Spatiotemporal localization of proteins in microorganisms via photoactivated localization microscopy

Dissertation

zur Erlangung des Doktorgrades der Naturwissenschaften

(Dr. rer. Nat.)

an der Fakultät für Biologie der

Ludwig-Maximilians-Universität München

Vorgelegt von

Giacomo Giacomelli

München, Oktober 2020

Die vorliegende Doktorarbeit wurde im Zeitraum von Oktober 2014 bis December 2020 in der Arbeitsgruppe bakterielle Zellbiologie (Prof. Dr. Marc Bramkamp) and der Ludwig-Maximilians-Universität München durchgeführt.

Erstgutachter/in:	Prof. Dr. Marc Bramkamp
Zweitgutachter/in:	Prof. Dr. Andreas Klingl

Tag der Abgabe: 26.10.2020

Tag der mündlichen Prüfung: 12.01.2021

Eidesstattliche Erklärung:

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfsmittel angefertigt worden ist. Die vorliegende Dissertation wurde weder ganz, noch in wesentlichen Teiler einer anderen Prüfungskommission vorgelegt. Ich habe noch zu keinem früheren Zeitpunkt versucht, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

I hereby confirm that I have written the accompanying thesis by myself without contribution from any sources other than those cited in the text. Moreover, I declare that I have not submitted or defended a dissertation previously without success. This thesis has not been presented to any other examining board.

Kiel, den 20.01.2021

Giacomo Giacomelli

“Tout le monde y croit cependant, me disait un jour M. Lippmann, car les expérimentateurs s'imaginent que c'est un théorème de mathématiques, et les mathématiciens que c'est un fait expérimental.”

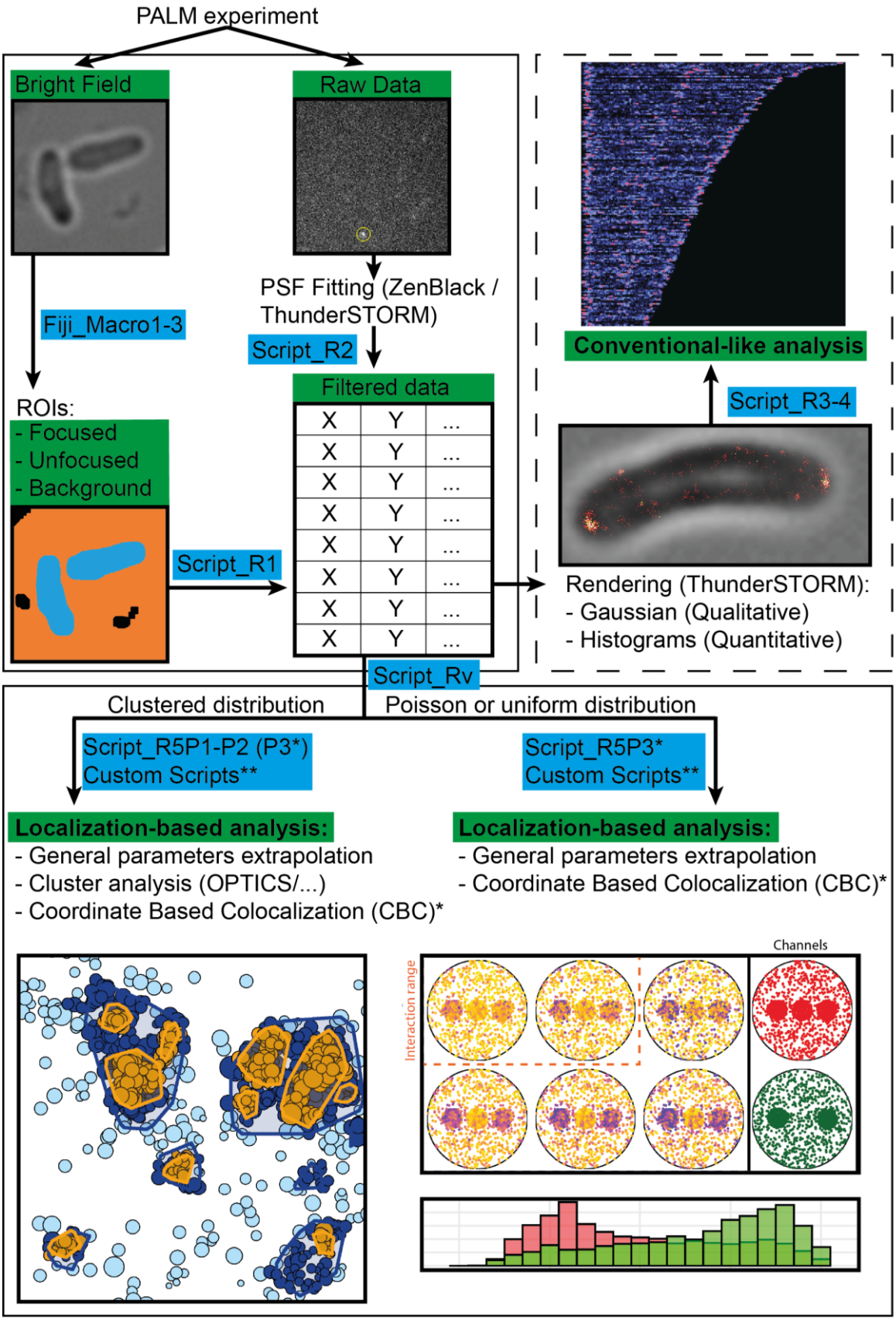
- Henri Poincaré, Calcul des probabilités (2nd ed., 1912), p. 171.

“Everybody believes in the exponential law of errors [i.e., the Normal distribution]: the experimenters, because they think it can be proved by mathematics; and the mathematicians, because they believe it has been established by observation.”

- Whittaker, E. T. and Robinson, G. "Normal Frequency Distribution." Ch. 8 in The Calculus of Observations: A Treatise on Numerical Mathematics, 4th ed. New York: Dover, pp. 164-208, 1967. p. 179.

Table of contents:

Graphical Abstract



*dual colour only
**case by case scenario

Abstract

Photoactivated localization microscopy (PALM) is a single molecule fluorescence microscopy technique (SMLM) that relies on the controlled activation and imaging of photo-activatable/convertible fluorescent proteins to determine their position with nanometer scale precision. The analysis of SMLM data is composed of two sequential aspects: the generation of a super-resolution table/image and the subsequent analysis. In recent years, several data analysis packages dedicated to the generation of super-resolved images have been developed. These packages have been extensively characterized and compared in a community-wide effort, therefore allowing researchers to identify optimal solutions for their experiments and providing software developers with a gold standard. On the contrary, the development of data analysis packages dedicated to the study of protein coordinates has been lagging behind, and no comprehensive approach has been developed to date.

Here, I present a combination of Fiji and R based scripts for the characterization, filtering and quality assurance of SMLM derived localizations.

Furthermore, I demonstrate that specific conventional image analysis techniques can be applied, both quantitatively and qualitatively, to super resolution images. I then apply these analysis tools exemplarily on the characterization of the spatio-temporal localization of a novel DNA repair system in *Corynebacterium glutamicum*, termed Dip (DNA damage induced protein) C.

Finally, I combine the multiple data analysis packages that I developed and/or adapted for the study of specific biological scenarios into a single cohesive pipeline, therefore providing a generalized and comprehensive approach toward the coordinate based analysis of the spatio-temporal localization of proteins in PALM and, in general, in SMLM. Each of the data analysis packages that comprise the pipeline is here presented together with the biological scenario that prompted its development. These include the study of magnetosome formation in *Magnetospirillum gryphiswaldense*, the study of the chromosome segregation machinery in *C. glutamicum* and the study of flagellar organization in *Trypanosoma brucei.*

Zusammenfassung

Die photoaktivierte Lokalisationsmikroskopie (PALM) ist eine Einzelmolekül-Fluoreszenzmikroskopie Technik (SMLM), die auf der kontrollierten Aktivierung und Aufnahme von photoaktivierbaren / konvertierbaren fluoreszierenden Proteinen beruht, um ihre Position mit einer Präzision im Nanometerbereich zu bestimmen. Die Analyse von SMLM-Daten besteht aus zwei aufeinander folgenden Aspekten: der Erzeugung einer Tabelle / eines hochauflösenden Bildes und der anschließenden Analyse. In den letzten Jahren wurden mehrere Datenanalysepakete entwickelt, die sich der Berechnung der hochaufgelösten Bilder widmen. Diese Pakete wurden in gemeinschaftsweiten Anstrengungen umfassend charakterisiert und verglichen, sodass Forscher eine optimale Lösung für eigene Experimente wählen können, während Softwareentwicklern einen Goldstandard zur Hand haben. Gegensätzlich wurde jedoch die Entwicklung von Datenanalysepaketen zur spezifischen Untersuchung von Proteinkoordinaten vernachlässigt, so dass in diesem Bereich keine umfassenden Instrumente existieren.

In dieser Arbeit präsentiere ich eine Kombination aus Fiji- und R basierten Skripten zur Charakterisierung, Filterung und Qualitätssicherung von SMLM Proteinkoordinaten.

Darüber hinaus zeige ich, dass bestimmte konventionelle Bildanalysetechniken sowohl quantitativ als auch qualitativ auf „Superresolution“ Bilder angewandt werden können. Im Folgenden verwende Ich diese Analysewerkzeuge dann beispielhaft zur Charakterisierung der räumlich-zeitlichen Lokalisierung eines neuartigen DNA-Reparatursystems in *Corynebacterium glutamicum*, welches ich DipC (DNA-Schaden-induziertes Protein) genannt habe.

Schließlich kombiniere ich die genannten Datenanalysepakete, die ich für die Untersuchung spezifischer biologischer Szenarien entwickelt und / oder angepasst habe, zu einer einzigen zusammenhängenden Arbeitsroutine. Diese bietet einen allgemeinen und umfassenden Ansatz für die koordinatenbasierte Analyse der räumlich-zeitlichen Lokalisierung von Proteinen aus PALM- und im Allgemeinen aus SMLM-Experimenten. Jedes der Datenanalysepakete, die in beschriebener Routine enthalten sind, wird hier zusammen mit dem biologischen Szenario vorgestellt, das zu ihrer Entwicklung geführt hat. Dazu gehören die Untersuchung der Magnetosomenbildung in *Magnetospirillum gryphiswaldense*, die Untersuchung der Chromosomensegregationsmaschinerie in *C. glutamicum* und die Untersuchung der Flagellenorganisation in *Trypanosoma brucei*.

Abbreviations (Alphabetical order):

BIC: Bayesian information criterion
BMC: Bacteria microcompartments
CM: Cytoplasmic mycelles
dil: Dioctadecylindocarbocyanine
FRAP: Fluorescence recovery after photobleaching
FRC: Fourier ring correlation
FREM: Fundamental resolution measure
HaloTag: Haloalkane dehalogenase tag
IMV: Intracellular membrane vescicles
LDL: Low-density lipoproteins
MC: magnetosome chain
MINFLUX: Minimal photon fluxes
mNeonGreen: monomeric NeonGreen
NA: Numerical aperture
OPTICS: Ordering points to identify the clustering structure
PAINT: Points accumulation for imaging nanoscale topography
PALM: Photoactivated localization microscopy
PA-mCherry: Photo activatable monomeric Cherry
PPA: Point pattern analysis
PSF: Point spread function
ROI: region of interest
SIM: Structured illumination microscopy
SIMFLUX: Method that combines centroid estimation and illumination pattern induced photon count variations in a conventional widefield imaging setup to extract position information over a typical micrometer-sized field of view (not a proper acronym, the term combines the SIM and MINFLUX acronyms)
SMLM: Single molecule localization microscopy
SNAP tag: Monomeric human O6-alkylguanine-DNA-alkyltransferase
SOMAmers: Slow Off-rate Modified Aptamers
SPT: Single particle tracking
STORM: Stochastic optical reconstruction microscopy

Chapter 1: Introduction

1.1. History of microscopy and its consequences on microbiology

The origin of microscopy, intended as the use of magnifying lenses to observe microbes, goes back to a period between 1665 and 1683, when Robert Hooke and Antoni van Leeuwenhoek were able to observe and describe microscopic organisms via the use of simple microscopes that magnified objects between 25 and 250 fold. This technological advancement allowed for the identification of microbes as the cause of infectious diseases (Leeuwenhoek 1695/1697) and resulted in a wave of high expectations that can be summarized by the 1665 statement made by Robert Hooke himself: "by the help of microscopes, there is nothing too small, as to escape our inquiry" (Hooke 1665). In the next centuries, thanks to technical advancements in the construction of light microscopes, Robert Koch was able to demonstrate that specific microbes are responsible for specific diseases (Koch 1912), while staining techniques such as Gram (1884) and halum hematoxylin (1886) staining helped to move beyond an exclusively morphological description of the observed microorganisms. It also became clear that Hooke had been too optimistic and that light microscopy resolution was insufficient for the observation of smaller biological structures such as viruses and most cellular organelles. The discovery and imaging of said biological structures became possible only following the development of electron microscopy. Specifically, Helmut Ruska visualized viruses via electron microscopy for the first time in 1940 (Ruska 1940), while most cellular organelles would be discovered even later.

Around the same time (1938), Edouard Chatton proposed for the first time to divide living beings into two major groups based on the presence/absence of the nuclear cell membrane, therefore forging the terms prokaryotes (primordial nucleus) and eukaryotes (good nucleus). This categorization became accepted in the early 1960s in no small part thanks to electron microscopy and genetics (Stanier and Van Niel 1962, Sapp 2005).

By the time the division between eukaryotes and prokaryotes had been accepted, it was already known that the presence of intracellular organelles was not an exclusivity of eukaryotic cells. Gas vacuoles in cyanobacteria had been observed already in 1895 by Heinrich Klebahn (Klebahn 1895), while bacterial microcompartments (BMC) had been imaged in 1956 by Dews and Nikowitz (Drews and Niklowitz 1956). Nevertheless, where eukaryotic cells were characterized by complex cell architecture including membrane-bound organelles and cytoskeleton, prokaryotes were smaller and generally characterized by an apparent lack of such complexity.

It does therefore not come as a surprise that, even decades later, prokaryotes would often be described as simple "bags of enzymes" where enzymes float randomly within the cell and act independently from each other.

In the following years electron microscopy contributed toward the identification of several new membrane-enclosed prokaryotic organelles such as cytoplasmic mycelles (CM), the intracellular membrane vescicles (IMV) that form upon overproduction of certain membrane proteins in *Escherichia coli* (Arechaga et al. 2000, Aboulwafa and Saier 2011), the membrane encapsulated magnets that compose the magnetosomes in magnetotactic bacteria (Gorby et al. 1988) and more. Each single one of these discoveries contributed toward the clearing of the "bags of enzymes" misconception, however, it is not until the Green Fluorescent Protein (GFP) revolution in 1994 (Chalfie et al. 1994) that it became possible to fully explore the inner mechanics of a prokaryotic cell. The use of fluorescence microscopy added, in fact, a second layer of complexity to prokaryotes. Imaging of fluorescently-tagged proteins not only highlighted the presence of several cytoskeletal like structures and scaffolds proteins, but also pointed toward a highly regulated spatio-temporal localization for such structures.

In the early days of fluorescence microscopy, spatio-temporal regulation of proteins could only be observed from the point of view of the whole protein population. The inability to distinguish between single proteins meant that studies of protein dynamics via time lapse microscopy and/or fluorescence recovery after photobleaching (FRAP) were limited to population analysis and were therefore unable to distinguish between the behavior of protein subpopulation.

It is not until the early 2000s that this problem got address with the development of super-resolution light microscopy techniques (Hell and Wichmann 1994, Betzig et al. 2006). While there exist several super-resolution techniques, we can separate them into two groups: single molecule localization microscopy (SMLM) and non-SMLM techniques. As implied by its name, the SMLM group is composed by those techniques that achieve an increase in resolution by isolating fluorophores and fitting their emission profile with a point spread function (PSF) while non-SMLM use a variety of different methods. The ability to image single proteins with a nanometer scale precision did not only drastically increase the imaging resolution from a structural point of view, but it also allowed for sub-population dynamicity studies. The importance of super-resolution techniques was immediately recognized by the scientific community and three of their developers were awarded in 2014 with the Nobel Prize in Chemistry: Eric Betzig, Stefan W. Hell and William E. Moerner.

1.2. Resolution and the pioneers of single molecule analysis

From a practical point of view, in fluorescence microscopy, resolution is defined as the smallest distance at which two fluorescent objects can be distinguished. Fluorescent proteins such as GFP and its derivatives are shaped like cylinders of approximately 3 nm in diameter and 4 nm in length (Hink et al. 2000). Under a fluorescent microscope, the light emitted by these proteins appears as a 2D pattern

(Airy pattern) where a central intensity maximum is surrounded by rings whose intensities decrease the farther you move from the emission centre. When two fluorescent proteins are emitting at the same time it will be possible to distinguish between them only when they are farther apart than the resolution limit.

From a quantitative point of view, different standards can be used to determine whether or not two points can be distinguished. One of such standards is the Rayleigh criterion (Rayleigh 2009), which specifies that the contrast between the maximum and minimum intensity between the centres of two points needs to be at least 26% lower than the maximum. The distance is defined as:

$$\Delta x = \frac{1.22\lambda}{2n\ sin\theta} = 0.61\frac{\lambda}{NA} \quad (1)$$

where Δx is the distance at which two points can be distinguished, λ is the wavelength of the light emitted by the fluorophores and NA ($n\ sin\vartheta$) is the numerical aperture. It follows that two GFP molecules would need to be farther apart than 210 nm (given an emission peak at 505 nm and NA of 1.46) in order to be able to distinguish between the two.

While the resolution limit for fluorescent proteins is generally smaller than the size of both eukaryotic and prokaryotic cells, it is 60 to 70 time larger than the fluorophores themselves. It follows that simultaneous observation of the whole protein population and ability to distinguish between single fluorophores was limited to those cases where the emitters were located farther from each other than the resolution limit. Before the development of SMLM, these cases were quite rare and mainly consisted of studies of low-density receptor on cell surfaces. In the early 1980s, Barak and Webb (Barak and Webb 1981, Barak and Webb 1982) were the first to use single particle tracking in the field of fluorescence microscopy in a series of studies concerning the internalization of dioctadecylindocarbocyanine low-density lipoproteins (dil(3)-LDL) in fibroblast cells. Later on, Gross and Webb (Gross and Webb 1986, Gross and Webb 1988) used similar techniques to analyse the dil-LDL receptor clustering.

1.3. Basis of single molecule microscopy (SMLM)

In chapter 1.2, it was explained that the existence of a resolution limit in fluorescent microscopy is dependent on the inability to separate the signal originated from separate emitters that are located closer to each other than a certain distance (Rayleigh criterion – Equation 1). Since SMLM techniques work in conditions where each diffraction limited area contains at most a single active fluorophore,

resolution as previously defined cannot be applied. In SMLM, instead, two different parameters are taken into account when talking about resolution: Localization Precision and Structural Resolution.

1.3.1. Localization precision

The diffraction-limited spot originated from a single emitter can be described as the probability to detect a point within a two dimensional space: the point spread function (PSF). In the ideal case, the PSF correspond to the Airy Disk (the bright central region of the Airy pattern). In SMLM, several methods have been proposed in order to fit the PSF of a molecule and calculate its centroid position with nanometer scale precision and multiple software have been developed for that purpose (Sage et al. 2019). The most common way to approximate the diffraction pattern is to fit it via a 2D Gaussian distribution. In a 2D plane, the intensity distribution of the fluorophore I(x,y) can be described as:

$$I(x,y) = I_0\, e^{-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}} \tag{2}$$

where I_0 is the maximum intensity of the distribution, x_0 and y_0 are the coordinates of the centroid of the Gaussian distribution and σ_x and σ_y its standard deviation in x and y, respectively. When using this approximation, the localization precision of a centroid on a single axis (Δx), given that all sources of noise except from shot noise are negligible (shot noise is a result of the particle character of light), is inversely proportional to the square root of the number of photons that are collected (N) and can be described as:

$$(\Delta x)^2 \geq \frac{{\sigma_x}^2}{\mathrm{N}} \tag{3}$$

Two other parameters are taken into consideration when calculating the localization precision: pixelation noise and background noise.

The first kind of noise is given by the uncertainty of each photon localization due to pixel size (pixels are finite in size and a photon can localize anywhere within it) and yields to Equation 3 as follow:

$$(\Delta x)^2 \geq \frac{{\sigma_x}^2 + p^2/12}{\mathrm{N}} \tag{4}$$

where p is the pixel size and the $p^2/12$ factor is the variance of a top hat distribution of size p. The addition of the pixilation noise results in an increase of the apparent size of the spot. Background noise (b) originates from the detection of photons that are not emitted by the fluorophore itself. In a situation where background noise across the area being analysed is constant, Equation 4 can be further modified to take into account the presence of such noise as shown by Thompon et al. (Thompson et al. 2002):

$$(\Delta x)^2 \geq \frac{{\sigma_x}^2 + p^2/12}{\mathrm{N}} + \frac{4\sqrt{\pi}{\sigma_x}^3 b^2}{p\mathrm{N}^2} \tag{5}$$

While the part of the localization precision dependent on background noise is inversely proportional to the collected number of photons, the remaining factor is inversely proportional to the square root of the collected number of photons. It follows that the influence of background noise is bigger for fluorophores that emit a low number of photons while events characterized by a high photon count will be dominated by shot noise (pixelation noise will also play a role but it is usually smaller than shot noise).

1.3.2. Structural resolution

Out of the two resolution criteria that are used in SMLM, structural resolution is far more stringent than localization precision. While, as explained in 1.3.1, localization precision mainly depends on the number of photons emitted by the fluorophore and the goodness of its PSF fit, structural resolution is limited by the Nyquist-Shannon sampling criterion (Nyquist 1928, Shannon 1949). According to said criterion, in order to obtain a specific resolution, the labelling frequency needs to be of at least two molecules per desired spatial scale.

The labelling frequency and quality in a biological sample is influenced by a multitude of factors. These factors include the number of molecules that comprise the label, their size, and the stoichiometry between the emitter and the molecule of interest. Labelling can be done via translational fusion of the protein of interest with fluorescent proteins/self-labelling protein tags (PA-mCherry, HaloTag (Los et al. 2008), SNAP-tag (Keppler et al. 2003)) or via the binding of said protein with specific tags (primary antibodies, nanobodies) which are either conjugated to fluorophores or bound to a second molecule which is in turn conjugated to a fluorophore (secondary antibodies, DNA-PAINT (Jungmann et al. 2010)). Since labels have a finite size, their presence will result in an increase of structural size which will in turn increase the distance between emitters, ultimately decreasing the maximum resolution that can be achieved. The decrease in resolution due to tagging ranges from

immunostaining (each antibody spans for 10-15 nm, meaning that the combination of primary and secondary antibodies has a length of almost 30 nm) to the use of fluorescent proteins (as previously mentioned a typical fluorescent protein can be described as a barrel of 3 nm in diameter and 4 nm in length). Further decrease in structural resolution is caused by the suboptimal stoichiometry between the label and its target. Since in most cases a 1:1 stoichiometry is not achieved, the imaged structure comprises of tagged and free target molecules which translates in an increase of average distance between labels.

1.3.3. Reconciling localization precision and structural resolution: Fundamental resolution measure (FREM) and the Fourier ring correlation (FRC) resolution

SMLM imaging is diffraction-unlimited and its resolution depends on a multitude of factors (fluorophore choice, labelling density, tag-fluorophore combination size, stoichiometry between molecule of interest and emitter). Two complementary approaches that improve from the mere concepts of localization precision and structural resolution have been proposed in order to determine the resolution of an image in SMLM: fundamental resolution measure (FREM) and Fourier ring correlation (FRC).

FREM ignores labelling density and is instead used to calculate the accuracy with which distances between two fluorophores can be estimated. This means that its use is limited to the study of structures characterized by low molecule numbers (Cases where the labelling density within the structure has no significance) (Ram et al. 2006). In all those cases where instead labelling density play a huge role (structures characterized by high number of molecules and quasi-continuous structure) FRC is the preferred method of choice. This approach describes the length scale below which the image cannot be resolved (Nieuwenhuizen et al. 2013). It is important to notice that the resulting resolution will always be larger than localization precision and structural density as previously described (1.3.1 and 1.3.2).

1.3.4. Improving resolution and current challenges

While it is common to obtain a final resolution of 20 nm via state of the art SMLM, the theoretical resolution limit achievable with these techniques correspond to the size of the emitter itself (a better resolution would fail to comply with the Nyquist-Shannon criterion) which would in turn lead to true molecular imaging.

This limit scenario can only be obtained when the following conditions are satisfied: a 1:1 stoichiometry between the molecules of interest and the fluorophores, an extremely small tag-fluorophore combination and a localization precision comparable in size to the emitter itself. Steps toward this goal include improvements of the tag-fluorophore combinations themselves (Ries et al.

2012, Raulf et al. 2014, Grimm et al. 2015, Strauss et al. 2018) and technical/analysis improvements (dualObjective STORM, MINFLUX, SIMFLUX) (Xu et al. 2012, Balzarotti et al. 2017, Cnossen et al. 2020). Studies involving the development and optimization of tag-fluorophore combinations greatly contributed to improve structural resolution and localization precision while technical/analysis studies allowed for further improvement in localization precision.

A recent example in tag enhancement came from the use of Slow Off-rate Modified Aptamers (SOMAmers: 21-28 kDa) as tags for high-resolution in situ DNA-PAINT imaging (Strauss et al. 2018). Before the use of SOMAmers, DNA-PAINT imaging was characterized by high localization precision (below 10 nm) and low structural resolution. The low structural resolution was a result of the use of antibodies as tags (antibodies are ~150 kDa in size and cannot guarantee a 1:1 stoichiometry ratio). While the use of SOMAmers had no effect from a localization precision point of view, it allowed for a great improvement in structural resolution due to decreased tag-fluorophore size and a 1:1 stoichiometry ratio between tag and molecule of interest.

The latest resolution improvement achieved via a technical enhancement consisted in the combination of SMLM techniques (DNA-PAINT and dSTORM) with the use of patterned illumination (Cnossen et al. 2020). The combination of these two techniques resulted in a near two fold improvement in FRC resolution (from 16.4 nm to 8.6 nm) with a value relatively close to the size of the probes themselves but that does not allow yet for true molecular imaging.

1.4. Photoactivated localization microscopy (PALM)

PALM is one of the most prominent SMLM techniques. Developed in 2006 (Betzig et al. 2006), this technique relies on finely tuned activation/switching and time-resolved localization of photo-activatable/switchable fluorophores to construct high resolution images.

1.4.1. Fluorophores

The fluorophores of choice in PALM are fluorescent proteins. The use of such fluorophores leads to both advantages and disadvantages. On one side it overcomes the need of extra experimental procedures such as membrane permeabilization and fixation that are in general not compatible with living systems, on the other side fluorescent proteins have a lower photon yield than organic synthetic dyes and the fusion between the protein of interest and the fluorescent protein may lead to partial/complete loss of function.

The fluorescent proteins that are used in PALM can be split into the following categories: photo-activatable (the protein can be activated from a non-fluorescent state), photo-switchable (the protein

can switch between two different states in a spontaneous or stimulated fashion) and photo-convertible (the protein can change their emission spectrum) (Table 1.1). Photo-conversion between different conformational states is usually stimulated by the use of UV light (a 405 nm laser line), with the amount of converted protein being proportional to the 405 nm laser intensity and pulse length. While in most cases the number of active molecules at any given time can be titrated via the regulation

Table 1.1: Proteins commonly used in PALM. The proteins are divided into photo-activatable, photo-convertible, photo-switchable and others. The maximum excitation and emission wavelength for two conformational state for each protein are shown. An empty conformational state in position 1 signifies that the protein starts in a dark conformation state while an empty conformation state in position 2 signifies that the protein starts in an active conformational state.

	State 1	State 2		
Protein acronym	**Ex/Em (nm)**	**Ex/Em (nm)**	**PC laser (nm)**	**Source**
Photo-activatable proteins (Irreversible conversion between State 1 and 2):				
PA-GFP	-/-	504/517	405	(Patterson and Lippincott-Schwartz 2002)
PA-TagRFP	-/-	562/595	405	(Subach et al. 2010)
PA-mCherry2	-/-	570/596	405	(Subach et al. 2009)
PA-mKate2	-/-	586/628	405	(Gunewardene et al. 2011)
Photo-convertible proteins (Irreversible conversion between State 1 and 2):				
PS-CFP2	400/468	490/511	405	(Chudakov et al. 2004)
mEos3.2	507/516	572/580	405	(Zhang et al. 2012)
mMaple3	489/505	566/583	405	(Wang et al. 2014)
Dendra2	490/507	553/573	405/488	(Turkowyd et al. 2017)
Photo-switchable proteins (Reversible conversion between State 1 and 2):				
Dronpa	-/-	503/518	405	(Andresen et al. 2008)
Dreiklang	-/-	515/529	365(on) 405(off)	(Brakemann et al. 2011)
NijiFP*	469/507	526/569	405(G-R/on-G) 440(on-R) 488(off-G) 561(off-R)	(Adam et al. 2011)
eYFP	513/527	-/-	405	(Ormo et al. 1996, Dickson et al. 1997)
mNeonGreen	506/517	-/-	-	(Shaner et al. 2013, Stockmar et al. 2018)

*NijiFP has three conformational states. The dark state (off) was omitted in the table.

of the photoconversion laser, this is not always the case. For instance, mNeonGreen is characterized by the lack of a photoconversion laser coupled with the tendency to spontaneously switch back and forth between a dark and an active state. It follows that these fluorophores will first need to be excited and brought to a dark state via the use of the imaging laser (i.e.: 488 nm for mNeonGreen), in what

can be called a bleaching phase, and then imaged as single molecules once they stochastically return to their active state. The aforementioned process is not without flaws. The bleaching phase usually results in the permanent loss of part of the fluorophore population while the rate of return to the active state of the remaining fluorophores cannot be controlled and can lead to a condition where multiple fluorophores are active within a single diffraction limited area. This generally translates to a decrease in structural resolution, meaning that it is not recommended to use proteins belonging to this category when doing single color PALM (Stockmar et al. 2018).

While the inability to accurately control the switching rate of proteins such as mNeonGreen is a downside when doing single color PALM, these fluorophores are of extreme importance when performing dual color PALM. The reasoning behind this conclusion is that most of the other fluorophores share the same photoconversion laser (405 nm laser line), meaning that both proteins of choice will be converted to an active state at the same time. Although this limitation can be circumvented from a technical perspective by combining a beamsplitter with either a dual-camera system or, alternatively, with the partitioning of the camera chip, systems lacking said features will be unable to image two fluorophores that share the same photoconversion laser. This problem does not arise when imaging a protein that uses a photoconversion laser in combination with a protein that switch spontaneously between fluorescent and dark state, as the photo-activatable protein can be fully imaged prior the imaging of the second protein without loss of signal.

1.4.2. Fixed and live samples

The time needed to image each single fluorescent protein within a cell via PALM is directly proportional to the number of fluorophores located within the cell itself and ranges from few seconds to hours. This means that in order to find the localization of the entire protein population at a given time it is first necessary to fix the sample. Sample fixation can be responsible for the formation of artefacts and an overall decrease in signal (Stockmar et al. 2018) and it is therefore necessary to appropriately optimize the fixation protocol in order to minimize such effects (Leyton-Puig et al. 2016). Once the sample is fixed, the imaging parameters depend exclusively on the characteristics of the fluorophore of choice, its abundancy and local density (Bach et al. 2017). The fluorophore localization and other pertinent parameters obtained by fitting each molecule signal are then collected into a table which will act as a starting point for the data analysis.

Live imaging differs from the imaging of fixed samples not only because the technical obstacles that need to be overcome for the two kind of imaging are different, but also because the imaging parameters depend on the hypothesis being tested. As previously mentioned, the time needed to image all the fluorophores that are comprised within a cell is proportional to their number. As, in regards to the average time frame needed for a PALM experiment, all proteins show a certain degree of dynamicity (moreover cell morphology itself can change within said time frame), it is not feasible to obtain a faithful representation of the whole protein population via *in vivo* PALM. It follows that live imaging in PALM is mainly used to perform single particle tracking (SPT). In order to precisely image protein kinetics it is necessary to isolate the signal drift originated from the protein movement from the drift caused by other sources such as cell growth, sample drift and cell motility. Since a fluorophore will usually bleach within few seconds (fluorescent proteins will bleach even faster, compared to synthetic dyes) changes in cell size due to cell growth do not influence protein kinetics measurements and no correction is therefore necessary. This is not the case for sample drift and cell motility. Sample drift is universal and is usually corrected via the use of fluorescent fiducials followed by drift correction (Thompson et al. 2002). The presence or absence of cell motility is instead dependent on the microorganism of choice, where sample immobilization is necessary only for motile organisms such as *Bacillus subtilis* (sample immobilization is usually performed by placing the sample on low fluorescence agarose pads) (de Jong et al. 2011, Stracy and Kapanidis 2017).

An extra factor that need to be taken in consideration when performing single particle tracking via PALM is the aim of the experiment. If this involves the study of the kinetic changes of a protein upon different conditions, the use of short frame times (exposure time 5-15 ms) and low laser intensity will maximize the track length and the range of speeds that can be followed at the cost of localization precision (Uphoff et al. 2013, Stracy et al. 2016). On the opposite side, if the study focuses on the localization of the immobile population, it may be better to use longer frame times (up to 100 ms) or implement a dark time (interval time between consecutive frames where the lasers are off) between frames at the cost of being unable to image fast moving fluorophores.

1.5. PALM data analysis

PALM and, in general, SMLM techniques differ from other fluorescent microscopy techniques also with respect to the downstream image analysis. As explained in 1.4.2 the typical output of a PALM experiment is a table containing fluorophore localizations and fluorophore associated parameters (Table 1.2). The data contained within the table can either be used to build rendered fluorescent images and be analysed in a conventional-like fashion via image analysis software such as Fiji (Schindelin et al. 2012, Rueden et al. 2017), or analysed via coordinate based methods such as point

pattern analysis (PPA) (Spielman 2017), ordering points to identify the clustering structure (OPTICS) (Ankerst et al. 1999) and coordinate-based co-localization (CBC) (Malkusch et al. 2012). While analysis of rendered images can be of use and will be shown in chapter 3, it also results in a loss of data information. It is therefore suggested to treat the data as coordinates and use the appropriate type of analysis.

Table 1.2: Exemplary table output obtained via the PSF fitting of single fluorophores signals in a PALM experiment. The information contained within this table represent only some of the information that can be extracted from the raw data images and vary according to the software used for the fitting of the raw images. Starting from the left: Index, Frame number (frame at which the signal is being perceived), Position X (localization of the centroid on the X axis in nanometers), Position Y (localization of the centroid on the axis), Precision (localization precision of the centroid), Photon number (number of photons emitted) and PSF (point spread function half width at $e^{-1}\sigma_x$).

Index	Frame #	Position X (nm)	Position Y (nm)	Precision (nm)	Photon #	PSF (nm)
1	1711	29688.5	40718.9	20.6	152	96.3
2	1721	31831.6	38985.2	29.3	99	98.0
3	2311	29507.0	38742.3	24.5	142	104.6
4	2441	29482.8	38718.8	25.6	128	102.9
5	1272	29515.8	38720.0	29.7	147	121.1
6	1277	30847.0	39485.8	39.1	168	140.5
n	1 …	…	…	…	…	…

1.5.1. Point pattern analysis (PPA)

PPA is a set of analysis methods dedicated to the study of points distributed on a map, where the map consists of an axial plane of n axis (in 2D PALM it is a 2D plane) (Spielman 2017). PPA methods can be divided into two groups: one dedicated to the determination of the number of points within a specific region and one dedicated to examine the relative positioning of points (e.g. nearest neighbor distribution function G). In PPA, identification of a cluster or, more in general, of a pattern is based on the identification of statistical deviation from a null hypothesis, where the basic null hypothesis correspond to complete spatial randomness (e.g. Poisson point process) and more advanced ones take into consideration points dependence/interaction (e.g. Gibbs models (Baddeley et al. 2013)).

Given a dataset containing a point pattern in an area of arbitrary shape (fluorophores localizations within a cell), it is possible to estimate its nearest neighbor distribution $G_e(r)$ (G_e is the experimental estimate). Since this estimate is affected by the inability to observe points outside the area of analysis (edge effect), edge correction methods (e.g. Hanish estimator (Hanisch 2007)) are implemented in order to minimize its estimate bias (Ripley 2012, Baddeley 2019).

A second nearest neighbor distribution, $G_t(r)$ (G_t is the theoretical estimate), is then calculated under the assumption that the points contained within the dataset follow a Poisson point process within the area being analysed (independent fluorophores randomly distributed within a cell). This function is defined as:

$$G_t(r) = 1 - e^{-\delta \pi r^2} \quad (6)$$

where δ is the expected density of points, and r is the distance at which G_t is estimated. G_e and G_t are then compared, with statistical deviation between the two estimates suggesting spatial clustering/spatial regularity in the experimental dataset (refusal of the null hypothesis).

While, when applied to 2D PALM data, the aim of PPA is the identification of clusters on a two dimensional plane, it is important to point out that the definition of "cluster" varies between different coordinate analysis methods. Specifically, while PPA identify clusters via the use of a null hypothesis for statistical inference, other cluster analysis methods, such as OPTICS, aim to identify a set of meaningful subclasses present within a dataset.

1.5.2. Ordering points to identify the clustering structure (OPTICS)

The aim of the OPTICS algorithm is to order the dataset in a way that represent its density-based clustering structure (Ankerst et al. 1999). In order to be able to understand density-based cluster-ordering, we first need to understand the density-based clustering method that is at the base of OPTICS: DBSCAN (Ester et al. 1996). The definition of cluster according to DBSCAN is that, given a dataset of points, a circular area of radius ε centred on each element of the cluster will need to contain at least a certain number of other points ($MinPts$). The resulting cluster will be composed of two type of points: objects whom area contain at least $MinPts$ (core points), and objects which do not satisfy the rule but that are found within the area of points that do (border points). As a result, all the points that do not satisfy the listed conditions will not be considered as belonging to clusters (Figure 1.1 A).

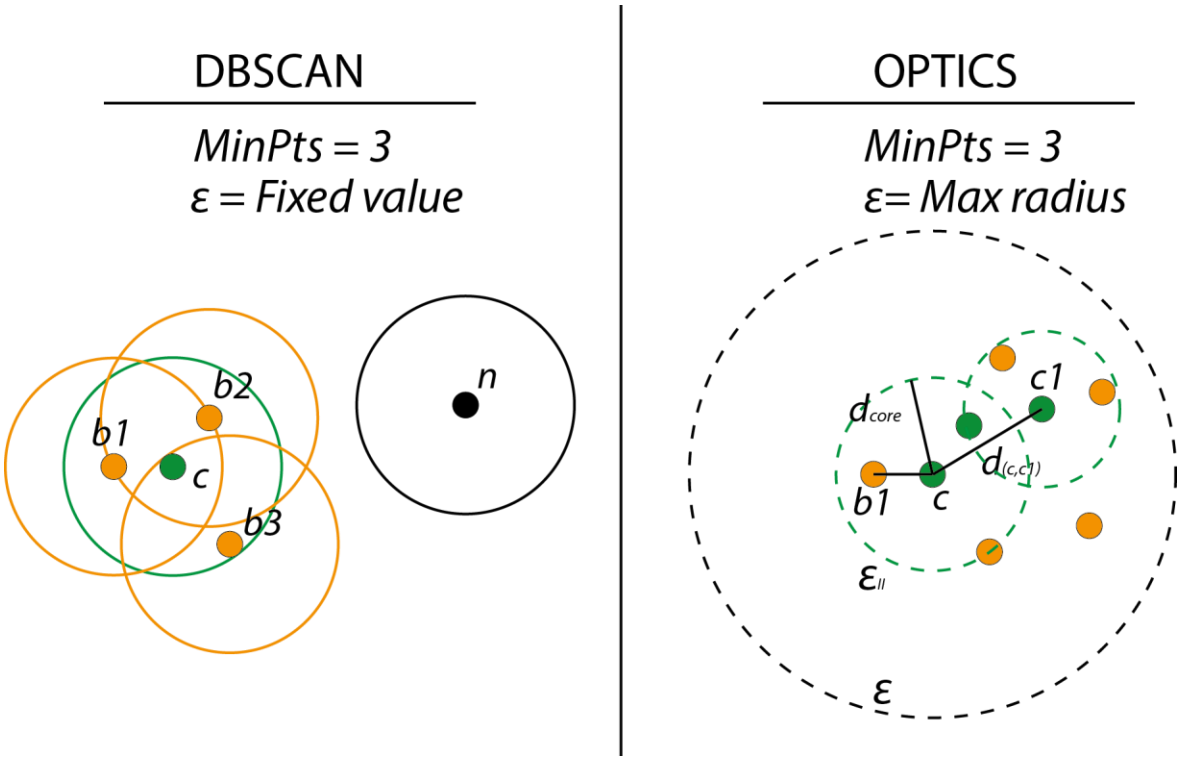


Figure 1.1: DBSCAN vs OPTICS. (A) In DBSCAN points are assigned to a cluster if they satisfy the conditions necessary to be a core point ($MinPts = 3, radius = \varepsilon$) or if their distance from a core point is smaller than ε. Core points (c) are highlighted in green, border points (b) in orange and points not belonging to clusters (n) in black. **(B)** In OPTICS ε is the maximum radius at which the $MinPts$ condition is tested for each point. The core distance of c (d_{core}) and reachability distance of c from c_1 ($d_{(c,c1)}$) are two extra parameters that are calculated in order to determine the points ordering.

OPTICS differentiates itself from DBSCAN by linearly ordering the points contained within the dataset such as nearest neighbors are next to each other within the ordering. ε, in this case, is the maximum distance to be considered and $MinPts$ is the minimum number of points necessary to form a cluster. Within the ordered dataset will also be annotated two new parameters for each point: core distance (d_{core}) and reachability distance ($d_{(c,c1)}$). These two parameters are defined as:

$$d_{core}(c) = \begin{cases} UNDEFINED & if\ |N_\varepsilon(c)| < MinPts \\ MinPts - th\ smallest\ distance\ to\ N_\varepsilon(c) & otherwise \end{cases} \quad (7)$$

$$d_{(c,c1)}(c) = \begin{cases} UNDEFINED & if\ |N_\varepsilon(c)| < MinPts \\ \max(dist(c, c_1), d_{core}(c)) & otherwise \end{cases} \quad (8)$$

where $N_\varepsilon(c)$ is the number of points contained within a circular area of radius ε centred on c and $dist(c, c_1)$ is the Euclidean distance between c and c_1 (Figure 1.1 B). Lastly, it is important to specify that reachability distance cannot be defined unless $c_1$is also a core point.

The reachability distances obtained via OPTICS can be then be plotted according to the determined ordering and understood graphically (Figure 1.2). Following the construction of the reachability plot, it is possible to set one or more threshold limits for ε (ε_{limit}) in order to isolate clusters of different minimum density (Figure 1.2).

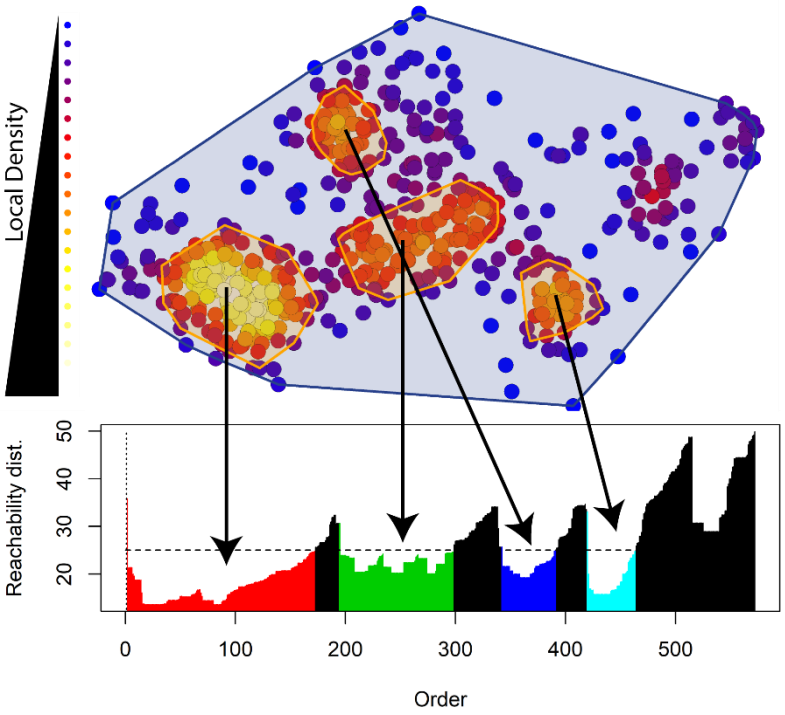


Figure 1.2: Reachability plots and their visual interpretation. Points are color coded according to their local density in order to simplify the optical recognition of clustering area. The blue overlay (containing all points) is obtained by using an ε_{limit} above 50 while the orange overlay is obtained by using an ε_{limit} of 25 (dashed horizontal line within the reachability plot). Generally speaking, the presence of a "valley" within a reachability plot correspond to a cluster, where a steep "valley" correspond to a fast increase in density and deeper "valleys" will be characterized by higher density.

1.5.3. Coordinate based co-localization (CBC)

Adaptation of existing mathematical algorithms for the study of co-localization in SMLM data was first proposed in 2012 (Malkusch et al. 2012). This method, differently from quantitative co-localization algorithms used for conventional fluorescence microscopy such as the Pearson correlation coefficient (Pearson 1896), does not output a single co-localization value for the whole image but rather calculates a CBC value for each single molecule of the two examined species (A and B).

Given a localization A_i belonging to the A species, the distributions of A and B around A_i ($D_{A_i,A}(r)$ and $D_{A_i,B}(r)$) are defined as:

$$D_{A_i,A}(r) = \frac{N_{A_i,A}(r)}{\pi r^2} * \frac{\pi R_{max}^2}{N_{A_i,A}(R_{max})} = \frac{N_{A_i,A}(r)}{N_{A_i,A}(R_{max})} * \frac{R_{max}^2}{r^2} \quad (9)$$

$$D_{A_i,B}(r) = \frac{N_{A_i,B}(r)}{N_{A_i,B}(R_{max})} * \frac{R_{max}^2}{r^2} \quad (10)$$

where $N_{A_i,A}(r)$ is the number of localizations of A within a radius r around A_i. The number of localizations is divided by area (πr^2) resulting in a localization density. This value is normalized by dividing the obtained density by the density of localizations of A observed at the maximum radius, which is characterized by $N_{A_i,A}(R_{max})$ and an area equal to πR_{max}^2. The same procedure is performed to calculate $D_{A_i,B}(r)$.

Using $D_{A_i,B}(r)$ and $D_{A_i,A}(r)$ it is then possible to calculate the Spearman´s rank correlation coefficient (S) for A_i:

$$S_{A_i} = \frac{\sum_{r_j=0}^{R_{max}} (O_{D_{A_i,A}}(r_j) - \bar{O}_{D_{A_i,A}}) * (O_{D_{A_i,B}}(r_j) - \bar{O}_{D_{A_i,B}})}{\sqrt{\sum_{r_j=0}^{R_{max}} (O_{D_{A_i,A}}(r_j) - \bar{O}_{D_{A_i,A}})^2} * \sqrt{\sum_{r_j=0}^{R_{max}} (O_{D_{A_i,B}}(r_j) - \bar{O}_{D_{A_i,B}})^2}} \quad (11)$$

where $O_{D_{A_i,A}}(r)$ is the rank calculated after Spearman, and $\bar{O}_{D_{A_i,A}}$ is the arithmetic average of $O_{D_{A_i,A}}(r)$ for $0 \leq r \leq R_{max}$. A final equation in then necessary to obtain the CBC value of A_i (C_{A_i}) from the Spearman´s rank correlation coefficient:

$$C_{A_i} = S_{A_i} * e^{(-\frac{E_{A_i,B}}{R_{max}})} \quad (12)$$

where $E_{A_i,B}$ is the distance between A_i and the nearest neighbor from B. The obtained value can range between -1 and +1 with a value of +1 (perfect correlation) corresponding to high probability of co-localization, a value of 0 (non-correlation) corresponding to low probability of co-localization and finally, a value of -1 (anti-correlation). Similarly to what happen when performing co-localization analysis in conventional fluorescence microscopy (Zinchuk and Zinchuk 2008), the anti-correlation value is of difficult interpretation from a biological point of view and particular care needs to be used when drawing conclusions from it.

1.6. Aim of the study

PALM and, more in general, single molecule localization microscopy techniques differentiate from conventional fluorescence microscopy and other superresolution techniques for their ability to precisely localize single molecules. While the ability to localize single molecules results in an increase in information output and potentially give us the ability to answer otherwise unsolved biological questions, as any other technique, it comprises flaws. Factors such as homogeneity of the data quality across a field of view, background fluorescence, fluorophores behaviour and cell orientation need to be taken in consideration when analysing the data, whether we are talking about clustering (OPTICS), co-localization (CBC) or other type of analysis. It is therefore necessary to try and minimize the effect that each of these background sources have on the data. A second, non-technical, source of variability depends on the relative novelty of SMLM techniques, with know-how concerning data analysis being relatively fragmented and several possible approaches being proposed in order to study the same phenomenon (Malkusch et al. 2012, Levet et al. 2019, Khater et al. 2020).

In this study I take a multistep combinatory analysis approach to PALM data in Fiji (Schindelin et al. 2012) and R (R_Core_Team 2020). Following a generalized filtering step dedicated to the polishing of the data, localizations are labelled and grouped according to specific parameters (Cell_ID, Cluster_ID, …) allowing for the application of a series of algorithms (PPA, OPTICS, CBC) in an ID specific/unspecific manner.

Each part of the script is thoroughly discussed in regards to its strengths, weaknesses and branching possibilities via examples of its applications across different microorganisms. These include the characterization of novel proteins and of the chromosome segregation machinery in *Corynebacterium glutamicum*, the study of the magnetosome formation in *Magnetospirillum gryphiswaldense* and the study of the flagellar organization in *Trypanosoma brucei brucei*.

Finally, by providing a step by step guide toward the polishing, analysis and interpretation of SMLM data, this dissertation does not intend to set an absolute standard for analysis but is rather aimed at easing the reader into the topic itself. Specifically, this work focuses on providing the reader with the ability to navigate, combine and extract the desired information with particular emphasis toward some of the most commonly required analysis methods (clustering, co-localization).

Chapter 2: Data quality and fluorophore dependent data filtering

Scripts and macros used within this chapter:

- Script_R1
- Script_R2
- Fiji macro 1
- Fiji macro 2

Results included within this chapter were published in:

- Bach J. N., Giacomelli G., Bramkamp M. (2017). "Sample Preparation and Choice of Fluorophores for Single and Dual Color Photo-Activated Localization Microscopy (PALM) with Bacterial Cells." Methods Mol Biol 1563: 129-141.

When testing a hypothesis, it is always desirable to try and minimize the amount of artifacts and background noise contained within the data, thus decreasing the chance to obtain false positives or false negatives. Data quality is influenced by two mechanics: experimental planning/procedures and raw data processing/filtering. As most aspects of experimental design are determined in a case by case scenario, I will only briefly discuss them. I will instead focus on the steps that I repute necessary in order to isolate high quality PALM data from artifacts/background noise with specific emphasis on raw data filtering.

In conventional fluorescence microscopy the quality of the data is assured by the use of a functional fluorescent fusion where the fluorophore is bright, monomeric and characterized by excitation and emission profiles that differ from the ones typical of the microorganism in use. Moreover, the imaged sample should be characterized by the absence of extracellular background signal and an appropriate cellular density (low enough so that each cell within the field of view can be observed separately).

While most of the general requirements necessary in order to obtain high quality data via conventional fluorescence microscopy (cellular density, lack of extracellular background, functional fluorescent fusion) also apply to PALM, the ideal fluorophore characteristics differ substantially between the two techniques. The ideal scenario for quantitative imaging of a fixed sample via PALM requires a fluorophore characterized by a 1:1 stoichiometry ratio with the protein of interest, a very high signal efficiency (fraction of fluorophore that can be detected within a cell), a high number of photons per switching cycle, the presence of a single on state before permanent bleaching and a defined lifetime, with little to no variation between fluorophores (Wang et al. 2014).

Since, at the present date, no fluorophore satisfies all of the conditions mentioned above and most biological samples show a certain degree of autofluorescence, correction and filtering methods are pivotal toward the improvement of the data quality. Here I approached the problem by establishing a

filtering pipeline based on a combination of fluorophore specific photo-characteristics (PSF width, photon count, persistence) and bright field derived regions of interest (ROIs). Since said pipeline can also be used for the determination of the photo-characteristic themselves (see Chapter 2.1), it can be readily applied even when the fluorophores properties are unknown.

2.1. Determination of PSF width and photon count filtering parameters

The first step within this pipeline is aimed at distinguishing molecules localizations derived from single fluorophores from those originated from background and/or multiple co-localizing fluorophores. Since, given a certain chemical environment, the number of photons emitted by a specific emitter and their PSF will have specific distributions, it is possible to isolate single fluorophore´s distribution and, based on that, establish filtering parameters (Figure 2.1). In order to determine the PSF width and photon count distribution for the fluorophores used within this dissertation (PAmCherry, Dendra2, mNeonGreen) I imaged cells containing the fluorophores of choice and emitter free cells with the same imaging parameters, where the parameters were aimed toward a low chance of having multiple active emitters within each diffraction limited area. This condition was achieved for PAmCherry and Dendra2 by appropriate tethering of the 405 nm activation laser and for mNeonGreen by bleaching the majority of the fluorophores previous the imaging. Since the analysis procedure used to determine the filtering parameters is consistent between different fluorophores, I will explain the rationale behind it only for one of the fluorescent proteins used: PAmCherry (filtering parameters for all fluorophores used can be found in Table 2.1).

Following the collection of the localizations, I applied two semi-automated, self-written Fiji macros (Giacomelli 2020) to the bright field channel in order to establish three regions of interest (ROI): focused (Fiji macro 1-2), unfocused (Fiji macro 1) and background (Fiji macro 1) (Figure 2.1 A, B). ROIs belonging to the focused category comprise cells that lie within the optical focus (focus at cell center) (blue), ROIs belonging to the unfocused category comprise cells clusters and non-cells particles that are perceived by the macros during the construction of the binary mask (black) and ROIs belonging to the background category comprise all the remaining field of view (orange) (Figure 2.1 A, B). I then extracted the ROIs contours via a third Fiji macro (Giacomelli 2020), converted them into the appropriate format in R (Script_R1) (Giacomelli 2020), and split the emitter localizations obtained via PALM accordingly (Script_R2 – as the script uploaded on the GitHub page is the final version, PSF width and photon count based filtering are performed before the assignment to specific ROIs). In presence of PAmCherry, localizations belonging to the focused category show a quasi-normal distribution for both photon count (99% localizations: $36<x<383$ photons) and PSF width (99% localizations: $47<x<215$ nm) (Figure 2.1 Ci, Di). Since these distributions differ from those obtained in absence of PAmCherry

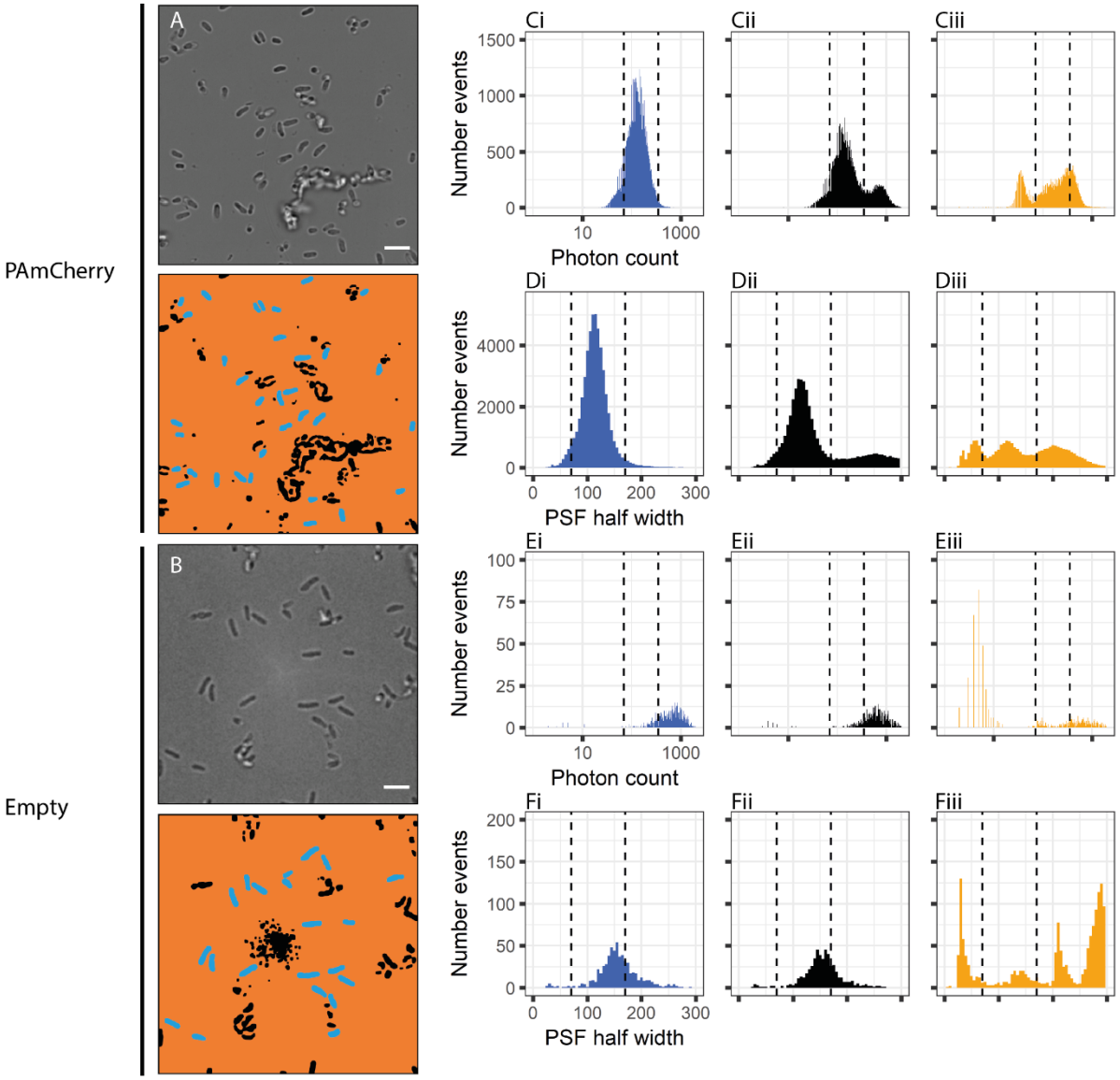


Figure 2.1: Photon count and PSF based filtering. WT cells (empty) and cells expressing the fluorophore of choice (PAmCherry) were imaged via PALM under the same imaging conditions. **(A,B)** The bright field image was transformed into a three colors binary mask (blue->focused, black->unfocused, orange->background). **(C-F)** PSF width (nm) and photon count were plotted for each condition and category (the histogram plots are color coded according to the category they belong to) and compared to determine the filtering parameters. The final filtering parameters are highlighted by dotted lines across all plots (Photon count: 70<x<350, PSF width: 70<x170 nm).

in both positioning and number of localizations (Figure 2.1 Ei, Fi), we can conclude that PAmCherry is the main source of localizations in the focused category.

Multiple peaks can be observed within the unfocused and background distributions in presence of PAmCherry (Figure 2.1 Cii-iii, Dii-iii) with the fluorophore being responsible for the majority of the signal. The presence of PAmCherry within these ROIs is corroborated by the fact that unfocused and background area include cells clusters (Figure 2.1 A) and that the distributions observed in presence and absence of the fluorophore differ (Figure 2.1 Cii-iii, Dii-iii, Eii-iii, Fii-iii), with the absence of the

fluorophore resulting in a lower number of localizations. PAmcherry is responsible for the subpopulation characterized by PSF and photon count in agreement with the focused category and for the one characterized by higher PSF width (>180 nm) and photon count (>350), where the second subpopulation is the result of multiple fluorophores emitting within a single diffraction limited area. The presence of overlapping fluorescence signals exclusively in background and unfocused area can be explained by the overlapping of cells within cells clusters, which results in a higher fluorophore activation chance per diffraction limited area at any given time.

A third, unique subpopulation can be observed within the background category (Figure 2.1 Ciii-Diii). With PSF width (<80 nm) and photon count (<70) lower than the ones that characterize PAmCherry, these localizations can be traced to autofluorescent particles.

Table 2.1: PSF width and photon count filtering parameters. Filtering parameters were determined as described in chapter 2.1 and applied to all PALM data acquired within this dissertation prior analysis.

Fluorophore	Photon count filtering (photons)	PSF width filtering (nm)
PAmCherry	70-350	70-170
Dendra2	100-500	80-160
mNeonGreen	70-500	60-180

Given the PSF width and photon count intervals that include 99% of the localizations perceived within the focused category (photons count: 36<x<383, PSF width: 47<x<215 nm), I performed a step-wise refining aimed at limiting loss of information while improving data quality. The lower limit of the photon count was increased to 70 (exclusion of the autofluorescent particles population) while the upper limit was decreased to 350 (decreased chance of overlapping fluorophores). Similarly, the lower limit of the PSF width was increased to 70 nm while the upper limit was decreased to 170 nm (Table 2.1).

It is important to keep in mind that application of the described filtering parameters to different fields of view or experiments result in variable data loss amount, where the percentage of data loss is proportional to the ratio between the amount of background and overlap signal compared to the amount of signal derived from single fluorophores.

2.2. Spatio-temporal grouping of fluorescent events

While, following the photon count and PSF width filtering process, we can be confident that each localization correspond to a single event, we still need to be able to distinguish between localizations originated from a fluorophore activation event and those caused by either the persistence of fluorophore over multiple frames or a blinking event (alternation of on and off state of the same fluorescent protein).

Under imaging conditions that prevent the activation of multiple molecules within a diffraction limited area at any given time, such as the ones being used during the collection of these datasets, the persistence of a fluorophore across multiple frames will be reflected by a sudden spike in the number of localizations across the experiment timespan (Figure 2.2 A). It follows that this phenomenon can be easily identified also in absence of a priori knowledge concerning the fluorophore photo-characteristics themselves. The same does not hold true for blinking. Since fluorophores can remain in an off state for relatively long time (few seconds) before returning to an active state (Annibale et al. 2011, Durisic et al. 2014), the presence of blinking does not necessarily translate in a spike in localizations. Identification of blinking becomes even more complex in presence of multimeric protein structures, where the appearance of spatially clustered localizations is to be expected even in the absence of blinking. Ultimately, blinking correction can only be applied with previous knowledge concerning the fluorophore blinking characteristics which are in turn determined by imaging the fluorophore at very low density (Annibale et al. 2011, Durisic et al. 2014). Since the aim of this chapter is to improve data quality in absence of previous knowledge concerning the fluorophores, the determination of spatio-temporal grouping parameters will be aimed toward the correction of fluorescence persistence rather than blinking.

The spatio-temporal grouping algorithm I used for this purpose is included within the ZEN 2.1 SP3 software (black)(64bit)(Version 14.0.4.201) associated with the Zeiss Elyra P.1 microscope and is based on three separate parameters: maximum consecutive frames (maximum number of consecutive frames where the fluorophore is emitting during a blinking event), off frames (expected number of frames between two blinking events of the same fluorophore) and search distance (maximum distance at which consecutive localizations are considered to be originated from the same fluorophore). It is worth mentioning that since fluorescence persistence imply the absence of multiple on-off switching,

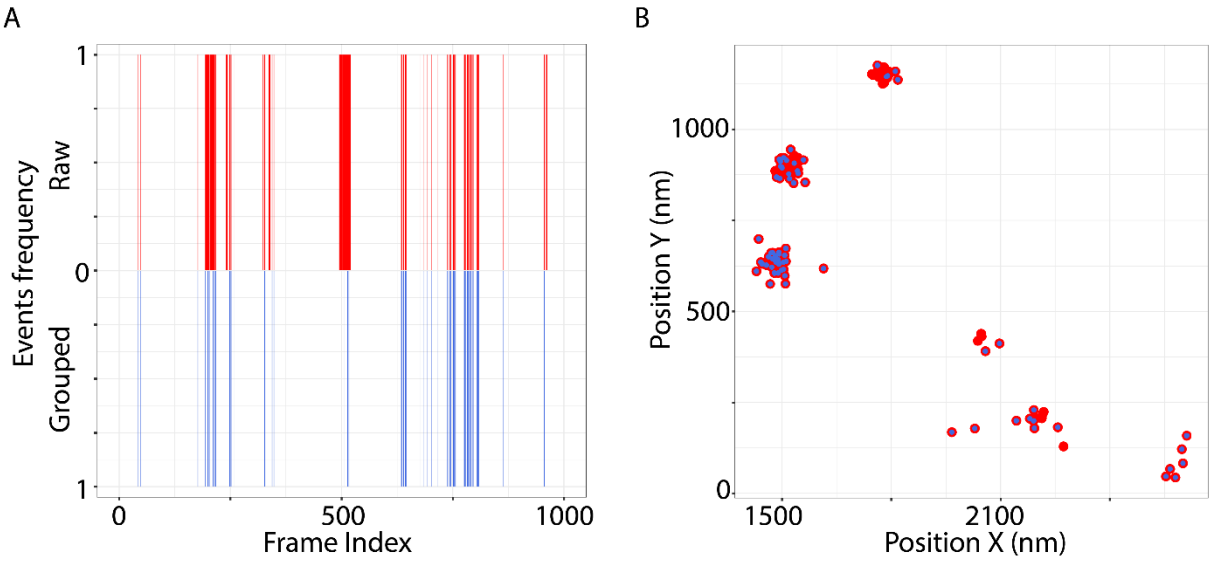


Figure 2.2: Spatiotemporal grouping of PAmCherry. (A) Fluorescent events frequency over 1000 frames in a ROI containing a single cell expressing PAmCherry before (red) and after grouping (blue). The temporal distribution of events across the experiment is not uniform with a high number of consecutive frames containing events at a Frame Index of 200 and 500. **(B)** Visual representation of the events recorded in **(A)** with overlapping between raw data (red) and grouped data (blue). Multiple events were grouped together when within 30 nm of each other in consecutive frames for up to 3 frames.

the off frames parameter will always be equal to 0 and the consecutive frames parameter is an estimate of the maximum number of frames a fluorophore can be active.

The search radius for fluorescence persistence was chosen to be on the same order of magnitude of the localization precision observed for the fluorophore itself (typical localization precision: 20 nm, search radius: 30 nm) while the consecutive frames parameter was stepwise increased until the localizations characterized by a frame number equal to the maximum allowed value accounted for less than 5% of the total localizations count (Table 2.2).

Table 2.2: Spatio-temporal grouping parameters. Grouping parameters were determined as described in chapter 2.2 and applied to all PALM data acquired within this dissertation prior analysis (the off frames are always set to 0).

Fluorophore	**Number of frames (% of persisting events)**					**Search radius (nm)**
	1 (%)	**2 (%)**	**3 (%)**	**4 (%)**	**5 (%)**	
PAmCherry	83.49	11.58	4.92	-	-	30
Dendra2	91.79	6.47	1.74	-	-	30
mNeonGreen	60.17	19.48	10.39	5.63	4.33	30

Given that localization precision is inversely proportional to the number of photons emitted by a fluorophore, as described in chapter 1.3.1, and that the fluorophores I used emit photons within the same order of magnitude, it does not come as a surprise that the search radius adopted for the grouping does not vary between them (Table 2.2). In contrast, the persistence length varies across the three fluorophores. Specifically, mNeonGreen events showed a higher degree of persistence (Table 2.2).

By combining careful experimental planning (i.e.: fluorophore choice, activation laser titration, …) and the mentioned filtering methods (ROI based, photon count based, PSF based, grouping based) it is possible to greatly reduce the degree of background signal, therefore laying the foundations for experimental reproducibility and a robust statistical analysis.

Chapter 3: Characterization of DipC spatio-temporal localization via conventional-like image analysis

Scripts and macros used within this chapter:

- Script_R3
- Script_R4

In order to apply conventional-like image analysis methods to PALM data, it is first necessary to extract the fluorophore localizations from the PALM result table (Table 1.2) and convert them back into an image. Two main approaches can be taken when converting the data from a list of localizations to an image: a PSF based approach (i.e.: Normalized Gaussian rendering) and a cumulative based approach.

In a PSF based approach each single localization will be represented as a point spread function, where the intensity and the width of the PSF can be extracted from the PALM table itself (photon count and PSF width). In a way, the resulting image will be extremely close to what one would obtain via conventional fluorescence microscopy, with the difference that the pixel size is not limited by the camera itself, that the fluorescence dynamic range is infinite and that the image resolution is higher (Figure 3.1 A). As we will see in chapter 3.1 this reconstruction approach results in greater loss of information compared to the second approach and is not suitable for most type of analysis.

While less visually appealing and simpler from a mathematical point of view, the cumulative based approach allows for quantitative analysis and is therefore generally preferred to Gaussian rendering.

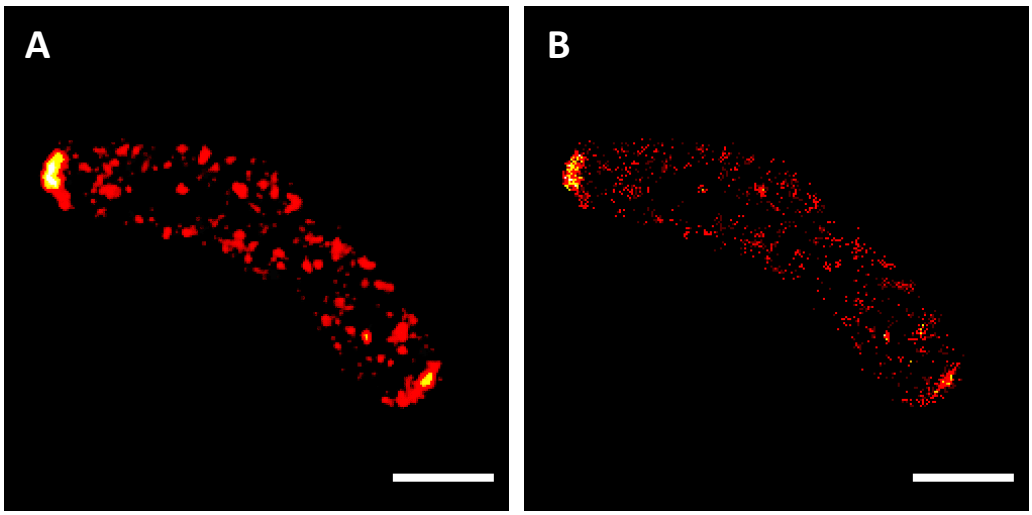


Figure 3.1: Conversion of PALM localizations to images for conventional-like analysis. DipC-PAmCherry localizations are extracted from the PALM result table and plotted on a 2D plane as either Normalized Gaussian distributions **(A)** or as Cumulative Histograms **(B)** via ThunderSTORM. Scale bar: 1 µm

Briefly, each pixel´s gray value corresponds to the number of localizations residing within the pixel itself (Figure 3.1 B). It follows that while the exact position of each fluorophore is lost, quantitative analysis is, to a certain degree, still feasible.

Both rendering methods are characterized by loss of information. However, the transformation of localization data into an image also translate into an easier visual interpretation and analysis. Such transformation is especially beneficial when the focus of the study is the analysis of the spatial organization of protein structures. It is important to notice that the increase in simplicity is not exclusively intrinsic of the data-type (pixel intensity rather than localization coordinate) but also derives from the greater availability and simplicity of use of image analysis software such as Fiji (Schindelin et al. 2012) when compared to the tools available for point pattern analysis (usually R and/or Matlab packages).

Here, I showcase the potential use of conventional-like image analysis in unique and general scenarios. Briefly, I exploit the septal/polar enrichment of two putatively interacting proteins in *Corynebacterium glutamicum*, DipC (cg0839) and DivIVA (cg2361), to unravel their relation via conventional-like image analysis. Moreover, I exemplify how fixed cell PALM imaging and conventional-like image analysis can be used to describe and compare changes in protein localization across a cell cycle and/or genomic background.

3.1. The identification of novel DivIVA interaction partners in *Corynebacterium glutamicum*: the DNA damage induced proteins C and D (DipCD)

In *C. glutamicum*, the most evident example of spatial organization is the DivIVA-dependent polar localization of the cell elongation machinery (Daniel and Errington 2003, Letek et al. 2008). DivIVA is a conserved protein among many gram-positive bacteria that is able to autonomously localize at regions characterized by high degrees of negative curvature, such as poles and septa, via a process called molecular bridging (Lenarcic et al. 2009). Briefly, DivIVA subunits form dimers which, in turn, can interact with each other to form a net-like structure characterized by an intrinsic curvature. As DivIVA subunits also interact with the lipid bilayer, the oligomeric structure is more likely to form/is more stable when located within regions characterized by an intrinsic curvature similar to the one of the oligomers themselves: poles and newly forming septa (Lenarcic et al. 2009).

DivIVA oligomers function as scaffold for a series of essential processes. Known mechanisms that rely on DivIVA for their localization in *C. glutamicum* include the cell elongation machinery, via the transglycosylase RodA (*cg0061*) (Sieger and Bramkamp 2014), and the chromosome segregation

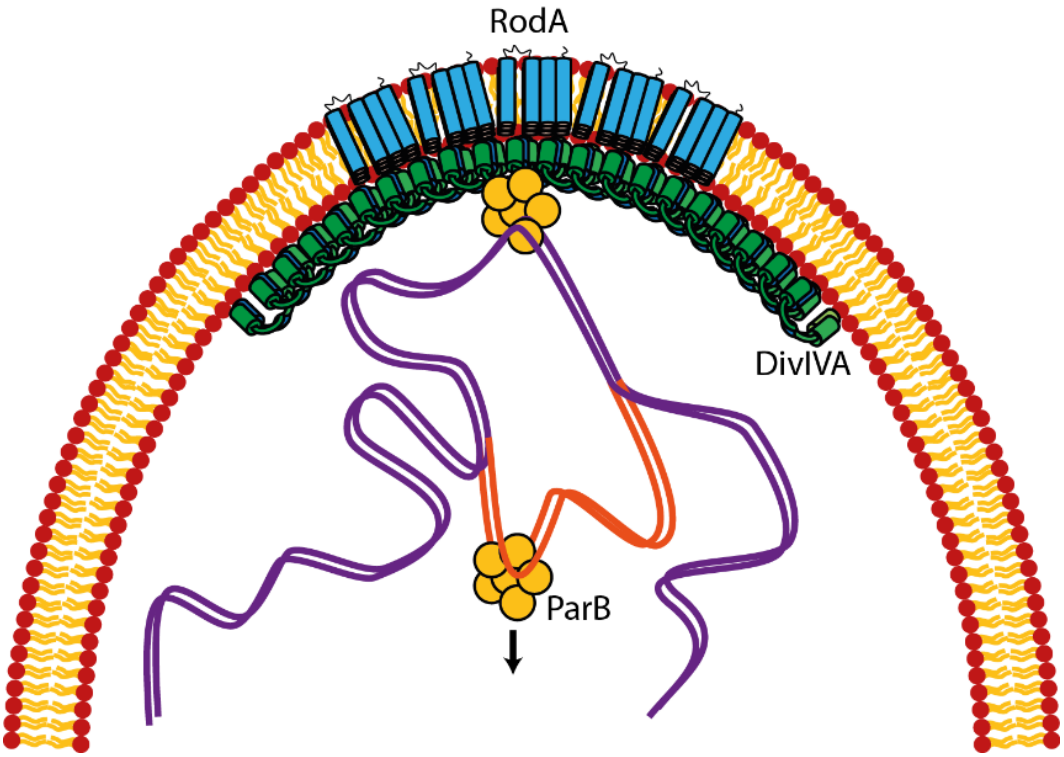


Figure 3.2: DivIVA dependent organization of the pole in *C. glutamicum*. The membrane associated DivIVA scaffold (in green) interacts with members of the elongasome (RodA – in blue) on the membrane side and with the chromosome segregation system (ParB – in yellow) on the cytoplasmic side.

machinery, via the ParB-mediated (*cg3426*) tethering of the ParABS system (Donovan et al. 2012, Bohm et al. 2020) (Figure 3.2). Members of the DivIVA superfamily are known to interact with a variety of different processes in different bacterial species (Kaval and Halbedel 2012). The ability of DivIVA members to interact with such a different array of proteins has been attributed to its least conserved region: the C-terminal region (Kaval and Halbedel 2012). As neither ParB nor RodA interact with said region (Donovan et al. 2012, Sieger and Bramkamp 2014), it is reasonable to hypothesize the existence of more DivIVA dependent pathways in *C. glutamicum*. Thus, I fished for novel DivIVA interaction partners via a pull-down experiment followed by mass spectrometry (full description of the method in Chapter 8.12.5-6). Briefly, I used colloidal Coomassie in order to compare the protein elution profiles obtained for the strain expressing the bait protein (strain C27: *C. glutamicum* RES167* *divIVA::divIVA-StrepTag*) and the control strain (*C. glutamicum* RES167*) (Figure 3.3). I then isolated the parts of the elution profile that differ between the C27 strain and the control (for both C27 and control strain), identified the proteins contained within said samples via mass spectrometry (mass spectrometry was performed by the MSBioLMU service unit) and finally compared the list of proteins obtained for each of the four C27 regions with the ones obtained for RES167*.

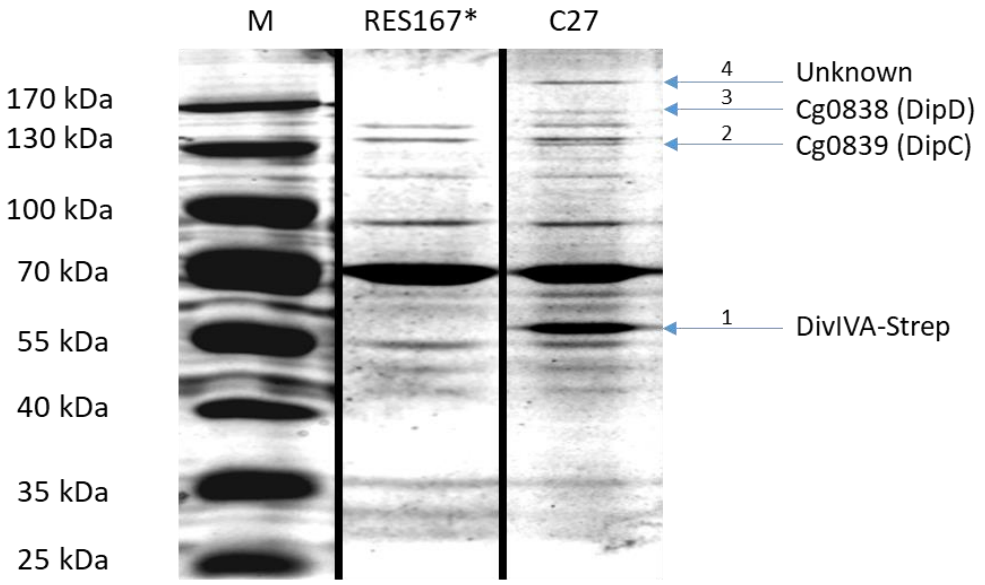


Figure 3.3: Pull-down elution profiles of RES167* and C27. While the elution profile of the control strain (*C. glutamicum* RES167*) contains several protein bands, it is possible to visually identify four unique protein bands within the C27 elution profile (numbered 1 to 4). The unique bands 1 to 3 can be attributed respectively to DivIVA-StrepTag (theoretical molecular weight: 39.67 kDa), Cg0839 (theoretical molecular weight: 139.33 kDa) and Cg0838 (theoretical molecular weight: 179.00 kDa). No significant difference could be found via mass spectrometry for band 4 between C27 and RES167*.

Out of the four protein bands unique for the C27 strain, one could be attributed to the bait protein (DivIVA-Strep), two could be attributed to putative proteins (Cg0838 and Cg0839) while for the last one it was not possible to find differences with the respective RES167* region (Figure 3.3).

Where the DipD (Cg0838) protein sequence is predicted to contain a DEAD, a DUF1998, a ResIII and a Helicase C motif, strongly suggesting that DipD is a helicase, the DipC (Cg0839) protein sequence is characterized by the lack of motifs/domains (the list of PFAM motifs contained within the two protein sequences is provided by the Kyoto Encyclopedia of Genes and Genomes - KEGG) (Kanehisa and Goto 2000, El-Gebali et al. 2019). DipC is also characterized by the lack of homology with other known protein sequences, with only closely related microorganisms expressing candidate homologous proteins (i.e.: *Corynebacterium ammoniagenes KCCM 40472, Corynebacterium aurimucosum DSM 44827, …*) (candidate homologous proteins were identified via CoryneRegNet 7) (Parise et al. 2020). Finally, both proteins are predicted to not contain transmembrane helices (prediction via the PredictProteins online tool) (Yachdav et al. 2014).

While the bioinformatic approach fails to provide meaningful data concerning DipC localization and function, it provides instead information concerning its expression. Specifically, it has been previously shown that a *lexA* box is located upstream *cg0841* (*dipA*) (Figure 3.4 A) (Jochmann et al. 2009), and microarray data suggest that both *dipA* and *dipC* undergo an increase in transcription upon

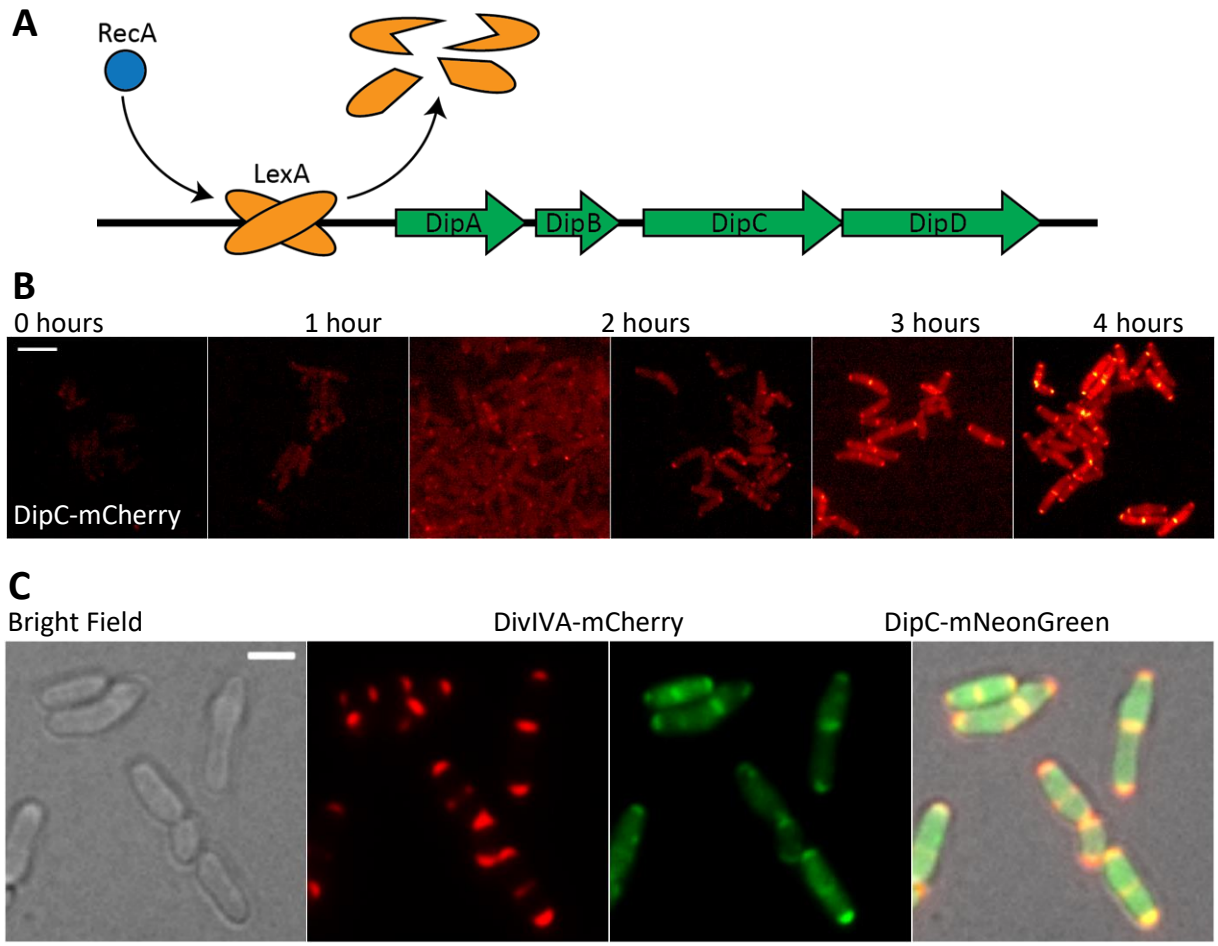


Figure 3.4: DipC is expressed upon DNA damage and localizes in a DivIVA fashion. (A) *dipC* transcription is regulated by a lexA box, a DNA sequence that is bound by LexA in absence of DNA damage inhibiting its transcription. **(B)** over-time expression of DipC-mCherry from its native locus (*C. glutamicum* RES167* *dipC::dipC-mCherry* - B1G2) following the addition of 200 ng/ml of mitomycin C. Scale bar: 5 µm **(C)** Co-expression of DivIVA-mCherry and DipC-mNeonGreen from their native loci (*C. glutamicum* RES167* *divIVA::divIVA-mCherry dipC::dipC-mNeonGreen* - B1F10) in presence of the DNA damage inducing agent mitomycin C (200 ng/ml). Scale bar: 2 µm

addition of the DNA damage inducing agent mitomycin C (Jochmann et al. 2009, Pauling et al. 2012). Based on said predictions I treated the *C. glutamicum* RES167* *dipC::dipC-mCherry* strain (B1G2) with 200 ng/ml of mitomycin C and followed the changes in DipC expression (Figure 3.4 B). Following the addition of the antibiotic, the B1G2 strain showed a gradual increase in fluorescence suggesting that *dipC* expression is upregulated in presence of DNA damage (Figure 3.4 B). Given the genetic proximity and shared regulatory behavior of *cg0839* and *cg0841*, I tentatively assigned *cg0841-38* to the same putative operon: the DNA damage induced proteins ABCD (*dipABCD*) operon. I included *dipD* within the putative operon due to its close genetic proximity to *dipC* (8 bp).

As DipC-mCherry localization is not homogeneous but rather characterized by a combination of low, diffusive cytoplasmic signal and, similarly to DivIVA, an enriched polar and septal signal (Figure 3.4 B), I constructed a strain that express fluorescently labeled versions of both DipC and DivIVA from

their native loci: RES167* *divIVA::divIVA-mCherry dipC::dipC-mNeonGreen* (B1F10) (Figure 3.4 C). The imaging of the two fluorescently labeled proteins via conventional fluorescence microscopy show a shared pattern in presence of DNA damage (Figure 3.4 C), in agreement with the putative interaction between DipC and DivIVA.

3.2. The septum as a reference: Normalized Gaussian rendering

The distance between the two sides of a septum in *C. glutamicum* is both smaller than the resolution achieved via conventional fluorescence microscopy and bigger than the resolution that can be achieved via PALM (Marienfeld et al. 1997). It follows that proteins such as DivIVA and its interaction partners, which localize on both sides of the septum, will appear as a single structure (Figure 3.4 C) when imaged via conventional localization microscopy and as two separate structures when imaged via PALM (Figure 3.5 A-B). I hereby combine the increase in resolution obtained via the use of PALM microscopy, conventional-like image analysis and the unique characteristics of the septum to determine the relative positioning of DivIVA and DipC.

This approach does not only shed light on the DipC-DivIVA relation but can also function as a general approach toward the study of septal organization. Given that the septum width distribution does not vary between the strains used, in this case *C. glutamicum* RES167 *dipC::dipC-PAmCherry* and RES167 *divIVA::divIVA-PAmCherry*, it is possible to image the tagged proteins via PALM independently, and use the reconstructed images to measure the distance between the two sides of the septum (Figure 3.5). It follows that each protein of interest can be tagged with the same fluorophore, therefore eliminating fluorophore dependent variability and reducing changes in imaging conditions.

Specifically, I imported the reconstructed images in Fiji and manually defined linear ROIs of 600 nm width spanning from pole to pole of cells containing a septum (Figure 3.5 A-B). I then extracted the intensity values along each ROI, where each value is the average across the width of the ROI itself at position X, and used them to plot the correspondent cell fluorescence profile (Figure 3.5 A-B). I then calculated the distance between the intensity maxima that surround each septum and fitted the resulting septal distances populations via normal distributions (Figure 3.5 C-D). Since the mean of the distribution for DivIVA (a membrane associated protein) is approximately 110 nm and is smaller than the one obtained for DipC (∼170 nm) I concluded that DipC proximity to the septum does not derive from a direct interaction with the membrane but rather due to protein dependent anchoring, possibly due to DivIVA (Figure 3.5 E). Similarly, this approach can be applied to multiple septum associated proteins in order to construct a more complex hierarchical structure.

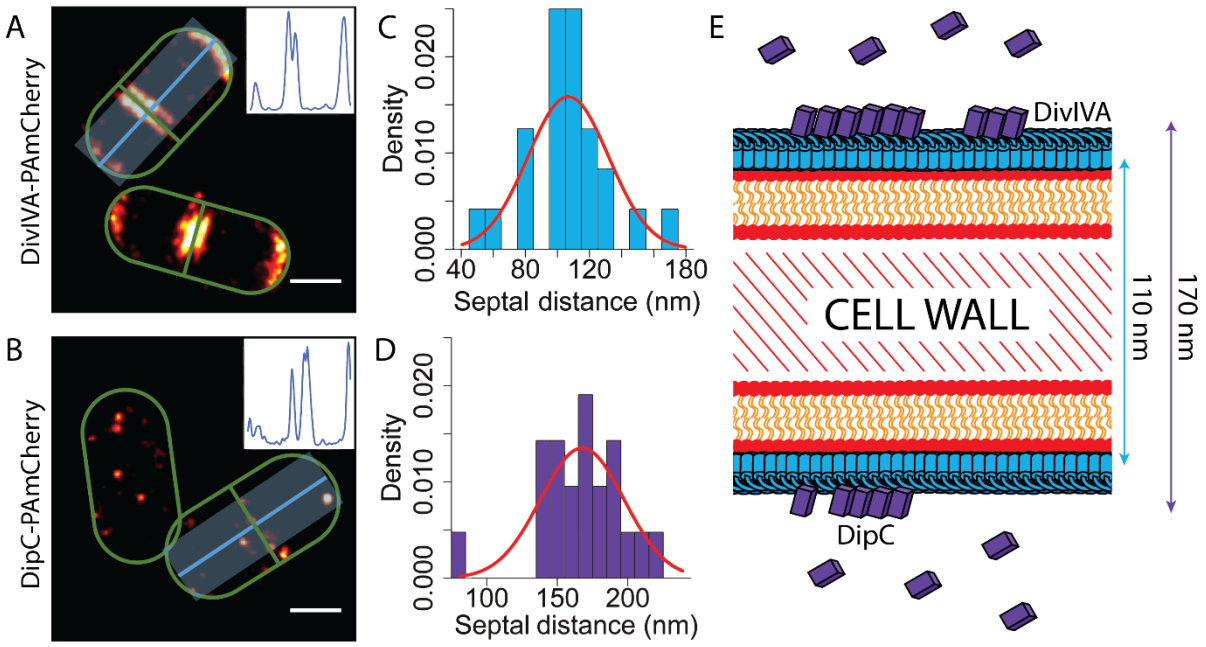


Figure 3.5: DipC is recruited to the septum by a protein, possibly DivIVA. Normalized Gaussian rendering of *C. glutamicum* RES167 *divIVA::divIVA-PAmCherry* **(A)** and RES167 *dipC::dipC-PAmCherry* **(B)** and associated fluorescence profile **(top-right corner).** The fluorescence profiles of cells presenting septa **(A-B)** were used to calculate the distance between fluorophores localizing on the two sides of the septa themselves. The septal distance population of DivIVA **(C)** and DipC **(D)** were fitted via a normal distribution and compared. The relative positioning of DivIVA and DipC is shown via a reductive model of the septum **(E)**. Scale bar: 0.5 µm.

3.3. Analysis of DipC mislocalization in a *divS* knock-out: Cumulative Histogram rendering and quantitative profiles

When using conventional fluorescence microscopy or Normalized Gaussian rendering the fluorescence profiles obtained from different fields of views/strains cannot be compared from a quantitative point of view. This means that the associated analysis is usually limited to the description of the changes in fluorescence profile and/or the tracking of a fluorescent spot over-time. In a Cumulative Histogram rendering the intensity values assigned to pixels correspond to the number of localizations that reside within the pixel itself (Figure 3.6 A). Assuming that all fluorophores have been imaged in the course of the PALM experiment, it is therefore possible to quantitatively compare Cumulative Histogram reconstructions obtained from different strains/in the course of different experiments from both an expression and a localization point of view (Figure 3.6).

The expression of genes related to DNA damage induced SOS response in bacteria is generally controlled by RecA and LexA (Witkin 1976, Walker 1984). In *C. glutamicum*, LexA has been demonstrated to directly control the expression of 48 SOS genes involved in a variety of physiological functions comprising the repair of the DNA damage and the survival of the cell during/following the repairs (Jochmann et al. 2009). The survival of the cell is ensured via a multitude of processes, including

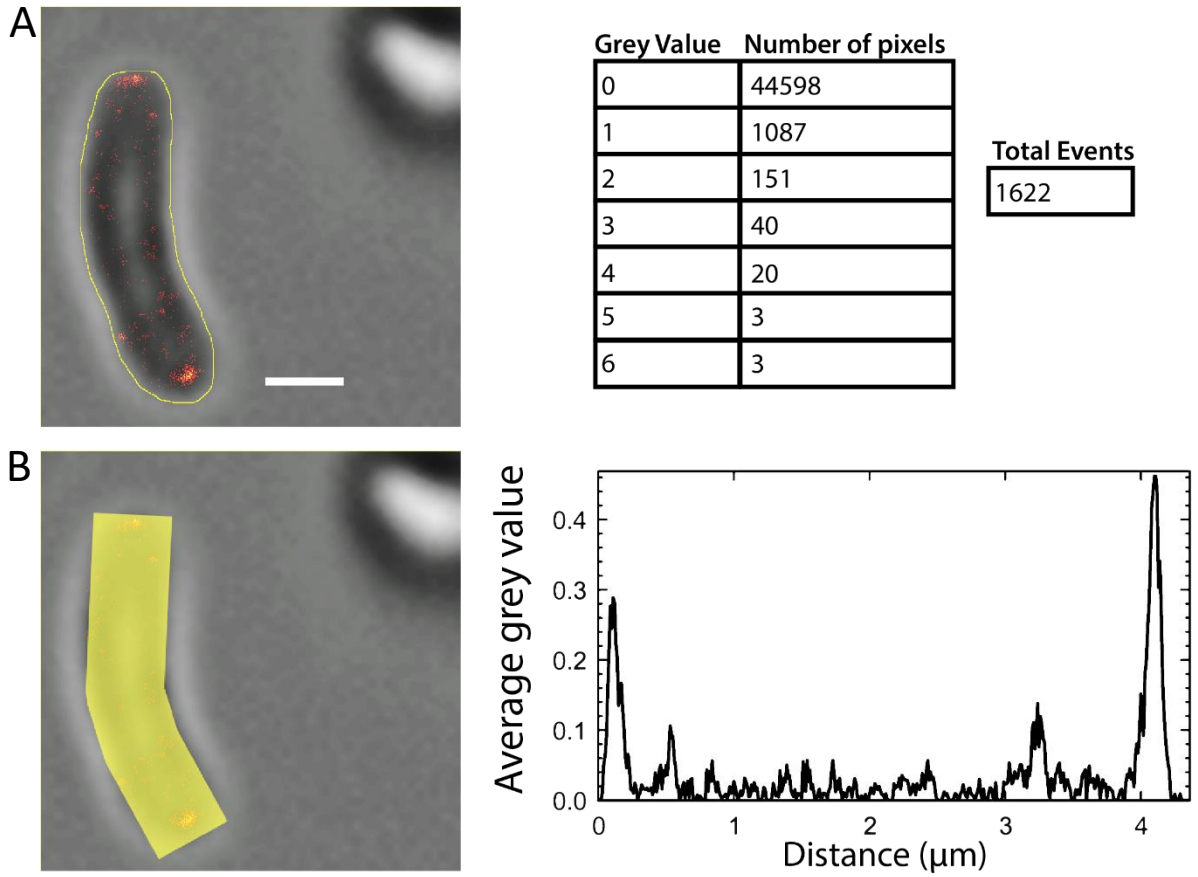


Figure 3.6: Analysis of Cumulative Histogram rendering data. (A) Example of extrapolation of quantitative information from cumulative histograms. The integration of the pixels´ grey values present within the focused ROIs obtained during the pre-analysis data filtering (2.1.1) equals to the number of fluorescence events perceived within the cell. **(B)** The determination and analysis of fluorescence profiles of Cumulative Histogram rendering data follows the same procedure used for Gaussian renderings in 2.2.1. The fluorescence profile **(right)** represents the average grey value obtained along the length of the cell for a segmented line of 1 µm width. Scale bar: 1 µm.

and not limited to the DivS-dependent inhibition of cell division (Ogino et al. 2008). As DipC is expressed during the DNA damage-dependent SOS response and localizes both in proximity of the cell elongation machinery and the divisome, it is possible for it to be involved in the regulation of said processes.

I therefore compared the localization and expression of DipC-PAmCherry during the DNA damage-induced SOS response in a wild type background strain (RES167 *dipC::dipC-PAmCherry*) (B3B3) and in a strain lacking the protein responsible for cell division inhibition during the SOS response: DivS (RES167 *ΔdivS dipC::dipC-PAmCherry*) (B5E2). While the two strains appear morphologically identical under standard growth conditions (Ogino et al. 2008), the knock out strain does not show the cell elongation phenotype typical of *C. glutamicum* cells exposed to DNA damage-inducing agents (Kijima et al. 1998, Ogino et al. 2008) (Figure 3.7 A-B). I used the focused ROIs obtained during the pre-analysis data filtering phase (Chapter 2.1) to measure the cell areas of the two strains. As the cell area

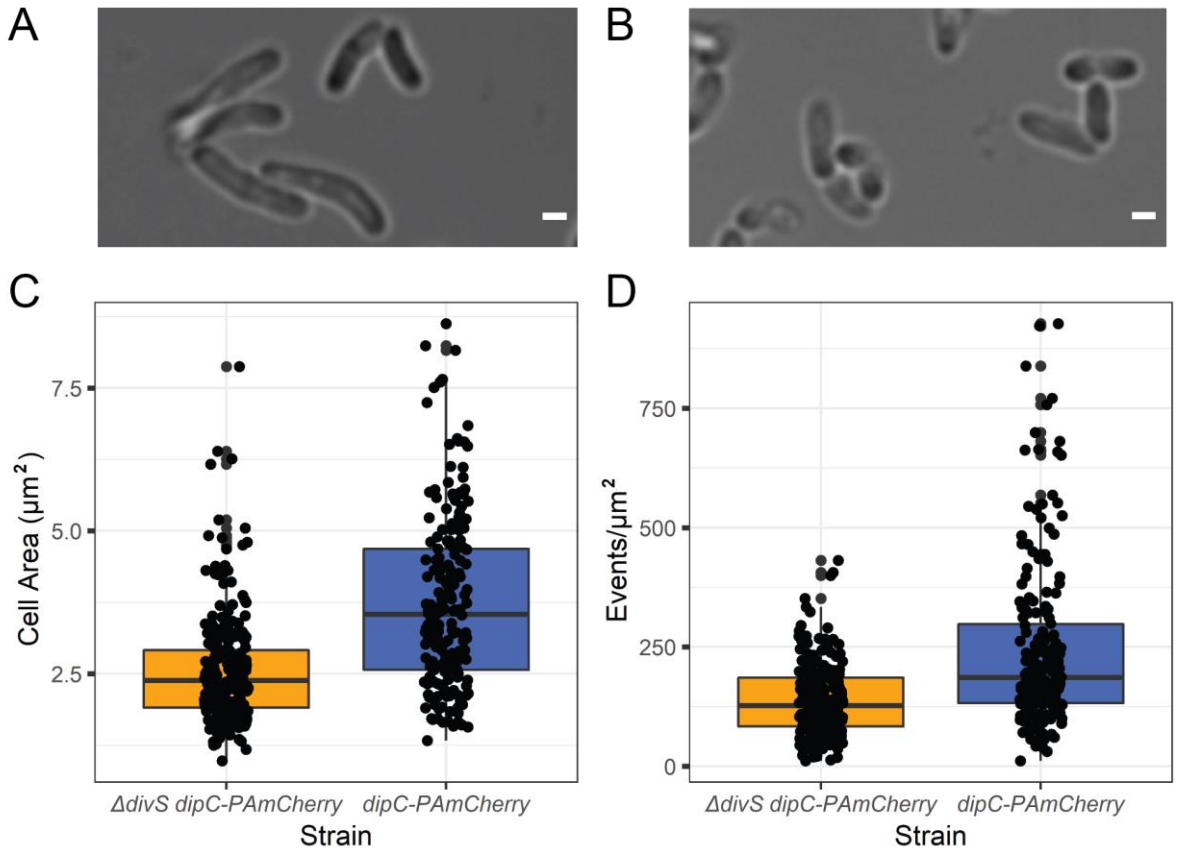


Figure 3.7: Absence of the cell division inhibitor protein DivS results in a decreased DipC protein density during the SOS response phase. *C. glutamicum* WT cells show an increase in cell length in presence of DNA damage due to inhibition of cell division **(A)** while cells lacking DivS lack such elongation phenotype **(B)**. As the difference in cell area size distribution is statistically significant **(C)**, I used DipC density to compare the protein expression levels across the different strains **(D)**. Both cell area and DipC density distributions were compared via one-sided Wilcoxon rank sum test. Scale bar: 1 µm.

distributions of the two strains differ significantly, with *ΔdivS* being characterized by smaller cells (one-sided Wilcoxon rank sum test with continuity correction, W = 34894, p-value < $2.2 * 10^{-16}$) (Figure 3.7 A-C), I normalized the number of events registered within each cell by the area itself ($events/\mu m^2$) before comparing the levels of DipC expression (Figure 3.7 D). The resulting fluorescence events density is statistically lower in the RES167 *ΔdivS dipC::dipC-PAmCherry* strain (one-sided Wilcoxon rank sum test with continuity correction, W = 33188, p-value < $1.21 * 10^{-14}$) suggesting that DivS may influence DipC levels.

Differently from the determination of the relative positioning of DipC and DivIVA (Chapter 3.2), the aim of the fluorescence profile analysis is here to provide information concerning the spatio-temporal localization of DipC in presence and absence of DivS during the SOS response. While cell fixation is a requirement for the complete imaging of the DipC protein population at a given time, it is also true that a fixed sample is a static representation of every possible state under which a cell can exist under the defined experimental conditions, whether we are talking about the stages of a bacterial cell cycle

(B, C, D period) or the stages a cell undergoes during the SOS response. Since the age of a cell is typically correlated to its length, it is possible to order the fluorescence profiles accordingly and get a first insight toward the spatio-temportal localization of the protein itself (Script_R3). DipC-PAmCherry, in agreement with previously shown data (Chapter 3.1 and chapter 3.2), is characterized by an increased

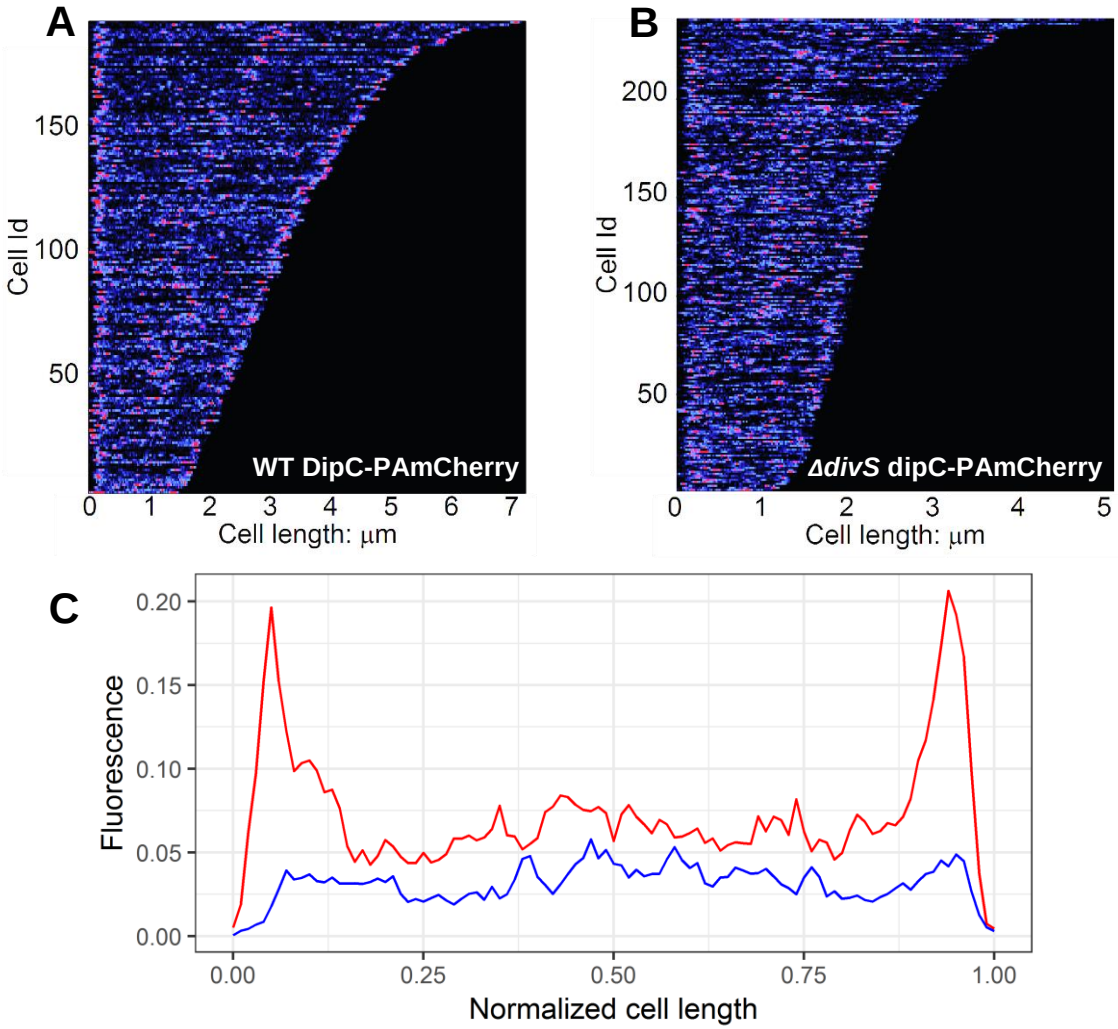


Figure 3.8: Visualization of the spatio-temporal localization of DipC-PAmCherry via demographs. (20 nm intervals). DipC-PAmCherry localization in presence in presence **(A)** and absence **(B)** of DivS in presence of mitomycin C (200 ng/ml). The correlation between absolute fluorescence intensity and the palette color is not maintained across cells (Fluorescence intensity is normalized so that each cell uses the full dynamic range permitted by the palette). Color Palette: (0) black -blue-red-white (max). Average fluorescence profiles of RES167 *dipC::dipC-PAmCherry* (red, n=192) and and RES167 *ΔdivS dipc::dipC-PAmCherry* (blue, n=235) **(C)**.

localization in polar and septal areas (Figure 3.8) in presence of mitomycin C. Surprisingly, its localization does not appear as well defined in absence of the cell division inhibitor protein DivS (Figure 3.8 B). In order to compare the average DipC enrichment areas for the two genetic background, the fluorescence profiles were normalized by cell length and averaged (Script_R4 – Figure 3.8 C). The

resulting averaged profiles confirmed the presence of a less pronounced polar localization in the knock out strain (Figure 3.8 C). As *C. glutamicum* divides asymmetrically (Messelink et al. 2020), the septal localization of DipC does not translate to a sharp increase in midcell fluoresce when normalized across the cell population (Figure 3.8 C). The fluorescence increase is instead spread across a broad area making it not suitable for in depth analysis (Figure 3.8 C).

Chapter 4: Single molecule localization microscopy data analysis

Scripts and macros used within this chapter:

- Fiji macro 1
- Fiji macro 2
- Script_R1
- Script_R2
- Script_Rv
- Script_R5P1
- Script_R5P2

Results included within this chapter were published in:

- Toro-Nahuelpan, M., et al. (2019). "MamY is a membrane-bound protein that aligns magnetosomes and the motility axis of helical magnetotactic bacteria." Nature Microbiology 4(11): 1978-1989.

The methods dedicated to the study of the spatial arrangements of points, whether it is on a 2D or a 3D area, have not been specifically developed for the study of PALM data and no consensus has yet been reached concerning which method is supposed to be used to answer specific questions. In the same way that multiple software packages have been developed in order to extract fluorophores localizations from PALM raw images (Sage et al. 2019), several scientific groups working with single molecules localization microscopy techniques developed their own analysis approach method to answer specific questions (Malkusch et al. 2012, Levet et al. 2019, Paul et al. 2019). I wrote a step-wise R script for the analysis of SMLM data (Giacomelli 2020) with the capability to extrapolate and combine a large variety of parameters (clustering, co-localization, events density, cell morphology) and to filter them from both a population and single cell point of view. The script is focused mainly toward the collection, translation (some of the methodologies were not available in R (Malkusch et al. 2012)) and integration of a series of known point pattern tools that, when combined, output a relatively complete description of said pattern. Finally, the separate part of the script can be used independently from each other and will be described as such in order to better exemplify their function within the script, however they greatly benefit from each other (i.e.: estimation of G(r) as a pre-requisite for OPTICS, cluster dependent co-localization analysis, …) and are meant to be used together.

4.1. About the distribution of points on a 2D plane and magnetosomes

In order to analyze the intracellular distribution of a protein of interest it is important to have two sets of data: the localization of the subunits, that can be obtained via SMLM, and the area in which these subunits localize (ROIs), which can be extrapolated via bright field/phase contrast images or other methodologies. As the two data sets are already available following the application of the filtering script (Script_R2), it is possible to perform a first analysis of the point pattern without further

data processing (Script_Rv). This first exploratory step consists of the estimation of the nearest neighbor distribution function $G(r)$ (edge corrected via one of the following estimators: border method estimator, Kaplan-Meier estimator (Baddeley and Richard 1997) and Hanisch estimator (Hanisch 1984)) and its comparison with the true value of G for a completely random process (Poisson). This step is a semi-mandatory prerequisite when the aim of the study is to characterize clusters (Figure 4.1) but it can also be used, for instance, to determine the modality and conditions under which a protein is able to form oligomers (Toro-Nahuelpan et al. 2019) (Figure 4.2).

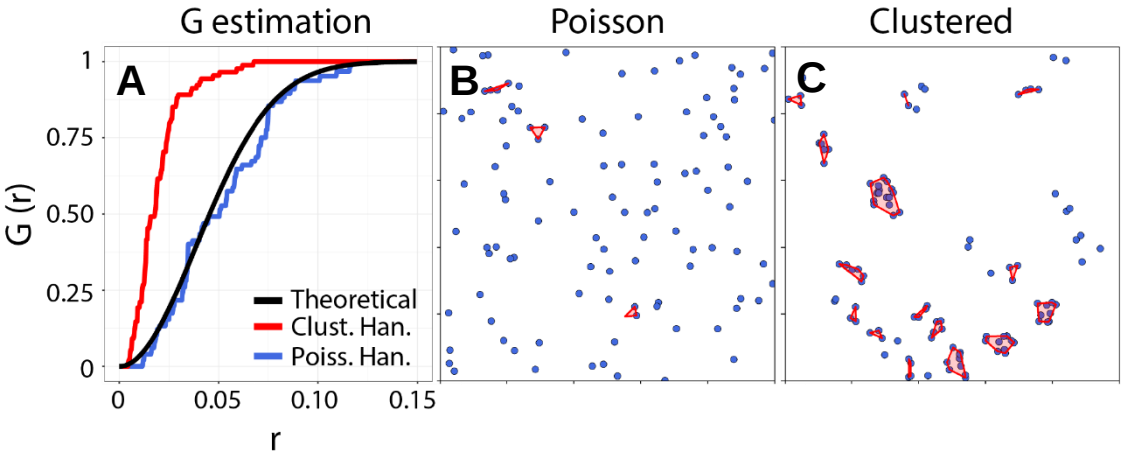


Figure 4.1: Clustering and point pattern distributions. Comparison of the $G(r)$ obtained from simulated data following a Poisson distribution (blue) and a clustered distribution (red – Matern cluster process) with a theoretical Poisson distribution (black) **(A)**. The $G(r)$ for both simulated dataset are corrected via the Hanisch estimator. Visualization of the simulated dataset and of the respective clusters identified via OPTICS (same parameters) **(B-C)**.

The necessity of determining the deviation of $G(r)$ from the G of a Poisson process prior to the identification and study of the protein clusters themselves (Figure 4.1 A) is caused by the fact that algorithms dedicated to the identification of clusters such as OPTICS focus exclusively on the identification of clusters without specifying whether the clusters themselves originated due to a Poisson process (Figure 4.1 B) or, in the case of a protein, due to a specific biological process (Figure 4.1 C). The ability to determine whether the heterogeneity in protein distribution observed within a cell is to be expected due to randomness or is rather the result of biological interactions is crucial when studying oligomerization.

In the course of the collaboration with Dr. Toro-Nahueln concerning MamY, a membrane bound protein responsible for the correct alignment of the magnetosome in *Magnetospirillum gryphiswaldense*, we could show via 3D-SIM and PALM that the natively expressed mCherry-MamY (strain FM52) preferentially localize along the geodetic axis (Figure 4.2A) (Toro-Nahuelpan et al. 2019). As single protein subunits are too small to be able to recognize cell membrane curvatures and bacterial two-hybrid data suggested that MamY subunits are able to interact with one another, it became

evident that oligomerization plays a pivotal role toward the recognition of the geodetic axis. We therefore set out to determine how MamY subunits come together to form said geodetically cell-spanning structure.

In order to follow the *de novo* emergence of MamY structures, we expressed Dendra2-MamY under the control of a tetracycline promoter (P_{tet}) in a *ΔmamY* strain (MT022) and exploited the resulting expression variability to study its clustering behavior via PALM. The estimation of the nearest neighbor distribution function $G(r)$, corrected via the Hanisch estimator, deviates from a Poisson distribution for all the average localizations densities tested (Figure 4.2 B-F) (average localizations density is here a proxy for protein concentration) highlighting the tendency of MamY to cluster at both high (Figure 4.2 B) and low (Figure 4.2 F) cellular concentrations. Next, I further processed the data with the goal to identify and characterize the nature of the clusters.

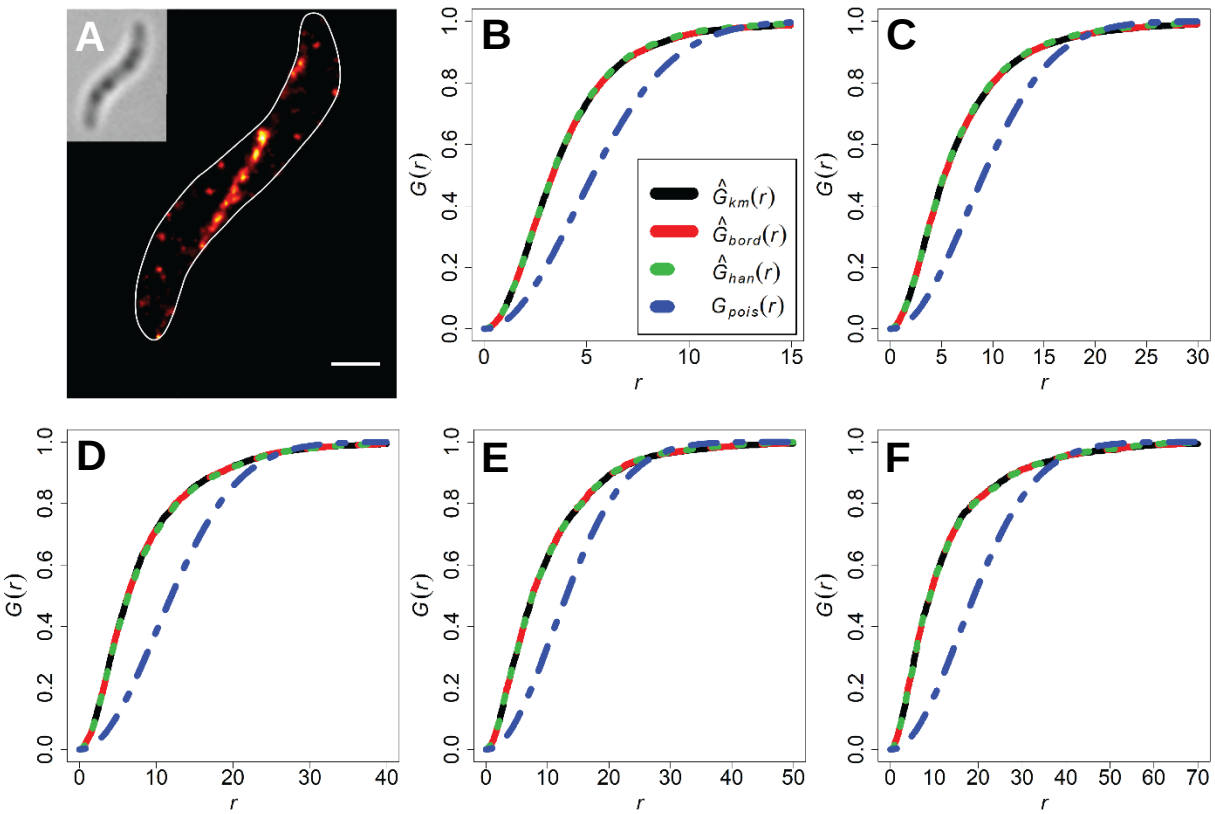


Figure 4.2: MamY form clusters at a wide variety of protein concentrations. Gaussian rendering of MamY-Dendra2 localizations in a *M. gryphiswaldense* WT background strain (strain eMT018) **(A)**. Estimation of ectopically expressed Dendra2-MamY nearest neighbor distribution $G(r)$ at five different protein concentrations (including the three available edge correction methods: border method, Kaplan-Meier and Hanisch and the associated theoretical Poisson distribution) in a *ΔmamY* background strain (strain MT022) **(B-F)**. The localizations density is shown in decreasing order from B ($\sim 7800\ localizations/\mu m^2$) to F ($\sim 600\ localizations/\mu m^2$). Scale bar: 500 nm.

4.2. Local density and nearest neighbor distance

Following the application of the filtering script (Script_R2) the data table is composed exclusively by filtered localization data and cell related parameters extrapolated from the ROIs used for the filtering itself (Cell ID, Cell Diameter, Cell Area). Script_R5P1 expands on said table by calculating a series of new localizations centered parameters such as nearest neighbor distances and local density (Figure 4.3 A-F), where local density is defined as the number of localizations contained within a squared area of lateral side equal to 50 nm centered on the localization itself.

I visualized the same Dendra2-MamY expressing cells that I used for the estimation of $G(r)$ in Chapter 4.1 in a local density dependent manner (Figure 4.3 A-F). In agreement with what was previously determined via the analysis of $G(r)$, I was able to observe signal heterogeneity for all the tested expression levels. Furthermore, each condition shows, to a certain degree, higher Dendra2-MamY local density in correspondence of the geodetic path of the cell (the tendency is more pronounced at higher protein concentrations). As local density and nearest neighbor distance are correlated, it is possible to observe the same pattern via the use of a nearest neighbor distance dependent color code (data not shown).

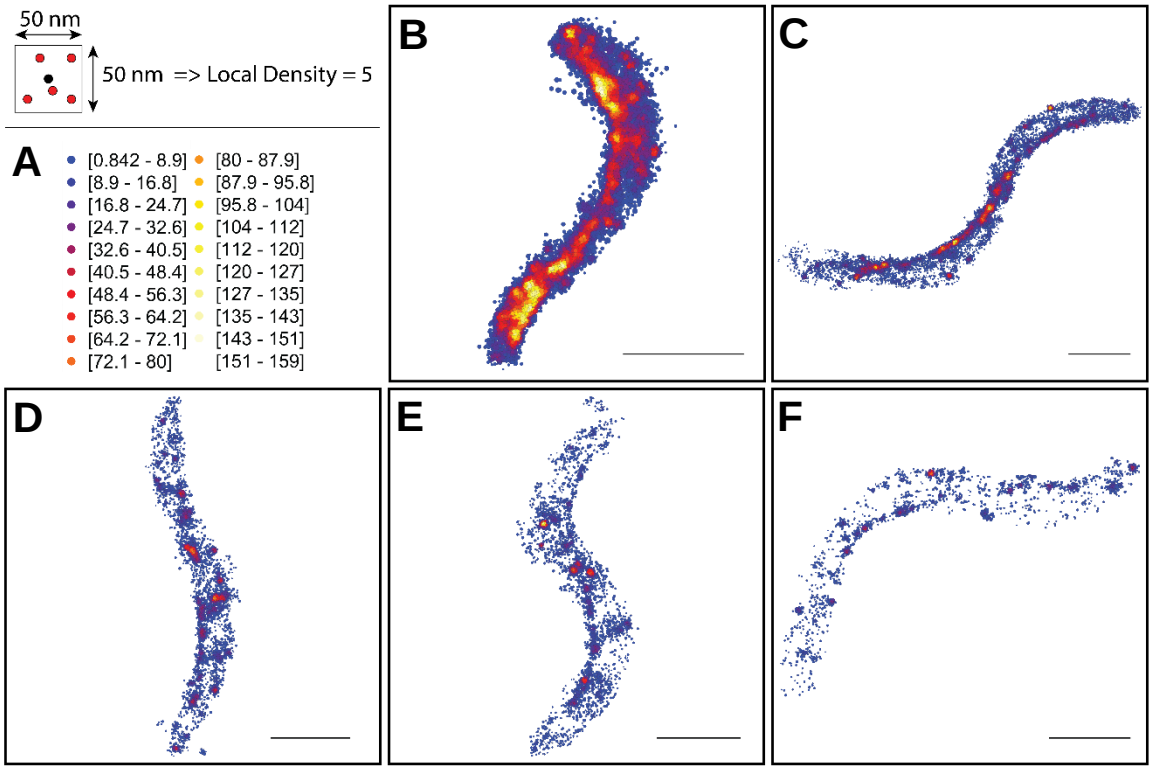


Figure 4.3: MamY clusters localize along the geodetic axis of the cell. Local density is expressed as the number of localizations residing within a square area (50 nm lateral side, 2500 nm^2) centred around the localization itself. The interval values shown in the legend are shared among the different plots **(A)**. Local density based visualization of MT022 cells ordered by decreasing Dendra2-MamY cellular concentration **(B-F)**. Scale bar = 1µm.

4.3. Cluster identification and characterization

So far, I showed that Dendra2-MamY subunits oligomerize across a variety of protein concentrations and that, at all protein concentrations tested, MamY is enriched along the geodetic path. I did not, however neither identified nor characterized the oligomers themselves. While it is possible to broadly deduce the localization of MamY clusters based exclusively on protein densities, the identification and characterization of oligomers allows for a deeper understanding of MamY clusters internal organization in relation, for example, to their size.

Following the application of Script_R5P1, the dataset is composed of cell related (Cell ID, Cell Diameter, Cell Area) and localization related (Nearest Neighbor Distance, Local Density) data categories. Script_R5P2 expands on said dataset by including clustering related categories (Cluster and Subcluster ID, Cluster and Subcluster Size, Subclusters/Cluster) and making readily available some commonly used filtering parameters (Localizations/Cell, Localizations/µm^2, Clusters and Subclusters/Cell, Clusters and Subclusters /µm^2).

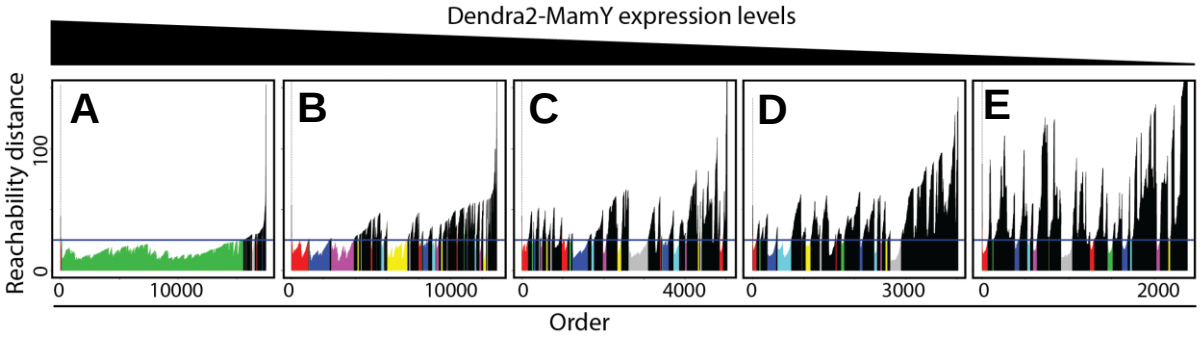


Figure 4.4: Identification of Dendra2-MamY clusters via OPTICS. Reachability plots of MT022 cells ordered by decreasing Dendra2-MamY cellular concentration **(A-E)** for $MinPts = 20$. Points lying below the ε_{limit} of 25 nm (blue horizontal line) are color coded and extracted as clusters via Script_R5P2.

Identification of clusters via OPTICS requires the selection of a parameter indicating the minimum number of points necessary to form a cluster ($MinPts$) and a threshold maximum distance at which these points need to be found (ε_{limit}) (Chapter 1.5.2). As the only parameter that changes across the five cells is the protein concentration itself, there is no reason to expect differences in Dendra2-MamY clustering requirements. It follows that a single combination of parameter and threshold is sufficient to identify the clusters across the different conditions. I therefore defined Dendra2-MamY clusters as a minimum of 20 points localizing within a circle of 50 nm in diameter ($MinPts = 20,\ \varepsilon_{limit} = 25nm$), where the maximum radius was chosen in order to maximize the ability of the algorithm to identify clusters across all the tested conditions (Figure 4.4). For simplicity of expression and clarity of form I will now refer to the separate Dendra2-MamY cellular concentration levels as “Expression Level” 1 to

5 (EL1-5), with EL1 being the one characterized by the highest protein expression (Figure 4.2 B, Figure 4.3 B and Figure 4.4 A) and EL5 being the one with the lowest (Figure 4.2 F, Figure 4.3 F and Figure 4.4 E).

Excluding EL1, where a single cluster includes the majority of the localizations (Figure 4.4 A), the reachability plots show a jagged appearance, with valleys (clusters) characterized by steep slopes (Figure 4.4 B-E). As the steepness of the slopes surrounding a valley can be used as a qualitative descriptor of the cluster border, with a steep slope mirroring a sharp cluster edge and a mellow one mirroring a gradient, we can conclude that Dendra2-MamY clusters are not the result of a protein gradient.

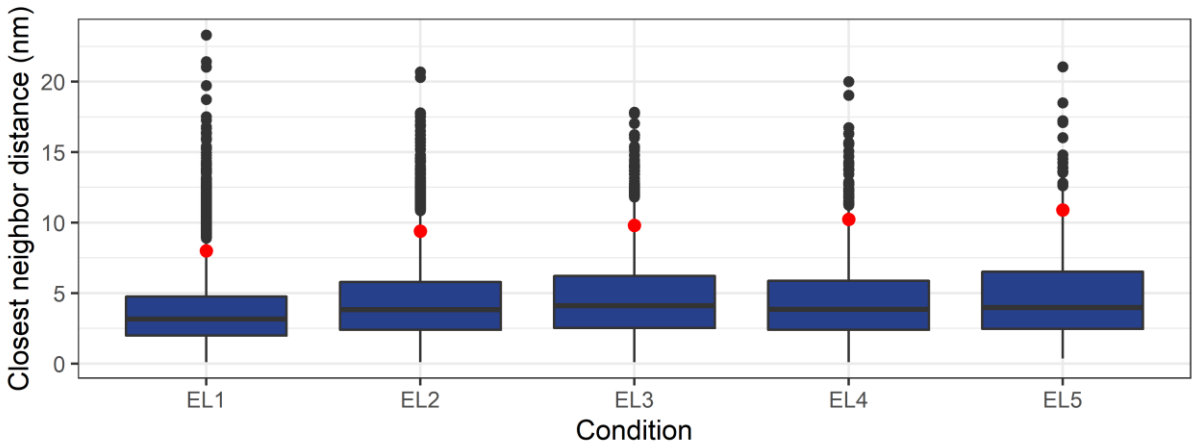


Figure 4.5: Intracluster nearest neighbor distance. Boxplot representing the nearest neighbor distance distribution of localizations belonging to clusters for EL1-5. The value separating the top 5% of the distances is labelled in red. Each boxplot is composed by the median (thick line), the interquartile range (accounts for 50% of the values), the whiskers (account for a maximum of 1.5xIQR) and the outliers (black dots).

I then compared the nearest neighbor distance distribution across the different conditions for the general localizations population and for localizations belonging to clusters. The five conditions significantly differ from each other (p-value < 0.05) when comparing samples derived from the general localization population (Table 4.1). This does not hold true when comparing localizations samples derived from clusters. EL3 to 5 show no significant difference with each other while EL1 differ significantly from all other conditions. EL2 shows an intermediate result (Table 4.1). While statistically significant, the difference in nearest neighbor distances within clusters between different expression levels does not translate in a drastic change of value (Figure 4.5). For all conditions tested the average nearest neighbor distance is below 5 nm, with 95% of the measurements below 11 nm (Figure 4.5). As the distances values are on the same scale of those expected for interacting molecules and do not drastically decrease with an increase in protein concentration, we can suggest that Dendra2-MamY clusters do not undergo significant cluster structure reorganization upon size increase.

In accordance with all the data collected, we can finally propose a model for the *de novo* oligomerization and localization of MamY. Briefly, MamY form oligomers at the membrane which will then relocate preferentially within the geodetic axis region, presumably by a membrane topology-sensing mechanism. Here, based on subunits availability, they will further increase in size while leaving their intracluster structure relatively unchanged. Moreover, the nearest neighbor distances observed for MamY subunits in clusters is within intermolecular interaction range, and bacterial two-hybrid data suggested that MamY subunits are able to interact with one another (Toro-Nahuelpan et al. 2019), supporting a model where the clusters are composed of directly interacting MamY subunits. Finally, the data do not support a filament-like assembly but rather favor a three-dimensional domain swapping mechanism.

Table 4.1: Nearest neighbor distance comparison. Multiple comparison test after Kruskal-Wallis (Siegel and Castellan 1988) of nearest neighbor distances for general and clustered localizations at five different expression levels (EL1-5). 600 localizations belonging to each expression level were sampled and used for the comparison. Sampling was repeated three times.

General population analysis					
Comparison	**Obs. Dif. 1**	**Obs. Dif. 2**	**Obs. Dif. 3**	**Crit. Dif. (p-value: 0.05)**	**Difference***
EL1-EL2	502.4525	440.8842	465.1650	140.3751	TRUE (3/3)
EL1-EL3	678.7250	691.0333	645.5733	140.3751	TRUE (3/3)
EL1-EL4	847.9742	856.3292	755.0342	140.3751	TRUE (3/3)
EL1-EL5	994.7858	1008.8575	956.8525	140.3751	TRUE (3/3)
EL2-EL3	176.2725	250.1492	180.4083	140.3751	TRUE (3/3)
EL2-EL4	345.5217	415.4450	289.8692	140.3751	TRUE (3/3)
EL2-EL5	492.3333	567.9733	491.6875	140.3751	TRUE (3/3)
EL3-EL4	169.2492	165.2958	109.4608	140.3751	TRUE (2/3)
EL3-EL5	316.0608	317.8242	311.2792	140.3751	TRUE (3/3)
EL4-EL5	146.8117	152.5283	201.8183	140.3751	TRUE (3/3)

Clustered population analysis					
Comparison	**Obs. Dif. 1**	**Obs. Dif. 2**	**Obs. Dif. 3**	**Crit. Dif. (p-value: 0.05)**	**Difference***
EL1-EL2	172.2375	115.8050	264.1400	140.3751	TRUE (2/3)
EL1-EL3	365.1542	304.7333	362.2100	140.3751	TRUE (3/3)
EL1-EL4	235.8375	186.5033	293.6350	140.3751	TRUE (3/3)
EL1-EL5	341.5417	295.3875	364.5941	140.3751	TRUE (3/3)
EL2-EL3	192.9167	188.9283	98.0700	140.3751	TRUE (2/3)
EL2-EL4	63.6000	70.6983	29.4950	140.3751	FALSE(0/3)
EL2-EL5	169.3042	179.5825	100.4516	140.3751	TRUE(2/3)
EL3-EL4	129.3167	118.2300	68.5750	140.3751	FALSE(0/3)
EL3-EL5	23.6125	9.3458	2.3841	140.3751	FALSE(0/3)
EL4-EL5	105.7042	108.8841	70.9591	140.3751	FALSE(0/3)

*TRUE if at least two out of three sampling comparison show a p-value < 0.05

Chapter 5: Intracluster heterogeneity analysis and ParB

Scripts and macros used within this chapter:

- Fiji macro 1
- Fiji macro 2
- Script_R1
- Script_R2
- Script_Rv
- Script_R5P1
- Script_R5P2

Results included within this chapter were published in:

- Bohm, K., et al. (2020). "Chromosome organization by a conserved condensin-ParB system in the actinobacterium *Corynebacterium glutamicum*." Nat Commun 11(1): 1485.

Before the advent of SMLM, clustering studies via fluorescence microscopy were limited to a broad determination of the cluster size and localization while all the data concerning the oligomeric structures themselves were derived from cryo-electron microscopy and by crystal structures (Zhang et al. 2007, Oliva et al. 2010).

Changes in a protein cluster structure, whether we are talking about an array of channels distributed on a membrane or a single oligomeric structure, usually involve a change in the average distance between subunits. SMLM data are therefore the perfect tool to study such phenomena. Script_R5P2 identify and characterize two types of cluster via OPTICS. The two cluster categories are automatically organized in a density dependent hierarchy, with a “lv1” cluster category having the lowest density and a “lv2” cluster category having the highest density. As OPTICS require exclusively a lower bound limit (ε_{limit}) for the identification of a cluster, “lv2” cluster will always be localized within

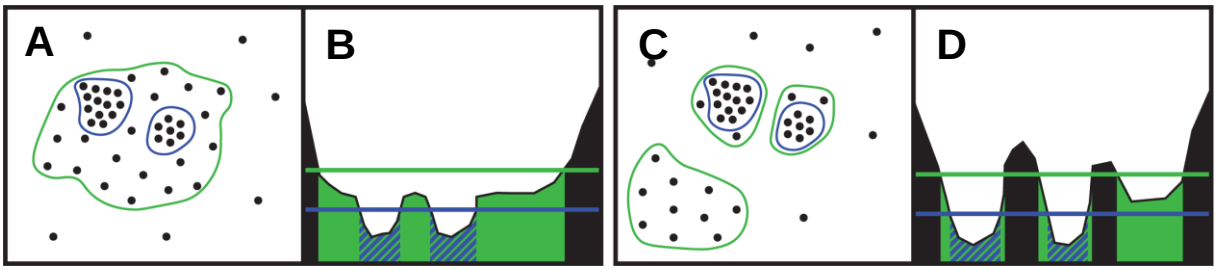


Figure 5.1: Schematics of subclusters and high density clusters. “Lv2” clusters (blue outline) can be generally divided into two separate types: subclusters **(A-B)** and high density clusters **(C-D)**. Subclusters visually appear as a variable number of small regions of higher density found within an already identified “lv1” cluster **(A)** while a high density cluster spans for the almost entirety of the “lv1” cluster to which it belongs **(C)**. In the ideal scenario, the two separate levels of clustering can be easily identified via the visualization/analysis of the reachability plots, where valleys characterized by different reachability distances correspond to different levels **(B, D)**.

a “lv1” cluster. Extra steps are therefore necessary to determine whether the two cluster categories need to be analysed separately or as a cluster-subcluster combination (Figure 5.1).

From a biological point of view, heterogeneity in clustering can be caused by a multitude of phenomena. Whether the changes in clustering are due to human-made genetic modifications (Bohm et al. 2020) or due to a change in substrate (Martins et al. 2019), the study of these changes can be of extreme value for the understanding of the underlying biological process. One of such processes is the ParB dependent *oriC* domain compaction in *C. glutamicum.*

5.1. ParB form nucleoprotein complexes in a *parS* dependent manner

While I previously mentioned ParB in chapter 3.1 as a DivIVA interaction partner involved in chromosome segregation (Ireton et al. 1994, Donovan et al. 2012), this interaction plays a role in only one of the ParB many functions: the tethering of the origin of replications to the polar regions. ParB is also involved in the segregation of the origin of replication over the existing nucleoid toward the center of the cell (Leonard et al. 2005, Zhang and Schumacher 2017, Bohm et al. 2020) and in the loading of the structural maintenance of chromosomes (SMC) proteins onto the DNA (Gruber and Errington 2009, Sullivan et al. 2009, Minnen et al. 2011, Bohm et al. 2020).

ParB dimers bind to 16 bp long DNA sequences (*parS*) located in proximity of the *oriC.* From here, ParB complexes have been shown to further extend by mean of spreading and bridging between dimers (Rodionov et al. 1999, Murray et al. 2006, Graham et al. 2014). Said spreading has been recently shown to be regulated in a CTP dependent manner (Osorio-Valeriano et al. 2019, Soh et al. 2019). While there exist ten *parS* sites in *C. glutamicum*, chromatin immunoprecipitation followed by sequencing (ChIP-seq) highlighted the presence of three separate nucleation zones (Bohm et al. 2020). As wild type-like growth and morphology are supported even in presence of a single *oriC* proximal *parS* sequence, this suggests that the presence of multiple ParB-binding sites is not a requirement but rather acts as an insurance toward the robustness of the chromosome segregation system (Bohm et al. 2020).

ParB clusters differ from the structures analysed in chapter 4.3. MamY oligomers are characterized by a 2D semi-constant internal structure that span between poles and follows the geodetic path (MamY is a membrane protein). As the geodetic path is spirally shaped, MamY oligomers periodically exit the focus plane causing a gradual fading of the structure (gradual decrease in local density) (Figure 4.3 B). On the other side, it has been recently proposed that ParB subunits nucleate around a single *parS* site in *E. coli* to form a spherical non-canonical liquid-liquid phase separated condensate (LLPS) (Guilhas et al. 2020). The description of ParB nucleoprotein structures is even more complex in *C. glutamicum* as there exist multiple closely located nucleation zones that combine into a variety of shapes when imaged via conventional-like fluorescence microscopy.

Here, I use PALM and intracluster heterogeneity analysis to show that ParB form highly heterogeneous nucleoprotein complexes. Furthermore, by comparing the clusters formed in presence of no, one or ten *parS* sites, I identify the source of said heterogeneity.

In details, I imaged and analysed *C. glutamicum parB::parB-PAmCherry* strains differing for the number of *parS* sites contained within their genome (Bohm et al. 2020). The three genomic backgrounds I used for this comparison contain ten (WT - CBK009), one (*parS*$_{2\text{-}10mut}$ - CBK029) and no *parS* sites (*parS*$_{1\text{-}10mut}$ - CBK087) and as such should be respectively characterized by three, one and no nucleation zones for each *oriC* (Bohm et al. 2020). While post-imaging analysis of ParB-PAmCherry data follows, the same procedure as the one used for Dendra2-MamY (Script_Rv, Script_R5P1, ScriptR5P2), the difference in the experimental question is reflected by a different analysis focus. Moreover, due to the use of a different tag (PAmCherry vs Dendra2) and of the different nature of the protein studied, some analysis parameters, such as photon count and PSF width filtering, are changed (Table 2.1).

ParB-PAmCherry nearest neighbor distribution function $G(r)$ differs from the one of a Poisson distribution in all the imaged genetic backgrounds (WT, *parS*$_{2\text{-}10mut}$ and *parS*$_{1\text{-}10mut}$) suggesting that ParB is able to form clusters also in absence of *parS* binding sites (Figure 5.2 A-C). In agreement with what can be observed via conventional fluorescence microscopy (Bohm et al. 2020), ParB enrichment can be observed at various locations along the cell length axis in presence of *parS* (Figure 5.2 E-F). On the contrary, while conventional live imaging of cells expressing fluorescently labeled ParB in absence of *parS* results in a homogeneously labeled cytoplasm, PALM imaging highlights the presence of clusters (Figure 5.2 D).

As shown before (Chapter 4), protein concentration has a great influence on the size and frequency with which clusters form. Protein concentration can be changed by altering its expression or degradation rate. One of the strains tested, *C. glutamicum parB::parB-PAmCherry parS*$_{1\text{-}10mut}$, is not only characterized by the lack of *parS*, but also by more than 20% DNA free cells and aberrant morphology, similarly to a *ΔparB* strain (Donovan et al. 2013, Bohm et al. 2020). As ParB concentration could be altered within anucleate cells, I compared protein concentration and cell morphology across the three strains.

Firstly, I tested whether the cell area size distribution of the three strains follow a normal distribution via the Shapiro-Wilk test (Royston 1995) (Figure16A-C). While the data from the WT (W = 0.98286, p-value = 0.4842) and parS$_{2\text{-}10mut}$ (W = 0.99067, p-value = 0.8816) cells can be treated as normally distributed, this is not the case for parS$_{1\text{-}10mut}$ cells (W = 0.93647, p-value = 0.02913). While, in fact, the quasi-symmetric cell division of WT and parS$_{2\text{-}10mut}$ cells is mirrored by a single normally distributed cell area size, the asymmetric cell division typical of the parS$_{1\text{-}10mut}$ cells is likely to result in

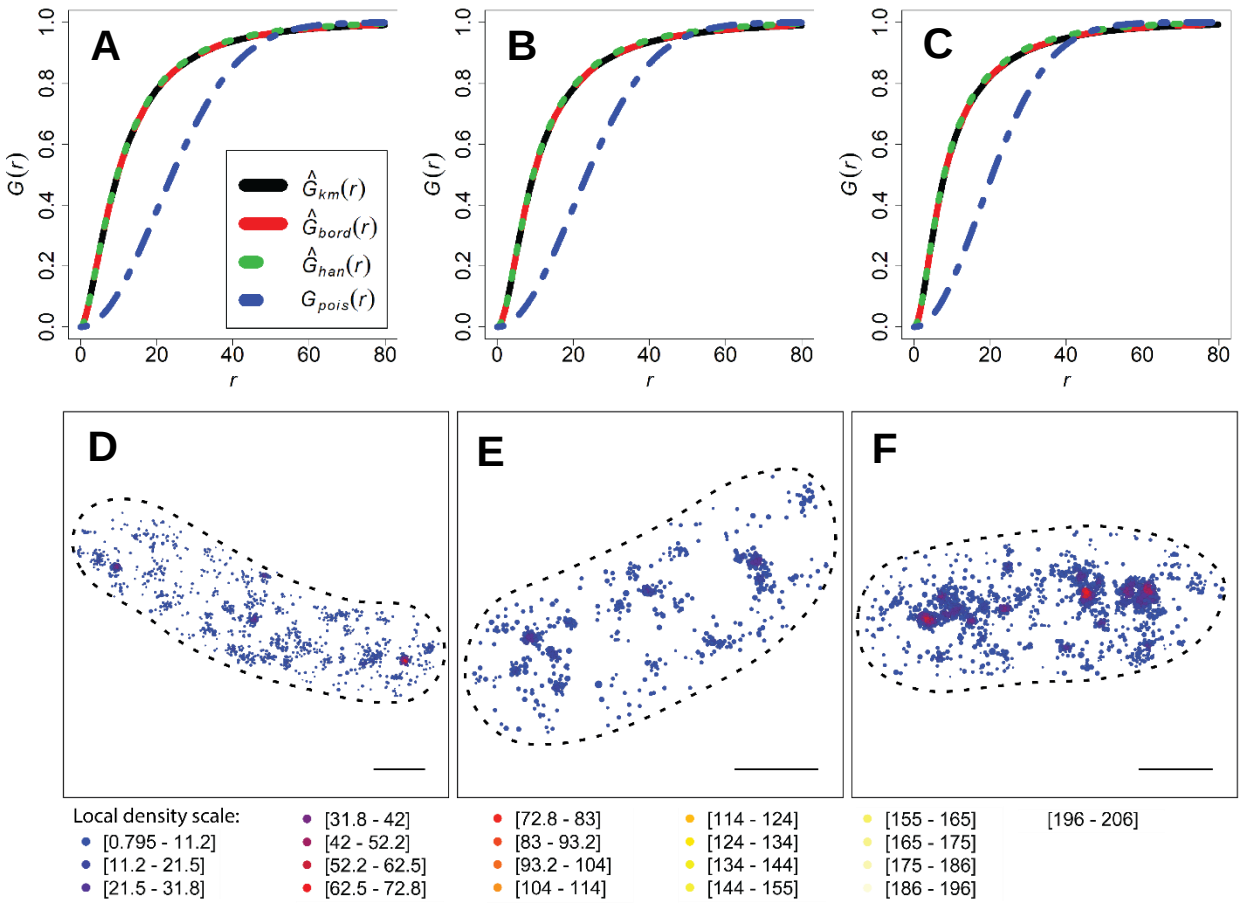


Figure 5.2: ParB-PAmcherry form clusters in all the genetic backgrounds analysed. ParB-PAmCherry nearest neighbor distribution function $G(r)$ in a *parS*$_{1\text{-}10mut}$ (CBK087) **(A)**, a *parS*$_{2\text{-}10mut}$ (CBK029) **(B)** and a WT (CBK009)**(C)** *C. glutamicum* background strain. The plots show three different edge correction methods: border method, Kaplan-Meier and Hanisch. The associated theoretical Poisson distribution is shown in blue. ParB-PAmCherry localizations distributions of representative *parS*$_{1\text{-}10mut}$ **(D)**, a *parS*$_{2\text{-}10mut}$ **(E)** and a WT **(F)** *C. glutamicum* cells. Local density is expressed as the number of localizations residing within a square area (50 nm lateral side, 2500 nm^2) centred around the localization itself. The interval values shown in the legend are shared among the different plots. Scale bar = 0.5 µm.

a multi-modal distribution. I therefore fitted the data with a Gaussian finite mixture model via the Mclust R package (Scrucca et al. 2016) and determined the number of normal components that optimize the Bayesian information criterion (BIC) for each strain (Figure 5.3 A-C). The BIC of the parS$_{1\text{-}10mut}$ strain is minimized in presence of two normal components (the likelihood ratio test returns a $\sim$10% chance to have obtained parS$_{1\text{-}10mut}$ cell area size data from a single component distribution). The subpopulation characterized by a smaller average size and smaller standard deviation (red line–Figure 5.3 A) accounts for almost a third of the cells and is mainly composed of minicells (anucleate cells and cells containing a truncated chromosome) while the subpopulation characterized by higher mean and higher standard deviation is composed of the cells from which the anucleate cells divided and of the cells that divided without defects (black line – Figure 5.3 A). Finally, comparison of the parS$_{2\text{-}}$

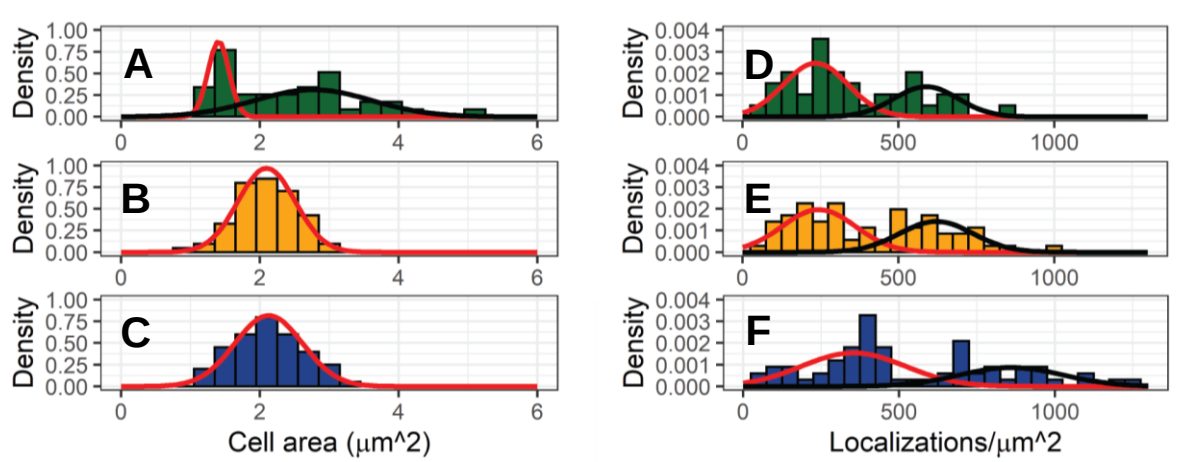


Figure 5.3: Cell area size and ParB-PAmCherry localizations density changes with the number of *parS* sites. Cell size area distribution of *parS*$_{1\text{-}10mut}$ **(A)**, *parS*$_{2\text{-}10mut}$ **(B)** and WT **(C)** cells. The cell area size distributions of *parS*$_{2\text{-}10mut}$ and WT are characterized by a single normal component (**B-C**) while *parS*$_{1\text{-}10mut}$ data are better described by two normal components **(A)**. The ParB-PAmCherry localization density of the three genetic backgrounds can be fitted in all cases by two normal components **(D-F)**. The normal components are represented as continuous red and black lines.

$_{10mut}$ and WT strains via Welch's t-test (two-sided, p-value = 0.6931) does not highlight differences in cell area size distribution between the two strains supporting the idea that cells containing a single *parS* site are wild-type like.

I then compared the average localizations density of cells belonging to the three different genetic backgrounds. The ParB-PAmCherry localizations densities can be described in all genetic background by two normal components (the likelihood ratio test returns a <10% chance to have obtained said data from a single component distribution for all strains) (Figure 5.3 D-F). In all cases, the component accounting for the majority of the population is the one characterized by a smaller average density (parS$_{1\text{-}10mut}$: ~64%, parS$_{2\text{-}10mut}$: ~58%, WT: ~64%) (Figure 5.3 D-F: red line). As the three set of data do not follow a normal distribution, I compared them via the multiple comparison test after Kruskal-Wallis (comparison between treatments). The average ParB-PAmCherry localizations density of the WT strain differs significantly from both the parS$_{1\text{-}10mut}$ and parS$_{2\text{-}10mut}$ strains (p-value<0.05) while no difference can be observed between the remaining two strains (p-value>0.05).

The changes in protein concentration observed across the strains cannot be attribute to changes in cell morphology (parS$_{1\text{-}10mut}$ and parS$_{2\text{-}10mut}$ protein concentrations do not differ), meaning that it is not necessary to filter the data in a morphology dependent manner prior further analysis.

The duality in expression observed for all three strains, with one subpopulation being characterized by approximately double the protein concentration (Figure 5.3 D-F) is not caused by a difference in morphology nor a change in number of *parS* sites (the two distributions are visible also

in the parS$_{1\text{-}10mut}$ strain). It may instead be caused by the duplication of the genetic area responsible for ParB expression following genome replication.

Finally, the shift toward higher values observed for the WT strain when comparted to the parS$_{2\text{-}10mut}$ strain can instead be attributed to the change in number of *parS* sites (no morphological difference can be observed between the two strains).

5.2. ParB and cluster heterogeneity analysis

If we assume that ParB in *C. glutamicum* behaves similarly to what observed in *E. coli*, with the formation of spherical non-canonical LLPS condensate surrounding a single *parS* site, clusters should be characterized by a defined size, shape and lack of heterogeneity. While this may hold true for the parS$_{2\text{-}10mut}$ strain, Chip-seq data show that each chromosome contains multiple, closely located, ParB nucleation zones, each corresponding to a defined propagation area (Bohm et al. 2020).

The presence of multiple *parS* sites within each nucleation zone, together with the proximity between each nucleation zone and the presence of multiple chromosomes within each cell means that the three-dimensional ParB nucleoprotein structures will appear as a dense highly heterogeneous area when imaged via 2D-PALM.

I first determined suitable clustering thresholds (ε_{limit}) for the WT genomic background and then applied them to the other strains. Specifically, given an exemplary WT background cell (Figure 5.4 C) I calculated its reachability plot ($MinPts = 32$) and, based on the shape of said plot, determined two separate thresholds ($\varepsilon_{1\_limit} = 50\ nm, \varepsilon_{2\_limit} = 35\ nm$) (Figure 5.4 A-B). The first threshold is aimed toward the identification of areas generally enriched in ParB ("lv1"), while the second threshold

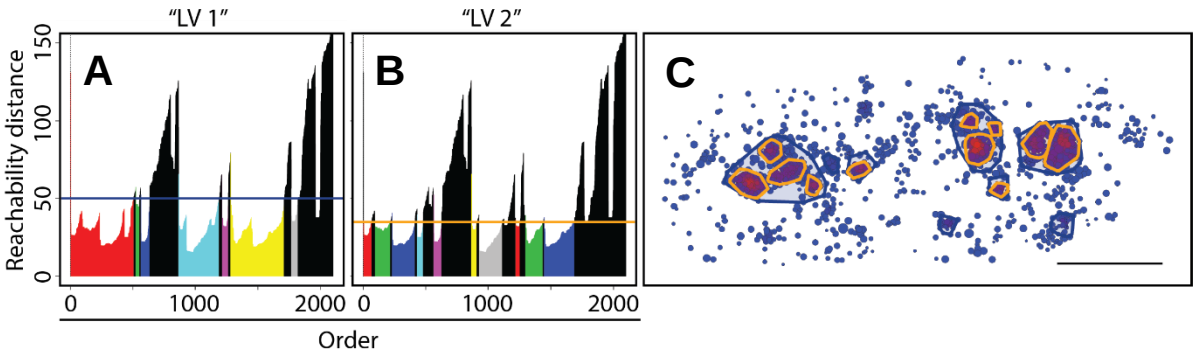


Figure 5.4: Identification of two levels of clustering for ParB-PAmCherry via OPTICS. Reachability plot for ParB-PAmCherry in a WT background *C. glutamicum* cell for $MinPts = 32$ **(A,B)**. Dual level clustering was determined via the use of two separate threshold levels ($blue\ line - lv1\_\varepsilon_{limit} = 50nm$ **(A)**, $orange\ line - lv2\_\varepsilon_{limit} = 35nm$ **(B)**). The clusters obtained can be visualized as convex areas overlapping the localizations themselves (Threshold line and cluster border share the same color) **(C)**. The color scale used for the localizations is the same of the one used in Figure 5.2. Scale bar= 0.5 µm.

focuses on the characterization of the heterogeneity contained within the "lv1" clusters (Isolation of the valleys present within the clusters identified within the "lv1" reachability plot).

While "Lv1" clusters mostly resemble in size and positioning the fluorescence profile shown by ParB via conventional fluorescence, I could also identify several smaller, randomly distributed clusters that are not observed in conventional fluorescence microscopy images (Figure 5.4 C). As it was not possible to determine a priori size threshold to distinguish between "lv1" clusters that originated from a *parS* dependent ParB nucleation event and those that originated in a different way, I separated the data based on a different criterion. Specifically, given that each *C. glutamicum* cell has typically at least two nucleoids (Bohm et al. 2017), I extracted the two larger "lv1" clusters from each cell and analysed them separately. From this point on I will refer to these clusters and the "lv2" clusters localizing within them as "lv1M" and "lv2M" respectively.

The "lv1M" clusters size data (cluster size is defined as the number of localizations composing the cluster) do not follow a normal distribution in any of the strains tested (p-value<0.05). I therefore compared the strains via multiple comparison test after Kruskal-Wallis. The size of "lv1M" clusters in the WT background strain is significantly higher than the two remaining strains (p-value < 0.05) while no difference can be observed between the "lv1M" clusters size distributions of cells containing a single or no *parS* sites (Figure 5.5 A).

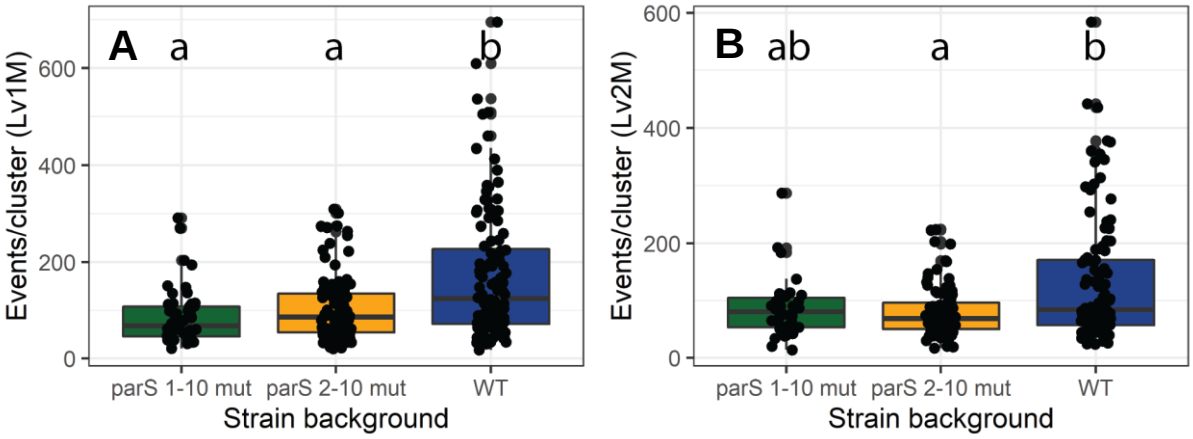


Figure 5.5: "Lv1M" and "Lv2M" ParB clusters are bigger in the WT background strain. Number of fluorescence events (ParB-PAmCherry localizations) composing a "Lv1M" cluster in a WT, $parS_{1\text{-}10mut}$ and $parS_{2\text{-}10mut}$ *C. glutamicum* background strain **(A).** Number of fluorescence events (ParB-PAmCherry localizations) composing a "Lv2M" cluster in a WT, $parS_{1\text{-}10mut}$ and $parS_{2\text{-}10mut}$ *C. glutamicum* background strain **(B)**. In both cases statistical significance is asserted via multiple comparison test after Kruskal-Wallis and is highlighted via lowercase letters **(a,b)**. Each boxplot is composed by the median (thick line), the interquartile range (accounts for 50% of the values), the whiskers (account for a maximum of 1.5xIQR) and the outliers (dots located outside the whiskers boundaries).

I performed the same comparison procedure for the "lv2M" clusters size data. Once again, the data did not follow a normal distribution in any of the strains tested (Shapiro-Wilk test: p-value < 0.05). The size of the "lv2M" clusters in the WT background is significantly higher than the strain containing a single *parS* site (multiple comparison test after Kruskal-Wallis, p-value < 0.05) while no statistical difference is found between $parS_{1\text{-}10mut}$ and the two remaining strains (Figure 5.5 B).

I then characterized parameters related to the heterogeneity of the "lv1M" clusters and compared the results across the three strains. When describing the heterogeneity based on subclustering we can observe that the proportion of "lv1M" clusters containing no subclusters is double for $parS_{1\text{-}10mut}$ and $parS_{2\text{-}10mut}$ strains compared to WT ($\sim 30\%$ against a 15%) (Figure 5.6 A). As the absence of subclusters can be attributed to the ParB structures three-dimensional positioning respective to the objective focus, these clusters were excluded from further analysis. The distribution of the number of subclusters (lv2M) localizing within each cluster (lv1M) does not follow a normal distribution in any of the strains analysed (Shapiro-Wilk test: p-value < 0.05). While we can observe more than two subclusters only for "lv1M" clusters belonging to the WT background strain (Figure 5.6 A), multiple comparison test after Kruskal-Wallis highlights no statistical difference between the strains (p-value > 0.05).

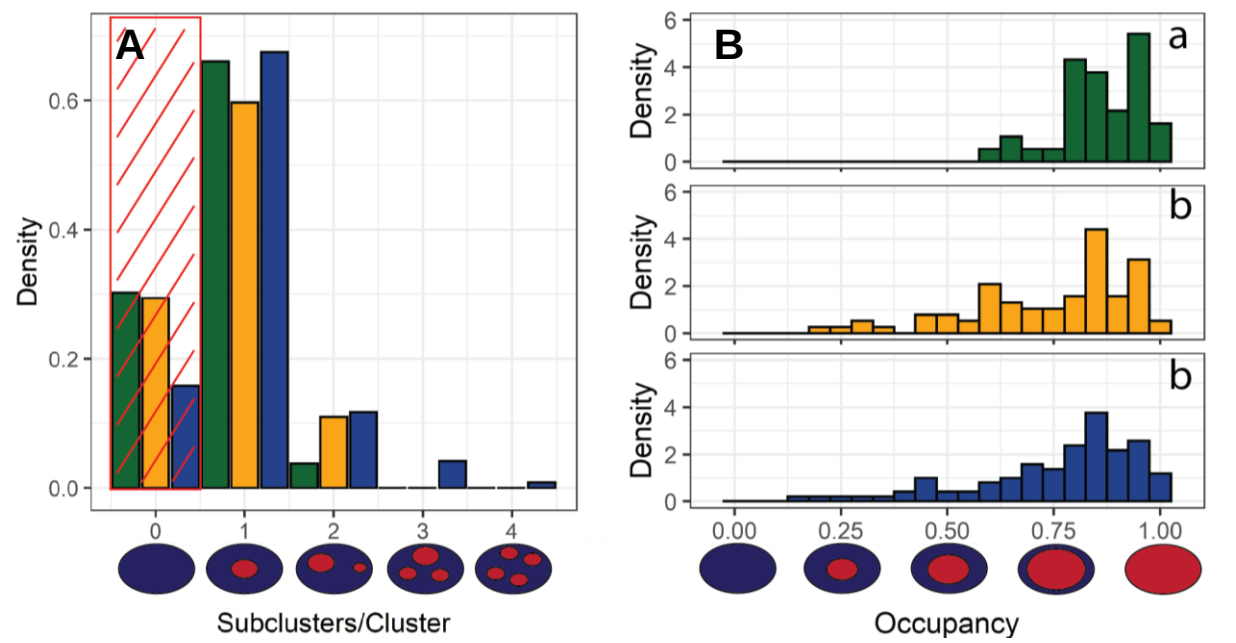


Figure 5.6: The WT background strain "Lv1M" clusters are characterized by higher heterogeneity. Number of distinct subclusters ("Lv2M") observed within a "Lv1M" cluster in a WT (blue), $parS_{1\text{-}10mut}$ (green) and $parS_{2\text{-}10mut}$ (orange) *C. glutamicum* background strain **(A)** Proportion of "Lv1M" cluster occupied by "Lv2M" clusters in a WT (blue), $parS_{1\text{-}10mut}$ (green) and $parS_{2\text{-}10mut}$ (orange) *C. glutamicum* background strain **(B)**.

As a second parameter to determine "lv1M" clusters heterogeneity I choose occupancy. Given a "lv1M" cluster and its subclusters ("lv2M$_k$"), the occupancy of "lv1M" is here defined as:

$$Occupancy_{lv1M} = \frac{\sum_{k=1}^{n} N_{lv2M_k}}{N_{lv1M}} \tag{13}$$

where k ranges from 1 to n, where n is the number of subclusters located within lv1M and N is the number of localizations of which a cluster is composed. Occupancy ranges from 0 (absence of subclustering), to 1 (subclusters occupy the entirety of the cluster) (Figure 5.6 B). The distribution of occupancy for the three strains analysed do not follow a normal distribution (Shapiro-Wilk test: p-value < 0.05). I therefore compared ParB cluster distributions via multiple comparison test after Kruskal-Wallis. The occupancy distribution for the *pars*$_{1\text{-}10mut}$ strain is significantly higher than the one of the other two strains (p-value < 0.05), while no difference can be observed between the *pars*$_{2\text{-}10mut}$ and the WT strains (p-value > 0.05).

To summarize, ParB concentration follows a bimodal distribution. While the number of *parS* sites have an effect on ParB expression levels (shift of the distribution toward higher values), the dual modality of the function is independent from the number of parS sites and is likely to be caused by the duplication of the *parB* genes following genome duplication.

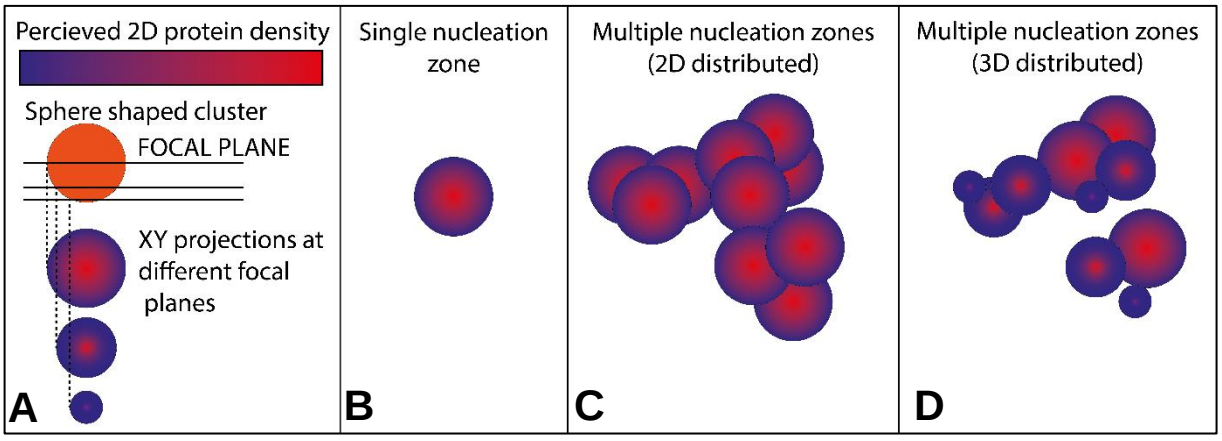


Figure 5.7: Interpretation of 2D-PALM data in regard to ParB heterogeneity. The protein density of a three-dimensional spanning cluster perceived via 2D-PALM depends on its position along the focal plane **(A)**. Spherical/Quasi-spherical clusters are characterized by a radial density gradient (*parS*$_{2\text{-}10mut}$) when imaged via 2D-PALM **(B)**. The partial overlap of multiple spherical clusters on a 2D plane result in an increase in heterogeneity **(C)**. Further complexity is derived by the three-dimensional localization of the clusters **(D)**

In absence of *parS*, ParB form “LV1M” clusters comparable in size and density with what observed in presence of a single *parS* site. These clusters are characterized by high occupancy and low internal heterogeneity, similarly to what it is expected for randomly localized spherical shaped clusters (Figure 5.7 B). The clusters observed in presence of a single *parS* site are characterized by a statically lower occupancy and low heterogeneity, in agreement with the increase chance of proximity between two ParB complexes due to genome duplication. The ParB “LV1M” clusters observed in the WT background strain are bigger than those observed in the $pars_{1\text{-}10mut}$ and $pars_{2\text{-}10mut}$ strain. Moreover, more than two high density regions within a single LV1M cluster can only be observed in the WT scenario in agreement with the presence of multiple, closely located, nucleation zone within each ParB nucleoprotein complex (Figure 5.7 D).

Chapter 6: Dual color imaging and co-localization analysis

Scripts and macros used within this chapter:

- Fiji macro 1
- Fiji macro 2
- Script_R1
- Script_R2
- Script_R5P3

The analysis approaches presented in the previous chapters show some of the ways in which the script I wrote can be used while highlighting the case to case dependent nature of SMLM data analysis. While analysis of dual color data does not differ from the one of single color data, with the exception of the calculation of the coordinate based co-localization value (CBC), the same is not true for the imaging itself.

When imaging a dual color sample via SMLM the following conditions need to be met: a buffer that allows for the biochemical functioning of both fluorophores, a combination of instrument and experimental design that allows for the independent imaging of the fluorophores, and a reference system that can be used to align the different channels.

The imaging of multiple organic dyes, or a combination of organic dyes and fluorescent proteins, requires the accurate selection of an appropriate imaging buffer (Endesfelder et al. 2011). This selection is not trivial as can be observed by the abundant literature entries on the topic (Nahidiazar et al. 2016, Glushonkov et al. 2018). While there is no lack of studies where fluorescent proteins have also being tested under various buffer conditions, the aim of such works focused on imaging fluorescent proteins in combination with organic dyes (Endesfelder et al. 2011). Fluorescent proteins themselves work under physiological conditions, meaning that buffers such as PBS or non-fluorescent growth media can be used during imaging. It follows that PALM imaging does not require the development of combination/fluorophore specific buffers (Shroff et al. 2008).

PALM imaging is only possible under conditions where fluorescent profiles can be separated in space and time. Currently there are three separate methods that allows for dual color PALM imaging. The first method involves the simultaneous activation and imaging of two photoactivatable fluorescent proteins characterized by different emission profiles, where the separation of the fluorescent profiles is guaranteed by the use of two cameras and different emission filter sets. The second and third methods involves the sequential imaging of the fluorescent proteins. In one case the two fluorophores can share the same emission profile and are selectively activated via Primed Photoconversion and UV-Photoactivation (Dempsey et al. 2015, Klementieva et al. 2016, Virant et al. 2017). In the last case the

fluorophores cannot share the same emission profile and only one of the two fluorophores is photoactivatable (Bach et al. 2017).

Finally, alignment between different wavelength is guaranteed by the presence of a reference system. The reference system usually consists of an array of multispectral fluorescent beads and can be used for lateral drift correction, calibration of a dual camera system and the calculation of a local weighted-mean matrix for aberration correction (Malkusch et al. 2012).

While both co-localization and cluster analysis can be theoretically applied directly on any point pattern (in our case the output of Script_R2), only co-localization analysis (Script_R5P3) can be directly used on such dataset. The artificially created inability to identify clusters (Script_R5P2) prior a more general analysis is voluntary. As previously detailed (Chapter 4), clustering algorithms are usually not able to autonomously differ between randomly and biologically derived clustering. It is therefore recommended to apply clustering algorithms exclusively in presence of data supporting the idea that the studied protein is indeed clustering (estimation of $G(r)$ via Script_Rv, visualization and characterization of the distribution in a density/nearest neighbor distance dependent manner via Script_R1). Co-localization analysis differs from clustering analysis on this aspect, as its aim is to test whether the relative positioning of each subunit within the two protein populations differ from what it is expected from two unrelated ones (similarly to what Script_Rv does for clustering).

6.1. Script_R5P3 basics and CBC values interpretation

Similarly to Script_Rv, where the presence of clustering is determined based on the deviation of $G(r)$ from the G of a Poisson process, the CBC value is expression of the deviation of the distribution of two protein populations in the neighborhood of the localization itself from the one of two non-co-localizing proteins (See chapter 1.5.3 for a more detailed description).

Although similar, there exist major differences in the way the two scripts construct their respective control datasets. In the case of Script_Rv the Poisson process shares with the experimental distribution the number of points and the area in which they are contained (shape included), while in the case of Script_R5P3 the control distributions share with the experimental ones the number of points and the maximum radius that is used to construct the distributions themselves (shape excluded). Moreover, while Script_Rv includes edge correction of the data, the edge effect is not taken into consideration when calculating the CBC value. It follows that, in order to avoid border artifacts (Figure 6.1), the maximum radius used for the calculation of the CBC value needs to be defined in relation to the size and shape of the area in which the protein localizes. In order to find a valid interval range for R_{max} I first simulated two Poisson point pattern distributions of equal intensity ($\lambda = 0.001\ \mathrm{localizations/nm^2}$) contained within a circular shaped area. As the point pattern distributions are independent from

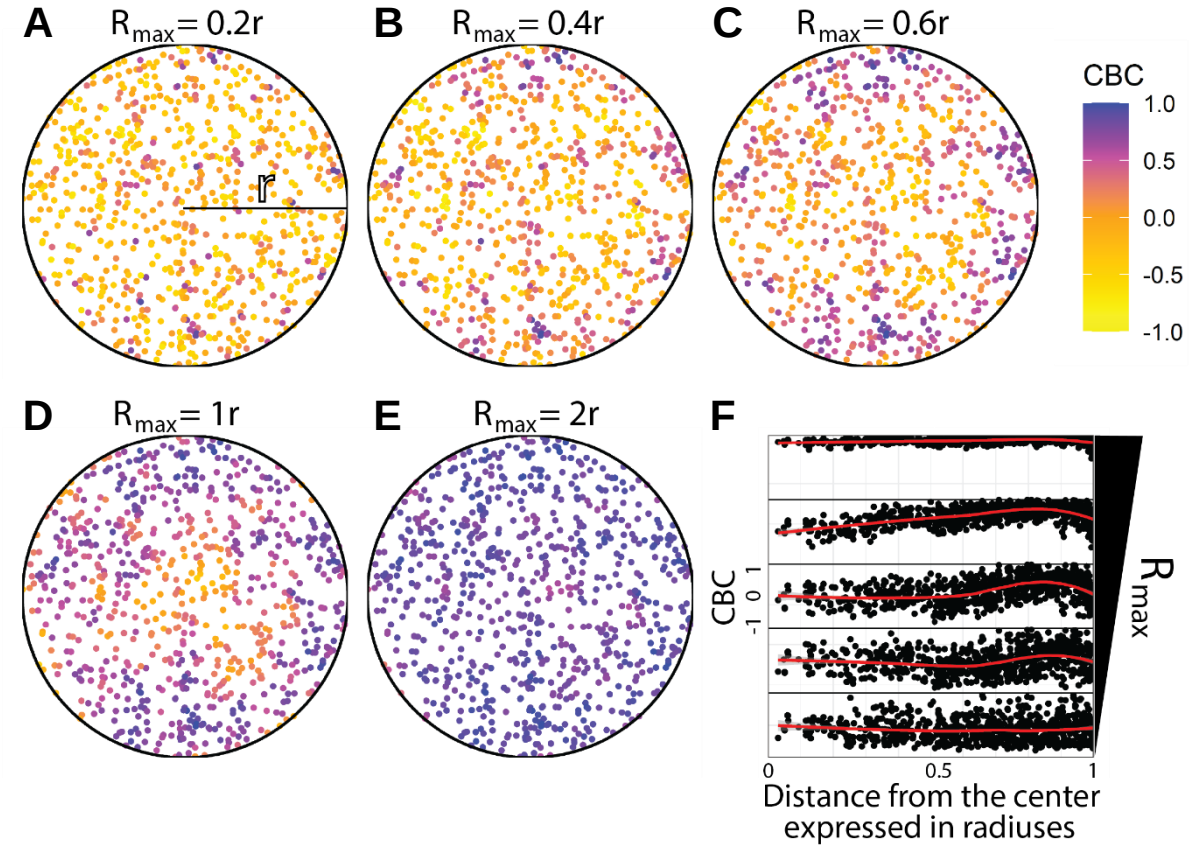


Figure 6.1: Effect of R_{max} on the calculation of CBC values. CBC dependent visualization of one out of two independent Poisson point pattern distributions of identical intensity ($\lambda = 0.001$) tested for co-localization **(A-E)**. The R_{max} used for the calculation of the CBC values is expressed as a fraction of the radius (r) defining the area that contains the point patterns. Distance from the center of the area and CBC values for the five R_{max} tested were plotted together with the locally estimated scatterplot smoothing (LOESS)(red line) **(F)**. $r = 500\ nm$, $interval\ width\ = 5\ nm$.

each other, only minor degree of co-localization is to be expected. I therefore calculated the CBC values for one of the two distributions at different R_{max} and visually compared the extent of border related artifacts (Figure 6.1 A-E). I could then broadly divide R_{max} values into three separate categories. The first category ($R_{max} \leq 0.2r$) is characterized by negligible border artifacts and LOESS estimator stably in the proximity 0 (Figure 6.1 A,F). The second category ($0.2r < R_{max} < 2r$) is characterized by perimetral increase of the CBC values with consequent deviation of the LOESS estimator from 0 for high distances from the center (Figure 6.1 B-D,F). The third category ($R_{max} \geq 2r$) is characterized by a strong compression of the CBC values and the LOESS estimator toward 1 (Figure 6.1 E-F). As the three categories are expression of the same phenomenon (CBC values increase due to border related artifacts) and passage between categories is gradual (Figure 6.1 B-D), the threshold values should be used as general guidelines rather than as laws.

An under-appreciation of the degree of co-localization can be instead observed for very small R_{max} in presence of overlapping structures. In order to test the magnitude of this effect, I simulated a series

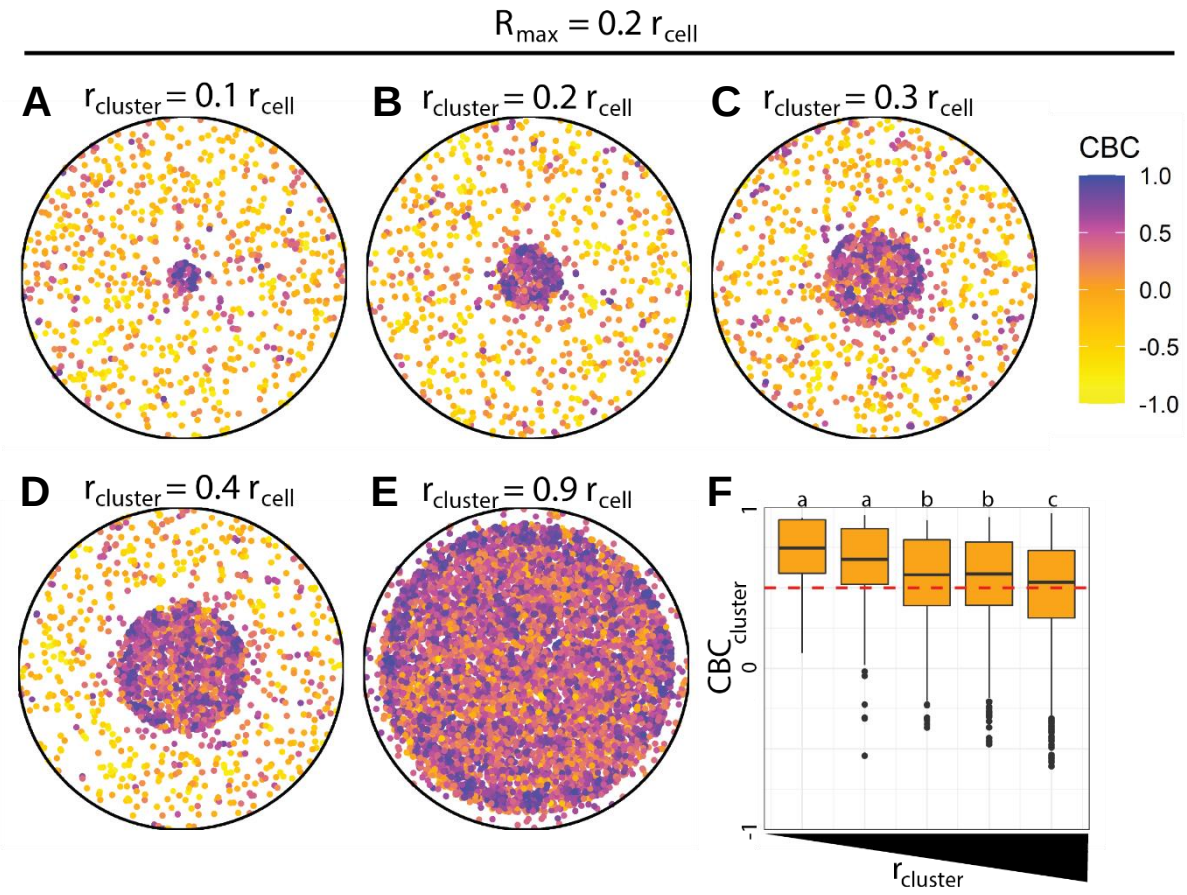


Figure 6.2: CBC values are affected by the size of overlapping clusters. CBC dependent visualization of ppA_i for five different $r_{cluster}$ **(A-E)**. Each ppA_i is composed of two separate Poisson point pattern distributions (one describing the cluster and one describing the cytoplasmic fraction). The R_{max} used for the calculation of the CBC values is constant across the five simulations. The CBC values for localizations residing within the clusters are visualized as boxplots **(F)**. Each boxplot is composed by the median (thick line), the interquartile range (accounts for 50% of the values), the whiskers (account for a maximum of 1.5xIQR) and the outliers (black dots). A horizontal dashed line at $CBC = 0.5$ is drawn as threshold for co-localization. Statistical significance between the CBC values distributions was tested via multiple comparison test after Kruskal-Wallis (p-value<0.05) (conditions that do not differ are grouped via cursive letters). $r_{cell} = 500\ nm$, $interval\ width = 5\ nm$.

of point patterns distributions pairs (ppA_i and ppB_i, where "i" assumes the value of the associated cluster radius) (Figure 6.2 A-E). Each pair simulates the distribution of two proteins, A and B, that form overlapping circular clusters of identical size within a circular shaped cell (each pair varies in $r_{cluster}$). ppA_i and ppB_i are respectively comprised of two overlapping Poisson point pattern distributions each, one that simulates a circular cluster ($\lambda_{cluster} = 0.01$ localizations/nm^2) and one for the cytoplasmic portion of the protein population ($\lambda_{cytoplasm} = 0.001$ localizations/nm^2, r_{cell}). The two clusters not only overlap but are merely a slightly shifted version of each other ($clusterB_{(x,y)} = clusterA_{(x+c,y+c)}$), and it is therefore expected to observe high CBC values for localizations belonging to these clusters. Hence, I compared the clusters CBC values obtained for ppA_i at different $r_{cluster}$ and constant R_{max} ($R_{max} = 0.2 r_{cell}$). None of the CBC values distributions followed a normal distribution

(Shapiro-Wilk test, p-value < 0.05) and were therefore compared with each other via multiple comparison test after Kruskal-Wallis. According to statistical significance (p-value < 0.05), the five conditions can be divided into three groups (Figure 6.2 F), with inverse relation between CBC values and $r_{cluster}$ size.

While the CBC values are indeed mostly above 0 for all conditions tested (Figure 6.2 F), CBC values are function of the deviation of the local relative distribution of ppA_i and ppB_i from their overall relative distribution (calculated at R_{max}). It follows that the average co-localization values will be decreased when the overall relative distribution is mostly calculated from within the co-localizing area ($R_{max} \leq 2r_{cluster}$) (Figure 6.2 C-E).

Finally, for microorganisms where cell width and protein clusters are on a comparable size scale (e.g.> bacteria, trypanosome flagellum) there may be no R_{max} interval that satisfies both conditions ($2r_{cluster} \leq R_{max} \leq 0.2r_{cell}$). Alternatively, it could be impossible to determine one of the two thresholds. As, under these conditions, interpretation of the absolute CBC values becomes less obvious, it is pivotal to compare said values with both positive and negative controls.

6.2. The parameters choice depends on the biological question

While, due to a lack in localization precision, co-localization cannot yet be used to describe the direct interaction between two protein populations (see chapter 1.3.4) (Dunn et al. 2011), it can be used to determine whether and in which proportion said proteins are associated to a certain biological element. As the term co-localization can be used to refer to a multitude of different situations (co-expression within a cell, polar/septal co-localization, co-association with a protein complex), each of which relies on a different size scale, the ability to determine CBC values for each of these scales grants the researcher a great degree of flexibility (Figure 6.3).

In chapter 6.1 I described how changes in one of the two parameters (R_{max}) used in co-localization analysis are mirrored by a change in CBC values (Figure 6.1, 6.2) and then established boundaries for R_{max} in relation to the study of co-localization intended as association of two proteins to the same cluster (Figure 6.2). As the same approach can be applied to different biological questions (interaction range, co-association to a protein complex, co-expression) (Figure 6.3) it is possible to generalize the boundaries obtained for R_{max} as follow:

$$2r_A \leq R_{max} \leq 0.2r_B \quad (14)$$

where r_A is the radius of the structure toward which we are testing for co-localization and r_B is the radius of the area in which the structure is confined. As the size scale changes with the biological question, r_A and r_B can each refer to clusters/cell structure, cells, or artificial ROIs.

The second parameter, interval width (or interval size), determines the resolution at which the localizations distributions are described. As a general rule, changes in interval width are not mirrored

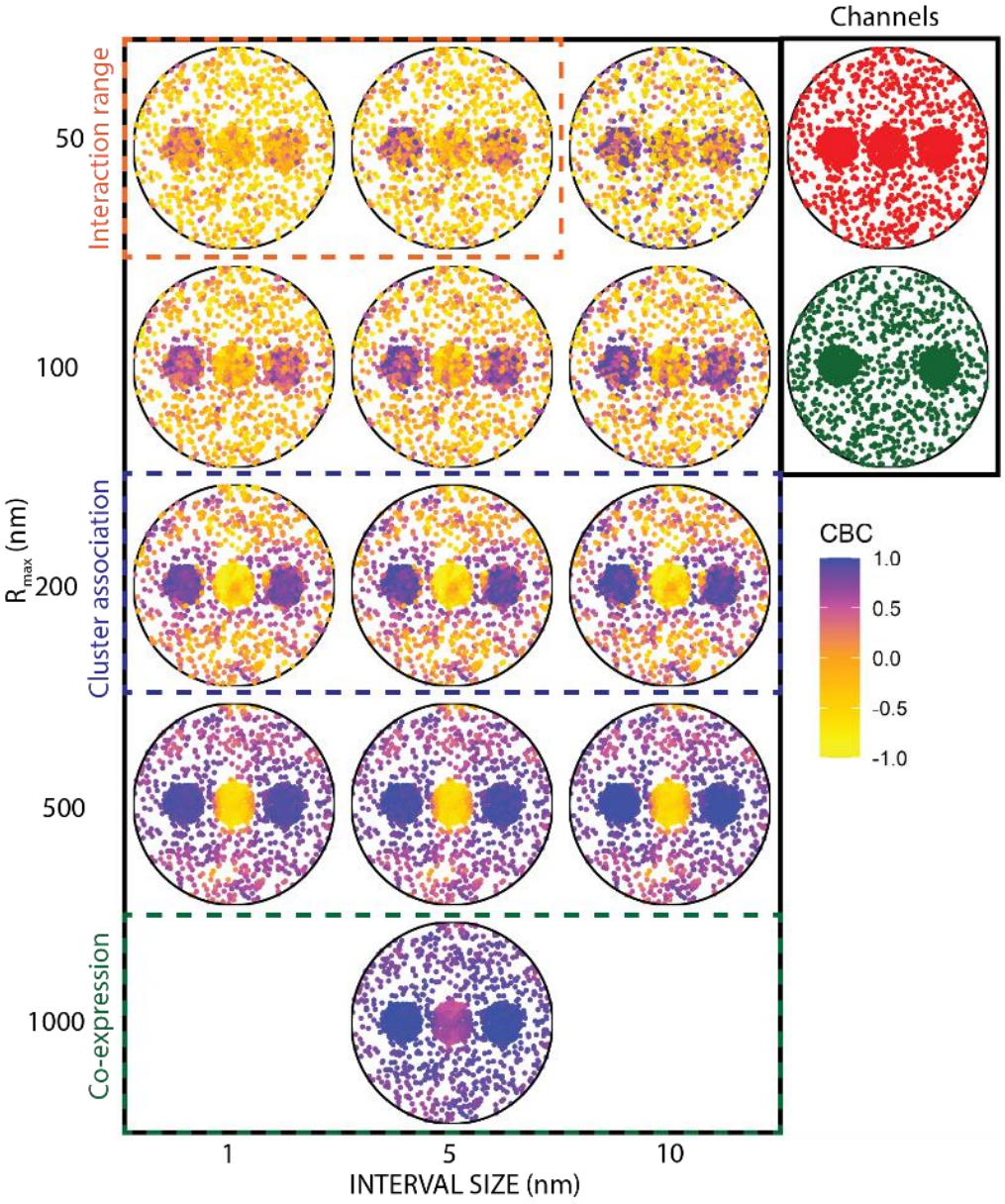


Figure 6.3: Co-localization at different size scales. A composite point process is simulated for each channel within a circular area of $r = 500\,nm$ (Top right). Each point process is respectively composed of a Poisson point pattern ($\lambda_{cytoplasm} = 0.001$ and $r_{cell} = 500\,nm$) that represent the cytoplasmic fraction and either two or three Poisson point patterns that represent the clustering fraction ($\lambda_{cluster} = 0.01$ and $r_{cluster} = 100\,nm$). Localizations are color coded according to their CBC value. The combination of parameters (R_{max} and interval size) suitable for the analysis of specific features (Co-expression, cluster association, Interaction range) are labeled accordingly (dashed color boxes).

by significant changes in CBC values so long as the interval width is sufficiently small compared to the R_{max} (Figure 6.3). An upper threshold can therefore be applied to interval size:

$$interval\ width \leq 0.1 R_{max} \tag{15}$$

While the use of extremely low interval sizes has no downside from a theoretical point of view, as a better resolved distribution will always allow for a better description of the CBC, it is computationally heavy and, beyond a certain point, yields no real benefit (Figure 6.3).

Chapter 7: Analysis of the flagellar tip organization in trypanosomes

Scripts and macros used within this chapter:

- Fiji macro 1
- Fiji macro 2
- Script_R1
- Script_R2
- Script_R5P3

Results included within this chapter will be published in:

- Bachmaier, S., et al. "Unpublished data."

Here, I apply the co-localization concepts that I developed in the previous chapter (Chapter 6) to the characterize the organization of a subcellular compartment: the flagellar tip of *Trypanosoma brucei*.

The flagellar tip of *Trypanosoma brucei* is here proposed to function as a cAMP microdomain essential for trypanosome vector transmission (Bachmaier, S., et al., "Unpublished data"). Biochemical and genetic data show that a newly identified cAMP response protein (CARP3) plays a major role in the maintenance of said domain. Furthermore, CARP3 has been found to co-localize with the adenylate cyclase 1 (AC1) and to be dependent in its localization on FLAM8 (Bachmaier, S., et al., "Unpublished data") (Figure 7.1 A). By using dual-color PALM and coordinate-based co-localization we not only corroborated the data concerning CARP3 interactions and dependencies, but also further characterized and modeled the cAMP microdomain itself. Briefly, I tested two different fluorescence

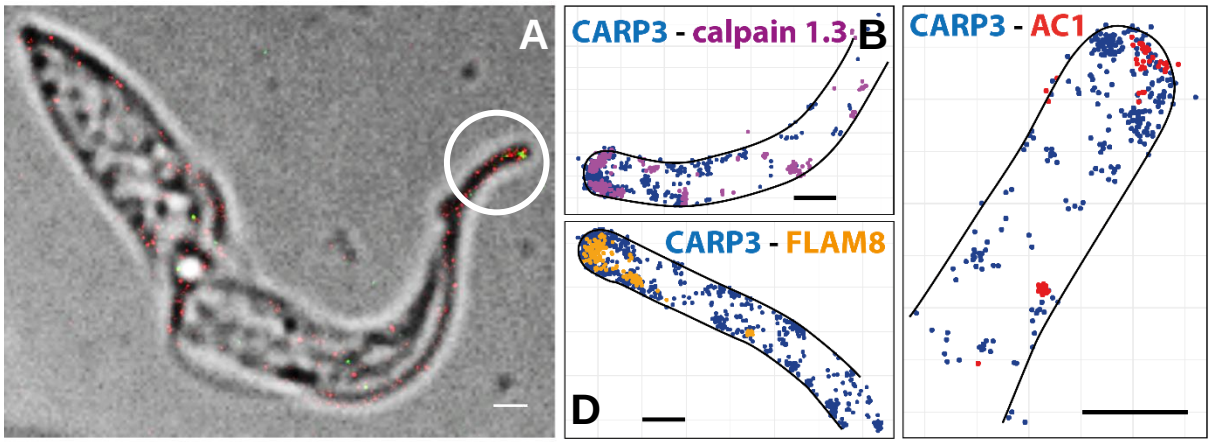


Figure 7.1: Dual-color PALM imaging of proteins enriched at the flagellar tip in trypanosomes. Overlap between bright field and normalized Gaussian rendering of CARP3-PAmCherry (red) and FLAM8-mNeonGreen (green). Both proteins are enriched at the flagellar tip (white circle). Scale bar: 1 µm **(A)**. CARP3-PAmCherry was imaged via PALM in combination with three other fluorescently labeled proteins: calpain 1.3-mNeonGreen **(B)**, AC1-mNeonGreen **(C)** and FLAM8-mNeonGreen **(D)**. Only protein subunits localizing within the flagellar tip, or in proximity of the tip, were tested for co-localization. Scale bar: 0.5 µm **(B, C, D)**

pairs for co-localization via PALM (Figure 7.1 C-D) and compared the obtained results with a positive (experimental) and negative (simulated) control. Furthermore, a previously known membrane associated protein that localize at the flagellar tip, calpain 1.3 (Liu et al. 2010), was also tested for co-localization in order to establish CARP3 relative position to the membrane (Figure 7.1 B). The imaged pairs always comprised the novel cAMP response protein (CARP3) tagged with PAmCherry and a second protein tagged with mNeonGreen (AC1, calpain 1.3, FLAM8 and CARP3 for the positive control). As the aim of the study is to characterize the adenylate cyclase complex located at the flagellar tip, only localizations residing within the terminal part of the flagella were analysed.

The flagellar tip of *Trypanosoma brucei* is characterized by a lateral width of approximately 1 micrometer ($r_B = 500\,nm$). It follows that the theoretical upper boundary for R_{max} is 100 nm ($0.2\,r_B$). However, the proteins that are tested for co-localization against CARP3 all appear to form structures that span for more than 100 nm ($r_A \geq 50\,nm$) (Figure 7.1 B-D), suggesting that the minimum value for R_{max} should be higher than 100 nm. It follows that it is not possible to establish a boundary for the determination of R_{max} that completely avoid border artefacts.

Further unreliability in the analysis derives from the fact that AC1, calpain 1.3 and FLAM8 are tagged with mNeonGreen, a non-photoactivatable fluorophore. As quantitative analysis of the localizations/clustering of mNeonGreen-tagged proteins is not reliable, the interpretation of the absolute CBC values is less obvious (Chapter 6.1).

I circumvented the unreliability caused by the inability to determine a boundary for R_{max} and by the use of mNeonGreen via the application of appropriate controls. Specifically, as a positive control for co-localization I imaged procyclic *T. brucei* AnTat 1.1 expressing CARP3-PAmCherry and CARP3-mNeonGreen. Based on the data obtained from the positive control I then simulated the negative control. Precisely a rectangular area of equal size to the one from the positive control and one micrometer in width was simulated. Within this area I simulated two Poisson point pattern distributions characterized by intensities ($\lambda_{PAm\_sim} = 0.000141\ \mathrm{localizations/nm^2}$, $\lambda_{mNeo\_sim} = 0.000021\ \mathrm{localizations/nm^2}$) in agreement with what observed within the positive control.

As I could only establish an upper threshold for R_{max}, I tested all proteins pairs for co-localization at different R_{max} (50, 100, 200, 300, 400, 500 nm) (Figure 7.2). All proteins pairs tested, simulations included, show a shift in the CBC distribution toward high values with an increase in R_{max} (Figure 7.2 A-E). While the shift is present also in the negative control (border artifact), the degree of the phenomenon varies greatly between the negative control and the remaining protein pairs, as shown by the comparison of the percentages of CBC values above 0.5 (CBC values of mNeonGreen tagged proteins toward CARP3) (Figure 7.2 F). Finally, I compared the CBC value distributions of all proteins

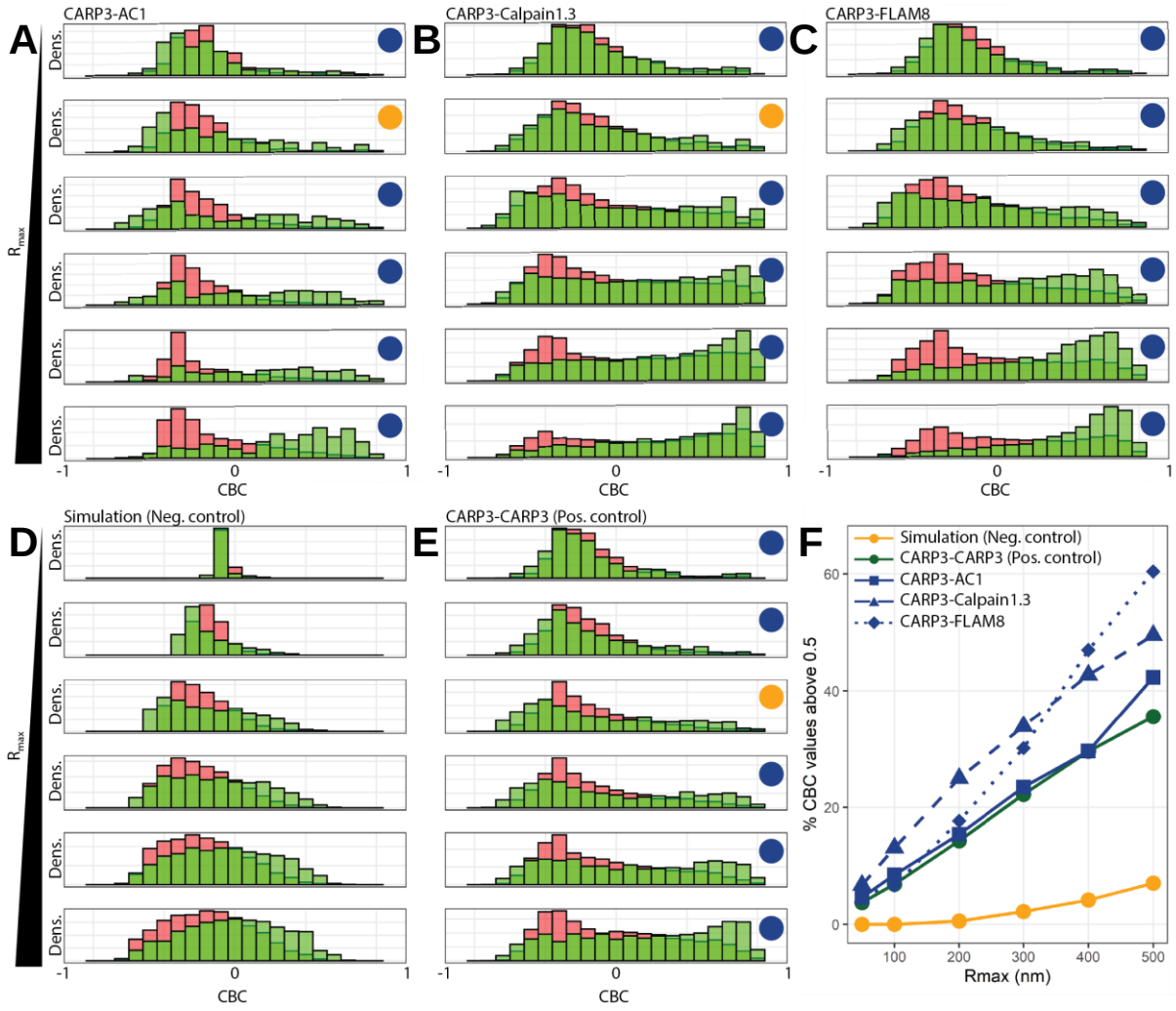


Figure 7.2: Co-localization analysis between CARP3 and three other components of the flagellar tip adenylate cyclase complex (AC1, Calpain1.3, FLAM8). CBC value density distributions for CARP3-AC1 **(A)**, CARP3-Calpain1.3 **(B)**, CARP3-FLAM8 **(C)**, negative control **(D)** and CARP3-CARP3 **(E)** calculated for six different R_{max} (50, 100, 200, 300, 400 and 500 nm). The CBC value density distributions for mNeonGreen tagged proteins toward PAmCherry tagged proteins is shown in green, while the opposite is shown in red. Distributions that differ significantly from the negative control for the respective R_{max} via multiple comparison test after Kruskal-Wallis (one-tailed, p-value<0.05) are marked with a blue dot, while those that do not differ are marked with an orange one. While the percentage of CBC values above 0.5 increases with the increase in R_{max} for all tested pairs (CBC values calculated for mNeonGreen tagged proteins against PAmCherry tagged protein), the percentages of CARP3-AC1, CARP3-Calpain1.3 and CARP3-FLAM8 are equal or greater than the ones of the positive control **(F)**. Interval width = 5 nm**.**

pairs (mNeonGreen tagged proteins toward CARP3) against the negative control via multiple comparison test after Kruskal-Wallis for all R_{max} tested (one-tailed). All protein pairs are characterized by significantly higher CBC values than the negative control for at least five of the R_{max} tested (p-value<0.05) (Figure 7.2 A-E, blue dots) supporting the idea that CARP3, AC1 and Calpain1.3 co-localize with CARP3.

As the various R_{max} used for the determination of the CBC values cover the study of interaction range ($\sim 50\,nm$), cluster association ($\sim 200\,nm$) and more (Figure 6.3) it is possible to further

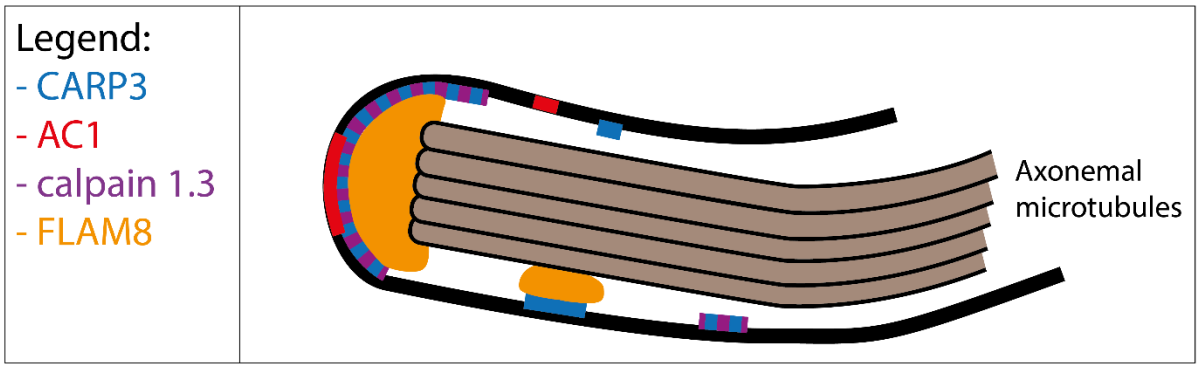


Figure 7.3: Organization of the flagellar tip in *Trypanosoma brucei*: CARP3, AC1, Calpain1.3, FLAM8. FLAM8 possibly transport and deposit CARP3 to the flagellar tip. Here, FLAM8 subunits form a structure that appear to span from the end of the axonemal microtubules to the membrane itself maintaining CARP3 localization. AC1 localizes within the flagellar tip membrane where it possibly interacts with CARP3. Finally, Calpain 1.3 forms distinct membrane associated patches, similarly to CARP3.

interpret the data. Precisely, the CARP3-AC1 pair is the closer in behavior to the positive control (Figure 7.2 F). The CARP3-FLAM8 and CARP3-calpain 1.3 pairs are also characterized by similar CBC values at low R_{max}, however these two pairs are characterized by a faster increase in CBC values for higher R_{max}. The sharper increase in CBC values for the CARP3-FLAM8 pair can be explained by the presence of two closely located FLAM8 fractions, one possibly involved with CARP3 transport and tip deposition (locating within interaction range of CARP3) (Bachmaier, S., et al., "Unpublished data") and one located in proximity of the axonemal microtubules (Rotureau et al. 2014) (Figure 7.3). Similarly, the sharper increase in CBC values for the CARP3-calpain 1.3 pair can be attributed to the presence of closely located membrane associated calpain 1.3 and CARP3 clusters within the flagellar tip (Figure 7.3).

Chapter 8: Discussion

8.1. An unbalanced effort

Following the development of single molecule localization microscopy techniques, there has been an extensive effort from the scientific community towards the development of data analysis packages. These packages mainly focused on the optimization and automatization of the image processing necessary for the extraction of the super-resolved localizations of fluorescent molecules from raw microscopy images (Sage et al. 2019). As algorithm efficiency and quality are pivotal toward achieving a resolving power close to electron microscopy, it soon became necessary to quantitatively compare the available packages. Results concerning the comparison of the different packages were first collected and published in 2015 by Daniel Sage et al. and later expanded in 2019 (Sage et al. 2015, Sage et al. 2019) with a total of 36 different packages being compared for multiple competition modalities.

While it is undeniable that these packages greatly improved the quality of the data extracted from the raw microscope images, SMLM data quantification and interpretation methods are lagging behind (Khater et al. 2020), meaning that a case by case approach has been generally used for the analysis of the point patterns obtained by said packages (especially when analysing non SPT data). In the same way that point pattern representations are fundamentally different from pixel intensity-based representations used in conventional fluorescent microscopy, the analysis approaches used in the two scenarios also differ. Analysis and interpretation of SMLM point patterns is generally more complex (Owen and Gaus 2013, Griffié et al. 2018) and often requires the development of new methods and/or the adaptation of analysis approaches that are encountered in different fields.

Currently, the scarcity in quantification and interpretation methods for SMLM data combined with a general lack of know-how concerning the analysis of point patterns within the fluorescence microscopy community act as a major bottleneck toward the achievement of the full potential by SMLM techniques.

Within this dissertation I described a series of novel, diverse biological scenarios that cover a broad spectrum of the possible challenges encountered in SMLM and the corresponding analytical methods that I developed/used for their analysis.

By providing a step by step guide toward the polishing, analysis and interpretation of SMLM data, this dissertation does not provide the reader with a full technical understanding of the statistics and analysis methods, but is rather aimed at easing the reader into the topic, enabling the interpretation and usage of said methods, therefore reducing said bottleneck.

8.2. Data quality

Even a technically correct data analysis leads to wrong conclusions/results if the underlying dataset is of insufficient quality. It follows that in order to obtain biologically meaningful information it is required to collect reliable datasets. The first step toward this purpose is the ability to distinguish between the localizations that originates from the fluorophore of choice and all other localizations. With that in mind, I constructed a combination of macros (Fiji macro 1-3) and scripts (Script_R1-2) that allows for the characterization and polishing of the obtained localizations in a ROI dependent manner (Chapter 2).

Briefly, the separation of the imaged field of view into separate categories (i.e.: cells within the objective focal plane, cells outside the objective plane and outside area) combined with a low fluorophore activation rate allows for the characterization of the localizations parameters within each separate category, therefore providing as an internal control for the identification and exclusion of, for instance, localizations derived from background fluorescence events.

8.2.1. Experimental controls and autofluorescence

The comparison of localizations originated in WT *C. glutamicum* cells and those expressing PAmCherry fusion constructs (Figure 2.1) is not only meant to serve as an example for the ROI dependent localizations polishing process, but also to highlight the importance of experimental design in SMLM. As most microorganisms exhibits some degree of autofluorescence (i.e.: chlorophyll, carotenoids, flavoproteins, …) (Croce and Bottiroli 2014, Kalaji et al. 2017, Surre et al. 2018), comparison between fluorescence profiles observed in presence and absence of the fluorophore of choice are essential. When imaging WT *C. glutamicum* cells under the same experimental conditions used for the imaging of PAmCherry we can observe autofluorescence (Figure 2.1 Ei,Fi). Most of the localizations that compose the autofluorescence population are characterized by higher PSF half width and photon count compared to PAmCherry (Figure 2.1) and it is therefore possible to exclude these localizations from further analysis. The remaining autofluorescence localizations cannot be distinguished from PAmCherry and act as a de facto noise signal.

In the example shown in chapter 2, the probability to register a localization that satisfy the filtering parameters in a given average sized WT *C.glutamicum* cell in any given frame is 0.0017 while this chance increases to 0.07 (~ 40 times higher) in cells expressing PAmCherry for the same experimental conditions. As autofluorescence signals are usually recorded with a semi-constant frequency across the length of the experiment, it is possible to increase the ratio between PAmCherry localizations and autofluorescence by either decreasing the number of frames imaged or increasing the PAmCherry photo-activation rate (photo-activation rate needs to be maintained below levels which causes multiple PAmCherry localizations to occur simultaneously within diffraction limited areas). A higher

ratio between will in turn translate in the ability to detect smaller variations in the tagged protein behavior.

Finally, case dependent filtering and/or ROI applications can be applied to those cases where autofluorescence does not localize homogeneously.

8.2.2. ROIs determination

In its current form, the detection of regions of interest via the Fiji macros 1, 2 and 3 comprises the conversion of a bright field image to a binary mask followed by semi-automated watersheding and ROIs coordinates extraction. The final outputs of these three macros are collected within a "Cells" folder and consist of a series of .txt files, one for each ROI, that contain the coordinates of the ROIs as a "tab" separated X/Y table (Giacomelli 2020). These outlines will then be imported in R and processed via the Script_R1 in order to make them compatible with Script_R2.

Development and implementation of automated and semi-automated cell segmentation, whether we are taking about bacterial or eukaryotic cells, is a rapidly evolving field. While, at first, parameters based software such as MicrobeJ and Oufti (Ducret et al. 2016, Paintdakhi et al. 2016) struggled to segment crowded fields of view, machine learning algorithms stood up to the challenge (Van Valen et al. 2016, Arganda-Carreras et al. 2017, Berg et al. 2019). I therefore deliberately implemented a simplistic segmentation method via the Fiji macros1-3. This approach can be easily replaced with alternative/more accurate segmentation methodologies so long as the final output of these segmentation methods include the ROIs coordinates.

Regardless of the segmentation method used for the identification of the ROIs, the establishment of boundaries is essential for the reliable analysis of the point patterns obtained in the course of a SMLM experiment. This does not only allow for the exclusion of localizations residing outside the ROIs from the dataset, but also for the association between localizations and ROIs (Script_R2). Moreover, point pattern analysis often requires border correction methods (Hand 2008, Baddeley et al. 2015), as absence of such correction can translate in artifacts (Chapter 6.1). Finally, the determination of morphological parameters (area, length, width) for the ROIs is a necessary step toward the normalization of the data (i.e.: localizations density) and can be instrumental for the division of the obtained localizations into multiple subsets. This is especially relevant when analysing bacteria, where cell size/length can be used as direct indicators of the cell cycle state or of a phenotype (i.e.: increase in cell length due to DNA damage).

8.3. The pitfall of visualization

Following the exclusion of localizations that reside outside the ROIs and the application of fluorophore dependent filtering parameters (Script_R2), the data comprise exclusively filtered

fluorophore localizations and associated ROIs. These data can either be converted back into a fluorescent-like image (i.e.: Normalized Gaussian rendering, Cumulative Histograms, …) or analysed as point patterns. It is however necessary to be keep in mind specific notions prior further processing. These notions are especially important for people that are already in possess of previous experience with fluorescence microscopy, as they may be tempted to analyse and interpret the data, especially when transformed into rendering, in the same way as conventional fluorescent images.

When describing the intracellular localization of a fluorescently labeled protein via conventional fluorescence microscopy there are few questions we can directly answer without the necessity to perform measurements: is the fluorescence homogeneously distributed or do we observe heterogeneity? Is the protein localizing in the cytoplasm or is it a membrane-integrated/membrane-associated protein? While these questions can also be answered via SMLM, the answer is not always as trivial and visualization of the data via rendering can be misleading. The scenario that best exemplify this phenomenon is the distinction between homogeneous and heterogeneous fluorescence.

8.3.1. Clustering should not be determined visually

When imaging a cell expressing a free monomeric fluorophore via conventional fluorescence microscopy it will appear as a uniform fluorescence signal, where the fluorescence intensity is function of the fluorophore density. If the same cell is imaged via SMLM, it will instead appear as a distribution of points on a plane. This distribution follows a homogeneous Poisson point process, also called purely random process (Chiu et al. 2013). Briefly, each localization/fluorophore is stochastically independent from the others and has the same probability to localize anywhere within the cell, while the number of points localizing within the cell is determined by the intensity (λ, expressed as localizations/area) of the Poisson process. As each point is independent, local distribution variability can still be observed

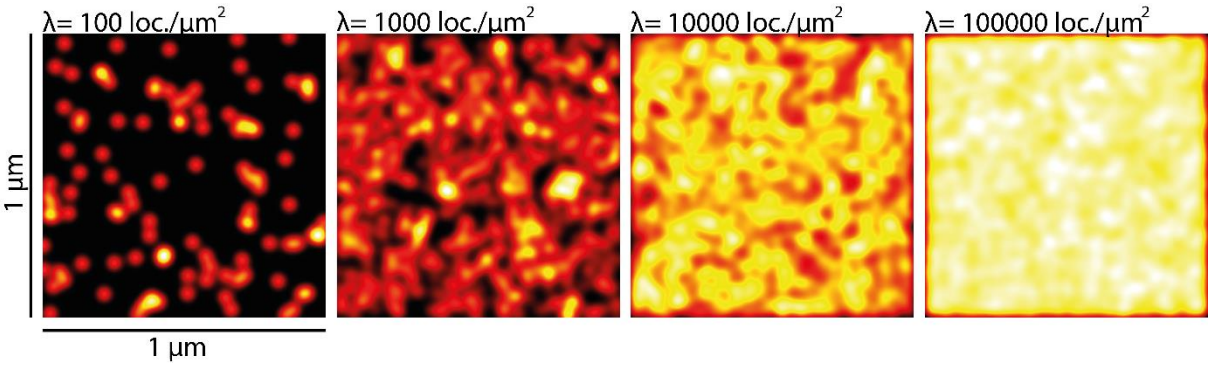


Figure 8.1: Normalized Gaussian rendering of four homogeneous Poisson point processes. Four homogeneous Poisson point processes of increasing density (λ = 100, 1000, 10000, 100000 localizations/µm^2) and 1 µm^2 in area were simulated via the "rpoispp()" command from the "spatstat" package. The localizations were then imported in ThunderSTORM and visualized as normalized Gaussian rendering. Lateral uncertainty was forced to 20 nm for the visualization (common uncertainty found within a PALM experiment). The visual dynamic range was maximized for each image individually.

and recognized by clustering algorithms (Figure 4.1). Moreover, rendering of randomly distributed localizations will translate to uniform fluorescence only in presence of high fluorophore density, while lower concentrations will appear spotty (Figure 8.1) (the field of view appears spotty even at 10000 localizations/μm^2, with 95% of the localizations characterized by a nearest neighbor distance < 10 nm).

Given the approximate size of a conventional GFP derived fluorophore barrel (3nm in diameter, 4 nm in length), a nearest neighbor distance smaller than 10 nm signify quasi-proximity and implies that the majority/entirety of the space is occupied by a single fluorescently labelled protein (barely any space left for other proteins). It follows that the Normalized Gaussian rendering of a cytoplasmic monomeric protein can never translate into uniform fluorescence and that homogeneity cannot be determined visually in SMLM.

Moreover, while the rendering of SMLM data usually appear spotty and theoretically preclude the viewer from easily determining whether a protein is uniformly/randomly distributed across the cytoplasm/membrane, this determination, when done via conventional fluorescence microscopy, does not give a complete view of the actual situation. In fact, a cell expressing sufficiently high levels of fluorescently labelled cytoplasmic protein will appear homogeneously stained whether the protein is monomeric or oligomeric.

We therefore need to rely on a different methodological approach in order to determine whether the tagged protein form clusters/localizes randomly. One such methods is the determination of the nearest neighbor distance function $G(r)$ (edge corrected) for the experimental process and its comparison with a theoretical function (Chapter 4.1). This analysis approach cannot only be used as a preliminary test for clustering but also to determine under which conditions a given protein undergoes clustering. Exemplary for the use of $G(r)$ toward such goal is the study of MamY clustering at variable protein concentrations (Chapter 4.1) (Toro-Nahuelpan et al. 2019).

Similarly, it is possible to determine whether the distribution of dimers/small oligomers follow a Poisson distribution by comparing their estimate of $G(r)$ with the one obtained for a Matern cluster process characterized by the same Poisson process intensity and average cluster size of the experimental oligomers (a Matern cluster process consist of a series of randomly distributed clusters of normally distributed size – “rMatClust” command) (Matern 1960, Matern 1986, Waagepetersen 2007).

While there exist other equally viable methods to determine deviation from a Poisson point process (Empty space function $F(r)$, Ripley's reduced second moment function - $K(r)$, summary function - $J(r)$), the estimation of $G(r)$ is both of simpler interpretation and can provide useful information concerning the nature of the clusters themselves.

The script responsible for the estimation and comparison of $G(r)$ (Script_Rv) selectively implement specific functions contained within the "spatstat" R package to the filtered localizations and ROIs. Moreover, with the purpose of improving the reader familiarity with the estimation and comparison of $G(r)$, I also included within Script_Rv the possibility to simulate uniform, Poisson and clustered point processes ("rMatClust").

8.3.2. Co-localization methods used in conventional fluorescence microscopy are not useful for SMLM data

A second analytical aspect where SMLM differs on a fundamental level from conventional fluorescence microscopy is the study of co-localization. While neither in SMLM nor in conventional fluorescence microscopy co-localization is synonym of interaction (Dunn et al. 2011), the resolution limit of light limits co-localization analysis in conventional fluorescence microscopy to the study of protein co-occurrence within subcellular compartments.

As the typical resolution in conventional fluorescence microscopy is approximately 200 nm (see chapter 1.2), co-localization between two fluorescently tagged proteins will translate in overlap of signal (Figure 3.4 C). It follows that when described statistically, co-localization analysis comprises the determination of an overall co-localization coefficients based on pixels intensity values across the fluorescent channels (Pearson Correlation Coefficient, Mander overlap coefficient, Fractional overlap) (Pearson 1896, Manders et al. 1992, Manders et al. 1993, Dunn et al. 2011).

In the same way that, due to an increase in resolution, the rendering of randomly distributed localizations obtained via a SMLM experiment does not result in a fluorescently homogeneous field of view (see chapter 8.3.1) (Figure 8.1), the imaging of two proteins that co-localize in conventional fluorescence microscopy (Figure 3.4 C) does not necessarily translate to signal overlap in SMLM (Figure 3.5).

In details, the Normalized Gaussian and histogram rendering derived from the localizations obtained from a SMLM experiment are characterized by higher resolution compared to conventional fluorescence microscopy images. Therefore, the signal overlap that is observed in conventional fluorescence microscopy translates into a series of different protein arrangements when imaged via SMLM. Two proteins may, in fact, localize within the same cell structure and show no overlap in their rendering due to low localization density (i.e.: two membrane proteins) or be associated to different parts of the same structure (i.e.: membrane protein and membrane-associated protein), resulting in a shift between the two colors. Consequently, if the aim of the co-localization analysis described here is to verify whether said proteins belong to the same structure, the analysis methods used for conventional fluorescence microscopy are not viable.

8.3.3. Why and when to use conventional-like image analysis

In chapter 3 I characterized DipC, a novel putative DivIVA interaction partner expressed during the DNA damage induced SOS response in *C. glutamicum.* Furthermore, I explored the advantages and limits of conventional-like image analysis in both qualitative and quantitative studies.

First, by combining qualitative fluorescence profiles and the unique positioning of DivIVA and DipC within the septal area I was able to determine that DipC is recruited to the septum indirectly, possibly by DivIVA. Then, by comparing quantitative DipC fluorescence profiles across genetically different strains I was able to show that DipC localization and expression levels are altered in absence of DivS, the protein responsible for division inhibition during the DNA damage induced SOS response in *C. glutamicum* (Ogino et al. 2008).

Fluorescence profiles can be calculated from both normalized Gaussian rendering and cumulative histogram rendering, with the latter being less appealing from a visual point of view but having the advantage of being quantitative (the pixel intensity value in a cumulative histogram rendering correspond to the number of localizations residing within the pixel).

While everything that was determined in chapter 3 via conventional-like image analysis can also be calculated via point pattern analysis methods, image analysis tools (Fiji, MicrobeJ, Oufti, BactMAP) (Schindelin et al. 2012, Ducret et al. 2016, Paintdakhi et al. 2016, van Raaphorst et al. 2020) are generally more common and user friendly. Therefore, a significant difference in the required effort and skills is currently necessary to achieve the same goal (whether this means the ability to use R, Matlab or other script driven statistical computing tools) (R_Core_Team 2020, The_Mathworks_Inc 2020). This may appear counterintuitive as I dedicated two separate scripts to the analysis of conventional-like image analysis in chapter 3.2., however visualization of data via Script_R3 can be replaced by other software such as MicrobeJ, Oufti and BactMAP, while Script_R4 can be replaced by the use of spreadsheets analysis tools such as Microsoft Excel (Microsoft_Corporation 2019).

Finally, it is likely that the development of SMLM focused point pattern analysis tools will eventually make the combined use of rendered SMLM images and conventional-like image analysis obsolete. However, this is currently not the case and conventional-like image analysis will still play a role in the close future.

8.4. Benefits and flaws of OPTICS, a quasi-parameter free approach to cluster analysis

Currently available point pattern clustering algorithms such as Density-based spatial clustering analysis with noise (DBSCAN) require difficult to determine parameters and output different results based on the parameters used (Nan et al. 2013, Deschout et al. 2014, Khater et al. 2020). OPTICS, on

the other side, requires a single parameter for the determination of the clustering organization of a point process (ε is not a necessary parameter as it is only required to fasten the computing) and is rather insensitive to the parameter itself (Ankerst et al. 1999). Moreover, visualization of OPTICS outputs via reachability plots is of easy interpretation (Figure 1.2). I therefore implemented OPTICS within the part of the SMLM analysis script I developed dedicated to cluster analysis.

8.4.1. Arbitrarity of thresholding and analysis complexity

Following the construction of a reachability plot, Script_R5P2 allows for the identification of two separate clustering levels via the manual establishment of thresholds (chapters 4.3 and 5) (Figure 8.2 A, C). As the OPTICS algorithm normally allows for automatic analysis of cluster-ordering and generate accordingly a hierarchical clustering structure (Ankerst et al. 1999)(Figure 8.2 B, D), the manual

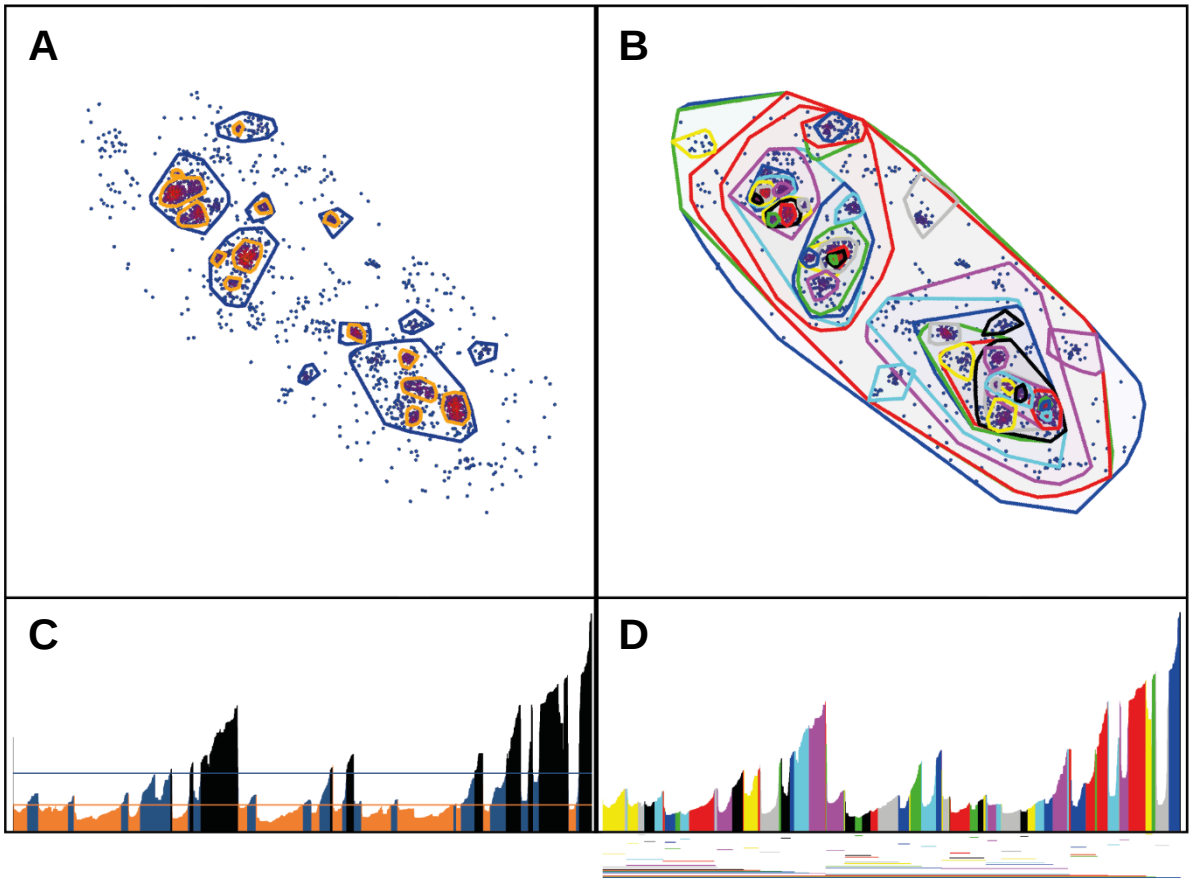


Figure 8.2: Manual and automated clustering of ParB-PAmCherry via OPTICS. The two reachability distance thresholds selected highlight general ParB enrichment areas (royalblue) and density variability within said areas (orange) (see chapter 5 for more details) **(A)**. While more detailed, hierarchical clustering visualization gives no hint concerning which clusters share similar characteristics **(B)**. Similarly, the interpretation of reachability plots in presence of two threshold levels is of immediate understanding **(C)** while the use of the same color for multiple hierarchical levels, combined with the high number of levels, makes it virtually impossible to interpret and recognize clusters within reachability plots originated from automated hierarchical clustering **(D)**.

selection of thresholds somehow limits OPTICS potential and can subject it to the arbitrarity of the user.

While the exact definition of the thresholds will always maintain a certain degree of arbitrarity, the approximate value each threshold is based on is determined visually according to the reachability plots (Figure8.2 C), as described in chapters 5 and 5.2, and is aimed toward the subdivision of the dataset into biologically relevant categories (enriched areas, cluster heterogeneity, identification of separate low/high density areas). It follows that this process excels in the determination and comparison of such features, as exemplified by the characterization of ParB clustering behavior in presence/absence of *parS* sites (Chapter 5.2)(Bohm et al. 2020). The segmentation of localizations via a finite number, in this case two, of manually established thresholds results in a loss of descriptive ability (Figure 8.2). However, hierarchical structuring focuses on identifying clusters rather than grouping (Figure 8.2 B). It follows that post-structuring analysis is required for the grouping of the obtained clusters into meaningful categories (a process that is once again partially arbitrary).

I therefore did not implement within Script_R5P2 the commands required for the automated determination of the hierarchical clustering structure, but rather focused instead into separating the dataset in three clustering categories (background, level_1 and level_2) and integrate labels with localizations specific data (obtained via Script_R5P1), therefore enabling the characterization of the clusters themselves and cluster dependent analysis.

8.4.2. Alternatives and future prospects

There currently exist multiple analysis approaches dedicated to the identification and characterization of clusters in SMLM. Generally, these methods can be divided into Statistical (estimation of $G(r)$), Bayesian, Density based (OPTICS), Voronoi based and Graph based (Khater et al. 2020). Each of these approaches has strengths and weaknesses. While some of these weaknesses cannot be circumvented (computational time/scalability), most of them can be bypassed via the use of appropriate biological and *in silico* controls. Moreover, while some algorithms are better apt at recognizing specific type of clusters compared to others (Khater et al. 2020), the lack of benchmark datasets dedicated to the assessment of the quality of the algorithms in regards to said clusters means that no full scale objective comparison has been made to this date. Currently, an ideal scenario would therefore require the freedom to implement and compare multiple clustering algorithms from the same software/platform, enabling the exploitation of each algorithm strength in a case by case scenario. However, it is not trivial to construct such a platform due to the speed at which the analysis methods have been evolving in the last years. Finally, the role of machine learning algorithms have been increasing, covering aspects ranging from the segmentation of clusters (Williamson et al. 2020)

to the classification of biological structures (Khater et al. 2019) and are expected to play an even bigger role in the years to come.

8.5. Benefits and flaws of CBC

Script_R5P3 currently implements and expand on the coordinate-based co-localization (CBC) approach based on the user defined input of interval width (nm) and R_{max} (Malkusch et al. 2012, Georgieva et al. 2016, Pageon et al. 2016). While neither this method nor other currently available co-localization methods can be used to determine protein-protein interaction (Dunn et al. 2011), the determination of CBC values can be used to infer on a variety of different biological scenarios, from the co-occurrence of two proteins within the same subcellular compartment, to the characterization of a microcompartment (Chapter 6). Furthermore, as each localization is characterized by a separate CBC value, it is possible to study the change in CBC values across a protein population relative to, for instance, its oligomeric state.

While the coordinate-based co-localization method is quite robust in regards to changes in local molecular density, it is also strongly influenced by the parameters used for its determination (Levet et al. 2019) (Figure 6.3). The dependency between results and input parameters is usually seen as a weakness for an analysis method, as it implies that the method itself is inconsistent. However, making such a claim for CBC analysis would not make sense, as changes in the input parameters do not correspond to an alteration in the results, but rather to a change in experimental question (Lagache et al. 2015), making this analysis method extremely flexible (Chapter 6).

In recent years, there have been developed several methods dedicated to the analysis of co-localization in SMLM experiments (Malkusch et al. 2012, Lagache et al. 2015, Pageon et al. 2016, Levet et al. 2019). While most of the available methods, including the one described within this dissertation, rely directly on coordinates for the determination of co-localization (Malkusch et al. 2012, Pageon et al. 2016), there has been recently proposed a method that is instead based on tessellation (Levet et al. 2019).

This analysis approach, named Coloc-Tesseler (CT), derives the spatial co-organization of two molecular species based on the overlap between Voronoi diagrams in a parameters free manner while also allowing for the determination of image-based co-localization coefficients, such as the Manders and Spearman's coefficients, in a straightforward manner (Levet et al. 2019). While the lack of parameters translates into a more consistent and straightforward analysis, it also means that it is not possible to determine the type of co-localization to which we are interested, therefore limiting its potential.

Overall, coordinate-based co-localization (CBC) requires a certain degree of knowledge concerning the proteins distributions (interaction range/cluster size/cell size) that are being analysed in order to determine appropriate boundaries for the parameters (chapter 6.3). However, as said knowledge can be provided by the script hereby presented prior to the co-localization analysis itself, the choice of acceptable parameters is quasi-trivial. Finally changes in CBC values due to a change in the experimental question are gradual, allowing a certain degree of tolerance in the choice of parameters, and quality of the data can be assured by the presence of negative and positive controls (Chapter 6.2).

8.6. An extra layer of analysis

The analysis of SMLM data differ from the analysis of fluorescence in conventional fluorescence microscopy similarly to how single cell analysis differ from whole population analysis. The ability to assign to each fluorophore parameters concerning its localization, clustering environment, co-localization status and concerning the cell it localizes in, means that a change in paradigm is necessary when approaching SMLM data for the first time.

For instance, in conventional fluorescence microscopy, Abbe limit and fluorescent background conveniently act as technical lower bound thresholds for both clustering and co-localization. The lack of this limits in SMLM means that it is up to us to establish new analysis thresholds. These limits are determined according to the distribution of the studied parameters (cluster size, density, cell size, …) and should ideally have biological significance (see chapters 5.1 and 5.2).

Furthermore, while within this dissertation I purposely refrained from integrating/combining multiple analysis aspects where unnecessary, it is not only common but also beneficial to study localization/co-localization in a clustering-dependent manner. With this in mind, I structured the data in a way that allows for the simultaneous application of multiple filtering parameters on a cell, cluster and localization level. As this increases the degree of control over the data themselves, it is possible to isolate and characterize structures/phenotypes that would be otherwise lost (Martins et al. 2019).

8.7. Conclusion and outlook

This work covers, via the use of multiple biologically relevant scenarios, pre- and post-processing of SMLM data while emphasizing the differences, benefits and limits of point pattern analysis.

I presented each module/script with possible alternatives while showing both their strength and weaknesses/limitations. It is worth mentioning that the pipeline presented within this dissertation is the result of a combination and optimization process that finds its origins within the necessity to determine specific protein features in the course of published (Martins et al. 2019, Toro-Nahuelpan et al. 2019, Bohm et al. 2020) and unpublished studies (Bachmaier, S., et al., “Unpublished data”) and

both the current version of the pipeline and the versions used for each individual study are collected within my GitHub page (Giacomelli 2020) and are freely available.

In its current version, the script uses as sole inputs a bright field image, for the determination of the ROIs, and a Zeiss Elyra P1 derived molecules localizations table (the table format is shown in the GitHub page). As not all table parameters are used within the script (Relevant parameters: index, first frame, position X, position Y, precision, number of photons, PSF width), it is possible to use localizations tables obtained from other microscopes/software, so long as the relevant parameters are available and the dataset is formatted accordingly.

The SMLM data filtering/analysis pipeline hereby presented allows for the characterization and comparison of a wide range of variables. Said variables describe and allow for the combination of multiple levels of analysis (single molecule, cluster, cell, field of view). Generally, I was able to filter raw localizations tables according to specific ROIs and in a fluorophore-dependent manner (photon count/PSF width). I was then able to convert the filtered data to a rendered image and analyse it via conventional-like image analysis (Script_R3-4), and/or directly process the filtered localizations table via point pattern analysis (Script_Rv, Script_R5P1-3). I then showcased multiple aspects of each part of the pipeline via experimental and/or simulated data with specific focus toward clustering/cluster heterogeneity and co-localization.

Specifically, using the conventional-like image analysis approach (Script_R3, Script_R4), I determined the relative positioning of DivIVA and DipC within *C. glutamicum* septal region, with the difference between DivIVA and DipC cross-septal distances supporting a model where DipC does not directly interact with the membrane but is rather recruited by a membrane associated protein, possibly DivIVA. I also used a variation of the same approach (histogram based rendering vs normalized Gaussian rendering) to determine in a semi-quantitative fashion the changes in DipC localization and expression in two different *C. glutamicum* genomic backgrounds (WT and *ΔdivS*) while exposed to the DNA inducing antibiotic Mitomycin C. Briefly, absence of DivS resulted both in a decrease in DipC signal and its delocalization.

Via the use of point pattern analysis methods (Script_Rv, Script_R5P1-2) I was then able to determine that the filament-like structures formed by the protein responsible for magnetosome alignment in *M. gryphiswaldense*, MamY, do not originate from a single nucleation event. They are instead the result of the independent localization of multiple oligomeric structures of similar molecular density that can recognize the geodetic membrane area. I subsequently showcased a separate aspect of cluster analysis (multi-leveled clustering) by comparing ParB clustering across different *C. glutamicum* genomic backgrounds (WT, $parS_{1\text{-}10mut}$ and $parS_{2\text{-}10mut}$). Briefly, while high density ParB

areas (LV1 clusters) are composed of a significantly higher number of localizations in the WT strain, suggesting that the degree of ParB propagation is dependent on the number of *parS* sequences, no statistical difference can be observed concerning the number of subclusters/cluster across the three strains. I could instead observe a significantly higher occupancy for the strain possessing no *parS* sites, supporting the idea of a decreased cluster heterogeneity in said strain (see chapter 5).

Based on simulated point patterns, I then established a general rule for the determination of co-localization in all its forms ($2r_A \leq R_{max} \leq 0.2r_B$) via CBC values (Malkusch et al. 2012) (ScriptR5P3). I hence determined CBC values for different R_{max} in a biologically relevant scenario, including a positive biological control and a negative simulated control. Briefly, I tested the cAMP response protein CARP3 for co-localization with itself and three other proteins involved in the organization of the flagellum tip in Trypanosoma brucei brucei (AC1, Calpain1.3, FLAM8) and compared them with the co-localization values obtained for the CARP3-NegativeControl pair. I proceeded to show that for all R_{max} tested (50 to 500 nm) AC1, Calpain1.3 and FLAM8 return CBC values similar or higher than the once returned by the positive control, suggesting that all these proteins are part of the same molecular complex.

Finally, as SMLM data analysis is still in its dawn, new computational methods are sure to be developed in the near future. While some of these methods will be aimed toward the extraction of specific biosignatures/have a niche field of use, others will overlap in function to existing analysis methods, possibly making the current version of this pipeline obsolete. While said increase in analysis methods has theoretically a positive effect, it will also make it extremely difficult for researchers being introduced for the first time to SMLM to navigate their way through the data. As an increase in methods without an objective way to compare these methods is somehow pointless, it is pivotal for the scientific community to establish in the near future benchmark SMLM data with known features with the sole aim of validating/comparing the separate methods, similarly to what was done concerning the identification of localizations from raw data (Sage et al. 2019).

Chapter 9: Material and Methods

9.1. Code availability

The complete collection of macros and scripts used in the analysis of SMLM data has been deposited in Github (https://github.com/GiacomoGiacomelli).

9.2. Reagents

Experiments were performed using chemicals purchased from the following companies: AppliChem (Darmstadt, Germany), Ibidi (Gräfelfing, Germany), Merck (Darmstadt, Germany), Roth (Karlsruhe, Germany), Sigma Aldrich Chemie (Steinheim, Germany), VWT (Radnor, PA, USA). Molecular biology products were further obtained from GE Healthcare (Chicago, IL, USA), Genaxxon Bioscience (Ulm, Germany), Invitrogen (Carlsbad, CA, USA), Macherey & Nagel (Düren, Germany), New England Biolabs (Ipswich, MA, USA) Roche (Rotkreuz, Switzerland), Thermo Fisher Scientific (Waltham, MA, USA) with purity grades "per analysis", unless otherwise noted.

9.3. Oligonucleotides and plasmids

Oligonucleotides used for the construction of plasmids and sequencing are listed in table 9.1. Plasmids used in the course of the doctoral study and/or used during the collaborations mentioned within this dissertation are listed in table 9.2. Construction of the plasmids is described in the next chapters.

Table 9.1: Oligonucleotides used for the construction and sequencing of plasmids

Restriction sites are underlined. STOP codons, thrombin cleavage site, HisTag, StrepTagII and linkers are indicated in bold.

Oligonucleotides (Plasmids construction)	Sequence 5´- 3´	Restriction Site
HindIII-5´AdhA-F	GCATCGAAGCTTATGACCACTGCTGCAC	HindIII
SalI-IsceIrec-5´AdhA-R	GCATCGGTCGAC**ATTACCCTGTTATCCCTA**CGGGCG GGTTTCAG	SalI
SalI-3´AdhA-F	GCATCGGTCGACGGCCAATTCATGGTGATC	SalI
XbaI-3´AdhA-R	GCATCGTCTAGATTAGAAACGAATCGCCACAC	XbaI
SalI-PAmCherry-F	CATGTCGACATGGTGAGCAAGGG	SalI
XbaI-STOP-PAmCherry-R	CATTCTAGA**TTA**CTTGTACAGCTCGTC	XbaI
SalI-mNeon-F	TATGTCGACATGGTGAGCAAGG	SalI
XbaI-STOP-mNeon-R	ATGTCTAGA**TTA**CTTGTACAGCTCGTC	XbaI
HindIII-3´Cg0839-F	CATAAGCTTTTAATGTGATTCGT	HindIII
SalI-3´Cg0839-R	CATGTCGACAATTGCAGTTGG	SalI
XbaI-Cg0839-Do-F	CATTCTAGAGGATCACTATGTCC	XbaI
XmaI-Cg0839-Do-R	TATCCCGGGCGGGATGAGGAAAG	XmaI
HindIII-DivS-Do-F	CGATGCAAGCTTGTGTCTGCTAGTCCCTTC	HindIII

SalI-DivS-Do-R	CGATGCGTCGACTCGGGAATAGATCAATGCTG	SalI
SalI-DivS-Up-F	CGATGCGTCGACTTCAAAACTCCCGCCTCAATTG	SalI
EcoRI-DivS-Up-R	CGATGCGAATTCGGATTGAAGTCCTGCAGCATC	EcoRI
SalI TAA StrepTag 3´Cg0839 R	GCATCGGTCGAC**TTACTTTTCAAACTGCGGATGGCT CCA**AATTGCAGTTGGGATTATATTG	SalI
HindIII 3´Cg0839 F	GCATCGAAGCTTCCACAAGCCACTGAG	HindIII
SalI-Cg0839-Do-F	GCATCGGTCGACGGATCACTATGTCCAACG	SalI
XbaI-Cg0839-Do-R	GCATCGTCTAGACGGGATGAGGAAAGAC	XbaI
OH1-Cg0839-Do-F	**TCTACCGACGCACTCATG**GGATCACTATG	-
SalI-Cg0839-Do-R	CATAGTCGACCGGGATGAGGAAAG	SalI
OH2-Cg0839-Up-R	**CATGAGTGCGTCGGTAGA**GGAATGAGGCC	-
HindIII-Cg0839-Up-F	CATAAAGCTTATGGTGCCATCG	HindIII
EcoRI-Cg0838-Do-R	CATG GAATTC CACCAGCCACGCTAC	EcoRI
OH1-Cg0838-Do-F	**TCTACCGACGCACTCATG**GAGTAGATAACACCAGCT ACC	-
OH2-Cg0839-Up-R	**CATGAGTGCGTCGGTAGA**GGAATGAGGCCTTTG	-
XbaI-Cg0839-Up-F	CATGTCTAGAGTGCCATCGATTTTG	XbaI
HindIII-Cg0838-Up-F	GCATCGAAGCTTCACCACAAGCCACTGAG	HindIII
SalI-Cg0838-Up-R	GCATCGGTCGACAGTGATCCTTAAATTGCAGTTGG	SalI
XbaI-5´Cg0838-F	GCATCGTCTAGAATGTCCAACGCACCTAAAAAG	XbaI
BamHI-5´Cg0838-R	GCATCGGGATCCCGGAGCAGTTGGTCAAAG	BamHI
XbaI-mCherry-R	GCATCGTCTAGACTTGTACAGCTCGTC	XbaI
XbaI-mNeon-R	ATGTCTAGACTTGTACAGCTCGTC	XbaI
HindIII-3´FtsZ-F	CAGAAGCTTGAAGTCAACGCAGCTGCA	HindIII
SalI-StrepTag-3´FtsZ-R	CATGGTCGAC**TTACTTTTCAAACTGCGGATGGCTCC A**CTGGAGGAAGCTG	SalI
SalI-FtsZ-Do-F	CATGGTCGACTTAAGAAGGAGAATAG	SalI
EcoRI-FtsZ-Do-R	CAGGAATTCGCACCCATGAGCGCATG	EcoRI
EcoRI-MurJ-Do-R	GCGGGAATTCGTTTGTGGAGCC	EcoRI
OH1-MurJ-Do-F	**TCTACCGACGCACTCATG**ATTACGCGTTTGTG	-
OH2-MurJ-Up-R	**CATGAGTGCGTCGGTAGA**ATCAGAATCGGAG	-
XbaI-MurJ-Up-F	CATGTCTAGATCTCACGTTGTG	XbaI
PstI-3´MurJ-F	CATCTGCAGGTGGGATGACGGTGC	PstI
SalI-3´MurJ-R	GATCGTCGACCCAACCAACAAGTTG	SalI
XbaI-MurJ-Do-F	CATTCTAGAATTACGCGTTTGTGA	XbaI
EcoRI-MurJ-Do-R	CATGAATTCGGTTTCCGCCCGTA	EcoRI
PstI-3´Cg0838-F	CATGCTGCAGCCAAACTGACATCG	PstI
SalI-3´Cg0838-R	CATGGTCGACCTCAGATTTTGGCATC	SalI
XbaI-Cg0838-Do-F	CATGTCTAGAATAACACCAGCTACCCAC	XbaI
BamHI-Cg0838-Do-R	CATAGGATCCTCGCACTTCACGC	BamHI
HindIII-3´Cg0841-F	CATGAAGCTTACCACGGGCTTTTTG	HindIII
SalI-3´Cg0841-R	CATGGTCGACGCTATCCAATTC	SalI
XbaI-Cg0841-Do-F	CATGTCTAGATAGCCGTAGTCG	XbaI
XmaI-Cg0841-Do-R	CATGCCCGGGAAAACGCACTGT	XmaI
HindIII-Cg0841-Up-F	GCATCGAAGCTTCGAAACGGACACAATTG	HindIII

SalI-Cg0841-Up-R	GCATCGGTCGACTGGATATTCACCCTTAGAATTG	SalI
XbaI-5´Cg0841-F	GCATCGTCTAGAATGCGTGGTGACGTTC	XbaI
BamHI-5´Cg0841-R	GCATCGGGATCCGATTCGCTGCGGATCATC	BamHI
HindIII-LexA-Up-F	CGATGCAAGCTTCAACGAGGAGAACAGCTTTTTCG	HindIII
SphI-LexA-Up-R	CGATGCGCATGCCTTCCGTTCCCTTCCTCTC	SphI
SphI-LexA-Do-F	CGATGCGCATGCGTCGCTTTTCAGGTTCCC	SphI
EcoRI-LexA-Do-R	CGATGCGAATTCGGCAATGGAGTACTTTGCTG	EcoRI
EcoRI-STOP-PAmCherry-R	CATGAATTC**TTA**CTTGTACAGCTCGTC	EcoRI
BamHI-5´GFP-F	CATGGATCCATGAGTAAAGGAGAAG	BamHI
EcoRI-STOP-3´GFP-R	CATGAATTC**TTA**TTTGTATAGTTCATCC	EcoRI
BamHI-mNeon-F	CATGGATCCATGGTGAGCAAGGGC	BamHI
EcoRI-STOP-mNeon-R	CATGAATTC**TTA**CTTGTACAGCTCGTCCATGCC	EcoRI
BamHI-PSmOrg.2-F	CATGGGATCCATGGTGAGCAAGGGCG	BamHI
EcoRI-STOP-PSmOrg.2-R	CATGGAATTC**TTA**CTTGTACAGCTCGTCCATG	EcoRI
SalI-5'Murj-F	CATGTCGACGTGGTGCGCTCGAC	SalI
BamHI-STOP-3'Murj-R	CATGGATCC**TTA**CCAACCAACAAGTTG	BamHI
BamHI-STOP-6His-3'MurJ-R	CATGGATCC**TTAATGGTGGTGGTGATGATG** CCAACCAACAAGTTG	BamHI
PstI-5´MurJ-F	GATACTGCAGGTGGTGCGCTCGAC	PstI
PstI-RBS-5´MurJ-F	GATCTGCAGCACACCTCCGATTCTG	PtsI
Sal-RBS-His-5'GFP-F	CATGTCGACATCGAAGGGAATCCGCAAATG**CATCA TCATCATCATCAT**ATGAGTAAAGGAGAAG	SalI
BamHI-3'GFP-R	CATGGATCCTTTGTATAGTTCATCC	BamHI
BamHI-THR-5'DivIVA-F	CATGGATCC**CTGGTTCCGCGTGGTTCC**ATGCCGTTG ACTCCAG	BamHI
SacI-STOP-3'DivIVA-R	CATGAGCTC**TTA**CTCACCAGATGGCTTG	SacI
SalI-RBS-5'DivIVA-F	CATGTCGACATCGAAGGGAATCCGCAAATGCCGTT GACTCCAG	SalI
BamHI-THR-3'DivIVA-R	CATGGATCC**GGAACCACGCGGAACCAG**CTCACCAG ATGGCTTG	BamHI
SacI-STOP-His-3'GFP-R	CATGAGCTC**TTAATGATGATGATGATGATG**TTTGTA TAGTTCATCC	SacI
BamHI-3'DivIVA-R	CCGGGATCCCTCACCAGATGGC	BamHI
BamHI-Dendra2-F	GAGGGATCCATGAACACCCCG	BamHI
SacI-STOP-Dendra2-R	CCGGAGCTC**TTA**CCACACCTGG	SacI
SalI-Cg0847-F	CATGGTCGACGTGACTGAAAAGTATCGTCC	SalI
BamHI-Cg0847-R	CATGGGATCCGTTAACGCAACGGGGAC	BamHI
SalI-Cg0842-F	CATGGTCGACATGGTTTCTTATAGCGTGCAC	SalI
BamHI-Cg0842-R	CATGGGATCCAACCCGCGGCAGC	BamHI
SalI-Cg0841-F	CATGGTCGACATGCGTGGTGACGTTC	SalI
BamHI-Cg0841-R	CATGGGATCCGCTATCCAATTCCTCTAAACG	BamHI
EcoRI-STOP-Dendra2-R	CATGGAATTC**TTA**CCACACCTGG	EcoRI
EcoRI-STOP-3'mCherry-R	CATGAATTC**TTA**CTTGTACAGCTCGTCC	EcoRI
SalI-RBS-5´FtsZ-F	CATGGTCGACATGACCTCACCGAAC	SalI
SacI-3´FtsZ-R	CATGGAGCTCCTGGAGGAAGCTGGG	SacI

SacI-linker-mCherry-F	CATGGAGCTC**GGGGGTGGCGGTTCAGGCGGTGGG GGTATG** GTGAGCAAGGGCGAG	SacI
EcoRI-STOP-3´mCherry-R	CATGGAATTC**TTA**CTTGTACAGCTCGTC	EcoRI
SacII-linker-5´PAmCherry-F	GCATCGCCGCGG**TCGAG**ATGGTGAGCAAGGGCGA G	SacII
NheI-linker-3´PAmCherry-R	GCATCGGCTAGC**TCCCTCGAG**CTTGTACAGCTCGTC CATGCC	NheI
NdeI-5´mNeon-F	GCATCGCATATGGTGAGCAAGGGCG	NdeI
BamHI-STOP-3´mNeon-R	GCATCGGGATCC**TTA**CTTGTACAGCTCGTCCATG	BamHI
XhoI-6His-F	GCATCGCTCGAGATGGGCCATCATCATC	XhoI
HindIII-GA-3´Neon-R	GCATCGAAGCTT**GA**CTTGTACAGCTCGTCC	HindIII
NdeI-5´DivIVA-F	CATGCATATGATGCCGTTGACT	NdeI
BamHI-STOP-3´DivIVA-R	CATGAATTC**TTA**CTCACCAGATGG	BamHI
BamHI-T-5´Cg0847-F	CATAGGATCC**T**GTGACTGAAAAG	BamHI
SalI-3´Cg0847	CATAGTCGACGTTAACGCAACG	SalI
Oligonucleotides (Sequencing)	**Sequence 5´- 3´**	
pEKEX2_seq_F	CATCGGCTCGTATAATGTGT	
pEKEX2_seq_R	CCGCTTCTGCGTTCTGATTT	
pk19mobsacB_seq_F	GCTTCCGGCTCGTATGTTG	
pk19mobsacB_seq_R	GCTGCAAGGCGATTAAGTTG	
StrepTag _seq_F	GGAGCCATCCGCAG	
FtsZdo_seq_R	GTGGTGCCTTTTGTCGTAC	
cg0839up_F_seq	GCAACCAATGAGAAGACAACCCACCGTATCTAG	
cg0839do_R_seq	CCTGATCATTTGCAAGAGC	
cg0838do_R_seq	GAAGATCCCGACAATGGCATCGCC	
Isce_ins_Fw_seq	GATGCTTTTAAAGGGAATTGTGTG	
Isce_ins_Rv_seq	GAGCAGTTTCAACACAATCCG	
divSup_F_seq	CTGCGCAGGAAGCC	
divSdo_R_seq	TATCGGCGCAGAGAATCC	

Table 9.2: Plasmids

Plasmid	Characteristics	Reference
pk19mobsacB	Integration vector, *ori*pUC, Km^r, mob sac	(Schäfer et al. 1994)
pk19mobsacB-IsceI-recognition-site	Integration vector, *ori*pUC, Km^r, mob sac, Insertion of the I-SceI meganuclease recognition site within the *adhA* gene	This study
pk19mobsacB-Cg0839-PAmCherry	Integration vector, *ori*pUC, Km^r, mob sac, *cg0839-PAmCherry*	This study
pk19mobsacB-Cg0839-mNeonGreen	Integration vector, *ori*pUC, Km^r, mob sac, *cg0839-mNeonGreen*	This study
pk19mobsacB-ΔdivS	Integration vector, *ori*pUC, Km^r, mob sac, deletion of *divS*	This study

pk19mobsacB-DivIVA-PAmCherry	Integration vector, *ori*pUC, Km^r, mob sac, *divIVA-PAmCherry*	This study
pk19mobsacB-DivIVA-mNeonGreen	Integration vector, *ori*pUC, Km^r, mob sac, *divIVA-mNeonGreen*	(Schubert et al. 2017)
pk19mobsacB-ParB-mNeonGreen	Integration vector, *ori*pUC, Km^r, mob sac, *parB-mNeonGreen*	(Bohm et al. 2020)
pk19mobsacB-ParB-PAmCherry	Integration vector, *ori*pUC, Km^r, mob sac, *parB-PAmCherry*	(Bohm et al. 2020)
pk19mobsacB-Cg0839-mCherry	Integration vector, *ori*pUC, Km^r, mob sac, *cg0839 (dipC)-mCherry*	This study
pk19mobsacB-Cg0839-StrepTagII	Integration vector, *ori*pUC, Km^r, mob sac, *cg0839 (dipC)-StrepTagII*	This study
pk19mobsacB-Δcg0839	Integration vector, *ori*pUC, Km^r, mob sac, deletion of *cg0839 (dipC)*	This study
pk19mobsacB-Δcg0838/39	Integration vector, *ori*pUC, Km^r, mob sac, deletion of *cg0838 (dipD) and cg0839 (dipC)*	This study
pk19mobsacB-mCherry-Cg0838	Integration vector, *ori*pUC, Km^r, mob sac, *mCherry-cg0838 (dipD)*	This study
pk19mobsacB-PAmCherry-Cg0838	Integration vector, *ori*pUC, Km^r, mob sac, *PAmCherry-cg0838 (dipD)*	This study
pk19mobsacB-mNeonGreen-Cg0838	Integration vector, *ori*pUC, Km^r, mob sac, *mNeonGreen-cg0838 (dipD)*	This study
pk19mobsacB-FtsZ-StrepTagII	Integration vector, *ori*pUC, Km^r, mob sac, *ftsZ-StrepTagII*	This study
pk19mobsacB-DivIVA-mCherry	Integration vector, *ori*pUC, Km^r, mob sac, *divIVA-mCherry*	(Donovan 2012)
pk19mobsacB-ΔmurJ	Integration vector, *ori*pUC, Km^r, mob sac, deletion of *murJ*	This study
pk19mobsacB-MurJ-mNeonGreen	Integration vector, *ori*pUC, Km^r, mob sac, *murj-mNeonGreen*	This study
pk19mobsacB-Cg0838-mNeonGreen	Integration vector, *ori*pUC, Km^r, mob sac, *cg0838 (dipD)-mNeonGreen*	This study
pk19mobsacB-Cg0841-mCherry	Integration vector, *ori*pUC, Km^r, mob sac, *cg0841 (dipA)-mCherry*	This study
pk19mobsacB-Cg0841-mNeonGreen	Integration vector, *ori*pUC, Km^r, mob sac, *cg0841 (dipA)-mNeonGreen*	This study
pk19mobsacB-mCherry-Cg0841	Integration vector, *ori*pUC, Km^r, mob sac, *mCherry-cg0841 (dipA)*	This study
pk19mobsacB-ΔlexA (in ΔdivS)	Integration vector, *ori*pUC, Km^r, mob sac, deletion of *lexA* in a *ΔdivS* genomic background	This study
pEKEx2	*E. coli-C. glutamicum* shuttle expression vector, P_{tac}, $lacI_q$, Km^r, pBL1 $oriV_{C.g.}$, pUC18 $oriV_{E.c.}$	(Eikmanns et al. 1991)
pEKEx2-mCherry	pEKEx2, *mCherry*	This study

pEKEx2-GFP	pEKEx2, *GFP*	This study
pEKEx2-mNeonGreen	pEKEx2, *mNeonGreen*	This study
pEKEx2-PS-CFP2	pEKEx2, *PS-CFP2*	This study
pEKEx2-PSmOrange2	pEKEx2, *PSmOrange2*	This study
pEKEx2-murJ-TAA	peKEx2, *murJ* (TAA)	This study
pEKEx2-murJ-6xHisTag	peKEx2, *murJ*-6xHisTag	This study
pEKEx2-RBS-MurJ-mNeonGreen	peKEx2, RBS_{murJ}, *murJ, mNeonGreen*	This study
pEKEx2-RBS-6xHisTag-GFP-Thr-DivIVA	peKEx2, RBS_{divIVA}, 6xHisTag-*GFP*-thrombin cleavage site-*divIVA*	This study
pEKEx2-RBS-DivIVA-Thr-GFP-6xHisTag	peKEx2, RBS_{divIVA}, *divIVA*-thrombin cleavage site-*GPF*-6xHisTag	This study
pEKEx2-RBS-DivIVA-mNeonGreen	peKEx2, RBS_{divIVA}, *divIVA-mNeonGreen*	This study
pEKEx2-RBS-DivIVA-Dendra2	peKEx2, RBS_{divIVA}, *divIVA-Dendra2*	This study
pEKEx2-RBS-DivIVA-GFP	peKEx2, RBS_{divIVA}, *divIVA-GFP*	This study
pEKEx2-RBS-DivIVA-PS-CFP2	peKEx2, RBS_{divIVA}, *divIVA-PS-CFP2*	This study
pEKEx2-RBS-DivIVA-PSmOrange2	peKEx2, RBS_{divIVA}, *divIVA-PSmOrange2*	This study
pEKEx2-Cg0847-Dendra2	peKEx2, *cg0847 (lcpA)-dendra2*	This study
pEKEx2-Cg0842-Dendra2	peKEx2, *cg0842-dendra2*	This study
pEKEx2-Cg0841-Dendra2	peKEx2, *cg0841 (dipA)-dendra2*	This study
pEKEx3	*E. coli-C. glutamicum* shuttle expression vector, P_{tac}, $lacI_q$, $Spec^r$, pBL1 $oriV_{C.g.}$, pUC18 $oriV_{E.c.}$	(Hoffelder et al. 2010)
pEKEx3-mCherry	pEKEx3, *mCherry*	This study
pEKEx3-RBS-FtsZ-linker-mCherry	pEKEx3, RBS_{ftsZ}, *ftsZ*-GGGGSGGGG-*mCherry*	This study
pCLTON1PamtR	Modified pCLTON1 expression vector	Gerd Seibold
pCLTON1PamtR-Dendra2	pCLTON1PamtR, *Dendra2*	Gerd Seibold
pCLTON1PamtR-sce(a)	pCLTON1PamtR, *sce(a)*	This study
pCLTON1PamtR-FtsZ-SW-PAmCherry	pCLTON1PamtR, *ftsZ-SW-PAmCherry* (sandwich)	This study
pAC4-AviTag™	Expression vector, AviTag™-MCS, *Ptrc* promoter, rrnB T1 terminator, rrnB T2 terminator, beta lactamase, $lacI_q$, Amp^r	Avidity
paC4-HisTag-mNeonGreen-AviTag™	paC4 AviTag™, 6x*HisTag-mNeonGreen*	This study
pAN4-AviTag™	Expression vector, MCS-AviTag™, *Ptrc* promoter, rrnB T1 terminator, rrnB T2 terminator, beta lactamase, $lacI_q$, Amp^r	Avidity
pET-16b	*E. coli* protein expression vector, $p_{T7\text{-}lac}$, Amp^R, N-*10xHisTag*, pBR322	Novagen

pET-16b-HisTag-mNeonGreen	pET-16b, *10xHisTag-mNeonGreen*	This study
pET-16b-HisTag-DivIVA	pET-16b, *10xHisTag-divIVA*	This study
pETDuet-1	Beta lactamase, p_{T7lac}, lacI, Amp^R	Novagen
pETDuet-1-ParB-eCFP	pETDuet-1, *parB-eCFP*	(Sieger et al. 2013)
pETDuet-1-ParB-eCFP-DivIVA-eYFP	pETDuet-1, *parB-eCFP divIVA-eYFP*	(Sieger et al. 2013)
pETDuet-1-Cg0847-eCFP-DivIVA-eYFP	pETDuet-1, *cg0847 (lcpA)-eCFP divIVA-eYFP*	This study
pETDuet-1-Cg0847-eCFP	pETDuet-1, *cg0847 (lcpA)-eCFP*	This study
Source plasmids		
Name	**Characteristics (source)**	**Reference**
pNCS-mNeonGreen	*pUC ori, SV40 ori, bla (mNeonGreen)*	Allele Biotechnology
pDendra2-N	*pUC ori, SV40 ori, PCMVIE, aph3 (dendra2)*	Evrogen
pmCherry2	*mCherry2* source plasmid	Laboratory collection (Shen et al. 2017)
pPAmCherry	*PAmCherry* source plasmid	Laboratory collection (Subach et al. 2009)
pBAD-PSmOrange2	*pBAD (PSmOrange2)*	(Subach et al. 2012)
pUC57-Kan-PS-CFP2	*pUC57, PSCFP2*	Genewiz (Synthesized)
pUC57-Kan-FtsZ-SW-mNeon	*pUC57, ftsZ-SW-mNeonGreen (sandwich)*	Genewiz (Synthesized)
pUC57-Kan-sce(a)	*pUC57*, *sce(a)*	Genewiz (Synthesized)

9.4. Bacterial strains

Bacterial strains and cell lines utilized in this study are listed in table 9.3. The table also contains bacterial strains used in the course of my doctoral studies that are not mentioned within this dissertation.

Table 9.3: Cell lines and bacterial strain utilized in this study

Dissertation strains and cell lines		
Name	**Characteristics**	**Reference**
E. coli DH5α	F^- ϕ80*lac*ZΔM15 (lacZYA-argF)*U169 recA1 endA1 hsdR17*(r_K- m_K+) *supE44 phoA thi-1 gyrA96 relA1 λ*$^-$	Invitrogen
E. coli NEB5α	fhuA2 a(argF-lacZ)*U169 phoA glnV44 a80a(lacZ)M15 gyrA96*	NEB
E. coli BL21	F-, ompT, $hsdS_B$ (r_B-m_B-), dcm, gal, λ(DE3), Cm^r	Thermo Fisher Scientific
***C. glutamicum* RES 167**	Restriction-deficient mutant, otherwise considered wild type	(Tauch et al. 2002)
C. glutamicum* RES 167	Restriction-deficient mutant, otherwise considered wild type (Undergone mutations_- tandem amplification of tus-Locus)	(Tauch et al. 2002)

C27	RES167* derivative, *divIVA::divIVA- StrepTag*	Lab. Collection
CDC010	RES167* derivative, *divIVA::divIVA-mCherry*	(Donovan 2012)
B1G2	RES167* derivative, *dipC::dipC-mCherry*	This study
B1F10	CDC010, *dipC::dipC-mNeonGreen*	This study
B2H6	RES167 derivative, *adhA:: adhA_ IsceI_rsSW*	This study
B3B3	B2H6 derivative, *dipC::dipC-PAmCherry*	This study
B5E2	B3B3 derivative, *ΔdivS*	This study
CBK009	RES167 derivative, *parB::parB-PAmCherry*	(Bohm et al. 2020)
CBK029	RES167 derivative, *parB::parB-PAmCherry*, $parS_{2\text{-}10\ mutated}$	(Bohm et al. 2020)
CBK087	RES167 derivative, *parB::parB-PAmCherry*, $parS_{1\text{-}10\ mutated}$	(Bohm et al. 2020)
***M. gryph* WT**	*M. gryphiswaldense* wild-type	(Schultheiss and Schuler 2003)
eMT018	*M. gryph WT*, conjugated with *pMT102, KmR*	(Toro-Nahuelpan et al. 2019)
MT022	*M. gryph ΔmamY*, *conjugated with pMT102,* KmR	(Toro-Nahuelpan et al. 2019)
Tbb ´Munich´	*Trypanosoma brucei brucei AnTat 1.1 'Munich'*	(Bachmaier et al. 2020)
Tbb ´Paris´	*Trypanosoma brucei brucei AnTat 1.1E 'Paris'*	(Bachmaier et al. 2020)
CARP-CARP	*Tbb ´Paris´ carp3^{a}::carp3^{a}-PAmCherry, carp3^{b}::carp3^{b}-mNeonGreen*	(Bachmaier, S., et al., "Unpublished data")
CARP-AC1	*Tbb ´ Paris´ carp3::carp3-PAmCherry, AC1::AC1-mNeonGreen*	(Bachmaier, S., et al., "Unpublished data")
CARP-FLAM8	*Tbb ´ Paris´ carp3::carp3-PAmCherry, FLAM8::FALM8-mNeonGreen*	(Bachmaier, S., et al., "Unpublished data")
CARP-Calpain1.3	*Tbb ´ Paris´ carp3::carp3-PAmCherry, calpain1.3::calpain1.3-mNeonGreen*	(Bachmaier, S., et al., "Unpublished data")

Strains: allelic replacements

Name	**Characteristics**	**Reference**
B1C8	RES167* derivative, *divIVA::divIVA-mNeonGreen*	This study
B1E8	RES167* derivative, *divIVA::divIVA-PAmCherry*	This study
B4A4	B3B3 derivative, *divIVA-divIVA-mNeonGreen*	This study
B4B7	B2H6 derivative, *divIVA-divIVA-mNeonGreen*	This study
B1C10	RES167* derivative, *parB::parB-mNeonGreen*	This study
B1E10	RES167* derivative, *dipC::dipC-mNeonGreen*	This study
B1F2	B1E8 derivative, *dipC::dipC-mNeonGreen*	This study
B2E9	RES167 derivative, *dipC::dipC-StrepTag*	This study
B1G4	B1C8 derivative, *dipC::dipC-mCherry*	This study
B2I6	B2H6 derivative, *dipC::dipC-mNeonGreen*	This study
B1H5	RES167* derivative, *ΔdipC*	This study
B2I8	B2H6 derivative, *ΔdipC*	This study

B2I3	B2H6 derivative, *ΔdipCD*	This study
B1H7	RES167 derivative, *dipD::mCherry-dipD*	This study
B4I4	B2H6 derivative, *dipD::mNeonGreen-dipD*	This study
B5B8	B2H6 derivative, *dipD::PAmCherry-dipD*	This study
B1G10	RES167* derivative, *ftsZ::ftsZ-StrepTag*	This study
B5C1	B2H6 derivative, *ftsZ::ftsZ-StrepTag*	This study
B5D5	B2H6 derivative, *ΔdivS*	This study
B5D7	B2I8 derivative, *ΔdivS*	This study
B5D9	B2I3 derivative, *ΔdivS*	This study
B5E2	B3B3 derivative, *ΔdivS*	This study
B5E4	B5B8 derivative, *ΔdivS*	This study
	Strains: Plasmid based expression	
Name	**Characteristics**	**Reference**
ggb1b1	RES167 derivative, *pEKEx2-mNeonGreen*	This study
ggb1b3	RES167 derivative, *pEKEx2-RBS-divIVA-mNeonGreen*	This study
ggb1b4	RES167 derivative, *pEKEx2-RBS-divIVA-Thr-gfp-6His*	This study
ggb1b6	RES167 derivative, *pEKEx2-RBS-6His-gfp-Thr-divIVA*	This study
ggb1c1	RES167 derivative, *pEKEx2-RBS-murj-mNeonGreen*	This study
ggb1c5	RES167 derivative, *pEKEx2-GFP*	This study
ggb1c7	RES167 derivative, *pEKEx2-RBS-divIVA-gfp*	This study
ggb1c9	RES167 derivative, *pEKEx2-RBS-divIVA-dendra2*	This study
ggb1d5	RES167 derivative, *pEKEx2-murj-6His*	This study
ggb1d6	RES167 derivative, *pEKEx2-murj-TAA*	This study
B1J7	RES167 derivative, *pEKEx2-RBS-divIVA-PSCFP2*	This study
B2A3	RES167 derivative, *pEKEx2-RBS-divIVA-PSmOrange2*	This study
B1J9	RES167 derivative, *pEKEx2-PSCFP2*	This study
B2A5	RES167 derivative, *pEKEx2-PSmOrange2*	This study
B2A1	RES167* derivative, *pEKEx2-mCherry*	This study
B2A7	B1C8 derivate, *pEKEx2-mCherry*	This study
B2C9	RES167 derivative, *pEKEx2-dipA-Dendra2*	This study
B2D1	RES167 derivative, *pEKEx2-cg0842-Dendra2*	This study
B2D3	RES167 derivative, *pEKEx2-lcpA-Dendra2*	This study
B3G5	B2H6 derivative, *pEKEx3-mCherry*	This study
B3G7	B2J4 derivative, *pEKEx3-mCherry*	This study
B2J4	B2H6 derivative, *pCLTON1PamtR-sce(α)*	This study
B2J10	B2H6 derivative, *pCLTON1PamtR-dendra2*	This study
B3C4	B2I6 derivative, *pCLTON1PamtR-sce(α)*	This study
B3C6	B2I8 derivative, *pCLTON1PamtR-sce(α)*	This study
B3C8	B2I3 derivative, *pCLTON1PamtR-sce(α)*	This study
B3D5	B3B3 derivative, *pCLTON1PamtR-sce(α)*	This study
B4I6	B4I4 derivative, *pCLTON1PamtR-sce(α)*	This study
B5C3	B5B8 derivative, *pCLTON1PamtR-sce(α)*	This study
B5C5	B5C1 derivative, *pCLTON1PamtR-sce(α)*	This study

9.5. Construction of bacterial plasmids and strains

Oligonucleotides, plasmids and strains mentioned below are listed in table 9.1, 9.2 and 9.3 Chemically competent *E. coli* NEB5α or DH5α were used for plasmids cloning (Green and Rogers 2013).

9.5.1. Allelic replacement (pk19mobsacB based constructs)

Derivatives of pk19mobsacB were used for clean allelic replacements in *C. glutamicum* (insertions, deletions). Up- and downstream homologous flanking sequences used for recombination span for $\sim$500 bp each.

Plasmid pk19mobsacB-IsceI-recognition-site was constructed for the insertion of the I-SceI meganuclease recognition site within the *adhA* gene. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs HindIII-5´AdhA-F/SalI-IsceIrec-5´AdhA-R and SalI-3´AdhA-F/XbaI-3´AdhA-R. The IsceI meganuclease recognition site sequence was included within the SalI-IsceIrec-5´AdhA-R primer. The amplified fragments were digested with HindIII/SalI and SalI/XbaI restriction enzymes and subsequently ligated in the pk19mobsacB plasmid.

C- terminal fluorescent of the DivIVA protein were obtained via the pk19mobsacB-DivIVA-PAmCherry plasmid. Fluorophore sequence was amplified using the corresponding source plasmid as template and the primers SalI-PAmCherry-F/XbaI-STOP-PAmCherry-R. The PCR product and pk19mobsacB-DivIVA-mNeonGreen were digested with the SalI/XbaI restriction. Finally, the digested vector and *PAmCherry* were ligated.

C-terminal fluorescent and non-fluorescent fusions of the DipC protein were obtained via the pk19mobsacB-Cg0839-PAmCherry, pk19mobsacB-Cg0839-mNeonGreen, pk19mobsacB-Cg0839-mCherry and pk19mobsacB-Cg0839-Strep plasmids. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs HindIII-3´Cg0839-F/SalI-3´Cg0839-R and XbaI-Cg0839-Do-F/XmaI-Cg0839-Do-R for the flurescent fusions and HindIII-3´Cg0839-F/SalI-TAA-StrepTag-3´Cg0839-R for the non-fluorescent fusion. Fluorophores sequences were amplified using the corresponding source plasmids as templates and the primers pairs SalI-PAmCherry-F/XbaI-STOP-PAmCherry-R (used for both PAmcherry and mCherry) and SalI-mNeon-F/XbaI-STOP-mNeon-R. The non-fluorescent tag sequence (StrepTagII) was included within the SalI-TAA-StrepTag-3´Cg0839-R primer. The 3´Cg0839 fragment was digested with HindIII/SalI restriction enzymes while the downstream homologous region was digested with XbaI/XmaI restriction enzymes prior ligation to pk19mobsacB. Finally, fluorophores PCR products were digested with SalI/XbaI and ligated to the construct.

N-terminal fluorescent fusions of the DipD protein were obtained via the pk19mobsacB-mCherry-Cg0838, pk19mobsacB-PAmCherry-Cg0838 and pk19mobsacB-mNeonGreen-Cg0838 plasmids.

Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs HindIII-Cg0838-Up-F/SalI-Cg0838-Up-R and XbaI-5´Cg0838-F/BamHI-5´Cg0838-R. Fluorophores sequences were amplified using the corresponding source plasmids as templates and the primers pairs SalI-PAmCherry-F/XbaI-mCherry-R (used for both PAmcherry and mCherry) and SalI-mNeon-F/XbaI-mNeon-R. The upstream homologous region was digested with HindIII/SalI restriction enzymes while the 5´Cg0838 fragment was digested with XbaI/BamHI restriction enzymes prior ligation to pk19mobsacB. Finally, fluorophores PCR products were digested with SalI/XbaI and ligated to the construct.

C-terminal fluorescent fusion of the DipD protein was obtained via the pk19mobsacB-Cg0838-mNeonGreen plasmid. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs PstI-3´Cg0838-F/SalI-3´Cg0838-R and XbaI-Cg0838-Do-F/BamHI-Cg0838-Do-R. mNeongGreen sequence was amplified using the corresponding source plasmids as template and the primers pairs SalI-mNeon-F/XbaI-STOP-mNeon-R. The 3´Cg0838 fragment was digested with PstI/SalI restriction enzymes while the downstream homologous region was digested with XbaI/BamHI restriction enzymes prior ligation to pk19mobsacB. Finally, *mNeonGreen* PCR product was digested with SalI/XbaI and ligated to the construct.

N-terminal fluorescent fusion of the DipA protein was obtained via the pk19mobsacB-mCherry-Cg0841 plasmid. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs HindIII-Cg0841-Up-F/SalI-Cg0841-Up-R and XbaI-5´Cg0841-F/BamHI-5´Cg0841-R. mCherry sequence was amplified using the corresponding source plasmid as template and the primers pairs SalI-PAmCherry-F/XbaI-mCherry-R. The upstream homologous region was digested with HindIII/SalI restriction enzymes while the 5´Cg0841 fragment was digested with XbaI/BamHI restriction enzymes prior ligation to pk19mobsacB. Finally, *mCherry* PCR products was digested with SalI/XbaI and ligated to the construct.

C- terminal fluorescent fusions of the DipA protein were obtained via the pk19mobsacB-Cg0841-mCherry and pk19mobsacB-Cg0841-mNeonGreen plasmids. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs HindIII-3´Cg0841-F/SalI-3´Cg0841-R and XbaI-Cg0841-Do-F/XmaI-Cg0841-Do-R. Fluorophore sequences were amplified using the corresponding source plasmids as templates and the primers pairs SalI-PAmCherry-F/XbaI-STOP-PAmCherry-R (used for mCherry) and SalI-mNeon-F/XbaI-STOP-mNeon-R. The 3´Cg0841 fragment was digested with HindIII/SalI restriction enzymes while the downstream homologous region was digested with XbaI/XmaI restriction enzymes prior ligation to pk19mobsacB. Finally, the fluorophores PCR products were digested with SalI/XbaI and ligated to the construct.

C-terminal fluorescent fusion of the MurJ protein was obtained via the pk19mobsacB-MurJ-mNeonGreen plasmid. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs PstI-3´MurJ-F/SalI-3´MurK-R and XbaI-MurJ-Do-F/EcoRI-MurJ-Do-R. mNeonGreen sequence was amplified using the corresponding source plasmids as template and the primers SalI-mNeon-F/XbaI-STOP-mNeon-R. The 3´MurJ fragment was digested with PstI/SalI restriction enzymes while the downstream homologous region was digested with XbaI/EcoRI restriction enzymes prior ligation to pk19mobsacB. Finally, mNeonGreen PCR product was digested with SalI/XbaI and ligated to the construct.

StrepTagII C-terminal fusion of the FtsZ protein were obtained via the pk19mobsacB-FtsZ-StrepTagII plasmids. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs HindIII-3´FtsZ-F/SalI-StrepTag-3´FtsZ-R and SalI-FtsZ-Do-F/EcoRI-FtsZ-Do-R. The StrepTagII sequence was included within the SalI-StrepTag-3´FtsZ-R primer. The 3´FtsZ fragment was digested with HindIII/SalI restriction enzymes while the downstream homologous region was digested with SalI/EcoRI restriction enzymes prior ligation to pk19mobsacB.

Deletion of the *cg0839* gene was obtained via the pk19mobsacB-Δcg0839 plasmid. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs HindIII-Cg0839-Up-F/OH2-Cg0839-Up-R and OH1-Cg0839-Do-F/SalI-Cg0839-Do-R. The upstream and downstream PCR products were then combined via overlap PCR using the primers pairs HindIII-Cg0839-Up-F/SalI-Cg0839-Do-R. The resulting PCR product was digested with HindIII/SalI restriction enzymes prior ligation to pk19mobsacB.

Deletion of the *cg0838-cg0839* genomic region was obtained via the pk19mobsacB-Δcg0838/39 plasmid. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs XbaI-Cg0839-Up-F/OH2-Cg0839-Up-R and OH1-Cg0838-Do-F/EcoRI-Cg0838-Do-R. The upstream and downstream PCR products were then combined via overlap PCR using the primers pairs XbaI-Cg0839-Up-F/EcoRI-Cg0838-Do-R. The resulting PCR product was digested with XbaI/EcoRI restriction enzymes prior ligation to pk19mobsacB.

Deletion of the *murJ* gene was obtained via the pk19mobsacB-ΔmurJ plasmid. Homologous regions were amplified from *C. glutamicum* RES167 genomic DNA using the primers pairs XbaI-MurJ-Up-F/OH2-MurJ-Up-R and OH1-MurJ-Do-F/EcoRI-MurJ-Do-R. The upstream and downstream PCR products were then combined via overlap PCR using the primers pairs XbaI-MurJ-Up-F/EcoRI-MurJ-Do-R. The resulting PCR product was digested with XbaI/EcoRI restriction enzymes prior ligation to pk19mobsacB.

Deletion of the *divS* gene was obtained via the pk19mobsacB-ΔdivS plasmid. Homologous regions were amplified from the *C. glutamicum* RES167 genomic DNA using the primers pairs HindIII-DivS-Do-F/SalI-DivS-Do-R and SalI-DivS-Up-F/EcoRI-DivS-Up-R. The downstream homologous region was digested with HindIII/SalI restriction enzymes while the upstream homologous region was digested with SalI/EcoRI restriction enzymes prior ligation to pk19mobsacB.

Deletion of the *lexA* gene in a *C. glutamicum ΔdivS* genomic background was obtained via the pk19mobsacB-ΔlexA plamid. Homologous regions were amplified from the *C. glutamicum* RES167 derivative adhA::adhA_ IsceI_rsSW *ΔdivS* (B5D5) genomic DNA using the primers pairs HindIII-LexA-Up-F/SphI-LexA-Up-R and SphI-LexA-Do-F/EcoRI-LexA-Do-R. The upstream homologous region was digested with HindIII/SphI restriction enzymes while the downstream homologous region was digested with SphI/EcoRI restriction enzymes prior ligation to pk19mobsacB.

9.5.2. Plasmid based expression (pEKEx2/3, pCLTON1PamtR, pET-16b, paC4….)

Derivatives of *E. coli*/*C. glutamicum* pEKEx2 and pEKEX3 shuttle vectors were used for the controlled expression of fluorophores, untagged and tagged proteins in *C. glutamicum*.

Expression of free mCherry was obtained via the pEKEx2-mCherry and the pEKEx3-mCherry plasmids. *mCherry* sequence was amplified using the corresponding source plasmid as template and the primers SalI-PAmCherry-F/EcoRI-STOP-PAmCherry-R (for the pEKEx2 construct) and SalI-PAmCherry-F/EcoRI-STOP-3'mCherry-R (for the pEKEx3 construct). The amplified PCR products were digested with SalI/EcoRI restriction enzymes and ligated to pEKEx2 and pEKEx3 respectively.

Expression of free GFP was obtained via the pEKEx2-GFP plasmid. *gfp* sequence was amplified using the corresponding source plasmid as template and the primers BamHI-5´GFP-F/EcoRI-STOP-3´GFP-R. The amplified PCR product was digested with BamHI/EcoRI restriction enzymes and ligated to pEKEx2.

Expression of free mNeonGreen was obtained via the pEKEx2-mNeonGreen plasmid. *mNeonGreen* sequence was amplified using the corresponding source plasmid as template and the primers BamHI-mNeon-F/EcoRI-STOP-mNeon-R. The amplified PCR product was digested with BamHI/EcoRI restriction enzymes and ligated to pEKEx2.

Expression of free PSmOrange2 was obtained via the pEKEX2-PSmOrange2 plasmid. *PSmOrange2* sequence was amplified using the corresponding source plasmid as template and the primers BamHI-PSmOrg.2-F/EcoRI-STOP-PSmOrg.2-R. The amplified PCR product was digested with BamHI/EcoRI restriction enzymes and ligated to pEKEx2.

Expression of free PS-CFP2 was obtained via the pEKEx2-PS-CFP2 plasmid. *PS-CFP2* sequence was extracted from pUC57-Kan-PS-CFP2 via restriction digestion (BamHI/EcoRI) and ligated to pEKEx2.

Expression of C-terminal fusions of the DivIVA protein were obtained via the pEKEx2-RBS-DivIVA-PS-CFP2, pEKEx2-RBS-DivIVA-PSmOrange2, pEKEx-RBS-DivIVA-mNeonGreen, pEKEx-RBS-DivIVA-Dendra2 and pEKEx2-RBS-DivIVA-GFP plasmids. RBS_{DivIVA}-*divIVA* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair SalI-RBS-5´DivIVA-F/BamHI-3´DivIVA-R, digested with SalI/BamHI restriction enzymes and ligated to pEKEx2. *Dendra2* was amplified from the corresponding source plasmid via the primers pair BamHI-Dendra2-F/SacI-STOP-Dendra2-R. *PS-CFP2* sequence was extracted from pUC57-Kan-PS-CFP2 via restriction digestion (BamHI/EcoRI). *PSmOrange2* and *mNeonGreen* were amplified from the corresponding source plasmids via the same primers pairs used for the respective free fluorophore expression. *Dendra2* PCR product was digested with BamHI/SacI restriction enzymes while the remaining fluorophores sequences were digested with BamHI/EcoRI restriction enzymes. All digested fragments were ligated to pEKEx2-RBS-DivIVA.

Expression of His-GFP-THR-DivIVA was obtained via the pEKEx2-RBS-6xHisTag-GFP-Thr-DivIVA plasmid. RBS_{DivIVA}-*6xHisTag-gfp* was amplified from the corresponding source plasmid using the primers pair SalI-RBS-His-5´GFP-F/BamHI-3´GFP-R. The RBS_{DivIVA}-*6xHisTag* sequence was included within the SalI-RBS-His-5´GFP-F primer. The fluorophore PCR product was digested with SalI/BamHI restriction enzymes and ligated to pEKEx2. *THR-divIVA* was amplified from *C. glutamicum* RES167 genomic DNA using primers pair BamHI-THR-5´DivIVA-F/SacI-STOP-3´DivIVA-R. The thrombin cleavage site (THR) was included within the BamHI-THR-5´DivIVA-F primer. The amplified PCR product was digested with BamHI/SacI restriction enzymes and ligated to the construct.

Expression of DivIVA-THR-GFP-His was obtained via the pEKEx2-RBS-DivIVA-Thr-GFP-6xHisTag plasmid. RBS_{DivIVA}-*divIVA* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair SalI-RBS-5´DivIVA-F/BamHI-THR-3´DivIVA-R. The RBS_{DivIVA} sequence was included within the SalI-RBS-5´DivIVA-F primer while the thrombin cleavage site (THR) was included within the BamHI-THR-3´DivIVA-R primer. The amplified PCR product was digested with SalI/BamHI restriction enzymes and ligated to pEKEx2. The *6xHisTag-gfp* sequence was amplified from the corresponding source plasmid using the primers pair BamHI-5´GFP-F/SacI-STOP-His-3´GFP-R. The *6xHisTag* was included within the SacI-STOP-His-3´GFP-R primer. The amplified fluorophore sequence was digested with BamHI/SacI restriction enzymes and ligated to the construct.

Expression of MurJ-6xHisTag was obtained via the pEKEx2-murJ-6xHisTag plasmid. *murJ-6xHisTag* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair PtsI-5´MurJ-F/BamHI-STOP-6His-3´MurJ-R. The *6xHisTag* sequence was included within the BamHI-STOP-6His-3´MurJ-R

primer. The amplified PCR product was digested with PtsI/BamHI restriction enzymes and ligated to pEKEx2.

Expression of MurJ was obtained via the pEKEx2-murJ-TAA plasmid. *murJ* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair SalI-5´MurJ-F/BamHI-STOP-3´MurJ-R. The amplified PCR product was digested with SalI/BamHI restriction enzymes and ligated to pEKEx2.

Expression of MurJ-mNeonGreen was obtained via the pEKEx2-RBS-MurJ-mNeonGreen plasmid. RBS_{MurJ}-*murJ* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair PstI-RBS-5´MurJ-F/SalI-3´MurJ-R. RBS_{MurJ} sequence was included within the PstI-RBS-5´MurJ-F primer. The amplified PCR product was digested with the PstI/SalI restriction enzymes and ligated to pEKEx2. *mNeonGreen* was amplified from the corresponding source plasmid using the primers pair SalI-mNeon-F/EcoRI-STOP-mNeon-R. The amplified PCR product was digested with the SalI/EcoRI restriction enzymes and ligated to the construct.

Expression of LcpA-Dendra2 was obtained via the pEKEx2-Cg0847-Dendra2 plasmid. *cg0847* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair SalI-Cg0847-F/BamHI-Cg0847-R. The amplified PCR product was digested with the SalI/BamHI restriction enzymes and ligated to pEKEx2. *Dendra2* was amplified from the corresponding source plasmid using the primers pair BamHI-Dendra2-F/EcoRI-STOP-Dendra2-R. The amplified PCR product was digested with the BamHI/EcoRI restriction enzymes and ligated to the construct.

Expression of Cg0842-Dendra2 was obtained via the pEKEx2-Cg0842-Dendra2 plasmid. *cg0842* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair SalI-Cg0842-F/BamHI-Cg0842-R. The amplified PCR product was digested with the SalI/BamHI restriction enzymes and ligated to pEKEx2. *Dendra2* was amplified from the corresponding source plasmid using the primers pair BamHI-Dendra2-F/EcoRI-STOP-Dendra2-R. The amplified PCR product was digested with the BamHI/EcoRI restriction enzymes and ligated to the construct.

Expression of DipA-Dendra2 was obtained via the pEKEx2-Cg0841-Dendra2 plasmid. *dipA* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair SalI-Cg0841-F/BamHI-Cg0841-R. The amplified PCR product was digested with the SalI/BamHI restriction enzymes and ligated to pEKEx2. *Dendra2* was amplified from the corresponding source plasmid using the primers pair BamHI-Dendra2-F/EcoRI-STOP-Dendra2-R. The amplified PCR product was digested with the BamHI/EcoRI restriction enzymes and ligated to the construct.

Expression of FtsZ-linker-mCherry was obtained via the pEKEx3-RBS-FtsZ-linker-mCherry plasmid. RBS_{ftsZ}-*ftsZ* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair SalI-RBS-5´FtsZ-F/SacI-3´FtsZ-R. RBS_{ftsZ} sequence was included within the SalI-RBS-5´FtsZ-F primer. The

amplified PCR product was digested with the SalI/SacI restriction enzymes and ligated to the pEKEx3. *linker-mCherry* was amplified from the corresponding source plasmid using the primers pair SacI-linker-mCherry-F/EcoRI-STOP-3´mCherry-R. The linker sequence was included within the SacI-linker-mCherry-F primer. The amplified PCR product was digested with the SacI/EcoRI restriction enzymes and ligated to the construct.

Expression of I-SceI was obtained via the pCLTON1PamtR-sce(a) plasmid. *sce(a)* sequence was extracted from pUC57-Kan-sce(a) via restriction digestion (KpnI/SacI) and ligated to pCLTON1PamtR.

Expression of FtsZ-SW-PAmCherry was obtained via the pCLTON1PamtR-FtsZ-SW-PAmCherry plasmid. *ftsZ-SW-mNeonGreen* sequence was extracted from pUC57-Kan-FtsZ-SW-mNeon via restriction digestion and ligated to pCLTON1PamtR (KpnI/SacI) and ligated to pCLTON1PamtR. For the substitution of the *mNeonGreen* sequence, *PAmCherry* sequence was amplified with the primers pairs SacII-linker-5´PAmCherry-F/NheI-linker-3´PAmCherry-R. Both pCLTION1PamtR-FtsZ-SW-mNeonGreen and the PCR product were digested with the SacII/NheI restriction enzymes and ligated.

Expression of HisTag-mNeonGreen was obtained via the pET-16b-HisTag-mNeonGreen plasmid. *mNeonGreen* was amplified from the corresponding source plasmid using the primers pair NdeI-5´mNeon-F/BamHI-STOP-3´mNeon-R. *10xHisTag* sequence was included within the pET-16b vector. The amplified PCR product was digested with NdeI/BamHI restriction enzymes and ligated to pET-16b.

Expression of HisTag-mNeonGreen-AviTag™ was obtained via the paC4-HisTag-mNeonGreen-AviTag™. *HisTag-mNeonGreen* sequence was amplified from pET-16b-HisTag-mNeonGreen using the primers pair XhoI-6His-F/HindIII-GA-3´Neon-R. AviTag™ sequence was included within the paC4 vector. The amplified PCR product was digested with XhoI/HindIII restriction enzymes and ligated to paC4.

Expression of HisTag-DivIVA was obtained via the pET-16b-HisTag-DivIVA plasmid. *divIVA* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair NdeI-5´DivIVA-F/BamHI-STOP-3´DivIVA-R. *10xHisTag* sequence was included within the pET-16b vector. The amplified PCR product was digested with NdeI/BamHI restriction enzymes and ligated to pET-16b.

Expression of LcpA-eCFP was obtained via the pETDuet-1-Cg0847-eCFP. *lcpA* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair BamHI-T-5´Cg0847/SalI-3´Cg0847. The amplified PCR product was digested with BamHI/SalI restriction enzymes. pETDuet-1-RodA-eCFP was also digested with BamHI/SalI in order to remove *rodA* sequence. Finally, digested vector and PCR product were ligated.

Expression of LcpA-eCFP was obtained via the pETDuet-1-Cg0847-eCFP-DivIVA-eYFP. *lcpA* was amplified from *C. glutamicum* RES167 genomic DNA using the primers pair BamHI-T-5´Cg0847/SalI-3´Cg0847. The amplified PCR product was digested with BamHI/SalI restriction enzymes. pETDuet-1-Cg0847-eCFP-DivIVA-eYFP was also digested with BamHI/SalI in order to remove *rodA* sequence. Finally, digested vector and PCR product were ligated.

9.5.3. Strains construction

pk19mobsacB based constructs were transformed into *C. glutamicum* competent cells via electroporation (Schäfer et al. 1994). The first crossover event necessary for allelic replacement via the pk19mobsacB plasmid was selected for via growth in presence of kanamycin. The second crossover event was selected for via growth in presence of 10% sucrose.

Insertion of the I-SceI meganuclease recognition site within the adhA gene was confirmed via colony PCR with the primers pair Isce_ins_Fw_seq/Isce_ins_Rv_seq followed by digestion of the PCR product via SalI. Deletions of *dipC*, *dipC-D* and *divS* were screened via colony PCR with the respective primers pairs cg0839up_F_seq/cg0839do_R_seq, cg0839up_F/cg0838do_R_seq and divSup_F_seq/ divSdo_R_seq. C-terminal fluorescent fusions of *dipC* were screened via colony PCR with a combination of cg0839do_R_seq and the respective 5´ forward fluorophore primer (SalI-PAmCherry-F and SalI-mNeon-F). C-terminal fluorescent fusion of *divIVA* was screened via colony PCR with the primers pair SalI-RBS-5´DivIVA-F/XbaI-STOP-PAmCherry-R. N-terminal fluorescent fusions of *dipD* were screened via colony PCR with a combination of cg0838do_R_seq and the respective 5´ forward fluorophore primer (SalI-PAmCherry-F and SalI-mNeon-F). C- terminal fusion of StrepTagII to *ftsZ* and *dipC* were screened via colony PCR with a combination of StrepTag_seq_F and the respective downstream reverse primer (FtsZdo_R_seq and cg0839do_R_seq).

pEKEx2, pEKEx3 and pCLTON1PamtR based constructs were transformed into *C. glutamicum* competent cells via electroporation (Schäfer et al. 1994). Once transformed, the plasmids were maintained by growing the strains in presence of the respective antibiotic (kanamycin for pEKEx2 and pCLTON1PamtR, chloramphenicol for pEKEx3).

pETDuet-1, pET-16b and pAC4 based constructs were transformed into chemically competent *E. coli* BL21 cells (Green and Rogers 2013). Once transformed, the plasmids were marinated by growing the strains in presence of the respective antibiotic (Carbenicillin for pETDuet-1, pET-16b and pAC4).

9.6. Culture conditions

E. coli NEB5α, DH5α and BL21 cells were grown at 37 °C in Lysogeny Broth (LB-Miller) medium at 200 rpm supplemented, when needed, with antibiotic ($50\,\mu g/mL$ kanamycin or $100\,\mu g/mL$ carbenicillin). Protein expression based on pETDuet-1 and pET-16b vectors was induced using variable

concentrations (0.1 to 1 mM) of isopropyl β-D-1-thiogalactopyranoside (IPTG) at an OD_{600} between one and two.

C. glutamicum cells were grown at 30°C and 200 rpm in Brain Hearth Infusion complex medium (BHI, Oxoid™) (chapter 3) or minimal salt medium CGXII (Keilhauer et al. 1993) supplemented with $120\ mM$ acetate (chapter 5) (Bohm et al. 2020). Cells were always preinoculated in BHI overnight. Cells grown in minimal medium were additionally re-diluted in CGXII supplemented with acetate and grown overnight following the preinoculation step. Cell cultures were adjusted to an OD_{600} of 0.5 for BHI based experiments and to an OD_{600} of 1 for growth in minimal medium. DNA damage was induced either by supplementing the medium with mitomycin C (final concentration: $200\ ng/mL$) or inducing the expression of the I-SceI meganuclease by supplementing the medium with anhydrotetracycline ($125\ ng/mL$) in cultures at an OD_{600} of one.

M. gryphiswaldense cells were grown at 30°C and 120 rpm under microoxic conditions in 2% oxygen aerated modified flask standard medium (FSM) (Chapter 4) (Heyen and Schuler 2003) containing $50\ \mu M$ ferric citrate. Medium was supplemented with anhydrotetracycline (final concentration: $50\ ng/mL$) for Dendra2-MamY induction. Growth of the strains was performed by Dr. Dr. Toro-Nahueln.

Pleomorphic *Trypanosoma brucei brucei* AnTat 1.1 'Munich' bloodstream forms (Bachmaier et al. 2020) were grown at 37°C and 5% CO_2 in modified HMI-9 medium (Vassella et al. 1997) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS) and 1.1% methylcellulose. Differentiation to the procyclic stage was initiated by density-dependent transformation of long slender bloodstream forms to growth-arrested short stumpy bloodstream forms (culture with starting density of $5 \times 10^5\ cells/mL$ was grown for 36 hours without dilution). Short stumpy forms were transferred into modified DTM medium (Vassella and Boshart 1996) complemented with 15% (v/v) heat-inactivated FBS at $2 \times 10^6\ cells/mL$, followed by addition of $6\ mM$ cis-aconitate and cultivation at 27°C. Procyclic forms of AnTat 1.1 'Munich' were grown at 27°C in SDM-79 medium (Brun and Schonenberger 1979) supplemented with 10% (v/v) heat-inactivated FBS and $20\ mM$ glycerol. Growth of the cell lines was performed by Dr. Bachmaier.

9.7. Sample fixation

C. glutamicum cells were fixed in 3% formaldehyde solution (36.5–38% in H_2O + 10–15% methanol, Sigma Aldrich) for 30 minutes at 30°C and 200 rpm. Cells were sedimented at 3200 x g for 3 minutes and resuspended in PBS supplemented with glycine (final concentration $10\ mM$) in order to quench the excess formaldehyde. Following the resuspension, the samples were incubated in the quenching

solution for 5 minutes. The washing was repeated three times. Finally, cells were diluted in buffer containing 50 mM Tris pH 7.4, 50 mM NaCl, 10 mM EDTA, and 0.5 M sucrose (TSEMS).

M. gryphiswaldense cells were fixed in 1% formaldehyde solution (36.5–38% in H_2O + 10–15% methanol, Sigma Aldrich) for 30 min at room temperature. Next, cells were spun down at 3000 x g for 10 min and the cell pellet was gently resuspended in 10 mL of HEPES buffer (10 mM, pH 7) and 10 mM glycine, incubated for 10 min and spun at 3000 x g for 10 min. Then, the cell pellet was washed with 10 mL of HEPES (10 mM, pH 7) and again spun at 3000 x g for 10 min. Finally, cells were resuspended in 300 µL of HEPES (10 mM, pH 7).

1.6×10^7 trypanosomes were fixed in 2% (w/v) paraformaldehyde (PFA) solution in water for 20 min at room temperature, washed 2-4 times with glycine supplemented PBS (1400 x g, 10 min, final concentration: 10 mM glycine) and finally resuspended in 50 µL PBS.

9.8. Sample preparation

Each well from an 8 well glass bottom µ-slide (Ibidi, Martinsried, Germany) was incubated with 200 µL of 0.1 % (w/v) poly-L-lysine in water for at least one hour at RT. Following the incubation, poly-L-Lysine solution was removed and each well was washed three times with 200 µL sterile filtered TMSEM or PBS. 5 µL fixed cells were mixed with 1.5 µL (1:5000) 100 nm diameter fluorescent TetraSpeck Microbeads (Thermo Fisher Scientific) and 200 µL buffer (TMSEM for *C. glutamicum* cells and PBS for trypanosomes cells) and loaded into an 8 well glass bottom µ-slide. Finally, the 8 well glass bottom µ-slides containing the fixed samples were centrifuged (3200 x g for *C. glutamicum* cells and 1400 x g for trypanosomes) for 10 min at 15°C in order to sediment the cells.

Agarose pads were prepared by using 65 µL gene frames (Thermo Scientific™) and 1.5 % low melting agarose solution (Biozym). 5 µL fixed cells (*M. gryphiswaldense*) were mixed with 1.5 µL (1:5000) 100 nm diameter fluorescent TetraSpeck Microbeads (Thermo Fisher Scientific). 2 µL of the mix was loaded on the agarose pad and covered with 18 x 18 mm² high-precision coverslips of 170 ± 0.005 µm thickness (Zeiss). Glass slides and coverslips were plasma-cleaned for 5 min and kept sealed until use.

9.9. Microscopy

9.9.1. Fluorescence microscopy

Images were acquired on an Axio-Imager M1 fluorescence microscope (Carl Zeiss) with an EC Plan Neofluar 100x/1.3 oil Ph3 objective and a Hamamatsu Orca-R2 camera. mCherry and PAmCherry tagged proteins were detected using the filter set 43 HECy 3 shift free (EX BP 550/25, BS FT 570, EM BP 605/70). mNeonGreen tagged proteins were detected using a BP 495-550 / LP 750 emission filter set.

9.9.2. PALM

PALM experiments were performed with an ELYRA P.1 (Zeiss) microscopy system. The microscope was equipped with a 405 nm Diode-Laser (50 mW), a 488 nm laser (200 mW), a 561 nm laser (200 mW) and a 640 nm laser (150 mW). The emission filter sets were the following: a 77 HE GFP+mRFP+Alexa 633 shift free (EX TBP 483+564+642, BS TFT 506+582+659, EM TBP 526+601+688), a 49 DAPI shift free (EX G 365, BS FT 395, EM BP 445/50), a BP 420-480 / LP 750, a BP 495-550 / LP 750, a LP 570 and a LP 655 filter set. All the imaging was performed with an alpha Plan-Apochromat 100x/1,46 Oil DIC M27 objective (Zeiss), in combination with a 1.6x Optovar. Images were recorded with an Andor EM-CCD camera iXon DU 897.

Imaging lasers were set in integration mode (active only while the camera is collecting photons) while the activation laser was left on for the all duration of the imaging.

PAmCherry and Dendra2 fusions were recorded at 50 ms exposure time and 200 EMCCD gain with a LP 570 emission filter set. The imaging laser (561 nm laser) was kept at constant power (20%) while the activation laser (405 nm laser) was linearly increased over the course of the experiment (increase rate dependent on the fusion). Experiment length (Number of frames) and maximum activation power were determined in order to activate and image all fluorophores by the end of experiment.

mNeonGreen fusions were recorded at 50 ms exposure time and 200 EMCCD gain with a BP 495-550 / LP 750 filter set. The imaging laser (488 nm laser) was kept at constant power (20%) while no activation laser was used. Samples were imaged until complete bleaching was achieved.

Zeiss ZEN 2.1 SP3 Black software was used for the determination of localizations coordinates. In details, a 2D x/y Gaussian fit was applied with a mask size of 9 pixels (1 pixel: 100 x 100 nm) and a signal to noise ratio of 6. 100 nm TetraSpeck microspheres were used as fiducials for lateral drift correction via the Zeiss ZEN 2.1 SP3 Black software. Parameters based data filtering and grouping were performed in a fluorophore specific manner as described in chapter 2.

The input localizations table for the script described within this study has been extracted via the Zeiss ZEN 2.1 SP3 Black software. It follows that, usage of the script via a different microscope requires adaptation of the table format. Input examples can be found on the associated GitHub page (https://github.com/GiacomoGiacomelli) (Giacomelli 2020).

9.10. Software

PALM imaging, PSF fitting, lateral shift correction, and spatio-temporal grouping were performed via the ZEN 2.1 SP3 (black) (64bit) software (Version 14.0.4.201).

Bright field based definition of all regions of interest (ROIs) and data extraction for conventional-like image analysis was performed via Fiji (Schindelin et al. 2012). Conversion of PALM localizations tables to Normalized Gaussian distributions or Cumulative Histograms was performed via ThunderSTORM (Ovesny et al. 2014). Fiji macros (Fiji_Macro_1, Fiji_Macro_2, Fiji_Macro_3, Fiji_Macro_4) were written in ImageJ Macro language (IJM).

R scripts (Script_R1, Script_R2, Script_R3, Script_R4, Script_R5P1, Script_R5P2, Script_R5P3, and Script_Rv) were written in RStudio (version 1.1.456) (RStudio_Team 2016) and run on R version 3.6.3 (2020-02-29) (R_Core_Team 2020). The following packages were used in the R scripts: gplots (Warnes et al. 2020), ggplot2 (Wickham 2016), plotrix (Lemon 2006), fields (Nychka et al. 2017), spatstat (Baddeley et al. 2015), pgirmess (Giraudoux 2018), RColorBrewer (Neuwirth 2014), dbscan (Hahsler et al. 2019), patchwork (Pederson 2019), mixtools (Benaglia et al. 2009), MASS (Venables and Ripley 2002) and agricolae (de Mendiburu 2020).

9.11. Molecular biology methods

9.11.1. Plasmids isolation from *E. coli*

Plasmid extraction from E. coli were performed via NucleoSpin® Plasmid Kit (Macherey-Nagel) according to the protocol provided by the manufacturer.

9.11.2. DNA amplification and colony PCRs

DNA fragments were amplified using Phusion® high fidelity polymerase (New England Biolabs) or ReproFast Polymerase (Genaxxon Bioscience) according to the protocol provided by the respective manufacturer. *C. glutamicum* RES167 genomic DNA or source plasmids (Table 8.2) DNA were used as PCR templates. Primers used are listed in table 8.1.

Successful allelic replacement in *C. glutamicum* and plasmid construction in *E. coli* were checked via colony PCR. Colony PCRs were performed via EconoTaq PLUS GREEN 2X Master Mix according to the protocol provided by the manufacturer with minor differences between the two bacteria. *C. glutamicum* cells were incubated at 95°C for 30 min while E. coli cells were incubated for 15 min. Primers used are listed in table 8.1.

9.11.3. DNA separation and purification

PCR products, colony PCR products and nucleic acids digested via restriction enzymes were run in 1 % agarose gels in TAE buffer ($40\ mM$ Tris $-$ Hcl pH $8,1\ mM\ EDTA, 20\ mM\ acetic\ acid$) at 90 V, stained within a $0.5\ \mu g/mL$ EtBr bath and visualized using a UV table. When necessary, nucleic acids were extracted and purified using a NucleoSpin® Gel Kit and PCR Clean-up Kit (Macherey-Nagel).

9.11.4. DNA quantification and sequencing

The concentration of PCR products and plasmids was determined using a UV/VIS spectrophotometer (BioDrop µLite, Serva). Plasmids sequencing was performed by the in-house sequencing service (Genomics service unit, Genetics, Faculty of Biology, LMU – Munich). Samples comprised 150-300 ng plamid DNA and 3 pmol sequencing primer in 5 mM TrisHCl (7 μL final volume).

9.11.5. Restriction digestion, dephosphorylation and ligation

All enzymatic reactions listed within this chapter were performed according to manufacturer´s protocols (New England Biolabs). PCR products and plasmids were digested with the respective restriction enyzmes (see chapter 8.5.3). Following plasmid digestion, re-ligation of the DNA ends was prevented via addition of alkaline phosphatase. Constructs were ligated overnight at 4°C with T4 DNA ligase.

9.12. Protein biochemical methods

9.12.1. *C. glutamicum* lysates preparation

X mL of cells (culture state depending on the aim of the experiment) were harvested and diluted to an OD_{600} of 1 to a final volume of 10 mL (PBS pH 7.4) and put on ice. Cells were washed in PBS (3200 x g, 10 min, 4°C) and finally resuspended in 1 mL PBS pH 7.4. Following ten cycles of cell disruption via FastPrep®-24 (MP Biomedicals) at 6.5 m/sec for 30 sec, cell debris was removed by centrifugation at 18000 g (15 min).

9.12.2. Polyacrylamide gel electrophoresis

Proteins contained within the cell lysates were separated by size via polyacrylamide gel electrophoresis (Laemmli 1970). Resolving (Tris-HCl pH 8.8, 10 % (v/v) acrylamide, 0.1 % (w/v) SDS, 0.05 % (w/v) ammonium persulfate, 0.05 % (w/v) tetramethylethylenediamine) and stacking gel (Tris-HCl pH 6.8, 4 % (v/v) acrylamide, 0.1 % (w/v) SDS, 0.05 % (w/v) ammonium persulfate, 0.13 % (w/v) tetranethylethylenediamine) were cast using a Mini-PROTEAN®Systems (Bio-Rad). Cell lysates were mixed with 4 x loading buffer (200 mM Tris-HCl pH 6.8, 50 % (w/v) glycerol, 10 % (w/v) SDS, 4 % β-mercaptoethanol, 0.08 % (w/v) bromophenol blue) to a final 1 x loading buffer concentration and run in electrophoresis chambers (Mini-PROTEAN®, Bio-Rad). The chambers were filled with running buffer (25 mM Tris, 0.192 M glycine, 3.5 mM SDS) and run at 90 V until the samples reached the interface between stacking and resolving gel. Voltage was then increased to 120 V and kept running for 40-60 min. Protein size was ascertained via the use of a pre-stained protein ladder (PageRuler™, Therm Fisher Scientific).

9.12.3. Colloidal Coomassie staining

The colloidal Coomassie staining solution was prepared according to Dyballa and Metzger (Dyballa and Metzger 2009). Briefly, 100 g of aluminum sulfate were dissolved in Milli-Q water, followed by the addition of 200 ml of ethanol, 0.4 of CBB G-250 and 47 ml of 85% orthophosphoric acid (the solution needs to be homogeneous before the addition of each subsequent ingredient) (the addition order needs to be kept as is). Finally, Milli-Q water is added to a final volume of 2 liters. SDS-gels were incubated with staining solution over-night, washed twice in Milli-Q water, destained up to 60 minutes in order to increase the contrast (destaining solution: 10% ethanol, 2% orthophosphoric, 88% Milli-Q water) and washed twice in Milli-Q water.

9.12.4. Western Blot

Successful allelic replacement was confirmed, when possible, via the use of Western Blots. Proteins separated via SDS-gels were transferred to methanol activated PVDP membranes. Electrophoretic chambers (Bio-Rad) were filled with transfer buffer (25 mM Tris-HCl pH 8.3, 0.2 M glycine, 20 % (v/v) methanol) and run at 200 mA for 3 h. Following the transfer, the PVDP membranes were incubated for one hour in 5 % skimmed milk powder in TBS-T (50 mM Tris-HCl pH 7.5, 150 mM NaCl, 0.1 % Tween 20) in order to block non-specific binding sites. PVDP membranes were then incubated with dilutions of the respective primary antibodies (1:2000 polyclonal rabbit anti-mCherry IgG (BioVision Inc.) for mCherry and PAmCherry fusions, 1:1000 monoclonal mouse anti-StrepTag IgG (Thermo Fisher Scientific) for StrepTagII fusions) in 5 % skimmed milk powder in TBS-T for one hour. Secondary anti-rabbit (ploYclonal goat IgG, alkaline phosphatase conjugate, Sigma-Aldrich®) and anti-mouse antibodies (polyclonal goat IgG, alkaline phosphatase conjugate, Sigma-Aldrich®) were diluted in 5 % skimmed milk powder in TBS-T for 1 h (final antibody concentration 1:10000). The membranes were then incubated for one hour in the secondary antibody dilution. The membranes were finally washed in TBS-T and developed. Detection of antibodies was performed via incubation of the membranes in the dark with chromogenic alkaline phosphate substrates 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium.

9.12.5. Protein identification via pull-down and mass spectrometry

For the identification of putative novel DivIVA interaction partners, strain C27 (*C. glutamicum* RES167* *divIVA::divIVA-StrepTag*) and a WT control (*C. glutamicum* RES167*) were cultivated in 50 ml of BHI medium using culture flasks pretreated with 0.5% sodium hypochlorite. Cells were harvested at an OD_{600} of 2 (exponential phase) and lysed as described in 8.12.1 (Lysis buffer: 50 mM TRIS/HCl pH 8.0, 200 mM NaCl, 100 mM KCl, 1 mM EDTA, 0.1% SDS, EDTA free protease inhibitor, DNAse). MagStrep “type3” XT beads (IBA) were used to specifically bind DivIVA-StrepTag in the course of the pull down. 40 µl (approximately 2 mg) of magnetic beads were used for each sample.

Beads were washed twice in 400 µl of lysis buffer at room temperature prior the addition of the respective lysate sample. The beads were then moved to 4°C and resuspended in 200 µl of a 1:1 mixture of lysis buffer and cleared lysate. The mixture was incubated 30 minutes on ice with occasional resuspension of the beads via gentle shaking. Following the incubation, the beads were washed three times in lysis buffer. Finally, the beads were incubated with elution buffer (Elution buffer: 50 mM TRIS/HCl pH 8.0, 200 mM NaCl, 100 mM KCl, 1 mM EDTA, 0.1% SDS, EDTA free protease inhibitor, 2 mM D-biotin) for 5 minutes with occasional shaking. The elution step was repeated twice.

For both C27 and the WT control, 10 µl of the cell lysate, clear lysate, washing step 1-3, elution step 1-2 and boiled beads were loaded in a 10% polyacrylamide gel and separated via gel electrophoresis. The gels were then stained via colloidal Coomassie as described in 8.12.3. Then, the elution profiles of C27 and WT were compared. The protein bands visible exclusively within the C27 elution profile and the correspondent area within the WT elution profile were physically removed from the gels and analysed via mass spectrometry.

9.12.6. Mass spectrometry

Mass spectrometry, including sample/protein digestion, was performed by the Mass Spectrometry of Biomolecules at LMU (MSBioLMU) service unit. The service unit used a Linear Trap Quadropole (LTQ) Orbitrap mass spectrometer (Thermo) combined with a nano liquid chromatography (LC) system including an autosampler (automated injection). RP C18 capillary columns were used for the LC.

9.13. Statistical analysis

All statistical tests performed outside and within the various scripts (Script_R1, 2, 3, 4, v, 5P1, 5P2, 5P3) were performed using R (R_Core_Team 2020).

Chapter 10: References

Aboulwafa, M. and M. H. Saier, Jr. (2011). "Biophysical studies of the membrane-embedded and cytoplasmic forms of the glucose-specific Enzyme II of the *E. coli* phosphotransferase system (PTS)." PLoS One 6(9): e24088.

Adam, V., et al. (2011). "Rational design of photoconvertible and biphotochromic fluorescent proteins for advanced microscopy applications." Chem. Biol. 18(10): 1241-1251.

Andresen, M., et al. (2008). "Photoswitchable fluorescent proteins enable monochromatic multilabel imaging and dual color fluorescence nanoscopy." Nat. Biotechnol. 26(9): 1035-1040.

Ankerst, M., et al. (1999). "OPTICS: Ordering Points To Identify the Clustering Structure." SIGMOD Rec. 28(2): 49–60

Annibale, P., et al. (2011). "Quantitative photo activated localization microscopy: unraveling the effects of photoblinking." PLoS One 6(7): e22678.

Arechaga, I., et al. (2000). "Characterisation of new intracellular membranes in *Escherichia coli* accompanying large scale over-production of the b subunit of F(1)F(o) ATP synthase." FEBS Letters 482(3): 215-219.

Arganda-Carreras, I., et al. (2017). "Trainable Weka Segmentation: a machine learning tool for microscopy pixel classification." Bioinformatics 33(15): 2424-2426.

Bach, J. N., et al. (2017). "Sample Preparation and Choice of Fluorophores for Single and Dual Color Photo-Activated Localization Microscopy (PALM) with Bacterial Cells." Methods Mol. Biol. 1563: 129-141.

Bachmaier, S., et al. "Unpublished data."

Bachmaier, S., et al. (2020). "Culturing and Transfection of Pleomorphic *Trypanosoma brucei*." Methods Mol. Biol. 2116: 23-38.

Baddeley, A. and D. G. Richard (1997). "Kaplan-Meier Estimators of Distance Distributions for Spatial Point Processes." Ann. Stat. 25: 263-292.

Baddeley, A., et al. (2015). Spatial Point Patterns: Methodology and Applications with R. London, Chapman and Hall/CRC Press. ISBN 9781482210200

Baddeley, A., et al. (2013). "Hybrids of Gibbs Point Process Models and Their Implementation." J. Stat. Softw. 55(11): 1-43.

Baddeley, A. J. (2019). "Spatial sampling and censoring". Stochastic Geometry: 37-78. ISBN 9780203738276

Balzarotti, F., et al. (2017). "Nanometer resolution imaging and tracking of fluorescent molecules with minimal photon fluxes." Science 355(6325): 606-612.

Barak, L. S. and W. W. Webb (1981). "Fluorescent low density lipoprotein for observation of dynamics of individual receptor complexes on cultured human fibroblasts." J. Cell. Biol. 90(3): 595-604.

Barak, L. S. and W. W. Webb (1982). "Diffusion of low density lipoprotein-receptor complex on human fibroblasts." J. Cell. Biol. 95(3): 846-852.

Benaglia, T., et al. (2009). "mixtools: An R Package for Analyzing Finite Mixture Models." J. Stat. Softw. 32(6): 1-29.

Berg, S., et al. (2019). "ilastik: interactive machine learning for (bio)image analysis." Nat. Methods 16(12): 1226-1232.

Betzig, E., et al. (2006). "Imaging intracellular fluorescent proteins at nanometer resolution." Science 313(5793): 1642-1645.

Bohm, K., et al. (2020). "Chromosome organization by a conserved condensin-ParB system in the actinobacterium *Corynebacterium glutamicum*." Nat Commun 11(1): 1485.

Bohm, K., et al. (2017). "Novel Chromosome Organization Pattern in Actinomycetales-Overlapping Replication Cycles Combined with Diploidy." mBio 8(3): e00511-17.

Brakemann, T., et al. (2011). "A reversibly photoswitchable GFP-like protein with fluorescence excitation decoupled from switching." Nat. Biotechnol. 29(10): 942-947.

Brun, R. and Schonenberger (1979). "Cultivation and in vitro cloning or procyclic culture forms of *Trypanosoma brucei* in a semi-defined medium. Short communication." Acta Trop. 36(3): 289-292.

Chalfie, M., et al. (1994). "Green fluorescent protein as a marker for gene expression." Science 263(5148): 802-805.

Chiu, S. N., et al. (2013). Stochastic Geometry and Its Applications, Wiley. Online ISBN: 9781118658222

Chudakov, D. M., et al. (2004). "Photoswitchable cyan fluorescent protein for protein tracking." Nat. Biotechnol. 22(11): 1435-1439.

Cnossen, J., et al. (2020). "Localization microscopy at doubled precision with patterned illumination." Nat. Methods 17(1): 59-63.

Croce, A. C. and G. Bottiroli (2014). "Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis." Eur. J. Histochem. 58(4): 2461.

Daniel, R. A. and J. Errington (2003). "Control of Cell Morphogenesis in Bacteria." Cell 113(6): 767-776.

de Jong, I. G., et al. (2011). "Live Cell Imaging of *Bacillus subtilis* and *Streptococcus pneumoniae* using Automated Time-lapse Microscopy." J. Vis. Exp. (53): 3145.

de Mendiburu, F. and Yaseen, M. (2020). "agricolae: Statistical Procedures for Agricultural Research." Research.R package version 1.4.0, https://myaseen208.github.io/agricolae/.

Dempsey, W. P., et al. (2015). "In vivo single-cell labeling by confined primed conversion." Nat. Methods 12(7): 645-648.

Deschout, H., et al. (2014). "Progress in quantitative single-molecule localization microscopy." Histochem. Cell Biol. 142(1): 5-17.

Dickson, R. M., et al. (1997). "On/off blinking and switching behaviour of single molecules of green fluorescent protein." Nature 388(6640): 355-358.

Donovan, C. (2012). "Cytoskeleton proteins involved in chromosome segregation and cell division in *Corynebacterium glutamicum*." Mathematisch-Naturwissenschaftlichen Fakultät, Universität zu Köln. PhD thesis. https://kups.ub.uni-koeln.de/4730/1/Catriona_Donovan_Thesis.pdf.

Donovan, C., et al. (2013). "Chromosome segregation impacts on cell growth and division site selection in *Corynebacterium glutamicum*." PLoS One 8(2): e55078.

Donovan, C., et al. (2012). "A synthetic *Escherichia coli* system identifies a conserved origin tethering factor in Actinobacteria." Mol. Microbiol. 84(1): 105-116.

Drews, G. and W. Niklowitz (1956). "Cytology of Cyanophycea. II. Centroplasm and granular inclusions of *Phormidium uncinatum*." Arch. Mikrobiol. 24(2): 147-162.

Ducret, A., et al. (2016). "MicrobeJ, a tool for high throughput bacterial cell detection and quantitative analysis." Nat. Microbiol. 1(7): 16077.

Dunn, K. W., et al. (2011). "A practical guide to evaluating colocalization in biological microscopy." Am. J. Physiol. Cell Physiol. 300(4): C723-742.

Durisic, N., et al. (2014). "Single-molecule evaluation of fluorescent protein photoactivation efficiency using an *in vivo* nanotemplate." Nat. Methods 11(2): 156-162.

Dyballa, N. and S. Metzger (2009). "Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels." J. Vis. Exp. (30): 1431.

Eikmanns, B. J., et al. (1991). "A family of *Corynebacterium glutamicum*/*Escherichia coli* shuttle vectors for cloning, controlled gene expression, and promoter probing." Gene 102(1): 93-98.

El-Gebali, S., et al. (2019). "The Pfam protein families database in 2019." Nucleic Acids Res. 47(D1): D427-D432.

Endesfelder, U., et al. (2011). "Chemically induced photoswitching of fluorescent probes--a general concept for super-resolution microscopy." Molecules 16(4): 3106-3118.

Ester, M., et al. (1996). "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise." KDD'96 Proceedings of the Second International Conference on Knowledge Discovery and Data Mining 96: 226-231.

Georgieva, M., et al. (2016). "Nanometer resolved single-molecule colocalization of nuclear factors by two-color super resolution microscopy imaging." Methods 105: 44-55.

Giacomelli, G. (2020). "GitHub Thesis Repository Giacomo Giacomelli.". https://github.com/GiacomoGiacomelli.

Giraudoux, P., et al.(2018). pgirmess: Spatial Analysis and Data Mining for Field Ecologists. https://CRAN.R-project.org/package=pgirmess.

Glushonkov, O., et al. (2018). "Optimized protocol for combined PALM-dSTORM imaging." Sci. Rep. 8(1): 8749.

Gorby, Y. A., et al. (1988). "Characterization of the bacterial magnetosome membrane." J. Bacteriol. 170(2): 834-841.

Graham, T. G., et al. (2014). "ParB spreading requires DNA bridging." Genes Dev. 28(11): 1228-1238.

Green, R. and E. J. Rogers (2013). "Transformation of chemically competent *E. coli*." Methods Enzymol. 529: 329-336.

Griffié, J., et al. (2018). "Dynamic Bayesian Cluster Analysis of Live-Cell Single Molecule Localization Microscopy Datasets." Small Methods 2(9): 1800008 .

Grimm, J. B., et al. (2015). "A general method to improve fluorophores for live-cell and single-molecule microscopy." Nat. Methods 12(3): 244-250, 243 p following 250.

Gross, D. and W. W. Webb (1986). "Molecular counting of low-density lipoprotein particles as individuals and small clusters on cell surfaces." Biophys. J. 49(4): 901-911.

Gross, D. and W. W. Webb (1988). Cell Surface Clustering and Mobility of the Liganded LDL Receptor Measured by Digital Video Fluorescence Microscopy. Spectroscopic Membrane Probes. ISBN: 9781351076791

Gruber, S. and J. Errington (2009). "Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in *B. subtilis*." Cell 137(4): 685-696.

Guilhas, B., et al. (2020). "ATP-Driven Separation of Liquid Phase Condensates in Bacteria." Mol. Cell 79(2): 293-303 e294.

Gunewardene, M. S., et al. (2011). "Superresolution imaging of multiple fluorescent proteins with highly overlapping emission spectra in living cells." Biophys. J. 101(6): 1522-1528.

Hahsler, M., et al. (2019). "dbscan: Fast Density-Based Clustering with R." J. Stat. Softw. 91(1): 1-30.

Hand, D. J. (2008). "Statistical Analysis and Modelling of Spatial Point Patterns by Janine Illian, Antti Penttinen, Helga Stoyan, Dietrich Stoyan." ISR 76(3): 458-458.

Hanisch, K. H. (1984). "Scattering Analysis of Point Processes and Random Measures." Math. Nachr. 117(1): 235-245.

Hanisch, K. H. (2007). "Some remarks on estimators of the distribution function of nearest neighbour distance in stationary spatial point processes." Series Statistics 15(3): 409-412.

Hell, S. W. and J. Wichmann (1994). "Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy." Opt. Lett. 19(11): 780-782.

Heyen, U. and D. Schuler (2003). "Growth and magnetosome formation by microaerophilic *Magnetospirillum* strains in an oxygen-controlled fermentor." Appl. Microbiol. Biotechnol. 61(5-6): 536-544.

Hink, M. A., et al. (2000). "Structural dynamics of green fluorescent protein alone and fused with a single chain Fv protein." J. Biol. Chem. 275(23): 17556-17560.

Hoffelder, M., et al. (2010). "The E2 domain of OdhA of *Corynebacterium glutamicum* has succinyltransferase activity dependent on lipoyl residues of the acetyltransferase AceF." J. Bacteriol. 192(19): 5203-5211.

Hooke, R. (1665). Micrographia. London, Mortyn and Allestry. https://doi.org/10.5962/bhl.title.904.

Ireton, K., et al. (1994). "spo0J is required for normal chromosome segregation as well as the initiation of sporulation in *Bacillus subtilis*." J. Bacteriol. 176(17): 5320-5329.

Jochmann, N., et al. (2009). "Genetic makeup of the *Corynebacterium glutamicum* LexA regulon deduced from comparative transcriptomics and in vitro DNA band shift assays." Microbiology 155(Pt 5): 1459-1477.

Jungmann, R., et al. (2010). "Single-molecule kinetics and super-resolution microscopy by fluorescence imaging of transient binding on DNA origami." Nano. Lett. 10(11): 4756-4761.

Kalaji, H. M., et al. (2017). "Frequently asked questions about chlorophyll fluorescence, the sequel." Photosynth. Res. 132(1): 13-66.

Kanehisa, M. and S. Goto (2000). "KEGG: kyoto encyclopedia of genes and genomes." Nucleic Acids Res. 28(1): 27-30.

Kaval, K. G. and S. Halbedel (2012). "Architecturally the same, but playing a different game: the diverse species-specific roles of DivIVA proteins." Virulence 3(4): 406-407.

Keilhauer, C., et al. (1993). "Isoleucine synthesis in *Corynebacterium glutamicum*: molecular analysis of the *ilvB-ilvN-ilvC* operon." J. Bacteriol. 175(17): 5595-5603.

Keppler, A., et al. (2003). "A general method for the covalent labeling of fusion proteins with small molecules in vivo." Nat. Biotechnol. 21(1): 86-89.

Khater, I. M., et al. (2019). "Caveolae and scaffold detection from single molecule localization microscopy data using deep learning." PLoS One 14(8): e0211659.

Khater, I. M., et al. (2020). "A Review of Super-Resolution Single-Molecule Localization Microscopy Cluster Analysis and Quantification Methods." Patterns 1(3): 100038

Kijima, N., et al. (1998). "Induction of only limited elongation instead of filamentation by inhibition of cell division in *Corynebacterium glutamicum*." Appl. Microbiol. Biotechnol. 50(2): 227-232.

Klebahn, H. (1895). "Gasvakuolen, ein Bestandteil der Zellen der Wasserblutebildenden Phycochromaceen. ." Flora (Jena) 80: 241.

Klementieva, N. V., et al. (2016). "Green-to-red primed conversion of Dendra2 using blue and red lasers." Chem. Commun. (Camb) 52(89): 13144-13146.

Koch, R. (1912). Gesammelte Werke von Robert Koch. Leipzig, Thieme. https://doi.org/10.5962/bhl.title.100138.

Laemmli, U. K. (1970). "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." Nature 227(5259): 680-685.

Lagache, T., et al. (2015). "Statistical analysis of molecule colocalization in bioimaging." Cytometry A 87(6): 568-579.

Leeuwenhoek, A. (1695/1697). “Arcana naturae detecta.” Delft, Krooneveldt H. https://doi.org/10.5962/bhl.title.62669.

Lemon, J. (2006). "Plotrix: a package in the red light district of R." R-news 6(4): 8-12.

Lenarcic, R., et al. (2009). "Localisation of DivIVA by targeting to negatively curved membranes." EMBO J 28(15): 2272-2282.

Leonard, T. A., et al. (2005). "Bacterial chromosome segregation: structure and DNA binding of the Soj dimer--a conserved biological switch." EMBO J 24(2): 270-282.

Letek, M., et al. (2008). "DivIVA is required for polar growth in the MreB-lacking rod-shaped actinomycete *Corynebacterium glutamicum*." J. Bacteriol. 190(9): 3283-3292.

Levet, F., et al. (2019). "A tessellation-based colocalization analysis approach for single-molecule localization microscopy." Nat. Commun. 10(1): 2379.

Leyton-Puig, D., et al. (2016). "PFA fixation enables artifact-free super-resolution imaging of the actin cytoskeleton and associated proteins." Biol. Open 5(7): 1001-1009.

Liu, W., et al. (2010). "Expression and cellular localisation of calpain-like proteins in *Trypanosoma brucei.*" Mol. Biochem. Parasitol. 169(1): 20-26.

Los, G. V., et al. (2008). "HaloTag: a novel protein labeling technology for cell imaging and protein analysis." ACS Chem Biol 3(6): 373-382.

Malkusch, S., et al. (2012). "Coordinate-based colocalization analysis of single-molecule localization microscopy data." Histochem. Cell Biol. 137(1): 1-10.

Manders, E. M., et al. (1992). "Dynamics of three-dimensional replication patterns during the S-phase, analysed by double labelling of DNA and confocal microscopy." J. Cell Sci. 103(3): 857-862.

Manders, E. M. M., et al. (1993). "Measurement of co-localization of objects in dual-colour confocal images." J. Microsc. 169(3): 375-382.

Marienfeld, S., et al. (1997). "Ultrastructure of the *Corynebacterium glutamicum* cell wall." Anton. Leeuw. Int. J. G. 72(4): 291-297.

Martins, G. B., et al. (2019). "Substrate-dependent cluster density dynamics of *Corynebacterium glutamicum* phosphotransferase system permeases." Mol. Microbiol. 111(5): 1335-1354.

Matern, B. (1960). Spatial Variation. Sweden, Statens Skogsforskningsinstitut. http://urn.kb.se/resolve?urn=urn:nbn:se:slu:epsilon-e-988.

Matern, B. (1986). "Spatial Variation." New York, Springer-Verlag. ISBN 9781461578925.

Messelink, J., et al. (2020). "Single-cell growth inference of *Corynebacterium glutamicum* reveals asymptotically linear growth." doi: https://doi.org/10.1101/2020.05.25.115055.

Microsoft_Corporation (2019). Microsoft Excel. https://office.microsoft.com/excel

Minnen, A., et al. (2011). "SMC is recruited to oriC by ParB and promotes chromosome segregation in *Streptococcus pneumoniae*." Mol. Microbiol. 81(3): 676-688.

Murray, H., et al. (2006). "The bacterial chromosome segregation protein Spo0J spreads along DNA from *parS* nucleation sites." Mol. Microbiol. 61(5): 1352-1361.

Nahidiazar, L., et al. (2016). "Optimizing Imaging Conditions for Demanding Multi-Color Super Resolution Localization Microscopy." PLoS One 11(7): e0158884.

Nan, X., et al. (2013). "Single-molecule superresolution imaging allows quantitative analysis of RAF multimer formation and signaling." Proc. Natl. Acad. Sci. U S A 110(46): 18519-18524.

Neuwirth, E. (2014). "RColorBrewer: ColorBrewer Palettes." https://CRAN.R-project.org/package=RColorBrewer.

Nieuwenhuizen, R. P., et al. (2013). "Measuring image resolution in optical nanoscopy." Nat. Methods 10(6): 557-562.

Nychka, D., et al. (2017). "fields: Tools for spatial data." Boulder, CO, USA, University Corporation for Atmospheric Research. https://CRAN.R-project.org/package=fields.

Nyquist, H. (1928). "Certain Topics in Telegraph Transmission Theory." Trans. AIEE 47(2): 617-644.

Ogino, H., et al. (2008). "DivS, a novel SOS-inducible cell-division suppressor in *Corynebacterium glutamicum."* Mol Microbiol 67(3): 597-608.

Oliva, M. A., et al. (2010). "Features critical for membrane binding revealed by DivIVA crystal structure." EMBO J 29(12): 1988-2001.

Ormo, M., et al. (1996). "Crystal structure of the *Aequorea victoria* green fluorescent protein." Science 273(5280): 1392-1395.

Osorio-Valeriano, M., et al. (2019). "ParB-type DNA Segregation Proteins Are CTP-Dependent Molecular Switches." Cell 179(7): 1512-1524 e1515.

Ovesny, M., et al. (2014). "ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging." Bioinformatics 30(16): 2389-2390.

Owen, D. M. and K. Gaus (2013). "Imaging lipid domains in cell membranes: the advent of super-resolution fluorescence microscopy." Front. Plant Sci. 4: 503.

Pageon, S. V., et al. (2016). "Clus-DoC: a combined cluster detection and colocalization analysis for single-molecule localization microscopy data." Mol. Biol. Cell 27(22): 3627-3636.

Paintdakhi, A., et al. (2016). "Oufti: an integrated software package for high-accuracy, high-throughput quantitative microscopy analysis." Mol. Microbiol. 99(4): 767-777.

Parise, M. T. D., et al. (2020). "CoryneRegNet 7, the reference database and analysis platform for corynebacterial gene regulatory networks." Sci Data 7(1): 142.

Patterson, G. H. and J. Lippincott-Schwartz (2002). "A photoactivatable GFP for selective photolabeling of proteins and cells." Science 297(5588): 1873-1877.

Paul, M. W., et al. (2019). "SMoLR: visualization and analysis of single-molecule localization microscopy data in R." BMC Bioinformatics 20(1): 30.

Pauling, J., et al. (2012). "CoryneRegNet 6.0--Updated database content, new analysis methods and novel features focusing on community demands." Nucleic Acids Res. 40(Database issue): D610-614.

Pearson, K. (1896). "VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia." Philos. T. R. Soc. A 187: 253-318.

Pederson, L., P. (2019). "patchwork: The Composer of Plots." https://CRAN.R-project.org/package=patchwork.

R_Core_Team (2020). “R: A language and environment for statistical computing” Vienna, Austrai, R Foundation for Statistical Computing. https://www.R-project.org/.

Ram, S., et al. (2006). "Beyond Rayleigh's criterion: a resolution measure with application to single-molecule microscopy." Proc. Natl. Acad. Sci. U S A 103(12): 4457-4462.

Raulf, A., et al. (2014). "Click chemistry facilitates direct labelling and super-resolution imaging of nucleic acids and proteins” supplementary information (ESI). See DOI: 10.1039/c4ra01027b. RSC Adv 4(57): 30462-30466.

Rayleigh (2009). "XXXI. Investigations in optics, with special reference to the spectroscope." The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 8(49): 261-274.

Ries, J., et al. (2012). "A simple, versatile method for GFP-based super-resolution microscopy via nanobodies." Nat Methods 9(6): 582-584.

Ripley, B. D. (2012). “Statistical Inference for Spatial Processes.” ISBN: 9780511624131

Rodionov, O., et al. (1999). "Silencing of genes flanking the P1 plasmid centromere." Science 283(5401): 546-549.

Rotureau, B., et al. (2014). "Flagellar adhesion in *Trypanosoma brucei* relies on interactions between different skeletal structures in the flagellum and cell body." J. Cell. Sci. 127(Pt 1): 204-215.

Royston, P. (1995). "Remark AS R94: A Remark on Algorithm AS 181: The W-test for Normality." J. R. Stat. Soc. C-Appl. 44(4): 547-551.

RStudio_Team (2016). “RStudio: Integrated Development Environment for R.” Boston, MA, RStudio, Inc. http://www.rstudio.com/.

Rueden, C. T., et al. (2017). "ImageJ2: ImageJ for the next generation of scientific image data." BMC Bioinformatics 18(1): 529.

Ruska, H., von Berries, B., Ruska, E. (1940). "Die Bedeutung der Übermikroskopie für die Virusforschung." Arch. ges Virusforsch. 1: 155-169.

Sage, D., et al. (2015). "Quantitative evaluation of software packages for single-molecule localization microscopy." Nat. Methods 12(8): 717-724.

Sage, D., et al. (2019). "Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software." Nat. Methods 16(5): 387-395.

Sapp, J. (2005). "The prokaryote-eukaryote dichotomy: meanings and mythology." Microbiol. Mol. Biol. Rev. 69(2): 292-305.

Schäfer, A., et al. (1994). "Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*." Gene 145(1): 69-73.

Schindelin, J., et al. (2012). "Fiji: an open-source platform for biological-image analysis." Nat. Methods 9(7): 676-682.

Schubert, K., et al. (2017). "The Antituberculosis Drug Ethambutol Selectively Blocks Apical Growth in CMN Group Bacteria." mBio 8(1).

Schultheiss, D. and D. Schuler (2003). "Development of a genetic system for *Magnetospirillum gryphiswaldense*." Arch. Microbiol. 179(2): 89-94.

Scrucca, L., et al. (2016). "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models." The R Journal 8(1): 289-317.

Shaner, N. C., et al. (2013). "A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum." Nat. Methods 10(5): 407-409.

Shannon, C. E. (1949). "Communication in the Presence of Noise." Proceedings of the IRE 37(1): 10-21.

Shen, Y., et al. (2017). "Engineering of mCherry variants with long Stokes shift, red-shifted fluorescence, and low cytotoxicity." PLoS One 12(2): e0171257.

Shroff, H., et al. (2008). "Live-cell photoactivated localization microscopy of nanoscale adhesion dynamics." Nat Methods 5(5): 417-423.

Siegel and Castellan (1988). “Non parametric statistics for the behavioural sciences.” New York, MacGraw Hill Int. ISBN: 9780070573574

Sieger, B. and M. Bramkamp (2014). "Interaction sites of DivIVA and RodA from *Corynebacterium glutamicum*." Front. Microbiol. 5: 738.

Sieger, B., et al. (2013). "The lipid II flippase RodA determines morphology and growth in *Corynebacterium glutamicum*." Mol. Microbiol. 90(5): 966-982.

Soh, Y. M., et al. (2019). "Self-organization of *parS* centromeres by the ParB CTP hydrolase." Science 366(6469): 1129-1133.

Spielman, S. E. (2017). "Point Pattern Analysis." International Encyclopedia of Geography: People, the Earth, Environment and Technology: 1-9. https://doi.org/10.1002/9781118786352.wbieg0849.

Stanier, R. Y. and C. B. Van Niel (1962). "The concept of a bacterium." Arch. Mikrobiol. 42: 17-35.

Stockmar, I., et al. (2018). "Optimization of sample preparation and green color imaging using the mNeonGreen fluorescent protein in bacterial cells for photoactivated localization microscopy." Sci. Rep. 8(1): 10137.

Stracy, M., et al. (2016). "Single-molecule imaging of UvrA and UvrB recruitment to DNA lesions in living *Escherichia coli*." Nat. Commun. 7: 12568.

Stracy, M. and A. N. Kapanidis (2017). "Single-molecule and super-resolution imaging of transcription in living bacteria." Methods 120: 103-114.

Strauss, S., et al. (2018). "Modified aptamers enable quantitative sub-10-nm cellular DNA-PAINT imaging." Nat. Methods 15(9): 685-688.

Subach, F. V., et al. (2009). "Photoactivatable mCherry for high-resolution two-color fluorescence microscopy." Nat. Methods 6(2): 153-159.

Subach, F. V., et al. (2010). "Bright monomeric photoactivatable red fluorescent protein for two-color super-resolution sptPALM of live cells." J. Am. Chem. Soc. 132(18): 6481-6491.

Sullivan, N. L., et al. (2009). "Recruitment of SMC by ParB-*parS* organizes the origin region and promotes efficient chromosome segregation." Cell 137(4): 697-707.

Surre, J., et al. (2018). "Strong increase in the autofluorescence of cells signals struggle for survival." Sci. Rep. 8(1): 12088.

Tauch, A., et al. (2002). "Efficient electrotransformation of *Corynebacterium diphtheriae* with a mini-replicon derived from the *Corynebacterium glutamicum* plasmid pGA1." Curr. Microbiol. 45(5): 362-367.

The_Mathworks_Inc (2020). MATLAB. Natick, Massachussetts. https://www.mathworks.com.

Thompson, R. E., et al. (2002). "Precise nanometer localization analysis for individual fluorescent probes." Biophys. J. 82(5): 2775-2783.

Toro-Nahuelpan, M., et al. (2019). "MamY is a membrane-bound protein that aligns magnetosomes and the motility axis of helical magnetotactic bacteria." Nat. Microbiol. 4(11): 1978-1989.

Turkowyd, B., et al. (2017). "A General Mechanism of Photoconversion of Green-to-Red Fluorescent Proteins Based on Blue and Infrared Light Reduces Phototoxicity in Live-Cell Single-Molecule Imaging." Angew. Chem. Int. Ed. Engl. 56(38): 11634-11639.

Uphoff, S., et al. (2013). "Single-molecule DNA repair in live bacteria." Proc. Natl. Acad. Sci. U S A 110(20): 8063-8068.

van Raaphorst, R., et al. (2020). "BactMAP: An R package for integrating, analyzing and visualizing bacterial microscopy data." Mol. Microbiol. 113(1): 297-308.

Van Valen, D. A., et al. (2016). "Deep Learning Automates the Quantitative Analysis of Individual Cells in Live-Cell Imaging Experiments." PLoS Comput. Biol. 12(11): e1005177.

Vassella, E. and M. Boshart (1996). "High molecular mass agarose matrix supports growth of bloodstream forms of pleomorphic *Trypanosoma brucei* strains in axenic culture." Mol. Biochem. Parasitol. 82(1): 91-105.

Vassella, E., et al. (1997). "Differentiation of African trypanosomes is controlled by a density sensing mechanism which signals cell cycle arrest via the cAMP pathway." J. Cell Sci. 110: 2661-2671.

Venables, W., N., and B. Ripley, D. (2002). "Modern Applied Statistics with S." New York, Springer. ISBN 9780387954578

Virant, D., et al. (2017). "Combining Primed Photoconversion and UV-Photoactivation for Aberration-Free, Live-Cell Compliant Multi-Color Single-Molecule Localization Microscopy Imaging." Int. J. Mol. Sci. 18(7): 1524.

Waagepetersen, R. P. (2007). "An estimating function approach to inference for inhomogeneous Neyman-Scott processes." Biometrics 63(1): 252-258.

Walker, G. C. (1984). "Mutagenesis and inducible responses to deoxyribonucleic acid damage in *Escherichia coli*." Microbiol. Rev. 48(1): 60-93.

Wang, S., et al. (2014). "Characterization and development of photoactivatable fluorescent proteins for single-molecule-based superresolution imaging." Proc. Natl. Acad. Sci. U S A 111(23): 8452-8457.

Warnes, G., R.,, et al. (2020). "gplots: Various R programing Tools for Plotting Data." http://CRAN.R-project.org/package=gplots.

Wickham, H. (2016). "ggplot2: Elegant Graphics for Data Analysis.", Springer-Verlag New York. https://ggplot2.tidyverse.org.

Williamson, D. J., et al. (2020). "Machine learning for cluster analysis of localization microscopy data." Nat. Commun. 11(1): 1493.

Witkin, E. M. (1976). "Ultraviolet mutagenesis and inducible DNA repair in *Escherichia coli*." Bacteriol. Rev. 40(4): 869-907.

Xu, K., et al. (2012). "Dual-objective STORM reveals three-dimensional filament organization in the actin cytoskeleton." Nat. Methods 9(2): 185-188.

Yachdav, G., et al. (2014). "PredictProtein--an open resource for online prediction of protein structural and functional features." Nucleic. Acids. Res. 42(Web Server issue): W337-343.

Zhang, H. and M. A. Schumacher (2017). "Structures of partition protein ParA with nonspecific DNA and ParB effector reveal molecular insights into principles governing Walker-box DNA segregation." Genes Dev. 31(5): 481-492.

Zhang, M., et al. (2012). "Rational design of true monomeric and bright photoactivatable fluorescent proteins." Nat. Methods 9(7): 727-729.

Zhang, P., et al. (2007). "Direct visualization of *Escherichia coli* chemotaxis receptor arrays using cryo-electron microscopy." Proc. Natl. Acad. Sci. U S A 104(10): 3777-3781.

Zinchuk, V. and O. Zinchuk (2008). "Quantitative colocalization analysis of confocal fluorescence microscopy images." Curr. Protoc. Cell Biol. Chapter 4: Unit 4 19.

List of figures:

List of tables:

Acknowledgements

I want to thank my doctoral supervisor Prof. Dr. Marc Bramkamp, who supported and guided me. We met around eight years ago and I hope that I am not the only one that is glad and benefitted from this meeting. You are great boss and a great person and you always found the time to support and guide us. I am looking forward to the next year together.

I want to thank all members of my thesis defence committee for their time and effort, especially Prof. Dr. Andreas Klingl, who not only agreed to be the second reviewer for my dissertation, but also helped me through these years, together with Prof. Dr. Kai Papenfort, as part of my TAC.

Moreover, I want to thank all the people I collaborated with in these years. You are too many to name, but you should know that I am eternally grateful for the possibility you gave me to work on so many unique and interesting projects. You provided me with ever new analytical challenges, and one could say that you are the core reason why I was able to build such comprehensive analytical tool.

A thank you also to the former (Francisca Rosa Mendez) and current (Nadine Hamze) Programme coordinators of the LSM graduate school for all the effort and work you put into giving us students the best, whether it is about courses of excursions.

Thank you also to all former and current lab members. Thank you Cat, you convinced me into joining the lab and I will never forget that, nor the great time together. Thank you Juri for Serious Sam. Thank you Boris, the Kalimotxo that I made with your expensive wine was great. Thank you Prachi for your great sense of humour. Thank you, Gustavo, for the TOTO intensifies and the 90s parties. Thank you, Helge, for the many evening of friendship and nerding. Thank you, Kati, for always presenting at the Tuesday meeting, we really miss that. Thank you Fabi for the chicken hearts soup. Thank you Lijun for the dubious looking treats you brought back from China, they tasted great, most of the time. Thank you, Abigail, for the memes and jokes and thank you and Manu for the great work in setting up the lab in Kiel. Thank you, Nadine and Karin, for keeping the lab functional whenever we are a mess. Thank you Derk for the unique ideas and crazy projects.

Thank you to our former lab neighbors, I miss you, I miss DnD nights and I miss the always available cold beer. Thank you especially to Theresa, Kristina and Astrid for giggles and joy, I hope you are visiting soon, FRIENDS!

Voglio ringraziare i miei genitori per tutto il supporto che mi hanno dato negli anni che ho passato all´estero e per quelli in cui mi avete cresciuto. Non sarei mai riuscito a raggiungere questo traguardo senza il vostro supporto. Siete fantastici e non vedo l´ora di rivedervi. Un grazie anche ai miei fratelli, siete sempre nei miei pensieri.

Finally, thank you Barbara for being here with me. You came to Kiel with me to start this new adventure and I plan to never make you regret that choice. For filling my life with happiness, love and board games, thank you.

Curriculum Vitae: