UCPH NanoScience
NanoVidenskab KU

tidsskrift.dk/index.php/NanoVidenskabKU

Thesis Paper

Structural dynamics of Cpf1 by FRET

Freja Jacobsen Bohr[a*] and Nikos Hatzakis[b]

Submitted June 2018, accepted July 2018

The RNA-guided endonuclease of class V Cpf1 *(Clustered Regularly Interspaced Short Palindromic Repeats* from *Prevotella and Francisella)* is a central element in prokaryotic immune mechanism, which use a CRISPR-RNA (crRNA) to locate and cleave viral DNA. Cpf1 allow genome-editing at a specific position specified by synthetic crRNA, hence a promising therapeutic agent to treat gene deficiencies. Our current understanding of Cpf1 structure and function primarily relies on crystal structures and cryoEM data, providing unique and invariant structures. In contrast to recent methods, the scope of this project is to expand our understanding of the dynamic structure of Cpf1, and understand how conformational changes and catalysis are related. To do this I used FRET (Förster Resonance Energy Transfer). Here, the work on this projects is presented, with the aim of 1) understanding the conformational changes of Cpf1 in free-form, binary complex (crRNA-bound) and tertiary complex (DNA-bound) in bulk, and correlating the conformational changes to intramolecular distances, thus providing insight to the mechanism of DNA-cleavage and 2) calibrate the distance using dsDNA as a rigid scaffold, preparing for single molecule FRET measurements on Cpf1. Ensemble measurements revealed conformational changes of Cpf1 upon binding DNA, however the assay needs to be optimized further to extract distinct distances. Spectrometric experiments revealed that excess dye in solution was a general problem in ensemble measurements, interfering with the results. To validate and setup the calibration I did both ensemble and smFRET measurements on dual-labelled dsDNA to address and eliminate the role of excess dye in solution. The inter-dye distance of dual-labelled dsDNA was determined to be 62.72 ± 0.93 Å. The simulated distance with Monte Carlo simulations was found to be 61.6 Å. This illustrates smFRET as a method to probe enzymatic motion, and thus could provide novel information about the mechanism of DNA-cleavage in Cpf1, paving the way for future genome-editing.

1 Introduction

In recent years RNA-guided endonucleases like CRISPR-Cas9 *(Clustered Regularly Interspaced Short Palindromic Repeats associated protein 9* from *Streptococcus pyogenes)*, has been of great interest as a powerful genome editing tool. Cas9 is a large, multi-domain protein that undergoes conformational changes induced by single-guided RNA and trans-activating RNA, followed by cleaving of viral target DNA.[1] Providing the Cas9-complex with synthesized guide RNA, allows you to edit the genome at a specific position. The CRISPR technology is applicable to many fields, such as stem cell- and gene therapy as a promising therapeutic option to treat inherited deficiencies such as cystic fibrosis or even cancer.[2,3,4]

In this thesis, the conformational changes of a similar RNA-guided endonuclease of class V Cpf1 *(CRISPR* from *Prevotella and Francisella 1)* has been investigated. Cpf1 is a central element in the prokaryotic immune mechanism, that uses crRNA to locate and cleave viral DNA. In order to deconvolute the DNA-targeting mechanism and hence understand the molecular details in the system, the structure of Cpf1 has recently been solved by our collaborating group at the Panum Institute using protein crystallization and cryoEM.[5]

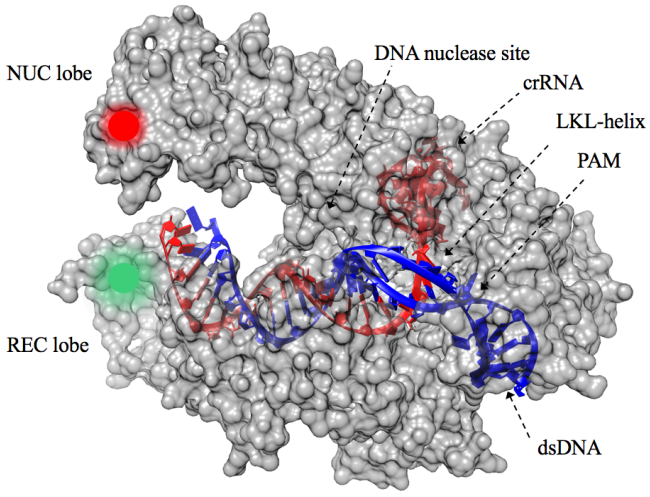

Figure 1: Crystal structure of Cpf1 after DNA cleavage. The transparent surface allows visualization of the crRNA-DNA complex. The NUC and REC lobes are shown, as well as the active site, PAM and LKL-helix. An acceptor fluorophore (red) is at position CYS1190, and a donor fluorophore is at position CYS473. These fluorophores are used for FRET. (PDB ID: 5MGA)

Cpf1 contains two specific hinge regions, the NUC and REC lobe, and displays an oval 'sea conch' structure, as shown in Fig. 1. crRNA-induced conformational changes primes the enzyme to a DNA-binding competent state in presence of magnesium, that acts as a cofactor.[6] Target DNA induces further conformational changes to the Cpf1-crRNA complex, allowing recognition by the PAM (protospacer adjacent motif). PAM is essential for target binding as it is recognized directly by Cpf1. PAM is not found in the bacterial CRISPR locus, but is a component of the invading DNA, this way Cpf1 can differentiate between host and viral DNA. After PAM recognition a LKL helix (loop-lysine helix–loop) is inserted into the target DNA, thus beginning to unzip the dsDNA-strand. Due to electrostatics the ssDNA pairs with the crRNA, forming a crRNA-target-DNA heteroduplex. If crRNA and target-DNA are complementary, the target DNA is cleaved with PAM at 5' end. DNA nuclease activity takes place in a pocket at the interface between the RuvC and NUC domains (Fig. 1).

Current consensus indicates, that the free form of Cpf1 (from now on Cpf1-apo or apo) is primarily in a relaxed fully open conformation. Upon binding to crRNA the enzyme is expected to transit to a semi closed state. Finally, binding of target DNA and recognition by PAM makes the enzyme go towards a more closed compact conformation. As to unwind the DNA-strand, the conformational equilibrium changes to a more open conformation, thus cleaving the strand (Fig. S1). Hence, DNA-bound/DNA-cleaved complex samples many different conformations. From CryoEM and crystal structures the estimated distance between the two lobes in each conformational state has been.[7]

Cpf1 recognizes a T-rich PAM, while Cas9 targets a G-rich PAM, which provides alternate targeting sites. Cpf1 cleaves DNA via staggered cut approximately 18-23 basepairs downstream from the PAM site. By contrast, Cas9 cuts DNA proximal to the PAM site via blunt cut, hence allowing disruption to the recognition sequence after repair. Consequently, Cpf1 has proven to be a smaller and much more simple endonuclease than Cas9, expanding genome editing capabilities.[8]

The scope of this project is to investigate conformational dynamics of Cpf1 using FRET, to provide insight to the mechanism of DNA-cleavage.

1.1 Using FRET to record conformational changes

Förster Resonance Electron Transfer (FRET) can be used as a spectrometric ruler by reporting distance-related molecular changes between two fluorophores (donor and acceptor).[9] In this thesis, two cyanine dyes, Cy3 and Cy5, was chosen as FRET dye pair, since their donor emission spectrum and acceptor absorbance spectrum overlap, thus allowing energy transfer between donor and acceptor, as seen in Fig. 2. The magnitude of energy transfer is distance dependent and happens through intermolecular dipole-dipole coupling. In occurrence of FRET, the donor signal

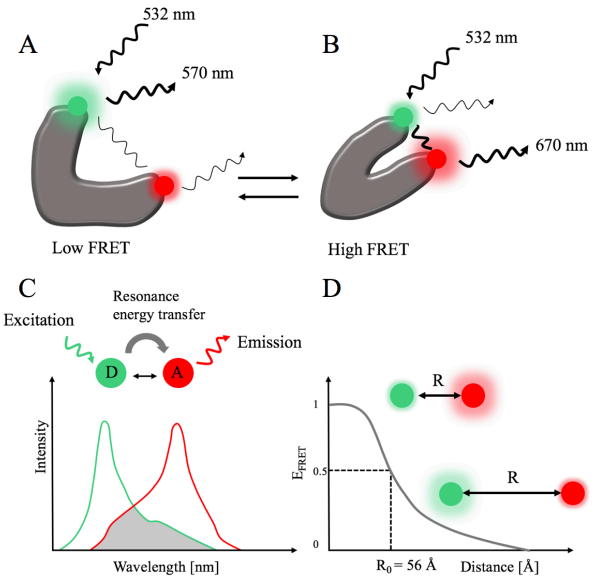

Figure 2: Illustration of how FRET works. **A:** Cy3 (green) and Cy5 (red) are attached to the enzyme at two flexible regions. **B:** Upon conformational changes and thus movement, the inter-dye distance between the fluorophores will change. The closer the two fluorophores are, the more energy will get transferred from Cy3 to Cy5, hence increasing the FRET signal. **C:** If donor emission spectrum (green) and acceptor excitation spectrum (red) overlap, the donor can transfer its resonance energy to the acceptor within a given distance. **D:** FRET efficiency as a function of distance. The Förster radius (R_0) can be found at 50 % FRET. When the dyes are in close proximity the FRET efficiency is high, and if they are far apart, the FRET efficiency is low.

is quenched, transferring energy to the acceptor, increasing the acceptor intensity (Fig. 2).[10] From the intensities of donor and acceptor (I_D and I_A) the FRET efficiency can be calculated as,

$$E_{FRET} = I_A / (I_A + I_D)$$

where I_D and I_A is the background and cross-talk corrected intensities for donor and acceptor, respectively. The FRET efficiency is given as the fraction of energy transfer occurring per donor excitation event. From E_{FRET} the intermolecular distance between donor and acceptor can be calculated using the Förster Radius (R_0), defined as the donor-acceptor distance obtained at 50 % E_{FRET} (here 56 Å[11]),

$$R = R_0 \times (1/E_{FRET} - 1)^{1/6}$$

The Förster radius is proportional to the donor quantum yield (Φ) or the emission efficiency of a given fluorophore, the overlap integral (J) between donor emission and acceptor excitation, the relative dipole orientation factor k^2 ($k^2 = 2/3$) and the refractive index of the medium (n)[12],

$$R_0 \sim k^2 \times \Phi \times J \times n^{-4}$$

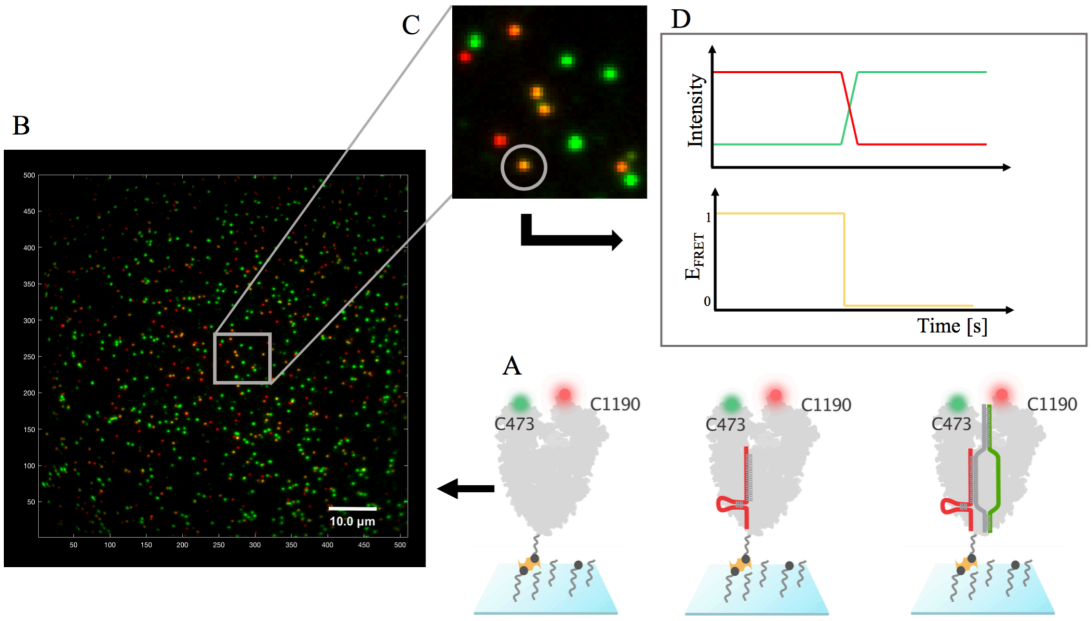

Figure 3: Experimental setup using TIRF microscopy for smFRET analysis. **A:** PLL-PEG surface with passivated Cpf1 through neutravidin-biotin linker. Here three different experimental conditions are shown: free-form (apo), Cpf1-crRNA binary complex and tertiary DNA-bound complex. **B:** Snapshot of the overlay of the red and green channel from dsDNA control experiment with same experimental setup (Fig. 6). The presented TIRF microscope setup provides data similar to what is shown. **C:** Zoom in focussing on donor-only (green), acceptor-only (red) and donor-acceptor (yellow). Colocalizing the D-only and A-only spots reveal FRET events as shown. **D:** Each colocalized donor and acceptor is related to an (idealized) FRET trace.

FRET is most sensitive to distance changes when the donor and acceptor are separated by a distance near the Förster Radius.[13] Thus, if fluorophores are attached to known sites within the enzyme, measurements of the energy transfer efficiency provide insight to inter- and/or intramolecular distances in the macromolecular length scale.[14]

1.2 Motivation for doing FRET studies on Cpf1

To investigate the dynamics and mechanism by which Cpf1 cleaves DNA, FRET studies are a valuable tool, as it is highly distance-sensitive. Ensemble FRET assays allow extraction of intermolecular distances between donor and acceptor in different environments, thereby providing crucial information about the conformational states. Protein dynamics are easily captured using FRET spectroscopy, as it has become a staple technique in recent years.[13] Additionally, ensemble assays are important in determining if a system is suited for smFRET measurements, and establishing the needed controls.

Shortcomings to the ensemble approach are that the extracted FRET distances provide only an average picture of the system, thereby potentially masking different states that might occur. Conformational dynamics can be observed in real-time by tracking changes in single molecule FRET efficiencies over time.[15] Hence, the single molecule approach allows investigation of the dynamic heterogeneity and subsequently disentangle the mechanism underlying conformational states of proteins. With single molecule techniques it is also possible to differentiate each molecule, yielding invaluable insight into distinctive molecular properties. The single molecule approach require complicated microscopy techniques, thus can be very challenging and expensive. Furthermore, single molecule data analysis is very time consuming, and requires numerous measurements to provide reliable and robust statistics, as the single molecule traces are hardly ever as the idealized ones portrayed in Fig. 3.

1.3 Total Internal Reflection Fluorescence microcopy

TIRF microscopy allow observation of thousands of single fluorescent events on surface-immobilized molecules, in a specific time-frame.[16] Immobilization of each molecule is necessary for real-time tracking using TIRF microscopy, as the diffusion of unrestricted molecules away from the field-of-view will make FRET measurements impossible.[17] Experimental setup for such TIRF smFRET experiment can be seen in Fig. 3. The enzymes, labelled with Cy3 and Cy5, are tethered to a PLL-PEG surface. Measuring smFRET through TIRF microscopy yields raw data as presented, that can be converted to FRET traces and thus providing FRET efficiencies and information about protein dynamics. For smFRET studies, it is crucial that the signal-to-noise ratio is sufficiently high to detect a single fluorescent molecule from the background. Because the evanescent field from the TIRF is restricted to a small volume, the signal-to-noise ratio is significantly improved than for other microscopes, and therefore yields the

required sensitivity for single molecule detection.[16] Consequently, the fluorescently labelled molecules must be tethered to the surface, so as they are located within the evanescent field.

1.4 Using ALEX to determine intermolecular distances

In order to resolve the conformational equilibrium changes of Cpf1, all experiments on the TIRF microscope were conducted using alternating laser excitation (ALEX). Using ALEX in smFRET studies has proven successful, as ALEX enables additional information and thus categorization of fluorescent traces by rapid switching between the two excitation wavelengths.[18] By doing so, it is possible to observe both fluorophores almost sequentially. Colocalizing donor and acceptor fluorescent signals in the raw image, reveal FRET events (Fig. 3). ALEX provides two fluorescence ratios; the FRET efficiency (E) and the stoichiometry (S) between the donor and acceptor fluorophores. The FRET efficiency is calculated from the background corrected intensity of donor and acceptor when exciting the donor, as previously described. The stoichiometry helps to ensure that a FRET signal arises from only one donor and one acceptor, and is calculated,

$$S = (I_A + I_D)/(I_A + I_D + I_A^*)$$

where I_A^* is the background corrected acceptor intensity when exciting the acceptor. Traces with only one acceptor and one donor are selected, defined by a bleaching step. Additionally, FRET measurements with ALEX allows for thermodynamic and kinetic analysis of the conformational changes.[19] The average number of background photons are calculated based on the uncorrected FRET traces, and subtracted from each channel separately.

Spectral cross-talk is usually a key problem in FRET studies using only one laser, since a part of the emitted donor photons are often detected in the acceptor-emission channel, due to spectral overlap. Therefore, the calculated energy transfer from donor to acceptor might be incorrect, thus resulting in wrong distances. ALEX provides a convenient way of accessing correction factors required for determining precise molecular distances.[20] Correction factors obtained are α and δ, to correct for blead-through between donor and acceptor channel and direct excitation of acceptor by donor-excitation laser, respectively. Differences in excitation intensities, quantum yields and detection efficiencies are accounted for by using the correction factor γ and β. Moreover, ALEX allows molecule sorting to exclude acceptor blinking and fluorophore bleaching.[21] α correction factor is calculated using background corrected FRET efficiency for donor-only traces (E^{DO}) and δ correction factor is determined from the stoichiometry of acceptor-only traces (S^{AO}),

$$\alpha = E^{DO} / (1 - E^{DO})$$
$$\delta = S^{AO} / (1 - S^{AO})$$

Incorporating α and δ correction factors allow calculation of the corrected acceptor fluorescence after donor excitation (F_A),

$$F_A = I_A - \alpha \times I_D - \delta \times I_A^*$$

The apparent FRET efficiency (E_{app}) and stoichiometry (S_{app}) is calculated, based on F_A. Correction factor γ is calculated as the normalization of fluorescence quantum yield and β is calculated as the normalization to equal excitation rates. In a homogenous approximation, γ and β can be determined by linearly fitting FRET populations to E_{app} and $1/S_{app}$ histograms, with y-intercept *a* and slope *b* as (Fig. 8A),

$$\beta = a + b - 1$$
$$\gamma = (a - 1)/(a + b - 1)$$

The corrected FRET efficiency and stoichiometry histogram is then calculated based on these four correction factors.[20] The expected FRET efficiency, <E> is deduced as the centre of a Gaussian fit to the corrected FRET efficiency. From the expected FRET efficiency, the distance is calculated as previously described.

Where single molecule measurements allow extraction of distinct heterogenetic behaviours from thousands of individual molecules, ensemble experiments yield the combined and hence average signal for millions of molecules. Even though this may mask distinct patterns within protein behaviour, the overall trend is observed - a trend that should be visible when comparing vast amounts of single molecule data with ensemble measurements. Hence a combination of both methods would allow validation and comparison of the achieved results.

1.5 Challenges when doing FRET

The major challenge in designing such FRET assays is to site-specifically label the enzyme with fluorophores, with high efficiency and precision to ensure that the reported FRET efficiency translates to the actual inter-dye distance. Cysteines are frequently used for site-specific labelling of proteins, because the thiol residues can react with a maleimide reagent (such as a fluorophore linker), resulting in a thioether.[20] Cpf1 was chemically modified, and all cysteines except three was removed.[7] The native surface exposed cysteines CYS1190 and CYS473 on the REC and NUC domains respectively, was used to monitor the conformational changes, linking the fluorophores to the protein. A third cysteine, C882, which is not solvent accessible, and hidden within the protein, could not be removed, because removal resulted in protein denaturation. Introducing multiple cysteines can contribute to protein misfolding due to non-native disulfide bridge formation, thus disrupt the DNA cleaving mechanism of Cpf1. Moreover, purification of labelled enzyme and separation from excess dye is crucial to ensure that reported FRET efficiencies are due to conformational changes and not to excess dyes in solution. Optimization of the purification

step can be done using e.g. size-exclusion chromatography to isolate enzyme from free dyes, however the yield might decrease drastically as the purity increases.[22]

Understanding the conformation and specific activity of proteins immobilized to surfaces are imperative, when doing smFRET experiments. Research has shown, that enzymes immobilized on the inert PLL-PEG/PEG-biotin surface remain fully active and that nonspecific adsorption is insignificant.[23]

1.6 Using dsDNA scaffolds as distance calibration

DNA labelled with Cy3 and Cy5 can act as a calibration control when doing ensemble and smFRET studies, as DNA can act as a rigid scaffold with no transitions. The 5' Cy3 is attached to the hydroxyl group of the ribose via a phosphodiester bond. The internal Cy5 dye is attached to the backbone via the phosphodiester bond of the base.[24] The primary amino group at the 3'-end can be used to attach a variety of modifiers, such as biotin for smFRET measurements. Moreover, labelling efficiency and purity are high compared to the enzyme. Additionally, DNA is inexpensive and fairly easy to work with, thus providing a technique to access the difficulties when doing FRET. The nature and dynamics of DNA is well-studied, and with its rigid scaffold and thus no transitions, dual labelled DNA has proven to be a great calibration of FRET efficiencies between donor and acceptor, and consequently intermolecular distance.[15]

2 Materials and Methods

For materials see supporting information.

2.1 Labelling of Cpf1

Labelled Cpf1 was provided by Montoya group. The enzyme is site-specifically labelled at positions CYS1190, CYS473 and CYS882, which can bind the fluorophores Cy3 and Cy5. The third labelling position CYS882 has shown not to have any effect on the FRET experiments, as it is hidden within the enzyme.[7] Unpublished (submitted) mass-spectroscopy data supports that the labelling efficiency on the third label is negligible. To model the actual distance between the two dyes, Monte Carlo simulations were performed. Simulations were done on crystal structures of Cpf1 (PDB ID: 5MGA), in FPS FRET positioning screening software.[25] The Monte Carlo simulations sample all possible dye-positions, and finds the average inter-dye distance (Fig. S3, A).

2.2 Ensemble FRET assay

2.2.1 Cpf1 experiment

1 µL 2.7 mg/mL Cpf1 (labelled with Cy3/Cy5) was divided into smaller eppendorf tubes. All fluorescence measurements were conducted at room temperature in reaction buffer (50 mM bicine, 150 mM KCl, pH = 8). The binary and ternary complexes between Cpf1, crRNA and DNA were assembled in reconstitution buffer by incubating Cpf1 with 200 nM RNA and 200 nM DNA for at least 10 min at 25 °C prior to measuring, which resulted in full complex formation. When RNA and/or DNA was present 5 mM $MgCl_2$ was added to the buffer before incubation.

2.2.2 dsDNA control experiment

Dual-labelled dsDNA (45 basepairs oligonucleotide) with Cy3/Cy5 was made from ssDNA/Cy3 (end-biotinylated) and ssDNA/Cy5 mixed in 1:1 molar ratio, to ensure high annealing efficiency. The oligonucleotides were annealing using the assigned annealing protocol.[26] The two dyes are separated by 17 basepairs DNA sequence, corresponding to a distance of 58.3 Å (measured using Chimera[27]). To model the actual distance between the two dyes, Monte Carlo simulations were performed on the crystal structure of dsDNA (Make-Na server[28]), in FPS FRET positioning screening software.[25] (Fig. S3, B).

Bulk fluorescence measurements were carried out on a T-format spectrofluorometer (Jasco FP-6200; Jasco, Easton, MD). For each FRET construct, a donor-only (Cy3-labelled) and acceptor only (Cy5-labelled) sample was prepared and its emission spectra at 530 nm and 640 nm excitation were collected, to correct for cross excitation an emission. Each experimental condition was measured at least 3 times for statistical robustness. Excitation at 532 nm uncovers spectra with emission peaks observed at 560 nm (donor) and 670 nm (acceptor). The FRET efficiency was calculated based on donor and acceptor intensity, for each experiment and thereby allow extraction of the distance between the fluorophores in each condition.

2.3 smFRET assay

2.3.1 dsDNA control experiment

The surface for smFRET experiments was made on passivated and cleaned cover glass[29] with a 1:100 mixture of 80 µL PLL-PEG and PLL-PEG-biotin. The surface was functionalized with 80 µL 0.1 g/L neutravidin. Prior to each experiment, 1 µM dsDNA was diluted 50,000 times and introduced to the microscope chamber and incubated for 2 min at 25 °C to ensure immobilization via biotin-neutravidin linker. Excess dsDNA was washed away with buffer (50 mM bicine, 150 mM KCl, 5 mM $MgCl_2$, pH = 8). An oxygen scavenging system / imaging buffer (1 U/mL PCD, 2.5 mM PCA, 2 mM Trolox) was prepared, and flushed into the chamber just before each measurement, to prevent fluorophore-blinking.[30]

All smFRET experiments were conducted on the TIRF microscope equipped with two EMCCD cameras. Donor (Cy3) and acceptor (Cy5) were excited using 532 nm (green) and 640 nm (red) solid state laser lines. A dual-cam setup was used to split the signal into two distinct channels. All experiments were conducted using alternating laser excitation (ALEX), with 100 ms exposure time, 300 EM gain and frame rate of 2 s^{-1}.[9,20]

2.3.2 Data treatment

Image analysis was performed using iSMS software package for smFRET, to extract traces for each co-localized donor/acceptor FRET pair. α and δ correction factors were extracted from iSMS.[31] Further data analysis to extract correction factors γ and β, and trace categorization was made on an in-house software written in Python by Johannes Thomsen.

3 Results and discussion

3.1 Ensemble FRET

All fluorescence ensemble measurements were obtained as intensity as a function of wavelength. The intensity is measured in arbitrary units, and therefore relative to each sample, hence the following results are normalized to 1.

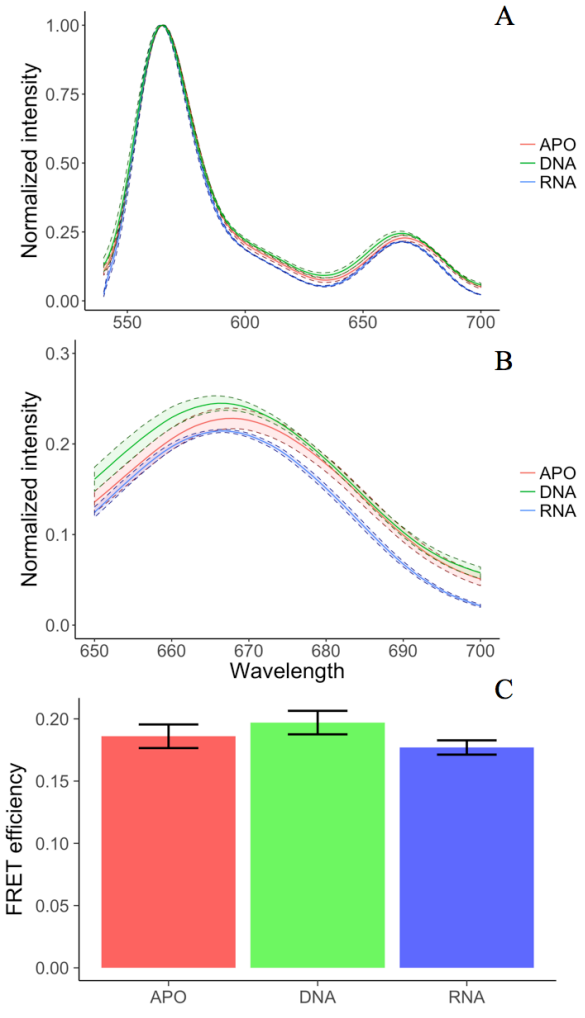

Figure 4: Results from ensemble FRET measurements on Cpf1. **A:** Normalized and corrected spectra for ensemble FRET measurements on free-form (apo) Cpf1, crRNA-bound and DNA-bound/DNA-cleaved. **B:** Zoom in of A in acceptor region. The spectrum indicates a transition from a more relaxed open conformation to a more compact conformation upon DNA binding. Excitation wavelength is 532 nm. **C:** Bar-chart of the FRET efficiencies.

Table 1: Ensemble FRET results for Cpf1

	Apo	crRNA	DNA
E_{FRET}	0.186 ± 0.009	0.177 ± 0.006	0.197 ± 0.009
R [Å]	71.64 ± 0.75	72.39 ± 0.48	70.80 ± 0.71

FRET efficiency and radius from ensemble FRET measurements on Cpf1 on free-form (apo), binary complex (crRNA) and tertiary complex (DNA). The results indicate that DNA bound form is in the most compact conformation. Errors are reported as S.D. of at least three measurements.

Exciting the donor fluorophore at 532 nm, and measuring emission intensity from 540 nm to 700 nm allows extraction of FRET efficiencies for each experimental condition (apo, crRNA/RNA and DNA). The acceptor intensity differs between apo, crRNA and DNA, hence the FRET efficiency and thereby distance varies as seen in table 1. Uncertainties in Fig. 4 are reported as the standard deviation from three-to-six experiments. The FRET efficiencies and distances were extracted as previously described, and the error determined with error propagation. The resulting FRET efficiency and distance can be seen in table 1. The ensemble approach allows investigation of the multiple conformations of Cpf1, thus recording conformational equilibrium shifts. Each sample was incubated for at least 10 minutes before measuring to ensure equilibrium between different conformational states, and consequently the reported distance will be a sum of distances for all possible conformations.

To quantify if distances are significantly different, I performed a Welch's t-test for unequal variance. This test assumes normality, and compares central tendency for two populations.[32] Performing the test with a significance threshold of $\alpha = 0.05$ revealed only one significant transition (Table S1.), showing that the conformational change, when DNA is added to the sample, is significant, with a p-value 3.37e-04. Thus, binding of DNA results in conformational changes of Cpf1, and the enzyme transits to a more closed state, with a 15 % increase in acceptor intensity compared to crRNA-bound. The inter-dye distance for the tertiary DNA-bound/DNA-cleaved state was found to be 70.80 ± 0.71 Å. I then went on to use an orthogonal method to further evaluate the distances using Monte Carlo (MC) simulations on crystal structure of Cpf1 after DNA-cleavage, which reveals an average dye-dye distance of 68.2 Å (Fig. S3, A), in good agreement with the experimental results, and thus I could further back the hypothesis from FRET measurement. The difference between the DNA-bound and the DNA-cleaved state cannot be distinguished in ensemble measurements, as the distinct states are masked when averaging. However, since the samples were incubated, I hypothesize that conformational equilibrium was reached prior to measuring. The distance of 70.80 ± 0.71 Å corresponds to an open conformation, thus represents the DNA unwinding and the DNA-cleaved state of Cpf1, and indicates that binding of DNA shifts the conformational equilibrium of

Cpf1 compared to crRNA-bound conformation. The results from table 1 indicate that the enzyme upon binding crRNA shifts the conformational equilibrium towards a more open conformational state, with a greater distance on average, although this difference was found not to be significant (Table S1.). It seems that the different conformational states are masked in ensemble FRET, making it impossible to distinguish between apo-form and crRNA-bound conformation. The low-FRET state in the apo protein (E_{FRET} = 0.186 ± 0.009) indicates the large conformational landscape of Cpf1 in the absence of crRNA and DNA and is in agreement with the hypothesized large conformational heterogeneity of the apo protein.

3.2 Evaluation of FRET

In order to ensure that the changes in FRET efficiencies arise from conformational changes and therefore converted accurately to distance, a calibration-control experiment using dual-labelled dsDNA with known inter-dye distance was performed. DNA is very rigid, hence a very prevailing measure for inter-dye distance in FRET experiments. The experiment was executed equivalently, and the emission spectrum recorded (Fig. S4). The hypothesis was further supported by MC simulations to extract dye-dye distances in the dual-labelled dsDNA. Different linker sequences can shift the distance and orientation of the fluorescent protein, causing changes in FRET efficiencies, and accentuating the importance in knowing the exact dye-dye distance.[33] MC simulations revealed a mean inter-dye distance of 61.6 Å (Fig. S3, B).

Converting FRET efficiencies into distance resulted in an inter-dye distance of 72.53 ± 0.46 Å, greater than the simulated dye-dye distance of 61.6 Å extracted from MC simulations, indicating that either the DNA-strands were annealed uncomplete, or that some additional fluorescent contribution is present when measuring.

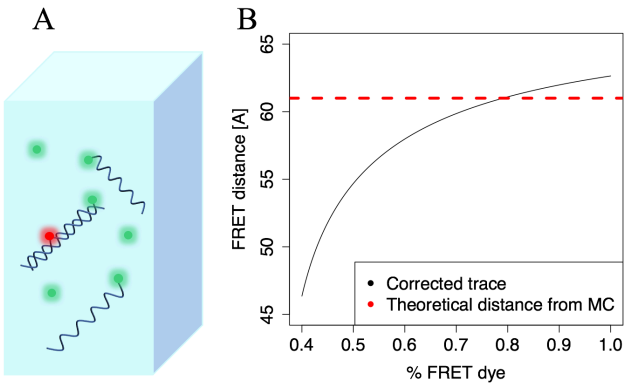

Figure 5: Correction for excess dye. **A:** Different types of dyes in solution. 1) Un-annealed ssDNA with Cy3-fluorophore, 2) free (unbound) Cy3-fluorophores and 3) dual-labelled dsDNA (so-called "FRET" dye) **B:** Correction for excess dye. The black line is the calculated distance assuming a given ratio of "FRET" dye (the percentage that actual contributes to the FRET efficiency and not due to excess dye). Intercept (red line = simulated distance) indicates the proportion of excess dye in solution that would correspond to the simulated distance of 61.6 Å.

According to the annealing protocol[26], the annealing is expected to be sufficiently high, so as the signal should not be disturbed from un-annealed ssDNA labelled with fluorophores.

Excess Cy3-fluorophores (Cy3) in solution will result in an increased donor intensity, hence less FRET efficiency and longer distance. This contribution to donor intensity could arise from unbound and thus free Cy3, or it could be un-annealed ssDNA with Cy3 (Fig. 5, A). Excess Cy5-fluorophores should not contribute to the intensity, as only laser excitation at 532 nm occurs. Using the equation:

$$I_D^{measured} = P_{cy3} \times I_{cy3} + P_{FRET} \times I_D^{actual}$$

it is possible to correct for excess dye in solution. I hypothesise that $I_D^{measured}$ (the measured donor intensity for dual-labelled dsDNA) corresponds to a sum of the contribution from excess dye, including unbound and un-annealed dyes ($P_{cy3} \times I_{cy3}$), and the contribution from actual FRET signal ($P_{FRET} \times I_D^{actual}$). Based on the equation, the actual donor intensity arising from FRET (I_D^{actual}) can be calculated. Calculating E_{FRET} and distance from I_D^{actual} instead of I_D allow to correct for excess dye (Fig. S5). Using the above equation, our findings suggest approximately 20 % excess dye in solution and thus 80 % "FRET" dye (Fig. 5, B). Consequently, having excess dye in solution whilst measuring FRET, may reveal the overall trend in the conformational changes, but fails to translate correctly into inter-dye distance, hence providing an incorrect depiction of the intermolecular distances of the different enzymatic states. Further on, excess dye could mask the smallest and most sensitive conformational equilibrium changes of Cpf1 in the ensemble approach.

These findings may provide an explanation as to why the emission spectrum for apo, crRNA- and DNA bound in Fig. 4 are very similar. I hypothesize that excess dye in solution contributes to the donor signal, masking the distinct difference between the different conformational states of Cpf1 in ensemble FRET measurements. It is impossible to differentiate between the relative population of multiple distributions and the shift of the distribution as a whole in ensemble measurements, consequently preventing extraction of the distinctive conformational equilibrium changes and corresponding intermolecular distances.

3.3 smFRET

As a setup for additional experiments with Cpf1, single molecule FRET is ideal to uncover distinct conformational changes otherwise masked in ensemble approach. To further understand the system, and calibrate for excess dyes in solution, single molecule FRET measurements on dsDNA were done. Dual-labelled dsDNA was passivated on a PLL-PEG surface using neutravidin-biotin linker (Fig. 6). Flushing with buffer ensures that only tethered DNA is present in the sample, thus removing unbound dyes and DNA.

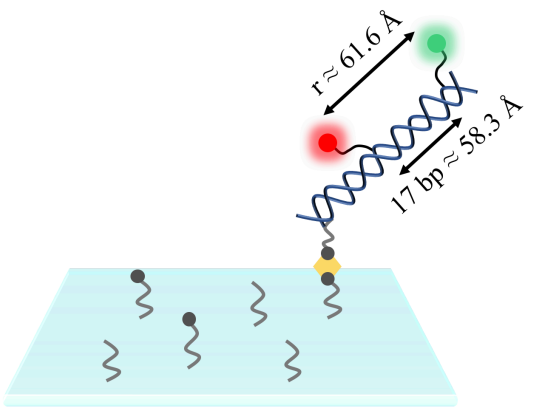

Figure 6: Experimental setup for the single molecule TIRF experiments on dual-labelled dsDNA. DNA is tethered to the PLL-PEG surface through biotin-neutravidin linker. The fluorophore linkers are separated by 17 bp corresponding to 58.3 Å. And the distance between the attached fluorophores is 61.6 Å found from MC.

smFRET experiments were conducted on a TIRF microscope with a dual-cam setup using ALEX. smFRET trajectories of hundreds of individual dsDNA molecules in steady state conditions were captured, and data treatment was finalized as presented earlier.

Counting each Cy3 and Cy5 in all raw movies, revealed 11.197 traces consisting of colocalized Cy3 and Cy5 (table S2). Comparing this number with the total number of Cy3 revealed 60.68 % non-colocalized Cy3. This number is even greater than the 20 % excess dye found in the distance calibration for ensemble experiments (Fig. 5, B). Surprisingly, the raw data showed equal proportions of Cy3 and Cy5. These findings indicate that Cy3-labelled ssDNA (end-biotinylated) and/or single-labelled dsDNA is tethered to the surface (Fig. S2). Emphasizing that labelling efficiency and annealing should be considered an important factor when doing ensemble measurements, as it can highly interfere with the results.

Using iSMS the raw traces were converted into distinct FRET traces (~ 200 traces) (Fig. 7). Further classification reveal 66 evident FRET traces. The traces are primarily categorized based on presence of a single bleaching step as well as a stoichiometry at 0.5, indicating a single FRET pair. A typical FRET trace can be seen in Fig. 7, where the green and red signal are cross-correlated, demonstrating that decrease in acceptor intensity leads to donor intensity increasing. For additional traces see Fig. S6.

The acceptor excitation channel verifies bleaching event for the acceptor, thus resulting in increasing donor intensity. The acceptor bleaching in combination with the donor bleaching and the stoichiometry reveals a certain FRET signal. Other factors taken into account are the standard deviation and median of stoichiometry, intensities and FRET efficiency. These guidelines are based on prior expertise in looking at FRET traces. The corrected FRET efficiency histogram (Fig. S7, B) is determined based on

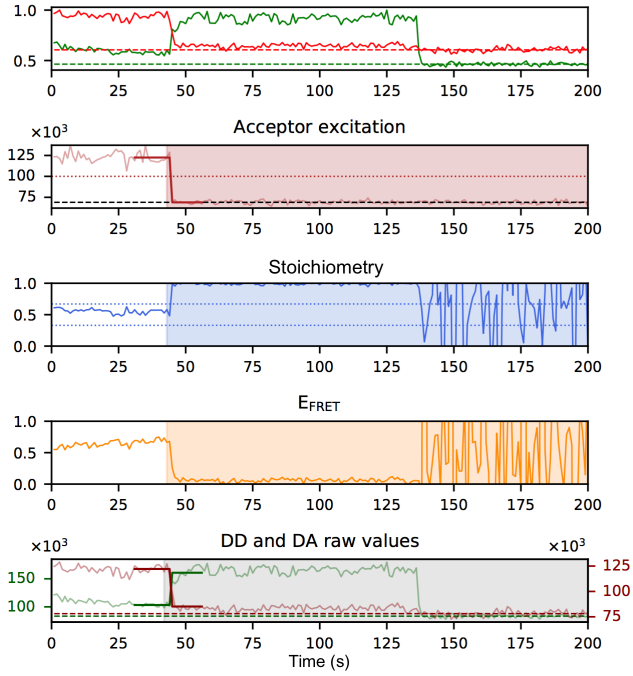

Figure 7: Representative single molecule trace using ALEX single molecule FRET measurements. ALEX provides a read-out for both the green and red channel, allowing for simultaneously observing both donor and acceptor. As the acceptor bleaches in a single bleaching step, the donor intensity increases, hence validating the trace as actual FRET with only a single FRET pair (S = 0.5). As the donor bleaches, the acceptor intensity is unchanged, indicating that both fluorophores are bleached.

the 66 FRET traces, from which the expected FRET efficiency is computed, and later converted to distance.

Correction factors were extracted from the 66 FRET traces (Fig. 8A). The α- and δ-correction factor is determined through iSMS, based on the baseline of each trace. The background can be set manually in iSMS, such that α and δ correct for blead-through and direct excitation. Additionally, β- and γ-correction factors are calculated as previously described, correcting for detection and excitation (table 2, Fig. 8A). Taking the correction factors into account, allow calculation of the expected FRET efficiency as the mean of a Gaussian fit to the non-binned FRET efficiency data, and thus distance is found (Fig. 8B). Although FRET efficiencies are not theoretically normally distributed, it has in practice been shown to be a robust method, with little discrepancy.[34]

Table 2: Correction factors

α	δ	β	γ
0.060	0.020	2.317	0.644

Correction factors from smFRET analysis on dual-labelled dsDNA. The correction factors correct for blead-through, direct excitation, differences in excitation intensities, quantum yield and detection efficiencies.

Fitting the probability density with a normal distribution results in a reduced chi-squared statistics $\chi^2/\nu = 2.27$. The reduced chi-square statistics is defined as chi-squared per degree of freedom.[35] Given the found chi-square the Gaussian fit seems to explain the data well. The residuals follow a normal distribution (Fig. S8) as expected from the central limit theorem, with error $\sigma = 0.17$, and mean $\mu = 0.00$. The standard error of the mean (SEM) for the distance is calculated to be 0.93 Å, which demonstrations that the calculated mean is in close proximity to the true mean of the population.[36] Our single molecule data reveals 0.34 ± 0.02 FRET, which predicts an inter-dye distance of 62.72 ± 0.93 Å (Fig. 8B), in great agreement with theory proposing an inter-dye distance of 61.6 Å from MC.

Individual smFRET trajectories are analyzed by hidden Markov modelling[37] (Fig. 9), to investigate transitions and revealed only a single populated state with no transitions, as expected due to the rigidity of the DNA scaffold that was investigated.

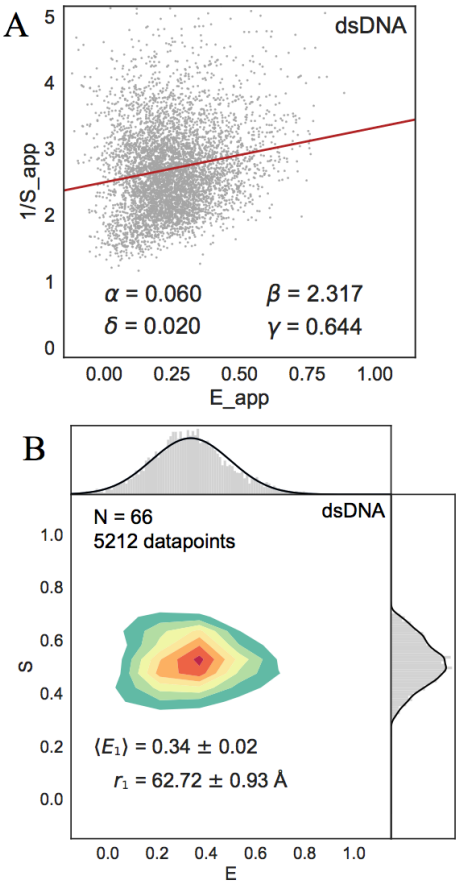

Figure 8: Results from the smFRET experiment **A:** Correction factors β and γ determined by linearly fitting E_{app} to $1/S_{app}$ histograms. **B:** Gaussian distributions were used to fit the histograms to identify individual populations of E_{FRET} states. Using the four correction factors to determine the FRET efficiency reveals the corrected two dimensional histogram showing stoichiometry and FRET efficiency and thus the expected efficiency and distance with standard error of the mean.

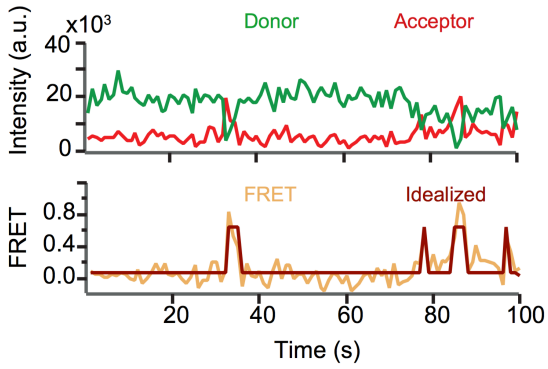

Figure 9: Individual smFRET traces fitted with hidden Markov modelling (HMM). The raw intensity traces for donor and acceptor can be seen at the top. The intensities are only shown until a bleaching event occurs. At the bottom is the calculated FRET trace, as well as the idealized (modelled from HMM). Each frame corresponds to one data point. This trace is from smFRET studies on Cpf1 (done by Johannes Thomsen and Simon Bo Jensen), showing multiple transitions.[7] Traces from dsDNA smFRET measurements show only one state, and no transitions as expected.

The stoichiometry populates at around 0.5 indicating that only a single FRET pair is present. The smFRET results are concluded based on 66 traces and 5212 data points in total, and thus considered statistical robust. Furthermore, the results illustrate that correction factors can be extracted, and consequently FRET efficiency and distance determined with great accuracy. This method can easily be translated to investigation of conformational changes of Cpf1, using the experimental setup shown in Fig. 3. smFRET analysis would provide further knowledge to the conformational equilibrium shifts of Cpf1, without the problem of excess dye, allowing for extraction of intramolecular distances at each conformational state.

In conclusion, FRET can be used as a spectroscopic ruler to report distance related conformational changes of molecules. The ensemble FRET measurements provided an easy way to access information about the conformational changes of Cpf1, showing that only the DNA-bound state is significantly different from the others. A control with dual-labelled dsDNA revealed that excess dye in solution complicates the ensemble FRET measurements, and thus may interfere with the results. smFRET experiments on the same dual-labelled dsDNA allowed us to extract dye-dye distance of 62.72 ± 0.93 Å in close proximity to MC simulated distance of 61.6 Å, thus demonstrating smFRET as a precise method to reveal conformational changes, and thus extract intermolecular distances without interference of excess dye.

3.2 Future perspectives

Detailed quantification of the conformational changes of Cpf1 was not carried out in the smFRET assay within the time frame of this project, especially since the microscope was not working the last two months of the project. However, could be implemented, using the described assay above (Fig. 3). Additional information about the mechanism of DNA binding and cleavage could be achieved with three-color FRET imaging. Labelling the target-DNA with an additional fluorophore, would allow single molecule tracking to ensure that the DNA is bound and consequently when it is cleaved. This would provide further knowledge to the mechanism by which Cpf1 binds and cleaves DNA, as the DNA-bound and DNA-cleaved state can now be differentiated. Understanding, how the DNA binds and ultimately when it is cleaved provide knowledge about the structure and function of Cpf1 and thus adds another dimension to our understanding of modern genome editing.

4 Conclusion

Investigations of the conformational changes of Cpf1 using Förster Resonance Energy Transfer to quantify the intramolecular dynamics were carried out. Firstly, ensemble FRET measurements of the conformational changes of Cpf1 upon adding crRNA and DNA were executed. Cpf1 was labelled with Cy3 and Cy5 at the two hinge regions (NUC and REC lobe). The assay revealed a significant conformational change when adding DNA, translating into a dye-dye distance of 70.80 ± 0.71 Å. This corresponds to the MC simulated distance of 68.2 Å, showing that the enzyme is in a DNA-bound/DNA-cleaved state. However, in ensemble assay it was not possible to directly differentiate between DNA-bound and DNA-cleaved conformation. Further examination of dual-labelled dsDNA as a rigid scaffold displayed approximately 20 % excess dye in solution, revealing difficulties with the ensemble assay. smFRET based approach was utilized to directly observe and quantify the conformation of dsDNA, and eliminated the role of excess dye. smFRET assay further supported that labelling efficiency and annealing of dsDNA should be considered an important factor in ensemble measurements. The inter-dye distance of dual-labelled dsDNA was proven to be 62.72 ± 0.93 Å – which resembles the simulated distance of 61.6 Å modelled from Monte Carlo simulations. Only one populated state with no transitions was found using hidden Markov model analysis. In this thesis I show, that FRET is a confident method to extract intermolecular distances, thereby could prove novel information about the conformational states of different biomolecules. This single molecule assay could be expanded to study the conformational states of Cpf1.

5 Acknowledgement

First, I would like to express my gratitude to my project supervisor, Associate Professor Nikos Hatzakis, Nanoscience-Center and Department of Chemistry, University of Copenhagen, for giving me the opportunity to be a part of his group. Special thanks, to our collaborators, Montoya group from the Panum Institute for providing the enzymes. Special thanks to Ph.d-student Simon Bo Jensen for introducing me to the laboratory and for assisting me on the microscope. My gratitude goes to Ph.d-student Johannes Thomsen for inviting me into his data-treatment universe, and lending me his script to sort and analyze smFRET traces. Special thanks to Ph.d-student Søren Schmidt-Rasmussen Nielsen, for supporting me all the way through, and always believing in me. Lastly, I would like to thank the entire Hatzakis Nano-enzyme-group for a lovely and fun working environment.

6 References

[a] Thorvaldsensvej 40, 1871 Frederiksberg C, University of Copenhagen, Denmark.; E-mail: frejabohr@hotmail.com
[b] T554, Thorvaldsensvej 40, 1871 Frederiksberg C, University of Copenhagen, Denmark.

1. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550,** 407–410 (2017).
2. Booth, C., Gaspar, H. B. & Thrasher, A. J. Treating Immunodeficiency through HSC Gene Therapy. *Trends Mol. Med.* **22,** 317–327 (2016).
3. Schwank, G. *et al.* Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13,** 653–658 (2013).
4. Kampmann, M. A CRISPR Approach to Neurodegenerative Diseases. *Trends Mol. Med.* **23,** 483–485 (2017).
5. Stella, S., Alcón, P. & Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **546,** 559–563 (2017).
6. Fonfara, I., Richter, H., BratoviÄ, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532,** 517–521 (2016).
7. Montoya, G., Hatzakis, N. & et. al. Conformational Activation of CRISPR-Cpf1 Catalysis and Endonuclease Function Recycling. *Submitted* (2018).
8. Fran, R. Real-time observation of DNA target interrogation and product release by the RNA-guided endonuclease CRISPR Cpf1 Digvijay. 443–447 (2005). doi:10.1016/B978-0-12-397169-2.00038-X
9. Rahul, R., Hohng, S. & Ha, T. A Practical Guide to Single Molecule FRET. *Nat. Methods* **5,** 507–516 (2008).
10. Sekar, R. B. & Periasamy, A. Fluorescence resonance energy transfer (FRET) microscopy imaging of live cell protein localizations. *J. Cell Biol.* **160,** 629–633 (2003).
11. Koch, A. *et al.* High-precision FRET analysis of the G-protein coupled receptor TGR5 in live cells. *Eur. J. Med. Res.* **19,** S12 (2014).
12. Iqbal, A. *et al.* Orientation dependence in fluorescent energy transfer between Cy3 and Cy5 terminally attached to double-stranded nucleic acids. *Proc. Natl. Acad. Sci.* **105,** 11176–11181 (2008).
13. Ma, L., Yang, F. & Zheng, J. Application of fluorescence resonance energy transfer in protein studies. *J. Mol. Struct.* **1077,** 87–100 (2014).
14. Hussain, S. An Introduction to Fluorescence Resonance Energy Transfer (FRET). *arXiv Prepr. arXiv0908.1815* (2009).
15. Di Fiori, N. & Meller, A. The Effect of dye-dye interactions on the spatial resolution of single-molecule FRET measurements in nucleic acids. *Biophys. J.* **98,** 2265–2272 (2010).
16. Tinoco, I., Gonzalez, R. L., Jr, I. T. & Jr, R. L. G. Biological mechanisms , one molecule at a time. 1205–1231 (2011). doi:10.1101/gad.2050011
17. Bavishi, K., Li, D., Eiersholt, S., Hooley, E. N. & Petersen, T. C. Direct observation of multiple conformational states in Cytochrome P450 oxidoreductase and their modulation by membrane environment and ionic strength. 1–9 (2018). doi:10.1038/s41598-018-24922-x
18. Kapanidis, A. N. *et al.* Alternating-laser excitation of single molecules. *Acc. Chem. Res.* **38,** 523–533 (2005).
19. Kapanidis, A. N. *et al.* Fluorescence-aided molecule sorting: Analysis of structure and interactions by alternating-laser excitation of single molecules. *Proc. Natl. Acad. Sci.* **101,** 8936–8941 (2004).
20. Hellenkamp, B. *et al.* Precision and accuracy of single-molecule FRET measurements - a worldwide benchmark study. 1–27 (2017).
21. Hohlbein, J., Craggs, T. D. & Cordes, T. Alternating-laser excitation: single-molecule FRET and beyond. *Chem. Soc. Rev.* **43,** 1156–1171 (2014).
22. van Vught, R., Pieters, R. J. & Breukink, E. Site-Specific Functionalization of Proteins and Their Applications To Therapeutic Antibodies. *Comput. Struct. Biotechnol. J.* **9,** e201402001 (2014).
23. Zhen, G. *et al.* Immobilization of the enzyme β-Lactamase on biotin-derivatized poly(L-lysine) -g-poly (ethylene glycol)-Coated sensor chips: A study on oriented attachment and surface activity by enzyme kinetics and in situ optical sensing. *Langmuir* **20,** 10464–10473 (2004).
24. Integrated DNA Technologies. *https://eu.idtdna.com/pages/Support/FAQs/where-on-a-nucleotide-are-cy-dyes-attached-*
25. Kalinin, S. *et al.* A toolkit and benchmark study for FRET-restrained high-precision structural modeling. *Nat. Methods* **9,** 1218–1225 (2012).
26. Bioneer. Anneal complementary pairs of oligonucleotides. *Tech. Tip* **0747,** 1–2 (2009).
27. Pettersen, E. F. *et al.* UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).
28. Stroud, J. Make-Na server. (2004).
29. Moses, M. E., Hedegård, P. & Hatzakis, N. S. Quantification of Functional Dynamics of Membrane Proteins Reconstituted in Nanodiscs Membranes by Single Turnover Functional Readout. *Methods Enzymol.* **581,** 227–256 (2016).
30. Vogelsang, J. *et al.* A reducing and oxidizing system minimizes photobleaching and blinking of fluorescent dyes. *Angew. Chemie - Int. Ed.* **47,** 5465–5469 (2008).
31. Preus, S., Noer, S. L., Hildebrandt, L. L., Gudnason, D. & Birkedal, V. ISMS: Single-molecule FRET microscopy software. *Nat. Methods* **12,** 593–594 (2015).
32. Kim, T. K. T-test as a Parametric Statistic. *Korean J. Anesthesiol.* **68,** 540–546 (2015).
33. Bajar, B. T., Wang, E. S., Zhang, S., Lin, M. Z. & Chu, J. A guide to fluorescent protein FRET pairs. *Sensors (Switzerland)* **16,** 1–24 (2016).
34. Yang, M. *et al.* The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. *Cell Rep.* **22,** 372–382 (2018).
35. Andrae, R., Schulze-Hartung, T. & Melchior, P. Dos and don'ts of reduced chi-squared. 1–12 (2010). doi:10.1088/0004-637X/733/1/66
36. Nagele, P. Misuse of standard error of the mean (SEM) when reporting variability of a sample. A critical evaluation of four anaesthesia journals. *Br. J. Anaesth.* **90,** 514–516 (2003).
37. McKinney, S. A., Joo, C. & Ha, T. Analysis of single-molecule FRET trajectories using hidden Markov modeling. *Biophys. J.* **91,** 1941–1951 (2006).