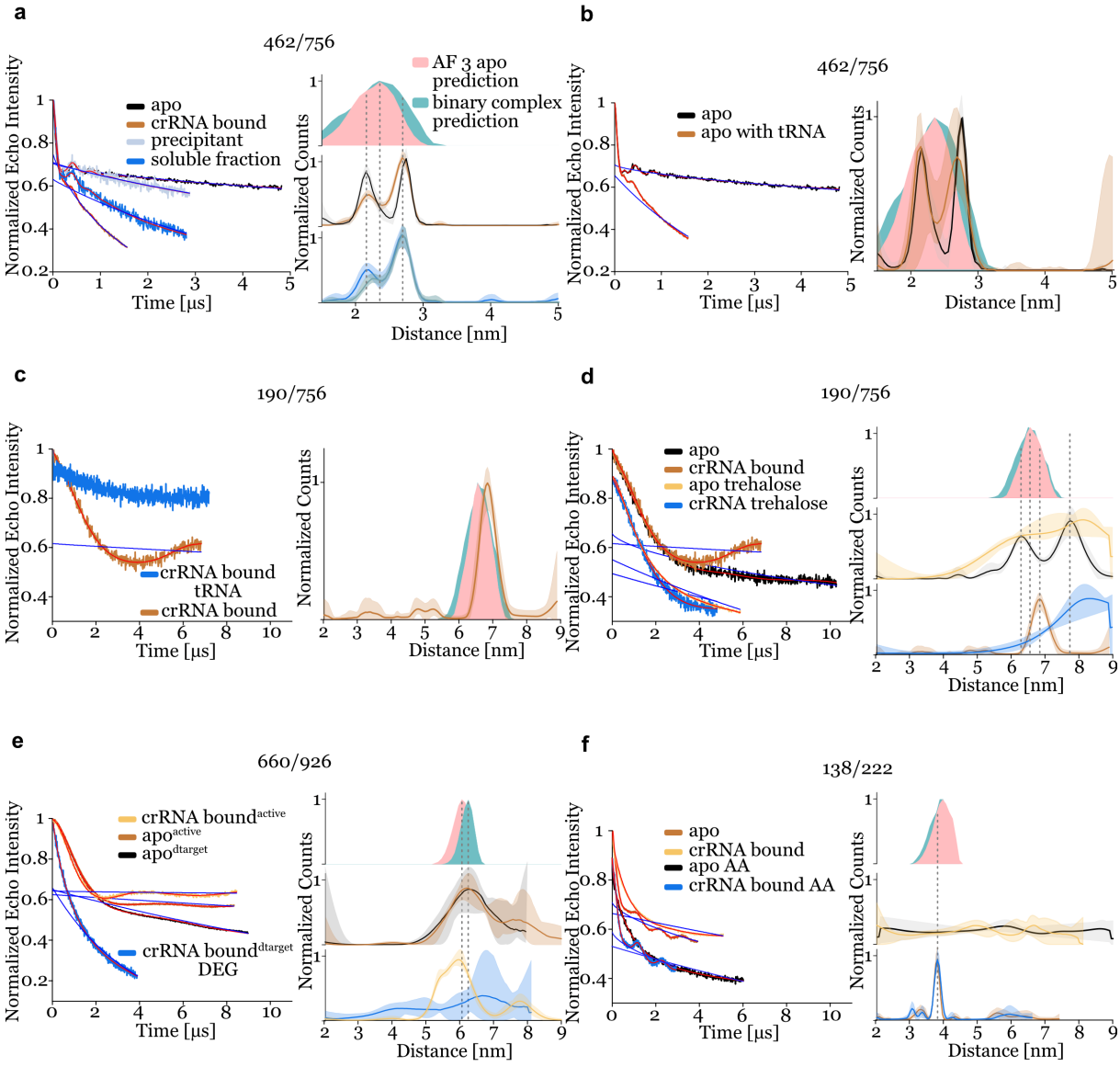


Figure 8.3-1: Full CDA analysis for testing the influence of different additives to reduce sample turbidity. PELDOR time traces and distance distributions of **a,** labelled 462/756 construct. Precipitation occurred during sample preparation, the precipitant and the soluble fraction were measured separately. The predictions of distance distributions on the basis of the AF3 prediction and on the experimental structure of the binary crRNA-bound complex are shown as colored areas. The predictions of the distance distributions are colored in the same way in a-d. The 95% confidence intervals are drawn as shaded areas in the same color of the distance distributions that it belongs to. **b,** *apo* 462/756 with and without tRNA. **c,** Labelled 190/756 construct with crRNA. As in b, the influence of tRNA was tested. The signal to noise ratio of time trace of the sample with tRNA was so low that no distance distribution could be calculated. **d,** Labelled 190/756 construct in its *apo* and crRNA bound states, with and without trehalose. The time traces of the samples with trehalose are shifted on the y-axis by 0.1. **e,** The labelled 660/926 construct in its active and dtarget form, with and without crRNA. The influence of d_4-ethylene glycol (d-EG) is analyzed. **f,** The labelled, cleavage active 138/222 construct with and without the amino acids L-glutamic acid and L-arginine (AA). The time traces of the samples with AA are shifted on the y-axis by 0.1. The time traces include the fits and the background function that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214].

8.3.2 Construct 138/222

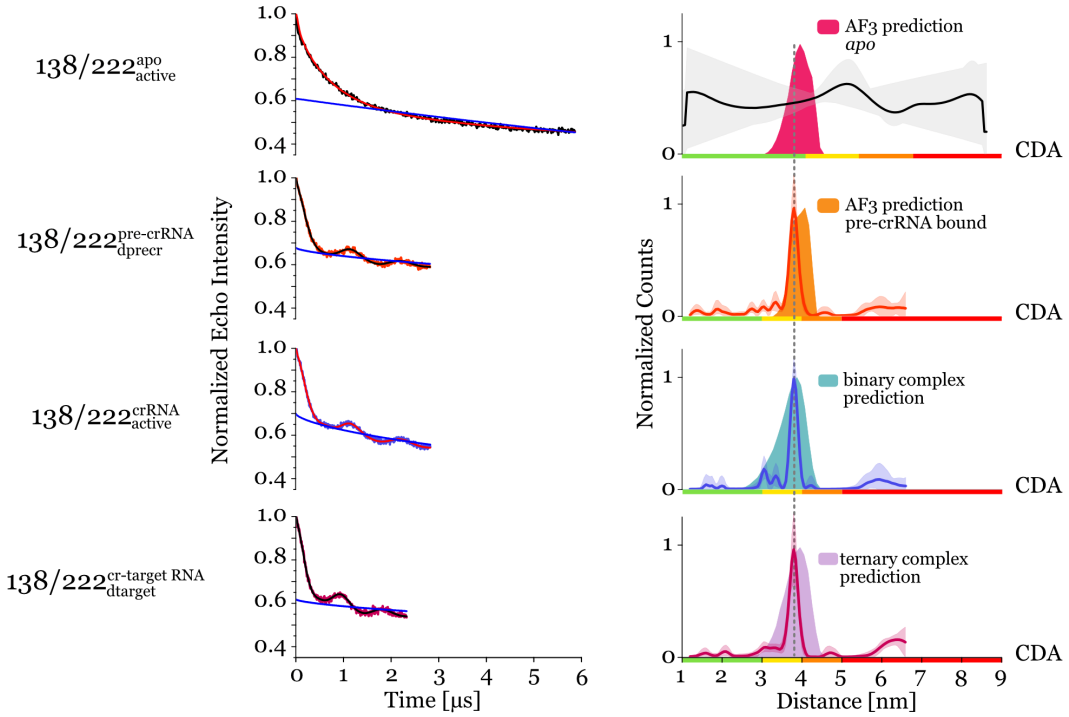


Figure 8.3-2: Full CDA analysis of 138/222. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits in red that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214] and the background function in blue. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models, the binary cryo-EM structure, and the ternary X-ray crystallographic structure are shown as shaded distributions.

8.3.3 Constructs to study REC lobe flexibility

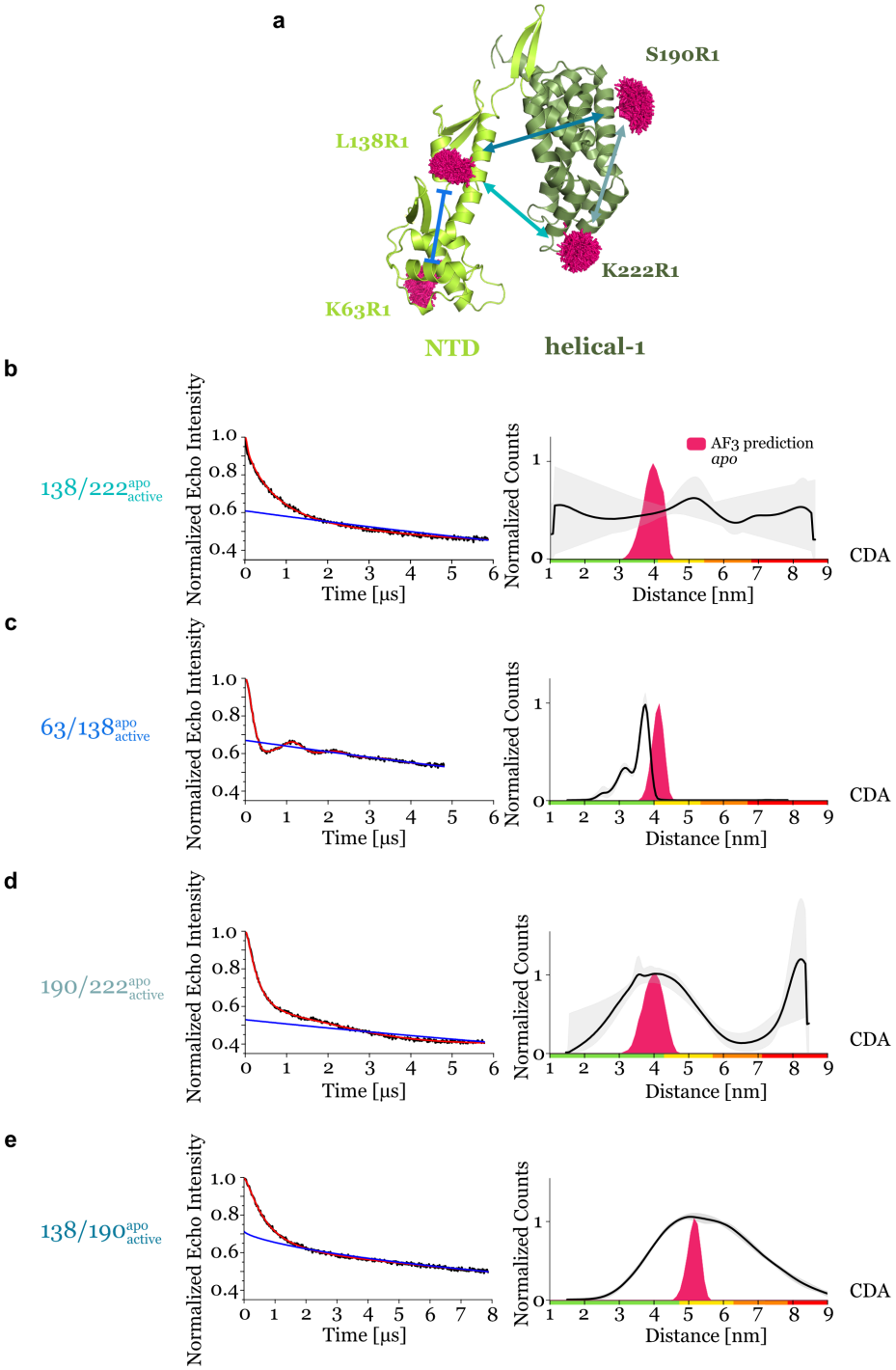


Figure 8.3-3: CDA analysis with background of different *apo* constructs in REC lobe. a, AF3 prediction of the REC lobe with MTSL rotamer clouds in magenta. The light green domain is the NTD and the olive-green domain is the helical-1 domain. The arrows indicate flexibility between two protein regions and a bar denotes rigidity between two protein regions. **b-e,** The PELDOR time trace and the distance distributions are shown from the *apo* state of the active double labelled protein constructs. The time traces include the fits in red that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214] and the background function in blue. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean

reliable, red: not reliable). The predictions of the distance distributions based on the AF3 model is shown as magenta-shaded distributions.

8.3.4 Construct 756/926

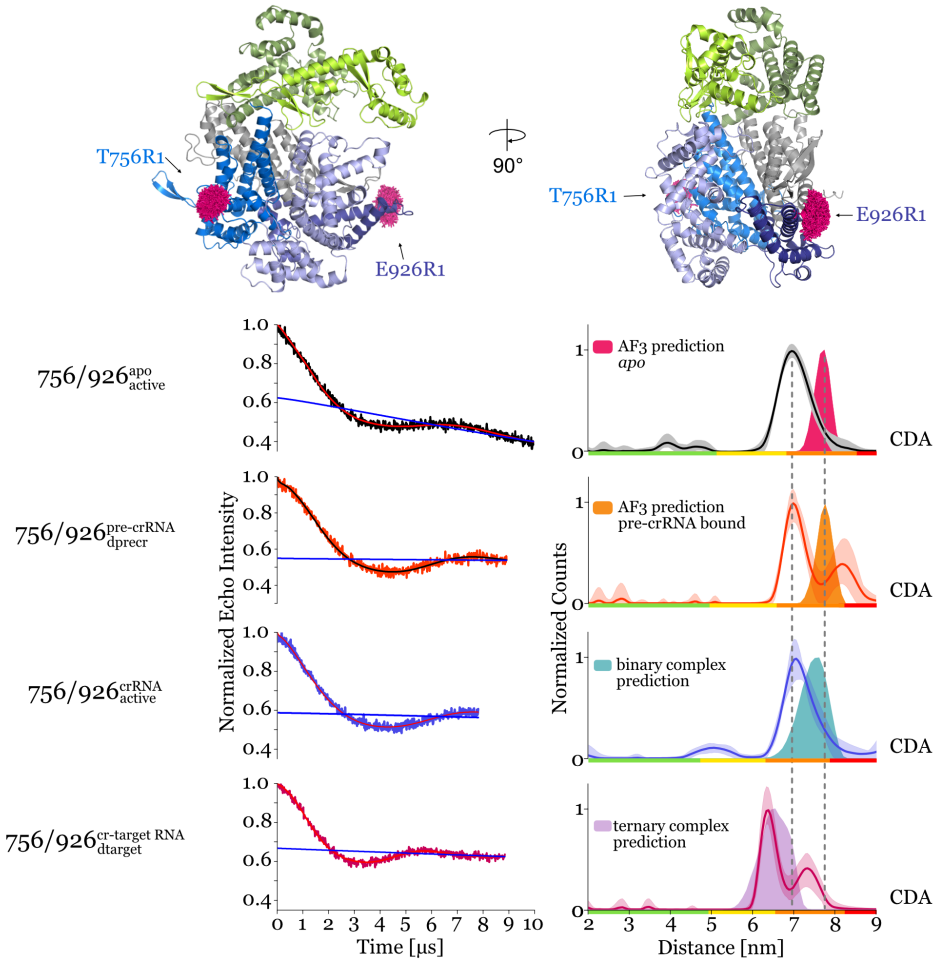


Figure 8.3-4: Full CDA analysis of 756/926. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 756 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits in red that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214] and the background function in blue. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models, the binary cryo-EM structure, and the ternary X-ray crystallographic structure are shown as shaded distributions.

8.3.5 Construct 462/926

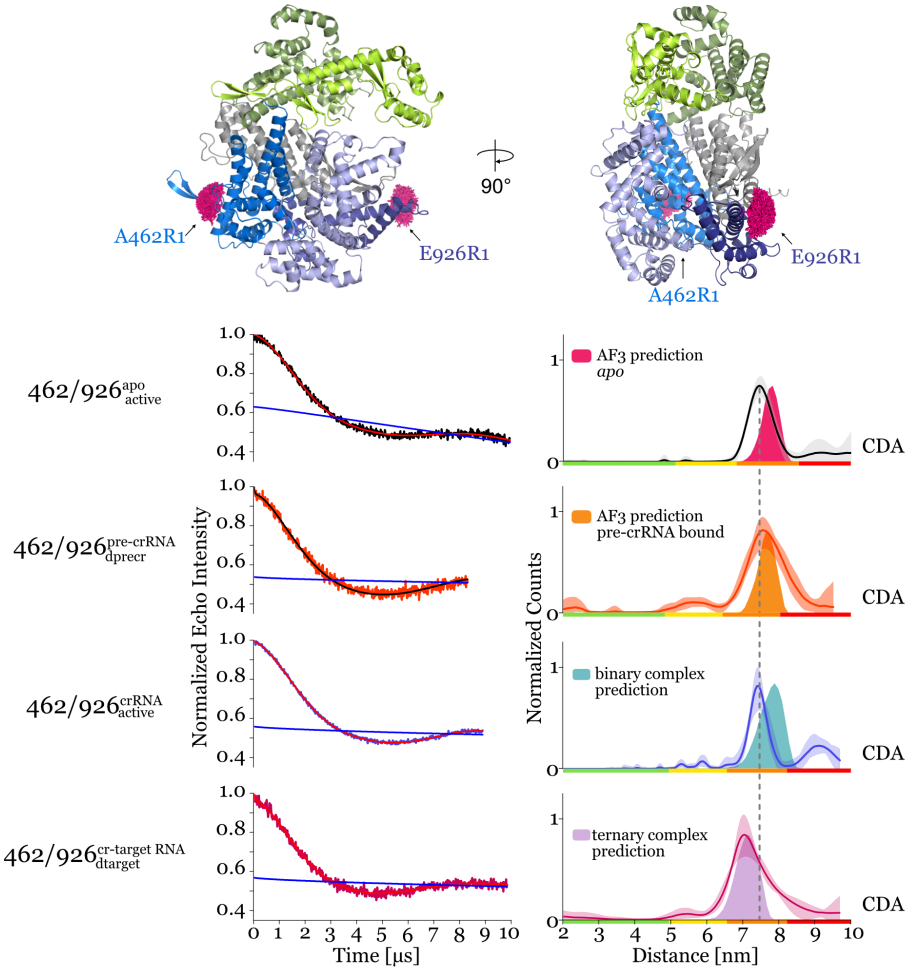


Figure 8.3-5: Full CDA analysis of 462/926. The AF3 structure of *apo LbuCas13a* is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 462 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits in red that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214] and the background function in blue. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models, the binary cryo-EM structure, and the ternary X-ray crystallographic structure are shown as shaded distributions.

8.3.6 Construct 660/926

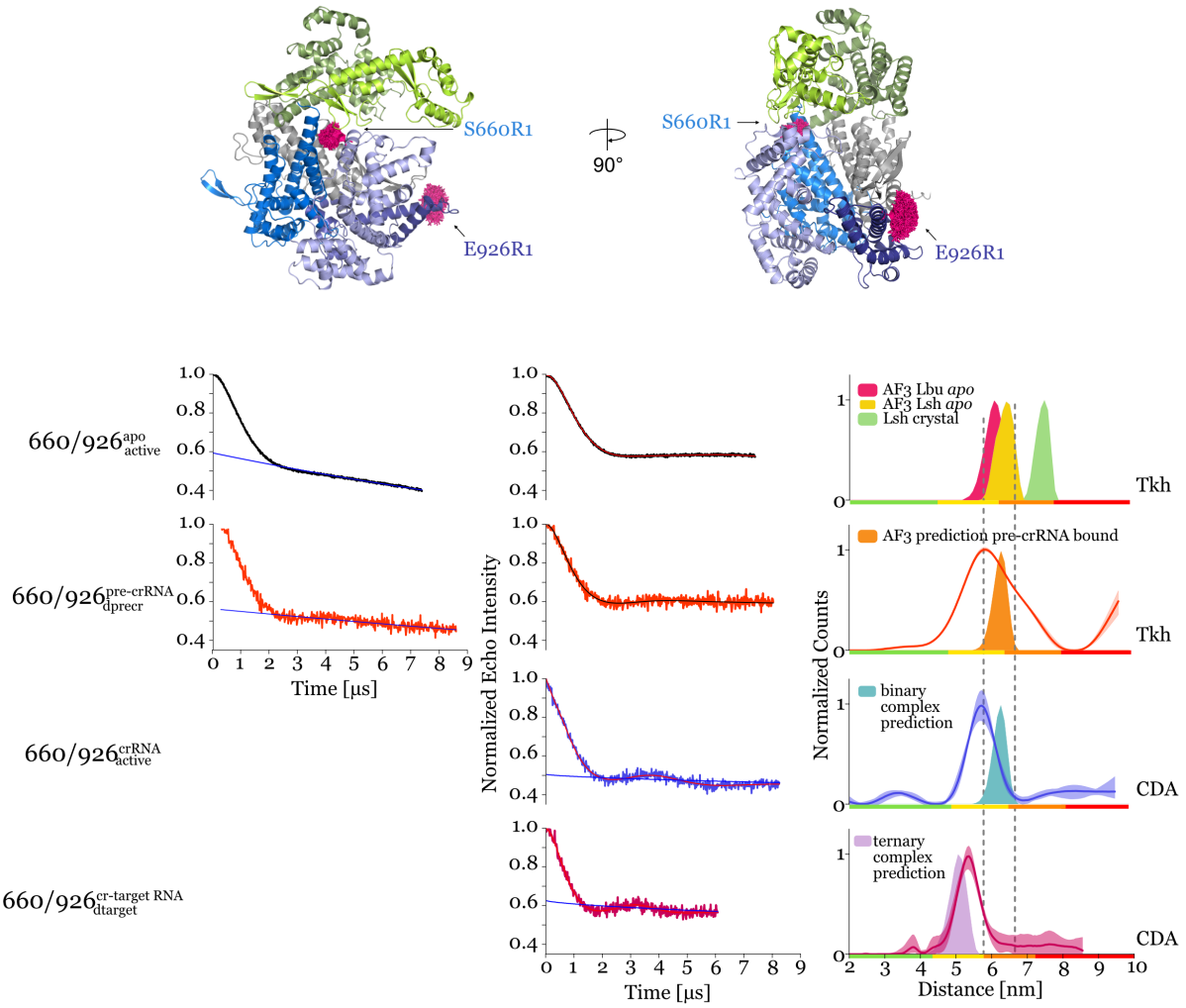


Figure 8.3-6: PELDOR analysis for the protein constructs labelled at 660/926. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 660 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through Tikhonov regularization and the ComparativeDeerAnalyzer (CDA)[214]. In Tikhonov regularization, an α was chosen that lies in the intersection between both legs of the L-curve. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models of *Lbu*Cas13a and *Lsh*Cas13a, the X-ray crystallographic structure from *apo Lsh*Cas13a, the binary cryo-EM structure of *Lbu*Cas13a, and the ternary X-ray crystallographic structure of *Lbu*Cas13a are shown as shaded distributions.

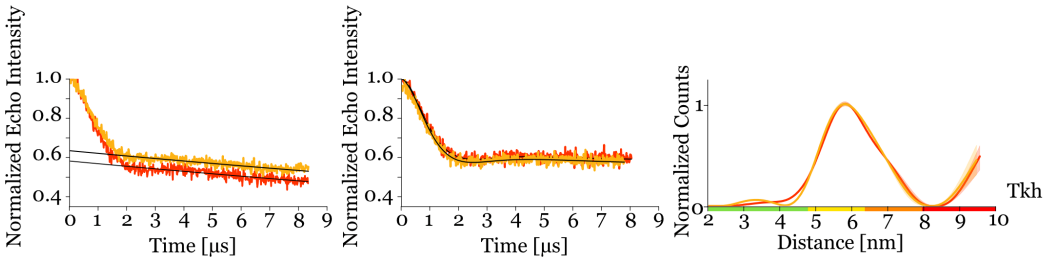


Figure 8.3-7: PELDOR analysis of the replicate dprecr 660/926 construct with pre-crRNA. The PELDOR measurement in dark orange is the same as the dark orange one in Figure 8.3-6 and the light orange measurement is the technical repeat. The time traces include the background functions (left) and the fits (middle) that were done with the software DeerAnalysis[276], through Tikhonov regularization. In Tikhonov regularization, an α was chosen that lies in the intersection between both legs of the L-curve. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable).

8.3.7 Construct 660/756

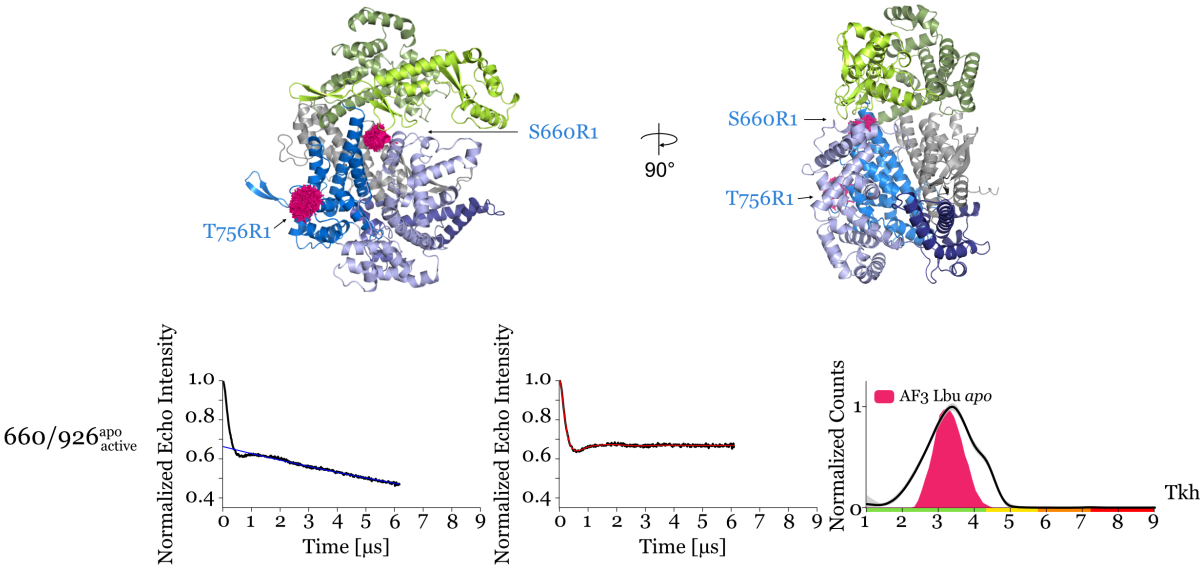


Figure 8.3-8: Tikhonov analysis of the protein construct labelled at 660/756 with PELDOR background function. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 660 and 756. The measurement shown is from the *apo* state of the active double labelled protein construct. The time trace includes the fits that were done with the software DeerAnalysis[276], through Tikhonov regularization. In Tikhonov regularization, an α was chosen that lies in the intersection between both legs of the L-curve. The distance distribution on the right contains the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The prediction of the distance distributions based on AF3 model of *Lbu*Cas13a is shown as shaded distributions.

8.3.8 Construct 190/756

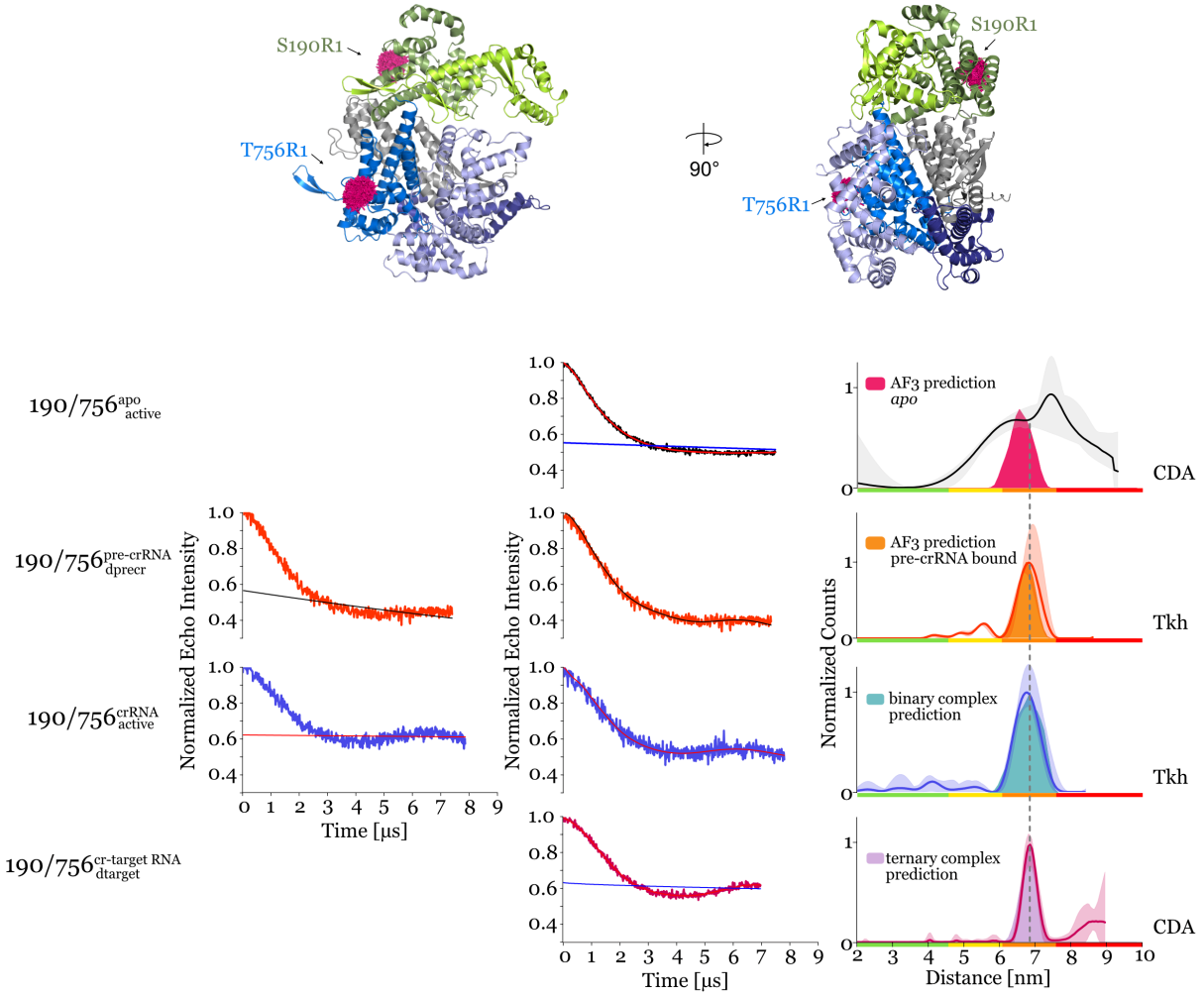


Figure 8.3-9: PELDOR analysis for the protein constructs labelled at 660/926. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 660 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the fits that were done with the software DeerAnalysis[276], through Tikhonov regularization and the ComparativeDeerAnalyzer (CDA)[214]. In Tikhonov regularization, an α was chosen that lies in the intersection between both legs of the L-curve. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models of *Lbu*Cas13a and *Lsh*Cas13a, the X-ray crystallographic structure from *apo Lsh*Cas13a (PDB: 5WTJ)[88], the binary cryo-EM structure of *Lbu*Cas13a, and the ternary X-ray crystallographic structure of *Lbu*Cas13a are shown as shaded distributions.

8.3.9 Construct 190/926

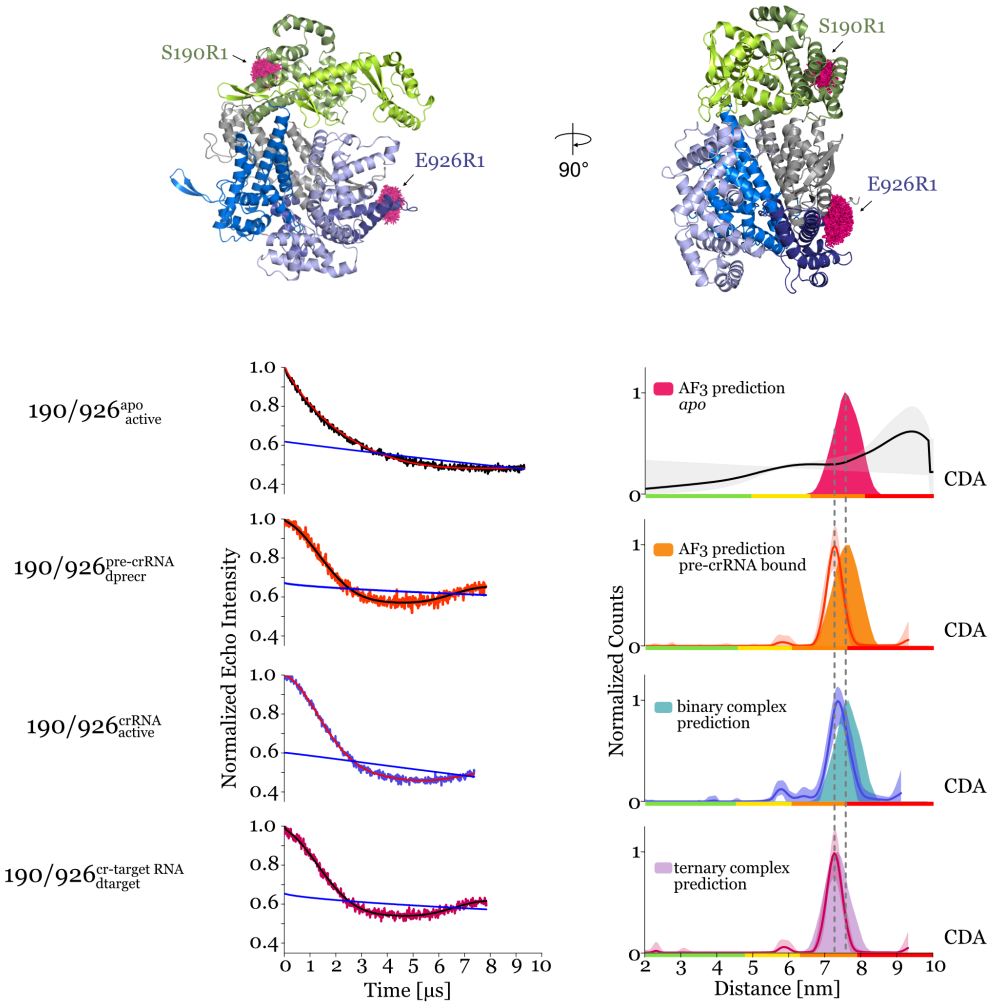


Figure 8.3-10: Full CDA analysis of labelled 190/926. The AF3 structure of *apo* *Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 190 and 926. From top to bottom measurements are shown from the *apo* state of the active double labelled protein construct, the dprecr construct with pre-crRNA, the active construct with crRNA, and the dtarget construct with cr- and target RNA. The PELDOR time trace and the distance distributions are shown for each construct. The time traces include the background function in blue and the fits that were done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. The distance distributions on the right contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predictions of the distance distributions based on AF3 models of *Lbu*Cas13a, the binary cryo-EM structure of *Lbu*Cas13a, and the ternary X-ray crystallographic structure of *Lbu*Cas13a are shown as shaded distributions.

8.3.10 Construct 138/926

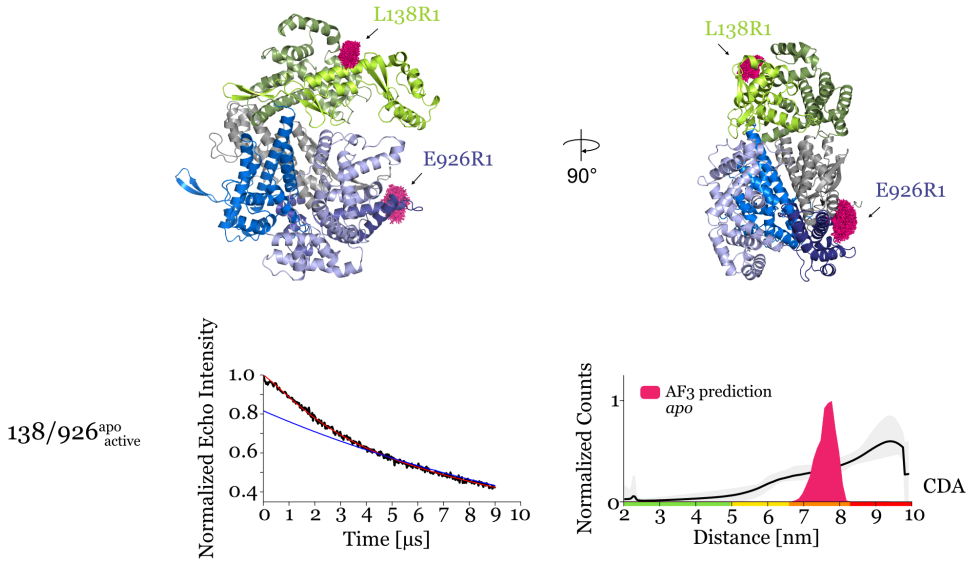


Figure 8.3-11: Full CDA analysis of labelled 138/926. The AF3 structure of *apo Lbu*Cas13a is shown with two orientations on top, with magenta MTSL rotamer clouds at the positions 138 and 926. The PELDOR time trace and distribution are shown from the *apo* state of the active double labelled protein construct. The time tracs includes the background function in blue and fit in red that was done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. The distance distribution on the right contains the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The prediction of the distance distribution based on AF3 model of *Lbu*Cas13a, is shown as magenta shaded distribution.

8.3.11 dprecr and dtarget single cysteine constructs for dimer studies

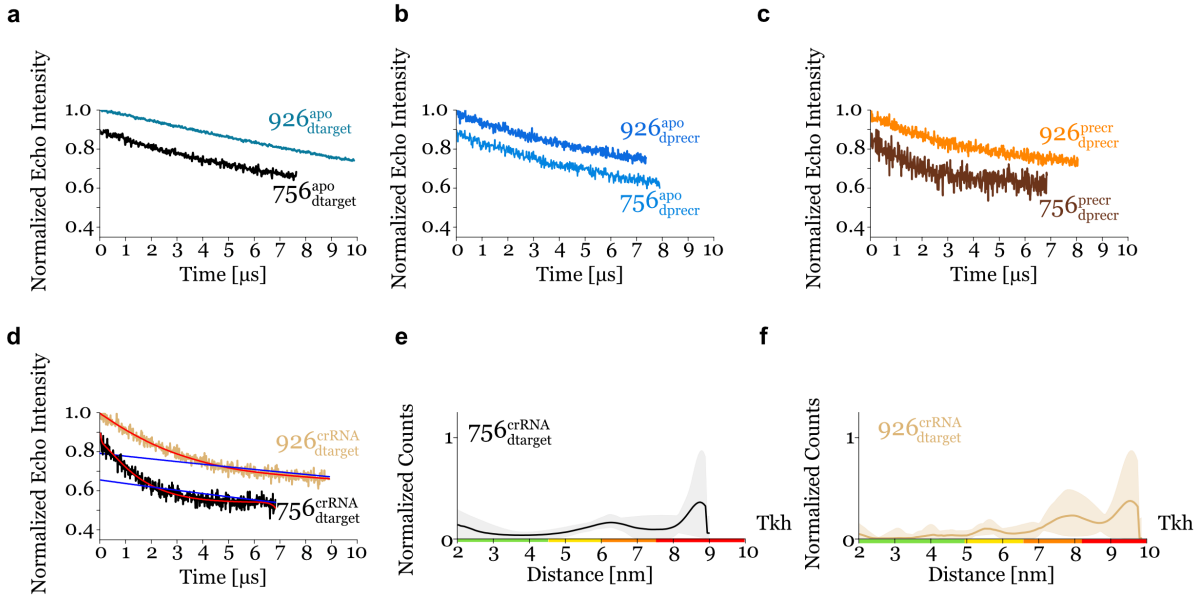


Figure 8.3-12: PELDOR spectroscopic analysis of single labelled 756 and 926 constructs. For visualization all time traces are shown with a y-axis offset of 0.1. **a,** Time traces of *apo* dtarget constructs. **b,** Time traces of *apo* dprecr constructs. **c,** Time traces of pre-crRNA bound dprecr constructs. **d,** Time traces, fits and background function of dtarget constructs with crRNA including their fit in red that was done with the software DeerAnalysis[276], through the ComparativeDeerAnalyzer (CDA)[214]. **e,** Distance distribution of the black time trace shown in d. The distance distribution contains the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). **f,** Distance distribution of the yellow/beige time trace shown in d. The representation is the same as in e.

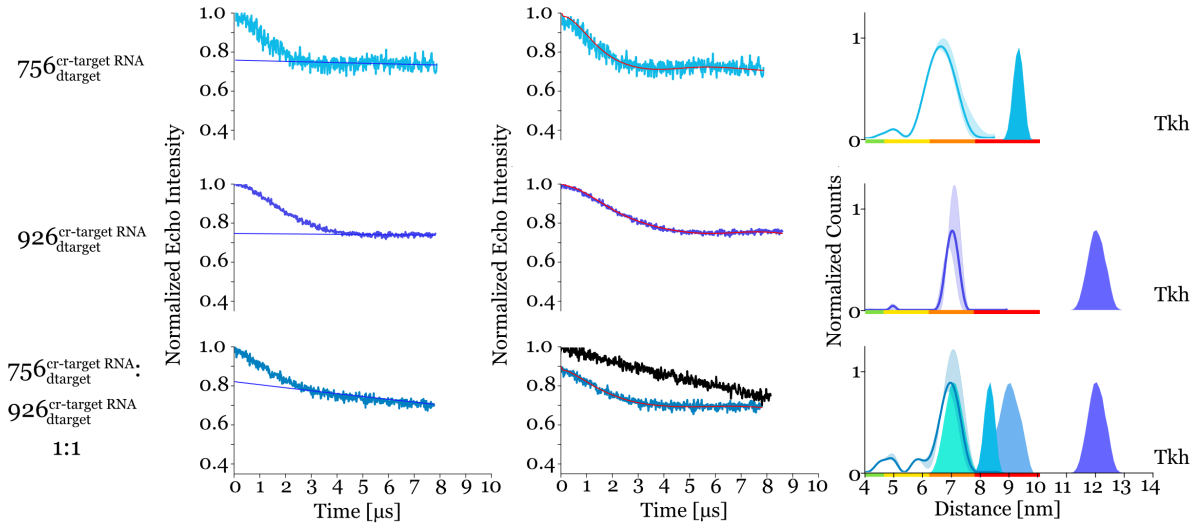
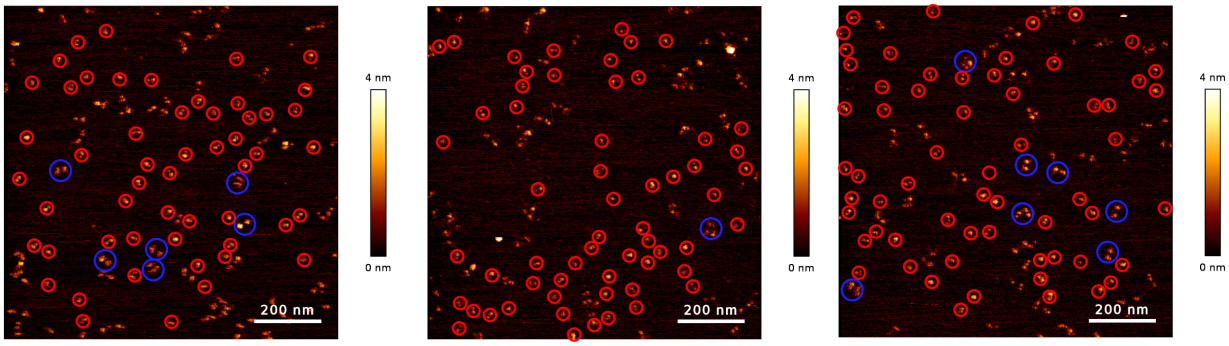


Figure 8.3-13: PELDOR analysis of the ternary complex with single labelled constructs 756 and 926. Time traces with their respective background (left), fit (middle), and distance distributions (right) are shown for different constructs. From top to bottom, the dtarget constructs 756 and 926 are shown first, with crRNA and target RNA bound to them. The third row shows the data for a 1:1 mixture of the dtarget 756 and 926 constructs in *apo* and with crRNA and target RNA. The fit of the respective time traces was done with the software DeerAnalysis[276], through Tikhonov regularization (Tkh). In Tikhonov regularization, an α was chosen that lies in the intersection between both legs of the L-curve. The distance distributions contain the 95% confidence interval as shaded area in the same color as the distribution, and the colored bars on the bottom, which describe the reliability of the distribution in the respective region (green: reliable distribution, yellow: width and mean are reliable, orange: mean reliable, red: not reliable). The predicted distributions are shown as shaded areas and are the same as in Figure 4.4-1 b.

8.4 Additional AFM data

a

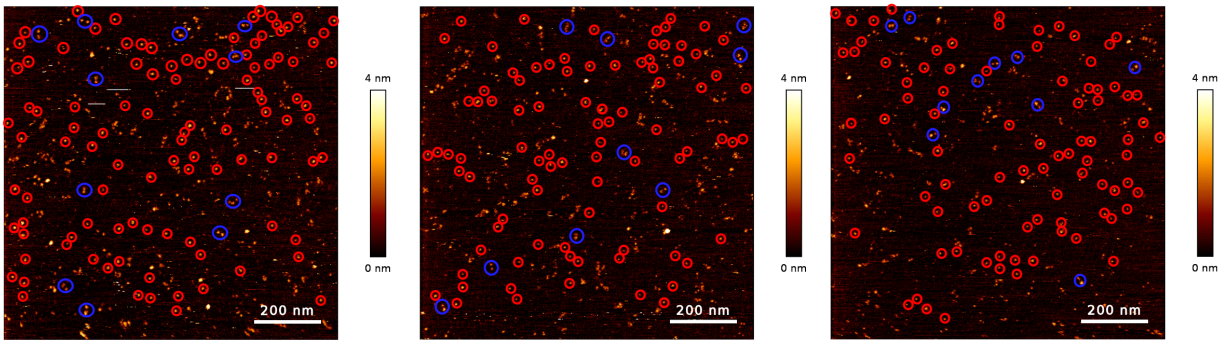


wt *Lbu*Cas13a apo
53 monomers
6 dimers
10% dimer

wt *Lbu*Cas13a apo
59 monomers
1 dimer
1.7% dimer

wt *Lbu*Cas13a apo
61 monomers
7 dimers
10% dimer

b

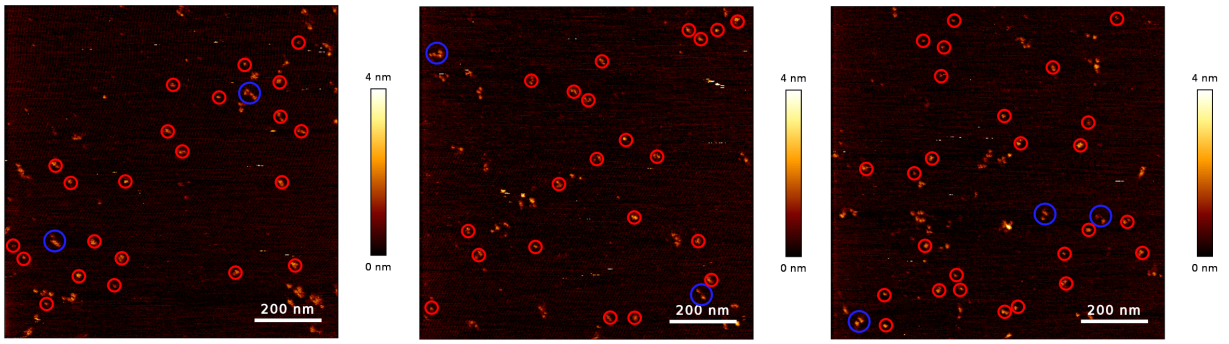


dprecr wt *Lbu*Cas13a + pre-crRNA
105 monomers
11 dimers
9.5% dimer

dprecr wt *Lbu*Cas13a + pre-crRNA
101 monomers
9 dimers
8.9% dimer

dprecr wt *Lbu*Cas13a + pre-crRNA
86 monomers
10 dimers
10.4% dimer

c

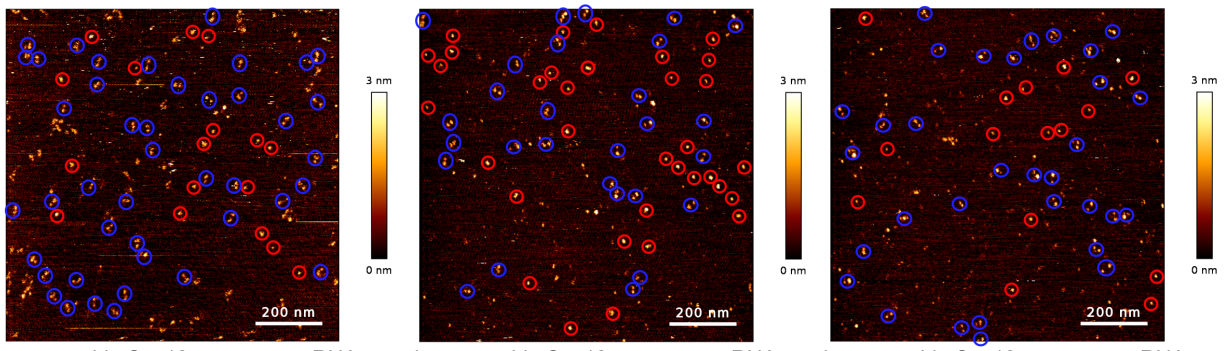


wt *Lbu*Cas13a + crRNA
22 monomers
2 dimers
8% dimer

wt *Lbu*Cas13a + crRNA
22 monomers
2 dimers
8% dimer

wt *Lbu*Cas13a + crRNA
26 monomers
3 dimers
10% dimer

d



dtarget wt *Lbu*Cas13a + cr-target RNA
17 monomers
43 dimers
71% dimer

dtarget wt *Lbu*Cas13a + cr-target RNA
35 monomers
28 dimers
41% dimer

dtarget wt *Lbu*Cas13a + cr-target RNA
15 monomers
33 dimers
68% dimer

Figure 8.4-1: AFM images of each *Lbu*Cas13a state. AFM images acquired from **a,** wt *Lbu*Cas13a without RNA. **b,** dprecr construct of wt *Lbu*Cas13a with pre-crRNA. **c,** wt *Lbu*Cas13a with crRNA. **d,** dtarget construct of wt *Lbu*Cas13a with target RNA. Blue circles highlight dimers and red circles highlight monomers.

8.5 PELDOR analysis reports from CDA

This part of the Appendix shows the PELDOR analysis reports that are given as output by the ComparativeDeerAnalyzer and the version is showed below:

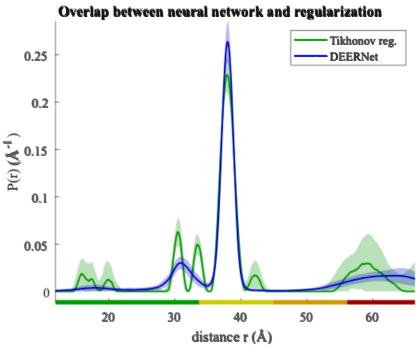
DEERNet Spinach SVN Rev 5662and DeerLab 0.9.1 Tikhonov regularization

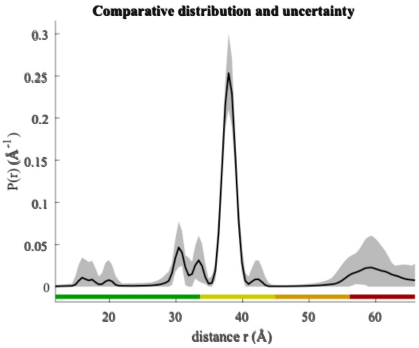
ComparativeDEERAnalyzer version 2.0

see: S. G. Worswick et al., DOI: 10.1126/sciadv.aat5218,L.Fabregas Ibanez et al., DOI: 10.5194/mr-1-209-2020

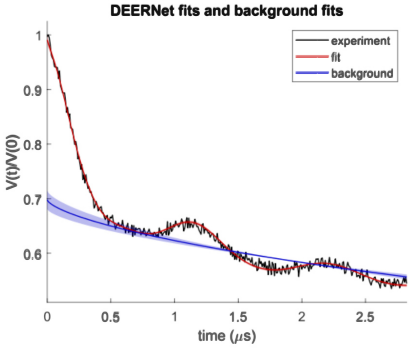
138/222$^{crRNA}_{active}$ with AA

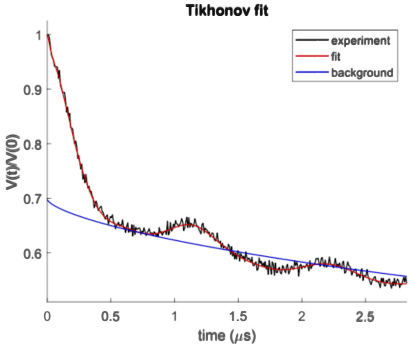
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.296

Signal-to-noise ratio: 42.3 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00768

From DEERNet fit: 0.00699

From Tikhonov fit: 0.00643

Zero time: 135 ns

Maximum time: 2824 ns

The last 2 % of the data was cut off

Time increment: 8 ns

Phase: -27.4 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.14

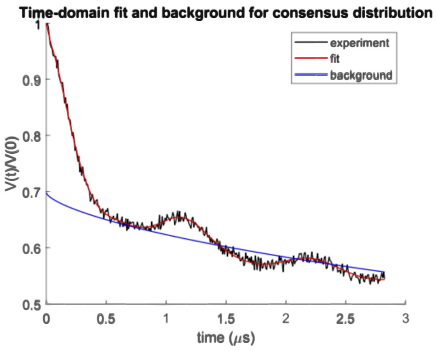
Reg. par. initial estimate by L-curve corner: 3.16

Overlap between DEERNet and regularization solutions: 0.785

Predicted overlap of consensus solution with ground truth: 0.66...0.83

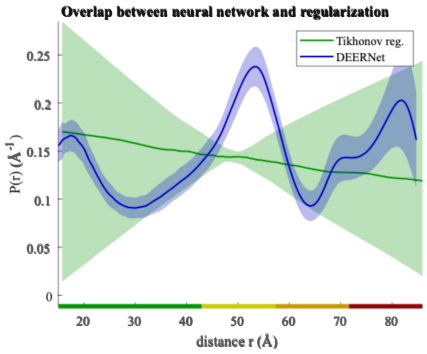
Mean distance: 37.9 Å

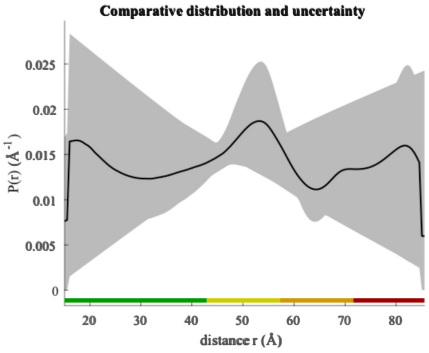
Distance standard deviation: 9.4 Å



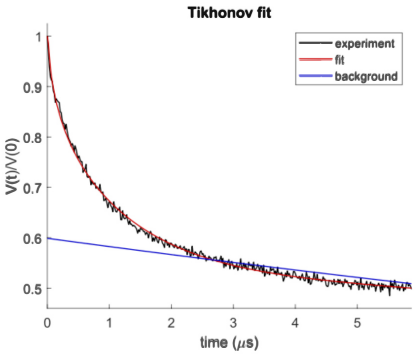
138/222$^{apo}_{active}$ with AA

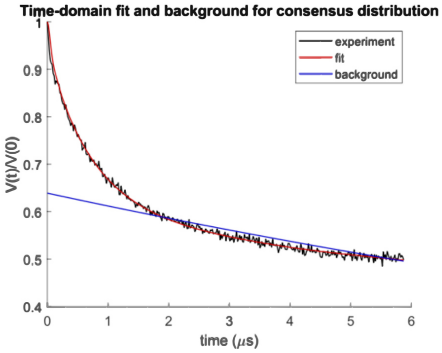
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.361

Signal-to-noise ratio: 60.7 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00570

From Tikhonov fit: 0.00724

Zero time: 133 ns

Maximum time: 5872 ns

Initial 1 point(s) of the data clipped

Time increment: 16 ns

Phase: -1.2 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 3230.50

Reg. par. initial estimate by lr: 6.31

Overlap between DEERNet and regularization solutions: 0.854

Predicted overlap of consensus solution with ground truth: 0.73...0.90

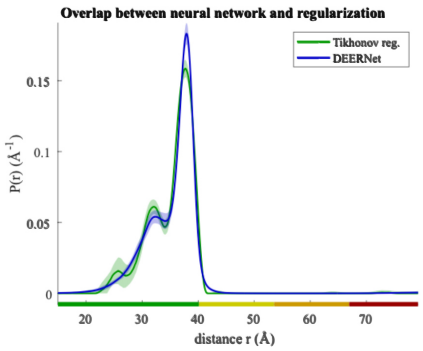
Mean distance: 50.2 Å

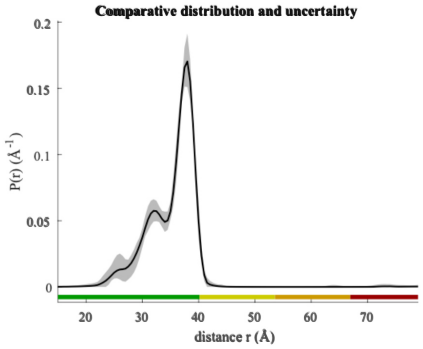
Single Gaussian provided different mean distance. Distribution may be incomplete.

Distance standard deviation: 20.2 Å

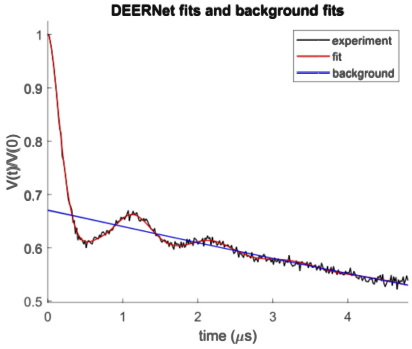
63/138$^{apo}_{active}$ with AA

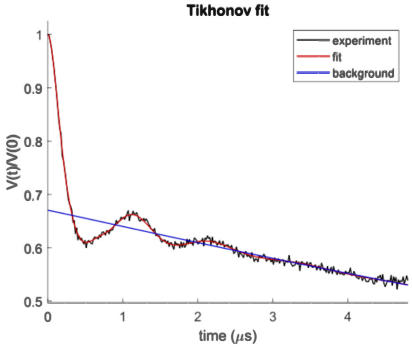
1. Distance distributions

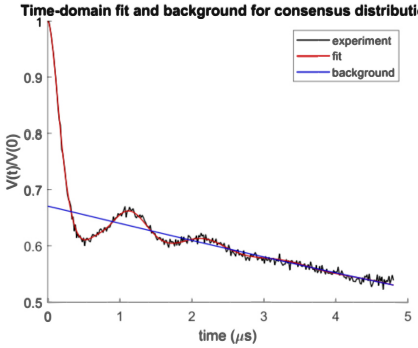




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.330

Signal-to-noise ratio: 63.1 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00580

From DEERNet fit: 0.00523

From Tikhonov fit: 0.00508

Zero time: 138 ns

Maximum time: 4800 ns

The last 1 % of the data was cut off

Time increment: 16 ns

Phase: -5.7 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.50

Reg. par. initial estimate by L-curve corner: 3.98

Overlap between DEERNet and regularization solutions: 0.926

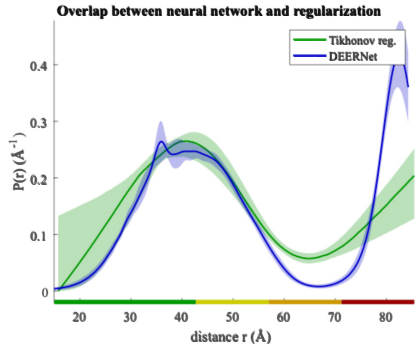
Predicted overlap of consensus solution with ground truth: 0.79...0.96

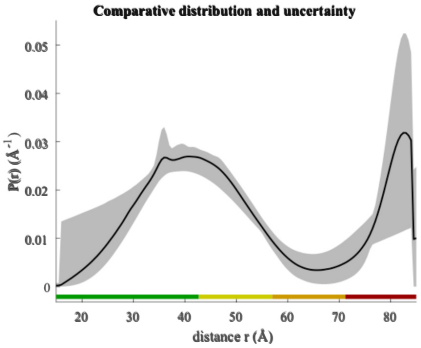
Mean distance: 35.8 Å

Distance standard deviation: 3.3 Å

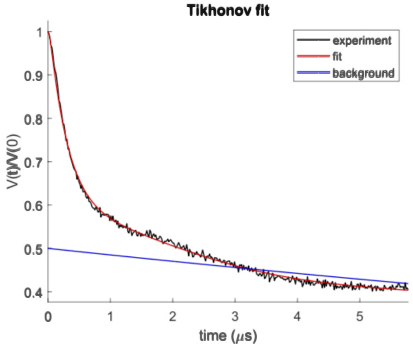
$190/222^{apo}_{active}$ with AA

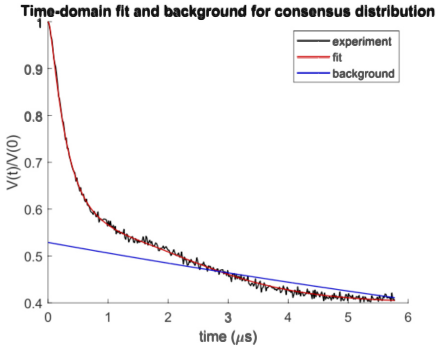
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.471

Signal-to-noise ratio: 77.3 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00533

From Tikhonov fit: 0.00675

Zero time: 135 ns

Maximum time: 5776 ns

The last 1 % of the data was cut off

Time increment: 16 ns

Phase: -3.6 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 142.09

Reg. par. initial estimate by lr: 25.12

Overlap between DEERNet and regularization solutions: 0.844

Predicted overlap of consensus solution with ground truth: 0.72...0.89

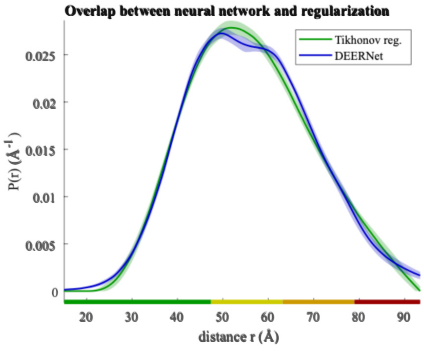
Mean distance: 51.2 Å

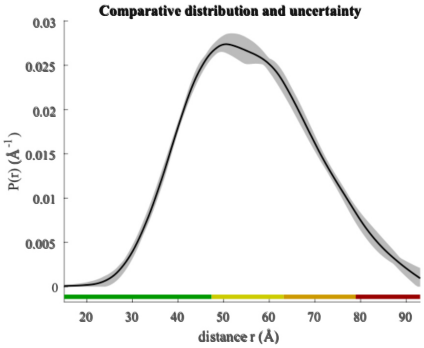
Single Gaussian provided different mean distance. Distribution may be incomplete.

Distance standard deviation: 18.8 Å

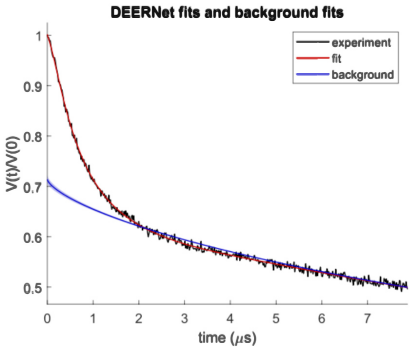
138/190$^{apo}_{active}$ with AA

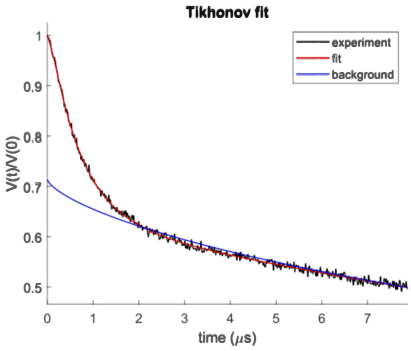
1. Distance distributions

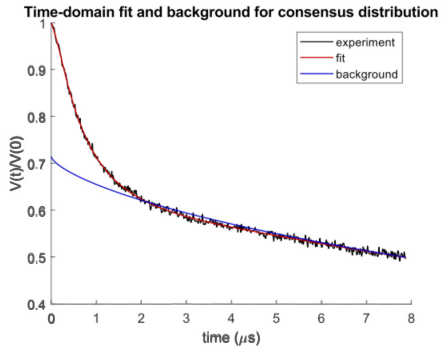




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.288

Signal-to-noise ratio: 55.1 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00513

From DEERNet fit: 0.00522

From Tikhonov fit: 0.00521

Zero time: 131 ns

Maximum time: 7872 ns

Time increment: 16 ns

Phase: 0.3 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 56.17

Reg. par. initial estimate by L-curve corner: 79.43

Overlap between DEERNet and regularization solutions: 0.977

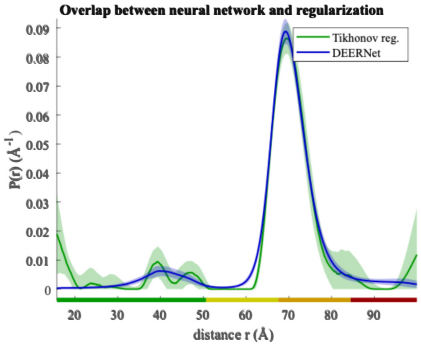
Predicted overlap of consensus solution with ground truth: 0.84...1.00

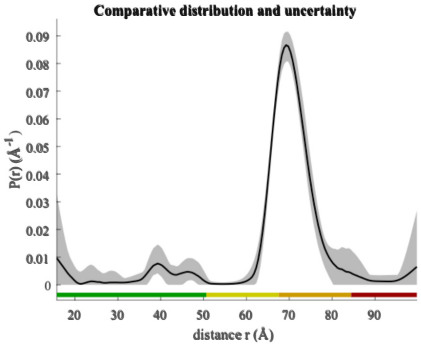
Mean distance: 55.4 Å

Distance standard deviation: 14.9 Å

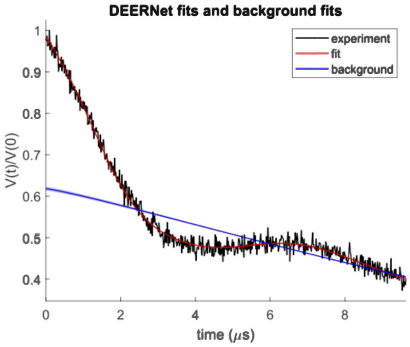
756/962$^{apo}_{active}$ with AA

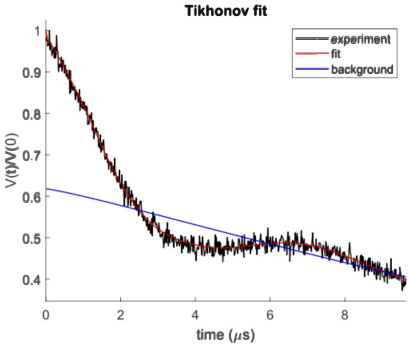
1. Distance distributions

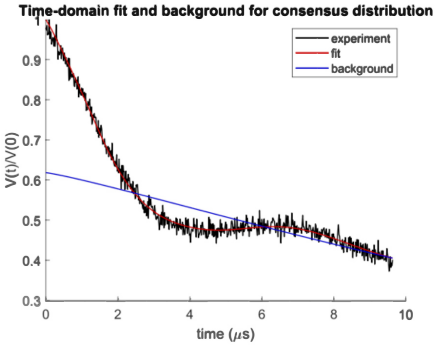




2. Fits of time-domain data







Modulation depth: 0.369

Signal-to-noise ratio: 25.8 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01353

From DEERNet fit: 0.01431

From Tikhonov fit: 0.01474

Zero time: 369 ns

Maximum time: 9632 ns

The last 5 % of the data was cut off

Initial 10 point(s) of the data clipped

Time increment: 16 ns

Phase: -3.2 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 3.13

Reg. par. initial estimate by L-curve corner: 50.12

Overlap between DEERNet and regularization solutions: 0.905

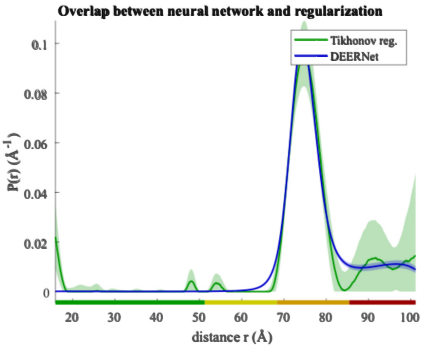
Predicted overlap of consensus solution with ground truth: 0.77...0.94

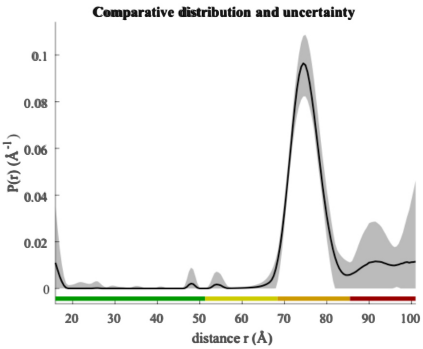
Mean distance: 69.8 Å

Distance standard deviation: 4.9 Å

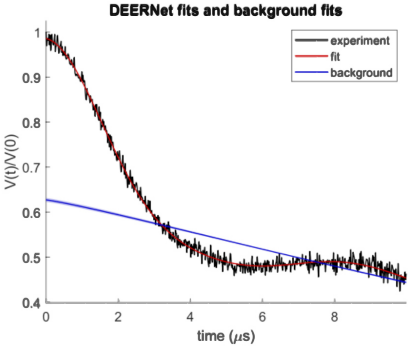
462/926$^{apo}_{active}$ with AA

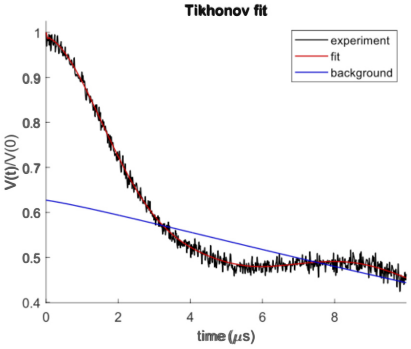
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.364

Signal-to-noise ratio: 32.2 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01053

From DEERNet fit: 0.01129

From Tikhonov fit: 0.01146

Zero time: 55 ns

Maximum time: 9984 ns

Time increment: 16 ns

Phase: 9.4 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.99

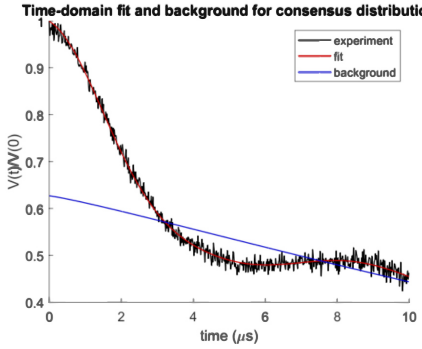
Reg. par. initial estimate by L-curve corner: 63.10

Overlap between DEERNet and regularization solutions: 0.906

Predicted overlap of consensus solution with ground truth: 0.77...0.95

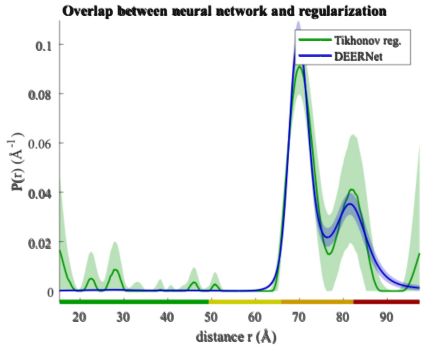
Mean distance: 76.8 Å

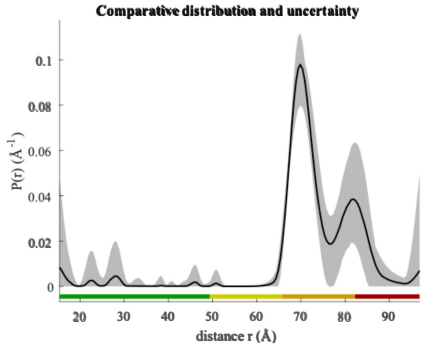
Distance standard deviation: 5.7 Å



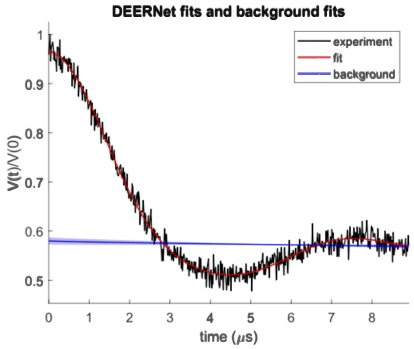
756/926$^{pre-crRNA}_{dprecr}$ with AA

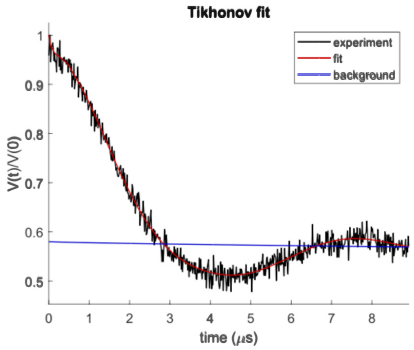
1. Distance distributions





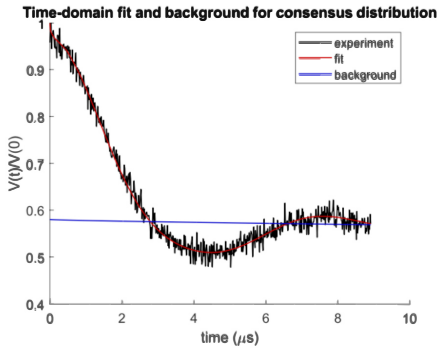
2. Fits of time-domain data





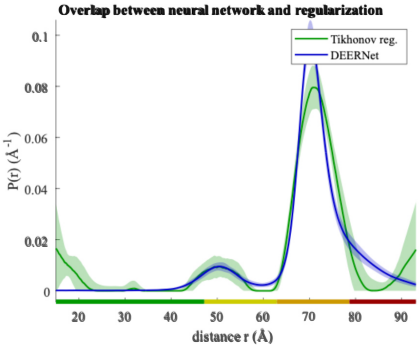
3. Experimental and processing parameters

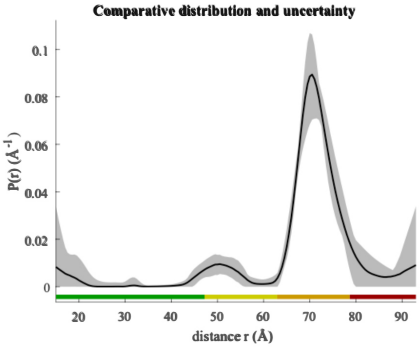
Modulation depth: 0.400
Signal-to-noise ratio: 26.5 (w.r.t. modulation)
Noise estimates normalized to maximum signal
From imaginary part: 0.01438
From DEERNet fit: 0.01512
From Tikhonov fit: 0.01527
Zero time: 122 ns
Maximum time: 8912 ns
Initial 2 point(s) of the data clipped
Time increment: 16 ns
Phase: -8.5 degree
Ensemble of 32 neural networks
Background separation by neural network
Background dimension: 3
Regularization parameter by best overlap with neural network solution
Regularization parameter used: 0.88
Reg. par. initial estimate by L-curve corner: 79.43
Overlap between DEERNet and regularization solutions: 0.874
Predicted overlap of consensus solution with ground truth: 0.74...0.92
Mean distance: 73.3 Å
Distance standard deviation: 7.0 Å



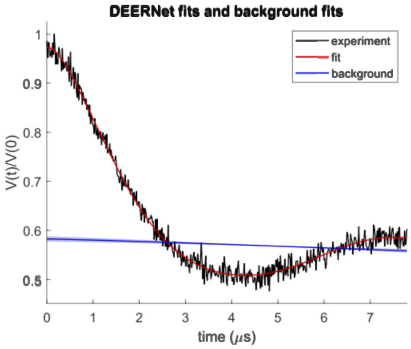
756/926$^{crRNA}_{active}$ with AA

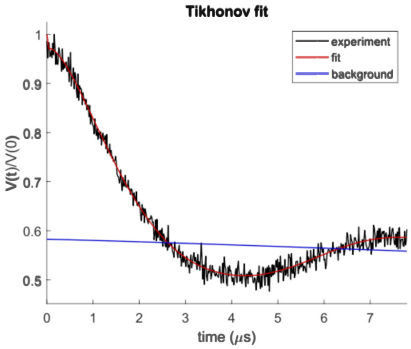
1. Distance distributions

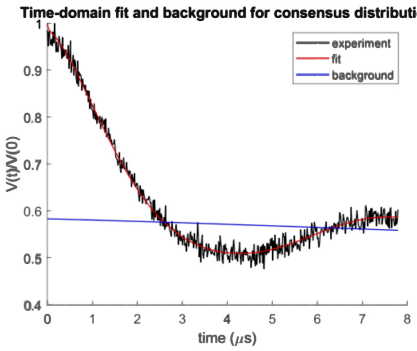




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.401

Signal-to-noise ratio: 28.4 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01456

From DEERNet fit: 0.01415

From Tikhonov fit: 0.01435

Zero time: 222 ns

Maximum time: 7792 ns

Initial 2 point(s) of the data clipped

Time increment: 16 ns

Phase: -19.5 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 3.94

Reg. par. initial estimate by L-curve corner: 63.10

Overlap between DEERNet and regularization solutions: 0.828

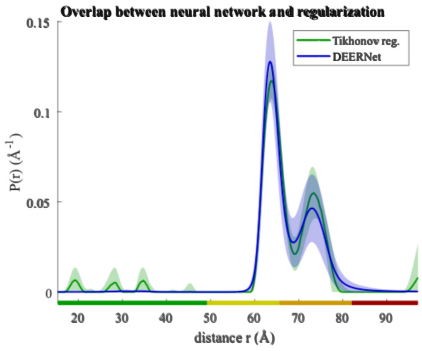
Predicted overlap of consensus solution with ground truth: 0.70...0.87

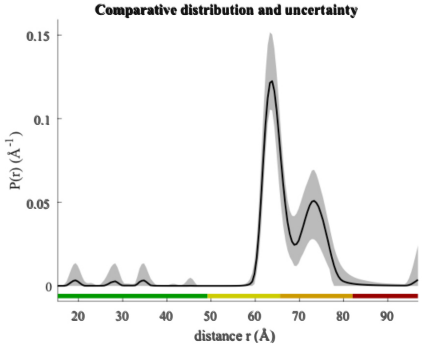
Mean distance: 71.0 Å

Distance standard deviation: 6.6 Å

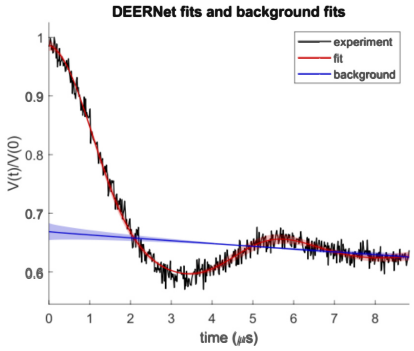
756/926$^{\text{cr-target RNA}}_{\text{dtarget}}$ with AA

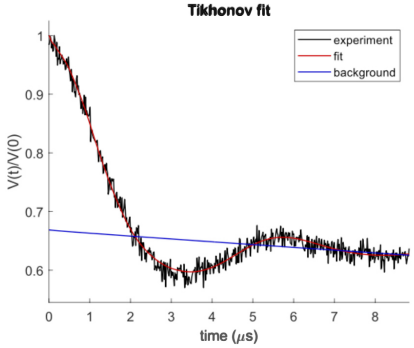
1. Distance distributions

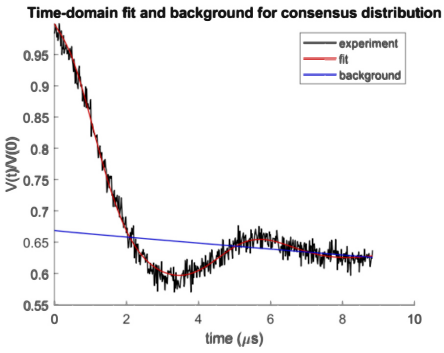




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.322

Signal-to-noise ratio: 27.8 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01183

From DEERNet fit: 0.01160

From Tikhonov fit: 0.01163

Zero time: 206 ns

Maximum time: 8832 ns

Initial 1 point(s) of the data clipped

Time increment: 16 ns

Phase: -30.3 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.70

Reg. par. initial estimate by L-curve corner: 31.62

Overlap between DEERNet and regularization solutions: 0.902

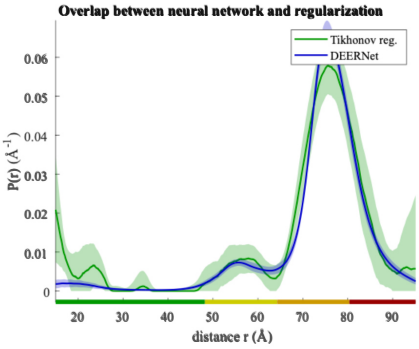
Predicted overlap of consensus solution with ground truth: 0.77...0.94

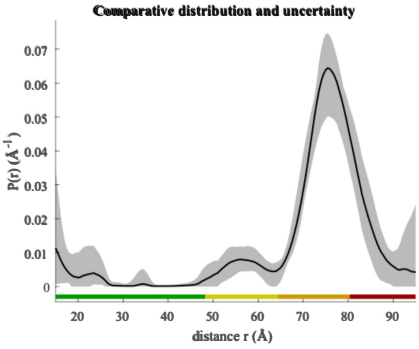
Mean distance: 66.8 Å

Distance standard deviation: 5.3 Å

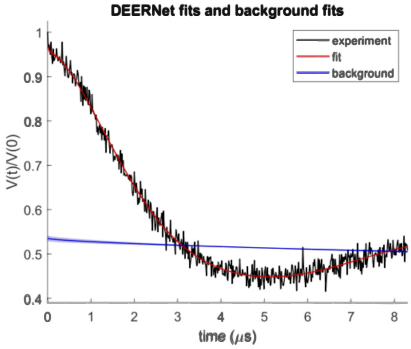
462/926$^{pre-crRNA}_{dprecr}$ with AA

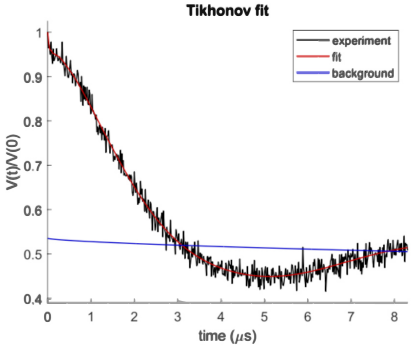
1. Distance distributions

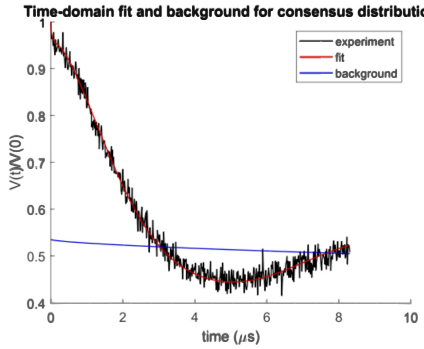




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.447

Signal-to-noise ratio: 27.4 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.02024

From DEERNet fit: 0.01631

From Tikhonov fit: 0.01659

Zero time: 237 ns

Maximum time: 8304 ns

Initial 10 point(s) of the data clipped

Time increment: 16 ns

Phase: -8.6 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 3.12

Reg. par. initial estimate by L-curve corner: 100.00

Overlap between DEERNet and regularization solutions: 0.899

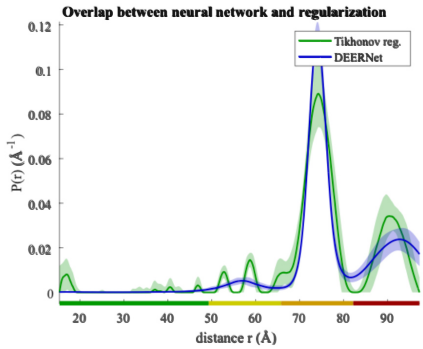
Predicted overlap of consensus solution with ground truth: 0.77...0.94

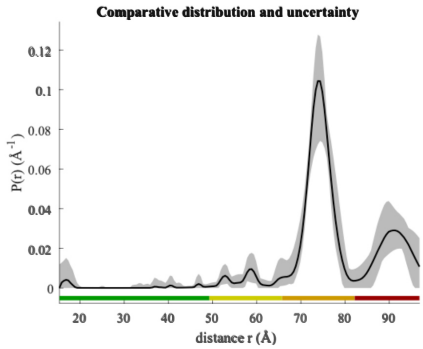
Mean distance: 75.7 Å

Distance standard deviation: 10.5 Å

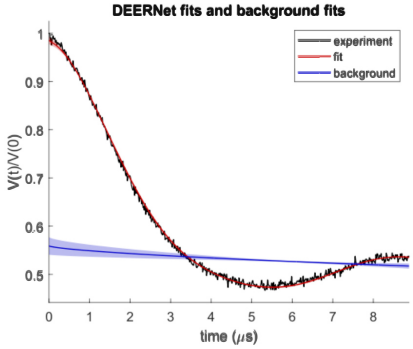
462/926$^{crRNA}_{active}$ with AA

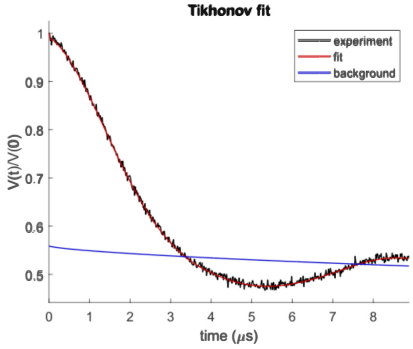
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.433

Signal-to-noise ratio: 71.2 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00492

From DEERNet fit: 0.00607

From Tikhonov fit: 0.00519

Zero time: 149 ns

Maximum time: 8880 ns

Initial 1 point(s) of the data clipped

Time increment: 16 ns

Phase: -3.5 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.22

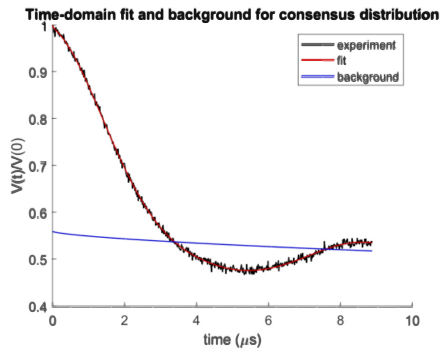
Reg. par. initial estimate by L-curve corner: 39.81

Overlap between DEERNet and regularization solutions: 0.805

Predicted overlap of consensus solution with ground truth: 0.68...0.85

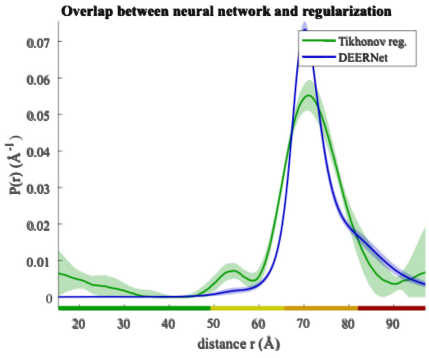
Mean distance: 76.8 Å

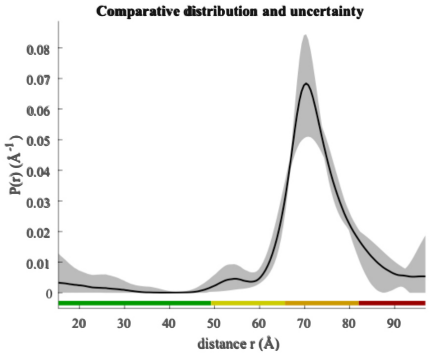
Distance standard deviation: 7.4 Å



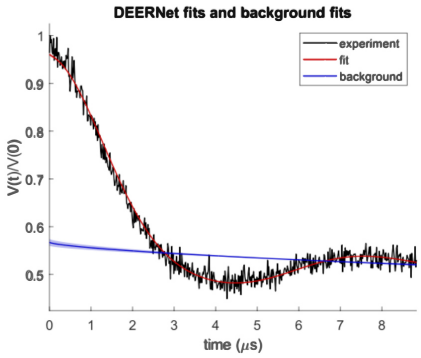
462/926$^{\text{cr-target RNA}}_{\text{dtarget}}$ with AA

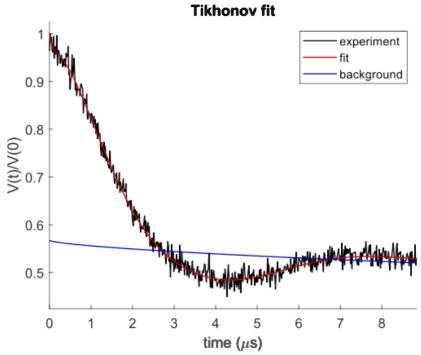
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.410

Signal-to-noise ratio: 26.4 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01572

From DEERNet fit: 0.01553

From Tikhonov fit: 0.01438

Zero time: 202 ns

Maximum time: 8832 ns

Time increment: 16 ns

Phase: -8.6 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 15.74

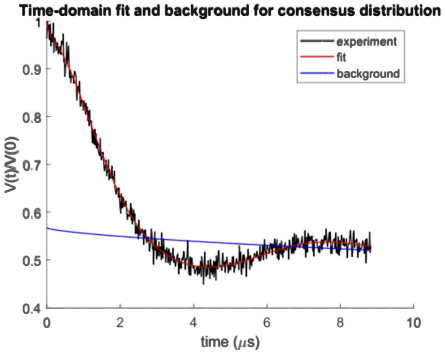
Reg. par. initial estimate by L-curve corner: 125.89

Overlap between DEERNet and regularization solutions: 0.822

Predicted overlap of consensus solution with ground truth: 0.70...0.87

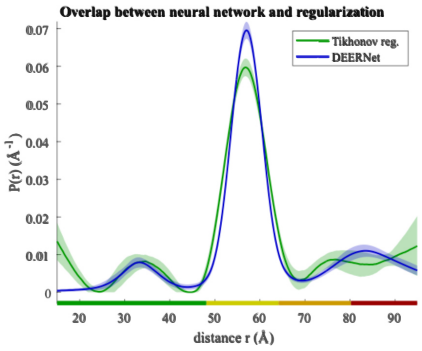
Mean distance: 71.6 Å

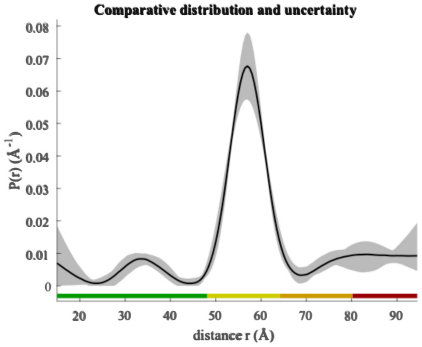
Distance standard deviation: 8.1 Å



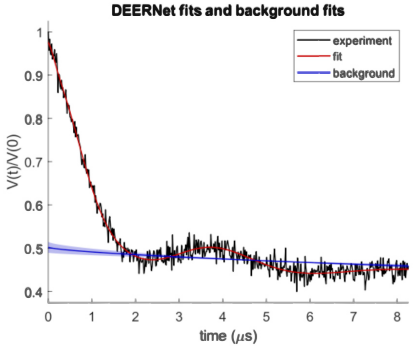
660/926$^{crRNA}_{active}$ with AA

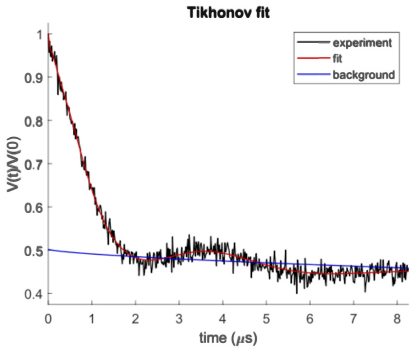
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.485

Signal-to-noise ratio: 31.7 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01476

From DEERNet fit: 0.01531

From Tikhonov fit: 0.01551

Zero time: 188 ns

Maximum time: 8256 ns

The last 1 % of the data was cut off

Initial 10 point(s) of the data clipped

Time increment: 16 ns

Phase: -21.6 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 19.91

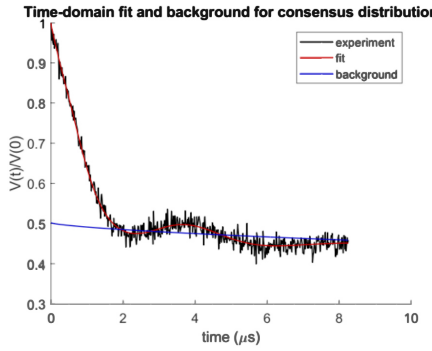
Reg. par. initial estimate by L-curve corner: 39.81

Overlap between DEERNet and regularization solutions: 0.876

Predicted overlap of consensus solution with ground truth: 0.75...0.92

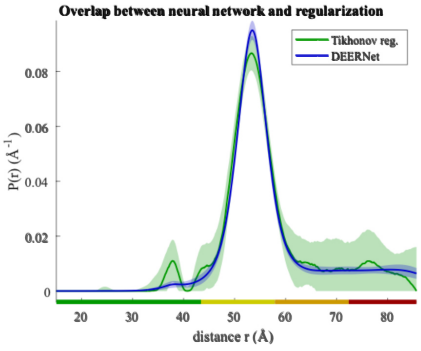
Mean distance: 59.1 Å

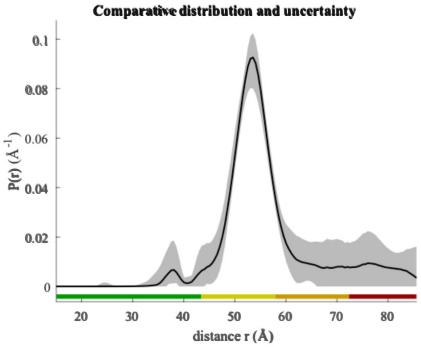
Distance standard deviation: 13.7 Å



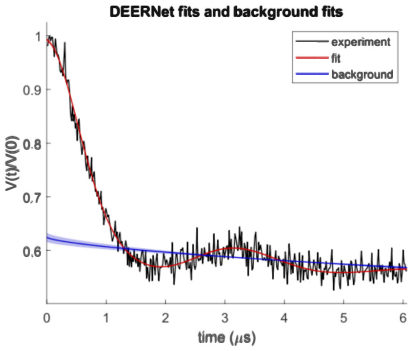
660/926$^{\text{cr-target RNA}}_{\text{dtarget}}$ with AA

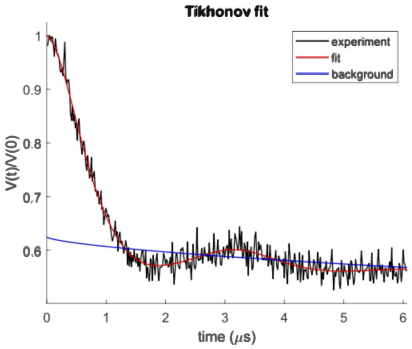
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.371

Signal-to-noise ratio: 21.6 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01666

From DEERNet fit: 0.01717

From Tikhonov fit: 0.01700

Zero time: 117 ns

Maximum time: 6064 ns

The last 5 % of the data was cut off

Initial 1 point(s) of the data clipped

Time increment: 16 ns

Phase: -19.7 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 1.76

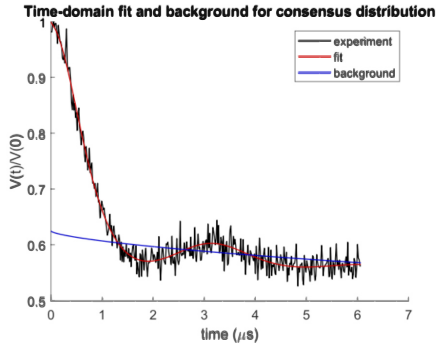
Reg. par. initial estimate by L-curve corner: 39.81

Overlap between DEERNet and regularization solutions: 0.936

Predicted overlap of consensus solution with ground truth: 0.80...0.97

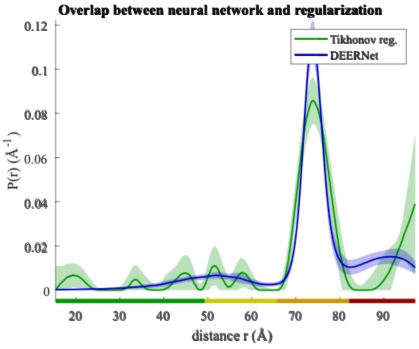
Mean distance: 54.9 Å

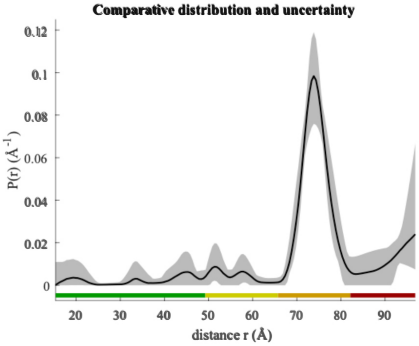
Distance standard deviation: 8.2 Å



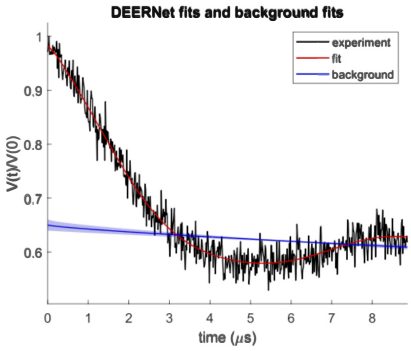
462/926$^{crRNA}_{active}$ with AA RNA from Liu et al.

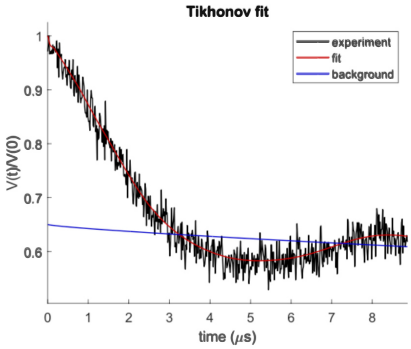
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Please consider improving signal-to-noise ratio (below 20)

Modulation depth: 0.338

Signal-to-noise ratio: 15.3 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.02207

From DEERNet fit: 0.02213

From Tikhonov fit: 0.02214

Zero time: 147 ns

Maximum time: 8864 ns

Time increment: 16 ns

Phase: -25.4 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 2.48

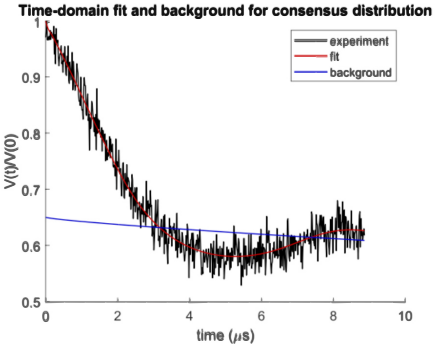
Reg. par. initial estimate by L-curve corner: 79.43

Overlap between DEERNet and regularization solutions: 0.772

Predicted overlap of consensus solution with ground truth: 0.65...0.82

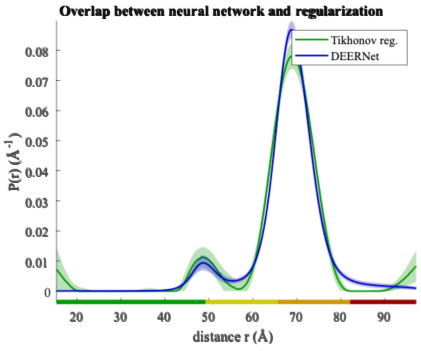
Mean distance: 74.8 Å

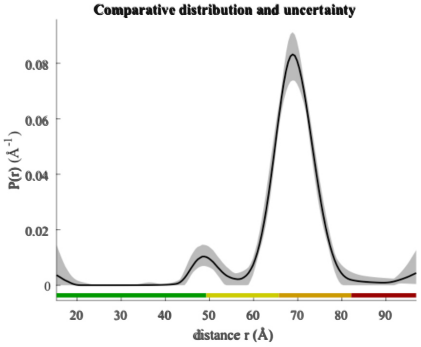
Distance standard deviation: 8.0 Å



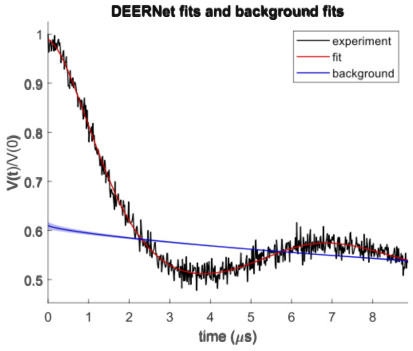
462/926 cr-target RNA dtarget with AA RNA from Liu et. al

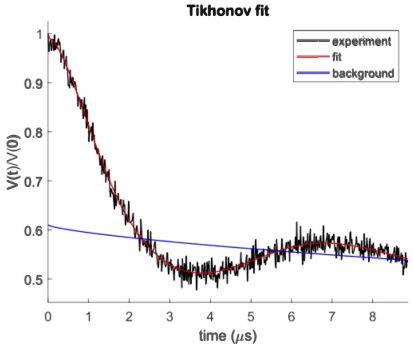
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.384

Signal-to-noise ratio: 26.3 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01465

From DEERNet fit: 0.01460

From Tikhonov fit: 0.01457

Zero time: 160 ns

Maximum time: 8864 ns

Initial 10 point(s) of the data clipped

Time increment: 16 ns

Phase: -14.1 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 5.58

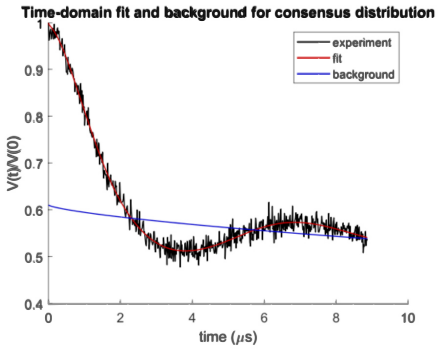
Reg. par. initial estimate by L-curve corner: 63.10

Overlap between DEERNet and regularization solutions: 0.910

Predicted overlap of consensus solution with ground truth: 0.78...0.95

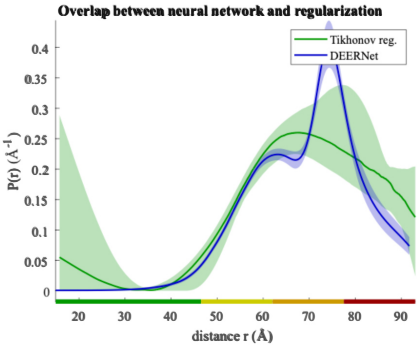
Mean distance: 68.5 Å

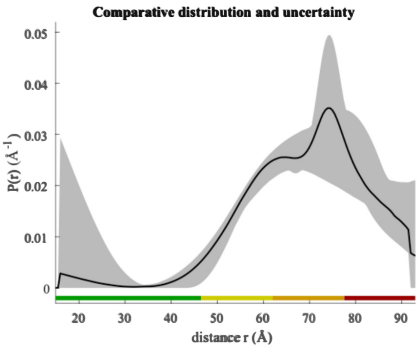
Distance standard deviation: 5.7 Å



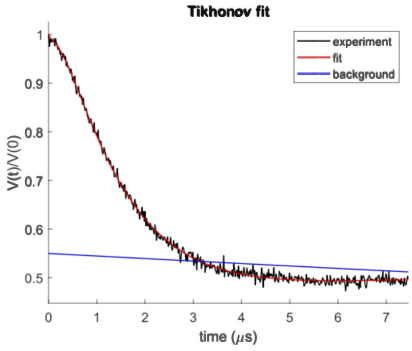
190/756$^{apo}_{active}$ with AA

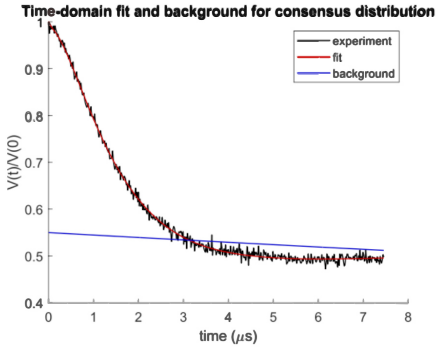
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.450

Signal-to-noise ratio: 57.7 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00774

From Tikhonov fit: 0.00773

Zero time: 94 ns

Maximum time: 7456 ns

Initial 2 point(s) of the data clipped

Time increment: 16 ns

Phase: -11.4 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 112.07

Reg. par. initial estimate by lr: 158.49

Overlap between DEERNet and regularization solutions: 0.879

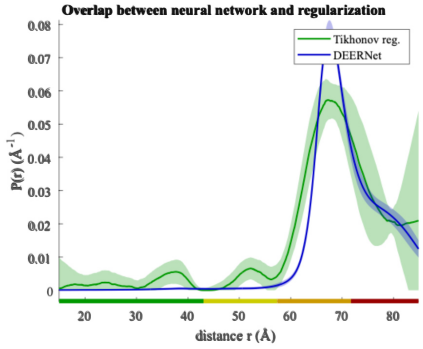
Predicted overlap of consensus solution with ground truth: 0.75...0.92

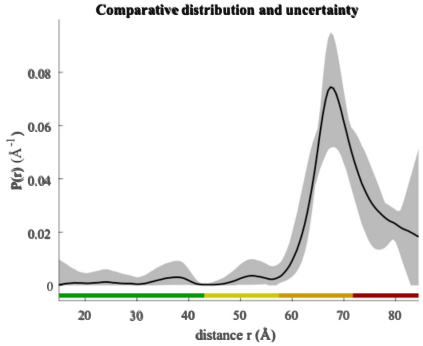
Mean distance: 71.0 Å

Distance standard deviation: 13.9 Å

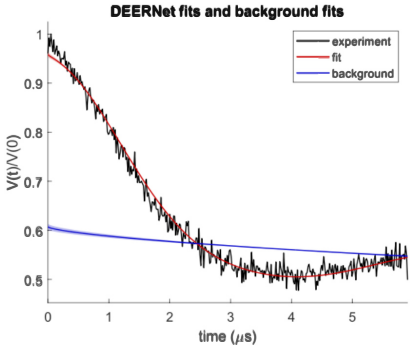
190/756$^{cr-target\ RNA}_{dtarget}$ with AA

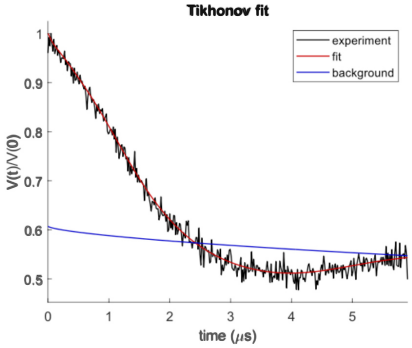
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.367

Signal-to-noise ratio: 22.6 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01340

From DEERNet fit: 0.01619

From Tikhonov fit: 0.01442

Zero time: 136 ns

Maximum time: 5904 ns

Initial 1 point(s) of the data clipped

Time increment: 16 ns

Phase: 0.2 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 7.89

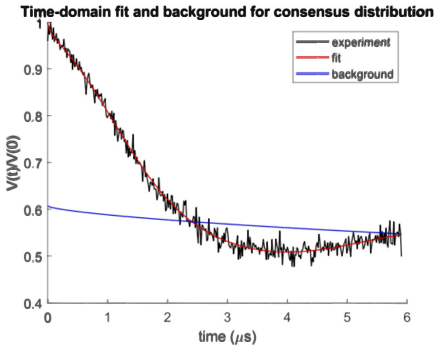
Reg. par. initial estimate by L-curve corner: 63.10

Overlap between DEERNet and regularization solutions: 0.822

Predicted overlap of consensus solution with ground truth: 0.70...0.87

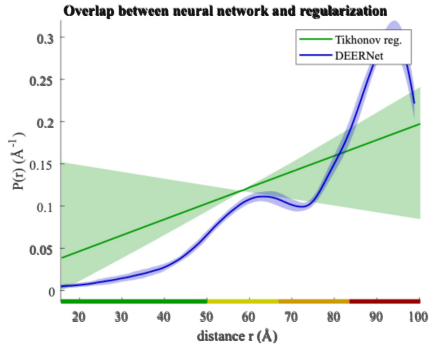
Mean distance: 70.0 Å

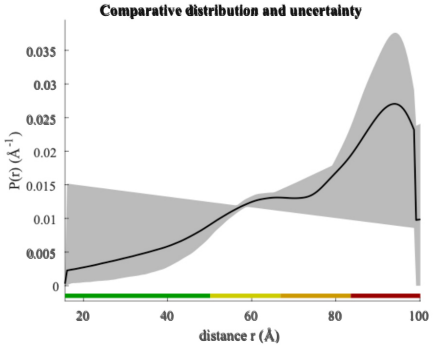
Distance standard deviation: 8.8 Å



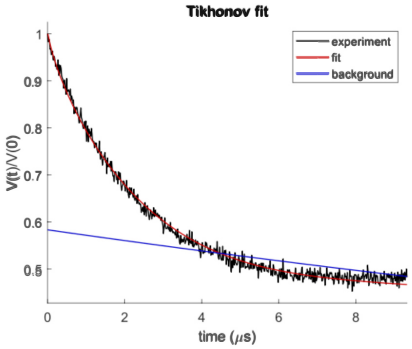
190/926$^{apo}_{active}$ with AA

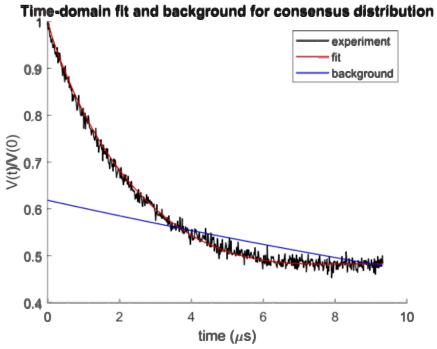
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.382

Signal-to-noise ratio: 29.0 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00923

From Tikhonov fit: 0.01203

Zero time: 180 ns

Maximum time: 9328 ns

Initial 3 point(s) of the data clipped

Time increment: 16 ns

Phase: -4.3 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 51321.57

Reg. par. initial estimate by lr: 50.12

Overlap between DEERNet and regularization solutions: 0.801

Predicted overlap of consensus solution with ground truth: 0.68...0.85

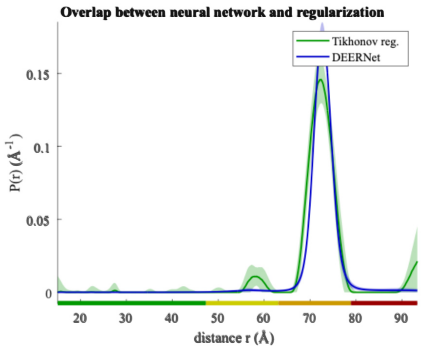
Mean distance: 72.2 Å

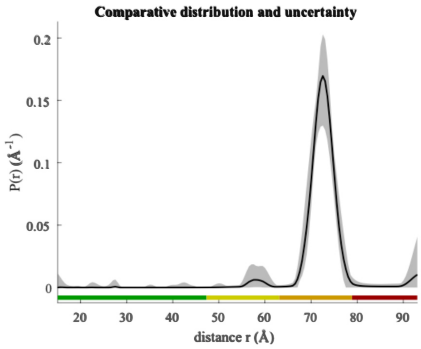
Single Gaussian provided different mean distance. Distribution may be incomplete.

Distance standard deviation: 20.8 Å

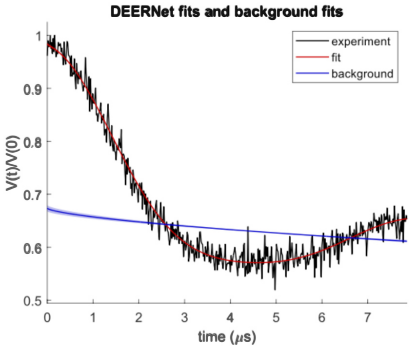
190/926 $^{\text{pre-tRNA}}_{\text{dprecr}}$ with AA

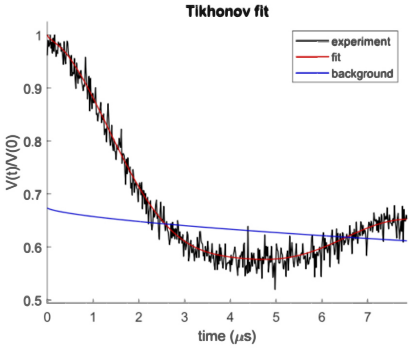
1. Distance distributions

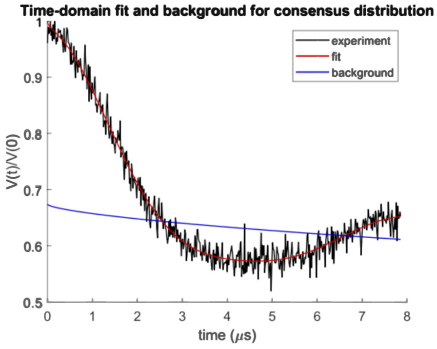




2. Fits of time-domain data







3. Experimental and processing parameters

Please consider improving signal-to-noise ratio (below 20)

Modulation depth: 0.315

Signal-to-noise ratio: 17.7 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01848

From DEERNet fit: 0.01785

From Tikhonov fit: 0.01764

Zero time: 156 ns

Maximum time: 7856 ns

Time increment: 16 ns

Phase: -17.8 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.88

Reg. par. initial estimate by L-curve corner: 39.81

Overlap between DEERNet and regularization solutions: 0.829

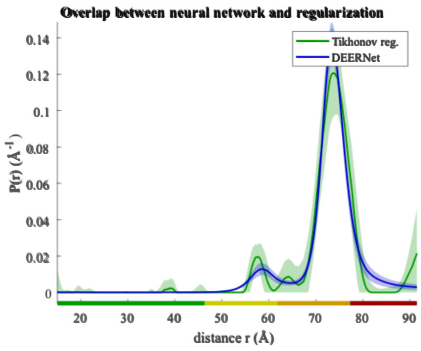
Predicted overlap of consensus solution with ground truth: 0.70...0.87

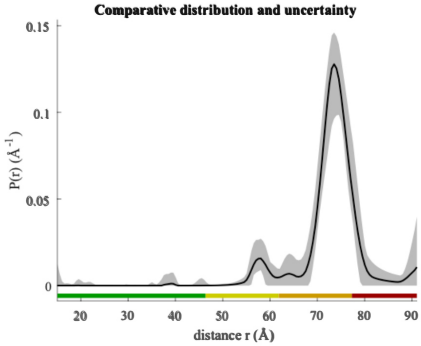
Mean distance: 72.4 Å

Distance standard deviation: 3.6 Å

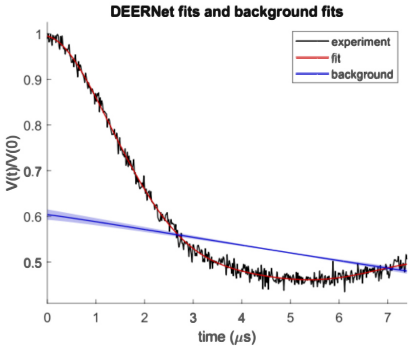
190/926$^{crRNA}_{active}$ with AA

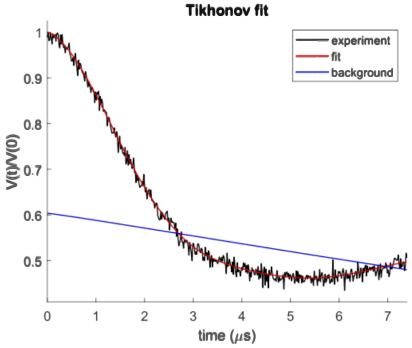
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.393

Signal-to-noise ratio: 36.1 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01096

From DEERNet fit: 0.01090

From Tikhonov fit: 0.01091

Zero time: 114 ns

Maximum time: 7392 ns

Initial 1 point(s) of the data clipped

Time increment: 16 ns

Phase: -16.8 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.39

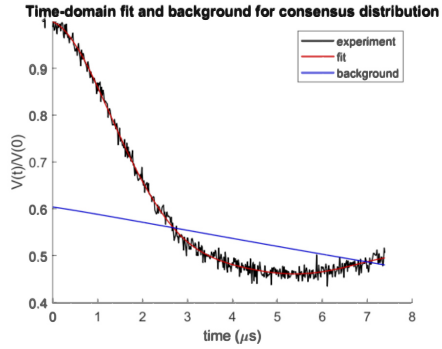
Reg. par. initial estimate by L-curve corner: 50.12

Overlap between DEERNet and regularization solutions: 0.874

Predicted overlap of consensus solution with ground truth: 0.74...0.92

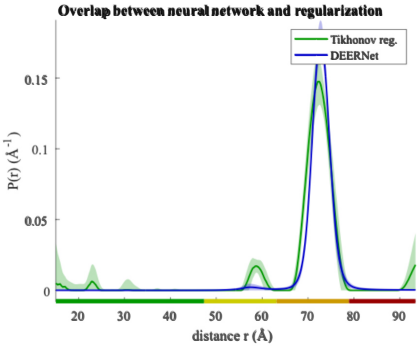
Mean distance: 73.5 Å

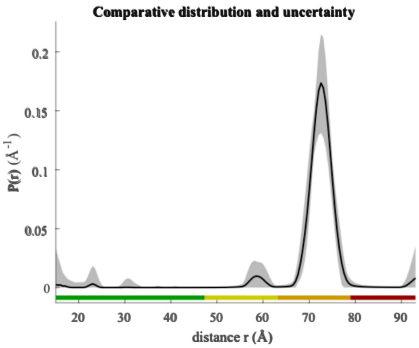
Distance standard deviation: 6.1 Å



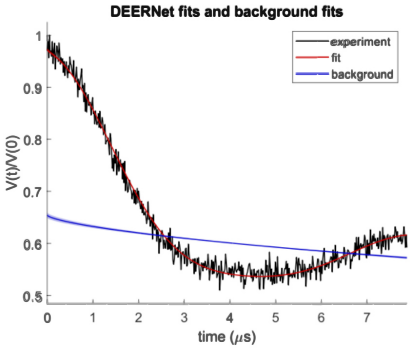
190/926$^{cr-target\ RNA}_{dtarget}$ with AA

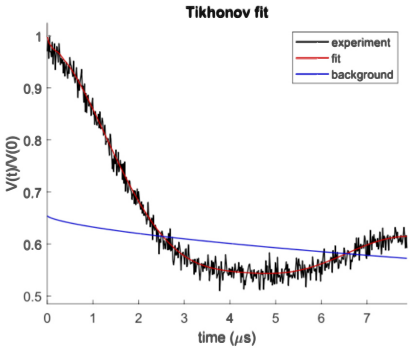
1. Distance distributions





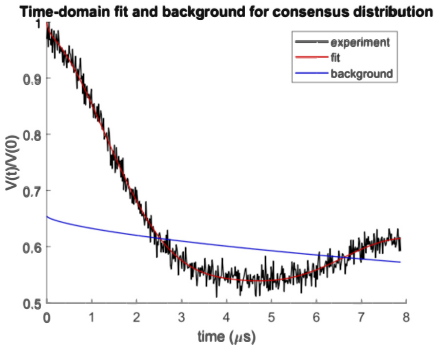
2. Fits of time-domain data





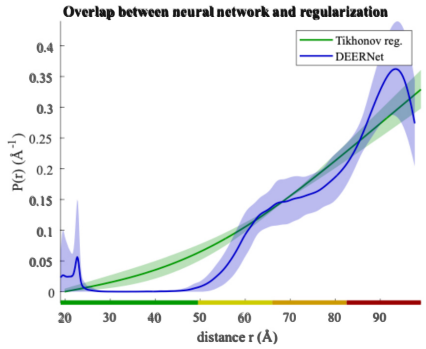
3. Experimental and processing parameters

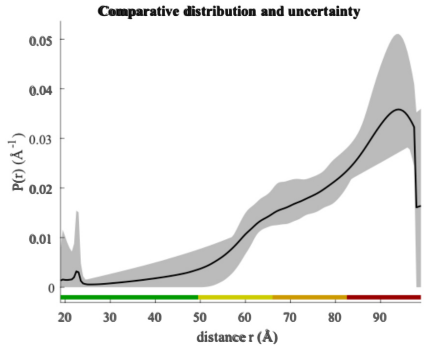
Modulation depth: 0.328
Signal-to-noise ratio: 22.1 (w.r.t. modulation)
Noise estimates normalized to maximum signal
From imaginary part: 0.01455
From DEERNet fit: 0.01485
From Tikhonov fit: 0.01435
Zero time: 167 ns
Maximum time: 7872 ns
Initial 2 point(s) of the data clipped
Time increment: 16 ns
Phase: 3.3 degree
Ensemble of 32 neural networks
Background separation by neural network
Background dimension: 3
Regularization parameter by best overlap with neural network solution
Regularization parameter used: 0.70
Reg. par. initial estimate by L-curve corner: 31.62
Overlap between DEERNet and regularization solutions: 0.816
Predicted overlap of consensus solution with ground truth: 0.69...0.86
Mean distance: 72.2 Å
Distance standard deviation: 3.7 Å



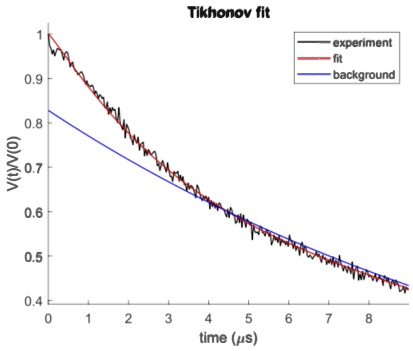
138/926$^{apo}_{active}$ with AA

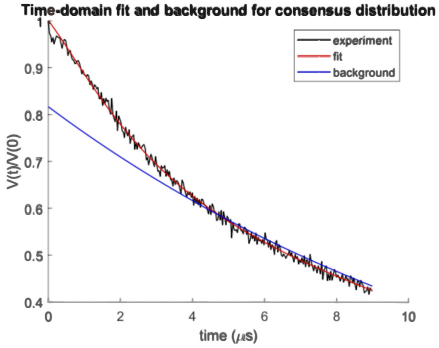
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Please consider improving signal-to-noise ratio (below 20)

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.183

Signal-to-noise ratio: 20.0 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00980

From Tikhonov fit: 0.01009

Zero time: 0 ns

Maximum time: 8992 ns

Time increment: 32 ns

Phase: -6.2 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 718.94

Reg. par. initial estimate by lr: 7.94

Overlap between DEERNet and regularization solutions: 0.872

Predicted overlap of consensus solution with ground truth: 0.74...0.91

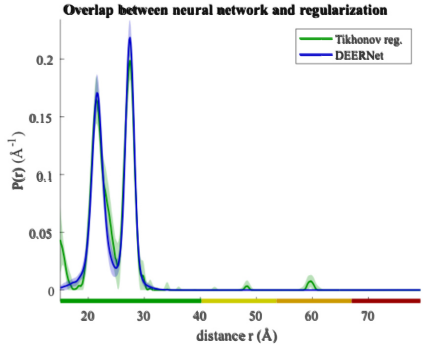
Mean distance: 78.8 Å

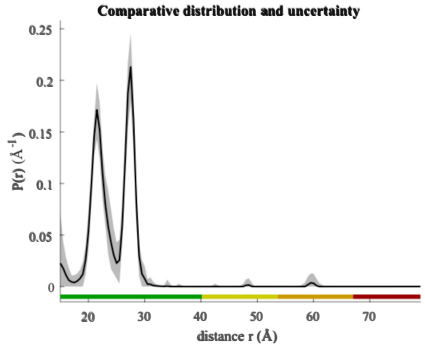
Single Gaussian provided different mean distance. Distribution may be incomplete.

Distance standard deviation: 15.6 Å

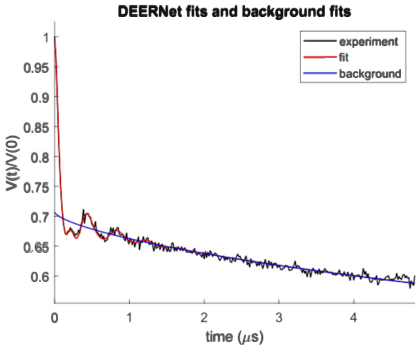
462/756 apo active

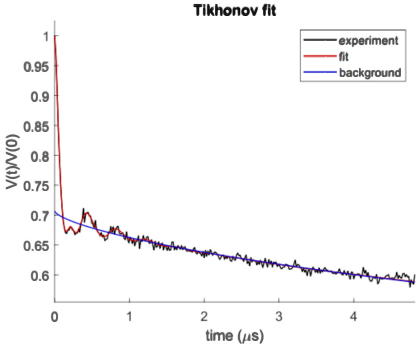
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.288

Signal-to-noise ratio: 57.4 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00481

From DEERNet fit: 0.00502

From Tikhonov fit: 0.00484

Zero time: 171 ns

Maximum time: 4816 ns

The last 1 % of the data was cut off

Time increment: 16 ns

Phase: -2.1 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.07

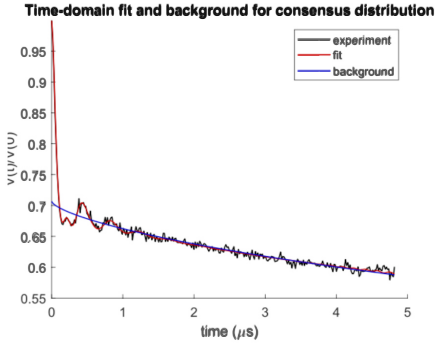
Reg. par. initial estimate by L-curve corner: 6.31

Overlap between DEERNet and regularization solutions: 0.898

Predicted overlap of consensus solution with ground truth: 0.77...0.94

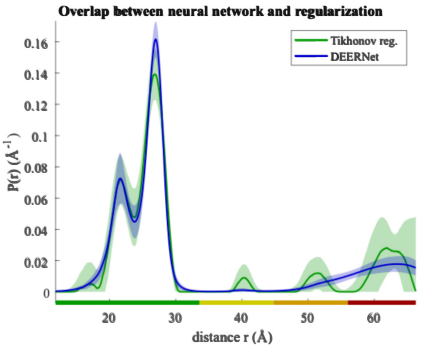
Mean distance: 23.8 Å

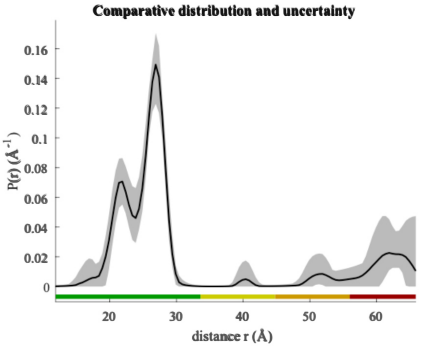
Distance standard deviation: 3.8 Å



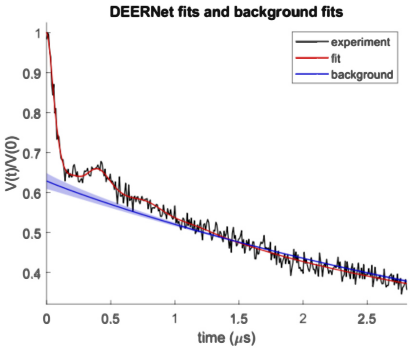
462/756 $^{crRNA}_{active}$

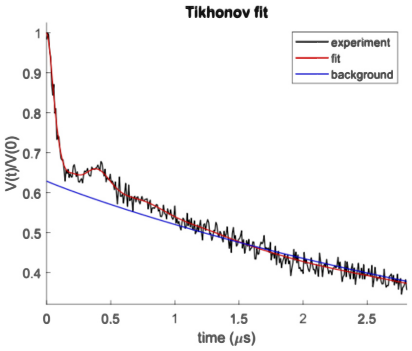
1. Distance distributions

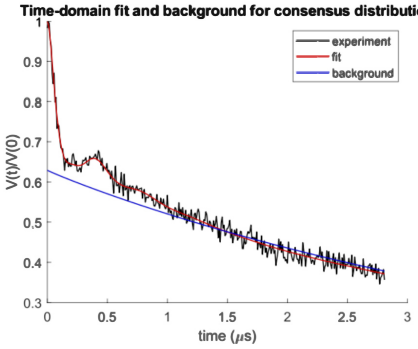




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.372

Signal-to-noise ratio: 25.1 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01483

From DEERNet fit: 0.01479

From Tikhonov fit: 0.01474

Zero time: 146 ns

Maximum time: 2.808000e+03 ns

The last 2 % of the data was cut off

Time increment: 8 ns

Phase: -8.0 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.99

Reg. par. initial estimate by L-curve corner: 15.85

Overlap between DEERNet and regularization solutions: 0.882

Predicted overlap of consensus solution with ground truth: 0.75...0.92

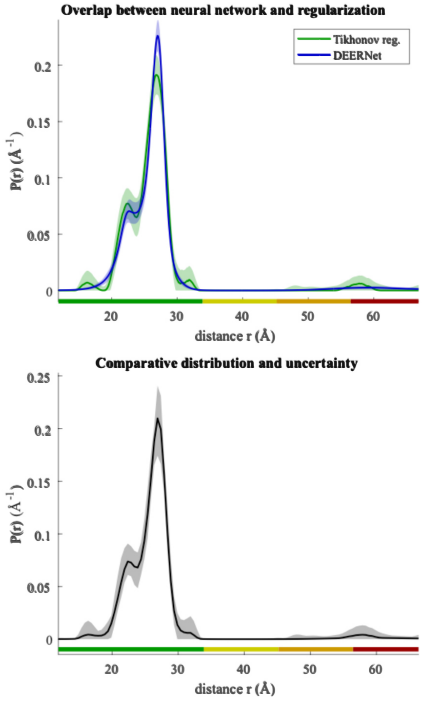
Mean distance: 32.0 Å

Single Gaussian provided different mean distance. Distribution may be incomplete.

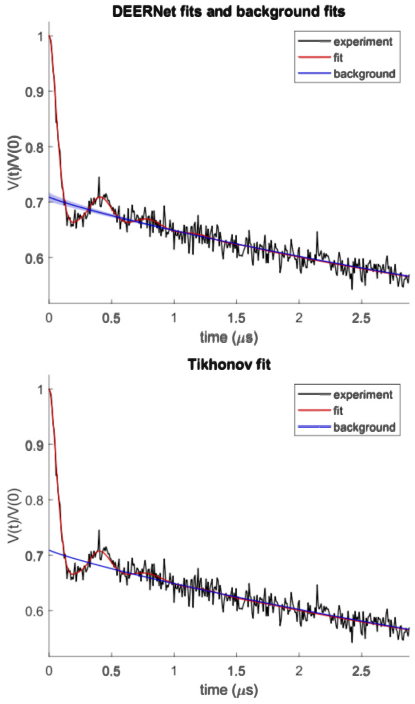
Distance standard deviation: 14.3 Å

precipitant of 462/756$^{\text{crRNA}}_{\text{active}}$

1. Distance distributions



2. Fits of time-domain data



3. Experimental and processing parameters

Modulation depth: 0.291

Signal-to-noise ratio: 21.1 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01289

From DEERNet fit: 0.01378

From Tikhonov fit: 0.01377

Zero time: 175 ns

Maximum time: 2880 ns

Time increment: 8 ns

Phase: 33.6 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.62

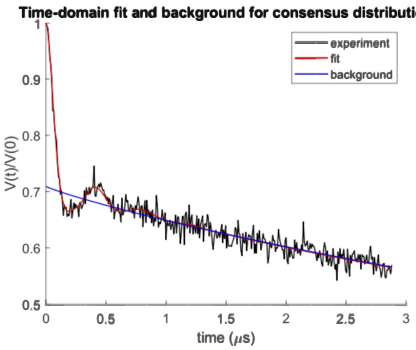
Reg. par. initial estimate by L-curve corner: 10.00

Overlap between DEERNet and regularization solutions: 0.911

Predicted overlap of consensus solution with ground truth: 0.78...0.95

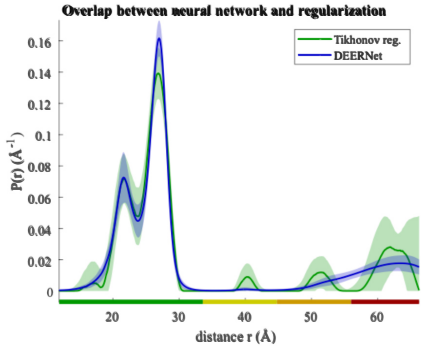
Mean distance: 25.9 Å

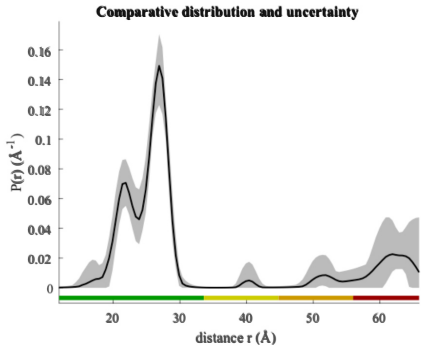
Distance standard deviation: 2.4 Å



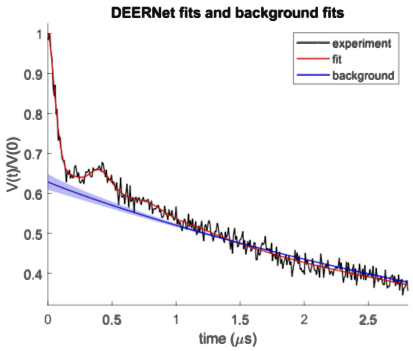
supernatant of 462/756 $^{crRNA}_{active}$

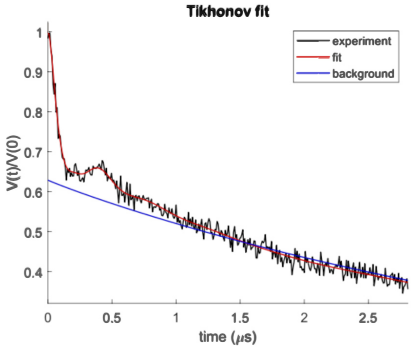
1. Distance distributions

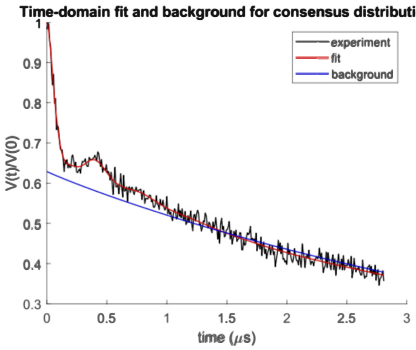




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.372

Signal-to-noise ratio: 25.1 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01483

From DEERNet fit: 0.01479

From Tikhonov fit: 0.01474

Zero time: 146 ns

Maximum time: 2.808000e+03 ns

The last 2 % of the data was cut off

Time increment: 8 ns

Phase: -8.0 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.99

Reg. par. initial estimate by L-curve corner: 15.85

Overlap between DEERNet and regularization solutions: 0.882

Predicted overlap of consensus solution with ground truth: 0.75...0.92

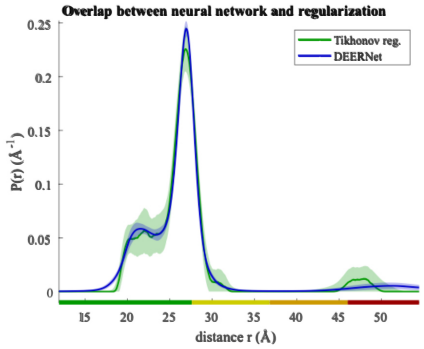
Mean distance: 32.0 Å

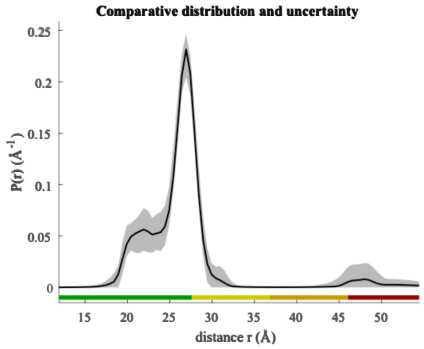
Single Gaussian provided different mean distance. Distribution may be incomplete.

Distance standard deviation: 14.3 Å

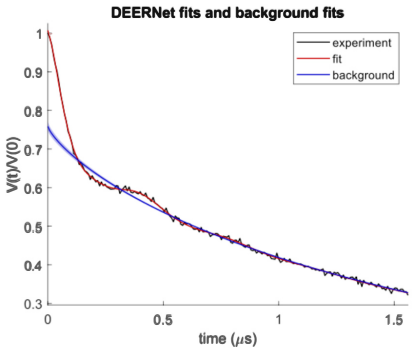
$462/756^{apo}_{active}$ with tRNA

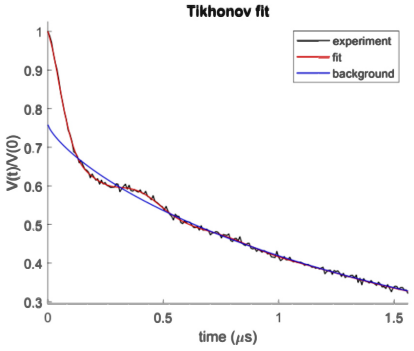
1. Distance distributions

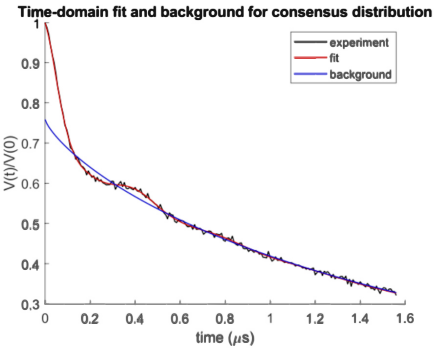




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.247

Signal-to-noise ratio: 50.7 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00691

From DEERNet fit: 0.00487

From Tikhonov fit: 0.00487

Zero time: 173 ns

Maximum time: 1560 ns

The last 1 % of the data was cut off

Time increment: 8 ns

Phase: -79.8 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.18

Reg. par. initial estimate by L-curve corner: 3.98

Overlap between DEERNet and regularization solutions: 0.927

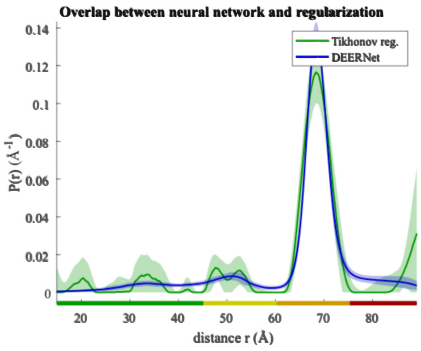
Predicted overlap of consensus solution with ground truth: 0.79...0.97

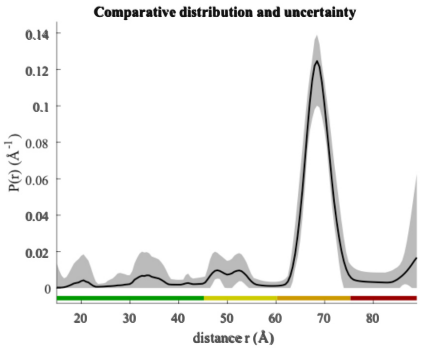
Mean distance: 26.0 Å

Distance standard deviation: 2.5 Å

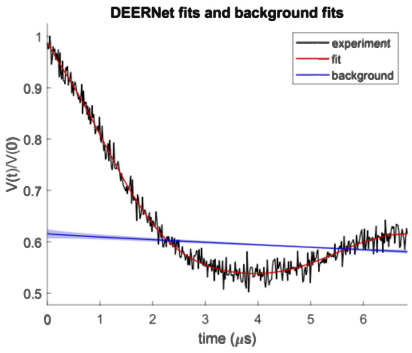
190/756 crRNA active

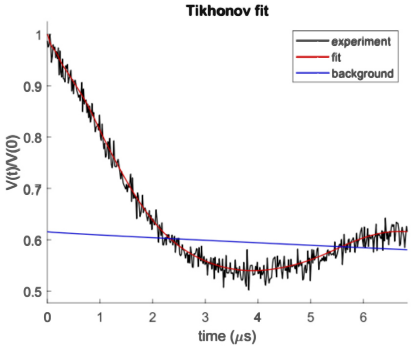
1. Distance distributions

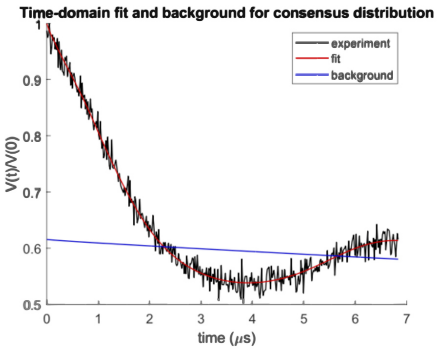




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.375

Signal-to-noise ratio: 26.3 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01551

From DEERNet fit: 0.01428

From Tikhonov fit: 0.01444

Zero time: 212 ns

Maximum time: 6832 ns

Initial 5 point(s) of the data clipped

Time increment: 16 ns

Phase: -23.6 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.88

Reg. par. initial estimate by L-curve corner: 39.81

Overlap between DEERNet and regularization solutions: 0.820

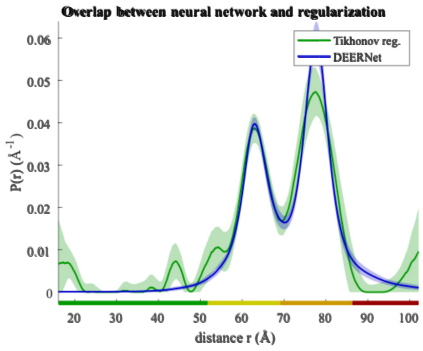
Predicted overlap of consensus solution with ground truth: 0.69...0.87

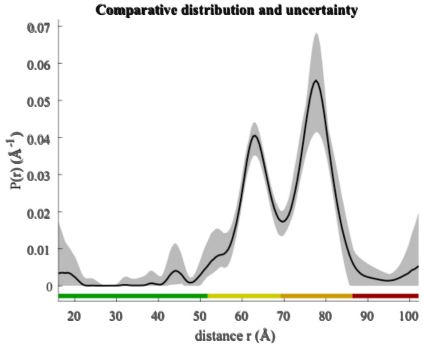
Mean distance: 68.0 Å

Distance standard deviation: 5.9 Å

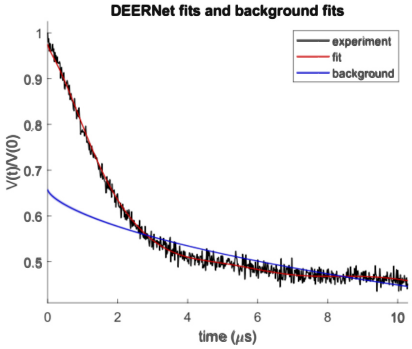
190/756 apo active

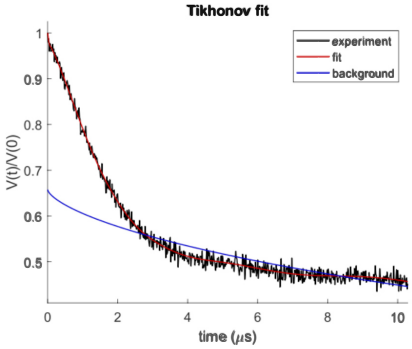
1. Distance distributions

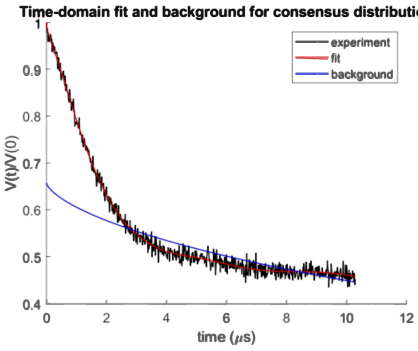




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.326

Signal-to-noise ratio: 28.6 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01036

From DEERNet fit: 0.01140

From Tikhonov fit: 0.01092

Zero time: 253 ns

Maximum time: 10288 ns

Initial 10 point(s) of the data clipped

Time increment: 16 ns

Phase: 6.8 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 1.97

Reg. par. initial estimate by L-curve corner: 125.89

Overlap between DEERNet and regularization solutions: 0.878

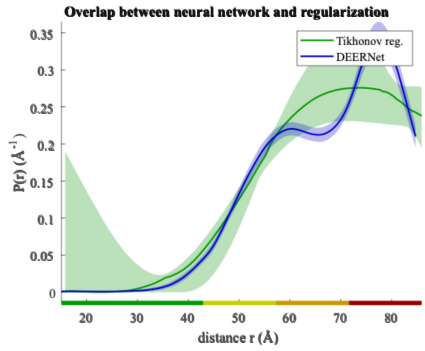
Predicted overlap of consensus solution with ground truth: 0.75...0.92

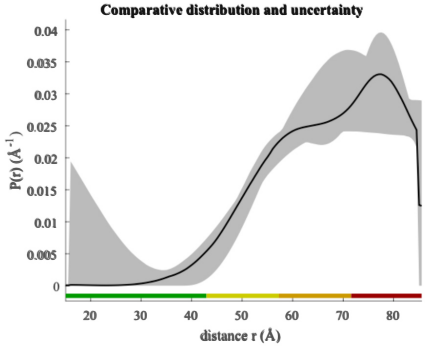
Mean distance: 71.0 Å

Distance standard deviation: 11.0 Å

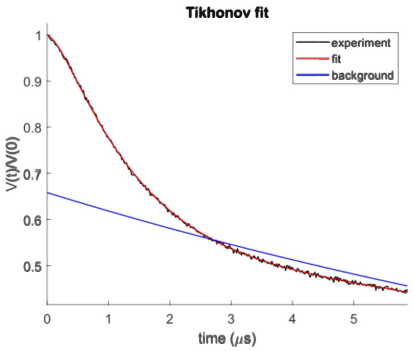
190/756$^{apo}_{active}$ with trehalose

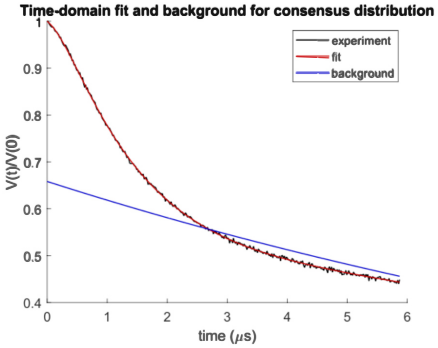
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.342

Signal-to-noise ratio: 121.2 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00286

From Tikhonov fit: 0.00596

Zero time: 141 ns

Maximum time: 5872 ns

Initial 5 point(s) of the data clipped

Time increment: 16 ns

Phase: -1.4 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 112.33

Reg. par. initial estimate by lr: 79.43

Overlap between DEERNet and regularization solutions: 0.932

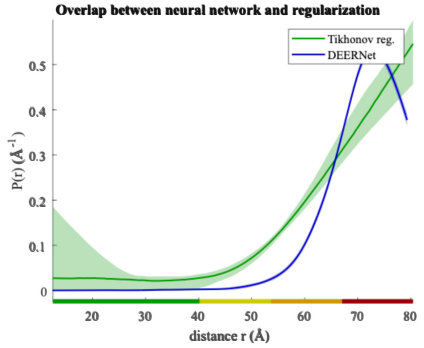
Predicted overlap of consensus solution with ground truth: 0.80...0.97

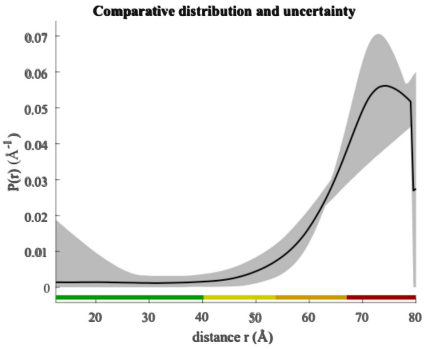
Mean distance: 72.8 Å

Distance standard deviation: 16.7 Å

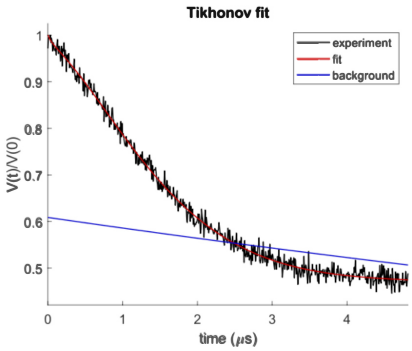
190/756$^{crRNA}_{active}$ with trehalose

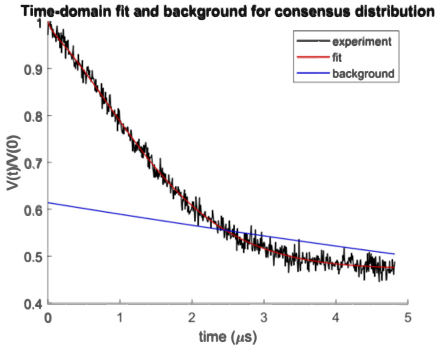
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.386

Signal-to-noise ratio: 27.6 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01331

From Tikhonov fit: 0.01303

Zero time: 233 ns

Maximum time: 4816 ns

Initial 5 point(s) of the data clipped

Time increment: 8 ns

Phase: 6.4 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 178.04

Reg. par. initial estimate by lr: 125.89

Overlap between DEERNet and regularization solutions: 0.783

Predicted overlap of consensus solution with ground truth: 0.66...0.83

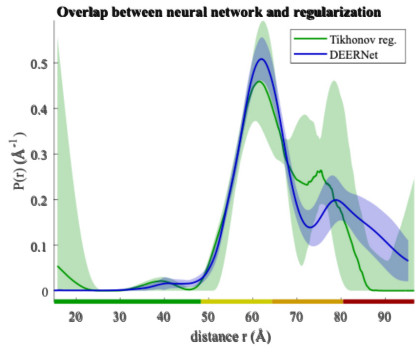
Mean distance: 67.8 Å

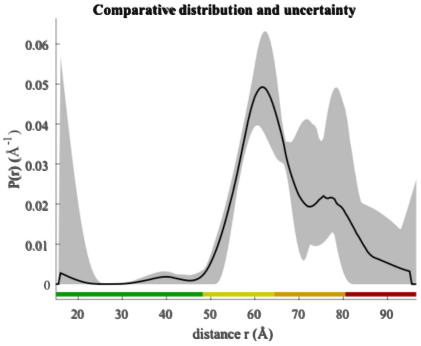
Single Gaussian provided different mean distance. Distribution may be incomplete.

Distance standard deviation: 11.3 Å

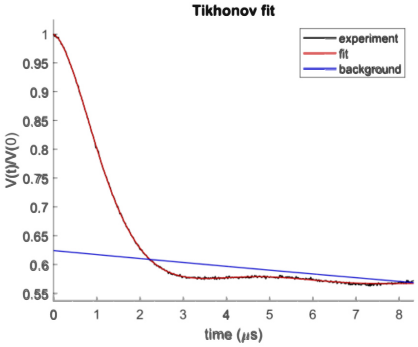
660/926 apo active

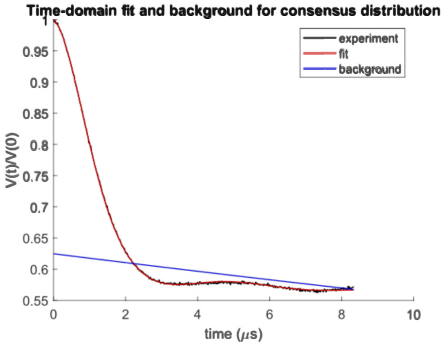
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.375

Signal-to-noise ratio: 224.3 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00171

From Tikhonov fit: 0.00220

Zero time: 166 ns

Maximum time: 8336 ns

The last 6 % of the data was cut off

Initial 8 point(s) of the data clipped

Time increment: 16 ns

Phase: -1.6 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 8.90

Reg. par. initial estimate by lr: 12.59

Overlap between DEERNet and regularization solutions: 0.853

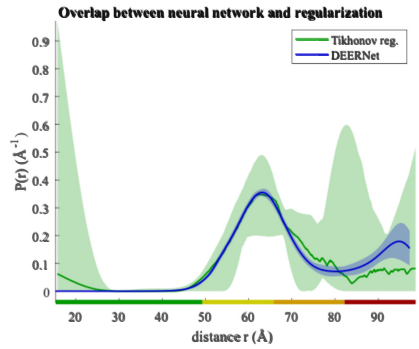
Predicted overlap of consensus solution with ground truth: 0.72...0.90

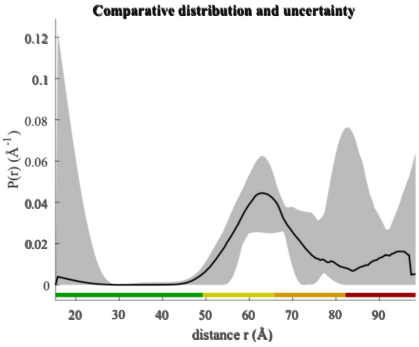
Mean distance: 66.2 Å

Distance standard deviation: 10.9 Å

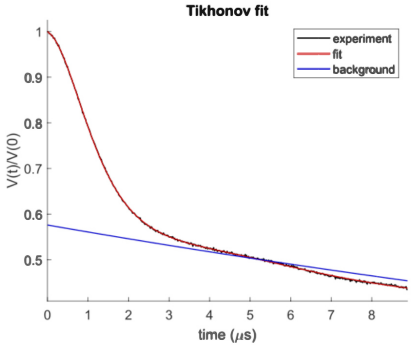
660/926$^{apo}_{dtarget}$

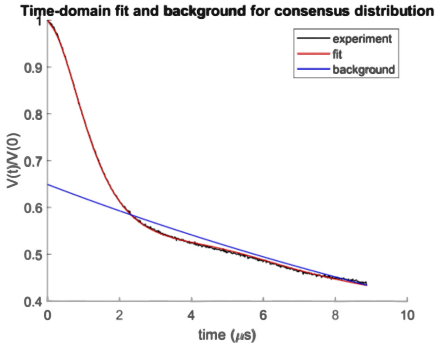
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.351

Signal-to-noise ratio: 201.1 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00152

From Tikhonov fit: 0.00374

Zero time: 142 ns

Maximum time: 8880 ns

Initial 5 point(s) of the data clipped

Time increment: 16 ns

Phase: 0.3 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 28.22

Reg. par. initial estimate by lr: 19.95

Overlap between DEERNet and regularization solutions: 0.881

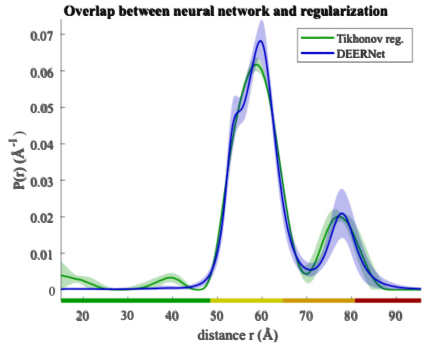
Predicted overlap of consensus solution with ground truth: 0.75...0.92

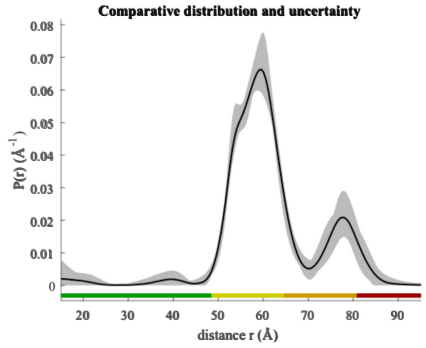
Mean distance: 68.0 Å

Distance standard deviation: 13.3 Å

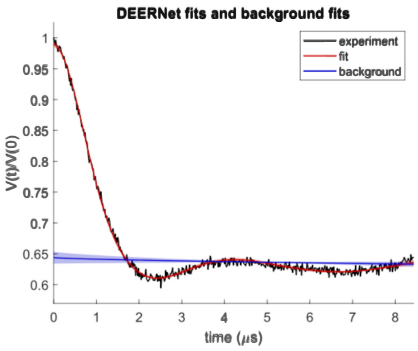
660/926 crRNA dtarget

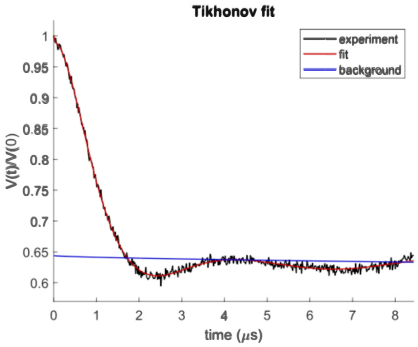
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

Modulation depth: 0.349

Signal-to-noise ratio: 69.3 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00455

From DEERNet fit: 0.00504

From Tikhonov fit: 0.00498

Zero time: 158 ns

Maximum time: 8432 ns

The last 5 % of the data was cut off

Initial 8 point(s) of the data clipped

Time increment: 16 ns

Phase: -6.1 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 7.05

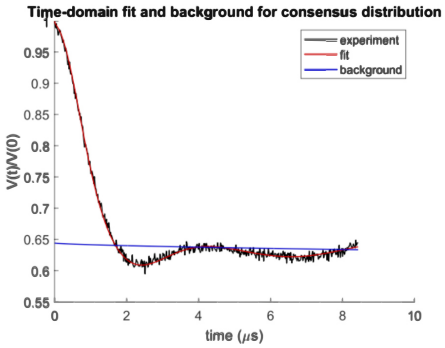
Reg. par. initial estimate by L-curve corner: 19.95

Overlap between DEERNet and regularization solutions: 0.925

Predicted overlap of consensus solution with ground truth: 0.79...0.96

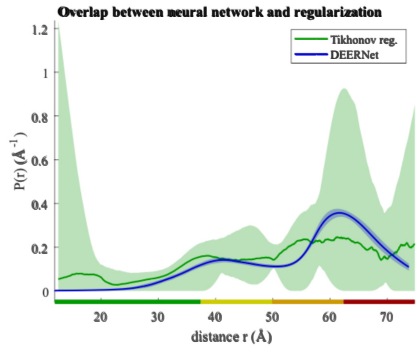
Mean distance: 60.9 Å

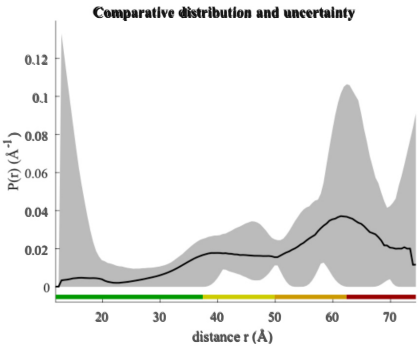
Distance standard deviation: 9.9 Å



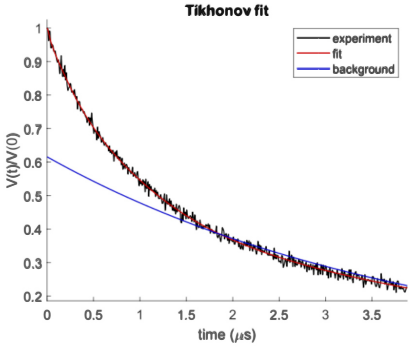
660/926$^{crRNA}_{dtarget}$ with DEG

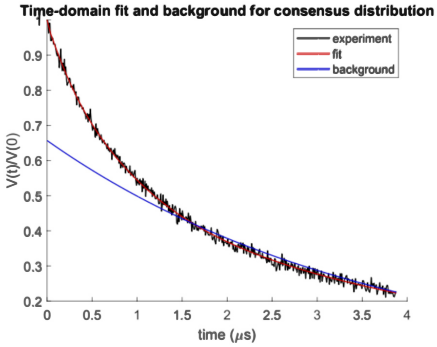
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.343

Signal-to-noise ratio: 31.2 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.01120

From Tikhonov fit: 0.01188

Zero time: 132 ns

Maximum time: 3880 ns

Time increment: 8 ns

Phase: -14.5 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 11.23

Reg. par. initial estimate by lr: 7.94

Overlap between DEERNet and regularization solutions: 0.818

Predicted overlap of consensus solution with ground truth: 0.69...0.86

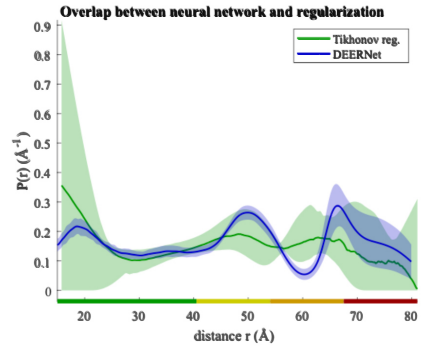
Mean distance: 53.2 Å

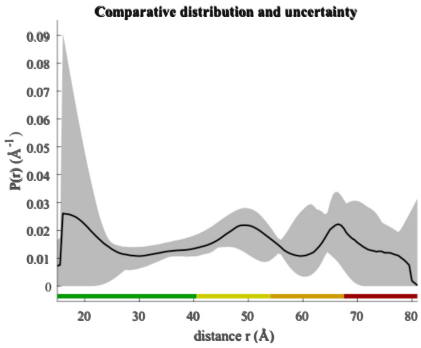
Single Gaussian provided different mean distance. Distribution may be incomplete.

Distance standard deviation: 14.1 Å

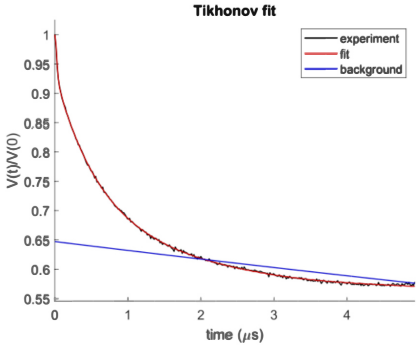
138/222 apo active

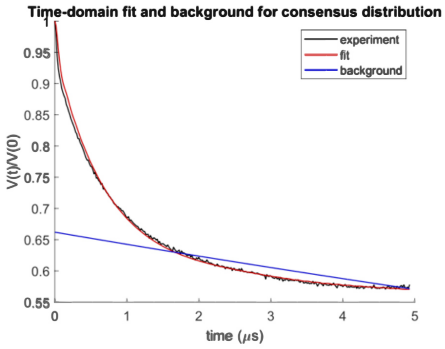
1. Distance distributions





2. Fits of time-domain data





3. Experimental and processing parameters

DEERNet background not provided, as it was considered unreliable.

Modulation depth: 0.338

Signal-to-noise ratio: 201.5 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00178

From Tikhonov fit: 0.00504

Zero time: 142 ns

Maximum time: 4928 ns

The last 8 % of the data was cut off

Initial 1 point(s) of the data clipped

Time increment: 16 ns

Phase: -4.8 degree

Ensemble of 32 neural networks

Background separation by DeerLab bilevel optimization

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 8.94

Reg. par. initial estimate by lr: 3.16

Overlap between DEERNet and regularization solutions: 0.851

Predicted overlap of consensus solution with ground truth: 0.72...0.89

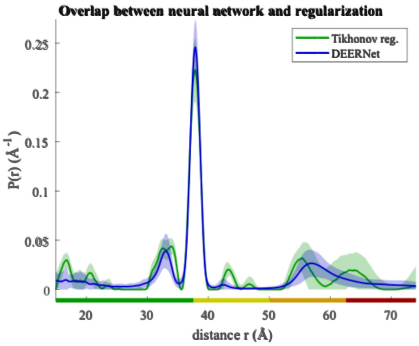
Mean distance: 46.2 Å

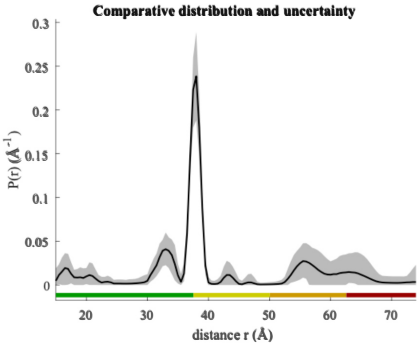
Single Gaussian provided different mean distance. Distribution may be incomplete.

Distance standard deviation: 18.7 Å

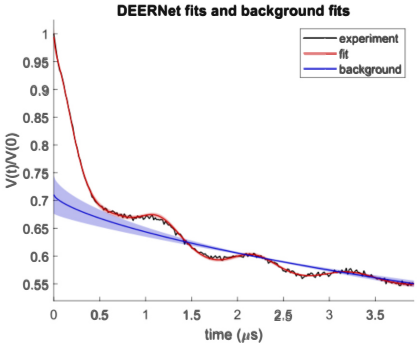
138/222crRNA active

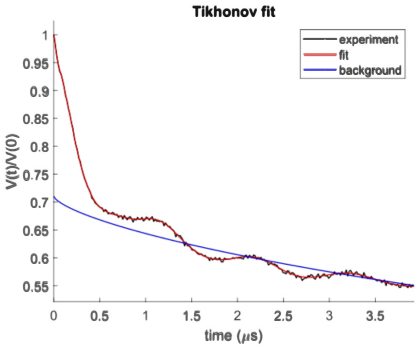
1. Distance distributions

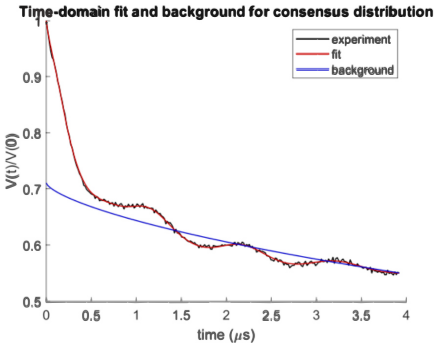




2. Fits of time-domain data







3. Experimental and processing parameters

Modulation depth: 0.285

Signal-to-noise ratio: 91.3 (w.r.t. modulation)

Noise estimates normalized to maximum signal

From imaginary part: 0.00283

From DEERNet fit: 0.00313

From Tikhonov fit: 0.00282

Zero time: 141 ns

Maximum time: 3920 ns

The last 10 % of the data was cut off

Initial 3 point(s) of the data clipped

Time increment: 16 ns

Phase: -19.1 degree

Ensemble of 32 neural networks

Background separation by neural network

Background dimension: 3

Regularization parameter by best overlap with neural network solution

Regularization parameter used: 0.06

Reg. par. initial estimate by L-curve corner: 0.50

Overlap between DEERNet and regularization solutions: 0.808

Predicted overlap of consensus solution with ground truth: 0.68...0.85

Mean distance: 37.6 Å

Distance standard deviation: 16.1 Å

References

1. Comeau, A. M. *et al.* Exploring the prokaryotic virosphere. *Research in microbiology* **159,** 306–313; 10.1016/j.resmic.2008.05.001 (2008).
2. Chanishvili, N. Phage therapy--history from Twort and d'Herelle through Soviet experience to current approaches. *Advances in virus research* **83,** 3–40; 10.1016/b978-0-12-394438-2.00001-3 (2012).
3. Keen, E. C. A century of phage research: bacteriophages and the shaping of modern biology. *BioEssays : news and reviews in molecular, cellular and developmental biology* **37,** 6–9; 10.1002/bies.201400152 (2015).
4. World Health Organisation. Fact sheets. Available at https://www.who.int/news-room/fact-sheets (2024).
5. European Center For Disease Prevention and Control. Zika virus disease - Annual Epidemological Report for 2021. Available at https://www.ecdc.europa.eu/sites/default/files/documents/zika-virus-disease-annual-epidemiological-report-2021.pdf (2024).
6. Centers for Disease Control and Prevention. History of Ebola Outbreaks. Available at https://www.cdc.gov/vhf/ebola/history/chronology.html (2024).
7. United Nations. Population. Accessed, 04.04.2024
8. NIH. National Library of Medicine- National Center for Biotechnology. Available at https://www.ncbi.nlm.nih.gov/ (2024).
9. Eugene V. Koonin, Kira S. Makarova & and Yuri I. Wolf. Evolutionary Genomics of Defense Systems in Archaea and Bacteria.
10. Dy, R. L., Richter, C., Salmond, G. P. C. & Fineran, P. C. Remarkable Mechanisms in Microbes to Resist Phage Infections. *Annual review of virology* **1,** 307–331; 10.1146/annurev-virology-031413-085500 (2014).
11. Wein, T. & Sorek, R. Bacterial origins of human cell-autonomous innate immune mechanisms. *Nature reviews. Immunology* **22,** 629–638; 10.1038/s41577-022-00705-4 (2022).

12. Bernheim, A. & Sorek, R. The pan-immune system of bacteria: antiviral defence as a community resource. *Nature reviews. Microbiology* **18,** 113–119; 10.1038/s41579-019-0278-2 (2020).

13. van Houte, S., Buckling, A. & Westra, E. R. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiology and molecular biology reviews : MMBR* **80,** 745–763; 10.1128/MMBR.00011-16 (2016).

14. Bernheim, A. *et al.* Prokaryotic viperins produce diverse antiviral molecules. *Nature* **589,** 120–124; 10.1038/s41586-020-2762-2 (2021).

15. Vivier, E. & Malissen, B. Innate and adaptive immunity: specificities and signaling hierarchies revisited. *Nature immunology* **6,** 17–21; 10.1038/ni1153 (2005).

16. Stern, A. & Sorek, R. The phage-host arms race: shaping the evolution of microbes. *BioEssays : news and reviews in molecular, cellular and developmental biology* **33,** 43–51; 10.1002/bies.201000071 (2011).

17. Makarova, K. S., Wolf, Y. I., Snir, S. & Koonin, E. V. Defense islands in bacterial and archaeal genomes and prediction of novel defense systems. *Journal of bacteriology* **193,** 6039–6056; 10.1128/JB.05535-11 (2011).

18. Doron, S. *et al.* Systematic discovery of antiphage defense systems in the microbial pangenome. *Science (New York, N.Y.)* **359**; 10.1126/science.aar4120 (2018).

19. Millman, A. *et al.* An expanded arsenal of immune systems that protect bacteria from phages. *Cell host & microbe* **30,** 1556-1569.e5; 10.1016/j.chom.2022.09.017 (2022).

20. Labrie, S. J., Samson, J. E. & Moineau, S. Bacteriophage resistance mechanisms. *Nature reviews. Microbiology* **8,** 317–327; 10.1038/nrmicro2315 (2010).

21. Foster, T. J. Immune evasion by staphylococci. *Nature reviews. Microbiology* **3,** 948–958; 10.1038/nrmicro1289 (2005).

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 beta-hairpin region in the recognition mechanism. *The Biochemical journal* **389,** 869–876; 10.1042/BJ20042107 (2005).

23. Folimonova, S. Y. Superinfection exclusion is an active virus-controlled function that requires a specific viral protein. *Journal of virology* **86,** 5554–5561; 10.1128/JVI.00310-12 (2012).

24. Lu, M. J. & Henning, U. Superinfection exclusion by T-even-type coliphages. *Trends in microbiology* **2,** 137–139; 10.1016/0966-842X(94)90601-7 (1994).

25. Parma, D. H. *et al.* The Rex system of bacteriophage lambda: tolerance and altruistic cell death. *Genes & development* **6,** 497–510; 10.1101/gad.6.3.497 (1992).

26. Snyder, L. Phage-exclusion enzymes: a bonanza of biochemical and cell biology reagents? *Molecular microbiology* **15,** 415–420; 10.1111/j.1365-2958.1995.tb02255.x (1995).

27. Lopatina, A., Tal, N. & Sorek, R. Abortive Infection: Bacterial Suicide as an Antiviral Immune Strategy. *Annual review of virology* **7,** 371–384; 10.1146/annurev-virology-011620-040628 (2020).

28. Cohen, D. *et al.* Cyclic GMP-AMP signalling protects bacteria against viral infection. *Nature* **574,** 691–695; 10.1038/s41586-019-1605-5 (2019).

29. Cook, G. M. *et al.* Ribonucleases in bacterial toxin-antitoxin systems. *Biochimica et biophysica acta* **1829,** 523–531; 10.1016/j.bbagrm.2013.02.007 (2013).

30. Meeske, A. J., Nakandakari-Higa, S. & Marraffini, L. A. Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. *Nature* **570,** 241–245; 10.1038/s41586-019-1257-5 (2019).

31. Tock, M. R. & Dryden, D. T. F. The biology of restriction and anti-restriction. *Current opinion in microbiology* **8,** 466–472; 10.1016/j.mib.2005.06.003 (2005).

32. Eugene V. Koonin, Kira S. Makarova & and Yuri I. Wolf. Evolutionary Genomics of Defense Systems in Archaea and Bacteria.

33. Bobadilla Ugarte, P., Barendse, P. & Swarts, D. C. Argonaute proteins confer immunity in all domains of life. *Current opinion in microbiology* **74,** 102313; 10.1016/j.mib.2023.102313 (2023).

34. Swarts, D. C. *et al.* DNA-guided DNA interference by a prokaryotic Argonaute. *Nature* **507,** 258–261; 10.1038/nature12971 (2014).

35. Anderson, J. *Handbook of experimental pharmacology* (Springer, Berlin, 2009).

36. Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science (New York, N.Y.)* **315,** 1709–1712; 10.1126/science.1138140 (2007).

37. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nature reviews. Microbiology* **13,** 722–736; 10.1038/nrmicro3569 (2015).

38. Koonin, E. V., Gootenberg, J. S. & Abudayyeh, O. O. Discovery of Diverse CRISPR-Cas Systems and Expansion of the Genome Engineering Toolbox. *Biochemistry* **62,** 3465–3487; 10.1021/acs.biochem.3c00159 (2023).

39. Toro, N., Mestre, M. R., Martínez-Abarca, F. & González-Delgado, A. Recruitment of Reverse Transcriptase-Cas1 Fusion Proteins by Type VI-A CRISPR-Cas Systems. *Frontiers in microbiology* **10,** 2160; 10.3389/fmicb.2019.02160 (2019).

40. Gleditzsch, D. *et al.* PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. *RNA biology* **16,** 504–517; 10.1080/15476286.2018.1504546 (2019).

41. Marraffini, L. A. & Sontheimer, E. J. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463,** 568–571; 10.1038/nature08703 (2010).

42. Veigl, S. J. Adaptive immunity or evolutionary adaptation? Transgenerational immune systems at the crossroads. *Biol Philos* **37**; 10.1007/s10539-022-09869-8 (2022).

43. Grissa, I., Vergnaud, G. & Pourcel, C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC bioinformatics* **8,** 172; 10.1186/1471-2105-8-172 (2007).

44. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell* **60,** 385–397; 10.1016/j.molcel.2015.10.008 (2015).

45. Barrangou, R. & Horvath, P. A decade of discovery: CRISPR functions and applications. *Nature microbiology* **2,** 17092; 10.1038/nmicrobiol.2017.92 (2017).

46. Behler, J. & Hess, W. R. Approaches to study CRISPR RNA biogenesis and the key players involved. *Methods (San Diego, Calif.)* **172,** 12–26; 10.1016/j.ymeth.2019.07.015 (2020).

47. Alan R. Davidson *et al.* Anti-CRISPRs: Protein Inhibitors of CRISPR-Cas Systems.

48. Pawluk, A., Davidson, A. R. & Maxwell, K. L. Anti-CRISPR: discovery, mechanism and function. *Nature reviews. Microbiology* **16,** 12–17; 10.1038/nrmicro.2017.120 (2018).

49. de Dong *et al.* Structural basis of CRISPR-SpyCas9 inhibition by an anti-CRISPR protein. *Nature* **546,** 436–439; 10.1038/nature22377 (2017).

50. Knott, G. J. *et al.* Broad-spectrum enzymatic inhibition of CRISPR-Cas12a. *Nature structural & molecular biology* **26,** 315–321; 10.1038/s41594-019-0208-z (2019).

51. Harrington, L. B. *et al.* A Broad-Spectrum Inhibitor of CRISPR-Cas9. *Cell* **170,** 1224-1233.e15; 10.1016/j.cell.2017.07.037 (2017).

52. Koonin, E. V., Gootenberg, J. S. & Abudayyeh, O. O. Discovery of Diverse CRISPR-Cas Systems and Expansion of the Genome Engineering Toolbox. *Biochemistry* **62,** 3465–3487; 10.1021/acs.biochem.3c00159 (2023).

53. Koonin, E. V. & Makarova, K. S. Origins and evolution of CRISPR-Cas systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **374,** 20180087; 10.1098/rstb.2018.0087 (2019).

54. Makarova, K. S. *et al.* Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nature reviews. Microbiology* **18,** 67–83; 10.1038/s41579-019-0299-x (2020).

55. Liu, T. Y. & Doudna, J. A. Chemistry of Class 1 CRISPR-Cas effectors: Binding, editing, and regulation. *The Journal of biological chemistry* **295,** 14473–14487; 10.1074/jbc.REV120.007034 (2020).

56. Niewoehner, O. *et al.* Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature* **548,** 543–548; 10.1038/nature23467 (2017).

57. Wang, J. Y., Pausch, P. & Doudna, J. A. Structural biology of CRISPR-Cas immunity and genome editing enzymes. *Nature reviews. Microbiology* **20,** 641–656; 10.1038/s41579-022-00739-4 (2022).

58. Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science (New York, N.Y.)* **379,** eadd8643; 10.1126/science.add8643 (2023).

59. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156,** 935–949; 10.1016/j.cell.2014.02.001 (2014).

60. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (New York, N.Y.)* **337,** 816–821; 10.1126/science.1225829 (2012).

61. Xiao, Y., Luo, M., Dolan, A. E., Liao, M. & Ke, A. Structure basis for RNA-guided DNA degradation by Cascade and Cas3. *Science (New York, N.Y.)* **361**; 10.1126/science.aat0839 (2018).

62. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513,** 569–573; 10.1038/nature13579 (2014).

63. Swarts, D. C., van der Oost, J. & Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Molecular cell* **66,** 221-233.e4; 10.1016/j.molcel.2017.03.016 (2017).

64. Liu, L. *et al.* The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170,** 714-726.e10; 10.1016/j.cell.2017.06.050 (2017).

65. Harrington, L. B. *et al.* Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science (New York, N.Y.)* **362,** 839–842; 10.1126/science.aav4294 (2018).

66. Yan, W. X. *et al.* Functionally diverse type V CRISPR-Cas systems. *Science (New York, N.Y.)* **363,** 88–91; 10.1126/science.aav7271 (2019).

67. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell* **60,** 385–397; 10.1016/j.molcel.2015.10.008 (2015).

68. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular cell* **60,** 385–397; 10.1016/j.molcel.2015.10.008 (2015).

69. Baltimore, D. Expression of animal virus genomes. *Bacteriological reviews* **35,** 235–241; 10.1128/br.35.3.235-241.1971 (1971).

70. Adler, B. A. *et al.* Broad-spectrum CRISPR-Cas13a enables efficient phage genome editing. *Nature microbiology* **7,** 1967–1979; 10.1038/s41564-022-01258-x (2022).

71. Xu, C. *et al.* Programmable RNA editing with compact CRISPR-Cas13 systems from uncultivated microbes. *Nature methods* **18,** 499–506; 10.1038/s41592-021-01124-4 (2021).

72. Behler, J. & Hess, W. R. Approaches to study CRISPR RNA biogenesis and the key players involved. *Methods (San Diego, Calif.)* **172,** 12–26; 10.1016/j.ymeth.2019.07.015 (2020).

73. Abudayyeh, O. O. *et al.* RNA targeting with CRISPR-Cas13. *Nature* **550,** 280–284; 10.1038/nature24049 (2017).

74. Leenay, R. T. & Beisel, C. L. Deciphering, Communicating, and Engineering the CRISPR PAM. *Journal of molecular biology* **429,** 177–191; 10.1016/j.jmb.2016.11.024 (2017).

75. East-Seletsky, A. *et al.* Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538,** 270–273; 10.1038/nature19802 (2016).

76. East-Seletsky, A., O'Connell, M. R., Burstein, D., Knott, G. J. & Doudna, J. A. RNA Targeting by Functionally Orthogonal Type VI-A CRISPR-Cas Enzymes. *Molecular cell* **66,** 373-383.e3; 10.1016/j.molcel.2017.04.008 (2017).

77. Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science (New York, N.Y.)* **360,** 439–444; 10.1126/science.aaq0179 (2018).

78. Jain, I. *et al. tRNA anticodon cleavage by target-activated CRISPR-Cas13a effector* (2021).

79. Wang, J. Y. & Doudna, J. A. CRISPR technology: A decade of genome editing is only the beginning. *Science (New York, N.Y.)* **379,** eadd8643; 10.1126/science.add8643 (2023).

80. Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science (New York, N.Y.)* **356,** 438–442; 10.1126/science.aam9321 (2017).

81. Patchsung, M. *et al.* Clinical validation of a Cas13-based assay for the detection of SARS-CoV-2 RNA. *Nature biomedical engineering* **4,** 1140–1149; 10.1038/s41551-020-00603-x (2020).

82. Barnes, K. G. *et al.* Deployable CRISPR-Cas13a diagnostic tools to detect and report Ebola and Lassa virus cases in real-time. *Nature communications* **11,** 4131; 10.1038/s41467-020-17994-9 (2020).

83. Zhang, J. *et al.* Highly Effective Detection of Exosomal miRNAs in Plasma Using Liposome-Mediated Transfection CRISPR/Cas13a. *ACS sensors* **8,** 565–575; 10.1021/acssensors.2c01683 (2023).

84. Gao, J., Luo, T., Lin, N., Zhang, S. & Wang, J. A New Tool for CRISPR-Cas13a-Based Cancer Gene Therapy. *Molecular therapy oncolytics* **19,** 79–92; 10.1016/j.omto.2020.09.004 (2020).

85. Wang, Q. *et al.* The CRISPR-Cas13a Gene-Editing System Induces Collateral Cleavage of RNA in Glioma Cells. *Advanced science (Weinheim, Baden-Wurttemberg, Germany)* **6,** 1901299; 10.1002/advs.201901299 (2019).

86. Blanchard, E. L. *et al.* Treatment of influenza and SARS-CoV-2 infections via mRNA-encoded Cas13a in rodents. *Nature biotechnology* **39,** 717–726; 10.1038/s41587-021-00822-w (2021).

87. Liu, L. *et al.* The Molecular Architecture for RNA-Guided RNA Cleavage by Cas13a. *Cell* **170,** 714-726.e10; 10.1016/j.cell.2017.06.050 (2017).

88. Wang, B. *et al.* Structural basis for self-cleavage prevention by tag:anti-tag pairing complementarity in type VI Cas13 CRISPR systems. *Molecular cell* **81,** 1100-1115.e5; 10.1016/j.molcel.2020.12.033 (2021).

89. Knott, G. J. *et al.* Guide-bound structures of an RNA-targeting A-cleaving CRISPR-Cas13a enzyme. *Nature structural & molecular biology* **24,** 825–833; 10.1038/nsmb.3466 (2017).

90. Kick, L. M., Wrisberg, M.-K. von, Runtsch, L. S. & Schneider, S. Structure and mechanism of the RNA dependent RNase Cas13a from Rhodobacter capsulatus. *Communications biology* **5,** 71; 10.1038/s42003-022-03025-4 (2022).

91. Meeske, A. J. *et al.* A phage-encoded anti-CRISPR enables complete evasion of type VI-A CRISPR-Cas immunity. *Science (New York, N.Y.)* **369,** 54–59; 10.1126/science.abb6151 (2020).

92. Wang, F. *et al.* Structural Basis for the Ribonuclease Activity of a Thermostable CRISPR-Cas13a from Thermoclostridium caenicola. *Journal of molecular biology* **435,** 168197; 10.1016/j.jmb.2023.168197 (2023).

93. Liu, L. *et al.* Two Distant Catalytic Sites Are Responsible for C2c2 RNase Activities. *Cell* **168,** 121-134.e12; 10.1016/j.cell.2016.12.031 (2017).

94. O'Connell, M. R. Molecular Mechanisms of RNA Targeting by Cas13-containing Type VI CRISPR-Cas Systems. *Journal of molecular biology* **431,** 66–87; 10.1016/j.jmb.2018.06.029 (2019).

95. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589; 10.1038/s41586-021-03819-2 (2021).

96. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*; 10.1038/s41586-024-07487-w (2024).

97. Kurakin, A. Scale-free flow of life: on the biology, economics, and physics of the cell. *Theoretical biology & medical modelling* **6,** 6; 10.1186/1742-4682-6-6 (2009).

98. Nam, K. & Wolf-Watz, M. Protein dynamics: The future is bright and complicated! *Structural dynamics (Melville, N.Y.)* **10,** 14301; 10.1063/4.0000179 (2023).

99. Berman, H. M. *et al.* Federating Structural Models and Data: Outcomes from A Workshop on Archiving Integrative Structures. *Structure (London, England : 1993)* **27,** 1745–1759; 10.1016/j.str.2019.11.002 (2019).

100. Schlundt, A., Tants, J.-N. & Sattler, M. Integrated structural biology to unravel molecular mechanisms of protein-RNA recognition. *Methods (San Diego, Calif.)* **118-119,** 119–136; 10.1016/j.ymeth.2017.03.015 (2017).

101. Rosenbach, H. *et al.* Expanding crystallization tools for nucleic acid complexes using U1A protein variants. *Journal of structural biology* **210,** 107480; 10.1016/j.jsb.2020.107480 (2020).

102. Lottspeich, F. & Engels Joachim W. *Bioanalytics : analytical methods and concepts in biochemistry and molecular biology /{u200B}* (ill, 1,110 p.,).

103. Taylor, G. L. Introduction to phasing. *Acta crystallographica. Section D, Biological crystallography* **66,** 325–338; 10.1107/S0907444910006694 (2010).

104. Callaway, E. The revolution will not be crystallized: a new method sweeps through structural biology. *Nature* **525,** 172–174; 10.1038/525172a (2015).

105. Kühlbrandt, W. Biochemistry. The resolution revolution. *Science (New York, N.Y.)* **343,** 1443–1444; 10.1126/science.1251652 (2014).

106. Brilot, A. F. *et al.* Beam-induced motion of vitrified specimen on holey carbon film. *Journal of structural biology* **177,** 630–637; 10.1016/j.jsb.2012.02.003 (2012).

107. Greber, B. J. *et al.* Ribosome. The complete structure of the 55S mammalian mitochondrial ribosome. *Science (New York, N.Y.)* **348,** 303–308; 10.1126/science.aaa3872 (2015).

108. He, Y. *et al.* Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533,** 359–365; 10.1038/nature17970 (2016).

109. Wu, M. & Lander, G. C. How low can we go? Structure determination of small biological complexes using single-particle cryo-EM. *Current opinion in structural biology* **64,** 9–16; 10.1016/j.sbi.2020.05.007. (2020).

110. Carroni, M. & Saibil, H. R. Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods (San Diego, Calif.)* **95,** 78–85; 10.1016/j.ymeth.2015.11.023 (2016).

111. Xue, L. *et al.* Visualizing translation dynamics at atomic detail inside a bacterial cell. *Nature* **610,** 205–211; 10.1038/s41586-022-05255-2 (2022).

112. Markwick, P. R. L., Malliavin, T. & Nilges, M. Structural biology by NMR: structure, dynamics, and interactions. *PLoS computational biology* **4,** e1000168; 10.1371/journal.pcbi.1000168 (2008).

113. Bax, A. Two-dimensional NMR and protein structure. *Annual review of biochemistry* **58,** 223–256; 10.1146/annurev.bi.58.070189.001255 (1989).

114. Chen, K. & Tjandra, N. The use of residual dipolar coupling in studying proteins by NMR. *Topics in current chemistry* **326,** 47–67; 10.1007/128_2011_215 (2012).

115. Hu, Y. *et al.* NMR-Based Methods for Protein Analysis. *Analytical chemistry* **93,** 1866–1879; 10.1021/acs.analchem.0c03830 (2021).

116. Dallmann, A. *et al.* Efficient detection of hydrogen bonds in dynamic regions of RNA by sensitivity-optimized NMR pulse sequences. *Angewandte Chemie (International ed. in English)* **52,** 10487–10490; 10.1002/anie.201304391 (2013).

117. Helmling, C. *et al.* Rapid NMR screening of RNA secondary structure and binding. *Journal of biomolecular NMR* **63,** 67–76; 10.1007/s10858-015-9967-y (2015).

118. Felli, I. C., Gonnelli, L. & Pierattelli, R. In-cell ^{13}C NMR spectroscopy for the study of intrinsically disordered proteins. *Nature protocols* **9,** 2005–2016; 10.1038/nprot.2014.124 (2014).

119. Peter, M. F. *et al.* Cross-validation of distance measurements in proteins by PELDOR/DEER and single-molecule FRET. *Nature communications* **13,** 4396; 10.1038/s41467-022-31945-6 (2022).

120. Schütz, S. & Sprangers, R. Methyl TROSY spectroscopy: A versatile NMR approach to study challenging biological systems. *Progress in nuclear magnetic resonance spectroscopy* **116,** 56–84; 10.1016/j.pnmrs.2019.09.004 (2020).

121. Verardi, R., Traaseth, N. J., Masterson, L. R., Vostrikov, V. V. & Veglia, G. Isotope labeling for solution and solid-state NMR spectroscopy of membrane proteins. *Advances in experimental medicine and biology* **992,** 35–62; 10.1007/978-94-007-4954-2_3 (2012).

122. Sun, S. *et al.* Solid-state NMR spectroscopy of protein complexes. *Methods in molecular biology (Clifton, N.J.)* **831,** 303–331; 10.1007/978-1-61779-480-3_17 (2012).

123. Drechsler, A. & Separovic, F. Solid-state NMR structure determination. *IUBMB life* **55,** 515–523; 10.1080/15216540310001622740 (2003).

124. Xie, H. *et al.* Solid-state NMR structure determination of a membrane protein in E. coli cellular inner membrane. *Science advances* **9,** eadh4168; 10.1126/sciadv.adh4168 (2023).

125. Peter, M. F. *et al.* Studying Conformational Changes of the Yersinia Type-III-Secretion Effector YopO in Solution by Integrative Structural Biology. *Structure (London, England : 1993)* **27,** 1416-1426.e3; 10.1016/j.str.2019.06.007 (2019).

126. Dumitru, A. C. & Koehler, M. Recent advances in the application of atomic force microscopy to structural biology. *Journal of structural biology* **215,** 107963; 10.1016/j.jsb.2023.107963 (2023).

127. Lyubchenko, Y. L. & Shlyakhtenko, L. S. AFM for analysis of structure and dynamics of DNA and protein-DNA complexes. *Methods (San Diego, Calif.)* **47,** 206–213; 10.1016/j.ymeth.2008.09.002 (2009).

128. Müller, D. J., Janovjak, H., Lehto, T., Kuerschner, L. & Anderson, K. Observing structure, function and assembly of single proteins by AFM. *Progress in biophysics and molecular biology* **79,** 1–43; 10.1016/S0079-6107(02)00009-3 (2002).

129. Puchner, E. M. & Gaub, H. E. Force and function: probing proteins with AFM-based force spectroscopy. *Current opinion in structural biology* **19,** 605–614; 10.1016/j.sbi.2009.09.005 (2009).

130. Fuentes-Perez, M. E., Dillingham, M. S. & Moreno-Herrero, F. AFM volumetric methods for the characterization of proteins and nucleic acids. *Methods (San Diego, Calif.)* **60,** 113–121; 10.1016/j.ymeth.2013.02.005 (2013).

131. Winzer, A. T., Kraft, C., Bhushan, S., Stepanenko, V. & Tessmer, I. Correcting for AFM tip induced topography convolutions in protein-DNA samples. *Ultramicroscopy* **121,** 8–15; 10.1016/j.ultramic.2012.07.002 (2012).

132. Dubrovin, E. V. Atomic force microscopy-based approaches for single-molecule investigation of nucleic acid-protein complexes. *Biophysical reviews* **15,** 1015–1033; 10.1007/s12551-023-01111-3 (2023).

133. Boldon, L., Laliberte, F. & Liu, L. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. *Nano reviews* **6,** 25661; 10.3402/nano.v6.25661 (2015).

134. Da Vela, S. & Svergun, D. I. Methods, development and applications of small-angle X-ray scattering to characterize biological macromolecules in solution. *Current research in structural biology* **2,** 164–170; 10.1016/j.crstbi.2020.08.004 (2020).

135. Piersimoni, L., Kastritis, P. L., Arlt, C. & Sinz, A. Cross-Linking Mass Spectrometry for Investigating Protein Conformations and Protein-Protein Interactions—A Method for All Seasons. *Chemical reviews* **122,** 7500–7531; 10.1021/acs.chemrev.1c00786 (2022).

136. Götze, M. *et al.* Single Nucleotide Resolution RNA-Protein Cross-Linking Mass Spectrometry: A Simple Extension of the CLIR-MS Workflow. *Analytical chemistry* **93,** 14626–14634; 10.1021/acs.analchem.1c02384 (2021).

137. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature protocols* **1,** 1610–1616; 10.1038/nprot.2006.249 (2006).

138. Poulsen, L. D., Kielpinski, L. J., Salama, S. R., Krogh, A. & Vinther, J. SHAPE Selection (SHAPES) enrich for RNA structure signal in SHAPE sequencing-based probing data. *RNA (New York, N.Y.)* **21,** 1042–1052; 10.1261/rna.047068.114. (2015).

139. Lai, W.-J. C. & Ermolenko, D. N. Ensemble and single-molecule FRET studies of protein synthesis. *Methods (San Diego, Calif.)* **137,** 37–48; 10.1016/j.ymeth.2017.12.007 (2018).

140. Kubitscheck & Bloeck, F. *Fluorescence Microscopy - From Principles to Biological Applications* (John Wiley & Sons, Inc, [Place of publication not identified], 2017).

141. Piston, D. W. & Kremers, G.-J. Fluorescent protein FRET: the good, the bad and the ugly. *Trends in biochemical sciences* **32,** 407–414; 10.1016/j.tibs.2007.08.003 (2007).

142. Iqbal, A. *et al.* Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proceedings of the National Academy of Sciences of the United States of America* **105,** 11176–11181; 10.1073/pnas.0801707105 (2008).

143. Lakowicz, J. R. *et al.* Distance distributions in proteins recovered by using frequency-domain fluorometry. Applications to troponin I and its complex with troponin C. *Biochemistry* **27,** 9149–9160; 10.1021/bi00426a012 (1988).

144. Shelke, S. A. & Sigurdsson, S. T. Site-Directed Spin Labeling for EPR Studies of Nucleic Acids. In *Modified Nucleic Acids,* edited by K. Nakatani & Y. Tor (Springer International Publishing, Cham, 2016), Vol. 31, pp. 159–187.

145. Schiemann, O. *et al.* Spin labeling of oligonucleotides with the nitroxide TPA and use of PELDOR, a pulse EPR method, to measure intramolecular distances. *Nature protocols* **2,** 904–923; 10.1038/nprot.2007.97 (2007).

146. Kar, M. *et al.* Phase-separating RNA-binding proteins form heterogeneous distributions of clusters in subsaturated solutions. *Proceedings of the National Academy of Sciences of the United States of America* **119,** e2202222119; 10.1073/pnas.2202222119 (2022).

147. Agam, G. *et al.* Reliability and accuracy of single-molecule FRET studies for characterization of structural dynamics and distances in proteins. *Nature methods* **20,** 523–535; 10.1038/s41592-023-01807-0 (2023).

148. Sakon, J. J. & Weninger, K. R. Detecting the conformation of individual proteins in live cells. *Nature methods* **7,** 203–205; 10.1038/nmeth.1421 (2010).

149. Hellenkamp, B. *et al.* Precision and accuracy of single-molecule FRET measurements-a multi-laboratory benchmark study. *Nature methods* **15,** 669–676; 10.1038/s41592-018-0085-0 (2018).

150. Lerner, E. *et al.* FRET-based dynamic structural biology: Challenges, perspectives and an appeal for open-science practices. *eLife* **10**; 10.7554/eLife.60416 (2021).

151. Google Deepmind and EMBL-EBI. AlphaFold Protein Structure Database. Available at https://alphafold.ebi.ac.uk/ (2024).

152. Baek, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science (New York, N.Y.)* **373,** 871–876; 10.1126/science.abj8754 (2021).

153. Evans, R. *et al.* Protein complex prediction with AlphaFold-Multimer; 10.1101/2021.10.04.463034 (2022).

154. Varadi, M. *et al.* AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic acids research* **50,** D439-D444; 10.1093/nar/gkab1061 (2022).

155. Akdel, M. *et al.* A structural biology community assessment of AlphaFold2 applications. *Nature structural & molecular biology* **29,** 1056–1067; 10.1038/s41594-022-00849-w (2022).

156. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596,** 590–596; 10.1038/s41586-021-03828-1 (2021).

157. Ruff, K. M. & Pappu, R. V. AlphaFold and Implications for Intrinsically Disordered Proteins. *Journal of molecular biology* **433,** 167208; 10.1016/j.jmb.2021.167208 (2021).

158. Bruley, A., Mornon, J.-P., Duprat, E. & Callebaut, I. Digging into the 3D Structure Predictions of AlphaFold2 with Low Confidence: Disorder and Beyond. *Biomolecules* **12**; 10.3390/biom12101467 (2022).

159. Versini, R. *et al.* A Perspective on the Prospective Use of AI in Protein Structure Prediction. *Journal of chemical information and modeling* **64,** 26–41; 10.1021/acs.jcim.3c01361 (2024).

160. Alderson, T. R., Pritišanac, I., Kolarić, Đ., Moses, A. M. & Forman-Kay, J. D. Systematic identification of conditionally folded intrinsically disordered regions by AlphaFold2. *Proceedings of the National Academy of Sciences of the United States of America* **120,** e2304302120; 10.1073/pnas.2304302120 (2023).

161. Escobedo, N. *et al. Ghost interactions: revealing missing protein-ligand interactions using AlphaFold predictions* (2023).

162. Barbarin-Bocahu, I. & Graille, M. The X-ray crystallography phase problem solved thanks to AlphaFold and RoseTTAFold models: a case-study report. *Acta crystallographica. Section D, Structural biology* **78,** 517–531; 10.1107/S2059798322002157 (2022).

163. Ceppi, I. *et al.* PLK1 regulates CtIP and DNA2 interplay in long-range DNA end resection. *Genes & development* **37,** 119–135; 10.1101/gad.349981.122 (2023).

164. McBride, J. M. *et al. AlphaFold2 can predict single-mutation effects* (2022).

165. Wayment-Steele, H. K. *et al.* Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature* **625,** 832–839; 10.1038/s41586-023-06832-9 (2024).

166. Del Alamo, D., Sala, D., Mchaourab, H. S. & Meiler, J. Sampling alternative conformational states of transporters and receptors with AlphaFold2. *eLife* **11**; 10.7554/eLife.75751 (2022).

167. Google DeepMind. AlphaFold3 server.

168. Baek, M. *et al.* Accurate prediction of protein-nucleic acid complexes using RoseTTAFoldNA. *Nature methods* **21,** 117–121; 10.1038/s41592-023-02086-5 (2024).

169. Sinha, S., Tam, B. & Wang, S. M. Applications of Molecular Dynamics Simulation in Protein Study. *Membranes* **12**; 10.3390/membranes12090844 (2022).

170. Bock, L. V., Igaev, M. & Grubmüller, H. Single-particle Cryo-EM and molecular dynamics simulations: A perfect match. *Current opinion in structural biology* **86,** 102825; 10.1016/j.sbi.2024.102825 (2024).

171. Bock, L. V., Gabrielli, S., Kolář, M. H. & Grubmüller, H. Simulation of Complex Biomolecular Systems: The Ribosome Challenge. *Annual review of biophysics* **52,** 361–390; 10.1146/annurev-biophys-111622-091147 (2023).

172. Drescher, M., Jeschke, G. & Bordignon, E. *EPR spectrsocpy. Applications in chemistry and biology* (Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, 2012).

173. Goldfarb, D. & Stoll, S. (eds.). *Handbook of EPR spectroscopy. Fundamentals and methods* (Wiley, Hoboken, N.J., 2018).

174. KUPROV, I. *Spin. From Basic Symmetries to Quantum Optimal Control* (Springer International Publishing, Cham, 2023).

175. Levitt, M. H. *Spin dynamics. Basics of nuclear magnetic resonance.* 2nd ed. (Wiley, Chichester, 2015).

176. Petasis, D. T. *EPR spectroscopy* (De Gruyter, Berlin, 2022).

177. Schweiger, A. & Jeschke, G. *Principles of pulse electron paramagnetic resonance* (Oxford University Press, Oxford [etc.], repr. 2005).

178. Tsvetkov, Y. D., Bowman, M. K. & Grishin, Y. A. *Pulsed electron-electron double resonance. Nanoscale distance measurement in the biological, materials and chemical sciences* (Springer, Cham, Switzerland, 2019).

179. Yang, Z. *et al.* Pulsed ESR dipolar spectroscopy for distance measurements in immobilized spin labeled proteins in liquid solution. *Journal of the American Chemical Society* **134,** 9950–9952; 10.1021/ja303791p (2012).

180. Pliotas, C. *et al.* Conformational state of the MscS mechanosensitive channel in solution revealed by pulsed electron-electron double resonance (PELDOR) spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **109,** E2675-82; 10.1073/pnas.1202286109 (2012).

181. Orwick-Rydmark, M. *et al.* Detergent-free incorporation of a seven-transmembrane receptor protein into nanosized bilayer Lipodisq particles for functional and biophysical studies. *Nano letters* **12,** 4687–4692; 10.1021/nl3020395 (2012).

182. Fleck, N. *et al.* SLIM: A Short-Linked, Highly Redox-Stable Trityl Label for High-Sensitivity In-Cell EPR Distance Measurements. *Angewandte Chemie (International ed. in English)* **59,** 9767–9772; 10.1002/anie.202004452 (2020).

183. Pierro, A. & Drescher, M. Dance with spins: site-directed spin labeling coupled to electron paramagnetic resonance spectroscopy directly inside cells. *Chemical communications (Cambridge, England)* **59,** 1274–1284; 10.1039/d2cc05907j (2023).

184. Wuebben, C., Vicino, M. F., Mueller, M. & Schiemann, O. Do the P1 and P2 hairpins of the Guanidine-II riboswitch interact? *Nucleic acids research* **48,** 10518–10526; 10.1093/nar/gkaa703 (2020).

185. Ackermann, K., Wort, J. L. & Bode, B. E. Nanomolar Pulse Dipolar EPR Spectroscopy in Proteins: CuII-CuII and Nitroxide-Nitroxide Cases. *The journal of physical chemistry. B* **125,** 5358–5364; 10.1021/acs.jpcb.1c03666 (2021).

186. Ackermann, K., Heubach, C. A., Schiemann, O. & Bode, B. E. Pulse Dipolar Electron Paramagnetic Resonance Spectroscopy Distance Measurements at Low Nanomolar Concentrations: The CuII-Trityl Case. *The journal of physical chemistry letters* **15,** 1455–1461; 10.1021/acs.jpclett.3c03311 (2024).

187. Malygin, A. A. *et al.* Exploring the interactions of short RNAs with the human 40S ribosomal subunit near the mRNA entry site by EPR spectroscopy. *Nucleic acids research* **47,** 11850–11860; 10.1093/nar/gkz1039 (2019).

188. Constantinescu-Aruxandei, D., Petrovic-Stojanovska, B., Schiemann, O., Naismith, J. H. & White, M. F. Taking a molecular motor for a spin: helicase mechanism studied by spin labeling and PELDOR. *Nucleic acids research* **44,** 954–968; 10.1093/nar/gkv1373 (2016).

189. Cooke, J. A. & Brown, L. J. Distance measurements by continuous wave EPR spectroscopy to monitor protein folding. *Methods in molecular biology (Clifton, N.J.)* **752,** 73–96; 10.1007/978-1-60327-223-0_6 (2011).

190. Hustedt, E. J. *et al.* Dipolar coupling between nitroxide spin labels: the development and application of a tether-in-a-cone model. *Biophysical journal* **90,** 340–356; 10.1529/biophysj.105.068544 (2006).

191. Schmidt, T., Wälti, M. A., Baber, J. L., Hustedt, E. J. & Clore, G. M. Long Distance Measurements up to 160 Å in the GroEL Tetradecamer Using Q-Band DEER EPR Spectroscopy. *Angewandte Chemie (International ed. in English)* **55,** 15905–15909; 10.1002/anie.201609617 (2016).

192. El Mkami, H., Ward, R., Bowman, A., Owen-Hughes, T. & Norman, D. G. The spatial effect of protein deuteration on nitroxide spin-label relaxation: implications for EPR distance measurement. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **248,** 36–41; 10.1016/j.jmr.2014.09.010 (2014).

193. Pievo, R. *et al.* A rapid freeze-quench setup for multi-frequency EPR spectroscopy of enzymatic reactions. *Chemphyschem : a European journal of chemical physics and physical chemistry* **14,** 4094–4101; 10.1002/cphc.201300714 (2013).

194. Collauto, A. *et al.* Rates and equilibrium constants of the ligand-induced conformational transition of an HCN ion channel protein domain determined by DEER spectroscopy. *Physical chemistry chemical physics : PCCP* **19,** 15324–15334; 10.1039/c7cp01925d (2017).

195. Hett, T. *et al.* Spatiotemporal Resolution of Conformational Changes in Biomolecules by Combining Pulsed Electron-Electron Double Resonance

Spectroscopy with Microsecond Freeze-Hyperquenching. *Journal of the American Chemical Society* **143,** 6981–6989; 10.1021/jacs.1c01081 (2021).

196. Weber, S. Transient EPR. In *eMagRes,* edited by R. K. Harris & R. L. Wasylishen (John Wiley & Sons, Ltd, Chichester, UK, 2007), pp. 255–270.

197. Murphy D., Editors: Jackson, S. D. & Hargreaves, J. S. J. Metal Oxide Catalysis. EPR Spectroscopy on Polycrystalline Oxide systems.

198. Weil, J. A. & Bolton, J. R. *Electron paramagnetic resonance. Elementary theory and practical applications*. 2nd ed. (Wiley, Hoboken, 2007).

199. Hesse, M., Meier, H. & Zeeh, B. *Spektroskopische Methoden in der organischen Chemie. 102 Tabellen*. 7th ed. (Thieme, Stuttgart, 2005).

200. Carlo Corvaglia, Editors: Marina Brustolon & Elio Giamello. Electron Paramagnetic Resonance - A Practitioners Toolkit. Principles and Applications of Electron Paramagnetic Spectroscopy.

201. van der Est, A. Continuous-Wave EPR. In *eMagRes,* edited by R. K. Harris & R. L. Wasylishen (John Wiley & Sons, Ltd, Chichester, UK, 2007), pp. 1411–1422.

202. Sahu, I. D. & Lorigan, G. A. Electron Paramagnetic Resonance as a Tool for Studying Membrane Proteins. *Biomolecules* **10**; 10.3390/biom10050763 (2020).

203. Goldfarb, D. Pulse EPR in biological systems - Beyond the expert's courtyard. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **306,** 102–108; 10.1016/j.jmr.2019.07.038 (2019).

204. Reginsson, G. W. & Schiemann, O. Pulsed electron-electron double resonance: beyond nanometre distance measurements on biomacromolecules. *The Biochemical journal* **434,** 353–363; 10.1042/BJ20101871 (2011).

205. Rabenstein, M. D. & Shin, Y. K. Determination of the distance between two spin labels attached to a macromolecule. *Proceedings of the National Academy of Sciences of the United States of America* **92,** 8239–8243; 10.1073/pnas.92.18.8239 (1995).

206. Jeschke, G., Speiss, H. W., Editors: Dolinsek, D., Vilfan, M. & Zumer, S. Novel NMR and EPR techniques. Distance Measurements in Solid-State NMR and EPR Spectroscopy.

207. Borbat, P. P. & Freed, J. H. Multiple-quantum ESR and distance measurements. *Chemical Physics Letters* **313,** 145–154; 10.1016/S0009-2614(99)00972-0 (1999).

208. Jeschke, G., Pannier, M., Godt, A. & Spiess, H. W. Dipolar spectroscopy and spin alignment in electron paramagnetic resonance. *Chemical Physics Letters* **331,** 243–252; 10.1016/S0009-2614(00)01171-4 (2000).

209. Milikisyants, S., Scarpelli, F., Finiguerra, M. G., Ubbink, M. & Huber, M. A pulsed EPR method to determine distances between paramagnetic centers with strong spectral anisotropy and radicals: the dead-time free RIDME sequence. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **201,** 48–56; 10.1016/j.jmr.2009.08.008 (2009).

210. Pannier, M., Veit, S., Godt, A., Jeschke, G. & Spiess, H. W. Dead-time free measurement of dipole-dipole interactions between electron spins. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **142,** 331–340; 10.1006/jmre.1999.1944 (2000).

211. Kurshev, V. V., Raitsimring, A. M. & Tsvetkov, Y. D. Selection of dipolar interaction by the "2 + 1" pulse train ESE. *Journal of Magnetic Resonance (1969)* **81,** 441–454; 10.1016/0022-2364(89)90080-2 (1989).

212. Milov, A. D., Salikhov, K. M. & Shirov, M. D. *Application of the double resonance method to double electron spin echo in a study of spatial distribiution of paranagmetic centers in solids* (1981).

213. Tsvetkov, Y. D., Bowman, M. K. & Grishin, Y. A. *Pulsed Electron–Electron Double Resonance* (Springer International Publishing, 2019).

214. Schiemann, O. *et al.* Benchmark Test and Guidelines for DEER/PELDOR Experiments on Nitroxide-Labeled Biomolecules. *Journal of the American Chemical Society* **143,** 17875–17890; 10.1021/jacs.1c07371 (2021).

215. Jeschke, G. & Polyhach, Y. Distance measurements on spin-labelled biomacromolecules by pulsed electron paramagnetic resonance. *Physical chemistry chemical physics : PCCP* **9,** 1895–1910; 10.1039/b614920k (2007).

216. Jeschke, G., Sajid, M., Schulte, M. & Godt, A. Three-spin correlations in double electron-electron resonance. *Physical chemistry chemical physics : PCCP* **11,** 6580–6591; 10.1039/b905724b (2009).

217. Borbat, P. P. & Freed, J. H. Pulse Dipolar Electron Spin Resonance: Distance Measurements. In *Structural Information from Spin-Labels and Intrinsic Paramagnetic Centres in the Biosciences,* edited by C. R. Timmel & J. R. Harmer (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), Vol. 152, pp. 1–82.

218. Gränz, M. *et al.* Dynamics of Nucleic Acids at Room Temperature Revealed by Pulsed EPR Spectroscopy. *Angewandte Chemie (International ed. in English)* **57,** 10540–10543; 10.1002/anie.201803682 (2018).

219. Abdullin, D. *et al.* Pulsed EPR Dipolar Spectroscopy on Spin Pairs with one Highly Anisotropic Spin Center: The Low-Spin FeIII Case. *Chemistry (Weinheim an der Bergstrasse, Germany)* **25,** 14388–14398; 10.1002/chem.201902908 (2019).

220. Abdullin, D., Florin, N., Hagelueken, G. & Schiemann, O. EPR-based approach for the localization of paramagnetic metal ions in biomolecules. *Angewandte Chemie (International ed. in English)* **54,** 1827–1831; 10.1002/anie.201410396 (2015).

221. Meyer, A. *et al.* 19F Electron-Nuclear Double Resonance Reveals Interaction between Redox-Active Tyrosines across the α/β Interface of E. coli Ribonucleotide Reductase. *Journal of the American Chemical Society* **144,** 11270–11282; 10.1021/jacs.2c02906 (2022).

222. Widder, P., Schuck, J., Summerer, D. & Drescher, M. Combining site-directed spin labeling in vivo and in-cell EPR distance determination. *Physical chemistry chemical physics : PCCP* **22,** 4875–4879; 10.1039/C9CP05584C (2020).

223. Kisseleva, N., Khvorova, A., Westhof, E. & Schiemann, O. Binding of manganese(II) to a tertiary stabilized hammerhead ribozyme as studied by electron paramagnetic resonance spectroscopy. *RNA (New York, N.Y.)* **11,** 1–6; 10.1261/rna.7127105 (2005).

224. Woody, A. Y., Eaton, S. S., Osumi-Davis, P. A. & Woody, R. W. Asp537 and Asp812 in bacteriophage T7 RNA polymerase as metal ion-binding sites studied by EPR, flow-dialysis, and transcription. *Biochemistry* **35,** 144–152; 10.1021/bi952037f (1996).

225. Borggräfe, J. *et al.* Time-resolved structural analysis of an RNA-cleaving DNA catalyst. *Nature* **601,** 144–149; 10.1038/s41586-021-04225-4 (2022).

226. Schiemann, O., Fritscher, J., Kisseleva, N., Sigurdsson, S. T. & Prisner, T. F. Structural investigation of a high-affinity MnII binding site in the hammerhead ribozyme by EPR spectroscopy and DFT calculations. Effects of neomycin B on metal-ion binding. *Chembiochem : a European journal of chemical biology* **4,** 1057–1065; 10.1002/cbic.200300653 (2003).

227. Altenbach, C., Marti, T., Khorana, H. G. & Hubbell, W. L. Transmembrane protein structure: spin labeling of bacteriorhodopsin mutants. *Science (New York, N.Y.)* **248,** 1088–1092; 10.1126/science.2160734 (1990).

228. Roser, P., Schmidt, M. J., Drescher, M. & Summerer, D. Site-directed spin labeling of proteins for distance measurements in vitro and in cells. *Organic & biomolecular chemistry* **14,** 5468–5476; 10.1039/c6ob00473c (2016).

229. Heubach, C. A. *et al.* Differentiating between Label and Protein Conformers in Pulsed Dipolar EPR Spectroscopy with the dHis-Cu2+ (NTA) Motif. *Chemistry (Weinheim an der Bergstrasse, Germany)* **29,** e202302541; 10.1002/chem.202302541 (2023).

230. Ghosh, S. *et al.* Cu2+-based distance measurements by pulsed EPR provide distance constraints for DNA backbone conformations in solution. *Nucleic acids research* **48,** e49; 10.1093/nar/gkaa133 (2020).

231. Gamble Jarvi, A., Bogetti, X., Singewald, K., Ghosh, S. & Saxena, S. Going the dHis-tance: Site-Directed Cu2+ Labeling of Proteins and Nucleic Acids. *Accounts of chemical research* **54,** 1481–1491; 10.1021/acs.accounts.0c00761 (2021).

232. Goldfarb, D. Gd3+ spin labeling for distance measurements by pulse EPR spectroscopy. *Physical chemistry chemical physics : PCCP* **16,** 9685–9699; 10.1039/c3cp53822b (2014).

233. Qi, M., Gross, A., Jeschke, G., Godt, A. & Drescher, M. Gd(III)-PyMTA label is suitable for in-cell EPR. *Journal of the American Chemical Society* **136,** 15366–15378; 10.1021/ja508274d (2014).

234. Yang, Y., Yang, F., Li, X.-Y., Su, X.-C. & Goldfarb, D. In-Cell EPR Distance Measurements on Ubiquitin Labeled with a Rigid PyMTA-Gd(III) Tag. *The journal of physical chemistry. B* **123,** 1050–1059; 10.1021/acs.jpcb.8b11442 (2019).

235. Shelke, S. A. & Sigurdsson, S. T. Site-Directed Nitroxide Spin Labeling of Biopolymers. In *Structural Information from Spin-Labels and Intrinsic Paramagnetic Centres in the Biosciences,* edited by C. R. Timmel & J. R. Harmer (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), Vol. 152, pp. 121–162.

236. Karthikeyan, G. *et al.* A Bioresistant Nitroxide Spin Label for In-Cell EPR Spectroscopy: In Vitro and In Oocytes Protein Structural Dynamics Studies. *Angewandte Chemie (International ed. in English)* **57,** 1366–1370; 10.1002/anie.201710184 (2018).

237. Torricella, F., Pierro, A., Mileo, E., Belle, V. & Bonucci, A. Nitroxide spin labels and EPR spectroscopy: A powerful association for protein dynamics studies. *Biochimica et biophysica acta. Proteins and proteomics* **1869,** 140653; 10.1016/j.bbapap.2021.140653 (2021).

238. Yang, Y. *et al.* In-Cell Trityl-Trityl Distance Measurements on Proteins. *The journal of physical chemistry letters* **11,** 1141–1147; 10.1021/acs.jpclett.9b03208 (2020).

239. Fleck, N. *et al.* Ox-SLIM: Synthesis of and Site-Specific Labelling with a Highly Hydrophilic Trityl Spin Label. *Chemistry (Weinheim an der Bergstrasse, Germany)* **27,** 5292–5297; 10.1002/chem.202100013 (2021).

240. Reginsson, G. W., Kunjir, N. C., Sigurdsson, S. T. & Schiemann, O. Trityl radicals: spin labels for nanometer-distance measurements. *Chemistry (Weinheim an der Bergstrasse, Germany)* **18,** 13580–13584; 10.1002/chem.201203014 (2012).

241. Gebhardt, C. *et al.* Molecular and Spectroscopic Characterization of Green and Red Cyanine Fluorophores from the Alexa Fluor and AF Series*. *Chemphyschem : a European journal of chemical physics and physical chemistry* **22,** 1566–1583; 10.1002/cphc.202000935 (2021).

242. Yagi, H. *et al.* Gadolinium tagging for high-precision measurements of 6 nm distances in protein assemblies by EPR. *Journal of the American Chemical Society* **133,** 10418–10421; 10.1021/ja204415w (2011).

243. Widder, P., Berner, F., Summerer, D. & Drescher, M. Double Nitroxide Labeling by Copper-Catalyzed Azide-Alkyne Cycloadditions with Noncanonical

Amino Acids for Electron Paramagnetic Resonance Spectroscopy. *ACS chemical biology* **14,** 839–844; 10.1021/acschembio.8b01111 (2019).

244. Wuebben, C. *et al.* Site-Directed Spin Labeling of RNA with a Gem-Diethylisoindoline Spin Label: PELDOR, Relaxation, and Reduction Stability. *Molecules (Basel, Switzerland)* **24**; 10.3390/molecules24244482 (2019).

245. Ackermann, K., Chapman, A. & Bode, B. E. A Comparison of Cysteine-Conjugated Nitroxide Spin Labels for Pulse Dipolar EPR Spectroscopy. *Molecules (Basel, Switzerland)* **26**; 10.3390/molecules26247534 (2021).

246. Zehnder, J., Cadalbert, R., Yulikov, M., Künze, G. & Wiegand, T. Paramagnetic spin labeling of a bacterial DnaB helicase for solid-state NMR. *Journal of magnetic resonance (San Diego, Calif. : 1997)* **332,** 107075; 10.1016/j.jmr.2021.107075 (2021).

247. Fleissner, M. R. *et al.* Site-directed spin labeling of a genetically encoded unnatural amino acid. *Proceedings of the National Academy of Sciences of the United States of America* **106,** 21637–21642; 10.1073/pnas.0912009106 (2009).

248. Braun, T. S. *et al.* Isoindoline-Based Nitroxides as Bioresistant Spin Labels for Protein Labeling through Cysteines and Alkyne-Bearing Noncanonical Amino Acids. *Chembiochem : a European journal of chemical biology* **21,** 958–962; 10.1002/cbic.201900537 (2020).

249. Hermanson, G. T. The Reactions of Bioconjugation. In *Bioconjugate Techniques* (Elsevier2013), pp. 229–258.

250. Miseta, A. & Csutora, P. Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Molecular biology and evolution* **17,** 1232–1239; 10.1093/oxfordjournals.molbev.a026406 (2000).

251. Belin, D. & Puigbò, P. Why Is the UAG (Amber) Stop Codon Almost Absent in Highly Expressed Bacterial Genes? *Life (Basel, Switzerland)* **12**; 10.3390/life12030431 (2022).

252. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annual review of biochemistry* **79,** 413–444; 10.1146/annurev.biochem.052308.105824 (2010).

253. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science (New York, N.Y.)* **292,** 498–500; 10.1126/science.1060077 (2001).

254. Chatterjee, A., Sun, S. B., Furman, J. L., Xiao, H. & Schultz, P. G. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli. *Biochemistry* **52,** 1828–1837; 10.1021/bi4000244 (2013).

255. Young, C. L., Britton, Z. T. & Robinson, A. S. Recombinant protein expression and purification: a comprehensive review of affinity tags and microbial applications. *Biotechnology journal* **7,** 620–634; 10.1002/biot.201100155 (2012).

256. Duong-Ly, K. C. & Gabelli, S. B. Affinity Purification of a Recombinant Protein Expressed as a Fusion with the Maltose-Binding Protein (MBP) Tag. *Methods in enzymology* **559,** 17–26; 10.1016/bs.mie.2014.11.004 (2015).

257. Parks, T. D., Leuther, K. K., Howard, E. D., Johnston, S. A. & Dougherty, W. G. Release of proteins and peptides from fusion proteins using a recombinant plant virus proteinase. *Analytical biochemistry* **216,** 413–417; 10.1006/abio.1994.1060 (1994).

258. Mertens, N., Remaut, E. & Fiers, W. Tight transcriptional control mechanism ensures stable high-level expression from T7 promoter-based expression plasmids. *Bio/technology (Nature Publishing Company)* **13,** 175–179; 10.1038/nbt0295-175 (1995).

259. Namdev, P., Dar, H. Y., Srivastava, R. K., Mondal, R. & Anupam, R. Induction of T7 Promoter at Higher Temperatures May Be Counterproductive. *Indian journal of clinical biochemistry : IJCB* **34,** 357–360; 10.1007/s12291-019-0813-y (2019).

260. Nam, H., Hwang, B. J., Choi, D.-Y., Shin, S. & Choi, M. Tobacco etch virus (TEV) protease with multiple mutations to improve solubility and reduce self-cleavage exhibits enhanced enzymatic activity. *FEBS open bio* **10,** 619–626; 10.1002/2211-5463.12828 (2020).

261. Carmignotto, G. P. & Azzoni, A. R. On the expression of recombinant Cas9 protein in E. coli BL21(DE3) and BL21(DE3) Rosetta strains. *Journal of biotechnology* **306,** 62–70; 10.1016/j.jbiotec.2019.09.012 (2019).

262. Bhawsinghka, N., Glenn, K. F. & Schaaper, R. M. Complete Genome Sequence of Escherichia coli BL21-AI. *Microbiology resource announcements* **9**; 10.1128/MRA.00009-20 (2020).

263. Swanson, M. S. & Dreyfuss, G. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Molecular and cellular biology* **8,** 2237–2241; 10.1128/mcb.8.5.2237-2241.1988 (1988).

264. Baronti, L., Karlsson, H., Marušič, M. & Petzold, K. A guide to large-scale RNA sample preparation. *Analytical and bioanalytical chemistry* **410,** 3239–3252; 10.1007/s00216-018-0943-8 (2018).

265. Hegg, L. A. & Thurlow, D. L. Residual tRNA secondary structure in 'denaturing' 8M urea/TBE polyacrylamide gels: effects on electrophoretic mobility and dependency on prior chemical modification of the tRNA. *Nucleic acids research* **18,** 2993–3000; 10.1093/nar/18.10.2993 (1990).

266. Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. & Lian, L.-Y. A simple method for improving protein solubility and long-term stability. *Journal of the American Chemical Society* **126,** 8933–8939; 10.1021/ja049297h (2004).

267. Walker, S. E. & Lorsch, J. RNA purification--precipitation methods. *Methods in enzymology* **530,** 337–343; 10.1016/B978-0-12-420037-1.00019-1 (2013).

268. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein engineering* **8,** 127–134; 10.1093/protein/8.2.127 (1995).

269. Dehouck, Y., Kwasigroch, J. M., Gilis, D. & Rooman, M. PoPMuSiC 2.1: a web server for the estimation of protein stability changes upon mutation and sequence optimality. *BMC bioinformatics* **12,** 151; 10.1186/1471-2105-12-151 (2011).

270. Rooman, M. dezyme. PoPMuSiC. Available at http://dezyme.com/.

271. Belin, D. & Puigbò, P. Why Is the UAG (Amber) Stop Codon Almost Absent in Highly Expressed Bacterial Genes? *Life (Basel, Switzerland)* **12**; 10.3390/life12030431 (2022).

272. Hagelueken, G., Abdullin, D., Ward, R. & Schiemann, O. mtsslSuite: In silico spin labelling, trilateration and distance-constrained rigid body docking in PyMOL. *Molecular physics* **111,** 2757–2766; 10.1080/00268976.2013.809804 (2013).

273. Gregor Hagelüken. mtsslSuite. Available at http://www.mtsslsuite.isb.ukbonn.de/.

274. Jain, N. K. & Roy, I. Effect of trehalose on protein structure. *Protein science : a publication of the Protein Society* **18,** 24–36; 10.1002/pro.3 (2009).

275. de Dong *et al.* The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **532,** 522–526; 10.1038/nature17944 (2016).

276. Jeschke, G. *et al.* DeerAnalysis2006—a comprehensive software package for analyzing pulsed ELDOR data. *Appl. Magn. Reson.* **30,** 473–498; 10.1007/BF03166213 (2006).

277. Lane, T. J. Protein structure prediction has reached the single-structure frontier. *Nature methods* **20,** 170–173; 10.1038/s41592-022-01760-4 (2023).

278. SIB Swiss institute of Bioinformatics. SIM Alignment Tool for Protein Sequences. Available at https://web.expasy.org/sim/ (2024).

279. Tim Mevs. preliminary title: Structural elucidation of LbuCas13a. master thesis. University of Bonn, 2024.

280. Lejla Tokic. Nachweis der Bindung von Cas13a an in vitro transkribierter CRISPR RNA. bachelor thesis. University of Bonn, 2020.

281. Kligun, E. & Mandel-Gutfreund, Y. The role of RNA conformation in RNA-protein recognition. *RNA biology* **12,** 720–727; 10.1080/15476286.2015.1040977 (2015).

282. 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 research* **49,** 5278–5293; 10.1093/nar/gkab286 (2021).

283. Nguyen, L. T. *et al.* Harnessing noncanonical crRNAs to improve functionality of Cas12a orthologs. *Cell reports* **43,** 113777; 10.1016/j.celrep.2024.113777 (2024).

284. Hu, G. & Zhou, H.-X. Magnesium ions mediate ligand binding and conformational transition of the SAM/SAH riboswitch. *Communications biology* **6,** 791; 10.1038/s42003-023-05175-5 (2023).

285. Misra, V. K. & Draper, D. E. The linkage between magnesium binding and RNA folding. *Journal of molecular biology* **317,** 507–521; 10.1006/jmbi.2002.5422 (2002).

286. Nguyen, G. T. *et al.* CRISPR-Cas12a exhibits metal-dependent specificity switching. *Nucleic acids research*; 10.1093/nar/gkae613 (2024).

287. Bock, L. V. & Grubmüller, H. Effects of cryo-EM cooling on structural ensembles. *Nature communications* **13,** 1709; 10.1038/s41467-022-29332-2 (2022).

288. Han, B.-G., Avila-Sakar, A., Remis, J. & Glaeser, R. M. Challenges in making ideal cryo-EM samples. *Current opinion in structural biology* **81,** 102646; 10.1016/j.sbi.2023.102646 (2023).

289. Liu, H. & Naismith, J. H. An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC biotechnology* **8,** 91; 10.1186/1472-6750-8-91 (2008).

290. Hett, T. & Schiemann, O. PELDOR Measurements on Nitroxide-Labeled Oligonucleotides. *Methods in molecular biology (Clifton, N.J.)* **2439,** 241–274; 10.1007/978-1-0716-2047-2_16 (2022).