Bayesian Cluster Analysis for Single-Molecule Localization Microscopy

Inaugural-Dissertation

to obtain the academic degree

Doctor rerum naturalium (Dr. rer. nat.)

submitted to the Department of Biology, Chemistry, Pharmacy

of Freie Universität Berlin

by

Saskia Kutz

2023

This doctorate study was conducted from October 2016 to July 2023 under the supervision of Prof. Dr. Helge Ewers in the Department of Biology, Chemistry and Pharmacy of Freie Universität Berlin.

1st Reviewer: Prof. Dr. Helge Ewers, Freie Universität Berlin

2nd Reviewer: Dr. rer. nat. Stephan Block, Freie Universität Berlin

Date of defence: February 26, 2024

Abstract

This thesis aims to develop and assess a software solution for Bayesian-based cluster analysis of super-resolution single-molecule localisation microscopy data and establish a methodology for imaging cells immersed in a liquid environment. Previous studies demonstrated that cluster analysis is susceptible to user bias, and advanced techniques require coding proficiency. Furthermore, previous research demonstrated the potential of DNA tethering for investigating clustering phenomena triggered by external stimuli without requiring cell adhesion. Nevertheless, DNA tethering adoption has been limited due to the requisite expertise in advanced chemistry. As a result, this dissertation focuses on developing a user-friendly software solution. The primary objectives are facilitating Bayesian-model-based cluster analysis and simplifying DNA-tethering. The designed DNA-tethering assay facilitates cell sample examination without surface adhesion. Consequently, researchers can investigate cell behaviour under specific and localised external stimuli in live-cell and single-molecule localisation microscopy modes. The software solution presented in this research builds upon previously published work, aiming to improve performance and accessibility. Additionally, it investigates the impact and limitations of different clustering algorithms on single-molecule fluorescence microscopy data. The software was implemented and evaluated using simulated and experimental data to validate its performance. Furthermore, the software could conduct cluster analysis on data acquired from the simplified DNA-tethering assay developed. This study presents a modular software tool with a user-friendly graphical user interface designed to facilitate cluster analysis of single-molecule localisation microscopy data. The software demonstrates improved computational and storage capabilities, empowering researchers to evaluate samples effectively. Furthermore, this software solution is suitable for all clustering-related single-molecule localisation microscopy research studies and can be customised with additional features.

Zusammenfassung

Das Ziel diese Dissertation ist, eine Softwarelösung für die bayessche Clusteranalyse von Einzelmolekul-Mikroskopie-Daten zu entwickeln und zu bewerten sowie die Etablierung einer Bildgebungsmethode von in fluider Umgebung befindlichen Zellen. Vorherige Studien haben gezeigt, dass Clusteranalysen durch Benutzer-Bias beeinflussbar sind und fortgeschrittene Techniken Programmierkompetenzen erfordern. Zudem wurde gezeigt, dass DNA-Tethering Potenzial hat, um durch externe Reize ausgelöste Clusterphänomene ohne Zelladhäsion zu untersuchen. Jedoch ist die Anwendung von DNA-Tethering aufgrund des erforderlichen fortgeschrittenen chemischen Fachwissens nicht weit verbreitet. Diese Dissertation konzentriert sich auf die Entwicklung einer benutzerfreundlichen Softwarelösung. Die Hauptziele sind die Vereinfachung der bayesschen Clusteranalyse und des DNA-Tetherings. Das entwickelte DNA-Tethering-Assay ermöglicht die Untersuchung von Zellproben ohne Oberflächenadhäsion. Dadurch kann das Verhalten von Zellen unter spezifischen und lokalen externen Reizen sowohl in lebenden Zellen als auch im Einzelmolekül-Modus untersuchen. Die präsentierte Softwarelösung baut auf bereits veröffentlichten Arbeiten auf und ihr Ziel ist, die Leistungsfähigkeit und Zugänglichkeit zu verbessern. Außerdem untersucht sie die Auswirkungen und Grenzen verschiedener Cluster-Algorithmen auf Einzelmolekul-Mikroskopie-Daten. Die Software wurde anhand simulierten und experimentellen Daten implementiert und evaluiert. Darüber hinaus wurden Clusteranalysen an Daten aus dem vereinfachten DNA-Tethering-Assay durchgeführt. Diese Studie präsentiert ein modulares Softwaretool mit einer benutzerfreundlichen grafischen Benutzeroberfläche, das die Clusteranalyse von Einzelmolekul-Mikroskopie-Daten erleichtert. Die Software zeigt verbesserte Rechen- und Speicherkapazitäten und ermöglicht eine einfache Auswertung von Proben. Darüber hinaus ist diese Softwarelösung für alle clusteringbezogenen Einzelmolekul-Mikroskopie-Studien geeignet und kann mit zusätzlichen Funktionen angepasst werden.

Declaration of Independence

Herewith I certify that I have prepared and written my thesis independently and that I have not used any sources and aids other than those indicated by me. I also declare that I have not submitted the dissertation in this or any other form to any other institution as a dissertation.

Contents

Abstract **v**

Zusammenfassung **vii**

Declaration **ix**

Acronyms **xv**

List of Figures **xxi**

List of Tables **xxiii**

1 Introduction **1**

2 Theory **7**

2.1 Biological background 7

2.1.1 CD95 receptor structure 8

2.1.2 CD95 ligand structure 9

2.1.3 Signalling pathways upon CD95 activation 9

2.2 Optics 11

2.2.1 The physics of fluorescence 11

2.3 Fluorescence microscopy 13

2.3.1 Single-Molecule Localisation Microscopy 15

2.4 Labelling methods 17

2.5 Cluster analysis 18

2.5.1 Centroid-based clustering 19

2.5.2 Hierarchical clustering 21

2.5.3 Density-based clustering 21

2.5.4 Ripley's K-, L-, and H-function . . . 22
2.5.5 Persistence-based clustering . . . 24
2.6 Bayesian Theory . . . 25
2.6.1 Bayesian data analysis for clustering . . . 26
2.6.2 Bayesian model based clustering for SMLM . . . 26

3 Material and Methods 29
3.1 Material . . . 29
3.1.1 Cell lines . . . 29
3.1.2 Cell culture . . . 30
3.1.3 Glass cleaning . . . 30
3.1.4 Supported lipid bilayer . . . 30
3.1.5 Cell tethering . . . 32
3.1.6 Cell staining . . . 32
3.2 Setups . . . 32
3.2.1 Spinning disk microscope . . . 33
3.2.2 Single-molecule localisation microscopy setup . . . 33
3.3 Computational methods . . . 34
3.3.1 Localisation Software . . . 34
3.3.2 Simulator for membrane protein dynamics . . . 35
3.3.3 Programming languages . . . 36
3.4 Computational power . . . 38

4 Results 41
4.1 BaClAva . . . 41
4.1.1 The structure of BaClAva . . . 41
4.2 Back-end improvements . . . 46
4.2.1 Influence of choice of algorithm - Ripley's, DBSCAN, ToMATo . . . 47
4.2.2 Sequential vs parallel computation . . . 49
4.2.3 Data storage formats . . . 52
4.3 Understanding clustering . . . 55
4.3.1 Influence of cluster sizes . . . 55
4.3.2 Influence of fluorophore density . . . 58

4.4 Clustering of biological proteins - GPI vs CD95 vs CCP 63
4.5 Tethering 66
4.5.1 The tethering procedure 66

5 Discussion **73**
5.1 BaClAva 73
5.1.1 Clustering 73
5.1.2 Restructuring of BaClAva 79
5.1.3 Simulation work 82
5.1.4 Cluster Analysis of SMLM of Cells 85
5.2 Tethering 87

6 Outlook **95**
6.1 BaClAva 95
6.2 Cell tethering 98

7 Conclusion **103**

8 Publication **107**

A Protocols **109**
A.1 Cell culture 109
A.1.1 Keeping cells in the cell culture 109
A.1.2 Cell seeding on glass slides 110
A.1.3 Transfections 111
A.2 Glass cleaning and channel preparation 112
A.2.1 Glass cleaning 112
A.2.2 Cleaning the channel top parts 112
A.2.3 Channel construction 113
A.3 Supported lipid bilayer formation 113
A.3.1 Lipid preparation 113
A.3.2 Supported lipid bilayer preparation 114
A.4 Tethering 115
A.5 Cell staining 116

A.5.1 Cell staining on glass 116
A.5.2 Staining of tethered cells 119

B Sequences **121**
B.1 DNA sequences 121
B.2 Live-cell tethering sequences 121

Acknowledgements **125**

Bibliography **127**

Acronyms

(*d*)STORM (*direct*) Stochastic Optical Reconstruction Microscopy.

$MgSO_4$ Magnesium sulfate.

2D two dimensional.

3D three dimensional.

AF647 Alexa Fluor 647.

AKT protein kinase B.

AOTF Acousto-Optic Tunable Filter.

API Application Programming Interface.

BaClAva Bayesian Cluster Analysis and visualization application.

BC O^6-benzylcytosine.

BDA Bayesian Data Analysis.

BG O^6-benzylguanine.

BSA Bovine Serum Albumin.

CCP clathrin-coated pit.

CD95L CD95 ligand.

CDF cumulative distribution function.

$cFLIP_L$ cellular FADD-like interleukin-1-β-converting enzyme-inhibitory protein.

CLC clathrin light chain.

CO_2 carbon dioxide.

CPU Central Processing Unit.

CRAN The Comprehensive R Archive Network.

CRD cysteine-rich domain.

CSV comma-separated values.

DBSCAN Density-Based Spatial Clustering of Applications with Noise.

DD Death Domain.

DISC death-inducing signalling complex.

DNA deoxyribonucleic acid.

DNA-PAINT DNA-Point Accumulation for Imaging in Nanoscale Topography.

EM electron multiplier.

EMCCD Electron Multiplying Charge Coupled Device.

EPS Encapsulated PostScript.

ERK extracellular signal-regulated kinase.

FACS Fluorescence Activated Cell Sorting.

FADD Fas-assiciated protein with death domain.

FBS Fetal Bovine Serum.

FWHM Full Width at Half Maximum.

GA Glutaraldehyde.

GFP Green Fluorescent Protein.

GPI Glycosylphosphatidylinositol.

GPU Graphics Processing Unit.

GUI graphical user interface.

hACT O^2-alkylcytosine-DNA-alkyltransferase.

hAGT O^6-alkylguanine-DNA-alkyltransferase.

HDF5 Hierarchical Data Format version 5.

HeLa human cervical cancer cells that were retrieved from Henrietta Lacks in 1951.

HILO highly inclined and laminated optical.

m-CD95L membrane-bound form of CD95L.

MMP matrix metalloproteases.

$NaBH_4$ Sodium borohydride.

NF-κB nuclear factor κB.

NH_4Cl Ammonium chloride.

NP nondeterministic polynomial.

OOP Object-oriented programming.

PALM Photo-Activation Localisation Microscopy.

PBS Phosphate Buffered Saline.

PDF Portable Document Format.

PFA Paraformaldehyde.

PI3K phosphoinositide 3-kinase.

PIP2 phosphatydilinositol-4,5-bis-phosphate.

PIP3 phosphatydilinositol-3,4,5-tris-phosphate.

PNG Portable Network Graphics.

POI protein of interest.

PSF Point Spread Function.

PTEN phosphatase and tensin homolog.

PyPI Python Package Index.

RAS rat sacroma.

RIPK receptor-interacting protein kinase.

ROI region of interest.

RTK receptor tyrosine kinase.

s-CD95L soluble form of CD95L.

SFK src-family kinase.

SH2 src homology domain 2.

SIM Structured Illumination Microscopy.

SLB supported lipid bilayer.

SLP self-labelling protein SLP.

SMAP Super-resolution Microscopy Analysis Platform.

SMLM Single-Molecule Localisation Microscopy.

SNR signal to noise ratio.

SSE sum of squared errors.

STED Stimulated Emission Depletion microscopy.

TDA Topological Data Analysis.

THD TNF homology domain.

TIFF Tagged Image File Format.

TIRF Total Internal Reflection Fluorescence.

TIRF total internal reflection fluorescence.

TMD Transmembrane Domain.

TNF Tumour Necrosis Factor.

ToMATo Topological Mode Analysis Tool.

List of Figures

2.1 Structure of the CD95 receptor and its ligand 8
2.2 CD95 signalling pathways 10
2.3 Jablonski diagram 12
2.4 Rayleigh criterion 13
2.5 Illumination modes 15
2.6 SMLM process 16
2.7 K-means clustering 20
2.8 DBSCAN clustering 22
2.9 ToMATo clustering 24

4.1 Flow diagram illustrating the three features of BaClAva 42
4.2 Schematic representation of the GUI-based software BaClAva. 43
4.3 Cluster algorithms compared with BaClAva 47
4.4 Verification of computation on a single dataset 50
4.5 Comparison of cluster statistics on a per-dataset level. 51
4.6 Comparison of cluster statistics on a per-cluster level. 51
4.7 Cluster algorithms of BaClAva compared for computational speed 52
4.8 Comparison of storing data in different file formats. 53
4.9 Comparison of storing data in different file formats. 57
4.10 K-means on datasets with two cluster sizes 59
4.11 The impact of the fluorophore density on clustering 60
4.12 Clustering three different target molecules with BaClAva in biological samples. 64
4.13 Verification of the clathrin-coated pit size 65
4.14 Tethering schematic and illustration of a T98G cell anchored to a SLB. . . 66

4.15 Quality assessment of SLBs 67
4.16 Time-lapse microscopy of tethered and untethered cells. 68
4.17 Time-lapse fluorescence imaging of anchored T98G CD95-GFP cells. . . . 69
4.18 SMLM reconstruction and clustering of tethered T98G cells. 70
4.19 Cluster analysis of tethered T98G CD95-GFP cells using dSTORM and Baclava. 71

B.1 Time laps of the cell in Fig 4.17 (A). 122
B.2 Time laps of the cell in Fig 4.17 (A) 123

List of Tables

3.1 Material for the cultivation of the cell line and work in the cell culture . . 30
3.2 Material used for the cleaning of the glass cover glasses 31
3.3 Materials used for the supported lipid bilayer 31
3.4 The material used for the cell tethering and their control experiments. . . 32
3.5 The material used for staining living cells or fixing and then staining them. 33
3.6 The laser of the TIRF microscope 34
3.7 List of R packages 37
3.8 List of Python packages 38
3.9 The computing resources and their specification 39

4.1 Overall cluster statistics for simulations with two input cluster sizes 56
4.2 Clustering on a dataset with two cluster sizes. 56

A.1 Electroporation Parameters for T98G cells 112

B.1 DNA-oligomer sequences 121

1

Introduction

Microscopy, a scientific technique employed to examine physical, chemical, and biological specimens to unravel their properties, has played a vital role in knowledge accumulation throughout the centuries [162]. Microscope design progressed through empirical experimentation until Ernst Abbe developed the theoretical foundations. In 1873, Abbe demonstrated that the diffraction of light, arising from its wave nature, imposes an intrinsic constraint on microscope resolution. He concluded that the minimum achievable resolution of a microscope is approximately half the wavelength of its light source [2]. Consequently, the resolution limit for visible light falls within the range of 150 nm to 300 nm.

The theoretical understanding of microscopy enabled manufacturing advancements, which led to the development of fluorescence microscopy in the early 20^{th} century by Oskar Heimstaedt [162]. Fluorescence microscopy enhances the signal-to-noise ratio by employing fluorescent dyes that selectively label biological specimens. Moreover, it is a relatively non-invasive technique. However, biological samples, comprising cells with sizes on the micrometre scale and subcellular components on the nanometre scale, require higher resolution than the resolution limit to visualise subcellular components and intracellular processes [118].

There are two distinct ways to bypass the resolution limit in fluorescence microscopy.

The first approach involves structured illumination schemes to acquire super-resolution images. Prominent examples include Stimulated Emission Depletion microscopy (STED) [63] and Structured Illumination Microscopy (SIM) [57]. These methods employ specific illumination patterns to surpass the diffraction limit and achieve enhanced spatial resolution. Single-Molecule Localisation Microscopy (SMLM) exploits the photophysical properties of fluorophores for super-resolution microscopy. Photo-Activation Localisation Microscopy (PALM)[65] is one of the pioneering methods in this field where fluorophores are selectively activated and localised. In (*direct*) Stochastic Optical Reconstruction Microscopy ((*d*)STORM) [92, 131], another technique under the SMLM umbrella, conventional fluorophores randomly switch between fluorescent and non-fluorescent states. [61, 91]

Consequently, only a small subset of fluorophores emit fluorescence at any given time, enabling spatial and temporal separation of the fluorophore signal. By fitting these positions obtained by PALM or (*d*)STORM using a Gaussian function, a super-resolution image with a lateral resolution of up to 20 nm can be reconstructed. In 2014, the Nobel Prize was awarded to Eric Betzig, William E. Moerner and Stefan W. Hell for their groundbreaking contributions in developing super-resolution fluorescence microscopy [111]. Their pioneering work has established the foundation for super-resolved imaging, revolutionising the way biological samples are visualised and expanding the possibilities for investigating cellular structures and processes at nanoscale resolutions.

DNA-PAINT (DNA-Point Accumulation for Imaging in Nanoscale Topography) has pushed resolution boundaries even further, achieving a spatial resolution below 5 nm. This technique utilises the transient binding of short DNA strands to label and localise target molecules [141]. Overall, advancements in super-resolution imaging techniques have revolutionised the field by enabling the visualisation of biological structures and processes with unprecedented detail and spatial resolution.

Over the past decade, cluster analysis has emerged as a prominent and extensively employed methodology for examining the spatial organisation of molecules of interest within their complex cellular milieu [76, 110]. Cluster analysis offers tools for exploring

the spatial distribution of molecules of interest within a complex cellular environment. The localisations obtained from fluorescently tagged molecules can be categorised into distinct clusters by using their spatial proximity to gain insights into their arrangements, interactions, and potential functional implications. Cluster analysis faces inherent challenges due to the pointillistic nature of single-molecule localisation microscopy. Cluster analysis in SMLM frequently leads to substantial errors, and results lack scientific substantiation. Furthermore, the accuracy and validity of clustering rely on the quality of the underlying data and the chosen clustering methods [110, 130].

This analytical approach is highly susceptible to subjective biases introduced by users, further compromising the reliability and objectivity of the findings. The choice of clustering parameters can significantly impact cluster results. As the decisions are often based on random user-defined criteria, the risk of introducing personal biases becomes evident, compromising the objectivity and reproducibility of the findings [110, 117, 130].

Consequently, this thesis presents a novel software solution named Bayesian Cluster Analysis and visualization application (BaClAva) based on the work of Rubin-Delanchy et al. [130], Griffié et al. [54], and Pike et al. [117]. This solution addresses cluster analysis limitations in single-molecule localisation microscopy with special focus on (*d*)STORM. By integrating advanced algorithms and statistical techniques, BaClAva aims to reduce the errors and subjective biases commonly encountered in cluster analysis by offering a user-friendly graphical user interface (GUI). The main results on BaClAva were published in Kutz et al. [82].

In SMLM and traditional wide-field or confocal microscopy, the typical imaging procedure involves the examination of cells cultured on glass substrates, occasionally featuring a thin coating. However, visualising specific cellular structures or molecules in or near the cell membrane in response to external stimuli is challenging. The membrane of interest for imaging is the cellular membrane that faces both the substrate and the microscope objective. Thus, the goal is to activate the cell exclusively in this area. Nevertheless, achieving precise and controlled activation through external stimuli, such as the introduction of ligands, poses a significant technical challenge when dealing with cells grown

on glass substrates because the stimulant either needs to diffuse to the target area while simultaneously stimulating a response from the cell via alternative regions or the cells must be cultivated in the presence of the stimulant, which can trigger an immediate response from the cells. Both scenarios prove suboptimal for a controlled experiment with well-defined activation initiation.

This thesis introduces a more sophisticated procedure for anchoring cells via DNA tethering [142, 150], designed to fully immerse cells in a fluid medium during imaging. Unlike conventional sample preparation methods, this approach involves anchoring the cells above the substrate, facilitating their immersion in a fluid environment and thereby preventing their adhesion to the surface. As a result, this technique offers the unique opportunity to exclusively capture and study the cellular response to external stimuli unaffected by unwanted influences.

The focus of this thesis is on cluster analysis of SMLM data, specifically investigating the spatial arrangement of the CD95 receptor in T98G cells. This research aimed to develop an analytical tool capable of analysing and characterising CD95 receptor clustering patterns. The user-friendly software BaClAva, based on Bayesian-engine-based cluster analysis [130], was developed to achieve this objective. BaClAva delivers a comprehensive and efficient platform for cluster analysis of SMLM data. In this way, BaClAva facilitates the exploration and interpretation of the spatial organisation of the CD95 receptor in T98G cells.

In addition to the development of BaClAva, this research also proposes a novel approach to DNA-tethering cells to supported lipid bilayers (SLBs) in channels. This innovative technique was designed to enable (*d*)STORM imaging of the CD95 receptor while allowing for controlled activation of the cells. By ensuring a controlled environment and eliminating potential confounding factors, the aim was to obtain accurate and reliable clustering information reflecting the spatial arrangement of the CD95 receptor in T98G cell.

This thesis is organised into seven chapters. After this introduction, Chapter 2 lays the foundation by presenting a theoretical background on the CD95 receptor, SMLM imaging

techniques, and labelling methodologies. It also introduces the clustering algorithms used and to be characterised for SMLM in this thesis. Chapter 3 delves into the materials and methods employed, encompassing information on the cell lines, chemicals, and coding libraries used throughout the research. Following that, Chapter 4 presents the results of the BaClAva software development process, including its application on simulated and cell samples and the implementation of the CD95 cell tethering technique for SMLM imaging. Chapter 5 discusses the results and their significance in the broader research field. Chapter 6 takes a forward-looking perspective, exploring potential future research directions. Among these are possible extensions for BaClAva, unanswered spatial and temporal organisation questions concerning CD95, and other research options that may yield valuable insights from BaClAva, the developed tethering approach, or both. Finally, the thesis concludes in Chapter 7 with a concise summary of the work undertaken in this study.

This thesis elucidates the user-friendly nature of BaClAva, highlighting its graphical user interface (GUI) and efficient computation capabilities. Additionally, various cluster algorithms are examined extensively to determine their suitability for analysing SMLM clusters of biological probes. Simulations of artificial SMLM clusters are employed to evaluate cluster algorithms. Furthermore, a novel tethering approach is introduced, offering a stable anchoring method for long-term imaging procedures such as live-cell imaging or SMLM. The CD95 receptor in T98G cells serves as the primary biological focus, with each experiment demonstrating its clustering nature within the cellular membrane.

2

Theory

This chapter introduces the foundations of this thesis, outlining the theoretical basis for understanding the research. This section describes the key receptor of interest and provides an in-depth presentation of the microscopy techniques employed throughout this research work. The chapter concludes by introducing the concept of clustering, emphasizing its importance, and presenting a range of clustering algorithms relevant to this study.

2.1 Biological background

Cancer is defined as the uncontrolled proliferation of cells arising from signalling pathway disruption within the cells. These cells evade the self-regulatory mechanisms that govern tissue growth. Cancerous cells can adapt to and manipulate their local environment, enabling them to thrive. Thus, cancer therapy aims to eradicate cancerous cells and their microenvironment but, at the same time, prevent the development of resistance to the treatment. [11, 98]

One of the receptors widely found in cancer cells is the CD95 receptor. To date, no mutation has been identified that results in loss of receptor function. Upon activation, this receptor shows a binary response, either inducing cell death or promoting cell survival.

Consequently, CD95 has emerged as a promising target for research and therapy, particularly for non-curable cancers, including glioblastoma multiforme, the most prevalent primary human brain cancer. [11, 98, 134]

2.1.1 CD95 receptor structure

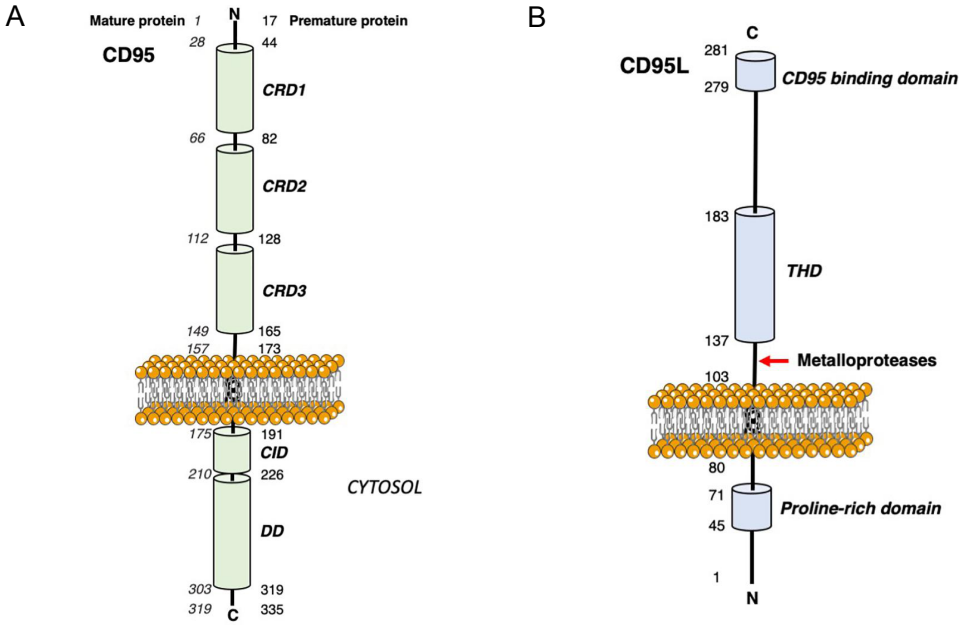

Figure 2.1: Structure of the CD95 receptor and its ligand. (A) An overview of the transmembrane protein CD95 embedded in the plasma membrane and its primary domains. CD95 receptors feature three repetitions of the cysteine-rich domain (CRD) in the extracellular domain. The Transmembrane Domain (TMD) forms a stable trimer with a rigid core and flexible ends. (B) A schematic representation of membrane-bound CD95L and its key domains. CD95L may undergo cleavage by metalloproteases, leading to the formation of its soluble derivative s-CD95L. [88]

The receptor CD95 is a Tumour Necrosis Factor (TNF) superfamily member and belongs to the TNF-subfamily of death receptors [20, 88]. CD95, also known as Fas, Apo-1, or TNFRSF6, is a type I transmembrane glycoprotein consisting of 319 amino acids [88]. Figure 2.1(A) illustrates the structure of this transmembrane receptor, featuring an elongated extracellular region, a Transmembrane Domain (TMD), and an intracellular domain. The N-terminus remains extracellular in type I transmembrane proteins, while the C-terminus resides intracellularly [20].

The monomers, homodimers, and homotrimers of CD95 can exist in a dynamic equi-

librium without being activated [41]. There is no direct involvement of the TMD in CD95 homodimerisation or homotrimerisation, but it may play an important role in the formation of CD95 clusters with a higher number of CD95 receptors. The Death Domain (DD) is located within intracellular domain of CD95 and serves as an interaction site in the apoptotic pathway. Nevertheless, the 3D structure of CD95 remains unknown due to its high aggregation tendency, domain flexibility, and TMD [20, 41].

2.1.2 CD95 ligand structure

The CD95 ligand (CD95L), also known as FasL, TNFSF6, or CD178, is a type II transmembrane protein, whose N-terminus is intracellular, and the C-terminus with the TNF homology domain (THD) is extracellular [20, 88], as depicted in Figure 2.1(B). The THD facilitates CD95L trimerisation. In its membrane-bound form (m-CD95L), CD95L activates the apoptotic pathway by stimulating CD95 [20]. By matrix metalloproteases (MMP) acting on m-CD95L, the soluble form is generated (s-CD95L). In contrast to m-CD95L, s-CD95L triggers non-apoptotic signalling pathways [20]. Depending on whether the ligand is membrane-bound or soluble, this distinction in activation modes represents a characteristic shared among various members of the TNF superfamily [20].

2.1.3 Signalling pathways upon CD95 activation

CD95 stimulation by its ligand CD95L induces several different signalling pathways, ultimately leading to either cell death or tumorigenesis [88, 98, 134], as illustrated in Figure 2.2. Tumorigenesis involves numerous processes, including cell survival, proliferation, and migration. Cancer cells, like glioma cells, tend to thrive when in contact with neighbouring cells. In contrast, individual cells display heightened vulnerability to cell death [11].

Cell death is primarily induced by the open-form stabilisation of Death Domain (DD) in response to ligand stimulation [134]. This process involves clustering CD95 receptors, eventually leading to the recruitment of Fas-assiciated protein with death domain (FADD) to the DD. Subsequently, the death-inducing signalling complex (DISC) is formed through the recruitment of procaspase-8 to the FADD [134]. The procaspase then self-cleaves,

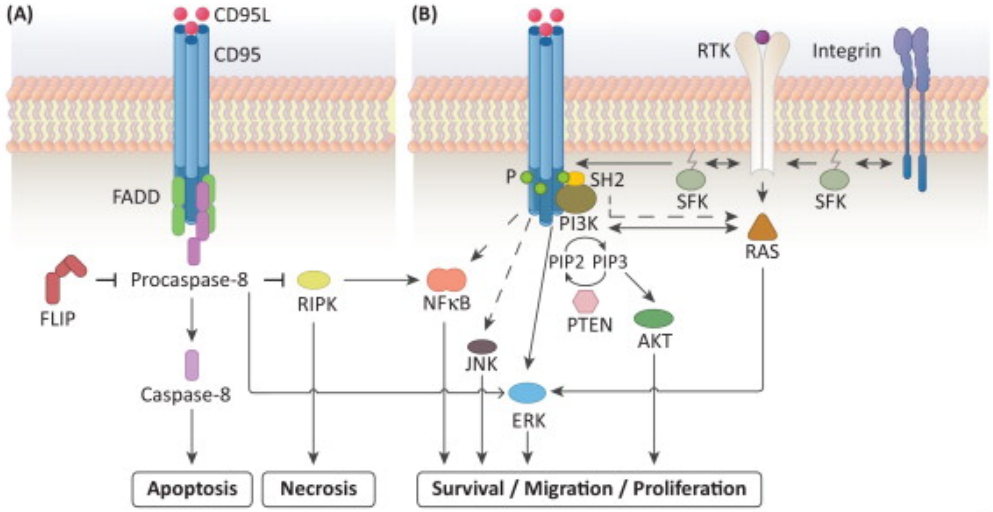

Figure 2.2: The schematic representation of the CD95 signalling network depicts the signalling pathways upon CD95 stimulation resulting in apoptosis or necrosis **(A)** and survival, migration, and proliferation **(B)**. Abbreviations: CD95 ligand (CD95L); Fas-assiciated protein with death domain (FADD); cellular FADD-like interleukin-1-β-converting enzyme-inhibitory protein (cFLIP$_L$); receptor-interacting protein kinase (RIPK); nuclear factor κB (NF-κB); receptor tyrosine kinase (RTK); src-family kinase (SFK), src homology domain 2 (SH2); phosphoinositide 3-kinase (PI3K); phosphatydilinositol-4,5-bis-phosphate (PIP2); phosphatydilinositol-3,4,5-tris-phosphate (PIP3); phosphatase and tensin homolog (PTEN); extracellular signal-regulated kinase (ERK), rat sacroma (RAS) [98]

producing caspase-8, which activates the apoptosis pathway. In addition to their role in apoptosis, FADD and caspase-8 also regulate a second pathway leading to cell death, namely necrosis [98], and contribute to tumorigenesis [88].

Upon CD95L activation, phosphorylation of CD95 by src-family kinases (SFKs) initiates various survival, proliferation, and migration pathways [98]. CD95 phosphorylation results in phosphoinositide 3-kinase (PI3K)-dependent protein kinase B (AKT) activation, promoting cell migration and invasion [98]. Interestingly, glioblastoma multiforme cells often possess phosphatase and tensin homolog (PTEN) mutations, leading to loss-of-function, which leaves the cell incapable of inhibiting this pathway. Additionally, PI3K can phosphorylate caspase-8, inhibiting apoptosis [98, 134].

Multiple pathways contribute to extracellular signal-regulated kinase (ERK) activation, leading to diverse cellular outcomes, including cell adhesion, migration, invasion, proliferation, and differentiation [11, 98]. ERK activation is achieved by recruiting src homology domain 2 (SH2) to CD95. Another route to ERK activation involves CD95

stimulation-induced activation of nuclear factor κB (NF-κB). Additionally, CD95 can interact with other membrane-bound molecules, such as receptor tyrosine kinases (RTKs), leading to tumorigenesis.

2.2 Optics

2.2.1 The physics of fluorescence

Similar to atom transitions, which are stimulated by excitation with light of a specific wavelength, molecules can also undergo exposure to excitation light. Due to the presence of two to hundreds of atoms, molecular physics possesses a higher level of complexity. Within molecules, electronic states exist, and they possess vibrational sub-states induced by the vibrational motion of each atom. Due to the ability of molecules to rotate around their bonds, these vibrational states also have rotational sub-states. Consequently, a quasi-continuum is observed at each electronic level. [10, 59, 136]

The process involved in a molecule emitting fluorescent light begins with an electron in a delocalised state. When the molecule absorbs a photon from an adequate excitation light source, such as a laser with the optimal wavelength and thus energy, it causes the electron to transition from its electronic singlet ground state to an excited state, as illustrated by the Jablonski diagram in Figure 2.3. In line with the Franck-Condon principle, the excitation process favours the vibrational level in the excited electronic state that exhibits the highest wave function overlap with the ground state. Subsequently, the molecule undergoes internal conversion, dissolving its vibrational and rotational energy within its local environment. As stated by Kasha's rule, electronic energy re-emission happens at the lowest vibrational level of the excited singlet state. A Stokes Shift takes place, which results in the emitted photon possessing lower energy than the absorbed photon, resulting in the emitted light having a longer wavelength and thus less energy. The Franck-Condon principle determines the vibrational level of the ground state. This rapid process, occurring within nanoseconds, is known as fluorescence. [10, 59, 137]

Inter-system crossing is an alternative pathway from the excited singlet state involving a

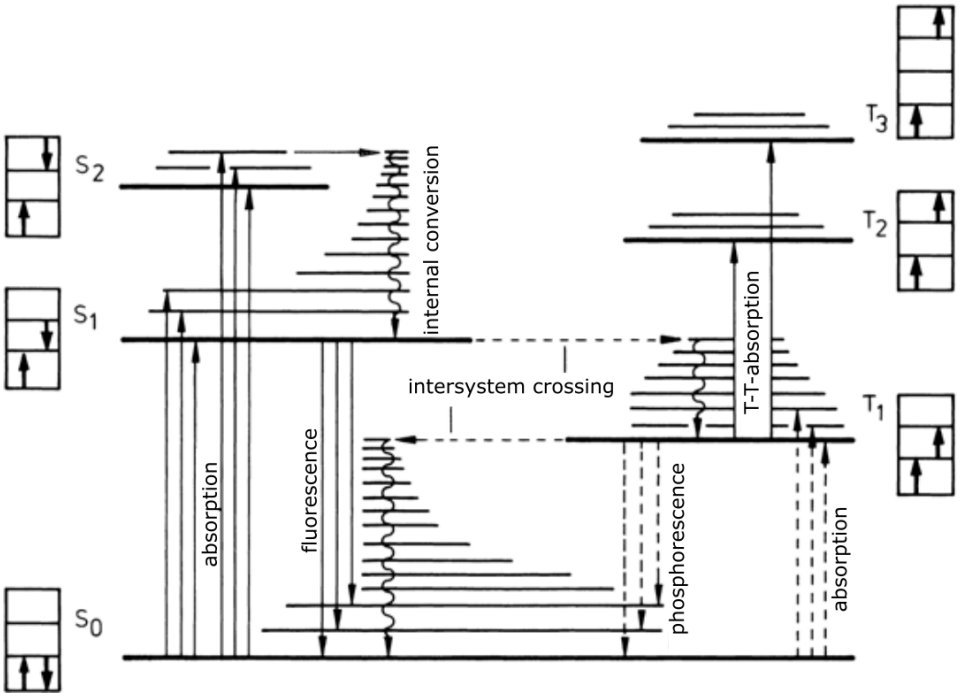

Figure 2.3: The Jablonski diagram illustrates the initial three singlet and triplet states in a molecule, along with the potential transitions that can occur. Furthermore, the diagram presents the electronic spin configuration for each electronic state, demonstrating the requirement of a spin flip to enable intersystem crossing. [59].

spin-flip of the excited electron into the metastable triplet state. Consequently, another spin flip is required to return the electron to the singlet ground state. However, this quantum-mechanical process has a very low probability and is referred to as forbidden. Hence, the electron experiences a more prolonged decay, typically lasting from microseconds to milliseconds. When energy is emitted as a photon during the decay process to the singlet ground state, it is called phosphorescence. The triplet state has a longer lifetime, which increases the probability of the molecule reacting with radicals, leading to its reduction or oxidation. Appropriate buffer conditions can return the molecule to its singlet ground state within seconds. However, if recovery is impossible, the molecule undergoes photobleaching. [59, 137]

2.3 Fluorescence microscopy

Optical microscopes serve as valuable tools for discerning structures beyond human vision. However, these microscopes are also subject to limitations caused by the wave nature of light. As highlighted by Abbe in 1873 for gratings [2], the spatial resolution, denoted by the aperture angle 2α, establishes the minimum distance, d, required for two spatially distinct objects to remain differentiable. This relationship is expressed in the equation:

$$d = \frac{\lambda}{2\,n\,\sin\alpha} = \frac{\lambda}{2\,\mathrm{NA}} \quad , \tag{2.1}$$

where λ represents the wavelength of the incident light, n corresponds to the refractive index of the immersion substance, and NA signifies the numerical aperture [59].

Fluorophores are regarded as point-like sources of light in fluorescent microscopy. When a long-pass filter eliminates excitation light, red-shifted emitted light from fluorophores can be detected exclusively with a camera or similar detector [36]. Microscope optics allow the sensor to capture the diffraction pattern of the emitted light. In conventional wide-field microscopy, this diffraction pattern is the microscope's Point Spread Function (PSF), see Fig. 2.4. A PSF with no aberration is called an Airy disc, with a central maximum and higher-order rings around the maximum [80, 106].

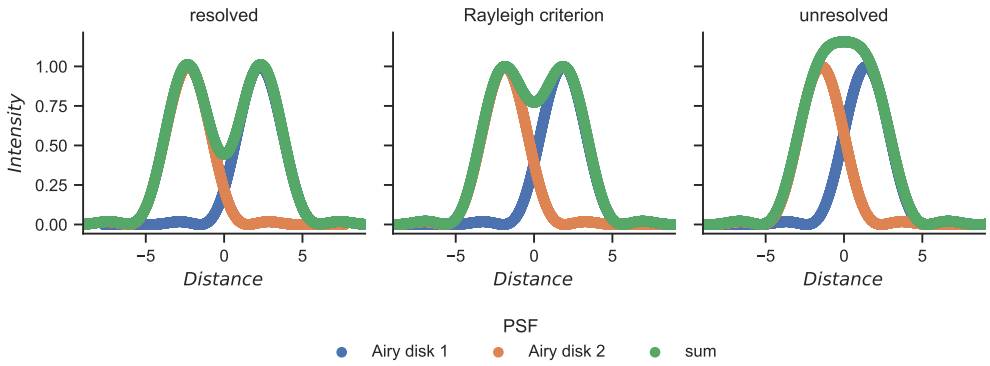

Figure 2.4: Rayleigh criterion. The Rayleigh criterion is illustrated in three intensity-distance diagrams, which display the intensity distribution of the PSFs for two point sources. From left to right, the distance between the two prominent peaks decreases until they eventually merge and become indistinguishable in their combined intensity profile. Only when the distance between the two peaks is equal to or larger than the Rayleigh criterion can the point sources be resolved as separate entities.

The Rayleigh criterion determines the resolution in fluorescent microscopy. It denotes the minimum distance required to differentiate two adjacent point sources. When the minimal distance is reached, the maximum of the first emission pattern coincides with the minimum of the second emission pattern, as shown in Figure 2.4. The lateral resolution, characterised by the minimal distance, d, to discern both distributions, is determined by the equation [1]:

$$d = 0.61\frac{\lambda}{NA} \quad , \tag{2.2}$$

where λ represents the wavelength of the emitted light from the two point sources. Resolution is limited by the objective aperture, NA, at a given wavelength. In this thesis, a NA $= 1.49$ objective and the fluorescent dye Alexa Fluor 647 (AF647) with an emission maximum of around 700 nm were used. Therefore, the minimum distance to resolve two AF647 dyes is 290 nm.

Apart from resolution, contrast is crucial when evaluating microscopes, as it indicates the ability to distinguish objects of interest from their backgrounds [36]. Fluorescent microscopy can obtain high contrast by labelling structures with dyes that exhibit high quantum yields. However, background noise from diverse sources can diminish the signal to noise ratio (SNR), including scattered excitation light, autofluorescence of cells, unspecific fluorophore binding, and detector readout [137].

Wide-field microscopy can benefit from employing various illumination techniques (Figure 2.5) to improve the signal-to-noise ratio (SNR). The standard illumination mode is the epi-mode (Figure 2.5 A), which uniformly excites the entire sample in width and depth [80]. However, this can lead to out-of-focus light, introducing noise to the resulting image. Another illumination method involves angling the excitation light beam at the critical angle between the glass and the sample as shown in Figure 2.5 B. This configuration allows for the selective detection of signals from regions close to the coverslip, reducing out-of-focus noise. In this setup, an evanescent electromagnetic field penetrates the sample. However, its intensity exponentially diminishes with depth,

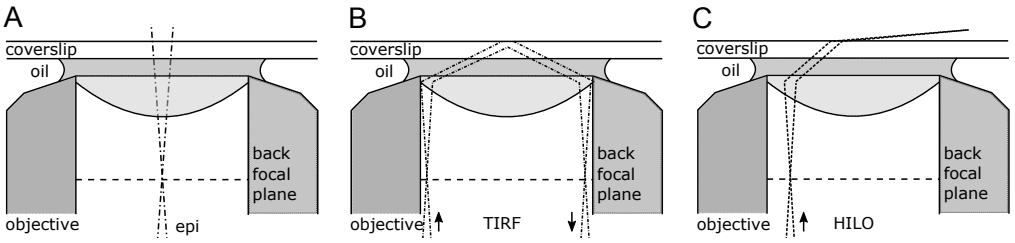

Figure 2.5: Illumination modes. (A) In the epi-mode, the excitation laser beam travels directly through the objective. The immersion medium used can be oil or water. The excitation light is focused at the centre of the back focal plane. (B) In the TIRF mode, the light beam is focused on the border of the back focal plane. Consequently, the light undergoes total internal reflection, resulting in only an evanescent field (not shown) that can penetrate the sample above the coverslip. (C) In the HILO mode, the excitation beam is focused on the back focal plane between the epi and TIRF focal points based on the required depth and beam width in the sample. To realise TIRF and HILO modes, an immersion medium with a refractive index matching that of the glass is required.

limiting the excitation to a region approximately 200 nm from the glass surface. This specific illumination mode is called total internal reflection fluorescence (TIRF) [17, 135].

An alternative technique involves positioning the illumination beam close to the critical angle (Figure 2.5 C), forming a highly inclined and laminated optical (HILO) sheet [18]. TIRF and HILO methods require oil-immersion objectives, where the refractive index of the immersion medium is matched to that of the objective lens and the coverslip [17]. These approaches provide enhanced SNR and reduced background noise, thus improving the overall quality of wide-field microscopy imaging.

2.3.1 Single-Molecule Localisation Microscopy

Single-Molecule Localisation Microscopy (SMLM) is an umbrella term for statistical fluorescent microscopy techniques that achieve sub-diffraction resolution. The most commonly used methods in this field are (*direct*) Stochastic Optical Reconstruction Microscopy ((*d*)STORM) [92], Photo-Activation Localisation Microscopy (PALM) [15], and DNA-Point Accumulation for Imaging in Nanoscale Topography (DNA-PAINT) [141].

(*d*)STORM is a technique that employs fluorophores and relies on blinking buffer conditions specific to the fluorophores [92]. In PALM, photoactivatable fluorescent proteins are expressed alongside the structure of interest within the specimen [15]. The third

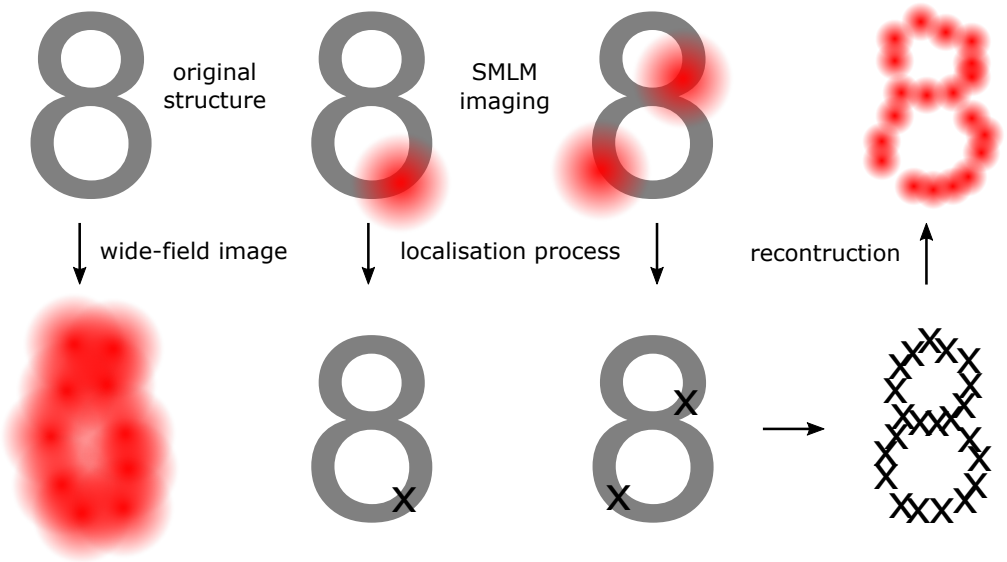

Figure 2.6: SMLM Process. Initially, a structure labelled with fluorophores and imaged in wide-field mode yields a blurred representation due to the resolution limit. However, employing SMLM techniques enables the temporal separation of fluorescent signals from individual dyes. The signals of each fluorophore can therefore be precisely located. Plotting the obtained coordinates reveals the underlying structure. The individual localisations and their associated localisation errors are utilised to reconstruct a super-resolved image.

method, known as DNA-PAINT, is based upon the application of photostable dyes linked to deoxyribonucleic acid (DNA) strands, with the blinking behaviour governed by the binding times of the DNA strands [141].

SMLM methods isolate the emission from individual fluorophores in time by manipulating the on and off times of the fluorophores (or the binding times for DNA-PAINT) as shown in Figure 2.6. This isolation enables the accurate fitting of a highly precise PSF model to the detected emission signal. In the case of (*d*)STORM, the objective is to control the fluorophores so that they spend more time in the dark than in the fluorescent state. However, they should remain in the fluorescent state long enough for the camera to capture an adequate number of photons to precisely determine the centre of the PSF. A high-resolution image can be reconstructed using fitted centre coordinates (localisations) and additional fitting parameters. Additionally, all localisation information is stored in the localisation table. [137]

2.4 Labelling methods

Various labelling methods are available for biological specimens using markers such as dyes or DNA strands [109]. One commonly used approach is immunofluorescence. Therefore, antibodies are combined with primary antibodies designed to target the protein of interest (POI). Secondary antibodies, labelled with an organic fluorophore or fluorescent dye, are directed against the host of the first antibody. Alternatively, labelling the primary antibody directly with a dye is possible [109]. Multiple secondary antibodies binding to a single primary antibody produce a stronger fluorescent signal. Consequently, it is common practice to stain immunofluorescence microscopy samples with primary and secondary antibodies.

The second labelling method involves cloning and expressing fluorescent proteins coupled to the protein of interest [86, 109]. This approach results in a highly specific fluorescent signal due to the combined expression of the POI and its label. However, the specificity may be compromised if either the fluorescent protein or the POI is not correctly folded or incorporated. There are antibodies available that target fluorescent proteins with high specificity. Some animals, such as camels, dromedaries, and llamas, have single-chain antibodies, whereas humans and other vertebrates have two-chain antibodies [13]. Nanobodies, antigen-binding fragments of Camelid antibodies, are particularly noteworthy in this context. They have a size of approximately 15 kDa, significantly smaller than the 150 kDa size of a two-chain antibody. Consequently, when nanobodies are used instead of antibodies, the distance between the structure of interest and the fluorescent dye reduces by at least 10 nm [13, 86].

Lastly, self-labelling protein SLP (SLP) methods deserve consideration. SNAP-tag, CLIP-tag, and HaloTag are among the widely recognised SLPs [93]. The fundamental concept involves expressing a fusion tag alongside the protein of interest, allowing the tag to establish a specific covalent bond with the labelled substrate. These SLP methods find application in both in vitro and in vivo settings, as neither the tag nor the substrate exhibit toxicity to cells.

The SNAP-tag and CLIP-tag are derived from human DNA-repair enzymes, namely O^6-alkylguanine-DNA-alkyltransferase (hAGT) and O^2-alkylcytosine-DNA-alkyltransferase (hACT) [93]. The hAGT and hACT enzymes have 207 amino acids and size of 21 kDa [42]. The engineered form of hAGT, known as SNAP-tag, specifically recognises and reacts with BG [75]. The catalytic cysteine of the SNAP-tag forms a stable covalent bond with the BG labelling group, releasing guanine [42]. Similarly, the CLIP-tag, an engineered version of hACT, functions complementarily to the SNAP-tag, interacting with O^6-benzylcytosine (BC) instead of BG, thus releasing cytosine [93].

The HaloTag is a mutant of a bacterial dehalogenase with no inherent reaction in mammalian cells [95]. This monomeric protein consists of 297 amino acids and has a size of 33 kDa [139]. Its substrate is an aliphatic hydrocarbon molecule commonly known as the HaloTag ligand. In contrast to the wild-type enzyme, which removes halogens from the substrate through a nucleophilic aspartate residue, the engineered HaloTag forms an irreversible covalent bond with any chloroalkane ligand [95, 139].

2.5 Cluster analysis

Cluster analysis aims to identify clusters or groups within a dataset based on specific measurements [64]. These clusters consist of data points that share similar characteristics and are distinguishable from others outside the cluster. In formal notation, the dataset is represented as $D = \{x_1, x_2, \ldots, x_n\}$, where x_i represents the i-th data point. As a result of clustering, these data points are divided into K distinct subsets, denoted by $C = \{C_1, C_2, \ldots, C_k, \ldots, C_K\}$. The cluster label, k, ranges from 1 to K [47]. The label k is assigned to data point x_i when it belongs to a cluster C_k. However, not every possible partitioning of dataset D is universally considered "excellent" or valuable. Moreover, there is no universal agreement on the definition of what constitutes a cluster or what characterises optimal clustering of D [64].

Cluster analysis starts with understanding the intrinsic properties of the data and determining the most appropriate clustering algorithm for the given task based on the data

context [3, 76]. There are various approaches to characterising clustering algorithms and clusters. Each data point is exclusively assigned to a single C_k in hard clustering. In contrast, soft clustering assigns each data point a degree of membership to each cluster [64].

Moreover, clustering can be classified into flat and hierarchical depending on the algorithm selected [3]. Flat clustering computes a single solution for the dataset. Hierarchical clustering produces multiple partitions at different levels of granularity in a single run. In hierarchical approaches, each cluster can be subdivided into smaller clusters or merged to form superclusters [64].

The type of clustering algorithm depends on the type of data. Aside from the data source (e.g., table, movie, speech, text), the extracted information may exist as feature vectors for each data point or in the form of (pairwise) relationships between data points [3]. Clustering algorithms can be model-based or cost-based. In the former case, a latent variable in a probabilistic data model determines the cluster label. In the latter case, the clustering algorithm aims to minimise the cost function. Cluster algorithms can also be classified as parametric or nonparametric[64]. In the parametric approach, the number of clusters K is required as input, necessitating the identification of the exact K or an estimate of K in advance. In contrast, the nonparametric regime allows for significant cluster size variations. Nonparametric clustering algorithms aim to distinguish small clusters from outlier data points, where outliers can be single data points or small clusters.

2.5.1 Centroid-based clustering

Centroid-based clustering algorithms represent some of the earliest computational clustering methods, with K-means being the most well-known. In centroid-based clustering, the number of clusters K is predetermined and known. The objective is to find the most optimal grouping of all objects in the data space, which is computationally challenging. MacQueen introduced the widely-used K-means algorithm in 1967 [96]. K-means groups data points from dataset D by optimising a specific objective function. Each group must include at least one data point, and each data point belongs to only one group. The algo-

rithm refines the clusters iteratively. It begins by selecting K points as initial centroids. It assigns each data point in D to the closest centroids using a proximity measure, typically the Euclidean distance. The centroids are then updated, and the process continues until the centroids remain unchanged or a predefined convergence criterion is met. K-means is guaranteed to converge to a local minimum, but minimising the objective function is an nondeterministic polynomial (NP)-hard problem. It is challenging to find optimal solutions to NP-hard problems in polynomial time [35, 161]. It is, therefore, common to use a less stringent rule, like stopping the process if only 1 % or less of all data points in D change their cluster membership [3].

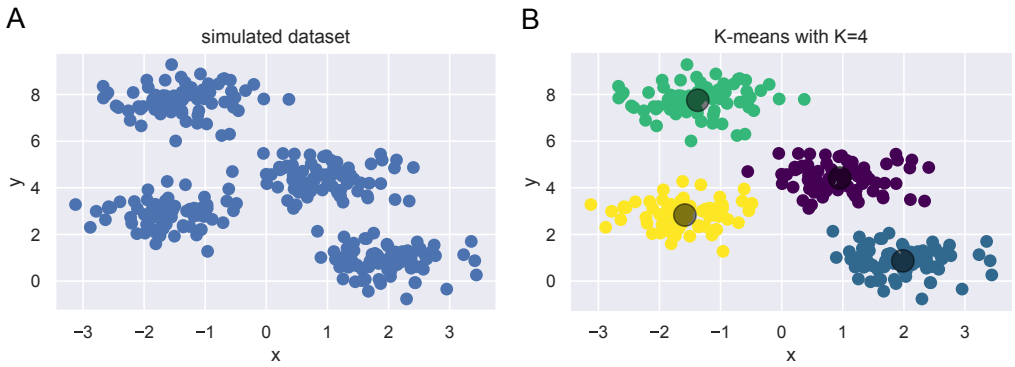

Figure 2.7: K-means clustering. Clustering with K-means requires the number of clusters K as input. A simulation of clusters (A) is partitioned using K-means with K=4 (B). The colours identify the clusters, while the dark grey dots represent the corresponding centroids.

In K-means, the objective function is thesum of squared errors (SSE). Dataset D is partitioned into a set of clusters $C = \{C_1, C_2, \ldots, C_k, \ldots, C_K\}$, where the centroid c_k belongs to the cluster C_k, as shown in Figure 2.7. The SSE of the cluster set C is defined as follows [3]:

$$\mathrm{SSE}(C) = \sum_{k=1}^{K} \sum_{x_i \in C_k} \|x_i - c_k\|^2 \tag{2.3}$$

with

$$c_k = \sum_{x_i \in C_k} \frac{x_i}{|C_k|} \quad . \tag{2.4}$$

Throughout the iterative process, the SSE consistently decreases, eventually reaching a local minimum. K-means tends to produce convex clusters, which may limit its suitability for spatial data featuring naturally elongated or non-linear configurations. In such cases, K-means may merge non-convex clusters or break them apart [3, 64].

2.5.2 Hierarchical clustering

In hierarchical clustering, a tree structure known as a dendrogram is constructed through a sequence of irreversible algorithmic steps. The dendrogram is divided into $n-1$ partitions of varying granularity, encompassing n data points in D [3]. Typically, agglomeration hierarchical clustering algorithms start with individual data points as clusters and merge or agglomerate them at each step [64]. However, hierarchical clustering can also be executed in a top-down (divisive) manner, where the data points are partitioned into a tree-like structure, starting with the entire dataset as a cluster. Subsequently, the selected flat algorithm calculates cluster sizes at each step.

2.5.3 Density-based clustering

Density-based clustering algorithms use data density as a parameter for identifying clusters. Consequently, these methods are nonparametric because they do not rely on assumptions about the number of clusters in the dataset D or the cluster distribution in D [64]. These algorithms are excellently suited to handling data points in continuous space. Clusters are regions where high-density data points are distributed, separated by sparsely distributed data points. Flexible cluster recognition allows diverse cluster shapes and is compatible with noise and outliers [3].

Ester et al. introduced Density-Based Spatial Clustering of Applications with Noise (DBSCAN) in 1996 [38]. DBSCAN identifies clusters of arbitrary shapes based on density, eliminating the need to determine the value of K before computation. The data points in dataset D can be categorised into three types: core points with a dense neighbourhood, border points belonging to a cluster lacking a dense neighbourhood, and noise points that do not belong to any cluster [3]. In DBSCAN, density is estimated by the number

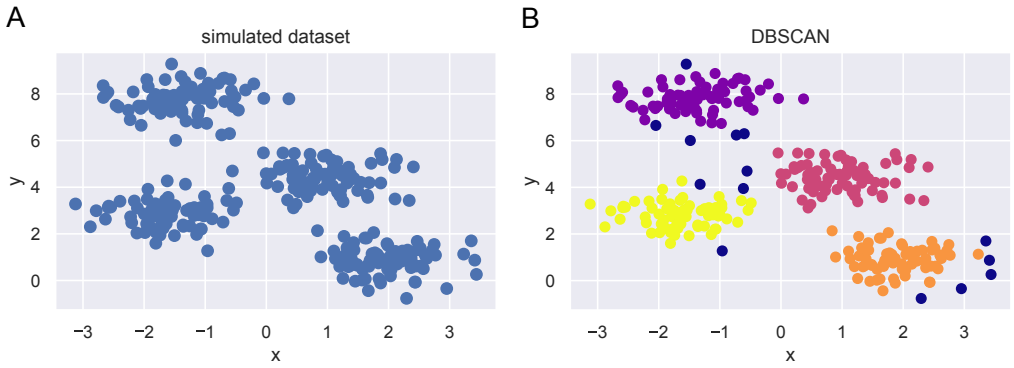

Figure 2.8: DBSCAN clustering. DBSCAN clustering is based on density estimation. A simulation of clusters (A) undergoes clustering with DBSCAN using $\epsilon = 0.5$ and a minimum of 4 data points in each cluster (B). Data points identified as noise are depicted in dark blue, while the other colours represent clusters.

of points within a fixed-radius neighbourhood around a data point p. Two points are considered connected when they are close. A DBSCAN cluster consists of core points with a neighbourhood area of radius ϵ and a minimum number of points.

Additionally, another data point q is directly density-reachable from the core point p if it is within ϵ distance of p. When p and q are density-reachable from another data point x_i in D, they are considered density-connected. Consequently, a cluster is defined as a set of density-connected points that are maximal in density reachability. In contrast, noise represents a set of points that do not belong to any cluster [3], as shown in Figure 2.8.

2.5.4 Ripley's K-, L-, and H-function

Clustering dense datasets is well suited to DBSCAN. However, a challenge remains in distinguishing between datasets with small clusters in noisy data or with random distribution of data points. Developed in 1977, Ripley's K-function compares a given distribution to a random distribution and tests the null hypothesis that the data points of the given distribution are randomly distributed [127]. Ripley's K-function is defined as follows [78]:

$$\mathrm{K}(r) = \frac{1}{n} \sum_{i=1}^{n} \frac{N_{p_i}(r)}{\rho} \quad , \tag{2.5}$$

where n denotes the total number of points, p_i represents the i-th point, ρ indicates the number of points per area, and N is the expected number of points within the distance r.

For random distributions, the expected distribution follows the random Poisson distribution with $\mathrm{K}(r) = \pi r^2$. The K-function can be normalised to

$$\mathrm{L}(r) = \sqrt{\frac{\mathrm{K}(r)}{\pi}} \quad , \tag{2.6}$$

so that $L(r) = r$ for the random Poisson distribution. Normalisation with

$$\mathrm{H}(r) = \mathrm{L}(r) - r \tag{2.7}$$

results in $H(r) = 0$ for the random Poisson distribution. For a given distribution, $H(r)$ greater than zero indicates clustering, while $H(r)$ smaller than zero indicates dispersion. In clustering, the maximum of $H(r)$ measures the domain radius, r_D, and the r-value represents the radius of maximal aggregation. It has been demonstrated that the radii do not necessarily have to be identical ($r_D \leq r \leq 2r_D$) [78].

Ripley's K-function and its normalisations L- and H-functions test whether the chosen dataset D or a subset thereof is a random distribution. Ripley's K-based clustering was developed to assign cluster labels to data points. Getis and Franklin express spatial point distribution as follows [45, 52, 130]:

$$\hat{\mathrm{K}}(r) = A \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\delta_{ij}}{n^2} \quad , \tag{2.8}$$

where δ_{ij} represents the distance r_{ij} between two points i and j, and if $r_{ij} < r$, then $\delta_{ij} = 1$; otherwise, $\delta_{ij} = 1$. The L-function for individual points is then given by

$$\mathrm{L}_i(r) = \sqrt{A \sum_{j=1}^{n} \frac{\delta_{ij}}{\pi(n-1)}} \quad . \tag{2.9}$$

The L-function assigns data points with an L-value below a threshold to the background. Data points are defined as clusters when their L-values exceed the threshold, and they

lie within $2r$ of each other [130].

2.5.5 Persistence-based clustering

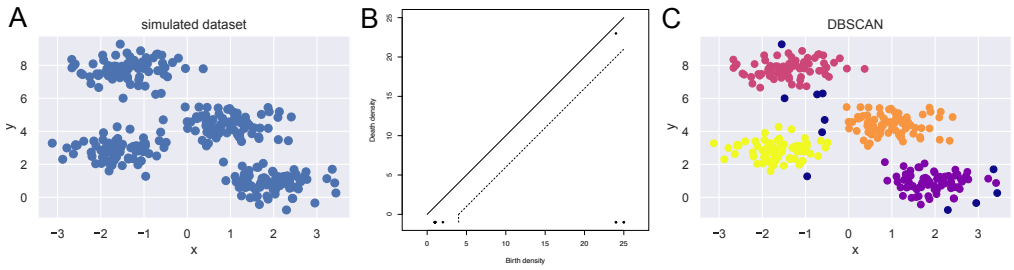

Figure 2.9: ToMATo clustering. ToMATo clustering is based on density estimation and hierarchical clustering. A simulation of clusters (A) undergoes clustering with ToMATo using a radius of 0.5 for density estimation and $\tau = 4$. The persistence diagram shows the birth and death dnsities for each detected peak (B). In the final cluster plot (C), data points identified as noise are depicted in dark blue, while the other colours represent clusters.

Topological Data Analysis (TDA) is a robust mathematical framework that examines a point cloud to reveal its topology or shape [113, 122]. In 2013, Chazal et al. introduced a persistence-based clustering method, Topological Mode Analysis Tool (ToMATo) [25]. Unlike DBSCAN, ToMATo uses a hierarchical approach and a density-based method to detect clusters of arbitrary shapes. The algorithm relies solely on a density estimate of dataset D and the pairwise distance between data points. As a result, it can be applied to any metric space $\mathbb{R}^d$ [26]. Figure 2.9 displays ToMATo clustering for a dataset with four clusters.

The persistence-based clustering algorithm starts with computing a density estimate f of dataset D. Local peaks correspond to cluster centres. Each data point in D is then assigned to a local peak by following the gradient vector field of f. The result is a collection of clusters or density modes. The maximum density of each mode is called birth density, f_b. The saddle point connecting a cluster with a neighbouring one represents its death density, f_d. Hence, every cluster has persistence

$$P = f_b - f_d \quad . \tag{2.10}$$

Cluster birth and death densities (f_b, f_d) can be visualised using persistence diagrams,

as shown in Figure 2.9 B. Cluster persistence is measured by the vertical distance from the point (f_b, f_d) to the diagonal. The persistence threshold, τ, distinguishes relevant peaks from irrelevant ones. The irrelevant peaks are located between the diagonal and the threshold line. A cluster with a persistence $P < \tau$ is merged with a neighbouring cluster with $P \geq \tau$. If no neighbouring cluster with $P \geq \tau$ exists, the cluster is classified as noise.[24, 25, 117]

2.6 Bayesian Theory

The Bayesian framework is a model-based approach to statistics. As part of its analysis, it incorporates prior information, which refers to information gathered before an experiment. It applies this prior information to the statistical problem at hand. In other words, Bayesian Data Analysis (BDA) quantifies possible events using probability functions [79]. The Bayesian theorem for conditional probabilities states that the

$$posterior(y|x) \propto likelihood(y|x) \cdot prior(y) \quad , \tag{2.11}$$

where y represents the unknown parameter, and x is the observed data from the current experiment [146]. The right-hand side of the equation consists of two components. The first is the likelihood function, $likelihood(y|x)$, which represents the sampling distribution for the parameter, excluding constant factors. Based on the observed data and proposed statistical models, it serves as the primary tool for assessing the plausibility of a given parameter value.

The second component is the prior distribution, $prior(y)$. It represents existing knowledge about parameter y before the current experiment. The prior distribution includes at least one tunable hyperparameter, the concentration parameter. The Gamma distribution is often chosen as the prior distribution due to its versatility and flexibility. The left-hand side of the equation presents the posterior distribution, $posterior(y|x)$. This distribution provides a statistical summary of the unknown parameter, drawing from past and current observations. Inference refers to the process of calculating the posterior probability [146].

Thus, BDA begins with determining a suitable likelihood function and a prior distribution for the statistical model. Both of these steps may require subjective information selection [79]. BDA can be divided into two primary sections: parametric and nonparametric Bayesian. In the parametric case, the models are well-defined and comprise a finite number of parameters. Nonparametric models are constructed over a broad parameter space. Rather than a fixed number of unknown parameters, the prior distribution encompasses the entire parameter space [64, 153]. Unlike parametric Bayesian data analysis, nonparametric Bayesian analysis cannot describe the prior distribution model in a closed form with fixed parameters [64, 79]. Instead, the prior distribution employs hyperparameters to select a suitable distribution from the continuous space of prior distributions. The Dirichlet process, introduced by Ferguson in 1973, is a reasonable choice in nonparametric Bayesian Data Analysis as it selects a discrete distribution [39].

2.6.1 Bayesian data analysis for clustering

Clustering methods based on Bayesian theory belong to the class of model-based clustering approaches [64]. After selecting the likelihood function and prior distribution, various posterior inference methods can be applied to characterise the observed clustering parameters and outcomes. Markov chain Monte Carlo simulations are commonly employed to refine unknown model parameters [64]. Alternatively, a deterministic approach utilises BDA to assess the results of clustering data using the same clustering algorithm (e.g. DBSCAN, K-means, Ripley's K-based, ToMATo) but with different input parameters [130].

2.6.2 Bayesian model based clustering for SMLM

A subjective choice of parameters is common in SMLM clustering [79]. Furthermore, there is often only a focus on coordinate values, disregarding localisation precision or other features stored in the localisation table [108]. Thus, the BDA approach can serve as a general mechanism to ascertain the optimal cluster result by considering the proposed setting, model, and algorithm and integrating supplementary information acquired dur-

ing the localisation process into the cluster analysis.

In this thesis, the BDA clustering approach and code are based on work by Rubin-Delanchy et al., published in 2015 [130]. Their work proposes employing the BDA approach to incorporate localisation precision into the analysis, thereby allowing for considering localisation inaccuracies and evaluating cluster results against a Bayesian model.

The general algorithm for this approach is as follows [130]:

1. Select a parameter range for each parameter of the selected clustering algorithm.
2. Generate cluster proposals for all possible parameter sets.
3. Evaluate cluster proposals using the Bayesian model.
4. Select the best proposal as the final cluster result.

The BDA process begins with selecting a suitable model. This thesis focuses primarily on membrane-bound proteins. Therefore, molecular positions can be represented as 2D coordinate positions within a single plane. The following paragraphs explain the Bayesian framework introduced by Rubin-Delanchy et al. [130]. The dataset $D = [x_1, x_2, ..., x_N]$ is associated with the variances $s_1^2, s_2^2, ..., s_N^2$. All data points are assumed to be independent, unbiased, and circular. The actual, unknown molecular coordinates, represented as $Z = [Z_1, Z_2, \ldots, Z_N]$, are perturbed by a Gaussian distribution $x_i = \mathcal{N}(Z_i, s_i^2 I_2)$. Here, $\mathcal{N}(\mu, \Sigma)$ denotes the Gaussian distribution characterised by a mean μ and a covariance Σ, with I_2 representing the identity matrix in the 2D plane.

The parameter of interest within BDA is the labelling $L = [l_1, l_2, \ldots, l_N]$ of the complete dataset D. D can be categorised into n subsets $C = [C_0, \ldots C_k, \ldots C_n]$, where C_0 represents the noise or background. The label $l_i = k$ is assigned to a data point x_i if it is part of a cluster C_k. According to Green and Richards [51], the prior probability $p(l)$ for the labelling of dataset D is defined as

$$p(l) = p_B^{n_0}(1-p_B)^{N-n_0}\frac{\alpha^m \Gamma(\alpha)\prod_{k=1}^{m}\Gamma(n_k}{\Gamma(\alpha+N-n_0)} \quad , \tag{2.12}$$

where p_B signifies the fixed probability for each data point to belong to the background, m is the count of cluster labels different from 0, and n_0 is the number of data points in the background subset C_0.

Cluster centres are uniformly spread across the region of interest. Additionally, background localisations are randomly distributed. In contrast, the cluster positions follow a Gaussian distribution with an unspecified mean μ and variance σ:

$$Z_i \propto N(\mu_k, \sigma_k^2 I_2) \quad , \quad i \in C_k \quad . \tag{2.13}$$

The variances σ_k are provided as a histogram, which the algorithm transforms into a continuous probability density.

In the Bayesian inference stage, the labelling is assigned a posterior probability

$$p(l|D) \propto p(l) \left[\prod_{i \in C_0} p_0(x_i) \prod_{k=1}^{m} \int p(\theta_k) \prod_{i \in C_k} p(x_i|\theta_k) d\theta_k \right] \quad , \tag{2.14}$$

where $\theta_k = (\mu_k, \sigma_k)$ signifies the parameter pair of cluster C_k, $p(\theta_k)$ stands for the prior probability of cluster C_k, $p(x_i|\theta_k)$ denotes the probability of data point x_i belonging to cluster C_k, and $p_0(x_i)$ represents the background probability for data point x_i. The integral is numerically estimated in an arbitrary rectangular subset R of $\mathbb{R}^2$ aligned to the axes. As is customary in facilitating BDA computation, the logarithm of the posterior probability is calculated as the final score for a cluster proposal. The cluster proposal with the highest score is selected as the final result [130].

3

Material and Methods

The following chapter lists and explains the microscopy setups and the material used for conducting the experiments. Furthermore, technical concepts of the computational work are presented.

3.1 Material

The material used for the experiments are listed in the following. For facilitate comprehension they are organised regarding their respective experiments.

3.1.1 Cell lines

Three different cell lines were used for the experiments: HeLa, CV-1, and T98G. HeLa cells are human cervical cancer cells that were retrieved from Henrietta Lacks in 1951 [84]). The cells have a stable transfection with clathrin light chain (CLC)-GFP. The CV-1 cell line is fibroblast kidney cells derived from a male adult African green monkey (**citation**). The T98G cells are patient-derived fibroblast brain cells. They are Glioblastoma Multiforme cells of a white adult human, and they have a hyperpentaploid chromosome count [145]. Furthermore, they are resistant to CD95 induced apoptosis [134]. The T98G cells were infected with a lentiviral construct to transfect the cells with the CD95-GFP plasmid or a transmembrane mutant. The cells were then sorted via Fluorescence Activated Cell Sorting (FACS) for the stably transfected clones.

3.1.2 Cell culture

The three cell lines described in section 3.1.1 were cultivated at 37 °C with 5 % CO_2, but with different media. The CV-1 cells were cultured in standard medium consisting of DMEM medium with 10 % Fetal Bovine Serum (FBS) and 1 % GlutaMax. For the HeLa CLC-GFP cells, the standard medium was supplemented with 1 % Penicillin-Streptomycin. The standard medium was supplemented with 1 % sodium pyruvate. All cells were regularly tested for mycoplasma [166] and only used when the test was negative. Table 3.1 shows the various materials used in the cell culture, the producing company. The protocols for cultivation, and splitting for all cell lines, and the transfection of CV-1 cells with Glycosylphosphatidylinositol-GFP are detailed in the appendix A.1.

Name	**Company**
Tissue Culture Testplate 6	TPP Techno Plastic Products AG
Opti-MeM (1X)	Gibco by Life Technologies GmbH
DMEM medium (1X)	Gibco by Life Technologies GmbH
Fetal Bovine Serum (FBS)	Thermo Fisher Scientific
GlutaMax (100X)	Gibco by Life Technologies GmbH
Trypsin-EDTA	Life Technologies GmbH
Penicillin-Streptomycin	Sigma-Aldrich
Phosphate Buffered Saline (PBS)	Sigma-Aldrich and Th. Greyer GmbH & Co. KG
Lipofectamine 3000	Invitrogen/Thermo Fischer Scientific

Table 3.1: Material for the cultivation of the cell line and work in the cell culture

3.1.3 Glass cleaning

Cells were seeded on either round or rectangular #1.5 thickness coverglass. The coverglass was first cleaned in an ultrasound bath in 2 % Hellmanex III for 20 min, rinsed extensively in Milli-Q water and afterwards sonicated further in 70 % ethanol. After drying the glasses, they were plasma cleaned for 30 min. The detailed the protocol is in the appendix A.2.1 and the material in table 3.2.

3.1.4 Supported lipid bilayer

Supported lipid bilayers were prepared for experiments with tethered cells. First, the lipids were mixed and unilamellar vesicles formed by sonication. Then, nice bilayers

Name	Company
round cover glass (18 mm diameter, #1.5)	VWR
High Precision Microcope Cover Glasses (25mm x 75mm, #1.5, 170 μm +/- 5 μm)	Paul Marienfeld GmbH & Co. KG
Hellmanex III	Hellma
Ethanol	VWR
MilliQ water filter	Merck
plasma cleaner	Diener electronic Plasma-Surface-Technology
ultrasound bath	Bandelin Sonorex

Table 3.2: Material used for the cleaning of the glass cover glasses

were formed out of them on clean glass slides. The detailed protocol is in appendix A.3.2 and the materials in table 3.3.

Name	Company
Bovine Serum Albumin (BSA)	Sigma-Aldrich
CD95L-T4-biotin	Apogenix AG
Trichloromethane/chloroform	Carl Roth
methanol	Th. Greyer GmbH & Co. KG
ultra-sound tip	Bandelin Sonoplus
18:1 Biotinyl PE 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(biotinyl) (sodium salt)	Avanti polar lipids/ Croda International Plc
18:1 PE CF 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-(carboxyfluorescein) (ammonium salt)	Avanti polar lipids/ Croda International Plc
16:0-18:1 PC (POPC) 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine	Avanti polar lipids/ Croda International Plc
18:1 PEG5000 PE 1,2-dioleoyl-*sn*-glycero-3-phosphoethanolamine-*N*-[methoxy(polyethylene glycol)-5000] (ammonium salt)	Avanti polar lipids/ Croda International Plc
diaphragm vacuum pump	vacuumbrand GmbH & Co
Avidin, NeutrAvidin	Invitrogen/Thermo Fischer Scientific

Table 3.3: Materials used for the supported lipid bilayer

3.1.5 Cell tethering

Cells were tethered to the supported lipid bilayer via DNA strands. On the one hand, a DNA strand was added to the bilayer via cholesterol. On the other hand, another strand was covalently coupled by a Halo tag to its ligand expressed by the cell and coupled to GPI. Incubating the cells on the bilayer tethers the cells to the bilayer. The protocols are in the appendix A.4 and the materials in table 3.4.

Name	**Company**
Live Cell Imaging Solution	Invitrogen/Thermo Fischer Scientific
Halo-tag-100-mer-DNA	biomers.net GmbH
cholesterol-20-mer-DNA	Eurogentec S.A.
Halo-JF549	Group of Professor Luke Lavis [56]
Centrifuge 5418R and 5417R	eppendorf
Thermomixer comfort	eppendorf
sticky-Slide VI 0.4	ibidi

Table 3.4: The material used for the cell tethering and their control experiments.

3.1.6 Cell staining

The cell lines used in the experiments express a protein of interest at the cell membrane coupled to GFP. Thus, the staining procedures are similar. They only vary depending on the orientation of the GFP (intracellular or extracellular) and whether the cells were tethered to a supported lipid bilayer or seeded directly on the glass slides. In all cases, the cells were fixed and quenched before and after the staining process. Permeabilisation and blocking depended on the kind of experiment. The detailed protocols are in the appendix A.5.

3.2 Setups

Microscopy experiments were performed using two microscopes. High-resolution SMLM experiments were carried out using a single-molecule localisation microscopy setup, while experiments for prolonged observation of living cells were performed on a spinning disk microscope.

Name	Company
Paraformaldehyde (PFA)	thermo scientific
Glutaraldehyde (GA)	Electron Microscopy Science
Bovine Serum Albumin (BSA)	Sigma-Aldrich
Sodium borohydride ($NaBH_4$)	Sigma-Aldrich
Ammonium chloride (NH_4Cl)	Carl Roth
Phosphate Buffered Saline (PBS)	Sigma-Aldrich
PIPES	Carl Roth
Magnesium sulfate ($MgSO_4$)	Th. Greyer GmbH
EGTA	abcr GmbH
Saponin	Carl Roth
ImageIT	Invitrogen/Thermo Fischer Scientific
Goat serum	Th. Greyer GmbH & Co. KG
FluoTag-Q anti-GFP labeled 1:1 with Alexa Fluor 647 (AF647)	NanoTag Biotechnologies GmbH

Table 3.5: The material used for staining living cells or fixing and then staining them.

3.2.1 Spinning disk microscope

The spinning disk microscope employed in this study was an inverted Olympus IX71 equipped with a CSU-X1 spinning disk system from Yokogawa. A 60x 1.42 NA oil objective was mounted on the microscope. Laser wavelengths and power outputs employed for illumination included a 491 nm laser from Cobolt with a power output of 100 mW, a 561 nm laser from Cobolt with a power output of 100 mW, and a 645 nm laser from Melles Griot with a power output of 500 mW. The entire imaging process was managed and regulated using MetaMorph software. A Hamamatsu EMCCD camera was used to acquire the images with a final pixel size in the sample of 91 nm.

3.2.2 Single-molecule localisation microscopy setup

A home build Total Internal Reflection Fluorescence (TIRF) microscope was used for all SMLM experiments. The microscope was first described in [4]. The TIRF microscope has four laser lines: 405 nm, 473 nm, 556 nm, and 639 nm. The lasers are controlled via an Acousto-Optic Tunable Filter (AOTF) (AOTFnc-400-650-TN S/N: 1001/350409, AA Opto Electronic). The images are acquired with a water-cooled and back-illuminated Photometrics EMCCD camera. The camera has 512×512 pixel with a pixel size of $16 \times 16\,\mu m$. With the Olympus 60x, 1.49 NA back focal plane TIRF objective and imaging

oil, the final pixel size in the sample is 96 nm. Before the acquisition, the EMCCD camera was calibrated. The microscope is controlled via a Micro-Manager 1.4.23 and a LabView 2015 program.

For (*direct*) Stochastic Optical Reconstruction Microscopy ((*d*)STORM) of the dye Alexa Fluor 647 (AF647) [82], the 639 nm laser illuminated the sample at powers of $0.008 - 0.015\,\mathrm{mW}/\mu\mathrm{m}^2$. Typically around 30 000 frames were acquired at an exposure time of $10 - 20$ ms in an imaging buffer of 1.5 % beta-mercaptoethanol and GLOX (2.5 mg/ml glucose oxidase, 0.2 mg/ml catalase, 200 mM Tris-HCl pH 8.0, 50 % glycerol) as imaging buffer (10:1).

Laser line	**Company**	**max. power output (mW)**
639 nm	Changchun New Industries Optoelectronics Tech. Co., Ltd.	700
556 nm	Laserglow technologies	240
473 nm	Laserglow technologies	142.9
405 nm	Cobolt MLD	360

Table 3.6: The laser of the home build TIRF microscope and their origin and maximal power output. The name of the laser is equivalent to its highest intensity wavelength.

3.3 Computational methods

The following section offers a description of two software tools, SMAP and FluoSim. The first one is one of the best-performing localisation platforms currently freely available [132]. The second one provides the possibility to simulate membrane molecules, their dynamics and combines them with fluorescent photophysical properties. Furthermore, this section offers a short presentation of the two programming languages Python and R.

3.3.1 Localisation Software

There exists a large variety of free and commercial software for the analysis of acquired SMLM images. In the 3D and 2D SMLM challenges 2018, which tested submitted tools on simulated and experimental data, the Super-resolution Microscopy Analysis Platform (SMAP) ranked one of the best-performing software packages available [132].

SMAP is a freely available modular software covering all steps of the 2D and 3D SMLM data analysis pipeline. It is written in the commercial programming language MATLAB, but it can also be installed as stand-alone software. SMAP is an analysis platform for 2D and 3D SMLM data with hundreds of plugins and can be extended with different algorithms. In addition, it has a GUI to control individual steps of single-molecule fitting, post-processing, rendering and advanced analysis. SMLM data analysis in SMAP is reproducible because it logs actions and their parameters. [124]

For correct detection and reduction of the fitting-error of SMLM data, SMAP uses camera metadata and other camera-specific settings. For the fitting, SMAP offers the option to use unmodified or custom-engineered PSFs. These PSF are fitted by maximum likelihood estimation with either a Gaussian or a spline-interpolated model. Post-processing of the fitted data can involve drift correction in 2D or 3D, merging localisations from consecutive frames or filtering based on localisation attributes. The rendering step allows the user to zoom in and move the sample in the GUI interactively. In addition, the advanced analysis tool offers options such as the measurement of resolution, cluster analysis or single-particle tracking analysis. [124]

For the analysis of SMLM images taken with the single-molecule localisation microscope setup described in section 3.2.2, the relevant camera and acquisition parameters stored in the metadata file were detected by SMAP. Additionally, the electron multiplier (EM) gain was set to 300, and the conversion factor for analogue to digital units to photons was set to 6.7. For the fitting step, the minimum distance for separate fitting was set to 7 pixels between two localisation candidates. For the PSF, the Gauss differential parameters were defined as sigma $= 1.2$, dynamic factor $= 1.7$, and free PSF with the workflow “set Cam parameters” [82].

3.3.2 Simulator for membrane protein dynamics

FluoSim is a software tool specifically for the dynamics of single membrane proteins. It is designed as a general starting point for simulations for live-cell and super-resolution

techniques[83].

The general procedure is to start with a 2D microscopy image. The regions of interest can then be populated with molecules and parameters for the kinetics, and trapping can be defined. The simulator updates all molecular localisations per time step based on a random number generator. The fluorescence signal of the sample depends on the photophysical characteristics and the labelling density of the chosen fluorescent organic dye or the expression level of the fluorescent protein. As a final step, Poisson-shot noise for the intensity fluctuations for individual fluorescent molecules and a Gaussian readout noise around the camera offset are added in the simulation of the desired output, e.g. for SMLM.[83]

3.3.3 Programming languages

Two of the most popular interpreted programming languages are R and Python [101, 156]. They run on all three standard operating systems (Windows, Apple, Linux) and are open-source [101, 105, 107].

R is a programming language specialised in statistical computing and graphics and is thus used for data analysis and the development of new techniques in statistics [94, 107]. R is a flexible language with an evolved tailored grammar enabling the efficient and concise manipulation of complex data structures [97, 156]. R is continuously improved and maintained by the R Development Core Team, the R Foundation and its community [107, 121]. Additionally to the R standard library, additional packages needed for this work are presented in table 3.7. Except for the RSMLM package, all packages are officially hosted on the The Comprehensive R Archive Network (CRAN) package repository [67] or the Bioconductor Project Package Repository [16]. The R package RSMLM was published by Pike et al. in 2020 [117] and has not been added to any official R repository. A corrected version hosted on the author's own Github repository account "saskiakutz/RSMLM" was used for this work. [70].

Python is one of the most relevant languages for scientific computing, machine learning,

package name	description	repository	reference
pacman	R package management tool for efficient installation and loading of pacakges from various repositories	CRAN	[126]
devtools	R package development tool	CRAN	[159]
tidyverse	collection of R packages for data analysis, including 'ggplot2', 'tibble', 'dplyr' and 'stringr'	CRAN	[158]
ggpubr	additional functionalities for 'ggplot2' creations and adaptation	CRAN	[73]
ggforce	extension to 'ggplot2' for composing advanced plots	CRAN	[116]
igraph	library for graph analysis	CRAN	[31]
RSMLM	library for SMLM data analysis, including ToMATo and DBSCAN	GitHub	[117]
geometry	computation of convex hulls, meshes and surface tesselation	CRAN	[58]
doParallel	parallel computing	CRAN	[30]
data.table	extension to the base data.frame for faster handling of data	CRAN	[37]
plyr	tools for working with data	CRAN	[157]
rhdf5	functionalities for working with hdf5 files in R	Bioconductor	[40]
tictoc	timing and recording for R scripts	CRAN	[68]

Table 3.7: The table shows the R packages used in this work. For each R library, a short description is offered as well as the hosting repository and their reference.

data analysis and general software development [101]. It is slower than compiled languages as Java or C++, but faster than specialised languages like R [97, 101]. Python is advanced and managed by the Python Software Foundation and its community [120]. Packages added to the basic Python modules are shown in table 3.8. All packages used are officially hosted on the Python Package Index (PyPI) [119] and, except for the rpy2 packages, also on the Anaconda repository [5].

Both languages are object-oriented and have interfaces to connect between them and other languages [101, 156]. Furthermore, the R standard library and the Python package Pandas provide dataframes with the same structure and features for tables [21, 101, 156], such as the localisation table for SMLM data. The code was developed in RStudio [129] and PyCharm [71] and the versions of the programming languages were Python 3.8 and R 4.1.0.

package name	description	repository	reference
PyQt5	linking to the Qt v5 cross-platform interface toolkit	PyPI, ANACONDA	[128]
h5py	functionalities for working with HDF5 files in Python	PyPI, ANACONDA	[28]
Pandas	data analysis tools, intrducing the dataframe concept to Python	PyPI, ANACONDA	[102]
rpy2	Python interface to R	PyPI	[43]
NumPy	data analysis tools for array computation	PyPI, ANACONDA	[60]
matplotlib	library for visualizations	PyPI, ANACONDA	[66]
seaborn	extension to 'matplotlib' for static graphics	PyPI, ANACONDA	[155]

Table 3.8: The table shows the Python packages used in this work. For each Python package, a short description is offered as well as the hosting repository and their reference.

3.4 Computational power

Different computer systems contributed to the data analysis and the development of code and software. The following table 3.9 lists their specifications and contributions. All code was developed on an Ubuntu laptop and deployed to high power computing Curta and an Ubuntu Analysis Computer. These systems mainly computed the data clustering because of their more extensive storage, memory, and processor capacities. The Windows computer was only used for Windows specific functionalities, such as correct pathways and links to R libraries.

Computer	Specifications
Curta	High Power Computing • run by ZEDAT at Freie University Berlin • OS: CentOS 7 • Batch-System: Slurm • Nodes: 170x Intel CPUs only, 12x Asus with GPUs • Processors: 2x Xeon Skylake 6130 (2.1 GHz, 16 cores, 22 MB Cache) per node • GPUs: 24x NVIDIA GTX1080Ti • Memory: 3/6/12/24 GB per core (96/192/384/768 GB per node) 103/50/8/4 nodes
Ubuntu Analysis Computer	• OS: 64-bit Ubuntu 20.04.3 LTS • GNOME 3.36.8 • Memory: 125.5 GiB • Processors: 20x Intel® Xeon(R) Silver 4210 CPU @ 2.20 GHZ • Graphics: NVIDIA Corporation TU102 [GeForce RTX 2080 Ti]
Ubuntu laptop computer	• OS: 64-bit Ubuntu 18.04.5 LTS • GNOME 3.28.2 • Memory: 7.7 GiB • Processors: 4x Intel® Core™ i5-6200U CPU @ 2.30 GHz processors • Graphics: Mesa Intel® HD Graphics 520 (SKL GT2)
Windows computer	• OS: 64-bit Windows 10 Home • Memory: 16.0 GB • Processor: x64-based AMD Ryzen 7 1700 Eight-Core Processor @ 3.00 GHz

Table 3.9: The computing resources and their specification. Curta is the central high power computing system of the Freie University Berlin [14]. The Ewers Lab owns the Ubuntu Analysis and the Linux laptop computer. The Windows computer is my personal one.

4

Results

4.1 BaClAva

The subsequent section presents the results of the software solution Bayesian Cluster Analysis and visualization application (BaClAva). The findings demonstrate the BaClAva optimisations alongside the characterisation of three clustering algorithms. Furthermore, SMLM simulations are used to determine the constraints of (*d*)STORM data. Using T98G cells to image and cluster CD95 receptors, BaClAva demonstrates its ability to handle SMLM data acquired from fixed and stained cellular specimens.

4.1.1 The structure of BaClAva

Bayesian Cluster Analysis and visualization application (BaClAva) is a software with a graphical user interface (GUI) designed to bring the advantages of Bayesian cluster analysis to the broader Single-Molecule Localisation Microscopy (SMLM) community. As shown in Fig. 4.1, BaClAva offers three combinable features for designing various workflows. The features Simulation, Clustering, and Filtering rely on different BaClAva modules. The simulation and clustering features are usable independently, depending on the experimental design. Only the third feature relies on the input of another feature and is only usable with data processed with the clustering feature.

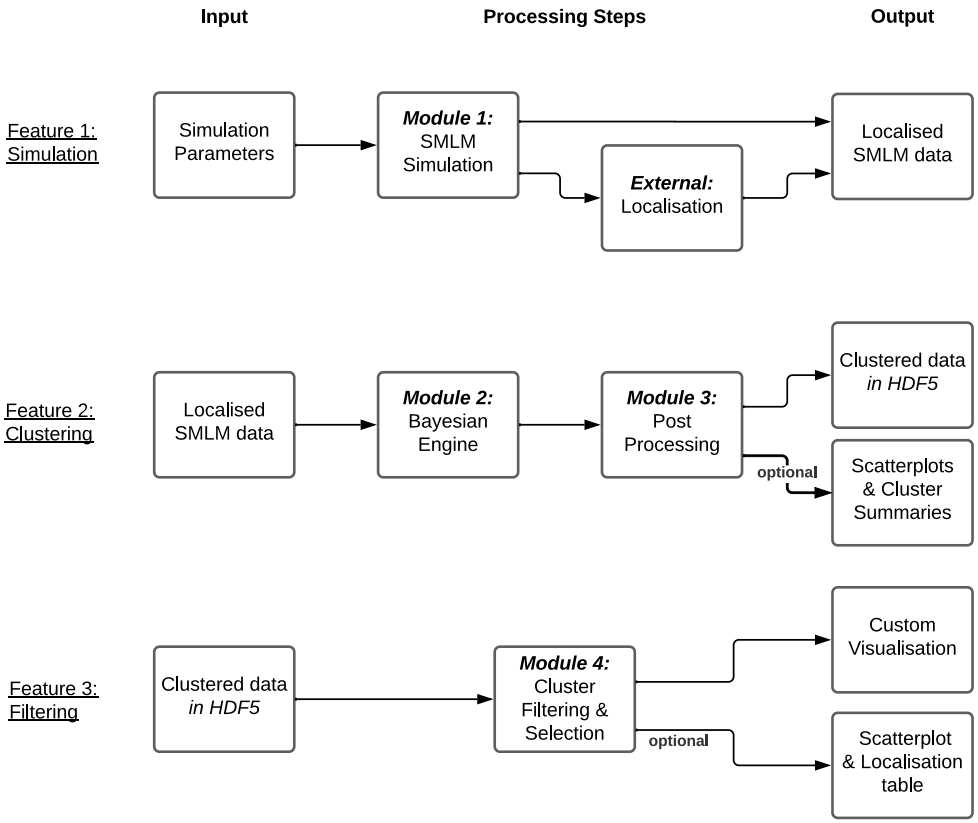

Figure 4.1: Flow diagram illustrating the three features of BaClAva. Each feature includes data input processing steps within BaClAva and data output. External software is required for Feature 1 if a localisation process is necessary after Module 1. Features 1 and 2 operate independently, while Feature 3 relies on the data output from Feature 2.

The software BaClAva consists of four modules for simulation, clustering, and analysis on batch and individual dataset levels. Each module of the software can be used independently via the GUI. Except for the last module, all modules are written in a batch-processing manner, meaning that BaClAva can compute the simulations, clustering or analysis of multiple datasets with a single run. A more in-depth schematic overview of BaClAva and its subroutines is shown in Fig 4.2.

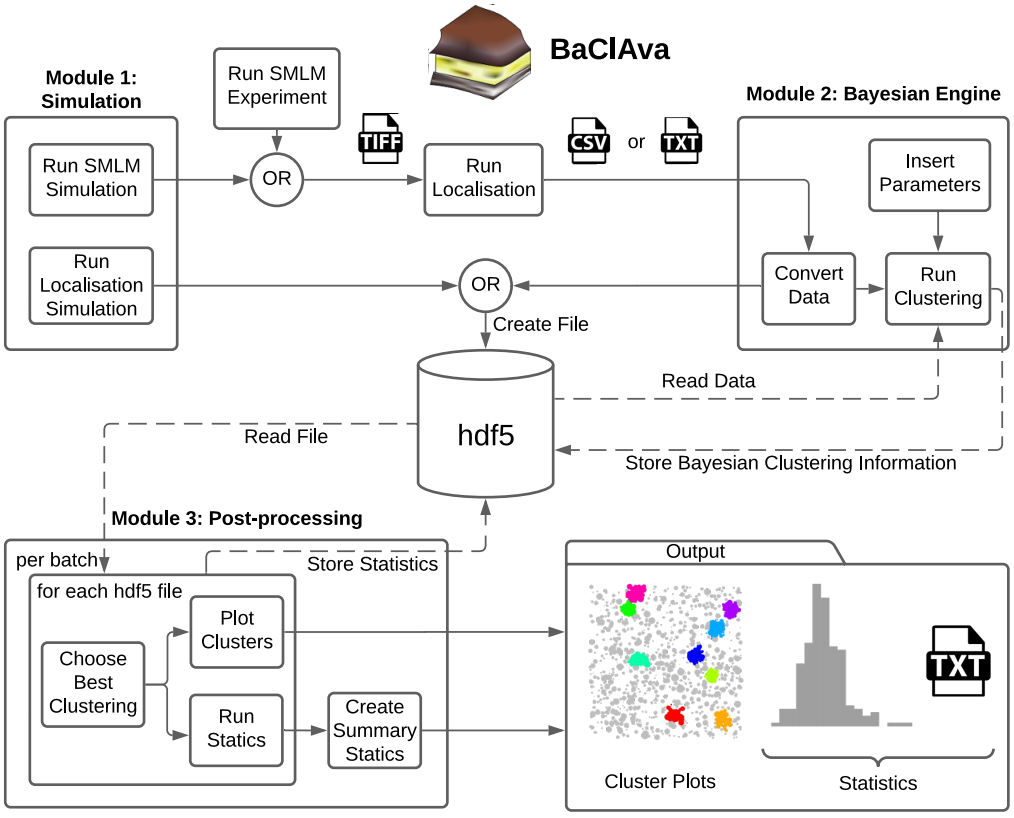

Figure 4.2: Schematic representation of the GUI-based software BaClAva. This diagram illustrates the independent software modules 1-3. Each module can handle single datasets or batches. Depending on the chosen sub-module, the simulation module stores data in TIFF or HDF5 format. Subsequently, the clustering module's Bayesian engine accepts CSV, text, or HDF5 files as input. Finally, the post-processing of the clustering output identifies the optimal clustering and exports scattering plots and statistical results, if specified. This schematic was published in Kutz et al. [82].

The first module, "Bayesian simulation", encompasses two sub-modules, but both aim at simulating SMLM data. Thought experiments are an essential cornerstone for the design of experimental SMLM assay, but especially for data processing-intensive work. Such experiments might offer crucial insights into the experimental setup and the data-processing strategies, such as the labelling density of the dye, resolution limitations and

acquisition length. The first sub-module is similar to FluoSim [83] and was developed by Roman Svetlitckii during his internship and Bachelor thesis. The idea of this module is to simulate proteins that both form clusters and can distribute randomly on the plasma membrane of a cell. In this module, the molecular ground truth positions are first simulated in two steps. First, the program places the user-defined clusters in the field of view without overlapping. The second step adds the background molecules with a selected density level. Afterwards, the SMLM experiment is modelled on top of the ground truth data taking into account experimental blinking statistics of the chosen dye, camera noise, and localisation accuracy. The result is an image stack of blinking fluorophores similar to the output of an SMLM experiment on a microscope and is exported as Tagged Image File Format (TIFF). Further steps of the Bayesian cluster analysis do not function without localisation software like SMAP. This Bayesian simulation sub-module operates alone or alongside microscopy data for the cluster analysis pipeline.

The second sub-module is based on Griffié et al. [54, 130] and offers the option to directly simulate the localisations and their localisation precision without ground truth information. Here, the idea is to simulate perfect PALM data where every molecule is localised only once during the experiment. The original work places the chosen cluster centres randomly without including the positions of the other cluster centres and their size in the field of view. In BaClAva, the cluster centres are at least two standard deviations apart from each other. Therefore, the clusters cannot completely overlap, making them distinguishable by eye. Further input parameters are the number of molecules per cluster and the model for the localisation distribution of an individual cluster. The implemented model for the distribution of localisations is a Gaussian distribution. Thus its standard deviation is the desired radius of the clusters. Additional parameters include the background percentage of localisation and the size of the region of interest (ROI). The localisation precision of the individual localisation is drawn from a gamma distribution. The background localisations' distribution is spatially random but can adapt on the x-axis via the parameters of a beta distribution. In both Bayesian simulation sub-modules, the software informs the user when problems during computation arise. This reporting is essential for the distribution of cluster centres because they have to be separated. Depending on the ROI or cluster size and the random cluster positioning, it might not be

possible to incorporate all clusters.

The second module is the clustering module, named the "Bayesian engine". The input is a single-molecule localisation dataset that has to include the columns for the 2D co-ordinates (X [nm] and Y[nm]) and the localisation precision as the standard deviation (STDEV[nm]). The GUI assists with the correct column selection by presenting the chosen dataset within the GUI in a shortened table format. It shows all columns but only the first five rows. The user can then easily select the necessary columns. Other crucial user selections are the clustering algorithm, whether the analysis should be computed sequential or parallel, and the parameter ranges. The parameter ranges are entered as the minimal, maximal, and step size spanning the parameter space.

The third module is the "Post Processing module". This module's main objective is to select the optimal cluster proposal automatically. Other important cluster analysis parameters, such as the number of clusters, the cluster area and the cluster density, are calculated and exported as two CSV files. The data can also get exported as cluster scatter plots, and the cluster analysis parameters as histograms and density plots. Furthermore, the GUI plotted the data for direct visualisation and manual manipulation.

The fourth module, "Cluster Filtering", visualises clustered and post-processed datasets in a per-dataset manner. The tool offers options to zoom in or out of the plotted dataset, select interesting sites and filter the data based on the cluster density and area. The data visualisation is automatically updated based on the user selections. If wished, the selected region of interest of the dataset could be stored as a cluster scatter plot and a CSV file.

BaClAva has a front-end and a back-end, as it is standard for object-oriented programmed software. The front-end part consists of the code for the GUI. Thus, the code encodes everything for visualisation and interactions with the user. The front-end of BaClAva is written in Python, but for the actual GUI part, some features within the Python code use the C++-syntax because of the GUI library PyQt5 (see table 3.8, [128]). The software has its main GUI class for building the software interface for the user and enables all

communication with the module classes. Each module has its class. Thus, they are independent and do not share any code or information.

The front-end code gathers the input information offered by the user and signals it to the back-end part of BaClAva. In contrast to the front end, the user cannot directly interact with the back-end part from the GUI. Therefore it is hidden from the user. The back-end checks the data for correctness and connects it from Python to the R code. Therefore, translating the Python objects to R objects. All the back-end computations for modules one to three are executed in R. Only module four has a Python-only backbone.

BaClAva presents the different modules as tabs in the GUI, and the user can always switch between them, but only one module can execute back-end computations at a time. The software uses threading enabling the user to interact with BaClAva constantly. After choosing all the necessary inputs, the user starts another threat for the back-end process by pressing the corresponding button. Without threading, the user could not interact with the window of BaClAva during computation. Starting another thread as soon as one thread starts is impossible until the calculation has finished because some computations are memory intensive.

Bayesian parameter selection remained consistent across all analyses. Specifically, a fixed background probability of 50 % and a concentration coefficient of 20 for the Dirichlet process were chosen a priori. The search sequences for determining the optimal cluster parameters were defined in intervals of 5. For instance, the sequence for the radius ranged from 5 to 300, while the threshold sequence spanned from 5 to 500.

4.2 Back-end improvements

Compared to the original code for Bayesian cluster analysis by Griffié et al.[54], BaClAva comprises several changes for faster computation and improved visualisation and storage. The following part presents the significant changes.

4.2.1 Influence of choice of algorithm - Ripley's, DBSCAN, ToMATo

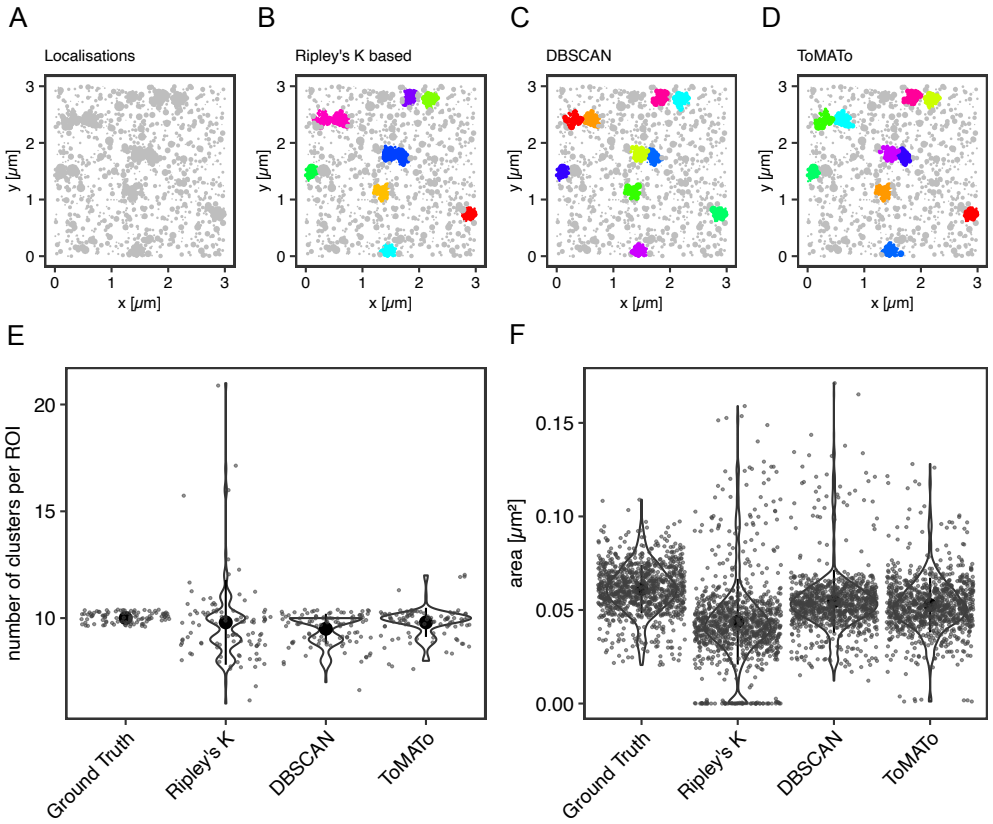

Figure 4.3: Cluster algorithms compared with BaClAva. For a simulated ground truth dataset containing 10 clusters (A), the clustering algorithms Ripley's-K-based (B), DBSCAN (C), and ToMATo (D) generate varying resultant clusters concerning their shapes, dimensions, and cluster counts. The colour assignment to clusters is random and is not comparable across algorithms. For 100 simulations with 10 clusters each, the violin plots illustrate the detected number of clusters (E) with the mean marked by a black circle. The results for Ripley's-K-based clustering were 9.8 ± 2.0, for DBSCAN 9.5 ± 0.7, and 9.8 ± 0.7 for ToMATo. Remarkable results are that the data spread was largest for Ripley's-K-based, and DBSCAN never identified more than 10 clusters. The cluster areas (F) calculated with the post-processing module resulted in $(0.061 \pm 0.013)\,\mu m^2$ for the ground truth data, $(0.044 \pm 0.013)\,\mu m^2$ for Ripley's-K-based, $(0.055 \pm 0.017)\,\mu m^2$ for DBSCAN, and $(0.053 \pm 0.015)\,\mu m^2$ for ToMATo (mean $\pm$ standard deviation). [82]

The initial research provided two clustering algorithms for the Bayesian engine: Ripley's K-based clustering and DBSCAN. BaClAva also includes ToMATo, an additional clustering algorithm sourced from the RSMLM package(table 3.7, [117]). Each of the three algorithms has three input parameters: the 2D coordinates of the dataset, and two cluster parameters, namely radius and threshold. The interpretation of the latter parameter varies among algorithms, as outlined in theory section 2.5.

For benchmarking the three cluster algorithms, 100 simulated images were generated, each with 10 clusters with 100 localisations. The cluster centres were at least 100 nm apart, and the random background density was 111 localisations per μm^2. Therefore, 50 % of all localisations were labelled as unclustered. The coordinates for the individual localisations were drawn from a normal distribution. The cluster centres were the mean with a standard deviation of 50 nm.

The results of the Bayesian analysis with the three cluster algorithms Ripley's-K-based, DBSCAN and ToMATo are shown in Fig 4.3. Fig 4.3 A shows an example of the simulated data without the cluster labels. By eye, it is possible to identify 10 clusters correctly. Fig 4.3 B, C, and D show the clustering scatter plots calculated by the three algorithms. The clusters are randomly colour-coded.

The simulated data encompasses 1000 clusters in total, but Ripley's-K-based clustering and ToMATo identified 980 and DBSCAN 950. On average, DBSCAN and ToMATo could identify 9.5 ± 0.7 and 9.8 ± 0.7 clusters per dataset, respectively. In contrast, the example of Ripley's-K-based clustering shows that it fails to separate adjacent clusters (Fig. 4.3 B). The mean number of clusters per dataset was 9.8 ± 2.0. The standard deviation of 2.0 illustrates the large spread in cluster numbers for Ripley's-K-based clustering. Therefore, the number of clusters and the cluster area were often not correctly identified. This problem was already shown by Pike et al.[117] and is due to Ripley's-K-based clustering not considering the local data point density. The summary violin plots aggregating all 100 simulations in Fig. 4.3 E and F show that DBSCAN and ToMATo can identify the correct number of 10 clusters and the resulting cluster area in most datasets, 61 % and 74 % of the cases, respectively.

However, within these 100 simulations, DBSCAN undercounted fewer than expected clusters in the datasets in 39 % of the cases but did not find more clusters than simulated. Whereas ToMATo also identified more clusters than expected in four and could not correctly identify all ten clusters in 22 cases. Overall, ToMATo had a misclassification rate of 0.26 and DBSCAN one of 0.39. In contrast, Ripley's-K-based clustering identified up to double the number of clusters simulated in 18 cases which is more than four times

more than ToMATo (Fig. 4.3 E).

Furthermore, this implementation of Ripley's-K-based clustering identified less than 10 clusters in 47 % of the samples. Therefore, the misclassification rate for Ripley's-K-based clustering was 0.65. As shown in Fig. 4.3 E, the distributions of cluster numbers were statistically significantly different to a normal distribution. Moreover, a Kruskal-Wallis rank-sum test revealed with a p-value of $8.751 \cdot 10^{-8}$ that the distributions have a significant difference. Finally, a Dunn's test with a Bonferroni adjustment was applied to determine the source of the difference. The Bonferroni correction modifies the family-wise error rate that results from the contribution of Type I errors in each distribution. As a result, the null hypothesis that the given distributions have the same median will be rejected for each p-value $p_i \leq \alpha/m$, where the overall significance level is $\alpha = 0.05$ and the number of hypotheses $m = 2$. The findings demonstrate that the group-wise comparisons between DBSCAN and Ripley's-K-based clustering, ground truth data and ToMATo, and Ripley's-K-based clustering and ToMATo are statistically significant; consequently, the null hypothesis cannot be dismissed in these circumstances. The null hypothesis is rejected according to the DBSCAN and ToMATo test result $p_{adjusted} = *$. The medians are, therefore, probably different. Furthermore, the DBSCAN and Ripley's-K-based clustering distributions have a $p_{adjusted} = ****$, providing a solid case for rejecting the null hypothesis.

The misclassification of data points results in cluster areas that differ from the simulated cluster areas. The average size of the ground truth data is $(0.061 \pm 0.013)\,\mu m^2$, but the cluster algorithms' findings for Ripley's-K-based clustering, DBSCAN, and ToMATo are $(0.044 \pm 0.023)\,\mu m^2$, $(0.055 \pm 0.017)\,\mu m^2$, and $(0.053 \pm 0.015)\,\mu m^2$, respectively. Plotting the data in a q-q-plot and using Shapiro-Wilk's method show that the area distributions are significantly different from a normal distribution. Furthermore, in the results for Ripley's-K-based clustering, 34 clusters have a cluster size of zero.

4.2.2 Sequential vs parallel computation

Even though Bayesian model-based clustering enhances the quality of clustering results in SMLM data, it is computationally more costly than a single cluster analysis with one

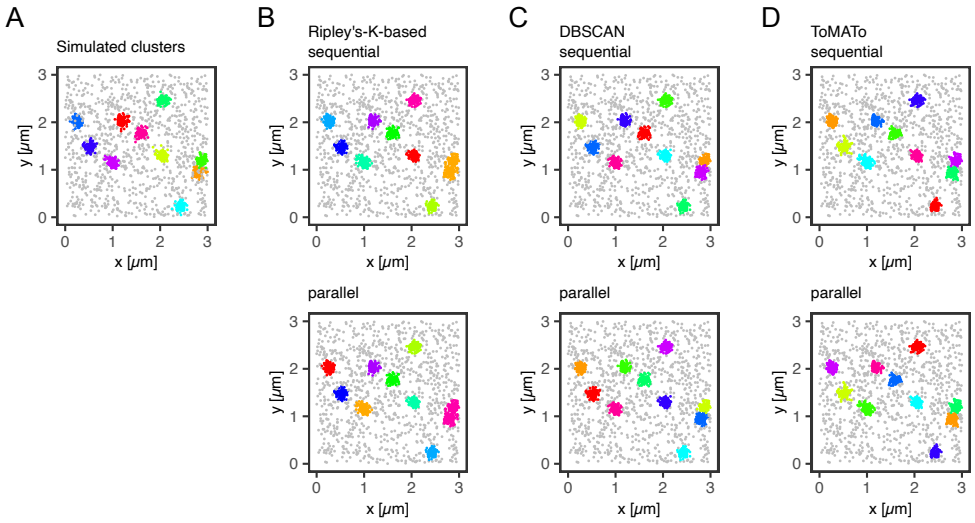

Figure 4.4: Verification of computation on a single dataset. A simulated dataset with ten clusters, each consisting of 100 data points and 50 % of all data points in the background (A), was clustered by Ripley's-K-based clustering (B), DBSCAN (C), and ToMATo (D). Clustering was carried out in both sequential (top) and parallel (bottom) modes for each clustering algorithm. The clustering outcomes for one method are unaffected by the selected computation. However the results of the other clustering algorithms differ. The colour assignment was arbitrary.

parameter tuple. Within BaClAva, the cluster proposal computation and the scoring process runs through all possible value combinations of the cluster parameter sequences. Programmatically, the original procedure consists of two nested for-loops. Even though the calculations of the proposals are independent of the others, they are computed sequentially on a single core. Therefore, it is a problem well-suited for parallelisation with a multi-core CPU. For example, the Ubuntu laptop computer (details: table 3.9) can calculate four cluster proposals in parallel with four available cores. The cores can start the subsequent computation independently from the other cores' calculations until all data is processed.

Since the code is slightly different for sequential and parallel computation, it must be established that the results for a cluster implementation are independent of the kind of computation. Out of the ten simulations with ten clusters each, 100 molecules per cluster and 50 % of all the molecules in the background, Fig. 4.4 shows one example. The clusters, except for their randomly assigned colours, look identical for sequential and parallel computation but can differ for different algorithms, as shown with the marked cluster. Comparing the post-processing output showed that the number of clusters per

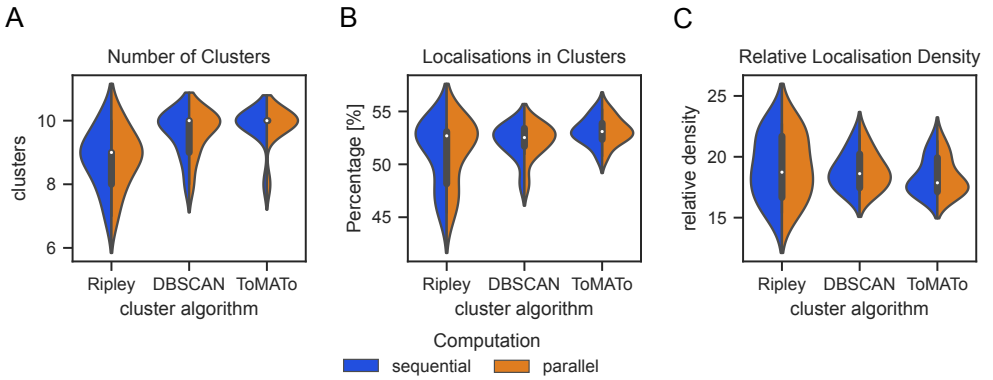

Figure 4.5: Comparison of cluster statistics on a per-dataset level. The half violin plots for 100 datasets with 10 clusters compare sequential and parallel computation for the number of clusters (A), the percentage of localisations in clusters (B), and the relative localisation density (clusters/background) (C). The results are identical for a given algorithm but differ for the three algorithms.

ROI, the percentage of localisations in clusters and the relative density (localisation in clusters/localisation outside of clusters) are identical for parallel and sequential computation (Fig. 4.5). Comparing the data on a cluster basis shows the same behaviour for the number of localisations per cluster, the cluster area and the cluster density (Fig. 4.6). In conclusion, the parallel and sequential computation results are identical for a given cluster algorithm.

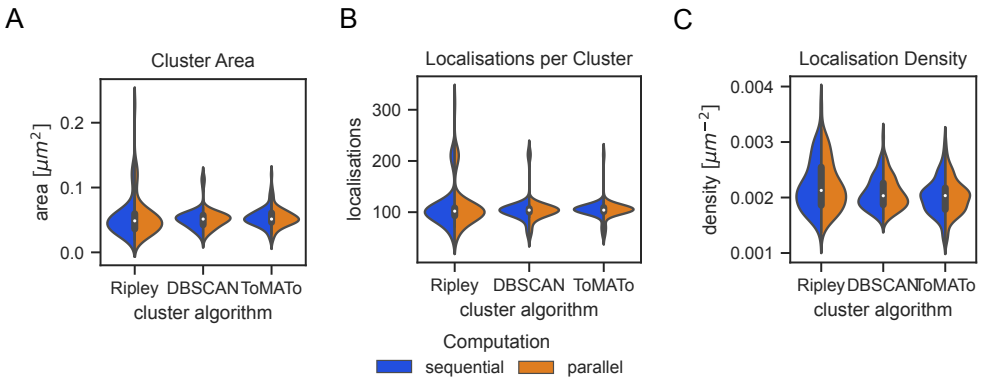

Figure 4.6: Comparison of cluster statistics on a per-cluster level. The half violin plots for 100 datasets with 10 clusters compare sequential and parallel computation for the area (A), the localisations (B), and the localisation density (C) per cluster. The per-cluster data is independent of the chosen computation mode but depends on the selected algorithm.

As shown in Fig. 4.7, parallelising the proposal computation reduces the computational runtime for all three algorithms on ten simulated datasets, as explained above. Analysing the results obtained in a sequential computation resulted in runtimes of

(25.78 ± 0.86) min and (28.45 ± 0.78) min for Ripley's-K-based clustering and DBSCAN, respectively (mean ± standard deviation). The ToMATo implementation had a runtime of (23.87 ± 0.23) min. In the parallel mode, the runtimes resulted in (10.41 ± 0.23) min and (11.90 ± 0.27) min for Ripley's-K-based clustering and DBSCAN, decreasing the runtime by 60 %. The computational time for clustering with ToMATo was (3.062 ± 0.720) min, which decreased by one magnitude. In summary, paralleling the clustering and scoring significantly reduces the computational time for the Bayesian cluster analysis in module 2.

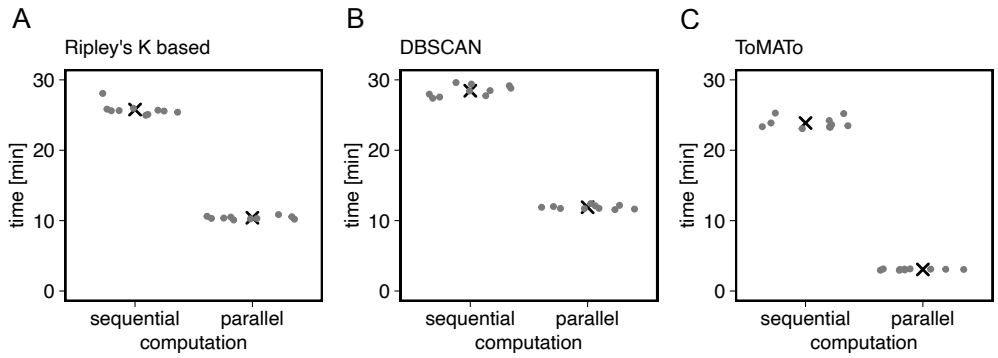

Figure 4.7: Cluster algorithms of BaClAva compared for computational speed. The computation time for sequential and parallel mode in the Bayesian engine module for (A) Ripley's-K-based, (B) DBSCAN, and (C) ToMATo clustering on a 4-cores computer with 3-cores selected as worker nodes. A black cross marks the mean over ten simulations. [82]

4.2.3 Data storage formats

In the original work by Griffié et al. [54], all data were stored as text files and plots in a folder structure. Text (.txt) and CSV (.csv) files are plain text formats and are the simplest way to store information. The data is stored in a row-vice manner and can be opened and read on all available platforms. Nonetheless, for every character in a file, one byte of storage is necessary. Thus, plain text formats can be very inefficient regarding memory requirements, especially for more complex data structures. Furthermore, metadata with the information, for example, about the most optimal cluster parameters, are stored in a separate file. By contrast, storing all the information on a single dataset in a database removes the complexity and amount of files. Moreover, it allows adding metadata to the relevant files within the database.

A database is a file with a more complex data structure. In the case of BaClAva, the introduced database is the Hierarchical Data Format version 5 (HDF5) file format with its file-ending .h5 because of its analogies to computer file systems. The data structure of an HDF5 file consists of groups, datasets, and metadata and can even incorporate plots. Here, a group is similar to a folder, and a dataset is comparable to a file. Thus, modules 2 and 3 automatically create a similar architecture to the original folder design presented in Griffié et al. [54] within an HDF5 file for every localised data sample. The metadata is then directly attached to the relevant dataset or group in the HDF5 file.

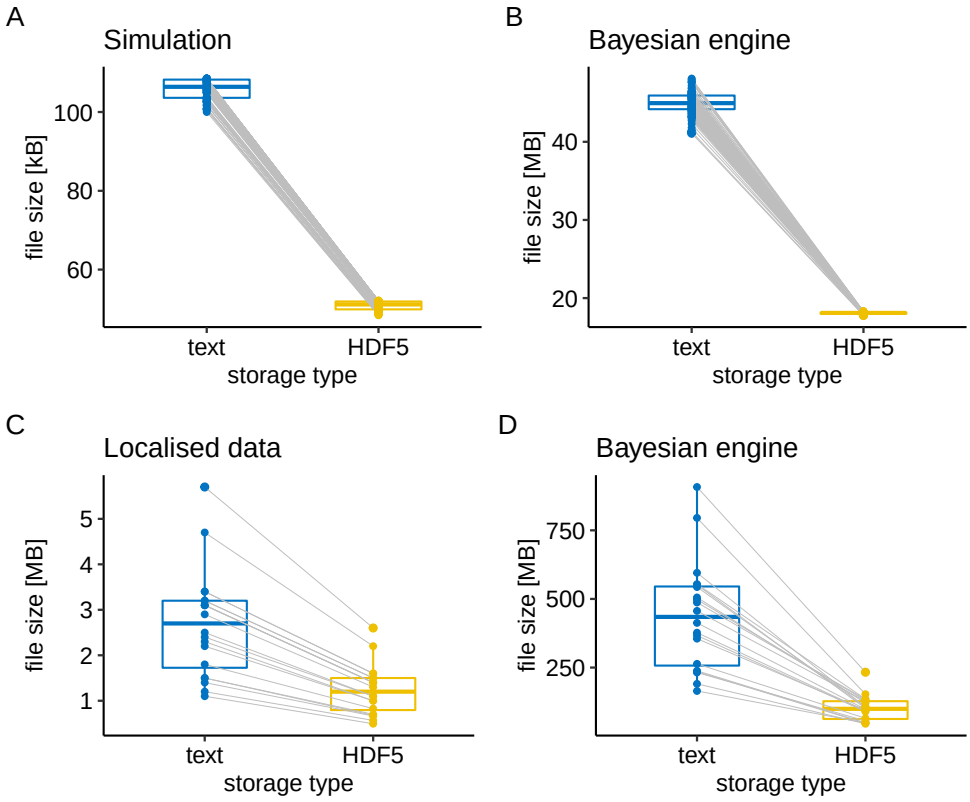

Figure 4.8: Comparison of storing data in different file formats. (A) Storing simulated data as either text or HDF5 files results in two file sizes. After applying ToMATo clustering with the Bayesian engine, the storage size increases in both cases (B). Microscopy data shows a broader range of file sizes after the localisation process (C) and clustering with the Bayesian engine plus Tomato (D).

Fig. 4.8 compares the 100 simulated and five microscopy experiment datasets. As shown in Fig. 4.8 A, the text files have, on average, a size of (105.9 ± 2.5) kB (mean $\pm$ standard deviation), whereas the HDF5 files storing the same data are 52 % smaller

with a size of (50.9 ± 1.1) kB. After the Bayesian run, the original software stored 6002 additional files for each dataset, adding up (44.9 ± 1.4) MB to the storage size (Fig. 4.8 B). In comparison, after the computation of module 2, the same information stored in the HDF5 format comprises (18.08 ± 0.11) MB for every dataset in a single HDF5 file. Thus, the code stores the additional data within the corresponding HDF5 files instead of adding thousands of files to the storage. On average, the storage size increases by (18.03 ± 0.11) MB. In summary, after the Bayesian run in module 2, the data stored in an HDF5 file has 40 % of the storage size of the same data stored as text files in a folder structure, and the simulated data is less than 3 % of the total storage size per dataset (text: (2.355 ± 0.062) % and HDF5: (2.815 ± 0.065) %).

The final step is the post-processing module, storing cluster plots, batch analysis histograms, density plots (PNG, EPS, and PDF) and the batch data as text files for both backbone structures. The backbone of BaClAva additionally stored the post-processing results in two CSV files. Two CSV files are necessary because some of the batch analysis results are on a per cluster level, i.e. the area or density per cluster, and some are on a per dataset level, i.e. the number of clusters or ratio clustered to the unclustered area. Thus, the number of rows differs. The CSV files were added to the output to facilitate further analysis because it is much simpler to import tabular data than vectorised data to other programs. BaClAva also offers the density per area of a cluster as an additional batch processing output and is not part of the original code by Griffié et al [54]. Thus, the number of files for the batch post-processing output is more in the case of BaClAva than in the original R code.

Nonetheless, this output only comprises 0.36 % of the total storage volume for BaClAva compared to 0.20 % for the adapted R code. As a result, the mean storage volume per dataset without the batch processing results in the case of text-files plus images is (45.3 ± 1.4) MB, and in the case of HDF5 files, (18.50 ± 0.11) MB, adding up to a total storage volume including the batch post-processing of 4.5 GB and 1.9 GB, respectively. In terms of files, there are 600 958 files in the text case and 667 in the HDF5 case.

Compared to the simulated datasets with their size in the kB range, datasets acquired

from microscopy samples are typically in the MB to GB range. Thus, performing the same analysis on experimental datasets checks whether the relationship between text and HDF5 files is consistent when scaling up. The original text files had a size of (2.7 ± 1.2) MB, and the HDF5 files had a size of (1.23 ± 0.54) MB (Fig. 4.8 C), meaning the HDF5 files were 46 % of the text file size. When applying the same Bayesian parameters in module 2 as for the simulated data above, the storage size for the text files of one experiment increased to (0.44 ± 0.20) GB, whereas the HDF5 files increased to (0.102 ± 0.046) GB (Fig. 4.8 D). Therefore after the Bayesian engine, the HDF5 storage size comprised 23 % of the text storage size for each dataset. After the final post-processing step, the storage size of the text files folder and HDF5 file plus the plots per sample increased by (2.7 ± 1.1) MB and (3.6 ± 1.5) MB adding up to (0.44 ± 0.20) GB and (0.105 ± 0.047) GB, respectively. In the end, the text file analysis consisted of 120 238 files and had a total storage size of 8.8 GB, whereas the HDF5 file analysis consisted of 167 files, resulting in a total storage size of 2.1 GB. Therefore, the HDF5 total storage size was 24 % of the text file storage size.

In conclusion, using HDF5 files for data storage decreases the necessary storage size and the stored number of files compared to saving the information as multiple single text files. On the one hand, the storage medium can accumulate more datasets; on the other hand, the lower number of files facilitates data organisation.

4.3 Understanding clustering

The following part explains how cluster algorithms behave when confronted with different cluster sizes and densities. Depending on the target molecule, the cluster sizes, background and localisation density within clusters might demonstrate a high variability in a single field of view.

4.3.1 Influence of cluster sizes

In the simulations in the previous section 4.2, all clusters had an input radius of 50 nm. Thus the next step is to see how the algorithms behave when confronted with different

statistic	ground truth	Ripley's-K-based	DBSCAN	ToMATo
total cluster count	1000	979	967	991
mean [μm^2]	0.045 ± 0.018	0.032 ± 0.019	0.037 ± 0.014	0.036 ± 0.015
{min, max} [μm^2]	{0.0080, 0.12}	{0, 0.20}	{0.0068, 0.15}	{0.000 68, 0.11}
25 % quantile [μm^2]	0.029	0.020	0.027	0.025
median [μm^2]	0.041	0.028	0.035	0.033
75 % quantile [μm^2]	0.061	0.041	0.046	0.047

Table 4.1: Overall cluster statistics for simulations with two input cluster sizes. Cluster statistics for 100 datasets with five clusters with a standard deviation of 50 nm and five with a standard deviation of 20 nm. The three cluster algorithms plus the Bayesian engine found different cluster results. Even the ground truth data has a spread due to the simulation's underlying Gaussian distribution.

cluster sizes in the same region of interest. Therefore, 100 datasets were simulated with ten clusters. However, five were drawn from a Gaussian distribution with a standard deviation of 20 nm and five from a Gaussian distribution with a standard deviation of 50 nm. Therefore, in total, 1000 clusters were simulated. Each cluster consisted of 100 localisations. As before, 50 % of the localisations were labelled as unclustered. The simulated datasets and their labels are referred to as ground truth data for the following comparisons. As shown in Table 4.1, the number of detected clusters varies between the three algorithms compared to the ground truth simulations. DBSCAN identified a total of 967 clusters with an area of $(0.037 \pm 0.014)\,\mu m^2$, Ripley's-K-based clustering determined 979 clusters with an area of $(0.032 \pm 0.019)\,\mu m^2$ and ToMATo 991 clusters with an area of $(0.036 \pm 0.015)\,\mu m^2$ (mean ± std).

statistic	ground truth	Ripley's-K-based	DBSCAN	ToMATo
mean cluster number	10	9.8 ± 1.6	9.67 ± 0.55	9.91 ± 0.51
{min, max}	{10, 10}	{7, 19}	{8, 10}	{8, 12}
25 % quantile	10	9	9	10
median	10	10	10	10
75 % quantile	10	10	10	10

Table 4.2: A simulated dataset contains 10 clusters, each composed of 100 data points. 5 clusters have a standard deviation of 20 nm, the remaining 5 have a standard deviation of 50 nm, and 50 % of all data points are within the background (A). The dataset was clustered using BaClAva plus Ripley's-K based clustering (B), DBSCAN (C), and ToMATo (D). In (B), all clusters appear to possess uniform dimensions, whereas (C) and (D) display two distinct cluster sizes. The colour assignment was random. The quantification of clusters per ROI (E) reveals a recurring tendency of the algorithms to determine the correct cluster count inaccurately. The violin plot for the cluster area (F) demonstrates that the algorithms underestimate cluster areas for both simulated cluster sizes.

On a dataset level, ToMATo detected 9.91 ± 0.51 clusters with a maximum of 12 and a minimum of 8. DBSCAN detected 9.67 ± 0.55 clusters with a minimum of 8 and a

maximum of 10, whereas Ripley's-K-based clustering identified 9.8 ± 1.6 clusters varying from 7 to 19 clusters per dataset. In all three cases, the median is ten clusters (Tab. 4.2). The Gaussian distribution of the simulated cluster sizes can also be identified in Fig 4.9, as the alpha-hull function in the post-processing module calculated all cluster sizes, including the ground-truth data. The density distribution for the ground truth data shows two distinct Gaussian peaks with an overlapping region, which are less pronounced for the results of the three different cluster algorithms. For Ripley's-K-based clustering and DBSCAN, the distributions depict one central peak with a shoulder for higher area values. The distribution for ToMATo reveals a second peak but at a smaller value than the simulated data. By overlapping the four distributions, it is already visible that none of the cluster algorithms could identify the correct cluster sizes for the larger clusters.

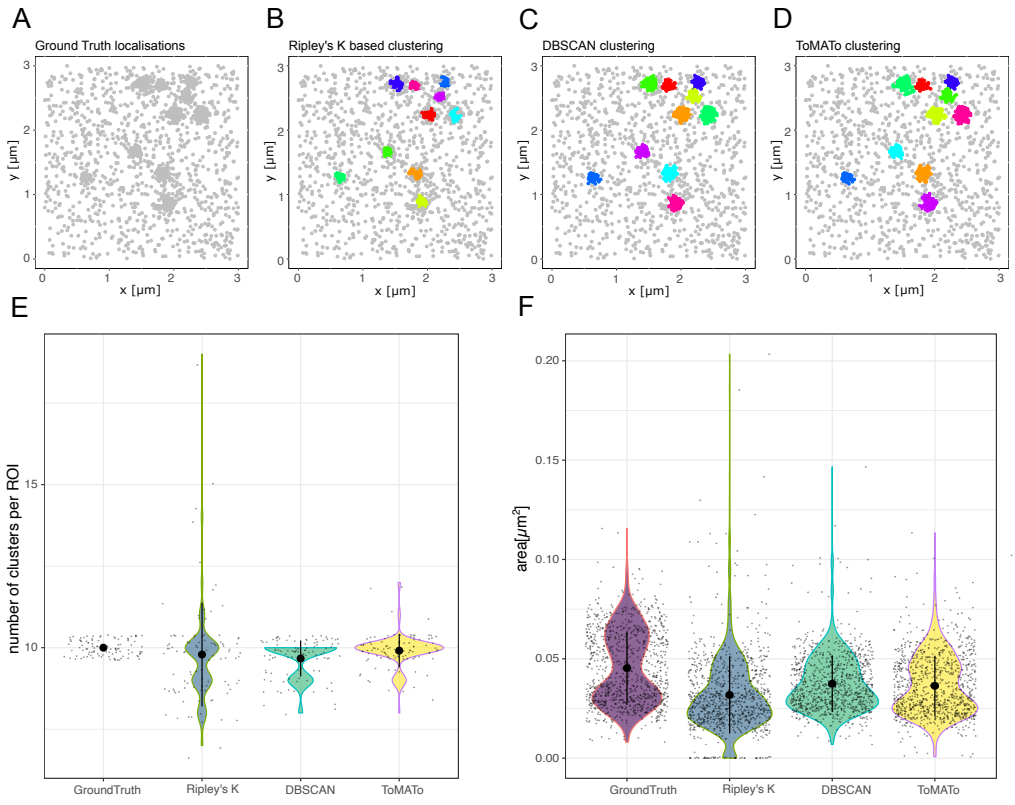

Figure 4.9: Comparison of storing data in different file formats. (A) Storing simulated data as either text or HDF5 files results in two file sizes. After applying ToMATo clustering with the Bayesian engine, the storage size increases in both cases (B). Microscopy data shows a broader range of file sizes after the localisation process (C) and clustering with the Bayesian engine plus Tomato (D).

Furthermore, the major peaks for the smaller cluster sizes are also shifted to smaller

values for the cluster algorithms compared to the simulation data (Fig. 4.9 F). The next step is to group the data for each algorithm and the ground truth data into two groups, one for the smaller and the larger cluster sizes. Figure 4.10 shows the results. Using K-means with $k = 2$ as explained in section 2.5.1, the analysis for the simulated ground data resulted in the areas $a_{G1} = (0.0308 \pm 0.0074)\,\mu m^2$ and $a_{G2} = (0.063 \pm 0.010)\,\mu m^2$ with 55.5 % and 44.5 % of the clusters, respectively. The analysis for ToMATo resulted in $a_{T1} = (0.0259 \pm 0.0067)\,\mu m^2$ and $a_{T2} = (0.052 \pm 0.010)\,\mu m^2$ with 59.1 % and 40.9 % of the clusters, respectively. For DBSCAN, the clustering resulted in $a_{D1} = (0.0290 \pm 0.0063)\,\mu m^2$ with 63.2 % of the clusters and $a_{D2} = (0.052 \pm 0.012)\,\mu m^2$ with 36.8 % of the clusters. The analysis of Ripley's-K-based clustering, clustered 71.4 % to an area $a_{R1} = (0.0228 \pm 0.0089)\,\mu m^2$ and 29.5 % of the clusters to $a_{R2} = (0.054 \pm 0.021)\,\mu m^2$. The corresponding Figure 4.10 shows that the grouped area distributions are Gaussian-like. All cases, including the simulated ground truth data, show a left-skewed distribution for the smaller area values and a right-skewed distribution for the larger area sizes.

4.3.2 Influence of fluorophore density

So far, the simulations were based on the idea that every localisation represents a single molecule, meaning that there is a one-to-one relationship between the ground truth molecules and the detected localisations. Nevertheless, this idea does not represent most SMLM techniques with blinking behaviour where a single molecule can be detected several times. The following will concentrate on (*d*)STORM with its most prominent organic fluorophore, Alexa Fluor 647 (AF647). For PALM, there are existing correction algorithms to account for the blinking behaviour [6, 7].

Let us consider an individual molecule of interest, labeled with a single AF647. In (*d*)STORM, the fluorophore blinks, thus being imaged multiple times over time. Due to the localisation inaccuracy, a point cloud represents the molecule, and clustering algorithms can identify such a point cloud as a cluster. Thus, this phenomenon is called self-clustering. To understand the influence of self-clustering on membrane molecules forming clusters or in a random distribution on the cell membrane, we prepared simula-

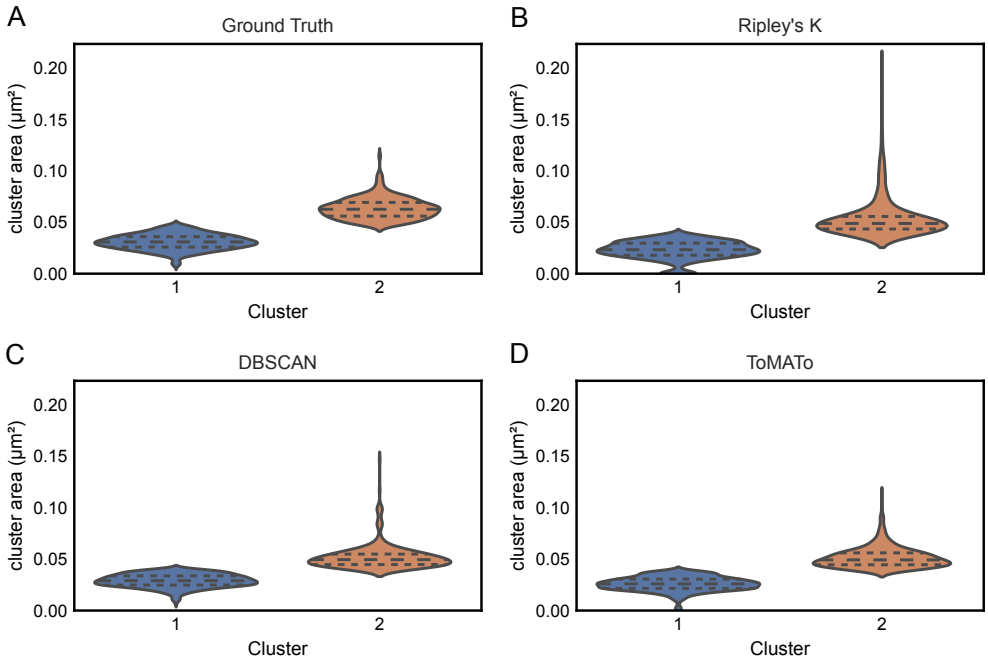

Figure 4.10: K-means on datasets with two cluster sizes. BaClAva analysed 100 simulations, each exhibiting two distinct cluster sizes. This analysis encompassed the computation of cluster areas for both the ground truth data and the results obtained after applying BaClAva and the three available clustering algorithms. These resulting datasets underwent K-means clustering, where K was set at 2. A violin plot depicts the results of each cluster for ground truth data (A), Ripley's-K based clustering (B), DBSCAN (C), and ToMATo (D).

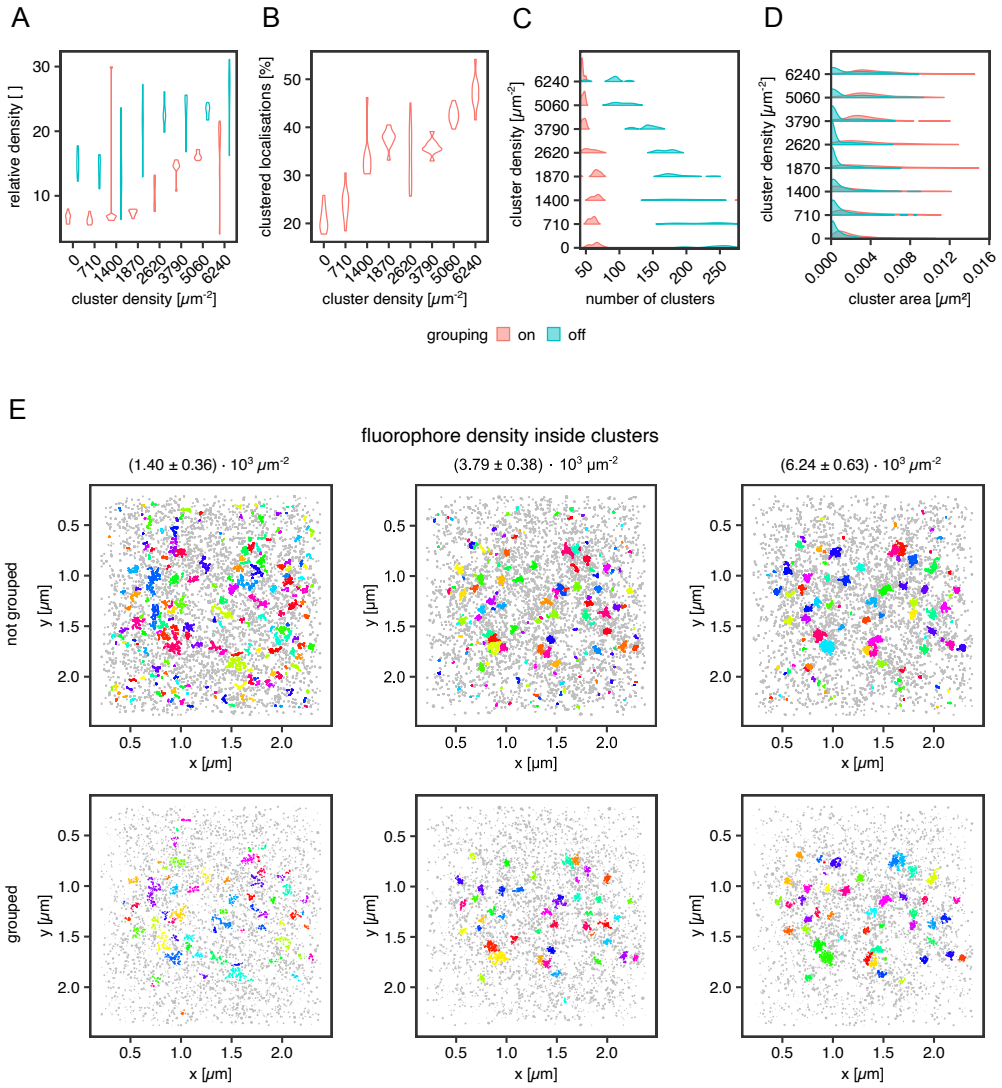

Figure 4.11: The impact of the fluorophore density on clustering. Violin plots for the relative density of the clusters vs the background (A) and the percentage of clustered localisations (B). Ridgeline plots for the number of clusters per ROI (C) and the areas of the clusters (D). All data were analysed with grouping on and off, except in (B). (E) Example of a simulation with 40 clusters analysed in SMAP with grouping set on/off. The BaClAva with ToMATo computed the optimal clustering results. The density within the cluster of the scatterplots per column was $(1.40 \pm 0.36) \cdot 10^3\ \mu m^{-2}$ (LEFT), $(3.79 \pm 0.38) \cdot 10^3\ \mu m^{-2}$ (MIDDLE) and $(6.24 \pm 0.63) \cdot 10^3\ \mu m^{-2}$ (RIGHT). [82]

tions of (*d*)STORM experiments in FluoSim (see section 3.3.2).

In the region of interest, 40 non-overlapping clusters were randomly distributed. The area of each cluster was 0.078 μm^2, meaning that the circular clusters had a diameter of 50 nm. Within the cluster area, the molecular density was increased $(0.71 \pm 0.25) \cdot 10^3$ μm^{-2} to $(6.24 \pm 0.63) \cdot 10^3$ μm^{-2}, translating to molecules per cluster ranging from 5.6 ± 1.9 up to 49.0 ± 4.9. The background non-clustered molecules were then spatially randomly distributed over the area of the ROI, excluding the areas assigned to the clusters. Thus for sparse clusters with (639 ± 49) molecules/μm^2, 94 % of all molecules were in the background. In contrast, 51 % of the molecules were generated as background molecules for clusters with a density of (346 ± 51) molecules/μm^2.

Assuming that the molecules were labelled with AF647, the blinking rates were $k_{on} = 0.01\,s^{-1}$ and $k_{off} = 10\,s^{-1}$, and the PSF had a Full Width at Half Maximum (FWHM) of 200 nm with a maximal intensity of 2007. The values for the camera (pixel size = 96 nm) and for the (*d*)STORM simulation (50 000 frames with an exposure time of 10 ms) were identical to an experiment on the microscope. The simulations were exported as image stacks and localised in SMAP (see section 3.3.1). The localisation process was performed for each dataset with and without grouping. Grouping decreases the number of localisations by merging localisations that are, in this case, maximal one frame off and within a circular area of the previous localisation with a radius of 35 nm.

Fig. 4.11 shows the results of the grouped versus the non-grouped data analysis. Figure 4.11A plots the ratio of molecules assigned to clusters to molecules assigned to the background for all analysed cluster densities. In both cases, the ratio increased with increasing cluster density, but the spread was smaller for the grouped data, showing that the grouping functionality in SMAP is efficient. The number of clusters in Fig 4.11C reveals that the number was smaller at every cluster density than the non-grouped data.

Furthermore, the number of clusters was constant up to (2.61 ± 0.39) localisations/μm^2 for grouped data and at higher concentrations, the number of clusters approached the simulated number of 40. In comparison, the non-grouped data show a significantly

higher number of clusters, decreasing with higher cluster density. Nevertheless, only the highest simulated density reached the actual simulated number of clusters. Consequently, the grouping removes clusters caused by single-blinking fluorophores. Therefore, a local density threshold for the cluster molecules must be surpassed to interpret the cluster results independent of blinking properties. As shown in Fig 4.11B, the percentage of clustered localisation increased with increasing fluorophore density. Thus, the relative enrichment inside the clusters than outside of them was higher for high cluster densities. Overall, to obtain the best results, at least 30 % of all localisations were in clusters, surpassing the relative density threshold of 10.

Besides the (relative) fluorophore density, the cluster area was a critical analysis objective. Similar to the density observations, the background data points also influenced the data distribution in the area, as shown in Fig 4.11D. For the grouped data, the cluster size increased starting from $(1.87 \pm 0.39) \cdot 10^3\,\mu m^{-2}$ to the simulated area of $0.078\,\mu m^2$. In comparison, in Fig 4.11D, the non-grouped data had many tiny clusters at all simulated concentrations, which was not present in the grouped data. Consequently, they might have been single-molecule clusters that the grouping correction corrected in SMAP. Starting from a concentration of $(2.62 \pm 0.39) \cdot 10^3\,\mu m^{-2}$, a second population emerged in the non-grouped data, equaling the main population in the grouped data.

Furthermore, the number of identified clusters decreased with increasing molecular concentration in the non-grouped datasets, starting from the same molecular concentration. Comparing the same datasets localised with and without grouping in Fig. 4.11E reveals that incorrect cluster assignments were removed or corrected by two means. Either by the grouping mentioned earlier (top row vs bottom row) or the increased molecule density within the simulated clusters (from left to right). At low densities, BaClAva often assigned multiple labels to the data points of a ground truth cluster, resulting in multiple clusters instead of a single one. On the other hand, increasing the molecular density in clusters enhanced the ground truth clusters in both conditions, resulting in a cluster assignment with a higher resemblance to the ground truth simulation and significantly reducing the influence of single-fluorophore blinking on the cluster identification for denser clusters. Therefore, grouping is an essential tool to minimise the influence of single-fluorophore

self-clustering on data analysis with density-based clustering approaches.

4.4 Clustering of biological proteins - GPI vs CD95 vs CCP

So far, all experiments were *in silico*, but performing SMLM experiments in intact cells determines the usability of BaClAva. The selected controls are standard in the SMLM field and are all expressed in or at the cell membrane. As a positive control for clusters in cells, the selected target was clathrin light chain (CLC) which is present as dozens of copies in a single clathrin-coated pit (CCP) with an approximate diameter of 150 nm [144]. Thus, CLC is a strongly clustered molecule. On the other hand, as a negative control, Glycosylphosphatidylinositol (GPI) was selected. As an outer membrane leaf molecule, it should be roughly homogenous distributed because it should not have any interactions with the cell's cytoskeleton. The target molecule of interest with an unknown distribution is the transmembrane receptor CD95 (section 2.1).

All samples were labelled following the same principle. The protein of interest was coupled to GFP, and the dye AF647 was bound to GFP via anti-GFP nanobodies. AF647 is the most commonly used and best-characterised (*d*)STORM dye available. Due to its blinking properties, expectations for GPI were a wide range of cluster sizes, as shown in the simulation experiments (section 4.3.2). On the other hand, the expectation for the CLC results was to show well-defined clusters because the strong signal of the clustered molecules should minimise the influence of blinking on the final clustering, as shown in the simulation work. Since the cells with the CD95 target molecule had not undergone treatment with their ligand CD95L, the experimental results show the pre-activation receptor distribution.

The reconstructed SMLM images showed three different spatial distributions for the target molecules (Fig. 4.12). For the GPI imaged in CV-1 cells, the full image of a part of the cell membrane shows a homogenous distribution of the grouped localisations (Fig. 4.12 A). As shown in the zoom-ins, there are only small clusters that most probably originate from the blinking of AF647. The positive control CLC was imaged in HeLa cells,

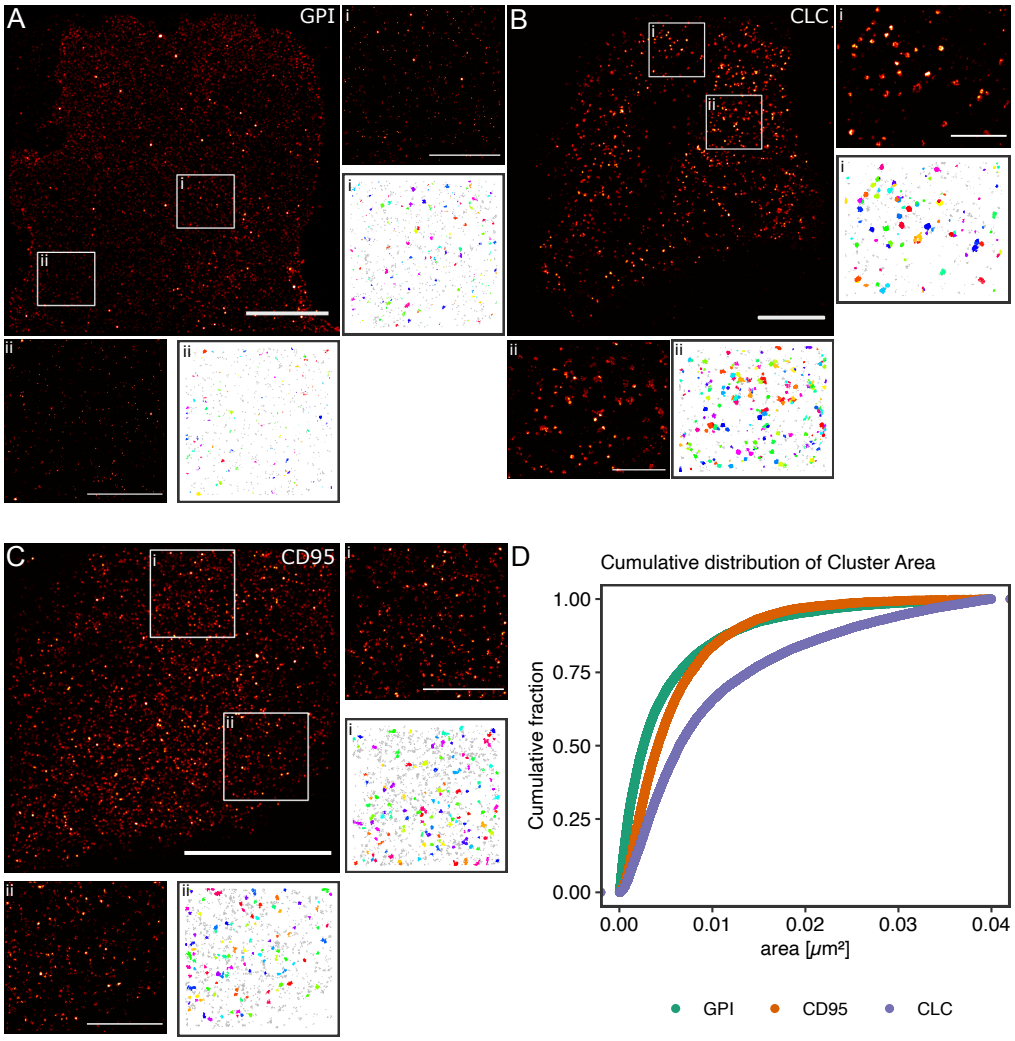

Figure 4.12: Clustering three different target molecules with BaClAva in biological samples.The negative control GPI-GFP in CV-1 cells (A), the positive control CLC-GFP in HeLa cells (B), and the sample with unknown distribution CD95-GFP in T98G cells were stained with anti-GFP nanobodies labelled with AF647 and imaged in (*d*)STORM mode. The large reconstructed images have a scale bar of 10 µm and the zoom-ins (i-ii) of 3 µm. Next to the reconstructed zoom-in is the corresponding cluster plot of the same area. (D) The cumulative distribution plot for the cluster areas of the three selected target molecules. [82]

and the reconstructions in Fig. 4.12 B show well-defined clusters and little background signal. Measuring several individual clathrin-coated pits, as shown in Fig. 4.13, revealed that the size of the clustered signal with a diameter $d_{CCP} = (126 \pm 24)\,\text{nm}$, measured over nine samples, agrees with the literature [144]. The reconstructed image of CD95 imaged in T98G cells in Fig. 4.12 C shows well-defined but smaller clusters similar to CLC with a higher background signal.

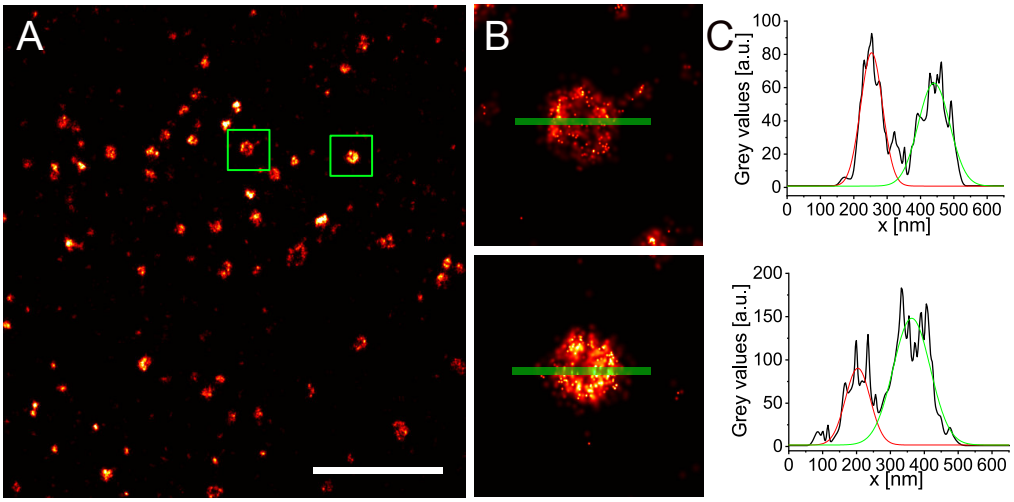

Figure 4.13: Verification of the clathrin-coated pit size. (A) The CLC-GFP in HeLa cells were stained with anti-GFP nanobodies labelled with AF647 and imaged in (*d*)STORM mode. The scale bar is 3 µm. (B) Zoom-ins two CCPs with an image size of 1 µm x 1 µm. (C) The corresponding line profiles for the two zoom-ins. Averaging over nine individual CCPs results in a diameter of (126 ± 24) nm. [82]

The cumulative distribution function (CDF) for the cluster area of several cells per target protein was calculated to compare the clustering behaviour for the three target proteins (Fig 4.12 D). As expected and already identifiable in the reconstructed SMLM images, the distributions for GPI and CLC are well-separated. The distribution for CD95 lies in between these two controls. According to a Kruskal-Wallis test, the three distributions are significantly different ($p < 2 \cdot 10^{-16}$). Together with the reconstructed images, the conclusion is that CD95 resides in small clusters and as homogenously distributed molecules in T98G cells. The CD95 clusters have approximately 0.54 molecules/nm^2 in the plane of the cell membrane.

4.5 Tethering

Cells do not exist in isolation from their surrounding environment. Instead, they interact with the glass surface on which they are seeded or with neighbouring cells with whom they come into contact. The following section presents a methodology that allows for the investigation of a signalling pathway or a particular aspect of it, free from any potential impact of the surrounding environment. In this method, individual cells are tethered to a supported lipid bilayer (SLB) via DNA strands, resulting in their isolation and preventing any physical interaction with the glass surface or neighbouring cells. By combining this DNA-tethering technique with standard SMLM and widefield imaging methods, the CD95 receptor on isolated T98G cells was observed.

4.5.1 The tethering procedure

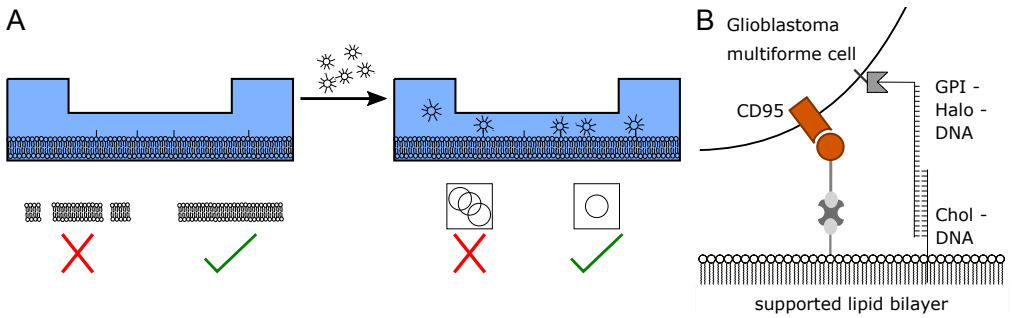

Figure 4.14: Tethering schematic and illustration of a T98G cell anchored to a SLB. (A) The supported lipid bilayer (SLB) was prepared within glass-bottomed channels as the first step in the tethering protocol. 20-mer DNA strands were attached to the SLB using cholesterol. Extensive checks were conducted to ensure the homogeneity of the bilayer. These checks identified any areas with faulty or incomplete formation, as cell attachment to the glass surface could occur in these regions. T98G CD95-GFP cells were labelled with 100-mer DNA strands using GPI-Halo and introduced to the channels with a well-prepared bilayer. The cells were then tethered to the bilayer by DNA coupling. A brief imaging test was performed before the experiment to assess cell mobility. When consecutive images were overlaid, it was possible to identify the immobilised cells while the non-tethered cells exhibited movement. (B) In the final experimental setup, a T98G cell was firmly anchored to a SLB using DNA strands, ensuring precise spatial positioning. The DNA strands on the cell side were linked through GPI-Halo, while the DNA-cholesterol construct was incorporated into the SLB on the opposite side. Activation of the CD95 receptor could be initiated by the ligand embedded within the SLB. Fluorescent imaging techniques were performed using the GFP coupled to CD95 receptors or AF647 coupled to the GFP via a nanobody (not shown).

Tethering cells to supported lipid bilayers (SLBs) within channels involves several steps, as illustrated in Figure 4.14. The protocol consisted of two main steps: preparation

of a supported lipid bilayer in channels and cell preparation. Firstly, an even coating of a supported lipid bilayer should be applied to the glass surface. In the experiments presented here, this bilayer should include a 20-mer oligomer via cholesterol for tethering and a quality assessment dye, carboxyfluorescein. The dye allows for checking that the bilayer is evenly coated and suitable for cell tethering, as seen in Figure 4.15 (A) for a well-prepared SLB and (B) for a poorly prepared SLB. A construct containing four CD95 ligands was introduced into the SLB to activate the CD95 receptor. Next, the cells must express the CD95 receptor genetically fused with GFP and GPI fused with a HaloTag. Finally, the 100-mer oligomer, containing a 20-base tethering sequence, must be covalently bound to the HaloTag. Upon addition of cells to the bilayer, they become tethered to it through the base pairing of the 20-mer complementary DNA strands. This process enables the study of cell behaviour within channels.

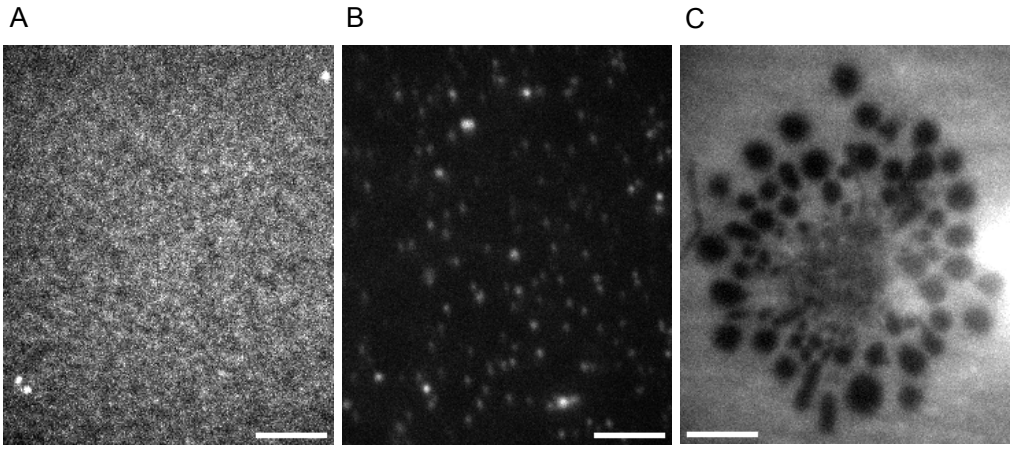

Figure 4.15: Quality assessment of SLBs. The SLB was incorporated with the dye carboxyfluorescein, which was activated to evaluate the quality of the prepared bilayer. (A) A carefully prepared SLB exhibits homogenous bilayer formation with only three distinct bright spots, indicating good SLB quality. (B) In contrast, an inadequately prepared SLB presents numerous bright spots, preventing the assessment of bilayer formation. (C) A cell attached to the glass surface after successfully penetrating the bilayer. The SLB appears as a bright region surrounding the cell, attributed to the presence of carboxyfluorescein within the SLB. For (A), the exposure time was 50 ms, while it was 100 ms for (B) and (C). Each image has a scale bar measuring 5 µm.

Cells in suspension tend to adopt a spherical shape because of their low surface-to-volume ratio compared to other 3D shapes [115, 149, 165]. Figure 4.16 shows living T98G cells tracked using the Janelia Fluor dye JF635 to label the GPI-Halo, which is only present on the cell membrane and not within the cell body. Imaging was performed to observe the behaviour of cells tethered to the SLB via complementary 20-mer DNA strands. Over

an extended observation period, the tethered cells remained stationary in the liquid environment (Fig. 4.16 A and B). However, non-tethered cells moved freely during the same period and were not labelled with Halo-JF635 (Fig 4.16 C and D).

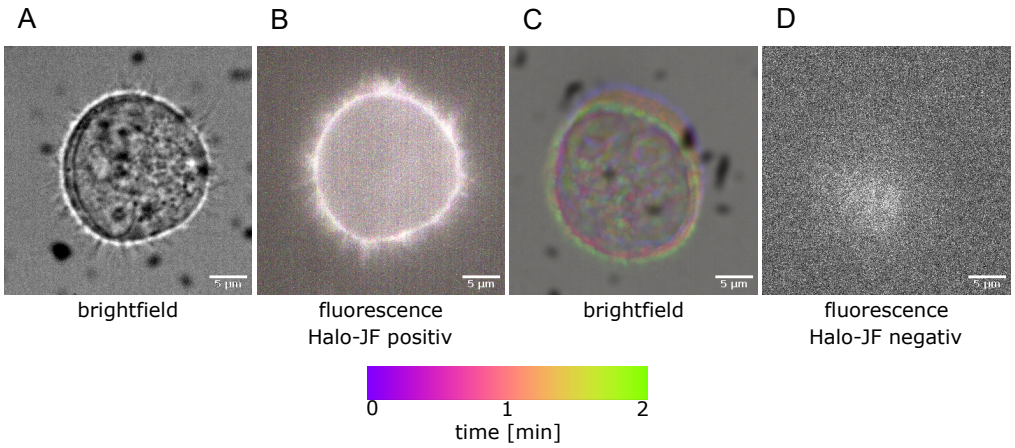

Figure 4.16: Time-lapse microscopy of tethered and untethered cells. (A) Brightfield image and (B) 635 nm channel image depict a T98G cell, positively transfected with GPI-Halo and tethered to the SLB, displaying noticeable protrusions. The fluorescent image (Halo-JF635) is a time-coloured projection spanning 2 minutes, showing perfect colour alignment resulting in a white appearance. In contrast, (C) presents the time-lapse sequence of an unanchored cell captured in the brightfield image, illustrating its movement throughout the 2 minutes. (D) The absence of fluorescent signals in the untethered cell is due to an unsuccessful GPI-Halo transfection. The images have a scale bar of 5 µm.

Ensuring the integrity of the bilayer was crucial in validating the tethering of immobile cells via DNA strands. As illustrated in Figure 4.15 C, several cells were identified as immobile and attached directly to the glass surface instead of being tethered via DNA strands. As a result, the tethering procedure includes examining bilayer quality when a stationary cell is detected to ensure proper tethering. The diagram in Figure 4.14 represents the inspection of bilayer quality as an integral step in the overall tethering process.

The gradual formation and development of filopodia can be observed through long-term imaging of the lower region of a tethered cell. Characterised by their elongated and tentacle-like structures, these filopodia extend outward from the cell surface [99]. They enable the cell to explore its microenvironment and serve as a reliable indicator of cellular viability. Figure 4.17 illustrates the behaviour of a tethered cell over 10 s and 5 s in Figure 4.17 A and B, respectively, highlighting the presence and development of filopodia. The detected signal in the image originates from GFP, which is coupled to the CD95 receptor.

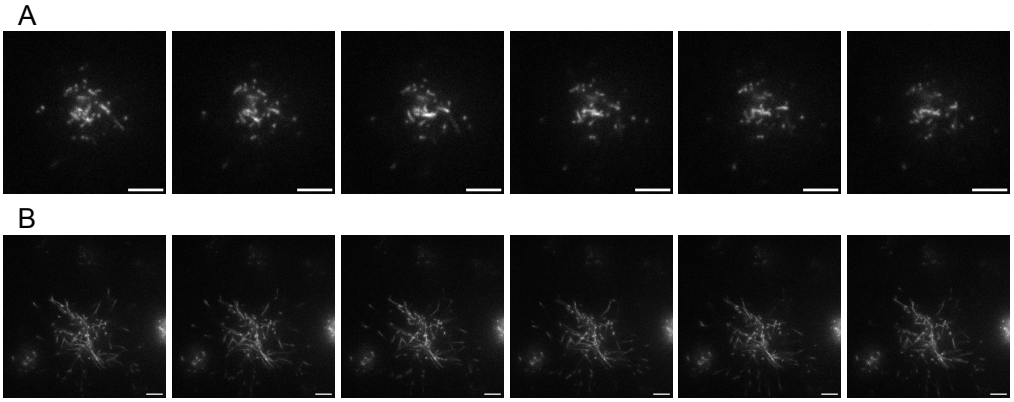

Figure 4.17: Time-lapse fluorescence imaging of anchored T98G cells. T98G CD95-GFP cells underwent fluorescence imaging following successful DNA-tethering to the supported lipid bilayer (SLB). Over 30 minutes of incubation, the cells developed extrusions on their cell surface, which moved during imaging. The time-lapse sequence of each cell commences with the leftmost image at 0 seconds and concludes with the rightmost image. In (A), a series of images capturing a single cell were taken over a 10-second duration, with each image having an exposure time of 100 ms. The interval between consecutive images was 2 s. In (B), another cell was imaged for 5 s with an exposure time of 50 ms. The frames depicted in this sequence were captured at 1-second intervals. The images include a scale bar measuring 5 µm.

Notably, the CD95 receptor is distributed both in the central cell membrane and within the developed filopodia.

In order to perform SMLM imaging of tethered cells, the standard protocol for fixing and staining cells with AF647 was followed. However, to maintain the tethering of cells to the supported lipid bilayer (SLB) for (*d*)STORM imaging, certain modifications were introduced to the protocol A.5. All procedures were conducted at room temperature. Moreover, the sample was continually suspended in the fluid to prevent damage or rupture of the supported lipid bilayer (SLB).

The main objective of conducting (*d*)STORM imaging was to examine the spatial distribution of the CD95 receptor in the lower region of the cells, as depicted in Figure 4.14 B. Data was collected to investigate CD95 receptors under conditions with and without the presence of CD95L on SLBs containing 0.8 % or 1.5 % biotin. The cells were incubated in the channels for 15 min or 30 min. Before the addition and tethering of the cells, the CD95 ligand was introduced to the SLB. This approach ensured that the interactions between the CD95 receptor and its ligand occurred exclusively in the lower region of the cells, providing a focused analysis of this specific cellular region. In addition, the bilayer

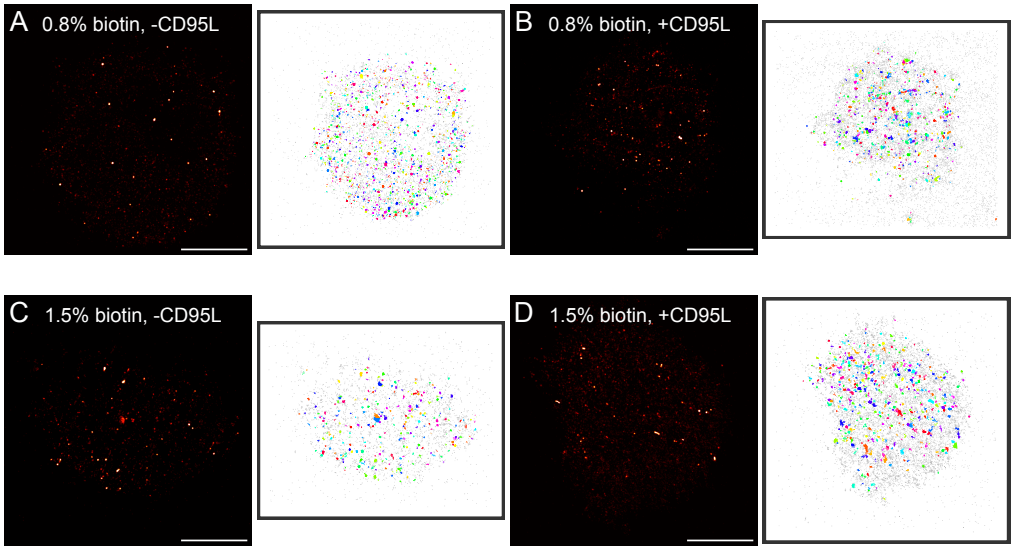

Figure 4.18: SMLM reconstruction and clustering of tethered T98G cells. T98G cells were tethered to a SLB and incubated for 15 min. Subsequently, the cells were fixed, stained with AF647-nanobodies, and imaged for 30 000 frames with an exposure time of 20 ms. The localised data was processed using BaClAva and the ToMATo algorithm. The T98G cells were exposed to four distinct conditions: (A) SLB with 0.8 % biotin concentration and without CD95L, (B) SLB with 0.8 % biotin concentration and with CD95L, (C) SLB with 1.5 % biotin concentration and without CD95L, and (D) SLB with 1.5 % biotin concentration and with CD95L. The reconstructed images are displayed on the right side, with a scale bar measuring 5 µm. The same data is presented on the left side but with cluster results highlighted in colour. Colours were assigned randomly to clusters.

uniformity was consistently assessed before imaging. Any irregularities identified near a cell resulted in excluding that specific cell from the imaging process. This control measure aimed to ensure the reliability and accuracy of the imaging results by eliminating any potential misleading characteristics that could arise from compromised bilayer integrity.

As the cells were situated above the lipid bilayer, the imaging approach utilised the highly inclined and laminated optical (HILO) mode, as described in the theoretical section 2.3. The resulting reconstructed images of the CD95 receptor within the tethered cells and an incubation time of 15 min are depicted in Figure 4.18. Clustering analysis was performed using BaClAva software and the ToMATo clustering algorithm to examine CD95 receptor spatial distribution. Figure 4.19 shows the clustering analysis results. This analysis revealed substantial variability across all cells, regardless of their treatment or genetic disparities associated with the CD95 receptor. Notably, even cells that underwent identical treatments and were imaged in the same channel exhibited significant heterogeneity in their clustering patterns.

The study successfully established a method for tethering living cells to a supported lipid bilayer through DNA, thereby enabling live imaging of the cells. Additionally, the cells were fixed and stained, allowing for SMLM imaging and subsequent cluster analysis. This innovative approach provides a valuable tool for investigating cellular dynamics and spatial distribution in a controlled experimental setup. Furthermore, this technique has the potential to be applied in numerous ways to study cellular processes in individual cells surrounded by a liquid environment.

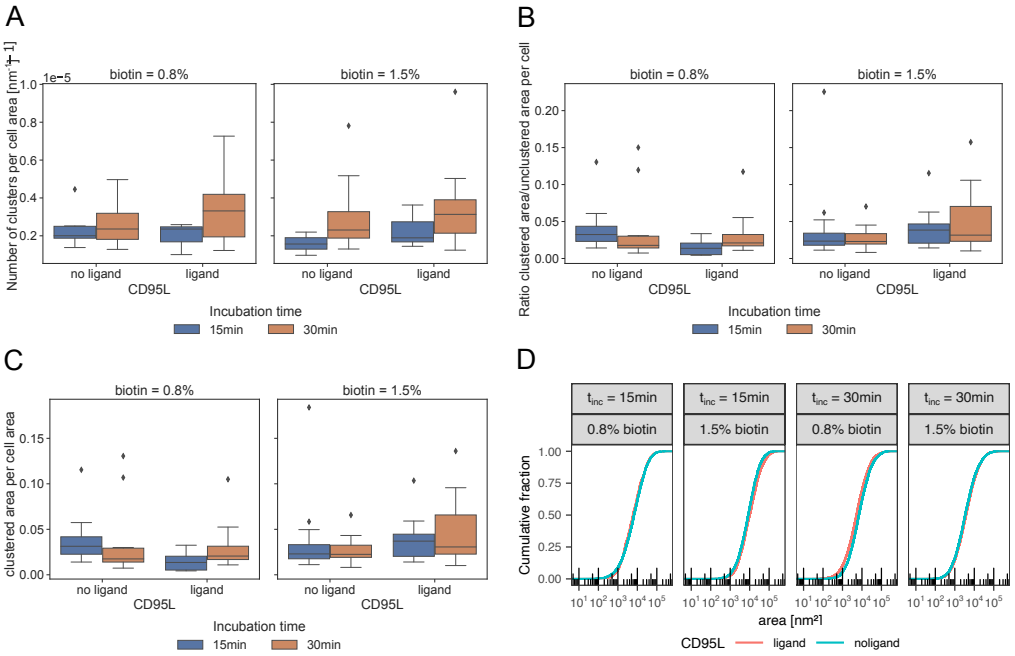

Figure 4.19: Cluster analysis of tethered T98G CD95-GFP cells using dSTORM and Baclava. In the dSTORM experiment, AF647 staining GFP via nanobodies was imaged in tethered T98G CD95-GFP cells under various conditions. Specifically, the biotin concentration in the bilayer was altered, and CD95L was either added or kept absent from the biotin. Cells were incubated for 15 or 30 minutes. Cluster analysis results, obtained using Baclava with the Tomato algorithm, are presented in three box plots. These plots display (A) the number of clusters per cell area, (B) the ratio of clustered area to unclustered area within a cell, and (C) the clustered area relative to the cell area. In (D), empirical cumulative distribution functions were used to analyse clustered area distributions.

5

Discussion

This chapter discusses the findings presented in the preceding chapter. The discussion starts with an analysis of the BaClAva results, followed by an analysis of the findings obtained for DNA-tethering.

5.1 BaClAva

The discussion regarding BaClAva is divided into several segments, commencing with an examination of the clustering of SMLM data as a whole. The development process of BaClAva is then discussed. Simulated and experimental SMLM data are evaluated as a final step.

5.1.1 Clustering

For cells to function efficiently and survive, the spatial and temporal organisation of molecular structures and expressed molecules on organelle surfaces, such as the cell membrane, is crucial. Our focus is solely on analysing the spatial organisation of molecules in cells since the cells used in the experiments were fixed. The organisation of membrane molecules has been analysed using cluster analysis with standard wide-field and confocal microscopy techniques [11, 49, 50, 74]. However, the diffraction limit of light poses a

limitation, preventing the detection of structures smaller than half the wavelength of the light source (see Theory section 2.3). Therefore, the results obtained using these methods are uncertain, and alternative techniques, such as electron and scanning-force microscopy, with a resolution comparable to protein sizes, must be used to interpret the results [151].

Super-resolution microscopy techniques, such as Single-Molecule Localisation Microscopy (SMLM), have been developed to overcome the limitation of fluorescence microscopy and the need for a complementary microscopy method. With SMLM, it is now possible to achieve resolutions of tens of nanometers in fluorescence microscopy images, which is crucial for analysing the spatial organisation of cellular molecules at the required finer granularity through cluster analysis. SMLM techniques represent a significant advancement in the study of membrane organisation, enabling cluster analysis to be performed with greater accuracy and precision and providing valuable insights into the spatial organisation of cellular molecules. [108]

Unlike standard wide-field, confocal [48, 112, 163], or Stimulated Emission Depletion microscopy microscopy [152], SMLM techniques such as PALM, (*d*)STORM, or DNA-PAINT produce a table as their primary output, from which a reconstructed image can be generated. This table comprises all information obtained by fitting the located fluorophores in the acquired frames, including their spatial coordinates and localisation error. We can use these coordinates to analyse the data using standard clustering techniques [108]. [62]

Among the clustering methods for tabular SMLM data, the nearest-neighbour algorithm (also known as the Clark-Evans statistic) [27] is one of the most straightforward. It evaluates whether the data is clustered by comparing it to randomly distributed data with the same density. However, this method does not provide any information about the spatial scale of clustering, the location of the clusters, or the number of localisations involved. [108]

Subsequent research by Owen et al. [114] using SMLM data has shown that Ripley's K-function on a given field of view outputs the length scale of cluster sizes and the

proportion of data points found in clusters. However, this method averages over the entire field of view and cannot provide information on the number or localisation of the clusters or details about individual clusters. Nevertheless, applying Ripley's K function to every data point in the dataset allows for estimations of regional clustering degrees and percentages of clustered localisations to be obtained.

The advent of SMLM techniques has prompted researchers to focus on addressing artefacts present in localised data [86]. These artefacts can stem from various sources, such as misaligned microscopes, improper sample preparation [140, 148], suboptimal laser power, sample drift [85, 100], buffer conditions leading to poor blinking behaviour of the fluorophores, and inadequate localisation parameters [6, 22]. A thorough investigation into these sources of artefacts has led to an improved understanding of high-quality imaging for more meaningful analysis of SMLM data.

However, clustering in SMLM techniques is subject to another source of artefacts, which is the foundation of these techniques: the blinking of fluorophores [6, 86]. When a single fluorophore blinks, its photons are detected as a blurred spot on the camera pixels and then fitted using localisation software. The simplest method involves fitting a 2D Gaussian to determine the peak coordinates, which are then recorded in the localisation table. A repeat of this procedure for different appearances of the fluorophore results in a cluster of localisations within the resolution limit of the SMLM experiment rather than identical coordinates being obtained when repeating the process for different appearances of the fluorophore. The merging of clusters can occur when two fluorophores are imaged in close proximity to each other, making their clusters indistinguishable from the cluster of a single fluorophore, especially given the non-uniform resolution in SMLM experiments [32].

To address this inherent issue, the group led by Gerhard J. Schütz proposed a filtration experiment in SMLM [12]. Data is collected for each concentration in this approach, wherein the dye density is modified. If the data is clustered, then the size of the clusters will eventually reach a plateau, while for unclustered data, the cluster sizes will continue to increase. This results in distinct filtration curves for the two cases. However, this

method is time-consuming.

The Schütz group recognised that the previous method is susceptible to subjective bias, prompting them to develop an improved technique [8]. Instead of filtration, this method relies on labelling the target with two spectrally separated dyes, and the cluster maps for the two colours are only correlated for clustered data. The data is transformed by applying a toroidal shift to one of the channels. The cumulative density functions of the nearest neighbour distance between the two channels are analysed to obtain different p-values, thus allowing a clear differentiation between clustered and unclustered situations without relying on prior information or user-defined parameters. However, it is necessary to label the target with two colours for this method to be effective, which could pose a challenge depending on the target.

Owen et al. introduced one of the initial per-cluster analysis approaches using a modified Ripley's-K-function algorithm, which determines clusters by applying a clustering threshold (cluster/background) and a search radius on the individual data points [114]. In this study, this algorithm is referred to as Ripley's-K-based clustering to distinguish it from clustering analysis using the previously mentioned Ripley's-K-function.

As previously mentioned, Ripley's-K-based clustering cannot accurately handle density fluctuations in a sample, leading to erroneous cluster size determinations. Thus, density-based methods are better suited for this purpose. DBSCAN is a method that depends on two search parameters: the radius and the minimum number of points within the search radius. This approach is less prone to errors. Valid points are considered part of a cluster, while the remaining points are assigned to the background. Unfortunately, the algorithm frequently struggles to differentiate between clusters humans can easily distinguish, merging them into one cluster. Pike et al. demonstrated in their publication that Ripley's-K-based clustering and DBSCAN are highly sensitive to slight changes in their input parameters, making it challenging to determine the optimal parameters for these methods. [117]

Pike et al. showed that persistence-based clustering is more robust to minor variations

in input parameters. Moreover, unlike density-based techniques like DBSCAN, methods falling under this category utilise density estimation of data points. The first application of persistence-based clustering to SMLM was introduced by Griffié et al. [52], followed by the development of the ToMATo cluster algorithm by Chazal et al. [24, 25], which was later adapted for SMLM by Pike et al. [117].

In the ToMATo algorithm [25, 117], local density maxima are identified in the density estimation. Creating a graph that connects neighbouring molecules enables the algorithm to calculate the density gradient along a path of the graph. As density decreases, corresponding to an increase in inter-point distance, the probability of reaching the cluster boundary also increases. The saddle point at the minimum in the density graph separates two clusters, defining the outer perimeter of each cluster and thereby separating them. Additionally, a threshold defines the persistence of a cluster from its centre into space. Clusters with a persistence lower than the threshold are assigned to neighbouring clusters or the background. The outcome of the process is the separation of partially overlapping clusters, resulting in an output that is less sensitive to input parameters. Nevertheless, this method has its limitations, as the persistence threshold is determined by the user in the persistence diagram of the dataset, which is susceptible to user bias, particularly in dense samples.

In addition to traditional clustering techniques, Voronoï tessellation [87] and machine learning [160] are emerging as compelling methods for cluster analysis in SMLM. Voronoï tessellation, for instance, is a method for detecting clusters by setting a threshold for polygonal regions and generating density region contours. This technique has proven highly effective in identifying clusters of various shapes and sizes and can also be used to delineate cellular boundaries.

On the other hand, machine learning has also demonstrated potential in SMLM cluster analysis [160]. Unlike other techniques, this method relies on the nearest neighbours of a localisation as inputs to the model. The number of nearest neighbours is determined by the number of input neurons, which should correspond to the expected size of the largest cluster. The advantage of this approach is that the model can be trained on

simple circular clusters and detect clusters of varying shapes. However, a limitation of this method is that clusters with more localisations than the number of input neurons will be divided into multiple clusters. Additionally, the user needs machine learning knowledge and expertise to train the model effectively and interpret the results accurately.

In order to understand the functional mechanisms responsible for cluster formation in cell biology, qualitative assays provide an initial insight. However, quantitative assays are necessary to extract cluster, cell, or batch characteristics (such as the distribution of cluster sizes or the number of clusters per cell area). The reliability and robustness of both qualitative and quantitative assays are essential for successful and dependable cluster analysis in cells.

Except for the machine learning approach, quantitative assays have a common requirement where the user must determine one (in the case of Voronoï tessellation) or, more commonly, two input parameters. The standard method involves testing various input parameters that the analyst deems reasonable and selecting the one that appears to be the most suitable. However, user bias heavily influences this approach, as no statistical method is utilised to establish cluster parameters [130]. Additionally, SMLM datasets of the same target molecule frequently differ from cell to cell. Consequently, the cluster parameters cannot be generalised, yet analysts usually apply the same set of parameters to all samples.

To address the issue of user bias in SMLM cluster analysis, the Owen group proposed a solution using the Bayesian engine [54, 130]. Instead of employing a single set of parameters, the user inputs a range and step size for each parameter of the chosen cluster algorithm. The Bayesian engine assesses all cluster results, known as proposals, generated by combining the input parameters and selects the optimal proposal based on a designated statistical model.

The model, which is the heart of the Bayesian engine, proposed by the Owen group was used for this work. It assumes that a cluster has a Gaussian distribution, meaning that there are more localisations in the centre of a cluster than at its periphery. This

model also implies that the underlying ground truth cluster structure must not have a well-defined border. This approach is well suited for the spatial organisation of molecules at the cell membrane because it is thought that the molecules do not have to be strictly bound to a specific cluster [147]. Nevertheless, this model is also suited for well-defined structures such as clathrin-coated pits (CCPs) because of the aforementioned statistical process in SMLM measurements.

Undoubtedly, data analysis using the Bayesian engine, which involves thousands of proposals, is slower than a one-shot cluster analysis with a fixed set of parameters. However, evaluating the cluster proposals against a specified model enhances confidence in the analysis and the resulting conclusions. Additionally, the analysis becomes more resilient and reproducible.

5.1.2 Restructuring of BaClAva

While the original R code developed by Griffié et al. [54] effectively showcases the approach, it requires users to possess some computer programming experience, limiting user-friendliness. In order to alleviate this issue and enhance user experience, a graphical user interface (GUI) was created in this study to guide users in selecting cluster algorithm implementations and computational modes available to them. It is worthwhile to note, however, that even with this improvement, users with coding expertise will still be able to add additional cluster algorithms or models. Therefore, the code structure has been restructured into an Object-oriented programming (OOP) structure that uses classes, methods, and functions to add objects without changing the existing code structure. The resulting software, Bayesian Cluster Analysis and visualization application (BaClAva), provides an accessible solution for users with varying coding experiences, as shown in Figures 4.1 and 4.2.

As part of the original protocol published by Griffié et al. in 2016 [54], users were required to modify the code and data directory. However, these modifications are no longer necessary with BaClAva's OOP structure. The GUI requests user input, obviating the need to know where this information is kept and processed in the code. This

enhancement in user experience could elevate the adoption of the Bayesian engine-based cluster analysis approach since it curtails the probability of improper usage due to the original complexity of the method. The findings obtained from the Bayesian engine in the algorithm suggested by Griffié et al. were stored as text files in a complex folder structure. By storing most of the data within Hierarchical Data Format version 5 (HDF5) files for each dataset, BaClAva simplifies the storage of Bayesian engine results, reducing the storage size as well as the number of files required for each dataset. The HDF5 file contains all the information collected by BaClAva, comprising localisations, proposals, scoring, and post-processing results like cluster identification, cluster area, and number. This structure enhances data reliability since HDF5 files are less prone to unintentional alterations than text files. BaClAva also offers user-friendly CSV file exports enabling cluster analysis in preferred programmes such as Excel or Origin (Fig. 4.2).

Additionally, BaClAva improves the exporting options for cluster plots in various formats compared to the original work by providing improved options for exporting cluster plots. Lastly, it offers more extensive cluster characteristic plots for a batch of datasets, including histograms and density plots. BaClAva provides the possibility of customising these plots within the program's interface.

One of the primary shortcomings of the original work was its speed, with the authors reporting a typical analysis taking up to 24 hours [130]. The computational time required by the original work could limit its adoption; however, BaClAva addresses this issue by implementing parallelisation, resulting in significant speed-up in computation, as demonstrated in Figure 4.7 even on a 4-core computer. BaClAva can analyse a batch of datasets with specified input parameters in a single run. Generally, there are two possible options for parallelisation: parallel computation of several datasets or parallel computation of a single dataset by parallelising the looping through the parameter ranges. In BaClAva, the latter option was chosen due to the ease of parallelisation afforded by its two nested for-loops. Parallelising datasets and combining both options could be a promising solution to tackle speed limitations. However, implementing these options may require access to a computational cluster with sufficient resources to handle multiple large datasets. Without such resources, memory limitations may hinder algorithm performance when

processing multiple datasets simultaneously. As each dataset takes up a constant portion of the available memory, the Bayesian engine's computation also requires additional memory to perform its operations. In contrast, BaClAva has already shown significant improvements in speed and performance in implementing parallelisation on a single dataset due to its nested for-loops, making it a valuable tool for researchers interested in cluster analysis without the need for extensive computational resources.

The modular nature of BaClAva allows programmers to extend the software with other cluster algorithms, analysis tools, or parameter options according to future research needs. The application is separated into a front-end and a back-end so that changes made in the front-end, i.e., the GUI setup, will not unintentionally impact the back-end calculations. Regarding changing data, this can only be done in the application's back-end, which is hidden from the user, where programmers can add different cluster algorithms and post-processing calculations. BaClAva has four main workflows, as shown in Fig. 4.1, but its modularity allows it to be extended with additional workflows in the future. The front-end of BaClAva and the back-end for module 4 are written in Python, while everything else is written in R. However, future modules could be written in any programming language thanks to the many available Application Programming Interface (API) libraries. Furthermore, this Bayesian engine version is MIT-licensed software [82], making it freely available for use and distribution.

Proper error handling and messages are critical in any program to improve the user experience and make debugging easier. BaClAva encapsulates the original code and all additional code blocks in try-and-catch blocks to ensure proper error handling and messages. This technique allows errors to be detected and meaningful messages to be displayed to the user.

For instance, if a user inputs all necessary input via the GUI but mistakenly selects a file without tabular data, they will receive an error message from BaClAva. This message will instruct them to check the file. Similarly, error catches have been implemented for correct input formats for the cluster parameter ranges. Proper error handling and messages improve user satisfaction and make the program more user-friendly.

In summary, BaClAva offers user-friendly software that enables users without knowledge of the underlying code structure to analyse SMLM datasets. Furthermore, BaClAva offers a wide range of export and analysis options tailored to standard notebook and desktop computers. BaClAva also implements parallelisation to speed up computation, and its modular design allows easy addition of upcoming algorithms or tools. In the analysis process, users can quickly identify and resolve problems due to error handling and informative error messages. Overall, BaClAva provides a streamlined and accessible solution for SMLM cluster analysis.

5.1.3 Simulation work

The typical output of an SMLM experiment comprises numerous frames consisting of detected photons from blinking fluorescent dyes such as AF647. Blinking refers to fluctuating a fluorophore between a bright and dark state. This results in multiple detections of the same molecule within a region with a radius defined by localisation precision. Consequently, a single fluorophore can cause self-clustering, a phenomenon where a single fluorophore forms a cluster by itself. If two molecules are in close proximity, their clusters may overlap due to blinking, detection, and localisation processes.

Since blinking is a stochastic process, and each localisation varies in precision, self-clustering and clustering of two or more fluorophores may be indistinguishable. Clustering localisations is a crucial step toward future methods for determining the number of underlying molecules of interest in a cluster associated with a biochemical signalling process. Therefore, it is necessary to confirm through separate experiments or biochemical methods that each molecule of interest is labelled with only one fluorophore. In this study, molecules of interest were genetically modified to express a GFP tag. The SMLM-fluorophore AF647 was then conjugated to GFP using an anti-GFP nanobody at a 1:1 ratio (AF647:nanobody) [125].

The optimal labelling ratio must be determined in advance to generate high-quality SMLM data for cluster analysis, and protective measures must be incorporated into the

experimental staining procedure to decrease non-specific background labelling in cluster analysis. These protective measures include introducing blocking agents, additional extensive washing steps, or modifying the nanobody-dye complex incubation time to avoid non-specific fluorophore binding to unintended targets [86]. These measurements are critical when analysing membrane organisation since molecules of interest can be located inside and outside clusters, and non-specific background labelling can significantly interfere with membrane organisation analysis.

In contrast to including a pre-clustering step to eliminate signals from non-specifically labelled molecules, which is not always feasible, the focus should be on optimising the staining procedure to reduce background labelling and ensure that only the molecules of interest are labelled with the fluorophore. By doing so, localisations outside of clusters can still be included in the analysis, as they may represent molecules of interest that are not part of the pre-defined clusters [55, 130]. These localisations may provide valuable information about the underlying biochemical processes, contributing to a better understanding of the system. However, if the clusters of interest are located within established structures, a suitable filtering algorithm might be utilised to eliminate unwanted background signals.

Self-clustering and specific and non-specific background signals pose challenges for clustering, where each data point is considered an individual contribution. Simulations can provide insight into their contributions, as shown in Figure 4.11. The simulations were based on the assumptions of a round ground truth cluster size, a fixed field of view, and all fluorophores binding only to the molecules of interest, representing well-prepared samples with no erroneous labelling and a meticulously conducted SMLM experiment, which includes optimal laser-power, buffer conditions, and a perfectly aligned microscope setup. Although the ground truth clusters were circular, the detection clusters were not necessarily circular (see also Fig. 4.3), so the results were reported in terms of area rather than radius.

Based on the simulations, the main conclusion is that there seems to be a threshold for the relative localisation density between inside and outside clusters. This threshold

is where individual dye blinking becomes negligible. The threshold is observed as a decrease in the number of small clusters with an increase in labelling density.

The detected clusters were overestimated in number and underestimated in area below the threshold, corresponding to low fluorophore density or low molecule density. In such conditions, several sub-peaks in the area of the ground truth clusters are observed, resulting in the ground truth cluster being split into multiple clusters. Increasing the density ratio enhances the likelihood of detecting the actual cluster size of the underlying cluster. Therefore, selecting appropriate fluorophore labelling density and imaging settings, such as the number of frames and optimal imaging conditions, is critical for accurate cluster identification and reliable conclusions.

So far, all data points have been considered in the analysis. However, a common and simple practice is to exclude localisations from multiple blinking events that occur temporally and spatially close to each other. These localisations are grouped and counted as a single localisation, leading to a significant decrease in the number of small background clusters. The simulations were performed using the blinking characteristics of the AF647 dye, which is a well-studied dye. Still, Jayasnghe et al. [69] have demonstrated that changing the imaging technique can help detect smaller clusters and clusters with lower densities when comparing (*d*)STORM and DNA-PAINT. Therefore, the inherent threshold depends on the dye and the imaging technique. Other crucial factors include the chosen localisation algorithm, the grouping or alternative method, and the clustering technique.

Simulations have become a valuable tool for understanding limitations and optimising experimental conditions for SMLM cluster analysis. Using simulations, researchers can evaluate the contribution of various factors, such as labelling density, imaging settings, and clustering algorithms, to the final cluster results. Informed decisions about the experimental setup can be made using simulation insights, leading to improved accuracy in the analysis by avoiding biases and errors. Ultimately, simulations can lead to more accurate and reliable cluster analysis, which is essential for understanding the underlying biological processes.

5.1.4 Cluster Analysis of SMLM of Cells

A simulation is much easier to conduct than a microscopy experiment on a cell. When conducting simulations, particular parameters can be the focus while keeping others constant or eliminating them to simplify the experiment. For instance, in the previous section, the laser power was held constant to maintain the blinking conditions of the selected dye AF647. Additionally, the clustering areas were perfectly circular, within a flat and homogenous area without compartments or non-target constituents. However, the situation in cells is considerably different, as known and unknown cellular processes and structural boundaries impact cellular molecules' temporal and spatial organisation. All of these factors could potentially interfere with the target process [133]. Furthermore, the blinking behaviour of dyes is affected by laser power and buffer conditions, with the latter changing over time. Studies have also suggested that the surrounding environment may influence the blinking behaviour [86].

For the classification of unknown target molecules, it is imperative to use well-established and reliable controls. Accordingly, the characteristics and restrictions of the probes have been comprehensively examined in the cellular milieu. This approach leads to a more precise comprehension of the clustering outcomes of the unknown target molecule by contrasting its behaviour to that of the recognised controls. Additionally, the choice of dye used can significantly impact the results when conducting cluster analysis. Binding and blinking properties can vary depending on the selected antibodies, nanobody and fluorophore, which can influence the clustering outcome. In order to minimise these effects, it is recommended to use the same antibody- or nanobody-fluorophore combination for both the controls and the target molecules. As a result, clustering differences can be attributed to the underlying target molecules rather than to dye properties such as binding or blinking. In our study, all probes had the same structure, consisting of a molecule of interest genetically tagged with GFP and then coupled to an anti-GFP-nanobody labelled with AF647. Therefore, any observed differences in clustering were due to the molecules being studied rather than the probes themselves.

As a positive control for clustering at the cell membrane, clathrin-coated pits (CCPs) have been extensively studied, and they appear as round formations with a radius of

approximately 80 nm (0.02 μm^2) in 2D experiments [144]. Moreover, their size can be measured independently of the clustering procedure, for example, by examining the line profile of CCPs in SMLM reconstructed images, making them an excellent target for clustering (Fig. 4.13). However, identifying an appropriate negative control for clustering at the cell membrane is challenging. Most potential candidates are typically involved in natural signalling processes, leading to cluster formation for downstream signalling due to the spatial and temporal reorganisation of molecules [12, 47, 72]. For instance, Griffié et al. [55] conducted a study which showed that many signalling molecules at the T-cell membrane form small clusters even without activation, possibly accelerating the signalling process upon activation. Moreover, cytoskeletal or structural factors may compartmentalise molecular organisation at the cellular membrane [81].

When selecting a suitable negative control for clustering at the cell membrane, choosing a target with minimal influence from cellular processes on its spatial organisation is crucial. An ideal probe for this purpose is Glycosylphosphatidylinositol (GPI). GPI is an artificially introduced target without natural interaction partners in cells [89]. It is only anchored to the outer membrane of the plasma membrane. Therefore, it experiences minimal influence from the cytoskeleton or other structures within the cell that might affect other membrane molecules and their organisation. Therefore, GPI is an excellent negative control candidate for membrane clustering experiments.

The reconstructed images obtained from (*d*)STORM data and their corresponding cumulative distribution functions (CDFs) for the clusters are clearly distinguishable for both controls, as demonstrated in Figure 4.12. The controls serve as reference points for both random distribution and clustering scenarios, and clusters are detected in both scenarios with varying distributions, particularly when sampling across multiple cells (Fig. 4.12A,B). In order to ensure the accuracy of the cluster analysis, all regions of interest were carefully selected to ensure a flat membrane and were not located at the 2D cell border. This decision was made because the chosen model and Bayesian engine were not specifically designed to detect borders and holes in 2D. The resulting data highlights the efficiency of the clustering analysis for SMLM data, as the two controls are distinctly separated (Fig. 4.2D).

When investigating a molecule of interest like CD95 (see Section 2.1), a transmembrane receptor with an unknown distribution within the plasma membrane, it is expected to fall somewhere between the two extremes represented by the GPI and CCP controls. If the molecule has a distribution similar to GPI, it is less likely to be clustered. However, if it resembles the CCP control, it is expected to have more clustered regions. The resulting CDF in Fig. 4.12D supports the idea that CD95 forms small clusters at the plasma membrane containing approximately 0.54 molecules per nm^2, as the CD95 receptors have a cumulative distribution function that falls between the two controls. This is consistent with previous studies that have shown non-activated CD95 to exist as monomers, homodimers, and homotrimers at the plasma membrane [88].

In summary, the developed BaClAva analysis procedure can identify small molecule clusters in SMLM data, which is essential for understanding molecule behaviour in plasma membranes like CD95. However, proper controls are crucial for accurate clustering interpretation; clathrin-coated pits (CCPs) and GPI are commonly used. Overall, BaClAva is an effective tool for analysing and characterising plasma membrane molecule distributions, but control selection is crucial for accurate interpretation.

5.2 Tethering

The experimental requirement of stabilising cells in space to facilitate experiments within a liquid environment while preventing contact with glass surfaces is not a novel concept [9, 104]. However, while cells in their natural environment are typically surrounded by other cells, investigating and comprehending biochemical processes can be highly complex when dealing with a large population of cells [86]. This complexity arises from the numerous interactions between the cells and the substrate on which they are cultivated [33, 44, 123, 154]. Hence, it is more rational to prioritise the study of an individual cell, given its inherent complexity and the multitude of aspects that still evade complete understanding. However, in most cases, the substrate is a glass surface, which can induce distinct cellular responses compared to the cell's original natural environment

[33, 46].

Consequently, researchers started exploring methodologies to visualise cells without adhesion to a substrate through adhesion molecules [167]. Instead, the objective was to have the cell surrounded by liquid and secured in its position using a suitable mechanism. Arthur Ashkin devised optical tweezers, also called optical traps, to immobilise particles in three-dimensional space [9]. Optical tweezers employ tightly focused lasers to trap dielectric particles. Ashkin was awarded the Nobel Prize in Physics in 2018 [77]. Optical tweezers have already been employed in single-cell investigations, including their application in combination with (*d*)STORM [34, 164]. However, the experimental laser setup becomes increasingly complex.

An alternative approach to trapping cells involves biochemical means. In the same way, as with optical tweezers, cells are trapped and surrounded by fluid. However, in this case, they are tethered to a supported lipid bilayer (SLB) using DNA strands. Consequently, this method can be combined with various microscopy techniques without the need for additional laser components in the microscopy setup. In addition, unlike optical tweezers, DNA tethering facilitates the investigation of cellular responses to molecules introduced into the supported lipid bilayer. [19, 150]

The study by Selden et al. demonstrated in 2012 [142] that the DNA strands on the cell side must have a minimum length of 60-80 bases to overcome glycocalyx, which forms the extracellular matrix of a cell containing glycolipids and glycoproteins [103]. This finding highlights the significance of DNA strand length in successful cell tethering and subsequent investigation of cellular responses to stimuli introduced into the SLB. Consequently, this technique allows for the controlled examination of cell reactions to external stimuli. Taylor et al. explored this method in 2017, examining its application to induce the clustering of molecules or detect the clustering of membrane molecules triggered by intracellular signals [150].

In their paper, the results presented by Taylor et al. demonstrated the reliability of tethering through imaging data, particularly in their Single Particle Tracking experiments.

Furthermore, they successfully demonstrated the viability of employing SNAP and CLIP-tag techniques to link DNA strands to a particular structure of interest or substrate. As a result, they established a versatile platform for studying clustering involving specific membrane proteins. [150]

Bloch et al. published a paper in 2016 [19] demonstrating that tethering DNA to support a SLB enabled the movement of vesicle-like structures. The authors employed cholesterol to incorporate DNA strands into the supported lipid bilayer and vesicles, enabling controlled tethering and subsequent investigation of vesicle dynamics. Due to the absence of a glycocalyx in the vesicles, DNA strands were maintained at uniform short lengths.

Hence, this study adopted a synergistic approach by integrating the methodologies employed in the above DNA-tethering studies. The study integrated the DNA characteristics and the method of cell-side coupling using self-labelling proteins, as elucidated by Taylor et al. [150], incorporating the DNA-strand length parameters, as determined by Selden et al. [142], into the experimental design. Furthermore, the study used the linking technique for the supported lipid bilayer, as demonstrated in the research conducted by Bloch et al. [19]. This combined approach allowed for a comprehensive exploration of the subject matter, considering the various aspects and insights from these previous studies.

Figure 4.14 illustrates the final experimental setup configuration. In this arrangement, a 20mer DNA strand is connected to the supported lipid bilayer using cholesterol, while a 100mer DNA strand is linked to the cell via GPI-Halo. The purpose of tethering in this study was not to induce or detect any specific cellular processes. Instead, the role of the DNA was solely to serve as an anchor, connecting the cells to the supported lipid bilayer. Cholesterol is a natural constituent of the cell membrane. Hence, GPI was selected as the biochemically introduced membrane anchor, considering previous investigations that established its minimal interaction with cellular molecules [89], as discussed in the previous section 5.1.4. The choice of GPI ensures successful cell tethering and enables the investigation of membrane protein clustering.

The results presented in section 4.5 provide strong evidence for the effectiveness of the derived DNA-tethering method. The figures 4.17 demonstrate that the cells were successfully anchored to the lipid bilayer and maintained their tethering connection throughout the experiment. Notably, the tethered cells exhibited signs of robust health, as indicated by the development of filopodia, which extended into the surrounding environment 4.17. This observation indicates that tethered cells could interact with their surroundings while maintaining stable anchoring. Such conditions are crucial for facilitating the controlled activation of cell membrane receptors. Furthermore, the DNA-tethering connection remained intact even when exposed to high laser power for extended periods during Single-Molecule Localisation Microscopy (SMLM). This observation underscores the stability and reliability of the DNA-tethering approach, particularly under challenging imaging conditions.

Studying the spatial organisation of CD95 receptors during their interactions with CD95L was the final objective of this study. In order to establish a well-defined binding site between the ligand and receptor, the CD95L construct was incorporated into the supported lipid bilayer. As a result, the region for the ligand-receptor interaction was the lower portion of the T98G cell membrane, where the cell was also tethered to the SLB. By preventing the mixing of the ligand with the surrounding fluid, ligand-receptor interactions were restricted to the desired area and not distributed over the entire cell surface. This approach was chosen to avoid the inherent complexity and challenges associated with imaging and analysing ligand-receptor interactions in a three dimensional (3D) environment. The controlled setup allowed for a more focused and manageable investigation than experiments involving liquid ligands.

As depicted in Figure 4.14, the cells are securely anchored to the lipid bilayer via DNA strands, establishing a robust connection between the cells and the bilayer. This tethering approach is essential because the ligand-receptor interaction alone lacks sufficient strength and reliability to withstand routine fluid exchange procedures, fixation and staining protocols, or the high laser powers employed in SMLM imaging. The DNA-tethering mechanism ensures the stability and reliability of the cell-bilayer connection, enabling subsequent experiments and analyses. As a result, the tethering procedure

allows for measurements both in the presence and absence of the CD95 ligand. Cells cannot maintain a secure and stable connection without the linkage to the supported lipid bilayer (SLB).

The analysis of the reconstructed images consistently showed that cells exhibited different distributions of data points irrespective of experimental conditions. Specifically, some cells displayed a dense distribution of data points, indicating a more compact spatial arrangement of CD95 receptors. In contrast, other cells displayed a dispersed distribution of data points, suggesting a relatively spacious CD95 receptor organisation. Furthermore, the cluster analysis results provided further insights into the spatial organisation of the localised data points. The analysis revealed a wide range of clustering sizes observed across all experimental conditions, supporting the previous observations. However, no significant differences in clustering patterns were observed that would allow for definitive and conclusive interpretation.

These findings suggest that the spatial organisation of CD95 receptors under the studied conditions was complex and multifaceted, indicating the need for further investigation. For deeper insights, it is necessary to investigate the spatial distribution of the CD95 ligand and examine its potential influence on the spatial organisation of its receptor. However, it was impossible to examine its spatial distribution due to limitations in the experimental conditions regarding the ligand.

There are several potential reasons for the inconclusive results obtained from the data analysis. Firstly, the T98G cell line used in this study is derived from a human glioblastoma multiform tumour from a white male [145]. As is common in cancer cells, T98G cells are polyploid, meaning some chromosomes exist in more than two sets [143]. This inherent genetic variability within the cell line could contribute to the complexity of the obtained results. However, the T98G cells in this study were transfected with the GFP-tagged CD95 receptor to ensure stable and permanent expression [82]. Nevertheless, it would be beneficial to examine the morphology of the cells and employ a sorting method to assess whether the transfected T98G cells form a homogeneous group.

Another factor to consider is the timing of cell fixation, as the cells were fixed 15 minutes after being tethered to the bilayer. Consequently, the cells appeared predominantly round, creating challenges distinguishing them from dying cells. This limitation is significant because it hinders cell health and viability assessment before fixation with prewarmed 4 % PFA and 0.2 % GA.

Moreover, altering the living conditions of cells, as done multiple times during the experimental procedures, inevitably induces stress. This stress factor may have contributed, at least in part, to the observed variability in each experimental condition. Furthermore, differences in settlement status within the channel may also have influenced the observed variability. Variations in ligand activation levels could affect cells in channels with CD95 ligands.

Exploring additional options, such as expanding the dataset or incorporating live-cell imaging techniques, could provide further opportunities to address the complexities encountered in this study. In particular, adopting live-cell SMLM imaging, in conjunction with temporal cluster analysis methods, presents a promising strategy to investigate dynamic changes in the spatial distribution of the CD95 receptor and its ligand over time. Capturing real-time images of receptor and ligand distributions could provide a more comprehensive understanding of their spatial organisation and interactions. Furthermore, conducting co-cluster analysis under conditions where the ligand is present would provide valuable insights into the interactions and dynamics between the CD95 receptor clusters and the ligand. These analyses, when performed in combination, have the potential to uncover the underlying factors that contribute to the observed cluster behaviour and spatial arrangements across all tested conditions. Integrating these analyses in a complementary way has the potential to reveal the underlying factors that lead to the observed cluster behaviour and spatial arrangements across all tested conditions. The acquisition of such detailed insights would significantly contribute to advancing the understanding of the spatial and temporal organisation of the CD95 receptor and its interaction with CD95L.

In summary, this study provides evidence for the successful application of DNA teth-

ering as a technique for SMLM imaging of membrane receptors on isolated cells in a fluid environment. The results demonstrate the feasibility of using this approach to obtain high-resolution data suitable for subsequent cluster analysis. However, despite the thorough analysis, no significant findings regarding observable differences in clustering patterns across the experimental conditions examined were revealed. This lack of definitive insights can be attributed to the significant variability observed in the acquired data. The fluctuations within the data points, possibly influenced by factors such as cell-to-cell variability and variability in the experimental procedures, likely contributed to the challenges in identifying distinctive clustering patterns. Nevertheless, these findings underscore the challenges and complexities associated with acquiring and analysing SMLM data, particularly in the context of cluster analysis. Consequently, they underline the need for further investigation to unravel the underlying factors that govern the clustering behaviour of membrane receptors under different conditions.

6

Outlook

This chapter offers a perspective on potential future directions and research prospects concerning BaClAva and cell tethering through DNA strands.

6.1 BaClAva

This investigation focused on cellular membranes, with the findings demonstrating the efficacy of simulations in optimising cell experiments for Single-Molecule Localisation Microscopy (SMLM). Simulations and cell experiments throughout the investigation were limited to a two-dimensional plane. Preparing the cell samples involved significant care to ensure the membrane was as flat as possible A. Consequently, the only ones subjected to observation and analysis were cell membrane regions where holes were not detectable in total internal reflection fluorescence (TIRF) mode.

When imaging on glass, the cell membrane is not perfectly flat [22, 87]. Instead, it is uneven, with areas closer to the glass surface and others exhibiting more significant curvature, influenced by the molecular composition, cell behaviour such as movement in a specific direction, and surroundings, including contact with other cells. In 2018, Jonas Ries and colleagues proposed a technique to extract three-dimensional data from two-dimensional raw data [90]. Combining this technique with Bayesian analysis in three dimensions can offer novel insights into molecular distribution on the cell membrane.

More specifically, this combination of methods could help determine whether clustering is limited to two dimensions and solely dependent on the reaction between molecular partners or if cell membrane curvature also contributes to the process. Griffié et al. [55] developed a 3D extension of Bayesian cluster analysis in 2017. However, as this research was exclusively centred on the two-dimensional cell membrane, the proposed 3D algorithm was not integrated into BaClAva.

BaClAva is not limited to clustering cell membranes and could be applied to clustering other molecules of interest, including septin structures and calcium channels in neurons. One approach to testing the cluster detection quality of these molecules involves simulating their theoretical or currently known spatial organisation using FluoSim [83]. Once data has been generated, Bayesian Cluster Analysis and visualization application (BaClAva) can process it. The clustering results may indicate that adjustments are necessary for the Bayesian Engine model to accurately identify the expected organisational structure of the target, as the model is designed for detecting Gaussian clusters. Furthermore, utilising clustering techniques within the cell, particularly in combination with the aforementioned 3D extension, could facilitate the identification of areas of increased density or absence of a target molecule.

This investigation concentrated on AF647, the most frequently used SMLM dye. Nonetheless, as discussed, DNA-PAINT could significantly improve cluster detection due to its high localisation precision ($\sim 1\,\mathrm{nm}$) and sub-5-nm spatial resolution [141]. DNA-PAINT could be particularly beneficial for detecting smaller clusters and distinguishing clusters in close proximity. Furthermore, DNA-PAINT could be a valuable tool for co-clustering analysis, as it allows the use of the same dye for detecting different targets, with the only difference being the DNA sequence for each target. By contrast, PALM or (*d*)STORM necessitate using a different dye with distinct imaging properties, leading to disparate localisation precisions for the molecules under investigation.

An exciting application of clustering is in live cell imaging, which was first demonstrated by Griffié et al. [53]. Compelling studies can be conducted using suitable dyes, such as observing cluster development after activation of the target molecule. Additionally,

studying the movement of grouped molecules is of significant interest. Questions that arise include how quickly clusters form, whether they existed before activation, whether they grow, and if there is a maximum cluster size. It is also interesting to investigate whether preexisting and activated clusters are stable and, if so, when and how they return to their original molecular distribution. Moreover, employing supplementary dyes would allow tracking and analysing the spatial arrangement and interaction of various target molecules, such as a receptor and its ligand, over a period of time.

The incorporation of data clustering and tracking can be a powerful method for gaining a deeper understanding of the spatial and temporal organisation of molecules within living cells. By analysing the clustering of tracking data, one can identify specific locations where a molecule remains stationary or where multiple molecules form clusters. Exploring such regions and their persistence can provide valuable insights into the dynamic behaviour of molecules within the cell. Another intriguing application of this integration is the ability to layer tracking data from one target molecule over clustering data from another. This enables the examination of the behaviour of the tracked molecule upon contact with the other.

The integration of clustering and tracking data allows the examination of how molecules behave when in contact. For instance, this approach can be used to investigate the interaction between the clustered transmembrane molecule CD95 and its tracked ligand CD95L. Comparing the movements of the tracked molecules before and after binding to the receptors and simultaneously analysing the development of the clusters can lead to a more comprehensive understanding of the interaction between the two molecules. Additionally, further research could explore the interaction between clustered membrane molecules and the underlying cytoskeleton of the cell. Ultimately, the integration of clustering and tracking data presents novel possibilities for investigating the behaviour of molecules in living cells.

BaClAva operates as self-contained software and is compatible with Windows, Linux, and Apple platforms (excluding ARM-based chips). However, chemistry, biology, and medical scientists typically use Fiji/ImageJ for data analysis [29, 138]. Hence, integrating Ba-

ClAva with Fiji/ImageJ provides an excellent opportunity to introduce more individuals to unbiased and model-dependent clustering of SMLM data. Two methods can be implemented to transform BaClAva into a Fiji/ImageJ plugin that utilises Java as its primary programming language [29]. The first method involves rewriting the GUI models in Java, which is possible because the GUI package [128] has a Java version. The second method involves using a Java-to-Python API that executes existing BaClAva software from a Java script. In either case, the backend of BaClAva will not require any modifications as both Java and Python have APIs that can interface with the R programming language.

Overall, BaClAva is a valuable tool for life sciences research that can facilitate a wide range of interesting studies. Its current capabilities include unbiased and model-dependent clustering of Single-Molecule Localisation Microscopy (SMLM) data, but it can be further enhanced with additional clustering and analysis options to provide scientists with a deeper understanding of their data. In addition, its integration with Fiji/ImageJ could increase its popularity and impact. Finally, with its robust clustering and analysis capabilities and user-friendly GUI, BaClAva has the potential to contribute to the field of SMLM significantly. BaClAva is a valuable resource that could greatly benefit the life sciences community.

6.2 Cell tethering

The outcomes of the presented tethering experiments illustrate the capacity of tethering living cells to a supported lipid bilayer (SLB) to enable controlled CD95 activation and facilitate fixed cell (*d*)STORM imaging. Unfortunately, as previously discussed in section 5.2, the clustering findings were inconclusive. There are, however, multiple potential directions for further investigation. Specifically, using live cell SMLM techniques is of particular interest to deepen our understanding of CD95 receptor organisation and the possible changes in its organisation upon activation. According to the experimental results, the receptor forms clusters without ligands. However, the stability of these clusters remains uncertain, and they will likely undergo temporal fluctuations in size [147]. It would be valuable to determine the extent of these fluctuations. An exciting

prospect would be to observe the behavioural patterns of individual CD95 receptors. The most intriguing research questions centre around the potential transition of receptors between clusters and the duration of their residence within a distinct cluster.

Furthermore, how does this behaviour manifest when the ligand activates the receptors? Dye labelling of the ligand would help answer this question. However, this particular aspect could not be pursued due to limitations in the present study. As a result, several questions emerge, primarily concerning the potential correlation between the number of ligands associated with a cluster and the size of said cluster. Considering the findings from previous studies [55, 130], which showed an increase in cluster size following activation of the target cells, it becomes intriguing to explore whether this phenomenon is also observable in the context of CD95 receptor clusters. Furthermore, it would be interesting to determine whether the ligand binds exclusively to a single receptor or changes its binding partners.

In the presented experiments, imaging was conducted following the tethering of cells and a meticulous examination of the supported lipid bilayer for any ruptures close to the cell. Unfortunately, this approach also meant that the cells had already undergone ligand stimulation or had the opportunity to acclimate to their new environment before the imaging process commenced. Consequently, it would be advantageous to precisely determine where a cell will tether to the SLB before introducing the cells. Knowledge of this location would enable the observation of the initial contact between CD95 receptors and the supported lipid bilayer. One possible approach could involve reducing the area of the SLB to match the camera's field of view, such as by using a grid structure created through a photolithographic technique [23]. This reduction in the area would facilitate a thorough inspection of the bilayer within the designated region before initiating the imaging process. The microscope setup could be adjusted to focus on the intended observation plane. Once the preparations are complete, the imaging process can be initiated by introducing the cells, thereby facilitating the complete capture of the sequence of events starting from the initial anchoring of the cells to the supported lipid bilayer. This comprehensive capture includes observing CD95 receptor contact with the ligand if it has been introduced to the SLB.

Moreover, visualising the complete process in three dimensions also holds significant advantages. A 3D imaging approach enables a comprehensive examination of CD95 receptor distribution throughout the entire cell membrane. By examining the spatial arrangement of these receptors in three dimensions, it becomes possible to acquire valuable insights into any potential adaptations that might arise in their distribution. Such adaptations could occur as a consequence of the cells being tethered to the SLB and the potential activation induced by CD95L.

As demonstrated in the results section, cells developed filopodia over time. Notably, filopodia possessed a detectable GFP signal emitted from GFP-labelled CD95 receptors, as illustrated in Figure 4.17. Filopodia formation and extension are fundamental processes that enable cells to explore and interact with their surrounding microenvironment actively. Consequently, conducting an in-depth investigation of the behavioural dynamics of CD95 receptors within these specialised structures presents a promising avenue for future research endeavours.

T98G cells were the primary focus of this thesis. However, the tethering technique employed in this study has broad applicability to various cell types capable of enduring a non-adherent state. This method could particularly benefit cell types that naturally exist in fluidic environments, such as red blood cells and macrophages. By anchoring these cell types to a supported lipid bilayer, a novel platform is created to investigate their molecular organisation and interactions within their immediate surroundings while immersed in a fluid medium at the same time.

This unique experimental setup offers distinct advantages, enabling the visualisation and analysis of dynamic cellular processes and the intricate biochemical events that occur when these cells interact with external stimuli. These processes, which may otherwise be challenging to study, can be observed and examined in real-time, shedding light on the underlying functional mechanisms of red blood cells and macrophages.

Extending the tethering technique to other cell types beyond T98G cells presents new

research possibilities. This expansion is potentially impactful in that it could enrich our understanding of cellular dynamics and signalling processes across a wide range of biological specimens. Moreover, such an approach can potentially contribute to multiple fields, including cell biology, immunology, and biomedical research.

7

Conclusion

In this thesis, an innovative software solution for cluster analysis named BaClAva was presented, which is based on a deterministic Bayesian engine. This software aims for user-friendliness and efficiency and was published in Kutz et al. [82]. This software builds upon the research conducted by Rubin-Delanchy et al. [130], Griffié et al. [54], and Pike et al. [117].

Moreover, the successful demonstration of the efficient tethering of T98G cells to a supported lipid bilayer (SLB) in live-cell imaging represents a significant achievement. This accomplishment enables the utilisation of Single-Molecule Localisation Microscopy (SMLM) imaging techniques on tethered cells, allowing one to study cellular dynamics upon activation by external.

The BaClAva software developed following Griffié et al. [54], and Pike et al. [117] research represents a significant advancement. Through simulations, the selected code modifications were demonstrated to enhance the speed and functionality of the original work. Notably, these improvements facilitated a more comprehensive analysis of the performance of the three selected cluster algorithms focused on: Ripley's-K-based clustering, DBSCAN, and ToMATo. Applying BaClAva in SMLM imaging successfully identified and quantified clusters in diverse sample types, including random, clustering, and novel

samples.

The employment of DNA strands for tethering cells to an SLB offers a powerful tool for examining specific cell activation events in isolation from potential influences from other cellular stimuli. Drawing upon the research conducted by Selden et al. [142], Taylor et al. [150], and Block et al. [19], a novel approach to DNA tethering was developed in this study. The viability of the tethering process was established by examining cell behaviour over time. Additionally, by employing staining techniques and capturing images in SMLM mode, the obtained data convincingly demonstrate the feasibility of imaging tethered cells using the high laser intensities requisite for SMLM imaging.

Before BaClAva was published in Kutz et al. [82], researchers seeking to conduct Bayesian-engine-based cluster analysis had to write their own code or use the code-based solution offered by Griffié et al. [54]. Regrettably, both required programming proficiency from the user or, at the very least, involved modifying lines of code. In contrast, BaClAva is a software solution with a user-friendly graphical user interface (GUI) that guides the user through the entire process. As a result, users are relieved of the burden of understanding the details of the underlying code or the precise operations of the data processing system. Furthermore, they are no longer troubled with concerns about inadvertently altering the code or data, thus leading to a higher degree of confidence in adopting Bayesian-engine-based clustering techniques. Further, the enhanced storage solution minimises storage space while maintaining high-precision data and effectively protects against user-induced data alterations.

The advancement of methodologies for investigating particular signalling pathways or their constituents holds significant importance in cell biology and medical research. Cell tethering is an attractive solution as it empowers researchers to regulate activation. Within the context of this thesis, the DNA-tethering technique, introduced by Selden et al. [142] and Taylor et al. [150], was modified to facilitate ease of implementation without compromising the capability for long-term cell tethering or the ability to withstand the high laser powers used in SMLM.

Certain limitations were encountered during the development and study of BaClAva. Firstly, the primary emphasis was placed on three specific cluster algorithms without exploring additional published algorithms or investigating alternative implementations of a chosen cluster algorithm. The object-oriented architecture of BaClAva makes it easy to integrate alternative cluster algorithms.

Regarding simulations, attention was exclusively directed towards circular clusters, as the underlying assumption was that membrane molecules within a cluster adhere to a two-dimensional Gaussian distribution. Nonetheless, it is worth considering that alternative cluster structures may also possess considerable interest and relevance.

The cluster results include all localisations fed into the software. Consequently, these results do not yield an accurate count or density of the molecules of interest within the clusters; instead, they provide information on the number of localisations and their corresponding density in the clusters. It is worth emphasising that the determination of the actual underlying number of molecules and their density remains an ongoing investigation within the field, particularly considering the statistical nature of (*d*)STORM, the SMLM technique employed in this thesis and the challenges involved in the localisation and clustering process. Consequently, this study did not specifically address determining the underlying molecule count and density, leaving it for further research.

The DNA-tethering study had certain limitations. Firstly, the study focused solely on tethering T98G cells to an SLB and imaging the spatial distribution of the CD95 receptor. Due to encountered challenges in establishing the tethering method, no alternative cell lines or targets were investigated.

The final experiment that examined the spatial organisation of the CD95 receptor after activation with its ligand was inconclusive. This thesis did not resolve this issue or provide conclusive evidence of successful CD95 receptor activation. Comprehensive investigations to address these concerns were considered. However, it was determined that such investigations would significantly extend the timeline. Consequently, it was deemed suitable to conclude the study at this stage.

The techniques introduced in this thesis offer significant potential for biochemical and medical research, with promising prospects for further development, particularly regarding the features of BaClAva. Potential extensions to BaClAva encompass a range of options, including exploring the third dimension, incorporating live-cell SMLM imaging, co-clustering analyses, and incorporating additional cluster algorithms or tools to enhance the analysis of obtained cluster results.

Due to their numerous functions, there is still much room for investigating the molecular spatial organisation within cells. However, the software solution BaClAva represents one solution that has the potential to contribute significantly by providing model-based cluster outcomes. BaClAva has been demonstrated to identify CD95 receptor clusters in cells grown on glass slides and in cells DNA-tethered to an SLB. Consequently, it is a valuable tool for analysing the spatial distribution of molecules of interest in various imaging settings.

The tethering approach can be combined with various fluorescence imaging techniques, provided DNA-tethering remains intact during preparation and imaging. DNA-tethering also offers a novel avenue for investigating non-adherent cells, eliminating the need to modify the existing microscopy setup. This capability opens up numerous opportunities to study how cells respond to external stimuli.

In conclusion, this thesis introduced BaClAva, a user-friendly software solution for cluster analysis in SMLM data. BaClAva provides researchers a valuable platform for examining and interpreting complex cluster patterns, revealing molecular organisation and spatial distribution. Additionally, this thesis presented an innovative approach to tethering cells to an SLB, enabling live-cell imaging and SMLM studies. The successful demonstration of efficient DNA-tethering to an SLB for live-cell and SMLM imaging showcased the potential of BaClAva and the DNA-tethering method in opening up novel research avenues. Therefore, these techniques can contribute to a better understanding of cellular processes, spatial organisation, and molecular interactions.

8

Publication

Part of the research on BaClAva was published in the following publication:

[82] Saskia Kutz, Ando C. Zehrer, Roman Svetlitckii, Gülce S. Gülcüler Balta, Lucrezia Galli, Susanne Kleber, Jakob Rentsch, Ana Martin-Villalba, and Helge Ewers. An efficient gui-based clustering software for simulation and bayesian cluster analysis of single-molecule localization microscopy data. Frontiers in Bioinformatics, 1: 47, 2021. ISSN 2673-7647. doi: 10.3389/fbinf.2021.723915.

Appendix A

Protocols

The following part consists of all protocols used for the microscopy experiments. The organisation of the protocols depends on their use. For example, there are protocols for cell culture, glass cleaning, fixed-cell staining on glass slides, channel preparation, bilayer preparation, live-cell tethering, and tethered-cell staining. They are stated in an itemised fashion to facilitate their future use.

A.1 Cell culture

Even though the cell lines used for this thesis needed different cell culture media, their cell culture handling was similar. Thus, the following protocols are generalised.

A.1.1 Keeping cells in the cell culture

The following protocol describes how to keep cells healthy and alive in the incubator. All cell lines used were kept in an incubator at 37 °C and 5% CO_2. For optimal results, the cells were regularly split, keeping their cell density around 50% to not stress the cells due to overcrowding or malnutrition.
The standard protocol for splitting:

1. warm up cell medium in a waterbath to 37 °C
2. warm up Trypsin-EDTA to room temperature (ca. 27 °C)

3. check cells under microscope
4. prepare cell medium, Trypsin-EDTA and PBS under the hood
5. suck up cell medium
6. wash cells 1-2 times with PBS, depending on the number of dead cells
7. keep PBS on cells for 5min to facilitate the next step
8. suck up PBS and add 1 ml of Trypsin-EDTA
9. incubate the cells at 37 °C in the cell culture incubator
10. check if cells have detached on the microscope
11. gently knock against the cell dish or flask to detach even more cells
12. stop the reaction by adding 5ml of cell medium
13. use a pipette to detach and separate all cells
14. keep a third of the fluid volume
15. add fresh medium to reach the final volume
16. incubate the cell for further use in the cell incubator

A.1.2 Cell seeding on glass slides

This protocol is designed for seeding cells on standard microscopy glass slides. Depending on the cell line and the type of glass slides, the number of cells per millilitre might need to be adjusted. The first part of the protocol is identical to section A.1.1. For properly cleaned glass slides see A.2.1

1. follow the protocol steps 1-13 in section A.1.1
2. take up to the entire fluid and move it into a falcon
3. add fresh medium to the remaining cells to reach the final volume
4. incubate the cell for future use in the cell incubator

5. count the cells density
6. add $8 \cdot 10^4$ cells/µl to the microscopy glass slides
7. add enough cell medium for a final volume of 2 ml of cell medium
8. incubate the cells for at least 12 h in the cell incubator

A.1.3 Transfections

The system for cell transfection depends on the cell line and plasmid with the protein of interest. Thus, transfection via lipofectamine is a valid option for CV-1 cells. On the other hand, electroporation results in higher transfection efficiency and less cell mortality for T98G.

GPI-GFP transfection via lipofectamin. Transfection via lipofectamine follows the standard protocol offered by Invitrogen/Thermofisher **ref ???**.

1. prepare mixture 1: OptiMEM and Lipofectamine 3000
2. prepare mixture 2: OptiMEM, GPI-GFP vector, and P3000
3. vortex each, mix and vortex again
4. incubate combined mixture for 15 min
5. transfect cells with 250 µl of the prepared mixture approximately 30 min after cell seeding
6. incubate the cells for 24 h at 37 °C in the cell incubator

GPI-Halo transfection using the Neon®System. The Neon®System (Thermo Fisher Scientific) was employed for plasmid transfection. This technology utilises electroporation, using a pipette tip as an electroporation chamber, to introduce plasmid DNA into cells. Electroporation involves the application of an electric shock to cells, temporarily increasing cell membrane permeability and facilitating the uptake of plasmid DNA into cells. The transfection reaction required $5.6 \cdot 10^5$ cells in a 10 µl volume. Parameters for electroporation were customised to the specific cell type, as shown in Table A.1 for T98G cells. Following transfection, the cells were cultivated in 1 ml of growth medium at 37 °C and 5 % CO_2.

Parameter	Value
Pulse voltage	1050 V1050V
Pulse width	40 ms
Pulse number	1

Table A.1: Parameters for T98G transfection with Neon® System

A.2 Glass cleaning and channel preparation

For optimal preparation of supported lipid bilayers well or performance of SMLM experiments, contaminations or impurities cannot cover the glass slides and channels. They would cause imperfect bilayer formation or background signal in the SMLM experiments, reducing the data quality. The following protocols describe the cleaning procedure of the glass slides and top part for the channels and the channel construction.

A.2.1 Glass cleaning

The following cleaning procedure removes contaminations and impurities on the surface of the glass slides, and the surface becomes more hydrophilic. A ceramic holder for glass slides facilitates the cleaning procedure. The slides can be stored for future use. Optimal results are obtained by repeating the plasma cleaning step before use.

1. incubate glass slides in 2% Hellmanex in an ultra-sound bath for at least 15 min
2. rinse glass slides 5 times with MilliQ water
3. incubate glass slides in 70% ethanol in an ultra-sound bath for at least 15 min
4. dry the class slides under a hood (15-30 min)
5. use a plasma cleaner for at least 15 min
6. store the glass slides free of any contamination

A.2.2 Cleaning the channel top parts

This step is only necessary if there is no sterile storage option for the channel top parts. Since the microscope **ref???** cannot hold **size** cover slides, the channel tops were cut into two identical pieces by a member of the workshop at the Department of Biology, Chemistry, Pharmacy at FU. Therefore the following procedure helps to clean the channels.

The sticky part of the channel tops should be kept clean and dry for the entire process. Otherwise, the final channels might become leaky because the ethanol destroyed the glue of the sticky surface.

1. place the channel tops up-side-down in a large cell culture dish
2. add 70% ethanol
3. use a cotton swab to clean the inside of the channels with 70% ethanol
4. remove the ethanol and dry the channel tops under the hood

A.2.3 Channel construction

For the channel construction clean glass slides (section A.2.1) and channel top parts (section A.2.2) are glued together to form the final channels. The glass slides and channel top parts should have the same size.

1. remove the cover of the sticky surface of the channels
2. attach a freshly plasma cleaned glass slide
3. gently press the two part together
4. keep the channel under pressure by adding weight (2 kg) on top of the glass slides for at least 2 h

A.3 Supported lipid bilayer formation

The supported lipid bilayer can be prepared on cleaned standard glass slides and well as in channels. First, the lipid composition is prepared and then added to the glass slides or into the channels.

A.3.1 Lipid preparation

The preparation of the supported lipid bilayer was established in the lab by Raluca Groza. The small unilamellar vesicles (SUVs) were prepared in a stock concentration ten times the concentration used in the experiments.

1. prepare bilayer buffer (10 m M HEPES, 200 m M NaCl, 2 m M $CaCl_2$, pH 6.8)
2. dilute the dry lipids in methanol:chloroform
3. calculate the amount of the lipids used depending on their ratio and concentration **???**
4. mix the lipids in a glass vial
5. evaporate under vacuum over night
6. resuspend the lipids in 1 ml bilayer buffer
7. vortex for 2 min at maximum speed
8. transfer the fluid into a 1.5 ml eppi
9. sonicate to form SUVs with the tip sonicator for 1 min at 70 % and one cycle on ice
10. centrifuge for 45 min at 14 000 G-force and 21 °C
11. fill supernatant into a new eppi
12. store SUVs at 4 °C

A.3.2 Supported lipid bilayer preparation

The following protocol describes how to prepare the supported lipid bilayer on plasma cleaned glass slides and in channels. The plasma cleaning step of the glass cleaning procedure should be (re-) done shortly before adding the SUVs for optimal results. In addition, the channels should be constructed directly before preparing the supported lipid bilayer within them. Since the supported lipid bilayer should not dry out, always remove only half of the solution, add fresh liquid, and pipette up and down to get a well-formed bilayer without any leftover vesicles. All steps are at room temperature

1. add the bilayer solution at a 10-fold dilution and incubate for 10 min
2. wash with bilayer buffer for 20 min
3. wash 5× with 1 % BSA in bilayer buffer
4. incubate with 1 % BSA in bilayer buffer for 30 min

5. wash with bilayer buffer for 5 min
6. CD95L samples: incubate with 1 µg of Neutravidin **???** for 20 min
7. wash with bilayer buffer for 5 min
8. CD95L samples: incubate with 0.2 µg ml^{-1} for 1 h
9. wash with bilayer buffer for 5 min
10. store samples at 4 °C

A.4 Tethering

The tethering of the cells via DNA consists of several steps. Before the tethering, The GPI-Halo transfection of the cells should be done 24 h according to the protocol in the section. This section outlines further preparation steps for the cells and the supported lipid bilayer.
Preparation of the supported lipid bilayer for tethering:

1. wash the supported lipid bilayer with bilayer buffer for 5 min
2. incubate with 300 nM of cholesterol-20mer-DNA for 40 min
3. wash with bilayer buffer for 10 min

Preparation of the cells for tethering 24 h after the transfection:

1. wash the transfected cells with PBS
2. suck off PBS and add 500 µl of Trypsin-EDTA
3. incubate the cells at 37 °C in the cell culture incubator
4. stop the reaction by adding cell medium to the cells
5. take up the entire fluid and move it into a 15 ml falcon
6. wash the cells on the plate again with cell medium and add it to the falcon
7. centrifuge cells for 5 min at 350 rcf and room temperature

8. suck off the fluid and add 400 µl of prewarmed live-cell medium
9. mix cells well with a pipette
10. transfer cells to a 1 ml eppi
11. add 1 µM of Halo-tag-100mer-DNA
12. incubate for 20-30 min in Thermomixer at lowest mix speed (300 rpm) and 37 °C
13. mix cells well with a pipette

In the final step, the cells are added to the bilayer:

1. wash bilayer 5 times with prewarmed live-cell medium
2. add the well-separated cells to the bilayer; best practise is to work on a microscope to adjust the desired amount of cells
3. incubate cells for the desired amount of time.
4. when working on a microscope: check if the cells are tethered by tipping a pipette-tip into the sample and watch the cells on the camera

A.5 Cell staining

The staining procedure for all experiments is similar because of the coupling of GFP to the proteins of interest. Therefore, the protocols mainly differ depending on the location of the protein of interest and its GFP.

A.5.1 Cell staining on glass

The procedure for staining cells on glass depends on whether they will be imaged live or fixed. In both cases, section A.1.2 provides the protocol for cell seeding.

Live-cell staining. It is possible to execute the staining of living cells directly on the microscope, thus performing all steps at room temperature. In addition, it is helpful to heat the metal holder for improved cell viability.

1. transfer glass slide with cell to metal holder

2. add 1 ml of prewarmed live-cell medium
3. focus on cells if staining is imaged on microscope
4. stain cells with 50 nM FluoTag-Q antiGFP nanobody labelled with Alexa Fluor 647 in PEM for 15 min
5. wash cells 5 times with live-cell medium
6. image cells in live-cell medium

The live-cell staining protocol could be used for tethered cells with external GFP as well.

Fixation and staining of cells. Fixation and staining procedures for cells seeded on glass vary depending on the protein of interest and its location in the cell. However, since all targets are coupled to GFP, the actual staining differs in the concentration of the nanobodies.

- CV-1 GPF-GFP cells:
    1. fixation with prewarmed 4 % PFA with 0.2 % GA in PBS for 20 min at 37 °C
    2. wash once with PBS
    3. quenching with freshly prepared 0.1 % $NaBH_4$ in PBS for 7 min at room temperature
    4. wash 5 times with PBS on a shaker
    5. blocking in three steps:
        (a) with ImageIT for 30 min
        (b) wash 3 times with PBS on shaker
        (c) with 4 % goat serum in 1 % in PBS for 1 h
    6. wash 3 times with PBS on shaker
    7. staining with 50 nM FluoTag-Q anti-GFP nanobodies labelled 1:1 with Alexa Fluor 647 for 1 h
    8. wash 3 times with PBS on shaker
    9. post-fixation with 4 % PFA and 0.2 % GA in PBS for 20 min at room temperature

10. post-quenching with 0.1 % $NaBH_4$ in PBS for 5 min at room temperature
11. wash 3 times with PBS on shaker
12. store at 4 °C

- HeLa CLC-GFP cells:

1. fixation with prewarmed 4 % PFA in PEM for 20 min at 37 °C
2. wash once with PEM
3. quenching with 50 mM NH_4Cl in PBS for 5 min at room temperature
4. wash 3 times with PEM on shaker
5. permeabilisation with 0.2 % saponin in PEM
6. wash 3 times with PEM on shaker
7. blocking with 4 % goat serum in 1 % of BSA in PEM for 1 h
8. wash 3 times with PEM on shaker
9. staining with 50 nM FluoTag-Q anti-GFP nanobodies labelled 1:1 with Alexa Fluor 647 for 30 min
10. post-fixation with 4 % PFA in PEM for 20 min at room temperature
11. wash once with PEM
12. post-quenching with 50 mM NH_4Cl in PBS for 5 min
13. wash 3 times with PEM on shaker
14. store at 4 °C

- T98G CD95-GFP:

1. fixation with prewarmed 4 % PFA and 0.2 % GA in PEM for 20 min at 37 °C
2. wash once with PEM
3. quenching with freshly prepared 0.1 % $NaBH_4$ in PEM for 7 min at room temperature
4. wash 3 times with PEM on shaker
5. permeabilise with 0.2 % saponin in PEM for 5 min
6. wash 3 times with PEM on shaker

7. blocking with 4 % goat serum in 1 % BSA in PEM for 1 h
8. wash 3 times with PEM on shaker
9. staining with 50 nM FluoTag-Q anti-GFP nanobodies labelled 1:1 with Alexa Fluor 647 for 30 min
10. wash 3 times with PEM on shaker
11. post-fixation with 4 % PFA and 0.2 % GA in PEm for 20 min at room temperature
12. wash once with PEM
13. post-quenching with 0.1 % $NaBH_4$ in PEM for 7 min
14. wash 3 times with PEM on shaker
15. store at 4 °C

A.5.2 Staining of tethered cells

The protocol for staining tethered cells is very similar to the protocols used for staining cells on glass. The main differences are that the samples consist of a bilayer that should be kept stable and that the probe remains at room temperature. This protocol describes the fixation and staining procedure of CD95-GFP after tethering the cells to the supported lipid bilayer. It is essential to keep fluid left on top of the bilayer at all times because it should not dry out.

1. wash cells 5 times with live-cell medium.
2. fixation with prewarmed 4 % PFA and 0.2 % GA in PEM for 20 min
3. wash 5 times with PEM
4. quench with freshly prepared 0.1 % $NaBH_4$ in PEM for 5 min
5. wash 5 times with PEM
6. permeabilize with 0.2 % saponin in PEM for 5 min
7. wash 5 times with PEM
8. staining with 50 nM FluoTag-Q anti-GFP nanobodies labelled 1:1 with Alexa Fluor 647 in PEM for 30 min

9. wash 5 times with PEM
10. post-fixation with 4 % PFA and 0.2 % GA in PEM for 20 min
11. post-quench with 0.1 % $NaBH_4$ in PEM for 5 min
12. wash 5 times with PEM
13. store at 4 °C

Appendix B

Sequences

Further details are given in this chapter about the DNA strand sequences and the image sequences used for the live-cell tethering experiments.

B.1 DNA sequences

For the tethering, oligomers consisting of DNA strands and a modification were used. The first one was integrated into the supported lipid bilayer and the second one was covalently bond via the HaloTag to the Halo-GPI of the transfected T98G cells.

Oligomer name	sequence (5′-..... - 3′)	5′-modification
Halo-tag-100-mer-DNA	TTT TTT TTT TTT TTT TTT TT TTT TTT TTT TTT TTT TTT TT TTT TTT TTT TTT TTT TTT TT TTT TTT TTT TTT TTT TTT TT ACT GAC TGA CTG ACT GAC TG	HaloTag Ligand (O2)
cholesterol-20-mer-DNA	CAG TCA GTC AGT CAG TCA GT	cholesterol
control-20-mer-AlexaFluor647	ACT GAC TGA CTG ACT GAC TG	Alexa Fluor 647

Table B.1: Sequences of the DNA-oligomers used to tether cells to the supported lipid bilayer.

B.2 Live-cell tethering sequences

The sequences in Fig. 4.17 only show selected images at every 2 seconds or every second. The following are more detailed sequences with shorter intervals between consecutive images.

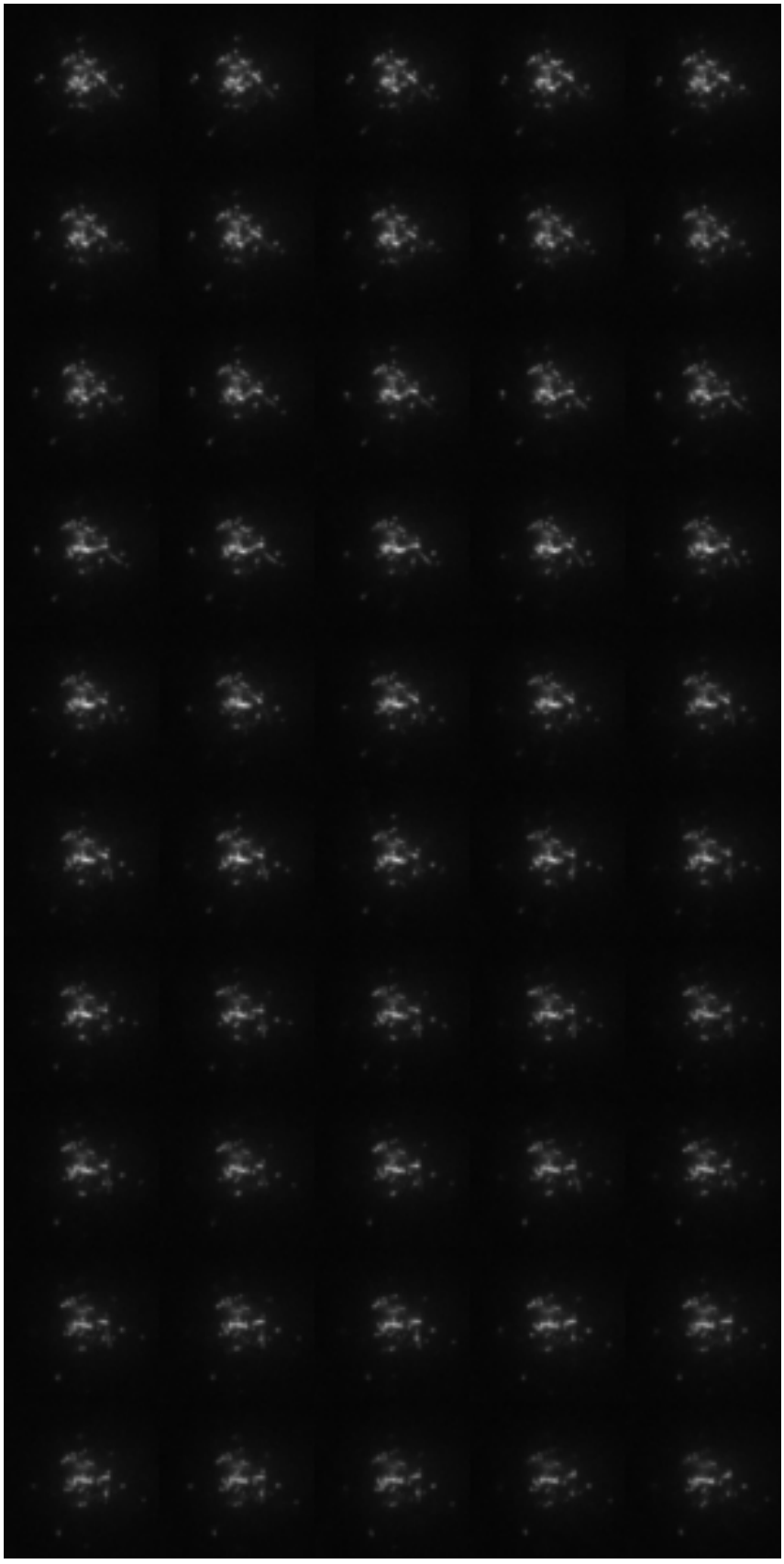

Figure B.1: Time laps of the cell in Fig. 4.17 (A). Here, the images have a interval of 200 ms. The sequence starts with the frame in the left top corner and ends with the one at the right bottom.

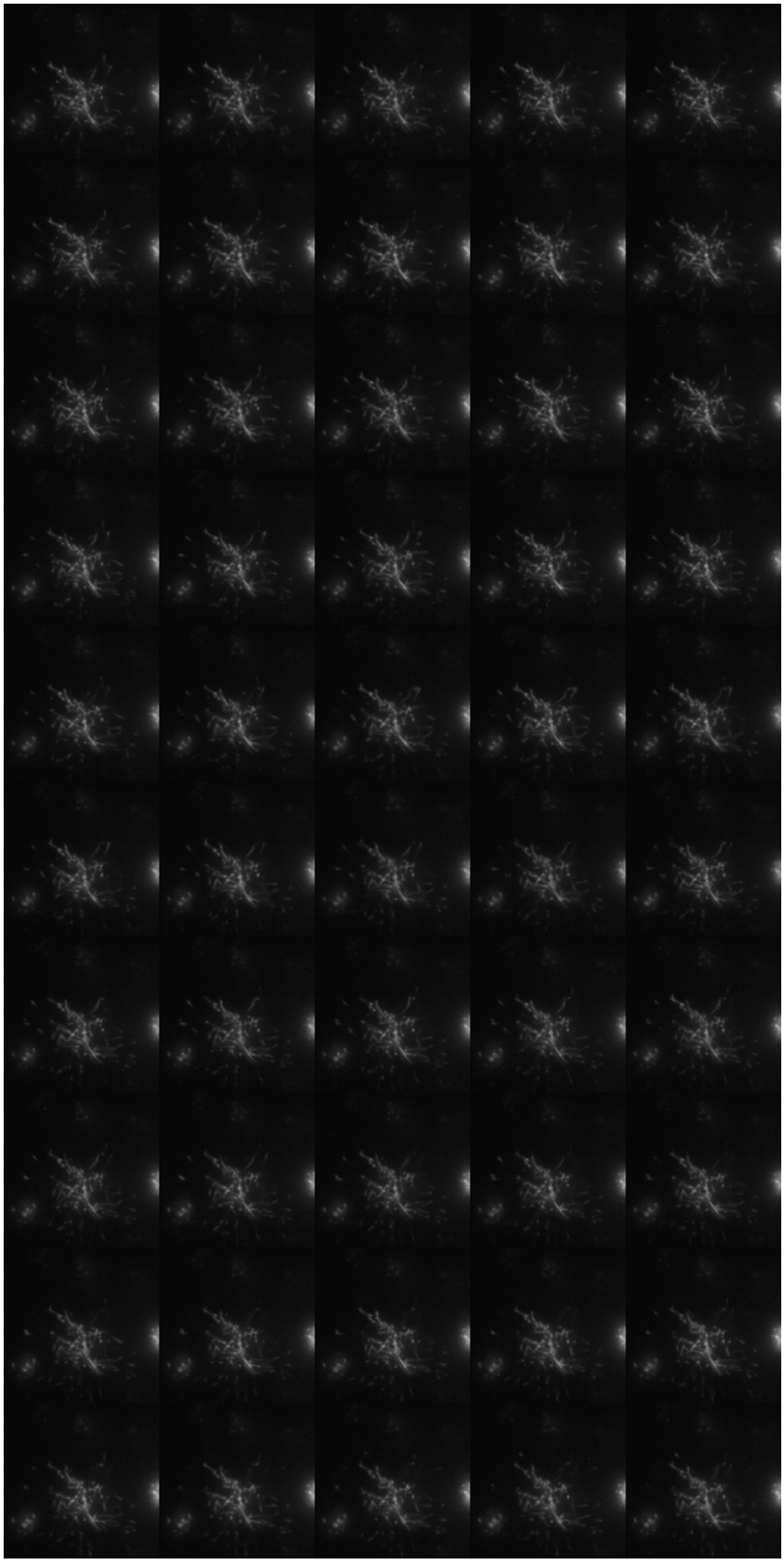

Figure B.2: Time laps of the cell in Fig. 4.17 (B). Here, the images have a interval of 100 ms. The sequence starts with the frame in the left top corner and ends with the one at the right bottom.

Acknowledgements

I express my gratitude to my supervisor, Prof. Dr Helge Ewers, for suggesting this project and allowing me to work on this project. His initial idea and the opportunity he provided to delve into the captivating realm of clustering within the broad spectrum of STEM, ranging from cell cultures to software development, have significantly influenced my personal and career path.

Ando, as my ally throughout this journey, I extend my heartfelt thanks for your unwavering commitment during the long and challenging days of cell tethering. Our discussions about clustering strategies and data preparation efforts have been enlightening. Your positivity and shared celebrations were the constant motivation.

Claire, thank you for sharing your profound expertise in cell culture and pivotal role in resolving our transfection challenges, significantly contributing to our success. Our insightful discussion and your valuable contributions were invaluable.

To my fellow lab members, your support and shared experiences were sources of strength. Your presence made the journey through experiments, highs, and lows more tolerable and memorable.

I want to express my deep appreciation to my family and friends for their strong belief in my scientific passion. Your continuous support and encouragement provided comfort during challenging times but also magnified my successes.

I want to thank the Berlin and Zurich swim teams for their efforts. During challenging periods when lab work seemed overwhelming, practices provided a welcome break. My

frustration was released, and I regained motivation for the upcoming experiments. Your camaraderie during training sessions and the enjoyable moments shared at competitions have been a source of positivity and inspiration.

I sincerely thank Claire, Amin, and Andreas for reading and providing valuable feedback on my work. Your insights have been invaluable in refining this thesis.

I am grateful to CyberMentor and my former mentor Christian for their pivotal roles in nurturing my passion for STEM and shaping my future career trajectory. Your unwavering support has enhanced my confidence and dedication to STEM.

A huge thank you goes out to everyone who has been a part of this journey over the past few years, whether you cheered for my project, dedicated time to swimming or offered support. Your time, energy, and unwavering motivation were the building blocks of this incredible adventure!

Bibliography

[1] LORD RAYLEIGH SEC. R. S. : XV. On the theory of optical images, with special reference to the microscope. In: *Philosophical Magazine Series 5* 42 (1896), Nr. 255, S. 167–195

[2] ABBE, Ernst: Beiträge zur Theorie des Mikroskops und der mikroskopischen Wahrnehmung. In: *Archiv für mikroskopische Anatomie* 9 (1873), Nr. 1, S. 413–418

[3] AGGARWAL, C.C. ; REDDY, C.K.: *Data Clustering: Algorithms and Applications*. CRC Press, 2018 (Chapman & Hall/CRC Data Mining and Knowledge Discovery Series). – URL `https://books.google.co.jp/books?id=cH50DwAAQBAJ`. – ISBN 9781315360416

[4] ALBRECHT, David ; WINTERFLOOD, Christian M. ; SADEGHI, Mohsen ; TSCHAGER, Thomas ; NOÉ, Frank ; EWERS, Helge: Nanoscopic compartmentalization of membrane protein motion at the axon initial segment. In: *Journal of Cell Biology* 215 (2016), 10, S. 37–46. – URL `www.jcb.org/cgi/doi/10.1083/jcb.201603108`. – ISSN 15408140

[5] ANACONDA: *ANACONDA Installers and Packages*. `https://repo.anaconda.com`. – Accessed: 2021-11-16

[6] ANNIBALE, Paolo ; VANNI, Stefano ; SCARSELLI, Marco ; ROTHLISBERGER, Ursula ; RADENOVIC, Aleksandra: Identification of clustering artifacts in photoactivated localization microscopy. In: *Nature Publishing Group* (2011). – URL `http://psicquic.goo-`

[7] ANNIBALE, Paolo ; VANNI, Stefano ; SCARSELLI, Marco ; ROTHLISBERGER, Ursula ; RADENOVIC, Aleksandra: Quantitative Photo Activated Localization Microscopy:

Unraveling the Effects of Photoblinking. In: *PLoS ONE* 6 (2011), 7, S. e22678. – URL https://dx.plos.org/10.1371/journal.pone.0022678. – ISSN 1932-6203

[8] ARNOLD, Andreas M. ; SCHNEIDER, Magdalena C. ; HÜSSON, Christoph ; SABLATNIG, Robert ; BRAMESHUBER, Mario ; BAUMGART, Florian ; SCHÜTZ, Gerhard J.: Verifying molecular clusters by 2-color localization microscopy and significance testing. In: *Scientific Reports* 10 (2020), Nr. 1, S. 1–12

[9] ASHKIN, Arthur: Optical trapping and manipulation of neutral particles using lasers. In: *Proceedings of the National Academy of Sciences* 94 (1997), Nr. 10, S. 4853–4860

[10] ATKINS, P.W. ; FRIEDMAN, R.S.: *Molecular Quantum Mechanics*. OUP Oxford, 2011

[11] BALTA, Gülce S. G. ; MONZEL, Cornelia ; KLEBER, Susanne ; BEAUDOUIN, Joel ; BALTA, Emre ; KAINDL, Thomas ; CHEN, Si ; GAO, Liang ; THIEMANN, Meinolf ; WIRTZ, Christian R. ; SAMSTAG, Yvonne ; TANAKA, Motomu ; MARTIN-VILLALBA, Ana: 3D Cellular Architecture Modulates Tyrosine Kinase Activity, Thereby Switching CD95-Mediated Apoptosis to Survival. In: *Cell Reports* 29 (2019), 11, S. 2295–2306.e6. – ISSN 2211-1247

[12] BAUMGART, Florian ; ARNOLD, Andreas M. ; LESKOVAR, Konrad ; STASZEK, Kaj ; FÖLSER, Martin ; WEGHUBER, Julian ; STOCKINGER, Hannes ; SCHÜTZ, Gerhard J.: Varying label density allows artifact-free analysis of membrane-protein nanoclusters. In: *Nature methods* 13 (2016), Nr. 8, S. 661–664

[13] BEGHEIN, Els ; GETTEMANS, Jan: Nanobody Technology: A Versatile Toolkit for Microscopic Imaging, Protein–Protein Interaction Analysis, and Protein Function Exploration. In: *Frontiers in Immunology* 8 (2017), S. 771. – URL https://www.frontiersin.org/article/10.3389/fimmu.2017.00771. – ISSN 1664-3224

[14] BENNETT, Loris ; MELCHERS, Bernd ; PROPPE, Boris: *Curta: A General-purpose High-Performance Computer at ZEDAT, Freie Universität Berlin*. http://dx.doi.org/10.17169/refubium-26754. 2020

[15] BETZIG, Eric ; PATTERSON, George H. ; SOUGRAT, Rachid ; LINDWASSER, O W. ; OLENYCH, Scott ; BONIFACINO, Juan S. ; DAVIDSON, Michael W. ; LIPPINCOTT-SCHWARTZ, Jennifer ; HESS, Harald F.: Imaging intracellular fluorescent proteins at nanometer resolution. In: *science* 313 (2006), Nr. 5793, S. 1642–1645

[16] BIOCONDUCTOR: *Bioconductor - Open Source Software for Bioinformatics*. https://www.bioconductor.org. – Accessed: 2021-11-16

[17] BIRK, Udo J.: *Hardware and Software*. Kap. 3, S. 83–116. In: *SuperResolution Microscopy*, John Wiley & Sons, Ltd, 2017. – URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527802074.ch3. – ISBN 9783527802074

[18] BIRK, Udo J.: *Localization Microscopy*. Kap. 5, S. 179–262. In: *SuperResolution Microscopy*, John Wiley & Sons, Ltd, 2017. – URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527802074.ch5. – ISBN 9783527802074

[19] BLOCK, Stephan ; ZHDANOV, Vladimir P. ; HOOK, Fredrik: Quantification of multivalent interactions by tracking single biological nanoparticle mobility on a lipid membrane. In: *Nano letters* 16 (2016), Nr. 7, S. 4382–4390

[20] BODMER, Jean-Luc ; SCHNEIDER, Pascal ; TSCHOPP, Jürg: The molecular architecture of the TNF superfamily. In: *Trends in Biochemical Sciences* 27 (2002), Nr. 1, S. 19–26. – URL https://www.sciencedirect.com/science/article/pii/S0968000401019958. – ISSN 0968-0004

[21] BRUCE, P. ; BRUCE, A.: *Practical Statistics for Data Scientists: 50 Essential Concepts*. O'Reilly Media, Inc., 2017. – URL https://www.oreilly.com/library/view/practical-statistics-for/9781491952955/. – ISBN 9781491952931

[22] BURGERT, Anne ; LETSCHERT, Sebastian ; DOOSE, Sören ; SAUER, Markus: Artifacts in single-molecule localization microscopy. In: *Histochemistry and cell biology* 144 (2015), S. 123–131

[23] CASTELLANA, Edward T. ; CREMER, Paul S.: Solid supported lipid bilayers: From biophysical studies to sensor design. In: *Surface Science Reports* 61 (2006), Nr. 10, S. 429–444

[24] CHAZAL, F. ; SILVA, V. de ; GLISSE, M. ; OUDOT, S.: *The Structure and Stability of Persistence Modules*. Springer International Publishing, 2016 (SpringerBriefs in Mathematics). – URL https://books.google.de/books?id=0QM6DQAAQBAJ. – ISBN 9783319425450

[25] CHAZAL, Frédéric ; GUIBAS, Leonidas J. ; OUDOT, Steve Y. ; SKRABA, Primoz: Persistence-Based Clustering in Riemannian Manifolds. In: *J. ACM* 60 (2013), nov, Nr. 6. – URL https://doi.org/10.1145/2535927. – ISSN 0004-5411

[26] CHAZAL, Frédéric ; MICHEL, Bertrand: An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists. In: *Frontiers in Artificial Intelligence* 4 (2021), S. 108. – URL https://www.frontiersin.org/article/10.3389/frai.2021.667963. – ISSN 2624-8212

[27] CLARK, Philip J. ; EVANS, Francis C.: Distance to nearest neighbor as a measure of spatial relationships in populations. In: *Ecology* 35 (1954), Nr. 4, S. 445–453

[28] COLLETTE, A.: *Python and HDF5*. O'Reilly Media, Inc., 2013. – URL https://www.oreilly.com/library/view/python-and-hdf5/9781491944981/. – ISBN 9781449367831

[29] COLLINS, Tony J.: ImageJ for microscopy. In: *Biotechniques* 43 (2007), Nr. S1, S. S25–S30

[30] CORPORATION, Microsoft ; WESTON, Steve: *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. : , 2020. – URL https://CRAN.R-project.org/package=doParallel. – R package version 1.0.16

[31] CSARDI, Gabor ; NEPUSZ, Tamas: The igraph software package for complex network research. In: *InterJournal* Complex Systems (2006), S. 1695. – URL https://igraph.org

[32] CULLEY, Siân ; ALBRECHT, David ; JACOBS, Caron ; PEREIRA, Pedro M. ; LETERRIER, Christophe ; MERCER, Jason ; HENRIQUES, Ricardo: NanoJ-SQUIRREL: quantitative mapping and minimisation of super-resolution optical imaging artefacts. In: *BioRxiv* (2017), S. 158279

[33] CURTIS, ASG: The mechanism of adhesion of cells to glass: a study by interference reflection microscopy. In: *The Journal of cell biology* 20 (1964), Nr. 2, S. 199–215

[34] DIEKMANN, Robin ; WOLFSON, Deanna L. ; SPAHN, Christoph ; HEILEMANN, Mike ; SCHÜTTPELZ, Mark ; HUSER, Thomas: Nanoscopy of bacterial cells immobilized by holographic optical tweezers. In: *Nature communications* 7 (2016), Nr. 1, S. 13711

[35] DIMRI, Sushil C. ; MALIK, Preeti ; RAM, Mangey: *Algorithms: Design and Analysis*. De Gruyter, 2021. – URL https://doi.org/10.1515/9783110693607. – ISBN 9783110693607

[36] DOBRUCKI, Jurek W. ; KUBITSCHECK, Ulrich: *Fluorescence Microscopy*. Kap. 3, S. 85–132. In: *Fluorescence Microscopy*, John Wiley & Sons, Ltd, 2017. – URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527687732.ch3. – ISBN 9783527687732

[37] DOWLE, Matt ; SRINIVASAN, Arun: *data.table: Extension of 'data.frame'*. : , 2021. – URL https://CRAN.R-project.org/package=data.table. – R package version 1.14.0

[38] ESTER, Martin ; KRIEGEL, Hans-Peter ; SANDER, Jörg ; XU, Xiaowei u. a.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *kdd* Bd. 96, 1996, S. 226–231

[39] FERGUSON, Thomas S.: A Bayesian Analysis of Some Nonparametric Problems. In: *The Annals of Statistics* 1 (1973), Nr. 2, S. 209 – 230. – URL https://doi.org/10.1214/aos/1176342360

[40] FISCHER, Bernd ; SMITH, Mike ; PAU, Gregoire: *rhdf5: R Interface to HDF5*. : , 2021. – URL https://github.com/grimbough/rhdf5. – R package version 2.36.0

[41] FU, Qingshan ; FU, Tian M. ; CRUZ, Anthony C. ; SENGUPTA, Prabuddha ; THOMAS, Stacy K. ; WANG, Shuqing ; SIEGEL, Richard M. ; WU, Hao ; CHOU, James J.: Structural Basis and Functional Role of Intramembrane Trimerization of the Fas/CD95 Death Receptor. In: *Molecular Cell* 61 (2016), 2, S. 602–613. – ISSN 10974164

[42] GAUTIER, Arnaud ; JOHNSSON, Kai ; O'HARE, Helen: *AGT/SNAP-Tag: A Versatile Tag for Covalent Protein Labeling*. Kap. 5, S. 89–107. In: *Probes and Tags to Study Biomolecular Function*, John Wiley & Sons, Ltd, 2008. – URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527623099.ch5. – ISBN 9783527623099

[43] GAUTIER, Laurent: *rpy2 - R in Python*. https://rpy2.github.io/index.html

[44] GEORGE, James N. ; WEED, Robert I. ; REED, Claude F.: Adhesion of human erythrocytes to glass: the nature of the interaction and the effect of serum and plasma. In: *Journal of cellular physiology* 77 (1971), Nr. 1, S. 51–59

[45] GETIS, Arthur ; FRANKLIN, Janet: Second-Order Neighborhood Analysis of Mapped Point Patterns. In: *Ecology* 68 (1987), Nr. 3, S. 473–477. – URL http://www.jstor.org/stable/1938452. – ISSN 00129658, 19399170

[46] GIEBEL, K-F ; BECHINGER, Clemens ; HERMINGHAUS, Stephan ; RIEDEL, M ; LEIDERER, Paul ; WEILAND, Ulrich ; BASTMEYER, Martin: Imaging of cell/substrate contacts of living cells with surface plasmon resonance microscopy. In: *Biophysical journal* 76 (1999), Nr. 1, S. 509–516

[47] GOWRISHANKAR, Kripa: Active Re-Modelling of Cortical Actin Regulates Spatiotemporal Organization of Molecules on a Living Cell Surface. In: *Biophysical Journal* 98 (2010), Nr. 3, S. 304a–305a

[48] GRÄF, Ralph ; RIETDORF, Jens ; ZIMMERMANN, Timo: Live cell spinning disk microscopy. In: *Microscopy Techniques: -/-* (2005), S. 57–75

[49] GRAKOUI, Arash ; BROMLEY, Shannon K. ; SUMEN, Cenk ; DAVIS, Mark M. ; SHAW, Andrey S. ; ALLEN, Paul M. ; DUSTIN, Michael L.: The immunological synapse: a molecular machine controlling T cell activation. In: *Science* 285 (1999), Nr. 5425, S. 221–227

[50] GRASSMÉ, Heike ; CREMESTI, Aida ; KOLESNICK, Richard ; GULBINS, Erich: Ceramide-mediated clustering is required for CD95-DISC formation. In: *Oncogene* 22 (2003), Nr. 35, S. 5457–5470

[51] GREEN, Peter J. ; RICHARDSON, Sylvia: Modelling heterogeneity with and without the Dirichlet process. In: *Scandinavian journal of statistics* 28 (2001), Nr. 2, S. 355–375

[52] GRIFFIÉ, Juliette ; BOELEN, Lies ; BURN, Garth ; COPE, Andrew P. ; OWEN, Dylan M.: Topographic prominence as a method for cluster identification in single-molecule localisation data. In: *Journal of biophotonics* 8 (2015), Nr. 11-12, S. 925–934

[53] GRIFFIÉ, Juliette ; BURN, Garth L. ; WILLIAMSON, David J. ; PETERS, Ruby ; RUBIN-DELANCHY, Patrick ; OWEN, Dylan M.: Dynamic Bayesian cluster analysis of live-cell single molecule localization microscopy datasets. In: *Small Methods* 2 (2018), Nr. 9, S. 1800008

[54] GRIFFIÉ, Juliette ; SHANNON, Michael ; BROMLEY, Claire L. ; BOELEN, Lies ; BURN, Garth L. ; WILLIAMSON, David J. ; HEARD, Nicholas A. ; COPE, Andrew P. ; OWEN, Dylan M. ; RUBIN-DELANCHY, Patrick: A Bayesian cluster analysis method for single-molecule localization microscopy data. In: *Nature Protocols* 11 (2016), Nr. 12, S. 2499–2514

[55] GRIFFIÉ, Juliette ; SHLOMOVICH, Leigh ; WILLIAMSON, David J. ; SHANNON, Michael ; AARON, Jesse ; KHUON, Satya ; L. BURN, Garth ; BOELEN, Lies ; PETERS, Ruby ; COPE, Andrew P. u. a.: 3D Bayesian cluster analysis of super-resolution data reveals LAT recruitment to the T cell synapse. In: *Scientific reports* 7 (2017), Nr. 1, S. 4077

[56] GRIMM, Jonathan B. ; BROWN, Timothy A. ; ENGLISH, Brian P. ; LIONNET, Timothée ; LAVIS, Luke D.: Synthesis of Janelia Fluor HaloTag and SNAP-Tag ligands and their use in cellular imaging experiments. In: *Super-Resolution Microscopy: Methods and Protocols* (2017), S. 179–188

[57] GUSTAFSSON, Mats G.: Surpassing the lateral resolution limit by a factor of two using structured illumination microscopy. In: *Journal of microscopy* 198 (2000), Nr. 2, S. 82–87

[58] HABEL, Kai ; GRASMAN, Raoul ; GRAMACY, Robert B. ; MOZHAROVSKYI, Pavlo ; STERRATT, David C.: *geometry: Mesh Generation and Surface Tessellation*. : , 2019.

– URL `https://CRAN.R-project.org/package=geometry`. – R package version 0.4.5

[59] HAKEN, H. ; WOLF, H.C.: *Molekülphysik und Quantenchemie: Einführung in die experimentellen und theoretischen Grundlagen*. Springer Berlin Heidelberg, 2013 (Springer Lehrbuch)

[60] HARRIS, Charles R. ; MILLMAN, K. J. ; WALT, Stéfan J. van der ; GOMMERS, Ralf ; VIRTANEN, Pauli ; COURNAPEAU, David ; WIESER, Eric ; TAYLOR, Julian ; BERG, Sebastian ; SMITH, Nathaniel J. ; KERN, Robert ; PICUS, Matti ; HOYER, Stephan ; KERKWIJK, Marten H. van ; BRETT, Matthew ; HALDANE, Allan ; RÍO, Jaime Fernández del ; WIEBE, Mark ; PETERSON, Pearu ; GÉRARD-MARCHANT, Pierre ; SHEPPARD, Kevin ; REDDY, Tyler ; WECKESSER, Warren ; ABBASI, Hameer ; GOHLKE, Christoph ; OLIPHANT, Travis E.: Array programming with NumPy. In: *Nature* 585 (2020), S. 357–362

[61] HEILEMANN, Mike ; LINDE, Sebastian van de ; MUKHERJEE, Anindita ; SAUER, Markus: Super-resolution imaging with small organic fluorophores. In: *Angewandte Chemie International Edition* 48 (2009), Nr. 37, S. 6903–6908

[62] HELL, Stefan W.: Far-field optical nanoscopy. In: *science* 316 (2007), Nr. 5828, S. 1153–1158

[63] HELL, Stefan W. ; WICHMANN, Jan: Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. In: *Optics letters* 19 (1994), Nr. 11, S. 780–782

[64] HENNIG, C. ; MEILA, M. ; MURTAGH, F. ; ROCCI, R.: *Handbook of Cluster Analysis*. CRC Press, 2015 (Chapman & Hall/CRC Handbooks of Modern Statistical Methods). – URL `https://books.google.de/books?id=tDc0CwAAQBAJ`. – ISBN 9781466551893

[65] HESS, Samuel T. ; GIRIRAJAN, Thanu P. ; MASON, Michael D.: Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. In: *Biophysical journal* 91 (2006), Nr. 11, S. 4258–4272

[66] HUNTER, John D.: Matplotlib: A 2D graphics environment. In: *Computing in science & engineering* 9 (2007), Nr. 3, S. 90–95

[67] INSTITUTE FOR STATISTICS AND MATHEMATICS OF WU (WIRTSCHAFTSUNIVERSITÄT WIEN): *The Comprehensive R Archive Network*. `https://cran.r-project.org/index.html`. – Accessed: 2021-11-16

[68] IZRAILEV, Sergei: *tictoc: Functions for Timing R Scripts, as Well as Implementations of Stack and List Structures*. : , 2021. – URL `https://CRAN.R-project.org/package=tictoc`. – R package version 1.0.1

[69] JAYASINGHE, Isuru ; CLOWSLEY, Alexander H. ; LIN, Ruisheng ; LUTZ, Tobias ; HARRISON, Carl ; GREEN, Ellen ; BADDELEY, David ; DI MICHELE, Lorenzo ; SOELLER, Christian: True molecular scale visualization of variable clustering properties of ryanodine receptors. In: *Cell Reports* 22 (2018), Nr. 2, S. 557–567

[70] JEREMY PIKE, Saskia K.: *RSMLM - R package for analysis of single molecule localization microscopy (SMLM) data*. `https://github.com/saskiakutz/RSMLM`. – Accessed: 2021-11-16, original work by Jeremy Pike, corrections by Luca Baronti and Saskia Kutz

[71] JETBRAINS S.R.O: *PyCharm: The Python IDE for Professional Developers*. – URL `https://www.jetbrains.com/pycharm/`. – Accessed: 2021-11-18

[72] KALAPPURAKKAL, Joseph M. ; SIL, Parijat ; MAYOR, Satyajit: Toward a new picture of the living plasma membrane. In: *Protein Science* 29 (2020), Nr. 6, S. 1355–1365

[73] KASSAMBARA, Alboukadel: *ggpubr: 'ggplot2' Based Publication Ready Plots*. : , 2020. – URL `https://CRAN.R-project.org/package=ggpubr`. – R package version 0.4.0

[74] KAUFER, Daniela ; FRIEDMAN, Alon ; SEIDMAN, Shlomo ; SOREQ, Hermona: Acute stress facilitates long-lasting changes in cholinergic gene expression. In: *Nature* 393 (1998), Nr. 6683, S. 373–377

[75] KEPPLER, Antje ; GENDREIZIG, Susanne ; GRONEMEYER, Thomas ; PICK, Horst ; VOGEL, Horst ; JOHNSSON, Kai: A general method for the covalent labeling of

fusion proteins with small molecules in vivo. In: *Nature biotechnology* 21 (2003), Nr. 1, S. 86–89

[76] KHATER, Ismail M. ; NABI, Ivan R. ; HAMARNEH, Ghassan: A review of super-resolution single-molecule localization microscopy cluster analysis and quantification methods. In: *Patterns* 1 (2020), Nr. 3, S. 100038

[77] KILLIAN, Jessica L. ; YE, Fan ; WANG, Michelle D.: Optical tweezers: A force to be reckoned with. In: *Cell* 175 (2018), Nr. 6, S. 1445–1448

[78] KISKOWSKI, Maria A. ; HANCOCK, John F. ; KENWORTHY, Anne K.: On the Use of Ripley's K-Function and Its Derivatives to Analyze Domain Size. In: *Biophysical Journal* 97 (2009), Nr. 4, S. 1095–1103. – URL https://www.sciencedirect.com/science/article/pii/S0006349509010480. – ISSN 0006-3495

[79] KRUSCHKE, John K.: Chapter 2 - Introduction: Credibility, Models, and Parameters. In: KRUSCHKE, John K. (Hrsg.): *Doing Bayesian Data Analysis (Second Edition)*. Second Edition. Boston : Academic Press, 2015, S. 15–32. – URL https://www.sciencedirect.com/science/article/pii/B9780124058880000027. – ISBN 978-0-12-405888-0

[80] KUBITSCHECK, Ulrich: *Principles of Light Microscopy*. Kap. 2, S. 23–83. In: *Fluorescence Microscopy*, John Wiley & Sons, Ltd, 2017. – URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527687732.ch2. – ISBN 9783527687732

[81] KUSUMI, Akihiro ; IKE, Hiroshi ; NAKADA, Chieko ; MURASE, Kotono ; FUJIWARA, Takahiro: Single-molecule tracking of membrane molecules: plasma membrane compartmentalization and dynamic assembly of raft-philic signaling molecules. In: *Seminars in immunology* Bd. 17 Elsevier (Veranst.), 2005, S. 3–21

[82] KUTZ, Saskia ; ZEHRER, Ando C. ; SVETLITCKII, Roman ; GÜLCÜLER BALTA, Gülce S. ; GALLI, Lucrezia ; KLEBER, Susanne ; RENTSCH, Jakob ; MARTIN-VILLALBA, Ana ; EWERS, Helge: An Efficient GUI-Based Clustering Software for Simulation and Bayesian Cluster Analysis of Single-Molecule Localization Microscopy Data. In: *Frontiers in Bioinformatics* 1 (2021), S. 47. – URL https://www.frontiersin.org/article/10.3389/fbinf.2021.723915. – ISSN 2673-7647

[83] LAGARDÈRE, Matthieu ; CHAMMA, Ingrid ; BOUILHOL, Emmanuel ; NIKOLSKI, Macha ; THOUMINE, Olivier: FluoSim: simulator of single molecule dynamics for fluorescence live-cell and super-resolution imaging of membrane proteins. In: *Scientific Reports* 10 (2020), 12, S. 1–14. – URL https://doi.org/10.1038/s41598-020-75814-y. – ISSN 20452322

[84] LANDRY, Jonathan J. ; PYL, Paul T. ; RAUSCH, Tobias ; ZICHNER, Thomas ; TEKKEDIL, Manu M. ; STÜTZ, Adrian M. ; JAUCH, Anna ; AIYAR, Raeka S. ; PAU, Gregoire ; DELHOMME, Nicolas u. a.: The genomic and transcriptomic landscape of a HeLa cell line. In: *G3: Genes, Genomes, Genetics* 3 (2013), Nr. 8, S. 1213–1224

[85] LEE, Sang H. ; BADAY, Murat ; TJIOE, Marco ; SIMONSON, Paul D. ; ZHANG, Ruobing ; CAI, En ; SELVIN, Paul R.: Using fixed fiduciary markers for stage drift correction. In: *Optics express* 20 (2012), Nr. 11, S. 12177–12183

[86] LELEK, Mickaël ; GYPARAKI, Melina T. ; BELIU, Gerti ; SCHUEDER, Florian ; GRIFFIÉ, Juliette ; MANLEY, Suliana ; JUNGMANN, Ralf ; SAUER, Markus ; LAKADAMYALI, Melike ; ZIMMER, Christophe: Single-molecule localization microscopy. In: *Nature Reviews Methods Primers* 1 (2021), Nr. 1, S. 39

[87] LEVET, Florian ; HOSY, Eric ; KECHKAR, Adel ; BUTLER, Corey ; BEGHIN, Anne ; CHOQUET, Daniel ; SIBARITA, Jean-Baptiste: SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. In: *Nature methods* 12 (2015), Nr. 11, S. 1065–1071

[88] LEVOIN, Nicolas ; JEAN, Mickael ; LEGEMBRE, Patrick: CD95 Structure, Aggregation and Cell Signaling. In: *Frontiers in Cell and Developmental Biology* 8 (2020), S. 314. – URL https://www.frontiersin.org/article/10.3389/fcell.2020.00314. – ISSN 2296-634X

[89] LI, Jia H. ; SANTOS-OTTE, Paula ; AU, Braedyn ; RENTSCH, Jakob ; BLOCK, Stephan ; EWERS, Helge: Directed manipulation of membrane proteins by fluorescent magnetic nanoparticles. In: *Nature communications* 11 (2020), Nr. 1, S. 4259

[90] LI, Yiming ; MUND, Markus ; HOESS, Philipp ; DESCHAMPS, Joran ; MATTI, Ulf ; NIJMEIJER, Bianca ; SABININA, Vilma J. ; ELLENBERG, Jan ; SCHOEN, Ingmar ; RIES,

Jonas: Real-time 3D single-molecule localization using experimental point spread functions. In: *Nature methods* 15 (2018), Nr. 5, S. 367–369

[91] LINDE, Sebastian van de ; HEILEMANN, Mike ; SAUER, Markus: Live-Cell Super-Resolution Imaging with Synthetic Fluorophores. In: *Annual Review of Physical Chemistry* 63 (2012), Nr. 1, S. 519–540

[92] LINDE, Sebastian van de ; LÖSCHBERGER, Anna ; KLEIN, Teresa ; HEIDBREDER, Meike ; WOLTER, Steve ; HEILEMANN, Mike ; SAUER, Markus: Direct stochastic optical reconstruction microscopy with standard fluorescent probes. In: *Nature protocols* 6 (2011), Nr. 7, S. 991–1009

[93] LISS, Viktoria ; BARLAG, Britta ; NIETSCHKE, Monika ; HENSEL, Michael: Self-labelling enzymes as universal tags for fluorescence microscopy, super-resolution microscopy and electron microscopy. In: *Scientific reports* 5 (2015), Nr. 1, S. 1–13

[94] LONG, JD ; TEETOR, P. ; SAFARI, an O'Reilly Media C.: *R Cookbook, 2nd Edition*. O'Reilly Media, Inc., 2019. – URL `https://www.oreilly.com/library/view/r-cookbook-2nd/9781492040675/`

[95] LOS, Georgyi V. ; ENCELL, Lance P. ; MCDOUGALL, Mark G. ; HARTZELL, Danette D. ; KARASSINA, Natasha ; ZIMPRICH, Chad ; WOOD, Monika G. ; LEARISH, Randy ; OHANA, Rachel F. ; URH, Marjeta ; SIMPSON, Dan ; MENDEZ, Jacqui ; ZIMMERMAN, Kris ; OTTO, Paul ; VIDUGIRIS, Gediminas ; ZHU, Ji ; DARZINS, Aldis ; KLAUBERT, Dieter H. ; BULLEIT, Robert F. ; WOOD, Keith V.: HaloTag: A Novel Protein Labeling Technology for Cell Imaging and Protein Analysis. In: *ACS Chemical Biology* 3 (2008), Nr. 6, S. 373–382. – URL `https://doi.org/10.1021/cb800025k`. – PMID: 18533659

[96] MACQUEEN, James u. a.: Some methods for classification and analysis of multivariate observations. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* Bd. 1 Oakland, CA, USA (Veranst.), 1967, S. 281–297

[97] MARIN, Jean-Michel ; ROBERT, Christian P.: *Bayesian Essentials with R*. Springer New York, 2014. – URL `https://link.springer.com/book/10.1007/978-1-4614-8687-9`

[98] MARTIN-VILLALBA, Ana ; LLORENS-BOBADILLA, Enric ; WOLLNY, Damian: CD95 in cancer: Tool or target? In: *Trends in Molecular Medicine* 19 (2013), 6, S. 329–335. – ISSN 14714914

[99] MATTILA, Pieta K. ; LAPPALAINEN, Pekka: Filopodia: molecular architecture and cellular functions. In: *Nature reviews Molecular cell biology* 9 (2008), Nr. 6, S. 446–454

[100] MCGORTY, Ryan ; KAMIYAMA, Daichi ; HUANG, Bo: Active microscope stabilization in three dimensions using image correlation. In: *Optical nanoscopy* 2 (2013), S. 1–7

[101] MCKINNEY, Wes: *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython, 2nd Edition*. O'Reilly Media, Inc., 2018. – URL `https://www.oreilly.com/library/view/python-for-data/9781491957653/`

[102] MCKINNEY, Wes u. a.: Data structures for statistical computing in python. In: *Proceedings of the 9th Python in Science Conference* Bd. 445 Austin, TX (Veranst.), 2010, S. 51–56

[103] MÖCKL, Leonhard: The emerging role of the mammalian glycocalyx in functional membrane organization and immune system regulation. In: *Frontiers in cell and developmental biology* 8 (2020), S. 253

[104] MOLLOY, Justin E. ; PADGETT, Miles J.: Lights, action: optical tweezers. In: *Contemporary physics* 43 (2002), Nr. 4, S. 241–258

[105] MOORE, A.D.: *Mastering GUI Programming with Python: Develop impressive cross-platform GUI applications with PyQt*. Packt Publishing, 2019. – URL `https://www.packtpub.com/product/mastering-gui-programming-with-python/9781789612905`. – ISBN 9781789615821

[106] MORTENSEN, Kim I. ; CHURCHMAN, L S. ; SPUDICH, James A. ; FLYVBJERG, Henrik: Optimized localization analysis for single-molecule tracking and super-resolution microscopy. In: *nature methods* 7 (2010), Nr. 5, S. 377–381

[107] NAGARAJAN, Radhakrishnan ; SCUTARI, Marco: *Bayesian Networks in R with Applications in Systems Biology*. New York : Springer, 2013. – URL `https:`

//link.springer.com/book/10.1007/978-1-4614-6446-4. – ISBN 978-1-4614-6445-7, 978-1-4614-6446-4

[108] NICOVICH, Philip R. ; OWEN, Dylan M. ; GAUS, Katharina: Turning single-molecule localization microscopy into a quantitative bioanalytical tool. In: *Nature protocols* 12 (2017), Nr. 3, S. 453–460

[109] NIENHAUS, Gerd U. ; NIENHAUS, Karin: *Fluorescence Labeling*. Kap. 4, S. 133–164. In: *Fluorescence Microscopy*, John Wiley & Sons, Ltd, 2017. – URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527687732.ch4. – ISBN 9783527687732

[110] NIEVES, Daniel J. ; PIKE, Jeremy A. ; LEVET, Florian ; WILLIAMSON, David J. ; BARAGILLY, Mohammed ; OLOKETUYI, Sandra ; DE MARCO, Ario ; GRIFFIÉ, Juliette ; SAGE, Daniel ; COHEN, Edward A. u. a.: A framework for evaluating the performance of SMLM cluster analysis algorithms. In: *Nature methods* 20 (2023), Nr. 2, S. 259–267

[111] NOBELPRIZE.ORG: *The Nobel Prize in Chemistry 2014*. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2014/. – Online; accessed 19-April-2016

[112] NWANESHIUDU, Adaobi ; KUSCHAL, Christiane ; SAKAMOTO, Fernanda H. ; ANDERSON, R R. ; SCHWARZENBERGER, Kathryn ; YOUNG, Roger C.: Introduction to confocal microscopy. In: *Journal of Investigative Dermatology* 132 (2012), Nr. 12, S. 1–5

[113] OUDOT, S.Y.: *Persistence Theory: From Quiver Representations to Data Analysis*. American Mathematical Society, 2015 (Mathematical Surveys and Monographs). – URL https://books.google.de/books?id=if8dCwAAQBAJ. – ISBN 9781470425456

[114] OWEN, Dylan M. ; RENTERO, Carles ; ROSSY, Jérémie ; MAGENAU, Astrid ; WILLIAMSON, David ; RODRIGUEZ, Macarena ; GAUS, Katharina: PALM imaging and cluster analysis of protein heterogeneity at the cell surface. In: *Journal of biophotonics* 3 (2010), Nr. 7, S. 446–454

[115] PALUCH, Ewa ; HEISENBERG, Carl-Philipp: Biology and physics of cell shape changes in development. In: *Current Biology* 19 (2009), Nr. 17, S. R790–R799

[116] PEDERSEN, Thomas L.: *ggforce: Accelerating 'ggplot2'*. : , 2021. – URL https://CRAN.R-project.org/package=ggforce. – R package version 0.3.3

[117] PIKE, Jeremy A. ; KHAN, Abdullah O. ; PALLINI, Chiara ; THOMAS, Steven G. ; MUND, Markus ; RIES, Jonas ; POULTER, Natalie S. ; STYLES, Iain B. ; MURPHY, Robert: Topological data analysis quantifies biological nano-structure from single molecule localization microscopy. In: *Bioinformatics* 36 (2020), 3, S. 1614–1621. – URL https://github.com/JeremyPike/RSMLM-tutorials.. – ISSN 14602059

[118] PLATTNER, H.: *Zellbiologie*. Thieme, 2011 (Reihe, NULTSCH BOTANIK). – URL https://books.google.de/books?id=TOgVZEMmSNwC. – ISBN 9783131652447

[119] PYTHON SOFTWARE FOUNDATION: *PyPI - The Python Package Index*. https://pypi.org. – Accessed: 2021-11-16

[120] PYTHON SOFTWARE FOUNDATION: *Python Software Foundation*. https://www.python.org/psf/. – Accessed: 2021-11-16

[121] R CORE TEAM: *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing (Veranst.), 2021. – URL https://www.R-project.org/. – Accessed: 2021-11-16

[122] RABADAN, Raul ; BLUMBERG, Andrew J.: *Topological Data Analysis for Genomics and Evolution: Topology in Biology*. Cambridge University Press, 2019

[123] RAJARAMAN, R ; ROUNDS, DE ; YEN, SPS ; REMBAUM, ANDA: A scanning electron microscope study of cell adhesion and spreading in vitro. In: *Experimental cell research* 88 (1974), Nr. 2, S. 327–339

[124] RIES, Jonas: SMAP: a modular super-resolution microscopy analysis platform for SMLM data. In: *Nature Methods* (2020), 8, S. 1–2. – URL http://www.nature.com/articles/s41592-020-0938-1. – ISSN 1548-7091

[125] RIES, Jonas ; KAPLAN, Charlotte ; PLATONOVA, Evgenia ; EGHLIDI, Hadi ; EWERS, Helge: A simple, versatile method for GFP-based super-resolution microscopy via nanobodies. In: *Nature methods* 9 (2012), Nr. 6, S. 582–584

[126] RINKER, Tyler W. ; KURKIEWICZ, Dason: *pacman: Package Management for R*. Buffalo, New York: , 2018. – URL `http://github.com/trinker/pacman`. – version 0.5.0

[127] RIPLEY, Brian D.: Modelling spatial patterns. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 39 (1977), Nr. 2, S. 172–192

[128]

[129] RSTUDIO: *RStudio: Open source and professional software for data science teams*. – Accessed: 2021-11-18

[130] RUBIN-DELANCHY, Patrick ; BURN, Garth L. ; GRIFFIÉ, Juliette ; WILLIAMSON, David J. ; HEARD, Nicholas A. ; COPE, Andrew P. ; OWEN, Dylan M.: Bayesian cluster identification in single-molecule localization microscopy data. In: *Nature methods* 12 (2015), Nr. 11, S. 1072–1076

[131] RUST, Michael J. ; BATES, Mark ; ZHUANG, Xiaowei: Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). In: *Nature methods* 3 (2006), Nr. 10, S. 793–796

[132] SAGE, Daniel ; PHAM, Thanh A. ; BABCOCK, Hazen ; LUKES, Tomas ; PENGO, Thomas ; CHAO, Jerry ; VELMURUGAN, Ramraj ; HERBERT, Alex ; AGRAWAL, Anurag ; COLABRESE, Silvia ; WHEELER, Ann ; ARCHETTI, Anna ; RIEGER, Bernd ; OBER, Raimund ; HAGEN, Guy M. ; SIBARITA, Jean B. ; RIES, Jonas ; HENRIQUES, Ricardo ; UNSER, Michael ; HOLDEN, Seamus: Super-resolution fight club: assessment of 2D and 3D single-molecule localization microscopy software. In: *Nature Methods* 16 (2019), 5, S. 387–395. – URL `https://doi.org/10.1038/s41592-019-0364-4`. – ISSN 15487105

[133] SAKA, Sinem K. ; HONIGMANN, Alf ; EGGELING, Christian ; HELL, Stefan W. ; LANG, Thorsten ; RIZZOLI, Silvio O.: Multi-protein assemblies underlie the mesoscale

organization of the plasma membrane. In: *Nature communications* 5 (2014), Nr. 1, S. 4509

[134] SANCHO-MARTINEZ, Ignacio ; MARTIN-VILLALBA, Ana: Tyrosine phosphorylation and CD95: a FAScinating switch. In: *Cell cycle* 8 (2009), Nr. 6, S. 838–842

[135] SANDERSON, Jeremy: *Total Internal Reflection Fluorescence Microscopy*. Kap. 23, S. 561–568. In: *Understanding Light Microscopy*, John Wiley & Sons, Ltd, 2019. – URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118696736.ch23. – ISBN 9781118696736

[136] SAUER, M. ; HOFKENS, J. ; ENDERLEIN, J.: *Handbook of Fluorescence Spectroscopy and Imaging: From Ensemble to Single Molecules*. Wiley, 2010

[137] SAUER, M. ; HOFKENS, J. ; ENDERLEIN, J.: *Basic Principles of Fluorescence Spectroscopy*. Kap. 1, S. 1–30. In: *Handbook of Fluorescence Spectroscopy and Imaging*, John Wiley & Sons, Ltd, 2011. – URL https://onlinelibrary.wiley.com/doi/abs/10.1002/9783527633500.ch1. – ISBN 9783527633500

[138] SCHINDELIN, Johannes ; ARGANDA-CARRERAS, Ignacio ; FRISE, Erwin ; KAYNIG, Verena ; LONGAIR, Mark ; PIETZSCH, Tobias ; PREIBISCH, Stephan ; RUEDEN, Curtis ; SAALFELD, Stephan ; SCHMID, Benjamin u. a.: Fiji: an open-source platform for biological-image analysis. In: *Nature methods* 9 (2012), Nr. 7, S. 676–682

[139] SCHMIDT, Marco F.: *Chemische Biologie – Adressierung neuer Wirkstoffziele*. S. 211–231. In: *Chemische Biologie: und Wirkstoffentwicklung*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2020. – URL https://doi.org/10.1007/978-3-662-61116-6_16. – ISBN 978-3-662-61116-6

[140] SCHNELL, Ulrike ; DIJK, Freark ; SJOLLEMA, Klaas A. ; GIEPMANS, Ben N.: Immunolabeling artifacts and the need for live-cell imaging. In: *Nature methods* 9 (2012), Nr. 2, S. 152–158

[141] SCHNITZBAUER, Joerg ; STRAUSS, Maximilian T. ; SCHLICHTHAERLE, Thomas ; SCHUEDER, Florian ; JUNGMANN, Ralf: Super-resolution microscopy with DNA-PAINT. In: *Nature protocols* 12 (2017), Nr. 6, S. 1198–1228

[142] SELDEN, Nicholas S. ; TODHUNTER, Michael E. ; JEE, Noel Y. ; LIU, Jennifer S. ; BROADERS, Kyle E. ; GARTNER, Zev J.: Chemically programmed cell adhesion with membrane-anchored oligonucleotides. In: *Journal of the American Chemical Society* 134 (2012), Nr. 2, S. 765–768

[143] SHABO, Ivan ; SVANVIK, Joar ; LINDSTRÖM, Annelie ; LECHERTIER, Tanguy ; TRABULO, Sara ; HULIT, James ; SPAREY, Tim ; PAWELEK, John: Roles of cell fusion, hybridization and polyploid cell formation in cancer metastasis. In: *World journal of clinical oncology* 11 (2020), Nr. 3, S. 121

[144] SOCHACKI, Kem A. ; DICKEY, Andrea M. ; STRUB, Marie-Paule ; TARASKA, Justin W.: Endocytic proteins are partitioned at the edge of the clathrin lattice in mammalian cells. In: *NATURE CELL BIOLOGY* 19 (2017)

[145] STEIN, Gretchen H.: T98G: An anchorage-independent human tumor cell line that exhibits stationary phase G1 arrest in vitro. In: *Journal of cellular physiology* 99 (1979), Nr. 1, S. 43–54

[146] STENNING, David C. ; DYK, David A. van: *BAYESIAN STATISTICAL METHODS FOR ASTRONOMY PART I: FOUNDATIONS*. S. 11–28. In: *Statistics for Astrophysics: Bayesian Methodology*, EDP Sciences, 2021. – URL `https://doi.org/10.1051/978-2-7598-2275-1.c005`

[147] STONE, Matthew B. ; SHELBY, Sarah A. ; NUNEZ, Marcos F. ; WISSER, Kathleen ; VEATCH, Sarah L.: Protein sorting by lipid phase-like domains supports emergent signaling function in B lymphocyte plasma membranes. In: *elife* 6 (2017), S. e19891

[148] TANAKA, Kenji A. ; SUZUKI, Kenichi G. ; SHIRAI, Yuki M. ; SHIBUTANI, Shusaku T. ; MIYAHARA, Manami S. ; TSUBOI, Hisae ; YAHARA, Miyako ; YOSHIMURA, Akihiko ; MAYOR, Satyajit ; FUJIWARA, Takahiro K. u. a.: Membrane molecules mobile even after chemical fixation. In: *Nature methods* 7 (2010), Nr. 11, S. 865–866

[149] TAUBENBERGER, Anna V. ; BAUM, Buzz ; MATTHEWS, Helen K.: The mechanics of mitotic cell rounding. In: *Frontiers in cell and developmental biology* (2020), S. 687

[150] TAYLOR, Marcus J. ; HUSAIN, Kabir ; GARTNER, Zev J. ; MAYOR, Satyajit ; VALE, Ronald D.: A DNA-based T cell receptor reveals a role for receptor clustering in ligand discrimination. In: *Cell* 169 (2017), Nr. 1, S. 108–119

[151] VEREB, G ; SZÖLLŐSI, J ; MATKO, J ; NAGY, P ; FARKAS, T ; VIGH, LMLWTA ; MATYUS, L ; WALDMANN, TA ; DAMJANOVICH, S: Dynamic, yet structured: the cell membrane three decades after the Singer–Nicolson model. In: *Proceedings of the national academy of sciences* 100 (2003), Nr. 14, S. 8053–8058

[152] VICIDOMINI, Giuseppe ; BIANCHINI, Paolo ; DIASPRO, Alberto: STED super-resolved microscopy. In: *Nature methods* 15 (2018), Nr. 3, S. 173–182

[153] WALKER, Stephen G.: *Bayesian nonparametric methods: motivation and ideas*. S. 22–34. In: HJORT, Nils L. (Hrsg.) ; HOLMES, Chris (Hrsg.) ; MÜLLER, Peter (Hrsg.) ; WALKER, Stephen G. (Hrsg.): *Bayesian Nonparametrics*, Cambridge University Press, 2010 (Cambridge Series in Statistical and Probabilistic Mathematics)

[154] WANG, Peng ; LI, Jie ; WEI, Qiang: Understanding the Interplay between Cell Force and Cell Adhesion Processes. In: *Engineered Regeneration* (2023)

[155] WASKOM, Michael ; BOTVINNIK, Olga ; O'KANE, Drew ; HOBSON, Paul ; LUKAUSKAS, Saulius ; GEMPERLINE, David C. ; AUGSPURGER, Tom ; HALCHENKO, Yaroslav ; COLE, John B. ; WARMENHOVEN, Jordi ; RUITER, Julian de ; PYE, Cameron ; HOYER, Stephan ; VANDERPLAS, Jake ; VILLALBA, Santi ; KUNTER, Gero ; QUINTERO, Eric ; BACHANT, Pete ; MARTIN, Marcel ; MEYER, Kyle ; MILES, Alistair ; RAM, Yoav ; YARKONI, Tal ; WILLIAMS, Mike L. ; EVANS, Constantine ; FITZGERALD, Clark ; BRIAN ; FONNESBECK, Chris ; LEE, Antony ; QALIEH, Adel: *mwaskom/seaborn: v0.8.1 (September 2017)*. September 2017. – URL `https://doi.org/10.5281/zenodo.883859`

[156] WICKHAM, H. ; GROLEMUND, G.: *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. O'Reilly Media, Inc., 2016. – URL `https://www.oreilly.com/library/view/r-for-data/9781491910382/`. – ISBN 9781491910368

[157] WICKHAM, Hadley: The Split-Apply-Combine Strategy for Data Analysis. In:

Journal of Statistical Software 40 (2011), Nr. 1, S. 1–29. – URL http://www.jstatsoft.org/v40/i01/

[158] WICKHAM, Hadley ; AVERICK, Mara ; BRYAN, Jennifer ; CHANG, Winston ; MCGOWAN, Lucy D. ; FRANÇOIS, Romain ; GROLEMUND, Garrett ; HAYES, Alex ; HENRY, Lionel ; HESTER, Jim ; KUHN, Max ; PEDERSEN, Thomas L. ; MILLER, Evan ; BACHE, Stephan M. ; MÜLLER, Kirill ; OOMS, Jeroen ; ROBINSON, David ; SEIDEL, Dana P. ; SPINU, Vitalie ; TAKAHASHI, Kohske ; VAUGHAN, Davis ; WILKE, Claus ; WOO, Kara ; YUTANI, Hiroaki: Welcome to the tidyverse. In: *Journal of Open Source Software* 4 (2019), Nr. 43, S. 1686

[159] WICKHAM, Hadley ; HESTER, Jim ; CHANG, Winston: *devtools: Tools to Make Developing R Packages Easier*. : , 2021. – URL https://CRAN.R-project.org/package=devtools. – R package version 2.4.2

[160] WILLIAMSON, David J. ; BURN, Garth L. ; SIMONCELLI, Sabrina ; GRIFFIÉ, Juliette ; PETERS, Ruby ; DAVIS, Daniel M. ; OWEN, Dylan M.: Machine learning for cluster analysis of localization microscopy data. In: *Nature communications* 11 (2020), Nr. 1, S. 1493

[161] WILLIAMSON, David P. ; SHMOYS, David B.: *An Introduction to Approximation Algorithms*. S. 3–26. In: *The Design of Approximation Algorithms*, Cambridge University Press, 2011

[162] WOLLMAN, Adam J. ; NUDD, Richard ; HEDLUND, Erik G. ; LEAKE, Mark C.: From Animaculum to single molecules: 300 years of the light microscope. In: *Open biology* 5 (2015), Nr. 4, S. 150019

[163] WRIGHT, Shirley J. ; WRIGHT, David J.: Introduction to confocal microscopy. In: *Cell Biological Applications of Confocal Microscopy, in Methods in Cell Biology* 70 (2002), S. 1–85

[164] YADAV, Amarjeet ; DUTTA, Anindita ; KUMAR, Pramod ; DAHAN, Yuval ; ARANOVICH, Alexander ; FEINGOLD, Mario: Optimal trapping stability of Escherichia coli in oscillating optical tweezers. In: *Physical Review E* 101 (2020), Nr. 6, S. 062402

[165] YEUNG, A ; EVANS, E: Cortical shell-liquid core model for passive flow of liquid-like spherical cells into micropipets. In: *Biophysical journal* 56 (1989), Nr. 1, S. 139–149

[166] YOUNG, Lesley ; SUNG, Julia ; STACEY, Glyn ; MASTERS, John R.: Detection of Mycoplasma in cell cultures. In: *Nature protocols* 5 (2010), Nr. 5, S. 929–934

[167] ZHANG, Hu ; LIU, Kuo-Kang: Optical tweezers for single cells. In: *Journal of the Royal Society interface* 5 (2008), Nr. 24, S. 671–690